sciendo

The Role of Climate Change News in Pharma Industry

Bogdan Ionut ANGHEL *
Bucharest University of Economic Studies, Bucharest, Romania
**Corresponding author, anghelbogdan17@stud.ase.ro*

Radu LUPU
Bucharest University of Economic Studies, Bucharest, Romania,
Romanian Academy, Bucharest, Romania[1]
radu.lupu@rei.ase.ro

Abstract. *In this article, we aim to study the role of climate change news in the European Big-Pharma companies returns. The goal of the research is to explain if the sentiment of the news posted on Twitter can explain the market movements, even if current global events may affect stock prices. The previous literature reveals the impact of climate change in the economy, but in the field of pharma companies, there is less information of how climate change affects the evolution in the stock market. We used the VADER methodology to examine and analyze the climate change news posted on Twitter. This allows us to extract the sentiment and use it to study the dependence between sentiment score and the evolution of the European Big-Pharma companies. To examine the relation, we use Ordinary Least Squares regressions and Quantile Regression. Our results reveal the fact that there is no significant dependency between the sentiment score of the news published on Twitter by the biggest newspapers and the price increase of the European Pharma companies stocks. We observe a small increase in dependence when we use the sentiment score with 1 day lag in the Ordinary Least Squares regressions and another improvement in results with Quantile Regression.*

Keywords: Climate Change, Sentiment Analysis, Stock Market, Big Pharma, Twitter

Introduction

In recent years, the focus of society on the environment and the effects of climate change has been a priority for the European Union especially. Also, the context for this study, with legislative changes in CO2 reduction, the COVID-19 pandemic and the crisis as a result of the Russian – Ukrainian war, affects stability and generates problems in the stock market.

Twitter became a good news source for the general population in the field of the stock market and political news. Previous research was made with Twitter posts as a database, which were analyzed with different Machine Learning models and used for different predictions and analysis. Many newspapers use Twitter and other social media platforms to share their news and engage people to discuss various topics.

The impact of climate change is expected to become more pronounced in the coming years, if the governments, companies and population will not take measures to stop and reduce pollution and the effects of it. Some of the implications of climate change are: rising global temperatures, rising sea levels, health impacts, economic impacts, political impacts, change the precipitation level, changes in the animals lifestyle.

In the context of the European Commission decision to adopt the European Green Deal, which promises to transform the EU into a modern, resource-efficient and competitive economy

[1] *Institute for Economic Forecasting*

DOI: 10.2478/picbe-2023-0004
© 2023 B. I. Anghel; R. Lupu, published by Sciendo.
This work is licensed under the Creative Commons Attribution 4.0 License.

by reducing to zero the net emissions of greenhouse gasses by 2050, we choose to focus our quantitative study on the European market.

Climate change can generate different risks for the pharma industry, which can influence the productivity and the profit of the companies. That fact can also be seen in the evolution of the value of the companies on the stock markets. Pharmaceutical companies are taking steps to mitigate the climate change risks, such as reducing the pollution in the manufacturing process, investing in renewable energy and improving energy efficiency, developing a climate-resilient production process.

The role of Machine Learning in this type of research is very high. The sentiment extraction from the database is done using the Supervised Learning Model and the outputs for our sentiment analysis classification are: negative, neutral or positive. Sentiment Analysis is an application of Natural Language Processing which automatically extracts and analyzes the database using Machine Learning Algorithms.

The economic context needs a careful selection of accurate and complex data from multiple sources. Therefore, we can generate many research questions, which require a clear approach to solve, using various quantitative methods.

We aim to adapt previous studies for sentiment analysis in the field of climate change, a topic with high importance nowadays. A medium-sized database will be extracted and structured using algorithms developed with Python libraries. The goal is to conduct a quantitative analysis for studying the correlation between news and the market movement. In the last years, the public attention was focused on the pharma sector, due to Covid-19 Pandemic. The period in which the effect of the sentiment could affect the European Big-Pharma Index may be another research question.

In this paper, we focus on the role of climate change news in the European Big-Pharma companies' returns. The goal of the research is to explain if the sentiment of the news posted on Twitter can explain the market movements, even if current global events may affect the stock prices.

Literature review

In the last few years, we can underline different changes in key trends of the Big-Pharma Industry. Gautam & Pan (2016) did an impressive analysis for the 1995-2015 period, but can't predict the problem of the Pharma Industry with climate change policies and social trends. (Belkhir & Elmeligi, 2019) found that Big-Pharma Industry is a higher intensity CO2 emitter, for example: pollution coefficients are higher than others non-climate-friendly industries like automotive. For the European Big-Pharma Industry, policies that want to reach a net-zero emission balance (Paris Agreement in 2015 and the European Climate Law with the European Green Deal) affects the value of the stocks and the growth rate. At this moment, the effects of the climate policy were disturbed by the Covid-19 crisis and the need to increase the production of a big variety of medical products. The high demand for medical products, the competition and the political decisions are the main problems that Big-Pharma companies are facing in this period. The European Pharma Industry may face problems in their business plans due to the involvement of the European Union policy for the reduction of CO2 emissions.

Climate change is a major discussion topic for most online news publications. The role and the power of the fact of online publications were demonstrated and explained in a lot of articles. Choi & Varian (2009) showed how these trends help in the prediction of the future stock prices.

It's possible to build a sophisticated forecasting model that can predict future trends in the stock market based on the sentiment of the news published on different social media platforms.

Haunschild et al. (2016) explained that climate change is caused by multiple natural factors, but in recent years, human activities (burning fossil fuel and pollution) are the main factor for increasing climate warming. This aspect was published also in an IPCC report, where they explain how climate change is the effect of the actions made by the population.

Li et al. (2011) used the keyword "Climate change" for exhaustive research on the topic of global climate change. The study can't give a unique definition for this phenomenon, it suggests that change in meaning year by year. Iii et al. (2019) contributed to the field of understanding climate change news. Also, they explained the implication in the stock market and developed a measure for the climate risk of a company.

Not only productivity, politics factors and the main shareholders can affect the market value of the Pharma Companies. The correlation between emotions via Twitter posts and stock price changes has been studied widely in the recent past. Bollen et al. (2011) investigated and showed how changes in public mood can be tracked and responded to a variety of social-cultural questions. Recent research suggests that news could be unpredictable, but various information can be extracted from social media content. Also, the possibility to study the sentiment in climate change news can underline the importance of news in the stock market movement.

Methodology

In this research, as input data, we employ daily logarithmic returns for 12 European Pharma Companies for the 1st January 2013 – 31st December 2022 time frame. The following tickers were used for the study: "BAYZF" (Bayer Aktiengesellschaft), "GMAB" (Genmab), "ROW.SW" (Roche Holding AG), "GSK" (GSK PLC), "VIFN.SW" (Vifor Pharma AG), "MRK" (Merk & Co), "JAZZ" (Jazz Pharmaceuticals PLC), "SNY" (Sanofi), "HZNP" (Horizon Therapeutics Public Limited Company), "LONN.SW" (Lonza Group AG), "NVO" (Novo Nordisk A/S), "AZN" (AstraZeneca PLC). In Table 1 is presented the statistical properties of logharitmical returns and information about the sample size for each selected stock in the study. The most valuable stock is Jazz Pharmaceuticals PLC, also with the biggest growth from the group, 1669,5% in 10 years with not the best climate.

Table 1. Descriptive statistics of the 12 European Pharma Companies included in the Pharma Index

Stock Ticker	BAYZF	GMAB	ROG.SW	GSK	VIFN.SW	MRK	JAZZ	SNY	HZNP	LONN.SW	NVO	AZN
count	2518.000000	2518.000000	2512.000000	2518.000000	2508.000000	2518.000000	2518.000000	2518.000000	2518.000000	2512.000000	2518.000000	2518.000000
mean	97.813165	19.583393	275.892277	42.461465	123.605482	64.975654	139.045763	47.137192	35.478820	296.578948	58.598906	40.568004
std	31.454228	12.593834	42.160937	5.128511	33.474698	14.504681	28.626902	4.707129	32.387994	203.042628	23.161299	12.073947
min	46.060001	1.400000	189.199997	28.660000	53.049999	38.979008	52.790001	37.060001	1.990000	46.962879	30.110001	22.334999
25%	66.339996	8.927500	244.000000	39.450001	97.537502	53.907443	127.272497	43.112500	14.052500	121.853395	44.342500	32.093750
50%	102.290001	17.869500	265.799988	41.434999	125.924999	60.376907	142.904999	48.105000	19.990000	254.799995	51.014999	36.275000
75%	127.260000	30.022501	308.125008	44.770000	148.000000	76.870003	157.045002	51.027499	55.782499	477.675003	65.367498	50.439999
max	155.850006	48.720001	400.549988	56.660000	192.800003	112.120003	193.020004	57.459999	119.910004	784.599976	135.339996	71.139999

The variation in sample size is the consequence of the bank holidays that are different in each European stock exchange. We replaced the NaNs values with the value from the next row using "bfill" method. We created an index with the mean of the previous stocks that we used in the

following study. In Figure 1 is represented graphically the evolution of the European Pharma Market from 2013 to the end of 2022.



Figure 1. Evolution of the European Big-Pharma Index

For the representative data of the Climate Change News, we were interested in collecting information via Twitter, from the most followed 40 online newspapers accounts. The list of the Twitter accounts used for data extraction is attached in the Appendix – Table 2. To collect the database, we used a scraping Python method, which sends requests to the Twitter server to provide all the posts that contain the selected keyword "Climate Change". For this time interval, we found 34.679 Tweets that contain "Climate Change"'. Using Python, we cleaned up the posts: remove the URL, replace the hashtags, emoticons and lowercase the entire text. On average, each Tweet in the database contains 25.64 words per post.

For this analysis we used Python, with libraries like: pandas, numpy, yfinance, snscrape, nltk. Our approach to automatic analysis of the relation between climate change news and the European Big-Pharma companies' returns.

Hutto & Gilbert (2014) presented VADER (Valance Aware Dictionary for sEntiment Reasoning), a simple rule based model for general sentiment analysis. The reason why we choose to use this method in our Python code, is that the algorithm is working well on social media style text, it is fast and doesn't require a training data set. The tool combines lexical futures with five general rules that embody grammatical and syntactical conventions that humans use when expressing on emphasizing sentiment intensity. In their analysis, Hutto & Gilbert (2014) showed the accuracy of the VADER method performs better and it's more efficient in analyzation of the sentiments in Tweets with 0.96% accuracy and the following method (NB) achieved a 0.84% accuracy. For the quantitative analysis, we used the compound score, as a metric that calculates the sum of all the lexicon ratings which have been normalized between -1 and 1. The positive sentiment score is between [0.05, 1], the neutral sentiment score is inside (-0.05, 0.05) interval and the negative sentiment score is [-1, -0.05]. Figure 2 shows the distribution of the Tweets' sentiment.

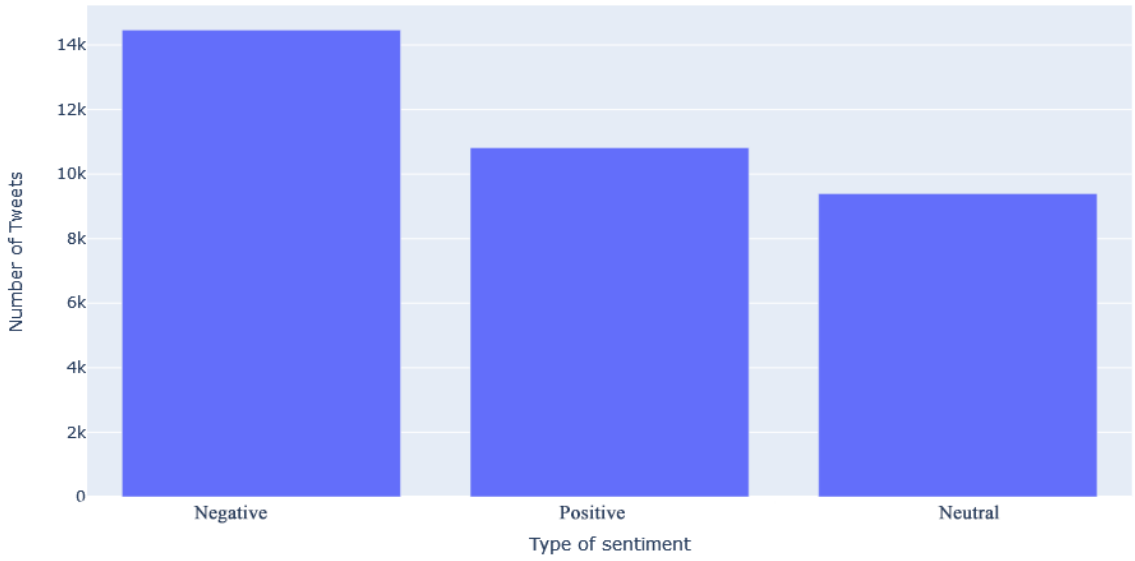


Figure 2. Classification of the sentiment-score analysis of the Tweets database

A Quantile is a cut point, or line of division, that splits a probability distribution into continuous intervals with equal probabilities. Most applied statistics view Quantile Regression as an elaboration of the linear regression model and associated estimation method of least squares. Quantile Regression is used to explain in a richer way the relationship between independent and dependent variables that are distant from the mean (Mosteller & Tukey, 1977). Koenker & Bassett (1978) developed a quantile regression methodology. They aim to extend the ideas to the estimation of conditional quantile functions, models in which quantiles of the conditional distribution of the response variable are expressed as functions of observed covariates. In their paper, Koenker & Hallock (2001) gives an exhaustive overview of quantile regression and its application of it in different situations. Also, a good explanation of the parameters and the results are presented in their paper.

The structure of the quantile regression model equation for the π (quantile regression formula is presented below).

$$Q_\tau(y_i) = \beta_0(\tau) + \beta_1(\tau)x_{i1} + \cdots + \beta_p(\tau)x_{ip} \quad i = 1, \ldots, n$$

The first step in our quantitative analysis was to use the Tweets database and apply the VADER method for outputting the sentiment analysis. After that, we clean up and normalize the sentiment score dataset.

The second step was to organize the tweets by the type of sentiment and present the distribution by interval in Figure 3.

The third step was preparing datasets for the quantitative analysis and running firstly Least Squares Regression with sentiment score as the dependent variable and Logarithmic Returns of the European Pharma Index as the independent variable. After that, we tried multiple lagged OLS regressions, with the dependent variable lagged by 1/2/5 days. The next step was to apply the Quantile Regression with sentiment score as the dependent variable and the European Pharma Index Price as the independent variable using (Koenker & Hallock, 2001) methodology.

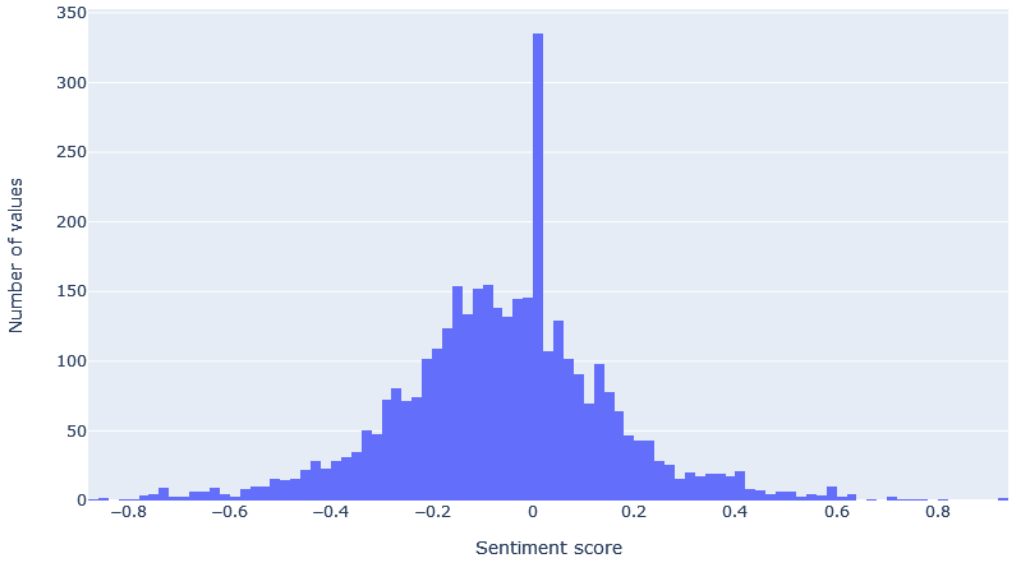


Figure 3. Distribution of the sentiment-score

Results and discussions

During the adoption of the Paris Agreement in 2015, it was a high volatility period. The measure for reducing CO2 affects the European Pharma Industry. In the first quarter of 2020, when the World Health Organization announced the beginning of the Covid-19 Pandemic, the Pharma Index slumped by 26,97%. After that, there was a growth in the pharmaceutical industry caused by the Covid-19 medicine demand in 2021, even if the public attention, authorities and media channels were focused on the reduction of carbon emissions.

We also note the database contains mostly negative sentiment elements (14465 tweets), positive posts (10819 tweets) and a number of 9396 neutral tweets. Figure 3 shows the distribution of the sentiment score. It can be seen a high number of negative sentiment tweets in the (-0.16, -0.06) interval. The largest interval is (0, 0.02) with 335 elements. We also notice the increase in the number of Twitter news posts by 40 publications. From 2015, year by year there was an upsurge in the number of news articles containing the keyword “Climate Change”. One of the main reasons was the adoption of the European Green Deal, which promises to transform the EU into a modern, resource-efficient and competitive economy by reducing to zero the net emissions of greenhouse gasses by 2050.

In total, there are 3651 observations used in all regression, for the period mentioned in the previous part (1st January 2013 – 31st December 2022). The main interesting observation for the Quantile Regression in comparison with the previous Ordinary Least Squares regressions is that the result improved, the R-squared is still at a very low level, the relationship between the news about climate change and the evolution of the Pharma Industry in Europe cannot be explained.

We estimate the quantile regression model for many quantiles between 0.05 and 0.95 and compare the best fit line from each of these models to Ordinary Least Squares results. In quantile regression results, the bandwidth value represents a parameter used to control the smoothness of the estimated quantile function. The bandwidth value of 0.05577 determines the "width" of the neighborhood of observations used to make local estimates of the quantile function. In this context, a bandwidth value of 0.05577 indicates a relatively narrow neighborhood of observations being

used for local quantile estimates, which may result in a more flexible but also potentially more noisy quantile estimate.

A sparsity value of 0.5031 indicates that approximately 50.31% of the data points have missing or zero values for the sentiment score. This could imply that there are significant gaps in the sentiment score data, or that many of the observations may not have a strong relationship with the European Pharma index. When fitting a quantile regression model to sparse data, it is important to consider how missing values are handled and to be cautious about drawing conclusions based on the estimated model. It may also be useful to consider additional data sources or imputation methods to fill in the missing values and improve the accuracy of the model.

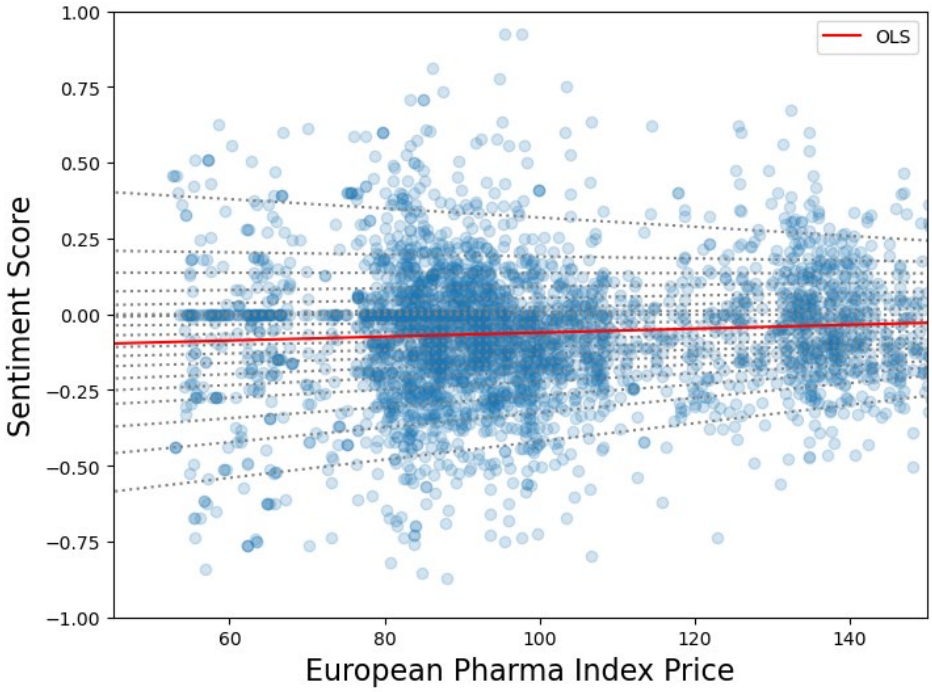


Figure 4. Comparison between the best fit lines for 20 quantile regression models to the least squares fit

We also note, as can be seen in Figure 4, the plot compares best fit lines for 20 quantile regression models to the least squares fit. The dispersion of the sentiment score decreased with the European Pharma Index Price increase. The sentiment score remains constant with the European Pharma Index price increasing.

In Figure 5 it can be seen the dotted black lines from a 95% pointwise confidence band around 10 quantile regression estimates (solid black line). The red ones represent Ordinary Least Squares regression results along with their 95% confidence interval. We also note that the quantile regression point estimates lie outside the OLS confidence interval, which suggests that the effect of the sentiment score on the European Pharma Index Price may not be constant across the distribution.

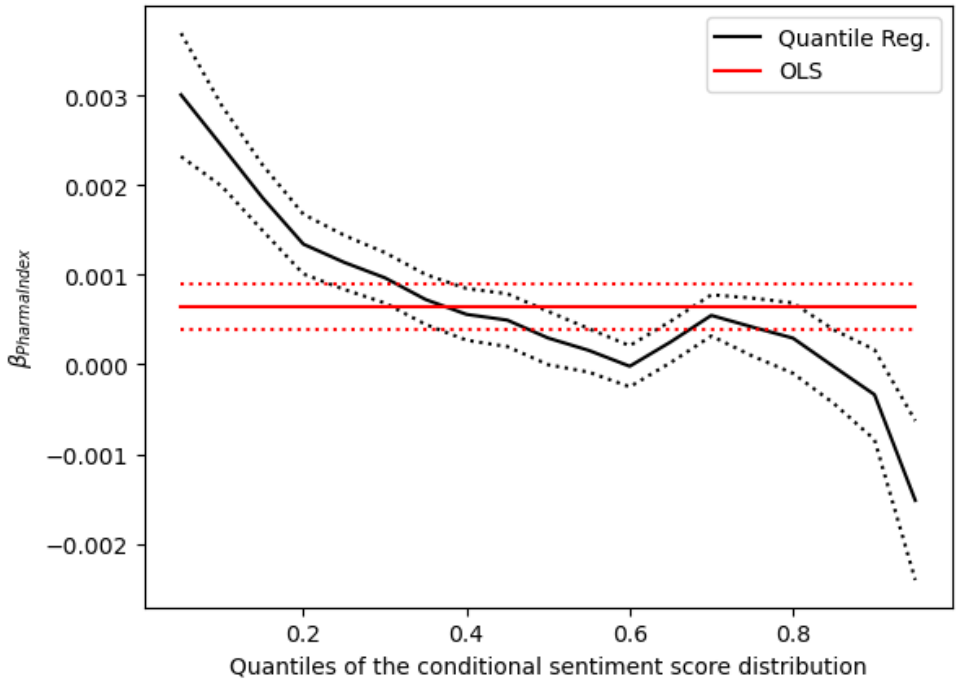


Figure 5. Quantile Regression Analysis of sentiment score

It can be seen an improvement by using the 1 day lag for the European Pharma Index. The consequences of the sentiment score variation, can be observed 1 day after the news posted on Twitter feed.

Conclusion

The recent literature in the field of climate change news, showed the importance of studying the sentiment of the news, which can help, using Machine Learning Tools, making predictions in different economic fields.

We selected, for our analysis, the most valuable 12 European Pharma companies by market capitalization. Also, the 40 online newspapers (see Appendix – Table 2) gave us a good starting point to create the database using scraping tools for extraction of Tweets with the keyword "Climate Change". The quantitative analysis was focused on a period spanning from 1st January 2013 to 31st December 2022. First of all, we created a Big-Pharma Index with the mean of the selected Pharma companies, we calculated the logarithmical return, then we solved the missing data and did the cleaning data process. After that, we extracted the dataset for this time interval and found 34.679 Tweets that contain the keyword. Using different Python algorithms, we cleaned up the Tweets: remove the URL, replace the hashtags, emoticons, other invalid characters and lowercase the entire text. We applied the VADER method (Hutto & Gilbert, 2014) on our Tweets database for outputting the sentiment analysis. After that, we cleaned up and normalized the sentiment score dataset. The Next step was preparing datasets for the quantitative analysis and run firstly Least Squares Regression with sentiment score as the dependent variable and Logarithmic Returns as the independent variable. After that, we tried multiple lagged Ordinary Least Squares regressions, with the dependent variable lagged by 1/2/5 days. Finally, we applied the Quantile

Regression with sentiment score as dependent variable and the European Pharma Index Price as the independent variable using (Koenker & Hallock, 2001) methodology.

The observation that one or more important results from the quantile regression point estimates lie outside the Ordinary Least Squares confidence interval is a significant finding. It suggests that the relationship between the sentiment score and the European Pharma Index Price may not be linear or constant across the entire distribution. This result highlights the importance of using a quantile regression approach, as it provides a more nuanced and detailed understanding of the relationship between the sentiment score and the European Pharma Index Price. In particular, the quantile regression approach allows us to examine the relationship between the sentiment score and the European Pharma Index Price at different points in the distribution, rather than just at the mean as in the traditional Ordinary Least Squares regression. This can provide valuable insights into the impact of the sentiment score on the European Pharma Index Price at different levels of the distribution.

The main contribution of our analysis is rooted in quantifying the climate change news and the dependence between the climate risk sentiment with the pharma industry, one of the biggest polluters at that moment. One of the impediments to our research was the period full of events that disrupted financial stability. For sure, due to the Covid-19 Pandemic, the European Pharma Industry failed to comply with long-term internal strategies, also it was hard to adapt to and respect the European Green Deal.

In future studies, we can use different methods for sentiment-score calculation, also we can collect a large database using more relevant keywords and use more news providers. In other terms, we can observe that year by year, public opinion is more focused on the climate change risk and searching for new ways to reduce the pollution. As we notice in the first part of the article, Big Pharma is one of the most polluting industries and it will be an interesting field to study the policies they will approach to combat the climate change risk, as well the impact of these on the stock market.

References

Belkhir, L., & Elmeligi, A. (2019). Carbon footprint of the global pharmaceutical industry and relative impact of its major players. *Journal of Cleaner Production*, *214*, pp. 185–194.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), pp. 1–8.

Choi, H., & Varian, H. (2009). Predicting the Present with Google Trends. *Economic Record, volume 88*

European Commission (2019). A European Green Deal *[Online] Retrieved January 22, 2023, from https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en*

Gautam, A., & Pan, X. (2016). The changing model of big pharma: Impact of key trends. *Drug Discovery Today*, *21*(3), pp. 379–384

Haunschild, R., Bornmann, L., & Marx, W. (2016). Climate change research in view of bibliometrics. *PLoS ONE*, pp. *11*(7)

Hutto, C. J., & Gilbert, E. (2014). *VADER:* A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text*, Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014*

Iii, R. F. E., Giglio, S., Kelly, B. T., Lee, H., & Stroebel, J. (2020). Hedging Climate Change News, The Review of Financial Studies, Volume 33, Issue 3, pp. 1182-1216
Koenker, R., & Bassett, G. (1978). Regression Quantiles. *Econometrica*, *46*(1), pp. 33–50.
Koenker, R., & Hallock, K. F. (2001). Quantile Regression. In *Journal of Economic Perspectives* Vol. 15, Issue 4, pp. 143- 156
Li, J., Wang, M. H., & Ho, Y. S. (2011). Trends in research on global climate change: A Science Citation Index Expanded-based analysis. *Global and Planetary Change*, volume *77*(1–2), pp. 13–20.
Mosteller, F., & Tukey, J. W. (1977). Data analysis and regression. A second course in statistics. *Dars*.
UNFCC International (2015). The Paris Agreement 2015*, [Online] Retrieved January 14, 2023, from https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement*

Appendix

Table 2. List of the online newspaper included in the sample for the climate change news database

Newspaper Name	Twitter Username	Newspaper Name	Twitter Username
The Guardian	guardian	**CBC NEWS**	CBCNews
The Daily Mail	MailOnline	**RT**	RT_com
The Daily Telegraph	Daily_Telegraph	**CNBC**	CNBC
The Independent	independent	**SKY News**	SkyNews
El Pais	el_pais	**BBC Breaking News**	BBCBreaking
Marca	marca	**BBC World**	BBCWorld
Financial Times	FinancialTimes	**CNN**	cnn
La Repubblica	repubblica	**CNN International**	cnni
The Sun	TheSun	**CNN Politics**	CNNPolitics
Le Monde	lemondefr	**The Washington Post**	washingtonpost
Daily Mirror	DailyMirror	**Newsweek**	newsweek
CNN Breaking News	cnnbrk	**France24**	France24_en
The New York Times	nytimes	**Politico**	politico
The Economist	TheEconomist	**Sky News Breaking**	SkyNewsBreak
The Wall Street Journal	WSJ	**Al Jazeera English**	AJEnghlish
Reuters	Reuters	**HuffPost**	huffpost
TIME	TIME	**China Xinhua News**	XHNews
NDTV	ndtv	**ABC News**	abcnews
The Associated Press	AP	**Al Jazeera Breaking News**	AJENews
The Telegraph	Telegraph	**Euronews**	euronews

Source: Authors' own research results.