The Institute for Data Intensive Engineering and Science

JOHNS HOPKINS UNIVERSITY

IDIES Annual Symposium 2016

The ‘Sixth’ Factor – Social Media Factor Derived Directly from Tweet Sentiments

Presented by Jim Kyung-Soo Liew, Ph.D.
Assistant Professor in Finance

JOHNS HOPKINS CAREY BUSINESS SCHOOL

Where business is taught with humanity in mind.

Contents

- Motivation
- Research question
- Investigations
- Conclusions
- Extensions

Motivation:

(1) Social Media platforms (e.g. Facebook, Twitter, LinkedIn, YouTube, etc.) generate tremendous amounts of data

(2) Anecdotal evidence shows that this “alternative” data source has influences on markets

Research question:

Is there a link between social media information and security markets?

First investigation…failure.

- In 2010, obtained StockTwits data set and asked NYU Stern MBAs to manually read tweets and classify them as either "+" or "-", ~50 students x 100 tweets = total 5,000 label data.
- Shipped "labeled" data to west coast startup that "trained" their classification models and labeled the rest of this data set
- Took all the data to Baruch (Stat Arb) students and asked them to join with intra-day prices by stock and we analyzed the results

➢ Conclusion – Tweet sentiments at that time was a weak contrarian indicator, not statistically significant. Ugh!

Second investigation

- In 2015, obtained tweet sentiments data. Downloaded IPOs information from Bloomberg, e.g. day, offer price/performance (both contemporaneous and predictive relationships)
- Sample included 300 IPOs

➢ Conclusion -- Found statistically significant relationship between IPO returns (first-open to first-close) and contemporaneously tweet sentiments. Also, documented a predictive relationship between prior day's tweet sentiments and IPO returns (first-offer to first-open).

Could we strengthen the results by examining more events?

Follow-on research:

- Publicly listed companies must disclose financial results each quarter, "quarterly-earnings reports" 4x per year. Sample size +6k-8k
- Examine Post-Earnings-Announcement-Drift (PEAD) well-known phenomenon in finance with many known variables contributing to explanation of "drift"
- We tested tweet sentiments in the presence of traditional PEAD variables

➢ Conclusion – Documented link between tweet sentiment and PEAD, found that negative tweet sentiment securities with positive earnings surprise generated significant excess returns for up to 15 days

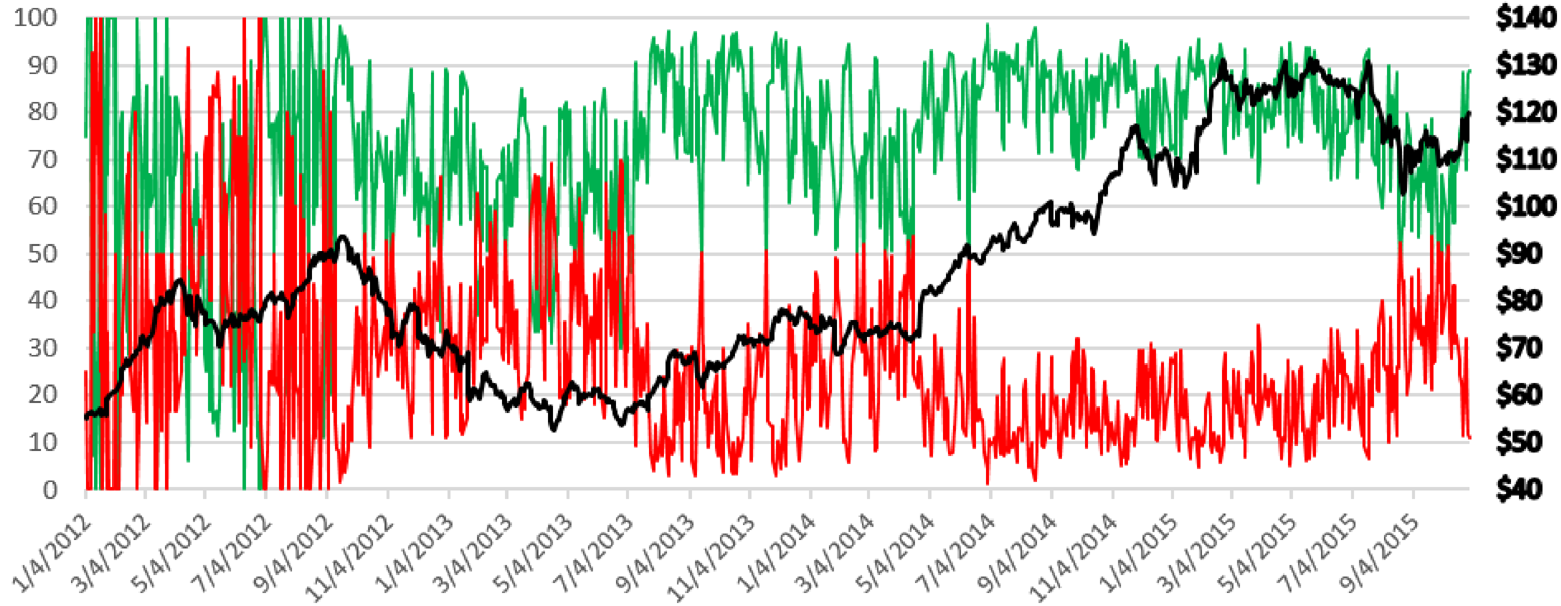
Third break-through investigation!

- “Twitter Sentiment and IPO Performance: A Cross-Sectional Examination” by Jim Liew and Garrett Wang, *Journal of Portfolio Management*, Summer 2016, Vol.42, No.4: pp. 129-135
- “Tweet Sentiments and Crowd-Sourced Earnings Estimates as Valuable Sources of Information Around Earnings Releases” by Jim Liew, Shenghan Guo, and Tongli Zhang, forthcoming *Journal of Alternative Investments* 2017.

- “The ‘Sixth’ Factor – Social Media Factor Derived Directly from Tweet Sentiments” by Jim Liew and Tamas Budavari, forthcoming *Journal of Portfolio Management* 2017. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2711825
- “Do Tweet Sentiments Still Predict the Stock Market?” by Jim Liew and Tamas Budavari, forthcoming *Alternative Investment Analyst Review* 2017. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2820269

“The ‘Sixth’ Factor – Social Media Factor Derived Directly from Tweet Sentiments”

Bullish % and Bearish % and AAPL Adjusted Close Price



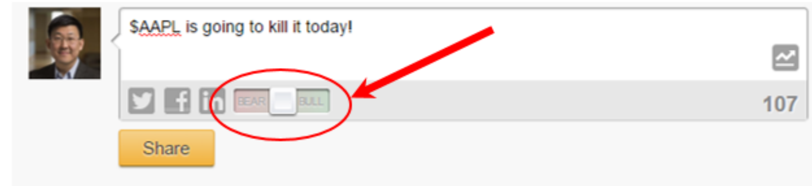
Bullish % and Bearish % constructed by counting the number of user-defined Bullish and Bearish tweets on "$AAPL" and dividing each by the total number.

Methodology:

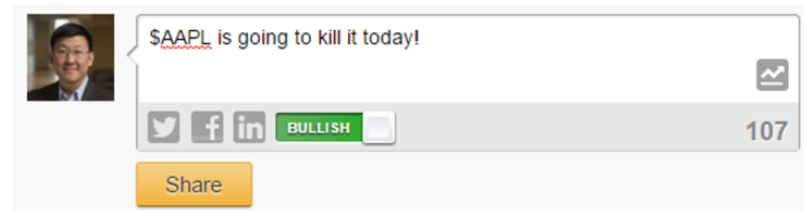
- Took sample of stocks with high tweet volumes
- Use the simplest measure of sentiment – User-defined “Bullish” or “Bearish” (Exhibit 1)

Exhibit 1

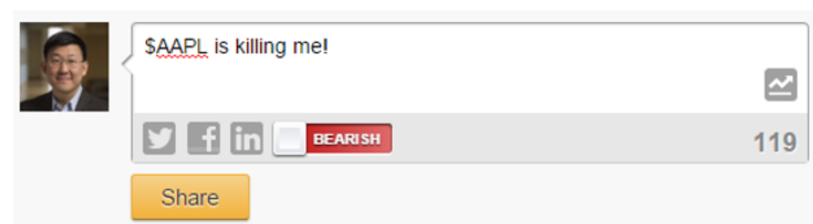
Panel A: StockTwits Sentiment Interface Sliding Bar



Panel B: Tweet Identified as “BULLISH”



Panel C: Tweet Identified as “BEARISH”



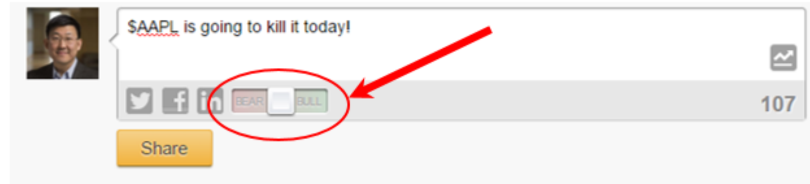
Methodology:

- Took sample of stocks with high tweet volumes
- Use the simplest measure of sentiment – User-defined "Bullish" or "Bearish" (Exhibit 1)
- Construct a time-series based on all user-labeled tweets. Tweets intra-day time-stamps aggregated to daily frequency
- Each stock has it's own Social Media Factor (idiosyncratic) and run the following regression below

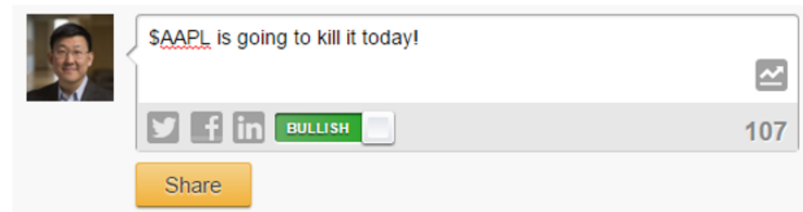
➢ If t-statistics are above 2.0 then it's statistically significant!

Exhibit 1

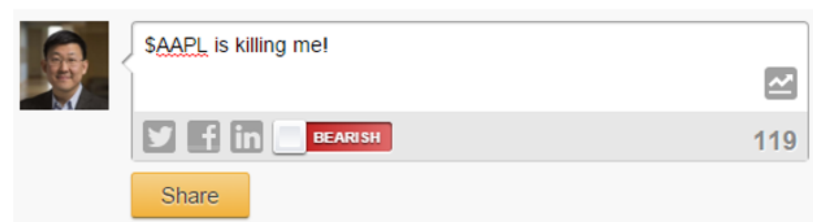
Panel A: StockTwits Sentiment Interface Sliding Bar



Panel B: Tweet Identified as "BULLISH"



Panel C: Tweet Identified as "BEARISH"



$$R_{i,t} - R_{f,t} = \alpha + \beta * (R_{m,t} - R_{f,t}) + s * (SMB_t) + h * (HML_t)$$
$$+ r * (RMW_t) + c * (CMA_t) + p * (SMF_{i,t}) + \varepsilon_{i,t}$$

Looking for t-stats above 2.0 “t-stat(p)”

Panel B: t-Statistics

Tickers	Begin Date	End Date	Nos.	t-stat(a)	t-stat(b)	t-stat(s)	t-stat(h)	t-stat(r)	t-stat(c)	t-stat(p)
AAPL	2013-01-03	2015-10-29	712	-0.17	13.48	0.11	4.10	8.59	-10.84	
AAPL	2013-01-03	2015-10-29	712	-9.70	12.58	0.06	4.22	7.90	-9.69	
FB	2013-01-03	2015-10-29	712	1.65	9.67	-0.57	-0.60	-4.05	-6.56	
FB	2013-01-03	2015-10-29	712	-6.96	8.88	-0.26	-0.35	-4.12	-5.78	
NFLX	2013-01-03	2015-10-29	712	2.19	7.23	0.95	-0.27	-0.94	-1.38	
NFLX	2013-01-03	2015-10-29	712	-9.10	6.28	0.50	-0.85	-1.01	-0.27	
YHOO	2013-01-03	2015-10-29	712	0.29	14.22	0.21	-2.48	-2.31	-1.37	
YHOO	2013-01-03	2015-10-29	712	-4.86	14.33	-0.05	-2.64	-2.70	-1.11	
AMZN	2013-01-03	2015-10-29	712	0.74	14.03	-1.41	0.22	-0.17	-6.78	
AMZN	2013-01-03	2015-10-29	712	-7.00	13.05	-1.71	0.30	-0.38	-6.11	
MU	2013-01-03	2015-10-29	709	0.75	13.53	1.57	-0.68	-1.57	-0.63	
MU	2013-01-03	2015-10-29	709	-4.44	13.45	1.34	-0.79	-1.73	-0.58	
BAC	2013-01-03	2015-10-29	712	0.23	23.62	-0.80	10.46	-7.76	-3.71	
BAC	2013-01-03	2015-10-29	712	-5.87	23.08	-0.65	10.16	-7.35	-3.74	
DIS	2013-01-03	2015-10-29	711	1.75	24.01	-2.34	-2.85	1.11	1.33	
DIS	2013-01-03	2015-10-29	711	-0.95	23.93	-2.32	-2.86	1.05	1.41	
PCLN	2013-01-03	2015-10-29	712	0.90	18.35	1.30	-1.04	-0.10	-4.91	
PCLN	2013-01-03	2015-10-29	712	-7.35	16.86	1.39	-1.36	-0.02	-3.94	
FSLR	2013-01-03	2015-10-29	712	0.33	9.19	2.50	4.48	-1.47	-3.81	
FSLR	2013-01-03	2015-10-29	712	-3.74	8.81	2.09	4.30	-1.34	-3.58	
MSFT	2013-01-03	2015-10-29	712	0.49	18.39	-2.67	2.98	7.33	-7.36	
MSFT	2013-01-03	2015-10-29	712	-2.43	18.02	-2.57	2.88	7.25	-7.19	
GOOG	2014-03-28	2015-10-29	402	-0.43	12.65	-1.37	2.00	1.83	-7.60	
GOOG	2014-03-28	2015-10-29	402	-2.80	11.86	-1.16	1.85	1.96	-7.20	
AAL	2013-12-09	2015-10-29	477	1.06	9.64	-0.61	-3.15	0.58	-0.15	
AAL	2013-12-09	2015-10-29	477	-4.49	9.44	-0.82	-3.51	0.72	0.09	
CHK	2013-01-03	2015-10-29	707	-1.19	10.86	2.88	8.89	-1.97	-1.82	
CHK	2013-01-03	2015-10-29	707	-3.19	10.94	2.95	8.68	-1.85	-1.58	
RIG	2013-01-03	2015-10-29	695	-1.44	9.84	2.61	8.79	-1.73	-0.46	
RIG	2013-01-03	2015-10-29	695	-3.55	9.75	2.55	8.38	-1.74	-0.27	
Mean			674	-2.31	13.67	0.12	1.98	-0.20	-3.52	
Median			712	-1.31	12.85	[illegible]	[illegible]	[illegible]	-3.65	

Looking for t-stats above 2.0 “t-stat(p)”

Panel B: t-Statistics

Tickers	Begin Date	End Date	Nos.	t-stat(a)	t-stat(b)	t-stat(s)	t-stat(h)	t-stat(r)	t-stat(c)	t-stat(p)
AAPL	2013-01-03	2015-10-29	712	-0.17	13.48	0.11	4.10	8.59	-10.84	
AAPL	2013-01-03	2015-10-29	712	-9.70	12.58	0.06	4.22	7.90	-9.69	9.82
FB	2013-01-03	2015-10-29	712	1.65	9.67	-0.57	-0.60	-4.05	-6.56	
FB	2013-01-03	2015-10-29	712	-6.96	8.88	-0.26	-0.35	-4.12	-5.78	7.36
NFLX	2013-01-03	2015-10-29	712	2.19	7.23	0.95	-0.27	-0.94	-1.38	
NFLX	2013-01-03	2015-10-29	712	-9.10	6.28	0.50	-0.85	-1.01	-0.27	10.58
YHOO	2013-01-03	2015-10-29	712	0.29	14.22	0.21	-2.48	-2.31	-1.37	
YHOO	2013-01-03	2015-10-29	712	-4.86	14.33	-0.05	-2.64	-2.70	-1.11	5.20
AMZN	2013-01-03	2015-10-29	712	0.74	14.03	-1.41	0.22	-0.17	-6.78	
AMZN	2013-01-03	2015-10-29	712	-7.00	13.05	-1.71	0.30	-0.38	-6.11	7.82
MU	2013-01-03	2015-10-29	709	0.75	13.53	1.57	-0.68	-1.57	-0.63	
MU	2013-01-03	2015-10-29	709	-4.44	13.45	1.34	-0.79	-1.73	-0.58	4.95
BAC	2013-01-03	2015-10-29	712	0.23	23.62	-0.80	10.46	-7.76	-3.71	
BAC	2013-01-03	2015-10-29	712	-5.87	23.08	-0.65	10.16	-7.35	-3.74	6.06
DIS	2013-01-03	2015-10-29	711	1.75	24.01	-2.34	-2.85	1.11	1.33	
DIS	2013-01-03	2015-10-29	711	-0.95	23.93	-2.32	-2.86	1.05	1.41	1.85
PCLN	2013-01-03	2015-10-29	712	0.90	18.35	1.30	-1.04	-0.10	-4.91	
PCLN	2013-01-03	2015-10-29	712	-7.35	16.86	1.39	-1.36	-0.02	-3.94	8.25
FSLR	2013-01-03	2015-10-29	712	0.33	9.19	2.50	4.48	-1.47	-3.81	
FSLR	2013-01-03	2015-10-29	712	-3.74	8.81	2.09	4.30	-1.34	-3.58	3.97
MSFT	2013-01-03	2015-10-29	712	0.49	18.39	-2.67	2.98	7.33	-7.36	
MSFT	2013-01-03	2015-10-29	712	-2.43	18.02	-2.57	2.88	7.25	-7.19	2.63
GOOG	2014-03-28	2015-10-29	402	-0.43	12.65	-1.37	2.00	1.83	-7.60	
GOOG	2014-03-28	2015-10-29	402	-2.80	11.86	-1.16	1.85	1.96	-7.20	2.77
AAL	2013-12-09	2015-10-29	477	1.06	9.64	-0.61	-3.15	0.58	-0.15	
AAL	2013-12-09	2015-10-29	477	-4.49	9.44	-0.82	-3.51	0.72	0.09	4.86
CHK	2013-01-03	2015-10-29	707	-1.19	10.86	2.88	8.89	-1.97	-1.82	
CHK	2013-01-03	2015-10-29	707	-3.19	10.94	2.95	8.68	-1.85	-1.58	3.04
RIG	2013-01-03	2015-10-29	695	-1.44	9.84	2.61	8.79	-1.73	-0.46	
RIG	2013-01-03	2015-10-29	695	-3.55	9.75	2.55	8.38	-1.74	-0.27	3.27
Mean			674	-2.31	13.67	0.12	1.98	-0.20	-3.52	5.49
Median			712	-1.31	12.85	[illegible]	[illegible]	[illegible]	-3.65	4.95

Yes!

Do Tweet Sentiments Still Predict the Stock Market?

Exhibit 1 – Number of Tweets by Year

Exhibit 1 - Number of Tweets by Year Across Bearish, Bullish, None, and Total
Bearish and Bullish labeled tweets are displayed. None contains tweets that are not labeled.
Total is the sum of Bearish, Bullish, and None.
The second panel contains cross-sectional percentages each year.

Year	Bearish	Bullish	None	Total
2011	15,269	46,252	1,114,984	1,176,505
2012	33,175	95,440	2,684,840	2,813,455
2013	173,799	606,998	5,229,252	6,010,049
2014	393,141	1,771,176	8,641,362	10,805,679
2015	654,371	2,596,087	12,361,854	15,612,312

36 million tweets!

	Percentages			
Year	Bearish	Bullish	None	Total
2011	1.3%	3.9%	94.8%	100%
2012	1.2%	3.4%	95.4%	100%
2013	2.9%	10.1%	87.0%	100%
2014	3.6%	16.4%	80.0%	100%
2015	4.2%	16.6%	79.2%	100%

Examples of Bullish and Bearish Tweets

Exhibit 1: Bullish and Bearish Tweets

Example of 5 Bullish Tweets

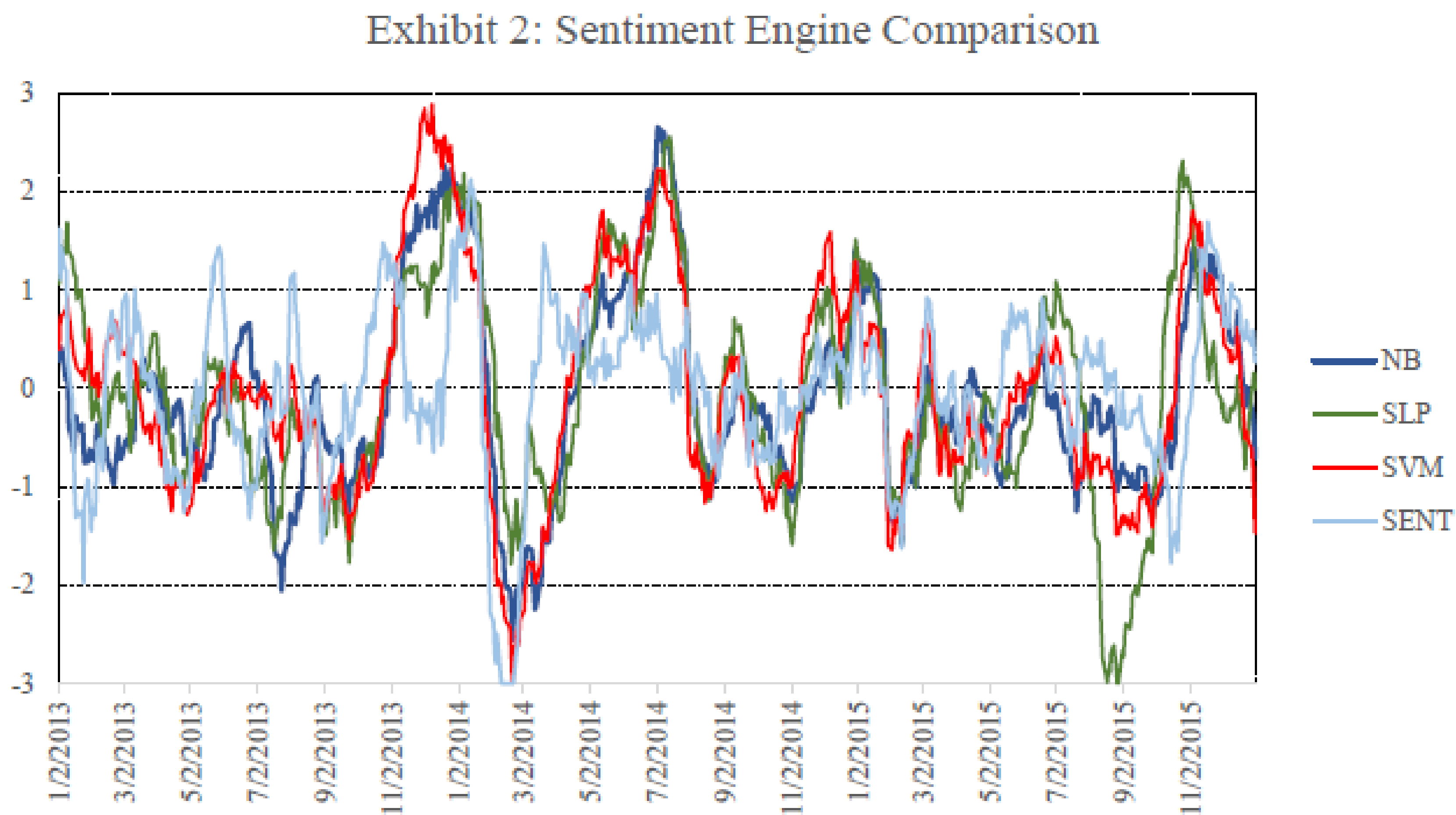
2013-01-02 13:54:05	$TNX breakout favored, lifting odds of risk-on trade near term $TLT $TBT $S PX $SPY #stocks #bonds. http://stks.co/nHho
2014-09-02 13:32:09	$DJIA looks to me good for a buy for the intra-day for an hour or 2 right now, o pened even while $SPX and $QQQ opened up.
2015-04-29 13:46:42	$SPY in with full margin. Free money. New highs coming up.
2015-11-30 15:54:16	$SPY told you to buy the dip. This week SP500 will head higher.
2015-12-01 08:35:57	$ES_F I hope we can start Wave iii up before the OPEN .... just a little push int o 2100

Example of 5 Bearish Tweets

2013-01-07 08:03:31	oops!! What happen to $GC_F and $CL_F FALLING HARD ... $ES_F LOD 1 454.75 $SPX 1434.44 KEY
2014-09-30 14:04:31	$SPY $QQQ Once this weak support is broken, markets will REALLY pull back.
2015-01-30 15:11:21	$SPX $SPY $ES_F Not to be the bearer of bad news. But, we're not done correcting. Bearish RSI divergence. http://stks.co/g1eYx
2015-06-01 12:59:05	$SPY market is completely manipulated. It's obvious. The Wall Street trader to grandma next door all agree, this market is smoke & mirrors
2015-12-01 13:00:48	$AMBA $AMZN $SPY $DJIA $XLF $MSFT $XLK $CMG I would not buy anything. Russia is about to have a WAR WITH ISLAMIC COUNTRY TURKEY. IM SELLING

Exhibit 2: Sentiment Engine Comparison

Exhibit 2a: Sentiment Engines Comparison smoothed by z-scoring MACD(20,60)



NB – Naive Bayes, SLP – Single Layer Perceptron, SVM – Support Vector Machine, SENT – Pattern's sentiment engine

Who’s driving?

Tweet sentiments vs Markets

Who's driving? Tweet sentiments vs Markets

Exhibit 4: Granger-Causality Results p-Values from 2013 to 2015

	Ho: Sentiment do not Granger-cause Mkt-RF				Ho: Mkt-RF do not Granger-cause Sentiment			
Lags	NB → Mkt-RF	SLP → Mkt-RF	SVM → Mkt-RF	SENT → Mkt-RF	Mkt-RF → NB	Mkt-RF → SLP	Mkt-RF → SVM	Mkt-RF → SENT
1	0.072*	0.767	0.034**	0.060*	0.022**	0.231	0.003**	0.040**
2	0.067*	0.851	0.025**	0.135	0.008**	0.089*	0.000**	0.023**
3	0.148	0.961	0.032**	0.120	0.000**	0.050**	0.000**	0.013**
4	0.284	0.520	0.032**	0.291	0.000**	0.013**	0.000**	0.014**
5	0.197	0.346	0.030**	0.021**	0.000**	0.003**	0.000**	0.020**

(p-value < 0.05: **, p-value < 0.10: *)

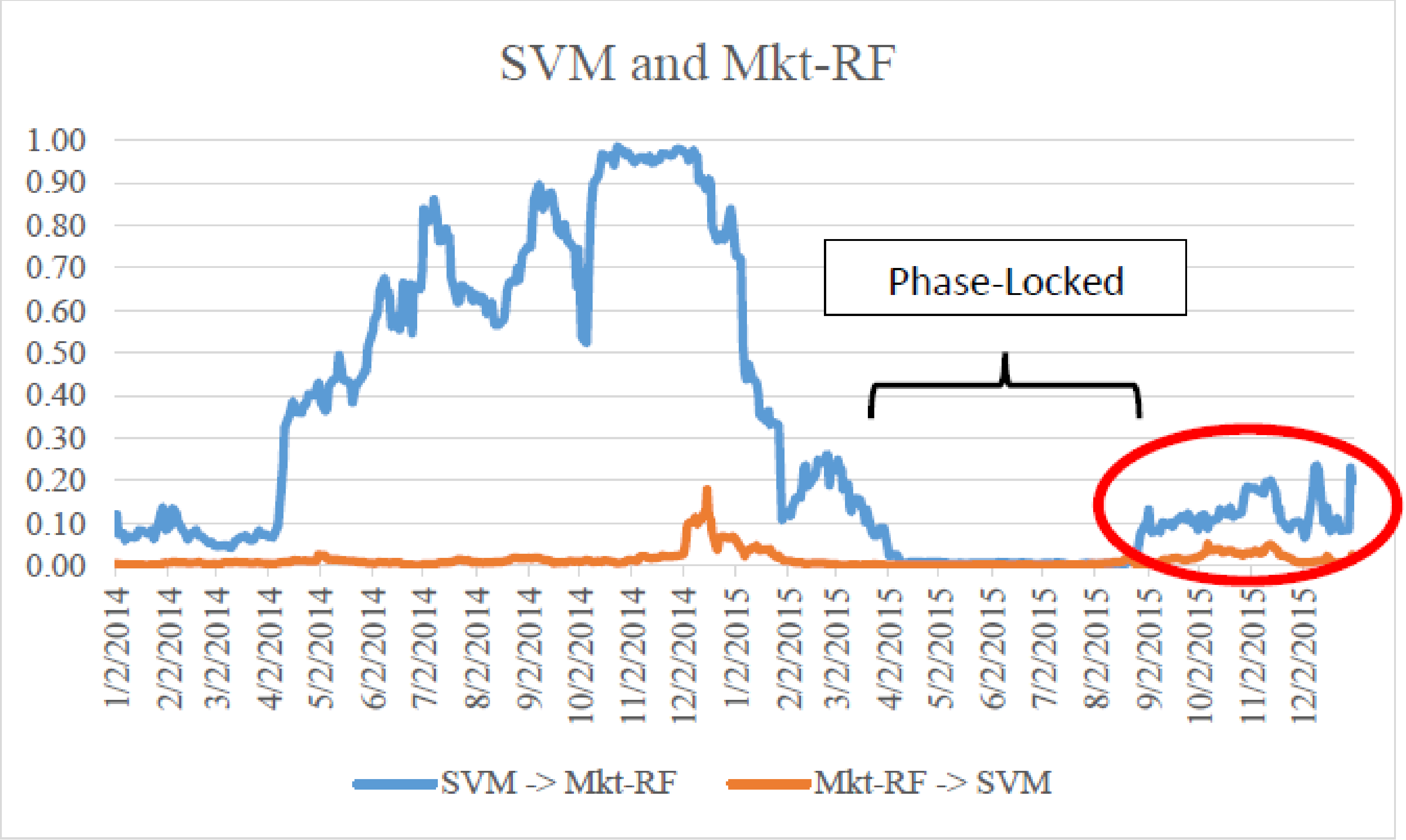
$$r_t = \beta_0 + \beta_1 r_{t-1} + \cdots + \beta_p r_{t-p} + \varepsilon_t \quad (3)$$

$$r_t = \beta_0 + \beta_1 r_{t-1} + \cdots + \beta_p r_{t-p} + \gamma_1 s_{t-1} + \cdots + \gamma_p s_{t-p} + \varepsilon_t \quad (4)$$

$$s_t = \gamma_0 + \gamma_1 s_{t-1} + \cdots + \gamma_p s_{t-p} + \varepsilon_t \quad (5)$$

$$s_t = \gamma_0 + \gamma_1 s_{t-1} + \cdots + \gamma_p s_{t-p} + \beta_1 r_{t-1} + \cdots + \beta_p r_{t-p} + +\varepsilon_t \quad (6)$$

Let’s examine the rolling p-values… “Phase-Locked” period when tweet sentiments actually drove markets!



Conclusions

- Finding links between equity price behavior around IPOs & earnings events and tweet sentiments gave us initial clues that something more substantial was lurking beneath the surface
- Linking the daily time-series of a given stock returns to its contemporaneous tweet sentiments was a significant break-through contribution
- Finally, documenting market-wide tweet sentiments “caused” markets to actually move gave hope for traders, but unfortunately the efficiency of the market prevailed. The trade is over!

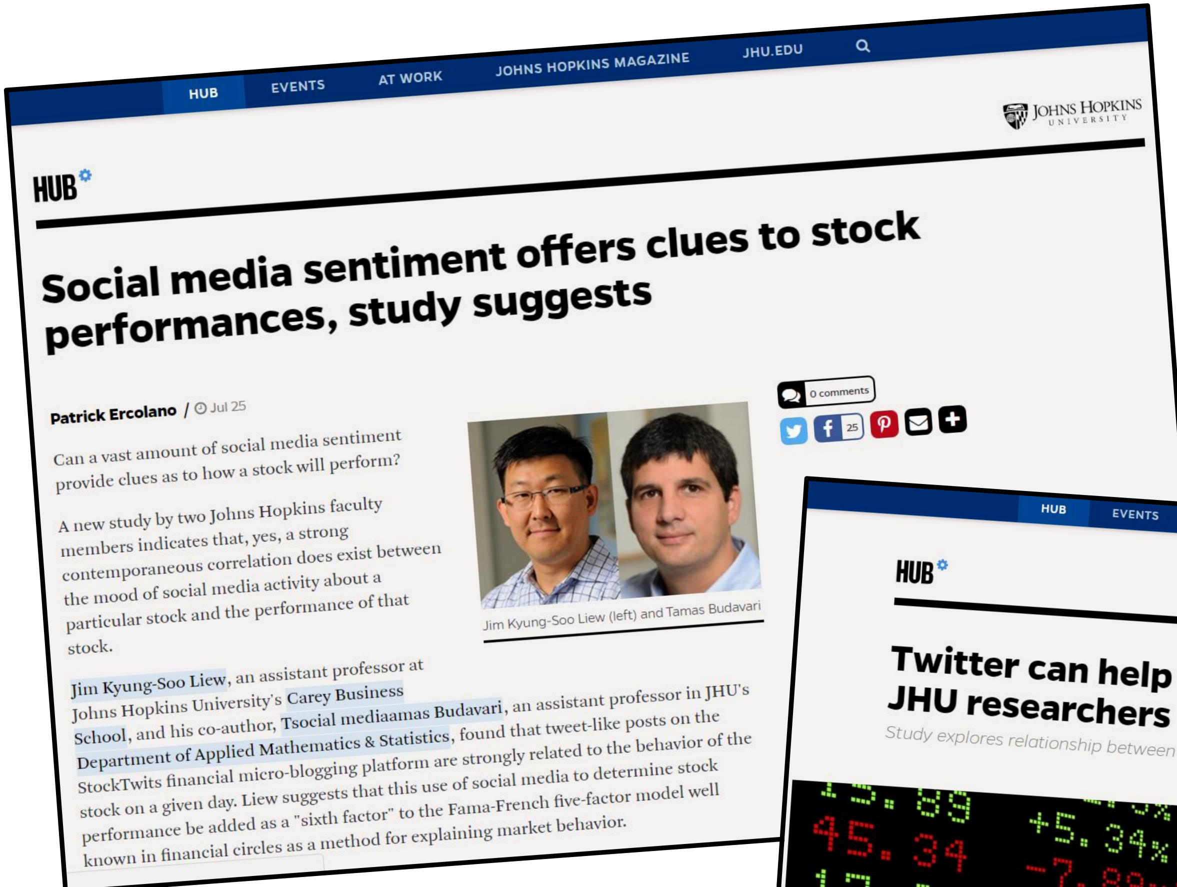
Extensions

- Machine Learning has amazing tools that can be readily applied to finance (we're offering "Big Data Machine Learning" at JHU Carey for the 1st time!)
- How about building a FinTech AI that trades the markets? Attempts to beat human traders.
- Mixing cross-collaboration, big data, huge computation resources, and domain expertise, results in explosive breakthroughs
- Input data would include: graphs, news, text, pictures, videos, sentiments, geo-positions of information, fundamentals, technical, uncertainty, etc.

WARNING! WARNING! WARNING!

The Terminator is coming to trade against you!

But not any time soon.



HUB EVENTS AT WORK JOHNS HOPKINS MAGAZINE JHU.EDU

JOHNS HOPKINS UNIVERSITY

HUB

Social media sentiment offers clues to stock performances, study suggests

Patrick Ercolano / Jul 25

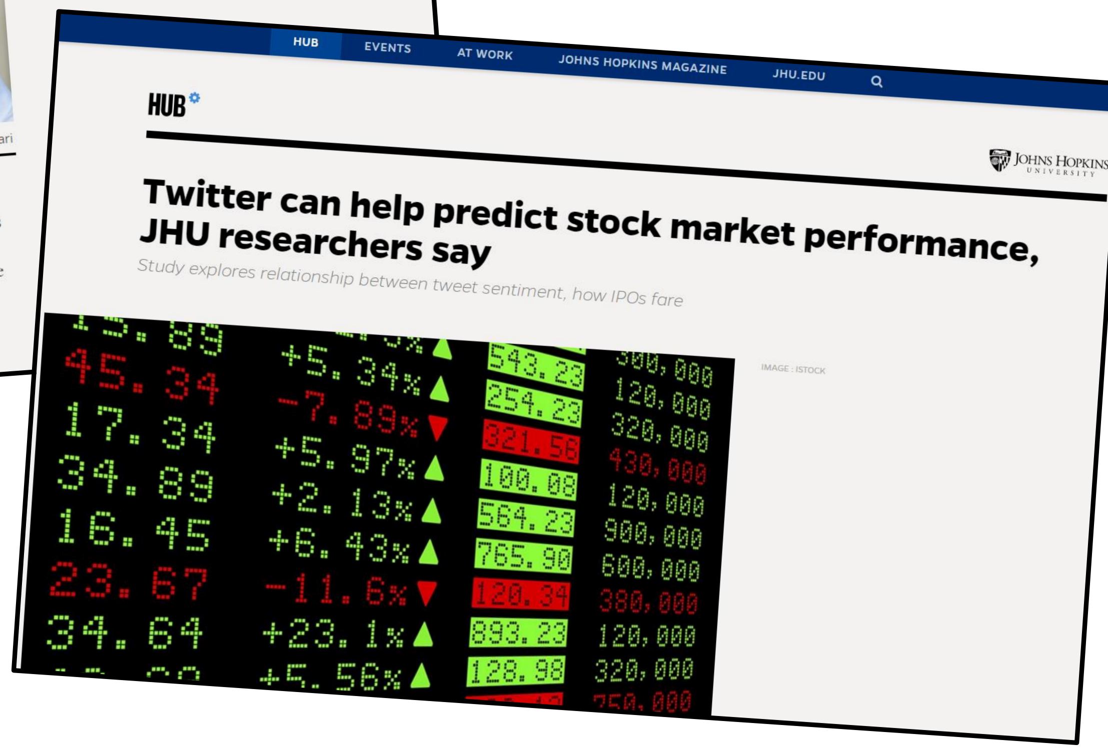
0 comments

Can a vast amount of social media sentiment provide clues as to how a stock will perform?

A new study by two Johns Hopkins faculty members indicates that, yes, a strong contemporaneous correlation does exist between the mood of social media activity about a particular stock and the performance of that stock.

Jim Kyung-Soo Liew (left) and Tamas Budavari

Jim Kyung-Soo Liew, an assistant professor at Johns Hopkins University's Carey Business School, and his co-author, Tsocial mediaamas Budavari, an assistant professor in JHU's Department of Applied Mathematics & Statistics, found that tweet-like posts on the StockTwits financial micro-blogging platform are strongly related to the behavior of the stock on a given day. Liew suggests that this use of social media to determine stock performance be added as a "sixth factor" to the Fama-French five-factor model well known in financial circles as a method for explaining market behavior.



HUB EVENTS AT WORK JOHNS HOPKINS MAGAZINE JHU.EDU

HUB

JOHNS HOPKINS UNIVERSITY

Twitter can help predict stock market performance, JHU researchers say

Study explores relationship between tweet sentiment, how IPOs fare

IMAGE: ISTOCK

Thanks for your time!