The Impact of Alternative Data and Social Media Sentiment on Stock Market Movements

Agrim Vishnoi & Aryan Jhaveri

Introduction

In the age of big data and rapid technological advancements, the proverb "knowledge is power" has never been more applicable, especially in the realm of finance. Investors and analysts have traditionally relied on standard metrics and reports to make informed decisions, but in this new world of technology, a substantial shift is underway. The emergence of alternative data sources provides an abundance of unconventional insights ranging from satellite data that captures real-time retail activity to social media sentiment that reflects public opinion. This new front of data promises to tap into hidden market trends and provide a competitive edge. However, the allure of these new sources is accompanied by a critical question: can they be trusted to provide accurate and unbiased answers, or do they simply introduce noise and create new risks into financial forecasting?

As we dive into this complex environment, the potential of alternative data sources to reshape stock market evaluations shouldn't be neglected. Alternative data refers to data gathered from nontraditional sources, e.g, credit card transactions, social media, satellite imagery, and product reviews (Liberto). Unlike traditional data, which often reflects past performance and is limited in scope, alternative data can offer real-time insights into consumer behavior, market sentiment, and even geopolitical events. For instance, the rise of social media platforms such as Twitter (now known as X) and Reddit has given them unprecedented influence over stock prices. This is especially evident in the dramatic rise and fall of companies like GameStop and AMC, which were driven more by online hype than by fundamental business metrics. Additionally, satellite imagery allows for the monitoring of infrastructure developments

and retail activity, offering a unique perspective on economic health that traditional data sources might miss.

Yet, with these advantages come significant challenges. The integration of such diverse and unstructured data sources into stock market analysis raises concerns about data quality, biases, and the potential for misinterpretation. For example, social media can amplify certain narratives, leading to the over- or undervaluation of stocks that don't accurately reflect their true intrinsic values. Similarly, while credit card transaction data provides insight into consumer spending patterns, it also raises privacy issues and questions about data accuracy.

One major aspect of the stock market significantly impacted by the rise of alternative data is stock market evaluation. As markets rapidly evolve due to alternative data's influence, evaluating a stock's intrinsic value becomes increasingly difficult. Our study sheds light on this critical shift in financial markets.

Literature Review

Reputed technology trend journalist Stephen Gossett explains how alternative data is transforming finance. He highlights that these types of data enable more accurate analysis, historically backed decisions, rewarding investments, and a competitive edge in the market. Though nontraditional, alternative data has proven effective for financial decision-making. According to Liberto, alternative data includes sources like satellite imagery, credit card transactions, and social media. This contrasts with traditional stock market data such as financial reports and earnings calls, and it's becoming increasingly essential.

Alternative data also plays a growing role in stock market evaluation. Traditionally, investors relied on two main forms of analysis: fundamental and technical (Chen). Fundamental analysis involves researching financial records, economic reports, company assets, and market share. Technical analysis, in contrast, focuses on price patterns to predict future movements. Alternative data influences both approaches. In technical analysis, it uncovers patterns or trends previously undetectable. For fundamental analysis, it can supplement or even replace traditional metrics to identify new investment opportunities.

Alternative data, as said previously, includes social media, web traffic, credit card transactions, and satellite imagery. When investors make decisions based on alternative data, they look at different types of sources to gain knowledge of distinct information. Credit card transactions are used to predict companies' revenues and profitability because they provide real-time data on how much consumers are spending on a company's product or service (Marenzi). Satellite imagery provides unique insights to a company. For example, it can provide data on retail activity, company construction projects, and potential natural disasters—all of which provide data on impactful areas of a company's success (Liberto). Web traffic data can be used to see spikes of interest in certain apps and websites. This could provide investors with an opportunity to invest in a company because of an increase in sudden attention. Lastly, social media provides a view of consumers' views on the company itself. This can be used to gauge the company's success based on how the customers themselves feel. Furthermore, it can provide investors with an indicator of future revenues. As you can see, there are many types of alternative data. Each of these different data sets provides investors with data on different aspects of the company.

A common phenomenon used when using social media as a form of alternative data is sentiment analysis. According to Professor Erik Cambria from NTU Singapore, sentiment analysis is a way of analyzing digital texts to determine the type of emotion that is displayed. This is heavily used as a way of analyzing alternative data sources and, more specifically, social media. For example, Associate Professor Kristian Bondo Hansen from Copenhagen Business School says that using sentiment analysis on alternative data, investors can gain insight into public views and understanding to make decisions on individual companies. In addition, sentiment analysis can be used to examine many forms of data, such as social media posts and comments, customer feedback, and product reviews. Sentiment analysis is a key method of analysis for multiple types of data and provides potential investors with the ability to see a valuable consumer perspective.

Alternative data may seem like a unique and safe way to collect data. However, even in this form of data, there are certain drawbacks to it. According to experienced writer Alex Woodie, alternative data has many different benefits but also has significant challenges. For instance, it has security concerns as it collects data from sensitive information such as credit card transactions and GPS data. Furthermore, alternative data can be unreliable many times. This is because there may not be accuracy in the data itself, as the data comes from unregulated sources and may not uphold the collection standards that traditional data does. Most importantly, there can be a heavy amount of bias from this source of data, specifically regarding social media. It also has a lack of uniformity when it comes to collecting data because these types of data aren’t straightforward. As you can see, alternative data has many positives, but just like with everything, it also has negatives.

As investors increasingly turn to alternative data to gain an edge, it becomes crucial to understand the true impact of these data sources on market stability and forecasting. This research paper intends to dive into the potential benefits and challenges of alternative data, seeking to determine if these tools genuinely improve our understanding or unnecessarily introduce uncertainty over financial markets.

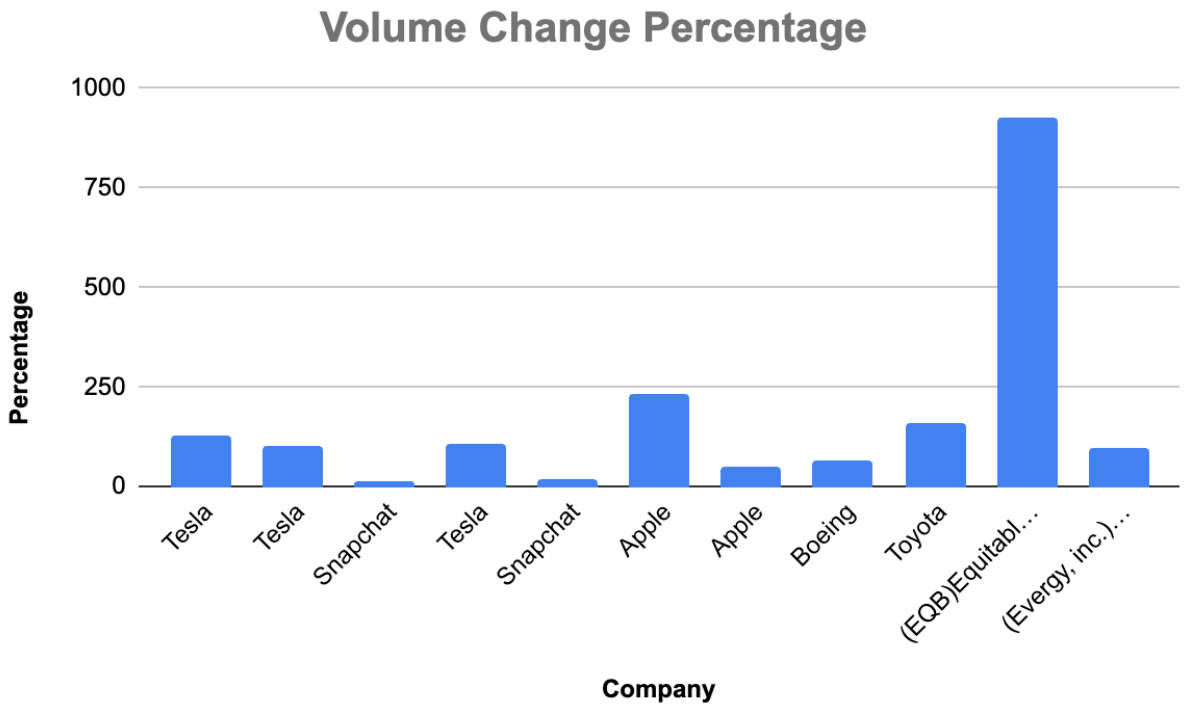
Data Collection

To further our understanding of the impact that alternative data can have on financial markets, we gathered a multitude of data from different sources to shed light on our research question. To gather our events, we used Twitter to find tweets specific to companies. We used tweets as our example of alternative data, as it is among the most prevalent forms of alternative data that are used. We would then see how these tweets affected the company's stock and its investors. Below is a summary table of each company's tweet we took, date, and the type of message tweeted (positive or negative connotation).

Company	Tesla	Tesla	Snap	Tesla	Snap	Apple	Apple	Boeing	Toyota	EQB	Evergy
Date	8/7/18	5/1/19	5/5/22	3/25/13	2/22/18	8/13/13	2/26/13	12/6/16	1/5/17	4/13/17	4/19/17
Tweet Connotation:	+	-	-	+	-	+	+	-	-	-	-

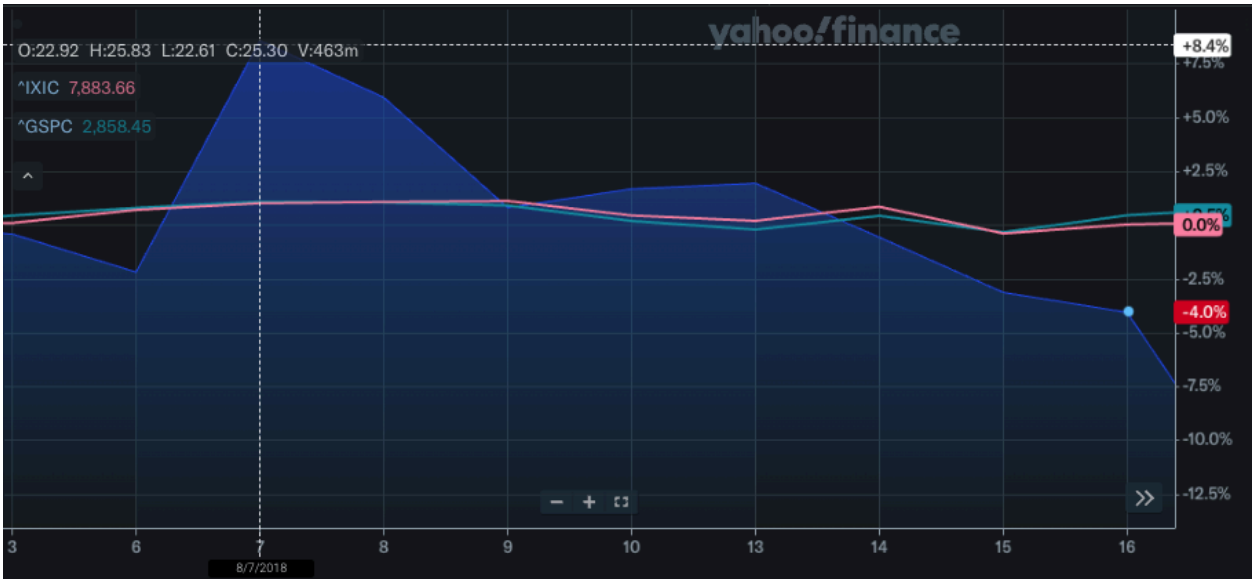
Pos(+)/Neg(-)											

After identifying the tweet, we examine how much the company's stock is affected by looking at how many shares are traded after the tweet is published. In other words, we focus on the volume of trading in the stock around the event. We record the initial volume two days before the tweet to avoid any news or interference with the prior volume. After the tweet is published, we record the volume on the day of the tweet. The volume percentage change is shown in the chart below.



The chart shows a clear volume increase for each company after the tweet. It shows that post-tweet, there was an increased amount of shares traded for each company. Our data shows

that there is a significant increase in volume percentage change for each event, ranging from 15% to nearly 660%. Each tweet had a different level of impact, but regardless, it still had a significant impact.



(Tesla 08/17/2018)

In addition, we recorded the average of daily volume for the 3 months prior to the tweet event. For example, if the tweet happened on 08/17/2017, then we took data from 05/01/2017 to 07/31/2017 to estimate the average daily volume over that time frame. These data were used to identify how many days after the event the volume was above the 3-month volume moving average. We counted how many days it was above the 3-month average volume in a 15-day trading period. This was used to display the full impact on a company and the duration of the volume impact from the event. For example, if the tweet is truly impactful to a company's stock value, then the data would be above the 3-month moving average for a significant period.

To effectively compare our volume data, we used different control variables for our data. Firstly, we used ETFs that are aligned specifically to the stock, which have a higher holding in that specific stock (Tesla and Evergy were not part of an ETF at the time of the tweet, so they do

not have a control variable as a specific ETF). Through this, we can see a more specific impact on how it affected the specific ETF. Additionally, we used the S&P 500 and Nasdaq for a more generalized control variable. We used these index funds to show the impact a specific tweet would have on a broader scale. The indices have essentially been included to show a comparison between the stock and its effect on the stock market as a whole. All three of these control variables are being used for both volume change percentage and days above the previous 3-month volume average.

Results and Analysis

The goal of our research paper was to investigate how sentiment on social media platforms such as Twitter could affect stock prices in real time. We charted the pricer actions of the companies that were impacted by it over several days before and following the posting of each tweet. We discovered particular tweets that correlated with significant changes in stock prices. In addition to monitoring the fluctuations in price, we also aimed to place these changes in the context of overall market performance. To do this, we placed it beside the price changes of the S&P 500 and the Nasdaq indices during the same period. This comparison helped us find the direct influence of the tweet by allowing us to assess whether the change in the price of a certain stock was unusual about the overall market.

One of these tweet-linked market events is shown in each chart that follows. For example, we examine the price performance of Tesla in the days following Elon Musk's tweet on May 1, 2020. This tweet stated that the "price of Tesla stock is too high." We utilized the same approach for all of the other companies in our dataset, such as Westar Energy, Equitable Group, Apple, Boeing, Toyota, Snapchat, and many more. We determined the change in the price of

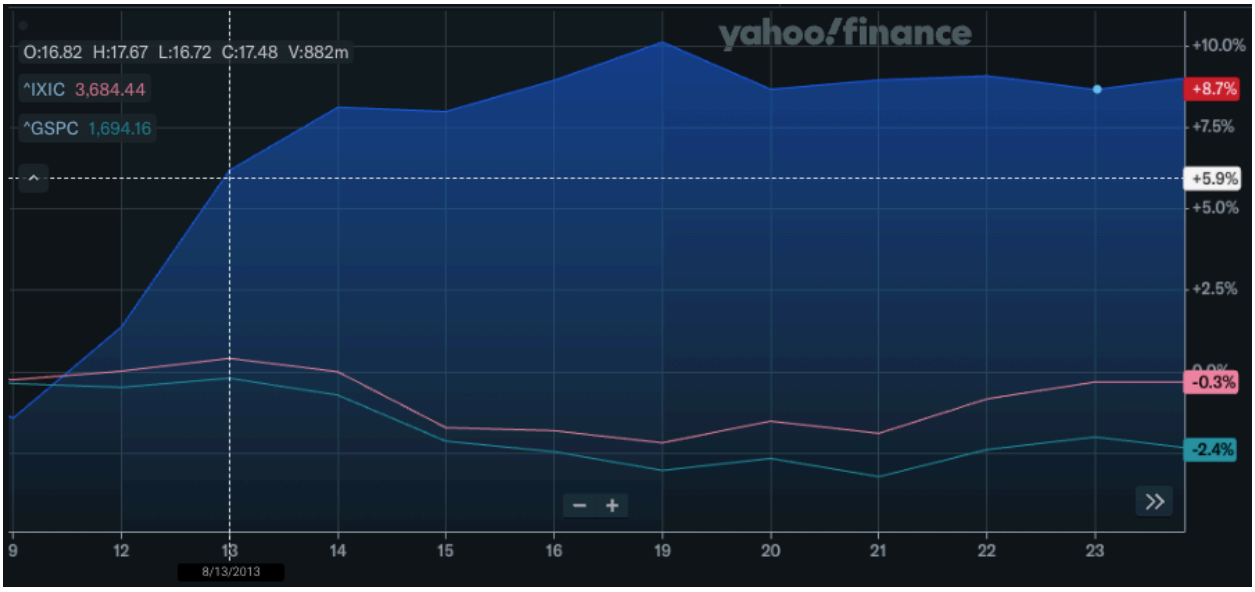
each stock and then compared it to the performance of the Nasdaq and S&P 500 indices over the same period. We were able to quantify “excess” movement that could be ascribed to the specific tweet using this method, regardless of the general market patterns.

(Tesla 03/25/2013)

(Equitable Group (now EQB Inc.) 04/13/2017)

(Westar Energy (now Evergy Inc.) 04/19/2017)

(Boeing 12/06/2016)

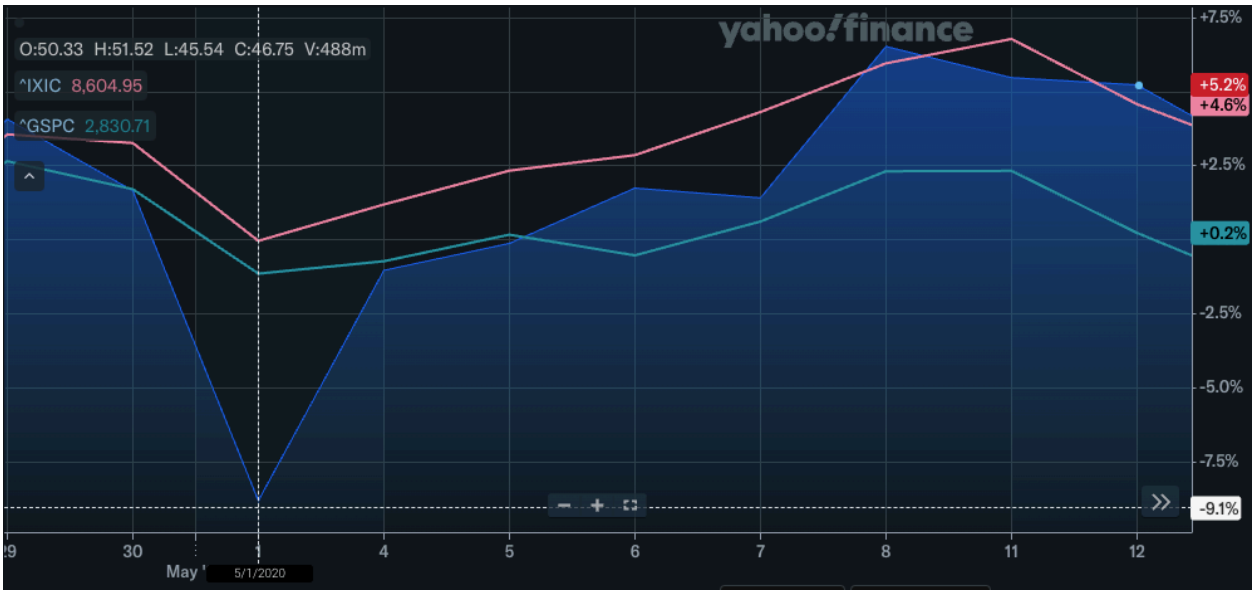


(Apple 08/13/2013)

(Apple 02/26/2013)

Due to the effects that they may bring, there are a few market events that are caused by tweets that may be especially important. The most notable tweet was Elon Musk's tweet on May 1, 2020, regarding Tesla stock being too high. In our graphic, it can be seen that the S&P 500

and Nasdaq indexes were relatively stable while Tesla's stock had a steep drop. It shows that a single tweet can override fundamentals and statistics, especially when it comes from a well-known person. Musk's statement had a greater impact on the market than any company data. It shows the influence of the internet in contemporary finance.



(Tesla 05/01/2020)

Another case of this was with Snapchat on February 22, 2018. Kylie Jenner posted a tweet criticizing the redesign of the app's interface. The result was a dramatic drop in Snapchat's stock, losing around 6% (or a billion dollars) in a single trading day. The dip occurred as market indices remained simple, which shows a selloff that was specific only to Snapchat's stock. The graph reflects the anomaly, as Snapchat's stock movement diverged from the S&P and Nasdaq. It reinforces the concept that influencer opinion can carry great weight, even though the company didn't publish or release any news themselves.

(Snap 02/22/2018)

(Snap 05/05/2022)

We also came across smaller but equally important incidents. An instance was Toyota on January 5, 2017, when President-elect Donald Trump attacked the company's plans to manufacture in Mexico. During the following trading days, Toyota's stock responded with a noticeable decline. However, the indices either showed slight increases or stayed steady.

(Toyota 1/5/2017)

Discussion

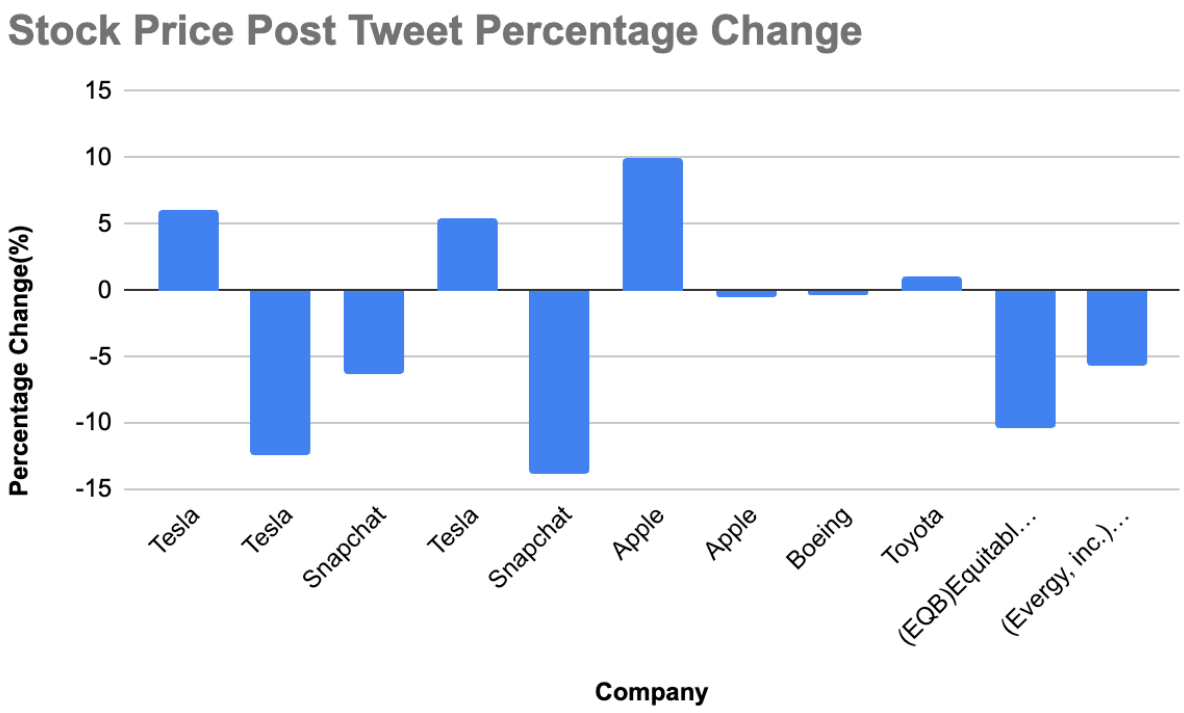
Our findings have immediate and interesting consequences. Initially, they verify that even in the absence of fresh financial data, emotional and psychological investor reactions—which are frequently brought on by the content or author of a tweet—can result in notable short-term volatility. This supports the idea that public personalities, such as CEOs, politicians, and celebrities, can actually affect investor behavior through social media alone and that markets are not entirely rational. This exposes investors to a new type of risk: sentiment-driven volatility that traditional financial measures are unable to measure.

Second, our data highlights the increasing importance of alternative data in contemporary investing, particularly real-time sentiment research. Conventional models based on fundamentals could miss sudden changes in sentiment that have an immediate effect on price movement. According to our research, investors who have access to tools that track social media trends are better able to predict and react to market movements. Investors can have this access through algorithms that identify powerful or emotionally charged tweets.

Lastly, we see a change in the dynamics of information. Although tweets are available to the general public, their market-moving potential can only be realized by those who have a way to assess and react to them effectively and instantaneously. Our graphs demonstrate that anomalies caused by tweets are statistically meaningful departures from index behavior rather than simply noise.

Conclusion

To further see the total impact on the actual stock of the company, we look at its price and the price percentage change. This research demonstrates how the tweet can impact the actual stock price and how it could impact investors. This is shown in the price percentage change chart below.



Comparing this to the summary table, you can see that almost all price changes (positive or negative) align with whether the tweet was a helpful or hurtful tweet against the company. Furthermore, we can see that there is a clear change in price, ranging from as small as 0.5% all the way to 15%.

Each graph and piece of data shows a different aspect of the impact of the tweet. As seen, specific tweets can have different levels of effect. However, the essential point is that their effect is noteworthy. From this, we can analyze and theorize its impacts so that we can even predict an alternative dataset's future influence.

The study's findings indicate a robust relationship between alternative data and stock market movements, particularly when examining social media tweets. The potential for social media to function as a stimulant for significant market activity is clear. An analysis of the provided graphs, which display price and volume changes in reaction to specific tweets,

demonstrates this. The volume change percentage chart indicates an increase in volume for several distinct companies after a tweet. The abrupt increase in volume suggests that tweets attract actual market involvement and significant response from investors. This appears to be independent of the tweet's topic or the business in question. The amount of the volume increases varies, indicating a wide variety of effects, with some tweets having a greater potential to attract investors than others.

In addition, the percentage change chart for stock prices after tweets shows that social media tweets also affect stock prices. The attitude expressed in a tweet often changes the direction of price movement. Usually, positive tweets imply that the price will increase, and negative tweets imply that the price will decrease. This supports the belief that investor sentiment, as it is communicated on social media, significantly influences the pricing of a company's stock. There needs to be a careful and thorough analysis undertaken to determine the meaningfulness of these market movements that are induced by tweets. The observed price and volume movements are often substantial from a statistical perspective. This indicates a strong correlation between stock market behavior and social media activity. The economic impact of these actions goes beyond just this, though.

There is no doubt that tweets can bring volatility as well as trading activity in the short term. This is why it's important for short-term traders to quickly profit from these swings. To determine if the influence of a tweet is short-lived or if it is long-lasting, our study makes sure to dive into the time component of these impacts on metrics such as the volume moving average. Looking at the data, it suggests that alternative data can have different types of impacts. They can have longer-lasting impacts on stock prices rather than just short-term spikes if a tweet's

influence is prolonged beyond the immediate aftermath. Additionally, a significant way to see the significance of these activities that are driven by tweets is their potential to be predicted.

The study goes more into the specifics of the tweet impacts and finds a few key traits that help make the difference between tweets that seem to be more or less meaningful. One important finding is that smaller and less-known businesses seem to be disproportionately affected by these tweets. These companies are extra vulnerable to the attention and sentiment that social media creates. This is due to things such as smaller initial trading volumes, as well as less coverage of the company. On the other hand, bigger and more well-known companies that tend to have higher trade volumes as well as closer observation by analysts might react to tweets in a more subdued manner.

Tweets from celebrities, powerful people, or those with large social media followings are usually more noteworthy. This is because their statements can quickly change the opinions of large numbers of investors and induce a lot of market movement. This phenomenon is shown by the example of tweets from public personalities, including Donald Trump and Elon Musk. This shows how the market can be affected by the statements made by powerful individuals. Future studies could examine elements like the specific content of a tweet, the amount of social media interactions, and the state of the market at the time the tweet was published.

In our research, we found a few distinct patterns. Firstly, the most impacted companies by the tweets are companies that are relatively smaller and less well-known. The companies that showed the most volume change were smaller. Secondly, tweets that came from celebrities and well-known people had a much bigger impact. For example, whenever President Donald Trump tweets about a company, that company is very heavily impacted.

Future researchers can use our research and scale it to a larger size. For example, we only have 11 tweets, but there are hundreds of thousands of tweets that impact companies. Future research can explore our ideas and concepts on a larger platform and focus on scalability.

Acknowledgment

We would like to express our deepest gratitude to Professor Todd Milbourn. His mentorship and guidance played an important role in shaping this research. His knowledge in the field of finance and data-driven decision-making helped refine our approach and gave us the foundation that we needed for our exploration into the intersection of alternative data and financial markets. Additionally, his encouragement and expertise inspired us to go beyond the conventional analysis and examine the impact of social media sentiment on market behavior.

References/Bibliography

A Practical Guide to Sentiment Analysis. SpringerLink, Springer International Publishing, 7 Feb. 2022, link.springer.com/book/10.1007/978-3-319-55394-8.

Chen, James. “Stock Analysis: Different Methods for Evaluating Stocks.” Investopedia, Investopedia, 11 Mar. 2023, www.investopedia.com/terms/s/stock-analysis.asp.

Dannemiller, Doug, et al. “Fueled by Better Information: Why Investment Management Should Embrace Alternative Data.” Deloitte Insights, Deloitte, 23 Nov. 2023, www2.deloitte.com/us/en/insights/industry/financial-services/financial-services-industry-predictions/2023/embracing-alternative-data-for-investment-decisions.html.

Gossett, Stephen. “What Is Alternative Data and Why Is It Changing Finance?” Built In, 12 July 2023, builtin.com/articles/alternative-data.

Hansen, Kristian Bondo. “Alternative Data and Sentiment Analysis: Prospecting Non-Standard Data in Machine Learning-Driven Finance - Kristian Bondo Hansen, Christian Borch, 2022.” Alternative Data and Sentiment Analysis: Prospecting Non-Standard Data in Machine Learning-Driven Finance, 20 Jan. 2022, journals.sagepub.com/doi/full/10.1177/20539517211070701.

Liberto, Daniel. “What Is Alternative Data?” Investopedia, Investopedia, 28 Feb. 2023, www.investopedia.com/what-is-alternative-data-6889002.

Marenzi, Octavio. “Alternative Data – a Closer Look at Credit and Debit Card Transaction Data.” Opimas, 14 Dec. 2022, www.opimas.com/research/883/detail/.

Technavio. “Alternative Data Market Analysis : US, China, Germany, UK, France - Size and Forecast 2024-2028.” Alternative Data Market Size, Share, Growth, Forecast Report 2028, Jan. 2024, www.technavio.com/report/alternative-data-market-industry-analysis.

Vallarino, Dr. Diego. “Gaining a Competitive Edge: The Role of Alternative Data Discovery in Quantitative Trading Field.” Medium, Medium, 9 June 2023, medium.com/@diegovallarino/gaining-a-competitive-edge-the-role-of-alternative-data-discovery-in-quantitative-trading-field-21ceced50fd7.

Woodie, Alex. “Alternative Data Isn’t so Alternative Anymore.” BigDATAwire, 22 Jan. 2024, www.datanami.com/2024/01/19/alternative-data-isnt-so-alternative-anymore/#:~:text=The%20most%20cited%20risk%20of,on%20the%20list%20at%2034%25.