QUANTIFYING THE EFFECTS OF ONLINE BULLISHNESS ON INTERNATIONAL FINANCIAL MARKETS

Huina Mao
School of Informatics and Computing
Indiana University, Bloomington, USA

ECB Workshop on Using Big Data for Forecasting and Statistics, April 7-8 2014, Frankfurt, Germany,

Our Research Team

Huina Mao
Indiana University, Bloomington
huinmao@indiana.edu

Dr. Scott Counts
Microsoft Research, Redmond
counts@microsoft.com

Dr. Johan Bollen
Indiana University, Bloomington
jbollen@indiana.edu

Investor Sentiment Theory

DeLong et al. (1990)

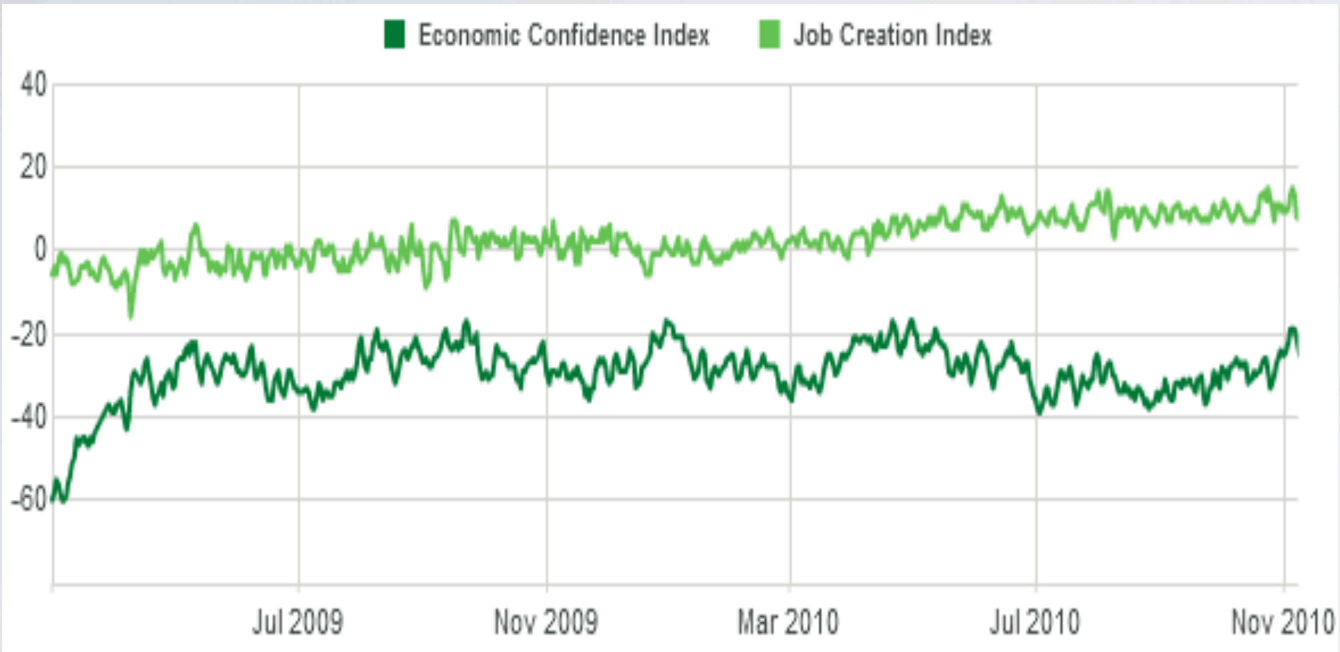
Introduced 24 years ago.

"Now, the question is no longer, as it was a few decades ago, whether investor sentiment affects stock prices, but rather **how to measure investor sentiment** and quantify its effects." (Baker and Wurgler 2007)

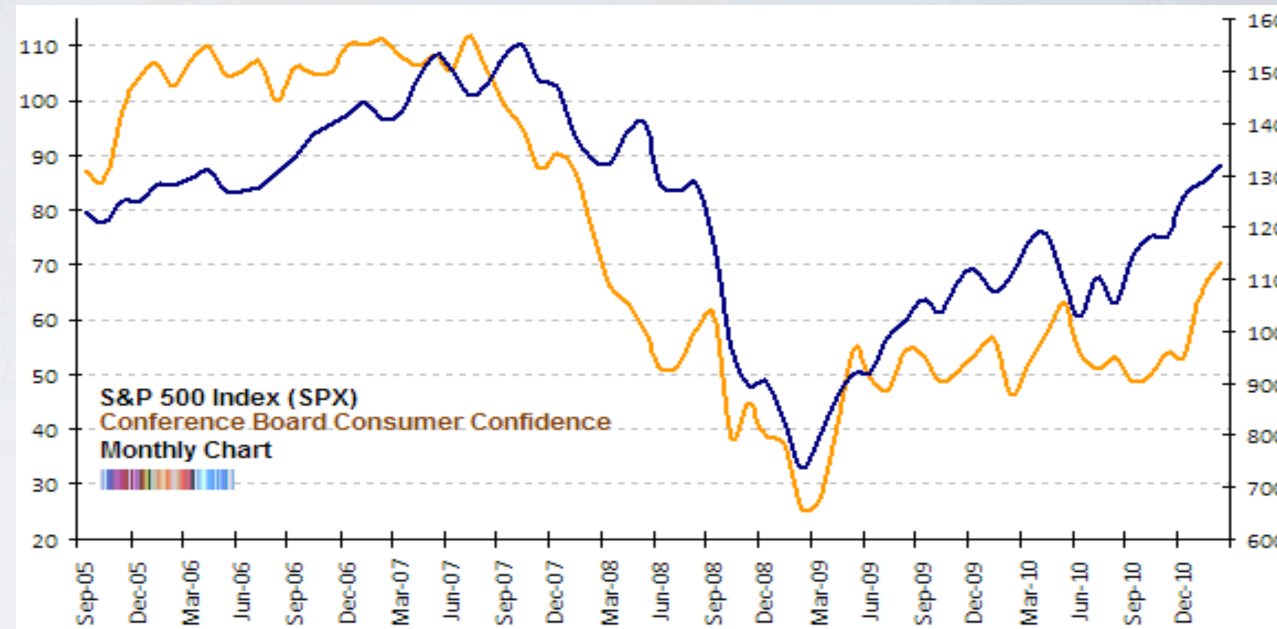
Baker and Wurgler, "Investor Sentiment in the Stock Market", Journal of Economic Perspectives, vol 21 (7), 2007

Surveys

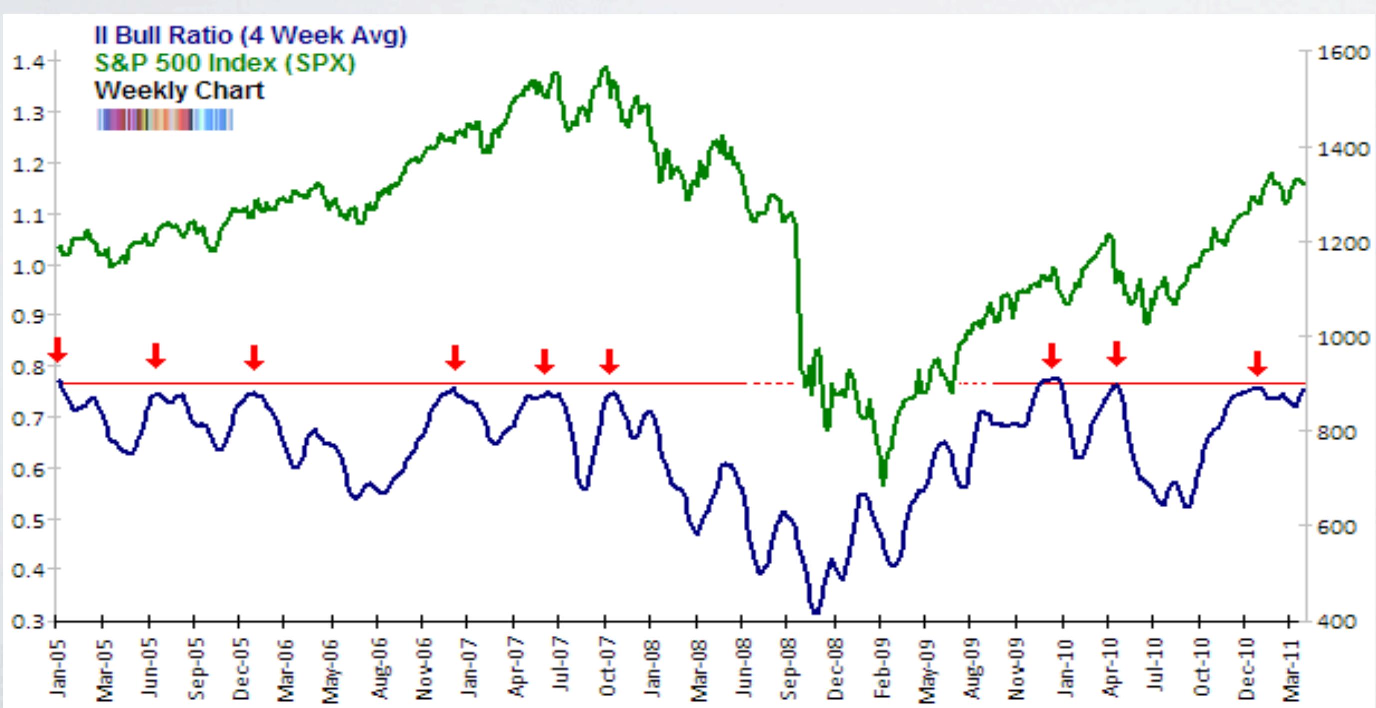
Gallup



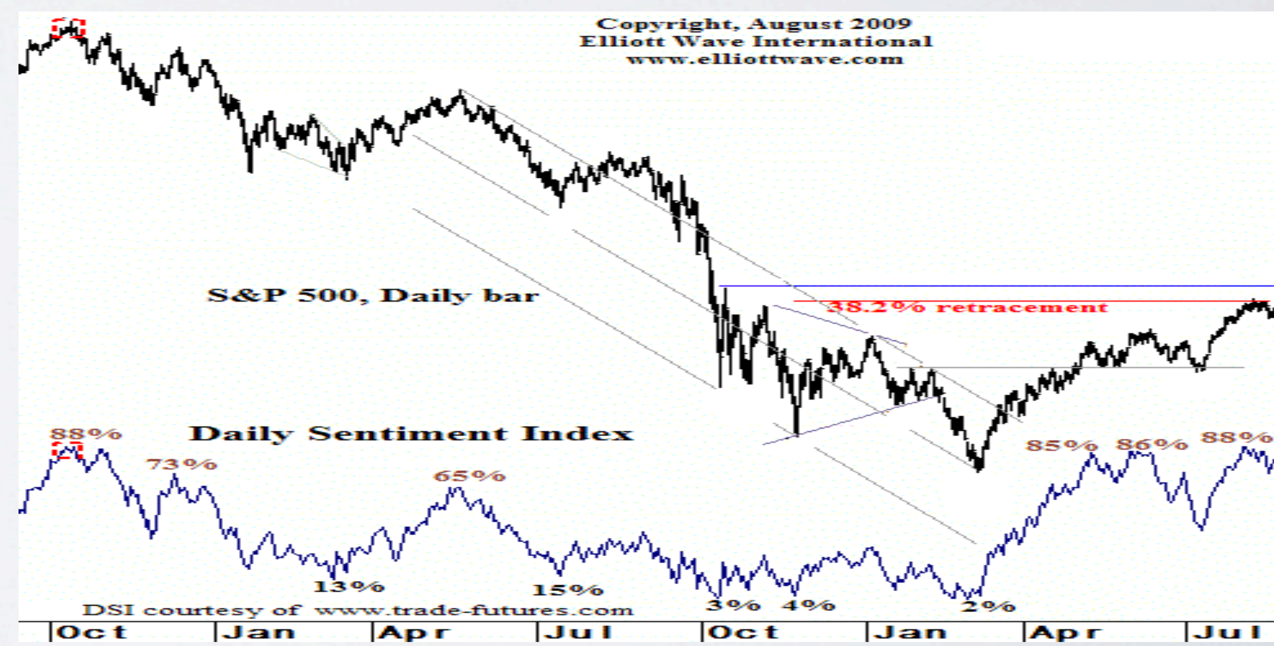
Consumer Confidence Index



Investor Intelligence



Daily Investor Sentiment



University of Michigan Consumer Sentiment Index

monthly poll: 500 telephone interviews & 5 questions

“Would you say that you (and your family living there) are better off or worse off financially than you were a year ago?”

"Now looking ahead—do you think that a year from now you (and your family living there) will be better off financially, or worse off, or just about the same as now?”

"Now turning to business conditions in the country as a whole—do you think that during the next twelve months we’ll have good times financially, or bad times, or what?"

......

Investor Intelligence (since 1963)

Weekly advisors sentiment report surveys: percentage of advisors' bullish, bearish views of over **100 independent investment newsletters**

Daily Sentiment Index (since 1987)

Interview small traders for their bullish or bearish feeling on US future markets

Pros and Cons of Surveys

Pros

Explicit

Representative samples

Straightforward-to-conduct

Controlled design

Cons

Small-scale

Reliability and validity may be an issue

Expensive-to-conduct

Low frequency (weekly, monthly, or annually)

Released with delay

Proxies of Investor Sentiment

Edmans et.al . "Sports sentiment and stock returns." *The Journal of Finance* 62.4 (2007): 1967-1998.

Hirshleifer et.al. "Good day sunshine: Stock returns and the weather." *The Journal of Finance* 58.3 (2003): 1009-1032.

Sports/weather $\longrightarrow$ Sentiment

Proxies of Investor Sentiment

Edmans et.al . "Sports sentiment and stock returns." *The Journal of Finance* 62.4 (2007): 1967-1998.

Hirshleifer et.al. "Good day sunshine: Stock returns and the weather." *The Journal of Finance* 58.3 (2003): 1009-1032.

Sports/weather ⟶ Sentiment

INDIRECT Indicator

Other Investor Sentiment Measurements

News Pessimism Index (Tetlock 2007):

Harvard Dictionary Negative Words

Stock Message Board Text Classification (Antweiler and Frank 2004):

Machine Learning Classifiers: Support Vector Machine and Naive Bayes.

Output: Bullish, bearish and neutral

References:

Tetlock, P. "Giving content to investor sentiment: The role of media in the stock market", 62 (3), pp: 1139--1168, The Journal of Finance, 2007

Antweiler, W. and Frank, M., "Is all that talk just noise? The information content of internet stock message boards" Th Journal of Finance, 59 (3), pp: 1259-1294, 2004

The Advent of Big Data

Literature review: Predicting socio-economic indicators from large-scale data

Google search predict flu (Ginsberg et.al 2008)

Blog sentiment predict stock market (Gilbert 2010)

Twitter predict box office (Asur 2010)

Google search predict unemployment claims (Ettredge 2005)

Google search predict car sales, travel, health (Choi 2009)

Google search reveals investor attention (Da 2011)

Yahoo search predicts consumer behavior (Goel 2010)

Mobile communication reveal economic prosperity (Eagle 2010)

Why Twitter?

(launched July 2006 by Jack Dorse)

Twitter Statistics (2013)

Total number of active registered Twitter users: **554,750,000**

Number of new Twitter users signing up everyday: **135,000**

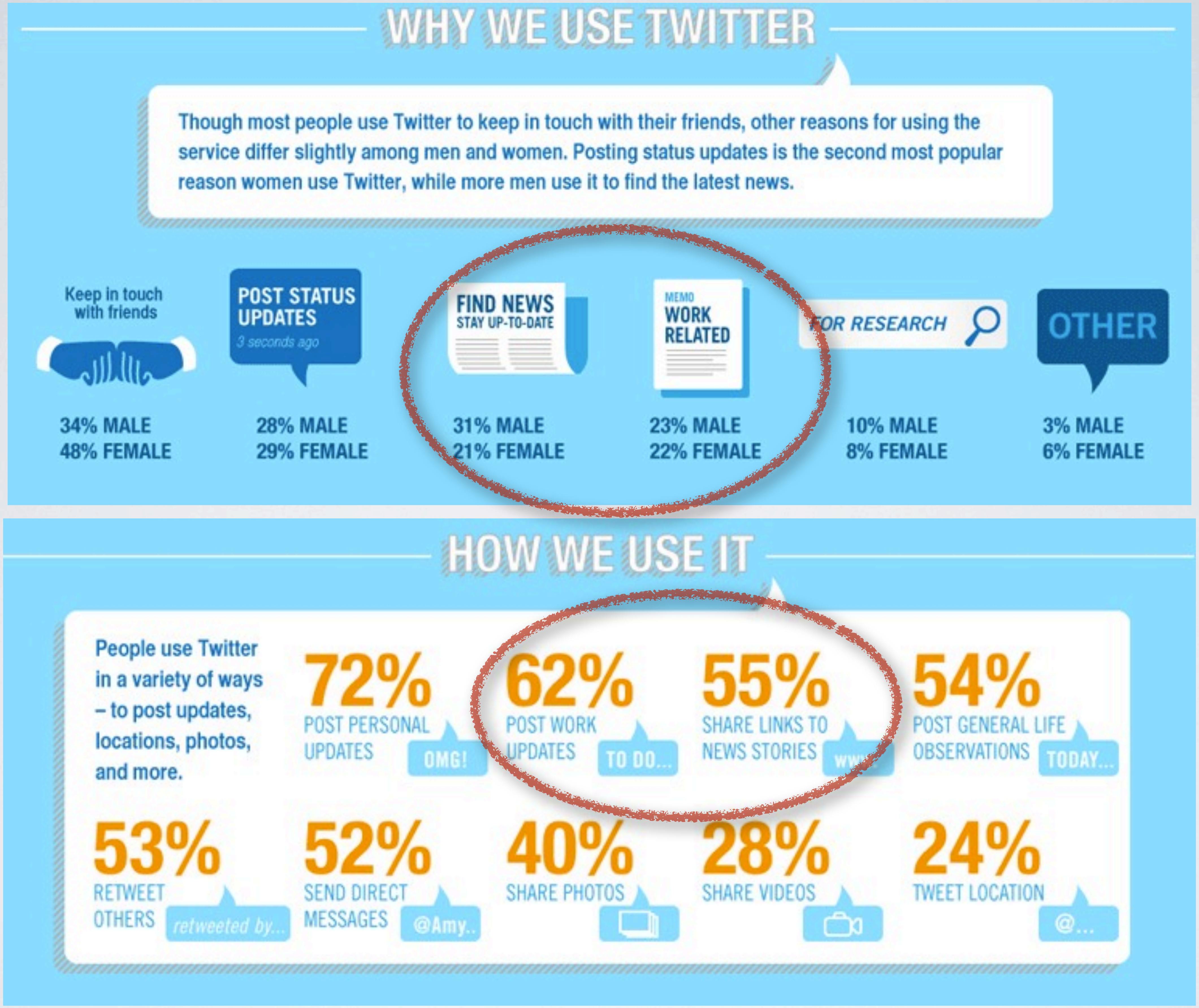
Number of Tweets that happen every second: **9,100**

Average number of tweets per day: **58 Million**

Number of Twitter search engine queries every day: **2.1 billion**

Percent of Twitters who do not tweet but watch other people tweet: **40%**

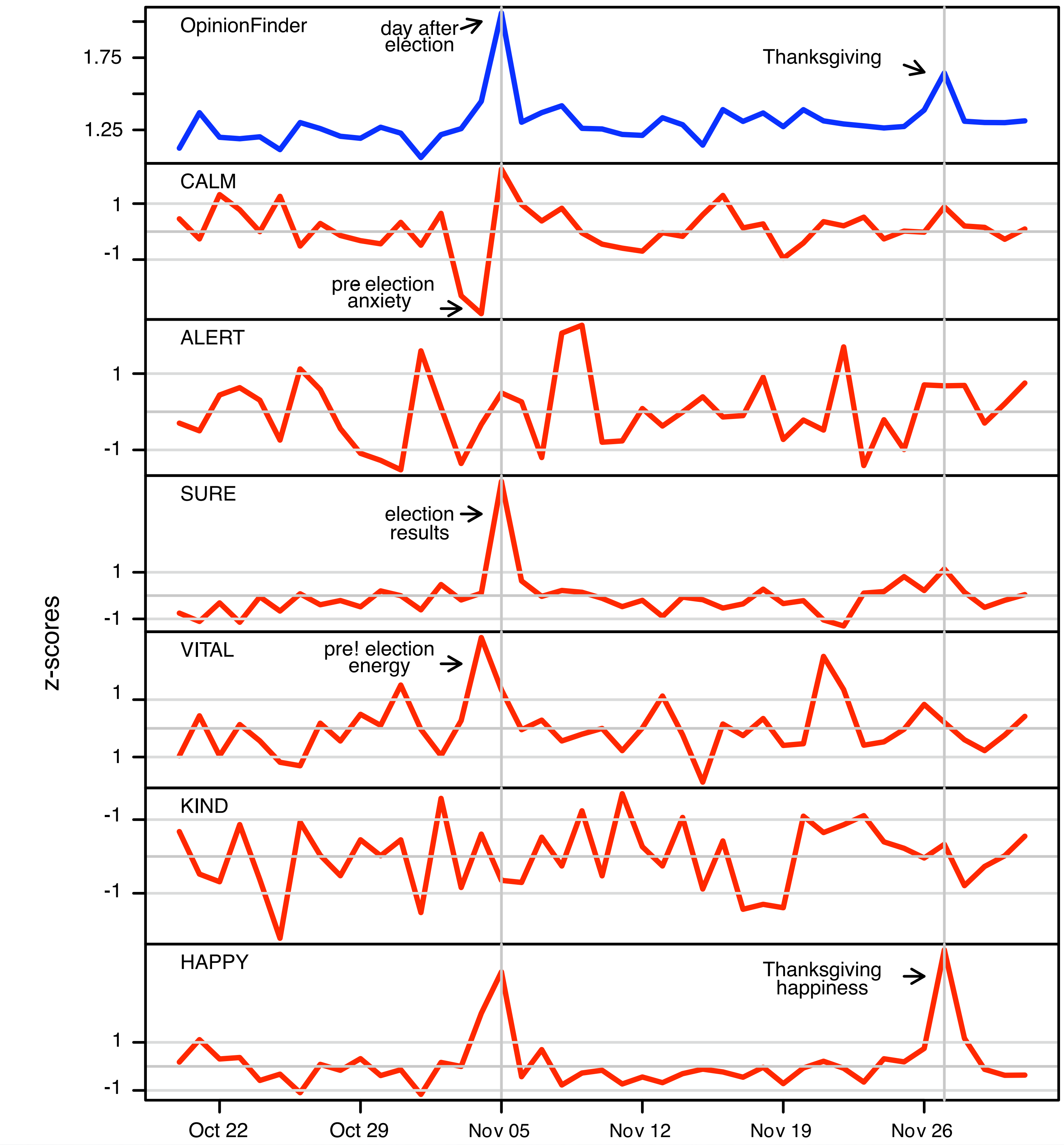
http://www.socialmediadd.com/Articles.asp?ID=248



http://www.socialmediadd.com/Articles.asp?ID=248

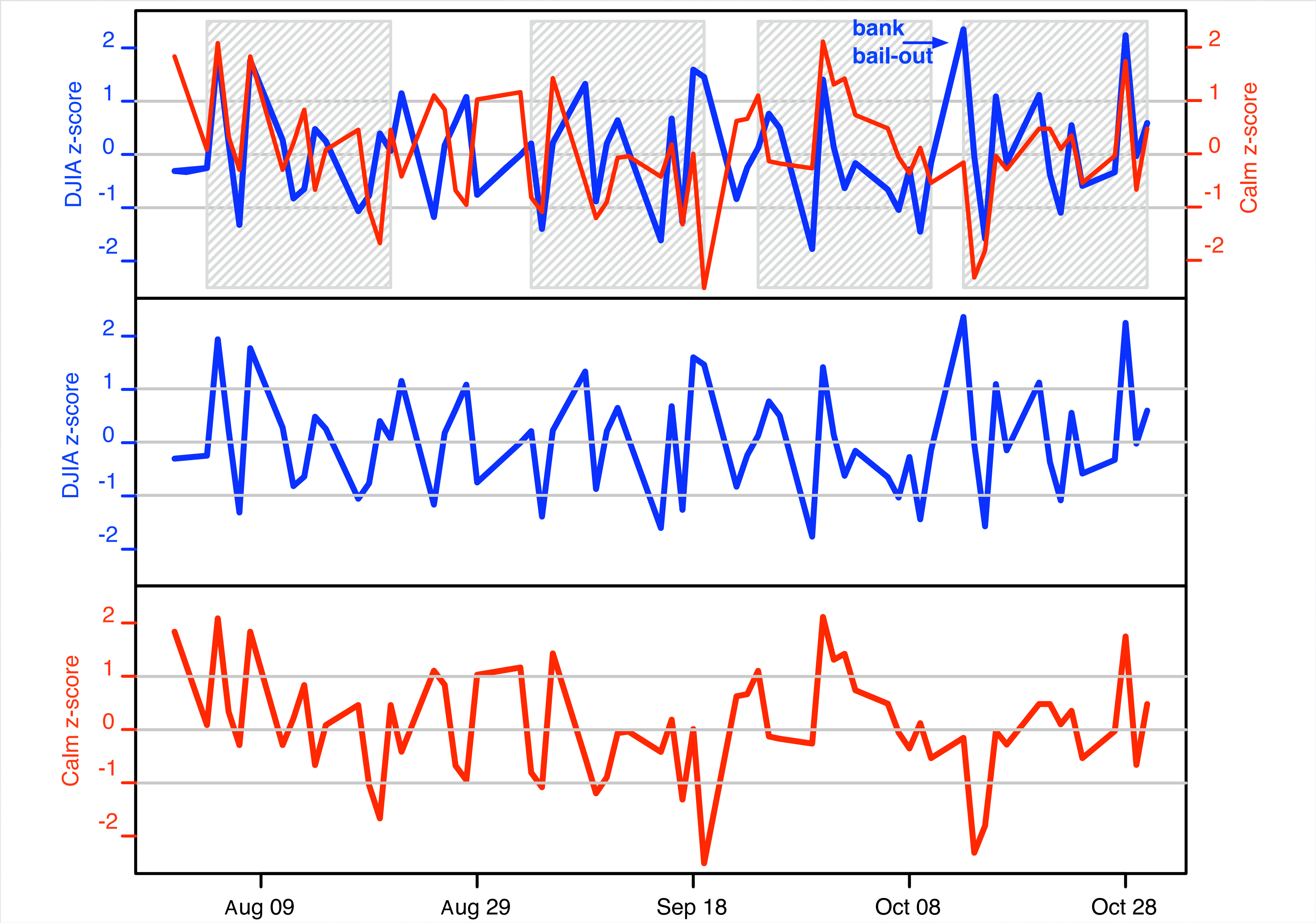
Twitter Mood and Stock Market

Social Mood From Twitter



Bollen, Mao, Zeng, "Twitter mood predict the stock market", Journal of Computational Science, vol 2(1), pp: 1-8,2011

Twitter Mood Predicts the Stock Market



Bollen, Mao, Zeng, "Twitter mood predict the stock market", Journal of Computational Science, vol 2(1), pp: 1-8,2011

Research Question:

How to measure investor sentiment from Twitter?

Investor Sentiment Surveys

Investor Intelligence (since 1963)

weekly advisors sentiment report surveys

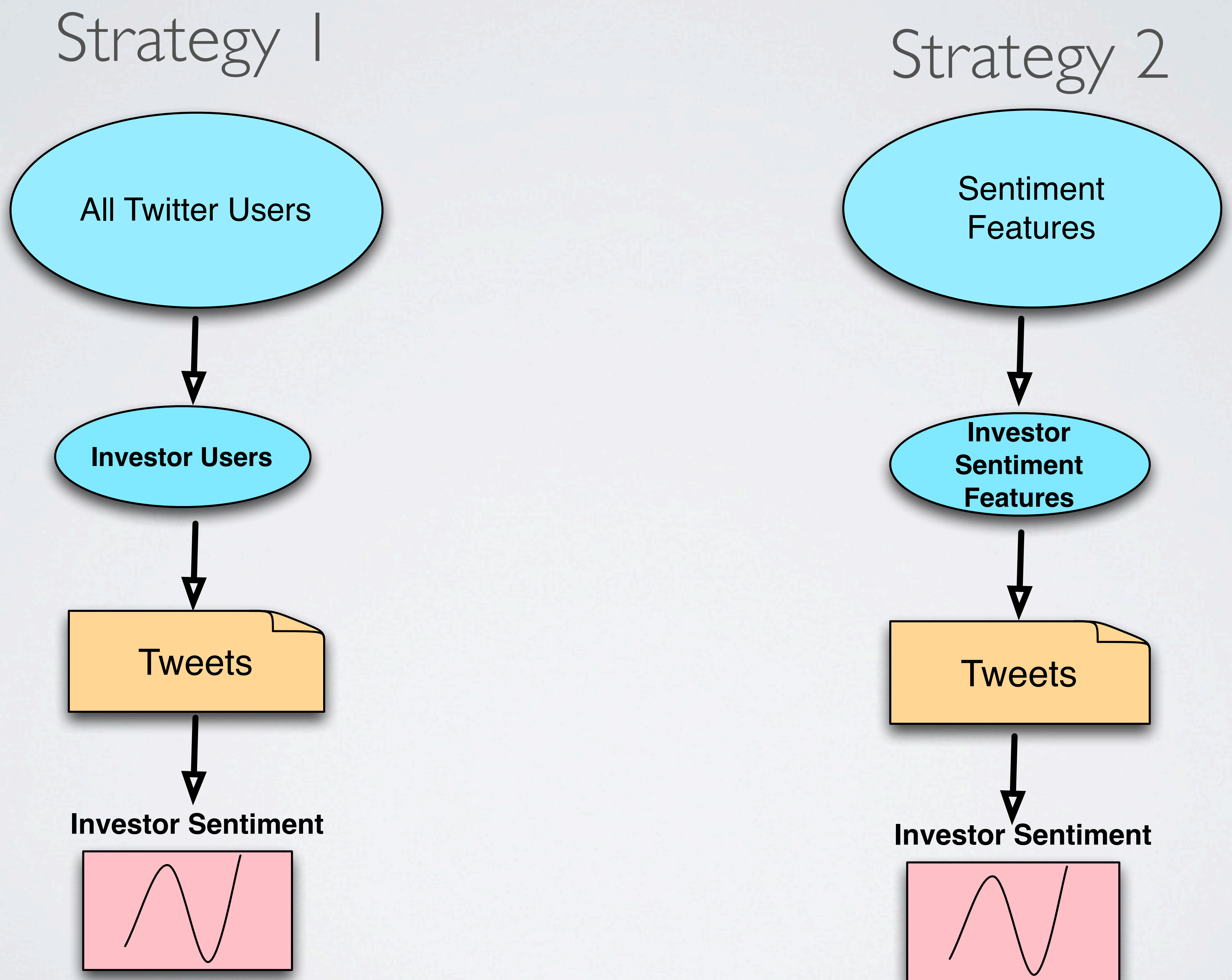
percentage of advisors' bullish or bearish views of over **100** independent investment newsletters

Daily Sentiment Index (since 1987)

daily polls

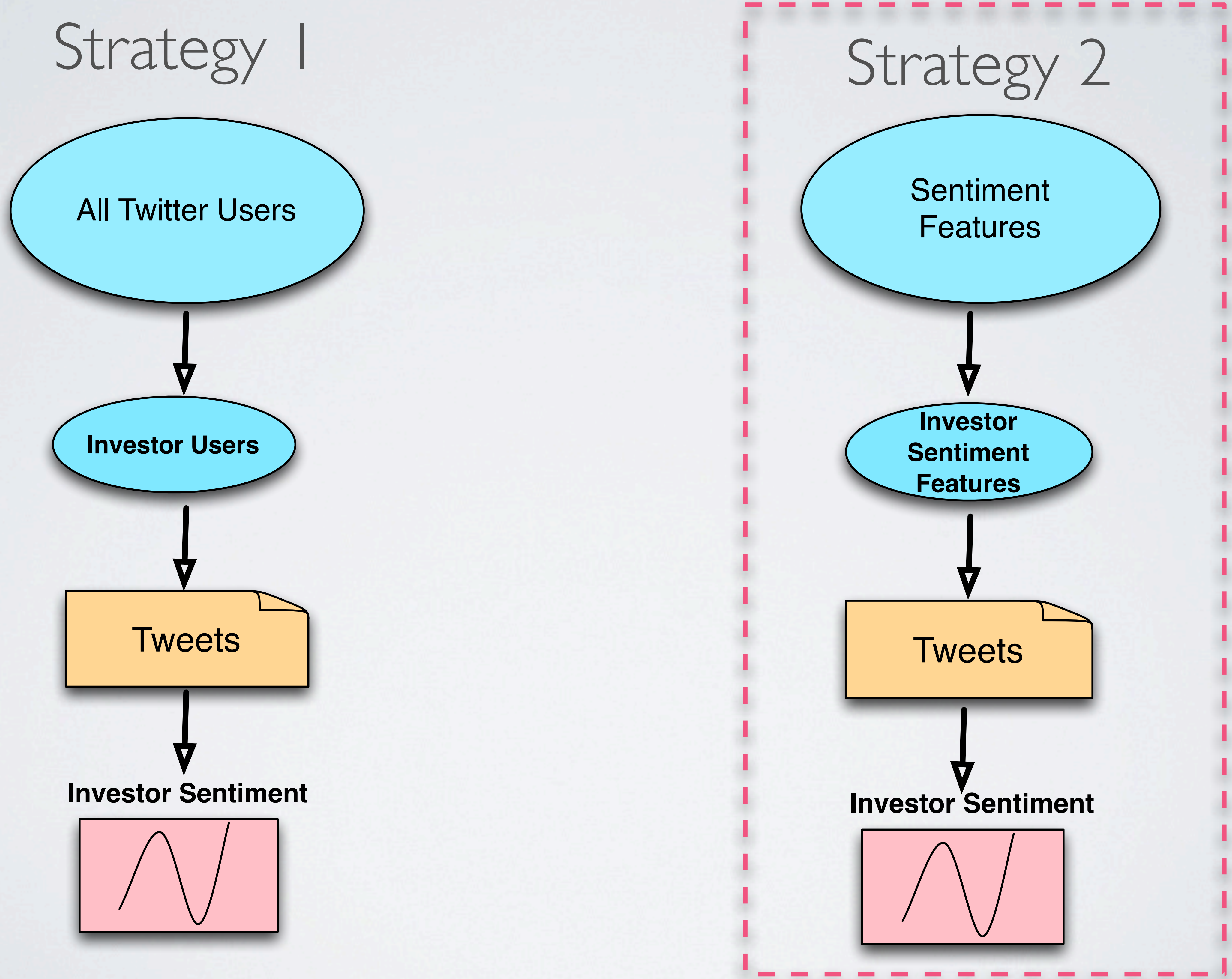
Interview traders for their bullish or bearish feeling on US future markets

Measuring Investor Sentiment



Measuring Investor Sentiment

We adopt:



Why we used 2 terms only: “bullish” and “bearish”?

Word	Frequency
signaling	46482
market	29088
remain	20544
sentiment	11900
territory	9870
pattern	8486
turning	8433
ahead	8241
chief	8079
outlook	7852

Google n-grams (LDC)
http://catalog.ldc.upenn.edu/LDC2006T13

The 2 words are very reliable and simple indicators of investment sentiment in online language

Proof:
- We looked up Google bigrams of form (bullish, x) and (bearish, x) (N=314M)
- Linguistic context is majority investor/sentiment-related

Twitter and Google Bullishness

Twitter Data

(daily)

From January 2010 to December 2012, N=45M/day

Bullish tweet = tweet that contains "**bullish**"
Bearish tweet = tweet that contains "**bearish**"

Bullish & bearish tweets: **0.31** million.

Google Data

(weekly)

From January 2007 to December 2012 (313 weeks)

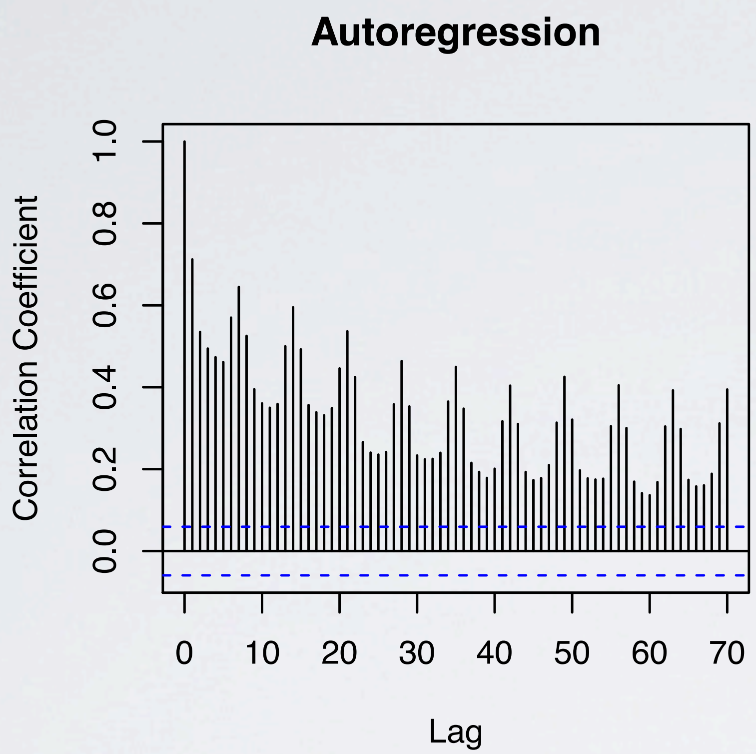
Bullish search = "**bull market**"
Bearish search = "**bear market**"

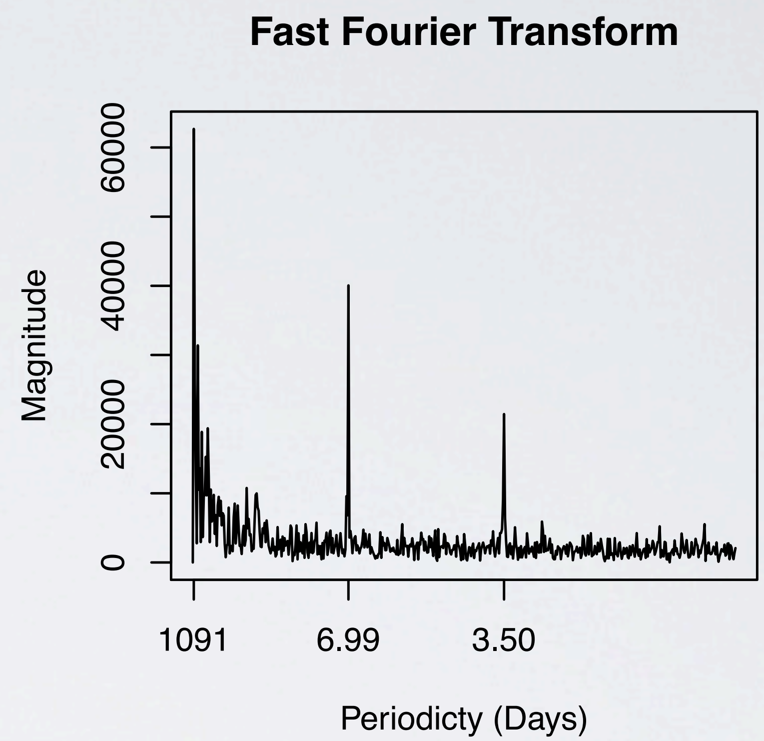
Search volumes from Google Trends.

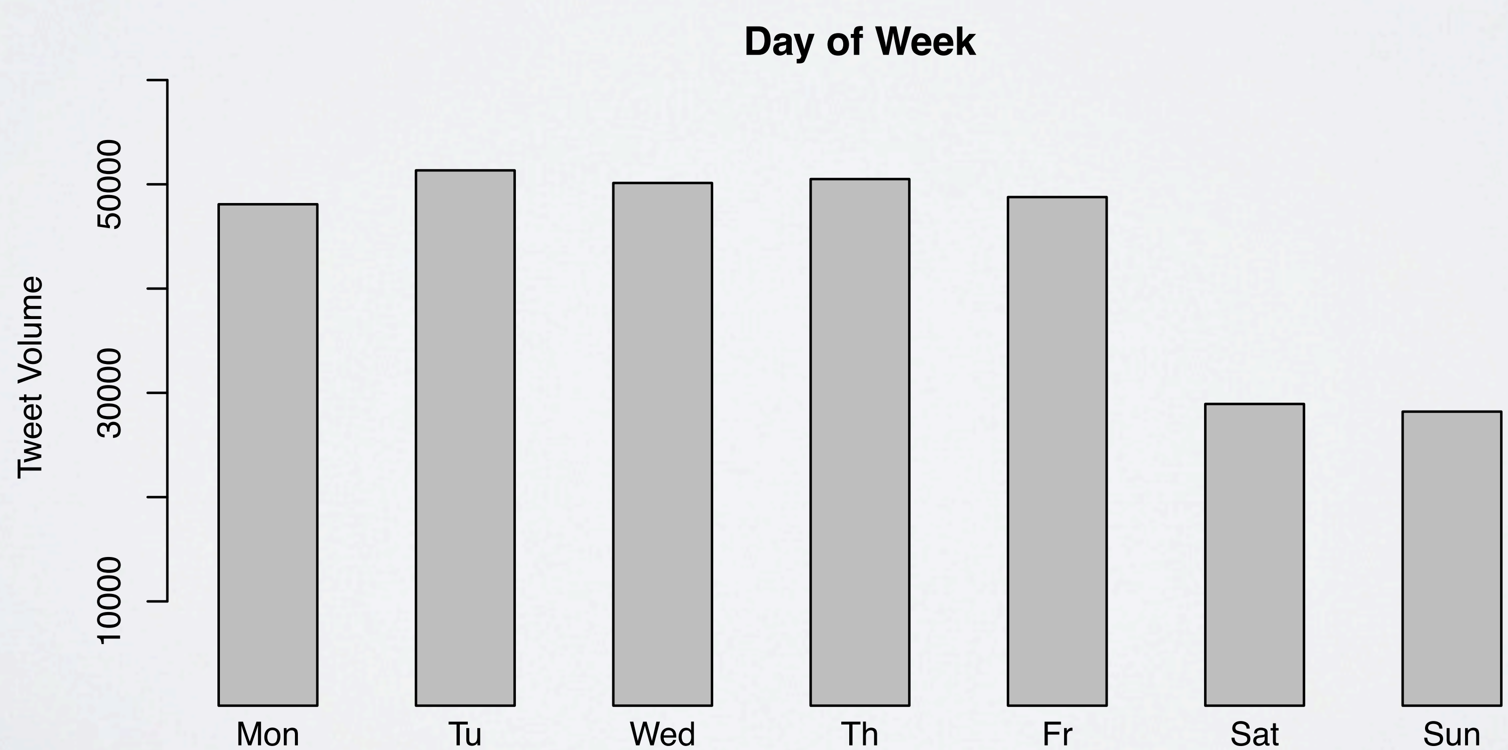
Twitter and Google Bullishness:

$$B_t = \ln\left(\frac{1 + ||\mathcal{B}_t||}{1 + ||\mathcal{R}_t||}\right) \qquad G_w = \ln\left(\frac{1 + ||\mathcal{B}_w||}{1 + ||\mathcal{R}_w||}\right)$$

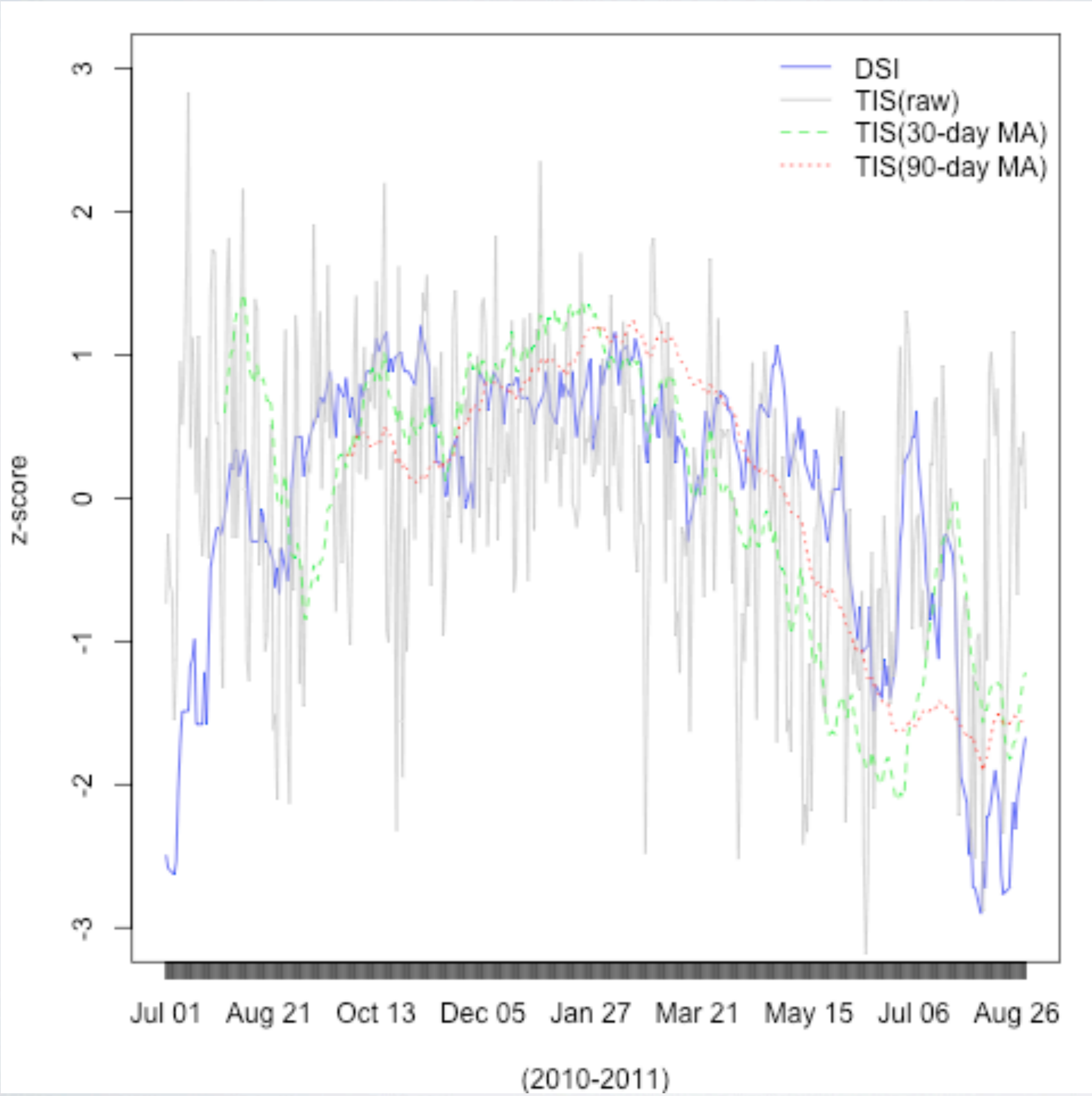
Bullish & Bearish Tweet Volume Weekly Pattern







Twitter Bullishness vs. Survey (Daily Sentiment Index)



Pearson correlation:
r=0.3 $p < 0.01$
30 day MA: 0.6
90 day MA: 0.7

Investor Sentiment Theory

DeLong et al. (1990)

Noise trader's irrationality can **drive the asset price** to deviate from its fundamental value **temporarily** after which it will **reverse** to the mean

Two Hypothesizes

If Twitter bullish and bearish tweets are only reactions to market changes, we may observe contemporaneous correlation, but **no** prediction.

If Twitter Bullishness is an indicator of investor sentiment, we may observe short-term lead and **reversal** in a long-run.

Vector Autoregression

$$R_t = \alpha + \sum_i^5 \beta_i R_{t-i} + \sum_i^5 \chi_i T^B_{t-i} + \sum_i^5 \delta_i Vol_{t-i} + \phi_i Exog_t + \epsilon_t$$

R: return; *T*: Twitter Bullishness; *Vol*: trading volume.

Exogenous variables: VIX, Daily Sentiment Index, and calendar controls.

Result 1

Predicting Daily Stock Returns of Dow Jones, S&P 500, Russell 1000 and Russell 2000 Using Twitter Bullishness.

Bullishness	DJIA		SP500		Russell1000		Russell2000	
Lag	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value
1	**12.56**	0.01⋆⋆⋆	**10.98**	0.05⋆⋆	**10.72**	0.05⋆⋆	**11.02**	0.05⋆⋆
2	2.27	0.67	2.61	0.65	2.46	0.67	2.66	0.65
3	2.18	0.69	3.69	0.53	4.037	0.48	4.58	0.43
4	-7.81	0.15	-8.10	0.16	**-9.99**	0.08⋆	**-10.28**	0.08⋆
5	-1.12	0.80	-1.28	0.79	-1.35	0.77	-1.37	0.78

Twitter Bullishness vs. Daily Sentiment Index

Linear Pearson Correlation Coefficient: **0.30 (p < 0.01)**

Daily Sentiment Index: beta_t-1 = **2.26 (p =0.1)**
Twitter Bullishness Index: beta_t-1 = **12.56 (p < 0.01)**

These two are related, but different.

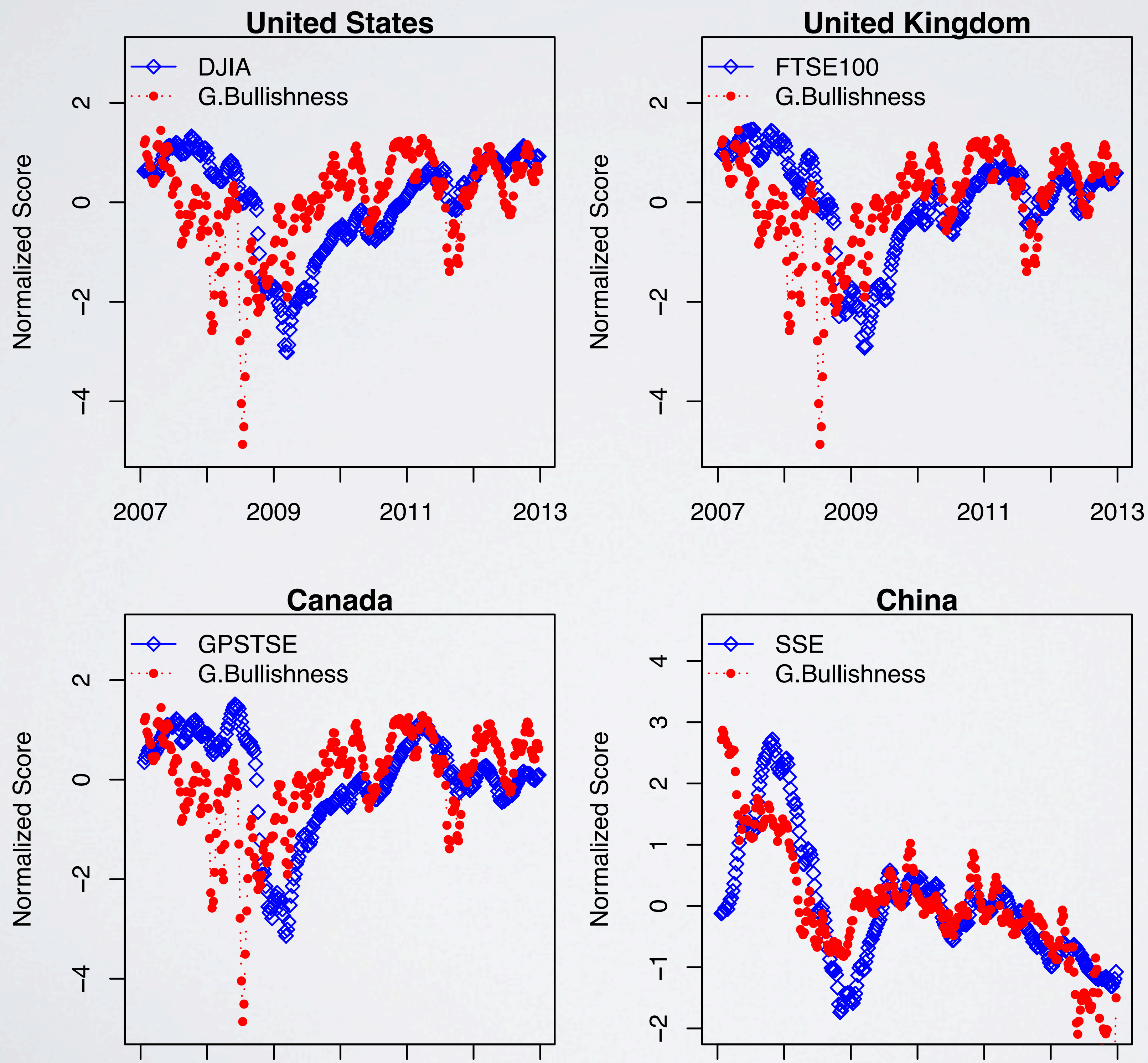
Result 2

Predicting Stock Returns of US, UK, CA, and CN Using Twitter Bullishness

Lag	US.DJIA		UK.FTSE		CA.GSPTSE		China.SSE	
	Coeff.	*p*-value	Coeff.	*p*-value	Coeff.	*p*-value	Coeff.	*p*-value
1	**13.18**	0.01⋆	**17.98**	0.0005⋆⋆	**14.08**	0.001⋆⋆	**8.73**	0.09⋆
2	1.30	0.81	**-10.39**	0.06⋆	-5.26	0.26	-3.16	0.571
3	3.03	0.57	**11.11**	0.04⋆	8.16	0.08	6.78	0.224
4	-8.79	0.10	**-9.85**	0.07⋆	**-11.35**	0.01⋆	-2.91	0.601
5	-2.31	0.60	-3.54	0.46	-1.799	0.64	-1.60	0.757

Result 3

Stock Market Price Against Google Bullishness



Result 4

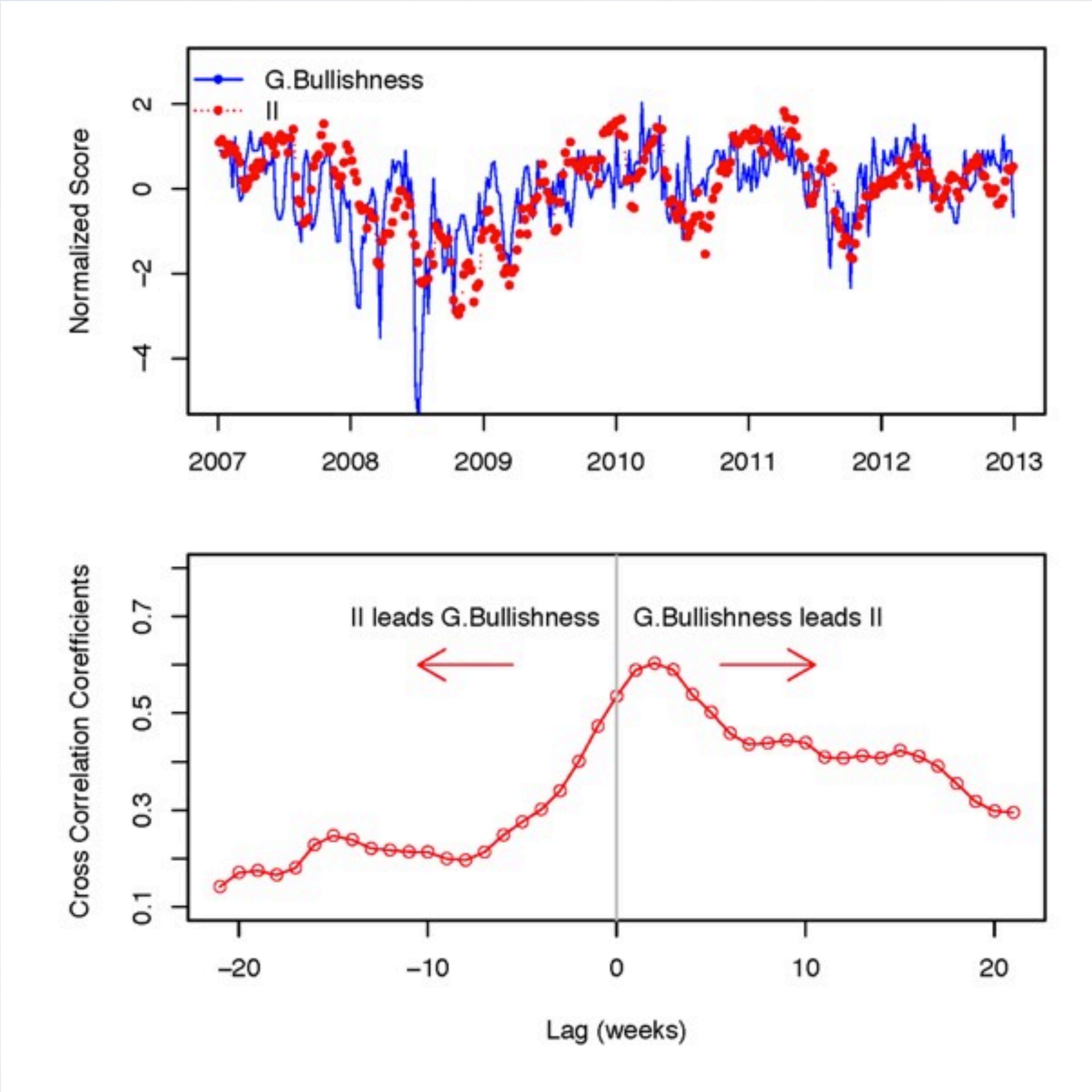
Predicting Weekly Stock Returns Using Google Bullishness

Bullishness	US.DJIA	UK.FTSE100	CA.GSPTSE	CN.SSE
ΔG^B_{w-1}	-21.48 (0.24)	18.36 (0.36)	3.84(0.84)	4.91 (0.87)
ΔG^B_{w-2}	6.65 (0.73)	23.68 (0.27)	16.09 (0.44)	20.0 (0.53)
ΔG^B_{w-3}	-19.92 (0.29)	0.14 (0.99)	1.83 (0.93)	-16.39(0.60)
ΔG^B_{w-4}	-17.71 (0.34)	8.40 (0.67)	-7.07 (0.71)	-25.84 (0.38)
G^B_{w-1}	-24.38 (0.32)	33.8(0.26)	13.93 (0.64)	25.11(0.71)
G^B_{w-2}	35.87 (0.21)	9.26(0.78)	24.54 (0.46)	47.40 (0.54)
G^B_{w-3}	-30.24 (0.29)	-32.76(0.32)	-14.29 (0.66)	-63.20 (0.41)
G^B_{w-4}	18.28 (0.44)	8.14(0.78)	-2.80 (0.92)	18.99(0.77)

Outside and inside the parentheses "()" are regression coefficients and p-values, respectively.

Result 4

Google Bullishness vs. Investor Intelligence (survey)



Pearson correlation
= **0.54** ($p < 0.01$)

Conclusion

Investor sentiment measurement from Twitter/Google complement (even substitute) survey measures.

Twitter Bullishness Index is predictive of daily stock returns.

Our results support investor sentiment theory.

Thank you!

Contacts

Huina Mao
Indiana University, Bloomington, USA
huinmao@indiana.edu

Dr. Scott Counts
Microsoft Research, Redmond, USA
counts@microsoft.com

Prof. Johan Bollen
Indiana University, Bloomington, USA
jbollen@indiana.edu