STATISTICAL ARBITRAGE TECHNIQUES USING SOCIAL AND NEWS SENTIMENT

ARUN VERMA

QUANT RESEARCH

BLOOMBERG

"You know, I've been dealing with these big mathematical models for forecasting the economy...if I could figure out a way to determine whether or not people are more fearful, or changing to more euphoric, and have a third way of figuring out which of the two things are working, I don't need any of this other stuff. I could forecast the economy better than any way I know how"

Alan Greenspan, November 2007

TALK OUTLINE

1. Extracting Sentiment from Story text
2. Sentiment Aggregation
3. Quantitative strategies
4. Topic Modeling and Sector level analysis

SENTIMENT ANALYSIS

News and social sentiment is captured as follows:

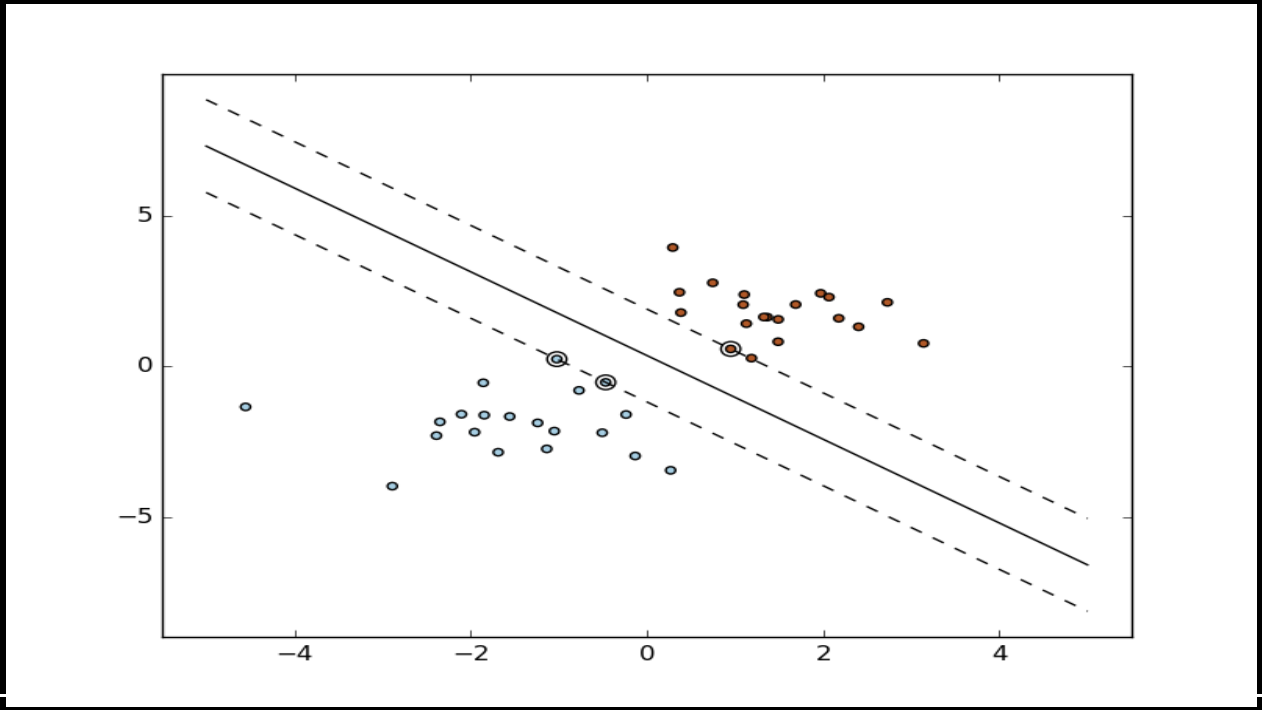
- Human experts assign positive/negative/neutral labels to stories in training set;
- The labeled data are fed into machine-learning models, such as support vector machine;
- Model assigns probability of being positive, negative and neutral to new stories Score: 1 for positive, -1 for negative, and 0 for neutral. We generate three way confidences - interpreted as probability of belonging to the label
- Company-level sentiment - weighted average of story-level sentiments – aggregated at daily and intra-day intervals.

SENTIMENT EXTRACTION & CLASSIFICATION

Support Vector machines help assign a category to each story - unfortunately standard SVMs can only do binary classification.

Use 3 SVMs to help with pairwise classifications, one each for positive/negative, positive/neutral and neutral/negative classification

The features are multidimensional – bag of words framework is used.

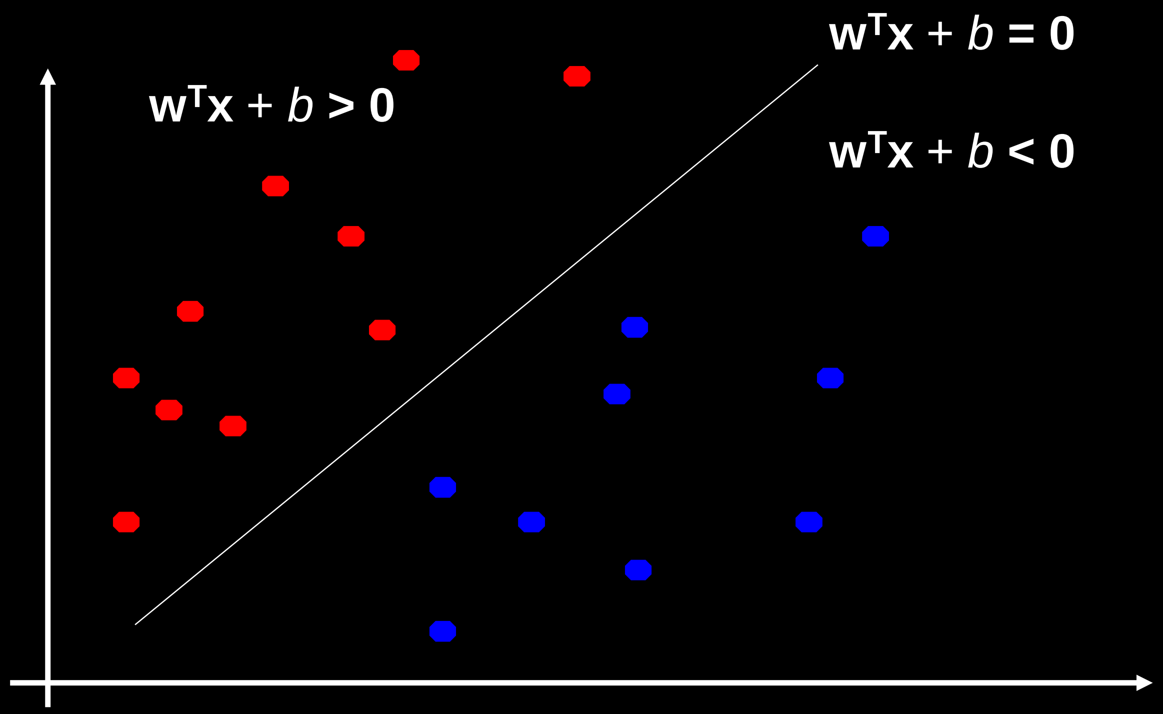


SVM FEATURES

- Keywords included in the story that are part of a (domain specific) dictionary are used as features
- This bag of words form the feature vectors in a very high dimensional space
- SVMs operate in this high dimensional space – we find a hyper plane vector which separates the two classes optimally (it’s a supervised machine learning process needing a lot of training data)

SVM BASICS

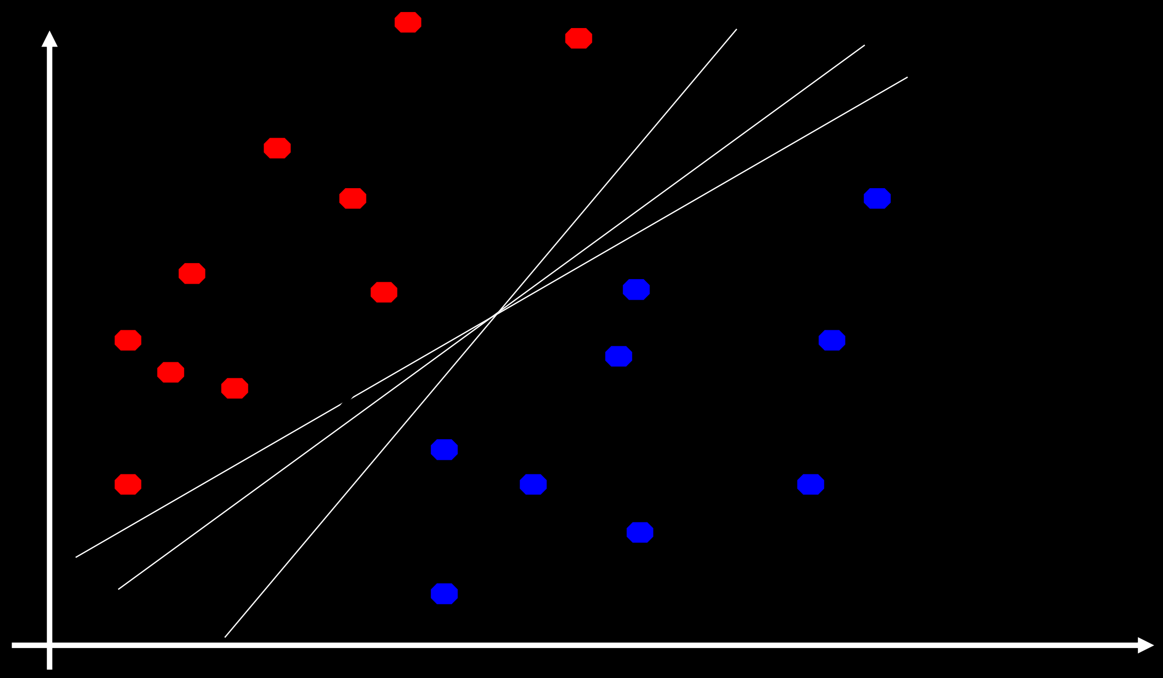
Binary classification can be viewed as the task of separating classes in feature space:



$f(\mathbf{x}) = \text{sign}(\mathbf{w}^T\mathbf{x} + b)$

SVM

Which of the linear separators is optimal?

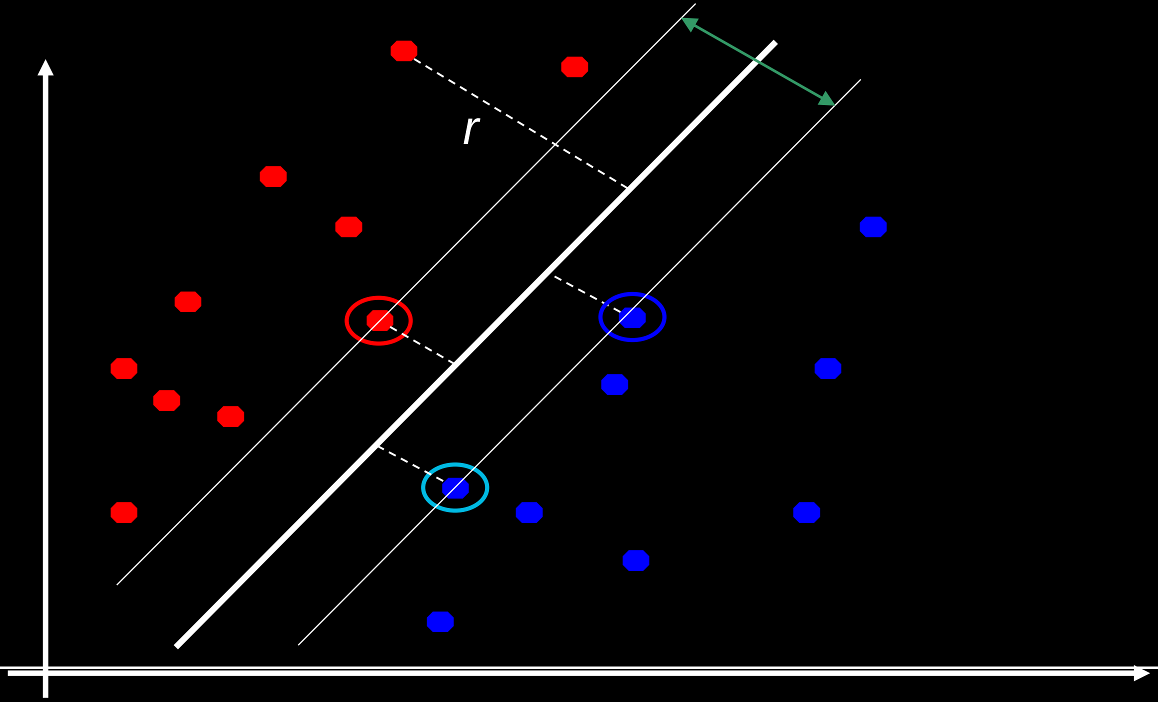


CLASSIFICATION MARGIN

Distance from example $\mathbf{x}_i$ to the separator is $r = \frac{w^T x + b}{\|w\|}$

Examples closest to the hyperplane are ***support vectors***.

Margin ρ of the separator is the distance between support vectors.



Bloomberg

SOLVING THE OPTIMIZATION PROBLEM

Find **w** and b such that
$\boldsymbol{\Phi}(\mathbf{w}) = \frac{1}{2}\, \mathbf{w}^T\mathbf{w}$ is minimized and for all $\{(\mathbf{x_i}, y_i)\}$
$y_i\,(\mathbf{w^T x_i} + b) \geq 1$

Need to optimize a *quadratic* function subject to *linear* constraints.

Quadratic optimization problems are a well-known class of mathematical programming problems, and many (rather intricate) algorithms exist for solving them.

The solution involves constructing a *dual problem* where a *Lagrange multiplier* α_i is associated with every constraint in the primary problem:

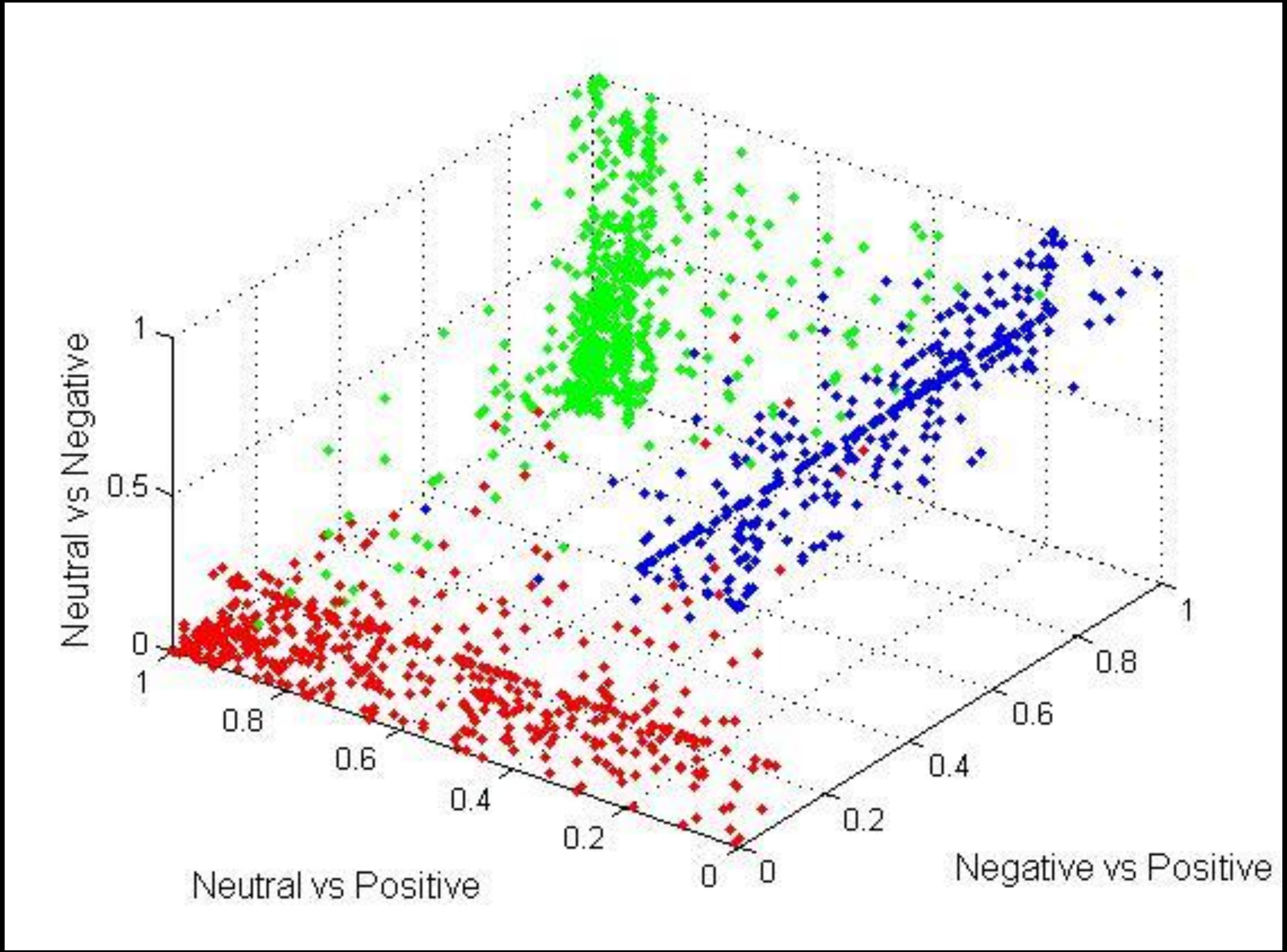
Find $\alpha_1 \ldots \alpha_N$ such that
$\mathbf{Q}(\boldsymbol{\alpha}) = \Sigma\alpha_i - \frac{1}{2}\Sigma\Sigma\alpha_i\alpha_j y_i y_j \mathbf{x_i}^T\mathbf{x_j}$ is maximized and
(1) $\Sigma\alpha_i y_i = 0$
(2) $\alpha_i \geq 0$ for all α_i

INTERPRETING SVM RESULT AS PROBABILITY

- Probability metric constructed based on distance from optimal hyperplane
- Logistic functional used:

$$P(y=1 \mid x) = \frac{1}{1+e^{\frac{w^T x+b}{\sigma}}}$$

THREE PAIRWISE SVM RESULTS



THREE WAY SENTIMENT METHODOLOGY

For a target story to be classified we run the three SVM on it to find the three values.

Define the Euclidean distance between two stories i, j :

$$d(i,j)^2 = (x(i) - x(j))^2 + (y(i) - y(j))^2 + (z(i) - z(j))^2$$

Use Nearest Neighbors: We define weights for an in-sample news story j as :

$$w(j) = \mathrm{e}^{-\frac{d(target,j)^2}{2\mathrm{B}^2}}$$

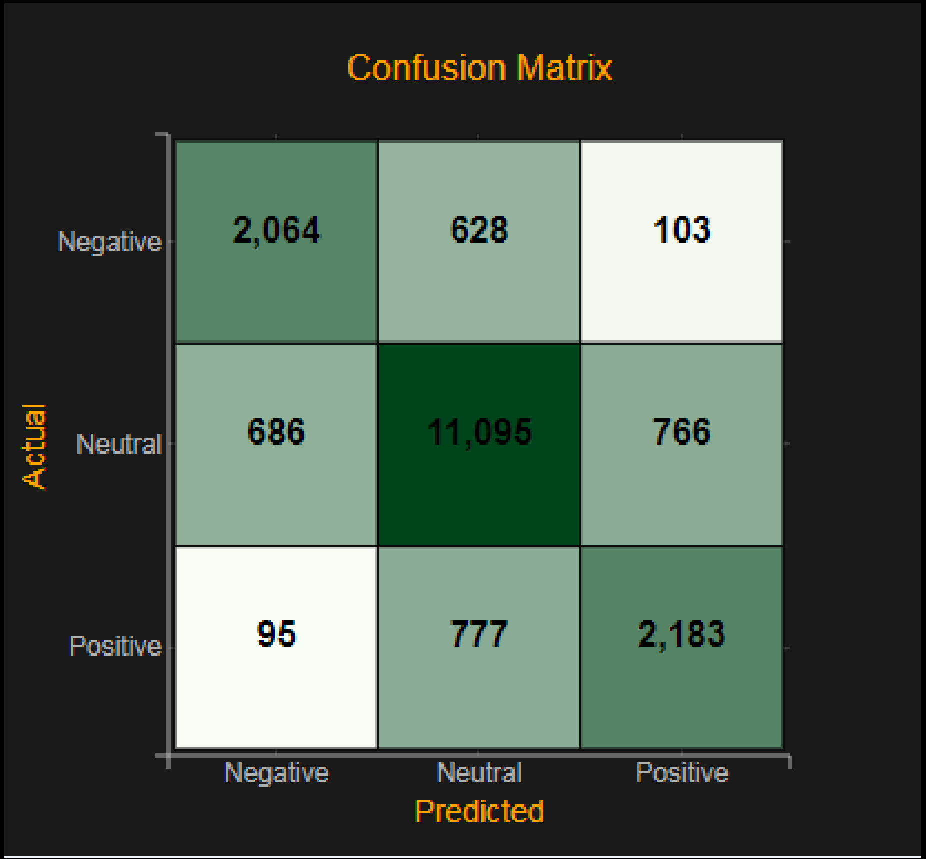
Where B is a bandwidth parameter. Normalize the weights as follows:

$$\overline{w}(j) = \frac{w(j)}{\sum w(j)}$$

Now, define the positive probability as $\sum_{j \in Pos} \overline{w}(j)$, negative probability as $\sum_{j \in Neg} \overline{w}(j)$ and neutral probability as $\sum_{j \in Neut} \overline{w}(j)$.

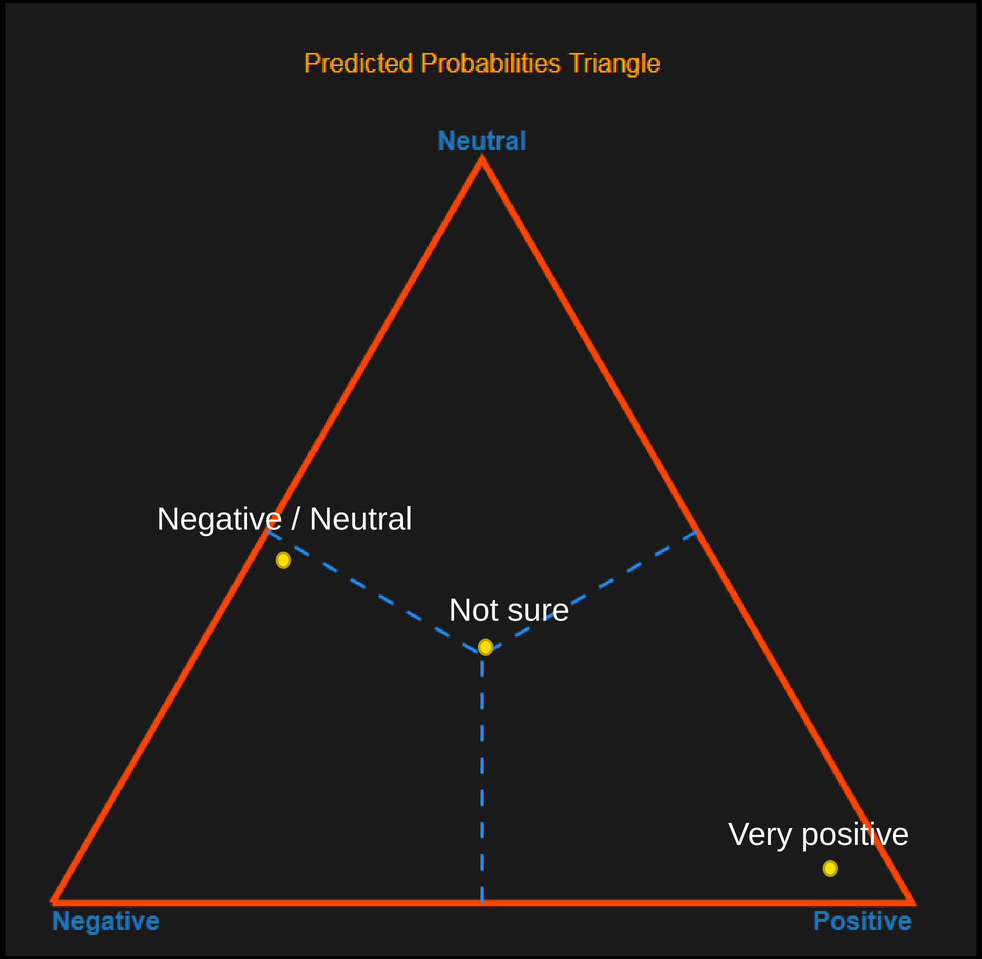
CONFUSION MATRIX

- K x K matrix where K = number of classes
- Cell[i, j] = number of samples whose:
  - Actual label = i
  - Predicted label = j
- Diagonal entries - correct predictions
- Off diagonal entries - misclassifications

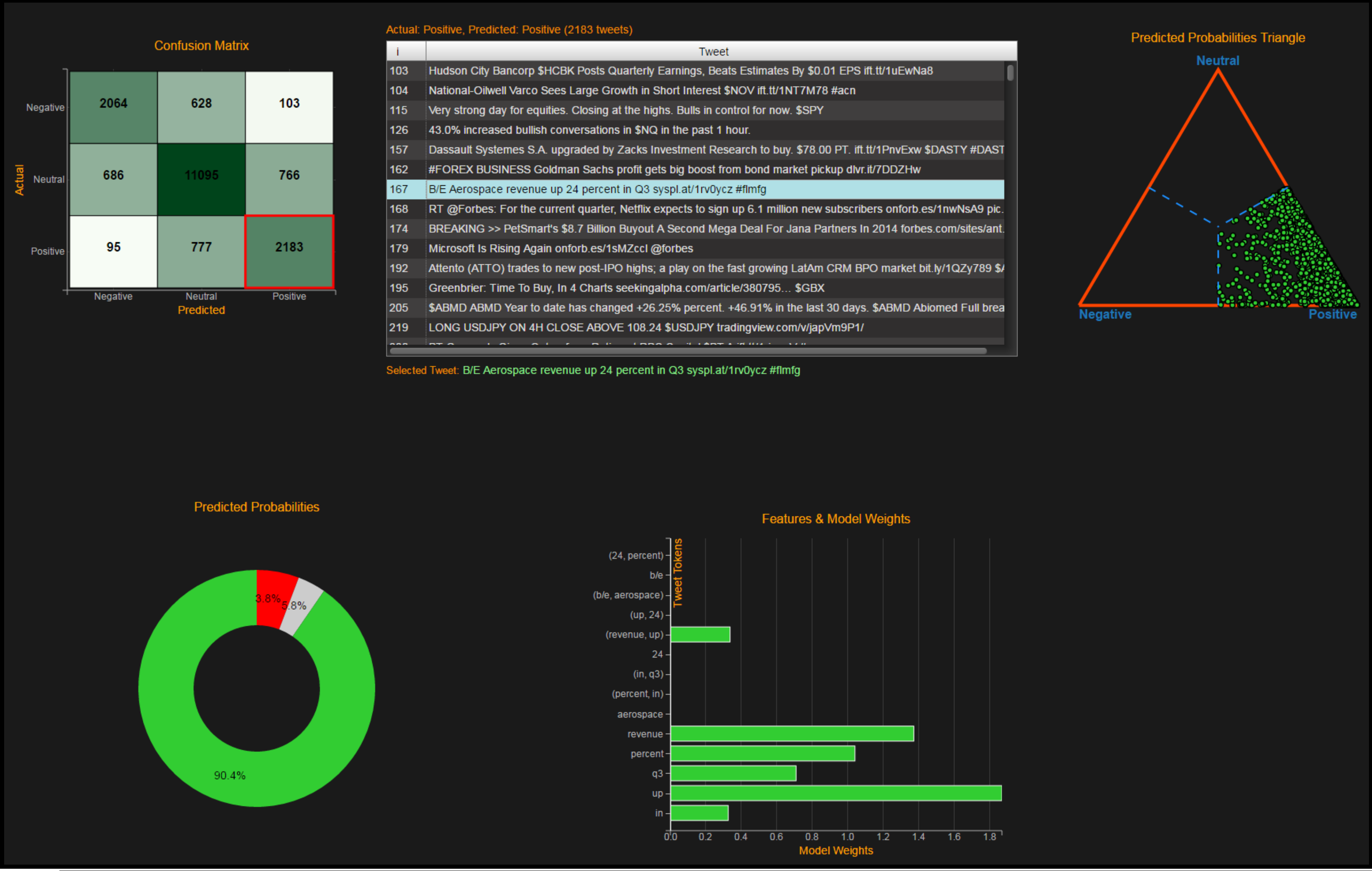


TRIANGLE VISUALIZATION

- Model returns 3 probabilities (which sum to 1)
- How can we visualize these 3d “points”?
- Points inside an equilateral triangle

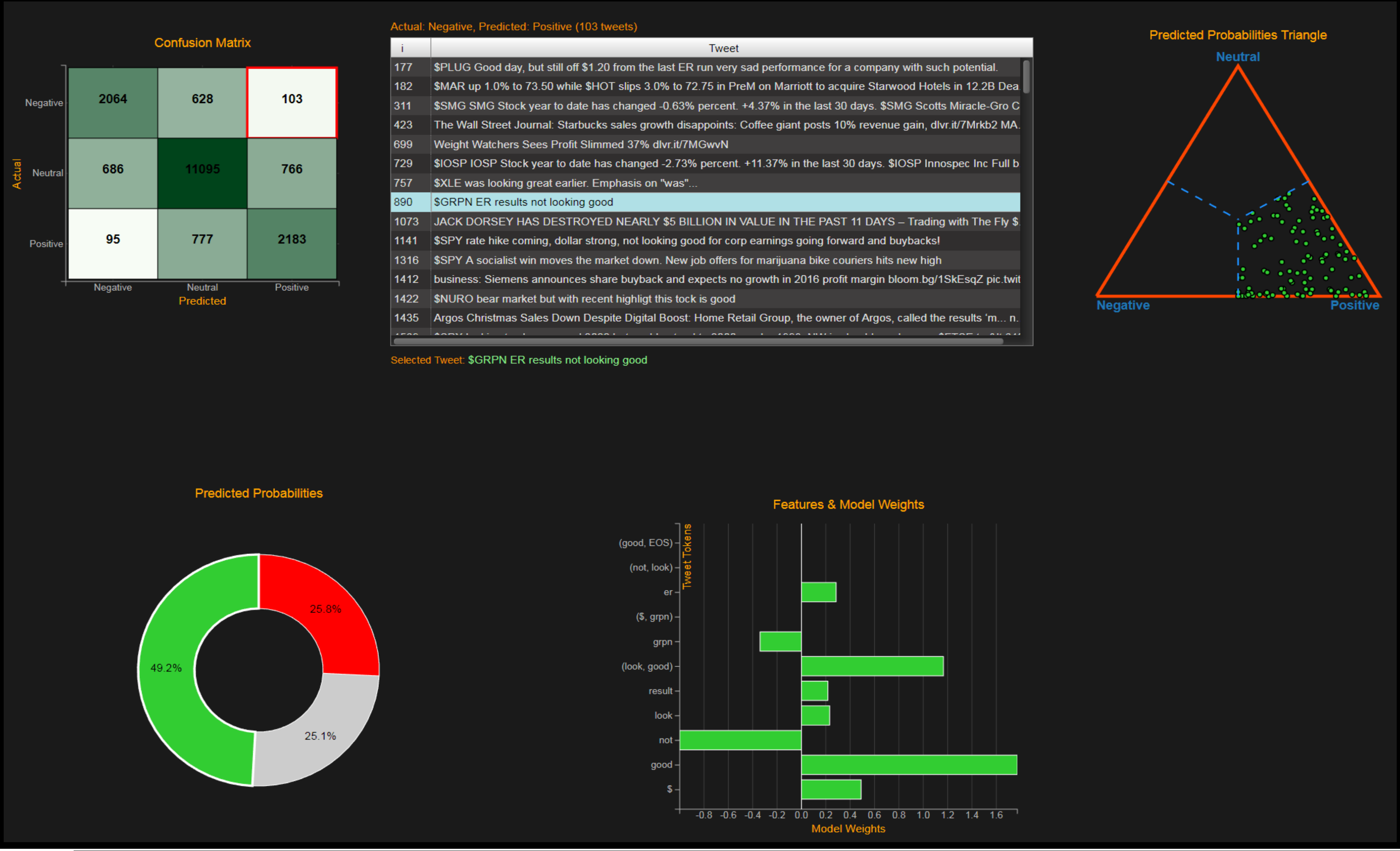


CLASSIFICATION OF TWEETS



Bloomberg

ANALYZE MISCLASSIFICATIONS



SENTIMENT AGGREGATION METHODS

We take all stories tagged with a particular company:

1) Each news story or tweet is scored with "confidences“ C_+, C_-, C_n for positive, negative and neutral sentiment respectively.
2) Story specific sentiment (polarity score) : $S^i = C_+^i - C_-^i$
3) Sentiment Average :

$$\mu = \frac{\sum_{i=1}^{N} S^i}{N} = \frac{\sum_{i=1}^{N}(C_+^i - C_-^i)}{N} = \overline{C_+} - \overline{C_-}$$

4) Sentiment dispersion = Inter-story variance + story specific dispersion

Simplifies to: $\overline{C_+} + \overline{C_-} - \mu^2$

DISPERSION CALCULATION

Sentiment Dispersion: To calculate the overall dispersion we need to track two components.

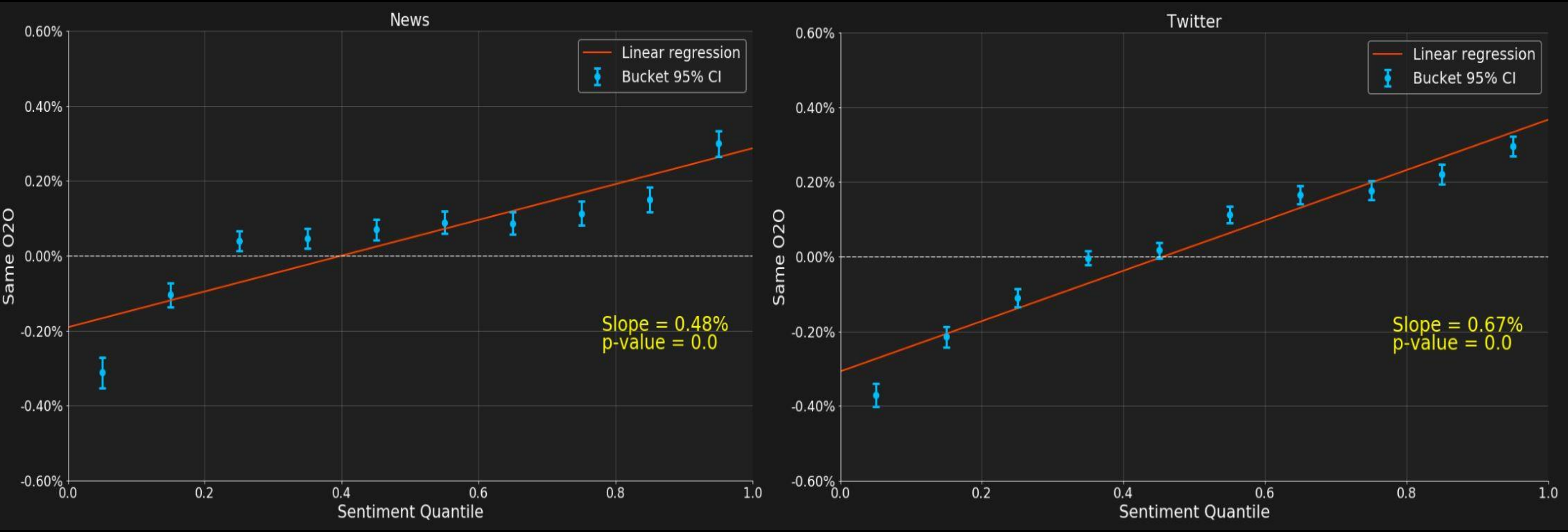
- **INTRA STORY VARIANCE** – Variance within one story due to classification uncertainty

The variance of the sentiment per story is defined as the variance of the trinomial probability distribution . $Var^i = \mathrm{E}\left[\left(\mathrm{X}^i -\right.\right.$

Romance should never begin with sentiment. It should begin with science and end with a settlement. — Oscar Wilde, An Ideal Husband

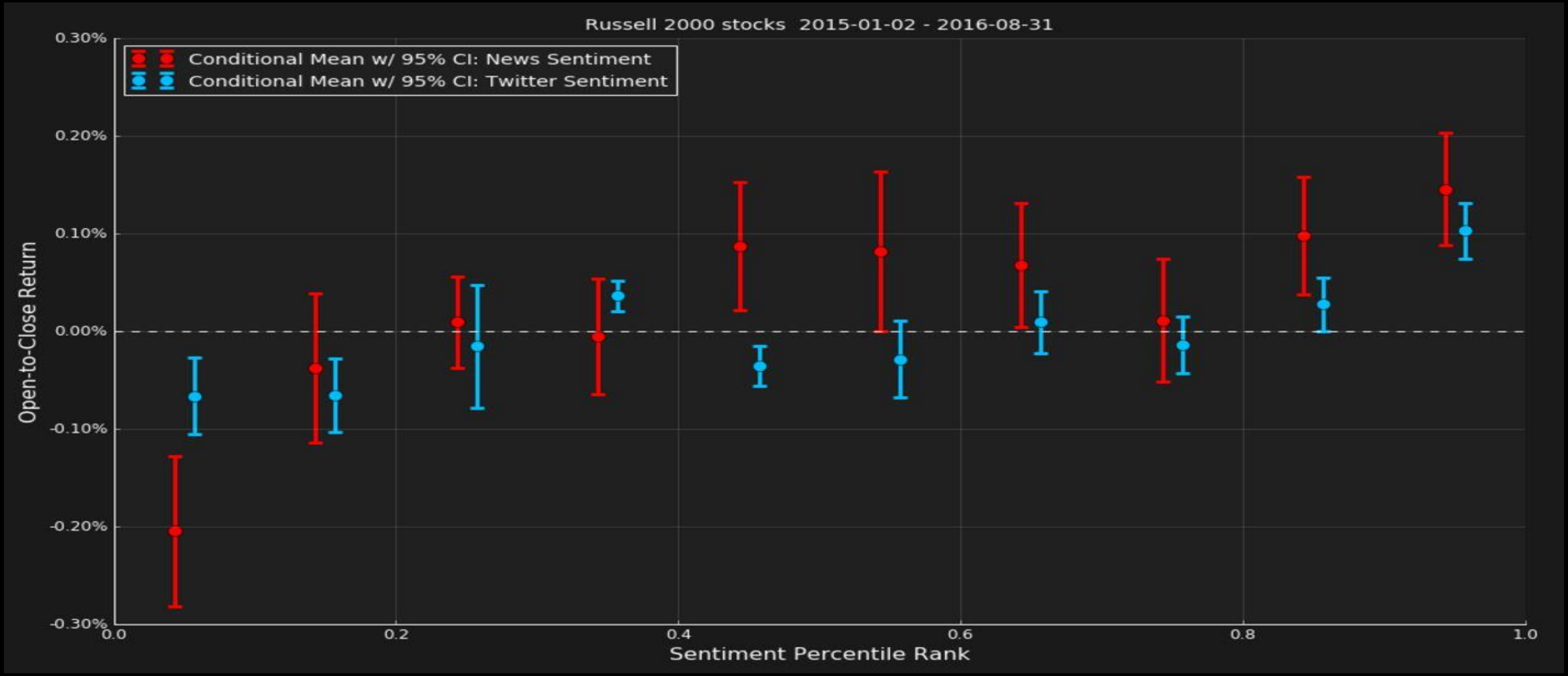
SENTIMENT VS CONTEMPORANEOUS RETURN

- The daily sentiment score is constructed using story level sentiment data over the 24 hours period, and calculated around 9:05 am every day.



TRADING SIGNAL (CONDITIONAL DRIFT)

Does market efficiently price in the sentiment information?



STRATEGY I: DAILY LONG-SHORT

Does market efficiently price in the sentiment information?

Strategy:

- Each day before market open, rank all stocks by their daily sentiment
- Three variations:

  1) Long top 1/3 and short bottom 1/3 stocks

  2) Long top 5% and short bottom 5% stocks

  3) Positions proportional to the difference of sentiment scores from its cross-sectional mean

- Create portfolio at market open and close out at market close

We backtest the strategy for Russell 2000 stocks from Jan-01-2015 to Jul-31-2016.

DAILY STRATEGY

The portfolio daily return can be computed as the following,

$$Ret_j = \sum_{i \in Long_j} w_{ij}^{Long} \left(\frac{P_{ij}^{close}}{P_{ij}^{open}} - 1 \right) - \sum_{i \in Short_j} w_{ij}^{Short} \left(\frac{P_{ij}^{close}}{P_{ij}^{open}} - 1 \right)$$

Where

Ret_j is portfolio return on day j;

P_{ij}^{close} is the close price of stock i on day j, P_{ij}^{open} is the open price of stock i on day j;

$Long_j$ is the basket of stocks to long on day j, w_{ij}^{Long} is the weight of stock i in $Long_j$;

$Short_j$ is the basket of stocks to short on day j, w_{ij}^{Short} is the weight of stock i in $Short_j$;

For HML portfolio, $w_{ij}^{Long} = \frac{1}{\#\ of\ Stocks\ in\ Long_j}$, $w_{ij}^{Short} = \frac{1}{\#\ of\ Stocks\ in\ Short_j}$

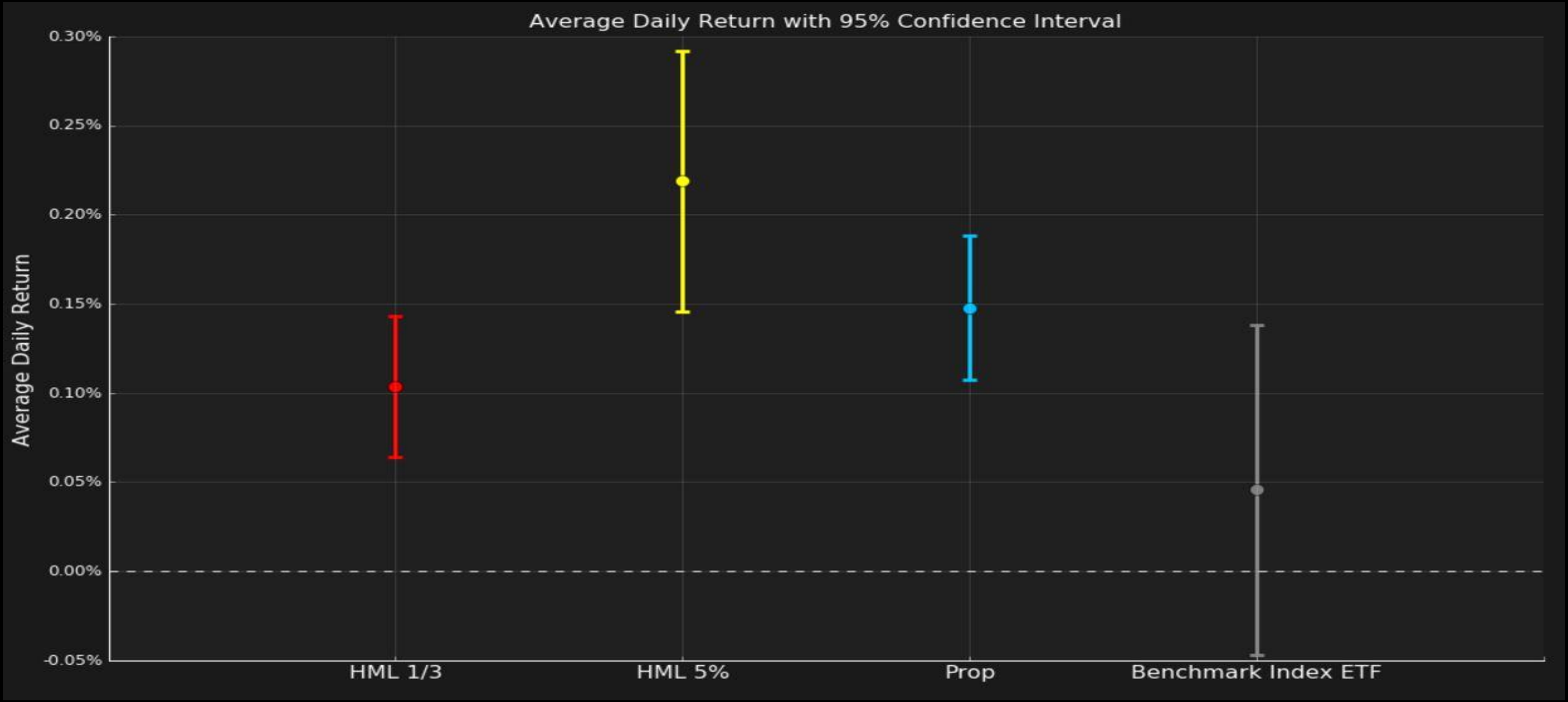
For proportional portfolio, $w_{ij}^{Long} = \frac{SS_{ij}^{Long} - \mu_j}{\sum_{i \in Long_j}(SS_{ij}^{Long} - \mu_j)}$, $w_{ij}^{Short} = \frac{SS_{ij}^{Short} - \mu_j}{\sum_{i \in Short_j}(SS_{ij}^{Short} - \mu_j)}$

μ_j is the cross-sectional mean of the company-level sentiment on day j;

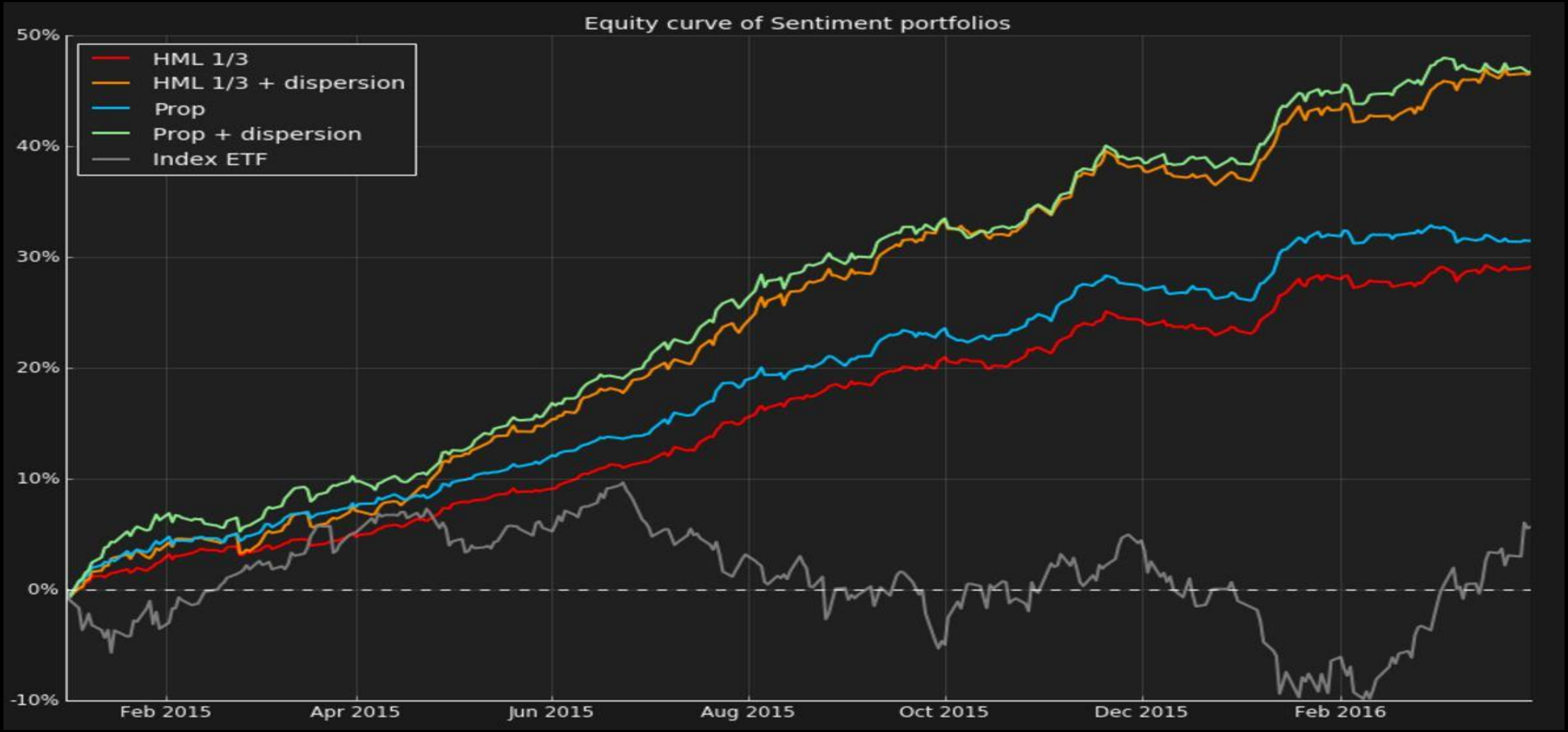
SS_{ij}^{Long} is the sentiment score of stock i in $Long_j$ on day j

SS_{ij}^{Short} is the sentiment score of stock i in $Short_j$ on day j

RUSSELL 2000 STOCKS



RUSSELL 2000 STOCKS



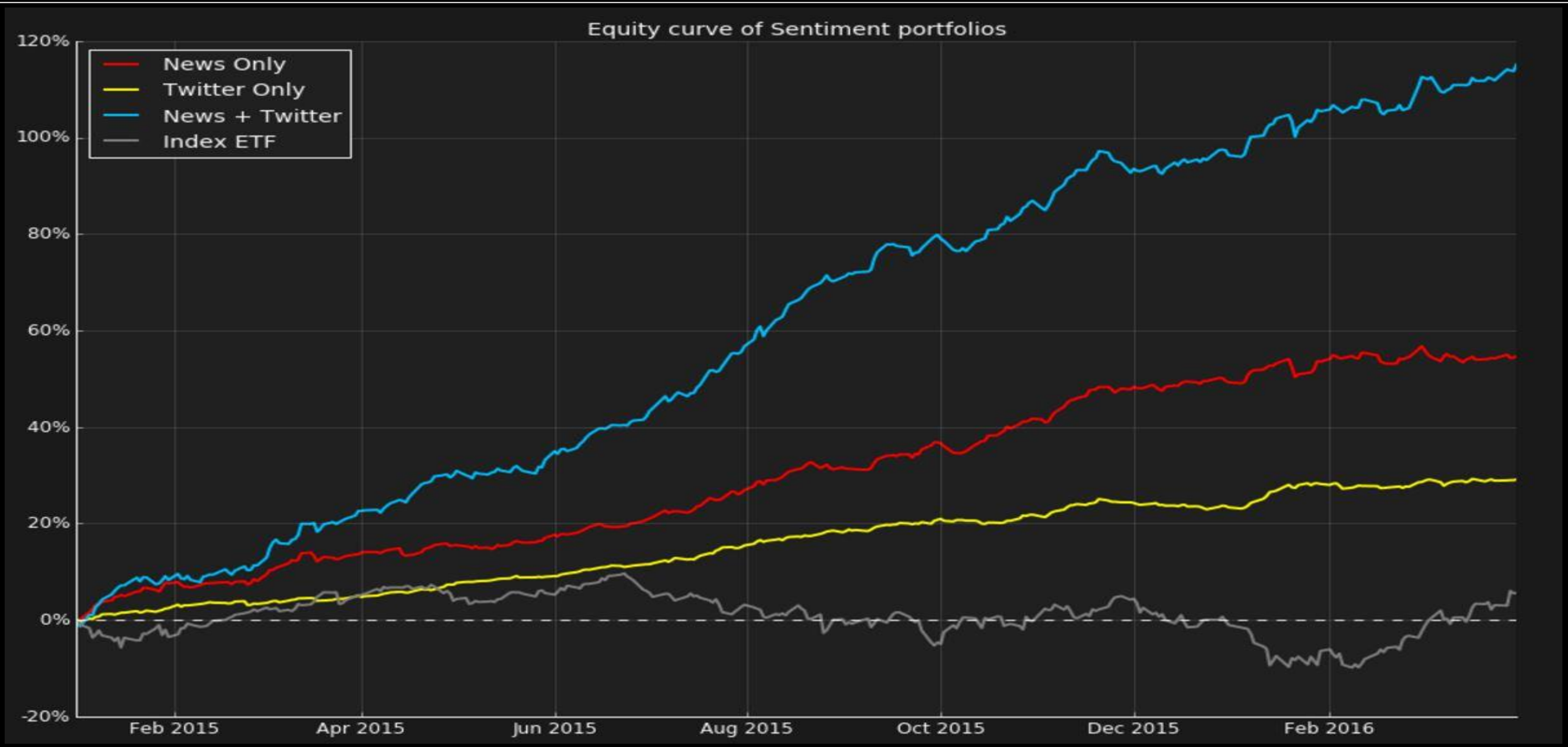
Bloomberg

RUSSELL 2000 STOCKS

Performance Statistics:

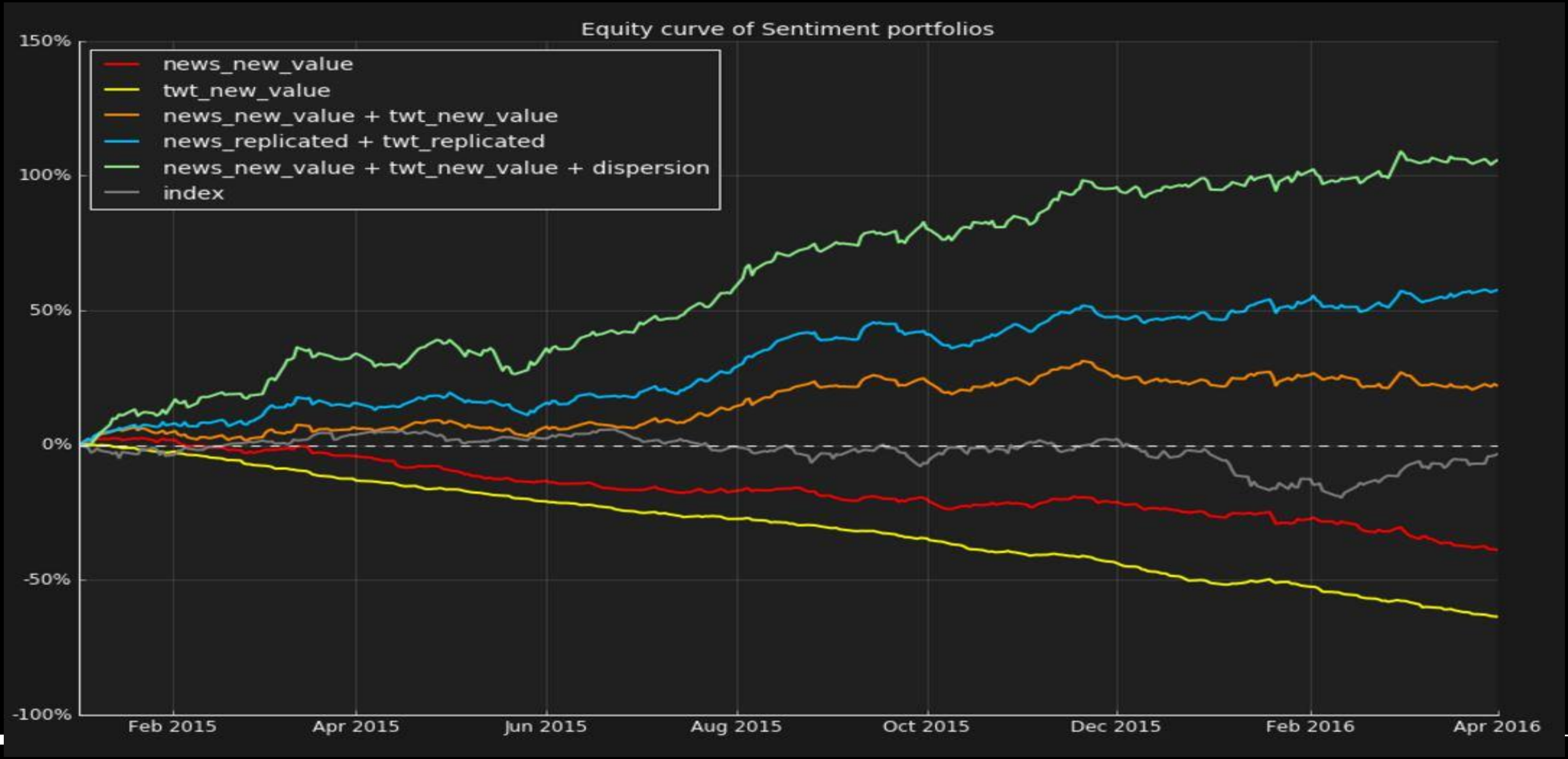
	Beta	Annualized Ret	Annualized Vol	Sharpe
HML 1/3	-0.04	23%	4%	5.37
HML 1/3 + dispersion	-0.08	38%	7%	5.48
Prop	-0.06	25%	5%	5.06
Prop + dispersion	-0.06	38%	7%	5.39
Index ETF (IWM)	1.00	5%	15%	0.30

RUSSELL 2000 –COMBINING NEWS & TWITTER



	Beta	Annualized Ret	Annualized Vol	Sharpe
News Only	-0.15	44%	9%	5.08
Twitter Only	-0.04	23%	4%	5.35
News + Twitter	-0.12	92%	15%	6.25
Index ETF (IWM)	1.00	5%	15%	0.30

RUSSELL 2000 – WITH TRANSACTION COST



STRATEGY II: EARNINGS EVENT DRIVEN

Does sentiment before earnings have predicting power for earning day return?

Strategy:

Each day, if there are companies to release earnings,

- Long positive sentiment and short negative sentiment;

  If portfolio has only long positions, short market ETF;

  If portfolio has only short positions, long market ETF;

- Create at market open and close out at the next market open.

Sentiment sources: News only, Twitter only, News/Twitter combined

We backtest the strategy for S&P 500 stocks from Jan-01-2015 to Jul-31-2016.

EVENTS DRIVEN STRATEGY

The portfolio daily return can be computed as the following,

$$Ret_j = \sum_{i \in Long_j} \frac{1}{N_j^{Long}} \left(\frac{P_{i(j+1)}^{open}}{P_{ij}^{open}} - 1 \right) \mathbb{1}\left(N_j^{Long} > 0\right) - \sum_{i \in Short_j} \frac{1}{N_j^{Short}} \left(\frac{P_{i(j+1)}^{open}}{P_{ij}^{open}} - 1 \right) \mathbb{1}\left(N_j^{Short} > 0\right)$$

Where

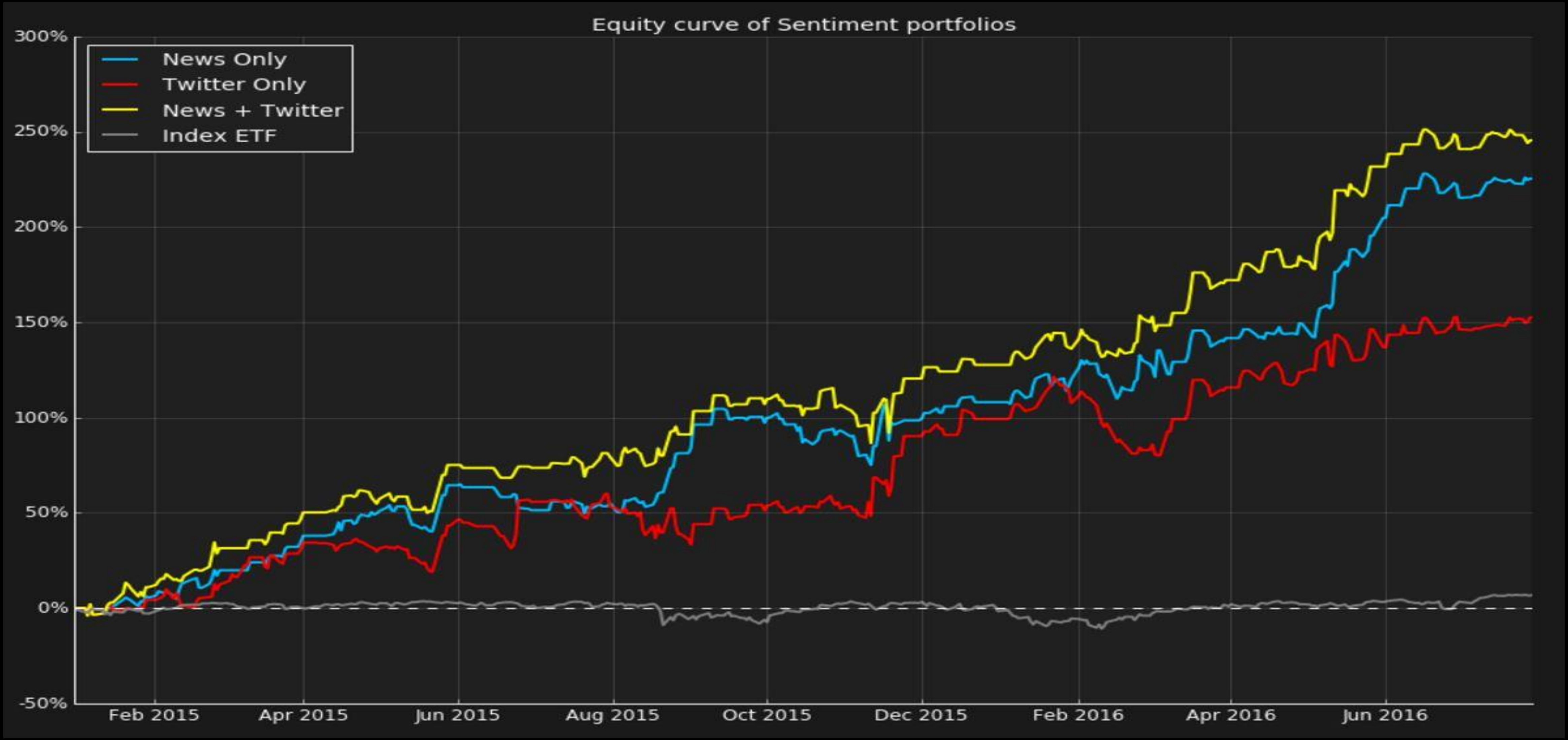
Ret_j is portfolio return on day j;

P_{ij}^{open} is the open price of stock i on day j, $P_{i(j+1)}^{open}$ is the open price of stock i on day j+1;

$Long_j$ is the basket of stocks to long on day j, N_j^{Long} is the number of stocks in $Long_j$;

$Short_j$ is the basket of stocks to short on day j, N_j^{Short} is the number of stocks in $Short_j$;

S&P 500 STOCKS



S&P 500 STOCKS

Performance Statistics:

	Beta	Annualized Ret	Annualized Vol	Sharpe	Average # of Short	Average # of Long
News Only	0.08	140%	59%	2.37	2	7
Twitter Only	-0.14	108%	61%	1.77	4	7
News + Twitter	0.12	156%	58%	2.68	2	5
Index ETF (SPY)	1.00	4%	15%	0.29	NaN	NaN

SHARPENING THE SENTIMENT SIGNAL – USING MEAN VARIANCE OPTIMIZATION

Instead of using sentiment directly to determine weights, we could use sentiment as an expected return for next day and optimize the portfolio

$$minimize\ w^T C w$$

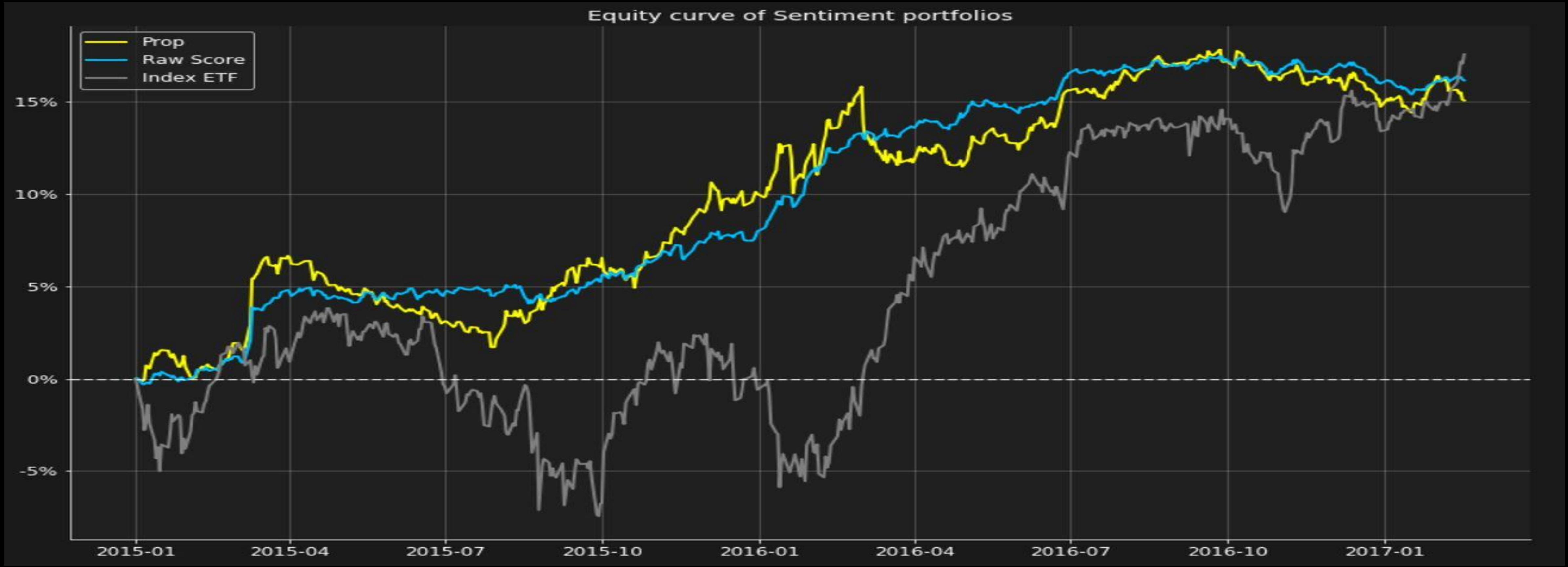
s.t. $w^T r \geq \mu$

$$w^T 1 = 0$$

$$r = S - E[S]$$

Bloomberg

MEAN VARIANCE OPTIMIZATION

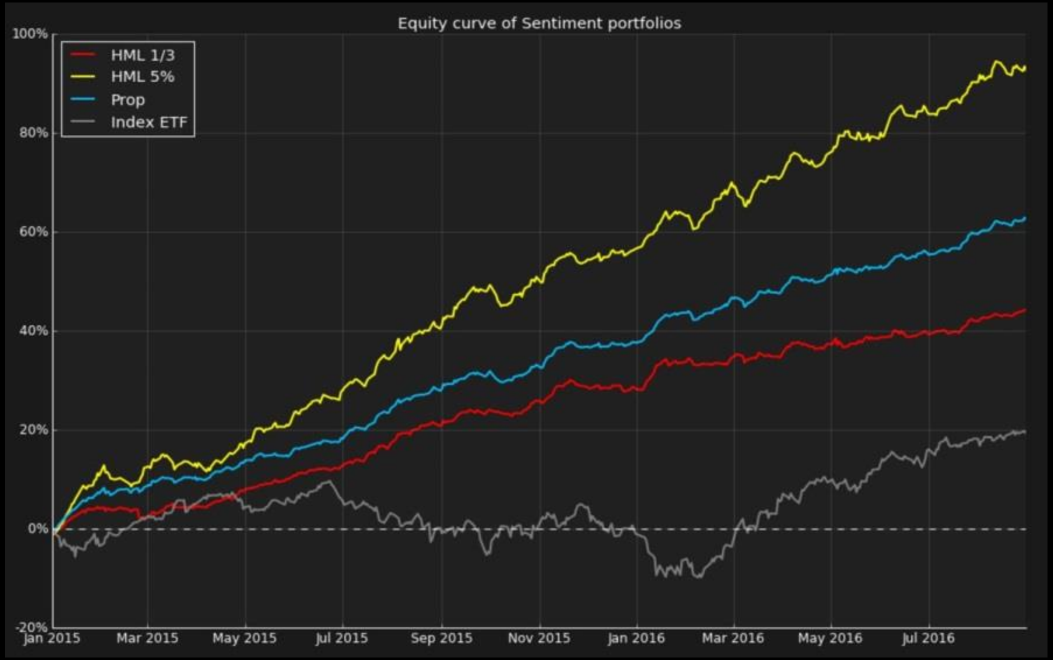


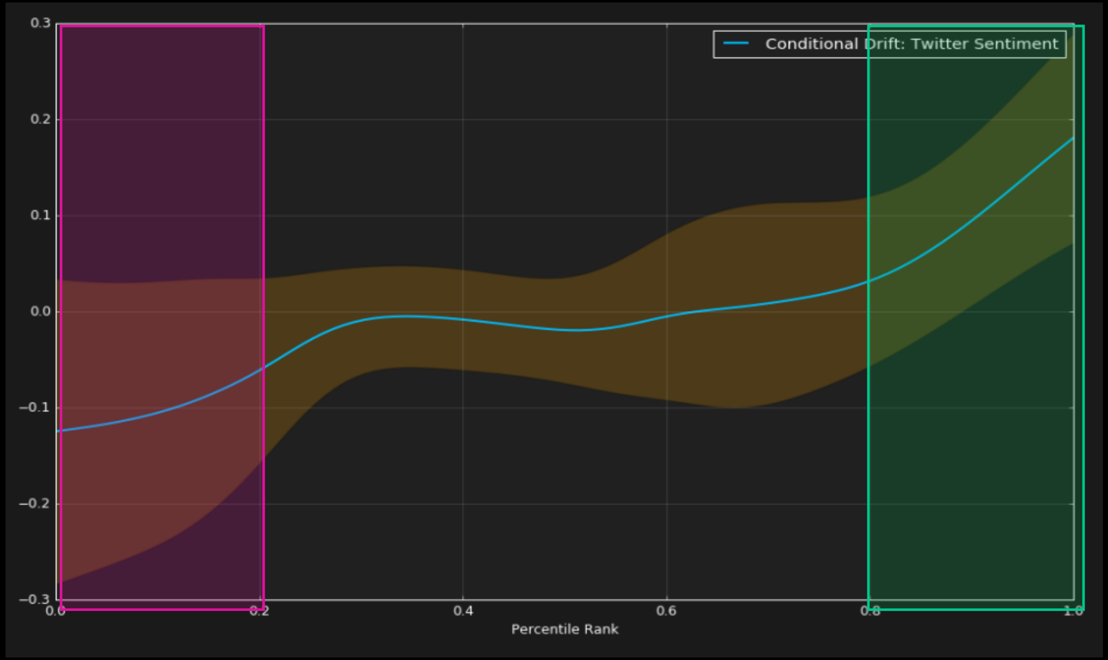
	Beta	Annualized Ret	Annualized Vol	Sharpe	Avg # of Short	Avg # of Long
Proportional	-0.10	7%	6%	1.28	169	133
Raw score	-0.02	8%	3%	2.54	162	136
Index ETF	1.00	8%	10%	0.80	NaN	NaN

Bloomberg

STRATEGY WORKS FOR SMALL CAPS

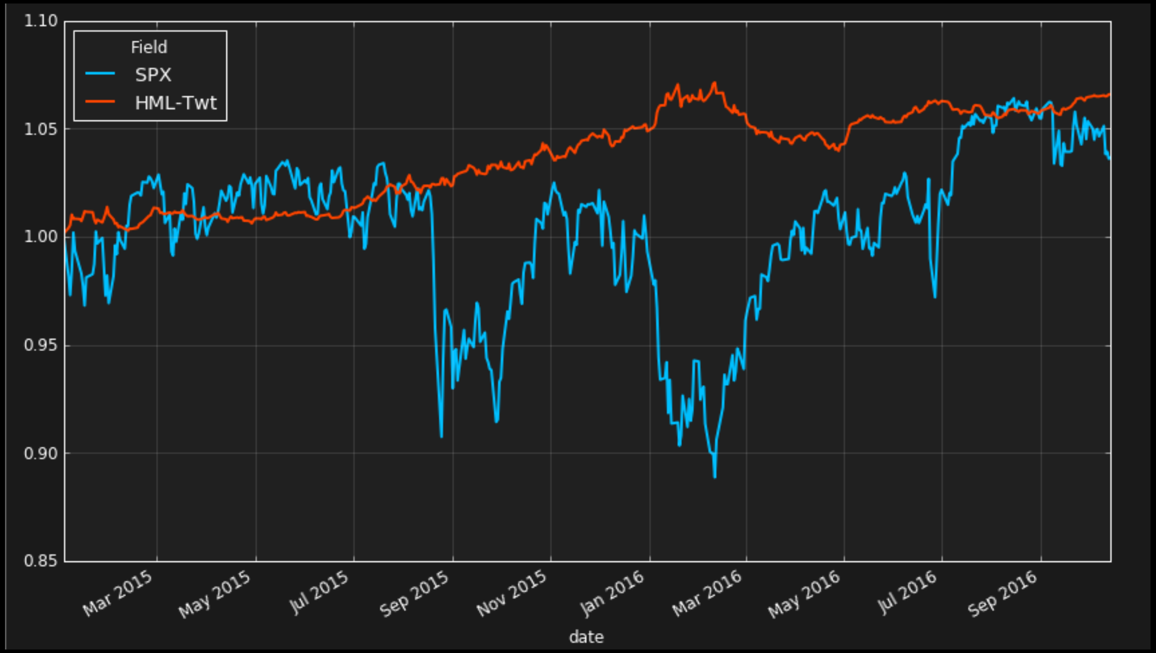
Russel 2000

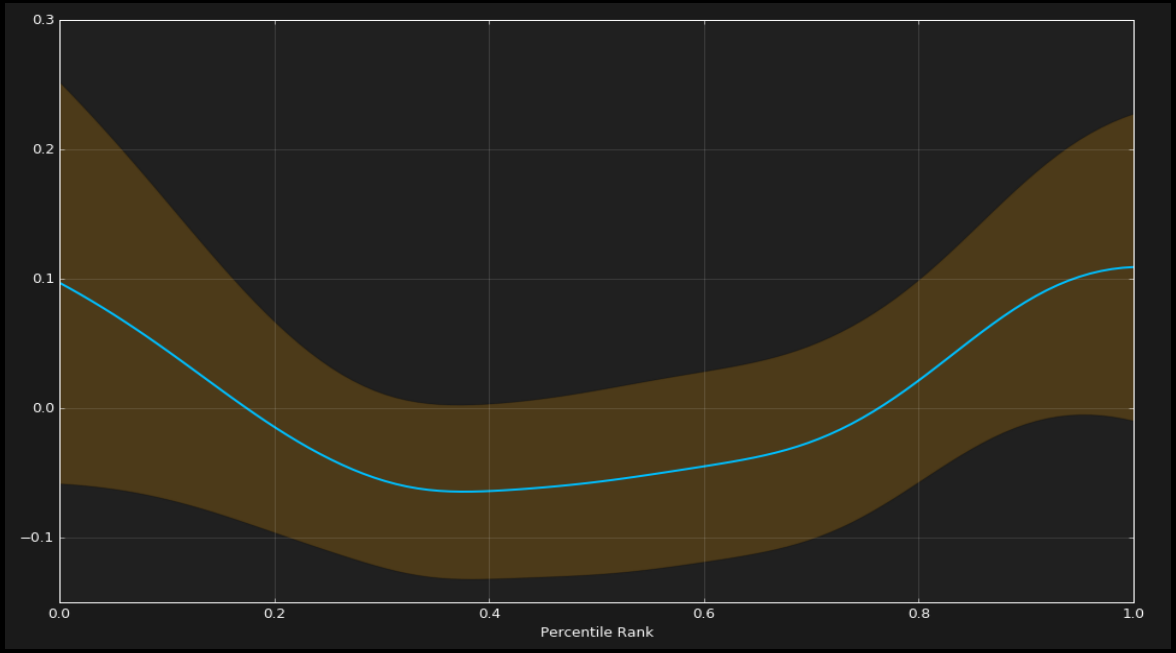




BUT NOT FOR SPX STOCKS

S&P 500

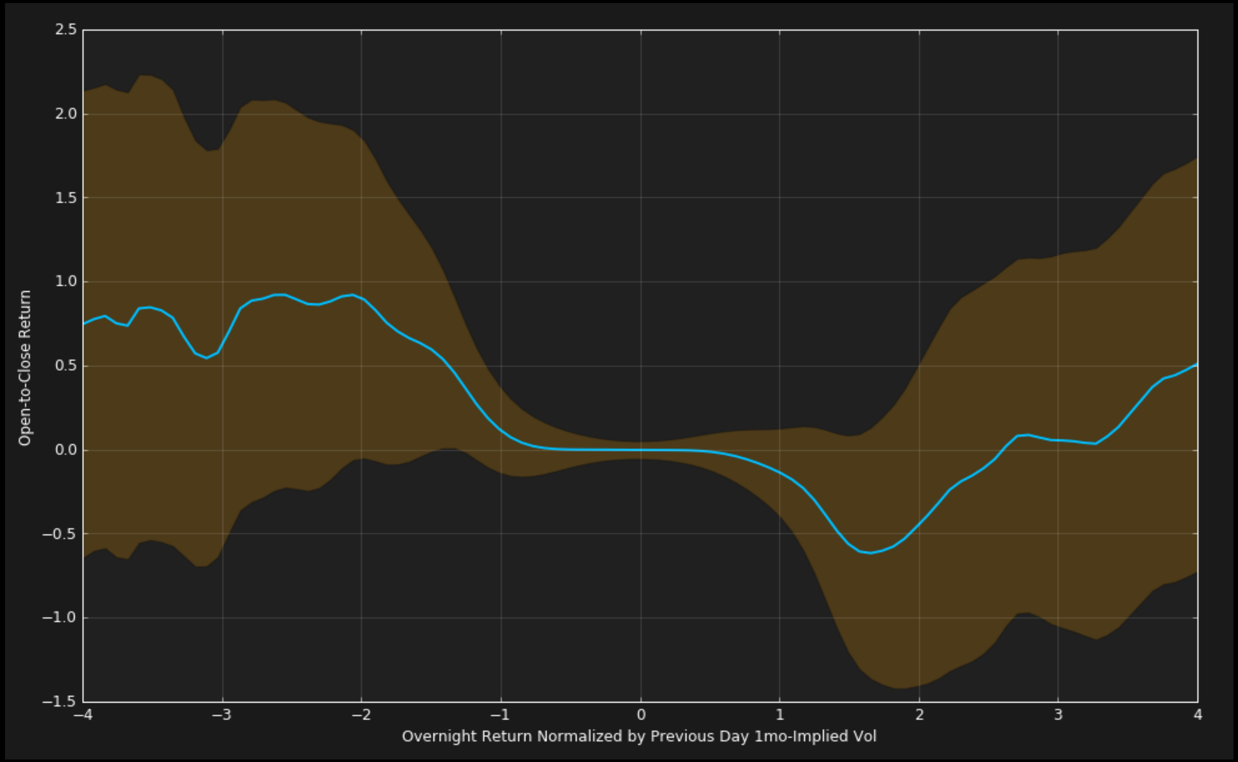




SENTIMENT USEFUL ONLY FOR SMALL CAP?

Here is a mean-reversion “signal” that works for S&P 500 stocks

Return(Open_t, Close_t) vs Return(Close_(t-1), Open_t)/30d_impvol_(t-1)

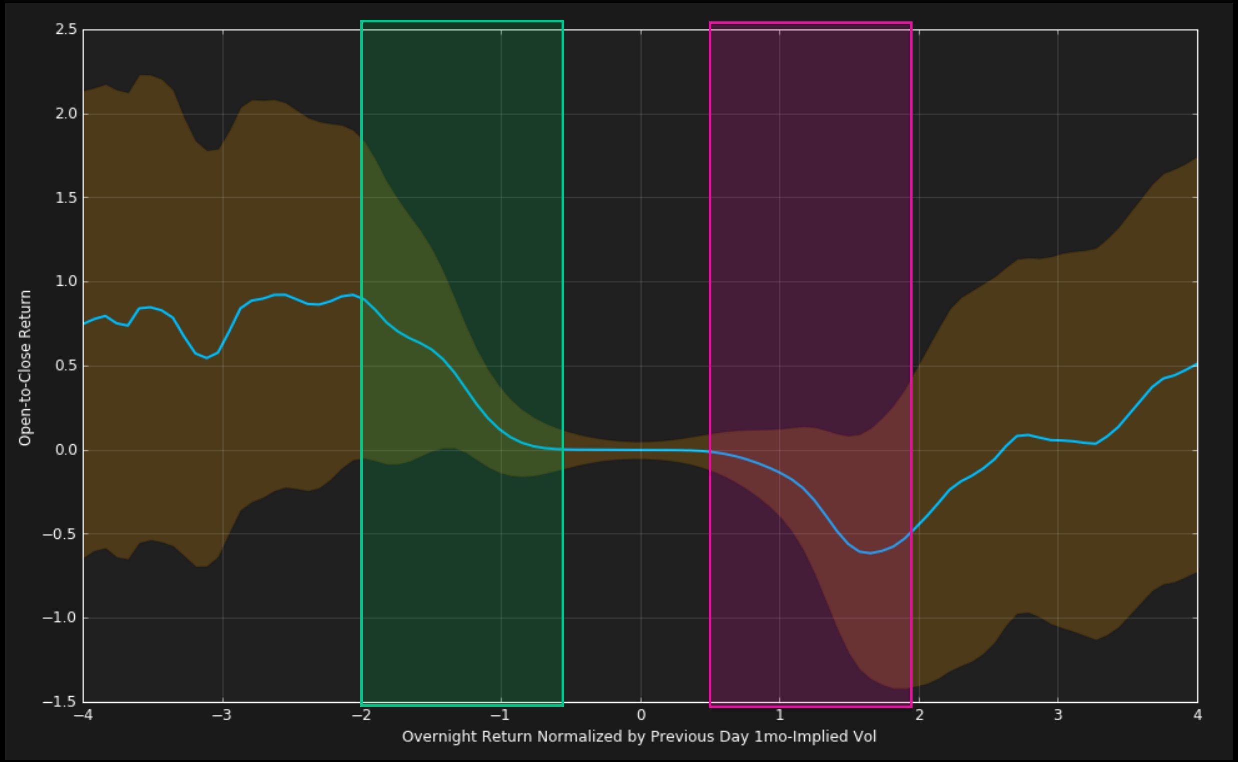


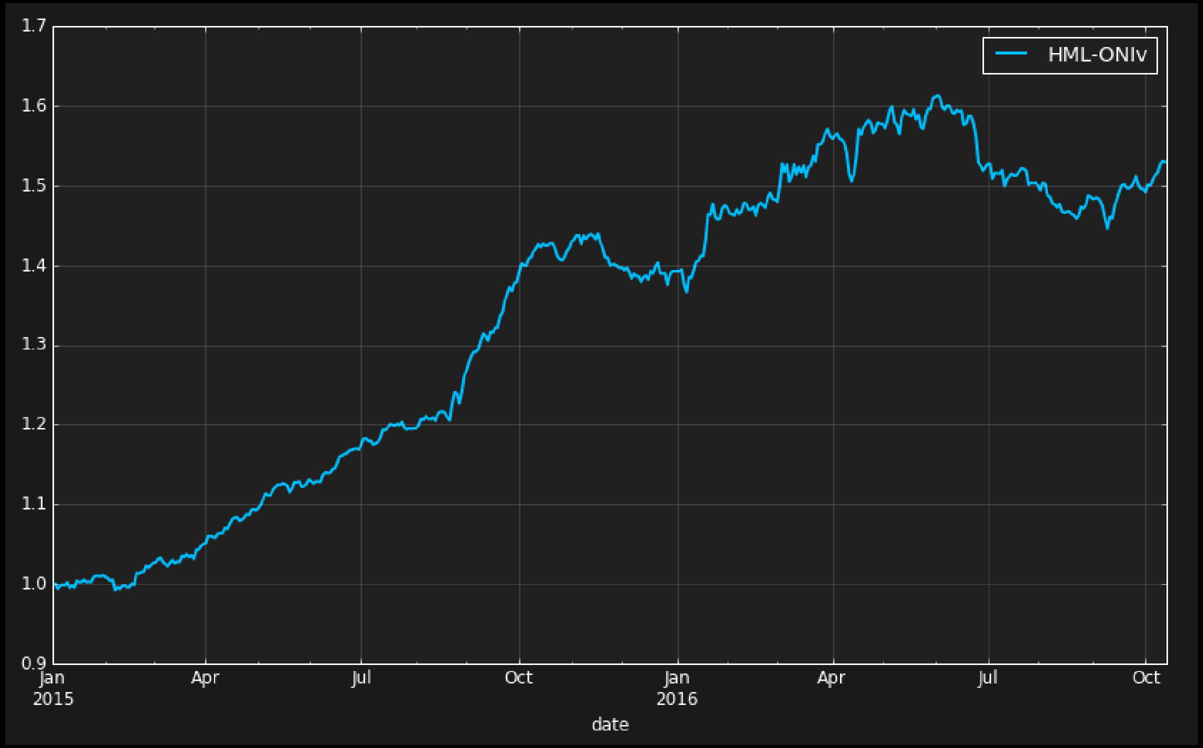
Bloomberg

MOMENTUM STRATEGY

Unless one can observe and trade immediately at open

Return(Open_t, Close_t) vs Return(Close_(t-1), Open_t)/30d_impvol_(t-1)

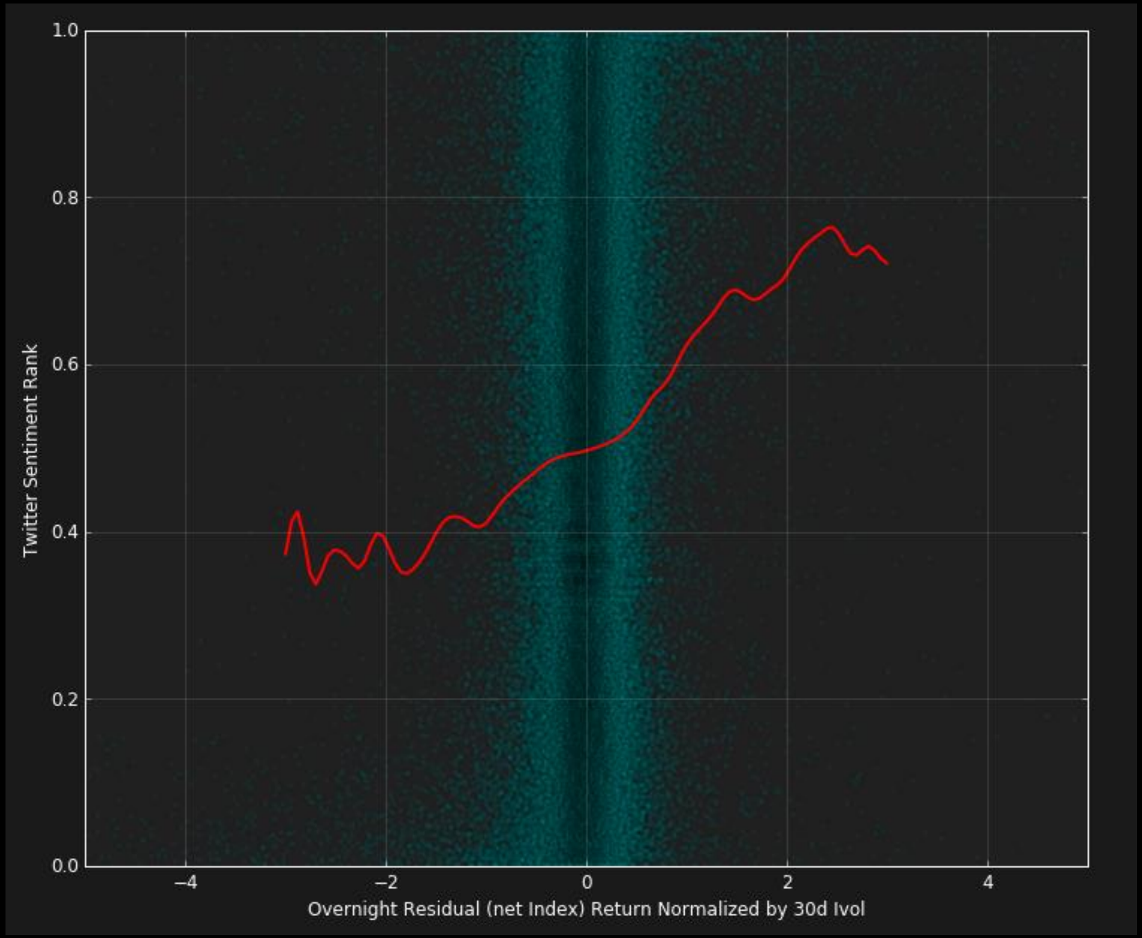




Bloomberg

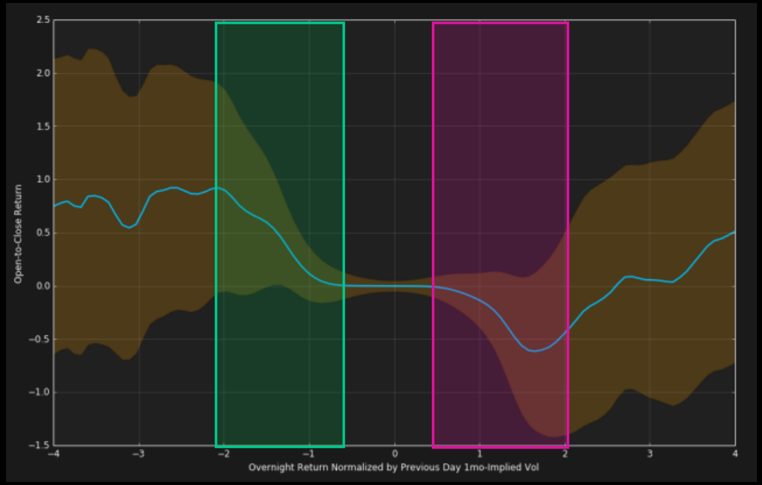
YET CORRELATED TO SENTIMENT

Twitter Sentiment Rank vs Vol-Normalized OV Return

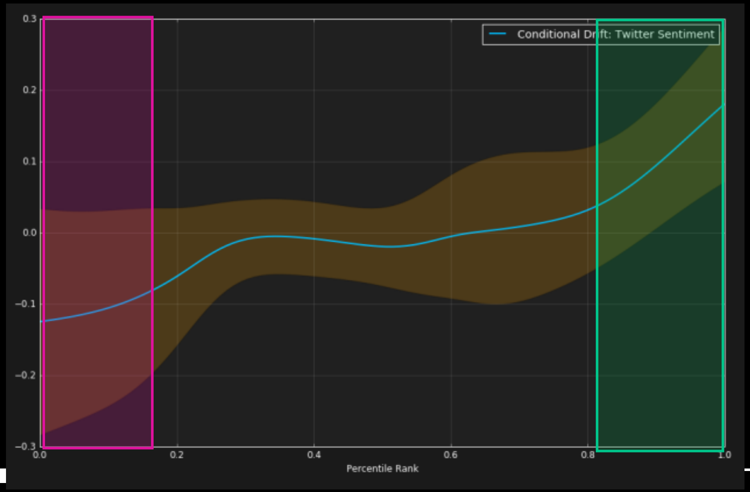


BUT WAIT A MINUTE…

Sentiment is Momentum while OV return is Mean Reversion signal!



Normalized OV returns
In S&P 500

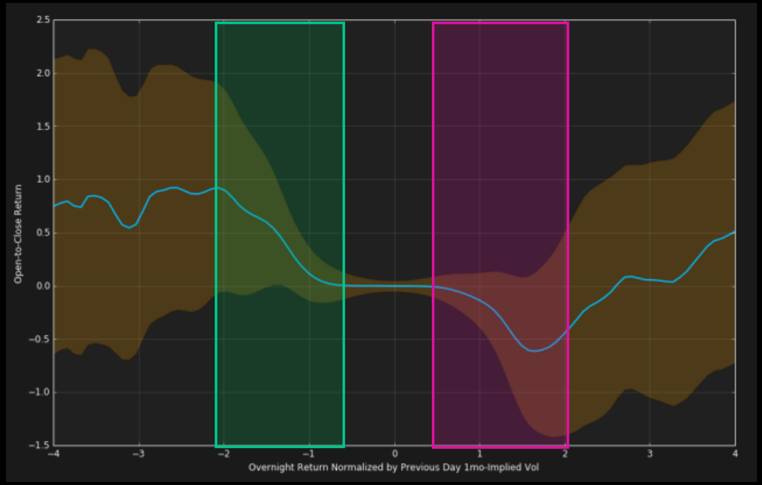


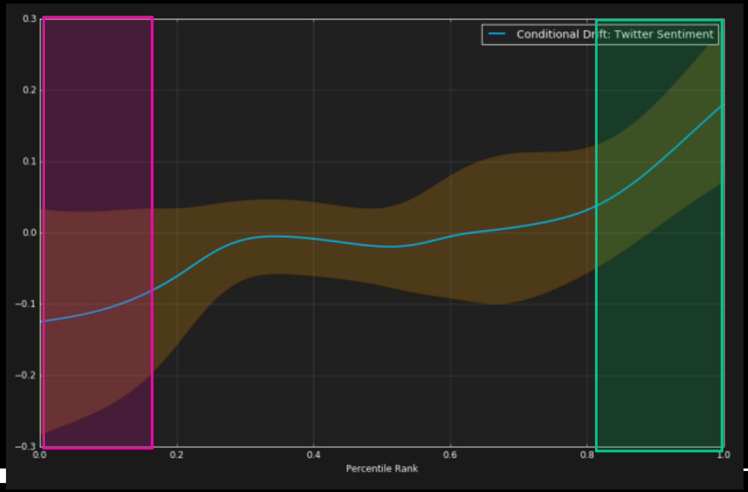
Sentiment in Russel 2000

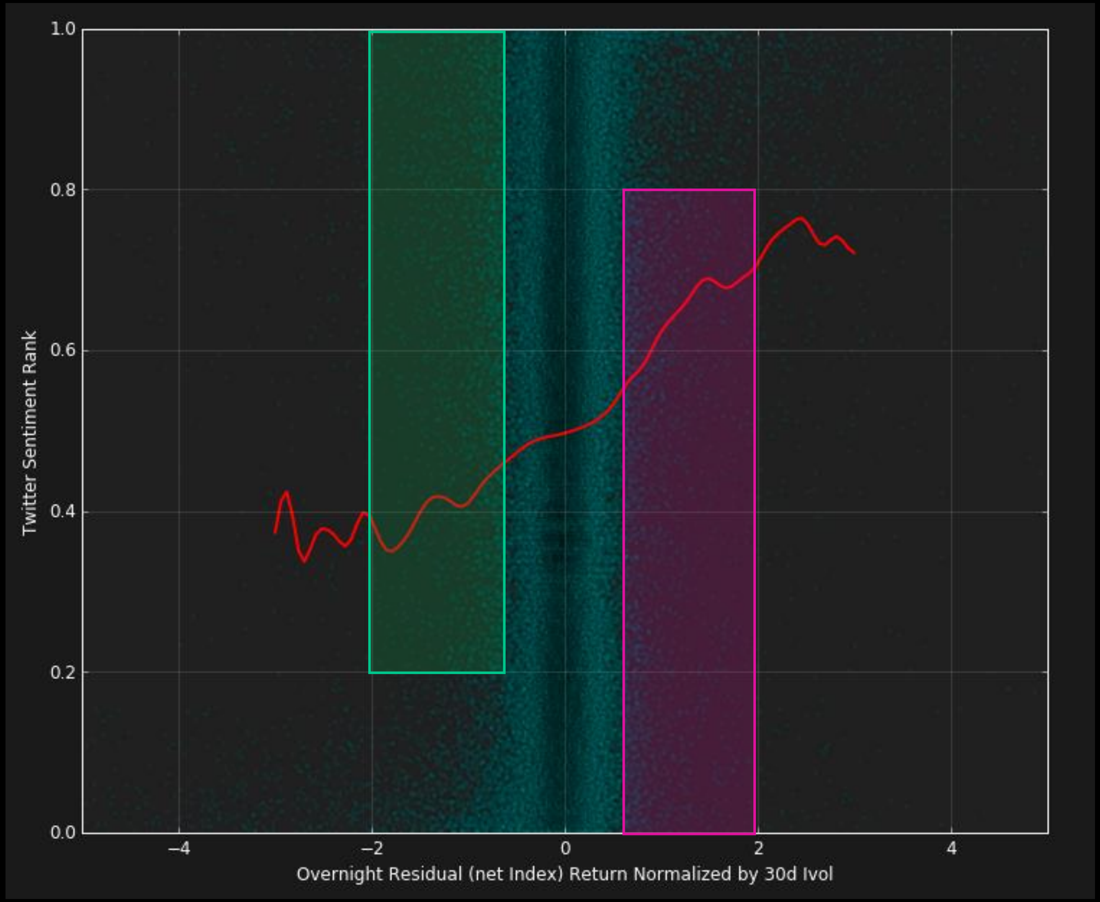
Bloomberg

HOW ABOUT WE TRY….

S&P 500 2sig Overnight movers with neutral-to-opposite sentiment







Up to 2sig OV moves, neutral-to-opposite sentiment

Bloomberg

SENTIMENT SEEMS TO ADD VALUE

S&P 500 2sig Overnight movers with neutral-to-opposite sentiment

- Even if the original overnight vs open-close strategy is a losing proposition if one were to take transaction costs into account, adding sentiment should make it less so

NEWS TOPIC CODES

One year (2016) of story-level data with topic codes

Use only stories that have a non-empty headline. This excludes tweets.

Look at one stock at a time. We looked at all current member of the S&P 500 for which data is available.

For each calendar day, compute the average Sentiment Score weighted by Sentiment Confidence, possibly including and excluding stories based on their topic codes.

Regress the daily excess return of the stock over the market on the same-calendar-day average sentiment score to obtain a measure of “sentiment impact”.

THE TEN "CONTROVERSIAL" TOPICS

ESGCONTROV: ESG Controversy Screening

LAW: Law

ESGRES: ESG Research

LITIGATE: Litigation

LAWPRAC: Legal Practice Areas

LAWSUITS: Lawsuits

IP: Intellectual Property

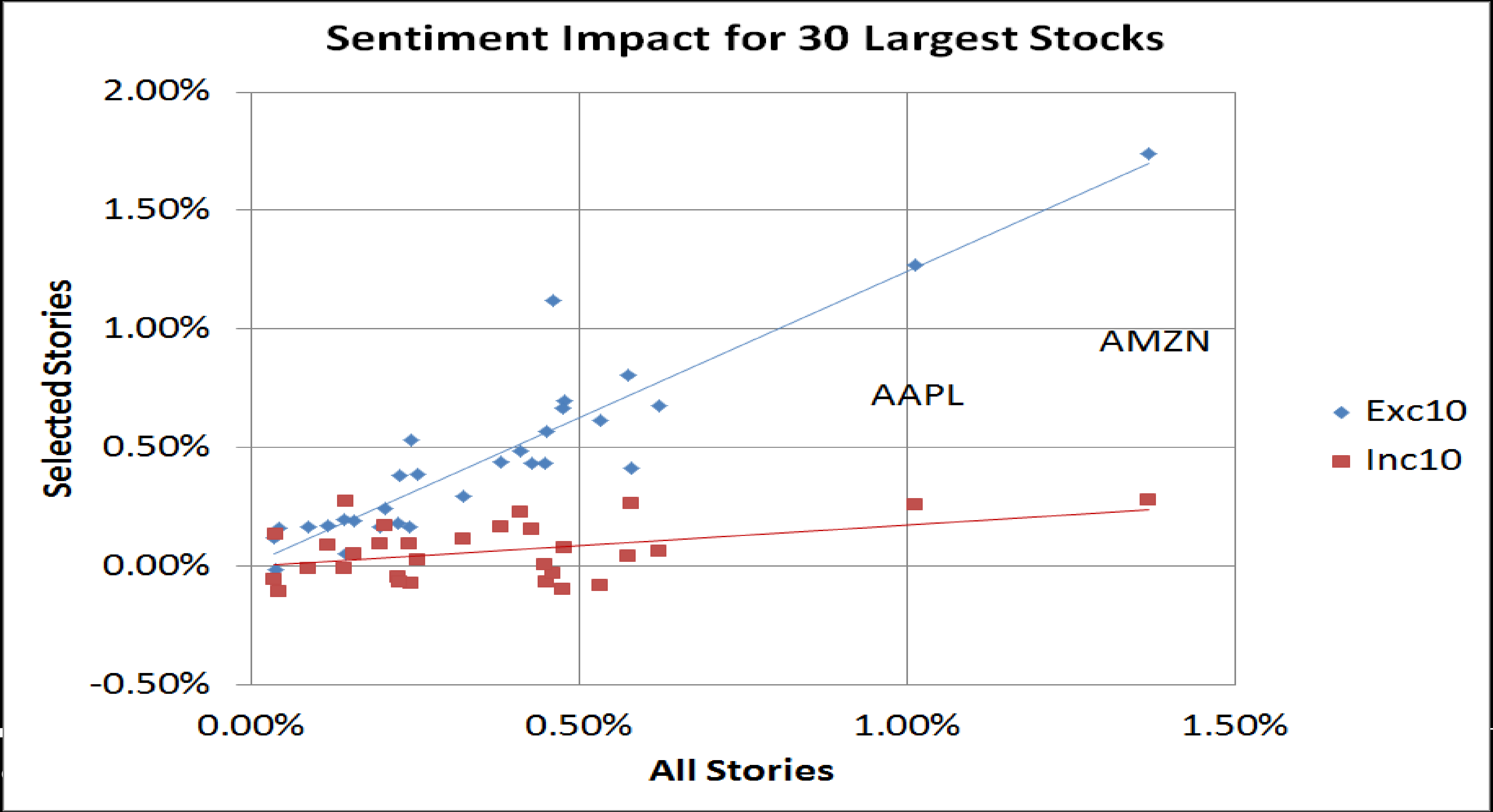
PATENT: Patents

CLASS: Class Action Lawsuits

CALVPOSS: Possible ESG Ledes

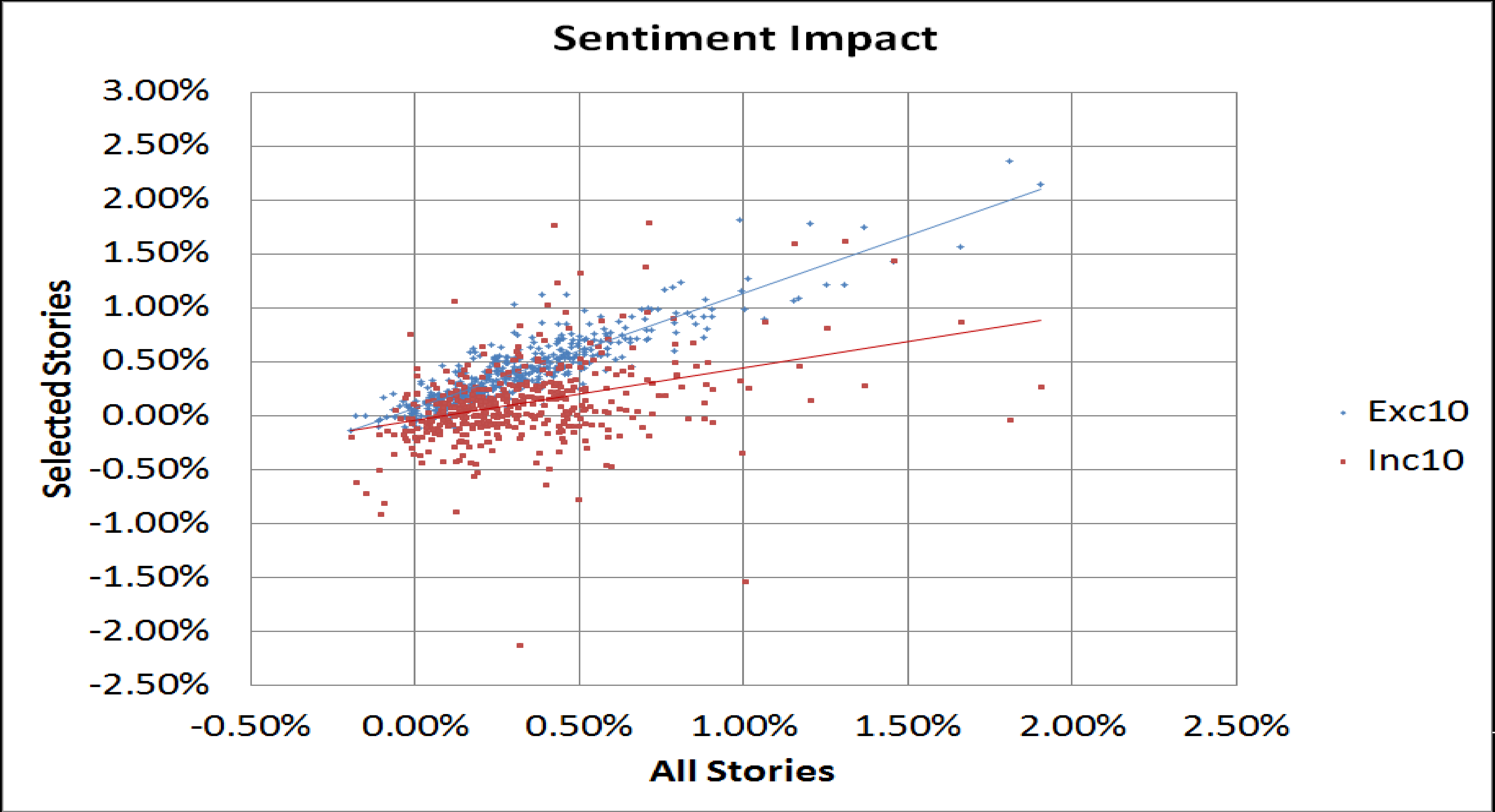
These ten codes were identified after staring at results from the 30 biggest (as of today) stocks, so the results for other stocks are kind of (OK not really) "out-of-sample".

30 BIGGEST STOCKS

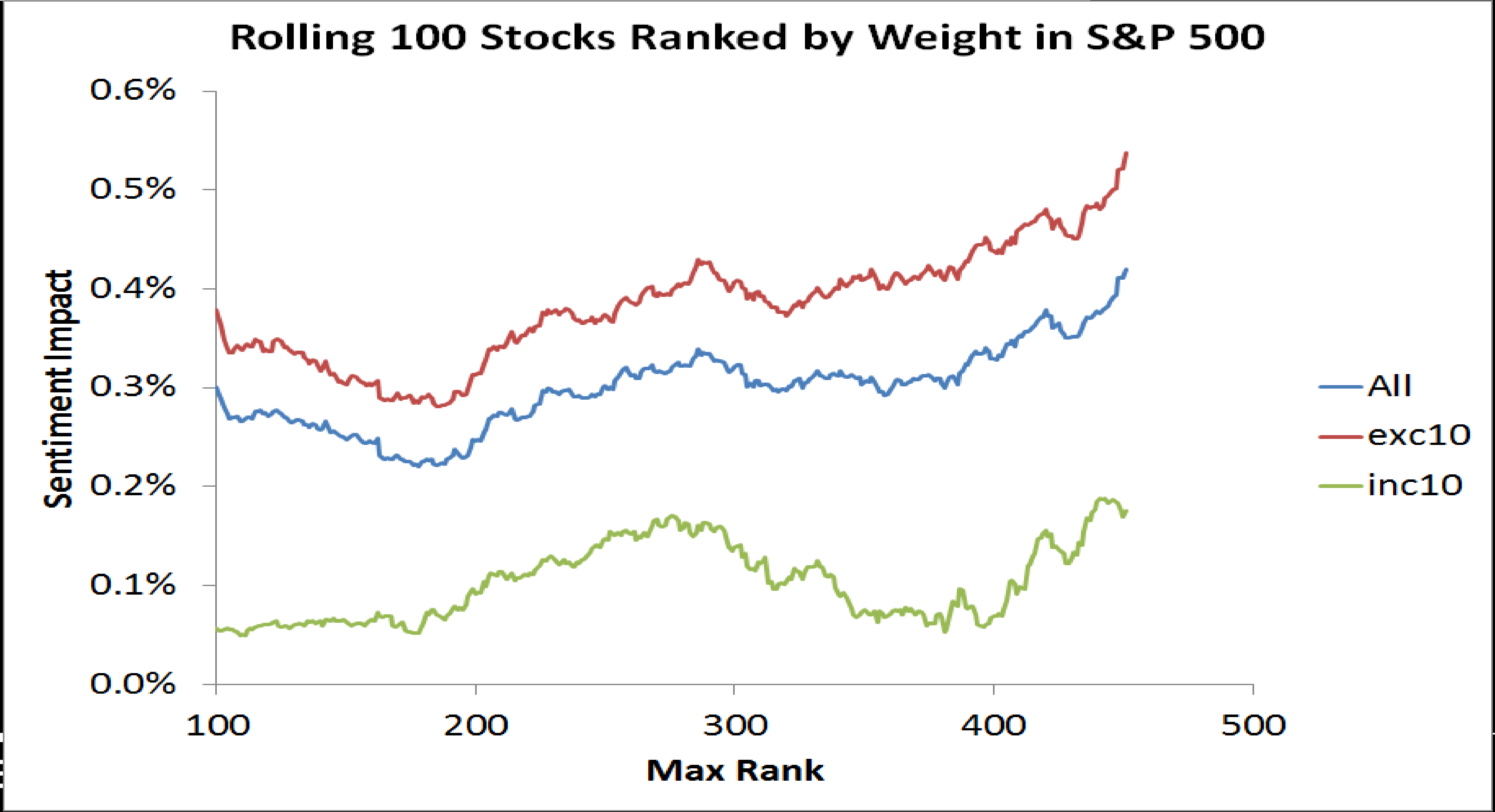


Bl

S&P 500 STOCKS

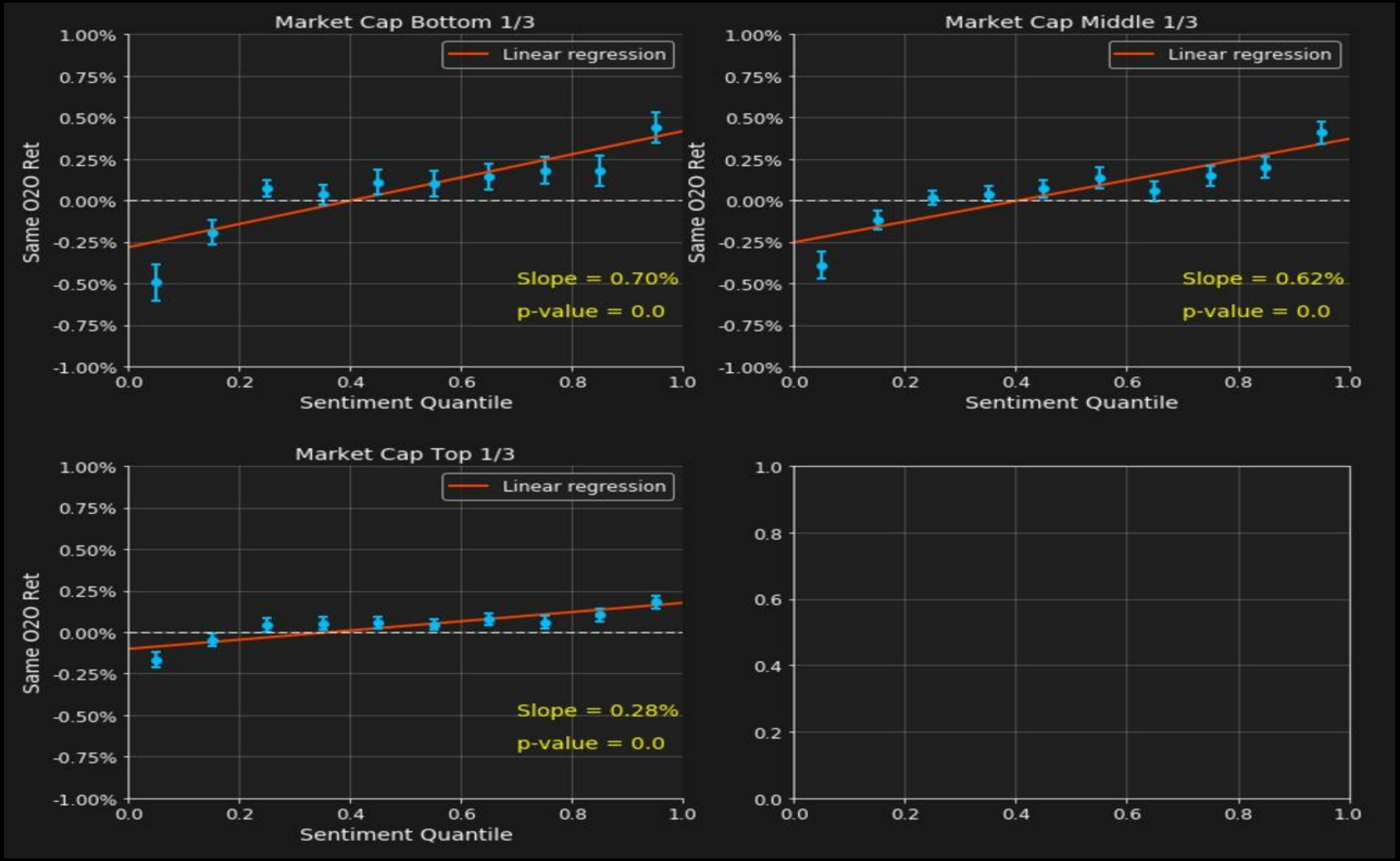


PATTERN IS CONSISTENT ACROSS LARGER AND SMALLER STOCKS IN S&P 500



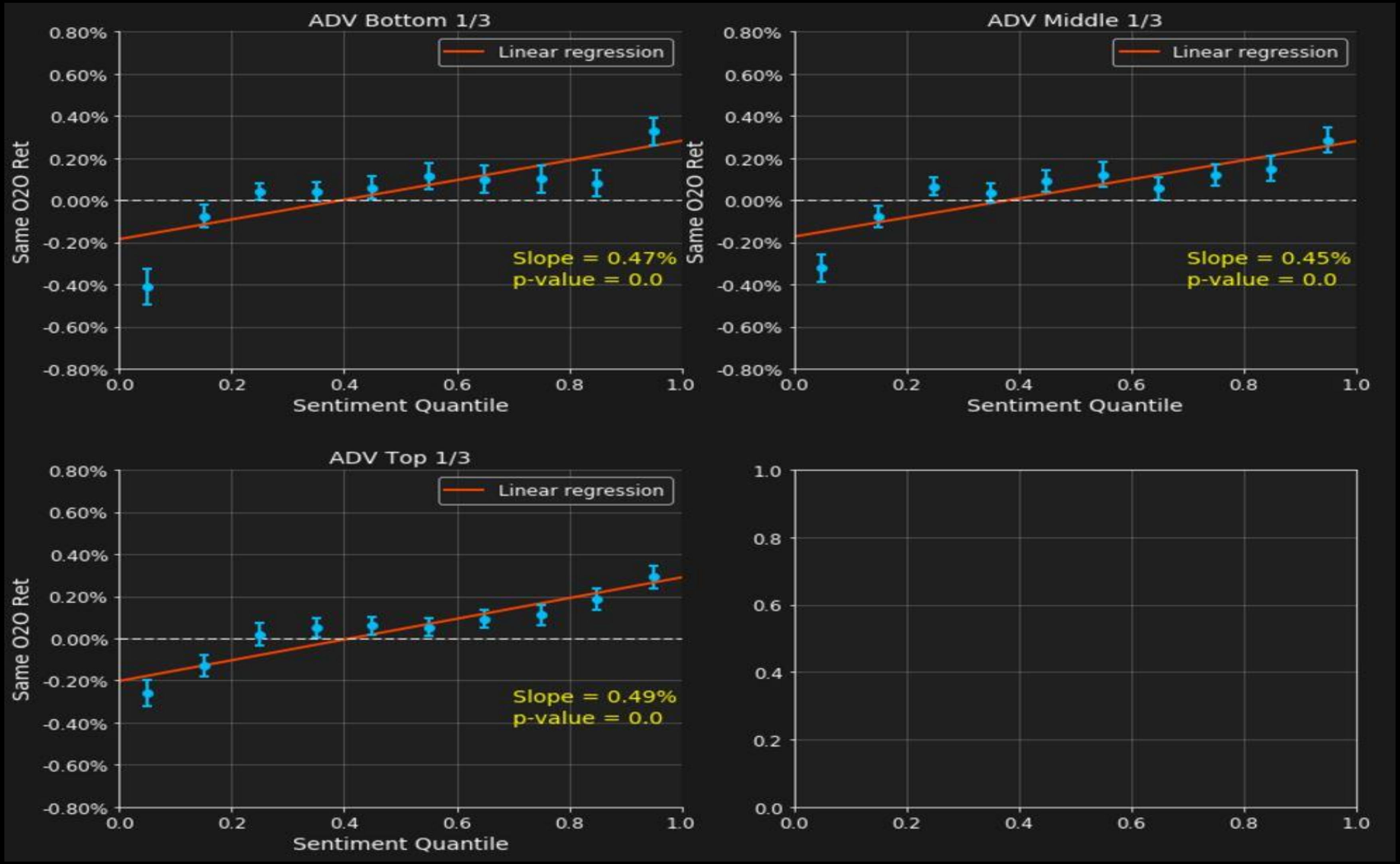
MARKET CAP (NEWS)

- Small cap stocks are more sensitive to news



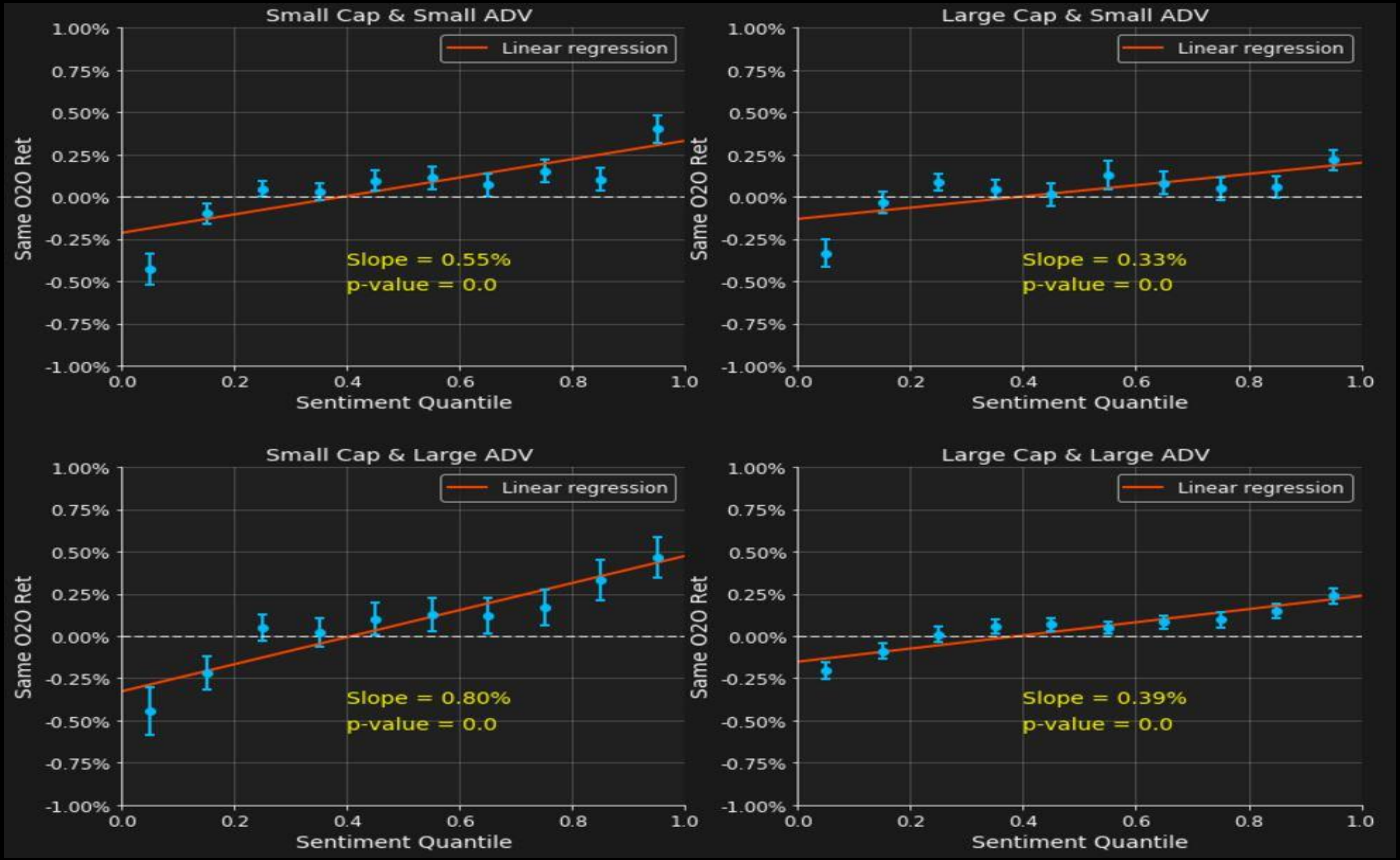
ADV (NEWS)

- 30-day average daily volume is estimated at the beginning of the testing period.
- Average daily trading volume doesn’t impact the sensitivity



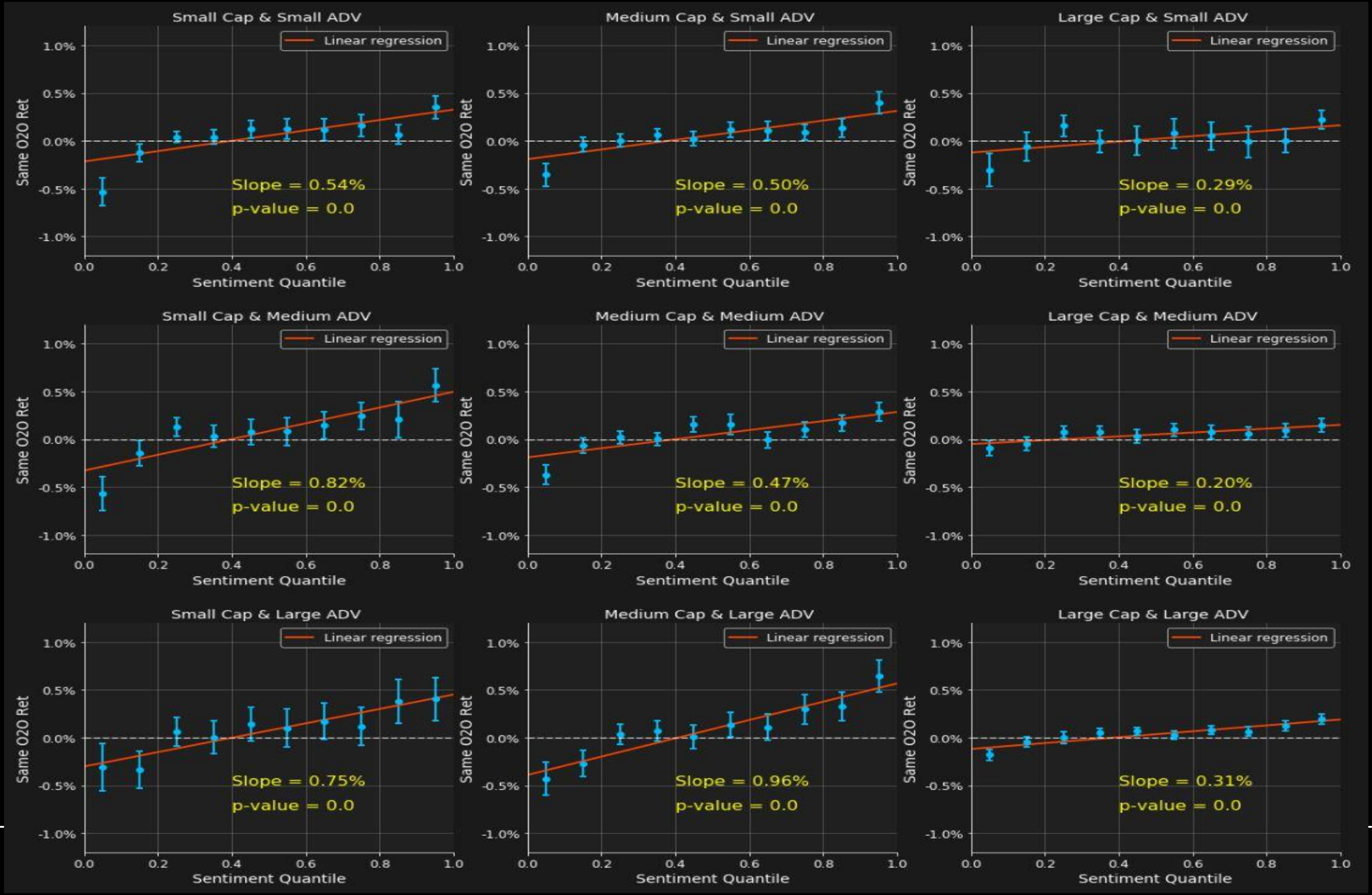
MARKET CAP & ADV (NEWS)

- Stocks are bucketed by median of their market cap and ADV, resp.
- Small cap stocks with larger trading volume are most sensitive to news



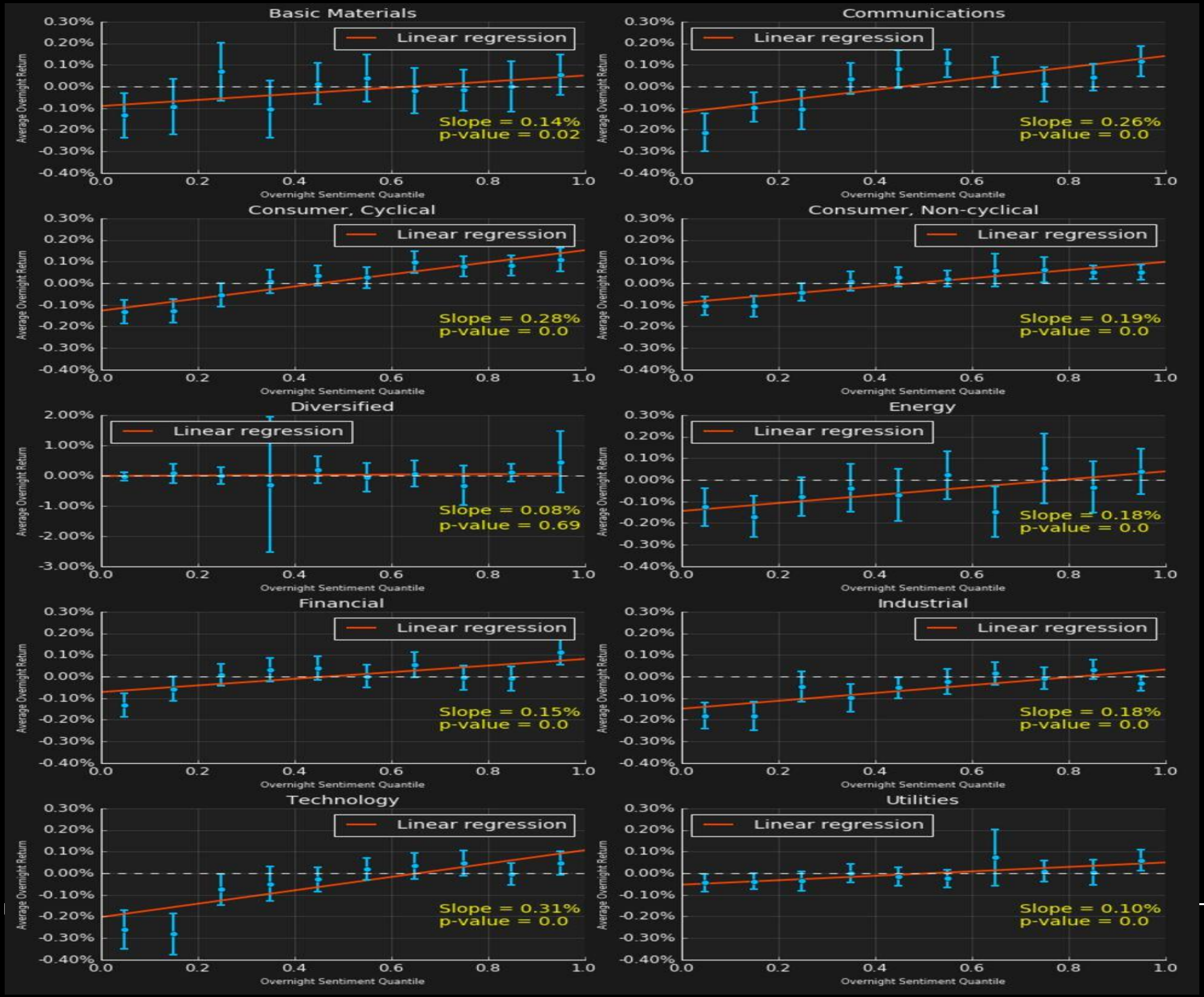
MARKET CAP & ADV (NEWS)

- Stocks are bucketed by quantiles [1/3, 2/3] of their market cap and ADV, resp.
- Small cap stocks with larger trading volume are most sensitive to news



Bloomberg

SECTOR (NEWS SENTIMENT)



- Communications, Consumer Cyclical, Technology have the strongest dispersion relationships.

DISCLAIMER

Sentiment without action is the ruin of the soul. — Edward Abbey

THANK YOU!