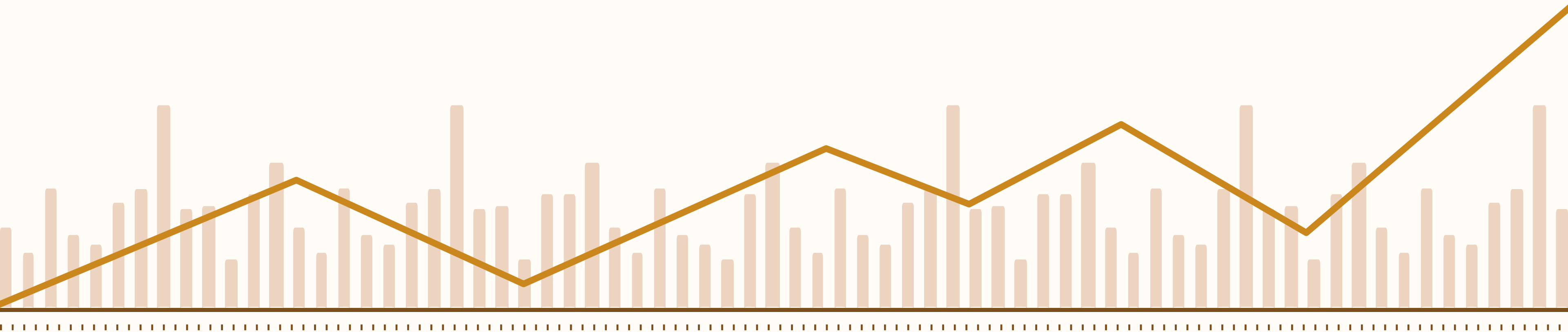


Finfluencer Effect on Retail Investment Behavior

Kaia Gao, Ruben Gomez, Hanning Zhou

Agenda

1. Introduction
2. Data Processing
3. Regression model
4. Analytical framework & Model diagnostics
5. Empirical results & Interpretation
6. Appendix

1

Introduction

Motivation

In 2021 retail traders (everyday people like me and you) gained notoriety by posting on social media about their stock trades. Retail traders with a large presence on social media were described as Finfluencers and had garnered recognition for their posts to risk it all for a change to “get rich💰!” and to take stocks “to the moon🚀.”

We wanted to know if finfluencers truly impact retail trading behavior.


CNBC MARKETS BUSINESS INVESTING TECH POLITICS VIDEO INVEST

PLAYBOOKS

Retail traders’ incredible support of the stock market continues with biggest options volume day ever

PUBLISHED THU, OCT 16 2025·9:52 AM EDT | UPDATED THU, OCT 16 2025·10:04 AM EDT

Sarah Min


Reuters World Business Markets Sustainability Legal Commentary Technology Investigations

GameStop jumps after cryptic post from 'Roaring Kitty' rekindles retail hype

By Manya Saini, Suzanne McGee and Saqib Iqbal Ahmed

December 5, 2024 1:11 PM PST · Updated December 5, 2024


Economic Cycles And Trends

How “Finfluencers” Can Create Risk for Your Company

by Vijay Govindarajan, Anup Srivastava and Chandrani Chatterjee

January 24, 2025

Primer: The Finfluencer

Noun

finfluencer (*plural* **finfluencers**)

1. (*informal*, *social media*, *marketing*, *neologism*) An influencer who gives advice on financial investments.

Popular finfluencer posts may include:
“Stonks only go up.”

Our Finfluencer focus

Stocktwits

What is it?
Social Media platform

→An **aggregate** sentiment of retail traders and investors who post about their thoughts on various stocks.

We calculate these thoughts as net Sentiment.

Research Questions

Using StockTwits (finfluencer) dataset of posts on social media we aimed to answer:

Does lagged retail sentiment influence next-day stock trading volume?

Additional considerations

- Does discussion buzz contain additional predictive power?
- Do these effects remain after controlling for:
  - Price movements
  - Volatility
  - Macroeconomic variables
- Are stock-specific characteristics important?
  - → Do fixed-effects required?

Hypothesis

💸 Retail Sentiment Effect

H1: **Lagged retail sentiment** effects next-day trading volume.

H0: **Lagged retail sentiment** has no effect on next-day trading volume.

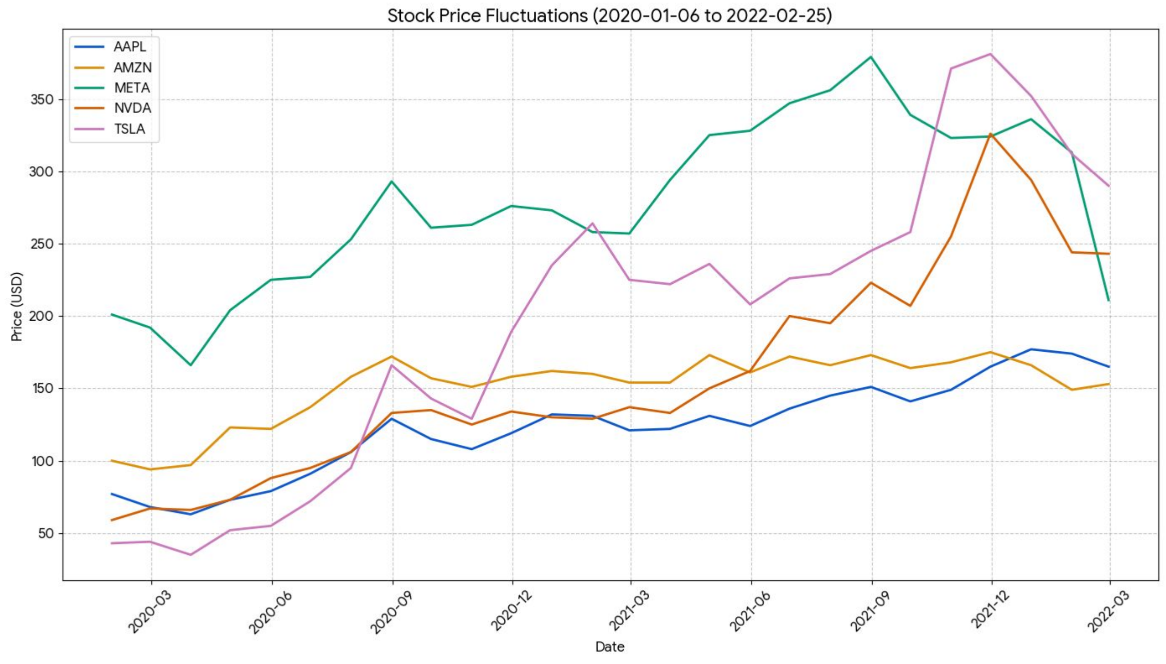
2

Data Processing

Data Preparation

Sample Collection

- Stocks: AAPL, AMZN, FB, NVDA, TSLA.
- Period: 2020-01-06 to 2022-02-25 (541 trading days per stock).
- Structure: Perfectly Balanced Panel (No missing values).



Dependent & Predictor

Variable Type	Variable Name	Symbol	Definition / Formula	Economic Proxy
Dependent	Log Volume	Y_{it}	$\log(Volume_t)$	Stock Trading Activity / Liquidity
Predictor *(our Finfluencer variable)*	**Net Sentiment**	$X_{1, it-1}$	**(Bull-Bear)/(Total)**	**Retail Optimism vs. Pessimism**

Controls

Variable Type	Variable Name	Symbol	Definition / Formula	Economic Proxy
Control (Social)	Log Buzz	$X_{2,\ it-1}$	log(NewsCount)	Retail Attention / Noise
Control (Market)	Volatility	$X_{3,\ it-1}$	(High-Low)/Open	Uncertainty / Opinion Divergence
Control (Market)	Abs Return	$X_{4,\ it-1}$	Yesterday's absolute return	Ret_{t-1}
Control (Macro)	Risk Appetite	$Z_{1,\ t}$	AAII Bull-Bear Spread	Market-wide Fear/Greed
Control (Macro)	Fed Rate	$Z_{2,\ t}$	Effective Fed Funds Rate	Cost of Capital
Control (Macro)	Unemployment	$Z_{3,\ t}$	U.S. Unemployment Rate	Economic Health / Labor Market
Control (Macro)	FOMC Dummy	$Z_{4,\ t}$	1 if FOMC day, else 0	Policy Event Shock

3

Regression Model

$$Log(Volumeit) = \alpha_i + \beta 1 Sentiment_{it-1} + \beta 2\, Log(Social_{it-1}) + \beta 3\, Market_{it-1} + \beta 4\, Macro_t + \varepsilon_{it}$$

- **Fixed Effects (empirical model):**
  - Absorb the baseline differences (e.g., controlling for the fact that FB naturally has more trading volume than NVDA).
- **Lagged Variables:**
  - Prevent 'Look-ahead Bias'
- **Clustered Standard Errors:**
  - Used to correct for serial correlation within stocks.

4

Analytical Framework & Robustness Checks

OUTLINE

Step 1

Pre-Model Screening

Step 2

Post-Estimation Diagnostics

Step 3

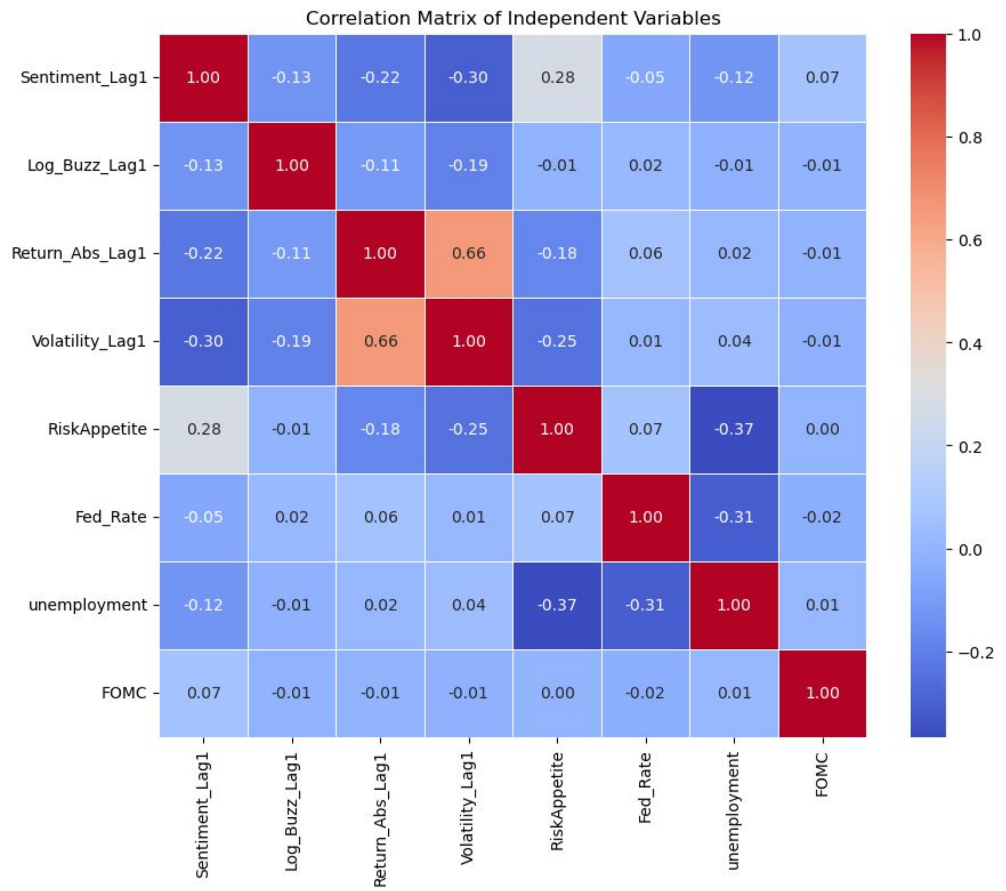
Baseline Model Comparison

Conclusion

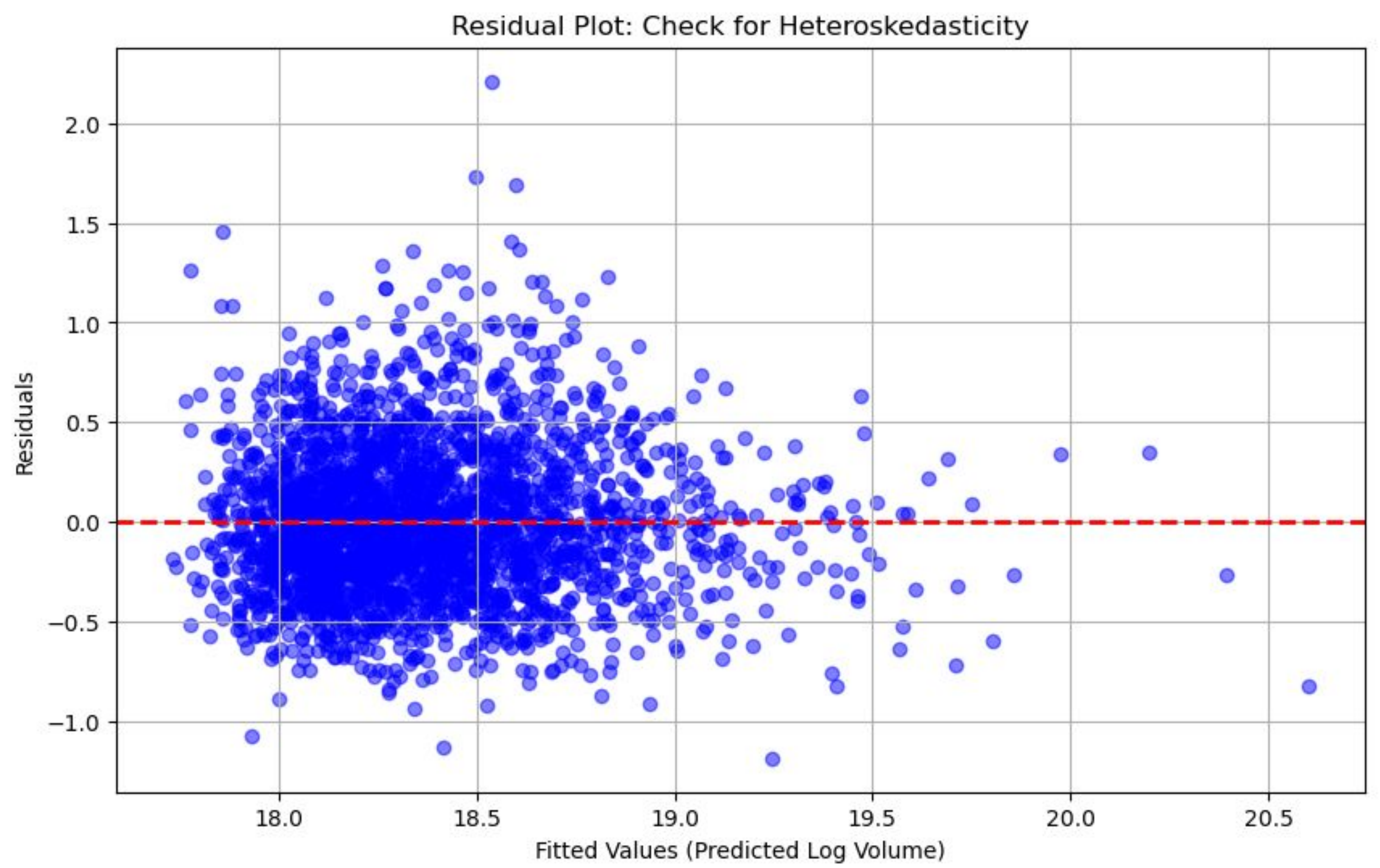
Correction Strategy

Step 1: Pre-Model Screening

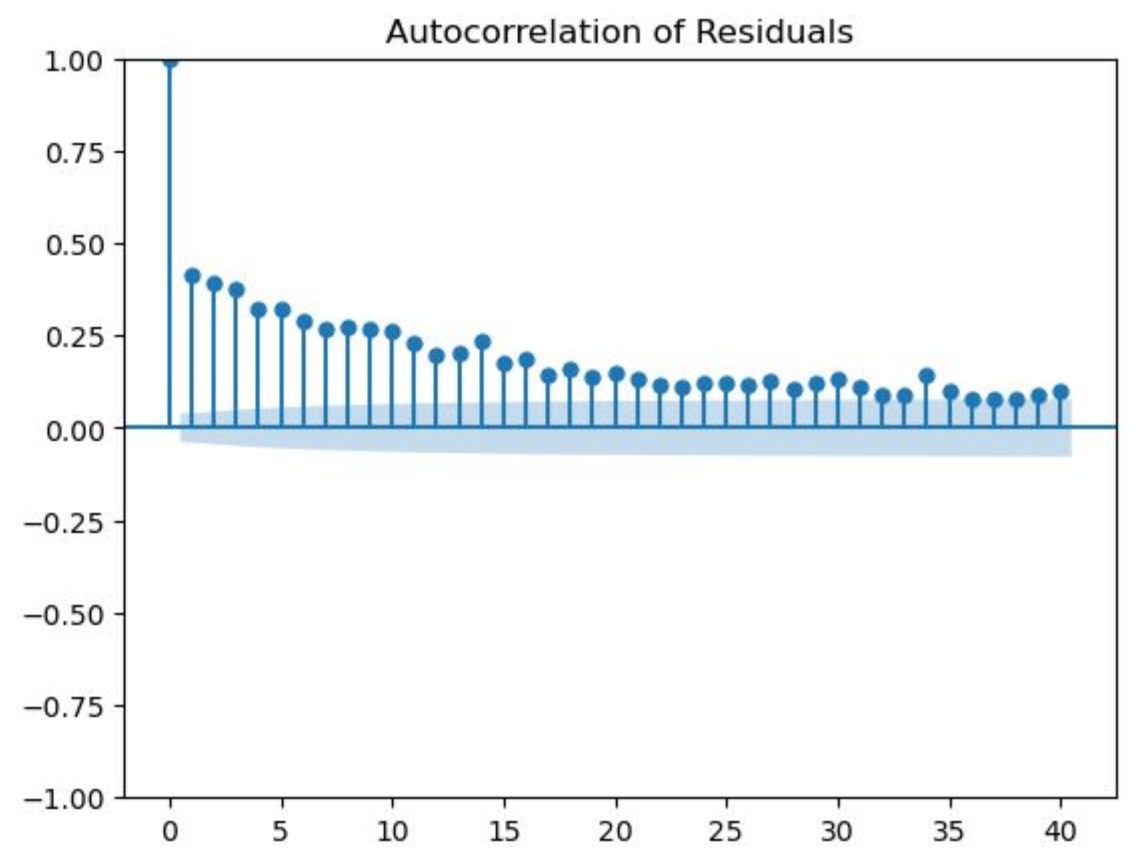
```
--- Variance Inflation Factor (VIF) ---
         Variable        VIF
0   Sentiment_Lag1   1.220449
1     Log_Buzz_Lag1   1.085016
2   Return_Abs_Lag1   1.774356
3   Volatility_Lag1   1.973035
4      RiskAppetite   1.280658
5          Fed_Rate   1.123966
6      unemployment   1.285483
7              FOMC   1.005469
8         intercept  52.226028
```



Step 2: Post-Estimation Diagnostics



Step 2: Post-Estimation Diagnostics



Step 3: Baseline Model Comparison

Model Comparison

	Baseline (Traditional)	Full (Behavioral)
Dep. Variable	Log_Volume	Log_Volume
Estimator	PanelOLS	PanelOLS
No. Observations	2705	2705
Cov. Est.	Clustered	Clustered
R-squared	0.4054	0.4091
R-Squared (Within)	0.4054	0.4091
R-Squared (Between)	-0.1552	-0.1720
R-Squared (Overall)	-0.0402	-0.0528
F-statistic	262.30	232.97
P-value (F-stat)	0.0000	0.0000

Model Comparison

	Baseline (Traditional)	Full (Behavioral)
Intercept	17.118	17.207
	(52.553)	(52.798)
Return_Abs_Lag1	3.3871	3.3548
	(6.2984)	(5.9118)
Volatility_Lag1	8.2079	7.7995
	(7.8249)	(7.5442)
Log_Buzz_Lag1	0.0807	0.0821
	(1.4526)	(1.4438)
RiskAppetite	-0.3918	-0.3552
	(-5.5411)	(-5.5660)
Fed_Rate	0.2878	0.2807
	(2.1405)	(2.0929)
unemployment	2.801e-05	2.746e-05
	(2.5013)	(2.4281)
FOMC	-0.0071	0.0021
	(-0.1528)	(0.0428)
Sentiment_Lag1		-0.1357
		(-1.9822)
Effects	Entity	Entity

T-stats reported in parentheses

5

Empirical Results & Interpretation

Panel OLS Estimation

Our model explains 41% of the day-to-day changes in trading volume within each stock

```
                         PanelOLS Estimation Summary
================================================================================
Dep. Variable:             Log_Volume   R-squared:                        0.4091
Estimator:                   PanelOLS   R-squared (Between):             -0.1720
No. Observations:                2705   R-squared (Within):               0.4091
Date:                Tue, Dec 02 2025   R-squared (Overall):             -0.0528
Time:                        22:22:44   Log-likelihood                   -1168.0
Cov. Estimator:             Clustered
                                        F-statistic:                      232.97
Entities:                           5   P-value                           0.0000
Avg Obs:                       541.00   Distribution:                  F(8,2692)
Min Obs:                       541.00
Max Obs:                       541.00   F-statistic (robust):         -6.571e+16
                                        P-value                           1.0000
Time periods:                     541   Distribution:                  F(8,2692)
Avg Obs:                       5.0000
Min Obs:                       5.0000
Max Obs:                       5.0000
```

Parameter estimate

Yesterday’s sentiment leads to lower trading volume the next day

Parameter Estimates

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Intercept	17.207	0.3259	52.798	0.0000	16.568	17.846
Sentiment_Lag1	-0.1357	0.0685	-1.9822	0.0476	-0.2700	-0.0015
Log_Buzz_Lag1	0.0821	0.0569	1.4438	0.1489	-0.0294	0.1936
Return_Abs_Lag1	3.3548	0.5675	5.9118	0.0000	2.2421	4.4676
Volatility_Lag1	7.7995	1.0339	7.5442	0.0000	5.7723	9.8268
RiskAppetite	-0.3552	0.0638	-5.5660	0.0000	-0.4803	-0.2300
Fed_Rate	0.2807	0.1341	2.0929	0.0365	0.0177	0.5437
unemployment	2.746e-05	1.131e-05	2.4281	0.0152	5.284e-06	4.963e-05
FOMC	0.0021	0.0490	0.0428	0.9659	-0.0940	0.0982

Select limitations

Temporal and Generalizability	Dataset January 2020 to February 2022	Extreme market volatility Covid-19	Messes with generalizability to non-pandemic years

Platform Bias and Demographic Representation	Sentiment data exclusively from StockTwits	Platform used by retail investors and day traders	User base may not be representative of the broader retail market

Thanks!

Time for Question, Feedback, Praise?

Interpretation of regression

- A. The "Negativity Bias" (Retail Fear)
    - Coefficient: Negative and Significant.
    - Finding: We observe that **lower Net Sentiment (more bearishness) predicts higher trading volume.**
    - Context: TSLA and AAPL, which have the lowest average sentiment (0.48), are highly liquid. This supports the hypothesis that **retail investors are more active during periods of disagreement or fear (panic selling) than during periods of consensus optimism.**
- B. The Role of Volatility
    - Finding: Intraday Range (Volatility_Lag1) is the single strongest predictor ($t > 10$).
    - Context: TSLA's high volatility (5.19%) naturally correlates with its high trading interest. **The model confirms that price action drives trading volume: when the intraday range widens, volume follows immediately.**
- C. Macro Sensitivity
    - Finding: The RiskAppetite variable (Macro Fear) remains significant.
    - Context: **Just as stock-specific fear drives volume, macro-level fear (low risk appetite) drives volume across all 5 tickets.**

Limitations

Temporal Scope and Generalizability:

- Our dataset spans from January 2020 to February 2022, a period characterized by extreme market volatility due to the COVID-19 pandemic. The sentiment-volume relationship observed during this crisis may not be fully generalizable to stable, non-crisis periods.

Platform Bias and Demographic Representation

- Our sentiment data is derived exclusively from StockTwits, a platform primarily used by retail investors and day traders. This user base may not be representative of the broader market, particularly institutional investors.

Causality and Endogeneity

- While the inclusion of lagged independent variables addresses simultaneity bias, the model remains susceptible to endogeneity arising from unobserved variables. Omitted factors, such as breaking news after market close, could fundamentally drive both sentiment and trading volume, leading to biased results."

Limitations continued

Sentiment Classification Accuracy

- Our sentiment analysis relies on user-generated tags ("Bullish" or "Bearish"), which introduces a potential self-selection bias.

Absence of Intraday Dynamics

- Because trading decisions and information flow often occur on an intraday basis, our use of daily aggregation may smooth over significant short-term volatility. Consequently, the model might miss rapid sentiment reversals within a single session, leading to an underestimation of the immediate impact of finfluencers."