Investigating the Relation Between Twitter Sentiment and Stock Prices

Max Adamski, Slawek Gilewski

Analysis of historical data

Dataset for analysis

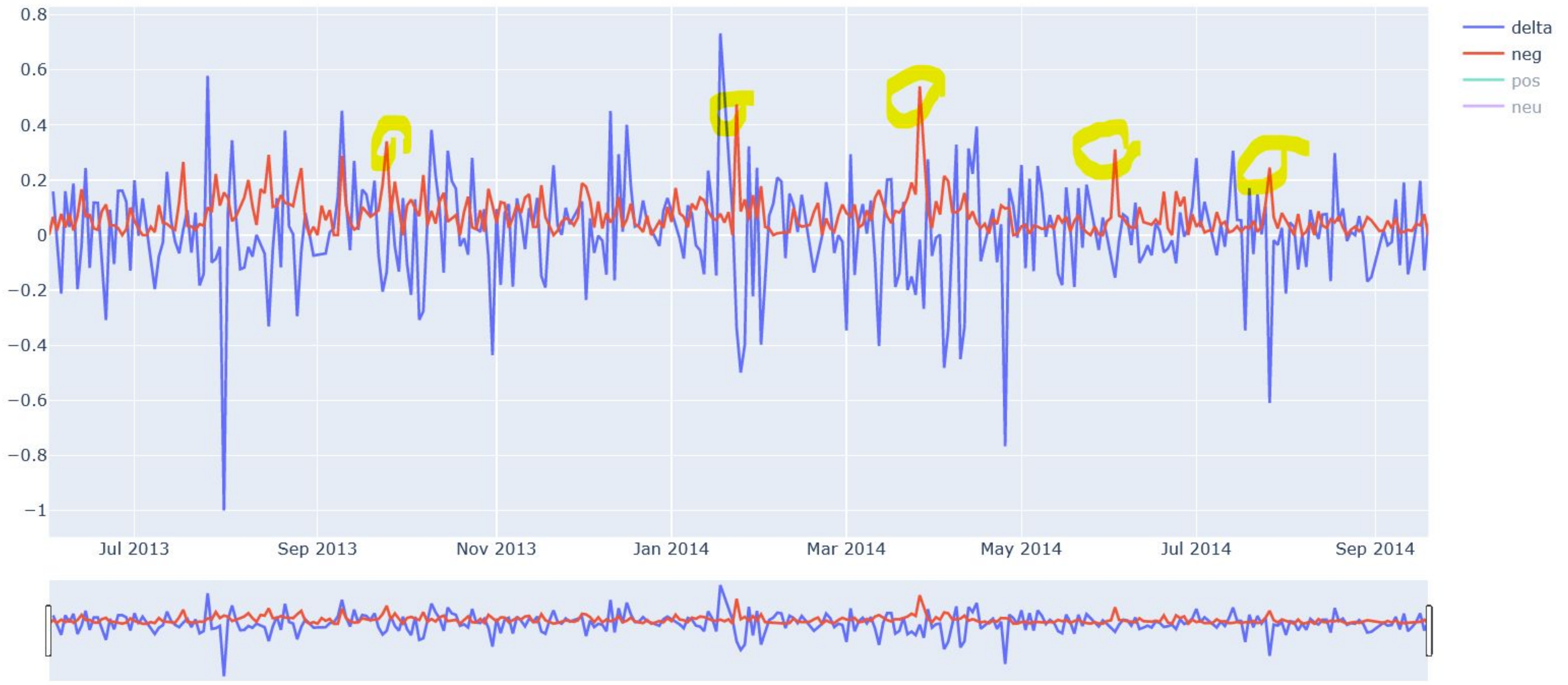
Ticker	Company	Tweets
TRV	Travelers Companies Corp	12,184
UNH	UnitedHealth Group Inc	15,020
UTX	United Technologies Corp	16,123
MMM	3M Co	17,001
DD	E I du Pont de Nemours and Co	17,340
AXP	American Express Co	21,941
PG	Procter & Gamble Co	25,751
NKE	Nike Inc	29,220
CVX	Chevron Corp	29,477
HD	Home Depot Inc	30,923
CAT	Caterpillar Inc	38,739
JNJ	Johnson & Johnson	40,503
V	Visa Inc	43,375
VZ	Verizon Communications Inc	45,177
KO	Coca-Cola Co	45,339
MCD	McDonald's Corp	45,971
XOM	Exxon Mobil Corp	46,286
DIS	Walt Disney Co	46,439
BA	Boeing Co	51,799
MRK	Merck & Co Inc	54,986
CSCO	Cisco Systems Inc	57,427
GE	General Electric Co	61,836
WMT	Wal-Mart Stores Inc	63,405
INTC	Intel Corp	68,079
PFE	Pfizer Inc	71,415
T	AT&T Inc	75,886
GS	Goldman Sachs Group Inc	91,057
IBM	International Business Machines Co	101,077
JPM	JPMorgan Chase and Co	108,810
MSFT	Microsoft Corp	183,184
Total		1,555,770

Table 1: The collected Twitter data for the 15 months period: the company names and the number of tweets.

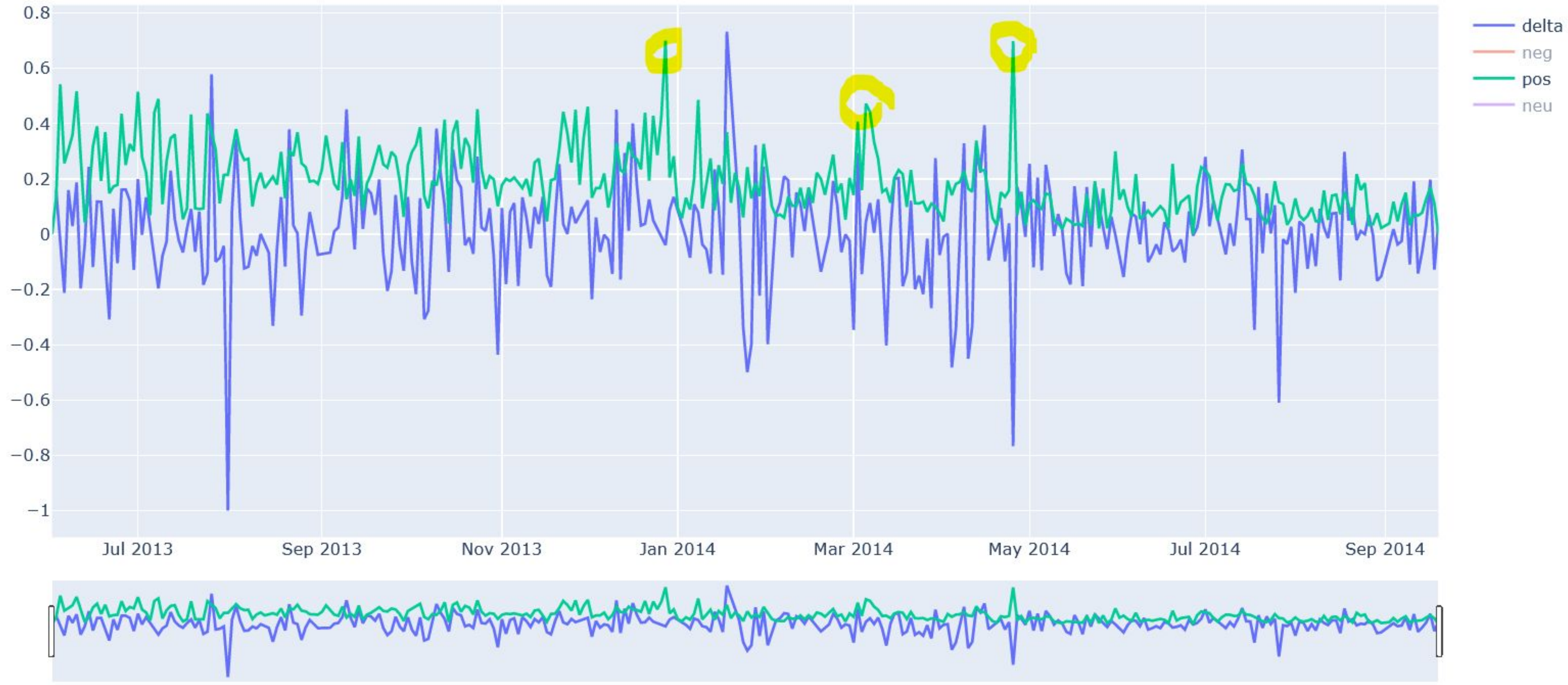
Source: Ranco et al., The Effects of Twitter Sentiment on Stock Price Returns

MSFT Price action & tweet sentiment

After removing the trend (negativity)



After removing the trend (positivity)



Small correlation (is not causation) :)

	pos	neu	neg
DeltaOpen	0.30	-0.13	-0.18
DeltaHigh	0.34	-0.13	-0.24
DeltaLow	0.30	-0.07	-0.29
DeltaClose	0.32	-0.06	-0.32

Positive gainers - Top 5

	symbol	pos	neg	neu
Microsoft	MSFT	0.330321	-0.283633	-0.095529
General Electric	GE	0.327986	-0.076051	-0.283461
CAT	CAT	0.319667	-0.351615	0.045353
Home Depot	HD	0.305868	-0.264657	-0.110991
IBM	IBM	0.290623	-0.402524	0.108387

Positive gainers - Bottom 5

	symbol	pos	neg	neu
Pfizer	PFE	0.098407	-0.108660	-0.040222
Verizon	VZ	0.092382	-0.229562	0.029501
Coca Cola	KO	0.088243	-0.146906	0.001384
AT&T	T	0.080441	-0.185863	-0.002700
Travelers	TRV	0.044055	-0.116357	0.045274

Negative losers - Top 5

symbol	pos	neg	neu
IBM	0.290623	-0.402524	0.108387
BA	0.242772	-0.357476	-0.010595
GS	0.252315	-0.352873	0.060282
CAT	0.319667	-0.351615	0.045353
V	0.174454	-0.328152	0.015777

Boeing Air

Goldman Sachs

Visa

Negative losers - Bottom 5

JP Morgan

Pfizer

Merck & Co

symbol	pos	neg	neu
JPM	0.224137	-0.125252	0.011951
TRV	0.044055	-0.116357	0.045274
PFE	0.098407	-0.108660	-0.040222
MRK	0.248310	-0.090426	-0.185661
GE	0.327986	-0.076051	-0.283461

Is any of this significant?

Go to https://arxiv.org/pdf/1506.02431.pdf to find the statistical significance tests

The Effects of Twitter Sentiment on Stock Price Returns

Gabriele Ranco[1], Darko Aleksovski[2,*], Guido Caldarelli[1,3,4], Miha Grčar[2], Igor Mozetič[2]
1 IMT Institute for Advanced Studies, Piazza San Francesco 19, 55100 Lucca, Italy
2 Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia
3 Istituto dei Sistemi Complessi (ISC), Via dei Taurini 19, 00185 Rome, Italy
4 London Institute for Mathematical Sciences, 35a South St. Mayfair, London W1K 2XF, UK
* darko.aleksovski@ijs.si

Live Twitter Pipeline

Tools

- TextBlob, VADER-Sentiment (sentiment analysis)
- Tweepy (Twitter API)
- Yfinance (Yahoo Finance API)
- Plotly, WordCloud (Visualization)

Pipeline

1. For each entity, stream related tweets
    a. Query company cash tag symbol or name (ex. “$MSFT OR microsoft”)

Pipeline

1. For each entity, stream related tweets
    a. Query company cash tag symbol or name (ex. “$MSFT OR microsoft”)
2. Clean text with regexes
    a. Remove stopwords and #hash $cash @mention tags
    b. Remove html &escapes; and urls

Pipeline

1. For each entity, stream related tweets
    a. Query company cash tag symbol or name (ex. “$MSFT OR microsoft”)
2. Clean text with regexes
    a. Remove stopwords and #hash $cash @mention tags
    b. Remove html &escapes; and urls
3. Get sentiment scores
    a. Vader: positivity, negativity, neutrality, polarity
    b. TextBlob: polarity, objectivity
    c. Remove objective tweets (often neutral sentiment anyway...)

Pipeline

1. For each entity, stream related tweets
    a. Query company cash tag symbol or name (ex. “$MSFT OR microsoft”)
2. Clean text with regexes
    a. Remove stopwords and #hash $cash @mention tags
    b. Remove html &escapes; and urls
3. Get sentiment scores
    a. Vader: positivity, negativity, neutrality, polarity
    b. TextBlob: polarity, objectivity
    c. Remove objective tweets (often neutral sentiment anyway...)
4. Compute statistics for a given timeframe
    a. And compare with price data from Yahoo finance

Current tweets about MSFT

MSFT stock & sentiment

Further research

TABLE 2.1 The Four Essential Types of Financial Data

Fundamental Data	Market Data	Analytics	Alternative Data
• Assets	• Price/yield/implied volatility	• Analyst recommendations	• Satellite/CCTV images
• Liabilities	• Volume	• Credit ratings	• Google searches
• Sales	• Dividend/coupons	• Earnings expectations	• Twitter/chats
• Costs/earnings	• Open interest	• News sentiment	• Metadata
• Macro variables	• Quotes/cancellations	• . . .	• . . .
• . . .	• Aggressor side		
	• . . .		

Source: Lopez, Advances in Financial Machine Learning

Thank you for attention :)

https://colab.research.google.com/drive/1wpHTakMfz0HRH8KEBAAmgzI8Jv3qbawR?usp=sharing