The Information Content of Pre-Open Social Media

Wachi Bandara

Pluribus Labs Research

wachi.bandara@pluribuslabs.com

April 4, 2016

Abstract: Social media discussion of financial instruments and markets contains significant information about the broad market as well as individual stock returns. While the importance of social media in event driven trades is well understood, it is less clear whether large message volumes, when aggregated, can have useful information about the future. We demonstrate the value of such information by using pre-open social media data to predict the day's price behavior. There is a positive relationship between pre-open social media sentiment and returns over the trading day and a negative relationship between uncertainty and the day's returns. Pre-open uncertainty expressed in social media also leads to greater trading volatility as well as higher trading volumes.

“Prediction is very difficult,” said Nils Bohr, “especially if it’s about the future.” Bohr, who in that immortal quote, it is said, refers to the use of overfitting in out-of-sample modeling, might well have been discussing finance. Indeed, prediction is the Holy Grail for financial market participants. There is a rich body of literature that investigates this issue in the return space (Rozeff [1984], Campbell and Shiller [1988], and Lettau and Ludvigson [2001] among many others). These studies have ranged from investigations that use different predictability horizons (monthly vs. quarterly) to those that limit prediction to direction (Christoffersen and Diebold [2006]) and to those that limit investigation to particular geographic areas (Jordan et al. [2014]). The broad range of methods and variables used in these studies that can make comparisons challenging. Predictability in the volatility space has been well-documented in the stochastic volatility literature (for instance, Heston [1993]); and the ARCH (Engle [1982]) and GARCH (Bollerslev [1986]) family of models. We now understand that volatility can be decomposed into predictable and unpredictable components. However, many models allow the use of exogenous variables to improve forecasting performance.

In this paper we investigate the information content of pre-open data derived from social media. We show that there is significant information in a pre-open social media sentiment indicator at both the index as well single-name stock level. This information is both economically and statistically significant and has explanatory power over returns, volatility and trading volume. To the best of our knowledge, this is the first paper to rigorously document the relationship between index and stock level pre-open sentiment and the corresponding day’s market behavior.

Traders and portfolio managers typically start their day well before market open, whereupon they consume news and other information pertinent to the trades they wish to execute that day. Often, traditional data sources – which are predominantly price-based – are limited to what is observable during market hours and are therefore several hours old at the start of a new trading day. Thus, general mood about the market or a particular firm could be an invaluable piece of information. Historically, accessing this information in a timely, systematized manner has been a prohibitively difficult task. However, the recent proliferation of discussions related to individual stocks and markets on social media sites has changed the information landscape, especially in the pre-open period.

Microblogging sites such as Twitter, Stocktwits and message boards have given market participants a public forum in which to express their thoughts and moods. The implication is that for the first time in history, we are able to assess the public mood of markets at massive scale and relatively low cost. The flip side of the scale benefit is the difficulty in filtering out noise from meaningful discussion. Social media messages are an amalgam of legitimate expressions of market-related opinions,[1] messages that contain no forward-looking information about the market[2] and spam, which have no informational content whatsoever.[3] Market participants have generally understood that there is immense value in having access to social media information as a means to gathering news more quickly. News proliferation on Twitter and Stocktwits is fast and reaches a large audience very quickly. However, there is still skepticism that large scale

[1] E.g. "$AAPL The steady march to a price target of 110 continues. I see 112-113 prior to earnings release in April."
[2] E.g. "$AAPL I bought some at 93. So what do you think?"
[3] E.g. "Some new post about $TSLA are on *xxxx.com* ! Read insights from investors."

social media data has significant value. Thus, there has been some reluctance by market participants to adopt social media as a viable signal.

In spite of this skepticism, there is a growing body of literature that documents its efficacy. One of the early efforts in this literature is Bollen et al. [2011] who find that the accuracy of return prediction is significantly improved by the inclusion of specific mood states gleaned from Twitter. Their sample period is about nine months and they show that return predictability is improved only with the inclusion of the "Calm" mood state, but no other. Subsequent papers have validated some of those findings. For instance, Sul et al. [2014] capitalize more directly on research aimed at conceptualizing emotions. Using the concepts of valence and arousal from the social psychology discipline they find that the cumulative emotional valence of tweets about a specific stock is related to the stock's subsequent returns. They note further that the "Calm" emotion used by Bollen et al. [2011] is related to neutral arousal and neutral valence. Chen et al. [2014], utilizing user-generated content from seekingalpha.com, find that views expressed in both the articles and comments on that platform predict stock returns and earnings surprises. Interestingly, they show that this return predictability holds even after controlling for the information content of newspaper articles and financial analyst opinions. More recently, Liew and Budavari [2016] show that sentiment derived from social media has explanatory power over idiosyncratic contemporaneous returns. Taken together, these findings show that the relationship between social media and stock returns is fertile ground for research.

In this paper we add to this literature by showing that there is significant information in pre-open social media content about the day's trading activity. There is a statistically significant relationship between returns, trading volumes and volatility at both the index as well as the individual stock level. There is unique information in the average social media sentiment, the number of observations and the variance in the sentiment.

Methodology

We perform our sentiment analyses using sentiment data generated by Pluribus Labs. These data inputs are real-time indicators of social media sentiment derived from Twitter and Stocktwits- two microblogging sites where market participants discuss expectations about markets and instruments. Pluribus Labs curates the Twitter and Stocktwits feeds extensively using proprietary spam filtering, NLP and sentiment scoring techniques.[4] In this paper we study the S&P500 index as well as the 100 stocks with the highest message volume.[5] Exhibit 1 shows the distribution of volumes on a log scale.

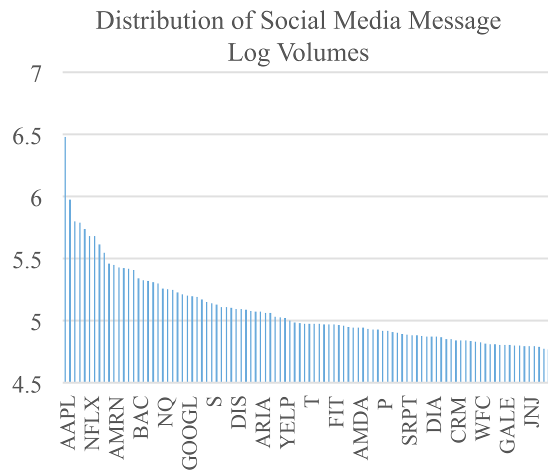
The chart on the left hand side shows the distribution of volumes by ticker, and the chart on the right hand side shows log message volumes by year. Clearly there is near exponential growth in message volumes year over year.

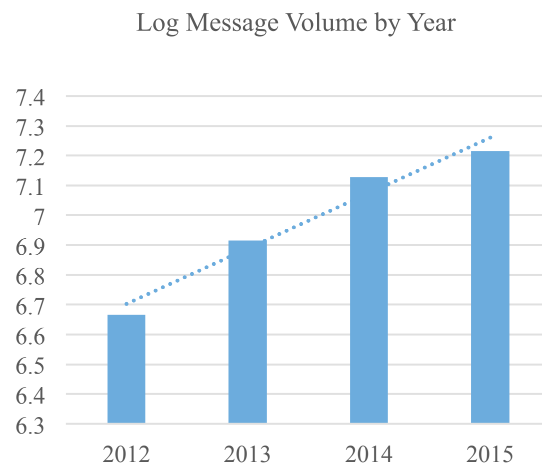
---

[4] For more information about these indices and other social media data in this paper, please contact info@pluribuslabs.com

[5] A full list of the tickers are: AA, AAL, AAPL, AMBA, AMGN, AMRN, AMZN, ARIA, BA, BABA, BAC, BBRY, BIDU, BIIB, BLUE, C, CELG, CHK, CMG, CRM, CSCO, CSIQ, CVX, CYBR, CYTX, DDD, DIA, DIS, EBAY, EXXI, F, FB, FCX, FEYE, FIT, FSLR, FXCM, GBSN, GDX, GE, GENE, GEVO, GILD, GLUU, GM, GMCR, GOOG, GPRO, GS, HIMX, HLF, IBB, IBIO, IBM, INO, INTC, ISR, JCP, JNJ, JPM, JUNO, KITE, KNDI, LL, LNKD, MCD, MNGA, MNKD, MSFT, MU, MXL, NFLX, NKE, NQ, PBMD, PCLN, PFE, PLUG, RAD, S, SBUX, SCTY, SHAK, SUNE, SWKS, SYN, T, TSLA, TWTR, UA, VRX, VZ, WMT, WTW, WYNN, XOM, YELP, YHOO, ZIOP, ZNGA

Exhibit 1: Summary Statistics of Message Volumes by Ticker and by Year





The pre-open period is defined to be between 6:00am and 9:00am EST.[6] Average sentiment, total message density and other metrics are computed at 9:00 am. Once the pre-open inputs are observed on a particular day, we measure returns, volatility and volume from market open to close over the same day.

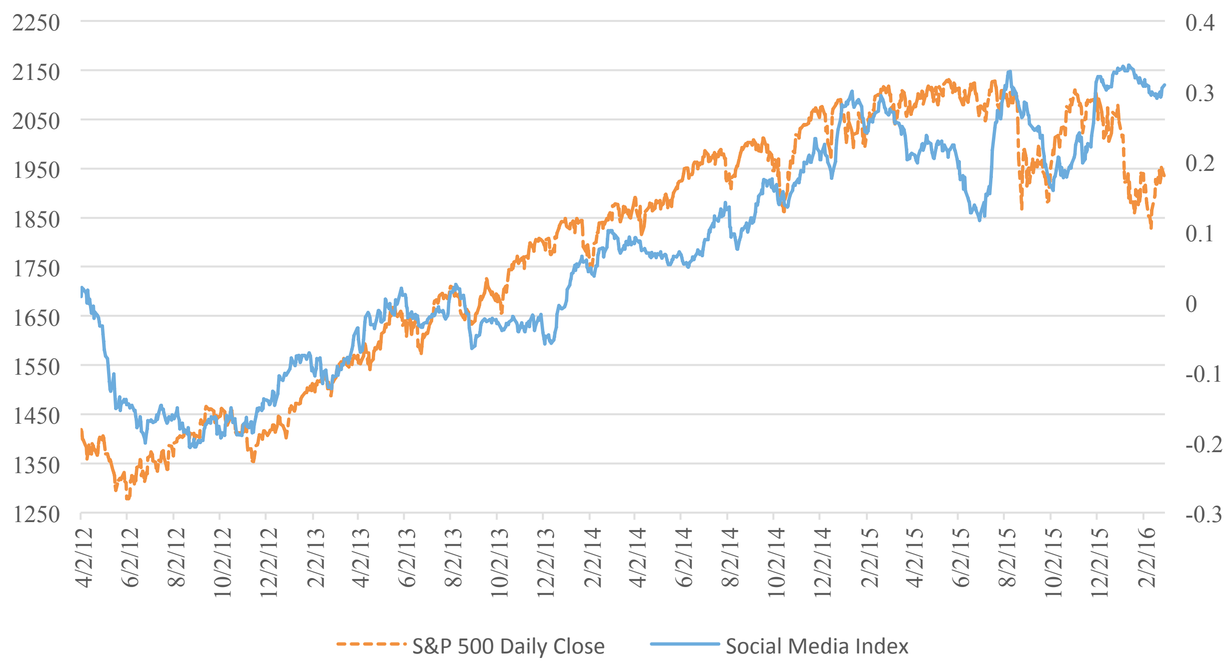
The sample period is January 1, 2012 to February 29, 2016. We choose this sample period because this is the longest period for which we have sufficient message volume to obtain robust estimates. We obtain stock level price data from Xignite and use tickdata.com for our S&P500 index data.

[6] Although market open is at 9.30am, we limit our pre-open data aggregation period to exclude the final 30 minutes before market open in order to simulate the time lag involved in dissemination and processing.

Pre-open Social Media Data and the S&P500 Index

We begin by investigating the ability of our social media sentiment indicator to predict the same day's return for the S&P500 index. Exhibit 2 shows the cumulative sentiment index and market close for the S&P 500 index. The index is created by rescaling the raw sentiment values during the pre-open period. It is clear that there is a strong contemporaneous relationship between the index and the market.

Exhibit 2: Pre-Market Sentiment Index and S&P 500 Daily Cumulative Returns



We use a regression framework to measure the relationship between pre-open social media information and daily returns; results are shown in Exhibit 3. We use close-to-close VIX change as a conditioning variable when measuring the efficacy of the social media variables. As expected, the coefficient on lagged VIX is negative- an increase in VIX generally leads to lower returns (Model 3). Model 1 shows that a one-standard-deviation increase in pre-open sentiment leads to an approximately 8 basis point increase in daily returns.[7] Model 2 shows that a one-standard-deviation increase in uncertainty measured by pre-open open raw sentiment volatility predicts a 12 basis point decrease in returns for the day. These effects are statistically and economically significant and are robust to the inclusion of lagged change in VIX in Model 4. The increasing adjusted-R^2 values when going from Models 1, 2 and 3 to 4 shows that each variable has unique explanatory power over returns. In each model we also control for annual and day-of-week affects.

Exhibit 3: S&P 500 Returns and Social Media Sentiment

	Dependent variable:			
	Return %			
	(1)	(2)	(3)	(4)
sentiment_mean_change	0.397***			0.332***
	(4.207)			(3.449)
sentiment_variance_change		-0.004***		-0.003***
		(0.001)		(0.001)
VIX_change_lagged			-0.051***	-0.049***
			(0.019)	(0.019)
Constant	-0.226	-0.197	-0.200	-0.220
	(0.137)	(0.137)	(0.138)	(0.136)
Year/Day Controls?	Yes	Yes	Yes	Yes
Observations	982	982	981	981
R^2	0.027	0.026	0.017	0.044
Adjusted R^2	0.018	0.017	0.008	0.033
F Statistic	2.976***	2.912***	1.862*	4.042***
Note:				*p**p***p<0.01

Next, we investigate whether pre-open social media metrics can predict subsequent realized volatility. Towards this end we use an exponential GARCH framework (Nelson [1991]). The results are provided in Exhibit 4 and show that an increase in sentiment volatility from the

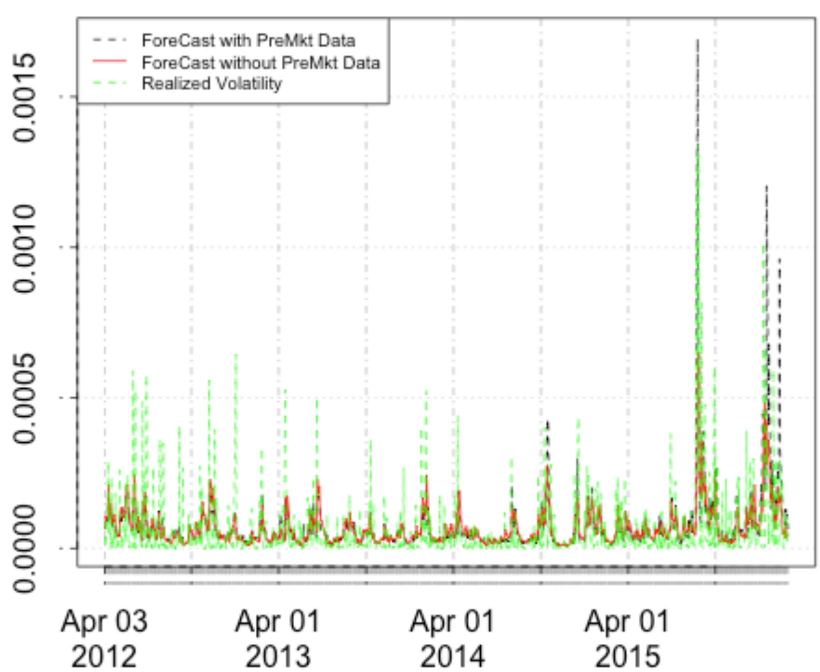
[7] The standard deviation of the normalized sentiment values we use in the regressions is 0.21. Mean unscaled pre-open variance is approximately 27.

previous day to today predicts an increase in daily volatility, as does the number of social media messages. Conversely, an increase in market sentiment predicts a decrease in volatility for the day, however, this relationship is not statistically significant. This shows that volatility has a strong relationship with pre-open uncertainty, but not with sentiment. Exhibit 5 shows a graphical representation of forecast and realized volatility. Note that the inclusion of external pre-open regressors in the variance equation improves the accuracy of forecasts, especially around the peaks.

Exhibit 4: Variance Equation Regressors in EGARCH Model

	Estimate	Std. Error	t value	Pr(> | t|)
change_sentiment_vol	0.095	0.024	3.926	0.0001
log_obs	0.031	0.003	10.200	0
sentimentIndex	-2.690	2.575	-1.045	0.296

Exhibit 5: Forecast and Realized Volatility

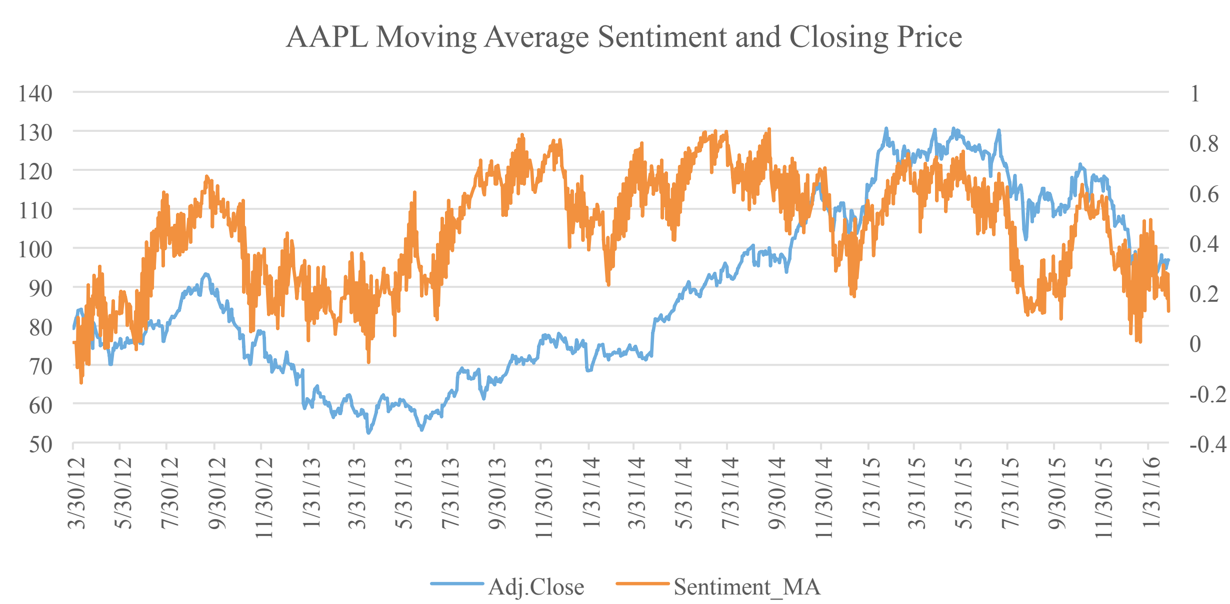


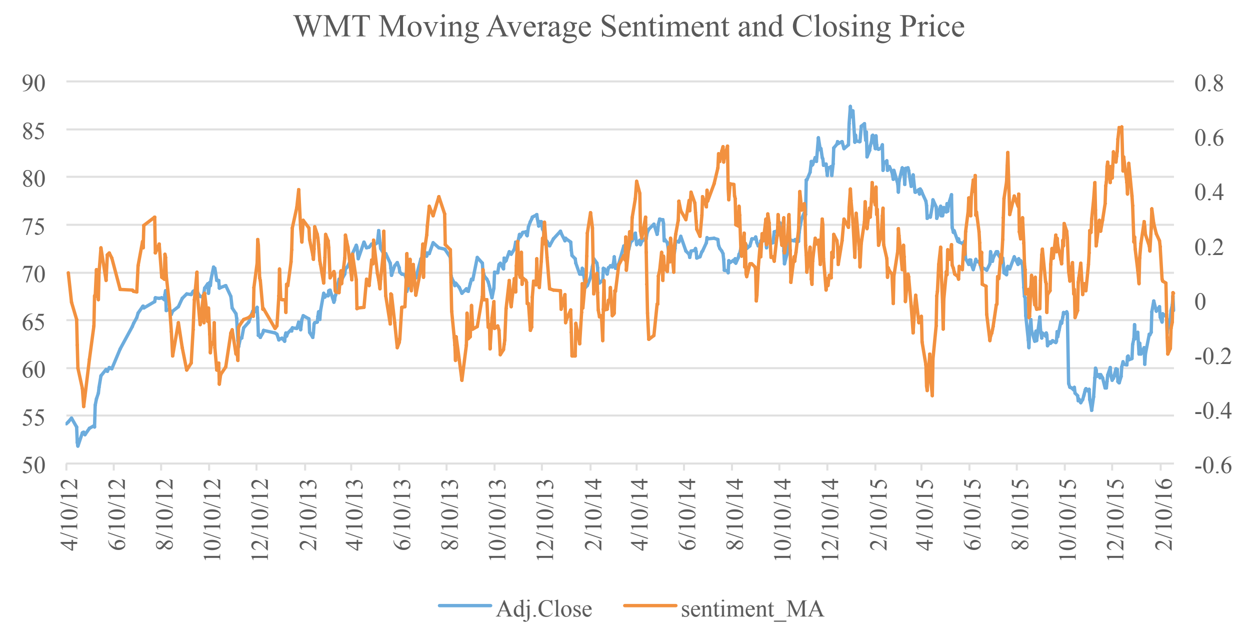
Pre-open Social Media Data and Individual Stock Returns

Next we investigate the relationship between pre-open sentiment and stock performance, focusing on returns, trading volume and volatility. Compared with the market index, social media messages about single name stocks are sparser. The median stock in the top 100 firms in terms of message count in our sample period with a $5 price filter has a social media signal on 539 days, or approximately every other day.

Exhibit 6 shows a graphical representation of the relationship between a weighted average sentiment index and closing prices for two stocks- Apple and Walmart. These are selected because they are very different firms- one a technology firm and the other a retailer- however, the relationship between sentiment and prices is clear.

Exhibit 6: Relationship Between Sentiment and Prices for AAPL and WMT





We conduct our study in a panel regression framework. For the sake of simplicity, we use the change in pre-open sentiment and variance from the previous day to today.[8] While pre-open sentiment is a proxy for general investor sentiment, sentiment variance is a proxy for uncertainty about a particular stock. Total message volume indicates how much conversation (in social media messages counts) there exists during one pre-open session and is – along with variance – a proxy for uncertainty.

[8] While it is possible to use metrics such as sentiment momentum or sentiment as an event detection tool, we avoid these metrics to demonstrate the efficacy of pre-open information in its simplest form. It is, of course, possible to create complex nonlinear data points using this data.

Exhibit 7: Return and Volumes vs Pre-open Sentiment

	Dependent variable:					
	Returns %		Daily Volume (Log)	Returns %		Daily Volume (Log)
	Panel A: All Firms			Panel B: Price Filtered Firms		
	(1)	(2)	(3)	(4)	(5)	(6)
sentiment_value_change	0.037**			0.039***		
	(0.018)			(0.015)		
sentiment_volatility_change			0.015***			0.013***
			(0.001)			(0.001)
total_message_volume	-0.002***	-0.002***	0.002***	-0.001**	-0.001**	0.002***
	(0.0003)	(0.0003)	(0.00005)	(0.0003)	(0.0003)	(0.00005)
volume_high_indicator* sentiment_value_change		0.092***			0.095***	
		(0.029)			(0.024)	
Day/Year Controls?	Yes	Yes	Yes	Yes	Yes	Yes
Stock Specific Controls?	Yes	Yes	Yes	Yes	Yes	Yes
Observations	33,275	33,275	87,069	29,827	29,827	70,370
R^2	0.010	0.010	0.666	0.004	0.004	0.608
Adjusted R^2	0.010	0.010	0.665	0.004	0.004	0.607
F Statistic	27.227***	27.720***	15,789.870***	9.257***	9.998***	9,084.436***
Note:						*p **p ***p<0.01

Exhibit 7 shows that there is significant predictive power in social media messages over the day's returns. In Model 1 we consider only change social media sentiment and total message volume. A unit change in sentiment leads to a .037% increase in returns for the stock. A one-standard-deviation increase in message volume (approximately 40 messages) leads to an 8 basis point decrease in return. As discussed earlier, single-name message volume can be sparse, thus generating a clean signal can sometimes be difficult. To account for this, we consider days where message volume is greater than the moving average number of messages over the past 60 days. The results are in Model 2. We see that the coefficient on sentiment change more than doubles, showing that with a greater amount of messages – i.e. with more information – we can generate a signal with more efficacy. In order to ensure that this result is not driven by small stocks which do not garner much investor attention, we filter out our list of firms to exclude firms whose stock value is below $5 at the start of the sample. This leaves us with a sample of 84 firms.[9] As we can see, the result holds even for price filtered stocks.[10]

In Models 3 and 6 we show that social media has significant information over the day's trading volume even after controlling for its autoregressive properties. A 10% increase in message volume leads to a 0.02% increase in trading volume. A unit increase in sentiment standard deviation leads to a 1.5% increase in trading volume. These results are all economically and statistically significant and robust to price filters.

Next, in Exhibit 8, we look at whether market level sentiment has an impact on individual stock prices. Once again, we use a panel regression framework and include market level sentiment and uncertainty indicators interacted with stock level indicators. The results provide some interesting insights into the nature of market perception and individual stock performance. After including market level variables, the coefficients on total stock-level message volume and stock-level sentiment change remain similar in sign and magnitude to the results shows in Exhibit 7.

---

[9] A full list of the price filtered firms are: AA, AAL, AAPL, AMBA, AMGN, AMZN, ARIA, BA, BABA, BAC, BBRY, BIDU, BIIB, BLUE, C, CELG, CMG, CRM, CSCO, CSIQ, CVX, CYBR, DDD, DIA, DIS, EBAY, F, FB, FCX, FEYE, FIT, FSLR, FXCM, GDX, GE, GILD, GM, GMCR, GOOG, GOOGL, GPRO, GS, HIMX, HLF, IBM, INO, INTC, JCP, JNJ, JPM, JUNO, KITE, KNDI, LL, LNKD, MCD, MSFT, MU, MXL, NFLX, NKE, PCLN, PFE, RAD, S, SBUX, SCTY, SHAK, SWKS, T, TSLA, TWTR, UA, VRX, VZ, WMT, WTW, WYNN, XOM, YELP, YHOO, ZIOP

[10] In unreported results we confirm that the result holds qualitatively and is statistically significant even when using using higher price filters.

However, we can see that market level sentiment also has an impact on individual stocks in a very intuitive way.

Exhibit 8: Stock Returns vs Pre-open Stock and Market Sentiment

	Dependent variable:
	Returns %
total_message_volume	-0.002***
	(0.0003)
index_sentiment_variance_change	0.004***
	(0.0005)
sentiment_value_change	0.045**
	(0.018)
index_sentiment_variance_change * sentimentIndex_change	-0.006***
	(0.002)
index_sentiment_variance_change * sentiment_value_change	0.002***
	(0.001)
Day/Year Controls?	Yes
Stock Specific Controls?	Yes
Observations	33,268
R^2	0.014
Adjusted R^2	0.014
F Statistic	31.138***
Note:	*p **p ***p<0.01

For instance, the interaction between index-level sentiment change and index-level sentiment volatility change has a negative sign. The implication here is that an increase in overall market sentiment with a corresponding decrease in disagreement about the market leads to an increase in stock returns.[11] Similarly, a decrease in market sentiment with an increase in disagreement leads to lower stock returns. The final interaction term in Exhibit 8 shows that the relationship between stock-level sentiment and market-level uncertainty has a positive sign. This indicates that if there is an increase in disagreement about the overall market but stock-level sentiment increases, this leads to an increase in stock returns. This is significant because it shows that stock-level sentiment can dominate market-level uncertainty. Similarly, even if market level uncertainty declines, a reduction in stock-level sentiment leads to a decrease in returns.

Next, we investigate whether pre-open social media data contains information about the day's volatility. In order to account for the clustering properties of volatility, we use an EGARCH model. EGARCH models are able to successfully model asymmetric responses in the conditional variance equation relative to other GARCH-type models (Taylor [2007]). We use two proxies for pre-open uncertainty derived from social media – variance and the number of messages. An increase in variance implies that there is greater disagreement expressed in the conversation about a stock, and an increase in the number of observations implies that there is more conversation about a stock. We insert each of these variables into the variance equation in the EGARCH model. We price-filter stocks at the $10 level and restrict to stocks for which we have a significant number of observations for the sample period. After the appropriate filters have been applied, we are left with 62 individual tickers. We consider both gross as well as residual returns with respect to the market and the Fama-French factors. Although the results are qualitatively similar, in Exhibit 9 we present results for residual returns with respect to the

[11] This effect is similar to what we witnessed in 2013, where the overall stock market rose approximately 30% with very little systematic uncertainty. Our result however, is not driven by 2013 as we control for year.

market. We show the coefficients for a change in the number of observations and change in pre-open sentiment volatility in the variance equation.

We can see that the change in both sentiment volatility and the number of observations have strong predictive power over the day's volatility. Taken together, the coefficient is positive in 85% of the cases, showing that an increase in message volume or disagreement leads to an increase in the day's volatility. Of the two variables, a change in the number of observations is a remarkably strong predictor of the day's volatility, with nearly all coefficients positive and statistically significant. A change in pre-open sentiment volatility also predicts an increase in daily volatility, although this effect is somewhat weaker. Additionally, there are no instances of the coefficient being negative and statistically significant; showing that even on a case-by-case basis, pre-open volatility indicators maintain their economic significance.

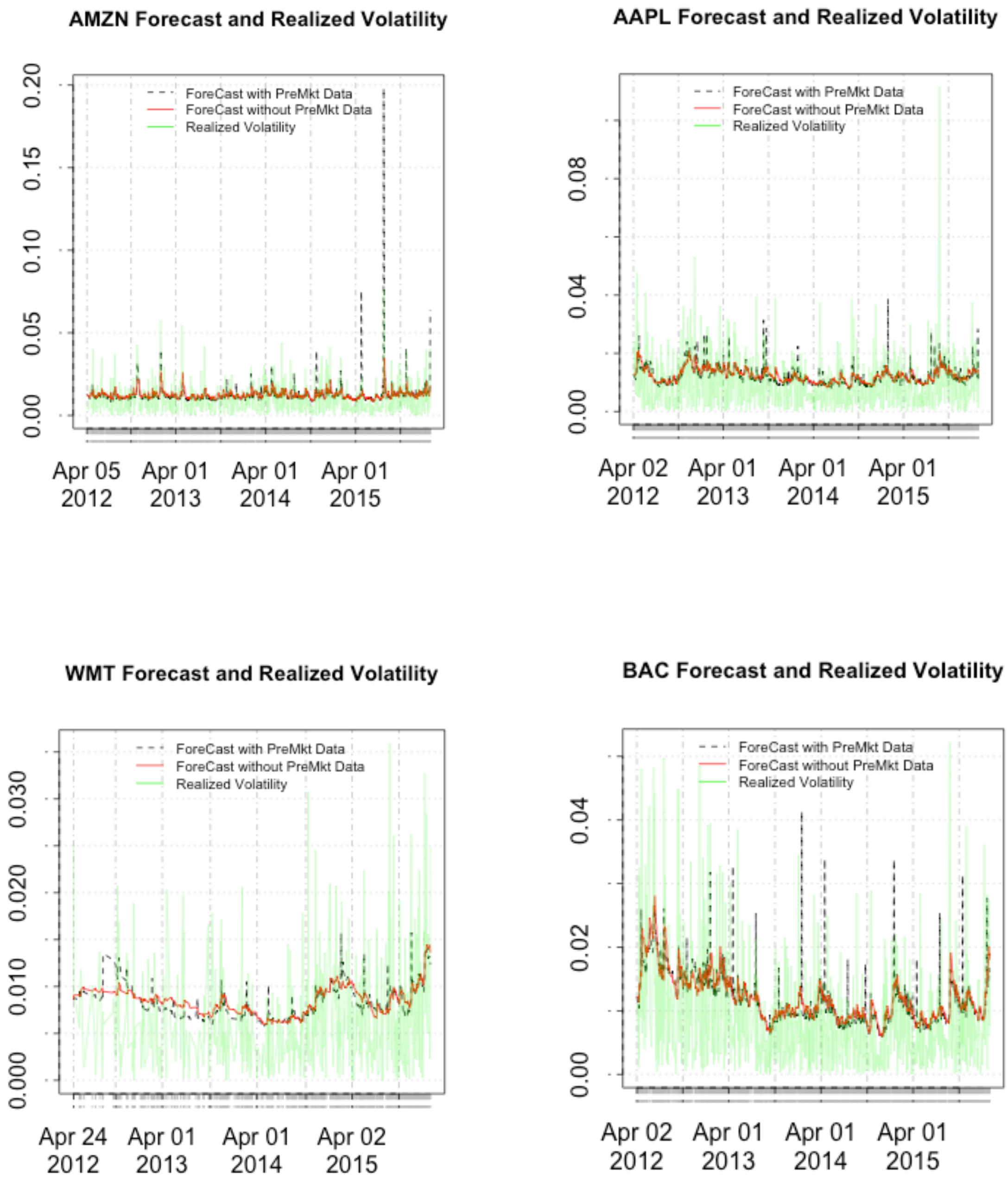
In Exhibit 10 we show a graphical representation of forecast volatility within an EGARCH framework and realized volatility.[12] AAPL, AMZN, WMT and BAC are used as illustrative examples. We can see that the volatility forecast that uses pre-open information is much better at predicting large spikes in volatility. This is intuitive since a high level of conversation around a stock typically means there is a great deal of uncertainty or interest around it. However, the improvement in accuracy does not apply to spikes alone. Pre-open sentiment has the effect of lowering predicted volatility during periods of low realized volatility and vice versa.

[12] We use the absolute value of returns as our volatility proxy. The results are qualitatively the same if we use a different metric like squared returns.

Exhibit 9: Pre-open Social Media Data in EGARCH Model

Ticker	Change in No. of Obs		Change in Volatility		Ticker	Change in No. of Obs		Change in Volatility	
	Estimate	t value	Estimate	t value		Estimate	t value	Estimate	t value
AA	0.009	3.097	0.029	1.414	INTC	0.014	4.454	0.031	0.968
AAPL	0.002	3.299	0.034	1.624	JCP	0.005	1.586	0.045	2.005
AMGN	0.029	3.225	-0.047	-1.394	JNJ	0.009	2.117	0.012	0.347
AMZN	0.007	4.150	0.010	0.522	JPM	0.003	2.525	0.037	1.622
ARIA	0.009	2.887	0.010	0.474	KNDI	0.014	2.151	-0.045	-0.604
BA	0.015	4.160	-0.003	-0.225	LL	0.002	0.325	0.105	4.130
BAC	0.007	3.295	0.019	0.767	LNKD	0.011	1.801	0.017	0.689
BIDU	0.010	2.369	0.017	0.695	MCD	0.006	2.254	0.036	1.142
BIIB	0.007	2.440	0.043	1.109	MSFT	0.004	0.968	0.015	0.552
C	0.002	1.094	0.022	1.283	MXL	0.070	2.351	-0.280	-1.605
CHK	0.025	4.484	-0.020	-0.720	NFLX	0.007	3.964	0.015	1.211
CMG	0.015	2.212	0.041	1.114	NKE	0.018	2.324	0.014	0.421
CRM	0.021	3.430	0.028	0.835	NQ	0.010	3.613	0.003	0.139
CSCO	0.014	3.999	-0.002	-0.067	PCLN	0.007	3.354	0.022	1.154
CSIQ	0.029	5.940	0.013	0.519	PFE	0.014	3.370	-0.025	-0.567
CVX	0.021	3.099	0.003	0.136	PLUG	0.004	2.598	0.046	1.375
DDD	0.016	3.482	0.017	0.817	RAD	0.010	2.650	0.092	1.941
DIA	0.029	1.153	-0.046	-1.318	SBUX	0.020	4.216	-0.010	-0.589
DIS	0.017	2.419	0.050	1.882	SWKS	0.030	3.905	0.027	1.003
EBAY	0.002	0.253	0.009	0.151	T	0.025	3.917	-0.001	-0.049
F	0.008	1.021	-0.004	-0.138	TSLA	0.004	4.206	0.055	3.660
FCX	0.017	2.066	0.037	1.577	UA	0.017	3.082	0.008	0.351
FSLR	0.013	3.342	0.059	2.423	VZ	0.004	1.154	0.070	2.668
GE	0.006	3.452	0.015	0.589	WMT	0.006	3.119	0.008	0.284
GILD	0.006	3.063	0.017	0.765	WTW	0.013	4.821	0.124	1.636
GLUU	0.008	1.698	0.026	0.958	WYNN	0.039	5.105	0.019	0.598
GM	0.013	4.238	0.018	0.638	XOM	0.012	3.400	0.022	0.828
GOOG	0.007	4.145	-0.010	-0.806	XOMA	0.014	3.058	-0.015	-0.456
GS	0.008	5.065	-0.012	-0.858	YELP	0.012	2.754	-0.007	-0.290
HLF	0.016	3.590	0.038	0.986	YHOO	0.007	3.622	0.034	1.227
IBM	0.013	2.170	-0.031	-1.970	ZNGA	0.024	5.468	-0.003	-0.611

Exhibit 10: Forecast Volatility vs |Returns| for Individual Stocks



Conclusion

This is the first paper to rigorously document the information content of pre-open social media information with respect to the day's stock price movement on the index and individual stocks. We show that pre-open information contains valuable and unique information about returns, trading volume and volatility.

We demonstrate in this paper that pre-open social media data could be a useful, unique data point for portfolio managers and traders to consider in informing their trades for the day. We show that there is positive predictive relationship between average sentiment and returns, and between sentiment disagreement and daily trading volume and volatility. The latter relationship is generally robust to well-known volatility clustering and asymmetry effects. There is also information about individual stocks in the general sentiment about the market. A stock with high sentiment during periods of high market uncertainty generally has positive returns.

We find that the explanatory power of social media – on returns in particular – doubles when there is a greater volume of messages present, indicating that we can extract a stronger signal from a greater number of messages. This is particularly important in an age when the use of social media is becoming more pervasive with the growing adoption of platforms such as Twitter and Stocktwits. The use and validity of methods such as those presented in this paper will grow more useful and important with time.

References

1. Bollen, Johan, Huina Mao, and Alberto Pepe. "Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena." *ICWSM* 11 (2011): 450-453.
2. Bollerslev, Tim. "Generalized autoregressive conditional heteroskedasticity."*Journal of econometrics* 31, no. 3 (1986): 307-327.
3. Campbell, John Y., and Robert J. Shiller. "The dividend-price ratio and expectations of future dividends and discount factors." *Review of financial studies* 1, no. 3 (1988): 195-228.
4. Chen, Hailiang, Prabuddha De, Yu Jeffrey Hu, and Byoung-Hyoun Hwang. "Wisdom of crowds: The value of stock opinions transmitted through social media." *Review of Financial Studies* 27, no. 5 (2014): 1367-1403.
5. Christoffersen, Peter F., and Francis X. Diebold. "Financial asset returns, direction-of-change forecasting, and volatility dynamics." *Management Science* 52, no. 8 (2006): 1273-1287.
6. Engle, Robert F. "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation." *Econometrica: Journal of the Econometric Society* (1982): 987-1007.
7. Heston, Steven L. "A closed-form solution for options with stochastic volatility with applications to bond and currency options." *Review of financial studies* 6, no. 2 (1993): 327-343.
8. Jordan, Steven J., Andrew J. Vivian, and Mark E. Wohar. "Forecasting returns: new European evidence." *Journal of Empirical Finance* 26 (2014): 76-95.
9. Lettau, Martin, and Sydney Ludvigson. "Consumption, aggregate wealth, and expected stock returns." *the Journal of Finance* 56, no. 3 (2001): 815-849.
10. Liew, Jim Kyung-Soo and Budavári, Tamás, “The 'Sixth' Factor -- Social Media Factor Derived Directly from Tweet Sentiments.” *Journal of Portfolio Management (Forthcoming)*
11. Nelson, Daniel B. "Conditional heteroskedasticity in asset returns: A new approach." *Econometrica: Journal of the Econometric Society* (1991): 347-370.
12. Rozeff, Michael S. "Dividend yields are equity risk premiums." *Journal of Portfolio management* (1984): 68-75.
13. Sul, Hong Keel, Alan R. Dennis, and Lingyao Ivy Yuan. "Trading on Twitter: The financial information content of emotion in social media." In *System Sciences (HICSS), 2014 47th Hawaii International Conference on*, pp. 806-815. IEEE, 2014.
14. Taylor, Stephen J. Modelling financial time series. 2007.