IIMB-WP N0. 644/2021

WORKING PAPER NO: 644

The Role of Reddit in the GameStop Short Squeeze

Abhinav Anand
Assistant Professor
Finance & Accounting
Indian Institute of Management Bangalore
Bannerghatta Road, Bangalore – 5600 76
abhinav.anand@iimb.ac.in

Jalaj Pathak
Assistant Professor
Finance & Accounting
Great Lakes Institute of Management
Chennai
jalaj.p@greatlakes.edu.in

Year of Publication – December 2021

The Role of Reddit in the GameStop Short Squeeze

Abhinav Anand*
Jalaj Pathak†

Abstract

Adapting recent innovations in text analysis on microblogging platforms, we present evidence that the tone of discussions on the subreddit `r/wallstreetbets` (WSB) displayed significant predictive associations with intraday GameStop returns, volatility, bid-ask spreads as well as volumes. Most importantly, we show that the comment distribution on the subreddit obeyed a power law, and that it was a tiny minority of 462 most influential subredditors whose posts most impacted the GME stock returns and volatility.

Keywords: Short Squeeze, GameStop, WallStreetBets
JEL Classification: D91, G14, G40

*IIM Bangalore, Finance & Accounting
†Great Lakes Institute of Management, Finance & Accounting

1 Introduction

Financial markets are impacted by several non-fundamental variables, ranging from the disclosure tone (Del Gaudio et al. 2020) to the tone of presidential candidates (Marinč et al. 2021). In this study we examine the following questions: were Gamestop (GME) returns, volatility, spreads and volumes indeed influenced by the tone of discussions on Reddit? Our paper is among the first few to be able to quantify the extent of the influence of such concentrated retail attention and is able to identify whether the impact was attributable equally among all retailers or if it was led by a few prominent leaders.

Using recent innovations in financial and microblogging text analysis, we examine the impact of WSB tone and show that it is significantly and positively associated with future GME return, volatility, bid-ask spreads and volumes. Further, we show that it was in fact, a tiny minority of 462 most influential users who had a major, disproportionate impact on the vicissitudes of the GME stock.

2 Methodology

The data for WSB and for intraday GME variables are downloaded using the 'pushshift' API and TAQ. We only examine the posts on the subreddit which mention GME using the words—"GME", "gme", "GAMESTOP", "gamestop" and "game stop". We adapt recent innovations in text analysis and segregate threads into a collection of sentences (Anand et al. 2021). Polar words/phrases are identified with weights +1/-1; and tone-modifying valence shifters are identified around each polar word/phrase. A preponderance of slang and emojis leads us to consult a wide range of lexicons including the standard LM dictionary, Mohammad & Turney (2010), Jockers (2017), MPQA (https://mpqa.cs.pitt.edu/) & Sentiword (https://www.aclweb.org/anthology/L06-1225/), which perform with more than 85% precision for microblogging content (Ghiassi & Lee 2018). We also em-

ploy Kralj Novak et al. (2015) which assigns polarity to emojis (especially money-related), and manually assign weights to idiosyncratic slang popular on WSB like 'paper hands', 'YOLO'; and terms like 'call', 'put', 'short squeeze' which have been assigned 0 weight in financial dictionaries. A short example follows below:

> *"my trade would have been up about $130k from oct 9 to oct 10, but failure to take proper action only allowed me to realize about $90k in realized profits in one day's time :("*

The LM dictionary based unigram approach yields a tone of 0.058, while that using the modified methodology is -0.06 due to the valence shifters 'but' and 'only', and the 'sad emoji'.[1]

3 Results and Analysis

Figures 1 and 2 show evidence of strong daily comovement of GME returns and volatility with the WSB subreddit tone, especially in late January and early February 2021. We present descriptive statistics for variables in the appendix.

We examine the impact of WSB subreddit tone on 1-min interval GME returns as follows:

$$GME_t = a_0 + b_n Tone_{t-n} + d * Controls_t + u_t \tag{1}$$

The dependent variables are GME return, volatility, bid-ask spreads and volume respectively. n ranges from 0 to 10, and controls include number of comments & upvotes for each thread; the lag of GME 1-min return; as well as the day of the month and week dummies. All standard errors are heteroskedasticity and autocorrelation consistent.

Table 1 presents the impact of WSB tone on GME variables. WSB tone from 4, 9 and 10 minutes past have a positive, statistically significant impact

[1] See the appendix for more details.

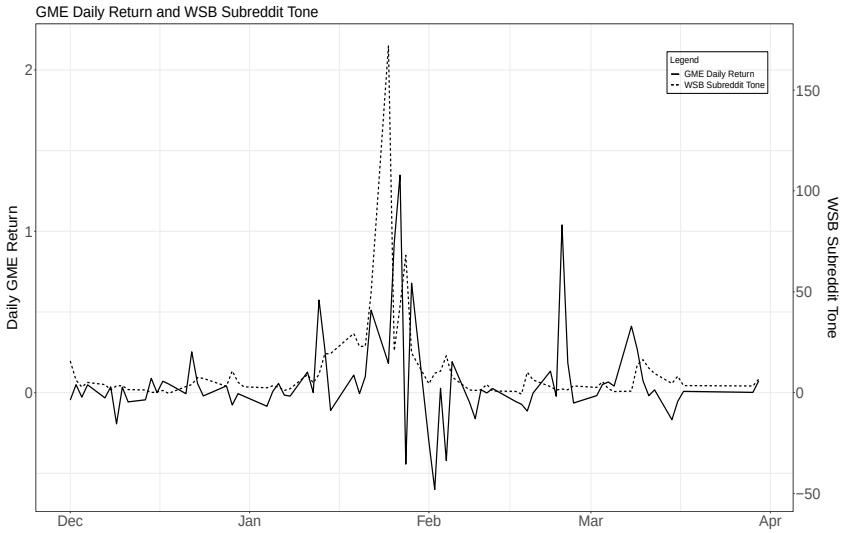

Figure 1: GME return (solid) and the WSB subreddit tone (dotted) comovement: Dec 2020–March 2021.

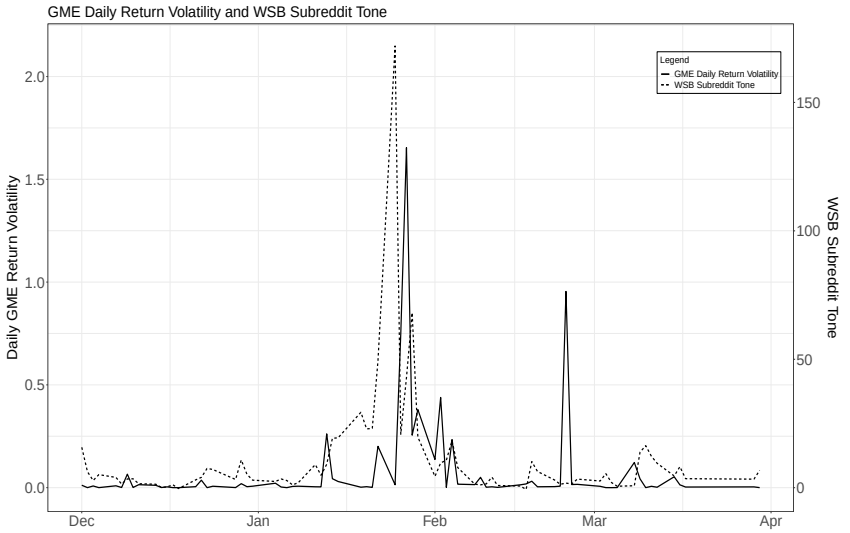

Figure 2: GME volatility (solid) and WSB subreddit tone (dotted) comovement: Dec 2020–March 2021.

on current 1-min GME returns. WSB tone from 5, 6, 9 and 10 minutes past have a positive, statistically significant impact on current 1-min GME volatilities.[2] Current 1-min WSB tone contemporaneously impacts bid-ask spreads in a positive, statistically significant manner; WSB tone from all 1-min intervals (except 3 and 4 minutes past) positively and significantly impact GME bid volumes; and WSB tones from 3, 4 and 5 minutes past positively and significantly impact current GME ask volumes.

4 Impact of 'influentials'

Pedersen (2021) suggests that securities markets exhibit large effects of thought leaders who display large follower counts. We identify as 'influentials', the top 5% most comment-provoking WSB redditors.[3] Our sample has 7702 thread posters, while the top 5% most comment-provoking authors—the influentials—are a group of 462 users who have prolifically contributed to other financial Reddit forums such as `'r/stockmarket'`, `'r/thetagang'`, `'r/Superstonk'`, `'r/WallstreetBreakers'` etc. Thus WSB influentials' posts garner high volume of comments on account of their strong association with online trading communities.

Table 2 presents influentials' impact on GME return and volatility. 4-min prior WSB tone from influentials' threads positively and statistically significantly impacts current GME 1-min returns; and 6-min prior WSB tone from influentials' posts impacts current 1-min GME volatility positively and significantly. The economic significance of influentials' posts are much higher than those for the benchmark results, as can be seen in table 3.[4]

Thus, the influentials' posts' tone disproportionately impacts aggregate tone which suggests that the comment distribution of WSB subreddit follows

[2]Calculated as in Andersen et al. (2007).

[3]We redefine influentials by their number of upvotes, and find no change in our results.

[4]We redefine influentials as the top 3, 6, and 9 percentiles of the most comment-provoking users and note that results remain similar.

a discrete power law.[5] We employ the methodology in Clauset et al. (2009) which yields a threshold estimate of $x_{\min} = 6.5$ and a scaling parameter of $\alpha = 2.68$. Figure 3 presents the fitted power law tail for the (log) comment distribution.

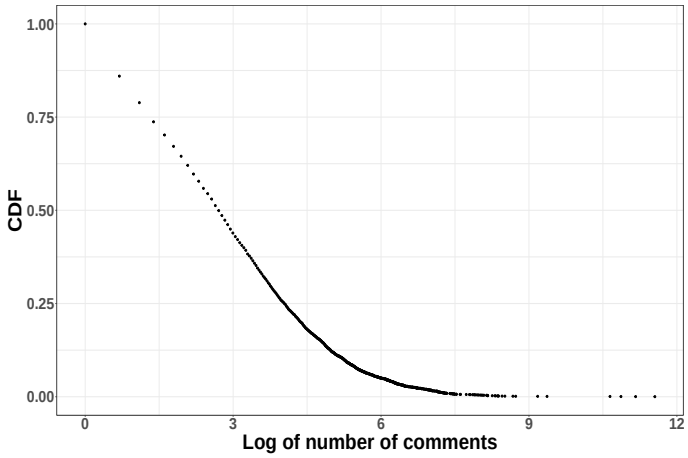

Figure 3: Cumulative distribution frequency (CDF) and the log of number of comments.

Further, we use bootstrapping with 500 simulations using MLE to estimate the scaling parameter. Figures 4 and 5 present the cumulative mean of α and $x_{\min}$ across the simulations.

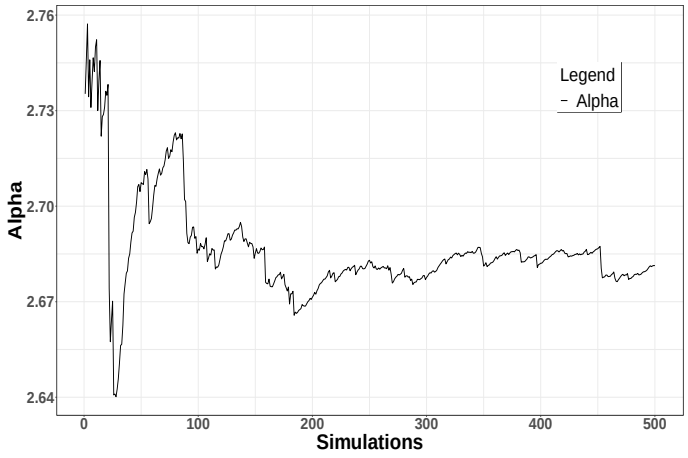

Figure 4: Cumulative sum of the scaling parameter (alpha) using 500 simulations.

[5]We rule out standard distributions like Normal, T and lognormal owing to poor fits.

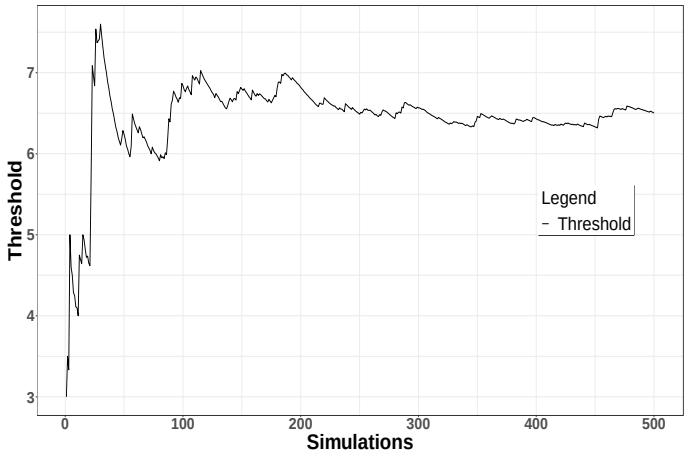

Figure 5: Cumulative sum of the threshold values using 500 simulations.

For our sample, the goodness-of-fit test yields a p-value of 0.48 and hence we cannot reject the null hypothesis that data are generated from a power law distribution.

For robustness, we run subsample regressions from Jan 15–Feb 15 2021 (most active subperiod), out-of-sample analysis for November 2020, as well as control for tone from the LM dictionary based unigram approach—all of which are presented in the appendix.

5 Conclusion

We examine the unprecedented short-squeeze episode triggered by mass-coordinated buying of GME stock and find that the tone of discussions on WSB subreddit have significant predictive association with the GME return, volatility, bid-ask spreads and volumes during December 2020–March 2021. We show that this impact is mostly due to a tiny minority of 462 most influential subredditors.

Table 1: Impact of intraday WSB tone on GME intraday variables

	$GME_t = a_0 + b_n Tone_{t-n} + d * Controls + u_t$										
	$n=0$	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$	$n=8$	$n=9$	$n=10$
Return	0.00043	−0.00024	−0.00084	0.00083	0.00152*	0.00101	0.00074	0.00075	0.00091	0.00207**	0.00177**
	(0.00058)	(0.00059)	(0.00068)	(0.00085)	(0.00077)	(0.00074)	(0.00070)	(0.00071)	(0.00080)	(0.00084)	(0.00076)
Volatility	−0.00002	0.00011	−0.00001	−0.00018	0.00018	0.00045**	0.00039**	0.00032	0.00031	0.00035*	0.00063***
	(0.00017)	(0.00016)	(0.00017)	(0.00020)	(0.00024)	(0.00020)	(0.00019)	(0.00022)	(0.00019)	(0.00020)	(0.00022)
Bid-Ask Spread	0.00068***	0.00034	0.00030	0.00032	0.00025	−0.00023	−0.00037	0.00007	−0.00007	−0.00014	−0.00002
	(0.00025)	(0.00026)	(0.00025)	(0.00023)	(0.00023)	(0.00022)	(0.00026)	(0.00022)	(0.00022)	(0.00021)	(0.00021)
Bid Volume	2,351.46*	1,722.16*	1,257.92*	850.81	607.10	732.88*	2,351.46*	2,351.46*	2,351.46*	2,351.46*	2,351.46*
	(1,217.63)	(926.69)	(731.78)	(539.47)	(412.73)	(406.07)	(1,217.63)	(1,217.63)	(1,217.63)	(1,217.63)	(1,217.63)
Ask Volume	378.36	448.57	655.91	691.21**	818.47***	830.53***	378.36	378.36	378.36	378.36	378.36
	(599.68)	(529.31)	(479.34)	(309.70)	(295.28)	(318.26)	(599.68)	(599.68)	(599.68)	(599.68)	(599.68)
Observations	2,428	2,428	2,428	2,428	2,428	2,428	2,423	2,422	2,421	2,420	2,419

This table presents the results from the regression of GME 1-min intraday variables on WSB intraday tone from December 2020 to March 2021. Standard errors are HAC and controls are as specified in equation (1).

Table 2: Impact of intraday WSB tone on GME (influentials)

	$GME_t = a_0 + b_n Tone_{t-n} + d * Controls + u_t$										
	$n = 0$	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$	$n = 7$	$n = 8$	$n = 9$	$n = 10$
Return	0.00344	0.00196	−0.00008	0.00366	0.00848**	−0.00423	0.00030	0.00141	−0.00063	−0.00156	0.00186
	(0.00244)	(0.00304)	(0.00214)	(0.00263)	(0.00351)	(0.00375)	(0.00301)	(0.00302)	(0.00235)	(0.00269)	(0.00199)
Volatility	−0.00060	0.00062	0.00037	0.00014	0.00053	0.00128**	−0.00067	−0.00005	0.00013	−0.00018	−0.00013
	(0.00040)	(0.00043)	(0.00054)	(0.00034)	(0.00049)	(0.00056)	(0.00065)	(0.00053)	(0.00051)	(0.00043)	(0.00049)
Bid-Ask Spread	−0.00111	0.00011	−0.00070	−0.00110	−0.00096	−0.00224**	−0.00076	0.00103	0.00195	0.00007	−0.00120
	(0.00096)	(0.00118)	(0.00077)	(0.00124)	(0.00092)	(0.00087)	(0.00081)	(0.00110)	(0.00190)	(0.00101)	(0.00128)
Bid Volume	3.66	−63.04	921.41	962.45	1,996.88*	−150.00	−567.02	−1,026.59**	−638.64	−994.87**	−101.74
	(406.50)	(569.03)	(798.14)	(764.36)	(1,029.72)	(489.04)	(436.41)	(440.00)	(489.75)	(482.20)	(483.75)
Ask Volume	−519.46	386.84	746.93	2,502.89	4,435.81**	1,959.53	−485.08	−1,456.82	−1,871.69***	−622.75	−30.21
	(665.64)	(684.73)	(768.12)	(1,824.11)	(2,055.10)	(1,841.15)	(718.93)	(908.76)	(647.39)	(548.90)	(708.19)
Observations	640	640	640	640	640	640	635	634	633	632	631

This table presents the results from the regression of GME 1-min intraday variables on WSB intraday tone from December 2020 to March 2021. Standard errors are HAC and controls are as in equation (1).

Table 3: Economic impact summary

Dependent variable	All users	Influentials
Return	0.06	0.28
Volatility	0.07	0.11
Bid-Ask Spread	0.06	0.09
Bid Volume	0.22	0.19
Ask Volume	0.09	0.32

This table presents the effect of a unit standard deviation movement in the independent variable (WSB tone) on a variety of dependent variables.

References

Anand, A., Basu, S., Pathak, J. & Thampy, A. (2021), 'Whose speeches impact European markets: ECBs' or the national central banks'?', *European Financial Management* . Forthcoming.

Andersen, T. G., Bollerslev, T. & Diebold, F. X. (2007), 'Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility', *Review of Economics and Statistics* **89**(4), 701–720.

Clauset, A., Shalizi, C. R. & Newman, M. E. (2009), 'Power-law distributions in empirical data', *SIAM Review* **51**(4), 661–703.

Del Gaudio, B. L., Megaravalli, A. V., Sampagnaro, G. & Verdoliva, V. (2020), 'Mandatory disclosure tone and bank risk-taking: Evidence from Europe', *Economics Letters* **186**, 108531.

Ghiassi, M. & Lee, S. (2018), 'A domain transferable lexicon set for Twitter sentiment analysis using a supervised machine learning approach', *Expert Systems with Applications* **106**, 197–216.

Jockers, M. (2017), *syuzhet: Extracts Sentiment and Sentiment-Derived Plot Arcs from Text*. R package version 1.0.6.

Kralj Novak, P., Smailović, J., Sluban, B. & Mozetič, I. (2015), 'Sentiment of emojis', *PloS one* **10**(12), e0144296.

Marinč, M., Massoud, N., Ichev, R. & Valentinčič, A. (2021), 'Presidential candidates linguistic tone: The impact on the financial markets', *Economics Letters* **204**, 109876.

Mohammad, S. & Turney, P. (2010), Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon, Association for Computational Linguistics, Los Angeles, CA.

Pedersen, L. H. (2021), 'Game on: Social networks and markets'. Working Paper.

6 Internet Appendix

6.1 Tone calculation methodology

Using the LM dictionary-based "bag-of-words" approach

$$\frac{(+1)[=\text{up}] + (-1)[=\text{failure}] + (+1)[=\text{profits}]}{17} = 0.058$$

Now, using the methodology borrowed from Anand et al. (2021), the tone is calculated as below:

Firstly, polar words/phrases are identified from the sentence followed by valence shifters around these polar words/phrases. Thus each sentence is divided into clusters with respect to polar words/phrases such as:

1. *my trade would have been **up** about $130k from oct 9 to oct 10, **but failure** to take proper action "sad face emoji"*

2. ***only** allowed me to realize about **$90k in realized **profits** in one day's time.*

Thus, the above sentence is divided into two clusters with ***but*** being a valence shifter (adversative conjunction) in the first cluster and ***only*** being a valence shifter (de-amplifier) in the second cluster.

The tone calculated is as follows:

$$(+1)[=\text{up}] + (-0.8)[=\text{but}] = +0.2$$

$$(-1)[=\text{failure}] = -1$$

$$(+1)[=\text{profits}] + (-0.8)[=\text{only}] = +0.2$$

$$(-0.55)[=\text{"sad face emoji"}] = -0.55$$

$$\frac{(+0.2)[=\text{first cluster}] + (-1)[=\text{second cluster}]}{19} +$$
$$\frac{(+0.2)[=\text{third cluster}] + (-0.550)[=\text{“sad face emoji”}]}{19} = -0.06$$

6.2 Descriptive Statistics

We produce the relevant descriptive statistics below:

Table 4: Descriptive Statistics

	Min	Mean	Median	Max	SD	IQR
GME Return	−0.124	0.0001	0.0001	0.081	0.012	0.008
GME Volatility	−0.029	0.0001	0.00009	0.020	0.003	0.002
GME Bid-ask Spread	−0.0001	0.004	0.003	0.060	0.004	0.003
GME Bid Volume	1	2362.7	670	64,859	5021.8	2807
GME Ask Volume	0	2350.2	722	36,721	4024.5	2439
WSB Tone	−1.665	0.172	0.120	2.186	0.378	0.293
Num Comment	0	136.5	9	104,178	2390.4	42
Num Upvotes	0	137.8	1	76,968	2011.4	3

Note: Summary statistics for all variables at the daily frequency.

6.3 Robustness

Table 5 presents the impact of GME tone on return for three robustness tests—subsample (Jan 15–Feb 15), out of sample (November) and adding LM tone as an additional control. We also get similar results for volatility, however we do not report them to ensure brevity.

We redefine influentials as the top 3, 6 and 9 percentiles of most comment-provoking users to gauge their posts' tone's impact on GME variables. We present results in table 6 only for return and volatility but note that other variables yield comparable results.

Table 5: Impact of intraday WSB tone on GME (robustness)

	$GME_t = a_0 + b_n Tone_{t-n} + d * Controls + u_t$										
	$n = 0$	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$	$n = 7$	$n = 8$	$n = 9$	$n = 10$
Subsample Analysis											
Return	0.00035	−0.00031	−0.00115	0.00079	0.00150*	0.00100	0.00096	0.00044	0.00104	0.00233**	0.00182**
	(0.00065)	(0.00066)	(0.00075)	(0.00093)	(0.00087)	(0.00080)	(0.00076)	(0.00079)	(0.00089)	(0.00093)	(0.00084)
Out of Sample Analysis											
Return	−0.00609	0.00216	0.00702	0.00302	0.00480	−0.00070	0.00404	0.00585**	0.00518*	0.00346**	0.00048
	(0.00980)	(0.00196)	(0.00490)	(0.00551)	(0.00712)	(0.00194)	(0.00517)	(0.00242)	(0.00305)	(0.00166)	(0.00485)
LM as additional control											
Return	0.00043	−0.00005	−0.00071	0.00104	0.00174**	0.00113	0.00092	0.00070	0.00117	0.00210**	0.00186**
	(0.00062)	(0.00064)	(0.00074)	(0.00093)	(0.00082)	(0.00079)	(0.00075)	(0.00074)	(0.00082)	(0.00084)	(0.00075)
Observations	1,804	1,804	1,804	1,804	1,804	1,804	1,804	1,804	1,804	1,804	1,804

This table presents the results from the regression of GME 1-min intraday returns on WSB intraday tone from December 2020 to March 2021. Standard errors are HAC and controls are as in equation (1).

Table 6: Impact of intraday WSB tone on GME (influentials)

	$GME_t = a_0 + b_n Tone_{t-n} + d * Controls + u_t$										
	$n=0$	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$	$n=8$	$n=9$	$n=10$
Influentials: Top 3 percentile of comment distribution											
Return	−0.00080	0.00426*	−0.00004	0.00173	0.00639**	−0.00261	0.00005	0.00223	0.00138	−0.00006	−0.00105
	(0.00213)	(0.00239)	(0.00236)	(0.00307)	(0.00262)	(0.00270)	(0.00230)	(0.00277)	(0.00189)	(0.00201)	(0.00319)
Volatility	−0.00060	−0.00031	0.00057	0.00013	0.00014	0.00088*	−0.00046	−0.00012	0.00027	0.00013	0.00007
	(0.00041)	(0.00040)	(0.00043)	(0.00037)	(0.00057)	(0.00053)	(0.00048)	(0.00043)	(0.00047)	(0.00036)	(0.00039)
Influentials: Top 6 percentile of comment distribution											
Return	0.00415*	0.00132	−0.00222	−0.00137	0.00921***	−0.00155	0.00123	0.00006	−0.00169	−0.00032	−0.00038
	(0.00235)	(0.00296)	(0.00306)	(0.00242)	(0.00351)	(0.00339)	(0.00300)	(0.00301)	(0.00210)	(0.00274)	(0.00217)
Volatility	−0.00038	0.00088*	0.00033	−0.00025	−0.00018	0.00143**	−0.00028	0.00014	−0.00008	−0.00035	−0.00005
	(0.00040)	(0.00045)	(0.00057)	(0.00054)	(0.00048)	(0.00061)	(0.00070)	(0.00055)	(0.00056)	(0.00043)	(0.00054)
Influentials: Top 9 percentile of comment distribution											
Return	0.00354*	0.00412*	−0.00138	0.00087	0.00185	0.00468**	−0.00035	−0.00350	−0.00184	−0.00420*	−0.00098
	(0.00215)	(0.00216)	(0.00282)	(0.00222)	(0.00235)	(0.00237)	(0.00292)	(0.00295)	(0.00298)	(0.00236)	(0.00229)
Volatility	0.00079	0.00106*	0.00095*	−0.00039	0.00005	0.00027	0.00084	−0.00010	−0.00100	−0.00067	−0.00103*
	(0.00056)	(0.00057)	(0.00054)	(0.00071)	(0.00060)	(0.00059)	(0.00062)	(0.00075)	(0.00076)	(0.00082)	(0.00060)

This table presents the results from the regression of GME 1-min intraday variables on WSB intraday tone from December 2020 to March 2021. Standard errors are HAC and controls are as in equation (1).