Journal of Research High School

Vol. 2024 (11) 484 – 490
ISSN 2688-3651 [online]

A Quantitative Analysis of the Media's Influence on Stock Prices and Overall U.S. Financial Market

Sean Magee[1] *

[1]Episcopal Academy, Newtown Square, PA, USA
*Corresponding Author: mages25@episcopalacademy.org

Advisor: David Tong, dtong@uchicago.edu

Received January 6, 2024; Revised July 16, 2024; Accepted August 24, 2024

Abstract

The prevalence of online newspapers and media sources has been coupled with the growth of accessible retail investing throughout the U.S.. Although there has been an increase in forecasting, there are no conclusive determinations regarding the influence of media on stocks and the broader market. The lack of current literature surrounding this contemporary topic of news media's relationship with the Market Efficiency Hypothesis necessitates new research. To assess the degree to which the media is able to affect the stock market, articles pertaining to Apple and S&P 500 from 7/28/22 to 2/03/23 were selected from the Wall Street Journal and Barrons. Each article was given a positivity rating, ranging from Extremely Positive to Extremely Negative, and was compared to the change in the corresponding stock's price. The research revealed a weak correlation with Apple, which had less data, and a mildly strong correlation with S&P 500. Using the correlation coefficient equation and statistical analysis, a 0.75 correlation factor was determined for S&P 500. Thus, these findings refute the Market Efficiency Hypothesis, showing how the stock does not trade at its "fair" value, and how modern factors such as instantaneous news and media influence its price. Overall, as news platforms continue to increase viewership, along with the ubiquity of social media platforms such as X, Reddit, and TikTok which continue to drive investing popularity and regularity, the influence that media has on the market will become paramount as new technology emerges.

Keywords: Market Efficiency Hypothesis, Rating Ratio, Classification Categories, News and Media

1. Introduction

With the emergence of new trading technologies and apps, convenience for retail investors has skyrocketed. Apps such as Robinhood, TD Ameritrade, Fidelity, TradingView and more have allowed for instantaneous transactions for the average working-class citizen. The combination of accessibility and an influx of media spotlight has allowed for a retail investor trading boom. In an attempt to build wealth, this upward trend of retail investors climaxed at a significant 77% market share of total stocks, mutual, and pension funds (Mackintosh, 2017). The most infamous result that highlighted this retail growth is the Reddit WallStreetBets incident, in which tens of thousands of retail investors executed a successful short squeeze on GameStop and AMC stock, which resulted in huge retail gains and huge hedge fund losses (Telesca et al., 2021). These revolutionary events bring curiosity to how retail investors make decisions, and how these decisions in turn affect the market. Further inspection of this topic raises the inquiry: through which media do investors consume market information, and how do the media and news influence individual stocks and perhaps the overall U.S. market? Amidst the prosperity of investors, major media outlets providing financial news have benefited, with reader counts at an all-time high. Accredited and reliable sources such as the Wall Street Journal, Barrons, The Economist, and others have grown to acquire millions of viewers a month (Castillo, 2023). With an immense outreach, these big media corporations have power over what daily viewers see, impacting what they make

Vol. 2024 (11) 484 – 490
ISSN 2688-3651 [online]

investments in, and as a result, the stock price itself. Over the past decade, with the onset of electronic news and media's prevalence, there has been an increase in the literature regarding digital news' relationship with the stock price. Different researchers have used different techniques to investigate this phenomenon, ranging from using models that digitally scan tone throughout articles to models that include time delay (MacDonald, 2022) (Ren, Y., Liao, F., & Gong, Y., 2020). However, there is a major inconsistency within this range of literature, with those who support the idea of news' correlation with the stock price, those who are against it, and those who are not able to generate solid evidence for their claim. There is especially a lack of literature regarding this topic when related to the Market Efficiency Hypothesis.

Thus, this research is dedicated to investigating and testing the hypothesis that media and news influence the variability of stocks and the overall U.S. market. According to this hypothesis, there is expected to be a relationship between the news rating and price change, in that when news is positive, the stock price is expected to increase, and vice versa. If the hypothesis is proven and statistically significant, then it contradicts the Market Efficiency Hypothesis, in that the stock is not fully reflective of all available information. To ensure valid results, proper steps using defined research objectives need to be taken. Firstly, the collection of data must be from a trustworthy source which is both reputable and frequently used. Secondly, the amount of data needs to be vast in order for the results to be significant. Thirdly, data analysis and statistical analysis must be conducted thoroughly and vigorously, to ensure the study's integrity, and reproducibility, and to show if the findings are statistically significant. Lastly, qualitative portions of the research such as interviews must be properly integrated and connected with the quantitative aspects in order to form a coherent analysis and argument. Thus, whether the hypothesis is proven or disproven, the findings are valid, due to the proper research steps taken throughout the study. However, it is also important to mention how the methodology of the rating process is slightly subjective and could raise questions regarding the validity of the study, the system used is largely replicable and reliable.

2. Methodology

It was hypothesized that news surrounding the sector of a stock, or the particular stock itself, changes how investors forecast the stock's potential value, subsequently driving the stock's price based on the buy/sell ratio of the stock. To examine this hypothesis, Apple stock and S&P 500 were chosen for the study. These two were selected since both have substantial market caps and trading volume, thus consequently receiving more news and media attention to assess correlation. Smaller companies were avoided due to the lack of data and the possibility of one factor, e.g. an earnings report, influencing the stock's movement and skewing the correlation. The S&P 500 was chosen as a benchmark compared to Apple as a "big tech" stock.

For this research, the timeframe of July 28th, 2022, to February 3rd, 2023, was determined to be sufficient regarding the amount of news for both Apple and the S&P 500, as 6 months is considered a sufficient duration to extract trends. For news and media collection, the Wall Street Journal (WSJ) and Barons were used because of their reputation and immense online traffic. For both WSJ and Barron, the news articles were sourced under the "News" section, after searching the stock's name or ticker symbol. It is also important to note that in WSJ, news articles were obtained from the "significant news only" section, disregarding extraneous news which did not relate to the specific company or index. This selection helps maintain the relevancy of the data and the integrity of the research.

After collecting all news articles from WSJ and Barrons in the timeframe specified above - 113 usable for Apple and 1572 for the S&P 500 - each article was given a rating and a classification. Rating was determined based on a scale of Extremely Positive, Positive, Slightly Positive, Neutral, Slightly Negative, Negative, and Extremely Negative. Although these ratings are subjective, the rating was given based on whether current and future implications were shone in a positive or negative light, and to the degree of the impact. In order to properly assess each article's positivity, the tone of terminology used throughout each article was accounted for. Articles that provide a drastic outlook specifically on the companies' future or stock price are given an "Extreme" label. However, articles that reference drastic outlooks on its sector but not the specific company are given the regular Positive or Negative label. Positive or Negative articles are denoted by choice of words, tone, and future predictions, while the "Slightly" label is given to articles that barely change the outlook of the company's future, or have limited tone. Finally, the "Neutral" label is

Vol. 2024 (11) 484 – 490
ISSN 2688-3651 [online]

given to articles that offered no deviating prediction and reported facts without tone. In each article, news that did not relate to the specific stock of interest was disregarded and was not factored into the rating assignment. Articles for both Apple and the S&P 500 contained a mix of reactionary news, predictions, and suggestions. Classification was determined based on a scale of Extremely Relevant, Relevant, or Barely Relevant. These classifications were given by assessing how pertinent the article was in regard to that specific stock, thus helping maintain the integrity of the research throughout the study.

Once every article relating to their respective stock had a rating and a classification, each article, based on its rating category, was sorted with the corresponding price change (calculated as a percentage based on the previous day's close) on that day. After this matching, the sum of all price differences was added together based on their rating category, and divided by the total number in that rating category, resulting in the average price difference for each rating category. This procedure was executed for each rating for the S&P 500 and Apple data. Based on our hypothesis, it was expected that the average stock price would change to correlate with the positiveness or negativeness of the news article.

Additionally, because S&P 500 had more data points allowing for better analysis of correlations between news and stock price fluctuations, a correlation between the two in the S&P 500 was estimated by calculating a rating ratio (rating ratio was calculated as the number of overall positive news divided by the number of overall negative news) which was then compared with the price change of the S&P 500 for each day it traded.

3. Quantitative Data and Analysis Results

3.1 Data Collected

After the ratings and classifications were assigned, the dates for both news and stock price changes were paired and compared. Over the course of the 6 months timeframe, 145 news articles for Apple were published, 113 of which were published on a trading day, thus generating 113 data points for analysis of Apple. For the S&P 500, 1572 articles were published, however, only 1447 were published on trading days and subsequently used in analysis. For both stocks, percent changes in stock price were totaled for each rating: Extremely Positive (Ex. Pos.), Positive (Pos.), Slightly Positive (Sli. Pos.), Neutral (Neutral), Slightly Negative (Sli. Neg.), Negative (Neg.), Extremely Negative (Ex. Neg.). The number of data points per rating (#Data Points) is listed in Tables 1 and 2 below as well as the average over the course of the study (AOS). The raw data calculated is delineated in Tables 1 and 2 below on the left. However, to ensure cleaner data, a separate data set is included. For this "cleaned" data set, price changes were calculated using Extremely Relevant and Relevant news classifications, disregarding all news articles deemed Barely Relevant. In addition, all rating categories with five or fewer data points were not included to exclude anomalies which maintained the integrity of the analysis.

Table 1. Apple Data Sets Over 6 Months Time Frame, Comparing Rating vs Average Price Change

Apple Raw Data Set			Apple Cleaned Data Set		
Rating	Avg Price Change	# Data Points	Rating	Avg. Price Change	# Data Points
AOS	+.21%	113	AOS	+.37%	41
Ex. Pos.	+.37%	6	Ex. Pos.	-	-
Pos.	+.52%	24	Pos.	+.56%	20
Sli. Pos.	-.55%	15	Sli. Pos.	-	-
Neutral	+.08%	23	Neutral	-	-
Sli. Neg.	+.99%	17	Sli. Neg.	-	-
Neg.	-.35%	25	Neg.	-.36%	21
Ex. Neg.	+1.36%	3	Ex. Neg.	-	-

Vol. 2024 (11) 484 – 490
ISSN 2688-3651 [online]

Table 2. S&P 500 Data Sets Over 6 Months Time Frame, Comparing Rating vs Average Price Change

S&P 500 Raw Data Set			S&P 500 Cleaned Data Set		
Rating	Avg Price Change	# Data Points	Rating	Avg. Price Change	# Data Points
AOS	+.04%	1447	AOS	+.15%	181
Ex. Pos.	+1.30%	10	Ex. Pos.	+1.08%	9
Pos.	+.59%	54	Pos.	+.68%	41
Sli. Pos.	+.12%	611	Sli. Pos.	+.43%	19
Neutral	-.01%	23	Neutral	+.13%	17
Sli. Neg.	-.03%	427	Sli. Neg.	+.04%	16
Neg.	-.28%	105	Neg.	-.19%	70
Ex. Neg.	-.80%	9	Ex. Neg.	-.80%	9

3.2 Apple Findings

As indicated in both data sets, there is a weak correlation between the rating of the news and the average price change for that rating. For the Raw Data Set, the Extremely Positive, Positive, Neutral, and Negative ratings had the predicted result, whereas the Slightly Positive, Slightly Negative, and Extremely Negative ratings did not show the expected result. However, when the data was cleaned (Ratings with less than 5 articles were removed and the "Barely Relevant Classification" was removed in the new calculation), there was a much stronger correlation with the Extremely Positive, Positive, Neutral, and Negative ratings. Only the Slight Negative ratings contradicted the hypothesis (Extremely Negative and Slightly Positive ratings were excluded to maintain data integrity). Overall, while both sets of data display a correlation between news ratings and stock price, the cleaned data does have a stronger correlation.

3.3 S&P 500 Findings

Similarly to Apple Data, the S&P 500 data also indicated a correlation; however, the S&P 500's correlation was much stronger than Apple's. This is most likely attributed to the vast amount of data the S&P 500 had (10 times more articles published than Apple) and the S&P 500's less susceptibility to volatile reporting, as not only does the S&P 500 have more reporting, but also because it's a composite of the top 500 companies which is a stable bet. In the raw data set, the ratings exhibited a direct relationship between the degree of the rating and the average price difference; as the categories' ratings became more severe, the average price change followed the severity. This correlation was strong for both, however, the cleaned data set's fewer data points hindered its correlation as opposed to the raw data: 1447 versus 181 data points. Overall, the S&P 500 had a strong correlation between the news articles released and its price change.

3.4 Correlation Findings

This graph represents the correlation between the rating ratio of positive to negative news and the price change. The X-axis represents the ratio of positive to negative news and was calculated as the number of overall positive news divided by the number of overall negative news. Rating ratios below 1.00 were days in which more negative than positive news was released. If the ratio was 3.00 on a given day, the positive to negative news articles could have been 9:3 for example. The Y-axis represents the percent change in the stock price of the S&P 500 for each day. A linear

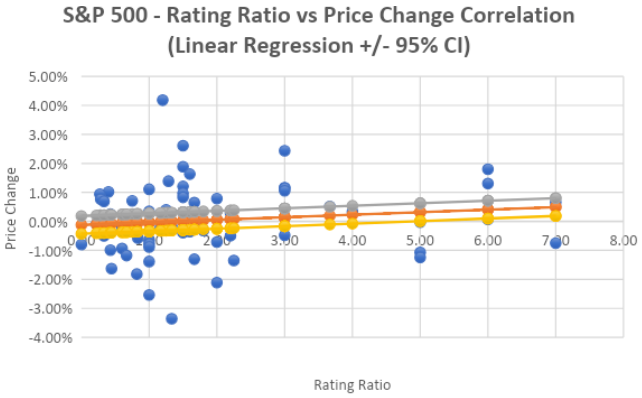

Figure 1: The correlation between the rating ratio and the price change in S&P 500.

Vol. 2024 (11) 484 – 490
ISSN 2688-3651 [online]

regression line was constructed with 95% confidence intervals, demonstrating the substantial correlative relationship between rating ratio and price change. It is important to note that the linear regression is constructed assuming linearity, independence, homoscedasticity, normality and multicollinearity between the dependent variable and the independent variable, i.e., the rating ratio and the price change in the S&P 500, respectively.

To estimate the correlation between the rating of a news piece and the price change of the stock, extremely positive, positive, and slightly positive rating categories were all reassigned as positive for the sake of data sufficiency and calculation. The same process was done with all negative rating categories. The counts of positive and negative news were summed and recorded for each day, revealing the 66 viable data points in which the S&P 500 traded in the timeframe specified for the study. Also, it is important to mention that each data point averaged 9.33 news articles per day. To calculate the rating ratio, the positive news count was divided by the negative news count. This method was selected as opposed to simply subtracting each count by the other, due to more accurate reflections of what the reader of the news would be exposed to. Days with larger data were better represented with a ratio than a subtraction, for example, a day with a 23 positive news count and an 18 negative news count would have a better comparison to a day with a 6 positive news count and a 1 negative news count using a ratio since both days would have a +5 count using the unselected subtraction method. Thus, using a ratio to represent the counts was more reflective of how the audience of Barrons and WSJ would perceive the news, using both counts as an anchor to determine the true meaning of the difference between the two. Using the correlation coefficient equation, a correlation of 0.75 was determined, reflecting a strong positive correlation. The correlation graph deceived this strong coefficient since while the 66 data points on the graph contributed to a higher calculation of the correlation coefficient (a lesser amount of data, such as 25 data points, would have yielded a weaker correlation), it made the trend look random. However, as an example, although the data points with a rating ratio of less than 1 seemed random, there were 12 data points with a negative price change, compared to only 7 data points with a positive price change. So, as the line of best fit demonstrated, there was a positive correlation despite the depicted intricacies. Overall, there was a mildly strong correlation between the rating of a news article and the price change of the S&P 500. This indeed supported the hypothesis that media and news had an influence on the stock market and its fluctuations.

4 Qualitative Data and Key Conceptual Factors

The underlying assumption upon which this research is based is that current news and media is not already factored into the stock's price. Therefore, when news articles are released, it is new information that has not previously been factored into the price by the majority of investors. Although larger corporations and capital firms have faster access to information, generally speaking, the public is not well-informed before news is released. There are also other important aspects that play a role such as the time delay between the release of news and the trading of the stock and previous predictions that affect the experiment (Mamaysky 2023). These aspects have been taken into account when giving an article a rating - acknowledging the previous predictions in relation to the "surprising" news along with the fact that most trading is done daily by investors when reacting to the news. Price change is calculated to include aftermarket news and trading by basing price change off of the previous day's close, not the current day's open.

This behavioral economics research revolves around the ever-controversial *Market Efficiency Hypothesis,* developed by Eugene Fama (Fama, 1970). The Market Efficiency Hypothesis states: when new information comes into the market, it is perfectly reflected in stock prices and thus it would be impossible to "beat" the market. (Fama 1970). However, this current research produced contradicting results and therefore opposes this theory since the data shows a correlation between news and corresponding stock price change. This result implies the reality of inequities in the market, whether it be specialized knowledge, lower trading costs, low management fees, or agency costs, and that one is able to profit from the advantage (Ang et al., 2011). To source second opinions regarding this hypothesis, thought leaders in the field such as Wharton Lecturer, Mr. David Erickson, and Johns Hopkins Founding Director of International Economics and Finance, Dr. Gordon Bodnar, were interviewed for their input. As Dr. Bodnar first makes clear, one of his key points is that market efficiency is fundamental to understanding finance but gets ridiculed by the press because they don't fully understand that it's not meant to be an undefiable theory. In other words, a stock's price is never going to be perfectly reflected at its true value, however, it is close enough to assume the application of this

Vol. 2024 (11) 484 – 490
ISSN 2688-3651 [online]

hypothesis. This is quintessential to this research and shows how the release of news articles (new information) is reflected in the stock's varying price, adjusted with time delay for investors. Commenting upon this research and methodology, "Efficient market hypothesis is the idea that since there's no pattern (of news), then the stock prices should behave pretty randomly, and as your research has shown, if a piece of positive information comes out, it generally causes the stock price to go up, then if a piece of negative information that people weren't aware of before comes out, it goes down". So, as pieces of positive and negative news instantaneously stream in and are available to the public, the stocks will reflectively vary based on such constant news. According to Dr. Bodnar, "[The Market Efficiency Hypothesis] Leads to an idea that if everyone is rational and if everyone knows about Apple in terms of what is currently available, but there's going to be something new coming out today or tomorrow, and that's going to change people's opinion about where the stock price is". The research conducted is merely testing the Market Efficiency theory and determining whether such a theory is applicable to a set of data.

However, Mr. Erickson takes a similar but slightly different path, arguing that "The market does not trade on perfect information". His thought of Market Efficiency is that due to the continuous influx of public information and news inhibits the ability for any stock to trade at its fair price. As Mr. Erickson has decades of experience in the area of finance, he shares his experience that "You see things mispriced from a technical standpoint. Do you see them correct themselves over time? It really depends." He suggests that there could be factors based on human error that inflate or deflate the price, misinterpreting its fair value. A prime example of this would be the GME short squeeze during early 2021, where retail investors saw hedge funds shorting the stock and subsequently intentionally bought so much of the stock that when the hedge funds' shorts contracts expired, it drove the price up even more. This short squeeze of GME defies the assertion that stocks trade mostly around their fair price, as retail investors sought to artificially inflate its stock price. Mr. Erickson also clarifies an important aspect of this research "It's not news generically, it's specific news that can impact a stock… and some news is significant to that stock and some news is not. It also has to do with how well followed the company is." Although these factors have been directly addressed through the methodology of the experiment, one flaw he points out is how the classification of the news article (positive vs negative) does not take into consideration the existing expectation of that news. An example Mr. Erickson gave is that if everybody assumed that Apple's quarterly would hit a certain expectation, but Apple's earnings increased only marginally, the stock would fall despite having that increase in earnings. Another reason that Erickson details as to why the Market Efficiency Hypothesis is faulty is because major institutions have faster access to news compared to retail investors. Because institutions can make trades based on newer information compared with a retail investor's time delay, the stock is not trading at its fair value because it's not an even playing field.

5 Conclusion

The research supports the idea that the market efficiency hypothesis does not function perfectly. It can be said that there is a direct correlation between news released and the subsequent change in stock price. Apple's data does show a weak correlation, but the S&P's superior amount of data shows such correlation and impact much more clearly and is almost a perfect representation of the expected result. Although this hypothesis of news impacting stock price appears plain, the rise of social media and instantaneous news will fundamentally have a great impact on how oncoming generations view behavioral economics. Occurrences such as GME and AMC, powered by the interconnectedness of social media and news, will certainly increase in frequency, defying our current understanding. Overall, as technology continues to advance, it is uncertain whether markets or even human action, will continue to follow conventional rational behavior.

6 Reflection

The research yielded a weak correlation for Apple and a mildly strong correlation for the S&P 500. However, the study does have some critiques to be had. Firstly, though a six-month period was a sufficiently long time frame for the S&P 500 yielding over 1500 data points, a longer time period would have been better to collect more data for Apple. This would have delivered cleaner results, decreasing the margin of error. One more point of a flaw within the

Vol. 2024 (11) 484 – 490
ISSN 2688-3651 [online]

nature of the study is that while it is conceded that the Rating system and Classification system are subjective but do have a repeatable methodology, it is important to note how closely Rating and Classification are related to each other, and therefore influence the given rating and classification of the article. Another inherent potential bias is reactions to news might not be directly from Wall Street Journal or Barrons but to general news released by the company and amplified by other media forms, such as Twitter and Reddit. Although the point of time delay may raise concerns, the way the stock price change was calculated took into account the possibility for news articles that were released after trading hours to be incorporated properly, by taking into account the closing price of the previous day's trade. Also, while two high-attention stocks are sufficient for this study to analyze the viability of the market efficiency hypothesis, it would have been more in-depth to include companies ranging across sectors and publicity share.

References

Castillo, L. (2023, October 31). Wall Street Journal Readership Statistics [Fresh Research]. https://blog.gitnux.com/wall-street-journal-readership-statistics/

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383-417. https://doi.org/10.2307/2325486

MacDonald, M. (2022, August). *Quantifying the impact of news on stock price changes*. Retrieved June 19, 2024, from https://www-2.rotman.utoronto.ca/insightshub/finance-investing-accounting/news-stock-swings

Mackintosh, P. (2020, December 17). *Who Counts as a Retail Investor?* Nasdaq. Retrieved November 26, 2023, from https://www.nasdaq.com/articles/who-counts-as-a-retail-investor-2020-12-17

Mamaysky, H. (2023, July 26). *AI and the Financial Markets: Closing the Gap between News and Stock Prices to Prevent Banking Crises*. Columbia Business School. Retrieved November 7, 2023, from https://leading.business.columbia.edu/main-pillar-21st-century-finance/21st-century-finance/ai-financial-m arkets

Ren, Y., Liao, F., & Gong, Y. (2020). *Impact of News on the Trend of Stock Price Change: an Analysis based on the Deep Bidirectiona LSTM Model*. Science Direct. Retrieved June 19, 2024, from https://www.sciencedirect.com/science/article/pii/S1877050920315829

Review of the Efficient Market Theory and Evidence [Review *Efficient Market Hypothesis*, by A. Ang]. (2011, April 27). Retrieved October 23, 2023, from https://www0.gsb.columbia.edu/faculty/aang/papers/EMH.pdf

Telesca, J. (2021, January 29). *GameStop Stock, Reddit and Robinhood: What You Need to Know How traders online drove the videogame retailer's stock to new highs and caused a crazy week in markets*. Wall Street Journal. Retrieved November 26, 2023, from https://www.wsj.com/articles/gamestop-stock-reddit-and-robinhood-what-you-need-to-know-11611960243