Charles University

Faculty of Social Sciences
Institute of Economic Studies

MASTER'S THESIS

The Impact of News on Videogame Stock Market Prices and Volatility

Author: **Bc. Veronika Mertová**

Study program: **Economics and Finance**

Supervisor: **PhDr. František Čech Ph.D.**

Academic Year: **2023**

Declaration of Authorship

The author hereby declares that she compiled this thesis independently; using only the listed resources and literature, and the thesis has not been used to obtain a different or the same degree.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, August 1, 2023 Veronika Mertová

Acknowledgements

I am very grateful to my thesis supervisor PhDr. František Čech Ph.D. for providing valuable comments and suggestions and for his time investment into improving the thesis. I would also like to express my gratitude to PhDr. Jiří Schwarz Ph.D. for providing his insight and suggesting thesis improvements during consultations. Lastly, I would like to thank my family for supporting me during my studies.

Abstract

The thesis investigates the impact of social media and news headline sentiment on stock prices, specifically comparing gaming firms to companies from other industries. Tweets and news headlines containing keywords referring to four selected gaming and four non-gaming companies were collected over 5 and 3 months, respectively. Both tweets and news collected came from the general users or media rather than focusing solely on financial ones. The data were aggregated into daily values. Daily stock price data were also collected for each examined company to derive returns and volatility. The data were analysed using a vector autoregression model in combination with Granger causality. The study found no significant differences between gaming and non-gaming sectors. The polarity of sentiment showed no effect on stock prices. However, when sentiment was divided into different emotions, some significance was observed, although the findings varied across individual firms regardless of their sectors. It was concluded that when using sentiment for market predictions, it is beneficial to either utilize specifically financial media or determine the specific type of sentiment that influences a particular stock.

JEL Classification	G14, G17, C32, C58
Keywords	Tweets, News Headlines, Gaming Industry, Sentiment Analysis, Emotions, Stock Returns, Volatility
Title	The Impact of News on Videogame Stock Market Prices and Volatility

Abstrakt

Tato práce zkoumá vliv sentimentu na sociálních sítích a v novinových titulcích na ceny akcií, konkrétně porovnává společnosti z herního průmyslu s firmami z ostatních odvětví. Tweety obsahující klíčová slova odkazující na čtyři vybrané firmy z herního průmyslu a čtyři jiné firmy byly sbírány po dobu 5 měsíců, novinové titulky potom 3 měsíce. Tweety a zprávy pocházely od běžných uživatelů a medií, nikoliv pouze z finanční oblasti. Pro každou zkoumanou společnost, byla sbírána také denní data o cenách akcií za účelem výpočtu výnosů a volatilit. K analýze dat byl využit model vektorové autoregrese v kombinaci s Grangerovou kauzalitou. Studie neprokázala žádný významný rozdíl mezi herním a neherním sektorem. Polarita sentimentu významně neovlivňovala tržní ceny. Nicméně, když byl sentiment rozdělen na jednotlivé emoce, byla pozorována určitá signifikantnost, výsledky se ale lišily u jednotlivých firem, bez ohledu na jejich odvětví. Byl učiněn závěr, že při využití sentimentu pro předpovídání trhu je vhodné využít výhradně finanční média nebo určit konkrétní typ sentimentu, který ovlivňuje zvolené akcie.

Klasifikace JEL	G14, G17, C32, C58
Klíčová slova	tweety, titulky zpráv, herní průmysl, analýza sentimentu, emoce, výnosy, volatilita
Název práce	Vliv zpráv na ceny a volatilitu akciového trhu videoherního průmyslu

Table of Contents

List of tables

List of plots

List of equations

Acronyms

AAA	large scale, blockbuster videogames, also called “triple-A”
ADF	Augmented Dickey–Fuller test
API	Application Programming Interface
CDPR	CD Projekt Red
DJIA	Dow Jones Industrial Average
DLC	Downloadable Content
GARCH	Generalized AutoRegressive Conditional Heteroskedasticity
VADER	Valence Aware Dictionary and sEntiment Reasoner
VAR	Vector Autoregressive model

Master’s Thesis Proposal

Author: Bc. Veronika Mertová

Supervisor: PhDr. František Čech Ph.D.

Proposed topic: The Impact of News on Videogame Stock Market Prices and Volatility

Motivation:

Evidence from recent years suggests that there is a relationship between reported news and movements on financial markets. According to Strycharz et al. (2018) the topics covered by media are influential on the stock market, both communication about financial news and communication to customer have the influence. Furthermore, it was found that emotionality predicted market fluctuation for some companies and that higher emotionality led to an increase in fluctuation. Other research shows that sentiment analysis of social media can be useful in predicting stock returns. (Bollen et al., 2011)

In December 2020 a polish game developer CD Project released its long-awaited game “Cyberpunk 2077.” The release was preceded with a heavy and ambitious marketing campaign and the game received a lot of attention on the social media which (alongside with the success of previous game) led to the stock price of CD Project steadily increasing. However, shortly before the release, the stock price started to plummet down. This was mainly due to the reviews being published, they stated that the game is in poor technical condition and does not fulfill the promises given by the developer. (Łuczuk and Maj, 2022) Moreover, CD Project was facing a huge fan backlash on the social media.

The goal of the proposed thesis is to explore further the influence of social media and reported news on the value of publicly traded videogame companies. The main focus will be on the potential difference a gaming community can have on this influence, as it can be assumed that gamers are more active and involved with their favorite products on social media as was demonstrated by the case of Cyberpunk.

Hypotheses:

1. Hypothesis #1: The videogame companies stock volatility shows the characteristics of leverage effect, volatility clustering and persistence
2. Hypothesis #2: The videogame companies stock prices change more following specific news/social media sentiment than prices of other types of companies
3. Hypothesis #3: The change in volatility of videogame companies stocks is higher following specific news/social media sentiment than change in stock volatility of other companies

Methodology:

In the thesis the vector autoregression model will be used to model the dependency between media variables and market fluctuations as well as between social media sentiment and market fluctuations. The models for video game industry and others will be compared. Augmented Dickey-Fuller test will be implemented to test for stationarity and Akaike’s information criterion to select appropriate number of lags.

To estimate volatility of the stocks GARCH family models will be used depending on data availability and their performance. Daily log returns will be used as stock price change variable.

Expected Contribution:

The videogame industry is fast-growing and becoming more significant in the financial markets. It is estimated to grow up to $268 billion annually by 2025 up from $178 billion in 2021. (statista, 2021) The previous research has discovered that media and social media can play a role in understanding financial market fluctuations. The company communication (both financial and to customer) can influence its market value.

This thesis will extend upon previous research by exploring these effects within a specific industry and will aim at providing insight to whether this market is more susceptible to the media sentiment. Such knowledge can be helpful to the companies when producing reports as well as potential investors.

Outline:

1. Introduction: motivation for the topic and main goals of the work
2. Literature review: go over the previous research focused on the influence of news and social media on stock values and stability
3. Data: describe the data sources and their processing
4. Methodology: describe and explain the used models and test
5. Empirical analysis: show and comment the results of the analysis, test the hypotheses
6. Conclusion: summarize the main findings and discuss the implementation

Core Bibliography:

Bollen, J., Mao, H. and Zeng, X. (2011), “Twitter mood predicts the stock market”, *Journal of Computational Science*, Vol. 2 No. 1, pp. 1–8.

Engle, R. (2001), “GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics”, *Journal of Economic Perspectives*, American Economic Association, Nashville, Vol. 15 No. 4, pp. 157–168.

Łuczuk, P. and Maj, S. (2022), “Can Stock Market Fluctuations Be Predicted? Wig.Games Index Secret on the Example of CD Project”, *Social Communication*, Vol. 8 No. 1, pp. 8–20.

Luetkepohl, H. (2007), “Econometric Analysis with Vector Autoregressive Models”, European University Institute

Nelson, D.B. (1991), “Conditional Heteroskedasticity in Asset Returns: A New Approach”, *Econometrica*, [Wiley, Econometric Society], Vol. 59 No. 2, pp. 347–370.

Strauß, N., Vliegenthart, R. and Verhoeven, P. (2018), “Intraday news trading: The reciprocal relationships between the stock market and economic news”, *Communication Research*, SAGE Publications, Los Angeles, CA, Vol. 45 No. 7, pp. 1054–1077.

Strycharz, J., Strauss, N. and Trilling, D. (2018), “The Role of Media Coverage in Explaining Stock Market Fluctuations: Insights for Strategic Financial Communication”, *International Journal of Strategic Communication*, Routledge, Philadelphia, Vol. 12 No. 1, pp. 67–85.

Zheludev, I., Smith, R. and Aste, T. (2014), “When Can Social Media Lead Financial Markets?”, *Scientific Reports*, Springer Nature, pp. 1–12.

1 Introduction

Over the past several decades, research on market price movements has increasingly incorporated investor sentiment as one of the explanatory variables. This approach aims to capture the overall attitude and emotional outlook of investors towards a particular asset or the market as a whole. Investor sentiment encompasses a range of emotions such as fear, joy, optimism, and pessimism, all of which can significantly influence trading decisions and subsequently impact asset prices and market volatility. While traditional financial theories and models have primarily focused on fundamental factors like earnings, interest rates, and economic indicators, they often fail to fully account for the irrational and unpredictable behavior exhibited by investors.

Advancements in technology, particularly the rapid growth of social media platforms and the widespread adoption of the internet, have created an opportunity to collect and analyze vast amounts of sentiment-related data. With the aid of sentiment analysis tools and machine learning algorithms, researchers can now go through an immense number of textual data sources, including news articles, social media posts, and online forums, to assess the prevailing sentiment among market participants. This explosion of information sharing has empowered individual investors and enthusiasts to have a more significant impact on asset prices, underscoring the need for a comprehensive investigation into the dynamics of investor sentiment.

The gaming industry, in particular, has witnessed exponential growth and transformation in recent years. With the advent of online gaming communities, forums, and social media groups dedicated to gaming discussions, gamers from all over the word have found a powerful tool to voice their opinions, experiences, and emotions related to gaming companies and their products. This interactive and passionate engagement sets the gaming community apart from traditional industries, potentially creating distinct patterns in the relationship between sentiment and stock price movements. The main goal of this thesis is to distinguish whether the effect of sentiment differs between the gaming sector and other ones. As gamers tend to be more active and vocal about their favourite products, it can be assumed that the sentiment expressed by the gaming community has a larger effect on the stock price movements of a matching gaming company. The thesis aims to collect and analyse data for four firms of gaming and four games of non-gaming industry and compare the effect the sentiment has on the returns and volatility of each one of them.

The basis for assuming differences in the sentiment effect between the gaming and other industries lies in the unique behavior of gamers on the internet and their reactions to the successes or failures of video game companies. An example of the impact of investor sentiment on the gaming industry can be seen in the research conducted by Łuczuk and Maj, (2022). The study focused on the Polish video game development company CD Projekt Red, where a failure to deliver on promises during a massive marketing campaign led to a significant backlash on social media and in the news. This backlash was followed by an almost instant price drop in the company's stock prices and even resulted in a lawsuit issued by investors.

To analyze the relationship between sentiment, stock returns, and volatility, the thesis employs the vector autoregression model in conjunction with Granger causality analysis, which offers a framework for exploring causal relationships. By utilizing sentiment data extracted from Twitter and news headlines, the research considers both individual opinions and mainstream media coverage, providing a comprehensive perspective on how different sources of sentiment may influence market dynamics. It is wiorth noting, that in contrast to previous research, this work collect data from general Twitter posts and news headlines, rather than just financial ones, as it might better capture the difference between the gamers and other internet users behaviour. The sentiment data were collected between October 20th, 2022, and March 10th, 2023, from Twitter, and headlines were collected from January 1st to March 31st, 2023. Price data were obtained from Yahoo Finance.

The thesis is structured as follows. The next chapter delves into the specifics of the gaming industry and its processes to provide readers with a deeper understanding of its unique characteristics compared to other industries. Chapter three reviews previous research and works that have dealt with sentiment analysis and its links to market performance, encompassing studies related to sentiment expressed both in newsprint and social media. In chapter four, the data collection and processing methods are described and explained in more detail, along with summary statistics. Chapter five presents the employed methodology, providing comprehensive details of the models used. The results of the analyses, hypothesis testing, and interpretation of findings are presented in chapter six. Finally, chapter seven concludes the thesis by summarizing the key findings and offering insights into potential limitations and suggestions for future research.

2 Background

2.1 Videogame industry performance

The relevance of videogame industry has been steadily increasing during the last decade. It has become the most profitable part of the entertainment sector and is expected to continue growing. Newzoo a company that specializes in analysing videogame industry, has expected the value to grow to $217 billion by 2023. This estimation was based on high expectations from gaming industry during the pandemic. Its performance has skyrocketed as revenues for publicly traded companies grew by almost 30% by June 2020. The growth was expected to continue at high rate as many people stayed at home and looked for a means of entertainment or interaction with friends. (Wijman, 2020) The industry has grown to $192 billion by the end of 2021 and during 2022 its value has fallen to $184 billion. That can, however, be considered a market correction following two years of growth fuelled by the pandemic. Newzoo expects 2022 to be the only corrective year for the gaming industry and anticipates further growth in the following years. They state that even though players have been spending less due to a tougher economic situation in 2022, there are now more customers in the industry than ever before. (Wijman, 2022)

The gaming industry is not homogenous and its various parts focus on different types of customers. Among these segments, mobile games stand out as the most successful, commanding 50% of the gaming market. This dominance can be attributed to the ease of entry for customers and the prevalent free-to-play monetization model employed by a significant portion of mobile games. Mobile games alone have generated $92 billion in 2022, they are followed by $52 billion by console games, $38 billion by PC games and lastly, $2 billion generated by browser games. (Clement, 2022a) Moreover, different monetization strategies play a crucial role in the digital gaming industry, where developers adopt diverse business approaches to maximize profits. These strategies will be further explored in later sections of this thesis.

2.2 Videogame marketing and social media

2.2.1 Price of development

There are various prices associated with releasing a videogame. The companies need to pay their employees, provide equipment and space and then sell the product efficiently which makes marketing a large portion of expenditures as well. The price is, of course, dependent on the size of a project. While some games that are uncomplicated and developed by a small team or even a single person cost less than a thousand dollars, games produced by a publicly traded company will be significantly more expensive to develop. Medium sized games can cost several hundred thousand dollars and prices for the largest or so-called "triple-A" titles can reach hundreds of millions of dollars. (Yury and Mickiewicz, 2018)

AAA or triple-A titles are huge projects that take the developers several years to complete with hundreds of employees working on them. For example, Grand Theft Auto 5 (known as GTA 5) cost $265 million to develop and release, the budget for development itself was $140 million, the rest is marketing expenses. (Selway, 2023) Only when the product us finished and released, the company generates income. The success is highly dependent on sales soon after the release since the life cycle of games is short and a rapid decay in sales can be expected after a few weeks. This is, however, mainly an issue for single player games, as the games with a large multiplayer component depreciate in slower rate. (Engelstätter and Ward, 2013) Developers who release single player content can then solve this issue by creating additional content for an already released game. The "DLC" or downloadable content will be discussed more in the monetization section of the thesis.

In order to maximize the sales in the period shortly after the release, the developers need to focus time and resources to marketing. For small developers, the expenditures for marketing start around 10% but can be as high as 50% of the project budget. (Denby, 2019) For larger projects, these expenditures rise even higher.

High prices of development, need for quality marketing and the fact that all revenue will be generated only after the release may lead developers to enter a contract with a publisher. The role of a publisher is to help with the pre-release expenses in exchange for the share of earnings. According to Statista, one of the largest videogame publishers in the world *Electronic Arts,* has spent $961 million dollars solely on marketing and sales expenditures. (Clement, 2022b) The global spending for videogame advertisements has reached over $5 billion in 2020.

2.2.2 Common marketing strategies

Videogames can benefit from a large variety of marketing strategies due to being a digital, easy to access product. Traditional and most expensive methods, such as TV commercials or billboards are used only by the largest producers. More popular and widely spread methods utilise gaming press, conferences, social media, game distribution platforms or participation marketing.

When promoting a game, it is desirable to establish contact with specialised media. Gaming focused press has high, even global reach and is therefore able to influence large number of potential customers long before the game is released. Direct communication with the media is then one of the first steps when promoting a newly created game title and creating a fanbase around it, especially for smaller developers. It is, however, necessary to make a good impression with trailers, screenshots or other material from the game. (Záhora, 2014) The companies should, therefore, try to deliver quality information about their product to the press in order to make researching it easier. For this purpose, so-called press kit is created by the developers and provided to the media. (Carriker, 2017)

As in other industries, conferences are organized that give the developers, publishers, media, hardware producers and other members of gaming industry an opportunity to get together in order to exchange information and form contracts between each other. Furthermore, they draw attention of gamers worldwide crating a space for new announcements, advertisements and showcases. Due to the interest of gamers, some of the industry-only conferences have gradually opened to public and focused more on the customers. Among the most popular conventions are *Gamescom* in Germany, *E3* in the USA and *Tokyo Game Show.* (Tedesco, 2019)

Developers are also incentivised to use a number of marketing tools provided for them by the distribution platforms on which their games are sold. For the retailers of physical copies, these include posters, statues or other visual attractions placed in the shop as well as providing some bonus (a small present, for example) for purchasing the game. The marketing methods of digital retailers are more sophisticated and in-depth as their digital environment allows for it. For example, one of the largest digital retailer *Steam* provides the developers with detailed guidelines on how to promote a game in their environment. It recommends setting up a product page early, so that the interested customers can add it to their watchlist. The product page should include easy to grasp, important information about the product and not be cluttered, it can then be used throughout the development to make announcements and monitor interest. Furthermore, *Steam* has detailed information about their users and can utilize a system of recommendations to target specific customers that are likely to enjoy given title. On the release day, the game will be featured on the "new releases" section of the store. After the release, especially after the decay in sales mentioned above, the developers can promote their game further by decreasing its price using a sale placing the game on a more prominent place in the store and sending emails to all who wishlisted it. Or, by releasing additional content for it.

Community marketing is essential for a gaming company, its purpose is to create a brand and build a community of engaged customers around it. It focuses on strengthening the relationship with existing customer rather than just attracting new ones. Therefore, a strong communication channels are required, preferably those, that allow customers to give feedback directly to developers, and social media are a perfect fit. In addition to popular social websites, the companies utilise their own space (such as forums or blogs) to create and maintain communities. Such approach lets developers receive ideas and criticisms from fans and makes them feel involved in the process.

There are even further advantages to creating communities, individual members can help each other to solve problems without direct involvement of the developer. It creates platform to share experience and opinions, helps the spread of news about the product and if a healthy community is created, it will contain loyal customers who are happy to be involved with the given brand. However, there are also drawbacks to this approach, especially when social media are involved. Firstly, a non-member can access the community posting false information or damaging the brands name. Furthermore, negative reviews or comments tend to influence popularity more than positive ones. (Baxi *et al.*, 2016)

If a gaming company is successful at creating a community around their product, they have access to one of the most powerful marketing tools in the field. Excited and involved gamers create and share new ideas, information and other content leading to a situation where the consumers rather than the authors themselves become the promoters of a brand. The companies can then encourage such creativity by providing additional incentives to the active community members and use their engagement as a marketing tool. This approach is known as participation marketing. (Poch and Martin, 2015) Research has shown that user-generated content can be more persuasive than professionally created content, however, can affect customers differently. So, the balance between the two is recommended. (Goh Khim-Yong *et al.*, 2013)

There forms of involving community in the gaming industry vary. The developers can include engagement tools directly to the game, the most basic form of this approach is in-game customization. Or they can provide the community with tools that help modify the game itself, that allows players to create "mods" which might include small changes to the game details or even major overhaul of core game mechanics. Majority of modders then state, that they find creating mods fun and enjoy improving the product for the players. (Poor, 2014)

The most popular and spread form od participation marketing is, however, videogame streaming – a live broadcast of gameplay acompanied by commentary. The broadcasters (or streamers) can be anyone who just enjoys playing games and decides to share their experience in this form, or they can be proffesionals who work from a streaming studios, spend several hours each day streaming and earn enough money this way to make it their living. The income of the professionals comes either directly from their fans or they can enter a contract with developers and be paid for streaming their games. (Johnson and Woodcock, 2017) Research has shown that people, who concider themselves gamers are more likely to believe user-generated content than the traditional marketing methods. Furthermore, if someone is familiar with specific creator, they are more likely to believe their opinion. (Foster, 2016)

2.3 Monetization

As mentioned above, videogames are a digital product and can therefore be monetized using different approaches than would be possible with a physical product. In order for a game to be successful commercially, the developers need to select a suitable form of monetization (or a combination of them) early in the game development cycle and implement it well. (Grotland, 2011) In the following section the most commonly used monetization strategies will be discussed.

2.3.1 Premium and DLCs

Premium is the term used to label a game that is sold once and grants the customer access to the whole product without any limitations. This approach is most common for titles developed for PC or consoles. However, as mentioned in section 1.2.1, it takes years to create a game that is most likely to be sold for full price only a few weeks after the release. Furthermore, the cost of creating a game has increased during the decades due to higher wage demands or cost of implementing new technologies. In has been an industry standard that a full price for a triple-A game is $60 for consoles and PC. With the release of new console generations at the end of 2019, *Sony* has increased this price to be $70 and was shortly followed by other large publishers, one of the last to keep the original ammount was *Microsoft*, who will be increasing to a new standard at the begining of 2023. (Stewart, 2022)

Apart from increasing the initial price of purchasing a game, developers can use other methods to generate more revenue from a released product. One of these methods is creating extensions for the game after it has been released. So-called "DLCs" (downloadable content) wary in size and content, from simply adding a few new items or game modes to completely reworking the game systems or adding hours of gameplay. This approach can prolong the game's lifecycle indefinitely. There are companies maintaining only one or two products through updates and DLCs and basing their bussiness plan on that, essentialy making the DLCs their primary product. According to *Steam* store page, there are over 80 DLCs available for *Euro Truck Simulator 2* that was released in 2012 and yet, the game maintains its high user base. (Steam, 2022)

2.3.2 Game as a service

In the recent years, gaming companies have tried to prolong the lifecycles of their games by continually updating, changing and improving the experience for players. This approach keeps the customer interested and engaged with the title for a significantly longer period. It is most used for titles with a multiplayer component or purely multiplayer ones and became known as "Games as service" in the industry. For games supported this way it is common to incorporate some form of microtransaction into their monetization model to make money of the game for as long as possible. (Schreier, 2017) Microtransactions will be discussed in one of the following sections.

2.3.3 Free to play

The term free to play is used for titles that a user can download and play in their entirety without any payment requirement. The developers can use various methods to generate revenue in a free to play title. One of them is incorporating in-game advertisement paid for by the advertising company. They can incentivise players to view the ads by providing an in-game bonus for those who watch. More commonly, however, some form of in-app purchase is available. Players can buy virtual goods, equipment, faster progress or visual changes using real world money in the game. (King and Delfabbro, 2019)

The free to play model suffers from a relatively low conversion rate, research found that only 1% to 5% of users ever makes a purchase. From those who do, majority pays between $1 and $5 per month which accounts to less than 15% of a game’s total revenue. Most of the revenue comes from the so-called "whales" who spend on average more than $25 monhtly, but form a low percentage of the whole user base. (Shi *et al.*, 2015)

2.3.4 Subscription

A subscription method of monetizing a product has become widespread in the recent years not only in gaming, but in the digital entertainment industry in general. It requires a user to pay regularly in order to use a given product. In the gaming castor, the subscriptions can be divided into two different categories. The first one is subscription to a specific game which provides either the ability to play the game (most famous example includes the *World of Warcraft*) or in-game bonusses such as faster progress, additional storage, unique cosmetics etc.

The second category are the game libraries which function similarly as *Netflix* or *Spotify*. When subscribed, user gains access to hundreds of games that can be downloaded and played for free while the subscription lasts, these include *Xbox game pass* and *PlayStation plus*. The third category focuses on gamers who may not have sufficient hardware to run modern games and provide the opportunity to play them through cloud. *GeForce NOW* is a service that allows users to play any game from their *Steam* library using remote hardware.(GeForce NOW, 2022) *Google* has launched similar service called *Stadia* in 2019, however, it will be shut down in January 2023. (Keane, 2022)

2.3.5 Microtransactions

The microtransaction model is being implemented into continually supported premium titles with "game as service" approach or free-to-play games. It offers players small purchase options (known as microtransactions) that provide the purchaser with additional content not available to non-paying players or speed up the process of gaining free content. (King and Delfabbro, 2019)

Depending on the title, developers and the platform for which it is developed, different types of in-game content can be purchased through microtransactions. Purely cosmetic ones, that only change the visage of a player’s character, mount, spells etc. but have no direct gameplay effect are considered acceptable by 69% of gaming community. On the other hand, purchasable items that provide in-game advantage in form of additional power, faster progression, or are in any way "pay-to-win" are disliked and concidered toxic. (Taylor, 2018)

Other forms of microtransactions do not offer one specific item but rather a package of various rewards that are unlocked by the player making progress in the game itself. This package is called "battle pass" or "season pass" and usually offers the best value a person can get for their money in exchange for requiering them to level the passs up to unlock the rewards that are already paid for. (Davenport, 2018) Some games include a free battle pass with an option to pay for a premium branch that includes better rewards. Furthermore, other passes will reward leveling up by providing the game's premium currency which can then be used to purchase other items. (Epps, 2022) As the rewards from the pass are usually unobtainable after a season matched with it ends, this strategy takes the advantage of the fear of missing out effect which can serve as an efficient way to keep the player hooked and is critisized by players for it. (Livingston, 2020)

Loot boxes are an item players can buy or earn through game activities; it is a cache that contains randomized items. None of the content is guaranteed meaning that players are spending money on the chance of getting a desired item rather than paying for it directly. Due to the randomness of drops, the loot boxes can be considered a form of gambling and are therefore regulated in some countries. The forms of regulation differ, they include a complete ban on using loot boxes or give the developer an obligation to reveal a percentage chance of receiving a specific item. (Straub, 2020)

Microtransactions are a difficult topic in the industry and their reception by the community differs based on the developer's approach toward them. For example, the developers of a popular free-to-play game *Path of Exile* are praised for their monetization system. They sell only cosmetic items and quality-of-life improvements (such as bigger storage space) and no purchase can make your character stronger in any way. Furthermore, they created a good game which players enjoy and are willing to support the team by making purchases in the store. (Loot and Grind, 2019)

On the other hand, there are studios that create sophisticated monetization schemes, which according to King and Delfabbro (2018) can be considered predatory. They can include systems that hide the true long-term cost of playing the game until the player is already committed either psychologically, financially or both. They are designed to encourage repeated spending by strategically placing rewards in a way that reinforces purchase behaviour and exploit the inequality of information between the provider and purchaser. These systems encourage players to spend increasing amounts of money which leads to a feeling they have already spent too much to quit playing and therefore spend even more.

The predatory monetization systems are common for free-to-play mobile games and appear only minimally on PC or Consoles and when they do, the backlash of community follows. A great example from recent year is the release of *Diablo Immortal* a mobile game playable also on PC developed by *Blizzard Entertainment*. It was released with a highly aggressive monetization that encouraged players to spend real world money in order to get stronger, the "whales" soon became so powerfull they could not be matched by any small-spender or a free-to-play user. (Welsh, 2022) This led to a quick exodus of majority of players, mainly those who came from PC and were used to games where in order to get better, you need to play and be good. This specific case will be more discussed in one of the following sections.

2.4 Videogame stock prices

The videogame stocks have underperformed in 2022 due to the lift of pandemic restriction and less people staying home than the previous years. However, the industry is a growing one and 2022 is considered just a correction year, the expected future growth is 12.9% annually through 2030. Constant technology upgrades lead to improving the gaming experience and the industry attracts more customers each year. (Bolton, 2022)

As mentioned in the price of development section, it takes substantial time, effort and money to make a videogame and the gaming companies are betting a lot on a successful release and good player acceptance. After the years spend in development, even a great game can be accepted negatively if the monetization model is chosen poorly, the marketing is too aggressive, or the developers do not take time to finalize the product and release it with bugs or in a bad technical condition. This is also true for publicly traded companies, because an unsuccessful release may mean a loss of investor trust as well. In the following section, two recent cases badly managed releases followed by a backlash and stock price drop will be introduced.

2.4.1 Case of CD Project RED

A renowned Polish developer *CD Project RED* (or *CDPR)* famous for their successful *Witcher* game trilogy has announced a new title set into a cyberpunk world in 2013, which was even before the release of their later top selling, award winning hit *Witcher 3*. The gamers had limited information about it until a huge marketing campaign began in June 2019 at the *E3* conference in Los Angeles. After the trailer was played at the presentation, Keanu Reeves (from the blockbuster movies *The Matrix* or *John Wick*) took the stage, and it was announced he will play one of the protagonists of *Cyberpunk 2077.* The main advertising channels for the game were characteristic black and yellow merchandise and billboards, trailers played in cinemas, television commercials and a huge social media campaign. The game has profiles on large social media and the developers have shown a lot of material on *YouTube.* The game's budget was about $315 million and 45% of it was used for advertising. With the campaign of enormous size, the developers tried to reach not only the gaming community, but the public in general and they succeeded. (Łuczuk and Maj, 2022) With the high expectations for the release date set to December 10, 2020, and the pandemic present, the stock prices of *CDPR* grew. (MarketWatch, 2022)

Cyberpunk 2077 was selling well, according to the *CDPR's* financial results for 2020, the game has sold over 13.7 million copies by the end of the year, by these numbers alone, one could assume the release was a success. However, compared to the expectations the gaming community and the investors had for so highly anticipated title, the result was considered a disappointment. Furthermore, the first reviews of the game were mostly critical, mainly due to the fact that it was released in a poor technical condition that made it nearly unplayable on lower end PCs and "current generation" consoles (referring to *PlayStation 4* and *Xbox One* as the "new generation" *PlayStation 5* and *Xbox Series X/S* were available on the market for only a few weeks) even though the game was sold specifically for the "current generation" without any dedicated version for the new. These problems were so severe they led to a temporary removal of the game from the *PlayStation store*. Concidering the extencive marketing campaign and the studio's communication prior to release, they managed to make the impression that the game was already a hit even before it was available. (Łuczuk and Maj, 2022) All of these factors led to the stock price decline, from zł443 ($101) in December 2020 to zł165 ($38) in August 2021, the price has not recovered to the pre-release value up to December 2022 when it is zł130 ($30). (MarketWatch, 2022)

Furthermore, *CDPR* was facing a lawsuit from their investors over the disastrous launch. They stated the game was virtually unplayable on the current gen consoles, that the *CDPR* was forced to offer refunds and subsequently, their reputation was damaged. And, most importantly, that the company released statements about its business and operations that were false, misleading or without a reasonable basis at relevant times. Therefore, the lawsuit claimed, the investors had suffered damages. In December 2021, the lawsuit was settled by *CDPR* paying $1.85 million settlement to the investors. (Plunkett, 2021)

2.4.2 Case of Activision Blizzard

In November 2018 at the *Blizzard's* conference *BlizzCon* a new instalment of the *Diablo* series was announced to the players worldwide. The announcement was met with a negative response from the fans who have long anticipated a PC sequel to the series, *Diablo Immortal*, however, was to be a mobile game. Following the announcement, a huge backlash by the gamers on social media occurred. The reveal trailer had been seen by over 3 million people scoring 443 thousand dislikes and only 17 thousand likes, players were signing a petition demanding the title to be cancelled and the hashtag *#NotMyDiablo* circulated the social media. (Taylor, 2018b) This fallout led to a drop of *Activision Blizzard* stock price from $49 in November 2018 to $42 in January 2019. (Yahoo Finance, 2022) After that, the company's stock price experienced a period of growth further accelerated by the pandemic that, in contrast to other industries boosted the gaming one.

On June 2, 2022 *Diablo Immortal* was released on mobile devices and even on PC, the PC version development surprised many players when it was announced in April 2022. It was to alow for cross-play and cross-saves meaning that PC players could play with anyone on mobile and also alternate between both devices playing the same character. (Goslin, 2022) The game was released with a higly aggressive monetization and as mentioned above, this approach is more common for mobile devices where majority of players came to accept it. That is, however, different for gamers who play primarily on PC and during the firs week of the release, *Diablo Immortal* was already facing a fan outrage caused mainly by the microtransactions system, but also poor technical condition of the PC version, which was the exact copy of the mobile one without any improvements. Gamers on Twitter and Reddit as well as streamers and other content creators were calling out the problems and expressing dismay even though the game in its core was good. (Gale General OneFile, 2022) The displeasure of fans and their outrage on social media was again followed by a stock price drop, from $78 in June to $73 in October 2022. (Yahoo Finance, 2022)

3 Literature review

This section will focus on describing sentiment analysis and its application to financial market data. Sentiment analysis, also known as opinion mining, is a technique that leverages natural language processing to determine the emotion or sentiment expressed in a given text. It has emerged as a powerful tool for understanding and interpreting textual data and is widely utilized by businesses, governments, and researchers. Given the vast number of internet users, social networks, news outlets, and other websites, sentiment analysis finds utility across various domains. This work primarily emphasizes its relevance within the financial markets.

The two primary approaches for sentiment analysis are lexicon-based and machine learning algorithms. The machine learning approach utilizes supervised learning techniques to learn complex patterns and make decisions based on empirical data. Consequently, it requires a substantial amount of labelled training data to develop a successful algorithm (Hassan Yousef et al., 2014). The other, lexicon-based approach relies on pre-existing sentiment dictionaries that contain words or phrases along with associated sentiment scores and therefore, does not require a robust training set. The overall sentiment of a text is determined by calculating a score based on the presence and score of the words from the lexicon. These lexicons can be curated manually, automatically using associations with so-called "seed words," or semi-automatically (Khoo and Johnkhan, 2018). However, a disadvantage of the lexicon-based approach is its real-world application, as the same word can have multiple meanings in different contexts. With a sufficient training set, a machine learning model is likely to outperform the lexicon approach.

Nevertheless, the performance of the lexicon-based approach can be enhanced by incorporating rules into the algorithm, resulting in a lexicon and rule-based approach. In addition to lexicons, rule-based systems utilize a set of predefined rules or heuristics to determine sentiment. These rules are based on linguistic patterns, sentence structures, or field-specific knowledge. For instance, a rule might state that a negation word can reverse the sentiment of the following word, scoring "not happy" as negative rather than positive. In this paper, VADER (Valence Aware Dictionary and sEntiment Reasoner), a lexicon and rule-based sentiment analysis tool specifically designed to analyse sentiments expressed in social media, developed by Hutto and Gilbert (2014), will be utilized.

3.1 Investor sentiment and stock prices

According to the Efficient Market Theory introduced by (Fama, 1970), all available information is instantly reflected in asset prices. Consequently, asset prices are to already incorporate all relevant information. However, (Malkiel, 2003) argues that the media can be biased, as they may have interests or preferences that influence the information they provide. Additionally, even unbiased information can be misinterpreted, resulting in not all information being accessible to everyone. Moreover, the popularity of published articles can attract more investor attention and cause short-term deviations from efficient market prices. This section will further explore the influence of investor sentiment on asset pricing.

Due to the inability of traditional financial models to capture, predict or even explain shock and connected dramatic changes in stock prices, behavioural economists began to implement models incorporating additional information based on certain assumptions. First of them, introduced by De Long *et al.*, (1990) is that investors are influenced by their sentiment. The second assumption, by Shleifer and Vishny, (1997) adds that since trading against sentimental investors can be risky, rational investors should not anticipate prices to strictly adhere to what a traditional model would imply. (Barber and Odean, 2000) examined trading patterns and behaviours of individual investors in their paper. While their research does not specifically target media, it suggests that media coverage can contribute to overreactions and heightened trading volume by individuals, potentially resulting in suboptimal trading outcomes. Since an individual's investing behaviour can impact market liquidity and, consequently, short-term price volatility, media coverage could, therefore, influence asset prices.

One of the earlier works specifically focused on investor sentiment is by (Brown and Cliff, 2005). They utilized a direct survey to construct a sentiment index and analysed its relationship with subsequent asset returns and market valuation ratios. The findings indicated that investor sentiment does have a significant impact on asset values. Excessive optimism was associated with the overvaluation of stocks and the overall market, while excessive pessimism with undervaluation. Additionally, the authors identified a connection between sentiment levels and future stock returns. Their research suggested that extreme sentiment levels are often followed by a reversal in stock prices. For instance, periods of high optimism tend to be followed by lower returns. Baker and Wurgler (2007) followed with research in which they assert that the impact of sentiment on stock prices is no longer a matter of debate, as it has been proven in several studies. But rather the question is now centred on understanding the nature of this effect and devising methods to measure it. In their research, they focused on identifying which stocks exhibit the highest sensitivity to investor sentiment. They concluded that more speculative stocks tend to be particularly responsive to sentiment. Furthermore, they discovered that stocks of firms that do not pay dividends, have low market capitalization, or exhibit higher volatility are associated with higher-than-expected returns during periods of low sentiment. Additionally, they observed that elevated optimism can result in overvaluation of stocks before a market crash.

3.2 News and stock prices

Most of the previous research examining the relationship between media sentiment and the stock market has primarily focused on financial news and has analysed sentiment within entire articles. One of the early papers that delved into this topic is by Tetlock (2007), which centred on the role of media coverage in shaping investor sentiment and its subsequent impact on stock market behaviour. The specific financial news articles investigated in the study were the "Abreast of the Market" columns published in The Wall Street Journal over a 16-year period from 1984 to 1999. These columns were selected due to their wide readership, established reputation among investors, and online availability.

To analyse the daily variation of the column, Tetlock employed the General Inquirer (GI), a quantitative analysis program developed by Stone *et al.* (1966), in combination with the Harvard psychosocial dictionary. Vector autoregressions were then employed to estimate the connections between measures of media sentiment and the stock market. The primary finding of the study was that pessimistic sentiment leads to temporary price downturns, which can result in increased investor activity, heightened volatility, and greater market volume in the short term, thereby influencing traders' emotions. This suggests that media sentiment does indeed influence investor sentiment, and subsequent changes in investor sentiment have an impact on market returns.

Rather than analysing general financial news articles, Engelberg (2008) focused specifically on articles that contained earnings announcements of the firms under examination. In his study, he differentiated between soft and hard measures of earnings. The hard measures were derived from the accounting data provided in the earnings announcement, while the soft measures were based on articles discussing those earnings. Engelberg also used a metric for abnormal returns, which allowed him to assess the performance of stocks following an earnings announcement. His findings indicated that soft earnings news held predictive power for larger changes in future time horizons. He further identified soft information that was relevant for analysts and had implications for future returns. Schumaker *et al.*, (2012) conducted another study aiming to distinguish the effect of different types of news. They investigated whether the objectivity or subjectivity of an article can affect its predictive power. Additionally, they examined whether the sentiment polarity of a subjective article plays a role in prediction accuracy. To facilitate their research, they developed a system called AZFinText, which collected price data from a publicly available database and financial articles from Yahoo! Finance. The system represented the articles using proper nouns and polarity and employed machine learning algorithms to create market predictions every 20 minutes. The authors' findings suggest that the subjectivity of articles can influence trading behaviour, and their system performed best in predicting market movements using negative subjective articles. Interestingly, they observed a downswing associated with positive articles and upswings linked to negative and neutral articles, which contradicts the conclusions of Tetlock (2007).

Li *et al.* (2014) employed a company-specific approach in their research, focusing on companies listed on the Hong Kong Stock Exchange. They utilized a news archive to gather articled between years 2003 and 2008, which contained both market news and company-specific news. The authors noted that there is correlation between the number of news articles and both stock weight and market capitalization, indicating that as a stock becomes larger, it tends to receive greater news coverage. They compared various approaches for constructing the sentiment variable. They examined the use of different sentiment dictionaries, specifically Harvard psychological dictionary and Loughran–McDonald financial sentiment dictionary, SenticNet (a publicly available lexical resource for sentiment analysis), and a bag-of-words approach. Their findings indicated that models incorporating sentiment analysis outperformed the bag-of-words approach. Additionally, they found that utilizing polarity (positive or negative sentiment) alone did not yield useful predictions. Furthermore, the study revealed only minor differences between the two sentiment dictionaries employed. Heston and Sinha, (2017) introduced a different approach, rather than simply utilizing polarity, they incorporated both individual news sentiment and aggregate weekly news sentiment in their study. They employed Thomson Reuters neural network to measure the sentiment of over 900,000 news articles. The primary finding of their research suggests that daily news sentiment is effective in producing short-term price predictions, particularly for a period of one or two days. On the other hand, aggregated weekly news sentiment has the potential to predict stock returns over a longer period of one quarter. Additionally, the study revealed that positive news has an immediate impact on stock prices, while negative news tends to result in a delayed reaction.

Xu *et al.*, (2022) conducted a study focused on exploring sentiment within managers' news reports in the Chinese market. They created a managers' sentiment variable based on more than 700 media reports generated by managers. This variable was then incorporated into a bivariate regression model to forecast returns. The findings of the study indicate that manager sentiment is a negative predictor of future market returns. Moreover, the research demonstrates that the manager sentiment variable contains additional and valuable information that can enhance the accuracy of predictions. Notably, the study highlights that the predictive power of manager sentiment is significantly higher during periods characterized by higher sentiment levels. In the same year, Balas *et al.*, (2022) employed a natural language processing approach to develop a classification model for predicting market behaviour, focusing on news headlines rather than entire articles. Unlike the previously mentioned works, their objective was to classify market behaviour as bullish or bearish. In their study, three different approaches were utilized to process the headlines: support vector machines, random forests, and Bidirectional Encoder Representations from Transformers (BERT). They concluded that the BERT model was able to yield the most promising results, achieving accuracy of 86.25% in classifying the market behaviour.

3.3 Social media and stock prices

One of the early works exploring the use of sentiment expressed by internet users to predict stock market behaviour was conducted by Tumarkin and Whitelaw (2001). They examined the relationship between information shared on financial forums and stock prices. The findings of their research indicate that the majority of the information found in forum postings is noise rather than valuable news. Additionally, the study revealed no predictive power associated with the volume of messages posted on these forums. However, the authors concluded that a reverse relationship exists, and that market information influences the activity of forum users. Chen *et al.*, (2014) conducted a study investigating the impact of opinions expressed on a popular social media platform, specifically designed for investors, on market prices. Their research findings suggest that the information conveyed in both comments and articles on this platform holds a significant predictive power concerning subsequent stock returns and earnings surprises. Notably, they demonstrate that this predictive power remains robust even when accounting for the effects of other information sources, such as financial news.

One of the first studies examining the relationship between Twitter sentiment and securities' returns was conducted by Tayal and Satya (2009). The primary objective of their research was to differentiate the predictive effects of sentiment expressed on micro-blogs, particularly Twitter, compared to blogs. They collected sentiment data from both Twitter and various relevant blogs. Their results consistently indicated that Twitter outperformed blogs in terms of predictive capability. Following that, Zhang *et al.*, (2011) sought to predict stock market indicators through Twitter post analysis. In contrast to solely using positive or negative sentiment, they extracted a broader range of emotions from the tweets. Their findings suggest that periods of high emotional outburst on Twitter might be followed by a subsequent decline in stock market indices. Bollen *et al.* (2011) also focused on sentiment expressed by Twitter users. They examined the correlation and predictive power of public mood, derived from a large collection of Twitter posts, on the closing price of the Dow Jones Industrial Average (DJIA). They measured six distinct emotions (Calm, Alert, Sure, Vital, Kind, and Happy) as well as positive or negative mood. To develop their model, they employed a combination of Granger analysis and neural networks. The research findings indicate that the positive or negative mood does not have a significant impact on the DJIA. However, they discovered that the emotion of calmness exhibited a notable effect and incorporating it into the model improved the accuracy of predictions. Consequently, they conclude that public calmness holds predictive power for the DJIA, rather than the overall positivity or negativity of public sentiment.

Rao and Srivastava, (2012) examined the relationship between Twitter sentiment and volatility, trading volume and stock prices. They collected and analysed over 4 million tweets between the years 2010 and 2011. The researchers utilized positive and negative tweet classification to create variables representing Bullishness and Agreement, and they incorporated the volume of tweets into their model. Their findings revealed a strong relationship between positive or negative sentiment expressed on Twitter and price movements of individual stocks or indices. Si *et al.* (2013) extended the sentiment approach by leveraging topic-based sentiment extracted from Twitter to create predictions for the stock market. The researchers employed Dirichlet Process Mixture to identify the daily set of topics discussed on Twitter. They then derived sentiment scores for each of these topics. By performing a regression analysis between these sentiment values and the stock index, they generated their predictions. This approach showed better predictive power compared to methods that did not incorporate topic information.

Smailović *et al.*, (2014) developed and applied an active learning approach to Twitter sentiment analysis. They implemented a support vector machine algorithm to classify continually steamed tweets as positive, negative, or neutral. The addition of neutral classifier improved the predictive power of the algorithm by increasing the correlation between sentiment value and the close price. Authors conclude that by augmenting a trading strategy with consideration of changes in sentiment value, one could improve the returns. (Kordonis *et al.*, 2016) then implement similar approach, but target specific popular tech firms. They conclude that changes in public sentiment gathered from Twitter do help predicting specific stock changes.

As the pandemic of COVID-19 broke out and heavily influenced the world economy, a work utilizing the sentiment analysis to explain market movement during such time periods was introduced by Valle-Cruz *et al.*, (2022). It was focused specifically on the effects of Twitter polarity on behaviour of financial indices during pandemics. They used a lexicon-based approach to compare the market reactions to sentiment during COVID-19 and H1N1 pandemics. The authors note that social media propagation, and therefore increased number of Twitter accounts, over the 11 years between the two pandemics plays a significant role in explaining the indices behaviour. They conclude that the stock price drops during recent years have been more dramatic due to there being more speculation, rumours and negative news. Other COVID-19 related research was conducted by Zammarchi *et al.*, (2023), who aimed to investigate the link between changes in opinions expressed on Twitter regarding Italy following the pandemic outbreak, as it was one of the first countries in Europe to be severely affected and impose a lockdown, and the subsequent decrease in values of FTSE-MIB index, the main Italian Stock index. The authors highlight the potential utility of using Twitter sentiment towards a specific country as a proxy for the perceived reputation of that country. Additionally, their findings suggest that shifts in sentiment scores can serve as an early indicator to detect changes in stock values.

4 Data

This section will focus on describing the process of gathering, cleaning, transforming, and aggregating the data. Summary statistics and plots will be utilized to provide information about the data structure.

For the purposes of this thesis, eight different publicly traded companies have been selected to gather financial and sentiment data. Half of these companies represent large and popular companies that are not directly in the video game business. Specifically, Toyota, Tesla, Amazon, and Apple have been chosen. While Toyota and Tesla are not involved in the video game market, Apple provides a mobile gaming platform and Amazon has a video game division and operates a video game streaming platform. However, these gaming attributes of Apple and Amazon are considered minimal enough to still classify them as non-gaming firms for the purposes of this research. It is important to note this distinction in case there is a significant difference in the price behaviour of their stocks.

The other half of the selected companies represents firms for which video games are the main form of business. Nintendo and Activision have been chosen to represent the largest players in the field, while Ubisoft and CD Projekt Red have been selected as representatives of semi-large gaming businesses. Companies that have a large gaming division but also do lucrative business in other fields, such as Microsoft or Sony, are not considered in this work. However, their inclusion could provide valuable additional insights in future research.

4.1 Twitter

Twitter is a popular online micro-blogging platform with a large user base and a significant volume of daily activity. As of 2022, Twitter has over 368 million monthly active users, and more than 500 million tweets are sent every day. Users on Twitter can post short messages, originally limited to 140 characters but expanded to 280 characters in 2017. This limit is rarely met as only 1% of tweets actually reach the character limit of 280, and with only 5% of tweets being longer than 190 characters. (Ruby, 2023) The character limitation on Twitter has made it a preferred platform for sharing opinions, emotions, and information. Users can engage in discussions and conversations, share their thoughts, and express their sentiments within the constraint of the tweet length, while companies and influencers can share information or promotional material. This made Twitter a great channel for participation marketing, as mentioned in section 1.2.2.

The availability of Twitter's Application Programming Interface (API) allows users to search for specific tweets using various parameters, such as language, location, or keywords. This feature along with the short nature of tweets makes Twitter a valuable source for gathering public opinions, trends, news, and sentiment.

4.1.1 Posts scrape

To gather Twitter data, a free developer Twitter account was utilized, which has a limitation of sampling tweets from the past 7 days. The data collection process employed the R package called "rtweet," developed by Kearney, (2019). This package provides functions to access Twitter's REST API and search for past tweets using specific keywords. A script was developed using this approach, which downloaded 2000 tweets per day for each of the selected companies.

The choice of collecting 2000 tweets daily was made for convenience, as the free API has a limit of gathering no more than 16 thousand tweets simultaneously, and the data was being collected for 8 separate companies. Only non-retweeted posts written in English that contained at least one of the relevant keywords were collected. The list of used keywords can be found in appendices.

In contrast to other previous works, this research collected all tweets containing the specified keywords, rather than solely focusing on those from financial media or mentioning finance. The goal was to capture general public opinions, not just opinions specifically related to the companies' finances. The approach was chosen because this thesis aims to investigate whether the high emotionality of gamers could have a significant influence on prices, even if the tweets were not explicitly concerned with stock prices.

The script was executed daily between October 20th, 2022 and March 10th, 2023. In total, over 1.7 million tweets were collected (the number of collected tweets for CDPR and Ubisoft was significantly lower due to smaller activity on Twitter regarding these companies) before undergoing the cleaning process. During the cleaning phase, duplicate tweets and those containing HTTPS links, which could be considered promotional, were removed from the dataset. This cleaning process aimed to ensure the quality and relevance of the collected tweets. During this initial cleanup, more than a half of tweets was removed from the dataset. This indicated that a significant number of collected posts were duplicates, which could be attributed to promotional campaigns encouraging followers to repost the tweet instead of simply retweeting it. Additionally, many tweets contained links, typically leading to a store or content creator's profile.

4.1.2 Sentiment analysis

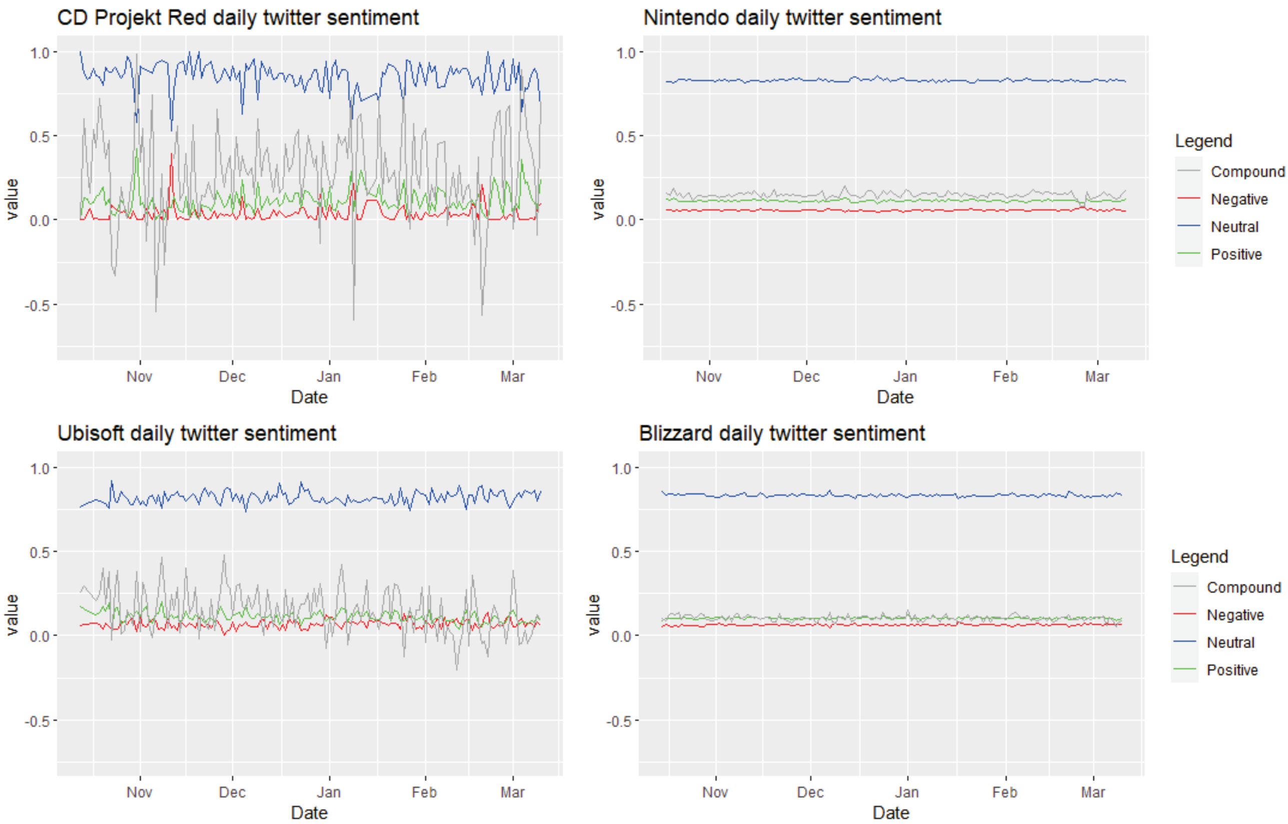
In order to assign polarity scores to individual tweets, a lexicon and rule-based sentiment analysis tool called VADER (Valence Aware Dictionary and sEntiment Reasoner), developed by Hutto and Gilbert, (2014) was utilized. They constructed and empirically validated a list of lexical features and assigned them a sentiment measure using a combination of quantitative and qualitative methods. In addition to this, they incorporated generalizable rules for grammatical and syntactical conventions into the system.

They described five heuristics to convey changes in the intensity of sentiment, which significantly improved the results compared to a typical bag-of-words model. These five rules consider word-order sensitive relationships between individual terms. The first rule accounts for the increased intensity of expressed sentiment by the inclusion of an exclamation point. The second rule considers capitalization to emphasize a word and increase its sentiment intensity. The third rule incorporates degree modifiers, such as adverbs or so-called booster words, which influence the sentiment value. For example, *"good"* is less intensive than *"really good."* The fourth rule includes a mechanism to distinguish contrastive conjunctions like *"but"* and adjust the sentiment accordingly. Lastly, the fifth rule captures a sentiment change caused by a negation flip word that precedes a polarity word by several other words.

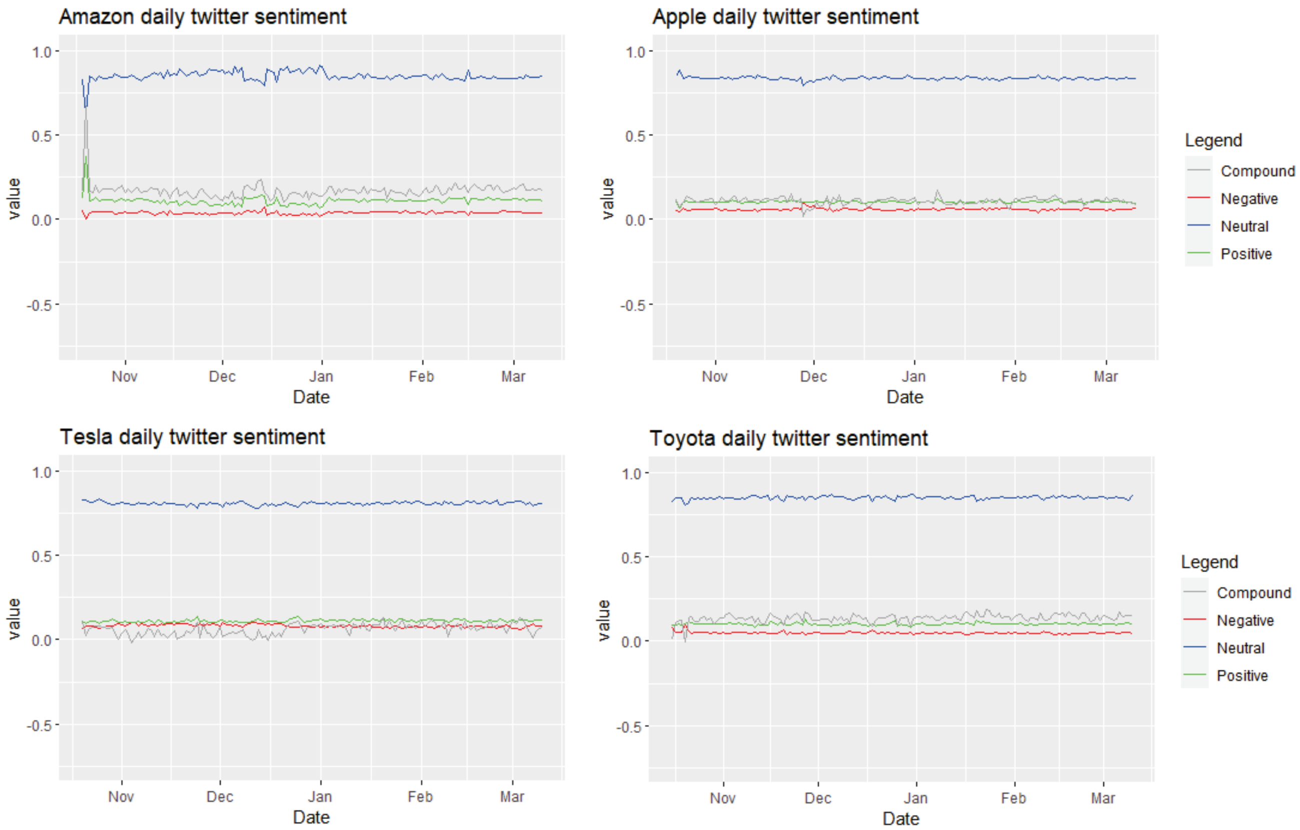
VADER is specifically attuned to the usage of micro-blog texts, and the authors incorporated additional lexical features commonly appearing in such environments. These features include a full list of emoticons, sentiment-conveying acronyms, and slang words used by internet users. VADER is capable of distinguishing and assigning sentiment values to these expressions, thus eliminating the need for standard text cleaning processes such as lowercasing, stop words removal, special characters removal, or tokenization.

Using the VADER tool, separate sentiment values were assigned to each individual tweet. These values represented positive, negative, neutral, and compound sentiment. The compound variable is a normalized, weighted composite sentiment score that will be utilized in the analysis. The compound score can be used to categorize sentiment as positive, negative, or neutral. Values above 0.05 are considered as positive sentiment, values below -0.05 are considered as negative sentiment, and values between -0.05 and 0.05 as neutral sentiment.

The computed sentiment values were then aggregated into daily sentiment time series for each company separately, employing two different approaches. The first aggregation approach involved calculating the mean of the sentiment values. The second approach utilized a weighted mean, where the weights were determined by the number of retweets for each collected tweet. The plots below display the mean daily sentiment scores for each of the researched companies.

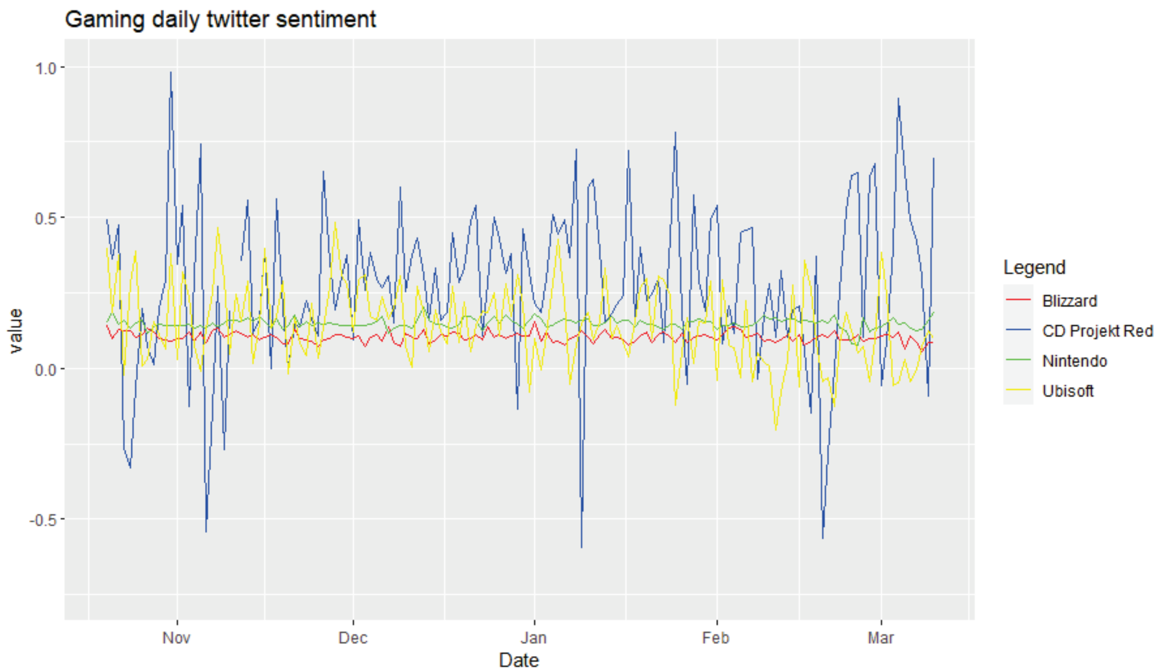


Plot 4.1: mean daily Twitter sentiment for gaming companies

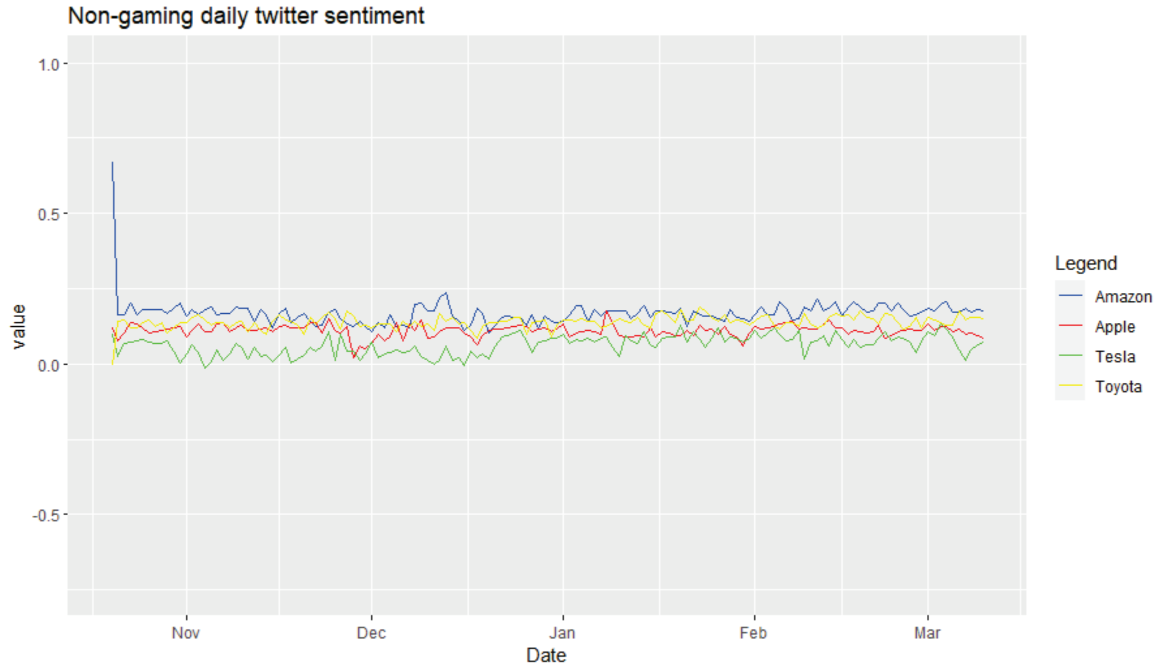


Plot 4.2: mean daily Twitter sentiment for non-gaming companies

These plots show the mean compound sentiment for gaming firms and non-gaming firms, respectively.



Plot 4.3: twitter compound sentiment for gaming firms



Plot 4.4: twitter compound sentiment for non-gaming firms

This table shows summary statistics of the daily compound sentiment time series for each of the selected firms:

Gaming:	**Activision**		**CD Projekt RED**		**Nintendo**		**Ubisoft**	
	mean	**w. mean**	**mean**	**w. mean**	**mean**	**w. mean**	**mean**	**w. mean**
min	0.054	-0.098	-0.84	-840.2	0.078	-0.003	-0.204	-15.66
1st Qu.	0.093	0.066	0.149	14	0.139	0.072	0.0526	2
median	0.103	0.098	0.291	63.6	0.149	0.087	0.148	7
mean	0.104	0.154	0.281	129.2	0.149	0.101	0.148	9
3rd Qu.	0.113	0.141	0.47	212.7	0.158	0.102	0.253	13
max	0.153	5.19	0.98	979.7	0.201	1.8	0.482	62.03

Non-gaming:	**Amazon**		**Apple**		**Tesla**		**Toyota**	
	mean	**w. mean**	**mean**	**w. mean**	**mean**	**w. mean**	**mean**	**w. mean**
min	0.104	0.021	0.019	-0.113	-0.015	-0.189	-0.005	-0.779
1st Qu.	0.149	0.079	0.098	0.049	0.038	0.004	0.127	0.080
median	0.168	0.101	0.111	0.069	0.069	0.025	0.139	0.101
mean	0.17	5.25	0.110	0.077	0.063	0.030	0.138	0.089
3rd Qu.	0.183	0.123	0.121	0.087	0.085	0.044	0.152	0.121
max	0.67	670	0.176	0.61	0.132	1.07	0.189	0.196

Note: in this table, all weighted mean values are multiplied by 1000 to improve readability

Table 4.1: summary statistics for the twitter daily sentiment values

As can be observed from the table 3.1, the weighted mean values are generally close to zero. This is likely because the most frequently retweeted posts often have promotional or informational nature and tend to have a neutral sentiment. Therefore, for our analysis, the simple mean approach appears to be more useful. Furthermore, both the plots and the table indicate that companies with a larger number of collected tweets tend to have the overall mean sentiment scores closer to zero compared to companies with fewer observations. This could be attributed to the overall sentiment on Twitter being closer to neutral.

Since the price data is not available for the weekend days, but the sentiment variable is, a data manipulation that would allow us to capture the effect of sentiment over the weekend is necessary. In order to do so, we merge weekend values into a single observation by taking the mean of the compound sentiment scores for Saturday and Sunday. The calculations for the weekend returns will be discussed in the financial data section.

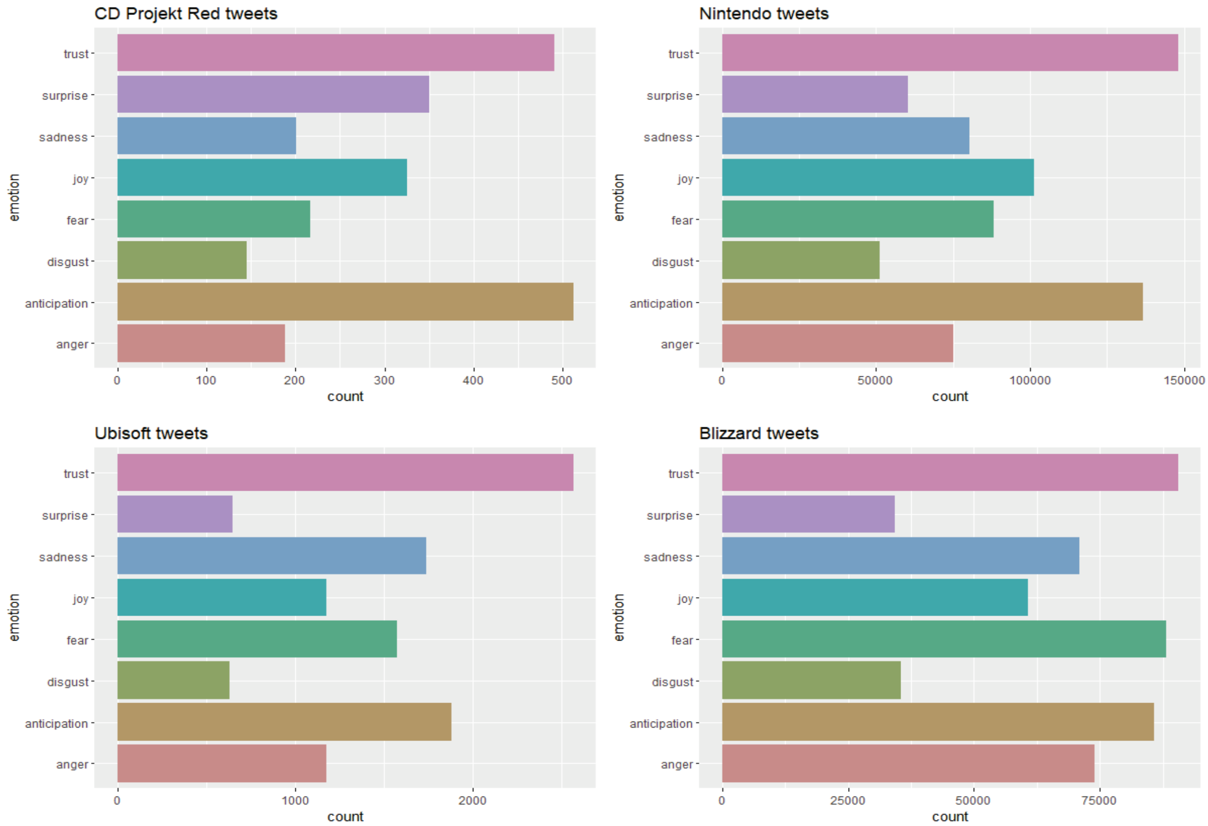
4.1.3 Emotion classification

As mentioned in the literature review, expanding the classification of tweets into a broader range of emotions can enhance the results and provide insights into the relationships between sentiment and stock market movements. In addition to using VADER to assign polarity values to the tweets, each tweet was also categorized based on the specific expressed emotion.

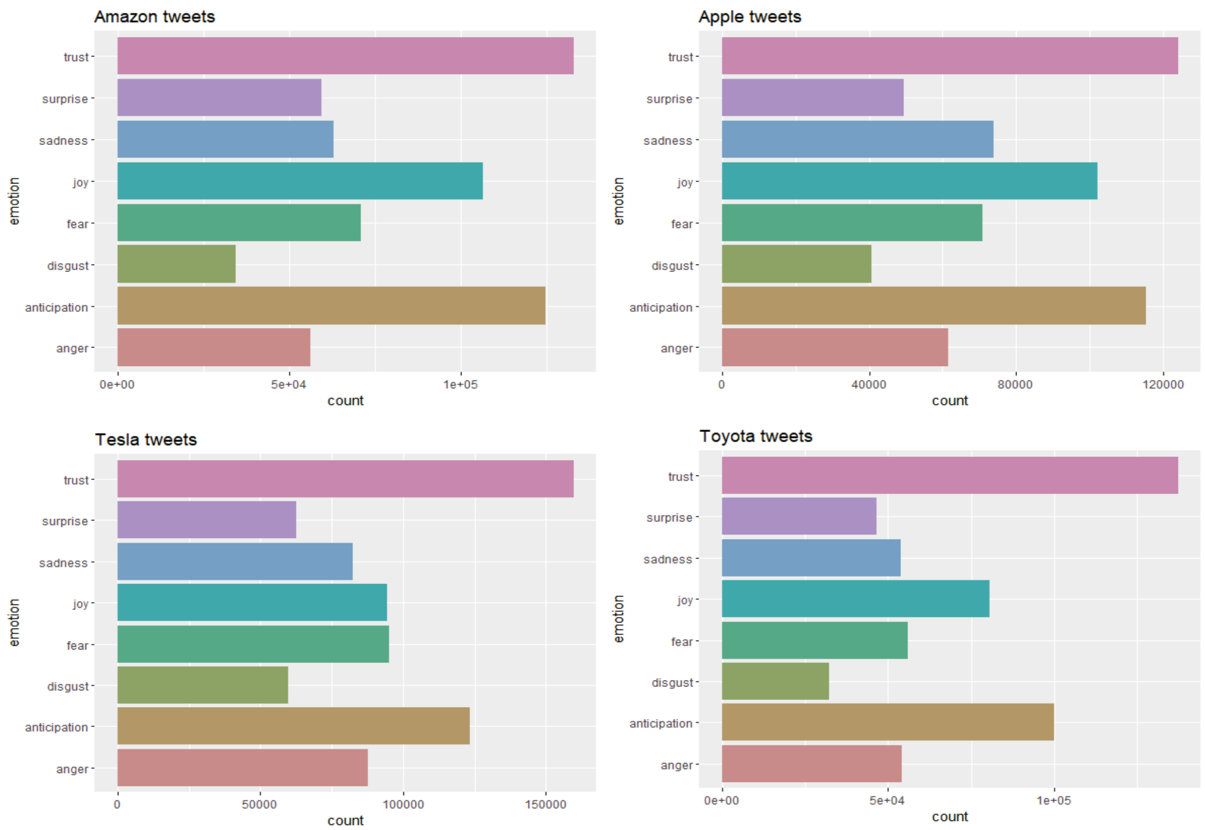
To classify the tweets, the NRC word-emotion classification lexicon included in the Syuzhet R package, developed by Jockers (2015) in 2015 but still actively maintained, was utilized. Despite the package's title, "Extracts Sentiment and Sentiment-Derived Plot Arcs from Text," which suggests a requirement for text with a plot, it has been successfully used in previous research to evaluate emotions in relation to stock market performance by Zammarchi *et al.*, (2023). The NRC lexicon assigns labels to words based on six possible emotions: "anger," "anticipation," "disgust," "fear," "joy," "sadness," "surprise," or "trust."

In contrast to VADER, the Syuzhet package does not have the ability to differentiate sentiment shifts caused by punctuation or the usage of capital letters. Consequently, prior to applying the NRC lexicon, several preprocessing steps were applied. These steps included converting the text to lowercase, removing non-ASCII symbols, and eliminating non-English words from the dataset.

Similar to the sentiment approach, each tweet was initially classified individually based on its expressed emotion. For the analysis, the mean and weighted mean scores were calculated for the tweets collected in a single day, with retweets serving as the weight. The plots below illustrate the overall distribution of emotions for each company. For this time-series, weekend values were merged into a single observation using mean, as well.



Plot 4.5: twitter emotion distribution for gaming firms



Plot 4.6: twitter emotion distribution for non-gaming firms

4.2 News headlines

Even though previous studies have primarily focused on gathering sentiment information from entire articles, there have been successful studies that specifically utilize news headlines, for example, Balas *et al.* (2022). News headlines can serve as a valuable source of both financial and business information. With the advancements in technology and widespread internet usage, accessing and collecting news headlines online has become simpler and more convenient. This accessibility enhances the efficiency of the data collection process.

4.2.1 Gathering news headlines

To gather relevant news headlines, the research utilized the Europe Media Monitor (EMM), an online service provided by the European Commission's Joint Research Centre (JRC). EMM is a software designed to automatically collect, analyze, and aggregate news articles from a wide range of online sources, including traditional and social media, in over 70 languages.

Specifically, the NewsBrief feature within EMM was employed. NewsBrief offers a concise and customizable overview of current news and trends, allowing users to quickly access selected information. It filters and categorizes news articles, catering to the needs of journalists, researchers, and policymakers. (EMM, 2018)

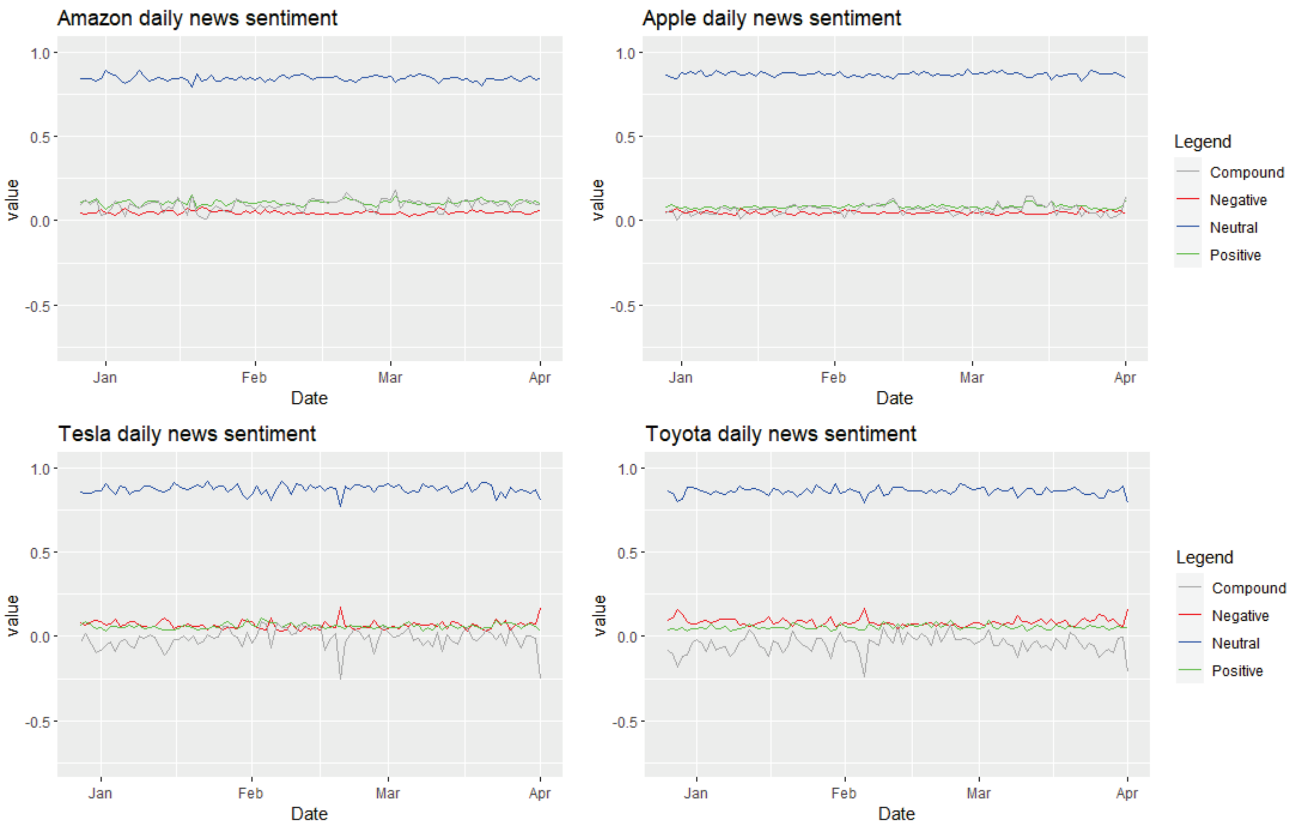
Using the specified keywords (same as for Twitter), a search was conducted on the NewsBrief webpage. A web scraping technique was then implemented to gather the headlines and compile them into a dataset. The headlines that included specified keywords and were published between January 1st and March 31st, 2023, were collected. The time difference between the Twitter and news datasets is caused by the maximum 3-month timeframe of NewsBrief's results display. In total, almost 300 thousand news headlines were collected and analysed.

4.2.2 Sentiment analysis

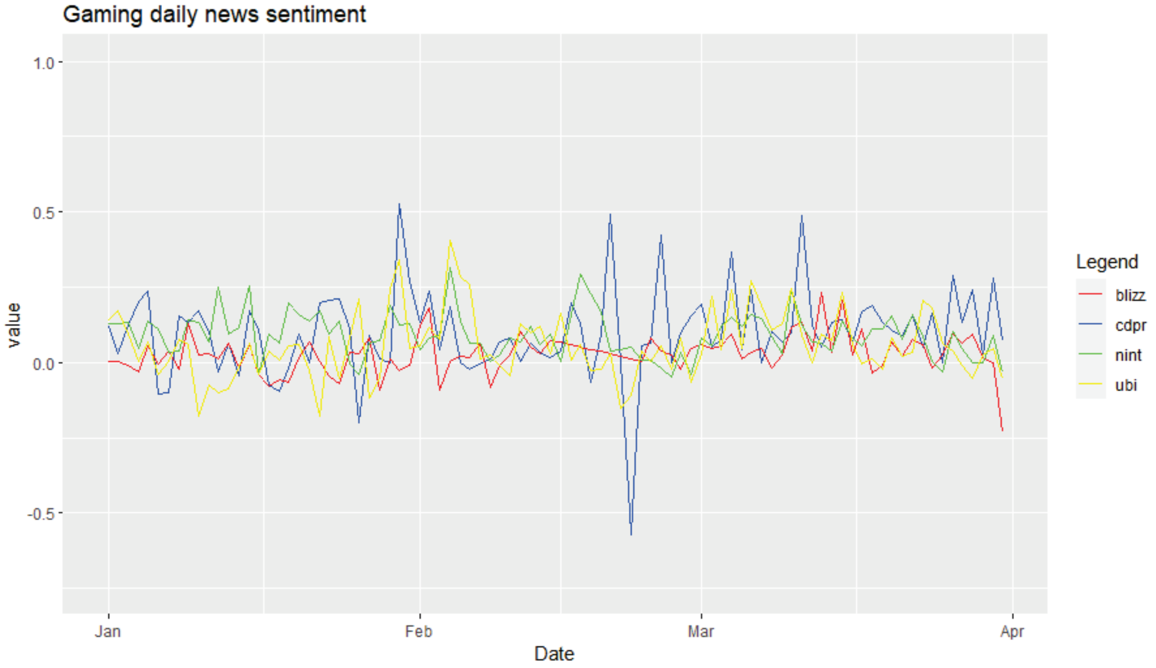
Even though VADER is specifically designed for sentiment analysis on social media, it has been proven effective when applied to other types of short textual documents, including news headlines. Therefore, VADER was utilized to compute sentiment values for the collected news headlines. The same approach as for tweets was used, with the exception that there is no weight variable, so only a simple mean aggregation was employed to calculate daily sentiment values. Same as for the tweets, the mean weekend values were computed for the news headlines.

The following plots display the daily sentiment values for each company, they are followed by a plot showing daily compound mean sentiment for gaming and non-gaming companies and a table containing summary statistics.

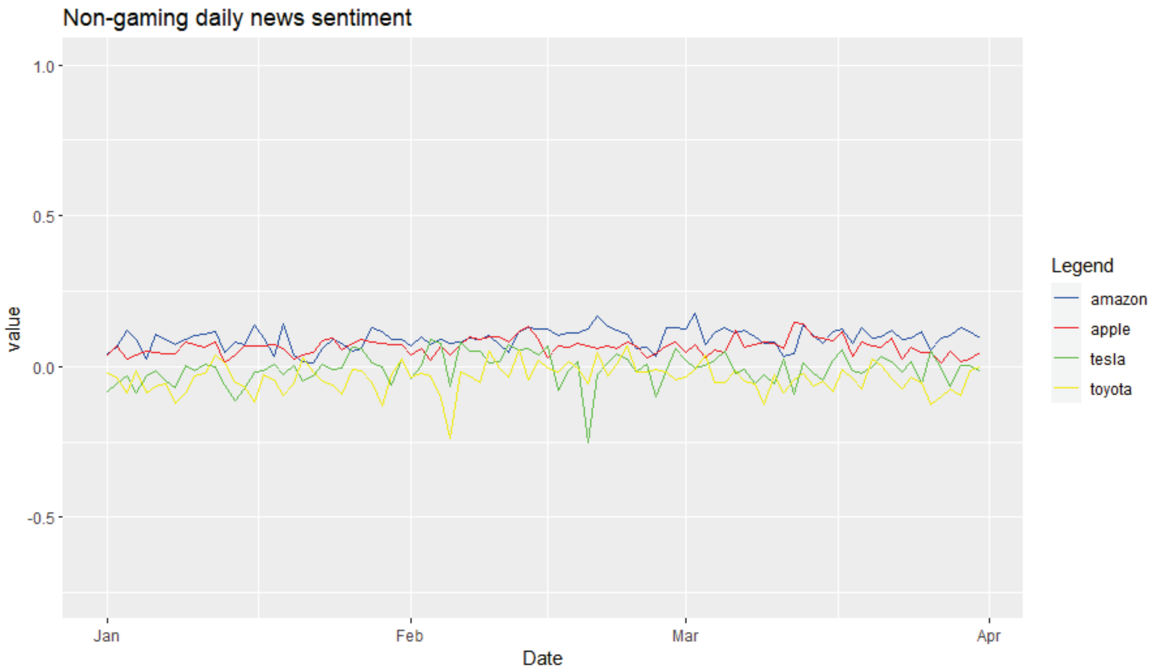
Plot 4.7: mean daily news sentiment for gaming companies



Plot 4.8: mean daily news sentiment for non-gaming companies



Plot 4.9: news compound sentiment for gaming firms



Plot 4.10: news compound sentiment for non-gaming firms

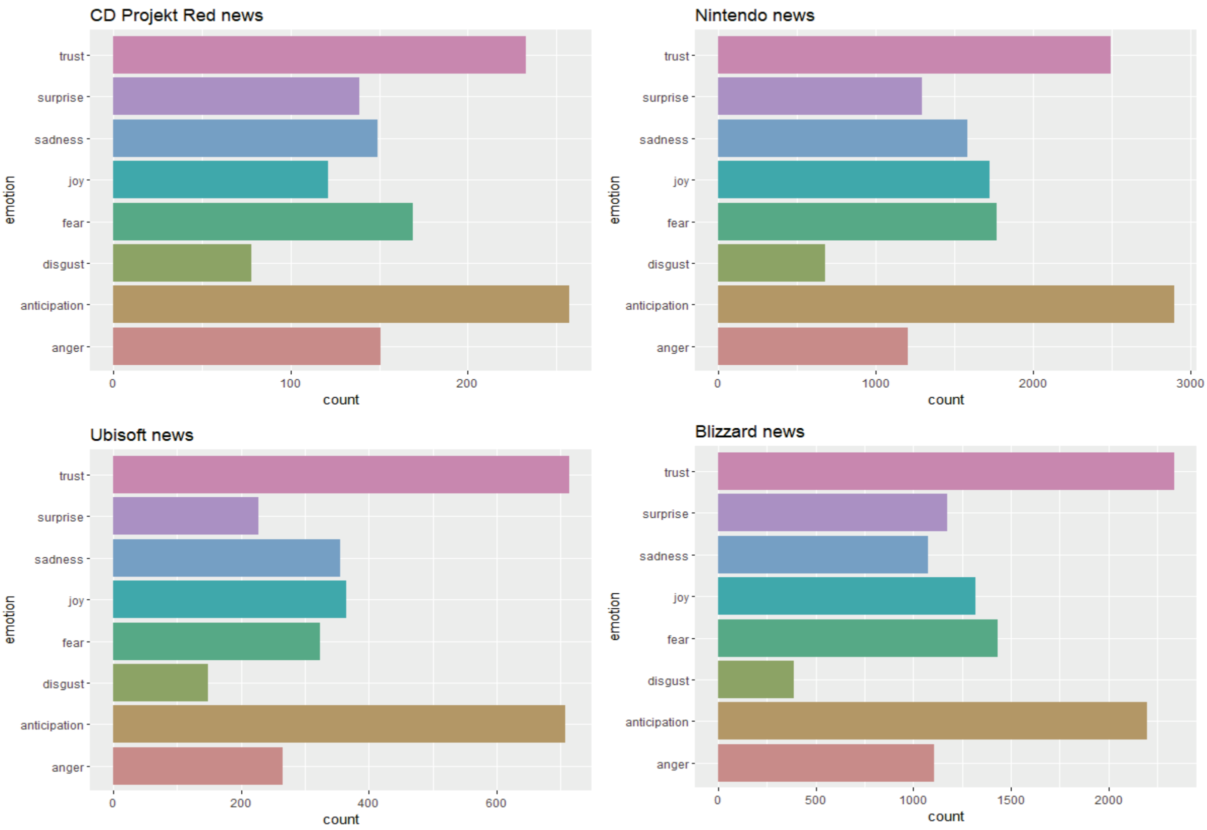
	Gaming:				Non-gaming:			
	Activision	**CDPR**	**Nintendo**	**Ubisoft**	**Amazon**	**Apple**	**Tesla**	**Toyota**
min	-0.340	-0.572	-0.049	-0.179	0.013	0.003	-0.254	-0.241
1st Qu.	-0.009	0.020	0.040	-0.007	0.074	0.046	-0.036	-0.074
median	0.025	0.095	0.082	0.047	0.094	0.067	-0.007	-0.038
mean	0.024	0.100	0.088	0.058	0.093	0.065	-0.012	-0.045
3rd Qu.	0.061	0.169	0.132	0.100	0.117	0.082	0.021	-0.014
max	0.232	0.526	0.314	0.406	0.178	0.146	0.090	0.069

Table 4.2: summary statistics for the mean news daily sentiment values

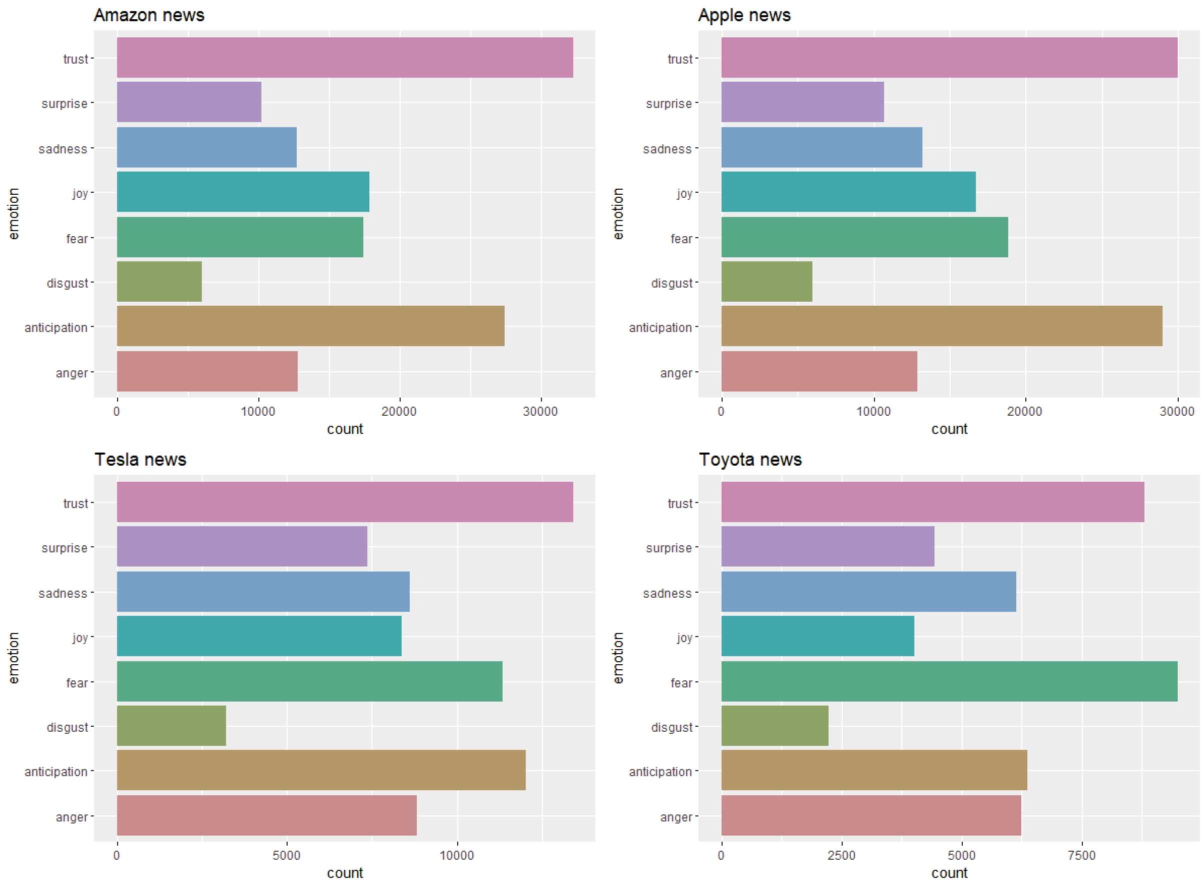
4.2.3 Emotion classification

To classify the emotions expressed in the headlines, the same approach as mentioned in the Twitter section was applied. However, since there is no weight variable for the headlines, the daily aggregate value was created using the simple mean, and the weekend value is a mean of Saturday and Sunday.

The following plots illustrate the distribution of emotions expressed in the headlines related to specific companies.



Plot 4.11: news emotion distribution for gaming firms



Plot 4.12: news emotion distribution for non-gaming firms

4.3 Financial data

The daily financial data for this thesis was primarily gathered from the Yahoo! Finance website for all companies except CD Projekt RED, for which the data was not available there. An alternative source, Investing.com, was used to obtain the CDPR data. Opening and closing prices for each company between January 1st, 2022 and March 31st, 2023 were collected. This time period was chosen to provide a comprehensive overview of price development and a solid basis for estimating volatility. Subsequently, only the relevant portion of the dataset was considered for each analysis and associated stationarity tests.

Daily log returns were computed for most days of the dataset (all except Mondays) using the formula:

$$r_t = log(Close_t) - log(Close_{t-1}) \tag{4.1}$$

However, to account for the effect of sentiment over the weekend, a different formula was used to calculate the over-the-weekend return:

$$r_{weekend} = log\left(Open_{Monday}\right) - log\left(Close_{Friday}\right) \tag{4.2}$$

To ensure consistency, the returns for Mondays were computed using the formula:

$$r_{Monday} = log\left(Close_{Monday}\right) - log\left(Open_{Monday}\right) \tag{4.3}$$

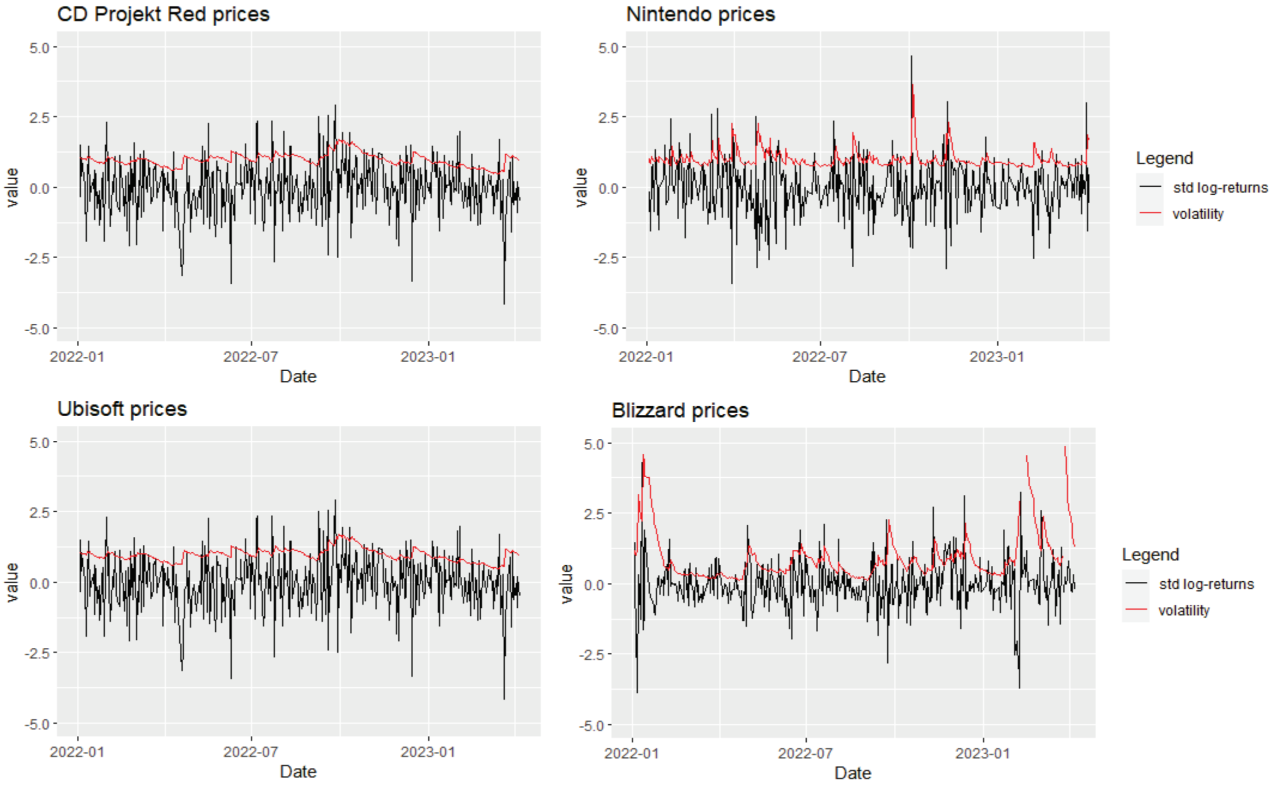
Using these formulas, a time-series of log returns was created and utilized to estimate daily volatility for each company using a GARCH model, which will be further explained in the methodology section. Prior to estimating the GARCH model, the log returns data was standardized to prevent computation errors caused by rounding low numbers.

The standardization:

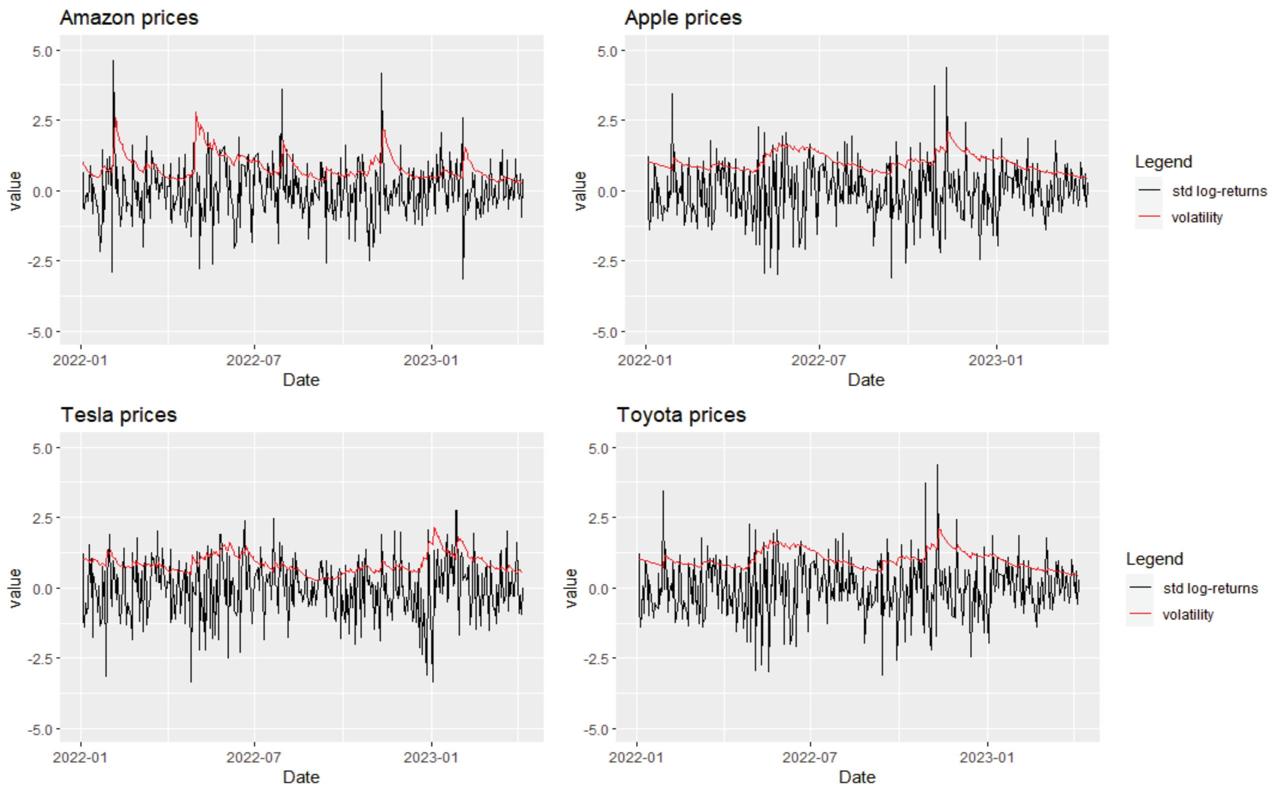
$$r_s = \frac{r_i - \mu}{\sigma} \tag{4.4}$$

Where r_s is the standardised log-return, r_i is non-standardised log-return, μ is the mean and σ is the standard deviation.

The following plots display the standardized log returns and estimated volatility for each of the researched companies.



Plot 4.13: gaming companies returns and volatility



Plot 4.14: non-gaming companies returns and volatility

The table below presents the summary statistics for standardized log returns and estimated volatility.

Gaming:	**Activision**		**CD Projekt RED**		**Nintendo**		**Ubisoft**	
	logrets	**volatility**	**logrets**	**volatility**	**logrets**	**volatility**	**logrets**	**volatility**
min	-3.877	0.125	-4.159	0.468	-3.414	0.729	-6.407	0.779
1st Qu.	-0.505	0.384	-0.571	0.850	-0.538	0.822	-0.442	0.810
median	-0.091	0.691	0.030	0.965	0.040	0.901	0.021	0.844
mean	0.000	1.012	0.000	0.976	0.000	1.000	0.000	0.898
3rd Qu.	0.380	1.094	0.601	1.091	0.583	1.046	0.436	0.910
max	5.807	8.416	2.902	1.703	4.670	3.645	3.878	3.049

Non-gaming:	**Amazon**		**Apple**		**Tesla**		**Toyota**	
	logrets	**volatility**	**logrets**	**volatility**	**logrets**	**volatility**	**logrets**	**volatility**
min	-5.387	0.294	-3.080	0.440	-3.343	0.270	-3.912	0.480
1st Qu.	-0.525	0.501	-0.578	0.776	-0.608	0.628	-0.572	0.733
median	0.000	0.690	-0.017	0.935	0.072	0.783	-0.041	0.913
mean	0.000	0.827	0.000	1.001	0.000	0.850	0.000	0.952
3rd Qu.	0.577	1.014	0.615	1.170	0.631	1.040	0.534	1.219
max	4.605	2.781	4.366	2.102	2.763	2.133	4.628	1.575

Table 4.3: summary statistics for standardised logrets and volatility

For the analysis, the non-standardized log returns and the volatility estimated from the standardized log returns were used.

5 Methodology

5.1 GARCH

When dealing with financial data, their specific features need to be considered. One of these features is volatility clustering, which refers to the occurrence of small and large price changes in clusters. Another feature is the leverage effect, which implies that volatility tends to be higher after negative price shocks compared to positive ones. Additionally, financial data typically follow a leptokurtic distribution, meaning they have heavier tails than a normal distribution. This indicates that extreme values are less likely, and there are more values around the mean. When estimating volatility of financial data, a model that can capture all these features should be chosen.

To estimate volatility in financial data, the General Autoregressive Conditional Heteroskedasticity (GARCH) model, introduced by Bollerslev, (1986) was utilized. It is a generalization of ARCH (q) model, which does not consider past volatility and therefore is not able to capture the effects of volatility clustering mentioned above. Furthermore, it assumes that both positive and negative shocks have the same effect on volatility which fails to consider the leverage effect. The GARCH model is the most common one to use when dealing with financial time series. The model's conditional variance is dependent on both the sign and magnitude of the returns.

The following equations shows the specification of GARCH (1,1) model:

$$a_t = \sigma_t \, \varepsilon_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{5.1}$$

Where a_t is mean corrected return $a_t = r_t - \mu_t$, σ_t represents volatility, ε_t is a sequence of independent and identically distributed random variables (i.i.d.) and $\varepsilon_t \sim N\,(0,1)$. Furthermore, $\alpha_0 > 0,\ \alpha_1 \geq 0,\ \beta_1 \geq 0$ and $(\alpha_1 + \beta_1) < 1$ hold. Stationarity is assumed for the series of a_t and σ_t.

GARCH (1, 1) model is the most frequently used form of a more widely specified GARCH (p, q) model:

$$a_t = \sigma_t \, \varepsilon_t$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i a_{t-i}^2 + \sum_{j=1}^{q} \beta_i \sigma_{t-i}^2 \tag{5.2}$$

Where $\alpha_i > 0,\ \alpha_i \geq 0,\ \beta_i \geq 0$ and $\sum_{i=1}^{\max(p,q)} \alpha_i + \beta_i < 1$ has to hold. The unconditional variance of a_t is finite and conditional variance σ_t^2 changes over time. In case of $q = 0$, the GARCH (p, q) is reduced to the previously mentioned ARCH (q).

In this thesis, the volatility was estimated by constructing the GARCH (1, 1) model using the standardised log-returns of each company separately and then fitting the model to the time series to get an approximation of the conditional volatility.

5.2 Stationarity tests

In order to estimate the models used in the thesis, namely the GARCH and vector autoregression models (introduced in the next section), it is necessary to assume stationarity. Stationarity implies that the mean and variance of a time series do not change over time. To assess the stationarity of the analysed time series, the Augmented Dickey-Fuller (ADF) test was employed.

To perform the ADF test, the following equation is estimated:

$$\Delta y_t = \alpha + \beta t + \theta y_{t-1} + \sum_{i=1}^{k}(\delta_i \Delta y_{t-i}) + e_t \tag{5.3}$$

Where y represents the tested variable, α is a constant, β is the coefficient of time trend, k is the lag order of the autoregressive process and e is and i.i.d. residual term. If the coefficients of α and β would be set to zero, this equation corresponds to modelling random walk.

After estimating the above equation, following test is conducted:

$$H_0: \theta = 0, \quad H_1: \theta < 0 \tag{5.4}$$

A test statistic is computed as:

$$DF = \frac{\hat{\theta}}{SE(\hat{\theta})} \tag{5.5}$$

Where $\hat{\theta}$ is the estimated coefficient of the lagged value of y and $SE(\hat{\theta})$ is the standard error of the coefficient. The test statistic is then compared to the critical values of the Dickey-Fuller distribution. (Wooldridge, 2020) The table of test results can be seen in the appendix.

5.3 VARs and Granger causality

The Granger causality test introduced by Granger, (1969) can be used to investigate the predictive power and significance of selected variables' lags. In order to perform the Granger causality analysis a Vector Autoregressive (VAR) model first needs to be estimated. VAR models are used to capture the relationship between multiple variables and their lagged values. The equations in VAR analysis consider not only the lags of the explained variable but also incorporate the lagged values of other variables in the system.

The following equations shows the form of bivariate VAR:

$$\begin{aligned} y_t &= c_1 + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{i=1}^{p} \beta_i x_{t-i} + \varepsilon_t \\ x_t &= c_2 + \sum_{i=1}^{p} \gamma_i x_{t-i} + \sum_{i=1}^{p} \delta_i y_{t-i} + \varepsilon_t \end{aligned} \tag{5.6}$$

Where c_1, c_2 represent the constant terms of the equations, y_t represents the stock's variables, x_t the sentiment variables, p is the number of lags and ε_t is the error term.

The extended form of previous model, multivariate VAR model is defined as:

$$\boldsymbol{y_t} = C + \sum_{i=1}^{p} A_i y_{t-i} + \varepsilon_t \tag{5.7}$$

Where y_t represents the vector of variables at time t, C is a vector of constant terms, A_i is a coefficient matrix and ε_t is the error term. (Lütkepohl, 2005)

For the purposes of the analysis, the bivariate VAR was used to estimate the effects of sentiment, while the multivariate VAR was applied to estimate the effects of various emotions expressed in the texts.

After the VARs model estimation, the Granger causality analysis was conducted. In order to perform the Granger causality analysis, restricted versions of the models above need to be constructed. The restricted versions of VAR model assume that the lagged values of the potential causal variables do not have a significant effect on the current variable of interest. Essentially, the restricted model imposes the assumption that there is no Granger causality between the variables. Therefore, to construct a restricted model, the potential causal variables are excluded from the model – their coefficients set to zero. The remaining variables are included in the restricted model.

Restricted model of the bivariate VAR:

$$\boldsymbol{y_t} = c_1 + \sum_{i=1}^{p} \alpha_i y_{t-i} + \varepsilon_t \tag{5.8}$$

The hypotheses of the bivariate Granger causality would therefore be:

$$H_0: \forall\, i, 1 \leq i \leq p: \beta_i = 0, \quad H_1: \exists\, i, 1 \leq i \leq p: \beta_i \neq 0 \tag{5.9}$$

In order to compare the fit between restricted and unrestricted VAR, the F-statistic is used:

$$F = \frac{(RSS_0 - RSS_1)/p}{RSS_1/(T-k)} \tag{5.10}$$

where RSS_0 and RSS_1 stand for the residual sum of squares of restricted and unrestricted model respectively. T is the number of observations, p the number of lags and k the number of variables in the VAR model. The F value is then compared to the selected critical value.

When using the Granger causality analysis, it is important to remember that the interpretation of causality is limited to statistical relationships and might not mean a true causal relationship in a deterministic sense. Therefore, caution is needed to interpret the results in the context of specific data and research questions.

6 Results

This chapter aims to present the empirical results obtained from the analyses conducted in the study. Hypotheses regarding this analysis will be presented and tested. The impact of both Twitter posts and news headlines sentiment variables on the security's returns and volatility will be discussed. Additionally, the influence of emotions expressed on Twitter and in news headlines on the returns will be presented and interpreted along with the Granger causality results. Subsequently, potential real-life explanations for the presented results will be provided to enhance the understanding of the findings. Furthermore, the chapter will discuss the limitations encountered during the study and suggest possible steps to improve the results.

6.1 Twitter sentiment

In this section the following hypotheses will be focused and tested:

1. Twitter sentiment expressed by general public has no impact on security returns.
2. Twitter sentiment expressed by general public has no impact on security volatility.

The following tables show the results of VARs analysis of logarithmic returns and twitter sentiment for specifically gaming companies. Two tables are presented, one using mean value for the sentiment, the other weighted mean as discussed in section 3.1.2.

VARs for: logrets ~ tweets mean sentiment, gaming firms

	Dependent variable:							
	blizz_logrets	**blizz_sent**	**cdpr_logrets**	**cdpr_sent**	**nint_logrets**	**nint_sent**	**ubi_logrets**	**ubi_sent**
logrets_1st_lag	-0.341***	0.064	-0.130	-0.871	-0.180	-0.163	-0.022	0.174
	(0.102)	(0.128)	(0.104)	(1.144)	(0.109)	(0.111)	(0.103)	(0.517)
sent_1st_lag	-0.040	0.026	0.002	0.095	0.108	0.354***	-0.011	0.103
	(0.084)	(0.106)	(0.010)	(0.105)	(0.103)	(0.104)	(0.020)	(0.100)
const	0.006	0.100***	-0.001	0.226***	-0.017	0.095***	-0.001	0.132***
	(0.009)	(0.011)	(0.003)	(0.038)	(0.015)	(0.015)	(0.004)	(0.020)
Observations	88	88	94	94	88	88	98	98
R^2	0.117	0.004	0.017	0.015	0.037	0.124	0.003	0.012
Adjusted R^2	0.097	-0.020	-0.004	-0.007	0.014	0.104	-0.018	-0.009
Residual Std. Error	0.012 (df = 85)	0.015 (df = 85)	0.023 (df = 91)	0.253 (df = 91)	0.014 (df = 85)	0.014 (df = 85)	0.027 (df = 95)	0.135 (df = 95)
F Statistic	5.657*** (df = 2; 85)	0.153 (df = 2; 85)	0.804 (df = 2; 91)	0.687 (df = 2; 91)	1.621 (df = 2; 85)	6.033*** (df = 2; 85)	0.158 (df = 2; 95)	0.572 (df = 2; 95)
Note:								*p<0.1; **p<0.05; ***p<0.01

Table 6.1: VAR results for mean twitter sentiment and logrets of gaming companies

VARs for: logrets ~ tweets weighted mean sentiment, gaming firms

	Dependent variable:							
	blizz_logrets	**blizz_sent**	**cdpr_logrets**	**cdpr_sent**	**nint_logrets**	**nint_sent**	**ubi_logrets**	**ubi_sent**
logrets_1st_lag	-0.331***	-0.0004	-0.129	-0.625	-0.200*	0.0002	-0.018	0.020
	(0.102)	(0.001)	(0.104)	(0.816)	(0.108)	(0.0002)	(0.104)	(0.042)
logrets_2nd_lag					-0.233**	-0.0002		
					(0.108)	(0.0002)		
sent_1st_lag	-14.459	0.066	-0.0001	0.059	24.053	-0.055	-0.090	0.073
	(17.359)	(0.107)	(0.013)	(0.102)	(46.715)	(0.107)	(0.253)	(0.102)
sent_2nd_lag					73.300	0.263**		
					(44.936)	(0.103)		
const	0.003	0.0001***	-0.001	0.092***	-0.011*	0.0001***	-0.001	0.008***
	(0.002)	(0.00001)	(0.003)	(0.021)	(0.006)	(0.00001)	(0.004)	(0.001)
Observations	88	88	94	94	87	87	98	98
R^2	0.122	0.008	0.017	0.010	0.097	0.092	0.002	0.008
Adjusted R^2	0.102	-0.016	-0.005	-0.012	0.053	0.047	-0.019	-0.013
Residual Std. Error	0.012 (df = 85)	0.0001 (df = 85)	0.023 (df = 91)	0.181 (df = 91)	0.014 (df = 82)	0.00003 (df = 82)	0.027 (df = 95)	0.011 (df = 95)
F Statistic	5.921*** (df = 2; 85)	0.324 (df = 2; 85)	0.776 (df = 2; 91)	0.441 (df = 2; 91)	2.192* (df = 4; 82)	2.070* (df = 4; 82)	0.081 (df = 2; 95)	0.383 (df = 2; 95)
Note:								*p<0.1; **p<0.05; ***p<0.01

Table 6.2: VAR results for weighted mean twitter sentiment and logrets of gaming companies

As shown in Tables 5.1 and 5.2, the VAR analysis does not yield any significant values that indicate that sentiment can influence returns for the examined firms. The only significant values observed in some cases pertain to the autoregressive process for log returns (log returns depend on their previous values). Additionally, some significant values suggest a possible reverse dependency, where returns might influence sentiment. However, since this is not the focus of this study, it is not further tested. Based on these results, we cannot reject the first hypothesis for any of the gaming firms.

Following tables show the same analysis for the non-gaming companies:

VARs for: logrets ~ tweets sentiment, non-gaming firms

	Dependent variable:							
	amazon_logrets	**amazon_sent**	**apple_logrets**	**apple_sent**	**tesla_logrets**	**tesla_sent**	**toyota_logrets**	**toyota_sent**
logrets_1st_lag	0.057	0.041	-0.094	-0.030	-0.008	0.011	-0.045	0.130
	(0.107)	(0.104)	(0.106)	(0.105)	(0.108)	(0.065)	(0.107)	(0.163)
sent_1st_lag	0.013	0.114^{**}	0.059	0.421^{***}	$\mathbf{0.282}^{*}$	0.479^{***}	-0.027	0.109
	(0.045)	(0.044)	(0.098)	(0.097)	**(0.159)**	(0.095)	(0.052)	(0.079)
const	-0.005	0.146^{***}	-0.008	0.063^{***}	-0.021^{*}	0.032^{***}	0.003	0.126^{***}
	(0.008)	(0.008)	(0.011)	(0.011)	(0.011)	(0.006)	(0.007)	(0.011)
Observations	91	91	91	91	91	91	91	91
R^2	0.004	0.075	0.012	0.176	0.036	0.238	0.006	0.030
Adjusted R^2	-0.018	0.054	-0.010	0.158	0.014	0.221	-0.017	0.008
Residual Std. Error	0.025	0.024	0.020	0.019	0.043	0.026	0.011	0.017
F Statistic	0.188	3.568^{**}	0.546	9.417^{***}	1.660	13.745^{***}	0.246	1.357

Note: *p<0.1; **p<0.05; ***p<0.01

Table 6.3: VAR results for mean twitter sentiment and logrets of non-gaming companies

VARs for: logrets ~ tweets weighted mean sentiment, non-gaming firms

	Dependent variable:							
	amazon_logrets	**amazon_sent**	**apple_logrets**	**apple_sent**	**tesla_logrets**	**tesla_sent**	**toyota_logrets**	**toyota_sent**
logrets_1st_lag	0.057	0.0001	-0.091	-0.0003	0.042	0.00001	-0.047	-0.0001
	(0.107)	(0.0001)	(0.106)	(0.0004)	(0.107)	(0.0002)	(0.107)	(0.0004)
sent_1st_lag	0.012	0.00001	6.518	-0.087	-2.599	0.083	-8.749	-0.090
	(0.037)	(0.00005)	(30.831)	(0.105)	(69.141)	(0.107)	(21.642)	(0.072)
const	-0.003	0.0001^{***}	-0.002	0.0001^{***}	-0.004	0.00003^{***}	0.0005	0.0001^{***}
	(0.003)	(0.00000)	(0.003)	(0.00001)	(0.005)	(0.00001)	(0.002)	(0.00001)
Observations	91	91	91	91	91	91	91	91
R^2	0.005	0.003	0.009	0.014	0.002	0.007	0.004	0.019
Adjusted R^2	-0.018	-0.020	-0.014	-0.009	-0.021	-0.015	-0.018	-0.003
Residual Std. Error	0.025	0.00003	0.020	0.0001	0.043	0.0001	0.011	0.00004
F Statistic	0.200	0.121	0.389	0.618	0.077	0.315	0.187	0.846

Note: *p<0.1; **p<0.05; ***p<0.01

Table 6.4: VAR results for weighted mean twitter sentiment and logrets of non-gaming companies

Granger causality tests for: logrets ~ tweets sentiment

F test and Wald χ^2 test based on VAR(1) model:

	F	df1	df2	p	Chisq	df	p
tesla_logrets <= tesla_com	3.16	1	88	.079	3.16	1	.075

Table 6.5: Granger causality test for tesla sentiment causing logrets

As shown in Tables 5.3 and 5.4, the results for non-gaming companies are similar to those of gaming ones. However, in Table 5.3, there is one value (shown in bold) that, at a 10% significance level, might indicate that Twitter sentiment concerning Tesla could influence its log returns. To further investigate this, a Granger causality test was conducted, and the results are presented in Table 5.5. Both the F-test and Wald χ^2 test show p-values lower than 0.1, allowing us to reject the first hypothesis specifically for Tesla at a 10% significance level.

To test the second hypothesis, the same models were run with volatility values instead of log returns. The volatility values were computed using standardized log returns and fitting a GARCH model, as described in Section 4.1. The following tables show the VAR results for gaming companies using both mean and weighted mean sentiment values. Any time series found to be non-stationary by the ADF (results of which can be seen in the appendix) as discussed in Section 4.2, was not considered for the analysis.

VARs for: volatility ~ tweets sentiment, gaming firms

	Dependent variable:			
	nint_volatility	**nint_sent**	**ubi_volatility**	**ubi_sent**
volatility_1st_lag	0.728***	0.012*	0.126	0.020
	(0.078)	(0.006)	(0.096)	(0.082)
volatility_2nd_lag			0.400***	-0.066
			(0.096)	(0.082)
sent_1st_lag	-1.080	0.280***	0.004	0.102
	(1.250)	(0.103)	(0.122)	(0.104)
sent_2nd_lag			-0.057	0.012
			(0.120)	(0.103)
const	0.415**	0.095***	0.423***	0.170*
	(0.184)	(0.015)	(0.111)	(0.095)
Observations	88	88	97	97
R^2	0.512	0.137	0.202	0.018
Adjusted R^2	0.500	0.116	0.168	-0.025
Residual Std. Error	0.170 (df = 85)	0.014 (df = 85)	0.160 (df = 92)	0.137 (df = 92)
F Statistic	44.579*** (df = 2; 85)	6.721*** (df = 2; 85)	5.833*** (df = 4; 92)	0.412 (df = 4; 92)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 6.6: VAR results for mean twitter sentiment and volatility of gaming companies

VARs for: volatility ~ tweets weighted mean sentiment, gaming firms

	Dependent variable:			
	nint_volatility	**nint_sent**	**ubi_volatility**	**ubi_sent**
volatility_1st_lag	0.710^{***}	0.0001^{***}	0.121	0.002
	(0.078)	(0.00001)	(0.095)	(0.007)
volatility_2nd_lag			0.401^{***}	-0.006
			(0.095)	(0.007)
sent_1st_lag	129.416	-0.111	-0.579	0.070
	(561.784)	(0.098)	(1.492)	(0.103)
sent_2nd_lag			-1.137	0.102
			(1.496)	(0.104)
const	0.261^{***}	0.00005^{***}	0.433^{***}	0.010
	(0.083)	(0.00001)	(0.109)	(0.008)
Observations	88	88	97	97
R^2	0.508	0.163	0.207	0.024
Adjusted R^2	0.496	0.143	0.173	-0.018
Residual Std. Error	0.171 (df = 85)	0.00003 (df = 85)	0.159 (df = 92)	0.011 (df = 92)
F Statistic	43.874^{***} (df = 2; 85)	8.263^{***} (df = 2; 85)	6.005^{***} (df = 4; 92)	0.577 (df = 4; 92)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 6.7: VAR results for weighted mean twitter sentiment and volatility of gaming companies

The results of VARs presented in Tables 5.6 and 5.7 demonstrate that volatility is not dependent on the sentiment expressed by general Twitter users. Similar to the results for log returns, an autoregressive process can be observed for volatility; however, no significant values indicate any influence of sentiment on volatility. Therefore, the second hypothesis cannot be rejected for any of the gaming firms.

Tables below show the same analysis for non-gaming firms. Note that a large portion of them was removed due to volatility being non-stationary in the selected time period.

VARs for: volatility ~ tweets sentiment, non-gaming firms		
	Dependent variable:	
	apple_volatility	**apple_sent**
volatility_1st_lag	0.966***	-0.0004
	(0.031)	(0.006)
sent_1st_lag	0.196	0.420***
	(0.514)	(0.097)
const	0.011	0.064***
	(0.068)	(0.013)
Observations	91	91
R^2	0.915	0.176
Adjusted R^2	0.914	0.157
Residual Std. Error	0.103	0.019
F Statistic	476.407***	9.369***
Note:		*p<0.1; **p<0.05; ***p<0.01

VARs for: volatility ~ tweets weighted mean sentiment, non-gaming firms		
	Dependent variable:	
	apple_volatility	**apple_sent**
volatility_1st_lag	0.966***	0.00000
	(0.031)	(0.00002)
sent_1st_lag	-42.471	-0.088
	(161.569)	(0.106)
const	0.037	0.0001***
	(0.038)	(0.00002)
Observations	91	91
R^2	0.915	0.008
Adjusted R^2	0.913	-0.015
Residual Std. Error	0.103	0.0001
F Statistic	475.961***	0.346
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 6.8: VAR results for mean and weighted mean twitter sentiment with volatility of non-gaming companies

Based on the results presented in Table 5.8, there is no evidence to indicate any relationship between Twitter sentiment expressed in general (not only financial) posts and the volatility of Apple stock. Additionally, the volatilities of other non-gaming companies were not stationary between October 20th and March 10th, which means that no conclusion can be drawn from them. As a result, there is insufficient evidence to reject the second hypothesis for any of the non-gaming firms.

The results and tests included in this section do not demonstrate any significant relationship between Twitter sentiment and market price movements of the given companies. This contrasts with the findings of some previous works, such as the study conducted by Rao and Srivastava, (2012), which suggests that social media sentiment does influence the market. However, it is important to note that there are some fundamental differences between this work and the aforementioned study. In this analysis, the focus is on Twitter posts concerning the firms in general terms, not just limited to financial posts. This might indicate that while sentiment can be useful for predicting financial market movements, it should primarily be gathered from sources specifically dealing with financial information related to the given company.

Furthermore, this work utilizes only daily aggregate data collected over a period of fewer than 5 months. A longer time period could potentially yield more robust results and might increase the likelihood of the volatility time series being stationary in more cases.

It is worth noting that no evident differences between gaming and non-gaming firms were observed in the analysis. This suggests that the impact of Twitter sentiment on market movements might not be significantly influenced by whether a company operates in the gaming sector or not.

6.2 News headlines sentiment

For this section, the hypotheses are analogous to those concerning Twitter:

1. Sentiment expressed in the news headlines has no impact on security returns.
2. Sentiment expressed in the news headlines has no impact on security volatility.

A comparable approach to the Twitter analysis was employed. However, as there is no weight variable for the headlines, only the mean is used to aggregate sentiment for each day. The results of VARs for both gaming and non-gaming firms' returns and news headlines sentiment are presented in the following tables:

VARs for: logrets ~ news sentiment, gaming firms

	Dependent variable:			
	cdpr_logrets	**cdpr_sent**	**nint_logrets**	**nint_sent**
logrets_1st_lag	-0.029	1.338*	0.049	0.607
	(0.130)	(0.734)	(0.139)	(0.807)
sent_1st_lag	-0.008	0.109	-0.002	0.143
	(0.022)	(0.126)	(0.023)	(0.136)
const	0.001	0.092***	-0.002	0.072***
	(0.004)	(0.022)	(0.002)	(0.014)
Observations	63	63	55	55
R^2	0.003	0.066	0.003	0.031
Adjusted R^2	-0.030	0.035	-0.036	-0.006
Residual Std. Error	0.026 (df = 60)	0.146 (df = 60)	0.011 (df = 52)	0.065 (df = 52)
F Statistic	0.085 (df = 2; 60)	2.125 (df = 2; 60)	0.065 (df = 2; 52)	0.828 (df = 2; 52)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 6.9: VAR results for mean news sentiment and logrets of gaming companies

VARs for: logrets ~ news sentiment, non-gaming firms

	Dependent variable:			
	amazon_logrets	**amazon_sent**	**tesla_logrets**	**tesla_sent**
logrets_1st_lag	0.154	0.092	0.112	0.107
	(0.133)	(0.201)	(0.149)	(0.142)
logrets_2nd_lag			-0.179	-0.171
			(0.141)	(0.134)
logrets_3rd_lag			0.044	0.284^{**}
			(0.129)	(0.123)
sent_1st_lag	-0.068	0.106	-0.050	0.138
	(0.088)	(0.134)	(0.150)	(0.142)
sent_2nd_lag			-0.108	0.072
			(0.140)	(0.134)
sent_3rd_lag			0.012	0.071
			(0.142)	(0.135)
const	0.009	0.087^{***}	0.008	0.0005
	(0.009)	(0.014)	(0.005)	(0.005)
Observations	58	58	56	56
R^2	0.032	0.016	0.070	0.188
Adjusted R^2	-0.003	-0.020	-0.044	0.089
Residual Std. Error	0.020 (df = 55)	0.030 (df = 55)	0.037 (df = 49)	0.035 (df = 49)
F Statistic	0.917 (df = 2; 55)	0.447 (df = 2; 55)	0.617 (df = 6; 49)	1.896 (df = 6; 49)
Note:				$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 6.10: VAR results for mean news sentiment and logrets of non-gaming companies

Tables 5.9 and 5.10 not show any evidence that the market returns of specific companies are dependent on the sentiment expressed in general (not just financial) news headlines concerning those companies, or the time series of their log-returns is not stationary. Similar to the results of tweets sentiment, there is insufficient evidence to reject the first hypothesis for either gaming or non-gaming firms.

The tables below present the same analysis with volatility instead of log-returns:

VARs for: volatility ~ news sentiment, gaming firms

	Dependent variable:	
	nint_volatility	**nint_sent**
volatility_1st_lag	0.615***	-0.056
	(0.111)	(0.051)
sent_1st_lag	-0.319	0.085
	(0.314)	(0.145)
const	0.363***	0.125**
	(0.111)	(0.051)
Observations	55	55
R^2	0.440	0.043
Adjusted R^2	0.419	0.006
Residual Std. Error	0.139	0.064
F Statistic	20.465***	1.157
Note:	*p<0.1; **p<0.05; ***p<0.01	

VARs for: volatility ~ news sentiment, non-gaming firms

	Dependent variable:	
	tesla_volatility	**tesla_sent**
volatility_1st_lag	0.851***	-0.088***
	(0.115)	(0.028)
volatility_2nd_lag	0.062	0.092***
	(0.116)	(0.028)
sent_1st_lag	0.388	0.145
	(0.492)	(0.121)
sent_2nd_lag	-0.053	0.141
	(0.491)	(0.120)
const	0.067	-0.004
	(0.053)	(0.013)
Observations	57	57
R^2	0.892	0.213
Adjusted R^2	0.884	0.152
Residual Std. Error	0.137	0.033
F Statistic	107.565***	3.514**
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 6.11: VAR results for mean news sentiment and volatility of gaming and non-gaming companies

Based on the results presented in Table 5.11, the autoregressive process for volatility can be observed; however, there is still no evidence of a relationship between the sentiment and volatility. Consequently, the second hypothesis cannot be rejected either.

As was the case for the Twitter sentiment, the contrast to previous works can again be observed. Even though previous works indicate the dependency of news and price movements, there was no evidence supporting that found in this analysis. This might be due to the usage of general headlines, not just those of financial news. Furthermore, only the headlines were examined, not the whole articles, however the financial only headlines have been proven useful in the past, specifically in the work of Balas *et al.*, (2022). And lastly, for the news analysis only daily data collected between January 1st and March 31st was used, longer sample period might prove useful in producing more robust results and might lead to more time series being stationary.

6.3 Twitter emotion

This section aims to present and interpret the results of analysing several different emotions and their effects on security returns and volatility. The emotions considered and their derivation are described in detail in section 3.1.3.

For the emotions expressed in Twitter posts the following hypotheses will be tested:

1. No specific emotion expressed in general Twitter posts has any impact on security returns.
2. No specific emotion expressed in general Twitter posts has any impact on security volatility.

The tables containing the results of VARs, specifically for regressions considering log-returns as the explained variable, are presented. Any bold text in the tables indicates a significant emotion variable for the specific firm, which will later be used in the Granger causality test for that company. Both gaming and non-gaming companies use the mean emotion variable and weighted mean emotion variable, where the weight is the number of retweets, as described in section 3.1.3.

logrets ~ tweet emotion:							
	non-gaming:			*gaming:*			
1st lag	**amazon**	**apple**	**toyota**	**cdpr**	**blizz**	**nint**	**ubi**
logrets	0.097	-0.153	-0.068	-0.195	-0.346***	-0.316***	0.012
	(0.123)	(0.115)	(0.110)	(0.118)	(0.104)	(0.118)	(0.110)
anger	0.012	**-0.197***	0.042	-0.009	-0.015	-0.105	-0.008
	(0.126)	(0.106)	(0.056)	(0.010)	(0.052)	(0.083)	(0.024)
anticipation	0.026	0.010	-0.011	0.003	0.051	-0.046	-0.002
	(0.076)	(0.056)	(0.033)	(0.005)	(0.042)	(0.043)	(0.012)
disgust	-0.138	**0.242****	-0.084	-0.013	0.050	0.00003	0.002
	(0.188)	(0.113)	(0.052)	(0.012)	(0.074)	(0.100)	(0.026)
fear	0.066	0.026	-0.025	**-0.016***	0.026	-0.051	0.023
	(0.101)	(0.078)	(0.051)	(0.008)	(0.057)	(0.064)	(0.019)
joy	-0.065	0.036	-0.063	0.003	-0.105	0.059	-0.002
	(0.080)	(0.066)	(0.039)	(0.008)	(0.074)	(0.075)	(0.022)
sadness	-0.118	0.046	-0.079	-0.001	-0.063	-0.088	-0.018
	(0.131)	(0.066)	(0.051)	(0.009)	(0.063)	(0.072)	(0.019)
surprise	0.067	-0.132	**0.105****	**-0.017****	-0.032	-0.123	0.003
	(0.137)	(0.084)	(0.049)	(0.008)	(0.061)	(0.091)	(0.026)
trust	0.046	0.006	0.020	0.005	0.019	0.066	0.003
	(0.048)	(0.048)	(0.020)	(0.006)	(0.045)	(0.059)	(0.009)
const	-0.001	-0.016	0.038	0.016**	0.022	0.105*	-0.003
	(0.037)	(0.074)	(0.032)	(0.008)	(0.042)	(0.057)	(0.015)
Observations	85	84	93	74	92	88	101
R^2	0.047	0.139	0.146	0.160	0.167	0.156	0.024
Adjusted R^2	-0.068	0.034	0.053	0.042	0.075	0.058	-0.072
Residual Std. Error	0.026 (df = 75)	0.020 (df = 74)	0.011 (df = 83)	0.021 (df = 64)	0.012 (df = 82)	0.014 (df = 78)	0.028 (df = 91)
F Statistic	0.409 (df = 9; 75)	1.324 (df = 9; 74)	1.572 (df = 9; 83)	1.352 (df = 9; 64)	1.820* (df = 9; 82)	1.597 (df = 9; 78)	0.250 (df = 9; 91)
Note:							*p<0.1; **p<0.05; ***p<0.01

Table 6.12: VAR results for mean Twitter emotions and logrets of gaming and non-gaming companies

granger causality tests for: logrets ~ tweet emotion

F test and Wald χ^2 test based on VAR(1) model:

	F	df1	df2	p	Chisq	df	p
cdpr_logrets <= cdpr_fear	3.65	1	64	.061	3.65	1	.056
cdpr_logrets <= cdpr_surprise	4.43	1	64	.039	4.43	1	.035
cdpr_logrets <= cdpr_fear|cdpr_surprise	2.92	2	64	.061	5.84	2	.054
apple_logrets <= apple_anger	3.46	1	74	.067	3.46	1	.063
apple_logrets <= apple_disgust	4.61	1	74	.035	4.61	1	.032
apple_logrets <= apple_anger|apple_disgust	3.32	2	74	.041	6.65	2	.036
toyota_logrets <= toyota_surprise	4.51	1	83	.037	4.51	1	.034

Table 6.13: Granger causality tests for significant Twitter emotions (mean) causing logrets

logrets ~ retweet-weighted tweet emotion:

	non-gaming:			*gaming:*			
1st lag	**amazon**	**apple**	**toyota**	**cdpr**	**blizz**	**nint**	**ubi**
logrets	0.103	-0.140	-0.064	-0.136	-0.293***	-0.243**	-0.007
	(0.119)	(0.112)	(0.106)	(0.108)	(0.105)	(0.117)	(0.110)
anger	-0.010	**-0.566****	**0.282***	0.009	-0.060	-0.070	-0.020
	(0.243)	(0.226)	(0.157)	(0.013)	(0.119)	(0.165)	(0.036)
anticipation	-0.013	0.061	-0.041	0.005	0.166	-0.075	-0.005
	(0.151)	(0.138)	(0.106)	(0.009)	(0.104)	(0.112)	(0.024)
disgust	-0.291	**0.745****	-0.231	0.005	0.177	0.045	0.023
	(0.374)	(0.301)	(0.168)	(0.019)	(0.140)	(0.228)	(0.046)
fear	0.192	0.050	-0.100	-0.015	0.100	-0.008	0.050
	(0.204)	(0.182)	(0.156)	(0.013)	(0.110)	(0.139)	(0.036)
joy	-0.129	0.053	**-0.238***	0.008	-0.233	0.095	0.005
	(0.156)	(0.172)	(0.122)	(0.011)	(0.154)	(0.172)	(0.039)
sadness	-0.327	0.062	-0.152	-0.008	-0.101	-0.025	-0.034
	(0.267)	(0.158)	(0.138)	(0.015)	(0.140)	(0.143)	(0.027)
surprise	0.212	**-0.374***	**0.384****	-0.006	0.063	-0.190	0.031
	(0.271)	(0.219)	(0.159)	(0.013)	(0.148)	(0.207)	(0.050)
trust	0.124	0.061	**0.107***	-0.002	-0.046	0.155	0.008
	(0.101)	(0.136)	(0.063)	(0.009)	(0.094)	(0.135)	(0.018)
const	0.009	-0.015	-0.010	0.001	-0.003	-0.015*	-0.011
	(0.014)	(0.010)	(0.006)	(0.004)	(0.005)	(0.008)	(0.008)
Observations	86	84	93	91	92	88	101
R^2	0.063	0.161	0.174	0.069	0.187	0.125	0.044
Adjusted R^2	-0.048	0.058	0.085	-0.035	0.098	0.024	-0.051
Residual Std. Error	0.026 (df = 76)	0.019 (df = 74)	0.011 (df = 83)	0.023 (df = 81)	0.012 (df = 82)	0.014 (df = 78)	0.027 (df = 91)
F Statistic	0.564 (df = 9; 76)	1.573 (df = 9; 74)	1.949* (df = 9; 83)	0.663 (df = 9; 81)	2.099** (df = 9; 82)	1.235 (df = 9; 78)	0.463 (df = 9; 91)
Note:							*p<0.1; **p<0.05; ***p<0.01

Table 6.14: VAR results for weighted mean Twitter emotions and logrets of gaming and non-gaming companies

granger causality tests for: logrets ~ retweet-weighted emotion

F test and Wald χ^2 test based on VAR(1) model:

	F	df1	df2	p	Chisq	df	p
apple_logrets <= apple_anger	6.27	1	74	.014	6.27	1	.012
apple_logrets <= apple_disgust	6.15	1	74	.015	6.15	1	.013
apple_logrets <= apple_surprise	2.91	1	74	.092	2.91	1	.088
apple_logrets <= apple_anger|apple_disgust|apple_surprise	3.42	3	74	.021	10.27	3	.016
toyota_logrets <= toyota_anger	3.23	1	83	.076	3.23	1	.072
toyota_logrets <= toyota_joy	3.78	1	83	.055	3.78	1	.052
toyota_logrets <= toyota_surprise	5.82	1	83	.018	5.82	1	.016
toyota_logrets <= toyota_trust	2.86	1	83	.094	2.86	1	.091
toyota_logrets <= toyota_anger|toyota_joy|toyota_surprise|toyota_trust	3.64	4	83	.009	14.57	4	.006

Table 6.15: Granger causality tests for significant Twitter emotions (weighted mean) causing logrets

Table 5.12 shows some significant emotion values for Apple, Toyota, and CD Projekt Red. In Table 5.13, the results of the Granger causality test for these significant emotions are presented. Based on these numbers, it can be observed that when considering the mean as the daily aggregator, the first hypothesis can be rejected at a 10% significance level in the following cases:

- Anger and disgust for Apple, both individually and jointly
- Surprise for Toyota
- Fear and surprise for CD Projekt Red, both individually and jointly

Moving on to Table 5.14, which considers the weighted mean of emotions, slightly different results can be observed. For Apple and Toyota, additional emotions are significant, while for CD Projekt Red, no emotion is significant in this case. The results of the Granger causality test presented in Table 5.15 show that these emotions do Granger cause log-returns for the specified firms. Based on that, the first hypothesis, when considering the weighted mean as the daily aggregator, can be rejected at a 10% significance level in the following cases:

- Anger, disgust and surprise for Apple, each individually and jointly as well
- Anger, joy, surprise and trust for Toyota, both individually and jointly

The differences between the two aggregate approaches are likely caused by a similar effect as mentioned in section 3.1.2. The most retweeted posts are likely to be promotional or advertisements, especially for gaming companies, as they aim to sell their digital products. Additionally, the number of tweets posted (and collected) for CD Projekt Red is lower than for either Tesla or Apple, making each popular tweet more impactful. The combination of a lower sample size and highly weighted tweets that might be of promotional nature rather than reflecting the actual emotions of users could be considered as the reason for the difference in the approaches. However, further research would be required to confirm the existence of such an effect.

The following tables present the results of the same models, using volatility as the dependent variable. Companies for which the volatility time series is non-stationary in the selected time period were not considered.

volatility ~ tweets emotion:

	non-gaming:	*gaming:*	
1st lag	**apple**	**nint**	**ubi**
volatility	0.987***	0.725***	0.183*
	(0.033)	(0.072)	(0.101)
anger	-0.660	**2.844*****	-0.208
	(0.550)	(0.917)	(0.147)
anticipation	-0.456	**1.570*****	-0.002
	(0.286)	(0.462)	(0.076)
disgust	-0.326	-1.158	-0.040
	(0.570)	(1.104)	(0.158)
fear	-0.087	-0.310	0.024
	(0.389)	(0.724)	(0.114)
joy	**-0.823****	-1.233	0.143
	(0.329)	(0.837)	(0.133)
sadness	**1.090*****	0.910	0.145
	(0.330)	(0.802)	(0.118)
surprise	**1.618*****	0.353	0.053
	(0.427)	(1.034)	(0.158)
trust	0.382	-0.871	**-0.113****
	(0.248)	(0.640)	(0.056)
const	-0.086	-0.855	0.742***
	(0.371)	(0.659)	(0.124)
Observations	84	88	101
R²	0.931	0.644	0.109
Adjusted R²	0.922	0.602	0.021
Residual Std. Error	0.098 (df = 74)	0.152 (df = 78)	0.170 (df = 91)
F Statistic	110.448*** (df = 9; 74)	15.652*** (df = 9; 78)	1.243 (df = 9; 91)
Note:			*p<0.1; **p<0.05; ***p<0.01

volatility ~ retweet-weighted tweets emotion:

	non-gaming:	*gaming:*	
1st lag	**apple**	**nint**	**ubi**
volatility	0.987***	0.704***	0.197*
	(0.034)	(0.076)	(0.100)
anger	-1.360	**3.953****	-0.350
	(1.213)	(1.922)	(0.219)
anticipation	-1.125	**3.463*****	0.051
	(0.719)	(1.256)	(0.145)
disgust	-0.700	-2.125	0.100
	(1.572)	(2.646)	(0.281)
fear	-1.040	-1.789	0.138
	(0.933)	(1.640)	(0.213)
joy	**-2.803*****	-2.391	0.240
	(0.888)	(1.985)	(0.234)
sadness	**2.975*****	0.361	0.212
	(0.812)	(1.668)	(0.162)
surprise	**4.398*****	0.095	0.290
	(1.131)	(2.470)	(0.299)
trust	**1.378***	-2.131	**-0.200***
	(0.722)	(1.537)	(0.110)
const	-0.012	0.329***	0.644***
	(0.069)	(0.113)	(0.102)
Observations	84	88	101
R²	0.930	0.596	0.163
Adjusted R²	0.921	0.549	0.080
Residual Std. Error	0.099 (df = 74)	0.161 (df = 78)	0.165 (df = 91)
F Statistic	108.695*** (df = 9; 74)	12.784*** (df = 9; 78)	1.968* (df = 9; 91)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 6.16: VAR results for mean and weighted mean Twitter emotions and volatility of gaming and non-gaming companies

granger causality tests for: volatility ~ tweets emotion

F test and Wald χ^2 test based on VAR(1) model:

	F	df1	df2	p	Chisq	df	p
apple_vola <= apple_joy	6.24	1	74	.015	6.24	1	.013
apple_vola <= apple_sadness	10.88	1	74	.001	10.88	1	<.001
apple_vola <= apple_surprise	14.37	1	74	<.001	14.37	1	<.001
apple_vola <= apple_joy|apple_sadness|apple_surprise	7.01	3	74	<.001	21.03	3	<.001
nint_vola <= nint_anger	9.63	1	78	.003	9.63	1	.002
nint_vola <= nint_anticipation	11.54	1	78	.001	11.54	1	<.001
nint_vola <= nint_anger|nint_anticipation	9.38	2	78	<.001	18.77	2	<.001
ubi_vola <= ubi_trust	4	1	91	.048	4	1	.045

Table 6.17: Granger causality tests for significant Twitter emotions (mean) causing volatility

granger causality tests for: volatility ~ retweet-weighted tweets emotion

F test and Wald χ^2 test based on VAR(1) model:

	F	df1	df2	p	Chisq	df	p
apple_vola <= apple_joy	9.96	1	74	.002	9.96	1	.002
apple_vola <= apple_sadness	13.42	1	74	<.001	13.42	1	<.001
apple_vola <= apple_surprise	15.12	1	74	<.001	15.12	1	<.001
apple_vola <= apple_trust	3.64	1	74	.060	3.64	1	.056
apple_vola <= apple_joy|apple_sadness|apple_surprise|apple_trust	4.92	4	74	.001	19.67	4	<.001
nint_vola <= nint_anger	4.23	1	78	.043	4.23	1	.040
nint_vola <= nint_anticipation	7.6	1	78	.007	7.6	1	.006
nint_vola <= nint_anger|nint_anticipation	5.64	2	78	.005	11.28	2	.004
ubi_vola <= ubi_trust	3.28	1	91	0.73	3.28	1	.070

Table 6.18: Granger causality tests for significant Twitter emotions (weighted mean) causing volatility

Based on Table 5.16, potential emotions affecting the volatility of specified securities were identified, and their effects were further tested using Granger causality analysis. The results of the Granger causality test are presented in Tables 5.17 and 5.18. Upon inspecting the p-values, it can be concluded that when considering the weighted mean, the second hypothesis can be rejected in the following cases:

- Joy, sadness, surprise and trust for Apple, individually and jointly, at 5% significance level
- Anger and anticipation for Nintendo, individually and jointly, at 5% significance level
- Trust for Ubisoft at 10% significance level

Furthermore, it can be observed in Table 5.16 that compared to the log-returns analysis, the differences between considering simple or weighted mean are minimal. Only one additional emotion, trust for Apple, became significant when using the weighted mean. All the other variables that were significant using only a simple mean are also significant in the weighted mean model. This indicates that indeed, for companies with larger amounts of Twitter posts (both collected and posted), the difference between the two approaches is lower, which makes intuitive sense as the effect of individual weights becomes less pronounced. Therefore, the difference observed when considering log-returns of CD Projekt Red is likely caused simply by the sample size rather than the firm being in the gaming industry. It also seems that for larger sample sizes, it can be beneficial to prefer the weighted mean over the simple one, as in both cases (log-returns and volatility), it increased or maintained the number of significant emotions.

6.4 News headlines emotion

The hypotheses formulated for the news headlines emotions resemble those for the Twitter emotions:

1. No specific emotion expressed in general news headlines has any impact on security returns.
2. No specific emotion expressed in general news headlines has any impact on security volatility.

In the following tables, the VARs of emotions expressed in headlines with log-returns and volatility, respectively, along with the Granger causality analysis of the significant emotions, are reported.

logrets ~ news emotion:

	non-gaming:			*gaming:*		
1st lag	**amazon**	**apple**	**tesla**	**cdpr**	**blizz**	**nint**
logrets	0.140	-0.238*	-0.126	-0.002	-0.474***	0.063
	(0.134)	(0.134)	(0.135)	(0.128)	(0.119)	(0.156)
anger	**0.248***	0.091		-0.022	0.031	**-0.050***
	(0.126)	(0.088)		(0.042)	(0.034)	(0.026)
anticipation	-0.094	0.004		-0.002	0.010	-0.003
	(0.094)	(0.065)		(0.023)	(0.014)	(0.020)
disgust	-0.032	-0.176		-0.004		-0.041
	(0.157)	(0.135)		(0.044)		(0.045)
fear	-0.054	0.030		0.021	**-0.053***	**0.045***
	(0.104)	(0.078)		(0.030)	(0.031)	(0.025)
joy	-0.176	0.061	0.075	**-0.092***	0.015	-0.012
	(0.127)	(0.078)	(0.145)	(0.049)	(0.031)	(0.024)
sadness		0.025	-0.043	0.039	0.026	-0.012
		(0.072)	(0.094)	(0.035)	(0.026)	(0.029)
surprise	-0.003	-0.014	-0.109	0.002	-0.002	-0.016
	(0.149)	(0.096)	(0.161)	(0.036)	(0.024)	(0.030)
trust	**0.138***	0.063		0.032	-0.006	**0.038***
	(0.080)	(0.059)		(0.025)	(0.019)	(0.020)
const	-0.005	-0.039*	0.017	-0.001	0.0005	-0.004
	(0.028)	(0.022)	(0.025)	(0.006)	(0.005)	(0.007)
Observations	62	60	61	66	58	60
R^2	0.137	0.140	0.027	0.132	0.319	0.168
Adjusted R^2	0.007	-0.015	-0.042	-0.008	0.207	0.018
Residual Std. Error	0.020 (df = 53)	0.013 (df = 50)	0.042 (df = 56)	0.025 (df = 56)	0.011 (df = 49)	0.011 (df = 50)
F Statistic	1.055 (df = 8; 53)	0.902 (df = 9; 50)	0.390 (df = 4; 56)	0.944 (df = 9; 56)	2.863** (df = 8; 49)	1.118 (df = 9; 50)
Note:				$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$		

Table 6.19: VAR results for news headlines emotions and logrets of gaming and non-gaming companies

granger causality tests for: logrets ~ news emotion

F test and Wald χ^2 test based on VAR(1) model:

	F	df1	df2	p	Chisq	df	p
amazon_logrets <= amazon_anger	3.86	1	53	.055	3.86	1	.049
amazon_logrets <= amazon_trust	2.94	1	53	.092	2.94	1	.086
amazon_logrets <= amazon_anger|amazon_trust	2.54	2	53	.089	5.07	2	.079
cdpr_logrets <= cdpr_joy	3.59	1	56	.063	3.59	1	.058
blizz_logrets <= blizz_fear	2.83	1	49	.099	2.83	1	.093
nint_logrets <= nint_anger	3.6	1	50	.064	3.6	1	.058
nint_logrets <= nint_fear	3.31	1	50	.075	3.31	1	.069
nint_logrets <= nint_trust	3.36	1	50	.073	3.36	1	.067
nint_logrets <= nint_anger|nint_fear|nint_trust	2.77	3	50	.051	8.32	3	.040

Table 6.20: Granger causality tests for significant news emotions causing logrets

After examining the table 5.19, it can be seen that some emotions appear significant for potentially influencing the returns. These were further tested using the Granger causality analysis, results of which can be seen in table 5.20. Based on the numbers reported in the tables, it can be concluded that for the news headlines, the first hypothesis can be rejected at 10% significance level in these cases:

- Anger and trust for Amazon, individually and jointly
- Joy for CD Projekt Red
- Fear for Activision Blizzard
- Anger, fear and trust for Nintendo, individually and jointly

An observation to note is the fact that for each examined gaming company, at least one of the emotions expressed in the news headlines Granger causes the returns, while from the selection of non-gaming companies, this is true only for Amazon. A potential explanation lies in the generality of the collected news headlines, as for the tweets, not only the financial news, but all the available headlines were collected. However, it is likely that when a universal media covers the videogame industry, it focuses more on its performance in business rather than reporting information that would be interesting only to the gaming community. This is not the case for the examined non-gaming companies, which are more likely to receive general media coverage. This could be further investigated by collecting emotion data from specific, specialized and non-specialized media and comparing the performance.

volatility ~ news emotion:

	non-gaming:		gaming:	
1st lag	**apple**	**tesla**	**nint**	**ubi**
volatility	0.940^{***}	0.912^{***}	0.697^{***}	0.188
	(0.027)	(0.049)	(0.091)	(0.125)
anger	0.285		0.417	0.027
	(0.225)		(0.288)	(0.388)
anticipation	0.079		0.116	0.167
	(0.177)		(0.221)	(0.175)
disgust	-0.051		-0.569	0.462
	(0.343)		(0.466)	(0.410)
fear	-0.301		**-0.501***	-0.034
	(0.207)		(0.274)	(0.348)
joy	-0.281	-0.628	-0.204	0.199
	(0.199)	(0.529)	(0.259)	(0.340)
sadness	0.247	0.090	**1.077*****	0.527
	(0.185)	(0.357)	(0.300)	(0.327)
surprise	0.149	-0.425	-0.108	-0.440
	(0.243)	(0.589)	(0.331)	(0.399)
trust	-0.121		0.121	-0.203
	(0.163)		(0.216)	(0.210)
const	0.672^{***}	0.161	0.237^{**}	0.075
	(0.127)	(0.106)	(0.104)	(0.068)
Observations	67	60	61	60
R^2	0.165	0.610	0.878	0.978
Adjusted R^2	0.034	0.540	0.869	0.974
Residual Std. Error	0.203 (df = 57)	0.119 (df = 50)	0.152 (df = 56)	0.033 (df = 50)
F Statistic	1.255 (df = 9; 57)	8.687^{***} (df = 9; 50)	100.461^{***} (df = 4; 56)	243.609^{***} (df = 9; 50)
Note:			*p<0.1; **p<0.05; ***p<0.01	

Table 6.21: VAR results for news headlines emotions and volatility of gaming and non-gaming companies

granger causality tests for: volatility ~ news emotion

F test and Wald χ^2 test based on VAR(1) model:

	F	df1	df2	p	Chisq	df	p
nint_vola <= nint_fear	3.34	1	50	.074	3.34	1	.068
nint_vola <= nint_sadness	12.88	1	50	<.001	12.88	1	<.001
nint_vola <= nint_fear|nint_sadness	6.74	2	50	.003	13.48	2	.001

Table 6.22: Granger causality tests for significant news emotions causing volatility

Tables 5.21 and 5.22 reveal sufficient evidence to reject the second hypothesis only in the case of fear and sadness for Nintendo, including their joint effect. It appears that the predictive power of emotions in the news headlines is weaker for volatility compared to returns. This stands in contrast to the findings for Twitter emotions, where the potential predictive power of individual emotions appeared to be comparable between both returns and volatility.

7 Conclusion

The initial goal of the thesis was to discover and describe whether the sentiment or emotions expressed on social media have different effects on stock market price movements based on whether a specific firm is part of the gaming industry or not. For both of these sectors, data was gathered individually for four different companies. Since the main objective was to determine whether the expressiveness of gamers in social media can lead to a more pronounced effect, all tweets and headlines were collected, rather than just those dealing specifically with finance, which was the approach of most previous works in the sentiment topic.

Between October 20th, 2022, and March 10th, 2023, over 1.7 million tweets were collected, each individually assigned a sentiment value using the VADER tool, and aggregated into daily values. To assign tweets with specific emotions, the NRC lexicon was employed, marking individual tweets with expressed emotions and aggregating the obtained values to daily levels. A similar approach was used for gathering sentiment and emotions data from news headlines, which were collected from the Europe Media Monitor between January 1st and March 31st, 2023, resulting in the gathering of almost 300 thousand headlines. Financial data for the selected companies were mainly collected from Yahoo Finance, specifically the daily opening and closing stock prices, from which daily logarithmic returns and volatility were computed. Subsequently, vector autoregression models were used to determine whether relationships exist between sentiment and market movements. If the variables were significant, a causal relationship was further tested using Granger causality analysis.

The results of the analyses, especially when compared to the outcomes of previous research, indicate that tweets produced by the general public and news from universal media hold less predictive power than specialized financial media or social media accounts when solely considering sentiment polarity. Some dependencies were discovered when employing a larger scale of emotions; however, each individual firm appeared to be influenced by different emotions, if at all. Regarding Twitter emotions, it appeared that when dealing with larger sample sizes, for example, firms with thousands of tweets written every hour, using weighted mean where retweets are the weight as the daily aggregator proved beneficial. Conversely, for firms with lower numbers of posted tweets, a simple mean performed better, likely due to popular tweets having a larger impact compared to cases with higher sample sizes. Nothing suggested that this difference was caused by the fact that it concerned a gaming company, as it did not hold for other larger firms.

Regarding emotions expressed in news headlines, they appeared to be less effective than those from Twitter, especially for non-gaming companies. However, for gaming companies, even though the number of significant emotions for each firm decreased or remained the same, some significant emotions were still identified. The likely reason is that the gaming industry is generally not often covered by universal media, unless it is a report of a specific business success or failure. In other words, the better performance of emotions in the news headlines on gaming companies might be attributed to a higher percentage of news concerning finance in that field. This statement would need to be further examined by research focused on either comparing general and financial media directly or examining the percentage of news concerning finance pertaining to gaming or non-gaming industries.

Overall, the most important finding of this thesis appears to be that when employing sentiment analysis into market models, one needs to either focus on the sentiment among investors or find specific channels, media, websites, etc., that are relevant for the examined company. It appears that, regardless of the business field, different emotions or combinations of them are relevant for each company. Additionally, different communication channels might be either less or more relevant for determining the sentiment influencing individual companies. In simple terms, one should either deeply focus on the specifics of one selected company and build a customized model or depend mainly on investor sentiment and financial reports.

Having said that, the results of this thesis would likely be improved and more conclusive with a larger sample period. Due to the usage of daily data over only several months, problems with time series stationarity were encountered, resulting in a rather large portion of data being disregarded. Furthermore, examining the change in log-returns, rather than just log-returns, could help uncover additional information.

In conclusion, the thesis might prove beneficial for anyone attempting to create market models using sentiment information from social media and news headlines, as it sheds light on the performance of general sentiment and how it might influence individual firms. It emphasizes the need for a deeper understanding of what sentiment channels are relevant for selected companies or the usage of exclusively financial media sentiment.

Bibliography

"Activision Blizzard, Inc. (ATVI) Stock Price, News, Quote & History - Yahoo Finance". (2022), , available at: https://finance.yahoo.com/quote/ATVI/ (accessed 29 December 2022).

Baker, M. and Wurgler, J. (2007), "Investor Sentiment in the Stock Market", *Journal of Economic Perspectives*, Vol. 21 No. 2, pp. 129–152, doi: 10.1257/jep.21.2.129.

Balas, V.E., Bansal, R.C., Mangipudi, S.K. and Dawn, S. (2022), "Sentiment Analysis of Stocks Based on News Headlines Using NLP", *Proceedings of the International Conference on Artificial Intelligence Techniques for Electrical Engineering Systems (AITEES 2022)*, Vol. 3, Atlantis Press Zeger Karssen, The Netherlands.

Barber, B.M. and Odean, T. (2000), "Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors", *The Journal of Finance*, [American Finance Association, Wiley], Vol. 55 No. 2, pp. 773–806.

Baxi, B., Panda, R. and Karani, A. (2016), "Community Marketing: an effective Marketing Strategy", *Romanian Journal of Marketing*, No. 3, pp. 2–11.

Bollen, J., Mao, H. and Zeng, X. (2011), "Twitter mood predicts the stock market", *Journal of Computational Science*, Vol. 2 No. 1, pp. 1–8, doi: 10.1016/j.jocs.2010.12.007.

Bollerslev, T. (1986), "Generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics*, Vol. 31 No. 3, pp. 307–327, doi: 10.1016/0304-4076(86)90063-1.

Bolton, N. (2022), "Top Video Game Stocks", *Investopedia*, 6 October, available at: https://www.investopedia.com/news/top-video-game-stocks-2017/ (accessed 29 December 2022).

Brown, G.W. and Cliff, M.T. (2005), "Investor Sentiment and Asset Valuation", *The Journal of Business*, The University of Chicago Press, Vol. 78 No. 2, pp. 405–440, doi: 10.1086/427633.

Carriker, M. (2017), “Making a press kit on your site”, 21 December, available at: https://www.gamasutra.com/blogs/MeganCarriker/20171221/312157/Making_a_press_kit_on_your_site.php (accessed 22 June 2021).

“CDR | CD Projekt S.A. Advanced Charts”. (2022), *MarketWatch*, available at: https://www.marketwatch.com/investing/stock/cdr/charts?countrycode=pl (accessed 29 December 2022).

Chen, H., De, P., Hu, Y. (Jeffrey) and Hwang, B.-H. (2014), “Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media”, *The Review of Financial Studies*, [Oxford University Press, The Society for Financial Studies], Vol. 27 No. 5, pp. 1367–1403.

Clement, J. (2022a), “Global gaming revenue by segment 2022”, *Statista*, 2 December, available at: https://www.statista.com/statistics/292751/mobile-gaming-revenue-worldwide-device/ (accessed 22 December 2022).

Clement, J. (2022b), “Electronic Arts marketing spend 2022”, *Statista*, 24 November, available at: https://www.statista.com/statistics/672141/electronic-arts-marketing-and-sales-spending/ (accessed 22 December 2022).

Davenport, J. (2018), “Battle passes are replacing loot boxes, but they’re not necessarily a better deal”, *PC Gamer*, 5 July.

De Long, J.B., Shleifer, A., Summers, L.H. and Waldmann, R.J. (1990), “Noise Trader Risk in Financial Markets”, *Journal of Political Economy*, University of Chicago Press, Vol. 98 No. 4, pp. 703–738.

Denby, L. (2019), “What should my indie game marketing budget be? - Game If You Are”, 23 October, available at: https://gameifyouare.com/2019/10/23/what-should-my-indie-game-marketing-budget-be/ (accessed 22 December 2022).

“‘Diablo Immortal’ Microtransactions Spark Fan Outrage”. (2022), *Gale General OneFile*, 6 June, available at: https://go-gale-com.ezproxy.is.cuni.cz/ps/i.do?p=ITOF&u=karlova&v=2.1&it=r&id=GALE%7CA706124245&inPS=true&linkSource=interlink&sid=bookmark-ITOF (accessed 29 December 2022).

Engelberg, J. (2008), “Costly Information Processing: Evidence from Earnings Announcements”, SSRN Scholarly Paper, Rochester, NY, 18 January, doi: 10.2139/ssrn.1107998.

Engelstätter, B. and Ward, M.R. (2013), “Strategic Timing of Entry: Evidence from Video Games”, *Discussion Paper No. 13-117*, p. 41.

Epps, D. (2022), “Marvel Snap’s season pass, explained”, *Polygon*, 27 October, available at: https://www.polygon.com/guides/23426835/marvel-snap-season-pass-premium (accessed 28 December 2022).

“Euro Truck Simulator 2 on Steam”. (2022), , available at: https://store.steampowered.com/app/227300/Euro_Truck_Simulator_2/ (accessed 25 December 2022).

“Europe Media Monitor (EMM) | Knowledge for policy”. (2018), *EMM*, available at: https://knowledge4policy.ec.europa.eu/online-resource/europe-media-monitor-emm_en (accessed 1 July 2023).

Fama, E.F. (1970), “Efficient Capital Markets: A Review of Theory and Empirical Work”, *The Journal of Finance*, [American Finance Association, Wiley], Vol. 25 No. 2, pp. 383–417, doi: 10.2307/2325486.

Foster, L. (2016), “Effects of Video Game Streaming on Consumer Attitudes and Behaviors”, *Electronic Theses and Dissertations*.

"GeForce NOW Cloud Gaming". (2022), *NVIDIA*, available at: https://www.nvidia.com/en-gb/geforce-now/ (accessed 25 December 2022).

Goh Khim-Yong, Heng Cheng-Suang, and Lin Zhijie. (2013), "Social Media Brand Community and Consumer Behavior: Quantifying the Relative Impact of User-and Marketer-Generated Content", *Information Systems Research*, The Institute for Operations Research and the Management Sciences (INFORMS), Vol. 24 No. 1, pp. 88–107, doi: 10.1287/isre.1120.0469.

Goslin, A. (2022), "Diablo Immortal now coming to PC along with mobile this June", *Polygon*, 25 April, available at: https://www.polygon.com/23040721/diablo-immortal-pc-mobile-release-date-crossplay (accessed 29 December 2022).

Granger, C.W.J. (1969), "Investigating Causal Relations by Econometric Models and Cross-spectral Methods", *Econometrica*, [Wiley, Econometric Society], Vol. 37 No. 3, pp. 424–438, doi: 10.2307/1912791.

Grotland, J. (2011), "Monetization Strategies: iOS + Android", 24 November.

Heston, S.L. and Sinha, N.R. (2017), "News vs. Sentiment: Predicting Stock Returns from News Stories", *Financial Analysts Journal*, Routledge, Vol. 73 No. 3, pp. 67–83, doi: 10.2469/faj.v73.n3.3.

Hutto, C. and Gilbert, E. (2014), "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text", *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 8 No. 1, pp. 216–225, doi: 10.1609/icwsm.v8i1.14550.

Jockers, M.L. (2015), "Syuzhet", HTML, .

Johnson, M.R. and Woodcock, J. (2017), "'It's like the gold rush': the lives and careers of professional video game streamers on Twitch.tv", *Information, Communication & Society*, Routledge, Vol. 22 No. 3, pp. 336–351, doi: 10.1080/1369118X.2017.1386229.

Keane, S. (2022), “Google to Shut Down Stadia, Refund All Purchases”, *CNET*, 1 December, available at: https://www.cnet.com/tech/gaming/google-stadia-is-shutting-down-in-2023-all-purchases-to-be-refunded/ (accessed 25 December 2022).

Kearney, M.W. (2019), “rtweet: Collecting and analyzing Twitter data”, *Journal of Open Source Software*, Vol. 4 No. 42, p. 1829, doi: 10.21105/joss.01829.

King, D.L. and Delfabbro, P.H. (2018), “Predatory monetization schemes in video games (e.g. ‘loot boxes’) and internet gaming disorder”, *Addiction*, Vol. 113 No. 11, pp. 1967–1969, doi: 10.1111/add.14286.

King, D.L. and Delfabbro, P.H. (2019), “Video Game Monetization (e.g., ’Loot Boxes’): a Blueprint for Practical Social Responsibility Measures”, *International Journal of Mental Health & Addiction*, Springer Nature, Vol. 17 No. 1, pp. 166–179, doi: 10.1007/s11469-018-0009-3.

Kordonis, J., Symeonidis, S. and Arampatzis, A. (2016), “Stock Price Forecasting via Sentiment Analysis on Twitter”, *Proceedings of the 20th Pan-Hellenic Conference on Informatics*, Association for Computing Machinery, New York, NY, USA, pp. 1–6, doi: 10.1145/3003733.3003787.

Li, X., Xie, H., Chen, L., Wang, J. and Deng, X. (2014), “News impact on stock price return via sentiment analysis”, *Knowledge-Based Systems*, Vol. 69, pp. 14–23, doi: 10.1016/j.knosys.2014.04.022.

Livingston, C. (2020), “Do you buy battle passes, and do you complete them?”, *PC Gamer*, 5 August.

Łuczuk, P. and Maj, S. (2022), “Can Stock Market Fluctuations Be Predicted? Wig.Games Index Secret on the Example of CD Project”, *Social Communication*, Vol. 8 No. 1, pp. 8–20, doi: 10.2478/sc-2022-0002.

Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, New York : Springer, Berlin.

Malkiel, B.G. (2003), “The Efficient Market Hypothesis and Its Critics”, *Journal of Economic Perspectives*, Vol. 17 No. 1, pp. 59–82, doi: 10.1257/089533003321164958.

“PATH OF EXILE Tips: Microtransactions That Are Totally Worth It”. (2019), *Loot and Grind*, 6 August, available at: https://lootandgrind.com/path-of-exile-tips-microtransactions-totally-worth-it/ (accessed 28 December 2022).

Plunkett, L. (2021), “CD Projekt Settles Cyberpunk 2077 Lawsuit, Will Pay $1.85 Million”, *Kotaku*, 17 December, available at: https://kotaku.com/cd-projekt-settles-cyberpunk-2077-lawsuit-will-pay-1-1848230221 (accessed 29 December 2022).

Poch, R.(1) and Martin, B.(2). (2015), “Effects of intrinsic and extrinsic motivation on user-generated content”, *Journal of Strategic Marketing*, Routledge, Vol. 23 No. 4, pp. 305–317, doi: 10.1080/0965254X.2014.926966.

Poor, N. (2014), “Computer game modders’ motivations and sense of community: A mixed-methods approach”, *New Media and Society*, SAGE Publications Ltd, Vol. 16 No. 8, pp. 1249–1267, doi: 10.1177/1461444813504266.

Rao, T. and Srivastava, S. (2012), “Analyzing Stock Market Movements Using Twitter Sentiment Analysis”, *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*, IEEE Computer Society, USA, pp. 119–123.

Ruby, D. (2023), “58+ Twitter Statistics For Marketers In 2023 (Users & Trends)”, *DemandSage*, 1 March, available at: https://www.demandsage.com/twitter-statistics/ (accessed 29 June 2023).

Schreier, J. (2017), “Top Video Game Companies Won’t Stop Talking About ‘Games As A Service’”, *Kotaku*, 30 May, available at: https://kotaku.com/top-video-game-companies-wont-stop-talking-about-games-1795663927 (accessed 25 December 2022).

Schumaker, R.P., Zhang, Y., Huang, C.-N. and Chen, H. (2012), “Evaluating sentiment in financial news articles”, *Decision Support Systems*, Vol. 53 No. 3, pp. 458–464, doi: 10.1016/j.dss.2012.03.001.

Selway, J. (2023), “How Much Did GTA 5 Cost to Make?”, *Game Rant*, 6 June, available at: https://gamerant.com/grand-theft-auto-5-development-cost-budget-breakdown/ (accessed 31 July 2023).

Shi, S.W., Xia, M. and Huang, Y. (2015), “From Minnows to Whales: An Empirical Study of Purchase Behavior in Freemium Social Games”, *International Journal of Electronic Commerce*, Routledge, Vol. 20 No. 2, pp. 177–207, doi: 10.1080/10864415.2016.1087820.

Shleifer, A. and Vishny, R.W. (1997), “The Limits of Arbitrage”, *The Journal of Finance*, [American Finance Association, Wiley], Vol. 52 No. 1, pp. 35–55, doi: 10.2307/2329555.

Si, J., Mukherjee, A., Liu, B., Li, Q., Li, H. and Deng, X. (2013), “Exploiting Topic based Twitter Sentiment for Stock Prediction”.

Smailović, J., Grčar, M., Lavrač, N. and Žnidaršič, M. (2014), “Stream-based active learning for sentiment analysis in the financial domain”, *Information Sciences*, Vol. 285, pp. 181–203, doi: 10.1016/j.ins.2014.04.034.

Stewart, M. (2022), “Xbox Series X/S First-Party Games Will Cost $70 Next Year”, *Game Informer*, 6 December, available at: https://www.gameinformer.com/2022/12/06/xbox-series-xs-first-party-games-will-cost-70-next-year (accessed 25 December 2022).

Stone, P.J., Dunphy, D.C. and Smith, M.S. (1966), *The General Inquirer: A Computer Approach to Content Analysis*, M.I.T. Press, Oxford, England, p. 651.

Straub, N. (2020), "Every Country With Laws Against Loot Boxes (& What The Rules Are)", *ScreenRant*, 5 October, available at: https://screenrant.com/lootbox-gambling-microtransactions-illegal-japan-china-belgium-netherlands/ (accessed 28 December 2022).

Tayal, D. and Satya, K. (2009), "Comparative Analysis of the Impact of Blogging and Micro-blogging on Market Performance", *International Journal on Computer Science and Engineering*, Vol. 1.

Taylor, I. (2018a), "Report: 69% of consumers find cosmetic microtransactions acceptable", *GamesIndustry.Biz*, 16 February, available at: https://www.gamesindustry.biz/cosmetic-microtransactions-are-an-acceptable-form-of-monetisation-among-69-percent-of-consumers-report-finds (accessed 28 December 2022).

Taylor, I. (2018b), "Activision Blizzard stock tumbles following Diablo Immortal backlash", *GamesIndustry.Biz*, 6 November, available at: https://www.gamesindustry.biz/activision-blizzard-stocks-tumble-amid-fan-backlash-over-diablo-immortal (accessed 29 December 2022).

Tedesco, L. (2019), "16 Biggest Gaming Conventions Ranked", *Game Rant*, 18 November, available at: https://gamerant.com/gamin-conventions-world-smallest-largest-ranked/ (accessed 23 December 2022).

Tetlock, P.C. (2007), "Giving Content to Investor Sentiment: The Role of Media in the Stock Market", *The Journal of Finance*, [American Finance Association, Wiley], Vol. 62 No. 3, pp. 1139–1168.

Tumarkin, R. and Whitelaw, R.F. (2001), "News or Noise? Internet Postings and Stock Prices", *Financial Analysts Journal*, CFA Institute, Vol. 57 No. 3, pp. 41–51.

Valle-Cruz, D., Fernandez-Cortez, V., López-Chau, A. and Sandoval-Almazán, R. (2022), "Does Twitter Affect Stock Market Decisions? Financial Sentiment Analysis During Pandemics: A Comparative Study of the H1N1 and the COVID-19 Periods", *Cognitive Computation*, Vol. 14 No. 1, pp. 372–387, doi: 10.1007/s12559-021-09819-8.

Welsh, O. (2022), "It's not just because it's exploitative that Diablo Immortal doesn't work", *Polygon*, 10 July, available at: https://www.polygon.com/23200136/diablo-immortal-exploitative-monetization-legendary-gems-rifts (accessed 28 December 2022).

Wijman, T. (2020), "Global Game Revenues Up an Extra $15 Billion This Year as Engagement Skyrockets", *Newzoo*, 4 November, available at: https://newzoo.com/insights/articles/game-engagement-during-covid-pandemic-adds-15-billion-to-global-games-market-revenue-forecast (accessed 22 December 2022).

Wijman, T. (2022), "The Games Market Will Decline -4.3% to $184.4 Billion in 2022; Long-Term Outlook Remains Positive", *Newzoo*, available at: https://newzoo.com/insights/articles/the-games-market-will-decline-4-3-to-184-4-billion-in-2022 (accessed 22 December 2022).

Wooldridge, J.M. (2020), *Introductory Econometrics: A Modern Approach*, Seventh edition., Cengage Learning, Boston, Massachusetts.

Xu, Y., Liang, C., Li, Y. and Huynh, T.L.D. (2022), "News sentiment and stock return: Evidence from managers' news coverages", *Finance Research Letters*, Vol. 48, p. 102959, doi: 10.1016/j.frl.2022.102959.

Yury, S. and Mickiewicz, D. (2018), "How much does it cost to make a video game?", *VironIT*, 19 June, available at: https://vironit.com/how-much-does-it-cost-to-make-a-video-game (accessed 22 December 2022).

Záhora, Z. (2014), *Marketing digitálních her*, Masarykova univerzita, Filozofická fakulta.

Zammarchi, G., Mola, F. and Conversano, C. (2023), “Using sentiment analysis to evaluate the impact of the COVID-19 outbreak on Italy’s country reputation and stock market performance”, *Statistical Methods & Applications*, doi: 10.1007/s10260-023-00690-5.

Zhang, X., Fuehres, H. and Gloor, P.A. (2011), “Predicting Stock Market Indicators Through Twitter ‘I hope it is not as bad as I fear’”, *Procedia - Social and Behavioral Sciences*, Vol. 26, pp. 55–62, doi: 10.1016/j.sbspro.2011.10.562.

Appendix A: keywords used to gather data

Gaming

CD Projekt Red	**Ubisoft**	**Nintendo**	**Blizzard**
CD Projekt	Ubisoft	Nintendo	Blizzard
CD Project	Assassin	Zelda	Activision
Cyberpunk	Far Cry	Switch	Overwatch
Witcher 3	Rainbow Six	Byonetta	Diablo
Adam Kicinski	Yves Guillemot	Super Mario	Warcraft
			Starcraft
			Call of Duty

Non-gaming

Amazon	**Apple**	**Tesla**	**Toyota**
Amazon	Apple	Tesla	Toyota
Audible	iPhone	Musk	Supra
Twitch	iMac		Prius
Jeff Bezos	iPad		Corolla
	iOS		

Keywords used to search both tweets and news

Appendix B: results of ADF tests

ADF for logrets from Oct 20 to Mar 10

gaming	adf value	p-value
blizz	-4.894	0.010
cdpr	-4.904	0.010
nint	-4.964	0.010
ubi	-3.660	0.031
non-gaming		
amazon	-3.950	0.014
apple	-5.350	0.010
tesla	-3.363	0.064
toyota	-4.664	0.010

ADF for volatility from Oct 20 to Mar 10

gaming	adf value	p-value
blizz	-2.834	**0.231**
cdpr	-2.441	**0.394**
nint	-3.985	0.013
ubi	-3.310	0.073
non-gaming		
amazon	-2.634	**0.314**
apple	-3.838	0.020
tesla	-2.559	**0.345**
toyota	-2.451	**0.389**

ADF for logrets from Jan 01 to Mar 31

gaming	adf value	p-value
blizz	-4.033	0.014
cdpr	-4.677	0.010
nint	-3.781	0.025
ubi	-2.877	**0.219**
non-gaming		
amazon	-3.292	0.080
apple	-4.256	0.010
tesla	-3.060	0.144
toyota	-2.350	**0.433**

ADF for volatility from Jan 01 to Mar 31

gaming	adf value	p-value
blizz	-2.351	**0.433**
cdpr	-1.328	**0.848**
nint	-2.934	**0.196**
ubi	-3.382	0.066
non-gaming		
amazon	-2.339	**0.437**
apple	-3.486	0.049
tesla	-3.619	0.038
toyota	-2.521	**0.363**

ADF tests for log-rets and volatility corresponding to time-series of tweets and headlines, bold values not stationary

ADF for compound tweets sentiment

gaming	adf value	p-value
blizz	-4,059	0,010
cdpr	-4,130	0,010
nint	-4,728	0,010
ubi	-5,963	0,010
non-gaming		
amazon	-3,544	0,041
apple	-4,179	0,010
tesla	-3,978	0,013
toyota	-4,528	0,010

ADF for compound weighted mean tweets sentiment

gaming	adf value	p-value
blizz	-6,084	0,010
cdpr	-3,959	0,014
nint	-4,466	0,010
ubi	-4,618	0,010
non-gaming		
amazon	-4,942	0,010
apple	-4,848	0,010
tesla	-4,758	0,010
toyota	-6,281	0,010

note: weights = n. of retweets

ADF for compound news sentiment

gaming	adf value	p-value
blizz	-2,337	**0,438**
cdpr	-4,663	0,010
nint	-4,009	0,014
ubi	-2,648	**0,311**
non-gaming		
amazon	-3,524	0,046
apple	-2,442	**0,395**
tesla	-3,441	0,055
toyota	-3,604	0,039

ADF tests for sentiment values of tweets and headlines, bold values not stationary

ADF for tweet emotions

gaming	emotion	adf value	p-value	non-gaming	emotion	adf value	p-value
blizz				**amazon**			
	anger	-4,822	0,010		anger	-3,243	0,084
	anticipation	-3,804	0,021		anticipation	-3,418	0,055
	disgust	-3,809	0,021		disgust	-3,431	0,052
	fear	-4,622	0,010		fear	-3,566	0,039
	joy	-5,047	0,010		joy	-3,778	0,022
	sadness	-3,270	0,079		sadness	-3,731	0,025
	surprise	-4,022	0,010		surprise	-3,326	0,070
	trust	-4,941	0,010		trust	-4,226	0,010
cdpr				**apple**			
	anger	-5,806	0,010		anger	-3,482	0,047
	anticipation	-4,198	0,010		anticipation	-5,189	0,010
	disgust	-4,225	0,010		disgust	-3,637	0,033
	fear	-5,539	0,010		fear	-3,234	0,085
	joy	-4,065	0,010		joy	-4,229	0,010
	sadness	-5,025	0,010		sadness	-4,038	0,010
	surprise	-4,766	0,010		surprise	-5,157	0,010
	trust	-4,740	0,010		trust	-4,375	0,010
nint				**tesla**			
	anger	-4,057	0,010		anger	-3,814	0,020
	anticipation	-4,232	0,010		anticipation	-4,746	0,010
	disgust	-4,036	0,010		disgust	-3,210	0,089
	fear	-3,385	0,060		fear	-2,724	**0,275**
	joy	-4,952	0,010		joy	-4,358	0,010
	sadness	-4,985	0,010		sadness	-3,573	0,038
	surprise	-4,519	0,010		surprise	-4,252	0,010
	trust	-4,493	0,010		trust	-4,134	0,010
ubi				**toyota**			
	anger	-5,349	0,010		anger	-4,318	0,010
	anticipation	-5,051	0,010		anticipation	-4,948	0,010
	disgust	-4,944	0,010		disgust	-4,706	0,010
	fear	-4,787	0,010		fear	-4,689	0,010
	joy	-5,098	0,010		joy	-5,537	0,010
	sadness	-5,322	0,010		sadness	-4,102	0,010
	surprise	-4,787	0,010		surprise	-5,906	0,010
	trust	-4,504	0,010		trust	-4,896	0,010

ADF tests for emotion values of tweet means, bold values not stationary

ADF for tweet emotions, weighted

gaming	emotion	adf value	p-value	non-gaming	emotion	adf value	p-value
blizz				**amazon**			
	anger	-4,078	0,010		anger	-5,964	0,010
	anticipation	-4,021	0,010		anticipation	-6,130	0,010
	disgust	-3,678	0,029		disgust	-6,247	0,010
	fear	-4,112	0,010		fear	-6,886	0,010
	joy	-3,824	0,020		joy	-4,853	0,010
	sadness	-3,958	0,013		sadness	-4,629	0,010
	surprise	-3,550	0,040		surprise	-5,994	0,010
	trust	-4,018	0,010		trust	-3,978	0,013
cdpr				**apple**			
	anger	-5,209	0,010		anger	-4,743	0,010
	anticipation	-4,064	0,010		anticipation	-4,805	0,010
	disgust	-4,238	0,010		disgust	-4,866	0,010
	fear	-4,773	0,010		fear	-4,679	0,010
	joy	-3,807	0,021		joy	-5,054	0,010
	sadness	-4,980	0,010		sadness	-5,023	0,010
	surprise	-4,639	0,010		surprise	-5,066	0,010
	trust	-4,811	0,010		trust	-5,091	0,010
nint				**tesla**			
	anger	-4,750	0,010		anger	-4,978	0,010
	anticipation	-4,991	0,010		anticipation	-4,181	0,010
	disgust	-4,633	0,010		disgust	-4,344	0,010
	fear	-4,727	0,010		fear	-4,457	0,010
	joy	-4,596	0,010		joy	-4,225	0,010
	sadness	-4,951	0,010		sadness	-4,481	0,010
	surprise	-5,121	0,010		surprise	-4,537	0,010
	trust	-4,749	0,010		trust	-4,716	0,010
ubi				**toyota**			
	anger	-5,307	0,010		anger	-3,998	0,011
	anticipation	-4,595	0,010		anticipation	-4,318	0,010
	disgust	-4,633	0,010		disgust	-4,261	0,010
	fear	-4,785	0,010		fear	-3,697	0,027
	joy	-4,876	0,010		joy	-4,468	0,010
	sadness	-4,388	0,010		sadness	-3,571	0,038
	surprise	-4,775	0,010		surprise	-4,694	0,010
	trust	-5,304	0,010		trust	-4,406	0,010

ADF tests for emotion values of tweet weighted means, weight = n. of retweets, bold values not stationary

ADF for news emotions

gaming	emotion	adf value	p-value	non-gaming	emotion	adf value	p-value
blizz				**amazon**			
	anger	-4,836	0,010		anger	-3,338	0,069
	anticipation	-3,545	0,042		anticipation	-5,373	0,010
	disgust	-1,442	**0,806**		disgust	-5,190	0,010
	fear	-4,543	0,010		fear	-4,241	0,010
	joy	-4,540	0,010		joy	-4,813	0,010
	sadness	-3,178	0,096		sadness	-2,908	**0,201**
	surprise	-4,508	0,010		surprise	-5,870	0,010
	trust	-3,987	0,013		trust	-3,446	0,052
cdpr				**apple**			
	anger	-4,486	0,010		anger	-3,535	0,043
	anticipation	-5,161	0,010		anticipation	-3,949	0,015
	disgust	-5,835	0,010		disgust	-3,213	0,090
	fear	-5,028	0,010		fear	-4,641	0,010
	joy	-4,444	0,010		joy	-3,738	0,025
	sadness	-3,335	0,070		sadness	-3,454	0,051
	surprise	-3,745	0,025		surprise	-5,170	0,010
	trust	-5,092	0,010		trust	-3,339	0,069
nint				**tesla**			
	anger	-4,226	0,010		anger	-0,861	**0,953**
	anticipation	-3,961	0,014		anticipation	-2,071	**0,547**
	disgust	-4,451	0,010		disgust	-1,817	**0,652**
	fear	-3,602	0,036		fear	-1,330	**0,853**
	joy	-4,227	0,010		joy	-3,688	0,030
	sadness	-4,382	0,010		sadness	-3,664	0,032
	surprise	-4,754	0,010		surprise	-3,260	0,082
	trust	-3,922	0,016		trust	-2,711	**0,283**
ubi				**toyota**			
	anger	-4,658	0,010		anger	-3,638	0,034
	anticipation	-3,210	0,091		anticipation	-4,418	0,010
	disgust	-4,282	0,010		disgust	-3,736	0,025
	fear	-4,985	0,010		fear	-3,626	0,035
	joy	-3,466	0,049		joy	-4,379	0,010
	sadness	-4,451	0,010		sadness	-3,666	0,031
	surprise	-3,380	0,063		surprise	-2,974	**0,174**
	trust	-4,100	0,010		trust	-4,833	0,010

ADF tests for emotion values of headlines means, bold values not stationary