INVESTIGATING THE EFFECT OF EARNINGS CALL SENTIMENT ON STOCK VOLATILITY

by

GEORGE DREEMER

University of Groningen

Faculty of Economics and Business

Pre-Msc Marketing

June 2022

TABLE OF CONTENTS

ABSTRACT

This paper investigates what effect earnings call sentiment has on stock volatility after the call. Our sample dataset consisted of 178 transcripts from 2016 to 2020 of 10 NASDAQ-listed firms. The cleaned text data was used as input for sentiment analysis with NLTK's VADER. We scraped stock prices, as well as the NASDAQ index price, 10 days before and after the call to calculate volatility and the mean volatility change after the earnings call. In 87% of cases, volatility increased after the call. Our multiple linear regression returned statistically insignificant results. Our study's theoretical implication is the importance of investigating more factors when it comes to finding out the effect of earning's call sentiment on volatility. Furthermore, managerially it shows that earnings calls are a decisive moment for volatility in the period after the call.

Keywords

Earnings conference calls · Sentiment analysis · Stock volatility

Seminar Supervisor

A. E. Tatar

Seminar Topic

Predict the outcome of Fund-Raising Event using Textual Data

INTRODUCTION

In this paper, we are investigating whether and how earnings call sentiments influence stock price after the meeting. We are interested in doing this research, as corporate disclosures, such as earnings calls, are an important piece in understanding a company's behaviour (Li, 2011). Additionally, previous research shows that quantitative information, such as financial statements, only gives us half of the picture, to complete our understanding we must also consider qualitative data (Arslan-Ayaydin et al. 2016). Out of all types of corporate disclosures, we are focusing on earnings conference calls, because they are a source of information coming directly from management, describing company performance and outlook (Brown et al. 2004; Matsumoto et al. 2011; Price et al. 2012). Furthermore, earnings calls are spoken, differing from other corporate disclosures that are usually in writing. They also include interaction between managers and call participants, such as analysts and investors. Both of these factors make earnings calls a unique opportunity for researchers to better understand managers' disclosure behaviour. Finally, it is established by researchers that markets react to earnings news (Beaver, 1968). Summing all of this up we can see that earnings calls play an essential role in building an understanding of a company and it is an important factor in understanding the market reactions.

We chose to investigate sentiment's effect on the stock volatility, as conference call tone conveys useful information to the stock market (Davis et al. 2015; Price et al. 2012). It is also important to note that managers may use sentiment to manipulate listeners, for example by being overly optimistic to cover for an impending downside for the company. This is demonstrated by an overwhelming body of research (Bozzolan et al. 2015; Cho et al. 2010; Huang et al. 2014; Davies et al. 2007; Patelli et al. 2014). In the case of manager manipulation during earnings calls, according to Hutton et al. (2009), opaque firms are more likely to experience a crash in price, than truthful firms. This brings us to the research question:

Research Question. What is the effect of earnings call sentiments on stock volatility?

This area of research has been a growing topic of interest in the financial research world and there are good reasons for that. On one side we have the increasing availability of unstructured textual data. Researchers have access to earnings call transcripts and audio, financial analyst reports, SEC filings and many more similarly rich textual datasets (Li, 2011).

Furthermore, advances in data mining, textual analysis and machine learning have greatly empowered researchers in their quest to understand such data (Core, 2001).

Our research focuses on earnings call transcripts from a dataset made available by Roozen et al (2021) through Data Verse. We split each transcript text into two parts, the Presentation, which we call PreQA, and the QA. We opted to split into these two parts specifically, because the presentation is a monologue of the executives, while the QA is an interactive portion of the call, where executives and analysts interact. To determine sentiment, we have used a lexicon-based sentiment algorithm called VADER through the NLTK library for Python. This algorithm, among other values, returns the compound sentiment score between -1 and 1. We use this value to create the two independent variables - PreQA C-Score and QA C-Score.

For our dependent variable, stock volatility change, we opted to scrape the stock prices of the respective companies through the Yahoo Finance API for Python. We used the methodology of Doran et al. (2010) to determine that 10 days before and after the earnings call date is an optimal period for the measurement. To calculate volatility from stock prices we created a formula similar to the one presented by Hayes (2021). After calculating daily volatility 10 days before and after the call, we took the mean of the values and subtracted the two means to get volatility change (Volatility Delta). We chose volatility change to define stock price success or failure because changes in risk appetite are an important determinant of asset prices (Bekaert et al. 2014). Furthermore, we used the same methods to create our control variable - NASDAQ Volatility Delta. It follows the same philosophy as our DV, but it scrapes the price of the NASDAQ Composite Index. We chose this as our control because all of the companies in our dataset are part of the NASDAQ and we aim to rule out any broader market volatility changes in our analysis. Finally, we created a multiple linear regression model with Scikit and Statsmodels Python libraries to determine the direction and strength of the relationship between our variables.

In the remainder of the paper, we will firstly cover the literature review, our hypothesis and conceptual model. Second, we talk about our sample dataset, data processing, sentiment analysis and volatility calculations. Next, we present our results and finally our conclusion.

LITERATURE REVIEW

Earnings Calls & Volatility

Our research investigates the relationship between earnings call sentiments and the subsequent success or failure of a stock, which we determine through the volatility change.

Earnings Call Structure. Generally, earning calls (ECs) are a form of verbal corporate disclosure that takes place every quarter on a conference call between corporate participants (e.g. CEO, CFO and VPs), analysts and investors. The structure of such a call includes three standard components: (1) a safe harbour statement, which warns participants that financial results may include forward-looking statements, this has the function of limiting a company's liability in the case that these statements turn out to not be true in the future; (2) after the present corporate participants each give their statements that they have prepared; (3) finally there is a Q&A session when analysts and other non-corporate participants can ask questions and interact with the corporate participants (Corporate Finance Institute, n.d.).

Earnings Calls' Market Influence. Earnings calls are an important method for communicating corporate disclosures, and informative to market participants, additionally, they trigger heightened trading and stock price responses (Bushee et al. 2003; Frankel et al. 1999). It is considered that ECs decrease information disparities between firms and investors (Brown et al. 2004; Matsumoto et al. 2011) and provide useful insights to market participants aiding them in their buy-sell decisions (Hobson et al. 2017; Jung et al. 2016). Finally, ECs have the potential to shed light on more information than written financial documents, as they facilitate the revelation of corporate participants' discretionary incentives and behaviour (Bloomfield 2008; Davies et al. 2007).

Earnings Call Attributes. Past research on ECs has already unearthed various attributes which affect and inform market participants who are listening to the call. The attributes range from aspects of the call itself (e.g. duration) to behavioural aspects such as the tone of the call. In regards to duration, Matsumoto et al. (2011) found that a longer call translates to a higher content of information and is more insightful to the market. Other research instead looked at spontaneity in the language of corporate participants during the call. They concluded that the market responds negatively whenever the pre-Q&A portion of the call is similar to the Q&A one because it indicates pre-scripted answers to analysts' questions (Lee, 2016). We also explored the research of Burgoon et al. (2016). They found that when managers have intentionally made a financial misstatement, their conference call language will reveal the truth. Finally, other researchers have found that the tone or sentiment of ECs could reflect the company's history or future outlook based on the optimism of the disclosure (Henry et al. 2016; Loughran et al. 2011). Some researchers, such as Davis et al (2015), have measured positive or negative words to determine sentiment, which is also the basis for our study.

Volatility. Stock volatility is a statistical measure which represents the extent to which asset prices deviate from the mean price (Hayes, 2021). To measure stock volatility we use the respective company's stock price scraped from Yahoo Finance, which we use to calculate volatility. Volatility is risk-neutral expected stock price variance computed from a panel of options prices (Bekaert et al. 2014). These are also known as fear indices (Whaley, 2000). They are forward-looking by nature, measuring the volatility market participants expect to see in the next 30 days (Whaley, 2009). Considering all of this we relate the success or failure of the stock based on the change in volatility after the call compared to it before. Volatility increase symbolises market participants changing their minds by buying and selling and deviating stock price around the mean more intensely, thereby increasing volatility.

Hypothesis Development

We distinguish between two types of earnings call sentiments: (1) sentiment of the presentation given by the corporate participants, which takes place before the Q&A session and (2) sentiment during the Q&A session, when analysts ask questions and interact with the corporate participants (Roozen et al 2021). These two distinct parts are used to create two independent variables, which we refer to as Pre-Q&A C-Score and Q&A C-Score in the model. We chose to separate the EC into these two portions specifically because they are inherently different. Firstly, since the pre-Q&A statements are preplanned, they are often edited by corporate lawyers to reduce litigation risk, for example by subduing positive statements. On the other hand, the Q&A itself is more spontaneous and analysts get to steer the narrative through their questions (Bochkay et al. 2020). Our dependent variable is volatility change which we create by calculating the mean volatility for 10 days before and after the earnings call (Doran et al. 2012). We then subtract them to get our dependent variable - Volatility Delta. We chose volatility to measure the success or failure of a stock, as by nature volatility is a forward-looking index (Whaley, 2008). This matches up well with the content of ECs because they also contain forward-looking statements (Corporate Finance Institute, n.d.).

To develop our two hypotheses, we firstly considered agency theory, according to which managers have a direct interest in pushing positive news disproportionately to negative news (Kim et al. 2016). Furthermore, positive language can reflect corporate participants' distortion of the truth (Yuthas et al. 2002). Finally, other researchers found that an optimistic tone can be used to hide bad news and overly inflate investors' company perception by overly emphasising good news (Cho et al. 2010). Therefore, considering the aforementioned and that the pre-Q&A portion is most planned (Bochkay et al. 2020), we expect that positive sentiment in this section might cause distrust and uncertainty in market participants, thereby increasing volatility. This points to our first hypothesis:

Hypothesis 1. The sentiment of pre-Q&A will be positively associated with volatility - when sentiment is positive, volatility increases.

It is important to mention that although we hypothesise this, according to business ethics, the authenticity and accuracy of financial information are fundamental to financial reporting (Frecka, 2008). Furthermore, market participants value ethical and transparent corporate disclosures (Jo et al. 2008). Therefore managers may aim to gain a reputation for ethical and honest behaviour, making their statements truthful (Jones, 1995). Therefore, although we remain with our hypothesis, we believe that there is an argument for a negative association between the variables as well.

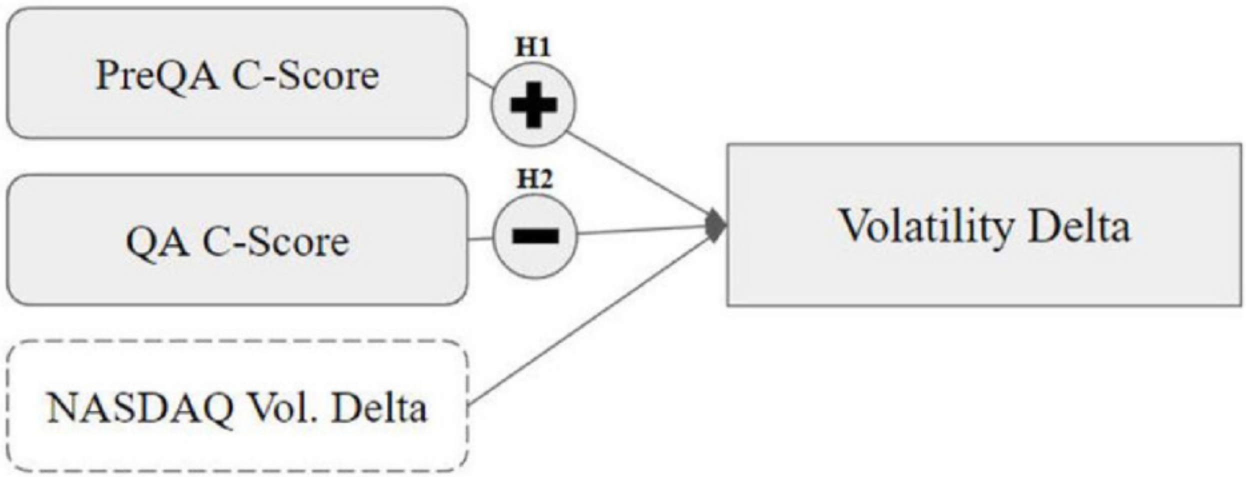
When we consider the spontaneous and interactive nature of the Q&A segment (Bochkay et al. 2020), as well as previously mentioned business ethics literature, we expect that the Q&A is more likely to present a more unfiltered take from the corporate participants. Furthermore, as the answers are relatively less scripted, managers get an opportunity to build their reputation of being truthful and disclose more of their vision (Jones, 1995). All of these factors emphasise the higher likelihood of the Q&A sentiment being more representative of reality. Therefore, we expect that a positive Q&A sentiment will negatively influence the volatility, bringing us to the second hypothesis:

Hypothesis 2. The sentiment of Q&A will be negatively associated with volatility - when sentiment is positive, volatility decreases.

We have created a conceptual model to illustrate the two relationships and have included our control variable in dashed line (see figure 1):

FIGURE 1:

Conceptual Model



METHODOLOGY

Our research question is about the effect of sentiment on stock volatility. However, the challenge lies in one word and that is the sentiment. By sentiment, we mean classifying whether a given piece of textual data is positive, negative or neutral (Monkey Learn, n.d.). We aim to understand the direction, strength and significance of the sentiment's effect on volatility. This section will cover the steps we will follow to answer our research question. We will address our sample, how we worked with the textual data to get to our independent variables, as well as our methods to reach the dependent and control variables through stock prices, finally, we will talk about our model.

Sample

Our sample dataset used to create our independent variables was provided by Roozen et al. (2021) through Data Verse. It consists of 188 earnings call transcripts of 10 NASDAQ-listed companies such as Apple, Intel and Amazon. The data was collected between 2016 to 2020 and has been made available in the form of uniformly structured text files. After an initial inspection of the files, the final sample was reduced to 178 files, due to 10 of the files containing unremovable artefacts that corrupted our preliminary data cleaning efforts. For our dependent and control variables, we obtained our data dynamically through the Yahoo Finance Python library. We scraped closing stock price values 10 days before and after the earnings call. These values were used for the formula in calculating the daily volatility, mean 10-day volatility and ultimately change in volatility after the call. We go more in-depth into each step in the following sections.

Text Data Cleaning & Sentiment Analysis

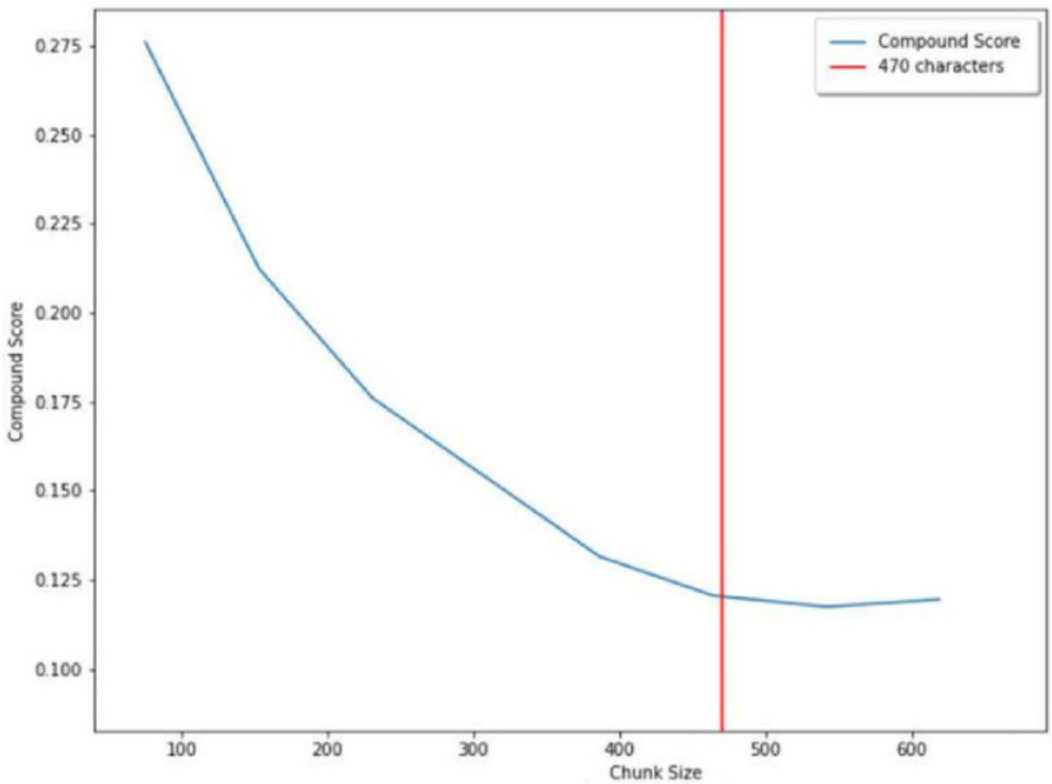
Cleaning. The first step toward the independent variables is to load the raw textual data and clean it. Cleaning refers to the process of preparing raw text for Natural Language Processing, the goal is to produce clean text suitable for a computer to analyse (Roldós, 2021). In the case of our raw text, we first split it into the pre-QA and QA, we do that by targeting titles indicating the beginning and end of these sections. Once split we face a multitude of unnecessary artefacts, which need to be removed. Examples of these are speaker titles, dashes, and new line symbols, among others. We clean these through our cleaning function, consisting of if statements targeting these issues. The final steps necessary before moving to the sentiment analysis are to normalise the text, tokenize it and remove stopwords. Normalising the text means removing capitalization (Roldós, 2021), while tokenizing is the process of dividing the text into word tokens, in the process also removing punctuation (NLP Stanford University, 2009). Finally, we remove stopwords, which are commonly used words in a language that do not add to sentiment analysis, such as "a" and "the" (Ganesan, 2019).

Sentiment Analysis. The end product of the aforementioned process is a long cleaned text consisting of upwards of 26000 characters or 3600 tokens, a length unsuitable for sentiment analysis. Generally, the maximum length for most analyzers is 512 tokens (Briggs, 2021). We are using VADER, which is a lexicon-based sentiment analyzer, meaning that it gives each word a sentiment rating based on its rating in the lexicon, which then builds the sentiment scores - positive, neutral, negative and compound (Hutto et al., 2014). The compound score lies between -1 and 1 signifying the intensity of the sentiment in either direction, while 0 signifies neutrality. This model is optimised for texts the size of social media posts (Mogyorosi, 2021). To meet this criterion with the dataset that we have, we divide our cleaned text into chunks with equal character lengths. We then run the analyzer on each chunk and then take the mean divided by the chunk size. We divide by the chunk size to reduce the mathematical bias of the chunk size within the equation. This gives us a total compound score for the pre-QA and the QA, which are our two independent variables.

Optimal Chunk Size. For us to determine the chunk size we plot the sentiment c-score for each respective chunk size up to 600. We look for the point where the change in compound score levels out, in other words where the model starts achieving consistent readings uninfluenced by chunk size. This is how we chose 470 characters for our chunk (see figure 2).

FIGURE 2:

Determining Optimal Chunk Size



Calculating Volatility

Stock Volatility. To calculate volatility we first obtain the daily stock closing price of the respective company 10 days before and 10 days after each earnings call. We then use this data in a formula that performs three operations: (1) first it takes a window of size 10 (days) and selects it to do mathematical computations on it, (2) we calculate the percent change between each consecutive value in the window, (3) we compute the standard deviation over the given axis. This method is similar to the one used by Hayes (2021). The formula results in a volatility value for each day in the window. Next, we compute the mean of the values in the windows before and after. We subtract these two values from each other to get the change in volatility - Volatility Delta, our DV. This variable signifies the change in volatility between the period before the earnings call and after it. Additionally, we create a dummy variable, called Volatility Increase, used only for descriptive statistics. It shows whether a given call resulted in a volatility increase, with 1 signifying an increase in the period after the call.

NASDAQ Volatility. As all of the companies in our dataset are listed on NASDAQ, it is important to see if the specific company stock was not simply influenced by an index-wide volatility change in that period. Therefore, we employ the aforementioned methods on scraped NASDAQ index price data in the same periods to create our control variable - NASDAQ Vol. Delta. It signifies the change in volatility for the index's price.

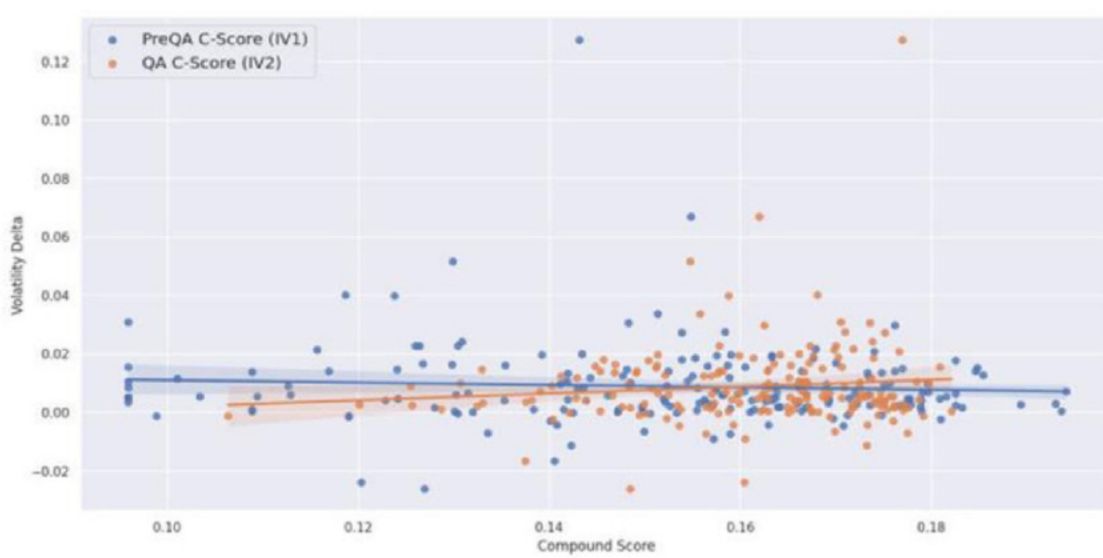
Model Specification

To investigate the relationship between earnings call sentiments and volatility, we employ a multiple linear regression model. We have chosen these variables specifically, as previous literature shows that earnings call sentiment influences stock price and it provides stock market participants with useful information for their buying-selling decisions (Hutton et al. 2009; Price et al. 2012). Furthermore, we have chosen multiple linear regression after looking at the four assumptions of linear regression (Bobbitt, 2020):

1. We have showcased the linearity of the relationship (see figure 3). We can see from this that there are some outliers in our data, which we need to address.

FIGURE 3:

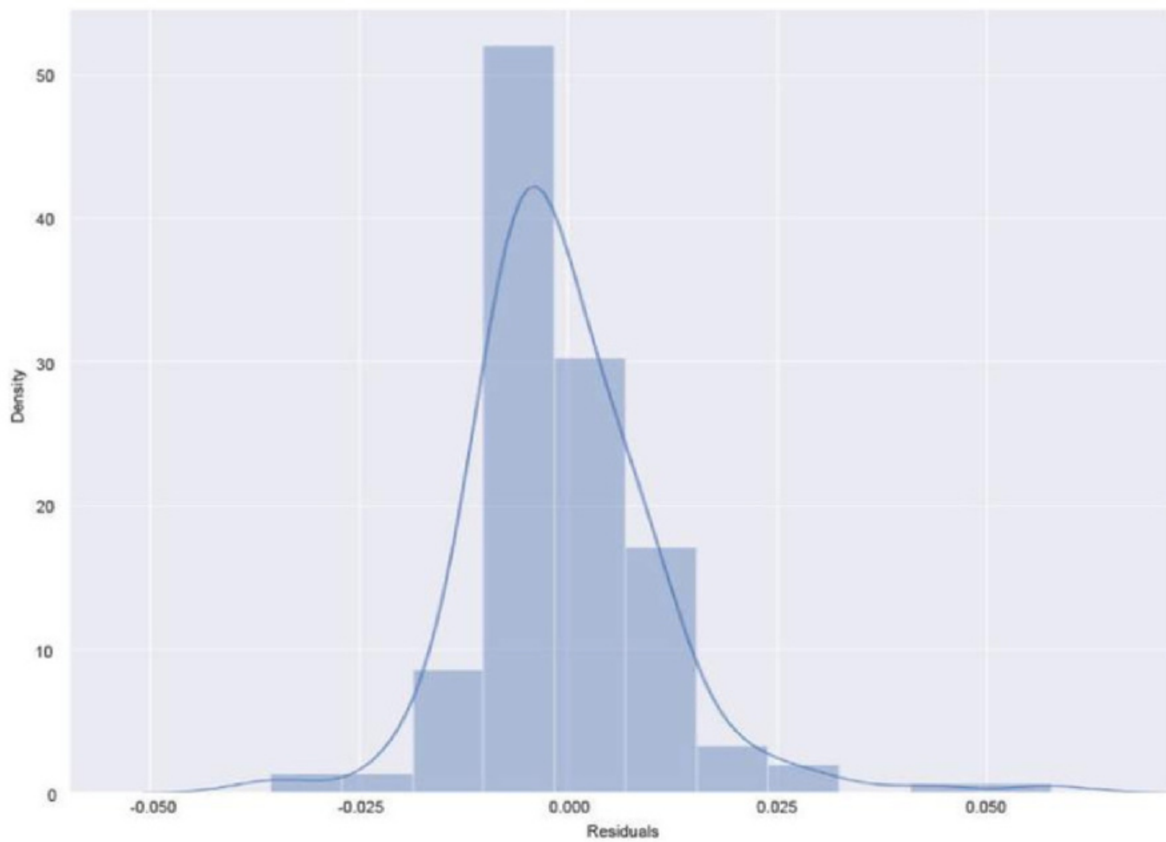
Linearity between DV and IVs



2. We determine that the distribution of residuals is slightly skewed and we attribute this to outliers in our data (see figure 4). We attempt to correct outliers in our regression model, but still, keep in mind the possible bias towards underestimation.

FIGURE 4:

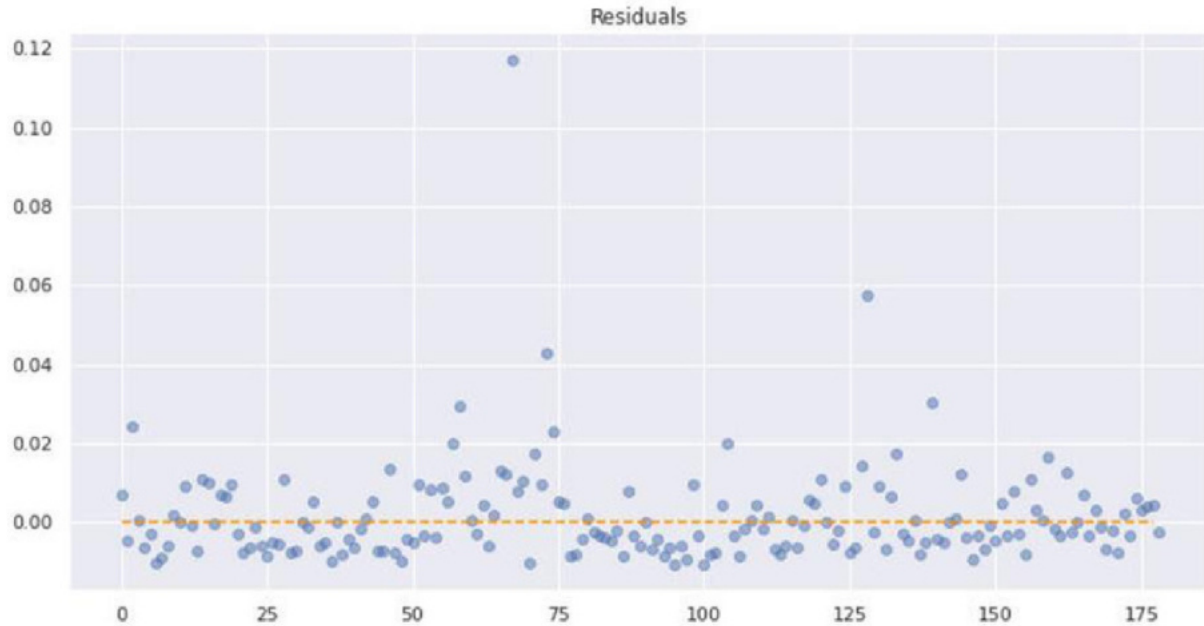
Normality of Residuals



3. Thirdly, we attempted to determine the presence of multicollinearity through a heatmap of the correlation and examine the variance inflation factor. This analysis did not work for us and we were unable to determine this third assumption.
4. Next, we looked at whether the residuals have constant variance at every level of X. Our goal was to see if we fulfilled the requirement of homoscedasticity. Compared to the latter assumption this one is more important as it affects the significance test of the coefficients. We were able to show proof of homoscedasticity (see figure 5):

FIGURE 5:

Homoscedasticity Graph



All of this leads us to build our multiple linear regression equation:

$$y_{Volatility\ Delta} = \beta_0 + \beta_1 x_{PreQA\ C-Score} + \beta_2 x_{QA\ C-Score} + \beta_3 x_{NASDAQ\ Vol.\ Delta} + \epsilon$$

RESULTS

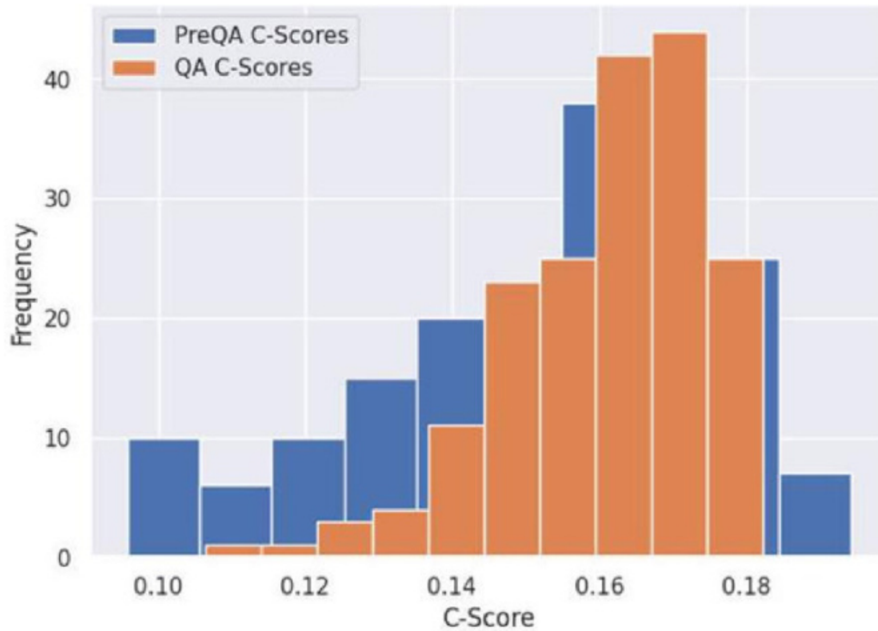
Descriptive Statistics

With our research question in mind, namely the effect of earnings call sentiment on stock volatility, we first look at the two most relevant descriptive statistics, first related to sentiment and the second to volatility:

1. The mean of the PreQA and QA sentiment compound scores are .15 and .16, respectively. This shows that our sentiment analyzer has rated most of the calls in our dataset as weakly positive. It could indicate that we do not have enough calls in our dataset that are negative or it may be a hint on the accuracy of the analyzer. Furthermore it might limit our ability to test for the negative sentiment and it could have an undesired effect on our regression model. To investigate further we plot the frequency of c-scores for both the PreQA and QA to visually represent the distribution (see figure 6):

FIGURE 6:

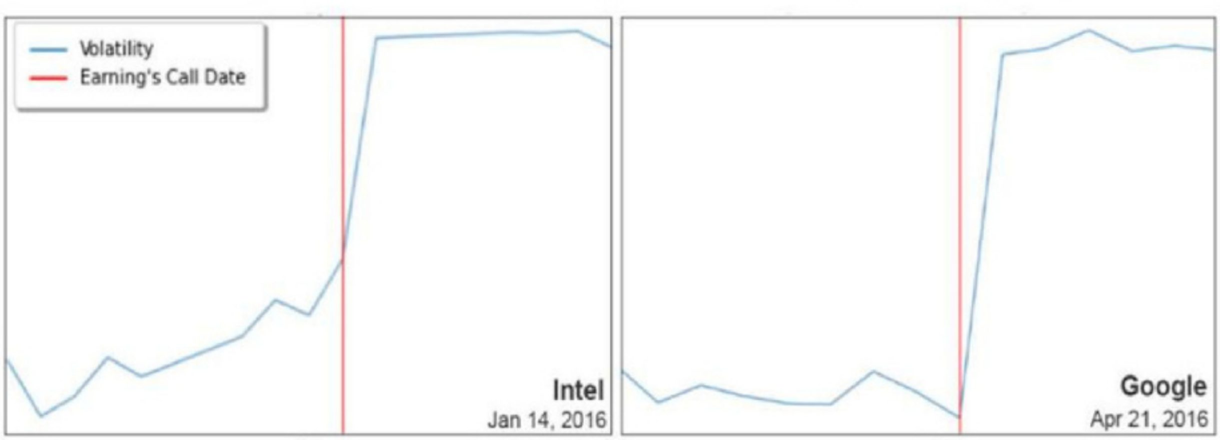
Frequency of Compound Scores



2. Our second descriptive statistic of interest is the mean of our dummy variable, Volatility Increase. It is set up to return one when there was an increase in volatility after the call. Its mean value is .87, showcasing that earnings calls have an effect on volatility in our sample, as 87% volatility increased in the period after. We plotted the volatility for each call and have 178 images. Provided below is an example of two such plots (see figure 7) and we have included additional images of the phenomenon in Appendix A.

FIGURE 7:

Volatility Before and After Earnings Calls of Intel (a) and Google (b)



(a) (b)

Main Results

Multiple linear regression was used to test if PreQA C-Score and QA C-Score significantly predicted Volatility Delta. We also incorporated NASDAQ Vol. Delta as our control variable. The fitted model was: Volatility Delta = - 0.002 - 0.039*(PreQA C-Score) + 0.100*(QA C-Score) + 0.242*(NASDAQ Vol. Delta). The regression equation was found to be insignificant ($F(3,175) = 1.133$, $p < 0.337$), with an R^2 of 0.019. It was found that the PreQA compound sentiment score predicted volatility delta insignificantly ($\beta = -0.039$, $p = 0.390$). Furthermore, the QA compound sentiment score predicted volatility delta insignificantly ($\beta = -0.104$, $p = 0.201$). If the QA result was significant, it would be in line with Hypothesis 2, as the coefficient shows a negative direction of the effect. However, given that our results are statistically insignificant we cannot confirm our hypotheses. Furthermore, the R-squared value tells us that the variables in our regression describe 1.90% of the variance in the dependent variable. Therefore, we can infer that we need to rethink the existing variables in our model, as well as the ones we could add to make it more robust.

CONCLUSION

Discussion

Earnings Call Effect On Volatility. When we set out with this study our main question was what the effect is of earnings call sentiments on volatility. Both of our hypotheses were unconfirmed as our model and variables were statistically insignificant. Although our results are insignificant and we cannot infer strength or direction of effects of sentiment on volatility, we have two outcomes that showcase an effect exists within our sample. Firstly, the descriptive statistics showcase that, in our sample, 87% of the time the volatility increased in the period after an earnings call. This shows that earnings calls have an impact and as previous research has shown, they trigger heightened trading and stock price responses (Bushee et al. 2003; Frankel et al. 1999). Heightened trading would increase the volatility, which is also showcased in our results. The next outcome is revealed when we plotted the linearity between the two sentiment variables and the change in volatility (see figure 3). If we look closely most of the data points fall above zero, which further supports our descriptive statistics.

Theoretical and Practical Implications

From a theoretical point of view, our results do not make any implications in terms of the relationships between the variables. However, our results show that only measuring sentiment is not enough to build a regression when investigating the effects of sentiment on volatility. This implies that there are other factors surrounding earnings calls that need to be accounted for.

The key implication for a practitioner, especially managers participating in earnings calls, is that the sentiment of their language is not enough to move the markets. However, the earnings call itself does move the markets through a combination of factors surrounding the call.

Limitations

Dataset. First, quantitatively we are limited, as our dataset contains 178 earning's call transcripts, while studies in this research area use much larger datasets. Qin et al. (2019) use an initial dataset of 2,243 transcripts reduced to 576 calls. Furthermore, the researchers have not only used the text but also the audio, which may provide more robust results. Finally, our dataset appears to have a predominantly positive sentiment, which may have impacted our regression model. Although we could also attribute the sentiment score results to the limitations of our analyzer, discussed in the next section.

Sentiment Analyzer. The research we conducted for this paper required multiple steps that influence and impact each other. Therefore, we cannot attribute our statistically insignificant results to the regression model alone. Another aspect is the sentiment analyzer. We used NLTK's VADER analyzer, which has the following limitations that potentially impacted our results for the independent variables:

1. The analyzer is specifically designed to determine the sentiment of social media posts, meaning that its accuracy may be impacted when used for financial language, as one used in earnings calls.
2. VADER employs a lexicon-based approach, meaning it gives a sentiment score to each word in our document based on its pre-prepared lexicon of words. As mentioned in (1), this analyzer is optimised for social media and the lexicon it has may be inadequate for our dataset.
3. Generally, sentiment analyzers are inherently limited for our purpose as they are designed for shorter texts. This required us to seek an alternative, namely to divide our text into chunks. It is not fully understood what the impact of this may be on sentiment score accuracy.

Regression Model. Before considering the results of our model, we look at the assumptions for a linear model and how our sample dataset fits into that. First, our normality of the residuals plot (see figure 4) shows that our model is biassed towards underestimation. This causes problems as it may shrink or inflate our confidence intervals. We attempted to fix the issue by addressing outliers, but ultimately it may be caused by insufficient linearity. The latter we see visualised in our linearity plot (see figure 3). Although there is linearity it could be stronger. These limitations lead us to believe we should explore other models, which we will address in the next section.

Future Research.

After carefully considering our limitations, we see a lot of room for improvement and exciting avenues for future research in this area.

Starting with our dataset, we are inspired by other researchers to want to include more data. In the information age, it is easy to get lost in the possibilities and to avoid that we follow the footsteps of Qin et al. (2019) and would like to include audio files. This opens the door to AI-powered sentiment analysis on voice with AssemblyAI (Foster, 2022). Furthermore, it may be beneficial to add data not directly related to our independent variable, such as news articles coming out before and after the earnings call. After scraping the articles with the Yahoo Finance API, we can run a sentiment analysis and use the scores as a control variable. Furthermore, it could be beneficial to control for other aspects of the call. If we refer back to the literature on earnings call attributes, we see multiple options. Firstly, according to Matsumoto et al. (2011), a longer call translates to a higher content of information and is more insightful to the market. Therefore it might be fruitful to include the duration of ECs as a control variable. Secondly, we might consider pre-QA and QA similarity. Lee (2016) found that the market responds negatively whenever the pre-Q&A portion of the call is similar to the Q&A one because it indicates pre-scripted answers to analysts' questions. There is a Pythonic solution to measuring similarity between texts as presented by Keshav (2020).

Next, we believe it would be beneficial to explore and test more sentiment analyzers. Specifically, analyzers that are trained on financial data and optimised for financial language. We have many avenues to explore, but one example is Google's BERT, as used by Briggs (2021). This analyzer has been further developed by Araci (2019) to create FinBERT. This language model is designed to tackle NLP tasks in the financial domain.

Finally, once we have addressed the issue of robustness of the data and the incompatibility of our sentiment analyzer, we want to improve our regression model. As with previous decisions, there are multiple considerations to be made. Firstly, our dataset showed the presence of outliers, although this can be addressed without a model change, we could change the model to quantile regression. This regression is suitable when linear regression requirements are not met or when the data contains outliers (Sharma, 2022). Any further choices on regression models could only be made after testing our new data for the assumptions.

We are excited to explore this area of research, improve our methods in our future academic endeavours and achieve insightful results.

REFERENCE LIST

Araci, D. (2019). FinBERT: Financial Sentiment Analysis with Pre-trained Language Models. *arXiv, 1908, 10063.*

Arslan-Ayaydin, Ö., Boudt, K., & Thewissen, J. (2016). Managers set the tone: Equity incentives and the tone of earnings press releases. *Journal of Banking & Finance, 72, S132-S147.*

Beaver, W. H. (1968). The information content of annual earnings announcements. *Journal of accounting research, 67-92.*

Bekaert, G., & Hoerova, M. (2014). The VIX, the variance premium and stock market volatility. *Journal of econometrics, 183(2), 181-192.*

Bloomfield, R. (2008). Discussion of "annual report readability, current earnings, and earnings persistence". *Journal of Accounting and Economics, 45(2-3), 248-252.*

Bobbitt, Z. (2020, January 8). *The Four Assumptions of Linear Regression*. Statology. Retrieved June 12, 2022, from https://www.statology.org/linear-regression-assumptions/

Bochkay, K., Hales, J., & Chava, S. (2020). Hyperbole or reality? Investor response to extreme language in earnings conference calls. *The Accounting Review, 95(2), 31-60.*

Bozzolan, S., Cho, C. H., & Michelon, G. (2015). Impression management and organizational audiences: The Fiat group case. *Journal of Business Ethics, 126(1), 143-165.*

Briggs, J. (2021, Mar 10). *How to Apply Transformers to Any Length of Text*. Medium. Retrieved May 24, 2022, from https://www.opinosis-analytics.com/knowledge-base/stop-words-explained/

Brown, S., Hillegeist, S. A., & Lo, K. (2004). Conference calls and information asymmetry. *Journal of Accounting and Economics, 37(3), 343–366.*

Burgoon, J., Mayew, W. J., Giboney, J. S., Elkins, A. C., Moffitt, K., Dorn, B., Spitzley, L. (2016). Which spoken language markers identify deception in high-stakes settings?

Evidence from earnings conference calls. *Journal of Language and Social Psychology, 35(2), 123-157.*

Bushee, B. J., Matsumoto, D. A., & Miller, G. S. (2003). Open versus closed conference calls: the determinants and effects of broadening access to disclosure. *Journal of accounting and economics, 34(1-3), 149-180.*

Cho, C. H., Roberts, R. W., & Patten, D. M. (2010). The language of US corporate environmental disclosure. *Accounting, Organizations and Society, 35(4), 431-443.*

Core, J. E. (2001). A review of the empirical disclosure literature: discussion. *Journal of accounting and economics, 31(1-3), 441-456.*

Corporate Finance Institute. (n.d.). *Earnings Call - Overview, Importance, and Structure.* Corporate Finance Institute. Retrieved April 7, 2022, from https://corporatefinanceinstitute.com/resources/knowledge/finance/earnings-call/

Davies, D., & Brennan, N. (2007). Discretionary disclosure strategies in corporate narratives: Incremental information or impression management. *Journal of Accounting Literature, 26(1), 116-196.*

Davis, A. K., Ge, W., Matsumoto, D., & Zhang, J. L. (2015). The effect of manager-specific optimism on the tone of earnings conference calls. *Review of Accounting Studies, 20(2), 639-673.*

Davis, A. K., Piger, J. M., & Sedor, L. M. (2012). Beyond the numbers: Measuring the information content of earnings press release language. *Contemporary Accounting Research, 29(3), 845-868.*

Doran, J. S., Peterson, D. R., & Price, S. M. (2012). Earnings conference call content and stock price: the case of REITs. *The Journal of Real Estate Finance and Economics, 45(2), 402-434.*

Foster, K. (2022, Feb 17). *Best APIs for Sentiment Analysis in 2022.* AssemblyAI. Retrieved June 16, 2022, from https://www.assemblyai.com/blog/best-apis-for-sentiment-analysis/

Frankel, R., Johnson, M., & Skinner, D. J. (1999). An empirical examination of conference calls as a voluntary disclosure medium. *Journal of Accounting Research, 37(1), 133-150.*

Frecka, T. J. (2008). Ethical issues in financial reporting: Is intentional structuring of lease contracts to avoid capitalization unethical?. *Journal of Business Ethics, 80(1), 45-59.*

Ganesan, K. (2019, April 6). *What are Stop Words?*. Opinosis Analytics. Retrieved April 27, 2022, from https://www.opinosis-analytics.com/knowledge-base/stop-words-explained/

Hartmann, J., Heitmann, M., Schamp, C., & Netzer, O. (2021). The power of brand selfies. *Journal of Marketing Research, 58(6), 1159-1177.*

Hayes, A. (2021, October 30). *Volatility Definition: Calculation & Market Examples.* Investopedia. Retrieved April 18, 2022, from https://www.investopedia.com/terms/v/volatility.asp

Henry, E., & Leone, A. J. (2016). Measuring qualitative information in capital markets research: Comparison of alternative methodologies to measure disclosure tone. *The Accounting Review, 91(1), 153-178.*

Hobson, J. L., Mayew, W. J., Peecher, M. E., & Venkatachalam, M. (2017). Improving experienced auditors' detection of deception in CEO narratives. *Journal of Accounting Research, 55(5), 1137-1166.*

Huang, X., Teoh, S. H., & Zhang, Y. (2014). Tone management. *The Accounting Review, 89(3), 1083-1113.*

Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Eighth International Conference on Weblogs and Social Media (ICWSM-14).*

Hutton, A. P., Marcus, A. J., & Tehranian, H. (2009). Opaque financial reports, R2, and crash risk. *Journal of Financial Economics, 94(1), 67-86.*

Jones, T. M. (1995). Instrumental stakeholder theory: A synthesis of ethics and economics. *Academy of management review, 20(2), 404-437.*

Jung, M. J., Wong, M. F., & Zhang, X. F. (2018). Buy‐side analysts and earnings conference calls. *Journal of Accounting Research, 56(3), 913-952.*

Keshav, A. (2020, February 27). *Measuring the Document Similarity in Python*. GeeksforGeeks. Retrieved June 22, 2022, from https://www.geeksforgeeks.org/measuring-the-document-similarity-in-python/

Kim, J. B., Wang, Z., & Zhang, L. (2016). CEO overconfidence and stock price crash risk. *Contemporary Accounting Research, 33(4), 1720-1749.*

Lee, J. (2016). Can investors detect managers' lack of spontaneity? Adherence to predetermined scripts during earnings conference calls. *The Accounting Review, 91(1), 229-250.*

Li, F. (2010). Textual analysis of corporate disclosures: A survey of the literature. *Journal of Accounting Literature, 29(1), 143-165.*

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10‐Ks. *The Journal of Finance, 66(1), 35-65.*

Matsumoto, D., Pronk, M., & Roelofsen, E. (2011). What makes conference calls useful? The information content of managers' presentations and analysts' discussion sessions. *The Accounting Review, 86(4), 1383-1414.*

Mogyorosi, M. (2021, Jan 13). *Sentiment Analysis: First Steps With Python's NLTK Library*. Real Python. Retrieved May 20, 2022, from https://realpython.com/python-nltk-sentiment-analysis/

Monkey Learn (n.d.). *Sentiment Analysis Guide.* Monkey Learn. Retrieved June 20, 2022, from https://monkeylearn.com/sentiment-analysis/

NLP Stanford University. (2009, April 7). *Tokenization*. Stanford University. Retrieved May 20, 2022, from https://nlp.stanford.edu/IR-book/html/htmledition/tokenization-1.html#

Patelli, L., & Pedrini, M. (2014). Is the optimism in CEO's letters to shareholders sincere? Impression management versus communicative action during the economic crisis. *Journal of Business Ethics, 124(1), 19-34.*

Price, S. M., Doran, J. S., Peterson, D. R., & Bliss, B. A. (2012). Earnings conference calls and stock returns: The incremental informativeness of textual tone. *Journal of Banking & Finance, 36(4), 992-1011.*

Qin, Y., Yang, Y. (2019). What You Say and How You Say It Matters: Predicting Stock Volatility Using Verbal and Vocal Cues. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 390–401.*

Roldós, I. (2021, May 31). *Text Cleaning for NLP: A Tutorial*. Monkey Learn. Retrieved May 19, 2022, from https://monkeylearn.com/blog/text-cleaning/

Roozen, D.; Lelli, F. Stock Values and Earnings Call Transcripts: a Dataset Suitable for Sentiment Analysis. *Preprints 2021, 2021020424.*

Sharma, P. (2022, Jan 19). *Different types of Regression Models*. Analytics Vidhya. Retrieved June 12, 2022, from https://www.analyticsvidhya.com/blog/2022/01/different-types-of-regression-models/

Whaley, R. E. (2000). The investor fear gauge. *The Journal of Portfolio Management, 26(3), 12-17.*

Whaley, R. E. (2009). Understanding the VIX. *The Journal of Portfolio Management, 35(3), 98-105.*

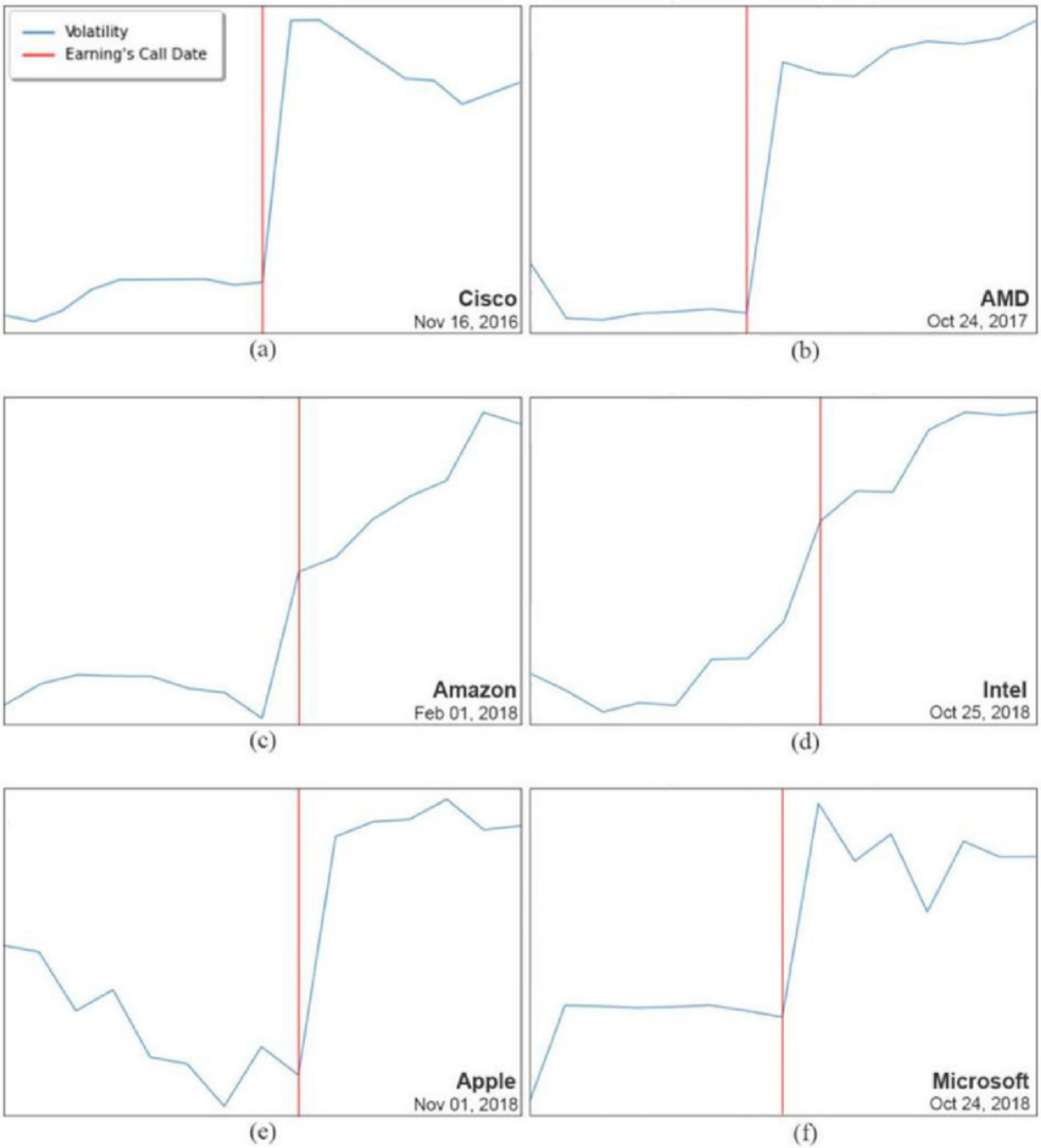
Yuthas, K., Rogers, R., & Dillard, J. F. (2002). Communicative action and corporate annual reports. *Journal of Business Ethics, 41(1), 141-157.*

Word Count: 5028

APPENDIX A

Volatility Plots

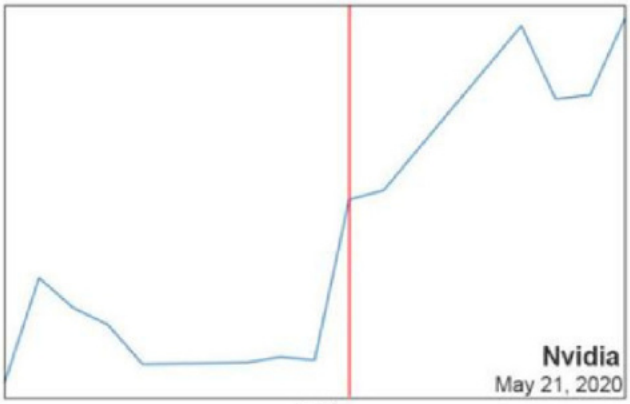
The provided plots do not include the x-axis, nor the y-axis for simplicity, as the intention is to show the increase or decrease. All full plots are accessible here: http://e.pc.cd/HvPotalK - with the following naming format: "YEAR-MONTH-DAY-TICKER_NAME-plot.png".



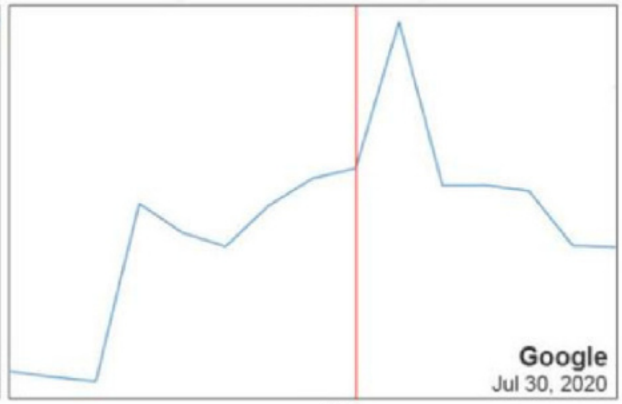
(a) (b)

(c) (d)

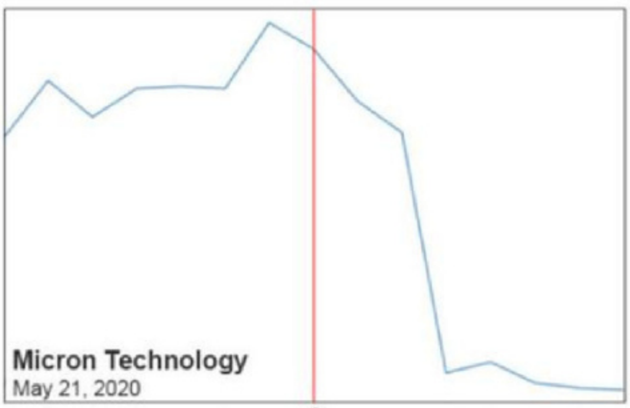
(e) (f)


Nvidia
May 21, 2020

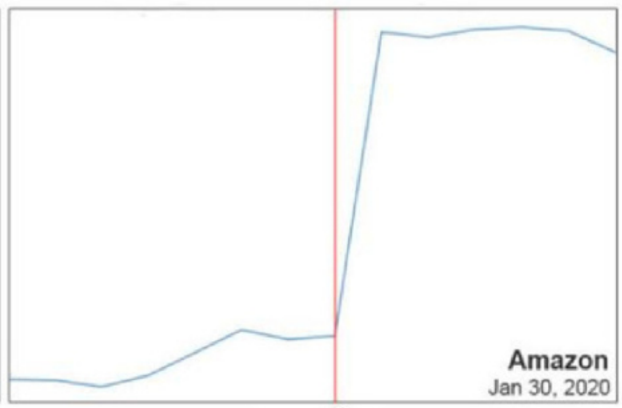
(g)


Google
Jul 30, 2020

(h)


Micron Technology
May 21, 2020

(i)


Amazon
Jan 30, 2020

(j)

APPENDIX B

Python Notebook Access

Collab notebook with our Python code and outputs is accessible here: http://e.pc.cd/5vPotalK