

# Can Sentiment Analysis and Options Volume Anticipate Future Returns?

Patrick Houlihan<sup>1</sup>  · Germán G. Creamer<sup>1</sup> 

Accepted: 4 May 2017 / Published online: 24 May 2017  
© Springer Science+Business Media New York 2017

**Abstract** This paper evaluates the question of whether sentiment extracted from social media and options volume anticipates future asset return. The research utilized both textual based data and a particular market data derived call-put ratio, collected between July 2009 and September 2012. It shows that: (1) features derived from market data and a call-put ratio can improve model performance, (2) sentiment derived from Stock-Twits, a social media platform for the financial community, further enhances model performance, (3) aggregating all features together also facilitates performance, and (4) sentiment from social media and market data can be used as risk factors in an asset pricing framework.

**Keywords** Social media · Investor sentiment · Behavioral finance · Machine learning

*“Prediction is very difficult, especially about the future.”*

Niels Bohr

## 1 Introduction

In today’s society, much human interaction takes place online through blogs, emails and chat boards, to name a few. Blogging websites like Twitter, have gained mass popularity and serve as a medium for communicating through a few sentences, embodying

---

✉ Patrick Houlihan  
phouliha@stevens.edu

Germán G. Creamer  
gcream@stevens.edu

<sup>1</sup> Stevens Institute of Technology, Hoboken, NJ, USA

the low social presence and high self-disclosure classification of Social Media as defined by [Kaplan and Haenlein \(2010\)](#). The nature of microblogs, being more to the point on a topic and less verbose (140-character limit for Twitter posts), make them prime candidates to extract sentiment for use in predictive analytics ([Bermingham and Smeaton 2010](#); [Ghiassi et al. 2013](#); [Martínez-Cámara et al. 2014](#); [Aisopos et al. 2016](#); [Saif et al. 2016](#)).

[Gruhl et al. \(2005\)](#) showed that blogs and other on-line social media websites are predecessors to 'real-world' behavior and the volumes of posts related to various products on Amazon's website are highly correlated with actual purchase decisions. [Pang and Lee \(2004\)](#) provided further support for social media data as a viable source to use in predictive analytics, which is validated by the fact that people are more inclined to share their opinions on social media websites to mere strangers. Extracting features from social media messages have proven to be a robust method for a variety of different labels. [Hennig-Thurau et al. \(2015\)](#) and [Asur and Huberman \(2010\)](#) leveraged Twitter messages and tweets related to a specific movie before its release date and showed a positive correlation between message volume and movie ticket sales. [Wu and Brynjolfsson \(2014\)](#) created an index of Google search queries related to housing prices and sales, which was shown to be a forward-looking indicator of the housing market trends. [Choi and Varian's \(2012\)](#) research showed that Google query search volume is a strong predictor of future economic activity in various industries. Also, Google trends data was leveraged to forecast weekly volatility by [Hamid and Heiden \(2015\)](#). These studies further validated the internet as a source for robust predictive data and behavior patterns.

Several studies related to capital markets suggested the volume of stock chatter messages were a predictor of volatility and next day returns ([Wysocki 1998](#); [Tumarkin and Whitelaw 2001](#); [Antweiler and Frank 2004](#); [Da et al. 2011](#); [Zhang et al. 2013, 2014](#); [Shen et al. 2016](#)). [Bollen et al. \(2011\)](#) extracted the mood state and sentiment of many users on a stock blogging site and presented highly predictive directional moves in the Dow Jones Industrial Average, two days out, with an 87.6% accuracy. Also, [Houlihan and Creamer \(2015\)](#) leveraged volume and sentiment as features from StockTwit messages and showed how they help explain continuation and reversal effects. Sentiment will be one of the main features used in this research.

Another way to capture the market sentiment is through the options market. [Anthony \(1988\)](#) has shown that increased trading in call options leads to next day gains in various underlying stocks that experienced a spike in call volume the day prior. The latter research would warrant using call option volume as a feature for a model to predict a label, such as future directional moves. [Chen and Lu \(2017\)](#) identified stocks with large decreases in option implied volatility experienced abnormal gains. [Cao et al. \(2003\)](#) find that option volume imbalances, specifically, short-term out of the money call option volumes, are predictors of pending takeovers. This finding points to a somewhat inefficient market, one where only informed traders have access to insider information before an announcement. However, this inefficiency can be leveraged as an indicator for a model that attempts to predict a label such as the next day directional move. [Billingsley and Chance \(1988\)](#) showed one such indicator, the put-call ratio, to yield abnormal gains when used in a trading strategy. The put-call ratio, PCR, is simply the total daily put volume divided by the daily call volume for a particular

equity. Intuitively, a ratio below 1.0 would point to a bullish indicator, whereas a ratio greater than 1.0 points to a bearish indicator. However, [Billingsley and Chance \(1988\)](#) show that a ratio of 0.7 is a better threshold. Additionally, not only is PCR suggestive of being a short-term indicator for near-term directional moves of stocks or indexes, but the PCR also seems to be more of a contrarian indicator than a conformist indicator. In fact, several other indicators are contrarian in nature, including short-term interest and VIX. [Hu \(2014\)](#) shows that imbalances between option volume and underlying volume predict future stock returns. [Pan and Poteshman \(2006\)](#) also show that volume for specific traders contained information about future prices. This latter study had access to a unique data set that showed new buyer volume that was broken out by various traders. Unique put-call ratios were derived using each particular trader. The data (1990–2001) was analyzed using a univariate regression, where the independent variables are the corresponding put-call ratios and the dependent variable is the next day risk-adjusted return. The results showed stocks with low put-call ratios derived from a particular trader (full-service) outperformed stocks with high put-call ratios by +40 basis points on the next day and 1% over the following week. The premise here is that informed, full-service investors trading the underlying stock instead of index options have firm, specifically related information rather than market-wide news. Also, stocks that went through periods of higher breadth (advancing issues relative to declining issues) rewarded investors with abnormal returns of 2.92% in 6 months and 4.95% in a 12-month period as shown by [Chen et al. \(2002\)](#). Also, [Houlihan and Creamer \(2014\)](#) formulated trader specific call-put ratios based on option contract volume and determined that specific traders have superior information over other traders as they showed higher Sharpe ratios with specific trader call-put ratios.

The contribution of this research suggests that sentiment extracted from social media messages and market data based call-put ratios contain information to forecast asset returns. In addition, we leveraged a unique dictionary which captures measurable mood states of authors. Sentiment is a crowd-sourced measure from the general investing community and behavior is in the form of overreactive and especially under-reactive effects observed by investors. Additionally, the call-put ratios represent traders whose sentiment and behavior can be captured through option volume data. Leveraging all features together yielded the highest monthly cumulative returns and annualized Sharpe ratios, suggesting the additional information generated by combining both sentiment and behavior from social media and market data improved asset return direction. Lastly, we validate several risk factors that help explain asset price returns.

## 2 Data

All raw data, price data, and micro-blogging messages were drawn from the period between July 2009 and September 2012. Additionally, time series were formed for all the various features (Table 1) and labels to create a matrix for all stocks used in the analysis. All features are derived on a stock by stock basis for each day.

### *Social media*

- Roughly 4.1 million messages were provided by StockTwits, a social media platform for the financial community consisting of 230,000 active members who

**Table 1** Financial and sentiment risk factors

Feature	Description	Type
Prior return	Asset return the day prior to event day	Market data
Event day return	Event day asset price return	Market data
Loughran and McDonald rating	Baseline—event day aggregated company specific (via cashtag) Z-score of the ratio of positive word count to negative word count using the <a href="#">Loughran and McDonald (2011)</a>	Sentiment
Loughran and McDonald volume	Event day aggregated company specific (via cashtag) Z-score of the ratio of positive word count to negative word count using the <a href="#">Loughran and McDonald (2011)</a>	Volume
Liu rating	Event day aggregated company specific (via cashtag) Z-score of the ratio of positive word count to negative word count using the <a href="#">Liu (2010)</a> word dictionary	Sentiment
Liu volume	Event day aggregated company specific (via cashtag) Z-score of the volume of messages with words present in the <a href="#">Liu (2010)</a> word dictionary	Volume
Pleasantness	Event day aggregated company specific (via cashtag) Z-score of the mean of Pleasantness words present in the message body using the <a href="#">Whissell et al. (1986)</a> word dictionary	Sentiment
Imagery	Event day aggregated company specific (via cashtag) Z-score of the mean of Imagery words present in the message body using the <a href="#">Whissell et al. (1986)</a> word dictionary	Sentiment
Activation	Event day aggregated company specific (via cashtag) Z-score of the mean of Activation words present in the message body using the <a href="#">Whissell et al. (1986)</a> word dictionary	Sentiment
MKT-RF	Event day market return minus risk-free rate	Market data
RF	Event day risk-free rate	Market data
SMB	Event day <a href="#">Fama and French (1993)</a> three-factor daily small minus big risk factor (firm size)	Market data
HML	Event day <a href="#">Fama and French (1993)</a> three-factor High minus low risk factor (firm valuation)	Market data
UMD	Event day <a href="#">Carhart (1997)</a> four-factor momentum risk factor (momentum)	Market data
Customer	Event day ratio of company specific (via cashtag) customer trader volume of option buy call options to summation of open buy call options and open buy put options	Sentiment
Broker dealer	Event day ratio of company specific (via cashtag) broker dealer trader volume of option buy call options to summation of open buy call options and open buy put options	Sentiment
Proprietary	Event day ratio of company specific (via cashtag) proprietary trader volume of option buy call options to summation of open buy call options and open buy put options	Sentiment
Professional	Event day ratio of company specific (via cashtag) professional trader volume of option buy call options to summation of open buy call options and open buy put options	Sentiment

This table shows the features used for regression analysis and the machine learning algorithms. Except for the prior event day return, all features are derived and will be used to predict post-event day asset price returns

discuss and exchange trading ideas, between July 13, 2009, and October 31, 2012. StockTwits also enabled its users to append tickers (CashTags) with a \$, that is \$TWTR, when discussing specific assets in messages, allowing for a simple regex match. This research uses only the following StockTwits fields:

- body—the message text.
- created\_at—datetime stamp of when messages were posted. Note: only messages whose timestamp of between 09:30 am EST and 4:00 pm EST were used in this analysis.
- symbols—list of tickers mentioned in message (cashtags).

### Market data

- Asset price data is from the University of Chicago’s Center for Research in Security Prices (CRSP) database. We assume an entry point at the market open price, and exit price at market close price, both per CRSP.
- Also used is a unique dataset provided by International Securities Exchange Holdings which consist of firm-wide daily option volume data broken out by various traders:
  - Customer—Option trade volume for traders acting on behalf of discount and full-service customers. This trader type dominates option volume.
  - Broker Dealer—Option trade volume for traders acting on behalf of institutional clients.
  - Proprietary—Option trade volume for proprietary traders acting on behalf of their firm.

## 3 Fama–MacBeth Regression Analysis

Before delving into the methodology, we first need to determine if the proposed features help explain the variability of asset price returns. Validating their explanatory power can be performed through the Fama–MacBeth regression estimation framework (Fama and MacBeth 1973). This method involves two regression steps. The first step consists of regressing (Formula 1) the proposed risk factors as the independent variables against each of the asset return series to compute each respective asset’s beta values.

$$R_i = \beta_{0,i} + \beta_{1,i} F_{1,i} + \dots + \beta_{m,i} F_{m,i} + \varepsilon_i \quad (1)$$

where  $R_i$ —excess returns for asset  $i$ ,  $F_{m,i}$ —risk factor  $m$  for asset  $i$ ,  $\beta_{m,i}$ —regression coefficient of asset  $i$  for factor  $m$ ,  $\varepsilon_i$ —residual of asset  $i$ .

Step two determines risk factor exposure of asset returns by running cross-sectional regressions (Formula 2) for each period of returns, against the betas, and with risk loading estimates  $\hat{\beta}$  for each asset calculated from step one.

$$R_t = \lambda_{0,t} + \hat{\beta}_{1,i} \lambda_{1,t} + \dots + \hat{\beta}_{m,i} \lambda_{m,t} + \eta_t \quad (2)$$

where  $R_t$ —excess returns for all assets at time  $t$ ,  $\hat{\beta}_{m,i}$ —risk loading estimates  $m$  from step 1 for asset  $i$ ,  $\lambda_{m,t}$ —slope  $m$  at time  $t$ ,  $\eta_t$ —idiosyncratic risk

The risk premium (exposure) for each factor is the average of the slopes ( $\lambda_{m,t}$ , Formula 3).

$$\hat{\lambda}_m = \frac{1}{T} \sum_{t=1}^T \lambda_{m,t} \quad (3)$$

where  $\lambda_{n,m}$ —period  $t$  slope for asset  $m$ ,  $\hat{\lambda}_m$ —risk exposure for factor  $m$ .

We run Fama–MacBeth regressions (Table 2) for well-known risk factors used in asset pricing models (APM), specifically, CAPM, Fama and French (1993) three-factor and Carhart (1997) four-factor to establish a baseline and understand the exposure the stocks have with these well-known risk factors. Next, we include the first sentiment factor which will act as the baseline; rating and volume derived from the Loughran and McDonald (2011) dictionary because of its popularity in the finance literature. We slowly build on this model by including the features from Table 1 in separate Fama–MacBeth regressions. Since we have nine features, not including the baseline, instead of running simulations for every possible subset ( $2^9 = 512$ ) of features, we add each one individually (Table 2) to the baseline model and run Fama–MacBeth regressions to determine their viability as risk factors.

The small-minus-big risk factor exhibited the smallest coefficient values (impact), suggesting the vast majority of stocks were not small cap stocks, but rather larger cap stocks. The evaluation of the financial risk factors (Table 2) using the Fama–MacBeth framework shows momentum (UMD) having the highest impact. These results indicate that the majority of stocks are exposed to short-term momentum effects that could be quickly shared by tweets. Also, sentiment derived from the Loughran and McDonald dictionary has a very low impact while the Liu dictionary and the Pleasantness and Activation parameters from the dictionary of affect in language (DAL) are the most important risk factors after UMD. The difference between the Loughran and McDonald and the Liu dictionaries can be explained because the first is optimized using large bodies of texts from financial reports while the second is optimized for succinct social media blogs as those used in this research. The largest impact values were observed with traders, suggesting option market behavior may drive underlying prices.

The risk premiums of the Liu dictionary and the DAL components have a negative relationship with return. This may indicate the overreaction of investors and the quick price reversal that follows any corporate news. Pearson correlation tests were run between the underlying volume and message volume (Fig. 1).

Over 70% of the stocks exhibited statistically significant correlations between underlying volume and message volume.

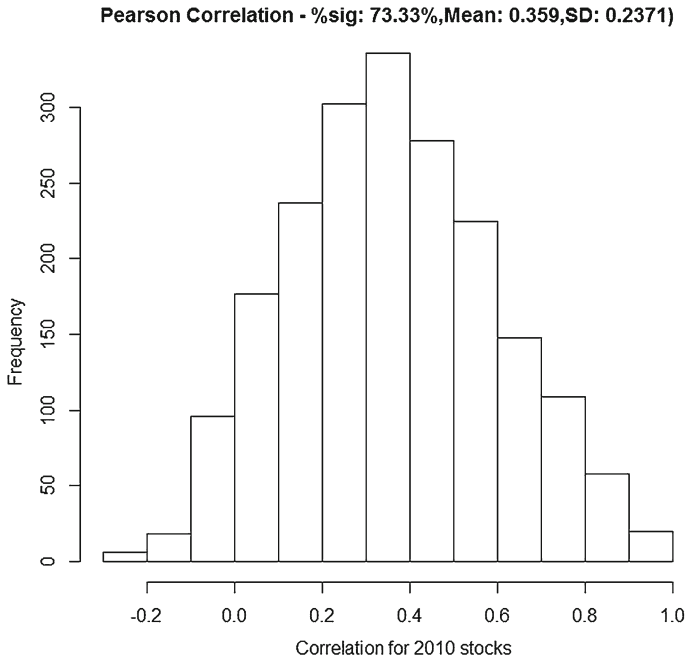
## 4 Methodology

With viable risk factors established, focus now shifts to their predictive capability. We take a machine learning approach through a majority vote, ensemble, method through leveraging five well-known classifiers to both train, validate and test a model to predict the assets price direction move, up or down, and in turn determine what position to take, long or short, respectively. Ensemble methods have been shown to outperform

**Table 2** Risk premium

Asset pricing model/Risk factors	Alpha	MKT-RF	SMB	HML	UMD	Loughran and McDonald	Liu	Pleasantness	Imagery	Activation
CAPM	0.0369	-0.0118								
Fama-French	0.0275	-0.0019	-0.0015	-0.0279						
Carhart	0.0119	0.0123	-0.0004	-0.0165	0.1297					
Loughran and McDonald	0.0116	0.0127	0.0002	-0.0162	0.1293	0.0010				
Liu	0.0062	0.0201	0.0014	-0.0176	0.1333		-0.0473			
Pleasantness	0.0119	0.0140	0.0002	-0.0146	0.1300			-0.0483		
Imagery	0.0118	0.0124	-0.0003	-0.0164	0.1296				-0.0112	
Activation	0.0137	0.0106	0.0002	-0.0154	0.1311					-0.0396
All	0.0068	0.0199	0.0023	-0.0197	0.1418	-0.0038	-0.0491	-0.0375	0.0043	-0.0425

This table shows risk premiums of financial and sentiment factors using the Fama MacBeth framework. Alpha is an average of the different assets for each model



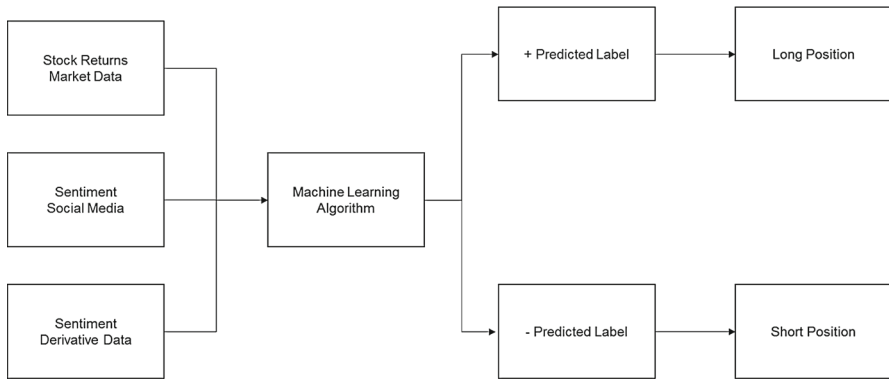
**Fig. 1** Underlying and message volume Pearson correlation. This figure shows a histogram of the Pearson correlation coefficients between underlying and message volume, the percent (% sig) that exhibited statistical significance, the mean and standard deviation (SD) of the correlation values

stand-alone classifiers (Dietterich 2000; Zhou et al. 2002; Maglogiannis 2007; Galar et al. 2012; Kanakaraj and Guddeti 2015). The machine learning classifiers chosen are listed below:

- **LogitBoost:** ensemble method of classification based on boosting that assigns more weight to the misclassified observations and minimizes the logistic loss (Friedman et al. 2000).
- **Naïve Bayes:** Bayesian parameter estimation method based on some known prior distribution (Russell et al. 2009).
- **AdaBoost:** adaptive boosting machine learning meta-algorithm used for improving performance and classifier accuracy by adding more weight to previously misclassified instances (Freund and Schapire 1997).
- **Logistic Regression:** logit based regression for categorical labels which has been shown to be an accurate classifier for binary labels (Cox 1972).
- **Bagging:** classifier that generates an aggregated predictor through multiple adaptations of a predictor; this has been shown to increase classifier accuracy by minimizing variance (Breiman 1996).

A multi-stage simulation process (Fig. 2) will be followed. Using 10-fold cross-validation, models will be trained using the above algorithms for each stock with the first 80% observations and tested with the remaining 20% observations (holdout). Splitting the dataset in this manner will prevent data snooping and adheres to the 80/20





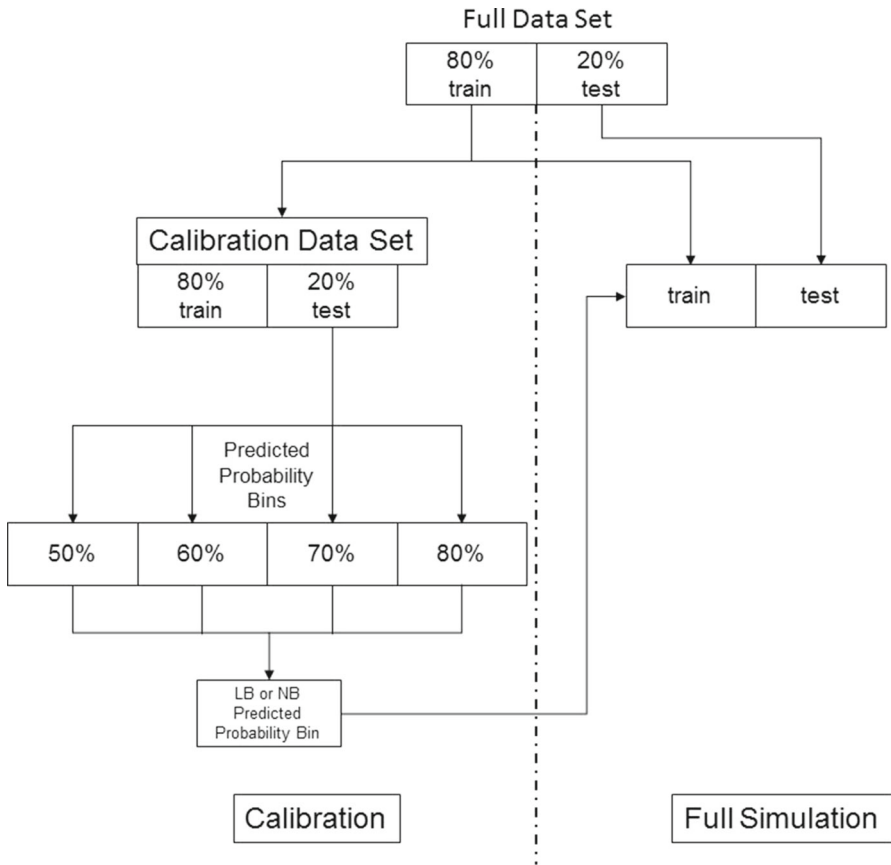
**Fig. 2** Trading strategy. This figure shows the trading strategy schema. We retrain and test the model for every feature set added and take long positions on labels predicted to be a positive one, 1, and short positions on labels predicted as zero, 0

Pareto principle. All labels have the return directional moves, up (1) or down (−1), for the next trading day. The train data will go through a calibration stage where it will be split 80% and 20% for train and test, respectively. The date ranges for train and test were, respectively: July 13, 2009, to March 10, 2012; 942 trading days, and March 11, 2012, to October 31, 2012, 235 trading days. The calibration stage will only be performed for the baseline case using the current and lagged (prior day) returns. The test stage of the calibration is further granulated into trading simulation bins between predicted probabilities of 50% and 80%, in steps of 5%, based on the forecasted return of directional moves and their respective predicted probabilities. Only assets with predicted probabilities greater than each respective bin are tradable securities or qualified assets. Based on these forecasts, our algorithm takes a long or short position, depending on the directional forecast of the label, positive or negative, respectively, on each qualified asset.

We simulate a daily trading strategy with our test dataset, taking a long or short position of the assets that have a positive or negative trend forecast, respectively. At the end of each day, we liquidate every position and calculate the daily return after transaction costs. Transaction costs open and close all positions for qualified stocks while taking into account the New York Stock Exchange rate of 0.0023 US dollars per share. The purpose of the calibration stage is to determine which predicted probability inherently achieves the highest performance. Once the best performing predicted probability is identified, we move forward with this value for the full simulation stage (Fig. 3).

We evaluate our models using the Sharpe Ratio, formula (4), average daily return and Matthews Correlation Coefficient (MCC), formula (5). The Sharpe ratio is known as the risk to variability ratio which adjusts the performance of an asset or portfolio by risk, volatility.

$$S = \frac{E[R - R_f]}{\sqrt{VAR[R]}} \quad (4)$$



**Fig. 3** Simulation flow. This figure shows the overall simulation flow. The full data set is first run through a calibration stage, left side, to determine the best performing algorithm and predicted probability bin. Once determined, the predicted probability bin is used throughout full simulation stage, right side

where  $R$ —return of asset or portfolio,  $R_f$ —risk-free rate through holding period.

MCC helps determine if the model is a robust predictor of the return direction (Matthews 1975). MCC is not only ideal for a binary label; it also overcomes the bias inherent in an unbalanced label count. Considering that markets tend to go up in the long run, return directional moves in the positive direction will outweigh moves in the negative direction. As a result, there will be a class label imbalance: more upticks (55%) than downticks (45%).

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{5}$$

where TP—true positive, forecasted true and actual true, TN—true negative, forecasted false and actual false, FP—false positive, forecasted positive and actual negative, FN—false negative, forecasted negative and actual negative.

We incrementally add features to the data set to determine the effect of certain features on model performance. The steps that run through the training and testing procedure for both the calibration and the main simulations are outlined below:

1. *Baseline features*: Use current and lagged return as features to forecast the direction of the next period return (label). This step will only be run for the calibration stage where we determine the ideal predicted probability bin and algorithm to use for the remaining steps.
2. *Baseline features and social media derived sentiment baseline feature*: Using the same baseline features, from 1, above, we include the baseline sentiment and volume feature derived from the Loughran and McDonald word dictionary; one simulation.
3. *Baseline features and first social media derived risk factor sentiment feature*: Using the same baseline features, from 1, above, we include the sentiment and volume feature derived from the Liu word dictionary; one simulation.
4. *Baseline features and market data derived sentiment*: Using the same baseline features, from 1, above, we include the aggregated ISE ratio and the individual trader ratios (customer, broker-dealer, proprietary and professional traders) according to Formula 6; five simulations.

$$\text{ISE} = \frac{\text{LONG CALLS}_{TC} (\text{Opening Position})}{\text{LONG PUTS}_{TP} (\text{Opening Position})} \quad (6)$$

where  $TC$  = trader specific call volume,  $TP$  = trader specific put volume.

The ISE call-put ratios are leading indicators of bullish or bearish market direction if the ratios are greater or less than 1 respectively.

5. *Baseline features and second social media derived risk factor sentiment feature*: Using the same baseline features, from 1, above, we include the sentiment and volume features derived from the Dictionary of Affect in Language; three simulations. [Agarwal et al. \(2009\)](#) showed that DAL accurately captured binary (positive or negative) sentiment from tweets and [Nguyen et al. \(2015\)](#) and [Xie et al. \(2013\)](#) successfully used semantic frames to predict future stock prices. Also, we take an approach similar to recent studies ([Cambria and White 2014](#); [Cambria et al. 2013](#); [Poria et al. 2014](#)) that leveraged dictionaries which expand meanings of words into multiple dimensions. We use a unique dictionary that contains multiple dimensions and extend these studies further by aggregating together with market data sentiment and additional sentiment measures (step 6). The DAL parameters are known as Pleasantness, Activation, and Imagery. These parameters, the additional three features, capture human emotion similar to Googles Profile of Mood states (six total emotional states) that were successfully used by [Bollen et al. \(2011\)](#), [Abu Bakar et al. \(2014\)](#), [Siganos et al. \(2014\)](#), [Kim and Kim \(2014\)](#), and [Danbolt et al. \(2015\)](#) to predict future directional moves in stocks. We score all messages using the DAL parameter scores by tokenizing each message and taking the average of each parameter for every message. Using this dictionary, we assume when authors write negative text they use more negative words than positive words and viceversa.
6. *Baseline features and all market data derived and social media derived statistics*; one simulation.

**Table 3** Predicted probability cutoffs

Predicted probability (%)	t + 1 Return	Total trades	Sharpe ratio	MCC	Standard Deviation
50	0.0102	3850	0.0046	-0.2363	0.0224
55	0.0790	1664	0.0308	-0.2406	0.0257
60	-0.0208	1243	-0.0078	-0.2413	0.0268
65	0.1719	884	0.0590	-0.2295	0.0291
70	0.0656	557	-0.0224	-0.2559	0.0292
75	0.1721	293	0.0719	-0.2454	0.0239
80%	0.3449	133	0.1337	-0.1920	0.0258
Anova	<<0.01**		<<0.01**	<<0.01**	

This table shows the annualized Sharpe ratios, average monthly return, t + 1 Return, and average Matthews correlation coefficient, MCC, for different predicted probability bins using an ensemble method with lagged log returns, current log returns and the baseline sentiment as features to predict the next period directional move, label. Also shown are associated *p* values of ANOVAs for comparison among the different predicted probability cut-offs

\*\* Represents 1% statistical significance

## 5 Results

To determine the ideal predicted probability bin for the validation stage, we run the machine learning ensemble method using the features from step 2. The 65% predicted probability bin yields a substantial number of trades: 884, an annualized Sharpe ratio of 0.2043, and an average monthly return of 0.1719 basis points (Table 3). We then move forward with 65% as the predicted probability bin to trade. All returns and Sharpe ratios show a significant difference at the 99% level.

The implementation of our forecast and trading strategy shows that the Loughran and McDonald dictionary outperforms the model based only on prior return (Table 4). However, the Liu dictionary and the components of the DAL (pleasantness, imagery, and activation) outperform the baseline Loughran and McDonald dictionary, suggesting that these are superior for our social media data set. The pleasantness sentiment parameter yields the largest Sharpe ratio (1.0139), return (2.29%), and MCC (-0.33). Out of the specific traders, the broker-dealer ratio exhibited the largest Sharpe ratio (0.4491), return (1.48%), and MCC (-0.32), suggesting the broker-dealer has superior information. This latter result is not surprising as broker-dealers have substantial resources at their disposal that also act on behalf of very sophisticated traders.

As in the case of the risk premiums of the Liu dictionary and DAL, all the sentiment indicators show negative MCCs. A forecasting model can capture this pattern and use it to anticipate the return direction.

A trading strategy based on the pleasantness category shows the largest positive and statistically significant alpha after adjusting by excess market return (MKT-RF), size (SMB), valuation (HML) and momentum effect (UMD) (Table 5). The pleasantness category more closely reflects the sentiment associated with every word. This characteristic explains its selection as a risk factor in our predictive models. Next, we

**Table 4** Return, volatility, and Sharpe ratio of trading strategies

Factors	Return (%)	Volatility	Sharpe ratio	MCC
Return	-1.04	0.0292	-0.3559	-0.2559
Loughran-McDonald	0.08	0.0227	0.0341	-0.3007
Liu	1.07	0.0227	0.4719	-0.2956
Customer	0.17	0.0272	0.0642	-0.2535
Broker dealer	1.48	0.0330	0.4491	-0.3225
Proprietary	0.77	0.0208	0.3685	-0.3040
Professional	0.68	0.0215	0.3160	-0.2101
ISE	2.78	0.0327	0.8508	-0.1291
Pleasantness	2.29	0.0226	1.0139	-0.3303
Imagery	0.18	0.0234	0.0775	-0.2850
Activation	0.84	0.0200	0.4215	-0.3078
All	4.09	0.0272	1.5003	-0.3156

This table shows the annualized Sharpe ratio, return, volatility, and average Matthews correlation coefficient (MCC) for the test period. Sharpe ratios and average returns are both adjusted for transaction costs and risk-free rate

combined all risk factors as features where the largest Sharpe ratio (1.5003), return (09%), and MCC (-0.32) was observed.

## 6 Discussion

The baseline simulation, step 2, yielded the worst results and the Liu dictionary, step 3, beat out the baseline dictionary. The Loughran and McDonald dictionary was optimized using large bodies of texts from financial statements while the Liu dictionary was optimized for succinct social media blogs, so this result is not surprising. Performance results further improved with the trader ratios, step 4, especially with the broker-dealer trader. This suggests that the behavior patterns of various trader types are a proxy of future returns of assets. Furthermore, it is typically the savvy investor type who trades derivative products, options, and who has access to both superior information and the means to trade, not only from a monetary perspective but also technological. This was most apparent in the simulation runs using the broker-dealer ratio which achieved the most robust performance out of all other ratios. Recall broker-dealer traders operate on behalf of institutional clients. Out of all trader types, institutional clients have access to both superior research and technology. When institutional clients channel through broker-dealers for trade execution, not only do broker-dealers gain access to information inherent in these trades, but also have access to their internal information and technology, which are not available to other traders.

Leveraging the customer ratio yielded the lowest performance out of all trader ratios. Again, this trader constitutes both discount and full service. The discount customer is most likely considered a noise trader (De Long et al. 1990a, b) and the full service could be considered a hybrid between noise and positive feedback traders. The

**Table 5** Risk-adjusted trading strategy return

Factors	Alpha	MKT-RF	SMB	HML	UMD	ADJ R-SQUARED
Return	-0.0275	-0.1255	0.5939*	-0.0430	0.4850	0.0026
	-0.0402	-0.0428	0.7453*	0.1967		0.0067
Loughran and McDonald	0.0440	0.3912*	-0.0307	-0.5425*	0.2045	0.0125
	0.0436	0.4299*	0.0426	-0.4347		0.0089
Liu	0.0931	0.3658*	0.0060	-0.3965	0.1547	0.0094
	0.0931	0.3966*	0.0659	-0.3150		0.0132
Customer	0.0664	0.3164*	0.2326	-0.5768	-0.2358	0.0082
	0.0673	0.2699	0.1255	-0.7021*		0.0085
Broker dealer	0.1750	-0.0599	1.3602*	-0.4425	0.1014	0.0165
	0.1732	-0.0456	1.4046*	-0.3899		0.0117
Proprietary	0.0559	0.3532*	0.1243	-0.2519	0.3725*	0.0109
	0.0438	0.4207*	0.2931	-0.0311		0.0169
Professional	0.0833	0.2442	0.4977*	-0.3468	0.0252	0.0199
	0.0831	0.2491	0.5064*	-0.3345		0.0182
All traders	0.1751	0.4137*	0.2002	-0.2150	0.0506	0.0091
	0.1732	0.4228*	0.2171	-0.1885		0.0117
Pleasantness	0.2188*	0.2695*	0.1719	-0.7909*	0.2136	0.0186
	0.2214*	0.3104*	0.2388	-0.6716*		0.0197
Imagery	0.0251	0.3695*	0.1806	-0.2642	-0.0800	0.0000
	0.0256	0.3539*	0.1510	-0.3136		0.0000
Activation	0.0591	0.1667	0.0016	-0.0461	0.3703*	0.0008
	0.0480	0.2395*	0.1348	0.1793		0.0069
All	0.3707	0.7115*	-0.4420	-1.2688*	0.3199	0.0395
	0.3730	0.772*	-0.3475	-1.129*		0.0376

The first regression of each sentiment category uses the Fama–French three-factor model to adjust the trading strategy returns (Alpha) by excess market return (MKT-RF), size (SMB) and valuation (HML). The second regression uses the Carhart four-factor model to evaluate the additional impact of the momentum factor (UMD)

\* Represents statistical significance at the 5% level

discount trader will not have access to superior information, and usually, constitutes the majority of the herd. Full service would have access to superior information, but the sheer numbers of discount far outweigh the full-service customer, which washes out any performance advantages that could have been observed if the option data was broken down by full service and discount. Proprietary trader performances yielded better results than the customer but slightly lower than the broker–dealer. Per [Pan and Poteshman \(2006\)](#), these traders possess little information about future stock prices and leverage the options market for hedging purposes. Overall, these results are promising, considering returns were adjusted for both transaction costs and market effects while residual alpha was still present. Future research will use the same framework by aggregating sentiment together for stocks in the same industry and sector.

## 7 Conclusion

This research shows the importance of both sentiment types extracted from social media messages and market data derived signals to forecast asset return. Both features contain a sentiment and behavioral aspect. Sentiment is an aggregated opinion of the general investing community, and the call-put ratios are sentiment for various trader ratios beyond what would be found on social media platforms. Social media provides information about the masses opinions and moods and a profile of the more conformist traders. The market data derived signal consists of customer, broker-dealer and proprietary traders who are not, besides customer, on social media outlets broadcasting their opinions to the world about stocks since there are strict SEC rules preventing them from doing so. However, we can capture their behavior through the option volume data. It is suggested that the broker-dealer trader may possess superior information with respect to all other traders as we saw the highest performance out of all simulations with this trader ratio used with lagged return, current return and sentiment. This research suggests that the additional information generated by combining both feature types, sentiment from both the masses and specific trader type behavior, from two forms, text and market data, improve the asset return prediction. This research shows the importance of sentiment extracted from social media messages and market data to both explain and forecast asset price returns. We demonstrate that sentiment extracted from social media and market data are valid additional risk factors in relation to the Fama-French and Carhart models. Furthermore, these results suggest that sentiment can be harnessed in a predictive analytics framework to realize positive residual alpha after adjusting for market effects.

**Acknowledgements** The authors would like to thank StockTwits for providing the messages. The authors also thank Shu-Heng Chen, Blake LeBaron, Jon Kaufman, David Starer, Hamed Ghoddsi, Khaldoun Khashanah, and three anonymous referees for suggestions and informal discussions about this research. The opinions presented are the exclusive responsibility of the authors.

## References

- Abu Bakar, A., Siganos, A., & Vagenas-Nanos, E. (2014). Does mood explain the monday effect? *Journal of Forecasting*, 33(6), 409–418.
- Agarwal, A., Biadys, F., & Mckeown, K. R. (2009). Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams. In *Proceedings of the 12th conference of the European chapter of the association for computational linguistics*, Athens, Greece, pp. 24–32.
- Aisopos, F., Tzannetos, D., Violos, J. & Varvarigou, T. (2016). Using n-gram graphs for sentiment analysis: an extended study on Twitter. In *Proceedings of the 2016 IEEE second international conference on big data computing service and applications*, Oxford, United Kingdom, pp. 44–51.
- Anthony, J. H. (1988). The interrelation of stock and options market trading-volume data. *The Journal of Finance*, 43(4), 949–964.
- Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance*, 59(3), 1259–1294.
- Asur, S., & Huberman, B. A. (2010). Predicting the future with social media. In *Proceedings of the 2010 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology (WI-IAT)*, Los Alamitos, CA, Vol. 1 (pp. 492–499).
- Bermingham, A., & Smeaton, A. F. (2010). Classifying sentiment in microblogs: Is brevity an advantage? In *Proceedings of the 19th ACM international conference on Information and Knowledge Management*, Toronto, CA (pp. 1833–1836).

- Billingsley, R. S., & Chance, D. M. (1988). Put-call ratios and market timing effectiveness. *The Journal of Portfolio Management*, 15(1), 25–28.
- Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1–8.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123–140.
- Cao, C., Griffin, J. M., & Chen, Z. (2003). Informational content of option volume prior to takeovers, Yale SOM Working Paper No. ES-31.
- Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, 28(2), 15–21.
- Cambria, E., & White, B. (2014). Jumping NLP curves: A review of natural language processing research. *IEEE Computational Intelligence Magazine*, 9(2), 48–57.
- Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52(1), 57–82.
- Chen, J., Hong, H., & Stein, J. C. (2002). Breadth of ownership and stock returns. *Journal of Financial Economics*, 66(2), 171–205.
- Chen, Z., & Lu, A. (2017). Slow diffusion of information and price momentum in stocks: Evidence from options markets. *Journal of Banking and Finance*, 75, 98–108.
- Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic Record*, 88(s1), 2–9.
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2), 187–220.
- Danbolt, J., Siganos, A., & Vagenas-Nanos, E. (2015). Investor sentiment and bidder announcement abnormal returns. *Journal of Corporate Finance*, 33, 164–179.
- Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The Journal of Finance*, 66(5), 1461–1499.
- De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990a). Positive feedback investment strategies and destabilizing rational speculation. *The Journal of Finance*, 45(2), 379–395.
- De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990b). Noise trader risk in financial markets. *Journal of Political Economy*, 98(4), 703–738.
- Dietterich, T. G. (2000). Ensemble methods in machine learning. In *Multiple classifier systems. MCS 2000. Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, Vol. 1857.
- Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *The Journal of Political Economy*, 81(3), 607–636.
- Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1), 3–56.
- Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119–139.
- Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: A statistical view of boosting. *The Annals of Statistics*, 28(2), 337–407.
- Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., & Herrera, F. (2012). A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(4), 463–484.
- Ghiassi, M., Skinner, J., & Zimbra, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with Applications*, 40(16), 6266–6282.
- Gruhl, D., Guha, R., Kumar, R., Novak, J., & Tomkins, A. (2005). The predictive power of online chatter. In *Proceedings of the eleventh ACM SIGKDD international conference on knowledge discovery and data mining*, Chicago, IL (pp. 78–87).
- Hamid, A., & Heiden, M. (2015). Forecasting volatility with empirical similarity and Google Trends. *Journal of Economic Behavior and Organization*, 117, 62–81.
- Hennig-Thurau, T., Wiertz, C., & Feldhaus, F. (2015). Does Twitter matter? The impact of microblogging word of mouth on consumers' adoption of new movies. *Journal of the Academy of Marketing Science*, 43(3), 375–394.
- Houlihan, P. & Creamer, G. G. (2014). Leveraging a call-put ratio as a trading signal. Howe School Research Paper No. 2015–49. Available at SSRN: <https://ssrn.com/abstract=2363475>.
- Houlihan, P. & Creamer, G. G. (2015). Leveraging social media to predict continuation and reversal in asset prices. Available at SSRN: <https://ssrn.com/abstract=2527968>.
- Hu, J. (2014). Does option trading convey stock price information? *Journal of Financial Economics*, 111(3), 625–645.
- Liu, B. (2010). Sentiment analysis and Subjectivity. *Handbook of Natural Language Processing*, 2, 627–666.



- Kanakaraj, M. & Guddeti, R. M. R. (2015). Performance analysis of Ensemble methods on Twitter sentiment analysis using NLP techniques. In *2015 Ninth IEEE international conference on semantic computing (ICSC)*, Anaheim, CA (pp. 169–170).
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, 53(1), 59–68.
- Kim, S. H., & Kim, D. (2014). Investor sentiment from internet message postings and the predictability of stock returns. *Journal of Economic Behavior and Organization*, 107, 708–729.
- Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, 66(1), 35–65.
- Maglogiannis, I. G. (2007). *Emerging artificial intelligence applications in computer engineering: Real word AI systems with applications in eHealth, HCI, information retrieval and pervasive technologies*. Amsterdam: Ios Press.
- Martínez-Cámara, E., Martín-Valdivia, M. T., Urena-López, L. A., & Montejo-Ráez, A. R. (2014). Sentiment analysis in Twitter. *Natural Language Engineering*, 20(01), 1–28.
- Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, 405(2), 442–451.
- Nguyen, T. H., Shirai, K., & Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications*, 42(24), 9603–9611.
- Pan, J., & Poteshman, A. M. (2006). The information in option volume for future stock prices. *Review of Financial Studies*, 19(3), 871–908.
- Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd annual meeting on association for computational linguistics*, Barcelona, Spain (p. 271).
- Poria, S., Cambria, E., Winterstein, G., & Huang, G. B. (2014). Sentic patterns: Dependency-based rules for concept-level sentiment analysis. *Knowledge-Based Systems*, 69, 45–63.
- Russell, S., Norvig, P., & Intelligence A. (2009). *Artificial Intelligence: A modern approach* (3rd ed.). Englewood Cliffs: Prentice-Hall.
- Saif, H., He, Y., Fernandez, M., & Alani, H. (2016). Contextual semantics for sentiment analysis of Twitter. *Information Processing and Management*, 52(1), 5–19.
- Shen, D., Zhang, W., Xiong, X., Li, X., & Zhang, Y. (2016). Trading and non-trading period Internet information flow and intraday return volatility. *Physica A: Statistical Mechanics and its Applications*, 451, 519–524.
- Siganos, A., Vagenas-Nanos, E., & Verwijmeren, P. (2014). Facebook's daily sentiment and international stock markets. *Journal of Economic Behavior and Organization*, 107, 730–743.
- Tumarkin, R., & Whitelaw, R. F. (2001). News or noise? Internet postings and stock prices. *Financial Analysts Journal*, 57(3), 41–51.
- Whissell, C., Fournier, M., Pelland, R., Weir, D., & Makarec, K. (1986). A dictionary of affect in language: IV. Reliability, validity, and applications. *Perceptual and Motor Skills*, 62(3), 875–888.
- Wu, L. & Brynjolfsson, E. (2014). The future of prediction: How Google searches foreshadow housing prices and sales. In A. Goldfarb, S. M. Greenstein, and C. E. Tucker (Eds). *Economic analysis of the digital economy*. University of Chicago Press, Chicago, IL, 89–118.
- Wysocki, P. D. (1998). Cheap talk on the web: The determinants of postings on stock message boards. *University of Michigan Business School Working Paper*, (98025).
- Xie, B., Passonneau, R. J., Wu, L., & Creamer, G. G. (2013). Semantic frames to predict stock price movement. In *Proceedings of the 51st annual meeting of the association for computational linguistics*, Sofia, Bulgaria (pp. 873–883).
- Zhang, W., Shen, D., Zhang, Y., & Xiong, X. (2013). Open source information, investor attention, and asset pricing. *Economic Modelling*, 33, 613–619.
- Zhang, Y., Feng, L., Jin, X., Shen, D., Xiong, X., & Zhang, W. (2014). Internet information arrival and volatility of SME PRICE INDEX. *Physica A: Statistical Mechanics and its Applications*, 399, 70–74.
- Zhou, Z. H., Wu, J., & Tang, W. (2002). Ensembling neural networks: Many could be better than all. *Artificial Intelligence*, 137(1), 239–263.