LIBERA UNIVERSITÀ INTERNAZIONALE DEGLI STUDI SOCIALI

"LUISS - GUIDO CARLI"

DIPARTIMENTO DI ECONOMIA E FINANZA

Corso di Laurea in Economia e Finanza

US Market Indexes reactions after COVID news.

RELATORE:
PROF. FEDERICO CARLO EUGENIO CARLINI

CORRELATORE:
PROF. ALBERTO CYBO – OTTONE

CANDIDATO:
ANDREA CRISPINO

MATRICOLA:
739811

Anno Accademico 2021-22

Contents

INTRODUCTION

Sentiment analysis has gained prominence as a technique for analysing text data to understand the underlying sentiment or emotion of the text. In recent years, sentiment analysis has been applied to various domains, including finance, where it has been used to predict stock market returns. In the context of the COVID-19 pandemic, understanding the relationship between sentiment and stock market returns has become even more important. The present study aims to investigate the relationship between sentiment and stock market returns during the COVID-19 pandemic. Specifically, we examine the relationship between COVID-19 cases, words of stress and anxiety in WSJ news, negative tone in WSJ news, and the returns of the S&P 500, Dow Jones, and VIX. The S&P 500 and Dow Jones are two of the most widely recognized and followed stock market indexes in the world, while the VIX is a measure of the volatility of the S&P 500. We hypothesize that there will be a negative relationship between the independent variables (COVID-19 cases, words of stress and anxiety in WSJ news, and negative tone in WSJ news) and the returns of the S&P 500 and Dow Jones. This is based on the assumption that negative sentiment in the news would cause investors to be less optimistic about the future of the stock market, leading to lower stock returns. Conversely, we expect a positive relationship between the independent variables and the VIX. This is because an increase in negative sentiment and uncertainty in the news would lead to higher market volatility, which is reflected in the VIX. To test these hypotheses, we conduct a sentiment analysis on news articles published by the Wall Street Journal (WSJ) from March 2020 to March 2021. We use multiple and univariate regression models to examine the relationship between the independent variables and the stock market returns of the S&P 500, Dow Jones, and VIX. Sentiment analysis has been widely used to analyse text data in various domains, including social media, customer feedback, and news articles. In finance, sentiment analysis has been applied to predict stock market returns and to analyse the impact of sentiment on financial markets. Previous studies have shown that sentiment in news articles can affect stock market returns. In the context of the COVID-19 pandemic, sentiment in news articles has been shown to be an important predictor of stock market returns. The present study builds examines the relationship between COVID-19 cases, sentiment in news articles and the stock market returns of the S&P 500, Dow Jones, and VIX during the COVID-19 pandemic. In addition to examining the impact of overall sentiment, we also examine the impact of specific types of sentiment, such as stress and anxiety.

Chapter I

RELATED LITERATURE REVIEW

The first year of COVID-19 pandemic shock was quite a rollercoaster for financial markets, starting from an extremely quick drawdown in the first two months, one of the worst in history, that leading to one of the fastest recoveries ever seen on the markets. As it's common sense, "money is scared", and when a systemic shock occurs capitals run away as fast as possible from stock markets. But what really scares the markets? And what drives stock to movements in value?

In this section, I present some research papers that actually serve as references for the most important assumptions that actually represent the foundations for the work I have completed.

The COVID-19 sentiment analysis literature focused on finding what was leading markets to poor returns in various territories and using numerous variables that could have had a significant impact on stock returns, during the worst and more tragical months of the pandemic outburst. ***Zhi Su, Peng Liu and Tong Fang (2021)*** constructed a fictional fear index based on Internet search volumes of the Chinese local search engine, which is called PIF (pandemic-induced fear index). The purpose of this analysis was to empirically investigate the effects of the fear produced by the pandemic on Chinese stock market returns. The analysis was conducted by using a multivariate regression approach with a reduced bias estimation to address the problems created by a small sample bias. What they found is that the effects of PIF was persistent in time, with a significative and negative impact on stock markets returns overtime. A very interesting outcome produced by the research is that investors' pessimistic behaviour, which means their Internet search history, increases pandemic induced fear leading to poor stock market returns, this is quite an interesting result because it states that what leads Chinese market is the fear of the pandemic itself rather than the number of cases/deaths released by the government or economic fundamentals, investors are irrational and for the actors who play in the market is more important a general sentiment of fear than numbers and statistics, at least in China.

"In theory, there is consensus that the efficiency with which firms operate and make use of their resources determines their future cash flows, and thus their respective stock prices." ***Daniel Neukirchen, Nils Engelhardt, Miguel Krause, Peter N. Posch (2022, p.1)***

In their 2022 paper named “Firm efficiency and stock returns during the COVID-19 crisis” they try to investigate whether this statement was still a valid hypothesis during COVID-19 pandemic crisis by using a simple OLS regression model, so they use stock returns during the first year of pandemic, the so called “collapse period”, as an independent variable and efficiency ratios as dependent ones. Their results proved that there is a strong and statistically significant relation between stock returns and the efficiency that a firm was able to provide during the period. In a time of uncertainty and pandemic fear of the market, educated investors moved their capitals to, hence, invested in, stock that were more likely to produce positive and stable cash flows, trying to protect themselves from a strong crisis which was affecting mostly the real economy, indeed the capacity of less efficient companies to deliver strong results in the medium run. They found that firms with higher scores in efficiency ratios, experienced an outstanding 9.44% of higher cumulative stock returns during the year and that a long-short portfolio composed by efficient firms (long) and inefficient firms (short) was able to deliver a strong return of 3.53% weekly, a great demonstration of how much pandemic changed the way of thinking among investors, it is better to invest in a highly efficient company which will less likely yield poor results in tough times to protect yourself from an unexpected shock affecting the market, as it was the COVID-19 march outburst.

Stock market is influenced by various sources and one of them is media communication, especially during press earnings release. ***David Ardia, Keven Bluteau, Kris Boudt (2021)*** have analysed this phenomenon in their paper “Media abnormal tone, earnings announcements, and the stock market”. Their methodology consists of measuring the usual tone of the press for a specific company and then capture an ‘abnormal tone’ which is often used by them after quarterly results are released, so they look for the effects that this abnormal tone has on the stocks of the specific firm constructing a variable called “cumulative abnormal tone” (CAT). Subsequently another variable is defined by excluding from CAT the information that can be derived from the firm release such as ratios and quantitative data, which is called “residual cumulative abnormal tone” (RCAT). The analysis is conducted on non-financial firms and the period starts in 2000 and ends in 2016. CAT is first analysed around earnings announcements. This prediction is consistent with the fact that CAT is largely driven by information disclosed by the firm, such as the earnings surprise, tone in the press release, and tone in the call. Since CAT is also based on past values, it implies that new information embedded in CAT incorporates the past as well. It is important to note, however, that a large portion of the variance of the CAT remains unaccounted for, which supports the use of RCAT to analyse incremental information from media tone. The next step is to examine whether RCAT is able to convey incremental information about a firm's stock price reaction around a press release. There is a significant contemporaneous correlation between increased RCAT and excess abnormal returns: one standard deviation increase is associated with 0.84% excess abnormal returns. In addition, RCAT is capable of predicting price movements in the future. The month following the announcement, stock price movements associated with RCAT appear to have reverted, with an excess abnormal return of -0.36% associated with a one standard deviation increase in RCAT.. Providing information about corporate earnings is a primary responsibility of the news media. As a result of abnormal pricing reactions surrounding earnings announcements, and 20 days afterward, future price dynamics can be predicted by observing the tone of media articles around earnings announcements. On aggregate, market participants overreact to media reports and earnings calls since they provide additional information not found in earnings press releases and earnings calls.

"The pandemic is causing huge impact on real economic activity, though the extent of actual impact is yet unknown." ***Badar Nadeem Ashraf (2020***). This essay examines how the COVID-19 pandemic affected the stock markets. Using daily data on COVID-19 confirmed cases, deaths, and stock market returns from 64 countries from January 22nd, 2020 to April 17th, 2020, he discovers that the increase in COVID-19 confirmed cases had a negative impact on stock markets. In other words, stock market returns decreased as more cases were proven. He also discovers that the stock markets responded more forcefully to an increase in confirmed cases than an increase in fatalities. Additionally, according to the data, the initial days following verified cases and then between 40 and 60 days later saw a significant decline in market activity. The panel data method was chosen over the traditional event study methodology to examine how changes in COVID-19 confirmed cases and deaths affect stock market returns for several reasons. As a matter of fact, COVID-19 outbreaks do not occur as a single instance but rather over a period of days. Also, panel data regression is more accurate at capturing the changes over time in the relationship between different variables. Lastly, panel data analysis separates time series variation from cross-sectional variation in order to minimize multicollinearity, heteroscedasticity, and estimate bias. As more cases are confirmed, the country's stock market returns fall. In general, the data show that the stock market has reacted swiftly to his COVID-19 pandemic, with this reaction changing over time in direct proportion to the severity of the infection.

COVID-19 had a very deep impact on markets and increased volatility is one of the main direct effects brought by the pandemic. “The COVID-19 and stock market volatility: An industry level analysis” ***Seungho Baek, Sunil K. Mohanty, Mina Glambosky (2020)*** study was concerned with understanding the transition from lower to higher volatility as shown by a Markov Switching AR model. Researchers use the MS-AR (1) model to confirm that the U.S. stock market has experienced a regime change with the introduction of COVID-19. US stock market risks have increased significantly standing at what these findings uncovered. According to an analysis of developments across 30 different industries, all businesses were experiencing increases in overall and idiosyncratic risk. The systematic risks of aggressive businesses like cars and office equipment reduced significantly, while those of defensive industries like telecom and utilities increased significantly. Businesses may be protected/exposed to demand shocks by customers with lower/higher price elasticity. In order to better understand the causes of day-to-day fluctuations in volatility, the total risk is regressed on economically significant variables determined by ML selection methods. The findings indicate that volatility swings are more responsive to COVID-19 news than economic indicators. Furthermore, bad news is feared twice as much as good news, as evidenced by the fact that fatalities are feared twice as much as recovery news. Upon hearing about COVID-19, the market reacted in a positive-negative asymmetrical manner. The investigation revealed that out of the 30 industries, the impact of restaurants, accommodation, and motels, as well as natural gas and petroleum, is the greatest. In contrast, the brewing and alcohol industries have not been as significantly impacted. Major oil corporations noticed a significant change in the overall demand and variations in oil pricing. Lockdowns limited travel and reduced crude oil consumption, which caused a sharp decline in crude oil prices in the second quarter of 2020. Contagious diseases may lead to social isolation and business closures in the food, accommodation, gaming, fashion, and transportation industries, which are more likely to be impacted by social isolation and business closures. In fields including food processing, brewing and distilling, healthcare, medicine, and pharmaceuticals, COVID-19 exposure is less common. Both conventional and specialized foods saw year over year sales growth in the supermarket sector. Increased socially or online distributed operations could lessen the COVID-19 shock.

Natural language processing (NLP) has recently been used in finance and economics, which has led to a significant evolution in the ability to study how news affects stock values. In the scientific paper called “Machine learning sentiment analysis, COVID-19 news and stock market reactions” ***Michele Costola Michael Nofer Oliver Hinz Loriana Pelizzon (2020)*** a machine learning toolkit known as the "Natural Language Toolkit," which extracts sentiment from news, is used by researchers to elaborate COVID-19 news. Their study covers the period from January through June 2020 and examines 203,886 online articles posted on MarketWatch.com, Reuters.com, and NYTimes.com about the epidemic. According to the findings, sentiment score was significantly correlated with market return. As a result, when the market returns are positive (negative) and the news is good (negative), sentiment scores increase (decrease). In addition, authors find that Reuters and MarketWatch news sources are negatively correlated with market returns based on the variance of sentiments and news sources, respectively. According to this research, stock market returns are negatively impacted both by the increase in unpredictability of sentiment and the increase in news arrival.

In “Public Concern and the Financial Markets during the COVID-19 outbreak” ***Michele Costola, Matteo Iacopini and Carlo Santagiustina (2020)*** three search engine data sources from Google Trends are used to measure public concern during COVID-19 outbreaks in Italy, Germany, France, Great Britain, Spain, and the United States. It appears that the dynamics of public concern in Italy are reflected in those in other countries. This study shows that Italian index returns better explain stock index returns than those in other countries. A time-varying analysis shows that the most severe effects are experienced during each step of the Italian lock-down process. Certain topics may be indexed by Google Trends based on the relative web search volume of certain words or phrases. There are GT indices at the regional or global level. Interpreting the GT index is straightforward. The more meaningful the GT Index is, the more public attention the topic is getting. A lot of research has been done in recent years, suggesting that the informational content of Google Trends data can be used to explain and predict various aspects of the economy and finance. Financial markets can benefit from GT indices based on equity-related terms. This is because the GT index is more diversified from the benchmark market and can be used to measure sentiment and early market signals. Macroeconomically, the GT index has been used to construct an indicator of economic uncertainty that accounts for a wide range of macroeconomic variables. In this paper, the authors retrieve the GT Country Indices (GT-COVID-19) from January 2020 to April 2020 on the topic of coronavirus. GT data was also used to monitor MERS, Chickenpox, and influenza. The impact on financial markets during the coronavirus outbreak is being explored at the national level as a proxy for public concern. It's an interesting case study because pandemics like COVID-19 are so virulent. This represents a large exogenous shock to the economy and financial system that could not have been reasonably foreseen. As of May 1, 2020, six countries were the hardest-hit countries in the world in terms of confirmed cases: the United States, Spain, Italy, the United Kingdom, Germany and France. The research draws three conclusions. Firstly, they find that the Italian GT-COVID-19 index produces the majority of the GT indices for all countries considered. Part of this lead-lag relationship is due to Italy's involvement in her COVID-19 outbreak and lockdown measures since World War II, making it the first country in Europe to suffer from the disease. Additionally, several European governments introduced travel restrictions to and from Italy in the first weeks of the outbreak. As the COVID-19 pandemic spread to other countries, new cases were reported, and lockdown measures were put in place. Therefore, the delayed responses of these indicators are probably caused by the above turns. The second step examines whether the GT-COVID-19 index returns are explained by the GT-COVID-19 index. Consequently, the delayed reaction of these indexes was likely caused by the aforementioned twists. A GT index can be interpreted as a measure of uncertainty and perceived risk related to Coronaviruses, given that epidemic diseases by definition are adverse events. It has been shown that GT-COVID-19 indices are an important component of explaining the dynamics of stock market returns in Italy, Spain and Germany. Interestingly, replacing the national GT index with the Italian index increases the exposure of all the markets considered to COVID-related societal concerns, disrupting the dispersion of equity index returns described. This highlights that the outbreak of the pandemic in Italy may have played a role in the public perception of the severity of the pandemic. Finally, in this context, a time-varying analysis is conducted to examine the impact of the GT-COVID-19 Index on financial markets over time. Interestingly, they identified the most severe impacts at each stage of the lockdown process in Italy, and while his COVID-19 presence in China was already known at the end of December 2019, the pandemic had spread to Europe. Little impact was seen before it started.

Micheal Donadelli, Renatas Kizys and Max Riedel in their "Globally Dangerous Diseases: Bad News for Main Street, Good News for Wall Street?" (2016) try to evaluate the balance between two contrasting effects of outbreaks of globally dangerous disease. Numerous studies have demonstrated that agents' emotions may have an impact on their financial decisions ***(De Long et al.,1990; Cen and Liyan-Yang,2013; Kaplanski and Levy,2015)***. The behavioural finance literature shows a strong correlation between stock returns and investor sentiment. Investor sentiment is often defined as a view about future cash flows and investment risks that is not supported by the facts ***(Baker and Wurgler 2007, p. 129)***. Regardless of how favourable or unfavourable an event is, it has a significant impact on stock market prices and investors' investment decisions. Early research has revealed, for instance, that sunshine, a well-known mood enhancer, positively corresponds with daily stock prices ***(Saunders, 1993; Hirschleifer and Shumway 2003)***. Lunar phases might have an impact on stock market performance, claim ***Yuan et al (2006)***. International athletic events, especially soccer games, have a big impact on investors' mood ***(Edmans et al. 2007; Kaplanski and Levy 2010)***. According to ***Kaplanski and Levy (2010)***, major aviation tragedies frequently result in negative attitude within two days of the tragedy. ***Hornat and Huizinga (2015)*** examine how statements regarding the launch of the European Financial Stability Facility affect the price of bank shares. The main contention of this study is that some interest groups, such as stock market participants, may interpret significant, catastrophic occurrences as good news. This study's main contribution is to investigate the relationship between investor sentiment and pharmaceutical stock prices using a novel mood variable (hereafter referred to as disease-related news, or DRNs), which is derived from WHO alerts and media reports on dangerous diseases that are spread around the world. In this regard, the hypothesis is that fear of globally severe diseases will have a negative impact on investor pessimism, but DRNs will have a favourable effect on investments in pharmaceutical stocks. The theory holds that, despite the spread of dread and unfavourable sentiment brought on by deadly diseases, investors predict pharmaceutical companies' cash flows to rise as they offer new medications intended to combat the new pandemic disease. Investor perception of the success of pharmaceutical companies may consequently have an impact on investment decisions. However, little research has been done on how pharmaceutical companies do on the stock market ***(Himmel-mann and Schiereck 2012; Theodossiou and Theodossiou 2014)***. There hasn't been a study on the connection between pharmaceutical stock returns and investor sentiment in the literature, apart from ***Huberman and Regev (2001)***. ***Huberman and Regev (2001)*** claim that investor fervour resulted in a significant advancement in cancer research through a case study. The stock price increased more than five times over the preceding five months when EntreMed's breakthrough was publicized in the popular press five months after it was published in Nature. Notably, their research does not examine how investor mood affects stock prices and returns, generalize to the entire pharmaceutical industry, or devise tactics for responding to breaking news about major world diseases. The goal of this study is to weigh the conflicting effects of international disease outbreaks. A pandemic disease may cause panic among stock market investors and the broader public, which has the effect of causing pharmaceutical stock sentiment to be negative. On the other side, a positive mood effect brought on by the outbreak of a pandemic disease is anticipated to benefit the price of pharmaceutical stock. In terms of methodology, capturing two competing effects (positive and negative) provides a unique framework that eliminates the possibility of spurious correlations. More specifically, the study explores the following questions related to pharmaceutical stocks that have never been explored before: Is there a correlation between the DRN-induced gauge of investor fear and lower stock market returns for pharmaceutical companies? What is the effect on pharmaceutical stock prices if a globally dangerous disease breaks out? Third, will investor sentiment (either optimism or pessimism) about pharmaceutical company performance be sustained over time? What is the difference between large pharma stock prices and small pharma stock prices for DRN? To address these issues, four different investment portfolios based on the stock prices of 102 publicly traded US pharmaceutical companies are constructed. Additionally, the S&P 500 Information Technology Index is considered solid. The empirical strategy uses two commonly used methods to assess how investor sentiment affects pharma stock prices post-aDRN. Event studies and regression analysis. After the DRN, they clearly see positive and sustained investor sentiment for returns on pharma stocks. Disease outbreaks create positive beliefs about R&D investments. It is also worth noting that investor sentiment is reflected in the presence of information ***(Palomino et al. 2009)***. Additionally, they employ the VIX as a proxy for investor fear to build an index that measures fear. Pharmaceutical stocks are always subject to significant negative impacts of the index. Small companies tend to be more sensitive to sentiment than larger companies. The results of this study are consistent with those of ***Qiu and Welch (2004)***, who found that small firm returns serve as predictors of investor sentiment under certain conditions ***(Edmans et al. 2007; Baker and Wurgler 2006)***. An important aspect of their study is that their results are supported by robustness checks. As a result, an investor can make a profitable trading strategy by taking a long position in the pharma stock portfolio and a short position in the VIX. It's important to notice that these strategies have important consequences. The authors of this paper examined whether DRNs are associated with a significant positive sentiment effect among investors interested in US pharmaceutical companies and, therefore, their stock prices, based on the abundance of recent behavioural finance studies which indicate that certain events (e.g., St. Patrick's Day, Yom Kippur War, Rosh Hashanah, International Sporting Games) may have a strong impact on investors' moods. There are only a few large pharmaceutical companies that are producing vaccines for globally dangerous diseases. According to this research, the number of vaccine producers in the United States has decreased dramatically since 1967 – there were 37 vaccine producers ***(Masignani et al. 2003)*** – to just over a dozen in 2016. These companies' cash flows are expected to be the basis for rational investors' trading strategies. Two reasons argue against rational trading. In the wake of DRNs, vaccine developers are likely to face high levels of uncertainty surrounding their distant cash flows. In addition, smaller pharmaceutical companies may not have enough resources to invest in large-scale research and development. New pharmaceutical products require an investment of at least USD 850 million for licensing ***(Masignani et al. 2003)***. The stock price of pharmaceuticals should not be affected by DRNs, therefore. It is therefore of no use to DRNscan if it doesn't alter investor sentiment about future performance and result in irrational trading ***(Kaplansky and Levy 2010)***. It enables to account for two conflicting sentiment effects that may be induced by fear and anxiety due to DRNs. Results are as follows. It is observed that pharmaceutical companies' stock returns are positively impacted by DRNs. In a portfolio of small stocks, this effect is relatively stronger. DRN effects also appears to last for several days. An ad hoc fear gauge index appears to have a significant negative effect on pharmaceutical companies' stock returns, assuming that DRNs can also generate panic and anxiety among international investors. Similarly, optimism and pessimism - induced by DRNs - are significantly more powerful in the pharmaceutical industry when compared to large stocks. This research has shown that the pharmaceutical industry's portfolio investment decisions are significantly influenced by optimism and pessimism. Investing institutions and individual investors, portfolio managers, analysts, and pharmaceutical companies will find these findings of utmost importance and practical value. Consequently, investment opportunities that can be exploited are identified. The results of the exercise demonstrate that trading strategies yield significant and positive returns when involving pharmaceutical stocks and the VIX volatility index. The results of this activity are also beneficial for portfolio managers who craft expert and knowledgeable counsel for investors. The pharmaceutical industry's investment prospects can be thoroughly analysed by financial specialists. Additionally, pharmaceutical companies can gain from rising market valuation and, consequently, from less expensive funding sources for R&D investment when a global disease breaks out and spreads quickly by issuing and selling new stocks to investors.

According to ***Kaplansky and Levi (2010)*** "Behavioural economic studies reveal that negative sentiment driven by bad mood and anxiety affects investment decisions and may hence affect asset pricing." In this study the effect of aviation disasters on stock prices are examined. The average market loss from an aviation disaster is over $60 billion on average, while the actual loss is estimated to be no more than $1 billion on average. In two days a price reversal occurs. Small and riskier stocks, as well as firms from less stable industries, seem to be more affected by it. After aviation disasters, implied volatility increases without actual volatility increasing as a result of this event effect. People who are anxious are more likely to be pessimistic concerning future returns, to take fewer risks, or both. Negative emotions such as anxiety can negatively impact investment decisions and results. In this study large-scale aviation disasters are examined. Investing in risky assets becomes less appealing during aviation disasters due to people's heightened anxiety. After an aviation disaster, stock market returns are likely to be negative. The study found a mean-reverting reversal effect two days after an aviation disaster accompanied by a significant event effect that negatively impacted stock prices for a short period. Investors' reactions to aviation disaster news can be interpreted in a number of ways:

1. Investors who do not act logically after hearing immediate news about aviation disasters (Lee, Shleifer, and Thaler, 1991) revert back to their normal behaviour after two days. Furthermore, sophisticated investors may profit from the relatively low prices, resulting in a price reversal.

2. When negative sentiment is present, investors display a greater degree of risk aversion than when it is absent. Within the framework of expected utility, results can be explained. Although moods affect preference and, in particular, risk aversion, the switch between them falls into behavioural economics.

Various ways are explored in this study to examine both event effects and reversal effects. Under rigorous robustness tests, the effect remains highly significant. The average return rates fall sharply in the first day after a disaster, when images and stories about the disaster abound in the media (rather than after disasters are widely known). As a result of this decline, the average daily return was almost 10 times higher during the observed period. Consequently, each aviation disaster causes a market loss of more than $60 billion on average, while the upper limit on economic loss due to these events is about $1 billion. Moreover, the event effect is followed by a reversal effect. It is about half as large as the decline on the first day's returns on the third day after the event. Also, the reversal effect occurs after the event effect. The first daily return on the 3rd day after the event is about half the size of the drop. This reversal trend continues for the next few days. The market fully returns to the median around 10 days after the drop. The coexistence of reversal and event effects can provide insight into the nature of behaviour. If the market decline was actually caused by a disaster and not by the influence of sentiment or fear, we would not expect a reversal effect. The main hypothesis is that excessive fear induces a price reversal. . A price reversal occurs when the fear subsides or when experienced investors capitalize on the effect. A nearly complete price reversal is further proof of the hypothesis. Several supplementary analyses were produced to further investigate the impact of the event. The first part, based on ***Baker and Wurgler's (2007)*** proposal to use his VIX and VXO versions of the Fear Index to measure market sentiment, correlated stock price declines with perceived volatility increases. Indicates that the Given that actual volatility has not increased, this suggests that fear influences perceptions of volatility after plane disasters. Moreover, as ***Baker and Wurgler (2006)*** predicted that the sentiment effect tends to be highly subjective and difficult to adjust, the magnitude of the sentiment effect may vary between portfolios constructed on the basis of volatility, size, and industry. to see if there is a difference. Inventory grows. This effect was robust and evident in all portfolios studied, with smaller firms, volatile stocks, and firms in volatile industries more likely to experience larger event effects, ***Baker and Wurgler (2006)*** are also consistent with the theory. Investor reactions to events can lead to flight to safety in a number of ways. This strategy is only viable for the short term. It can also be used to reallocate some investments from riskier assets to safer assets such as short-term securities and US dollars that are considered safer assets. To test whether the effect has ripple effects on bond and currency markets, we examine various maturities of government bonds and exchange rates for the US dollar. The change in price in the expected direction is insignificant. In this case, the flight to safety can occur across different assets, resulting in dilute impact on each asset. Alternatively, investors hold more cash in their day-to-day trading activities and postpone investing in risky assets. The results show that information flow plays an important role in the investor's decision-making process, has a psychological impact on investors, and can be modified over time to make the market more efficient. increase. One day after the plane crash occurs, they observe the event effect for two days. The correction process began three days after the market opened. Their analysis also found a strong correlation between the magnitude of an event's impact and its association with a catastrophe for US investors. In particular, disasters targeting Americans will have the strongest impact, followed by disasters targeting relatively weaker Europeans, and finally disasters targeting others. In addition to public attention, media coverage and rapid flow of information, this association may be related. Over the last 30 years, it has been observed that the impact of events on the rate of information flow increases faster than in the last 30 years. This result is consistent with the fact that detailed news has become available much more quickly over the last 30 years than in previous periods. Similarly, the event effect of a catastrophe on the continental United States was found to have a faster impact on the US market than if the effect corresponded to a catastrophe farther away. Finally, they show that the impact is much weaker for transport, industrial, and other disasters. This is because these disasters most likely lack the psychological resonance found in the case of aircraft disasters. According to the study, a plane crash is followed by a negative stock market return that backfires two days later. They turn to behavioural economics for an explanation, because the temporary drop in the stock market is more than 60 times his direct economic loss. In fact, psychology research shows that exposure to media coverage of plane crashes induces moodiness, anxiety, and fear, making people more pessimistic, risk-averse, or both. will be as a result, the study's hypothesis is that heightened fears after a plane crash led to a short-term decline in demand for risky assets, affecting stock prices. The stock market will reverse once the fear subsides or experienced investors capitalize on its effects.

Staying in the context of investor's sentiment, according to ***Kaplansky, Levi et al. (2015)*** in "Do happy people make optimistic investors?", investment plans, return and risk expectations, and subjective mood-makers are queried. Return and risk expectations are systematically affected by non-economic factors, but the return effect is most important. In addition to economic factors, other factors also influence investment plans. There is a significant correlation between sports performance and general emotions. The winter blues hypothesis is supported by lower yield expectations in fall-affected people. Correlations between stock prices and non-economic factors (such as general investor sentiment) have been found in several market-based empirical studies. Various approaches are used in the literature to define investor sentiment. The ***Baker and Wurgler (2007)*** definition is probably the most widely used. Investment sentiment is defined as "an investor's unsubstantiated beliefs about future cash flows and risks". This study therefore independently analyses the direct relationship between sentiment and risk and return expectations. In this article, some hypotheses are tested at the individual investor level rather than through aggregated empirical studies of market sentiment. The first analysis examines the relationships between mood-shaping factors reported by individuals and parameters related to general mood (mood). A second objective is to examine the relationship between sentiment-forming factors and individuals' subjective assessments of the market in terms of expected return (risk) and volatility (uncertainty). The final part of the study examines the impact of sentiment on actual investment plans. Using data collected from a sample of people who actually trade in the stock market, the study analyses, at the individual level, sentiment-forming factors, subjective estimates of future returns and risks, and investment plans. It is the first study to investigate a direct relationship between Individually. Based on the data collected from retail investors, it is unlikely that the observed correlations were caused by other common factors when using data from retail investors who completed the questionnaire on different months, days and times. very low. Sentiment therefore appears to be causal to individual return expectations, as we show. A cross-sectional analysis was performed using the Longitudinal Internet Research for Social Sciences (LISS) panel of his CenterER data from the University of Tilburg covering approximately 5,000 households in the Netherlands. About 900 people were sent his three waves of questionnaires a year after screening those holding stocks in their portfolios. Dutch and US stock markets were analysed for subjective return and risk forecasts for the next month and next year. Some mood-shaping factors have been reported on an individual basis, while others have been correlated with stock prices in previous studies. Aside from general sentiment (which has never been tested before), recent scores for an individual's favourite sports team, perceptions of the current weather, and whether they suffer from seasonal affective disorder (SAD), in particular, All factors. Consider and consider. The impact of SAD season on mood can be tested using three questionnaires distributed throughout the year. The fact that the questionnaire was completed on different days of the week allows the authors to use the days specified in the questionnaire to test the impact of weekday mood effects. In contrast to market-based empirical studies, they also have data on trade-free weekends, which are generally associated with positive sentiment. Finally, each participant provided information about their plans to buy and sell stocks for the next month. As a result, each individual simultaneously reports on subjective stock market expectations, feelings about the current mood, and investment plans. Individuals' expectations of future stock market returns are strongly related to several mood-shaping factors in addition to individual mood (measured using common sentiment variables). Stocks are expected to produce higher returns on average and are bought rather than sold when sentiment is more positive. Using the cumulative distribution of expected returns for all individuals, they found that high noneconomic mood scores had systematically higher subjective return expectations than subjects with low mood values. As a result, the cumulative distributions of the two groups do not overlap. There are many interesting results from this study. First, expected risk has a much smaller impact on expected return than sentiment. They found that in only a quarter of the cases, lower volatility expectations are associated with positive sentiment. Additionally, the authors note that investor expectations, recent trading activity, and future investment plans are generally consistent among retail investors. After all, psychology studies show that weekend moods are relatively optimistic. Weekend sentiment, however, does not appear to be associated with high-yield expectations. Most subjects are found to participate in laboratory experiments making irrational, contradictory, and unhelpful decisions. Lack of rational investor behaviour has also been shown to correlate with non-economic factors such as weather conditions, seasons, and sporting events. This study investigates the relationship between subjective return and risk expectations, emotions created by various sentiment creating factors (which contribute to mood and general feelings), and the investment plans of stock market investors. For the study, a representative sample of approximately 5,000 households was surveyed, and 1,465 questionnaires were completed by investors. The statistical analyses were based on these households' responses. SAD is also testable because the survey was conducted over three different seasons. Happier subjects are optimistic not only about the Dutch stock market, but also about the US stock market. Researchers found that the better you felt overall and the better your favourite sports team performed in the days before you completed the survey, the higher your expectations of stock market returns. The authors say that when the compiler combines these individual factors into a sentiment index, the overall sentiment is positive, the expected return is high, and the expected risk is low. In most cases the return effect is more severe and significant, while the risk effect is in the expected direction, but only in one significant case. Return expectations are influenced by SAD, which has been shown to correlate with sentiment. Coefficients for SAD patients in different seasons are very different. This information, coupled with the fact that the SAD coefficient is negative in the fall, suggests that patients with SAD should expect lower returns in the fall than at other times of the year. As a result, SAD is an important factor in shaping subjective expectations. Finally, in line with previous psychological research on the effects of weekdays, weekday moods are lower than weekend moods. However, this weekday deterioration in sentiment is not in line with lower return expectations. While some results are consistent with intuition, the magnitude of this phenomenon is surprising. For comparison, the cumulative distribution of expected returns for those feeling good is shifted significantly to the right compared to the distribution for those feeling bad. The result is her FSD of the expected return distribution induced by individual emotions, not just average effects.

Enrico Onali (2020) is one of the first researchers to address the increase in volatility of stock market indexes in the US in his “Covid-19 and stock market volatility” academic paper. The author examines the impact of Covid-19 cases and related deaths on the US stock market using changes in trading volume and volatility expectations, as well as day-of-the-week influences (Dow Jones and S&P 500 indices). Based on GARCH (1,1) analysis of data from April 8, 2019, to April 9, 2020, the results show that stock market returns in the United States do not change in response to changes in the number of cases and deaths in the United States and six major countries affected by the Covid-19 crisis, with the exception of China's reported cases. Some countries appear to have a positive effect on the conditional heteroscedasticity of Dow Jones and S&P 500 returns. According to VAR models, deaths in Italy and France are negatively correlated with stock market returns, while VIX returns are positively correlated. According to Markov-Switching models, the VIX's negative impact on stock market returns will have tripled by the end of February 2020. Why are Covid-19-related deaths and cases affecting the stock market in the United States? The author addresses this question using the GARCH (1,1) model, the conditional mean equation, and the conditional heteroscedasticity equation, while taking changes in reported deaths and cases into account. His projections account for volatility expectations as well as changes in trading volume (proxied by log returns for VIX prices). Conditional heteroscedasticity is influenced by mortality and cases in Italy and Spain, despite little evidence that this influences Dow Jones and S&P 500 returns. There is a small but rapidly growing body of literature on the effect of Covid-19 on the stock market ***(Gormsen & Koijen (2020)***, ***Yilmazkuday (2020)***, and ***Baker et al. (2020)***). According to ***Yilmazkuday (2020)***, the number of deaths related to Covid-19 has an impact on S&P 500 stock prices. ***Baker et al. (2020)*** discovered that Covid-19 news has a much greater impact on stock market volatility than other similar diseases by analysing textual data. For the first time, this paper estimates the conditional mean and volatility of Dow Jones index returns in relation to daily reported Covid-19 cases and deaths. The data in the report cover cases and deaths in the United States, China, France, Iran, Italy, Spain, and the United Kingdom up to April 9, 2020. Furthermore, the impact of the Covid-19 crisis on volatility expectations and stock market returns has been investigated. The research question addressed in this paper should be noted by both investors and policymakers. According to a Wall Street Journal article, the Dow Jones Industrial Average will fall by more than 12% on March 16, 2020, making it the Dow Jones Industrial Average's second worst day in its 124-year history. There are several possible explanations for the market's extraordinary volatility, but none are sufficient to explain the phenomenon. "High-volatility periods are typically caused by economic and political instability." In an effort to calm the stock market, the US government implemented a series of measures in March 2020. According to ***Gormsen and Koijen (2020)***, economic relief programs announced on March 13 and a fiscal stimulus package announced on March 24 failed to improve expectations for short-term growth (2020). The CBOE's volatility index (VIX), also known as the "fear index," has risen as a result of the Covid-19. This has resulted in significant drops in the Dow Jones and the S&P 500. As of 9 April 2020, the VIX had fallen to its lowest level since 6 March, when it fell to 41.67, its lowest level since 27 March. Given these findings, it appears that the stimulus package announced on March 24th may have contributed to a decrease in volatility expectations. Despite the fact that Covid-19 cases and deaths occurred outside of the United States, his findings suggest that the change in the number of cases and deaths had an impact on the Dow Jones and S&aP 500. As long as conditional volatility changes are considered, there is little evidence that these changes have influenced stock returns. In the relationship between volatility expectations and stock market returns, the last week of February 2020 marks the transition from a low volatility regime to a high volatility regime. This paper investigates the impact of Covid-19 on US stock market returns using the Dow Jones and S&P 500 indices as proxies. Covid-19 is a major global crisis that has affected a large portion of the population in countries such as China, Italy, Spain, the United Kingdom, Iran, and France in the first three months of 2020. Aside from the number of reported cases (in logs), there is no effect on US stock market returns for China. Conditional heteroscedasticity, on the other hand, had a positive impact on the Dow Jones and S&P 500 returns. According to VAR models, reported deaths in Italy and France have a negative impact on Dow Jones returns, while VIX returns have a positive impact. The negative impact of VIX on stock market returns tripled at the end of February, according to Markov-Switching models.

Smales (2020) has analysed the effects of COVID-19 on stock markets through how much investors have been paying close attention to the COVID-19 outbreak and its consequences. Its effects have hit every industry. According to the author, the increase in investor attention could have explained stock returns across sectors during this unusual period. He follows **Da et al. (2015)** methodology by using Google search volume (GSV) as an indicator of investor attention across 11 sectors. According to the findings, increased COVID-19 attention has a negative impact on US stock market returns. However, some sectors appear to benefit from the increased attention. Due to household and government spending during the crisis, consumer staples, healthcare, and information technology outperformed the overall market during this period. The information discovery hypothesis would explain such results if investors were seeking information online to understand COVID-19's impact on stock sector performance. COVID-19 has affected a large number of people due to its significant economic and social impact. Retail investors have lost a significant amount of money in a short period of time as a result of the transformation of the health crisis into a financial crisis, risk has skyrocketed ***(Zhang et al. 2020)***, and market volatility has reached virtually unprecedented levels. Due to the increasing number of confirmed cases, market liquidity and value have declined significantly ***(Baig et al. 2020)***, outstripping any previous public health emergency response ***(Schell et al. 2020)***. Initially, Chinese stocks ***(Awadhi et al. 2020)***, US stocks with international exposure ***(Ramelli and Wagner 2020)***, and countries affected by the 2003 SARS outbreak ***(Ru et al. 2020)*** were the most affected, but they later spread across the globe and were negatively portrayed by the media ***(Ali et al. 2020)***. ***(Haroon and Rizvi 2020)***. By the end of the sample period (end-May 2020), global markets had recovered a significant portion of the losses sustained at the height of the crisis (mid-March 2020), but the extent of the economic fallout and when markets will revert to normal remained unknown. They contribute to the discussion of how COVID-19 affects financial markets by discussing it through the lens of investor interest. Investors face cognitive constraints as a result of the vast amount of information available to them ***(Kahneman 1973)***, a problem that is more prevalent than ever in the age of the Internet and social media. As a result, investors must select from selective information to process ***(Peng and Xiong 2006)***. If this is done, investors may underreact when important information is delayed and overreact when irrelevant information is given undue weight ***(Lim and Teoh 2010, Barber and Odean 2013, Hirshleifer 2015)***. Two main theories have been proposed to explain why investor attention affects stock returns. Both consider the concept of scarce attention. The first is based on the premise of asymmetric choice, as proposed by ***Odeon (1999)*** and ***Barber and Odean (2008)***. The issue is that retail investors lack the resources to consider all possible investments when purchasing stocks, so they concentrate on the stocks they are most interested in acquiring. Retail investors, on the other hand, tend to see only stocks they already own when trying to sell them. Institutional investors, who have access to more resources, can shorten stocks, and screen investment universes, do not face the same constraints. Alternatively, researchers have argued that investor attention promotes information discovery, causing market pressure and increasing efficiency ***(Vlastakis and Markellos 2012 and Vozlyublennaia 2014)***. Furthermore, this second theory is consistent with the idea that in order for prices to respond to new information, investors must pay attention to it ***(Huberman and Regev 2001)***. This study attempts to empirically resolve the tension between the two competing theories, and the results show that the hypothesis of information discovery is more likely to be true. Several indirect indicators are offered by ***Da et al. (2011)*** to measure investor attention, but Google Search volume offers several advantages over other indirect measures. He used Google Search Volume (GSV) as a proxy for investor attention to investigate whether specific stock market sectors are likely to be impacted by increased investor attention during the crisis. The study investigates the impact of investor attention across stock sectors during the COVID-19 pandemic, a simple but critical research question. Is the effect variable? Several studies have used GSV to examine investor interest in financial markets. To accomplish this, the authors examined stock markets in North America ***(Da et al. 2011, 2015, Ding and Hou 2015, Bijl et al. 2016, Tang and Zhu 2017)***, Europe ***(Bank et al. 2011, Aouadi et al. 2013, Kim et al. 2019)***, and Asia ***(Takeda and Wakao 2014, Tantaopas et al. 2016).*** When investors pay attention (GSV rises), more information flows into the marketplace ***(Smith 2012)*** and is incorporated into prices more quickly ***(Andrei and Hasler 2014)***, resulting in increased trading activity ***(Preis et al. 2010 Bank et al. 2011, Takeda and Wakao 2014 Kim et al. 2019)*** and higher volatility ***(Vlastakis and Markellos 2012, Aouadi et al.***

2013, Dimpfl and Jank 2015, Goddard et al. 2015, Kim et al. 2019). Unless new information is released, spikes in GSV generally follow contemporaneous market responses that are then reversed ***(Da et al. 11, 2015, Tang and Zhu 2017, Heyman et al. 2019, Smales 2020)*** unless new information is released ***(Cheng et al. 2019)***. According to ***Vozlyublennaia (2014)***, market volatility increases as a result of investor attention, which improves market efficiency. GSV does not appear to have a consistent influence on stock returns. The existing literature shows a positive influence on returns ***(Bank et al. 2011, Da et al. 2011, Tang and Zhu 2017)***, a negative influence ***(Vozlyublennaia 2014, Da et al. 2015, Bijl et al. 2016, Chen 2017)***, and no influence ***(Preis et al. 2010, Takeda and Wakao 2014, Kim et al. 2019)***. The authors provide additional evidence of a negative relationship to back up their claim. ***Takeda and Wakao (2014)*** and ***Da et al. (2011, 2015)*** examined individual firms rather than stock market sectors. GSV-return heterogeneity has received little attention. ***Da et al. (2011)***, for example, argue that small stocks and those traded more frequently by retail investors have a stronger relationship, whereas ***Da et al. (2015)*** demonstrate how beta, volatility, and downside risk (which is typically small) play a significant role. None of the studies investigate whether the stock sectors of firms influence the relationship. To fill this research gap, GS is examined across various stock market sectors. The sample period used, from December 2019 to May 2020, is relatively short, but it provides a unique opportunity to see how the world's media, financial markets, and general population focused their attention on the single theme of "coronavirus" during a period of extreme market stress. Such times are interesting to study because they are associated with increased sensitivity to news ***(Garcia 2013)***, as well as anxiety and sadness ***(Smith and Ellsworth, 1985)***. The returns of stocks in the 11 GICS-defined US stock sectors are a major focus of the analysis. According to the literature, investor attention has primarily been focused on broad market indexes and individual stocks as a whole. He can examine investor attention in greater depth by focusing on sectors. A coronavirus pandemic, for example, poses a significant risk to some industries that stand to benefit (those providing home-based services), while others, namely international travel and tourism, would suffer. The author focuses on US stocks because it is estimated that the US market capitalization represents approximately 50% of all global market capitalization. A brief examination of global sectoral returns (based on MSCI sector indices) concludes the analysis as a robustness check and indication of generalizability. This paper makes three contributions. First, it contributes to the debate over the direction of the GSV-return relationship by providing empirical evidence to support the investor attention theory. Second, looking at GSV in relation to returns helps the authors to understand that the relationship between the two is not simple or straightforward. In conclusion, the authors believe they are the first to study this specific relationship in connection to the unprecedented market circumstances brought about by the outbreak of COVID-19. The information discovery hypothesis seems to provide a better justification for the findings that were presented in the empirical analysis. Retail investors are using Google searches to understand the coronavirus' impact on stock sector returns. They observed that when investor awareness about the coronavirus increased (as indicated by a rise in Global Search Volume, or GSV), it was generally linked to negative equity returns. However, there might be some benefits associated with changes in household spending and gains in government spending. These sectors can be considered “essential” and, in some cases, benefit from government spending that is relatively stable when compared to other sectors. Interestingly, there appears to be no relationship between the number of reported cases and deaths related to COVID-19 and stock market returns. This suggests that investors may be more interested in looking forward to future information instead of being focused on the current levels of COVID-19 cases. The researchers do not only attribute stock return heterogeneity during the crisis to investor attention, but provides evidence to support the claim that it is a significant contributing factor. The outbreak of COVID-19 has had an unprecedented impact on global stock markets, with the pandemic causing widespread panic and no sector escaping its effects. As the virus continued to spread and more and more everyday activities were impacted by political decisions related to the pandemic, investors began to take more of an interest in COVID-19. The findings from this study would suggest that when there is a rise in investor interest, as indicated by an increase in Google searches, it has a negative effect on stock returns in the United States market, as well as globally. This could help to explain why there are differences in returns across different sectors of the stock market. Some sectors are seeing an increase in interest, while others are facing more difficult times ahead, depending on how much they rely on household and government spending. Outstanding performance is usually concentrated among a few individuals. This finding suggests that individual investors are behaving rationally and in line with the information discovery hypothesis. They are looking for data that will help them comprehend the effect of COVID-19 on the relative moneymaking capability of different equity sectors.

Chapter II:

HISTORICAL INTRODUCTION AND DESCRIPTIVE STATISTICS

2.1 Market indices: from the beginning to their dissemination

Stock indices are the most extensively used economic data on a global scale. Stock indices make it feasible to monitor the price movement of stocks and other financial assets traded on certain stock exchanges, such as the New York Stock Exchange, across time. They are developed, computed, and distributed by the commercial sector. Numerous of these indicators are computed and published daily or multiple times per day during business hours by stock exchanges, investment firms, and other financial organizations, as well as by newspapers. These indexes indicate in a sensitive and sudden manner the market's sensitivity to shifting economic and political situations. The aforementioned developments are also reflected in the price fluctuations of industrial stocks, technology, transportation, and utilities. Equity indices are not envisioned as "portfolio indices portfolio," i.e., as indications of the performance of real portfolios or of the actual holdings of assets by an individual or entity like a pension fund. However, they are commonly used as a reference for making investment choices, managing investment portfolios, and analysing their performance. In addition, they serve as a baseline for assessing the risk associated with various asset types. They are commonly used by investment fund managers and researchers to determine long-term rates of return, to test financial theories such as the 'Capital Asset Pricing Model,' and to investigate economic cycle changes. Since the early 1970s, stock market indices have also served as a benchmark for a new form of shares known as Exchange Traded Funds (ETFs), such as the S&P/TSX 60 Index Fund. ETFs are traded in the United States and other regions of the world. These collective index-based funds are traded like common stocks. In other words, they are similar to mutual funds in that they are meant to imitate or track the fluctuations of certain stock market indexes, such as the S&P 500 and the Nasdaq 100. Market indices are of vital significance to macroeconomists, financial economists, and all participants in the financial sector since they allow for the observation and analysis of the behaviour of financial players and the development of the economy. Financial players and the development of the economy, assuring uniformity and consistency of analysis throughout the medium to long term. In order to evaluate the performance of stock markets, a vast array of indexes have been developed throughout history. Recent years have seen a remarkable expansion of this diversity. In economics lectures and textbooks, the mechanism of index creation has not yet received enough consideration. Indeed, time series of data are utilized for reasons that are often vastly different from those for which these indices were designed; as a result, a lack of vigilance may lead to findings that are potentially misleading. For instance, the recent return to the study of the development of financial markets over the long term, long-term returns on investments, and international comparisons of market performance appears to be based on rather uncertain premises and rather uncertain foundations, e.g., the studies of (Goetzman 2000) and (Rajan 2003). After the introduction of indicators by successful journalists at the end of the 19th century, they attracted the attention of quite disparate groups of individuals ***(Hautcoeur,2006)***. Indeed, the building of the index became a complicated and specialized work due to the fact that academics of such indicators had varying objectives and interests. Surprisingly, the study of indices did not rely originally on the significance of the stock market in the world of finance (for the financing of companies, for the portfolio choices of savers or for the decisions of investment banks). The majority of early attention came from economists who saw the stock market primarily as a barometer or indication of the macroeconomic climate.

2.2 Financial Indices

Prior to the First World War, portfolio strategy was already an almost independent literature, at least in the business press. However, it lacked a serious theoretical and quantitative basis ***(Hautcoeur,2006)***. The growing importance of finance in business and personal savings directed towards its development as an academic discipline. In the United States, business schools became involved and developed doctoral programmes and academic journals on economics. ***(Wharton 1921, Harvard 1922, Harvard Business Review 1922, Journal of Business University of Chicago 1928).*** Academics and students began to investigate issues such as investor behaviour (portfolio optimisation) or managers' choices (among various financial resources). This did not happen in France, ***(Hautcoeur,2006)*** where business schools were deeply separated and distrustful of the academic community. ***(Fisher 1925)*** was again a pioneer in trying to evaluate the performance of various investments and studying the possibility of predicting the stock market. Another example is ***(Jackson 1928)***. But the greatest impetus came from Alfred Cowles, who created the Cowles Commission in 1931, in part to understand his own failure as an investor ***(Wilson & Jones 1987)***. An important task of the Cowles Commission was to calculate retrospectively for the years 1871-1925 a stock market index consistent with the Standard & Poor's stock market index consistent with the Standard & Poor 500, published since 1926 (Cowles, 1938). This initiative was also followed and replicated by the CRSP in Chicago. Indeed, it began to record the prices and market transactions of US listed companies on a daily basis in the United States. For empirical studies to proliferate, it had to wait for the CRSP database and the development of computers. Thereafter, the theory continued into the 1950s without much empirical verification ***(H. Markowitz 2005, Bernstein 1992)***.

2.3 Methodological choices in the construction of stock indices

The construction of statistical indices encounters a series of methodological problems. These are understood and gradually resolved over the course of history. There were extensive debates and issues in the period prior to World War I, and they were concluded with a book by ***Irving Fisher ("The Making of Index Numbers - A Study of Their Varieties, Tests, and Reliability" 1922)***. However, ***(Armatte 2004)*** argued that these debates, primarily regarding consumer price indexes, could negatively impact the construction of stock market indexes. In particular, it was noted that the weighting issues of stock indexes were very relevant. In fact, improper weighting methods would lead to substantial errors due to the huge size differences between companies and significant price fluctuations between stocks. Reconstructing the precise weighting methods is a very difficult task. In fact, weighting methods were often not described but implicit solutions were adopted for problems of which full awareness was not had. Below is a taxonomy of the decisions that are made in the construction of an index:

• The choice of market that the index aims to synthesize. It can represent the entire financial market of a national economy or that of a specific institution (US market or New York Stock Exchange). ***(Hautcoeur,2006).***

• The choice of publication frequency.

• The choice of the type of prices adopted for the construction of the index: the average prices of a certain period (day, week, month) or a particular price (for example, the closing price of the last Friday of the month).

• The choice of financial instrument: stocks, bonds, specific securities such as preferred stocks, convertible bonds, multiple voting stocks. Legal changes could influence indexes through these choices.

• The choice of the number of securities included in the index and the selection method (including the modification of the list). A critically important aspect is the diversity in terms of liquidity of securities. In fact, this aspect must be taken into account in the construction of the index in order to avoid substantial errors. However, this has not always been measured or published.

• The choice of subgroups to distribute the various securities (industries, legal considerations such as nationality, the market of quotation, the type of security, etc.).

• The choice of weighting method (constant or variable) and weighting variables (market capitalization, liquidity, floating, etc.).

• The method for maintaining and updating the index over time.

Overall, the construction of stock market indices is a complex process that requires careful consideration of a wide range of factors to ensure the accuracy and relevance of the index. The goal is to create an index that accurately represents the performance of the market or markets it aims to track, while also being easy to understand and use for investors and analysts. It's worth noting that different indices have different objectives and use different methodologies. For example, an index that aims to represent the overall performance of the stock market will likely include a broad range of securities, while an index that aims to track the performance of a specific sector or industry will likely only include securities from that sector or industry. Similarly, different weighting methods may be used to give more or less importance to certain securities, depending on the index's objective and the data available. It's also worth noting that the construction of indices can be influenced by legal and regulatory changes. For example, a change in the laws governing the listing and trading of securities could affect the securities included in an index and the weighting methods used. The choice of the weighting method is one of the most important decisions in the construction of an index. Constant weighting methods, where each security has the same weight, are easy to understand and calculate but can be less accurate in reflecting the performance of the market. On the other hand, variable weighting methods, where the weight of each security is determined by a variable such as market capitalization or liquidity, can be more accurate but can be more complex to calculate and understand. Finally, it's important to note that the construction of stock market indices is an ongoing process, as new data and research become available, and as market conditions change. Therefore, it's important to regularly update and maintain the index to ensure that it continues to accurately represent the performance of the market.

2.4 Definition of market index

The market index is a collection of stocks that is meant to represent the performance of a segment of the overall financial market. Its value is derived from the market value of the individual equities that make up the investment. There are a few different methods that may be used to compute these indices.[1]

"The stock market indices differ primarily by the universe from which they are constructed and, secondarily, by their method of construction. Broadening the underlying portfolio makes it possible to approximate the market portfolio of the single-period CAPM, which in principle contains all the assets of the economy, weighted by their relative values. However, broadening the underlying portfolio can generate pitfalls. There is an econometric trade-off as the index is broadened from the largest capitalization stocks to include ever smaller ones. The more inclusive is the index, the closer it is to the ideal of an index that includes all wealth. The cost of completing the index, however, is the lowered continuity and reliability of the prices of smaller stocks" ***(E.Barone, Tracking Euro Stoxx 50, 2022)***

The value of some indexes is contingent on the kind of weights used, i.e., the proportion of each market portfolio stock. The following is a taxonomy of the primary building methods:

- market-cap weighting
- float-weighting
- revenue- weighting
- basic weighting

It is crucial to keep in mind that the value of each market index is calculated in a method that is exclusive to that index alone, and that this fact is what determines the index's worth. It is possible to calculate it using a number of approaches, such as taking the expenditures of its individual components and averaging or weighting them. This suggests that higher price changes in individual companies will have a stronger influence on the value of the index because of the increased correlation between the two variables. Instead, if market capitalization weighting were used, companies that have a large market capitalization would have a greater effect on the value of the index. This demonstrates that the value of the index as well as the oscillations it

[1] https://www.investopedia.com/terms/m/marketindex.asp

experiences are very sensitive to the techniques that are used to construct it. Investors pay close attention to the movement of important index prices in order to assess bigger market trends. These indices, in addition to serving as an indication of the market, also act as a representation of the financial market or a portion of it. When it comes to monitoring the state of the stock market in the United States, the three indexes that see the most action are the Dow Jones Industrial Average (DJIA), the Standard & Poor's 500 Index, and the Nasdaq Composite Index.[2] The stock market and the bond market use indications that are analogous to one another. The Bloomberg U.S. Aggregate Bond Index is a proxy for the U.S. bonds market, and it is provided by Bloomberg, which is one of the main providers in the world. It is important to note that direct investing is not permitted in some indices. As a consequence of this, these indices act as market benchmarks and provide the basis for indexed funds and passive strategies. The value of the portfolio is reflected by the market index, which fluctuates based on the characteristics of the various reference markets. Each indication is crafted by using a tried-and-true approach that was developed by the provider and is kept up to date by them. The vast majority of investors rely on indices to monitor the performance of markets and inform their investment decisions. These elements have a significant connection to the asset management sector, which is why there is such a connection. In point of fact, the business that deals in funds uses them as benchmarks for measuring performance. In addition, these indicators are used to produce indexed funds that are designed to replicate a reference index. This provides investors with the opportunity to get exposure to the reference index. In recent years, there has been an increase in popularity of investing strategies that include a low level of active management, such as index funds. Because of this, a significant amount of debate takes place over the objectives of these managers in the production of these measures. Due to the fact that these managers' individual performance is based on these measurements, they have a strong interest in the success of these metrics.

2.5 The main stock market indices: concepts and fundamentals

The growth and development of international finance with operations and investments across borders have increasingly led to the use of foreign market indices as a tool to monitor market trends around the world.

[2] https://www.investopedia.com/terms/m/marketindex.asp

2.5.1 The Dow Jones Industrial Average (DJIA)

The Dow Jones Industrial Average (DJIA) is an index comprised of the 30 largest blue-chip corporations in the United States. It was born in 1896. This index only examined 20 titles before to 1928. Initially, the DJIA was computed using a straightforward arithmetic average of prices. That is, the number of components is equal to the price of the titles divided by 30. The percentage change in the index will then equal the average percentage change in the prices of the index's 30 constituent titles. The aforementioned change in the DJIA index represents the return, minus dividends, of a portfolio constructed by purchasing one unit of each of the 30 stocks that comprise the index.[3]

Therefore, the portfolio's worth equals the total of the prices of the 30 titles. The Dow represents a portfolio consisting of

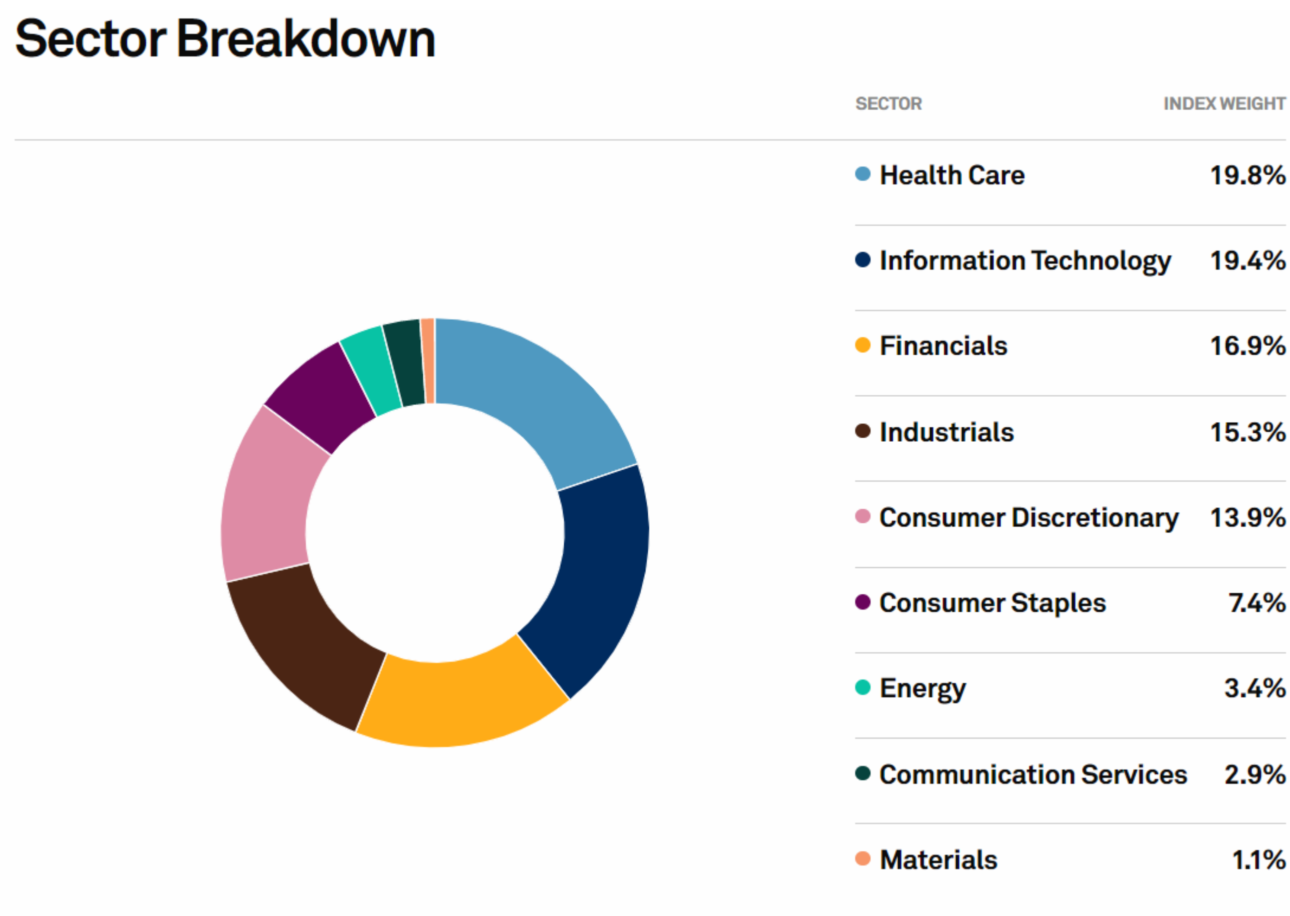

Figure 1. DJIA sector breakdown

[3] https://en.wikipedia.org/wiki/Dow_Jones_Industrial_Average

one share of each component of the index, with an investment in each stock proportionate to the stock's price. The DJIA is considered a price-weighted average index for this reason. The Dow Jones Transportation Average is the oldest index on the U.S. market; the DJIA is the second-oldest (DJTA). The DJIA was designed only for the aim of monitoring the economy and representing the U.S. market. When the index was founded in 1896, just twelve firms were included. These industries included railways, cotton, natural gas, tobacco, sugar, and oil. Beginning in the 20th century, the performance of industrial enterprises was primarily tied to economic growth rates. This reinforced the DJIA's dependency on the American economy. Indeed, this dependency enhances the index's ability to serve as a proxy for the overall economy. Even today, many investors associate a healthy and expanding economy with a high performance of the Dow Jones, and vice versa. Given the correlation between this index and the economy, as the economy evolves, the index's components change as certain firms become less important to the shifting economic trends. As a result, some firms are eliminated from the index and others are added based on the business realities that most accurately represent these economic developments. At this time, it is appropriate to indicate that a firm may be withdrawn from the DJIA if its market capitalisation decreases significantly. As a price-weighted average index, the Dow Jones gives higher-priced companies a larger weight. Consequently, percentage changes in the value of stocks with higher prices have a larger effect on the computation of the index's final value. Charles Dow determined the index's average by summing the prices of the 12 firms that included it and dividing by the number of companies that comprised the index. The outcome was a straightforward mean. The components of the index have changed over time due to acquisitions, stock splits, etc., thus a computation using a simple average was no longer relevant. The Dow Divisor is a constructed divisor used to divide a simple average. It is a constant used to compute the impact of one percentage point on each of the index's 30 constituent components. The Dow Divisor value on June 6, 2022, was 0.151727525954. Unlike the S&P 500, the Dow is not computed using an arithmetic average and does not rely on market capitalisation. Instead, it represents the sum of the prices of its constituent stocks divided by the Dow Divisor. In fact, a one-percentage-point change in the price of each index component has the same percentage-based effect on the index.

$$DJIA_t = \pi r^2 = \sum_{i=1}^{n} {P_{it}}/{dd_t}$$

n this expression, P_{it} is the price in $ of the i-th stock for i = 1, 2, ..., 30 at time t, and dd_t is the Dow Divisor. Below is the mathematical formulation of the latter:

$$dd_{t+1} = dd_t \, {\sum c_{t+1}^{a}}/{\sum c_t}$$

The stock price at time t is denoted by c_t, and the stock price at time (t + 1) is denoted by c_{t+1}^{a}, where dd_{t+1} is the divisor determined at time (t + 1) after the capital operations of the businesses in the basket, ddt is the value of the previous divisor prior to modification, and c_t is the closing price of the stock at time t. The Dow Divisor allows the DJIA to continue tracking the same stocks throughout time. This process is necessary due to the many stock splits, spin-offs, and other changes that have happened among Dow components since the index's inception. The divisor enables the analysis of various market trends, including dividend payments and the aforementioned occurrences. The DJIA's real value is not affected by these events since the Dow divisor is controlled. Key market developments and events throughout the years have caused significant changes in the value of the Dow divisor. In September of this year (2019), one dollar is worth around $0.0147, which is a huge difference from its value of $16.67 in 1928. Changing the divisor in this way guarantees that the index may be compared across different time periods. The Wall Street Journal is responsible for ensuring the correctness of the Dow divisor so that the DJIA may be compared to its past performance. The sum of the component values may be affected by factors such as stock splits or changes in the composition of the companies making up the index. This causes a change in the Dow divisor such that the last two quotations before and after the event are identical. This eliminates a potential performance drop in the index. The value of the Dow divisor has decreased as a result of most corporate actions, including as stock splits and spin-offs. The index has come a long way from its early days, when the entire number of DJIA members served as the divisor. However, due to market-altering occurrences, the Dow divisor has been tweaked by hand to guarantee a fair evaluation of the DJIA. After a number of adjustments, the current divisor is now less than 1. Therefore, the divisor is now a multiplicator as well. On the other hand, this makes any possible index rebuilding more difficult and less clear. Index providers regularly make muddled adjustments to the divisor, among other things.[4]

2.5.2 Limitations and negative aspects

Due to the fact that it consists of just 30 significant U.S. firms, the Dow has critics who argue it does not fairly represent the economy as a whole. It has been suggested that the sample size is too small and that varied company kinds are being overlooked. Some have suggested that the S&P 500, which includes a far larger number of companies, is a more realistic representation of the US economy. In actuality, the latter comprises 500 distinct enterprises. In addition, the stock price alone does not effectively represent the characteristics of a firm. In reality, market capitalization more accurately reflects these characteristics. This limitation results in distortions. A company with a low stock price and a high market capitalization would be given less weight than a company with a high stock price and a low market capitalization. Consequently, the perceived size of a corporation would be overstated.

2.5.3 S&P 500

S&P 500 is an index formed in 1957 to track the performance of the 500 biggest publicly listed companies on the New York Stock Exchange (NYSE) and the NASDAQ Composite. Standard & Poor's is a firm that gathers and disseminates economic data, issues credit ratings, and publishes stock market indices. Initial designers of the index envisioned it as a Laspeyres index weighted by market capitalisation.[5] To calculate the actual value of the S&P 500 index at a certain time t, the following formula is used.

$$P_t^f = \sum_{i=1}^{n} P_{it} q_{it} f_{it} \Big/ d_t$$

Here, p_{it} stands for the price of i-th stock, q_{it} for the total number of shares of shares that are now outstanding, and f_{it} for the percentage of stocks that are traded publicly. The term "fraction of a business's shares that are traded on an exchange" refers to the ratio of the total number of shares held by a firm to the number of shares that are actively traded on an exchange. The index

[4] https://www.investopedia.com/terms/d/dow-30.asp
[5] https://www.investopedia.com/terms/s/sp500.asp

divisor is denoted by the sign d_t in mathematical notation. After conducting in-depth research on each constituent, the S&P 500 Index Committee determines the Investable Weight Percentage Adaptation Factor (Investable Weight Percentage Adaptation Factor). This last goal is to get rid of firm leaders and administrators, whose shares are seldom sold to the public. When calculating the index, the denominator, also known as the "dividend," plays a vital part in the daily adjustments and updates that are performed. When the stock market closes for the day at a certain value and stays at that level throughout the night, it should open for trading the following day at the same value. If the make-up of the index components were to shift after the market had closed but the same stock quotes continued to be used, the divisor would need to be updated in order to maintain the status quo of the index when the market reopened the following day. The formulas that may be used to change the DJIA's divisor are somewhat comparable to this one. In addition, the market value after adjusting for float is denoted by the symbol MV_t. This value is calculated as of the end of business on the t-th trading day prior to the index rebalancing. As a direct result of this, the revised market value on the first trading day after the components of the index have been rebalanced is denoted by the symbol MV_{t+1}. The following is the definition of the new fraction d_{t+1}:

$$d_{t+1} = d_t \frac{MV_{t+1}}{MV_t}$$

where dt is the old divisor of the index. Therefore, the value of the new divisor is obtained by equating the value of the index before the change of composition ($\frac{MV_t}{d_t}$) to the value of the index after the change of composition ($\frac{MV_{t+1}}{d_{t+1}}$). Subsequently, being MV_t, d_t and MV_{t+1} known variables, it is sufficient to solve the equation for d_{t+1}, that is the unknown variable.

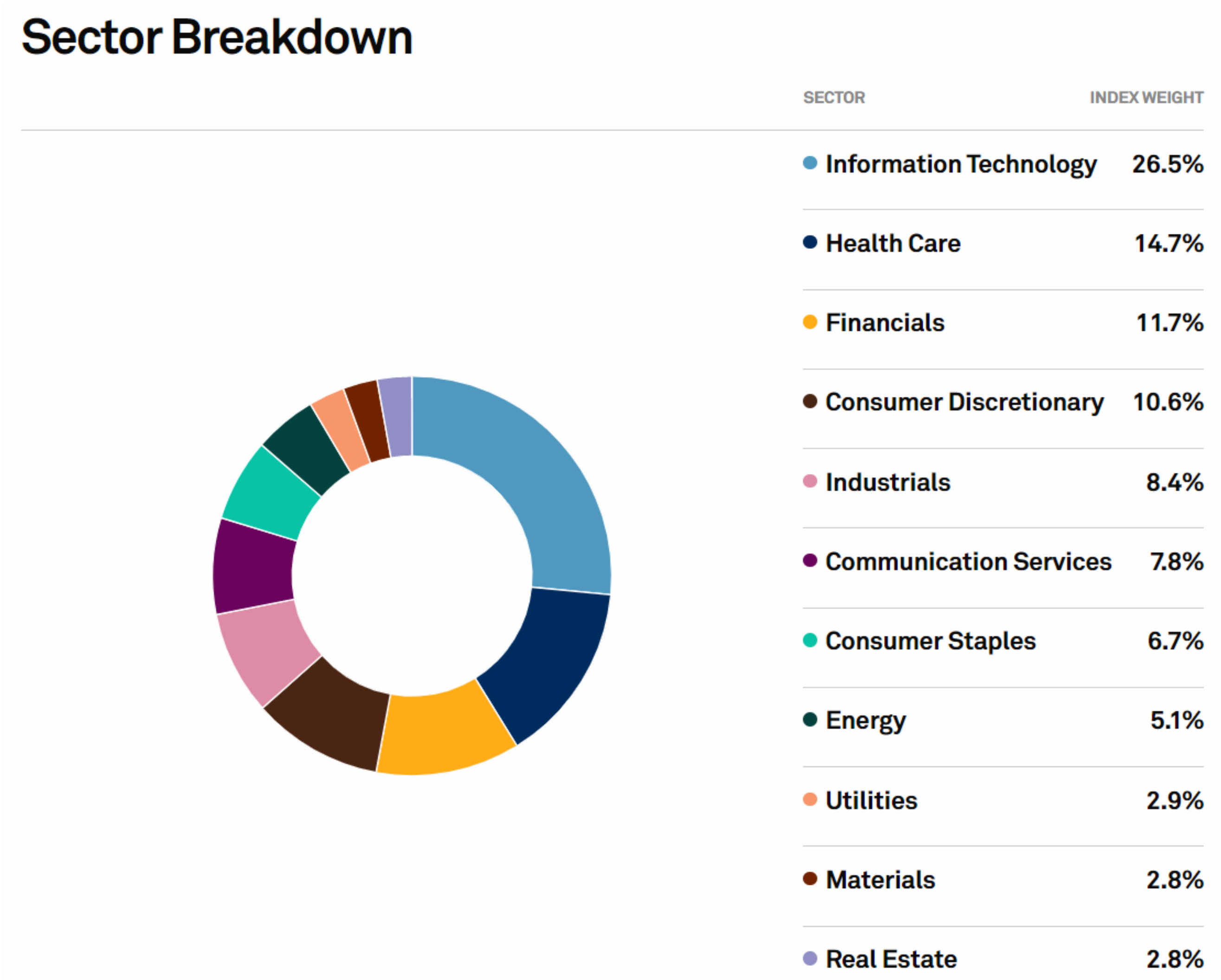

Figure 2. SP500 sector breakdown

The set of stocks that make up the S&P 500 is designed to represent the majority of the composition of the US economy. As a result, the value of the index and the various stocks within it are closely watched by market operators, as their performance represents an indicator of the health of the US economy. As previously specified, the exact combination and weightings of the various components of the S&P 500 are modified based on changes in the economy. Therefore, the indicator undergoes changes in its relative composition in function of the performance and evolution of the economy itself. The components of the S&P 500 are selected by a committee and are considered representative of the sectors that make up the US economy. To be part of the indicator, a company must meet certain dimensional requirements. These are based on liquidity criteria, including a market capitalization of at least 14.6 billion dollars. The value of the S&P

500 index is calculated as the sum of the market capitalization of all 500 stocks divided by a certain factor, usually referred to as the index divisor. The market capitalization of the entire index can be consulted on the Standard & Poor's website. The exact number of the divisor is considered proprietary information of the company, although its value is approximately nine billion. Again, there are some information opacities regarding the construction of indices. This causes a sort of black box in the construction of these making their reproduction almost impossible.

The S&P is established as a reference indicator for the US stock market. Indeed, it has been given the term "bellwether, a leading indicator of an economic trend" meaning an indicator that suggests the direction of the economy. The S&P is also a fundamental vehicle for investors who want exposure to the broad market of US indexed funds. In fact, the appreciation of the S&P 500 tends to follow the growth of the US economy. The fluctuations in the prices of the indicator also tend to accurately reflect periods of turbulence in the US economy. As a result, the long-term chart of the history of the prices of the S&P 500 serves as a reading of investor sentiment on the US economy. The S&P 500 began trading in 1957. It is considered a better index than the DJIA for two reasons, it considers a greater number of companies (500) and has a weighting based on market capitalization. Therefore, it takes into account the differences in the size of the companies in the index and thus represents a more accurate picture of the economy. One thing that appears appropriate to specify is that in recent years market indices are weighted based on the free float, that is, the floating freely available for trading. This precaution takes place because some stocks held by governments or families are not actually traded on markets. Therefore, considering market capitalization as a weighting methodology would be a distortion. However, this issue is more felt for indices related to the European and Japanese markets where a high fraction of stocks is held in non-traded portfolios.

2.5.4 VIX

Volatility is one of the elements that investors take into consideration when choosing their market investments. Typically, situations in which the market is less volatile are also more predictable. A contained volatility is what fund managers and institutional investors consider optimal for their investments. When markets are subject to panic selling, typically caused by some news, volatility increases. That's why the VIX index is also called "fear index". However, this is not a technically correct definition. According to the ***"White Paper Cboe Volatility Index" (CBOE, 2019)*** The VIX index has the very important task of giving everyone who operates on the market a common reference for volatility. A single number, equal for everyone, that synthetically expresses expectations for fluctuations in the next four weeks. That's why, even if it's not a stock index, it is still considered one of the most important in the world.

The VIX index, also known as the CBOE Volatility Index, is a measure of the implied volatility of the S&P 500 index. It is calculated using the prices of options on the S&P 500 index. The VIX index is often used as a measure of market risk and investor sentiment. The VIX index is calculated using a combination of call and put options on the S&P 500 index. The options used in the calculation are those that are closest to the at-the-money (ATM) strike price and that have at least 8 days until expiration. The VIX index is calculated using the following formula:

$$\sigma^2 = \frac{2}{T}\sum_i \frac{\Delta K_i}{K_i^2} e^{RT}\, Q(K_i) - \frac{1}{T}[\frac{F}{K_0} - 1]$$

It is necessary to know what each variable in this equation means:

- T is the remaining time, expressed in minutes, before the option expires.
- $\sum_i()$ indicates that the sum of everything that comes after should be taken for each valid option.
- ΔK_i is half the distance between two strike prices. If one option has a strike price of 440 and the next one has a strike price of 441, that means that ΔK_i is equal to 0.50.
- K_0 is the first out-of-the-money strike price, that is the closest to the current value of the index.
- K_i is the strike price that we are currently considering in the summing process, since the calculation is repeated for all valid strike prices.
- R is the risk-free interest rate, that is how much one would gain by holding cash in the option until its expiration. The number used is the return of 3-month US bonds.
- Q(K_i) is half the spread between the bid and ask for the option that is being considered.

Once this first part of the calculation is done, the square root of the result is then multiplied by 100.

As contrast to actual forward-looking volatility, the market's expectation of future volatility in the S&P 500 index options is what the VIX gauges. This expectation is measured over the next 30 trading days. According to the VIX index, the implied 30-day volatility of the S&P 500 index is provided as a percentage. This volatility is measured over the previous 30 days. A lower number for the VIX index suggests a higher amount of expected volatility, whereas a higher reading for the index indicates a higher level of expected volatility. It is important to keep in mind that the VIX gauges expected volatility rather than actual volatility. This suggests that while the VIX may show a high level of expected volatility, actual volatility may be lower or larger than anticipated even if the VIX may signal a high amount of anticipated volatility. In addition, the VIX is a measurement of the volatility of the S&P 500 index rather than the entire risk of the market. As a result, it ought not to be used by itself as a measurement of market risk but rather in conjunction with a number of other indicators. As a measure of market sentiment and an indicator of potential danger, the VIX index is often used by market analysts, investors, and traders. Real-time trading is available for futures and options contracts that are based on this index on the CBOE exchange. The VIX index, which measures the expected 30-day volatility of the S&P 500 index, was used as a barometer of market worry during the COVID-19 pandemic. This indicator was created to measure the volatility of the stock market. The VIX index reached all-time highs, which is an indication of enormous market concern and volatility. This was caused by the rapid spread of the virus, which in turn caused substantial economic uncertainty. The VIX index reached its final value of 16.04 on February 20, 2020. This number represented a level of predicted volatility that was considered to be on the low end of the spectrum. The VIX index, on the other hand, began a precipitous upward trend as the virus continued to spread and widespread economic fear resulted. On March 16, 2020, the VIX index hit a record high of 82.69, signalling a significant degree of anticipated market volatility and panic. The VIX index continued to reflect increased levels of market volatility and worry for the whole of the course of the outbreak. The fluctuation of the VIX index during the course of the disease's progression exemplifies the dynamic nature of the economy. During the course of the outbreak, one method for determining the sentiment of the market is to consult the VIX index. The value of options on the S&P 500 index falls at times of excessive market volatility because investors seek refuge in assets that are seen to be less risky. These include gold, bonds, and the US currency. The volatility index, often known as the VIX, may be used as a tool to evaluate the efficiency of various government stimulus programmes and other initiatives. Initiatives such as these could help to calm markets and reduce the VIX's volatility during times of heightened market fear.

2.6 Descriptive charts and graphs

In this section I present some descriptive statistics and graphs related to the variables I used to build my model. Starting with some graphs we can observe some interesting features.

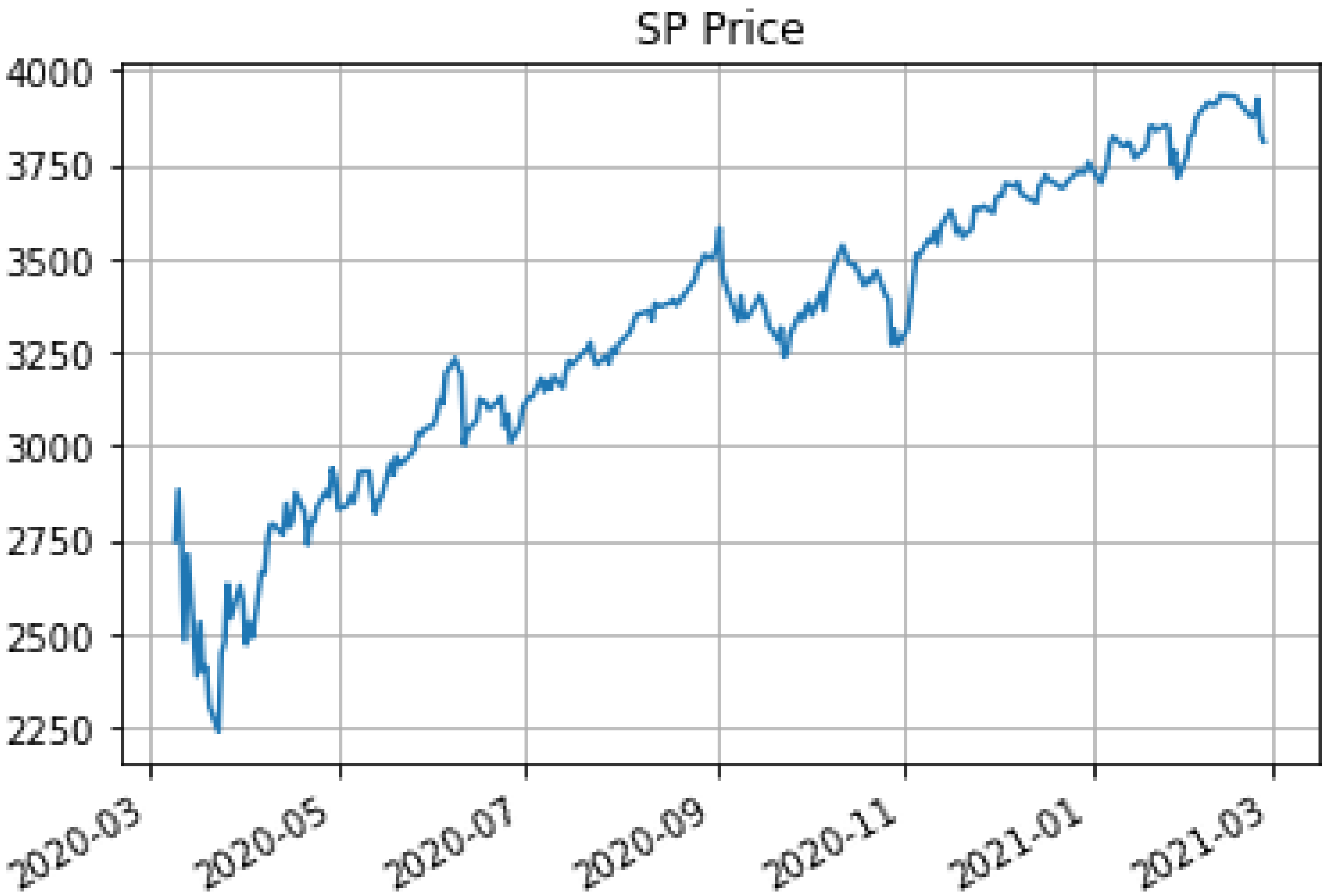

Figure 3. SP500 price path 03/2020 – 03/2021

The S&P500 index experienced significant fluctuations during the period from March 2020 to February 2021, which was analysed in this study. The index reached its decade-low level in a matter of days and experienced a sharp decline in value. However, it was able to recover and achieve its full recovery throughout the year. This period was characterized by extreme volatility in the stock market and marked a significant event in the financial industry. The S&P500 index, being one of the most widely followed and significant indicators of the stock market performance, provides a clear picture of the impact of the pandemic on the financial market. This recovery, despite the initial decline, highlights the resilience and stability of the financial market and the ability of the stock market to overcome challenges and recover. It is also an indication of the confidence of investors in the long-term growth potential of the economy. The study aimed to understand the relationship between the S&P500 index and various independent variables and to gain insights into the factors that drive the performance of the index. The recovery of the S&P500 index is an important case study in the field of finance and provides valuable insights into the behaviour of the stock market during a crisis.

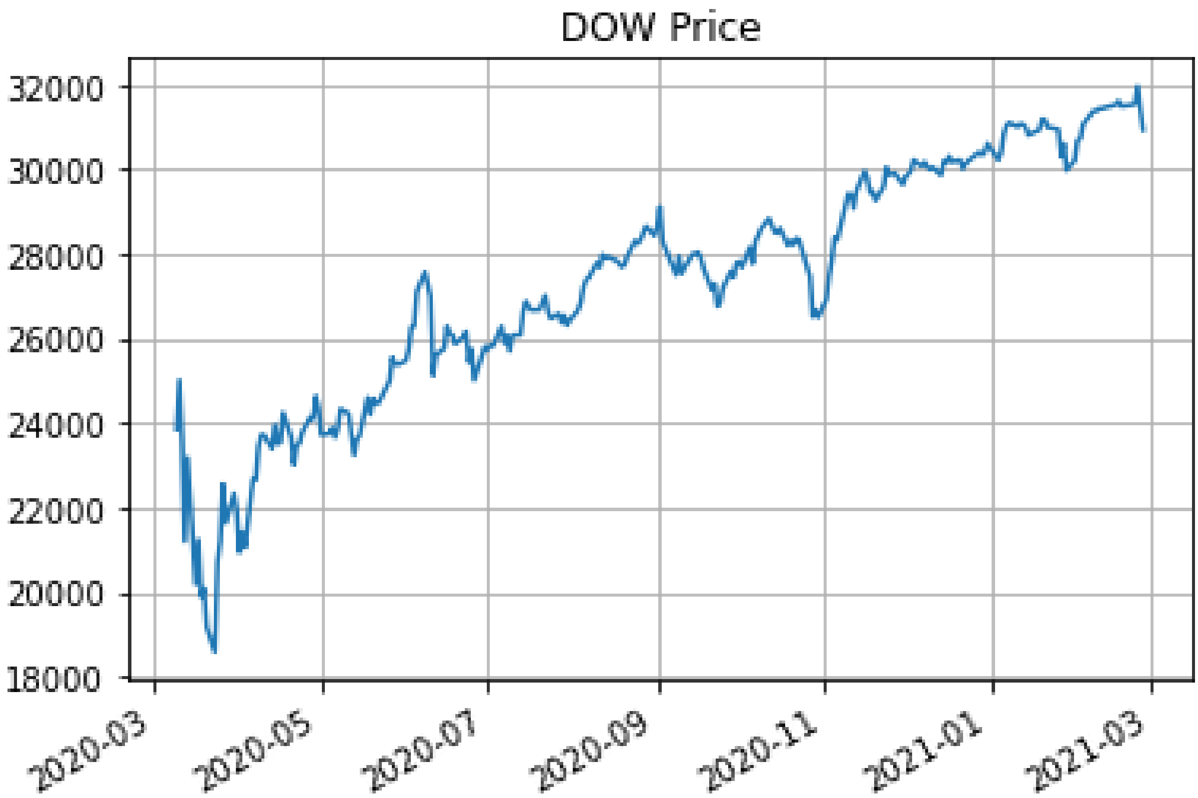

Figure 4. DJIA price path 03/2020 – 03/2021

The Dow Jones Industrial Average, which is another widely followed stock market index, followed a similar path as the S&P500 during the analysis period. This is not unexpected, given the historical relationship between these two indices. In fact, their correlation has been historically high, as we will explore in more detail later in the analysis. The movement of the Dow Jones is often seen as a reflection of the overall health of the U.S. stock market, making it an important indicator for investors and economists alike. During the time frame of the analysis, which ran from March 2020 to February 2021, the Dow Jones experienced both significant losses and impressive gains, ultimately mirroring the trajectory of the S&P500.

Figure 5. VIX price path 03/2020 – 03/2021

The Volatility Index, commonly known as the VIX, provides a clear representation of the level of fear present in the US market. During the first year of the COVID-19 pandemic, the VIX price path demonstrates this correlation clearly. In the initial months of the pandemic, the VIX reached a decade high, reflecting the uncertainty and fear in the market caused by the sudden outbreak of the virus. However, the VIX quickly retraced back, showing a decrease in market fear as the situation appeared to stabilize. Despite this initial stabilization, there were still moments of heightened fear in the market as the pandemic continued to spread and lead to more COVID-19 cases and related deaths. These periods are reflected in the VIX graph as spikes, showing a corresponding increase in volatility levels. This correlation between increased COVID-19 cases and increased market fear highlights the impact that the pandemic has had on the US market. It is important to note that while the VIX provides a clear representation of market fear, it does not necessarily indicate causality. Nevertheless, the relationship between the VIX and COVID-19 cases is undeniable and provides valuable insight into how the pandemic has affected the US market. This information can be used to make more informed investment decisions and anticipate future market movements.

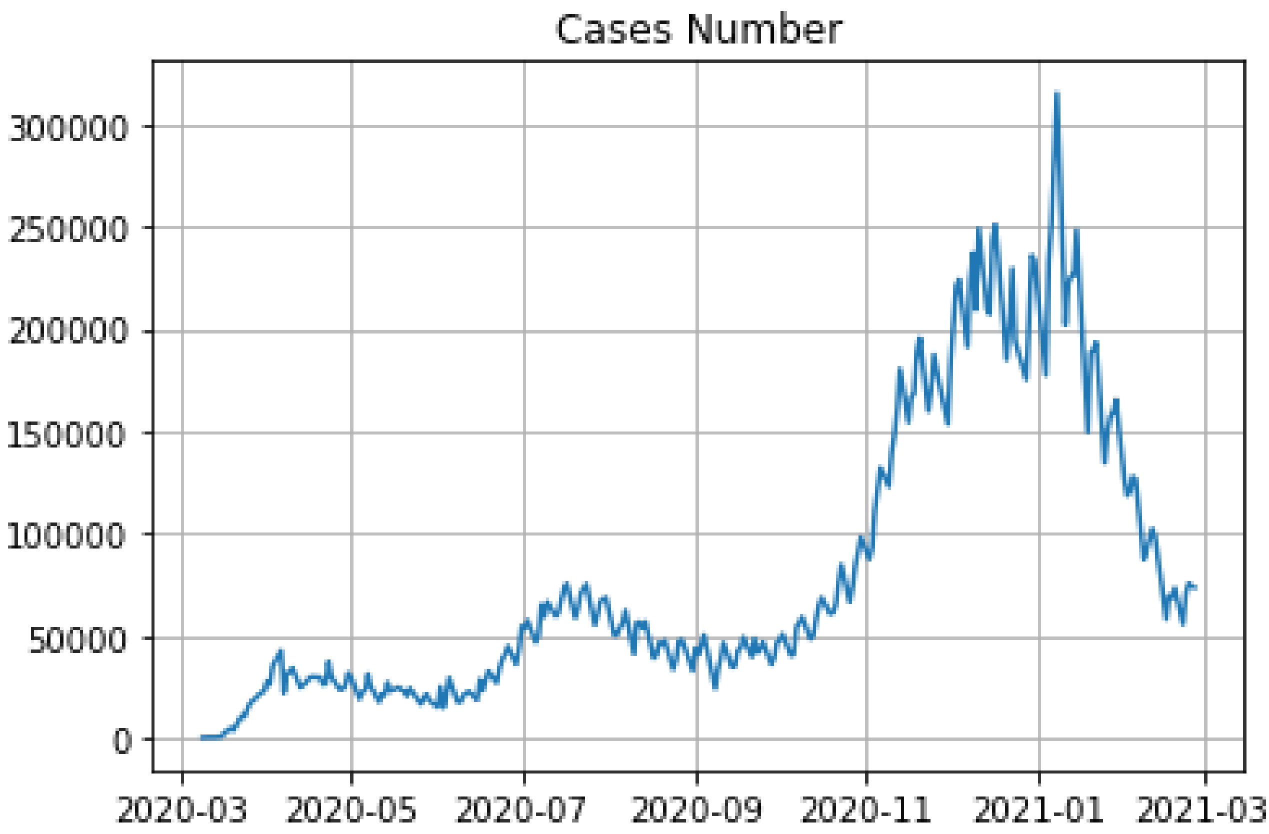

Figure 6. Graph of the path of cases number

The daily number of COVID-19 cases is displayed in this graph, and it is evident that a sharp increase in cases leads to a spike in the VIX price. This observation may seem contrary to what was seen in the graphs for the S&P500 and Dow Jones indices, however, the markets were able to recover quickly and strongly from the initial effects of the pandemic. It was actually the sustained worsening of conditions, rather than a single shock, that had a lasting impact on the markets, as we will delve into further in the following chapter. This graph serves as a visual representation of the relationship between the spread of the virus and market volatility. It is well established that increased uncertainty, brought on by the escalation of the pandemic, leads to higher levels of volatility in the financial markets. By analysing this graph, it becomes clear that the correlation between the number of cases and VIX price is significant at certain points in time.

Correlation Matrix

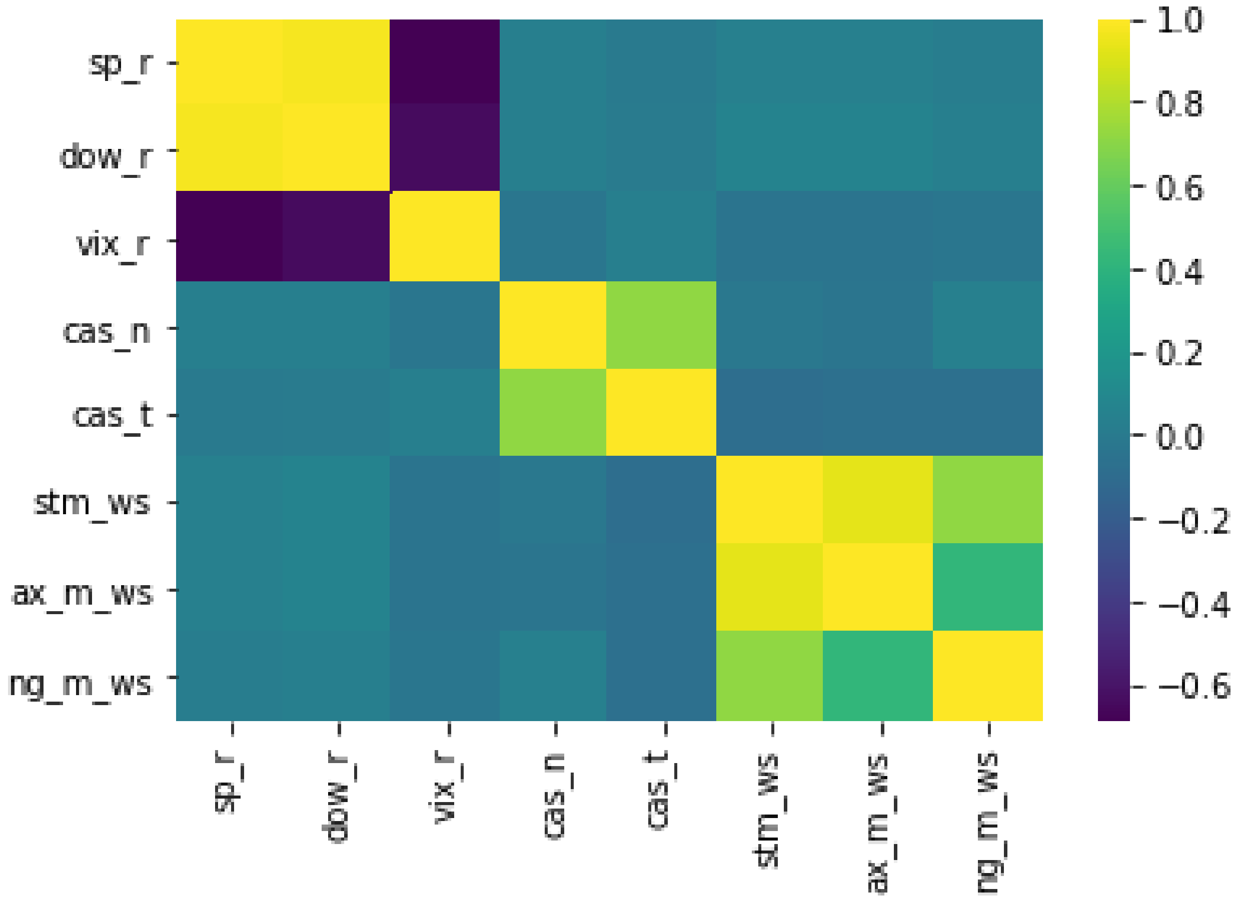

Figure 7. Correlation matrix of the variables

The figure presented here displays a correlation matrix that summarizes the relationships between the dependent and independent variables used in the two models under analysis. As can be observed, there is a negative correlation between the returns of the stock market indexes, such as S&P500 and DOW JONES, and the independent variables related to the tone of anxiety and negativity in COVID-19 related news in the Wall Street Journal. This correlation suggests a potential relationship between these variables, however, it is important to note that correlation does not necessarily imply causation. The VIX, a measure of market volatility, is also negatively correlated with the independent variables in the analysis. At first glance, this might seem contrary to what was observed in the previous graph, where VIX was seen to be highly volatile. However, this apparent discrepancy can be explained by the fact that VIX and the independent variables are generally uncorrelated, but can become highly correlated in moments of increased volatility. This is evident in the graph, where times of spikes in COVID-19 cases coincide with spikes in market volatility, resulting in a strong correlation between the two variables. Overall, the correlation matrix provides valuable insights into the relationships between the variables used in the analysis and highlights the complex interplay between the spread of COVID-19, public sentiment, and financial markets. This information is crucial for understanding the impact of the pandemic on financial markets and for making informed investment decisions. It is important to remember that the correlation matrix only provides a snapshot of the relationships between the variables at a certain point in time and that these relationships can change over time as the situation evolves.

2.7 Descriptive statistics

	Variables	Count	Mean	Std	Min	25%	50%	75%	Max
Y	sp_r	248	0.10%	2.13%	-12.77%	-0.53%	0.24%	1.04%	8.97%
	dow_r	248	0.07%	2.26%	-13.84%	-0.57%	0.15%	0.91%	10.76%
	vix_r	248	-0.16%	8.77%	-26.62%	-4.91%	-1.22%	2.86%	48.02%
X1	cas_n	248	80183	88.01%	218	29112	52068	118934	315115
	cas_t	248	8739725	98.40%	650	1857495	5969113	12917261	28370721
X2	stm_ws	248	19	17.67%	11	17	19	21	36
	ax_m_ws	248	13	19.88%	8	11	13	14	27
	ng_m_ws	248	6	21.99%	2	5	6	7	10

Figure 8. Descriptive statistics table

Here we can see a table of content in which are summarized some basic statistics about the variables used in my two regression models, divided in dependent variables (Y) and independent variables (X1, X2).

- The Y variables in this analysis represent the daily returns of three of the most widely followed financial indices: the S&P500, the DOW JONES, and the VIX. These indices serve as barometers of the overall health and performance of the stock market, providing valuable insight into market trends and conditions. Starting with the S&P500 and DOW JONES, these two indices exhibited a relatively stable path during the period of analysis, with low standard deviations of around 2% and means in the same range. This stability is a positive sign, indicating that the stock market was operating in a consistent and predictable manner during this time. The low standard deviations are particularly noteworthy, as they suggest that the stock market was not subject to significant fluctuations, which can be a source of uncertainty and unpredictability for investors.

On the other hand, the VIX index was much more volatile, with a standard deviation of around 9%. This high level of volatility is due in part to the spikes in fear that occurred during the year, such as during spikes in COVID-19 cases. When fear is high, investors tend to sell off assets, leading to rapid and significant changes in stock prices. This increased volatility is reflected in the VIX index, which is specifically designed to measure market volatility and provide insight into the level of risk associated with stock investments. When comparing the maximum drawdowns of these indices, it is evident that VIX suffered a much greater loss in its worst trading day (-26.62%) compared to the losses suffered by the market indexes (12.77% and 13.84%). This difference is significant, as it highlights the impact of the COVID-19 pandemic on the stock market. The initial outbreak of the pandemic caused widespread fear and uncertainty, leading to a rapid sell-off of assets and a decline in stock prices. This decline was much more pronounced in the VIX index, which provides a clearer picture of the level of fear and volatility in the market. It is worth noting that the stock market is an inherently complex and dynamic system, subject to a wide range of factors that can impact its performance. In addition to the COVID-19 pandemic, other factors such as global economic conditions, political events, and technological developments can also play a role in shaping market trends and conditions. Despite these challenges, the stock market remains a crucial component of the global financial system, offering investors the opportunity to grow their wealth over the long term. In conclusion, the Y variables in this analysis provide a comprehensive picture of the daily returns of three of the most widely followed financial indices. By analysing these variables, it is possible to gain a deeper understanding of the stock market and the factors that impact its performance. Whether viewed in the context of the COVID-19 pandemic or in the broader context of global financial conditions, these variables are a valuable resource for investors and analysts alike.

- The X1 variables are a representation of the daily spread of COVID-19 cases, both in terms of daily cases and the total number of cases over time. These variables provide crucial insight into the unpredictability of the pandemic and its impact on society. A visual analysis of the X1 variables highlights the large volatility of COVID-19 cases over time, with standard deviations of around 88% and 98%. This high level of volatility demonstrates the inconsistent nature of the spread of the virus. At various points throughout the year, the number of cases would skyrocket to extremely high levels, putting a strain on healthcare systems and causing widespread panic. These high-intensity periods would be followed by periods of low intensity where the number of cases would drop to almost insignificant levels. This unpredictability of the spread of COVID-19 has made it extremely challenging for public health officials and governments to control and contain the spread of the virus. The X1 variables also highlight the importance of monitoring the spread of the virus and implementing effective measures to control its spread. The large standard deviation of the X1 variables demonstrates that even small changes in the number of cases can result in a significant impact on the overall trend of the spread of the virus. As such, it is crucial to remain vigilant and proactive in addressing the spread of the virus, even during periods of low intensity, to prevent a resurgence in cases. In conclusion, the X1 variables provide a clear representation of the unpredictability of the COVID-19 pandemic and its impact on society. The large standard deviation of these variables emphasizes the importance of consistent monitoring and proactive measures to control the spread of the virus. It is clear that the spread of COVID-19 has been a challenge for the world and highlights the need for continued efforts to control its spread and reduce its impact on society.

- The X2 variables in this analysis are an important set of derivatives that are derived from Wall Street Journal articles during a specific period of time. These variables are designed to capture the tone of anxiety and negativity in COVID-19 related news published in the Wall Street Journal. The first derivative, Ax_m_ws, represents the daily time series of the average tone of anxiety in the news articles. It is calculated by counting the number of anxiety-related words in COVID-19 related news published in the Wall Street Journal on a daily basis. This information can provide valuable insight into the level of fear and uncertainty present in the market, as well as the sentiment of investors and analysts regarding the pandemic and its potential impact on the stock market. The second derivative, Ng_m_ws, represents the daily time series of the average negative tone in COVID-19 related news. This derivative is calculated by counting the number of negative words present in the news articles on a daily basis. Negative news can have a significant impact on investor sentiment and market performance, as it can cause fear, uncertainty, and a decrease in market confidence. The average negative tone in COVID-19 related news can provide valuable information about the level of negativity present in the market and help investors and analysts to make more informed decisions. Finally, the third derivative, Stm_ws, is a summary of the first two derivatives, representing the sum of the daily values of Ax_m_ws and Ng_m_ws. This variable provides a comprehensive representation of the anxiety and negativity present in COVID-19 related news published in the Wall Street Journal. By considering both the level of anxiety and negativity in the news, this variable can provide valuable insight into the overall sentiment and tone of the market, as well as the potential impact of the pandemic on the stock market. It is important to note that simple statistics such as mean and standard deviation may not provide much insight into these variables, due to their low numbers and the potential for large fluctuations. A small change in the numbers can result in a significant impact on the mean and standard deviation, which can skew the results of statistical analysis. However, despite these limitations, the X2 variables have been shown to have strong explanatory power in dependent variables, making them a valuable tool for understanding the relationships between news sentiment and market performance. By leveraging the information provided by these X2 variables, investors and analysts can make more informed decisions about the stock market, considering not only the financial data but also the sentiment and tone of news related to COVID-19. This type of analysis can provide valuable insight into the relationships between news sentiment, economic conditions, and stock market performance, helping to identify trends and opportunities in the market. In conclusion, the X2 variables are an important set of derivatives that provide valuable information about the tone of anxiety and negativity in COVID-19 related news published in the Wall Street Journal. By considering these variables in conjunction with stock market data, economic conditions, and other relevant information, investors and analysts can make more informed decisions about the stock market and achieve better returns.

Chapter III

METHODOLOGIES AND RESULTS PRESENTATION

3.1 Methodologies

To address issues of scale and comparability, I transformed the independent and dependent variables in my analysis by taking the natural logarithm of their values. By doing so, I was able to put the variables on a similar scale, which made it easier to compare their relative magnitudes and to assess their relationships with one another. In particular, the dependent variable in my analysis was stock returns, which I transformed by taking the natural logarithm of its values. This allowed me to convert the returns into a continuous and normally distributed variable, which made it easier to analyse their behaviour and to model their relationships with the independent variables. In addition, I also put the logged stock returns into growth rate to match the format of the independent variables. The independent variables were also transformed by taking the natural logarithm of their values. This was done to ensure that the variables were on a similar scale and to avoid issues of non-linearity. By transforming the independent variables in this way, I was able to more accurately model their relationships with the dependent variable and to identify which variables had the greatest explanatory power over stock returns. Overall, the transformation of the variables by taking the natural logarithm was an important step in my analysis, as it allowed me to put the data on a similar scale and to more accurately model their relationships. By doing so, I was able to more accurately identify the most important independent variables and to generate more accurate predictions of stock returns. However, it is important to note that the transformation of the variables is only one aspect of the analysis and should not be relied upon in isolation. Other factors, such as the choice of independent variables and the modelling approach, also play a critical role in the accuracy and reliability of the results. As such, it is important to take a comprehensive and systematic approach to data analysis and to consider a range of factors when making sentiment analysis.[6]

In my analysis, I utilized two different multiple linear regression models to better understand the relationship between the sentiment of news articles and stock market returns. The first model was built with lagged independent variables, while the second one used a Rolling Simple Moving Average (RSMA) on the independent variables. The first multiple linear regression

[6] From now on, every time I name independent variables they have to be intended as log growth rates time series.

model was designed to capture the effects of news sentiment on stock market returns that may not have been immediately apparent after the release of news related to the COVID-19 pandemic. To accomplish this, I lagged the independent variables by a certain number of business days. This allows us to consider the impact of the news on stock market returns over a longer period of time, and to account for any delayed effects that may have occurred. The second multiple linear regression model was designed to capture the persistence of the lagged effect of news sentiment on stock market returns. To do this, I used an RSMA on the independent variables. This technique helps to identify any trends in the relationship between news sentiment and stock market returns over time, and to determine whether the impact of news sentiment on stock market returns is sustained over a longer period of time. By using both lagged independent variables and an RSMA, my analysis provides a comprehensive examination of the relationship between news sentiment and stock market returns. This allows us to better understand the dynamics of this relationship, and to make more informed predictions about how changes in news sentiment may impact stock market returns. My analysis demonstrates the importance of considering the lagged effects of news sentiment on stock market returns, as well as the persistence of these effects over time. The use of multiple linear regression models, both with lagged independent variables and an RSMA, provides a more complete picture of the relationship between news sentiment and stock market returns. Multiple linear regression is a statistical technique used to determine the linear relationship between two or more independent variables and a dependent variable. In the context of sentiment analysis, multiple linear regression can be used to determine how the news coverage of the COVID-19 pandemic and the Wall Street Journal has impacted the returns of the S&P 500 and the Dow Jones Industrial Average. Sentiment analysis refers to the process of computationally determining the subjective opinions, emotions, or evaluations expressed in text. It has become an increasingly popular tool in finance and economics, as it can help us understand how market sentiment impacts financial outcomes. For example, sentiment analysis can be used to gauge the public's perception of a company or industry, and how this perception is reflected in stock returns. In the case of the COVID-19 pandemic and the Wall Street Journal, sentiment analysis can be used to determine how news coverage of these events has impacted stock market returns. The S&P 500 and the Dow Jones Industrial Average are two of the most widely used stock market indices in the world, and changes in their returns are closely watched by investors and analysts alike. Multiple linear regression can be used to model the relationship between the sentiment of news articles and the returns of the S&P 500 and the Dow Jones Industrial Average. The sentiment of news articles can be quantified using a variety of techniques, including word frequency analysis, lexiconbased sentiment analysis, and machine learning algorithms. These sentiment scores can then be used as independent variables in a multiple linear regression model, with the returns of the S&P 500 and the Dow Jones Industrial Average as the dependent variable. The advantage of using multiple linear regression for this type of analysis is that it allows us to account for the potential impact of other variables that might influence stock market returns. For example, it may be that the news coverage of the COVID-19 pandemic and the Wall Street Journal is not the only factor affecting stock market returns. Other variables, such as economic data releases, monetary policy announcements, and geopolitical events, could also be impacting the stock market. Multiple linear regression allows us to control for the influence of these other variables, and to isolate the impact of the sentiment of news articles on stock market returns. This can help us understand the relationship between news sentiment and stock market returns more accurately and can also help us to identify other variables that might be driving stock market returns. Sentiment analysis is a critical tool in finance and economics for understanding the relationship between news sentiment and stock market returns. In this analysis, two multiple linear regression models were utilized to examine the relationship between Wall Street Journal news and COVID-19 case numbers on the S&P 500 and Dow Jones Industrial Average (DJIA) stock market returns. The first model used lagged independent variables, while the second model used a Rolling Simple Moving Average (RSMA) on the independent variables. The first multiple linear regression model used lagged independent variables to capture the delayed effects of news sentiment on stock market returns. The goal of this model was to understand the impact of news sentiment on stock market returns over a longer period of time, considering any delays that may have occurred. To do this, the independent variables, such as Wall Street Journal news sentiment and COVID-19 case numbers, were lagged by a certain number of trading days. The second multiple linear regression model used an RSMA on the independent variables to capture the persistence of the lagged effect of news sentiment on stock market returns. This model aimed to understand the trend of the relationship between news sentiment and stock market returns over time, and to determine whether the impact of news sentiment on stock market returns is sustained over a longer period of time. By using an RSMA on the independent variables, the analysis was able to identify any trends in the relationship between news sentiment and stock market returns and make predictions about how changes in news sentiment may impact stock market returns in the future. Both multiple linear regression models were fit to the S&P 500 and DJIA stock market returns, with the independent variables being the sentiment of Wall Street Journal news and the number of COVID-19 cases. By using two different models, the analysis was able to provide a comprehensive examination of the relationship between news sentiment and stock market returns, and to better understand the dynamics of this relationship.

3.1.1 Models

The model with lagged independent variables is constructed as it follows:

$Y_t = \propto + \beta_0 X^1_{t-i} + \beta_1 X^2_{t-i} + \epsilon_t \quad \epsilon_t \sim N\,(0, \sigma^2)$; i = 3,5

i.e., 3 and 5 represent the number of trading days by which the variables were lagged.

The model with Rolling Simple Moving Average of independent variables, instead, is constructed this way:

$Y_t = \propto + \beta_0 X^1_i + \beta_1 X^2_i + \epsilon_t \quad \epsilon_t \sim N\,(0, \sigma^2)$; i = 5, 15, 20, 25, 30

Where X^1_i and X^2_i are time series of the RSMAs computed every day on variables, where i represents the duration of the rolling window.

i.e., 5, 15, 20, 25, 30 represent the duration in trading days of the rolling window of the SMA.

In conclusion, multiple linear regression is a useful tool for sentiment analysis in finance and economics and can be applied to the relationship between news sentiment and stock market returns. By controlling for the influence of other variables and isolating the impact of news sentiment, multiple linear regression can help us to better understand the relationship between the news coverage of the COVID-19 pandemic and the Wall Street Journal and the returns of the S&P 500 and the Dow Jones Industrial Average.

3.1.2 Variables

I put a large number of different combinations of variables through a rigorous and systematic testing process in order to find out which independent elements had the greatest capacity to explain stock returns. This was done so that I could identify those components. To be more specific, I used a multi-stage selection process in order to zero down on the independent factors that were the most important. In order to determine which of the approximately 10,000 possible variable combinations offered the best accurate projection of future stock returns, this technique required doing an analysis. In order to further improve my research, I conducted further experiments in which I examined hundreds of different combinations of lagging and rolling window strategies in order to identify the most effective approach. I was able to convey the fluidity and complexity of the nature of changes in stock prices by using a number of different time delays in conjunction with rolling windows. With the use of this strategy, I was able to establish not only the most significant independent variables but also the optimal window widths and time delays to apply to each variable. I looked at 106 different time series as potential candidates for inclusion in my study and examined them. These time series had a number of variables, and of course, each variable was in some way related to the COVID database. Because I looked at such a large number of different variables, I was able to determine the factors that had the greatest impact on stock returns and accurately portray the complex interplay between them. Throughout the whole of the inquiry, I used a technique that was analytical and iterative in nature in order to find the factors and parameter settings that would provide the best results. To be more specific, I disregarded the outputs that were the least successful at each level of the study so that I could focus on the combinations that had the most potential. Because of this methodology, I was able to gradually improve the quality of my study and identify the components that had the greatest capacity to explain stock returns. It is vital to exercise caution and not rely only on statistical analysis in any endeavor; this is true even if the stepwise selection strategy proves to be an effective method for identifying factors that significantly affect the outcome of the study. Other factors, such as changes in the economy and the state of the market, may also be factors that impact stock returns. I found the most significant independent elements for predicting stock returns by adhering to a strict and meticulous technique. I was able to uncover the variables with the highest explanatory power and capture the complexity of stock price movements by assessing a large number of components and attempting to use a variety of approaches.

3.2 Results presentation

In this section I'll present the results of my analysis with different tables that describe the output of the multiple regression models which are the objective of the analysis.

3.2.1 Lagged Model

Here I present the results of my lagged model. [7]

SP 500 3 days lag results

SP500 3 Days Lag	coef	std err	t	P>|t|
Intercept	0.0018	0.001	1.334	0.183
cas_n	-0.1805	0.050	-3.613	**0.000**
stm_ws	-0.0375	0.016	-2.324	**0.021**
R2	0.071			
# obs	245			

SP500 3 Days Lag	coef	std err	t	P>|t|
Intercept	0.0026	0.001	1.820	0.070
cas_t	-0.3437	0.113	-3.038	**0.003**
stm_ws	-0.0357	0.016	-2.192	**0.029**
R2	0.056			
# obs	245			

SP500 3 Days Lag	coef	std err	t	P>|t|
Intercept	0.0018	0.001	1.365	0.173
cas_n	-0.1814	0.050	-3.638	**0.000**
ax_m_ws	-0.0325	0.013	-2.510	**0.013**
R2	0.074			
# obs	245			

SP500 3 Days Lag	coef	std err	t	P>|t|
Intercept	0.0026	0.001	1.846	0.066
cas_t	-0.3437	0.113	-3.043	**0.003**
ax_m_ws	-0.0307	0.013	-2.358	**0.019**
R2	0.059			
# obs	245			

Figure 9. Table of results using 3 days lagged independent variables.

[7] In this section every time I name independent variables they have to be intended as lagged by 3 and 5 trading days.

Dow Jones 3 days lag results

DOW 3 Days Lag	coef	std err	t	P>|t|
Intercept	0.0015	0.001	1.084	0.279
cas_n	-0.1836	0.053	-3.444	**0.001**
stm_ws	-0.0364	0.017	-2.112	**0.036**
R2	0.063			
# obs	245			

DOW 3 Days Lag	coef	std err	t	P>|t|
Intercept	0.0024	0.002	1.594	0.112
cas_t	-0.3622	0.120	-3.006	**0.003**
stm_ws	-0.0344	0.017	-1.988	**0.048**
R2	0.052			
# obs	245			

DOW 3 Days Lag	coef	std err	t	P>|t|
Intercept	0.0016	0.001	1.109	0.268
cas_n	-0.1845	0.053	-3.463	**0.001**
ax_m_ws	-0.0309	0.014	-2.236	**0.026**
R2	0.065			
# obs	245			

DOW 3 Days Lag	coef	std err	t	P>|t|
Intercept	0.0024	0.002	1.615	0.108
cas_t	-0.3624	0.120	-3.011	**0.003**
ax_m_ws	-0.0291	0.014	-2.093	**0.037**
R2	0.054			
# obs	245			

Figure 10. Table of results using 3 days lagged independent variables.

The results of my research show that every one of the coefficients in this set of tables is statistically significant and has a value that is distinct from zero. This was determined by comparing the coefficients to the values shown in the content tables that came before and after it. This demonstrates that the variables chosen to be independent have a strong explanatory ability for predicting future changes in the values of the variable that is dependent on them. In particular, I came to the conclusion that the overall number of COVID occurrences was twice as helpful in defining the fluctuations in the dependent variable as the daily number of cases. According to these results, it would seem that the overall number of instances is a more reliable predictor of future changes in the dependent variable than the daily number of cases is. This conclusion was reached after analysing the tables. In addition, I noticed that the coefficient values for all of the WSJ news variables were quite similar to one another. The findings provide credence to the idea that both play a comparable part in forecasting changes in the variable that is being predicted over time. The fact that each of the WSJ news variables has the same value for its coefficient indicates that all of these factors are significant predictors of shifts in the dependent variable, which includes news mood and news tone. It is essential to emphasize that statistical study has its limits, and it is possible that stock returns may also be affected by other variables, such as the circumstances of the market or unplanned occurrences.

As a result, the purpose of the lagging variables in my study is to estimate the influence that their changes have had, although belatedly, on the variables that are being investigated. According to the findings of my investigation, I have come to the conclusion that the consequences of such an impact might start at anywhere between three and five days. My research has led me to the conclusion that there is no correlation between a delay of only one or two days in the independent variables and the production of results that are statistically significant. Surprisingly, the same results were shown even when the components were delayed by more than six days. This would imply that the effects of the causes on the dependent variable are delayed by a few days at the very least. Because they emphasize the need of choosing an adequate lag time for independent variables, these results have significant repercussions for my statistical research. It's possible that making incorrect assumptions about the relationships between variables might result from an inadequate evaluation of the lag time. In addition, the data point to the possibility that there is a delay between the occurrence of an event and the influence that it has on the variable that is being studied. It is vital to take temporality into account if one is conducting an investigation into the influence of explanatory factors on dependent variables. This indicates that we should not anticipate this kind of news release to effect financial returns either sooner or later than three to five trading days after the announcement. Instead, we should make these expectations based on the timing of the announcement. This data lends support to the concept that recent news pertaining to COVID might have an impact on stock prices at the present moment. As such, there is a connection between it and the results of the second portion of the section about RSMA model research.

SP 500 5 days lag results

SP500 5 Days Lag	coef	std err	t	P>|t|
Intercept	0.0018	0.001	1.388	0.166
cas_n	-0.2074	0.050	-4.159	**0.000**
stm_ws	-0.0307	0.016	-1.907	**0.058**
R2	0.080			
# obs	245			

SP500 5 Days Lag	coef	std err	t	P>|t|
Intercept	0.0024	0.001	1.712	0.088
cas_t	-0.3165	0.114	-2.780	**0.006**
stm_ws	-0.0292	0.016	-1.779	**0.076**
R2	0.044			
# obs	245			

SP500 5 Days Lag	coef	std err	t	P>|t|
Intercept	0.0019	0.001	1.406	0.161
cas_n	-0.2078	0.050	-4.168	**0.000**
ax_m_ws	-0.0250	0.013	-1.933	**0.054**
R2	0.080			
# obs	245			

SP500 5 Days Lag	coef	std err	t	P>|t|
Intercept	0.0025	0.001	1.727	0.085
cas_t	-0.3169	0.114	-2.784	**0.006**
ax_m_ws	-0.0237	0.013	-1.795	**0.074**
R2	0.045			
# obs	245			

Figure 11. Table of results using 5 days lagged independent variables.

Dow Jones 5 days lag results

DOW 5 Days Lag	coef	std err	t	P>|t|
Intercept	0.0017	0.001	1.169	0.243
cas_n	-0.2256	0.053	-4.255	**0.000**
stm_ws	-0.0279	0.017	-1.629	**0.105**
R2	0.079			
# obs	245			

DOW 5 Days Lag	coef	std err	t	P>|t|
Intercept	0.0023	0.002	1.516	0.131
cas_t	-0.3445	0.121	-2.844	**0.005**
stm_ws	-0.0262	0.017	-1.503	**0.134**
R2	0.042			
# obs	245			

DOW 5 Days Lag	coef	std err	t	P>|t|
Intercept	0.0017	0.001	1.188	0.236
cas_n	-0.2260	0.053	-4.264	**0.000**
ax_m_ws	-0.0236	0.014	-1.718	**0.087**
R2	0.080			
# obs	245			

DOW 5 Days Lag	coef	std err	t	P>|t|
Intercept	0.0023	0.002	1.532	0.127
cas_t	-0.3447	0.121	-2.847	**0.005**
ax_m_ws	-0.0222	0.014	-1.580	**0.115**
R2	0.043			
# obs	245			

Figure 12. Table of results using 5 days lagged independent variables.

3.2.2 RSMA Model

Here I present my output extracted from the Rolling Window Simple Moving Average model. [8]

SP500 and Dow Jones 5 days SMA results

SP500 5 Days SMA	coef	std err	t	P>|t|
Intercept	0.0023	0.001	1.639	0.102
cas_n	-0.1159	0.051	-2.270	**0.024**
stm_ws	-0.1708	0.072	-2.385	**0.018**
R2	0.050			
# obs	245			

SP500 5 Days SMA	coef	std err	t	P>|t|
Intercept	0.0027	0.001	1.849	0.066
cas_n	-0.1211	0.051	-2.373	**0.018**
ng_m_ws	-0.0609	0.029	-2.068	**0.040**
R2	0.045			
# obs	245			

SP500 5 Days SMA	coef	std err	t	P>|t|
Intercept	0.0023	0.001	1.638	0.103
cas_n	-0.1218	0.051	-2.388	**0.018**
ax_m_ws	-0.1233	0.060	-2.072	**0.039**
R2	0.045			
# obs	245			

DOW 5 Days SMA	coef	std err	t	P>|t|
Intercept	0.0022	0.001	1.473	0.142
cas_n	-0.1345	0.054	-2.486	**0.014**
stm_ws	-0.1852	0.076	-2.440	**0.015**
R2	0.056			
# obs	245			

DOW 5 Days SMA	coef	std err	t	P>|t|
Intercept	0.0027	0.002	1.701	0.090
cas_n	-0.1402	0.054	-2.592	**0.010**
ng_m_ws	-0.0660	0.031	-2.116	**0.035**
R2	0.050			
# obs	245			

DOW 5 Days SMA	coef	std err	t	P>|t|
Intercept	0.0022	0.001	1.465	0.144
cas_n	-0.1411	0.054	-2.610	**0.010**
ax_m_ws	-0.1315	0.063	-2.084	**0.038**
R2	0.049			
# obs	245			

Figure 13. Table of results using 5 days rolling window for SMA of independent variables

[8] In this section every time I name independent variables they have to be intended as a time series of RSMA

SP500 and Dow Jones 10 days SMA results

SP500 10 Days SMA	coef	std err	t	P>|t|
Intercept	0.0026	0.001	1.856	0.065
cas_n	-0.1008	0.052	-1.954	**0.052**
stm_ws	-0.2351	0.082	-2.871	**0.004**
R2	0.060			
# obs	245			

SP500 10 Days SMA	coef	std err	t	P>|t|
Intercept	0.0034	0.002	2.246	0.026
cas_n	-0.1124	0.051	-2.203	**0.029**
ng_m_ws	-0.0933	0.035	-2.664	**0.008**
R2	0.055			
# obs	245			

SP500 10 Days SMA	coef	std err	t	P>|t|
Intercept	0.0026	0.001	1.834	0.068
cas_n	-0.1078	0.052	-2.089	**0.038**
ax_m_ws	-0.1695	0.069	-2.446	**0.015**
R2	0.051			
# obs	245			

DOW 10 Days SMA	coef	std err	t	P>|t|
Intercept	0.0025	0.001	1.679	0.094
cas_n	-0.1188	0.055	-2.174	**0.031**
stm_ws	-0.2500	0.087	-2.879	**0.004**
R2	0.064			
# obs	245			

DOW 10 Days SMA	coef	std err	t	P>|t|
Intercept	0.0035	0.002	2.141	0.033
cas_n	-0.1303	0.054	-2.412	**0.017**
ng_m_ws	-0.1032	0.037	-2.781	**0.006**
R2	0.062			
# obs	245			

DOW 10 Days SMA	coef	std err	t	P>|t|
Intercept	0.0025	0.002	1.643	0.102
cas_n	-0.1269	0.055	-2.318	**0.021**
ax_m_ws	-0.1764	0.074	-2.398	**0.017**
R2	0.055			
# obs	245			

Figure 14. Table of results using 10 days rolling window for SMA of independent variables

SP500 and Dow Jones 15 days SMA results

SP500 15 Days SMA	coef	std err	t	P>|t|
Intercept	0.0028	0.001	1.960	0.051
cas_n	-0.1009	0.051	-1.964	**0.051**
stm_ws	-0.2546	0.086	-2.974	**0.003**
R2	0.062			
# obs	245			

SP500 15 Days SMA	coef	std err	t	P>|t|
Intercept	0.0038	0.002	2.361	0.019
cas_n	-0.1101	0.051	-2.153	**0.032**
ng_m_ws	-0.1053	0.039	-2.678	**0.008**
R2	0.056			
# obs	245			

SP500 15 Days SMA	coef	std err	t	P>|t|
Intercept	0.0029	0.001	2.039	0.043
cas_n	-0.1050	0.051	-2.048	**0.042**
ax_m_ws	-0.2079	0.073	-2.831	**0.005**
R2	0.059			
# obs	245			

DOW 15 Days SMA	coef	std err	t	P>|t|
Intercept	0.0026	0.001	1.743	0.083
cas_n	-0.1205	0.055	-2.209	**0.028**
stm_ws	-0.2590	0.091	-2.848	**0.005**
R2	0.064			
# obs	245			

DOW 15 Days SMA	coef	std err	t	P>|t|
Intercept	0.0037	0.002	2.171	0.031
cas_n	-0.1294	0.054	-2.384	**0.018**
ng_m_ws	-0.1093	0.042	-2.618	**0.009**
R2	0.059			
# obs	245			

DOW 15 Days SMA	coef	std err	t	P>|t|
Intercept	0.0028	0.002	1.808	0.072
cas_n	-0.1251	0.054	-2.296	**0.023**
ax_m_ws	-0.2086	0.078	-2.674	**0.008**
R2	0.060			
# obs	245			

Figure 14. Table of results using 15 days rolling window for SMA of independent variables

SP500 and Dow Jones 20 days SMA results

SP500 20 Days SMA	coef	std err	t	P>|t|
Intercept	0.0031	0.001	2.182	0.030
cas_n	-0.0947	0.051	-1.848	**0.066**
stm_ws	-0.3032	0.089	-3.401	**0.001**
R2	0.072			
# obs	245			

SP500 20 Days SMA	coef	std err	t	P>|t|
Intercept	0.0045	0.002	2.738	0.007
cas_n	-0.1040	0.051	-2.042	**0.042**
ng_m_ws	-0.1347	0.042	-3.194	**0.002**
R2	0.067			
# obs	245			

SP500 20 Days SMA	coef	std err	t	P>|t|
Intercept	0.0033	0.001	2.276	0.024
cas_n	-0.0989	0.051	-1.935	**0.054**
ax_m_ws	-0.2516	0.077	-3.279	**0.001**
R2	0.069			
# obs	245			

DOW 20 Days SMA	coef	std err	t	P>|t|
Intercept	0.0030	0.002	1.966	0.050
cas_n	-0.1138	0.054	-2.092	**0.038**
stm_ws	-0.3113	0.095	-3.288	**0.001**
R2	0.074			
# obs	245			

DOW 20 Days SMA	coef	std err	t	P>|t|
Intercept	0.0045	0.002	2.590	0.010
cas_n	-0.1222	0.054	-2.264	**0.024**
ng_m_ws	-0.1432	0.045	-3.203	**0.002**
R2	0.072			
# obs	245			

DOW 20 Days SMA	coef	std err	t	P>|t|
Intercept	0.0031	0.002	2.042	0.042
cas_n	-0.1187	0.054	-2.185	**0.030**
ax_m_ws	-0.2544	0.082	-3.121	**0.002**
R2	0.070			
# obs	245			

Figure 15. Table of results using 20 days rolling window for SMA of independent variables

SP500 and Dow Jones 25 days SMA results

SP500 25 Days SMA	coef	std err	t	P>|t|
Intercept	0.0031	0.001	2.196	0.029
cas_n	-0.0911	0.052	-1.764	**0.079**
stm_ws	-0.3018	0.092	-3.294	**0.001**
R2	0.069			
# obs	245			

SP500 25 Days SMA	coef	std err	t	P>|t|
Intercept	0.0045	0.002	2.691	0.008
cas_n	-0.1008	0.051	-1.964	**0.051**
ng_m_ws	-0.1346	0.045	-3.017	**0.003**
R2	0.063			
# obs	245			

SP500 25 Days SMA	coef	std err	t	P>|t|
Intercept	0.0033	0.001	2.293	0.023
cas_n	-0.0949	0.051	-1.844	**0.066**
ax_m_ws	-0.2532	0.079	-3.216	**0.001**
R2	0.067			
# obs	245			

DOW 25 Days SMA	coef	std err	t	P>|t|
Intercept	0.0030	0.002	2.001	0.046
cas_n	-0.1094	0.055	-1.996	**0.047**
stm_ws	-0.3149	0.097	-3.239	**0.001**
R2	0.073			
# obs	245			

DOW 25 Days SMA	coef	std err	t	P>|t|
Intercept	0.0046	0.002	2.571	0.011
cas_n	-0.1185	0.054	-2.177	**0.030**
ng_m_ws	-0.1450	0.047	-3.065	**0.002**
R2	0.069			
# obs	245			

DOW 25 Days SMA	coef	std err	t	P>|t|
Intercept	0.0032	0.002	2.074	0.039
cas_n	-0.1142	0.055	-2.089	**0.038**
ax_m_ws	-0.2590	0.084	-3.098	**0.002**
R2	0.069			
# obs	245			

Figure 15. Table of results using 25 days rolling window for SMA of independent variables

SP500 and Dow Jones 30 days SMA results

SP500 30 Days SMA	coef	std err	t	P>|t|
Intercept	0.0032	0.001	2.231	0.027
cas_n	-0.0905	0.052	-1.754	**0.081**
stm_ws	-0.3090	0.092	-3.360	**0.001**
R2	0.071			
# obs	245			

SP500 30 Days SMA	coef	std err	t	P>|t|
Intercept	0.0046	0.002	2.740	0.007
cas_n	-0.0998	0.051	-1.944	**0.053**
ng_m_ws	-0.1385	0.045	-3.083	**0.002**
R2	0.064			
# obs	245			

SP500 30 Days SMA	coef	std err	t	P>|t|
Intercept	0.0034	0.001	2.331	0.021
cas_n	-0.0946	0.051	-1.841	**0.067**
ax_m_ws	-0.2597	0.079	-3.287	**0.001**
R2	0.069			
# obs	245			

DOW 30 Days SMA	coef	std err	t	P>|t|
Intercept	0.0031	0.002	2.027	0.044
cas_n	-0.1090	0.055	-1.991	**0.048**
stm_ws	-0.3204	0.098	-3.281	**0.001**
R2	0.074			
# obs	245			

DOW 30 Days SMA	coef	std err	t	P>|t|
Intercept	0.0046	0.002	2.561	0.011
cas_n	-0.1184	0.054	-2.174	**0.031**
ng_m_ws	-0.1447	0.048	-3.034	**0.003**
R2	0.068			
# obs	245			

DOW 30 Days SMA	coef	std err	t	P>|t|
Intercept	0.0033	0.002	2.114	0.036
cas_n	-0.1138	0.055	-2.084	**0.038**
ax_m_ws	-0.2666	0.084	-3.176	**0.002**
R2	0.071			
# obs	245			

Figure 16. Table of results using 30 days rolling window for SMA of independent variables

The analysis that was performed as a result provides valuable insights into the connection between the COVID-19 cases and the influence that these cases have on the stock market. The research concentrates on the first year of the pandemic and provides a comprehensive understanding of how news and COVID-19 cases affected the market throughout this time frame. The results of the investigation demonstrate that each and every coefficient contained within the set of tables possesses statistical significance and is distinct from zero. This suggests that the chosen independent variables have a strong explanatory ability for predicting future changes in the values of the variable that is being studied (the dependent variable). It was found that the negative impact of the total number of cases had the same level of consistency throughout all of the rolling windows that were examined. This finding demonstrates that cases have had a significant impact on market indices for a considerable amount of time. Over a longer length of time, the independent variables had a higher impact on the outcome of the study. Each panel represents a different rolling window that was used in the calculation of the simple moving average. According to the findings, the affects that were detected continued for at least five trading days, which is equivalent to one trading week. In addition, it was found that the effects lasted for as long as 30 trading days, which is equivalent to six trading weeks. Realizing that news is published on a daily basis and that a single news item or case report does not have a significant impact on the market is a crucial part of staying abreast of current events. Rather, throughout the difficult year that was analysed, the general negative and anxious mood had an effect on the markets. In addition, the study came to the conclusion that incidences of COVID-19 had a large impact on the stock market, one that persisted for an extended period of time. The fact that the number of instances always has a negative effect, regardless of the rolling window being used, is suggestive of the fact that the number of cases has a considerable influence on market indices over an extended period of time. Other elements, of course, had a role in the market as well; nevertheless, it was seen that these aspects played an increasingly important role throughout the course of time. During the course of the analysis year, the model that was employed in the research was helpful in addressing the issue of the persistence of variable effects across time. It would appear that news and COVID-19 instances had a consistent impact on the market during the first year of the pandemic. As a consequence of this, one might draw the conclusion that price fluctuations on the market during this time period were very sensitive to the news and COVID-19 occurrences, and that this sensitivity persisted for a considerable amount of time. The findings of this study are important because they demonstrate the significant impact that COVID-19 had on the stock market during the first year that the epidemic was active. The study gives light on the sensitivity of the market to news and COVID-19 events, as well as how this sensitivity persisted for a significant amount of time. A comprehensive look is taken at the connection between the COVID-19 cases and the impact those cases have on the stock market. The analysis highlights the market's sensitivity to news and COVID-19 cases as well as the prolonged influence COVID-19 had on the market throughout the first year of the pandemic.

3.3 VIX Results

Here I report the results of the part of my regression analysis that focuses on the log returns of the VIX, that measures market volatility, and the independent variables used in previous analyses. However, I tried using only one variable at a time due to unsatisfactory results when both models were employed. I decided to focus on the log returns of the VIX was based on the fact that this variable may provide a more accurate representation of market volatility than the S&P500 and Dow Jones indices. This is because the VIX is specifically designed to measure market volatility, while the other two indices are broader measures of stock market performance. In this case, the dependent variable is the log returns of the VIX, while the independent variables are the same ones used in previous analyses. The results of the analysis indicate that the use of only one independent variable at a time resulted in better results than using two as the precedent cases.

3.3.1 Lagged Model

VIX 3 days lag results

VIX 3 Days Lag	coef	std err	t	P>|t|		VIX 3 Days Lag	coef	std err	t	P>|t|
Intercept	-0.0020	0.006	-0.361	0.719		Intercept	-0.0021	0.006	-0.373	0.710
stm_ws	0.1272	0.068	1.860	**0.064**		**ax_m_ws**	0.1007	0.055	1.833	**0.068**
R2	0.014					**R2**	0.014			
# obs	245					# obs	245			

Figure 17. Table of results using 3 days lagged independent variables.

Here I present a table that shows the results of a statistical analysis of the impact of COVID-related news on the VIX, DOW, and S&P 500. The P-value is somewhat high, but the coefficient can still be considered distinct from zero, indicating that COVID-related news had an impact on these indices. While the analysis suggests that the number of COVID cases did not have a statistically significant delayed impact on VIX returns[9], the results do suggest that the sentiment of news articles, specifically stress and anxiety words in the Wall Street Journal, had a delayed impact on the VIX. The analysis indicates that the effect of news on the VIX was delayed by three days, suggesting that there is a lag between the news cycle and the impact on the VIX.

[9] Cases numbers were not included because their coefficients were statistically not significant considering P-value.

Additionally, the coefficients in this analysis are positive, indicating an increase in volatility which is associated with bad market conditions, which is consistent with previous analyses of market volatility.

3.3.2 RSMA Model

VIX 5 days SMA results

VIX 5 Days SMA	coef	std err	t	P>|t|
Intercept	-0.0028	0.006	-0.496	0.620
cas_n	0.3854	0.212	1.820	**0.070**
R2	0.013			
# obs	245			

VIX 5 Days SMA	coef	std err	t	P>|t|
Intercept	-0.0037	0.006	-0.650	0.516
stm_ws	0.4567	0.251	1.817	**0.071**
R2	0.013			
# obs	245			

VIX 5 Days SMA	coef	std err	t	P>|t|
Intercept	-0.0054	0.006	-0.900	0.369
ng_m_ws	0.1856	0.101	1.838	**0.067**
R2	0.014			
# obs	245			

VIX 5 Days SMA	coef	std err	t	P>|t|
Intercept	-0.0033	0.006	-0.570	0.569
ax_m_ws	0.2581	0.210	2.226	**0.024**
R2	0.006			
# obs	245			

Figure 18. Table of results using 5 days rolling window for SMA of independent variables

VIX 10 days SMA results

VIX 10 Days SMA	coef	std err	t	P>|t|
Intercept	-0.0028	0.006	-0.496	0.620
cas_n	0.3854	0.212	1.820	**0.070**
R2	0.013			
# obs	245			

VIX 10 Days SMA	coef	std err	t	P>|t|
Intercept	-0.0060	0.006	-1.033	0.302
stm_ws	0.8683	0.334	2.599	**0.010**
R2	0.027			
# obs	245			

VIX 10 Days SMA	coef	std err	t	P>|t|
Intercept	-0.0092	0.006	-1.440	0.151
ng_m_ws	0.3534	0.145	2.443	**0.015**
R2	0.024			
# obs	245			

VIX 10 Days SMA	coef	std err	t	P>|t|
Intercept	-0.0061	0.006	-1.029	0.304
ax_m_ws	0.6394	0.284	2.255	**0.025**
R2	0.021			
# obs	245			

Figure 19. Table of results using 10 days rolling window for SMA of independent variables

VIX 15 days SMA results

VIX 15 Days SMA	coef	std err	t	P>|t|
Intercept	-0.0028	0.006	-0.496	0.620
cas_n	0.3854	0.212	1.820	**0.070**
R2	0.013			
# obs	245			

VIX 15 Days SMA	coef	std err	t	P>|t|
Intercept	-0.0081	0.006	-1.377	0.170
stm_ws	1.1851	0.352	3.368	**0.001**
R2	0.045			
# obs	245			

VIX 15 Days SMA	coef	std err	t	P>|t|
Intercept	-0.0146	0.007	-2.204	0.028
ng_m_ws	0.5801	0.162	3.572	**0.000**
R2	0.050			
# obs	245			

VIX 15 Days SMA	coef	std err	t	P>|t|
Intercept	-0.0084	0.006	-1.407	0.161
ax_m_ws	0.9214	0.303	3.041	**0.003**
R2	0.037			
# obs	245			

Figure 20. Table of results using 15 days rolling window for SMA of independent variables

VIX 20 days SMA results

VIX 20 Days SMA	coef	std err	t	P>|t|
Intercept	-0.0028	0.006	-0.496	0.620
cas_n	0.3854	0.212	1.820	**0.070**
R2	0.013			
# obs	245			

VIX 20 Days SMA	coef	std err	t	P>|t|
Intercept	-0.0073	0.006	-1.232	0.219
stm_ws	1.0176	0.366	2.779	**0.006**
R2	0.031			
# obs	245			

VIX 20 Days SMA	coef	std err	t	P>|t|
Intercept	-0.0127	0.007	-1.865	0.063
ng_m_ws	0.4901	0.173	2.834	**0.005**
R2	0.032			
# obs	245			

VIX 20 Days SMA	coef	std err	t	P>|t|
Intercept	-0.0078	0.006	-1.287	0.199
ax_m_ws	0.8240	0.316	2.611	**0.010**
R2	0.027			
# obs	245			

Figure 21. Table of results using 20 days rolling window for SMA of independent variables

VIX 25 days SMA results

VIX 25 Days SMA	coef	std err	t	P>|t|
Intercept	-0.0028	0.006	-0.496	0.620
cas_n	0.3854	0.212	1.820	**0.070**
R2	0.013			
# obs	245			

VIX 25 Days SMA	coef	std err	t	P>|t|
Intercept	-0.0076	0.006	-1.279	0.202
stm_ws	1.0437	0.370	2.823	**0.005**
R2	0.032			
# obs	245			

VIX 25 Days SMA	coef	std err	t	P>|t|
Intercept	-0.0122	0.007	-1.763	0.079
ng_m_ws	0.4646	0.180	2.586	**0.010**
R2	0.027			
# obs	245			

VIX 25 Days SMA	coef	std err	t	P>|t|
Intercept	-0.0085	0.006	-1.398	0.164
ax_m_ws	0.8990	0.319	2.816	**0.005**
R2	0.032			
# obs	245			

Figure 22. Table of results using 25 days rolling window for SMA of independent variables

VIX 30 days SMA results

VIX 30 Days SMA	coef	std err	t	P>|t|
Intercept	-0.0028	0.006	-0.496	0.620
cas_n	0.3854	0.212	1.820	**0.070**
R2	0.013			
# obs	245			

VIX 30 Days SMA	coef	std err	t	P>|t|
Intercept	-0.0079	0.006	-1.315	0.190
stm_ws	1.0640	0.374	2.845	**0.005**
R2	0.032			
# obs	245			

VIX 30 Days SMA	coef	std err	t	P>|t|
Intercept	-0.0132	0.007	-1.874	0.062
ng_m_ws	0.4998	0.184	2.718	**0.007**
R2	0.030			
# obs	245			

VIX 30 Days SMA	coef	std err	t	P>|t|
Intercept	-0.0085	0.006	-1.391	0.166
ax_m_ws	0.8873	0.323	2.748	**0.006**
R2	0.030			
# obs	245			

Figure 23. Table of results using 30 days rolling window for SMA of independent variables

Here, instead, I present the results associated with various rolling windows in which the analysis was performed. The first thing to notice is that cases numbers are present in tables and so their coefficients are statistically significant, this indicates that the effect of cases on VIX is more persistent rather then delayed, i.e. it kicks in right after the release and is very persistent during time, especially if the numbers have a consistent path over time. The same can be said about negative tone which wasn't present in the lagged model but is present here, generally speaking, as we've said before, results like this indicates that the effects, positive in the case of VIX, are persistent over time, and are very much influenced by their own consistency, i.e. if the bad mood is consistent over time, its effects will be much more consistent over time, as we can see in the precedent tables from the fact that even in a 30 days rolling window SMA, the effects are still have a veery big impact on markets volatility, represented by the VIX in this case.

CONCLUSIONS

During the first year of the pandemic, I carried out an in-depth investigation on the connection that exists between the performance of the stock market and the number of cases and news stories concerning COVID-19. This was a difficult moment for humanity, as the epidemic had a significant effect on every facet of our existence. As a result, we faced many challenges. I observed that there was a substantial association between the overall mood and sentiment of the population and the returns on the stock market as I was going through the data. On the other hand, the effects were delayed due to the nature of the variables, as the markets are not influenced by individual pieces of news but rather are concerned about the persistence of them. In the beginning phases of the pandemic, there was a distinct pattern of sudden manifestations of dread and uncertainty over health and the economy. As a consequence of this, the markets went through a big bear market and entered a negative region in terms of returns. I observed that the effects of COVID-19 related news and the number of instances on the financial markets were long-lasting as I dug deeper into the data. It took a significant amount of time for the benefits to become noticeable, but once they did, they were constant. In the course of carrying out this study, I'm aware of a number of limitations and potential areas for enhancement. For example, a new set of factors may be taken into consideration, or an even more complex model, such as the SIR model, which is designed specifically for pandemics, could be employed instead. In addition, dummy variables could be incorporated into the analysis in order to get further comprehension of the data. However, I believe that the analysis that I carried out with the variables and models that I used is the best that I could have done to provide a robust and consistent overview of one of the primary drivers of stock markets during the pandemic. I came to this conclusion because I believe that the analysis that I carried out is the best that I could have done. As I combed through the information, I noticed a distinct pattern developing. The consequences of

COVID-19 related news and the number of instances had a substantial impact on the stock markets, but it took some time for those repercussions to manifest. This should not have come as a surprise given that the markets are not easily moved by a single piece of news but rather by patterns and trends that develop over time. Because of this, serious damage to the markets did not occur until the epidemic had already been going on for a considerable amount of time. When we looked at the statistics, it was very evident that the overall mood and feeling of the people had a substantial impact on the stock markets. This was especially true during times of high volatility. The markets took a knock whenever there was an atmosphere of uncertainty and fear throughout the world. In contrast, the markets began to recover whenever there was a sense of optimism and hope in the air. This was especially clear, for example, when it came to the case of vaccine announcements, which typically had a favourable effect on the stock markets. In spite of the constraints of the study, I feel that the findings are sound and that they provide useful insights into the relationship between COVID-19 related news and the number of instances and stock market returns. The research can assist investors better understand the impact of pandemics on the economy and take necessary efforts to alleviate the effects of these pandemics by highlighting the delayed and long-lasting effects of these variables. The findings of the study indicate that the connection between COVID-19-related news stories, the number of instances, and the performance of the stock market is a convoluted one that involves multiple factors. Once they take hold, the consequences, notwithstanding the possibility that they will be delayed, will be profound and long-lasting. As a result of this, it is essential to pay close attention to trends and patterns throughout time, and to take actions that are appropriate in order to limit the consequences of pandemics on the economy.

APPENDIX

```
#import useful libraries
import statsmodels.formula.api as smf
from itertools import product
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sn

##import dataset and select trading days from the series in order to use just significant data for the regression
dataset: pd.DataFrame = pd.read_excel('C:/Users/andrea/Dropbox/Il mio PC (LAPTOP-65PVQTIR)/Desktop/TESI/dati.xlsx', index_col='date')
business_day = pd.date_range(start='2020-03-09', end='2021-02-26', freq='B')
dataset = dataset.loc[business_day]

##dropping NaNs in order to exclude "not a number" observations
dataset.dropna(inplace=True)

##importing raw dataset to compute correlation Matrix
datasetcorr: pd.DataFrame = pd.read_excel('C:/Users/andrea/Dropbox/Il mio PC (LAPTOP-65PVQTIR)/Desktop/TESI/datasetforcorr.xlsx', index_col='date')
business_day = pd.date_range(start='2020-03-09', end='2021-02-26', freq='B')
datasetcorr = datasetcorr.loc[business_day]
datasetcorr.dropna(inplace=True)

##correlation matrix with heatmap
corrMatrix = datasetcorr.corr()
sn.heatmap(corrMatrix, annot=False, cmap= 'viridis')
plt.show()

##excel data mapipulation to obtain growth rate in every variable

#re-import growth rate dataset
dataset: pd.DataFrame = pd.read_excel("C:/Users/andrea/Dropbox/Il mio PC (LAPTOP-65PVQTIR)/Desktop/TESI/datos.xlsx", index_col='date')

#computing descriptive stats
statistics = datasetcorr.describe(include='all')
statistics.to_csv('statistics.csv')

## compute moving average to capture persistence in the effects of independent variables
x = dataset.iloc[:, [ 10, 11, 12]]
SMA = x.rolling(30, min_periods=1).mean()
dataset1 = dataset.iloc[:, [0, 1, 3]]
dataset1 = dataset1.join(SMA)

##after dozens of generations I have selected the significant variables to delivery final results, here I extract them from the initial dataset
y = dataset1.columns[:2]
x1 = dataset1.columns[2:3]
x2 = dataset1.columns[3:]

## defining the variables and the formula that the for cycle will compute in order to generate ols regression results
combo = list(product(y, x1, x2))
regression = [s[0] + " ~ " + s[1] + "+" + s[2] for s in combo]

##creation of a blank excel sheet as a destination for results
from openpyxl import load_workbook
path = "C:/Users/andrea/Dropbox/Il mio PC (LAPTOP-65PVQTIR)/Desktop/TESI/cose utili/ma30.xlsx"
book = load_workbook(path)
writer = pd.ExcelWriter(path, engine = 'openpyxl')
writer.book = book

```

```python
##this is the actual cycle that runs the regressions, and stores them in the excel sheet, it has been optimized by adding Rsquare and N of observations to make the results more complete
res = []
for formula in regression:
    reg = smf.ols(formula=formula, data=dataset1).fit()

    tab = pd.DataFrame(pd.read_html(reg.summary().as_html())[1])
    tab.columns = tab.loc[0]
    tab = tab.iloc[1:]
    tab = tab.rename(columns={np.nan: ' '})
    tab = tab.set_index(' ')
    others = pd.DataFrame(
         {
             'R2': [reg.rsquared, *[np.nan] * (tab.shape[1] - 1)],
             '# obs': [reg.nobs, *[np.nan] * (tab.shape[1] - 1)]
         },
         index=tab.columns,
    ).transpose()
    tab = pd.concat([tab, others])
    ## this stores results in the excel
    res.append(tab)
    res[-1].to_excel(writer, sheet_name = formula.replace(" ", "").replace("~", "="))

writer.save()
writer.close()
```

This is the python code I used to compute SP500 and Dow Jones RSMA regressions.

```python
##import useful libraries
import statsmodels.formula.api as smf
from itertools import product
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sn

##import dataset and select only trading days from the series in order to use just significant data for the regression
dataset: pd.DataFrame = pd.read_excel('C:/Users/andrea/Dropbox/Il mio PC (LAPTOP-65PVQTIR)/Desktop/TESI/dati.xlsx', index_col='date')
business_day = pd.date_range(start='2020-03-09', end='2021-02-26', freq='B')
dataset = dataset.loc[business_day]

##dropping NaNs in order to exclude "not a number" observations
dataset.dropna(inplace=True)

##computing descriptive stats
stats = dataset.describe(include= 'all')
descstats = stats
descstats.to_csv('descstats.csv')

##excel data mapipulation to obtain growth rate in every variable

#re-import growth rate dataset
dataset: pd.DataFrame = pd.read_excel("C:/Users/andrea/Dropbox/Il mio PC (LAPTOP-65PVQTIR)/Desktop/TESI/datos.xlsx", index_col='date')

##compute shift to capture lagged effects of independent variables
dataset1= dataset.iloc[:, [3, 4, 10, 12]]
dataset1=dataset1.shift(periods=2,fill_value=0)
dep= dataset.iloc[:, :2]
dataset1=dataset1.join(dep)

##after dozens of generations I have selected the significant variables to delivery final results, here I extract them from the initial dataset 0
y = dataset1.columns[4:]
x1 = dataset1.columns[:2]
x2 = dataset1.columns[2:4]

## defining the variables and the formula that the for cycle will compute in order to generate ols regression results
combo = list(product(y, x1, x2))
regression = [s[0] + " ~ " + s[1] + "+" + s[2] for s in combo]

##creation and setting of a blank excel sheet as a destination for results
from openpyxl import load_workbook
path = "C:/Users/andrea/Dropbox/Il mio PC (LAPTOP-65PVQTIR)/Desktop/TESI/shift3.xlsx"
book = load_workbook(path)
writer = pd.ExcelWriter(path, engine = 'openpyxl')
writer.book = book

## this is the actual cycle that runs the regressions, and stores them in the excel sheet, it has been optimized by adding Rsquare and N of observations to make the results more complete
res = []
for formula in regression:
    reg = smf.ols(formula=formula, data=dataset1).fit()
    tab = pd.DataFrame(pd.read_html(reg.summary().as_html())[1])
    tab.columns = tab.loc[0]
    tab = tab.iloc[1:]
    tab = tab.rename(columns={np.nan: ' '})
```

```
    tab = tab.set_index(' ')
    others = pd.DataFrame(
        {
            'R2': [reg.rsquared, *[np.nan] * (tab.shape[1] - 1)],
            '# obs': [reg.nobs, *[np.nan] * (tab.shape[1] - 1)]
        },
        index=tab.columns,
    ).transpose()
    tab = pd.concat([tab, others])
    ## this stores results in the excel sheet
    res.append(tab)
    res[-1].to_excel(writer, sheet_name = formula.replace(" ", "").replace("~", "="))

writer.save()
writer.close()
```

This is the code I used to compute lagged regressions on SP500 and Dow Jones.

```python
#importing useful libraries
import statsmodels.formula.api as smf
from itertools import product
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sn

##import dataset and select trading days from the series in order to use just significant data for the regression
dataset: pd.DataFrame = pd.read_excel('C:/Users/andrea/Dropbox/Il mio PC (LAPTOP-65PVQTIR)/Desktop/TESI/dati.xlsx', index_col='date')
business_day = pd.date_range(start='2020-03-09', end='2021-02-26', freq='B')
dataset = dataset.loc[business_day]

##dropping NaNs in order to exclude "not a number" observations
dataset.dropna(inplace=True)

##excel data mapipulation to obtain growth rate in every variable

#re-import growth rate dataset
dataset: pd.DataFrame = pd.read_excel("C:/Users/andrea/Dropbox/Il mio PC (LAPTOP-65PVQTIR)/Desktop/TESI/datos.xlsx", index_col='date')

## compute moving average to capture persistence in the effects of independent variables
x = dataset.iloc[:, [ 10, 11, 12]]
SMA = x.rolling(5, min_periods=1).mean()
dataset1 = dataset.iloc[:, [2, 3]]
dataset1 = dataset1.join(SMA)

##after dozens of generations I have selected the significant variables to delivery final results, here I extract them from the initial dataset
y = dataset1.columns[:1]
x1 = dataset1.columns[1:]

## defining the variables and the formula that the for cycle will compute in order to generate ols regression results
combo = list(product(y, x1))
regression = [s[0] + " ~ " + s[1] for s in combo]

##creation of a blank excel sheet as a destination for results
from openpyxl import load_workbook
path = "C:/Users/andrea/Dropbox/Il mio PC (LAPTOP-65PVQTIR)/Desktop/TESI/cose utili/vix5DaysSMA.xlsx"
book = load_workbook(path)
writer = pd.ExcelWriter(path, engine = 'openpyxl')
writer.book = book

##this is the actual cycle that runs the regressions, and stores them in the excel sheet, it has been optimized by adding Rsquare and N of observations to make the results more complete
res = []
for formula in regression:
    reg = smf.ols(formula=formula, data=dataset1).fit()

    tab = pd.DataFrame(pd.read_html(reg.summary().as_html())[1])
    tab.columns = tab.loc[0]
    tab = tab.iloc[1:]
    tab = tab.rename(columns={np.nan: ' '})
    tab = tab.set_index(' ')
    others = pd.DataFrame(
        {
            'R2': [reg.rsquared, *[np.nan] * (tab.shape[1] - 1)],
            '# obs': [reg.nobs, *[np.nan] * (tab.shape[1] - 1)]
        },
        index=tab.columns,
    ).transpose()
```

```
    tab = pd.concat([tab, others])
    ## this stores results in the excel
    res.append(tab)
    res[-1].to_excel(writer, sheet_name = formula.replace(" ", "").replace("~", "="))

writer.save()
writer.close()
```

This is the code I used to compute RSMA regressions on VIX

```python
#importing useful libraries
import statsmodels.formula.api as smf
from itertools import product
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sn

##import dataset and select trading days from the series in order to use just significant data for the regression
dataset: pd.DataFrame = pd.read_excel('C:/Users/andrea/Dropbox/Il mio PC (LAPTOP-65PVQTIR)/Desktop/TESI/dati.xlsx', index_col='date')
business_day = pd.date_range(start='2020-03-09', end='2021-02-26', freq='B')
dataset = dataset.loc[business_day]

##dropping NaNs in order to exclude "not a number" observations
dataset.dropna(inplace=True)

##excel data mapipulation to obtain growth rate in every variable

#re-import growth rate dataset
dataset: pd.DataFrame = pd.read_excel("C:/Users/andrea/Dropbox/Il mio PC (LAPTOP-65PVQTIR)/Desktop/TESI/datos.xlsx", index_col='date')

##compute shift to capture lagged effects of independent variables
dataset1= dataset.iloc[:, [3, 4, 10, 12]]
dataset1=dataset1.shift(periods=3,fill_value=0)
dep= dataset.iloc[:, 2:3]
dataset1=dataset1.join(dep)

##after dozens of generations I have selected the significant variables to delivery final results, here I extract them from the initial dataset
y = dataset1.columns[-1:]
x1 = dataset1.columns[:-1]

## defining the variables and the formula that the for cycle will compute in order to generate ols regression results
combo = list(product(y, x1))
regression = [s[0] + " ~ " + s[1] for s in combo]

##creation of a blank excel sheet as a destination for results
from openpyxl import load_workbook
path =" C:/Users/andrea/Dropbox/Il mio PC (LAPTOP-65PVQTIR)/Desktop/TESI/cose utili/vix3DaysLag.xlsx"
book = load_workbook(path)
writer = pd.ExcelWriter(path, engine = 'openpyxl')
writer.book = book

##this is the actual cycle that runs the regressions, and stores them in the excel sheet, it has been optimized by adding Rsquare and N of observations to make the results more complete
res = []
for formula in regression:
    reg = smf.ols(formula=formula, data=dataset1).fit()

    tab = pd.DataFrame(pd.read_html(reg.summary().as_html())[1])
    tab.columns = tab.loc[0]
    tab = tab.iloc[1:]
    tab = tab.rename(columns={np.nan: ' '})
    tab = tab.set_index(' ')
    others = pd.DataFrame(
         {
             'R2': [reg.rsquared, *[np.nan] * (tab.shape[1] - 1)],
             '# obs': [reg.nobs, *[np.nan] * (tab.shape[1] - 1)]
         },
         index=tab.columns,
    ).transpose()
```

```
    tab = pd.concat([tab, others])
    ## this stores results in the excel
    res.append(tab)
    res[-1].to_excel(writer, sheet_name = formula.replace(" ", "").replace("~", "="))

writer.save()
writer.close()
```

This is the code I used to compute lagged regressions on VIX.

Bibliography

Ali, and et al. 2020. “Coronavirus (COVID-19) – an epidemic or pandemic for financial markets.” *Journal of Behavioral and Experimental Finance,.*

Andrei, and Hasler. 2014. “Investor attention and stock market volatility.” *Review of Financial Studies.*

Aouadi, and et al. 2013. “Investor attention and stock market activity: evidence from France.” *Economic Modelling.*

Armatte. 2004. *Historia de la Probabilidad y la Estadistica.* Madrid: Delta Publicaciones.

Ashraf, Badar Nadeem. 2020. “Stock markets' reaction to COVID-19: Cases or fatalities?” *Elsevier.*

Baig, and et al. 2020. “Deaths, panic, lockdowns and US equity markets: the case of COVID-19 pandemic.” *Working Paper, SSRN: 3584947.*

Baker, and et al. 2020. “The unprecedented stock market reaction to covid-19.” *White paper - Becker Friedman Institute for economics at UChicago.*

Bank, and et al. 2011. “Google search volume and its influence on liquidity and returns of German stocks.” *Financial Markets and Portfolio Management.*

Barber, and Odean. 2008. “All the glitters: the effect of attention and news on the buying behaviour of individual and institutional investors.” *Review of Financial Studies,.*

Barber, and Odean. 2013. “The behaviour of individual investors.” *Handbook of the Economics of Finance, Elsevier.*

Barone. 2022. “Tracking Euro Stoxx 50.”

Bernstein. 1992. “Capital ideas.” *Free Press.*

Bijl, and et al. 2016. ““Google searches and stock returns”.” *International Review of Financial Analysis.*

Boudt, Kris, David Ardia, and Keven Bluteau. 2021. “Media abnormal tone, earnings announcements, and the stock market.” *Journal of Financial Markets.*

CBOE. 2019. “White Paper Cboe Volatility Index.”

Cen, and Liyan-Yang. 2013. “Investor sentiment, disagreement, and the breadth return relationship.” 1076-1091.

Chen. 2017. “Investor attention and global stock returns.” *Journal of Behavioral Finance.*

Chen, James. 2020. *Investopedia.* 25 December. https://www.investopedia.com/terms/d/dow-30.asp.

Cheng, and et al. 2019. “Investor attention and stock price movement.” *Journal of Behavioural Finance.*

Cowles. 1938. “Common-Stock Indexes.” *Principia Press.*

Da, and et al. . 2015. “The sum of all FEARS investor sentiment and asset prices.” *Review of Financial Studies,.*

Da, and et al. 2011. “In search of attention.” *Journal of Finance.*

Dimpfl, and Jank. 2015. “Can internet search queries help to predict stock market volatility?” *European Financial Management.*

Ding, and Hou. 2015. “Retail investor attention and stock liquidity.” *Journal of International Financial Markets, Institutions and Money.*

Fischer. 1925. “Stocks vs Bonds .” *American Review of Reviews.*

Fisher. 1922. “"The Making of Index Numbers - A Study of Their Varieties, Tests, and Reliability".” *Riverside Press.*

Garcia. 2013. “Sentiment during recessions.” *Journal of Finance.*

Glambosky, Mina, Seungho Baek, and Sunil Mohanty. 2020. “COVID-19 and stock market volatility: An industry level analysis.” *Elsevier.*

Goddard, and et al. 2015. “Investor attention and FX Market volatility.” *Journal of International Financial Markets, Institutions and Money,.*

Gormsen, and Koijen. 2020. “Coronavirus: Impact on stock prices and growth expectation.” *SSRN Electronic Journal.*

Gormsen, and Koijen. 2020. “Coronavirus: Impact on stock prices and growth expectations.” *SSRN Electronic Journal.*

Haim Levy, Guy Kaplanski. 2010. “Sentiment and stock prices: The case of aviation disaster.” *Elsevier.*

Haroon, and Rizvi. 2020. “COVID-19: media coverage and financial markets behaviour – a sectorial enquiry.” *Journal of Behavioral and Experimental Finance.*

Hautcoeur. 2006. “Why and how to measure stock market fluctuations? The early history of stock market indices, with special reference to the French case.” *halshs-00590522.*

Heyman, and et al. 2019. "“Investor attention and short-term return reversals”." *Finance Research Letters.*

Himmelmann , and Schiereck. 2012. “Drug approval decisions: A note on stock liquidity effects.” *Journal of Empirical Finance.*

Hirshleifer. 2015. “Behavioral finance.” *Annual Review of Financial Economics.*

Horvath, and Huizinga. 2015. “Does the European Financial Stability Facility bail out sovereigns or banks? An event study.” *Journal of Money, Credit and Banking.*

Huberman, and Regev. 2001. “Contagious speculation and a cure for cancer: A non-event that made stock prices soar.” *Journal of Finance.*

Jackson. 1928. “Common and Preferred Stocks as Investments.” *The Journal of Business of the University of Chicago.*

Kahnemann. 1973. “Attention and Effort.” *Prentice-Hall, Englewood Cliffs.*

Kenton, Will. 2022. *Investopedia.* 2020 February. https://www.investopedia.com/terms/s/sp500.asp.

Kim, and et al. n.d. “Google searches and stock market activity: evidence from Norway.” *Finance Research Letters2019.*

Krause , Miguel, Daniel Neukirchen, and Nils Engelhardt. 2022. “Firm efficiency and stock returns during the COVID-19 crisis.” *Elsevier.*

Lim, and Teoh. 2010. ““Limited attention”, in Baker, H.K. and Nofsinger, J.R. (Eds).” *Behavioural Finance.*

Liu, Peng, Tong Fang , and Zhi Su. 2021. “Pandemic-induced fear and stock market returns: evidence from China.” *Elsevier.*

Markowitz. 2005. “Market Efficiency: A Theoretical Distinction and So What?” *Financial Analyst Journal.*

Norli, Edmans, and Garcia. 2007. “Sports sentiment and stock returns.” *Journal of Finance.*

Odean. 1999. ““Do investors trade too much?”.” *American Economic Review.*

Onali, Enrico. 2020. “Covid-19 and stock market volatility.”

Pellizzon , Loriana, Michele Costola , and Micheal Nofer. 2020. “Machine learning sentiment analysis, Covid-19 news and stock market reactions.” *Econstor.*

Peng, and Xiong. 2020. “Investor attention, overconfidence and category learning.” *Journal of financial economics.*

Preis, and et al. 2010. “Complex dynamics of our economic life on different scales: insights from search engine query data.” *Philosophical transactions of The Royal Society* .

Ramelli, and Wagner. 2020. ““Feverish stock price reactions to COVID-19”.” *Review of Corporate Finance Study.*

Rappuoli, Masignani, and Lattanzi. 2003. “The value of vaccines. .” *Vaccine 21, Supplement 2.*

Riedel, Max, Micheal Donadelli, and Renatas Kizys. 2016. “Globally Dangerous Diseases: Bad News for Main Street, Good News for Wall Street.” *House of Finance.*

Ru, and el al. 2020. “What do we learn from SARS-CoV-1 to SARS-CoV-2: evidence from global stock markets.” *Working Paper.*

Santagiustina , Carlo, Michele Costola, and Matteo Iacopini. 2020. “Public concern and the Financial Market during the COVID-19 outbreak.”

Saunders. 1993. “ Stock prices and Wall Street weather.” *The Journal of Finance.*

Schell, and et al. 2020. “This time is indeed different: a study on global market reactions to public health crisis.” *Journal of Behavioral and Experimental Finance.*

Shumway, Hirshleifer. 2003. “Good day sunshine: Stock returns and the weather.” *Journal of Finance.*

Smales, Lee A. 2021. “Investor attention and the response of the US stock market sectors to the COVID-19 crisis.” In *Review of Behavioural Finance* . Emerald Publishing Limited.

Smith. 2012. “Google internet search activity and volatility prediction in the market for foreign currency.” *Finance Research Letters.*

Smith, and Ellsworth. 1985. “Patterns of cognitive appraisal in emotion.” *Journal of Personality and Social Psychology,.*

Takeda, and Wakao. 2014. “Google search intensity and its relationship with returns and trading volume of Japanese stocks.” *Pacific-Basin Finance Journal.*

Tang, and Zhu. 2017. “How security prices respond to a surge in investor attention: evidence from Google search of ADRs.” *Global Finance Journal,.*

Tantaopas, and et al. 2016. “Attention effect via internet search intensity in Asia-Pacific stock markets.” *Pacific-Basin Finance Journal.*

Theodossiou, A., and P. Theodossiou. 2014. “Stock return outliers and beta estimation: The case of U.S. pharmaceutical companies.” *Journal of International Financial Markets, Institutions and Money.*

Unknown. 2023. *Wikipedia.* 9 February. https://en.wikipedia.org/wiki/Dow_Jones_Industrial_Average.

Veld , Chris, Guy Kaplansky, and Haim Levy. 2015. “Do happy people make optimistic investors?” *Cambridge University Press.*

Vlastakis, and Markellos. 2012. “Information demand and stock market volatility.” *Journal of Banking and Finance.*

Vozlyublennaia. 2014. “Investor attention, index performance, and return predictability.” *Journal of Banking and Finance.*

Waldmann, De Long, and Bradford. 1990. “Noise trader risk in financial markets.” *Journal of Political Economy.*

Welch, Qiu. 2004. “Investor sentiment measures.” *NBER Working Paper No. 10794.*

Wilson, and Jones. 1987. *An Incomplete Education.* Ballantine Books.

Wurgler, Baker. 2006. “Investor sentiment and the cross-section of stock returns.” *The Journal of Finance.*

Wurgler, Baker. 2007. “Investor sentiment in the stock market.” *Journal of Economic Perspectives.*

Yilmazkuday. 2020. “Covid-19 effects on the s&p 500 index.” *SSRN Electronic Journal.*

Young, Julie. 2022. *Investopedia.* 28 April. https://www.investopedia.com/terms/m/marketindex.asp.

Zhang, and el al. 2020. “Financial markets under the global pandemic of COVID-19.” *Finance Research Letters.*

Zhang, Palomino Renneboog. 2009. “Information salience, investor sentiment, and stock returns:.” *Journal of Corporate Finance* .

Zhu, Yuan Zheng. 2006. “Are investors moonstruck? lunar phases and stock returns.” *Journal of Empirical Finance.*

ABSTRACT

A method that analyses text data in order to discover the underlying emotion or mood of the text is known as sentiment analysis, and it has recently garnered a lot of popularity as a means of doing so. In recent years, sentiment analysis has been used to a wide number of areas, one of which is the financial industry, where it is utilised to forecast the returns of the stock market. In light of the COVID-19 outbreak, it is even more important to have a solid understanding of the relationship that exists between mood and the performance of the stock market. The purpose of this study is to investigate the connection that exists between investor sentiment and the performance of the stock market during the COVID-19 epidemic. In specifically, we study the association between COVID-19 instances, terms of stress and anxiety in WSJ news, negative tone in WSJ news, and the returns of the S&P 500, Dow Jones, and VIX. The S&P 500 and Dow Jones are two of the most commonly identified and watched stock market indexes in the world, whilst the VIX monitors the S&P 500's volatility. We have a hypothesis that the independent variables (COVID-19 cases, terms of stress and anxiety in WSJ news, and negative tone in WSJ news) will have a negative association with the returns of the S&P 500 and Dow Jones. This hypothesis is supported by the data. This is based on the theory that investors who are exposed to negative emotions in the news are more likely to have pessimistic expectations on the future of the stock market, which in turn leads to lower returns on investments. Our expectation is that there will be a positive correlation between the independent variables and the VIX. This is as a result of the fact that a spike in pessimism and uncertainty in the news would increase market volatility, as shown by the VIX. In order to validate these hypotheses, we will conduct a sentiment analysis on all of the news articles that were published in the Wall Street Journal (WSJ) between March 2020 and March 2021. Using multivariate and univariate regression models, we have investigated the relationship that exists between unrelated factors and the returns on the stock markets of the S&P 500, Dow Jones, and VIX. Several areas, including social media, customer feedback, and news stories, have utilised sentiment analysis to analyse textual data. Research on people's feelings has been applied to the study of finance in order to estimate stock market returns and evaluate the impact that sentiment has on financial markets. Previous studies have shown that the tone of news articles can have an effect on the results of the stock market. It has been established, in the context of the COVID-19 outbreak, that the tone of news articles is a strong predictor of returns on the stock market. In this analysis, the link between COVID-19 cases, the sentiment of news articles, and the returns of the S&P 500, Dow Jones, and VIX stock market indexes during the pandemic are investigated. We analyse the

effect of specific categories of emotion, such as stress and anxiety, in addition to the effect of general sentiment as a whole.

I carried out an in-depth investigation of the link between the performance of the stock market and the number of COVID-19 cases and news headlines during the first year of the outbreak. This investigation was carried out during the first year of the epidemic. This was a difficult moment for humanity, as the epidemic had a significant effect on every facet of our existence. As a result, we faced many challenges. As a direct consequence of this, we encountered a great deal of resistance. I discovered that there was a substantial correlation between the overall mood and emotion of the population as a whole and the performance of the stock market while I was going through the data and organising it. On the other hand, the effects were delayed because of the nature of the variables. The markets are not influenced by individual news items; rather, it is the persistence of the news that has an effect on the markets. In the early phases of the pandemic, there was a distinct pattern of sudden expressions of fear and uncertainty regarding health and the economy. These displays followed a definite pattern. As a direct consequence of this, the markets went through a prolonged period of bearish conditions and moved into a region with declining returns. As I dove deeper into the data, I saw that the effects of COVID-19-related news and the number of instances on the financial markets were long-lasting. This was something that struck me as particularly interesting. As they did become obvious, the benefits continued to materialise in a reliable manner even after a significant amount of time had passed. As a result of carrying out this investigation, I am now aware of a number of potential shortcomings as well as areas where improvements could be made. For instance, a new group of factors may be taken into consideration, or a model that is even more complicated, like the SIR model, which was developed specifically for pandemics, may be utilised instead. In addition, dummy variables could be incorporated into the research design in order to acquire a more in-depth comprehension of the data. Yet, I am of the opinion that the study that I carried out with the variables and models that I used is the best that I could have done in order to offer a solid and consistent account of one of the primary factors that influenced the stock market during the pandemic. I reached this choice because I believe that the analysis I undertook was the most thorough conceivable. As I combed over the data, I noted the emergence of a certain pattern. Even if the impacts of the revelation concerning COVID-19 and the number of cases didn't immediately manifest themselves in the financial markets, those announcements had a substantial impact on the markets nonetheless. One piece of news is unlikely to have a significant impact on the markets; rather, the markets are driven by recurring patterns and trends that develop over time. As a consequence of this, major market damage did not become apparent until after the outbreak had already been going on for a considerable amount of time. When we looked at the numbers, it was obvious that the general sentiment and emotions of the populace had a huge impact on the stock markets. This became clear when we analysed the data. This was particularly true during instances of great volatility. The markets incurred losses anytime there was a global climate of uncertainty and fear. On the other hand, whenever there was a sense of optimism and hope in the air, the market began to recover. This was made abundantly clear in the instance of vaccine announcements, which frequently produced a favourable effect on the stock markets. In spite of the fact that the study had certain shortcomings, I think the findings are sound and they offer some very helpful insights into the connection between COVID-19-related news and the number of cases as well as the returns on the stock market. The research has the potential to assist investors in better comprehending the impact that pandemics have on the economy and in taking the necessary efforts to limit the repercussions of these pandemics by highlighting the delayed and long-lasting consequences of these variables. The findings of the study indicate that the connection between COVID-19-related news stories, the number of occurrences, and the performance of the stock market is a complicated one that is impacted by a variety of different elements. The effects will be significant and long-lasting once they take hold, notwithstanding the possibility that they may be delayed. This is because of the potential for the delay. As a consequence of this, it is essential to pay close attention to trends and patterns throughout the course of time and to take the appropriate actions in order to limit the economic effects of pandemics.