Norwegian School of Economics
Bergen, Spring 2021

Private and Public Sanctions

Investigating cartel decisions using Twitter data

Knut Henrik Tjøstheim and Arne Bjørnevik Nygaard

Supervisor: Evelina Gavrilova-Zoutman

Master thesis, Economics and Business Administration

Major: Business Analytics

NORWEGIAN SCHOOL OF ECONOMICS

Acknowledgements

This thesis is written as the final part of the MSc degree in Economics and Business Administration at the Norwegian School of Economics. The process of writing this thesis have been both educational and challenging. First we want to thank all the friends we made during our years at NHH. We also want to give a special thanks to our supervisor, Associate Professor Evelina Gavrilova Zoutman, for her valuable insights and constructive criticism throughout the process of writing this thesis. We also want to thank Haraldsplass Diakonale Sykehus, who aided critical help when one of us suffered from severe thrombocytopenia in the last phases of this thesis. Lastly, we want to thank Twitter for granting us access to their API through the Academic Research program.

Norwegian School of Economics

Bergen, May 2021

Knut Henrik Tjøstheim	Arne Nygaard
Your name here	Your name here

Abstract

This thesis investigates whether the announcement of cartel decisions by the European Commission provides new information for investors and if Twitter data can be used to explain abnormal returns. The dataset consists of 39 cartel cases involving 124 different companies from January 2010 to May 2021. Using a standard event study methodology, we find evidence that supports previous studies findings and confirm that variables such as fines, geographic location, and the size of a company impact abnormal returns in relation to the European commission's cartel decision. These variables are confirmed important by the use of single-factor regression and decision trees. The Twitter variables were not found to have any explanatory power on abnormal returns. A statistical significant cumulative abnormal return in the event window [-15,15] of -2.29% was found in the sample containing all fined companies. We also find that companies that receive immunity from the European Commission have no significant cumulative abnormal returns on average.

Keywords – European Cartels, Event Study, Sentiment Analysis, Twitter

Contents

List of Figures

List of Tables

1 Introduction

Ideally, the penalty for committing corporate fraud would equal the social cost of the crime. The fines imposed by the court does, however, often only represent a small percentage of this cost. If the size of the punishment is too small, the chance that corporations repeat their actions increases. However, fines are not the only means of punishment. Private sanctions from the public entail a potentially significant cost for the corporations. Legal and economic literature covering the topic of corporate fraud agree that private sanctions often can deter corporate misbehaviour as much as public sanctions like fines.

While fines by nature are quantified, private sanctions from the loss of reputation and standings in society are not and can only be estimated. Following Mariuzzo et al. (2020a), who look at the relationship between public and private sanctions on EU cartel cases with the help of newspaper sentiment, we try in this thesis to replicate parts of their results and to evaluate the intangible costs of private sanctions by the use of a dictionary-based Twitter sentiment analysis, event study methodology and newer data.

The thesis bases itself on the assumptions of the efficient market hypothesis, which states that people are rational investors and that stock prices should reflect all available information (Fama, 1970). With this theory of economics in mind, we aim to find out if the relative difference in public sentiment and coverage on Twitter between companies can explain differences seen in stock performance around the cartel decision date. Our expectations in advance were that the markets would react negatively to the decision and that the extent of negative Twitter coverage would influence the stock returns.

Our main finding is that Twitter data is not suitable for estimating private sanctions. As the relative differences in Twitter sentiment between companies do not explain the differences in abnormal returns on the day. Fines and other firm characteristics have some explanatory power, and we find abnormal returns comparable to those of previous studies.

Our thesis contributes to the overall study of cartel convictions and deterrence theory. We are to our knowledge the first that have tried to use sentiment analysis on Twitter data for the purpose of measuring private sanctions.

The thesis is structured as follows: We start off in chapter 2 by describing the institutional setting and the legal framework that prohibits cartel competition in Europe. Then we proceed by providing a review of relevant literature and a short description of Twitter and social listening. The data we used is presented in chapter 3, with research questions and hypothesis following in chapter 4. In chapter 5 we describe the main methodologies that are being used. In chapter 6 the empirical findings and analysis are presented. The results of the analysis are discussed in chapter 7, before we summarise and conclude the thesis in chapter 8.

2 Background

2.1 European Cartel - Institutional Setting

EU anti-trust policies had its early beginnings in 1957 when West Germany, Belgium, France, Luxembourg and the Netherlands signed the treaty of Rome, forming the European Economic Community EEC (1957). The goal of the treaty was to create a single economic area with free competition between member states. The treaty also established the Court of Justice of the European Union and the European Commission. In the beginning the European commission mainly consulted the national competition authorities in each member state and it first got its mandate to impose sanctions on infringements with the introduction of the Council Regulation 17 in 1962. The treaty of Rome evolved into the European Union which got established in 1993, and was created as a three-pillar structure with the EEC as remaining part. The EEC was abolished at the treaty of Lisbon in 2009 which formed EU in its current state. The underlying treaty is now called the Treaty on the Functioning of the European Union (TFEU) which together with the treaty on European Union (TEU) creates the constitutional basis of the EU (Publications Office of the European Union, 2015). It also covers the its' competition laws.

Article 101 and 102 in the TFEU regulates illegal antitrust behaviour in the European Union. Article 101 states that anti-competitive agreements are forbidden, examples of behaviour it prohibit is price fixing and market sharing agreements (European Union, 2008). Article 102 prohibits the abuse of a dominant market position. It is the European Commission that lead the investigation of cartels in EU. According to their website, a cartel is "a group of similar, independent companies which join together to fix prices, to limit production or to share markets or customers between them". This leads to less incentives for the companies to provide better or cheaper products, ending in higher prices or worse quality products for the customers.

Cartels are hard to spot because of their illegal nature. There are several ways for an investigation to start in the EU:

1. Investigations can start by a leniency application from one of the cartel members. The leniency notice from 1996 secures that there is an incentive for the cartel

members to be first at reaching out to the European Commission. Companies that reach out to the Commission with important information about a cartel which they have participated in may receive full or some reduction from fines (European Commission, 1996). Normally, the first participant that apply for leniency will receive full reduction from their fines, while other participants can receive some reduction if they add significant value helping the case. Given the potential benefits, the leniency notice is a significant tool to provide the Commission with insider information.

2. In 2017, the Commission started a new tool to make it possible for individuals to provide information about past, ongoing or planned anti-competitive breaches according to Article 101, it is called the Whistleblower Tool (European Commission, 2017a). The anonymity of the whistleblower will be guaranteed with a special-designed encrypted message system that allows communication between the whistleblower and the Commission. It works well along with the leniency notice as it retrieves information from individuals, whereas the leniency notice focus on retrieving information from companies.

3. A complaint from citizens and firms about suspected infringements of Article 101. A formal complaint can be filled on the Commission's website and can lead to further investigation from the Commission (European Commission, 2017b)

4. Sector investigations and inquiries from the Commission when it believes that a market is not working the way it should be and believes that breaches according to the competition rules might be one of the main factors.

The Commission normally starts of by conducting an initial investigation phase. This can include surprise inspections on the premises of the suspected companies or the request of information (European Commission, 2017b). When the initial phase ends, they decide whether they want to pursue an in-depth investigation or not. If they decide to continue the investigation, the news will be published on their home site. This statement is anonymized and generally only include information about which sector that is under investigation. The commission continues by trying to settle the case. From 2008 it became possible for companies to receive a 10% settlement fine reduction if they completely acknowledge their involvement in the cartel (Laina & Laurinen, 2013). If the commission is not able to settle the case, then the investigation continues until a conclusion is made.

When the commission reaches a conclusion, a press release with key information about the case is published. This includes the fine for each company involved, and their respective reductions. This is the phase that our study focuses on, since this is often the first time that the public receives information about which companies that are involved.

Fines have two main objectives, to deter and to punish. The Commission considers the sales value of the involved companies and the duration of the infringement when setting the fine. The fine can be adjusted depending on the circumstances of the case, repeat offending is an example of something that can lead to an increased fine (European Commission, 2011). Depending on the cooperation, further reductions through the leniency program can be obtained.

2.2 Literature review

The relationship between public sanctions and private sanctions has been a topic in economic and legal literature for a long time. Much of the early work has been done by economists working within the field of deterrence theory. Believers of the deterrence theory argue that people and corporations choose to obey or violate laws after calculating all the possible gains and consequences of their actions. The general consensus from the studies that address the topic of corporate crime is that private sanctions from the loss of standings and stigma in society can deter corporate misbehaviour as good or better than formal legal sanctions. The literature often distinguishes between offences that are considered to affect "related-parties" and "third-parties". "Related-parties" are cases where customers are directly affected by the fraudulent behaviour of a company, while "third-party" offences happen when the public is indirectly affected by a corporation's misbehaviour. Cartel cases are mostly considered to be related party offences.

Some noticeable literature in the field of "related-party" offences have been done by Jarrell & Peltzman (1985), Karpoff & Lott (1993) and Alexander (1999). Their studies looked at how the markets respond to corporations that are being sentenced for fraudulent and cheating behaviour. In these studies, private sanctions ending in losses due to a worsened reputation have been estimated indirectly, as reputation is considered to be an intangible asset. Their methods involved decomposing stock prices into the effects of public sanctions, a readjustment effect, and a residual, which is explained as reputational loss.

A big part of literature is focusing on if private sanctions can work as a deterrent to corporate crime. Bosch & Eckard (1991) studied collusion in United States. By the use of event study methodology, they calculated the abnormal returns of 127 firms that were indicted in the period 1962-1980. They found the cumulative average abnormal return for the firms to be 1.08% post indictment. The authors hypothesize that the reaction may be explained by legal costs, loss of reputation and forgone monopoly profits.

Aguzzoni et al. (2013) did a similar study to the one done by Bosch & Eckard (1991), but on cartel investigations in the EU. Their study looked at 180 companies that were sentenced for cartel participation between 1979 and 2009. They found a statistically significant cumulative average abnormal rate of return (CAAR) of negative 3.57% in relation to the infringement decision by the European Commission. Their study also looked at the stock price drop surrounding the initial investigations on corporate premises and found it to be statistically significant. The total combined effects of the infringement decision and the surprise investigation weighted by market capitalization were between -3.03% and -4.55%. They estimated that only up to 8.9% of the total loss could be explained by the fine amount and conjectured that most of the loss was due to the ending of illegal activities.

A third study was done by Günster & van Dijk (2016). This study looked at a sample set consisting of 253 firms fined by the European commission between 1974-2004. Their result shows a CAAR of -1.85% around the final verdict, which was statistically significant. They concluded that fines and legal costs could explain around 25% of lost market capitalization. The remaining portion was explained by reputational impairment and anticipated profitability decreases. Factors they found to determine the severity of the stock price reduction were the magnitude of the fine, the duration of the infringement, and most importantly, the media attention covering the investigation events.

Ulrich (2018) investigated the effect of cartel fines in the European Union on share prices, dividend payments and management compensations between 2001-2018. He found a significant cumulative abnormal return of -2.89% in his primary sample over the event window [-25,10]. His study finds that the extent of the stock price reduction depends on the fine, country of incorporation and firm size.

Most similar to our work is the study conducted by Mariuzzo et al. (2020a). In their study, they looked at cartels convicted by the European Commission between 1992-2015. They studied the relationship between public and private sanctions and used sentiment analysis on newspaper articles to approximate the magnitude of the reputational effect. They found evidence supporting the findings of Aguzzoni et al. (2013) and confirmed that cartel members are more hurt at detection than at decision date. With the use of event study methodology, regression trees and sensitivity analysis, they found that fines are a key variable that can explain some of the loss in firm valuation on a short window around the decision, while reputational sanctions are more important in explaining value losses on a larger time frame. The results of their analysis also support the idea that the sentiment of the media coverage can work as a substitute to fines. They also conclude that private sanctions are less effective if there are no public sanctions.

We seek to contribute to these previous works by testing whether Twitter data can explain some of the variations in cumulative abnormal returns between companies while testing if fine still is a significant factor when looking at more recent cases.

2.3 Twitter and Social listening

Twitter has grown to become an important platform for information and opinion sharing since its beginning in 2006. Every Twitter user has the opportunity to share their thoughts and opinions about all kinds of topics through a tweet containing up to 280 characters. The tweets can be distributed and read by people from all over the world. Today Twitter has more than 199 million active users, which combined tweet more than 500 million tweets every day (Twitter, 2021b) (Internetlivestats, 2021). It is thus an enormous database covering all kinds of topics.

The use of Twitter data in economic research has increased in the last decade. As an opinion source, Twitter is benefiting from the fact that the aggregation of tweets from many users cancels out individual misconceptions and thus presents a possible more reliable perception of the event than traditional news media. Because of this, it is often used to understand stakeholders' view on corporations. Multiple studies have attempted to predict stock prices by looking at Twitter volume and sentiment, and it has shown itself capable of predicting index growth with a high degree of certainty Bollen et al. (2011).

Twitter data is also being used to study public opinion about news events, and it has been used to look at the relationships between Twitter mentions and election results (Tumasjan et al., n.d.). In addition to being a valuable source of information for academic researchers, modern-day corporations are spending money and resources on social listening to monitor both their own brand and products as well as the competitors' and the general market. Since twitter data provides a real-time evaluation of a company's sentiment, it gives us a unique opportunity to compare a normal sentiment to the sentiment around the event. This is different to the more static opinion found in more traditional news media.

Even though Twitter data can be useful in many ways, it also has its limitations, which can make it somewhat unsuitable as an opinion source.

1. The 280-character limitation sets a limit to the amount of information each tweet can contain.
2. It is difficult to collect the tweets other than by hashtags and user references, and the number of tweets is so large that it becomes difficult to collect the most relevant ones.
3. Individual tweets can be wrong, misleading, and hard to interpret for natural language algorithms.
4. Sampling bias – Users tend to be in the age between 20 and 40 years old, and some parts of society are more represented. Twitter is therefore not representative of the general population.

3 Data

We have collected and used four different datasets in this thesis. The data contains information about all publicly listed companies sentenced for cartel participation by the European commission from 2010-2021. We have chosen to only use data from this period due to Twitter's short lifespan and somewhat limited use before 2010. In this chapter we will describe how we collected the data, choices we have made and show summary statistics for each dataset.

3.1 Cartel data

The cartel dataset contains specific information about all the relevant cartels. We manually gathered the data from the EU commissions webpage by querying for cases with decision date after 01.01.2010, this resulted in a list of 46 cartel cases (European Commission, 2021). By reading prohibition decisions, press releases and summary decisions on each cartel we got the number of involved companies, the duration of the cartel and could label the cartel type. Not all cartels were relevant for our study. Cartels were regarded as relevant if they had at least one publicly listed company. We also excluded two cases that we identified as having a decision before 2010. We were left with a total of 39 cartels in our dataset after the manual processing.

Table 3.1 below shows the summary statistics of the cartel dataset. The cartels differ in type and size and a cartel can be classified with multiple cartel labels. We used four different cartel labels, price fixing, quota allotment, market share allocation and bid rigging. The labeling was done based on the information we got through the different articles made by the commission.

Table 3.1: Cartel summary statistics

Variables	N	Mean	Sd	Min	Max
Size	39	6.7	4.8	3	26
Bid rigging	3	0.07	0.26	0	1
Market share allocation	12	0.3	0.46	0	1
Price fixing	34	0.87	0.33	0	1
Quota allotment	11	0.28	0.44	0	1
Duration (years)	39	7.1	5.86	1	35

3.2 Company data

Along with the cartel dataset we collected data on all the publicly listed firms that were participating in the cartels. We used Google search and Yahoo Finance to identify if a firm was publicly listed or not. Companies that either had been delisted before, or listed after the infringement decision, were not included. Individual fines after reduction were added for each company, as well as each cartel members' decision date. For many of the cases both subsidiaries and parent companies were fined by the European Commission. We kept both parent and subsidiary as separate entities if they both had stock information. If only the parent company was listed on a stock exchange, we only included the parent company in the dataset and added the fine of the subsidiary to the parent company's fine. In the cases where one subsidiary was not public and had multiple parent companies, we divided its' fine between the parent companies evenly, or according to their ownership shares if it was available. In the cases where the European Commission fined the same company for participation in multiple cartels on the same day, we combined the fines and only included the company once in the dataset. In total 164 company-case pairs were gathered, 124 of these were unique companies. As extra information, we added the country were the headquarter of each company was located and the associated continent. In addition to this we added the revenue and economic sector classification for each respective company from Refinitiv Eikon datastream. We used the revenue from one year before the decision for each company.

Table 3.2 contain summary statistics of this dataset. In total, 132 of the companies were fined, and 32 received immunity from the Commission. Companies vary in size and are mostly located in either Europe or Asia.

Table 3.2: Company summary statistics

Variables	N	Mean	Sd	Min	Max
American companies	20	0.12	0.32	0	1
Asian companies	68	0.42	0.49	0	1
European companies	71	0.43	0.5	0	1
Other companies	5	0.03	0.17	0	1
Immunity	32	0.2	0.4	0	1
Non-Immunity	132	0.8	0.4	0	1
Size of fine over revenue (non immunity)	132	1.96%	2.30%	0.0008%	0.14%
Fine over revenue >= 1%	49	0.3	0.45	0	1
Fine over revenue <1%	115	0.7	0.45	0	1

Statistics about which country the involved firms are coming from is shown in figure 3.1. We see that Japan by a clear margin is the country with the most companies in our sample.

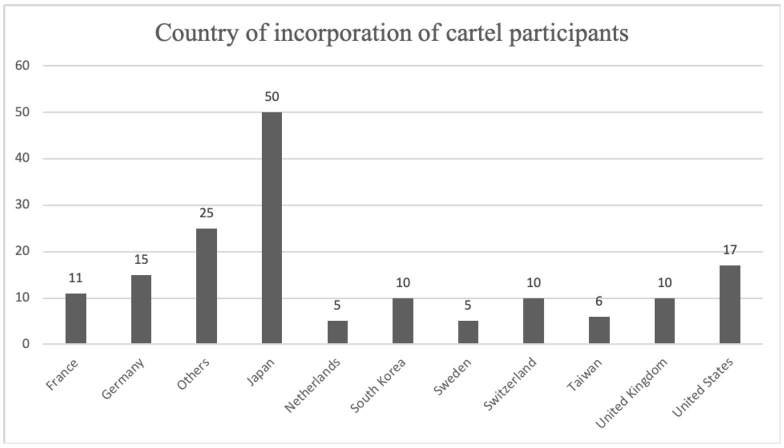

Figure 3.1: Companies classified by the location of their headquarter

Figure 3.2 shows the distribution of convicted companies on a yearly timeline. 2010 is the year with the most convicted companies and also the year with most cartel cases (6).

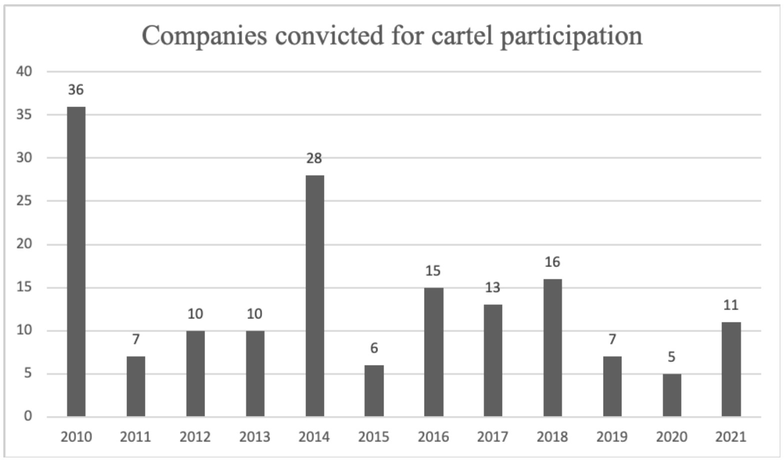

Figure 3.2: The number of public companies convicted for cartel participation by year in our dataset

The table below contain the companies that have repeatedly offended EU competition laws three or more times in 2010-2021, they are all large and well known companies. In total 25 companies are repeat offenders in our dataset.

Table 3.3: Repeat offenders

Company	Cases
Panasonic	4
JP Morgan	4
UBS	4
Hitachi	3
RBS	3
Denso	3
Samsung	3
Philips	3

3.3 Stock data

Daily stock data were downloaded for all companies in our dataset. The stock data was collected with the use of Yahoo Finance API in R. To cover all the cartel cases in our dataset, share prices were retrieved from 01.01.2009 to 26.05.2021. The adjusted closing price of each stock were used as it adjusts for dividends and splits. For stocks that were traded on multiple exchanges, for example on both NYSE and TSE, the home country's stock exchange was used. In addition to the daily stock data, a corresponding local market index for every company were also added to the dataset. A full list of stock tickers and index tickers can be found in table A4.1 in the appendix.

The data were lagged with one day for all companies and indexes that were listed on the east Asian stock exchanges (South Korea, Taiwan, Japan, Singapore). This was done to adjust for the difference in stock market opening hours because of time zone differences. Decision time is not available for all cases, but time zone differences make it probable that the first trading day on the east Asian markets after the decision will be the upcoming day.

3.4 Twitter data

The largest collection of data is the Twitter dataset. To gain access to Twitter, a Twitter developer account was obtained through an application for academic research. API keys and authentication tokens were provided with the account, which made it possible to extract historical tweets using the programming language R.

The Twitter academic research API uses the V2 endpoint, which is new in 2021 (Twitter, 2021a). None of the old R packages that we could find worked on this endpoint. In order to extract the data, a self-made loop was thus created in R. The loop iterated through every search word in a pre-made query list, previously made in Excel. For every search word, the corresponding case number was matched up against the same number in another sheet containing a range of dates from 70 days before to one month after the decision date. This gave us a timeline of tweets for each company over a 100-day period.

The Twitter developer account has some limitations which made the collection of tweets a somewhat time-consuming process. The maximum number of tweets per query is sat to 500 by Twitter and the maximum iteration rate is 900 queries every quarter (Twitter, 2021c). We worked around the iteration limit by including a three second sleep timer between every iteration of the loop. Most of the companies had less than 500 tweets in the daily timespan, but for larger companies like Sony and Samsung 500 tweets were not close to cover the daily Twitter volume. To overcome this issue, a nested loop was made to work around the 500-tweet limitation. If more than 200 tweets were collected after the first query, then another loop was initiated from the time of the last gathered tweet to collect more from that day. As some of the companies in our dataset gets thousands of tweets written about them every day and we had limited capacity, we decided to collect at most 2000 tweets per company per day.

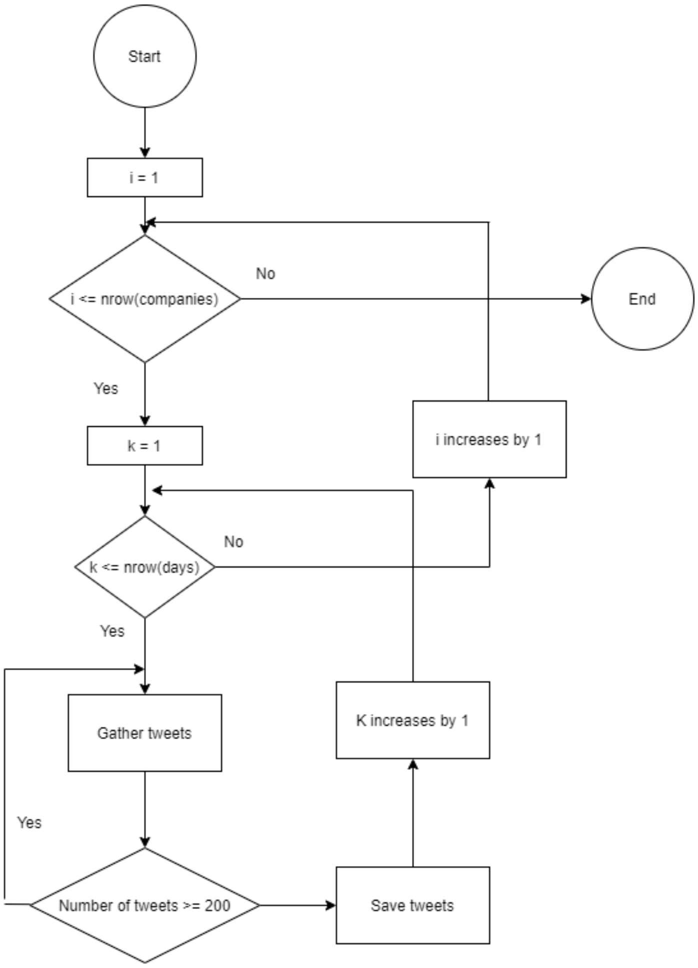

Figure 3.3: Flowchart of the tweet downloading process

Central to the process of collecting tweets was the choice of query names. Companies are referred to by many different names on Twitter, which makes it difficult to obtain all the relevant tweets about a company. As the academic account was limited to 10 million tweets, one query name was used for most companies. It was necessary to use multiple query names on companies which were identified as more uncommon to secure that a reasonable amount of tweets were collected. As a general rule the same company names as the EU commission used in their press release were used in the query. In the cases where some names were too similar to other words, or were the commission used abbreviations when naming the company, then a manual search for the name was done on Twitter. This was done to clarify if the name was used on Twitter or not. A full list of the query words can be found in the appendix. We searched for the mention of the word and not the hashtag. The reasoning is that we did not want to be limited by the hashtag and wanted all the general tweets about the company. We also did not query specifically for the cashtag of the company, as many of the firms were thought to be too small for this to result in any tweets.

Queries were only done for English tweets. This was mainly done because of practical reasons like capacity limitations and difficulties associated with multilingual sentiment analysis. Only including English tweets is a weakness of this study, but the main language used on Twitter is English, so it should not have too much to say. Retweets were kept as a person that retweets a message probably share the opinions of the original tweet. Weekends were included even though it is a non-trading day, as news that was to be shared during the weekend could affect stock prices when the stock exchange opens on Mondays. The time for tweet gathering was sat between 07:00 GMT to 19:00 GMT as it covers the opening hours in the European stock markets and the time when the Commission is most likely to publishes its press releases. We had to specify the time of the day because we wanted the most relevant tweets and only collected 2000 a day. The API searches by default from the latest time to the earliest time of the day, meaning it will find tweets from 19:00 first.

Before the tweets could be used in the analysis they needed some preliminary cleaning. Twitter messages contains many types of signs and other things which were not needed in our analysis. We also removed all mentions, links, numbers, punctuations, digits and symbols that are not a part of the English language. Duplicates in the cases where more than one query name were used for a single company were also removed. Lemmatisation and stemming were not done as the packages used in R for sentiment analysis works around these problems (Alex, 2019).

Summary statistics about the Twitter data can be seen inn table 3.4. Approximately 5 million tweets were downloaded in total, with large variations between companies.

Variables	N	Mean	Sd	Minimum Tweets	Max tweets
American companies	424781	22357	24824	25 - Trane Inc	76340 - Carpenter
Asian companies	2105328	31899	60934	1 - Holy Stone Enterprise	198099 - Samsung
European companies	2121693	30749	44383	1 - Ercros	184602 - MAN
Other companies	77365	15473	19056	79 - CSAV	44107 - Whirlpool
Motor companies	1028317	15581	31504	5 - Nachi-Fujikoshi	184602 - MAN
Electronic companies	2197623	53601	70356	1 - Holy Stone Enterprise	198099 - Samsung
Financial companies	1138545	4542	40078	1877 - Credit Agricole	160386 - Barclays
Other companies	364682	13507	37708	1 - Ercros	186485 - Panasonic

Table 3.4: Aggregated tweet statistics

Figure 3.4 shows the distribution of tweets including the word "cartel" from all the days we collected tweets. It shows that there are mentiones of cartels in the days before the decision, but most of the talk happens on the day with the actual event. The large amount of tweets 25 days before the decision is from one case.

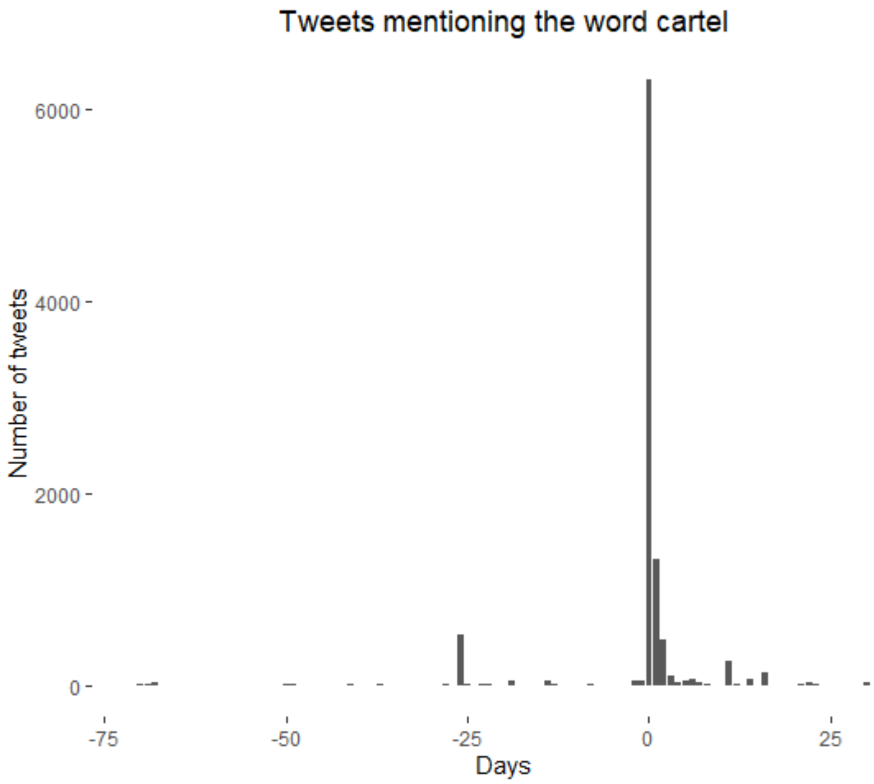

Figure 3.4: A timeline showing the number of tweets that mention the word cartel

3.5 Data limitations

The main limitations of our data is related to the Twitter dataset, some of which come as a consequence of limitations in the company and cartel dataset. Ideally we would have had an even larger sample size of companies and cartels, as a larger sample size generally makes the results more credible. The main limitation of the Twitter dataset, is that we dont collect tweets written in another language than English. This makes it likely that companies that are originating from English speaking countries are more represented in our Twitter dataset than companies from non-English speaking countries. This is especially true for small unknown companies located in non-English speaking countries outside of the EU. This result in us getting an inaccurate representation of the real mood/sentiment of some of the companies, which may affect the results of the analysis. Ideally we would have gathered all tweets about all the companies in the days close to the event, but because of uncertainty connected to capacity limitations this was not possible. Alternatively a sampling feature that only collected a percentage of tweets every day would have been a viable alternative, but this is not a feature Twitter supports for the historical archive at the moment. In addition to this we see the evolution of Twitter as a potential limitation. Twitter as a platformed have evolved much from 2010 to 2021 and there might be differences in how it is used now compared to ten years ago. The doubling from 140- to 280 characters in November 2017 is an example of a change that may affect the analysis. We will therefore test this in the robustness check by only looking at tweets created after 2017 and see if it changes the results.

Another potential problem with the Twitter data is the variation between companies and cartel cases when it comes to their tweet amount. Figure 5 shows the companies categorized by how many total tweets they have about them. The companies that have very few tweets may effect the result largely in both directions because variations here could lead to much larger abnormal mood. This will be taken into account in both the primary method and in the robustness check. In the figure bellow very few equals less than 100 tweets, few less than 1000 tweets, medium less then 10 000 tweets, many less than 100 000 tweets and very many is more than 100 000 tweets.

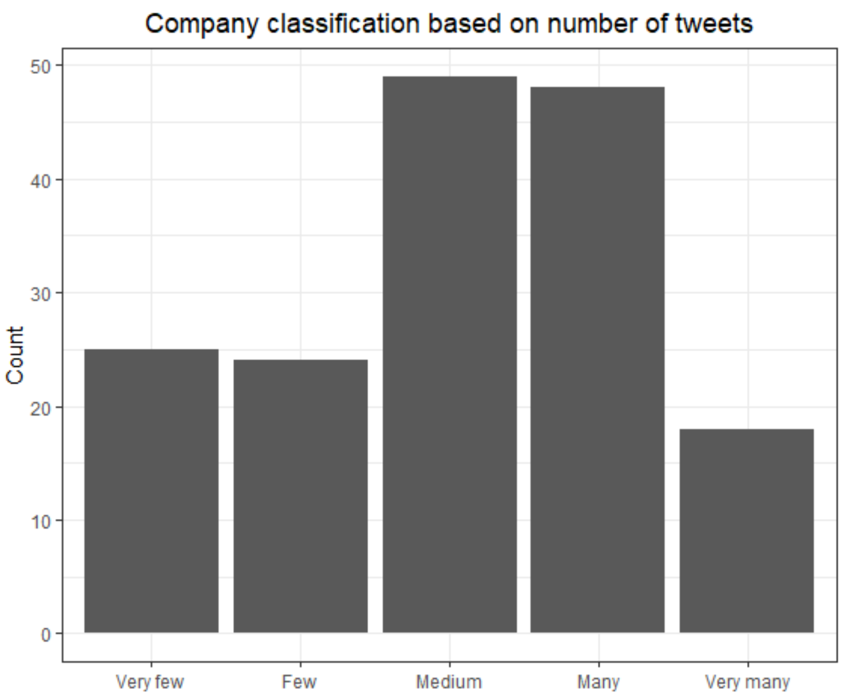

Figure 3.5: Companies classified by their number of tweets

4 Research Questions and Hypotheses

In this section, we present the research questions and the hypothesis that we want to test. The research questions aim to validate previous studies' findings and to investigate the possible use of Twitter data in explaining abnormal returns.

Research question 1: Does cartel convictions create abnormal returns? This has been researched before, and we aim to validate previous findings on the matter with newer data. If the cartel sentencing provides new information, then the expected outcome is to find abnormal stock return on the event day and in the event windows surrounding the event day.

Research question 2: Can Twitter data be used to estimate private sanctions? In theory information availability should affect prices in some direction. The logic being that the the more people that know about something, the larger the potential reaction. This make sense because an efficient market is based on information availability.

To help us answer the research questions, we have created three main hypotheses that test our expected findings:

Hypothesis 1: The announcement of cartel decisions is associated with an abnormal stock price reaction. Previous research has found small, but significant negative abnormal returns on both the decision day and in narrow event windows around the decision. We therefore expect the returns to be negative on the decision day. This will be tested through the event study methodology.

Hypothesis 2: The announcement of the cartel decision is associated with an abnormal Twitter mood reaction. We find it reasonable to think that sentiment will be lower at decision than the companies' average sentiment score. We expect the public to react negatively on Twitter to cartel convictions. The magnitude of the reaction will depend on how informed investors are about the negative effects of cartels.

Hypothesis 3: Twitter sentiment and the count of cartel tweets can explain differences in abnormal returns between companies The goal is to find out if Twitter variables can explain abnormal stock returns. This will be tested with cross-sectional regressions and decision trees.

5 Methodology

5.1 Event Study

To assess the consequences of the European Commission's infringement decision on the companies in the dataset, an event study methodology following the market model will be utilized as described by MacKinlay (1997). The intuition behind the event study is that we by calculating the abnormal stock returns around the event, may isolate and measure the event specific effects by comparing actual and expected returns around the event. In order to obtain the abnormal returns we use a benchmark return calculated with the use of the local market index for each respective stock. The underlying assumption of the market model is that there exists a linear relationship between the stock return and its associated market return and that markets are at least semi strong.

The market model as it is defined by MacKinlay (1997):

$$R_{it} = \alpha_i + \beta_i R_{mt} + \epsilon_{it} \tag{5.1}$$

$$E(\epsilon_{it}) = 0 \qquad Var(\epsilon_{it}) = \sigma_\epsilon^2$$

Here R_{it} is the normal return for security i at time t and R_{mt} is the market return at time t of the corresponding market index. The α and β parameters are estimated over the estimation window by Ordinary Least Squares regression, and the ϵ_{it} is the estimator of the abnormal returns. The size of β_i shows the stock's sensitivity to the chosen market index.

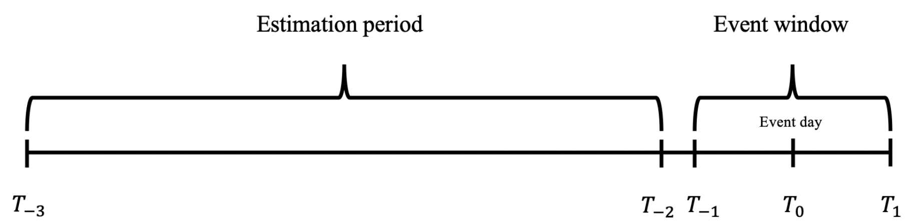

Figure 5.1: Event Study timeline

Market- and stock returns can be calculated in two ways, either by calculating simple returns or by calculating the natural logarithm of the returns. The difference between the methods are according to Wooldridge (2013) small when the results are close to 0. We have chosen to use the formula for simple returns in this thesis. Normal returns are given by the following formula:

$$R_{it} = \frac{P_{it} - P_{it-1}}{P_{it-1}} \tag{5.2}$$

Abnormal returns (AR) are calculated for each security i for every day t in the event window and is defined as the difference between actual returns R_{it} and the estimated normal returns R_{mt}. The α_i and β_i are the estimated coefficients from the estimation of the normal returns.

$$AR_{it} = R_{it} - \hat{\alpha}_i - \hat{\beta}_i R_{mt} \tag{5.3}$$

We have used an estimation window of 120 trading days, starting 180 trading days before the event and ending 60 trading days before the event. There is no set length for estimation windows in literature, but this should be a good window balancing the trade off between improved estimation accuracy and potential parameter shifts.

We use several event windows in this study. Some windows include days before the event and is motivated by the possibility of information leakage.

In order to study the impact of cartel convictions we have aggregated the results over each security and event window. For the different event windows we start of by calculating the cumulative abnormal returns (CAR) for company i in the duration of the event window L:

$$CAR_{iL} = \sum_{i=1}^{L} AR_{it} \tag{5.4}$$

The last thing we do with the stock data is to aggregate the results over the different event windows. We fist calculate the average abnormal return AAR_t and the cumulative average abnormal return $CAAR_{NL}$. The average abnormal return is the the average return of all securities on day t in the event window. The cumulative average abnormal return is calculated by taking the sum of the CAR for every company i over the event window L and dividing it by the total number of company event pairs N.

$$AAR_t = \frac{1}{N}\sum_{i=1}^{N} AR_{it} \tag{5.5}$$

$$CAAR_L = \frac{\sum_{i=1}^{L} CAR_{iL}}{N} \tag{5.6}$$

or alternatively

$$CAAR_L = \sum_{i=1}^{L} AAR_t \tag{5.7}$$

The statistical test used to validate the significance of the abnormal returns is the BMP test by Boehmer et al. (1991). This test is explained in appendix A7.

5.2 Twitter sentimental analysis

Sentiment analysis is a natural language processing technique which uses computational linguistics and textual analysis to analyze the subjective information from a text (Liu & Zhang, 2012) (Mooney et al., 2005). There are two main methods that are used when researchers are conducting sentiment analysis's: machine learning and dictionary-based analysis (Kearney & Liu, 2014). The machine learning approach uses a prelabeled dataset to train and learn patterns which it uses to classify the unlabeled data. The dictionary approach uses a predefined dictionary containing words which are classified as either positive or negative and uses it to evaluate the meaning of a text.

We have in this thesis conducted a dictionary-based sentiment analysis. There are two main reasons for this. Firstly, previous research on sentiment analysis suggests that there are small differences in using a machine learning approach over the much simpler dictionary approach on social media data (Hutto & Gilbert, 2015). Secondly, none of the prelabeled datasets that we found were evaluated to be large and good enough for our use. To validate our results we have chosen to use two sentiment packages in R. The results of a sentiment analysis are depending on which dictionary that is being used, because of the difference between the included words and how the words are weighted. The two packages we use are the sentimentr and VADER package. Inspired by Mariuzzo et al. (2020a) we use two different methods to reflect the difference between how an assumed layperson and someone who is familiar with the bad effects of cartel behaviour would interpret the tweets. We use the same reasoning as stated in their paper. Their reasoning is that if we can show that people that are more aware of the seriousness of cartels have a larger impact on the abnormal returns and therefore the valuation of a business, then that could work as an important policy message that improving competition culture through increased public awareness of the downside of anti-competitive behaviour can improve the deterrent effect of competition policy.

5.2.1 SentimentR and VADER

The sentimentr package is a lexicon-based sentiment analysis package which calculates the sentiment of a tweet by evaluating the individual words in the tweet and to some degree the context of the word. It does this by using valance shifters. Valance shifters are words which impact the interpretation of the following words. One of these shifters are "negators" which flips the sign of a polarized word, example "I do not like" will be treated as negative instead of positive. It also uses "amplifiers" and "de-amplifiers" which intensifies or reduce the impact of a polarized word. The last thing it does is to use "adversative conjunctions" which overrule the previous clause that contained a polarized word, example: "I like it but it is not worth it" (Rinker, 2019).

We use a customised lexicon that combines the standard lexicon "Syuzhet" made by Jockers (2015), with a lexicon made by Mariuzzo et al. (2020b). The Syuzhet lexicon is a very general dictionary that contain 10748 words, while the lexicon made by Mariuzzo contain 608 domain specific words that are relevant to our analysis (cartel, collusion, price fixcing etc). The main reason for using the customised lexicon is to see how a person that is familiar with the gravity of cartel behaviour would interpret the tweets. The polarity and intensity score of each word in the custom dictionary can take a value from -1 to 1.

To see how an assumed layperson would interpret the tweets we use the Vader package in R. Vader is a lexicon and rule-based sentiment analyst tool which is specifically attuned to sentiments expressed in social media. The algorithm uses the Vader lexicon (2014), a lexicon that is empirically validated by ten independent human judges. It also uses degree modifiers like the sentimentr package. The dictionary contains 7520 words and emojis where each is scored on a scale between −1 and 1 (Hutto, 2014b). We use VADER in the robustness section of this thesis.

5.3 Variable creation

From the collected tweets we create three different variables that will be tested. The two first are simple and similar to the newspaper variables tested by (Bosch & Eckard, 1991) and (Mariuzzo et al., 2020a), while the last is made to see if changes to normal sentiment affects stock returns. The three variables are:

1. Count of tweets containing company name and cartel in the event window
2. The sentiment of the tweets containing company name and the word cartel on a continues sentiment scale.
3. Abnormal sentiment

5.3.1 Abnormal Sentiment

In order to find out if Twitter sentiment can explain some of the abnormal stock returns in relation to the cartel conviction, we have chosen to calculate the abnormal sentiment for each company in many of the same event windows used for the abnormal stock returns. This is a novel method that is similar to the constant mean model used in event studies, but we do not look at the returns of sentiment. The reason we do not look at the returns of sentiment is because of the fact that sentiment can be negative which complicates the calculation and interpretation of returns. Instead we use the difference between the average sentiment of all tweets on a day (mood of the day) and the average mood for all the days in an estimation period. This makes it possible for us to calculate the abnormal mood, which can be aggregated over event windows and companies. The theoretical advantage of such a variable is that it makes it possible to compare the sentiment between the companies as we look at the change in relation to the company's own normal sentiment.

The normal or average mood for each company is created from an estimation window starting 70 days- and ending 16 days before the event. Only companies that have tweets in at least 60 out of the 100 days that tweets were collected from was included. This was done to deal with potential extreme outliers created by companies that have a very low Twitter presence. All tweets were classified into positive (1), neutral (0), or negative (-1) based on the evaluation done by sentimentr or VADER. This was done to remove the neutral tweets that are not polarising in any direction. The threshold for negative and positive tweets were set to -0.3 and 0.3 for sentimentr. There is no consensus on where the threshold should go in literature, so we decided on the threshold after looking at the distribution of the scores of all the tweets. Tweets that had a higher value than the threshold value were classified as positive and given a score of 1, while the tweets that were lower than the threshold were classified as negative and assigned a value of -1. Tweets in between the values were classified as neutral and given a value of 0. For VADER we used the suggested threshold made by the creator of the package, all tweets lower or equal to -0.05 were classified as negative, while all higher or equal to 0.05 were classified as positive.

The mood on Twitter on day t for company i is given by the following formula. The formula have been used in other studies looking at the relationship between stock prices and Twitter sentiment (Ranco Gabriele, 2015).

$$MOOD_{it} = \frac{M_{it,pos} - M_{it,neg}}{M_{it,pos} + M_{it,neg}} \tag{5.8}$$

Here $M_{it,pos}$ is the number of positive tweets in a day, while $M_{it,neg}$ is the number of negative tweets in a day. The mood can thus be positive or negative for any given company on any given day.

We calculate the average mood $AMOOD_i$ for each company by taking the sum of all the $MOOD_{it}$ in the estimation period EW and dividing by the number of days in the period.

$$AMOOD_i = \frac{1}{EW}\sum_{t=1}^{EW} MOOD_{it} \tag{5.9}$$

The abnormal mood on day t for company i in the event window is calculated by subtracting the mood of a day from the average mood calculated from the estimation period. We divide by the absolute value of the average mood to get the value in percent.

$$ABMOOD_{it} = \frac{MOOD_{it} - AMOOD_i}{|AMOOD_i|} \quad (5.10)$$

Cumulative abnormal mood is calculated as the rolling sum of all the abnormal mood in the length of the event window L.

$$CABMOOD_i = \sum_{t=1}^{L} ABMOOD_{it} \quad (5.11)$$

$$CAABMOOD_L = \frac{\sum_{i=1}^{L} CABMOOD_{iL}}{N} \quad (5.12)$$

We classify each company into one of three categories according to how negative the abnormal mood was on Twitter. The polarity of an event is derived from the ABMOOD in event window [0,2]. The distribution of the polarity is bell shaped (A9.1) and we set the cutoff at the 25 percentile for the negative events, 75 percentile for the positive events and categorise the rest as neutral. The justification for our selected cutoff values is that sentiment should be regarded in relative terms, at least in the context of related events. Sentiment polarity has no absolute meaning, and provide in our case just an ordering of events according to how much they differ from their own "normal" sentiment (Ranco Gabriele, 2015).

5.4 Cross sectional regression

To test whether our created variables poses any determinant power on the cumulative abnormal returns, we use cross sectional regression (James et al., 2014). Cross sectional regression is a tool often used in combination with event studies and have, for example, been used by both Ulrich (2018) and Aguzzoni et al. (2013). In cross sectional linear regression both the dependent and independent variables are associated with the same period in time. We plan to mostly use single factor ordinary least squares regressions in the form of different binary independent variables to see which that affects the abnormal returns. In addition to our created variables we will test other control variables, some of which have been found to be significant in earlier studies.

5.5 Decision trees

Decision trees are used as a robustness test after the cross sectional regression. The trees identifies which of the predictors that are the most useful and finds the interaction between predictors. The trees are plotted with simple if or if not questions, which negates the need for normal assumptions, for example linearity and parametric statistics. In the tree, the top node shows the most important variable for the outcome variable. This same logic works for all further branches of the tree. The further down in the tree, the less important for the outcome of the CAAR (James et al., 2014). We use regression trees and not classification trees as the outcome variable is continuous.

6 Analysis

6.1 Estimating the abnormal rate of return

The first analysis address the first research question with associated hypothesis. The analysis is done on the main sample which includes all companies in our dataset and on 16 subsamples. The CAAR in percent for each sample with significance and other statics can be found in table 6.2 below. The results from the main sample which includes all companies in our dataset can be seen in figure 6.1. From the graph we see that there is a steady drop in average abnormal returns from around 11 days before the event until 10 days after the event. The drop is according to the t statistics from the BMP test not significant for the largest event window [-25,25], but the second largest window of [-15, 15] have a CAAR of -1.63% and is significant on a 10% level. The event day itself is notably not significant even though the AAR on the day is mildly negative (-0,28%). The event window capturing the effect of the conviction [0,10] is significant and shows a CAAR of -1.24%. Between 40 and 50 percent of the companies are yielding positive CAR in the different event windows. This should contribute to increased variance which can explain why some of the windows are not significant. Overall the findings are quite similar to those of Mariuzzo et al. (2020a) who also found small negative abnormal returns around the decision date.

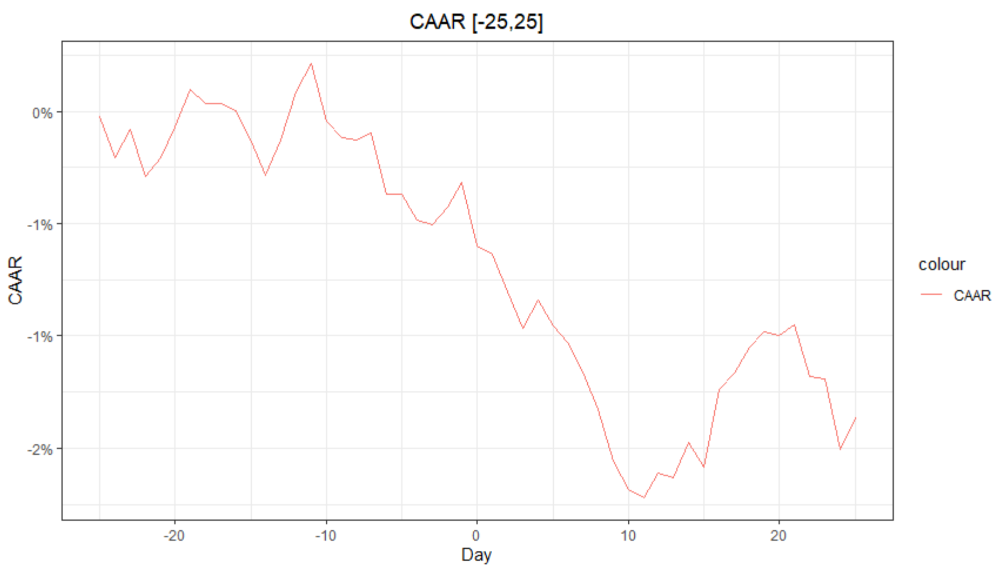

Figure 6.1: CAAR of entire sample

In figure 6.2 we have isolated the companies that received immunity from the European Commission in one sample and the companies that received fines in another. We see from the graph that the two samples behave similar until around five days before the event day. This could imply that there is some information leakage in the days before the conviction. The companies who received immunity have positive, but insignificant CAARs in all event windows. Based on the graph it looks like an event window of [-5,0] could have been significant. The percentage of companies that yield positive CAR in this sample is between 47% and 63% in the different event windows. In contrast to the companies that received immunity, the sample of fined companies have negative CAAR in all event windows. Window [0,10] is especially significant with a CAAR of -1.69%. The percentage of companies with positive CAAR in this sample is lower compared to the immunity sample, ranging from 40% to 45%. The results are consistent with the findings of Ulrich (2018), and could suggest that companies that are not penalized do not suffer significant abnormal returns on average from the indictment, whereas companies that are fined do. The reaming subsamples are only containing companies that received a fine from the European Commission. This was done to isolate effects that otherwise would have been affected by companies that received immunity.

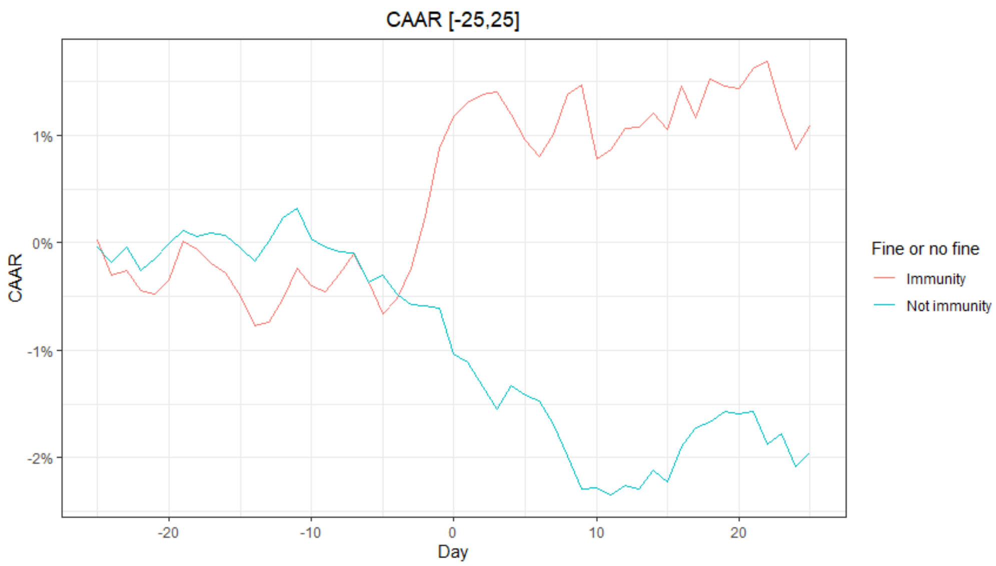

Figure 6.2: Sample of immune and fined companies

Figure 6.3 shows the CAAR of fined companies split into two samples depending on the size of the fine as a percentage of revenue. From the graph we see that the two samples correlate right up until the event day. Multiple event windows are significant for companies where the fine amount for a larger share than 1% of revenue. The percentage of companies that have a positive CAR in the different event windows are low compared to the other samples, with positive CAR being 37% over the [-25,25] window and only 29% in the [-1,1] window. The AAR on the event day itself is -0.96%, non significant and 37% of the companies have positive AR on the day. For the companies that receive a fine less than 1% of their annual revenue, the drop in CAAR is smaller. Only the [0,10] event window is significant and the drop in CAAR is only -0.91%.

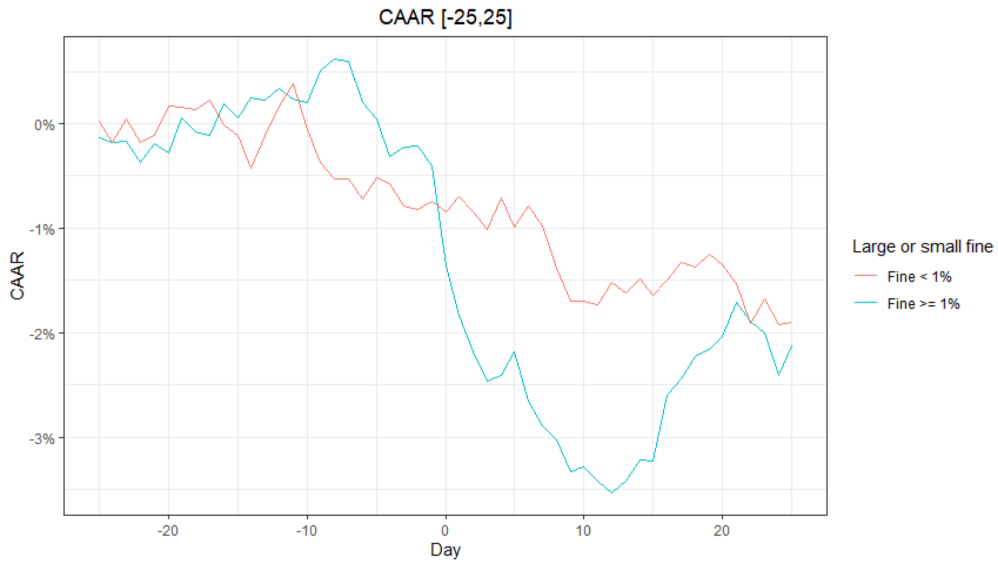

Figure 6.3: Companies categorised by fine as a percent of revenue

Figure 6.4 shows the CAAR of the companies sampled by continent. From the graph we see that there is a considerably larger drop in stock prices in the Asian sample compared to the European and American sample for all event windows. The asian sample has significant negative abnormal returns of -4.72% over the large event window [-25,25]. The European companies react on average negatively on the event day, with significant negative CAAR in two event windows after the event day, [0,2] and [0,10]. This indicates an after-effect of the conviction, but on the longer event windows there are not significant results. The North American companies have no significant event windows, and are even showing a positive CAAR in the more narrow windows around the event. Overall the results are consistent with the results of Ulrich (2018) who also found the Asian companies to have significant negative returns in the long event windows.

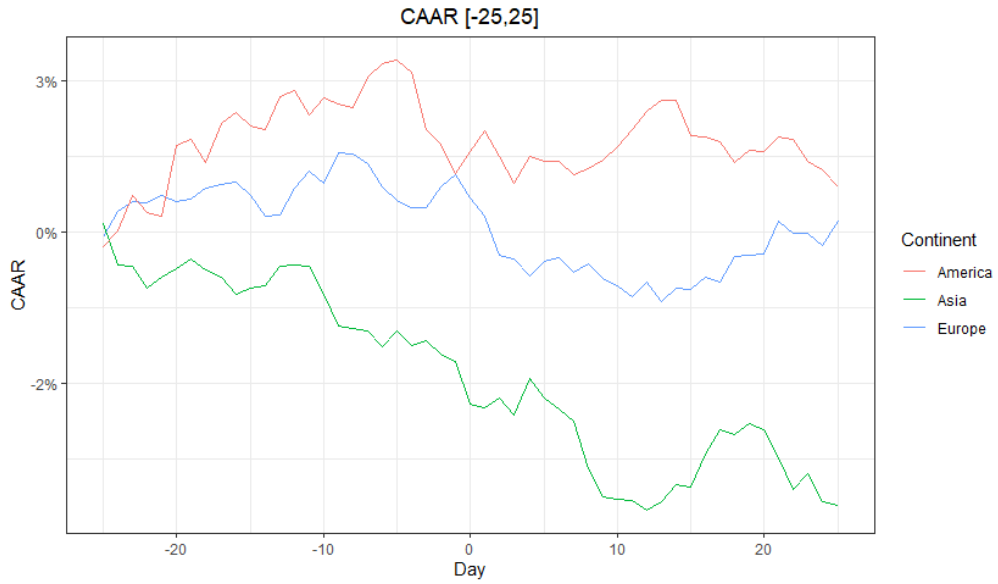

Figure 6.4: CAAR categorised by continent

In figure A11.2 we have one sample containing the 25% of companies with the lowest revenue and one with the 25% of companies with the highest revenue. From the graph, we see that the CAAR of the companies is behaving quite similar up until two days before the event. The small companies show a large drop in CAAR two days before the event, suggesting information leakage. Both event window [-5,5] and [0,10] are significant and have a CAAR of -1.99% and -2.73%. There are no significant event windows for the larger companies, which matches well with what we see on the graph.

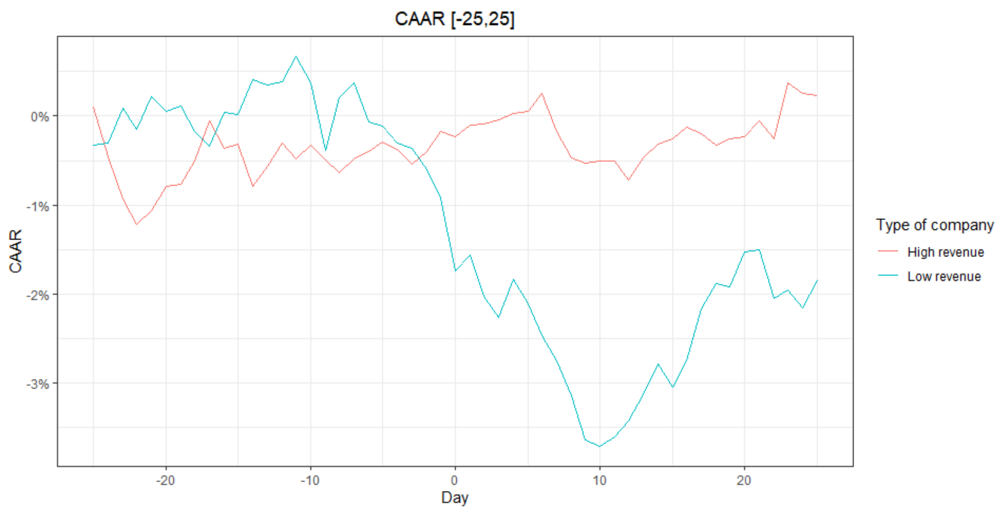

Figure 6.5: CAAR categorised by firm size

Figure 5 shows the CAAR categorised by economic sector. The financial sector is the only sector with positive CAAR over the whole event window, while firms categorized as industry have most negative CAAR. The CAAR of the finance sector is reminiscent of the European and American CAAR in figure 4, which makes sense as there are few financial companies from Asia in the dataset. There is no significant event windows for the finance sector, which makes sense because of the stable CAAR from 15 days before the event until the end of the period. The industrial sector experience a large drop of -7.70% over the duration of the [-25,25] window, the drop is significant on a 1% level. Non of the more narrow event windows or the event day itself is significant. The consumer cyclical sector have non significant CAAR in the long event windows, but experience a significant drop in the event windows which starts around the event day. The technology sector also suffers significant drops around the event day.

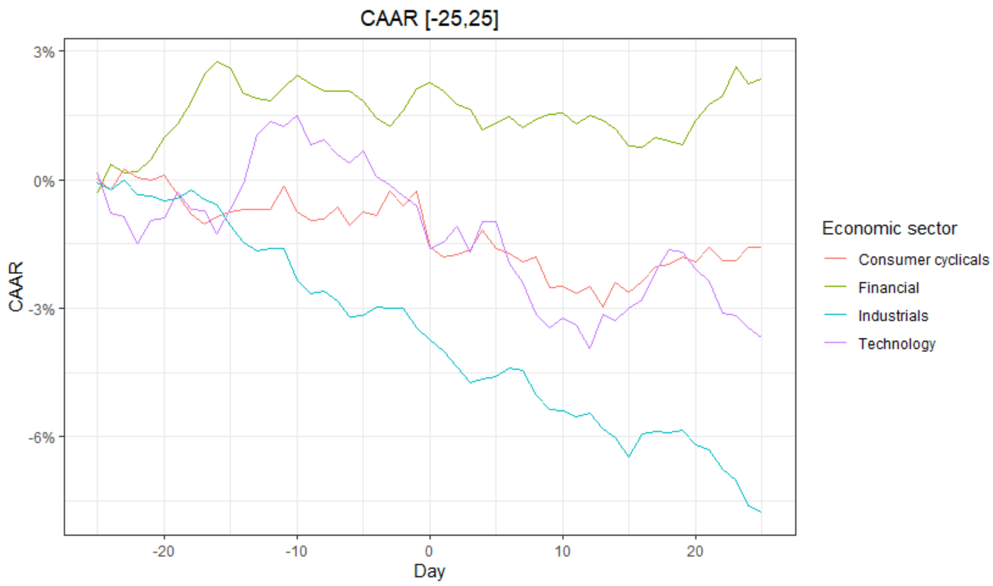

Figure 6.6: CAAR categorised by economic sector

The last figure show the companies classified into groups according to their abnormal mood on the day of the decision. From the graph we see that the companies with the worst abnormal mood on Twitter are declining in the larger event window [-25,25], while the companies with a positive mood experience increases their returns. This would have been a more interesting observation if it hadn't been for the fact that the companies with an average mood have the most decline during the event window.

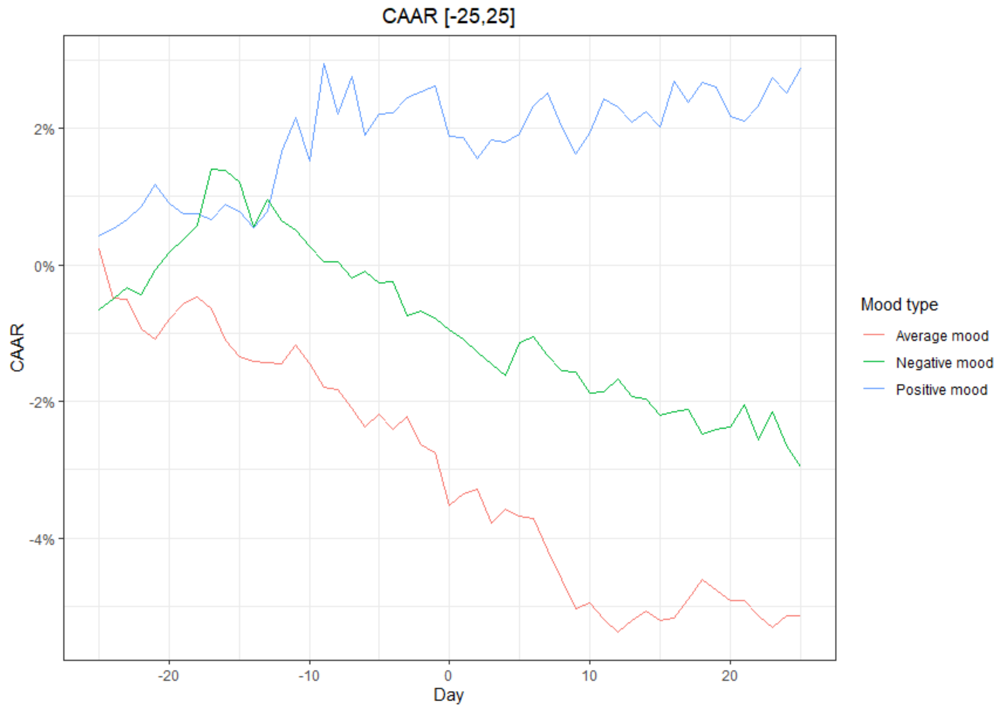

Figure 6.7: CAAR categorised by abnormal mood

To summarize, we see that EU cartel convictions cause significant abnormal returns in some, but not all subsamples. The null hypothesis of zero cumulative abnormal returns is thus conditionally rejected by the results of the event study. The magnitude of abnormal returns seem to be conditioned on several factors and will be further tested in section 6.3 and 6.4

Table 6.1: Abnormal return with different event windows and subsamples

Sample	Event window	CAAR	T-value	P-value	Observations	Positive CAR
All companies	(-25) to + 25	-1,48 %	-0,94	0,35	156	46 %
	(-15) to 15	-1,63 %	-1,84	0.07*	156	46 %
	(-10) to 0	-0,60 %	-0,82	0,413	159	43 %
	(-5) to 5	-0,62 %	-1,42	0,16	160	47 %
	(-1) to 1	-0,21 %	-0,65	0,52	164	45 %
	0	-0,28 %	-0,7	0,48	164	43 %
	0 to 2	-0,49 %	-1,51	0,133	164	48 %
	0 to 10	-1,24 %	-2,91	0.004***	160	42 %
Immunities only	(-25) to + 25	0,57 %	-0,01	0,99	30	63 %
	(-15) to 15	1,10 %	0,41	0,684	30	63 %
	(-10) to 0	1,48 %	1,28	0,21	31	55 %
	(-5) to 5	1,59 %	1,11	0,28	31	58 %
	(-1) to 1	1,06 %	1,34	0,19	32	62 %
	0	0,29 %	1,08	0,29	32	47 %
	0 to 2	0,50 %	1,03	0,31	32	56 %
	0 to 10	0,64 %	0,32	0,75	31	48 %
Excluding immunities	(-25) to + 25	-1,97 %	-1,07	0,29	126	41 %
	(-15) to 15	-2,29 %	-2,43	0.02**	126	41 %
	(-10) to 0	-1,11 %	-1,79	0.08*	128	40 %
	(-5) to 5	-1,16 %	-2,44	0.02**	129	44 %
	(-1) to 1	-0,52 %	-1,5	0,14	132	41 %
	0	-0,42 %	-1,28	0,2	132	42 %
	0 to 2	-0,73 %	-2,24	0,03	132	45 %
	0 to 10	-1,69 %	-3,54	0.001***	129	40 %
Fines >= 1% of revenue	(-25) to + 25	-1,99 %	-0,86	0,39	49	37 %
	(-15) to 15	-3,29 %	-2,05	0.05**	49	37 %
	(-10) to 0	-1,57 %	-0,98	0,332	49	43 %
	(-5) to 5	-2,23 %	-3,17	0.003***	49	33 %
	(-1) to 1	-1,62 %	-2,55	0.014**	49	27 %
	0	-0,95 %	-1,3	0,2	49	37 %
	0 to 2	-1,77 %	-2,6	0.012**	49	33 %
	0 to 10	-2,91 %	-3,54	0.001***	49	31 %
Fines <1% of revenue	(-25) to + 25	-1,25 %	-0,57	0,57	107	50 %
	(-15) to 15	-0,88 %	-0,93	0,35	107	50 %
	(-10) to 0	-0,17 %	-0,24	0,81	110	43 %
	(-5) to 5	0,09 %	0,02	0,98	111	53 %
	(-1) to 1	0,39 %	1,41	0,16	115	53 %
	0	0,00 %	0,52	0,6	115	46 %
	0 to 2	0,06 %	0,39	0,7	115	54 %
	0 to 10	-0,50 %	-1,25	0,21	111	47 %
European companies	(-25) to + 25	0,30 %	0,44	0,66	50	42 %
	(-15) to 15	-1,61 %	-0,87	0,39	50	42 %
	(-10) to 0	-0,28 %	0,1	0,92	51	49 %
	(-5) to 5	-1,04 %	-1,41	0,16	52	42 %
	(-1) to 1	-0,49 %	-0,39	0,7	54	43 %
	0	-0,37 %	-0,33	0,74	54	44 %
	0 to 2	-1,325	-2,33	0.023**	54	37 %
	0 to 10	-1,71 %	-2,05	0.05**	52	40 %
Asian companies	(-25) to + 25	-4,72 %	-3,66	0.001***	56	38 %
	(-15) to 15	-3,35 %	-3,19	0.002***	56	38 %
	(-10) to 0	-1,87 %	-3,2	0.002***	56	29 %
	(-5) to 5	-1,18 %	-1,91	0.06**	56	45 %
	(-1) to 1	-0,89 %	-2,82	0.01***	57	32 %
	0	-0,72 %	-3,4	0.001***	57	32 %
	0 to 2	-0,615	-1,79	0.08*	57	46 %
	0 to 10	-2,27 %	-3,66	0.001***	56	34 %
North American companies	(-25) to + 25	0,75 %	1,73	0,11	15	60 %
	(-15) to 15	-0,51 %	0,44	0,67	15	60 %
	(-10) to 0	-0,89 %	0,13	0,9	16	56 %
	(-5) to 5	-1,67 %	-0,8	0,44	16	44 %
	(-1) to 1	0,23 %	0,76	0,46	16	56 %
	0	0,37 %	1,6	0,13	16	69 %
	0 to 2	0,29 %	0,94	0,36	16	62 %

	0 to 10	0,23 %	0,56	0,58	16	62 %
Industrials	(-25) to + 25	-7,70 %	-4,52	0.0001***	37	30 %
	(-15) to 15	-5,40 %	-3,9	0.0004***	37	30 %
	(-10) to 0	-1,39 %	-1,3	0,2	37	49 %
	(-5) to 5	-1,42 %	-1,39	0,17	37	41 %
	(-1) to 1	-0,99 %	-1,34	0,19	37	35 %
	0	-0,29 %	-0,66	0,51	37	38 %
	0 to 2	-0,91 %	-1,46	0,15	37	41 %
	0 to 10	-1,92 %	-2,4	0.02**	37	41 %
Consumer cyclicals	(-25) to + 25	-1,60 %	-0,86	0,4	21	38 %
	(-15) to 15	-1,88 %	-1,2	0,24	21	38 %
	(-10) to 0	-0,83 %	-0,38	0,71	21	52 %
	(-5) to 5	-0,86 %	-1,1	0,28	21	43 %
	(-1) to 1	-1,20 %	-1,89	0.07*	21	33 %
	0	-1,31 %	-1,22	0,24	21	38 %
	0 to 2	-1,47 %	-2,28	0.03**	21	33 %
	0 to 10	-2,25 %	-2,12	0.05**	21	43 %
Financials	(-25) to + 25	2,64 %	1,73	0,11	17	35 %
	(-15) to 15	-2,30 %	-1,7	0,11	17	35 %
	(-10) to 0	-0,14 %	-0,32	0,75	19	42 %
	(-5) to 5	-0,63 %	-1,12	0,28	20	50 %
	(-1) to 1	0,45 %	0,87	0,39	23	57 %
	0	0,14 %	1,02	0,32	23	57 %
	0 to 2	-0,36 %	-1,02	0,32	23	52 %
	0 to 10	-0,63 %	-0,67	0,51	20	35 %
Technology	(-25) to + 25	-2,84 %	-1,8	0.08*	27	44 %
	(-15) to 15	-2,30 %	-0,99	0,33	27	44 %
	(-10) to 0	-3,10 %	-2,79	0.01***	27	19 %
	(-5) to 5	-1,65 %	-1,68	0,1	27	37 %
	(-1) to 1	-1,08 %	-2,09	0.05**	27	30 %
	0	-1 %	-2,48	0.02**	27	33 %
	0 to 2	-0,51	-0,7	0,49	27	56 %
	0 to 10	-2,85 %	-3,06	0.01***	27	26 %
25% companies with worst sentiment (most negative)	(-25) to + 25	-2.45%	0.04	0.97	25	40 %
	(-15) to 15	-3.72%	-1.78	0.09*	25	40 %
	(-10) to 0	-1.3%	-0.33	0.74	26	46 %
	(-5) to 5	-0.99%	-0.96	0.35	26	54 %
	(-1) to 1	-0.41%	-0.6	0.55	28	46 %
	0	-0.18%	0.27	0.79	28	54 %
	0 to 2	-0.51%	-0.96	0.35	28	50 %
	0 to 10	-0.83%	-0.73	0.47	26	46 %
25% companies with best sentiment (most positive)	(-25) to + 25	2.32%	0.82	0.42	19	58 %
	(-15) to 15	0.98%	0.68	0.51	19	58 %
	(-10) to 0	0.36%	0.46	0.65	20	45 %
	(-5) to 5	-0.29%	-0.64	0.53	20	45 %
	(-1) to 1	-0.67%	-1.44	0.17	20	40 %
	0	-0.73%	-3.33	0.004***	20	25 %
	0 to 2	-1.05%	-2.12	0.05**	20	40 %
	0 to 10	-0.99%	-1.24	0.23	20	45 %
25 % companies with highest revenue	(-25) to + 25	0.13%	0.49	0.63	31	52 %
	(-15) to 15	-0.14%	0.14	0.89	31	52 %
	(-10) to 0	0.09%	0.57	0.57	33	42 %
	(-5) to 5	0.34%	0.28	0.78	33	52 %
	(-1) to 1	0.3%	0.89	0.38	33	52 %
	0	-0.07%	0.27	0.79	33	55 %
	0 to 2	0.08%	0.26	0.8	33	61 %
	0 to 10	-0.36%	-0.48	0.63	33	45 %
25 % companies with lowest revenue	(-25) to + 25	-1.52%	0.21	0.83	38	34 %
	(-15) to 15	-3.07%	-1.3	0.2	38	34 %
	(-10) to 0	-2.11%	-1.39	0.17	38	37 %
	(-5) to 5	-1.99%	-1.98	0.06*	38	37 %
	(-1) to 1	-0.97%	-0.89	0.38	38	34 %
	0	-0.82%	-0.56	0.58	38	45 %
	0 to 2	-1.11%	-1.06	0.3	38	45 %
	0 to 10	-2.73%	-2.56	0.14**	38	34 %

6.2 Abnormal sentiment results

Event study analysis on corporate sentiment is performed on the base sample and on five sub samples using sentimentr, a table showing the different samples and event windows can be found in table A10.1 in the appendix. The figure under shows the Cumulative abnormal average mood of the entire sample. From the graph we see that the CAABMOOD is stable and hovering around 0% from 15 days to the day before the decision by the European Commission. This indicates that, on average, there is no significant information leakage that gains traction on Twitter. On the day of the decision there is a large drop of -91% which means that the sentiment is almost twice as bad as the company's average normal sentiment calculated from the estimation period. The following three days are also very negative. Over the event window [0,2] the CAABMOOD is -200% and it is -325% in the [0,10] event window. This means that most of the negative reaction comes within the first few days after the decision. Not all companies are experiencing a negative CABMOOD as a result of the decision, 23% of companies have a positive CABMOOD on the day of the decision and this percentage is increasing to 28% in the [0,10] window.

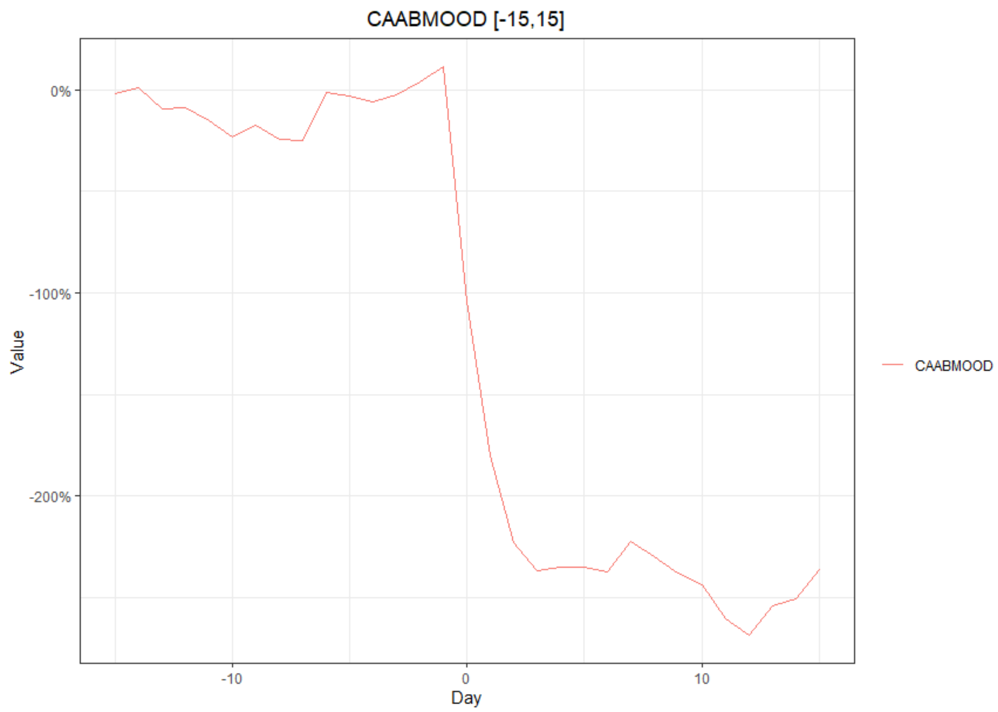

Figure 6.8: CAABMOOD of sentiment all companies

In figure 6.9, we have grouped the companies into those who received fines and those who received immunity. From the graph we see that both groups are yielding a negative

CAABMOOD over the long event window [-15,15]. It is interesting that the immunity sample are experiencing such large losses in sentiment during the event window. The sample experience a drop in sentiment close to -300% in the 15 days prior to the event, but only -56% on the day of the event. The companies that receive fines are behaving in a more logical way. The drop in mood is mainly happening on the day of the event, and then stabilize after a few days. We see it as probable that the long slow decline in mood seen in the companies that receive immunity is due to the fact that all companies are weighted equally and that there is a small sample pool.

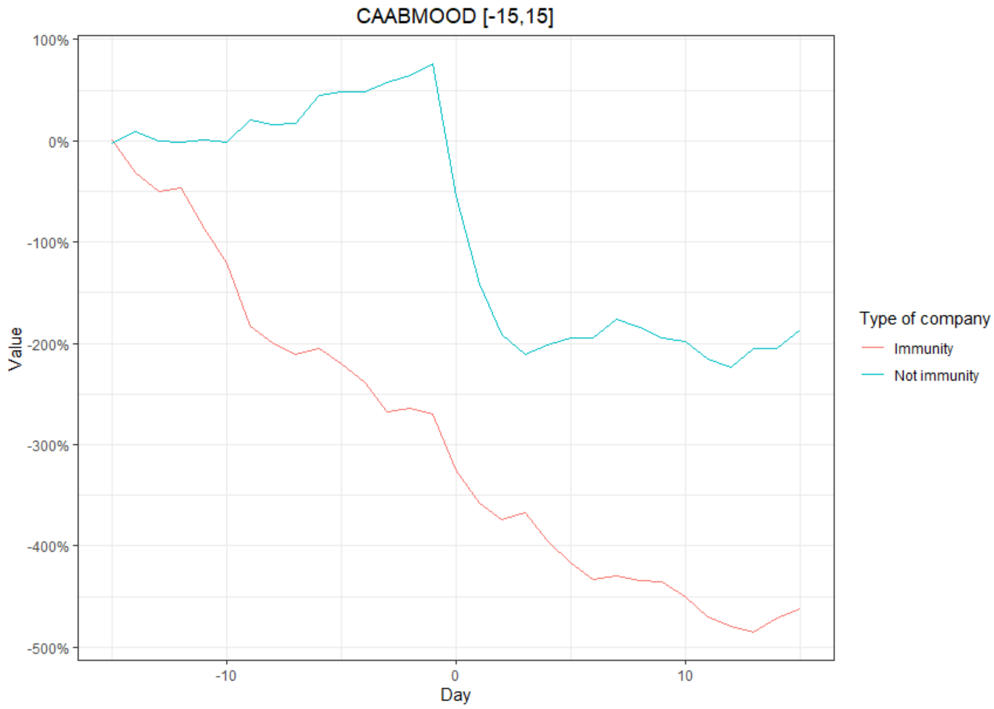

Figure 6.9: CAABMOOD of sentiment grouped by fine and immunity

Figure 6.10 shows the CAABMOOD for 3 subsamples; companies from North America, Europe and Asia. All three samples show a significant fall on and in the days after the event day, with North American companies having the most decline in mood. This is different from the abnormal stock returns, where North American companies showed neutral to positive CAAR in the event windows.

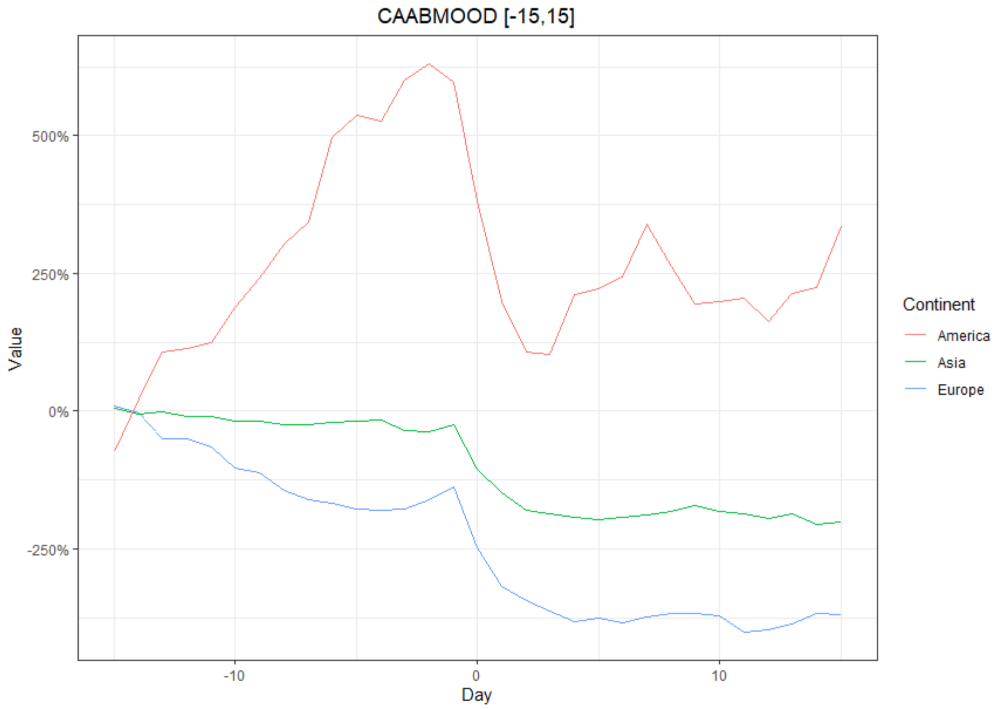

Figure 6.10: CAABMOOD of sentiment categorised by continent

Table 6.2 shows that the days with significantly abnormal mood is the event day and the upcoming days. This indicates that most of the event windows seen in table A.10.1 are significant because of these days. To conclude, most companies seem to experience abnormal negative mood on the days of the event. Immune companies are also affected negatively because of the event, though it is more challenging to explain their reaction. The relationship between abnormal mood and abnormal returns, among other variables, is explored in the next subchapter.

Day	AABMOOD	T-value	P-value	Observations
-15	19%	1.26	0.21	100
-14	-6%	-0.93	0.35	100
-13	-2%	-0.65	0.52	100
-12	13%	0.66	0.51	101
-11	2%	-0.39	0.7	100
-10	7%	0.13	0.9	95
-9	-5%	-0.84	0.4	100
-8	-8%	-1.24	0.22	103
-7	-5%	-0.61	0.54	101
-6	14%	0.68	0.5	102
-5	-5%	-0.44	0.66	100
-4	6%	0.19	0.85	99
-3	-19%	-1.49	0.14	92
-2	-4%	-0.31	0.76	102
-1	18%	1.3	0.2	103
0	-91%	-7.42	0.00001***	107
1	-65%	-6.02	0.00001***	106
2	-42%	-4.43	0.00002***	106
3	-27%	-2.68	0.009***	99
4	-19%	-1.45	0.15	96
5	-28%	-2.46	0.016**	97
6	-37%	-2.54	0.013**	97
7	-7%	-0.48	0.63	98
8	-6%	-0.56	0.58	97
9	-8%	-0.98	0.33	96
10	-1%	-0.53	0.6	90
11	-14%	-1.06	0.29	96
12	-12%	-1.28	0.2	95
13	-2%	0.28	0.78	100
14	-7%	-1.13	0.26	98
15	-1%	-0.62	0.54	97

Table 6.2: Daily Average abnormal mood

6.3 Regression results

Cross-sectional single-factor regression analysis is used to test for determinants of abnormal stock price for different event windows. The null hypothesis is that there is no linear relationship between the outcome variable and the variables that are tested. T-statistics are used to decide whether the null hypothesis can be rejected or not.

Table 6.3 shows a description of all the variables that are tested. They can be categorized into five different categories; Twitter data, fine characteristics, company size, business sector and country of origin. The first 5 variables are those which tries to find out if Twitter data has a significant impact the stock return on the event day.

Variable	**Description**
Abnormal sentiment	Abnormal sentiment score in same event window as outcome variable
25% percentile mood	The 25% of companies with most negative abnormal sentiment (binary)
75% percentile mood	The 25% of companies with most positive abnormal sentiment (binary)
Sentiment score cartel tweets	Sentiment score on only tweets including the word "cartel"
Count cartel tweets	Number of tweets including the word "cartel"
European	Company is located in Europe (binary)
Asian	Company is located in Asia (binary)
American	Company is located in Europe (binary)
Fine %	Fine as percentage of yearly revenue
Fine over 1%	The fine is greater than 1% of yearly revenue (binary)
Fine over 3%	The fine is greater than 3% of yearly revenue (binary)
log(Revenue)	Logaritm of yearly revenue
Industrial	The company is classified as industrial according to Thomson Reuters (binary) economic sector classification
Consumer cyclicals	The company is classified as consumer cyclicals according to Thomson Reuters economic sector classification (binary)
Financial	The company is classified as financial according to Thomson Reuters economic sector classification (binary)
Technology	The company is classified as technology according to Thomson Reuters economic sector classification (binary)
Number of years cartel	Number of years from cartel started to the cartel ended
Price fixing	Cartel characteristics (binary)
Market share allocation	Cartel characteristics (binary)
Bid rigging	Cartel characteristics (binary)
Quota	Cartel characteristics (binary)

Table 6.3: Variable description

The coefficients of the different regressions are presented in table 6.4, with intercept and R squared for significant variables in tables 6.5 and 6.6. All coefficients are presented and statistically significant variables are labeled with stars at 10%, 5% and 1% significance level. The regression shown in this chapter was done on companies which received a fine, as this is the most interesting group because of the significant fall in abnormal stock returns around the event date. The same table for immune companies can be seen in the appendix table A8.1, with close to none significant results.

The regressions confirm that fines normalised by revenue can explain some of the drop in returns. Larger fines in percentage of revenue leads to larger negative abnormal return. A fine of at least 1% or 3% are correlated with more negative abnormal return of 0.8% and

2.1% respectively on the event day, and are also significant for other event windows.

Only one of the Twitter data variables seems to have any impact on the abnormal stock return around the event day for any of the periods. The 25% percentile of companies with most positive abnormal return does 4.8% better than the rest on average, significant at a 5% level.

The only other variable to have significant impact on the abnormal stock return is the revenue. Higher revenue seems to correlate with better abnormal return, indicating that smaller companies are punished harder than large companies.

The full regression output for fine over revenue, abnormal sentiment, tweets including "cartel" and sentiment score on those tweets can be found in the appendix for event window [0,10], [-1,1] and on the event day. The only significant factor, fine over revenue, explains 5,6%, 7.7% and 10% of the outcome in the three windows, indicating that the fine is most effective on the day that it is announced.

Coefficient	[-15,15]	[-10,0]	[-5,5]	[-1,1]	0	[0,2]	[0,10]
Abnormal sentiment	0.001	-0.0004	0.001	0	-0.001	-0.001	0.001
25% percentile mood	-0.015	-0.001	0.001	0.003	0.006	0.002	0.01
75% percentile mood	0.048**	0.021	0.01	-0.001	-0.002	-0.005	0.007
Sentiment score cartel tweets	-0,007	-0,001	0,017	-0,025	-0,025	-0,024	-0,025
Count cartel tweets	0	0	0	0	0	0	0
European	0,011	0,014	0,002	0,001	0,001	-0.01*	0
Asian	-0,019	-0,014	0	-0,007	-0,005	0,002	-0,01
American	0,02	0,002	-0,006	0,008	0,009	0,012	0.022*
Fine %	-0,52	-0.431**	-0.321*	-0.371***	-0.372***	-0.337***	-0.462***
Fine over 1%	-0,016	-0,007	-0.017**	-0.018***	-0.008*	-0.017***	-0.02**
Fine over 3%	-0,017	-0,018	-0.025**	-0.025***	-0.021***	-0.025***	-0.023*
log(Revenue)	0.008*	0.005*	0.006**	0.003**	0,002	0.004***	0.007***
Industrial	-0.044***	-0,004	-0,004	-0,007	0,002	-0,003	-0,003
Consumer cyclicals	0,005	0,003	0,003	-0,008	-0,011	-0,009	-0,007
Financial	0	0,011	0,006	0,012	0,007	0,004	0,013
Technology	0	-0.025**	-0.006	-0.007	-0.007	0.003	-0.015
Number of years cartel	0	0,0001	0,001	0	0	0	0
Price fixing	-0,006	-0.028**	-0,003	-0,009	-0,009	-0,006	0,005
Market share allocation	-0,002	0,001	0,005	0,004	0,002	0,004	0,004
Bid rigging	0.13*	0.086***	0,021	0,007	0,011	0,001	0,022
Quota	-0,026	-0,009	0,005	0,003	0,001	0,003	-0,006

Table 6.4: Single-factor regression on companies that received fine

Intercept	[-15,15]	[-10,0]	[-5,5]	[-1,1]	0	[0,2]	[0,10]
75% percentile mood	-0.038***						
European						-0,003	
American							-0.02*
Fine %		-0,005	-0,007	0,0003	0,001	-0,002	-0.01**
Fine over 1%			-0,005	0,001	-0,001	-0,001	-0.009*
Fine over 3%			-0.008*	-0,002	-0,001	-0,004	-0.014***
log(Revenue)		-0.13**	-0.143**	-0.082**		-0.107***	-0.186***
Industrial	-0,01						
Technology		-0,006					
Price Fixing		0,013					
Bid rigging	-0.028***	-0.014***					

Table 6.5: Intercept for significant variables

	[-15,15]	[-10,0]	[-5,5]	[-1,1]	0	[0,2]	[0,10]
75% percentile mood	0.062						
European						0.026	
American							0.024
Fine %		0.039	0.026	0.077	0.105	0.069	0.056
Fine over 1%			0.031	0.07	0.022	0.069	0.042
Fine over 3%			0.033	0.072	0.066	0.077	0.028
log(Revenue)	0.025	0.027	0.042	0.031		0.057	0.071
Industrial	0.053						
Technology		0.038					
Price Fixing		0.039					
Bid rigging	0.085	0.099					

Table 6.6: R squared for significant variables

6.4 Regression tree results

To dig further into the importance and interactions of the variables, regression trees has been created. Our main variables have been used in the regression trees; abnormal mood, sentiment score on cartel tweets, number of cartel tweets, fine, continent and economic sector. Figure 6.11 and 6.12 shows trees for the event windows [-1,1] and [0,10], for the companies that received a fine. Regression trees for companies with fines, and regression trees for other event windows, can be found in the appendix.

The most significant predictor in both regression trees is the fine as percentage of revenue. When the fine is at least 1.7% of the annual revenue in the event window, the abnormal return is on average negative by 4% if the company is categorised as either Consumer cyclicals or technology, elsewise it is negative by 1.1%. If the fine is less than 1.7%, then further inequalities from cartel sentiment, economic sector and geographical placement has to be explored to find the expected abnormal return. The relative variable importance in this tree is 48% fine, 25% economic sector, 16% sentiment on cartel tweets, 8% geographical placement, 2% number of tweets including cartel and 1% abnormal sentiment.

For the other event window, [0,10], a fine over 4.7% of annual revenue is correlated with a negative abnormal stock return of 6.4%, while companies with smaller fines has to go through further branches. The relative variable importance in this tree is 47% fine, 20% economic sector, 14% abnormal sentiment, 9% geographical placement and 3% sentiment on cartel tweets.

This tree shows once again that the fine is more important than the other predictors when it comes to explaining the abnormal returns around the event day.

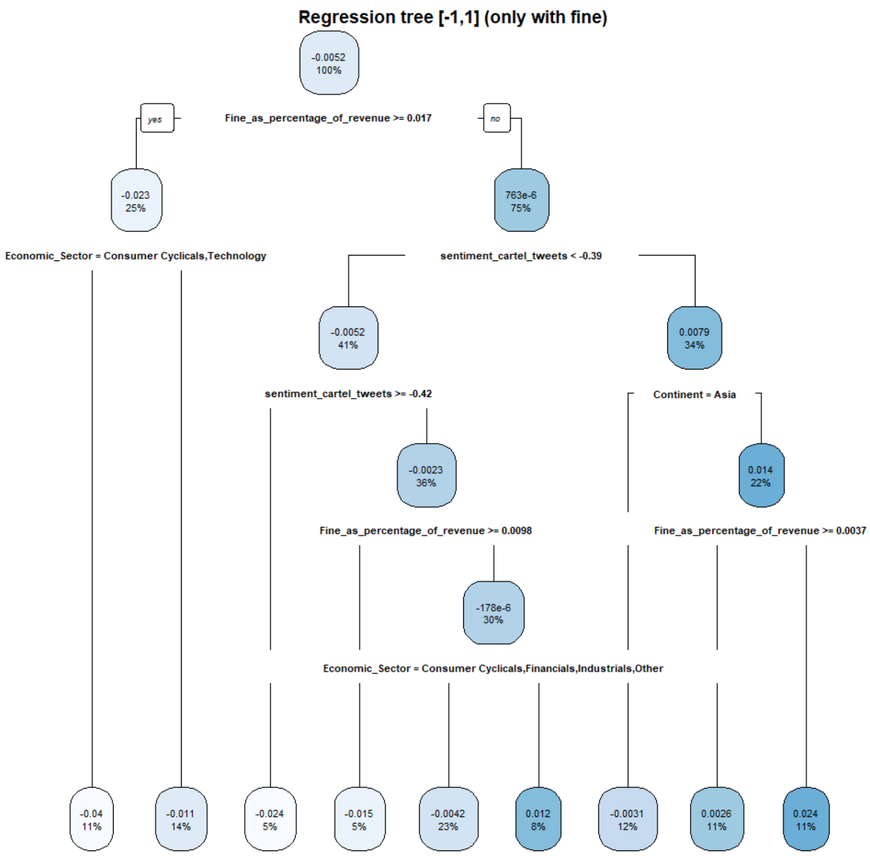

Figure 6.11: Regression tree [-1,1]

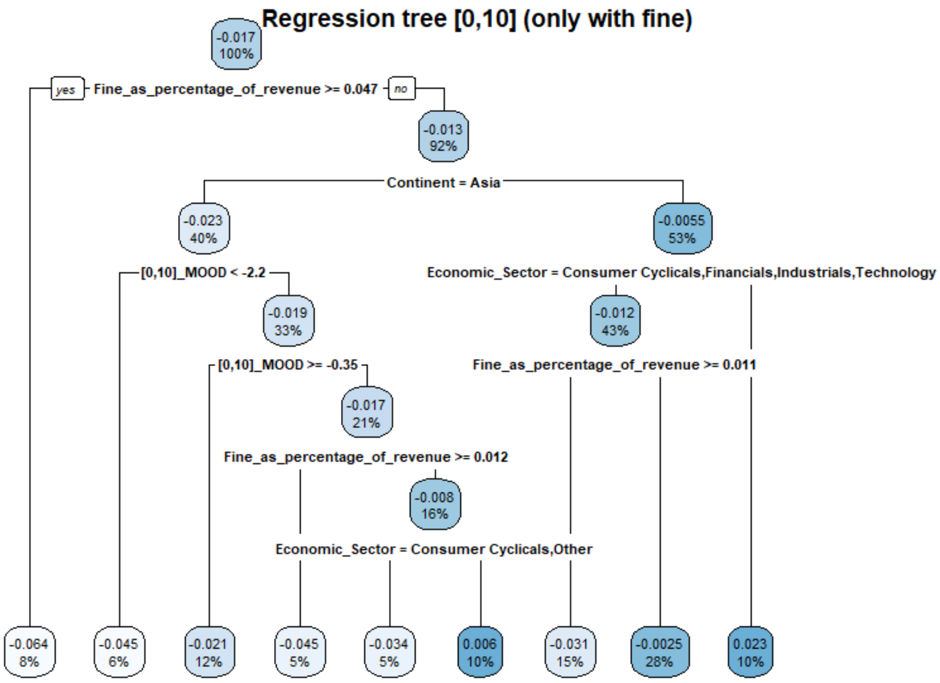

Figure 6.12: Regression tree [0,10]

6.5 Robustness checks

The first thing we wanted to test for in our robustness check was our dictionary. This was done to see if the result would change if we tried another method. The VADER package was used for this, as described in the methodology section.

Figure 5.12 shows the CAABMOOD over the same event window, [-15,15] as used by the sentimentr package. It looks very different from the results of the sentimentr method, and shows that normal words don't capture the real meaning of the tweets that are being tweeted. The plot for the VADER AABMOOD with 95% confidence interval each day can be found in the appendix.

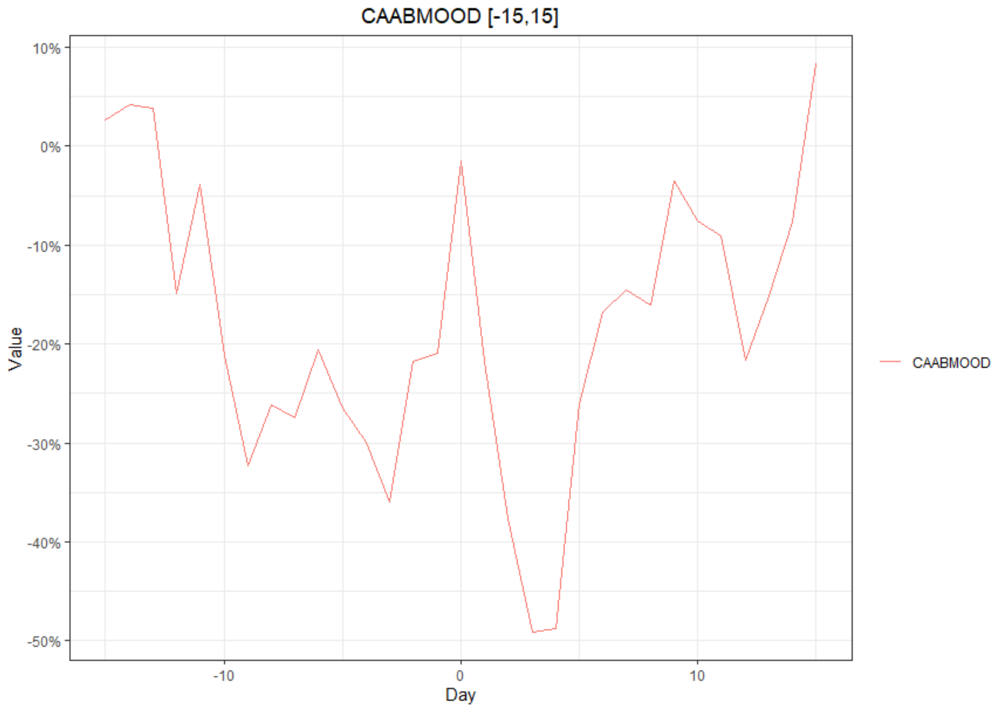

Figure 6.13: CAAR of sentiment all companies

The interpretation of the tweets containing the word cartel could be one of the reasons why the two methods have so different results. As mentioned in the methodology chapter, the sentimentr algorithm with its custom dictionary will to a larger extent indicate what a person that understands economical terminology thinks when he is reading a tweet, while the VADER sentiment shows what a "normal" person thinks. The VADER dictionary have a much lower average sentiment score in the tweets containing the word cartel. In sentimentr, the average score of these tweets during the event day is -0.36, while with the alternative VADER corpus the mean score is 0.14.

Table 5.13 shows the results when regressing abnormal sentiment with VADER on abnormal stock return. It is not possible to reject the null hypothesis about zero correlation between abnormal Twitter mood/sentiment and abnormal stock returns.

	Dependent variable:					
	Abnormal stock return event day		aret3day		tendayafter	
	(1)	(2)	(3)	(4)	(5)	(6)
Abnormal sentiment	-0.002 (-0.006, 0.002)					
Sentiment score on cartel tweets		-0.002 (-0.045, 0.041)		0.033 (-0.008, 0.074)		-0.017 (-0.085, 0.052)
abnormal sentiment [-1,1]			-0.001 (-0.003, 0.002)			
abnormal sentiment [0,10]					0.001 (-0.001, 0.002)	
Constant	-0.004 (-0.010, 0.001)	-0.004 (-0.012, 0.004)	-0.005* (-0.011, 0.0002)	-0.008** (-0.016, -0.0002)	-0.015*** (-0.025, -0.006)	-0.009 (-0.022, 0.004)
Observations	96	83	96	83	93	80
R2	0.008	0.0001	0.003	0.030	0.013	0.003
Adjusted R2	-0.003	-0.012	-0.008	0.018	0.002	-0.010
Residual Std. Error	0.028 (df = 94)	0.030 (df = 81)	0.028 (df = 94)	0.029 (df = 81)	0.046 (df = 91)	0.047 (df = 78)
F Statistic	0.746 (df = 1; 94)	0.008 (df = 1; 81)	0.260 (df = 1; 94)	2.480 (df = 1; 81)	1.166 (df = 1; 91)	0.226 (df = 1; 78)
Note:						*p<0.1; **p<0.05; ***p<0.01

Figure 6.14: CAAR of sentiment all companies

The Twitter data has limitations as mentioned in the data limitation section. Companies with few tweets could affect the result in one way or another, and changes in how Twitter have been used over the years could also affect the outcome.

In table 3,4 and 5 in the appendix we regress the three main Twitter variables on abnormal stock return with three different data samples; cases before 08.11.2017, cases after 08.11.2017 and companies with at least 10 000 tweets. As the tables show, most of the variables are still insignificant with only the sentiment score on cartel tweets being significant in some of the event windows. The coefficients of the variables in these windows still indicates that more negative tweets correlates with more positive abnormal stock return, which is the opposite of what we would expect, we interpret this as being a random sampling issue.

7 Discussion

We begin the discussion by repeating our first research question; does cartel convictions create abnormal returns? Based on the results of the analysis, there seems to be some evidence suggesting that cartel convictions create negative cumulative abnormal returns around the event day. However, this is conditional on several factors like the size of the fine as a percent of revenue, immunity, company size, and geographic location. Only the event window [0,10] is significantly negative when looking at the primary sample containing all companies. Thus, it is difficult to falsify a null hypothesis of zero cumulative average abnormal returns for the main sample containing all convicted companies. This does not mean that investors are not punishing companies for cartel participation, as the results differ if the fined companies are isolated.

The reason why the immune companies don't have negative abnormal returns is probably due to an anticipation effect. Markets may have anticipated a fine and priced in the expectations of the penalty in advance of the decision. The mildly positive reaction can thus be a reaction to the lack of a fine by the EU Commission, even though it is not significant. The fact that it, on average, is no negative price reaction for companies that receive immunity should work as an additional incentive for companies to apply for an immunity application, as they benefit from both no fine and no reduction in stock returns.

For the companies that receive a fine from the Commission, the results are significant and negative in multiple event windows. The regressions and the decision trees confirm that receiving a fine affect company returns negatively. This makes sense as a fine directly affects a company's earnings and reduces the company's total market value. However, the fine amount does only explain 10.5% of the variation seen in the cumulative abnormal returns in the sample seen in table A8.1. The remaining differences in returns have in previous studies been explained by the loss of extra cartel revenue and an assumed reputational factor. This leads us to the second research question, can Twitter data be used to estimate private sanctions?

Based on the analysis results, it seems clear that the variables created to measure private sanctions cannot explain the differences in abnormal stock returns between companies. The question is, therefore, why do they not? If one assumes a connection between reputational damage and public sentiment, one would expect that a company that receive more negative coverage on Twitter relative to other companies, would have worse stock returns. The results of the analysis don't show any such relation. As we see it, there are three possible reasons why the Twitter variables do not provide any explanatory power.

1. Data issues
2. Method issues
3. It works and there is no relation between Twitter sentiment and abnormal returns

The first explanation have already been touched upon in the data limitation section 3.5. Better query words and a larger sample of companies and tweets might have given us better data that could have affected the results. However the robustness checks show that the conclusion doesn't change when the assumptions changes, as the variables have no more explanatory power when tried on smaller and more specified samples. This leads us to believe that the data issue is not that relevant. It is, therefore, more likely that improvements to the sentiment analysis could affect the results. It might be that the algorithms over or underestimate the real sentiment of the tweets creating inaccurate sentiment scores for each company, which would affect the explanatory power of the variables. We tried with two different methods and they yielded very different results. This was not surprising considering that we tested whether "a layperson" or more financial aware people have the greater effect on returns. The results show that neither of them affect the market, but this could also be due to the algorithm miss-classifying tweets.

The most interesting explanation is that their is no relation between Twitter mood and abnormal returns. The variables only containing the cartel specific tweet sentiment, and the variable of the count of cartel specific tweets is the closest this thesis comes to testing variables similar to those used in earlier studies. Previous studies have found that the sentiment and the count of articles from the combined news media can influence the performance of stocks (Mariuzzo et al., 2020a) (Günster & van Dijk, 2016). That the Twitter variables have no explanatory power is, therefore, a bit surprising. There should be a connection between what is written by the traditional news media and what is written on Twitter (Gan et al., 2020). Increased newspaper coverage should generate more tweets, and the language used in the articles should also affect the opinions expressed on Twitter, the inverse should also apply. Because of this one would think that Twitter sentiment and the count of tweets would be as good if not even better at measuring private sanctions, as it contain the collective thoughts of large parts of the public.

The fact that the variables don't have any explanatory power could mean that Twitter is a less ideal source for measuring private sanctions. It would mean that while allot is written on Twitter, it has small to none effect on the share price of companies, at least related to cartel events. Even though some researchers have found the sentiment on social media to have predictive power(Wolf & Bergdorf, 2019), some studies argue that market/stock returns and volatility is exerting a stronger impact on investor sentiment than the other way around (Gan et al., 2020). The relationship between positive/negative Twitter sentiment and real-world outcomes is thus not fixed (Lim & Tucker, 2019). Some events generate a lot of negative attention on Twitter, and the stocks still perform better than the stocks of similar companies going through similar events with less Twitter attention. The reason why Twitter sentiment might not affect stock prices could potentially be due to the differences in the demographic on social media and in the stock market. While more than 50% of Twitter users are under the age of 35 (Statista, 2021), only 1.4% of the total stock value at least in the US is owned by the same age group(Federal Reserve, 2019) (USA facts, 2019). This could mean that even a considerable outrage on Twitter would not affect the returns, as the people who dictate the market might not be paying that much awareness to what is happening on the platform. It might also be that investors care more about the potential loss of revenue as a consequence of the cease of cartel participation, than they do about the loss in revenue because of a reputational blow, but this is just speculation.

Before we end the discussion, we want to reflect on some of the other variables that we tested in the analysis. Firstly, geographic location seems to matter for the abnormal returns of the companies. This is partly explained by differences in media coverage by previous studies (Ulrich, 2018). We agree that media coverage might be a possible explanation, as it is reasonable to think that news decrease in force over longer distances. However, it is interesting that North American companies are experiencing the most abnormal mood and coverage on Twitter on average (according to our sentimentr evaluation), when they have no abnormal returns. This could mean that other media types have more power to influence stock returns in cartel specific cases, but it could also mean that media coverage, in general, is less important than previously assumed by earlier studies. Identifying which media type that imposes the most reputational damage on a firms value could be a proposed future research question. We also see that small companies are more hurt by the conviction than large companies. This is probably due to investors thinking that the chance of bankruptcy is higher for this group. Thus, being part of a cartel provides an additional downside for small companies when caught, as they in general get penalised more by investors.

8 Conclusion

In this thesis, we have investigated the effect of cartel convictions in the European Union on stock returns. A common event study methodology has been conducted on a sample of 164 companies involved in 39 cartel cases from 2010 to 2021. The aim of the study was to find out if the decision by the European Commission results in abnormal stock returns and to identify whether variables created from Twitter data could explain differences in the abnormal returns between companies. In order to do this, we collected tweets on all companies from 70 days before to 30 days after the decision by the European Commission, which we used to create three different variables. Based on our methods, data, and analysis, we have found no proof that differences in abnormal Twitter mood, the count of cartel specific tweets, or the sentiment in these tweets can explain the differences in abnormal returns.

In addition to trying the new Twitter variables, we have also validated the findings of previous studies. We found the maximum cumulative average abnormal return for companies that received a fine to be -2.29% over a period of 15 days before to 15 days after the penalization, this suggests pre-event information leakage. By analysing subsamples over multiple event windows, we have confirmed the findings of earlier studies and agree with them that variables such as fine as a percent of revenue, firm size, and to some degree the country of incorporation and economic sector matter for the magnitude of abnormal return. Cross sectional regression and decision trees on different event windows further confirm the significance of these variables. We also find that companies that receive immunity from the European Commission on average are not penalised with lower stock returns by investors.

References

Aguzzoni, L., Langus, G., & Motta, M. (2013). The effect of eu antitrust investigations and fines on a firm's valuation. *The Journal of Industrial Economics*, *61*(2), 290–338. Retrieved from http://www.jstor.org/stable/43305471

Alex, S. (2019). When (not) to lemmatize or remove stop words in text preprocessing [Computer software manual]. ModelOp. Retrieved from https://opendatagroup.github.io/data%20science/2019/03/21/preprocessing-text.html

Alexander, C. (1999). On the nature of the reputational penalty for corporate crime: Evidence. *The Journal of Law and Economics*, *42*, 489-526.

Boehmer, E., Masumeci, J., & Poulsen, A. B. (1991). Event-study methodology under conditions of event-induced variance. *Journal of Financial Economics*, *30*(2), 253-272. Retrieved from https://www.sciencedirect.com/science/article/pii/0304405X9190032F doi: https://doi.org/10.1016/0304-405X(91)90032-F

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1-8. Retrieved from https://www.sciencedirect.com/science/article/pii/S187775031100007X doi: https://doi.org/10.1016/j.jocs.2010.12.007

Bosch, J.-C., & Eckard, E. W. (1991). The profitability of price fixing: Evidence from stock market reaction to federal indictments. *The Review of Economics and Statistics*, *73*(2), 309–317. Retrieved from http://www.jstor.org/stable/2109522

EEC. (1957). *Treaty of rome.* (https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=LEGISSUM%3Axy0023)

European Commission. (1996). Commission Notice on the non-imposition or reduction of fines in cartel cases ,No C 207/4 (39), pp. 4-6. *Official Journal of the European Communities*.

European Commission. (2011). Fines for breaking EU Competition Law. Retrieved from http://ec.europa.eu/competition/cartels/overview/factsheet_fines_en.pdf

European Commission. (2017a). Antitrust: Commission introduces new anonymous whistleblower tool. *Press Release IP/17/591, 2017-03-16, Brussels.*.

European Commission. (2017b). Antitrust - Procedures in anticompetitive agreements (Article 101 TFEU cases). Retrieved from http://ec.europa.eu/competition/antitrust/procedures_101_en.html

European Commission. (2021). Competition cases. Retrieved from https://ec.europa.eu/competition/elojade/isef/

European Union. (2008). *Treaty on the functioning of the european union article 101.* (https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A12008E101)

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383–417. Retrieved from http://www.jstor.org/stable/2325486

Federal Reserve. (2019). *Survey of consumer finances (scf).* (https://www.federalreserve.gov/econres/scfindex.htm)

Gan, B., Alexeev, V., Bird, R., & Yeung, D. (2020). Sensitivity to sentiment: News vs social media. *International Review of Financial Analysis*, *67*, 101390. Retrieved from https://www.sciencedirect.com/science/article/pii/S105752191930273X doi: https://doi.org/10.1016/j.irfa.2019.101390

Günster, A., & van Dijk, M. (2016). The impact of european antitrust policy: Evidence from the stock market. *International Review of Law and Economics*, *46*, 20-33. Retrieved from https://www.sciencedirect.com/science/article/pii/S0144818815000721 doi: https://doi.org/10.1016/j.irle.2015.12.001

Hutto, C. (2014b). Vader-sentiment-analysis [Computer software manual].

Hutto, C., & Gilbert, E. (2015, 01). Vader: A parsimonious rule-based model for sentiment analysis of social media text..

Internetlivestats. (2021). *Twitter usage statistics.* (https://www.internetlivestats.com/twitter-statistics/)

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2014). *An introduction to statistical learning: With applications in r.* Springer Publishing Company, Incorporated.

Jarrell, G., & Peltzman, S. (1985). The impact of product recalls on the wealth of sellers. *Journal of Political Economy*, *93*(3), 512–536. Retrieved from http://www.jstor.org/stable/1832006

Jockers, M. L. (2015). Syuzhet: Extract sentiment and plot arcs from text [Computer software manual]. Retrieved from https://github.com/mjockers/syuzhet

Karpoff, J. M., & Lott, J. R. (1993). The reputational penalty firms bear from committing criminal fraud. *The Journal of Law Economics*, *36*(2), 757–802. Retrieved from http://www.jstor.org/stable/725807

Kearney, C., & Liu, S. (2014). Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis*, *33*, 171-185. Retrieved from https://www.sciencedirect.com/science/article/pii/S1057521914000295 doi: https://doi.org/10.1016/j.irfa.2014.02.006

Laina, F., & Laurinen, E. (2013). The eu cartel settlement procedure: Current status and challenges. *Journal of European Competition Law Practice*, *4*, 302-311.

Lim, S., & Tucker, C. S. (2019). Mining twitter data for causal links between tweets and real-world outcomes. *Expert Systems with Applications: X*, *3*, 100007. Retrieved from https://www.sciencedirect.com/science/article/pii/S2590188519300071 doi: https://doi.org/10.1016/j.eswax.2019.100007

Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In C. C. Aggarwal & C. Zhai (Eds.), *Mining text data* (pp. 415–463). Boston, MA: Springer US. Retrieved from https://doi.org/10.1007/978-1-4614-3223-4_13 doi: 10.1007/978-1-4614-3223-4_13

MacKinlay, A. C. (1997). Event studies in economics and finance. *Journal of Economic Literature*, *35*(1), 13–39. Retrieved from http://www.jstor.org/stable/2729691

Mariuzzo, F., Ormosi, P. L., & Majied, Z. (2020a). Fines and reputational sanctions: The case of cartels. *International Journal of Industrial Organization*, *69*(C). Retrieved from `https://ideas.repec.org/a/eee/indorg/v69y2020ics0167718720300060.html` doi: 10.1016/j.ijindorg.2020.1

Mariuzzo, F., Ormosi, P. L., & Majied, Z. (2020b). private and public sanctions [Computer software manual]. (`https://github.com/PeterOrmosi/private_v_public_sanctions`)

Mooney, R., Brew, C., Chien, L.-F., & Kirchhoff, K. (Eds.). (2005, October). *Proceedings of human language technology conference and conference on empirical methods in natural language processing.* Vancouver, British Columbia, Canada: Association for Computational Linguistics. Retrieved from `https://www.aclweb.org/anthology/H05-1000`

Publications Office of the European Union. (2015). *The Treaty of Lisbon: introduction.* Retrieved from `https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=LEGISSUM:ai0033&from=EN`

Ranco Gabriele, C. G. G. M. M. I., Aleksovski Darko. (2015). The effects of twitter sentiment on stock price returns. doi: https://doi.org/10.1371/journal.pone.0138441

Rinker, T. W. (2019). sentimentr: Calculate text polarity sentiment [Computer software manual]. Buffalo, New York. Retrieved from `http://github.com/trinker/sentimentr` (version 2.7.1)

Statista. (2021). *Distribution of twitter users worldwide as of january 2021, by age group.* (`https://www.statista.com/statistics/283119/age-distribution-of-global-twitter-users/`)

Tumasjan, A., Sprenger, T., Sandner, P., & Welpe, I. (n.d.). In *Proceedings of the fourth international aaai conference on weblogs and social media.*

Twitter. (2021a). Enabling the future of academic research with the Twitter API. Retrieved from `https://blog.twitter.com/developer/en_us/topics/tools/2021/enabling-the-future-of-academic-research-with-the-twitter-api.html`

Twitter. (2021b). *Q1 2021 letter to shareholders.* (`https://s22.q4cdn.com/826641620/files/doc_financials/2021/q1/Q1'21-Shareholder-Letter.pdf`)

Twitter. (2021c). Rate limits. Retrieved from `https://developer.twitter.com/en/docs/twitter-api/rate-limits`

Ulrich, F. (2018). *The effects of eu cartel fines on firm performance, dividend policy and management compensation* (Doctoral dissertation). doi: 10.13140/RG.2.2.12187.31526

USA facts. (2019). *What percentage of americans own stock?* (`https://usafacts.org/articles/what-percentage-of-americans-own-stock/`)

Wolf, F., & Bergdorf, O. (2019). *Twitter sentiment and stock returns.* (Student Paper)

Wooldridge, J. M. (2013). Introductory econometrics : A Modern Approach, Mason, OH. *South-Western Cengage Learning*, *5th.*

Appendix

A1 Cartel characteristics

Table A1.1: Cartel characteristics

Case	Cartel name	Size	Quota	Market share allocation	Price fixing	Bid rigging	Cartel started	Cartel ended	Duration years	Date convicted
38344	Pre-stressing steel	17	1	1	1	0	1984	2002	18	30/06/2010
38511	DRAMS	10	1	0	1	0	1998	2002	4	19/05/2010
38866	Animal Feed Phosphates	6	0	1	1	0	1969	2004	35	20/07/2010
39092	Bathroom fittings & fixtures	17	0	0	1	0	1992	2004	12	23/06/2010
39258	Airfreight	14	0	0	1	0	1999	2006	7	09/11/2010
39309	LCD	6	0	0	1	0	2001	2006	5	08/12/2010
39437	TV and computer monitor tubes	8	0	1	1	0	1996	2006	10	05/12/2012
39462	Freight Forwarding	15	0	0	1	0	2002	2007	5	28/03/2012
39574	Smart card chips	4	0	1	1	0	2003	2005	2	03/09/2014
39579	Consumer detergents	3	0	0	1	0	2002	2005	3	11/04/2011
39600	Refrigeration compressors	5	0	1	1	0	2004	2007	3	07/12/2011
39605	CRT glass bulbs	4	0	0	1	0	1999	2004	5	19/10/2011
39610	Power cables	26	1	1	0	0	1999	2009	10	02/04/2014
39639	Optical disc drives	8	0	0	0	1	2004	2008	4	21/10/2015
39748	Automotive wire harnesses	5	1	0	1	0	2000	2009	9	10/07/2013
39801	Polyurethane foam	5	0	0	1	0	2005	2010	5	29/01/2014
39824	Trucks	6	0	0	1	0	1997	2011	14	19/07/2016
39861a	Yen interest rate derivatives	1	0	0	1	0	2007	2010	3	04/12/2015
39861	Yen interest rate derivatives	6	0	0	1	0	2007	2010	3	04/12/2013
39881	Occupant safety systems	5	0	1	1	0	2004	2010	6	22/11/2017
39904	Rechargeable batteries	4	0	0	1	0	2004	2007	3	12/12/2016
39914a	Euro interest rate derivatives	3	0	0	1	0	2005	2008	3	07/12/2016
39914	Euro interest rate derivatives	5	0	0	1	0	2005	2008	3	04/12/2013
39920	Braking systems	3	0	0	1	0	2007	2011	4	21/02/2018
39922	Automotive bearings	6	1	0	1	0	2004	2011	7	19/03/2014
39924	Swiss franc interest rate derivatives	4	0	0	1	0	2007	2007	1	21/10/2014
39960	Thermal systems	6	0	1	1	0	2005	2009	4	08/03/2017
40009	Maritime car carriers	5	1	0	1	0	2006	2012	6	21/02/2018
40013	Lighting systems	3	0	0	1	0	2004	2007	3	21/06/2017
40018	Car battery recycling	5	0	0	1	0	2009	2012	3	08/02/2017
40028	Alternators and starters	3	1	0	1	0	2004	2010	6	27/01/2016
40098	Blocktrains	3	0	1	1	0	2004	2012	8	15/07/2015
40113	Spark plugs	3	1	0	1	0	2001	2011	10	21/02/2018
40135	Foreign exchange spot trading	6	0	0	1	0	2007	2013	6	16/05/2019
40136	Capacitors	9	1	0	1	0	1998	2012	14	20/03/2018
40299	Closure systems	3	0	1	1	0	2009	2012	3	29/09/2020
40324	EGB	7	1	0	1	1	2007	2011	4	20/05/2021
40346	SSA Bonds	5	1	0	1	0	2010	2015	5	28/04/2021
40410	Ethylene	4	0	0	0	1	2011	2017	6	14/07/2020
40481	Occupants safety systems 2	3	0	1	0	0	2001	2011	10	05/03/2019
55555	PC video games	6	0	1	0	0	2007	2018	11	20/01/2021

A2 Twitter query words

Table A2.1: Twitter query words

Query name	Case	Query name	Case
Denso	40113	Philips	39639
Bosch	40113	Lite-On	39639
Westlake	40410	Sony	39639
WestlakeChem	40410	Quanta storage	39639
Orbia	40410	JTEKT	39922
Clariant	40410	Nachi-Fujikoshi	39922
Celanese	40410	AB SKF	39922
Sanyo	40136	SKF	39922
Hitachi Chemical	40136	NTN Corporation	39922
hitachi	40136	Carpenter	39801
Matsuo	40136	Recticel	39801
Nichicon	40136	ABB Ltd	39610
Nippon Chemi-Con	40136	ABB	39610
BOSCH	39920	Nexans	39610
CONTINENTAL	39920	NKT	39610
UBS	40135	NKT A/S	39610
Barclays	40135	Prysmian	39610
RBS	40135	Safran	39610
Citigroup	40135	Sumitomo	39610
JP Morgan	40135	Hitachi metals	39610
Bank of Tokyo	40135	Hitachi	39610
MUFGEMEA	40135	J-power	39610
MUFG Bank	40135	SWCC	39610
Mitsubishi UFJ Financial Group	40135	Shova holdings	39610
Bank of Tokyo-Mitsubishi	40135	Mitsubishi materials	39610
Autoliv	40481	Mitsubishi cable industries	39610
Magnalnt	40299	Taihan Electric Wire	39610
Magna	40299	Taihan	39610
UBS	39861	ICAP	39861a
RBS	39861	Renesas	39574
Deutsche Bank	39861	Infineon	39574
JPMorgan	39861	Whirlpool	39600
Citigroup	39861	Panasonic	39600
Behr	39960	ArcelorMittal	38344
Denso	39960	Voestalpine	38344
Valeo_group	39960	Micron	38511
Valeo	39960	Hynix	38511
Panasonic	39960	Infineon	38511
Wilhelmsengroup	40009	Samsung	38511
Wallenius Wilhelmsen	40009	Samsung Semiconductor	38511
Philips	**39574**	Renesas	38511
Samsung	**39574**	NEC	38511
JP Morgan	39924	NEC Corporation	38511
UBS	39924	Hitachi	38511
Credit Suisse	39924	Mitsubishi Electric	38511
RBS	39924	Mitsubishi	38511
Sony	**39904**	Nanya	38511
Panasonic	39904	Toshiba	38511
Sanyo	**39904**	Kemira Oyj	38866
Samsung SDI	39904	Kemira	38866
Barclays	39914	Yara	38866
Societe Generale	39914	Tessenderlo	38866
RBS	39914	Ercros	38866
HSBC	39914a	Quimica	38866
Crédit Agricole	39914a	FMC Corporation	38866
JP Morgan	39914a	Trane Inc	39092
MAN	39824	Masco Corporation	39092
Volvo	39824	Villeroy & Boch	39092
Daimler	39824	Samsung	39309

Iveco	39824	LG Display	39309
DAF	39824	Chimei InnoLux	39309
Samsung SDI	39437	AU Optronics	39309
Philips	**39437**	HannStar Display Corporation	39309
LG Electronics	39437	Kuehne + Nagel	39462
Technicolor	**39437**	Deutsche Post	39462
Panasonic	39437	United Parcel Service	39462
Toshiba	**39437**	DSV	39462
Air Canada	**39258**	Johnson Controls	40018
Airfrance	**39258**	Recylex	40018
Air France	**39258**	Campine	40018
KLM	**39258**	Hitachi Chemical	40136
British Airways	**39258**	Hitachi	40136
Cathay Pacific	**39258**	Sanden	39960
LAN Chile	**39258**	K-Line	40009
Qantas	**39258**	Kawasaki Kisen Kaisha	40009
SAS	**39258**	Holy Stone Enterprise	40136
Singapore Airlines	**39258**	Fujikura	39610
Lufthansa	**39258**	NGK spark plugs	40113
Denso	40028	NGK	40113
Hitachi	40028	Mitsui	40009
Melco	40028	NYK Line	40009
Mitsubishi Electric	40028	Nippon Yusen	40009
Henkel	39579	CSAV	40009
Procter & Gamble	39579	Vapores	40009
Unilever	39579	NSK Ltd	39922
Kuehne + Nagel	40098	NSK	39922
Kuehne & Nagel	40098	#Kuehne	39462
Valeo	40013	Kuehne & Nagel	39462
#Hella	40013	Hannstar	39303
Asahi glass	39605	Villeroy	39092
Nippon electric	39605	#Kuehne	40098
Schott AG	39605	Deutsche Bank	40346
TOKAI RIKA	39881	Bank of America	40346
AUTOLIV	39881	Credit Agricole	40346
TOYODA GOSEI	39881	Credit Suisse	40346
Sumitomo	39748	NatWest	40324
Furukawa	39748	Nomura	40324
Leoni	39748	UBS	40324
Bandai Namco	55555	UniCredit	40324
Focus Home	55555	Furukawa	39610
Capcom	55555	Procter	39579

A3 Company characteristics

Table A3.1: Company characteristics

Company	Cartel	Economic Sector	Country	Fine over revenue	Total number of tweets	Abnormal return % decision date	Abnormal mood % decision date	Cartel tweets	Sentiment score cartel tweets
ArcelorMittal	38344	Other	Luxembourg	0,001701	2276	-0,01679		19	-0,28423
Voestalpine AG	38344	Other	Austria	0,001892	72	0,003453		1	-0,48507
Infineon	38511	Technology	Germany	0,018731	3507	-0,02681	-1,28421	86	-0,07121
Micron Technology	38511	Technology	United States	0	2819	-0,00523	-0,12052	6	-0,07228
Hitachi	38511	Other	Japan	0,00651	17736	-0,02313	0,094732	2	-0,17442
Hynix Semiconductor	38511	Technology	South Korea	0,01107	766	-0,00845	0	8	-0,26058
Mitsubishi Electrics	38511	Other	Japan	0,000593	24631	-0,01064	-0,04069	1	-0,17206
Nanya Electronics	38511	Technology	South Korea	0,001883	1837	0,009755	-1,79091	2	-0,2266
NEC Corporation	38511	Technology	Japan	0,000334	17324	-0,01221	0,125825	3	-0,04662
Renesas	38511	Technology	Japan	0,000822	693	-0,02624		1	-0,17678
Samsung	38511	Technology	South Korea	0,001783	193299	0,001083	-0,6402	96	-0,42116
Toshiba Corporation	38511	Other	Japan	0,000347	72768	-0,02813	-0,21321	3	-0,1684
Ercros	38866	Other	Spain	0,026061	1	-0,03131		1	0,158114
FMC Corporation	38866	Other	United States	0,007241	65	0,006201			
Kemira Oyj	38866	Other	Finland	0	45	-0,02276			
Quimica	38866	Other	Chile	0,002765	409	0,010042			
Tessenderlo	38866	Other	Belgium	0,04	31	-0,00325		5	-0,45421
Yara	38866	Other	Norway	0	4746	-0,02418	-1,54569		
Masco Corporation	39092	Consumer Cyclicals	United States	0	43	0,012801			
Trane Inc	39092	Industrials	United States	0,027385	25	-0,01014			
Villeroy & Boch	39092	Consumer Cyclicals	Germany	0,1	898	-0,23288	-2	6	-0,33475
Air Canada	39258	Industrials	Canada	0,003425	9568	-0,05123	-2,18638	8	-0,39273
Air France	39258	Industrials	France	0,007631	7306	-0,02069	1,068688	13	-0,31214
British Airways	39258	Industrials	United Kingdom	0,010495	20196	-0,01225	-1,95729	28	-0,48033
KLM	39258	Industrials	Netherlands	0,005305	10647	-0,02069	0,737533	3	-0,49334
LAN Chile	39258	Industrials	Chile	0,003261	102	0,008169	0	0	0
Lufthansa	39258	Industrials	Germany	0	9223	-0,02461	0,592217	5	-0,07261
Qantas	39258	Industrials	Australia	0,000985	32668	-0,00961	-2,02314	17	-0,42509
SAS	39258	Industrials	Sweden	0,01604	32454	0,013719	-0,24817	13	-0,25095
Cathay Pacific	39258	Industrials	Hong Kong	0,009918	3066	-0,00119	-1,45698	2	-0,48507
Singapore Airlines	39258	Industrials	Singapore	0,009731	6807	-0,00542	-1	3	-0,49502
AU Optronics	39309	Technology	Taiwan	0,015438	567	-0,02953			
Chimei InnoLux	39309	Technology	Taiwan	0,086498	45	-0,05921			
HannStar Display Corporation	39309	Technology	Taiwan	0,007565	0	0,02622			
LG Display	39309	Technology	South Korea	0,017752	3353	-0,00179	-1,66611	118	-0,10283
Samsung	39309	Technology	South Korea	0	198099	0,013218	0,020747	50	-0,02491
Philips	39437	Technology	Netherlands	0,021028	117669	-0,01289	-1,52925	600	-0,43584
Technicolor	39437	Consumer Cyclicals	France	0,011197	9291	-0,02128	-0,31459	4	-0,23619
LG Electronics	39437	Technology	South Korea	0,012752	9282	-0,01492	-2,3172	578	-0,4569
Panasonic	39437	Technology	Japan	0,002279	170727	-0,00612	-1,46423	240	-0,42535
Samsung SDI	39437	Technology	South Korea	0,041947	350	-0,01817		88	-0,31969
Toshiba Corporation	39437	Other	Japan	0,000926	147902	-0,00402	-0,21227	30	-0,39963
Deutsche Post	39462	Industrials	Germany	0	1629	-0,00304	0,69666	4	-0,12236
DSV	39462	Industrials	Denmark	6,47E-05	1690	0,01275	-3,88921	1	-0,65465
Kuehne + Nagel	39462	Industrials	Switzerland	0,003354	8	0,003285		1	-0,70711
United Parcel Service	39462	Industrials	United States	0,000258	3227	-0,00034	-2,03226	15	-0,30783
Infineon	39574	Technology	Germany	0,021542	3843	-0,01325	-2,74767	369	-0,53002
Philips	39574	Technology	Netherlands	0,000864	125565	0,002691	-1,69716	553	-0,54196
Renesas	39574	Technology	Japan	0	706	-0,05123		14	-0,35142
Samsung	39574	Technology	South Korea	0,000222	192806	0,014487	-0,47172	5	-0,48885
Henkel	39579	Other	Germany	0	2391	-0,00504	-0,08289	14	-0,34592
Procter & Gamble	39579	Other	United States	0,003597	10082	0,006827	-0,83995	108	-0,38402
Unilever	39579	Other	United Kingdom	0,002405	17312	0,001564	0,371305	246	-0,37716
Whirlpool S.A	39600	Consumer Cyclicals	Brazil	0,01654	44107	0,006603	-0,14282	5	-0,55036
Panasonic	39600	Technology	Japan	0,000115	186485	-0,02084	-0,22816	11	-0,4481
Asahi glass	39605	Consumer Cyclicals	Japan	0,003951	116	-0,00907		3	-0,72077
Nippon Electric Glass	39605	Technology	Japan	0,014664	101	-0,03695		0	0
ABB Ltd	39610	Industrials	Switzerland	0	36121	-0,00417	-0,07543	4	-0,32767
Nexans	39610	Industrials	France	0,01053	743	0,023901	0	13	-0,24173
NKT A/S	39610	Industrials	Denmark	0,001833	20998	0,060341	-1,36053	1	-0,25
Prysmian	39610	Industrials	Italy	0,014384	765	0,026331		86	-0,22618
Safran	39610	Industrials	France	0,000591	5155	-0,01293	0,291298		
Fujikura	39610	Other	Japan	0,006267	937	-0,01157			
Furukawa	39610	Industrials	Japan	0,00342	1736	-0,001	0,235294		
Hitachi Metals	39610	Industrials	Japan	0,002074	26346	-0,01279	0,071746		
J-power	39610	Other	Japan	0,001265	39	-0,00471			
Mitsubishi Materials	39610	Other	Japan	0,000273	60	-0,00425			
Sumitomo	39610	Consumer Cyclicals	Japan	0,00053	2413	-0,00307	0,327751		
SWCC Shova Holdings	39610	Industrials	Japan	0,002139	925	-0,0163			
Taihan Electric Wire	39610	Industrials	South Korea	0,003554	8	0,001584			
Philips	39639	Technology	Netherlands	0	117490	0,004654		1	-0,22613
Lite-On	39639	Technology	Taiwan	0	304	0,017143		1	0,01409
Quanta storage	39639	Technology	Taiwan	0,017524	24	0,001186		1	0,01409
Sony	39639	Technology	Japan	0,000539	175000	0,000285	-0,36958		
Leoni	39748	Industrials	Germany	0,000362	6526	0,019991	-1,48689	13	-0,51659
Furukawa	39748	Industrials	Japan	0,000426	3736	-0,02037	-0,51843		
Sumitomo	39748	Consumer Cyclicals	Japan	0	1592	0,020274		1	-0,33955
Carpenter	39801	Other	United States	0,044503	76340	0,014384	-0,18501	1	-0,27735
Recticel	39801	Other	Belgium	0,027609	71	0,114968		2	-0,65507
DAF	39824	Industrials	Netherlands	0,042776	17490	-0,01373	-0,89333	183	-0,07759
Daimler	39824	Consumer Cyclicals	Germany	0,006749	20020	-0,0008	-2,14649	209	-0,13219

Company	Cartel	Economic Sector	Country	Fine over revenue	Total number of tweets	Abnormal return % decision date	Abnormal mood % decision date	Cartel tweets	Sentiment score cartel tweets
Iveco	39824	Industrials	Italy	0,020681	5215	-0,00856	-2,0674	158	-0,0464
MAN	39824	Industrials	Germany	0	184602	0,008446	-1,58629		
Volvo	39824	Industrials	Sweden	0,020054	85952	0,014934	-0,43289	206	-0,10809
Citigroup	39861	Financials	United States	0,000932	15167	0,001286	-4,49382	173	-0,45391
Deutsche Bank	39861	Financials	Germany	0,013605	18808	0,001676	-3,06815	37	-0,0648
JP Morgan	39861	Financials	United States	0,000984	62525	0,00801	-4,61151	167	-0,46487
RBS	39861	Financials	United Kingdom	0,011773	79561	0,003909	-2,76004	299	-0,34625
UBS	39861	Financials	Switzerland	0	24868	0,012305	-1,54804	49	-0,03069
ICAP	39861a	Financials	United Kingdom	0,016845	4431	-0,00982	-1,38506	94	-0,35644
Autoliv	39881	Consumer Cyclicals	Sweden	0,000846	1564	0,005886	-1,16999	10	-0,41465
TOKAI RIKA	39881	Consumer Cyclicals	Japan	0,00045	17	-0,00656			
TOYODA GOSEI	39881	Consumer Cyclicals	Japan	0,001698	190	-0,00096			
Panasonic	39904	Technology	Japan	0,000671	103066	0,001671	-0,24525	213	-0,4749
Samsung SDI	39904	Technology	South Korea	0	1823	0,011626		26	-0,28136
Sanyo	39904	Industrials	Japan	0,140885	6471	0,006287	-2,10114	136	-0,52066
Sony	39904	Technology	Japan	0,000483	176548	-0,00204	-0,31858	39	-0,5906
Barclays	39914	Financials	United Kingdom	0	160386	-0,00837	-0,66209	60	-0,22615
Societe Generale	39914	Financials	France	0,003185	3220	-0,0022	0,747187		
Credit Agricole	39914a	Financials	France	0,001829	5113	-0,01078	-2,20363	267	-0,23715
HSBC	39914a	Financials	United Kingdom	0,000422	92903	0,020371	-2,92244	193	-0,30416
JP Morgan	39914a	Financials	United States	0,003628	16242	-0,01314	-2,52266	9	-0,40164
Bosch	39920	Consumer Cyclicals	Germany	0,02123	77728	-0,01332	0,16687	4	-0,23162
CONTINENTAL	39920	Consumer Cyclicals	Germany	0,001	117504	0,004046	-0,1211	6	-0,3434
AB SKF	39922	Industrials	Sweden	0,041255	4724	-0,00397	0,181595	8	-0,19721
JTEKT Corporation	39922	Industrials	Japan	0	32	-0,0034			
Nachi-Fujikoshi Corporation	39922	Industrials	Japan	0,002752	5	0,020431			
NSK Ltd	39922	Consumer Cyclicals	Japan	0,010408	5790	0,005126	-0,04064		
NTN Corporation	39922	Industrials	Japan	0,045609	9	0,000575			
Credit Suisse	39924	Financials	Switzerland	0,000305	28913	-0,00067	-1,45137	150	-0,43951
JP Morgan	39924	Financials	United States	0,000941	26785	0,002474	-7,02758	23	-0,59171
RBS	39924	Financials	United Kingdom	0	113283	0,001951	-0,25765	115	-0,28645
UBS	39924	Financials	Switzerland	0,000423	34277	0,006066	-2,39343	162	-0,43471
Behr	39960	Consumer Cyclicals	United States	0,009335	6202	0,024017	-2,26442	13	-0,47097
Valeo	39960	Consumer Cyclicals	France	0,00162	1849	-0,00424	-1,52963	11	-0,46271
Denso	39960	Consumer Cyclicals	Japan	8,85E-06	1463	0,006155			
Panasonic	39960	Technology	Japan	0	58908	0,008798	-0,1914	1	-0,4
Sanden	39960	Consumer Cyclicals	Japan	0,027296	1264	-0,00938		3	-0,35591
CSAV	40009	Industrials	Chile	0,066882	79	-0,01692		11	-0,22863
Wallenius Wilhelmsen	40009	Industrials	Norway	0,073293	196	-0,01011		2	-0,49419
K-Line	40009	Industrials	Japan	0,004545	513	-0,01084			
Mitsui	40009	Industrials	Japan	0	3441	-0,00204	0,17148	2	-0,27732
NYK Line	40009	Industrials	Japan	0,008828	199	0,000198		12	-0,24918
Hella	40013	Consumer Cyclicals	Germany	0,001637	507	-0,00754		2	-1
Valeo	40013	Consumer Cyclicals	France	0	3034	0,011318	-1,375	25	-0,46065
Campine	40018	Other	Belgium	0,047748	51	0,000135		7	-0,44801
Johnson Controls	40018	Industrials	Ireland	0	2085	-0,00693	0,076087		
Recylex	40018	Industrials	France	0,069986	23	-0,01156		5	-0,63151
Denso	40028	Consumer Cyclicals	Japan	0	2319	0,000702	-0,67514	9	-0,27704
Hitachi	40028	Other	Japan	0,006859	25899	-0,01516	-1,6919	100	-0,25429
Melco	40028	Other	Japan	0,003437	3056	-0,0063	-2,08377	73	-0,23892
Kuehne + Nagel	40098	Industrials	Switzerland	0	18	-0,00213			
Bosch	40113	Consumer Cyclicals	Germany	0,030969	77733	-0,01349	0,16687	4	-0,23162
Denso	40113	Consumer Cyclicals	Japan	0	2842	-0,00032		4	-0,1502
NGK Spark Plugs	40113	Consumer Cyclicals	Japan	0,009719	17061	0,007767	0,35989	2	-0,28181
Barclays	40135	Financials	United Kingdom	0,006432	103228	-0,00318	-3,20672	352	-0,36668
Citigroup	40135	Financials	United States	0,003776	19292	0,008384	-3,06077	217	-0,36966
JP Morgan	40135	Financials	United States	0,002052	43370	0,005979	-2,74139	50	-0,41039
RBS	40135	Financials	United Kingdom	0,013579	85225	-0,00024	-2,427	400	-0,39336
UBS	40135	Financials	Switzerland	0	35466	-0,00225	-0,60159	25	-0,23018
Bank of Tokyo-Mitsubishi	40135	Financials	Japan	0,082876	2306	0,008751	-2,3054	45	-0,40632
Hitachi	40136	Other	Japan	0,004036	55318	0,001371	-0,06031	12	-0,27708
Holy Stone Enterprise	40136	Technology	Taiwan	0,001959	1	0,023637			
Matsuo	40136	Technology	Japan	0,022248	1977	-0,02555	-2,71429	2	-0,3103
Nichicon	40136	Industrials	Japan	0,087908	40	-0,01478		2	-0,3103
Nippon Chemi-Con	40136	Industrials	Japan	0,101927	8	-0,03292		3	-0,302
Sanyo	40136	Industrials	Japan	0	1585	-0,00036	0,115578	1	-0,1409
Magna	40299	Consumer Cyclicals	Canada	0	58227	-0,00065	-1,40252	3	-0,40094
NatWest	40324	Financials	United Kingdom	0	28444	0,008105	-1,80259	7	-0,076
UBS	40324	Financials	Switzerland	0,006286	35220	0,002266	-2,03984	113	-0,28256
UniCredit	40324	Financials	Italy	0,002966	2686	-0,00824		104	-0,24738
Nomura	40324	Financials	Japan	0,007837	30814	-0,00616	-3,44884	109	-0,27505
Bank of America	40346	Financials	United States	0,000145	63638	-0,00069	2,623853	41	-0,25551
Credit Agricole	40346	Financials	France	5,54E-05	1877	0,005056	-2,07935	38	-0,37366
Credit Suisse	40346	Financials	Switzerland	0,000409	42785	0,013513	-1,22761	39	-0,37011
Deutsche Bank	40346	Financials	Germany	0	54876	0,103755	-0,6078	30	-0,32989
Celanese	40410	Other	United States	0,01472	689	0,022035		5	-0,28176
Clariant	40410	Other	Switzerland	0,039271	1348	0,005184	0,045701	6	-0,34697
Orbia	40410	Other	Mexico	0,003536	192	0,035916		5	-0,28176
Westlake	40410	Other	United States	0	20365	0,024418	-1,45081	2	-0,06259
Autoliv	40481	Consumer Cyclicals	Sweden	0,026831	555	0,000441		11	-0,2252
Focus Home	55555	Technology	France	0,020225	71053	-0,02037	-0,81276		
Bandai Namco	55555	Consumer Cyclicals	Japan	5,73E-05	13107	-0,01235	-2,14207		
Capcom	55555	Technology	Japan	0,000592	158223	-0,01556	-1,11401		

A4 Stock ticker information

Table A4.1: Companies with stock tickers and market index

Company name	Yahoo ticker	Market index	Datastream ticker
AB SKF	SKF-B.ST	^OMX	W:SKFB
ABB Ltd	ABBN.SW	^SSMI	S:ABBN
Air Canada	AC.TO	^GSPTSE	C:AC
Air France	AF.PA	^FCHI	F:UTA
ArcelorMittal	MT	^NYA	U:MT
Asahi glass	5201.T	^N225	J:AG@N
AU Optronics	2409.TW	^TWII	TW:ADT
Autoliv	ALIV-SDB.ST	^OMX	U:ALV
Bandai Namco	7832.T	^N225	J:N@MB
Bank of America	BAC	^NYA	U:BAC
Bank of Tokyo-Mitsubishi	8306.T	^N225	J:KYTB
Barclays	BARC.L	^FTSE	BARC
Behr	MAS	^NYA	U:MAS
Bosch	BOSCHLTD.NS	^NSEI	IN:BOH
British Airways	BAY	^FTSE	BAY
Campine	CAMB.BR	^BFX	B:CAM
Capcom	9697.T	^N225	J:CAPO
Carpenter	XTY.F	^GDAXI	U:CRS
Cathay Pacific	0293.HK	^HSI	K:CATH
Celanese	CE	^NYA	U:CE
Chimei InnoLux	3481.TW	^TWII	TW:INN
Citigroup	C	^NYA	U:C
Clariant	CLN.SW	^SSMI	S:CLN
CONTINENTAL	CON.DE	^FCHI	D:CON
Credit Agricole	ACA.PA	^FCHI	F:CRDA
Credit Suisse	CSGN.SW	^SSMI	S:CSGN
CSAV	VAPORES.SN	^IPSA	CL:VPR
DAF	PCAR	^IXIC	@PCAR
Daimler	DAI.DE	^GDAXI	D:DAI
Denso	6902.T	^N225	J:DE@N
Deutsche Bank	DBK.DE	^GDAXI	D:DBK
Deutsche Post	DPW.DE	^GDAXI	D:DPW
DSV	DSV.CO	^OMX	DK:DSV
Ercros	ECR.MC	^FCHI	E:ECR
FMC Corporation	FMC	^NYA	U:FMC
Focus Home	ALFOC.PA	^FCHI	F:ALFO
Fujikura	5803.T	^N225	J:GG@N
Furukawa	FUWAY	^N225	J:FU@N
HannStar Display Corporation	6116.TW	^TWII	TW:HDC
Hella	HLE.DE	^GDAXI	D:HLE
Henkel	HEN3.DE	^GDAXI	D:HEN
Hitachi	6501.T	^N225	J:LK@N
Hitachi Metals	5486.T	^N225	J:HM@N
Holy Stone Enterprise	3026.TW	^TWII	TW:HSE
HSBC	HSBA.L	^FTSE	HSBA
Hynix Semiconductor	000660.KS	^KS11	KO:HYI
ICAP	TCAP.L	^FTSE	TCAP
Infineon	IFX.DE	^GDAXI	D:IFX
Iveco	CNHI.MI	FTSEMIB.MI	I:CNHI
Johnson Controls	JCI	^GSPC	U:JCI
JP Morgan	JPM	^NYA	U:JPM
J-power	9513.T	^N225	J:EPDC
JTEKT Corporation	6473.T	^N225	J:OE@N
Kemira Oyj	KEMIRA.HE	^OMX	M:KEMR
K-Line	9107.T	^N225	J:KK@N
KLM	AF.PA	^FCHI	F:UTA
Kuehne + Nagel	KNIN.SW	^SSMI	S:KNIN
LAN Chile	LFL.F	^GDAXI	CL:LAN
Leoni	LEO.DE	^GDAXI	D:LEO

LG Display	034220.KS	^KS11	KO:LGL
LG Electronics	066570.KS	^KS11	KO:JHD
Lite-On	2301.TW	^TWII	TW:LOT
Lufthansa	LHA.DE	^GDAXI	D:LHA
Magna	MG.TO	^GSPTSE	C:MG
MAN	MAN.DE	^GDAXI	D:MAN
Masco Corporation	MAS	^NYA	U:MAS
Matsuo	6969.T	^N225	J:MSUC
Melco	6503.T	^N225	J:UM@N
Micron Technology	MU	^NYA	@MU
Mitsubishi Electrics	6503.T	^N225	J:UM@N
Mitsubishi Materials	5711.T	^N225	J:LM@N
Mitsui	9104.T	^N225	J:MO@N
Nachi-Fujikoshi Corporation	6474.T	^N225	J:FK@N
Nanya Electronics	2408.TW	^KS11	TW:NYT
NatWest	NWG.L	^FTSE	NWG
NEC Corporation	6701.T	^N225	J:NJ@N
Nexans	NEX.PA	^FCHI	F:NXS
NGK Spark Plugs	5334.T	^N225	J:KS@N
Nichicon	6996.T	^N225	J:NP@N
Nippon Chemi-Con	6997.T	^N225	J:PJ@N
Nippon Electric Glass	5214.T	^N225	J:LO@N
NKT A/S	NKT.CO	^OMX	DK:NKT
Nomura	8604.T	^N225	J:NM@N
NSK Ltd	6471.T	^N225	J:NSKC
NTN Corporation	6472.T	^N225	J:NTN
NYK Line	9101.T	^N225	J:NY@N
Orbia	ORBIA.MX	^MXX	MX:CSB
Panasonic	6752.T	^N225	J:MI@N
Philips	PHIA.AS	^N100	H:PHIL
Procter & Gamble	PG	^NYA	U:PG
Prysmian	PRY.MI	FTSEMIB.MI	I:PRY
Qantas	QAN.AX	^AXJO	A:QANX
Quanta storage	6188.TWO	^TWII	TW:QSI
Quimica	SQM	^NYA	U:SQM
RBS	NWG.L	^FTSE	NWG
Recticel	REC.BR	^BFX	B:REC
Recylex	RX.PA	^FCHI	F:RX
Renesas	6723.T	^N225	J:RENE
Safran	SAF.PA	^FCHI	F:SGM
Samsung	005930.KS	^KS11	KO:SGL
Samsung SDI	006400.KS	^KS11	KO:SCT
Sanden	6444.T	^N225	J:SAEN
Sanyo	5958.T	^N225	J:SYAM
SAS	SAS.ST	^OMX	W:SAS
Singapore Airlines	C6L.SI	^STI	T:SAIR
Societe Generale	GLE.PA	^FCHI	F:SGE
Sony	6758.T	^N225	J:SO@N
Sumitomo	SSUMY	^N225	J:SUEL
SWCC Shova Holdings	5805.T	^N225	J:SHEW
Taihan Electric Wire	001440.KS	^KS11	KO:TWR
Technicolor	TCH.PA	^FCHI	F:TCH
Tessenderlo	TESB.BR	^BFX	B:TES
TOKAI RIKA	6995.T	^N225	J:TI@N
Toshiba Corporation	6502.T	^N225	J:TS@N
TOYODA GOSEI	7282.T	^N225	J:TYGS
Trane Inc	TT	^NYA	U:TT
UBS	UBSG.SW	^SSMI	S:UBSG
UniCredit	UCG.MI	FTSEMIB.MI	I:UCG
Unilever	ULVR.L	^FTSE	ULVR
United Parcel Service	UPS	^NYA	U:UPS
Valeo	FR.PA	^FCHI	F:FR
Villeroy & Boch	VIB3.DE	^GDAXI	D:VIB3
Voestalpine AG	VOE.VI	^ATX	O:VAS
Volvo	VOLV-B.ST	^OMX	W:VOBF

Wallenius Wilhelmsen	WAWI.OL	^OSEAX	N:WWL
Westlake	WLK	^NYA	U:WLK
Whirlpool S.A	WHRL4.SA	^BVSP	BR:NS4
Yara	YAR.OL	OSEBX.OL	N:YARA

A5 Regression output

Regression Results

	Dependent variable:			
	Abnormal stock return event day			
	(1)	(2)	(3)	(4)
Fine as % of revenue	-0.372*** (-0.558, -0.185)			
Abnormal mood		-0.001 (-0.005, 0.004)		
Sentiment score including cartel			-0.025 (-0.059, 0.010)	
Tweets including cartel				0.00000 (-0.00004, 0.0001)
Constant	0.001 (-0.004, 0.007)	-0.006 (-0.014, 0.002)	-0.012* (-0.026, 0.001)	-0.004 (-0.011, 0.003)
Observations	132	87	101	101
R2	0.105	0.001	0.020	0.0004
Adjusted R2	0.098	-0.011	0.010	-0.010
Residual Std. Error	0.026 (df = 130)	0.029 (df = 85)	0.030 (df = 99)	0.030 (df = 99)
F Statistic	15.284*** (df = 1; 130)	0.079 (df = 1; 85)	1.970 (df = 1; 99)	0.040 (df = 1; 99)
Note:				*p<0.1; **p<0.05; ***p<0.01

Figure A5.1: Single factor on event day

Regression Results

	Dependent variable:			
	Abnormal stock return [-1,1]			
	(1)	(2)	(3)	(4)
Fine as % of revenue	-0.371*** (-0.592, -0.151)			
Abnormal mood [-1,1]		-0.00003 (-0.002, 0.002)		
Sentiment score including cartel			-0.025 (-0.062, 0.013)	
Tweets including cartel				0.00001 (-0.00004, 0.0001)
Constant	0.0003 (-0.006, 0.007)	-0.006* (-0.013, 0.001)	-0.014* (-0.029, 0.001)	-0.006 (-0.014, 0.001)
Observations	132	87	101	101
R2	0.077	0.00001	0.017	0.002
Adjusted R2	0.070	-0.012	0.007	-0.008
Residual Std. Error	0.031 (df = 130)	0.028 (df = 85)	0.033 (df = 99)	0.033 (df = 99)
F Statistic	10.876*** (df = 1; 130)	0.001 (df = 1; 85)	1.686 (df = 1; 99)	0.239 (df = 1; 99)
Note:				*p<0.1; **p<0.05; ***p<0.01

Figure A5.2: Single factor on event window [-1,1]

Regression Results

	Dependent variable:			
	Abnormal stock return [0,10]			
	(1)	(2)	(3)	(4)
Fine as % of revenue	-0.462*** (-0.791, -0.133)			
Abnormal mood [0,10]		0.001 (-0.001, 0.002)		
Sentiment score including cartel			-0.025 (-0.082, 0.032)	
Tweets including cartel				-0.00001 (-0.0001, 0.0001)
Constant	-0.010** (-0.019, -0.001)	-0.013** (-0.024, -0.003)	-0.023** (-0.045, -0.001)	-0.014** (-0.025, -0.002)
Observations	129	85	98	98
R2	0.056	0.013	0.008	0.0003
Adjusted R2	0.049	0.002	-0.003	-0.010
Residual Std. Error	0.046 (df = 127)	0.046 (df = 83)	0.049 (df = 96)	0.050 (df = 96)
F Statistic	7.592*** (df = 1; 127)	1.131 (df = 1; 83)	0.750 (df = 1; 96)	0.026 (df = 1; 96)
Note:				*p<0.1; **p<0.05; ***p<0.01

Figure A5.3: Single factor on event window [0,10]

A6 Robustness checks on twitter data variables

Table A6.1: "Only cases after 08.11.2011

Coefficient	[-15,15]	[-10,0]	[-5,5]	[-1,1]	0	[0,2]	[0,10]
Abnormal sentiment	0.003	0	0.003*	0.001	0	-0.003	0
Sentiment score cartel tweets	0.258	0.143	0.041	0.060	-0.036	-0.055	-0.084
Count cartel tweets	0	0	0	0	0	0	0

Table A6.2: Only cases before 08.11.2017

Coefficient	[-15,15]	[-10,0]	[-5,5]	[-1,1]	0	[0,2]	[0,10]
Abnormal sentiment	0		0.001	0	-0.001		0.001
Sentiment score cartel tweets	-0.016	-0.009	0.014	-0.031	-0.026	-0.024	-0.024
Count cartel tweets	0	0	0	0	0	0	0

Table A6.3: Only companies with at least 10000 tweets

Coefficient	[-15,15]	[-10,0]	[-5,5]	[-1,1]	0	[0,2]	[0,10]
Abnormal sentiment	0.001	0	0.001	0	-0.001	0	0.001
Sentiment score cartel tweets	-0.038	0.037	0.055	0.005	-0.007	0.006	-0.019
Count cartel tweets	0	0	0	0	0	0	0

A7 BMP-test

It is important to use statistical tests to see if there are significant evidence for abnormal returns after the event study. In our analysis of abnormal stock return, the original BMP test have been utilized. The original BMP test Boehmer et al. (1991) is a standardized cross-sectional method which is robust to increased variance due to events. It uses standardized abnormal returns to try and decrease the impact of highly volatile returns. This method weights the more volatile abnormal returns (AR) less than the others. It is used is many similar studies, for example Aguzzoni et al. (2013) and Ulrich (2018). Standardized abnormal returns (SAR) is made by:

$$SAR_{it} = \frac{AR_{it}}{S_{ar_{it}}} \tag{.1}$$

where the standard deviation is calculated as:

$$S_{AR_{it}} = \sqrt{S^2_{AR_i} * (1 + \frac{1}{M_i} + \frac{R_{mt} - \overline{R}_m}{\sum_{t=T_0}^{T_1} *(R_{mt} - \overline{R}_m)^2}} \tag{.2}$$

It adjust for a forecast error which is neccessary because the event window is out-of-sample predictions. The t-statistics for the original BMP-test are:

$$z_{BMP,T} = \frac{ASAR_t}{\sqrt{N} S_{ASAR_t}} \tag{.3}$$

When testing for cumulative average abnormal return, the standard deviation changes to:

$$S^2_{CARi} = \sqrt{S^2_{AR_i} * (L_i + \frac{L^2}{M_i} + \frac{(\sum_{t=T_1+1}^{T2}(R_{mt} - \overline{R}_m))^2}{\sum_{t=T_0}^{T_1} *(R_{mt} - \overline{R}_m)^2}} \tag{.4}$$

L is the number of days in the event window, while M is the number of days in the training period. The t-value with standardized cumulative abnormal return is calculated as:

$$z_{BMP} = \sqrt{N} * \frac{\overline{SCAR}}{S_{\overline{SCAR}}} \tag{.5}$$

SCAR is the averaged standardized abnormal return (ASAR) from all firms N, with standard deviation calculated in the same manner as with the average standard abnormal return.

For the event study on abnormal Twitter mood, the standard cross-sectional test has been used. Here the t-statistics are the equation below when testing for average abnormal return:

$$t_{AAR_t} = \sqrt{N} * \frac{AAR_t}{S_{AAR_t}} \tag{.6}$$

The standard deviation comes from:

$$S^2_{AAR_t} = \frac{1}{N-1} * \sum_{i=1}^{N}(AR_{i,t} - AAR_t)^2 \tag{.7}$$

When testing for cumulative abnormal return the formulas are the same, except for the change from abnormal return to the cumulative abnormal return in the event window in all equations:

$$t_{CAAR_t} = \sqrt{N} * \frac{CAAR_t}{S_{CAAR_t}} \tag{.8}$$

The standard deviation comes from:

$$S^2_{CAAR_t} = \frac{1}{N-1} * \sum_{i=1}^{N} (CAR_{i,t} - CAAR_t)^2 \tag{.9}$$

A8 Regression coefficients for immune companies

Table A8.1: Immune companies regression

Coefficient	[-15,15]	[-10,0]	[-5,5]	[-1,1]	0	[0,2]	[0,10]
Abnormal sentiment	-0.0003	-0.004	-0.007**	-0.001	-0.006	0.005	-0.002
25% percentile mood	0	0	0	0	0	0	0
75% percentile mood	-0.001	-0.018	-0.011	-0.023	-0.014	-0.013	-0.006
Sentiment score cartel tweets	-0.056	-0.003	-0.024	-0.068	-0.004	0.009	0.048
Count cartel tweets	0.0004	0.0002	-0.00003	0.0002	0.0001	0.0002	0.0001
European	0.019	0.002	-0.013	-0.008	-0.0003	-0.0002	0.005
Asian	-0.008	0.012	0.003	0.001	-0.002	-0.007	-0.026
American	-0.026	-0.029	0.023	0.015	0.006	0.015	0.041
log(Revenue)	-0.003	-0.009	-0.012	0.001	0.005	0.005	-0.007

A9 Mood distribution and event windows of mood

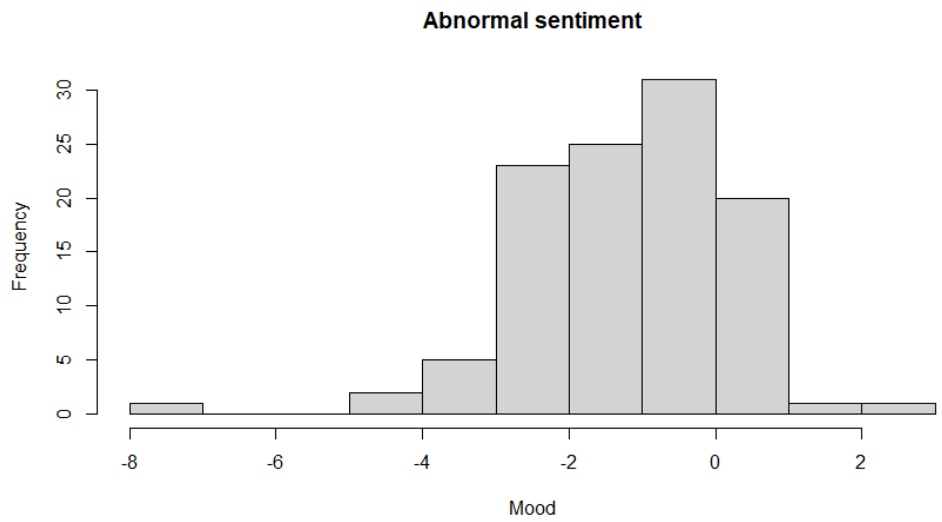

Figure A9.1: distribution of companies according to abnormal mood

A10 Full abnormal mood table

Table A10.1: Abnormal mood table for different event windows and subsamples

Sample	Event window	CAAR	T-value	P-value	Observations	Positive CAR
All companies	(-15) to 15	-233.00%	-2.14	0.034**	108	37%
	(-10) to 0	-90.00%	-1.89	0.061*	108	41%
	(-5) to 5	-232.00%	-4.36	0.00003***	108	29%
	(-1) to 1	-184.00%	-7.02	0.00001***	108	21%
	0	-115.00%	-8.73	0.00001***	108	20%
	0 to 2	-233.00%	-8.22	0.00001***	108	18%
	0 to 10	-251.00%	-4.17	0.00006***	108	30%
Immunity	(-15) to 15	-414.00%	-2.41	0.03**	22	45%
	(-10) to 0	-213.00%	-2.18	0.041**	22	45%
	(-5) to 5	-195.00%	-2.2	0.04**	22	41%
	(-1) to 1	-94.00%	-2.42	0.024**	22	27%
	0	-56.00%	-3.44	0.002***	22	27%
	0 to 2	-104.00%	-2.83	0.01***	22	23%
	0 to 10	-168.00%	-2.13	0.05**	22	41%
Not immunity	(-15) to 15	-189.00%	-1.45	0.156	85	34%
	(-10) to 0	-59.00%	-1.09	0.28	85	39%
	(-5) to 5	-240.00%	-3.77	0.0003***	85	26%
	(-1) to 1	-205.00%	-6.57	0.00001***	85	20%
	0	-113.00%	-8.22	0.00001***	85	19%
	0 to 2	-265.00%	-7.82	0.00001***	85	16%
	0 to 10	-273.00%	-3.69	0.0004***	85	27%
European	(-15) to 15	-367.00%	-3.14	0.003***	54	35%
	(-10) to 0	-184.00%	-3.39	0.001***	54	35%
	(-5) to 5	-220.00%	-4.44	0.00005***	54	26%
	(-1) to 1	-165.00%	-6.45	0.00001***	54	19%
	0	-111.00%	-6.94	0.00001***	54	22%
	0 to 2	-209.00%	-6.62	0.00001***	54	17%
	0 to 10	-235.00%	-3.67	0.0006***	54	30%
American	(-15) to 15	286.00%	0.52	0.61	15	33%
	(-10) to 0	240.00%	1.01	0.33	15	53%
	(-5) to 5	-302.00%	-1.13	0.28	15	13%
	(-1) to 1	-415.00%	-3.06	0.008***	15	7%
	0	-215.00%	-3.73	0.002***	15	7%
	0 to 2	-491.00%	-4.06	0.001***	15	7%
	0 to 10	-422.00%	-1.31	0.21	15	13%
Asian	(-15) to 15	-150.00%	-1.63	0.11	37	41%
	(-10) to 0	-64.00%	-1.49	0.145	37	43%
	(-5) to 5	-166.00%	-2.78	0.009***	37	38%
	(-1) to 1	-111.00%	-4.2	0.0001***	37	30%
	0	-81.00%	-4.9	0.00002***	37	24%
	0 to 2	-159.00%	-4.15	0.0002***	37	24%
	0 to 10	-159.00%	-3.21	0.003***	37	35%

A11 AAR for stock return and AABMOOD for mood

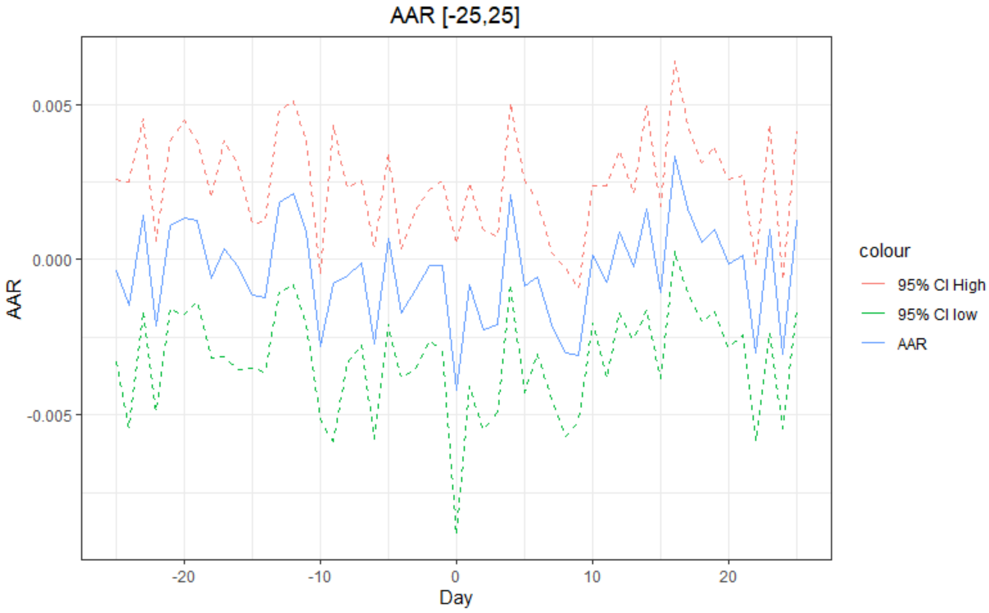

Figure A11.1: AAR with confidence interval for stock prices

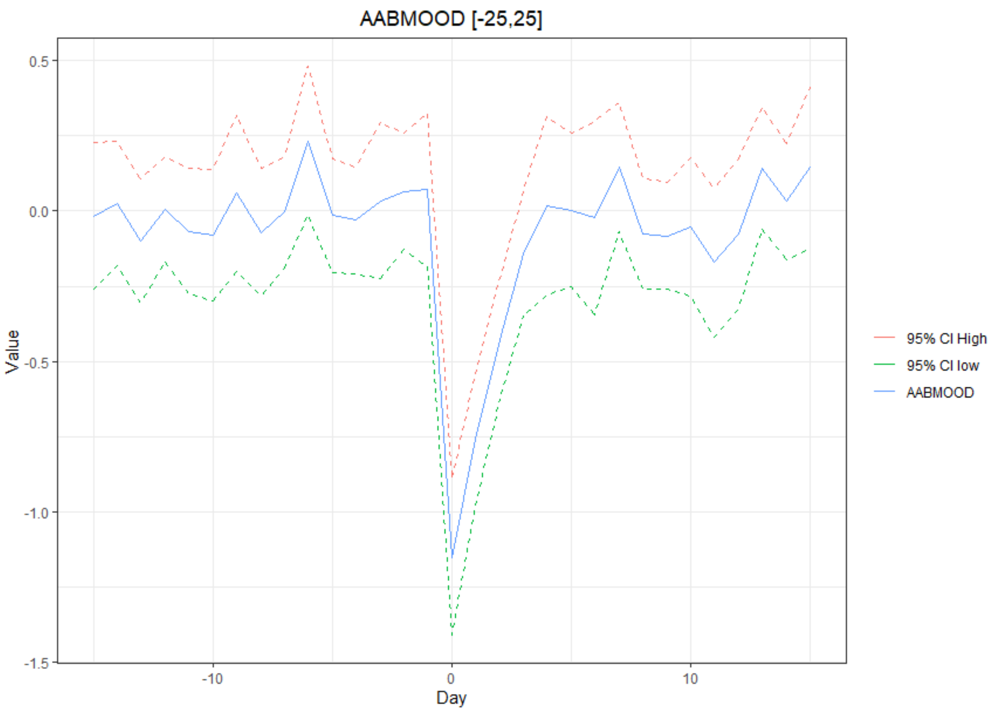

Figure A11.2: AABMOOD with confidence interval