

Twitter and the Financial Markets

By Syeam Bin Abdullah
a1850041



How do Interactions on Public Forums on the Internet Affect Financial Markets?

- Can Twitter data be used as a proxy for predicting price movement in the stock market? -> Hypothesis: It can.





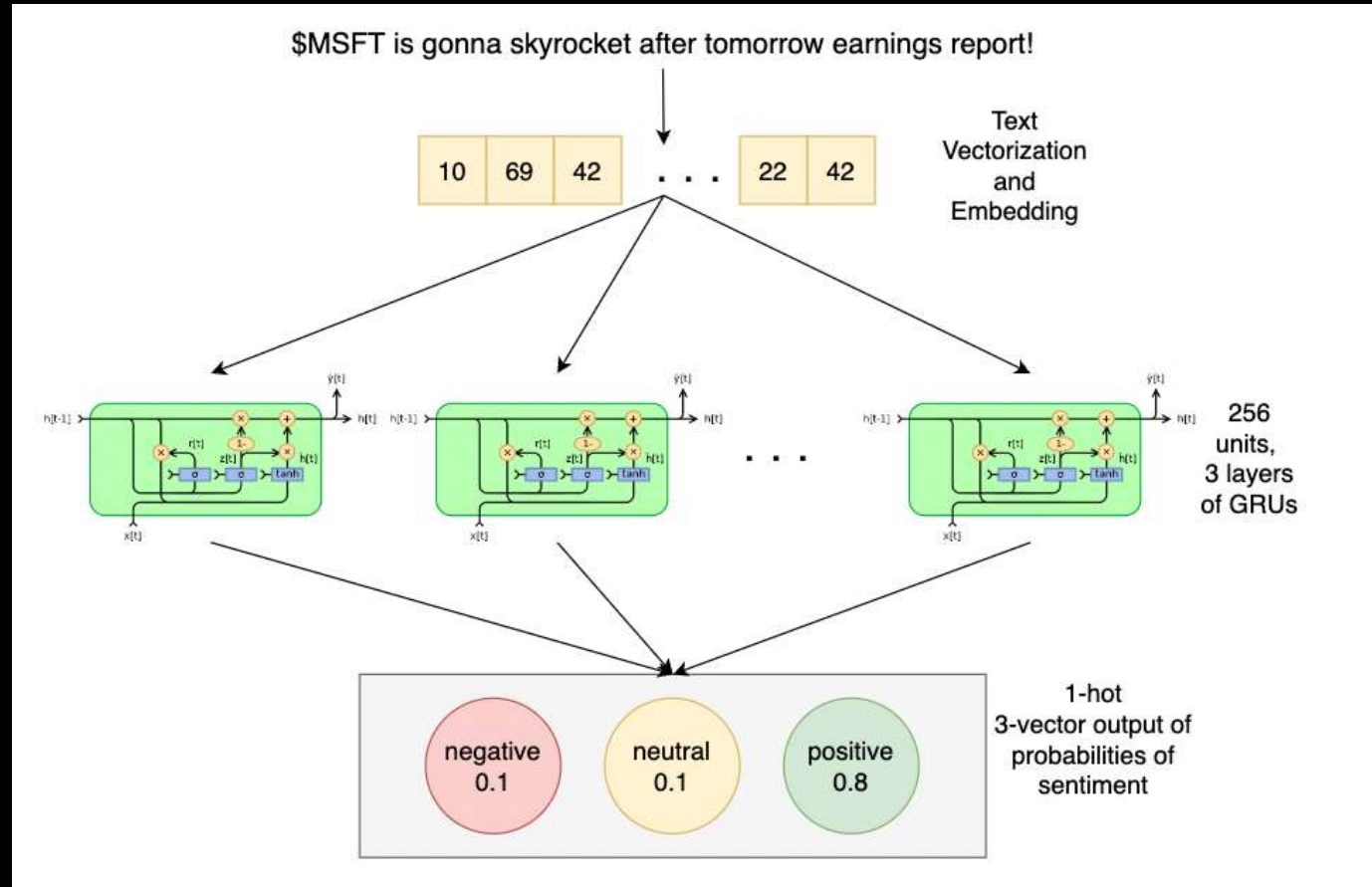
Why do this? How is this relevant?

- Seeking Alpha – crushing the market \$\$\$
- Further understand how social behavior affects the financial markets
 - what primary sources of information affect the way the markets move?
- Great study on behavioral economics in the real world.

Methods and Analysis

- Tweets from 2019 from Twitter were scraped (tweet passed through RNN – Recurrent Neural Network which spits out sentiment vectors). Python notebook left running on Kaggle.
- Maximum 5000 tweets/day from 28th Feb. to 17th Dec.
- ~150000 tweets scraped
- Daily stock close, open, high, and low price, percentage change and the trading volume for 15/30 stocks from DJIA – Yahoo Finance (using yfinance)
- Python with Pandas and matplotlib

$$P_d = \frac{t_d^+ - t_d^-}{t_d^+ + t_d^-}$$



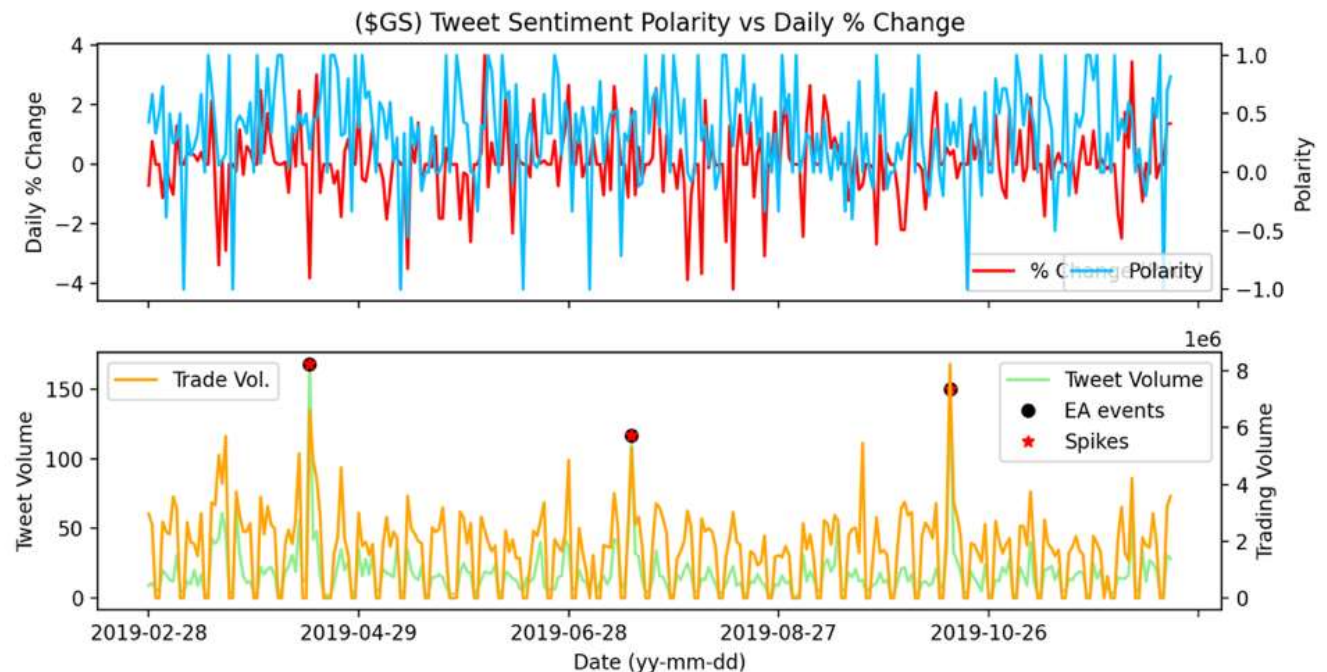
Ticker
AAPL
AXP
BA
CAT
CSCO
CVX
NKE
GS
IBM
INTC
KO
MMM
MSFT
V
XOM

Ticker: V										
	Date	trade volume	price %	abs. price %	tweets	sentiment	mode	polar	neutral	volume
0	2019-02-28	6250200.0	0.611304	0.611304	['content': '\$V', 'date': '2019-02-28', 'us...	0.500000	positive		0	4.0
1	2019-03-01	7016800.0	0.911447	0.911447	['content': '\$BAC \$C \$JPM banks should be cal...	0.500000	positive		0	13.0
2	2019-03-02	0.0	0.000000	0.000000	['content': 'For The Period Between 2008-03-1...	1.000000	positive		0	4.0
3	2019-03-03	0.0	0.000000	0.000000	['content': '\$KR Board Members Lost. They are...	0.000000	positive		0	2.0
4	2019-03-04	6979700.0	-1.010224	1.010224	['content': 'Popular: \$SPY, \$SPX, \$WDAY, \$ZNG...	0.777778	positive		0	12.0
..
288	2019-12-13	8558100.0	1.368809	1.368809	['content': '\$v+/w;6gQ8FjVDymf?', 'date': '20...	0.714286	positive		0	16.0
289	2019-12-14	0.0	0.000000	0.000000	['content': 'Mis TOP5 mejores compañías largo...	1.000000	positive		0	3.0
290	2019-12-15	0.0	0.000000	0.000000	['content': '\$v?67[93y9j6t8>qwJ}', 'date':...	0.000000	positive		0	4.0
291	2019-12-16	8249900.0	0.594166	0.594166	['content': 'Close last 1/2 \$V breakeven. no...	0.714286	positive		0	9.0
292	2019-12-17	10208700.0	-0.386595	0.386595	['content': 'U.S. stocks I currently hold.					

Methods and Analysis

- **No pearson corr. Between polarity and % change in \$.** Moderate correlation for volume and abs. return.
- Tweet vol. spike -> drastic price movement. Happened a lot during Earnings Announcements.
- Spike event detection (when tweet volume $> 3\sigma$).
- When $P_d < 0.2$, the event is classified as negative
- When $P_d \in [0.2, 0.6]$, the event is classified as neutral
- When $P_d > 0.6$, the event is classified as positive
- **Hypothesis: Tweet volume spike events cause drastic price movement, and sentiment polarity can help determine the direction of movement.**

Ticker	Pearson's Coefficient (3-day lag)		Granger-Cause	
	$p(T_d, R_d)$	$p(P_d, R_d)$	$T_d, R_d $	P_d, R_d
AAPL	0.54097055	0.13682904	0.07229056	0.59026017
AXP	0.38281115	-0.0136842	0.34207641	0.4862495
BA	0.53689419	0.21964648	0.00211434	0.00618897
CAT	0.36237887	-0.0037319	0.33528004	0.37728408
CSCO	0.63943819	0.05066817	0.0002447	0.09032849
CVX	0.53928483	0.07089701	0.27989501	0.96972585
NKE	0.40970968	0.11126223	0.2083516	0.33788428
GS	0.44757651	0.1413741	0.34391292	0.76900414
IBM	0.48040743	0.06589693	0.19323828	0.31509673
INTC	0.57046204	0.05734465	0.00267764	0.93779514
KO	0.62380723	0.10780703	0.14557281	0.76411895
MMM	0.69571285	0.03243169	0.58812186	0.40503183
MSFT	0.38762406	0.12691773	0.29717994	0.06553935
V	0.42416731	0.03335867	0.42826096	0.4062307
XOM	0.51330223	-0.0091482	0.05092554	0.39836057



Event Study

- Chosen method: Event Study from Financial Econometrics
- Looks at an event's effect on abnormal returns from securities over an event window (I chose 5 before to 10 days after event). Tests for statistical significance.
- Expected Normal Return and Variance, were calculated with data from 2018 (same time span as 2019 dataset).
- If test-statistic $\theta \gg 1$ then we must reject H_0 .

Abnormal Return: $AR_{i,d} = R_{i,d} - E[R_{i,d}]$

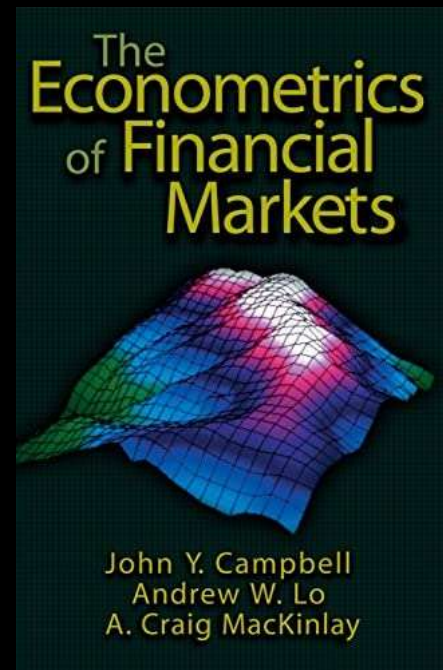
Averaged AR
(across firms): $\overline{AR}_\tau = \frac{1}{N} \sum_{i=1}^N AR_{i,\tau}$

Cumulative
Abnormal
Return: $CAR(\tau_1, \tau_2) = \sum_{\tau=\tau_1}^{\tau_2} \overline{AR}_\tau$

Variance
(across events): $var(CAR(\tau_1, \tau_2)) = \frac{1}{N^2} \sum_{i=1}^N (\tau_2 - \tau_1) \sigma_{\epsilon_i}^2$

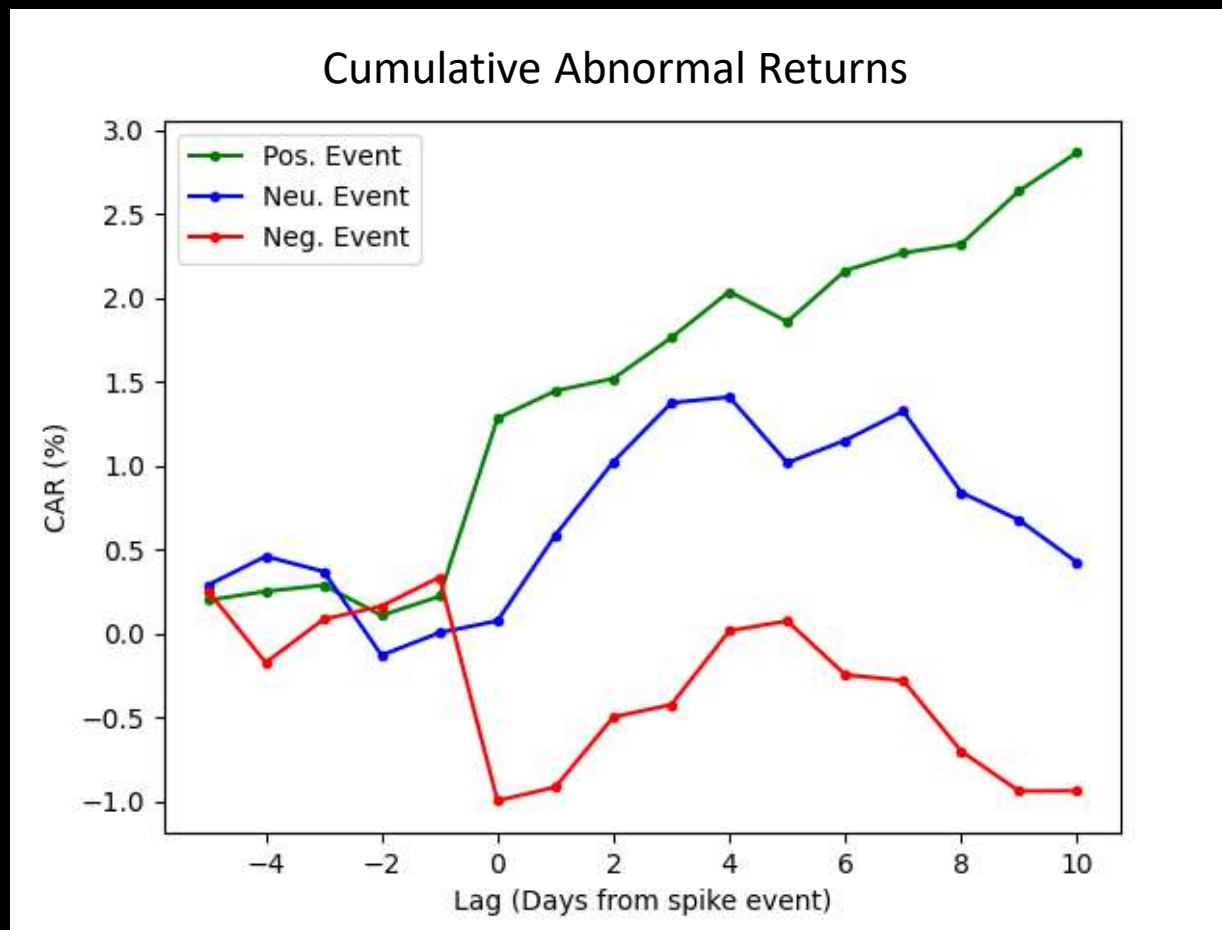
$$\frac{CAR(\tau_1, \tau_2)}{\sqrt{var(CAR(\tau_1, \tau_2))}} = \hat{\theta} \sim \mathcal{N}(0,1)$$

$$AR_{i,d} \sim \mathcal{N}(0, \sigma^2(AR_{i,d}))$$



Results

- The variance of the CAR (%) is $\sim 0.025\%$, whilst the abnormal return for a positive event on day 0 is $\sim 1.28\%$ ($\sim -1\%$ during negative events), which results in $\hat{\theta} \approx 8.01$ ($\hat{\theta} \approx -6.32$ during a negative event).
- Polarity of the tweets regarding the respective stock do appear indicate which direction the stock moves after and shortly after event.
- Therefore, the hypothesis is true!!!!
... STONKS!!!!





stonks