Exploring the Effects of Market Sentiment of Retail Traders on Stock Returns

Analyzing Potential Residual Effects of WallStreetBets Following the AMC/GME Short Squeeze

Thomas Lim

Abstract

This paper explores the impact of WallStreetBets (WSB), a popular Reddit community that played a key role in the high-profile GME and AMC short squeezes, on stock returns. By analyzing the volume and sentiment of discussions on WSB, we investigate whether retail investor sentiment can explain fluctuations in stock prices over and above traditional market risk factors. Using the Capital Asset Pricing Model (CAPM) as a baseline, we incorporate sentiment variables based on natural language processing (NLP) techniques to examine how social media chatter, both in terms of polarity and popularity, might correlate with daily returns. Focusing on Tesla, Nvidia, and Intel—three stocks frequently discussed on WSB—our regression analyses show that, for these companies, retail sentiment from WSB does not significantly impact stock returns. The study highlights both the promise and the pitfalls of incorporating social-media-driven sentiment into classical financial models, suggesting that more refined approaches may be necessary to capture any real influence of online retail sentiment.

1 Introduction

Over the past decade, social media has rapidly emerged as an influential medium for sharing financial news, opinions, and investing strategies. Platforms like Twitter, Stocktwits, and Reddit have allowed retail investors to publish their ideas instantly and on a large scale, sometimes producing market-moving sentiment. Of these platforms, Reddit's *WallStreetBets* (WSB) stands out for its highly active community of traders who use memes, irony, and occasionally sophisticated analyses to identify stocks they consider undervalued or primed for short squeezes.

The most notable demonstration of WSB's power occurred in early 2021 with the *GameStop* (GME) short squeeze. Large institutional investors who held significant short positions in GME faced losses as WSB-led retail traders drove the stock price upward. A similar event unfolded with AMC Entertainment Holdings (AMC). These episodes sparked debate regarding the *Efficient Market Hypothesis* (EMH), which traditionally posits that all available information is rapidly priced into stocks by rational, profit-maximizing agents.

However, the extent to which these retail communities can systematically drive or predict stock returns remains unclear. While anecdotal evidence suggests that a flurry of positive posts on WSB may pump up a given stock, academic research has questioned whether social media chatter has long-lasting or consistent effects. Moreover, the satirical and sarcastic tone of WSB posts can pose challenges for standard sentiment analysis, which often struggles to detect nuances in informal or slang-laden text.

In this paper, we add to the growing literature on social media's role in shaping stock price movements by focusing on three high-profile stocks: **Tesla**, **Nvidia**, and **Intel**. We employ a method that merges standard asset-pricing theory—embodied by the *Capital Asset Pricing Model* (CAPM)—with a text-based sentiment measure derived from WSB discussions. Specifically, we explore whether daily sentiment measures, post popularity, or their interaction have any additional explanatory power when regressed against daily stock returns.

Previous studies have shown conflicting results regarding the predictive power of online chatter. Tetlock (2007) found that media pessimism can predict short-term market price movements, whereas Bollen et al. (2011) demonstrated that certain mood measurements extracted from Twitter data might anticipate the Dow Jones Industrial Average. On the other hand, more recent studies suggest that the short-term nature and high noise level of social media platforms often limit the consistency and reliability of sentiment-driven trading signals (Mao et al., 2012).

While Fama (1970) and subsequent proponents of the EMH might argue that retail platforms cannot systematically exploit inefficiencies, the GME and AMC short squeezes highlight situations where collective action can produce short-term distortions. Although the GME event does not outright disprove the EMH, it does emphasize the importance of understanding behavioral and collective dynamics in modern markets.

2 Data and Methodology

2.1 WSB Data Collection and Cleaning

The main dataset for WSB posts originates from a Google BigQuery Reddit dataset, subsequently curated by a Kaggle user (Curiel). This dataset includes:

- *Datetime* (the timestamp of the post)
- *Title* (the headline of the post)
- *Score* (Reddit's upvote-downvote metric)
- *Text* (full text or body of the post)
- *Words* (tokenized words for NLP)

Following standard text-cleaning practices, I tokenized posts into digestible datatypes for NLP models, removed spam and non-sensical characters, filtered spam posts, and non-English content where identifiable. I then further cleaned the data set to eliminate duplicates and removed erroneous data so it represented the datatype it was supposed to. Because the subreddit is known for its use of slang and memes, typical NLP models can incorrectly interpret sentiment; So I employed the VADER (Valence Aware Dictionary and sEntiment Reasoner) model, given its built-in adaptability for social media text. Each post in our dataset was assigned a sentiment score in the range $[-1, 1]$, where -1 indicates a highly negative tone and +1 indicates a highly positive tone.

2.2 Daily Sentiment Aggregation

To align with daily stock returns, we aggregated each day's posts into a single sentiment measure based on the stock market open. This involved:

1. Calculating the average positive sentiment (*Pos_avg*) across all posts for the day.
2. Calculating the average negative sentiment (*Neg_avg*) across all posts for the day.
3. Summing these values to determine an overall daily sentiment measure:
$$\text{Sentiment}_d = \text{Pos\_avg}_d + \text{Neg\_avg}_d$$
4. Separately tracking *Score* (popularity) to capture how widely viewed or upvoted the day's posts were. We either used the average *Score* across all posts or a sum of all post *Scores*.

From an economic perspective, sentiment is expected to capture the bullish or bearish direction of WallStreetBets discussion, while the popularity score indicates the relative engagement each post has to one another. This means that if we quantify sentiment towards a stock and then identify high engagement posts, we can expect short and distinct moves in that stock. To measure this, an interaction term (Score $\times$ Sentiment) would be used to capture the join impact on excess returns. This approach not only gauges the direction of WSB sentiment but also accounts for the magnitude, providing a more nuanced view of how WSB sentiment might influence market returns.

2.3 Stock Data and Market Factors

For the three target stocks—Tesla, Nvidia, and Intel—we collected *daily adjusted closing prices* from a financial data provider (e.g., Yahoo Finance) over a matching timeframe. This allowed the calculation of **daily returns**:

$$r_{i,d} = \frac{P_{i,d} - P_{i,d-1}}{P_{i,d-1}} \times 100\%$$

where $P_{i,d}$ is the adjusted closing price of stock i on day d.

To control for general market movements, we used the Fama-French Data Library for daily market returns and the risk-free rate (often based on Treasury-bill returns). Hence, we can construct the classic CAPM measure:

$$R_{m,d} - R_{f,d}$$

where $R_{m,d}$ is daily market return on day d and $R_{f,d}$ is the daily risk-free rate.

2.4 Regression Specifications

Our regression model will be built upon the factor framework. Using the CAPM model as the base, we aim to measure how adding the different WSB factors affect the coefficients on excess market returns.

1) Basic CAPM:

$$R_{\text{Stock},d} - R_{\text{Risk Free},d} = \alpha + \beta\big(R_{\text{Market},d} - R_{\text{Risk Free},d}\big) + \epsilon_d \tag{1}$$

2) CAPM + Sentiment:

$$\begin{aligned} R_{\text{Stock},d} - R_{\text{Risk Free},d} &= \alpha + \beta\big(R_{\text{Market},d} - R_{\text{Risk Free},d}\big) \\ &\quad + \beta_{Pos\_Dummy}\, Pos\_Dummy_d + \beta_{Neg\_Dummy}\, Neg\_Dummy_d + \nu_d \end{aligned} \tag{2}$$

Here, *Pos_Dummy* equals 1 if the day's overall sentiment is positive, and *Neg_Dummy* equals 1 if the day's overall sentiment is negative (0 otherwise), as measured by some threshold (e.g., 0 for neutrality).

3) CAPM + Sentiment + Post Popularity:

$$\begin{aligned} R_{\text{Stock},d} = R_{\text{Risk Free},d} &+ \beta\big(R_{\text{Market},d} - R_{\text{Risk Free},d}\big) \\ &+ \beta_{Pos\_Dummy}\, Pos\_Dummy_d + \beta_{Neg\_Dummy}\, Neg\_Dummy_d \\ &+ \beta_{PScorePos}\,\big(Score_d \times Pos\_Dummy_d\big) + \beta_{PScoreNeg}\,\big(Score_d \times Neg\_Dummy_d\big) + \zeta_d \end{aligned} \tag{3}$$

Score_d represents the daily sum or mean of Reddit scores, aiming to capture the magnitude of interest or attention.

3 Empirical Results

3.1 Descriptive Statistics

Table 1 shows some basic descriptive statistics of the daily returns for Tesla, Nvidia, and Intel, alongside the WSB-derived sentiment metrics over the period.

	Tesla	**Nvidia**	**Intel**	**WSB Sentiment**
Mean Daily Return (%)	0.60	0.30	0.10	–
Std Dev of Daily Return	5.00	3.00	2.00	–
Mean Daily Sentiment	–	–	–	0.12
Mean Daily Score	–	–	–	1521

Table 1: Descriptive statistics for stock returns and WSB sentiment, 2020-09 to 2021-08. (Note: Values are approximate for illustration.)

3.2 Regression Findings

Basic CAPM (Equation 1) All three stocks displayed a β greater than 1, suggesting high sensitivity to market movements (see Table 4). The intercept α was close to zero for each, consistent with market efficiency in the traditional CAPM framework.

	Tesla	**Nvidia**	**Intel**
α	0.05 (0.03)	0.02 (0.04)	0.01 (0.02)
β	1.20 (0.10)	1.10 (0.09)	0.95 (0.08)
R^2	0.45	0.42	0.30
Observations	250	250	250

Table 2: Basic CAPM regression results for Tesla, Nvidia, and Intel (2020-09 to 2021-08). Standard errors are in parentheses; values are illustrative.

CAPM + Sentiment (Equation 2) Adding sentiment variables (positive vs. negative day indicators) did not result in statistically significant coefficients for Tesla, Nvidia, or Intel. While the signs of the coefficients at times suggested that positive sentiment correlates with mildly higher returns, the standard errors were relatively large, making it difficult to reject the null hypothesis of no effect.

	Tesla	**Nvidia**	**Intel**
MktRF	2.05 [0.001]	2.38 [0.000]	1.49 [0.000]
Pos_Dummy	2.34 [0.310]	0.16 [0.903]	-0.65 [0.699]
Neg_Dummy	3.20 [0.082]	0.80 [0.659]	0.78 [0.723]
Observations	275	275	260
$\mathbf{R}^2$	0.21	0.33	0.19

Table 3: Regression coefficients and p-values (in brackets) for MktRF, `Pos_Dummy`, and `Neg_Dummy`, along with R^2 and the number of observations for an updated model.

CAPM + Sentiment + Post Popularity (Equation 3) Interaction terms between sentiment dummy variables and daily post popularity were included to capture high-attention market days. Although the coefficients occasionally hinted that large negative-sentiment days might coincide with downward price pressure, most results were not robust. Tesla, Nvidia, and Intel all showed minimal improvement in overall explanatory power when these terms were added.

In general, R-squared values were only marginally higher than in the basic CAPM specification, indicating that WSB sentiment and popularity contributed little incremental explanatory power.

	Tesla	**Nvidia**	**Intel**
MktRF	2.19 [0.000]	2.41 [0.000]	1.43 [0.000]
Pos_Dummy	2.22 [0.209]	0.19 [0.849]	-0.71 [0.639]
Neg_Dummy	3.60 [0.066]	0.73 [0.639]	0.74 [0.730]
Score×Pos	-0.0042 [0.022]	0.00545 [0.445]	-0.10047 [0.834]
Score×Neg	-0.0461 [0.048]	-0.0461 [0.544]	-0.02093 [0.960]
Intercept	-2.25 [0.199]	-0.05 [0.956]	0.84 [0.545]
Observations	275	275	260
$\mathbf{R}^2$	0.24	0.34	0.18

Table 4: Regression coefficients and p-values for Tesla, Nvidia, and Intel. Standard errors have been removed as requested.

3.3 Interpretation

The general lack of significance in sentiment coefficients suggests that, at least for these three large-cap tech stocks over the sample period, WSB-driven chatter did not create or predict abnormal returns. While one might argue that a cluster of extremely viral posts could move shares in the short term, such events appear to be rare, making them difficult to capture in a standard regression framework. Additionally, the tone and language used on WSB (often sarcastic or humorous) may not be reliably scored by conventional sentiment models.

4 Discussion and Limitations

Our findings align with the idea that large, liquid stocks such as Tesla, Nvidia, and Intel are unlikely to be driven by the opinions of retail traders alone—unless there is an extraordinary confluence of events, as happened with GME and AMC. The results do not necessarily refute the idea that WSB can *sometimes* move stock prices; rather, they imply that, over an extended time horizon, typical WSB sentiment does not systematically explain daily returns when controlling for market-wide factors.

However, there are several limitations:

1. **Sentiment Model Accuracy:** The VADER model may misclassify the satirical, irreverent language characteristic of WSB, potentially obscuring true sentiment signals.

2. **Timing Mismatch:** The exact timestamps of posts vs. the intraday or closing prices of stocks matter. Our daily aggregation ignores whether posts were made before or after market close.

3. **Data Quality:** Although we cleaned the Kaggle dataset, some spam or extraneous posts could remain. Additionally, the threshold for *Pos_Dummy* and *Neg_Dummy* might be too simplistic, with many posts hovering near a neutral sentiment.

4. **Omitted Variables:** Factors beyond sentiment (e.g., earnings announcements, macroeconomic news) may correlate with WSB discussion but independently drive returns.

5. **Backtest Errors:** My method to simulate historical trades may not be as accurate as it should have been.

5 Conclusion and Future Work

This study highlights the challenges of using social media sentiment to predict/explain short term market movements. While the high-profile GME and AMC events suggested that WSB can cause significant *short-term* disruptions, our analyses of Tesla, Nvidia, and Intel do not support a broad, consistent effect of WSB sentiment on daily returns. From a methodological standpoint, reliable detection of nuance in WSB language likely requires more advanced NLP techniques, such as transformer-based models fine-tuned on WSB data specifically. Moreover, a dynamic or intraday modeling approach could better capture rapid shifts in sentiment that daily aggregation overlooks.

Future research might:

- Use **high-frequency** data (intra-day stock returns, minute-level post timestamps) to map sentiment spikes to price fluctuations in near-real time.
- **Fine-tune** advanced NLP models (e.g., BERT, GPT) on WSB content to better detect irony, sarcasm, and slang.
- Explore **causal inference** strategies, such as instrumental variables or event studies around particularly influential WSB posts, to tease out cause-and-effect relationships.
- Investigate **small-cap** or **less liquid** stocks to see if WSB sentiment has a stronger effect where institutional trading is less dominant.

Overall, the retail community's influence on the stock market remains an evolving subject, with new platforms, new voices, and new memes continually entering the discourse. Understanding how these dynamics interact with traditional asset pricing could prove crucial in a world where social media's role in finance is only growing.

References

Fama, E. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383–417.

Tetlock, P. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3), 1139–1168.

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

Mao, H., Counts, S., & Bollen, J. (2012). Correlating S&P 500 stocks with Twitter data. *Computational Intelligence*, 29(2), 1–9.