Review Article

International Journal of Advanced Computer Research, Vol 6(22)
ISSN (Print): 2249-7277 ISSN (Online): 2277-7970
http://dx.doi.org/10.19101/IJACR.2016.622012

Emotional speculative behavior in the option market

Francesco Corea*
Department of Economics and Finance, LUISS Guido Carli University, Rome, Italy

Received: 25-November-2015; Revised: 28-December-2015; Accepted: 10-January-2016

Abstract
Social media data have been proved to be effective in augmenting stock price forecasting models before ([8], [12]), but given the intrinsic speculative nature of traders who may use these innovative datasets, it appears more reasonable to investigate the relation between the Twitter data and the stock option prices. The underlying hypothesis is indeed that speculative trading strategies as the ones based on social media inference are may be more effective if evaluated on speculative instruments instead of simple stock prices. Consistent with previous works, it has been then studied on an intraday basis for three major technology stocks over a two-month period the relation between investors' sentiment and basic financial products. A set of different variables has been created to include different interactions between sentiment and option prices, and a statistical selection model has been put in charge of identifying the most relevant correlations. The results are quite mixed: social media data seem to be indeed useful for predicting some option prices, but no others, and in particular are able to better explain single companies' option price oscillations rather than the ones related to general indexes such as the Nasdaq-100.

Keywords
High frequency trading, Options, Sentiment analysis, Stepwise regression, Twitter.

1.Introduction

The amount of data produced nowadays is increasing exponentially, and as everyone already knows more than 90% of data available today have been generated in the last two-four years [30]. Even if a great portion of these new data comes from the Internet of Things, from sensors, and from mobile applications, another relevant quantity is generated by social networks and through web contents. The analysis of these data [1] may have different impact in different fields: it could indeed help in predicting the presidential election's outcome [34], understanding a disease spread [14], or assessing the success of a new music album release [18]. No matter the field of applications, though, social media data seem to be fundamental to predict the future [3].

The study of the investors' sentiment is deeply rooted in the financial literature [4,5] but the use of new sources of information has given traders and institutional investors a new important way to gain a competitive advantage [15,16].

It seems to be clear that Twitter represents one of the most important sources of social media data, given the standard format of the posting activities – every tweet is limited to 140 characters – and for the fast diffusion it had in the last few years, although many other informative channels have been exploited as well. Hence, it makes sense to split the preexisting works into three branches, based on the use of solo Twitter for financial markets applications, on the use of distinct social media data, or for a different final business purpose. In the first group, pioneering works have to be attributed to [8], who started a flow of research that has been then adopted and enhanced by others [25]. They focused on interpreting and deducing human emotions from microblogging activity, in order to provide insights on the movements of the Dow Jones Industrial Average Index [8, 22]. Similarly, [27] discovered a positive correlation between the stock volume and the Twitter volume, while Corea and Corea and Cervellati [12,13] selected three major technology stocks and built an indicator for predicting the variations in the Nasdaq-100. Brown [10] instead enquired the importance of the user's reputation as a driver for stock market changes rather that focusing on the tweet content. Ruiz et al. [28] concentrated their effort in explaining hidden correlations between

*Author for correspondence

microblogging activity and financial time series, while Oh et al. [26] and Sprenger et al. [31] put their emphasis on the microblogging informative power. Finally, Mao et al. [23] readapted the previous approach to analyze international financial markets, mixing both Twitter and Google data.

In the second branch, i.e., the use of social media data different from Twitter, Lavrenko et al. [21] a few decades ago and Schumaker et al. [29] later on analyzed how financial news release impacts on the financial markets. Tetlock [32, 33] dug more into this field focusing particularly on the effect of negative news, and Barber et al. [6] showed instead how analysts' recommendations may be exploited for extrapolating future trends. Other meaningful progresses in the field have been done by Antweiler et al. [2], and Koski et al. [20], that used stock message board as a primary source of information, while web search [9] and blogs [17] have been alternatively considered in formulating stock market predictions.

The last group – the one that concerns different business applications for sentiment analysis started with Fisher and Statman (2000) [19], in a work in which they used investors' sentiment for asset tactical allocation, and then varies from movie revenue forecasting [24] to the commercial sales prediction [11].

Hence, the literature is quite vast, and in order to give a contribution, the aim of this work is to analyze both for single stocks and stock index (i.e., the Nasdaq-100), how option price movements could be explained by changes into the tweets' sentiment. The analysis will be performed on an intraday basis, because on a daily one has been proved to be ineffective [12], and the paper will have the structured as follows: section 2 will take care of the data gathering and the methodology used. In section 3, there are going to be showed some achievements of the models proposed, while section 4 will finally draw the conclusions, providing suggestions for future researches and some insights from the current study.

2.Methodologies and dataset construction

For consistency and comparability with previous works [8, 12], the same large technology stocks have been analyzed, i.e., Google, Apple, and Facebook, as well as general stock indexes such as the Nasdaq-100 [13]. Contrarily to what previously done, though the following analysis has been focused on the study of the relationship between the social expressions and the option prices, and in particular on the impact of the former one on the latter. Hence, the stock prices for the three stocks and for the Nasdaq-100 have been extracted from Bloomberg, and the option prices have been derived through the following Black and Scholes set of equations (the notation is the standard one used in the literature):

$$Call = SN(d_1) - K^{-rT}N(d_2) \qquad (1)$$

$$Put = K^{-rT}N(-d_2) - SN(-d_1) \qquad (2)$$

$$d_1 = \frac{\ln\left(\frac{S}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}} \qquad (3)$$

$$d_2 = d_1 - \sigma\sqrt{T} \qquad (4)$$

Where the 3-months interest rate has been obtained from the Federal Reserve website, and the options have consistently a 3-months expiration time.

On the other side, the tweets, with relative sentiment and klout scores, have been acquired from DataSift. The sentiment score measures the level of positivity (or negativity) of the human opinion explicated in the text, while the klout score canalizes in a single value the degree of social influence of an individual. It varies between 1 and 100, and to a higher value corresponds a higher influence power. The time period considered for the analysis goes from September 24th to November 21st 2014, and it was possible to collect the data second by second, and eventually aggregate them on a minute basis.

For the content of the messages, two relevant choices have been taken for the sake of the study: First of all, only English tweets have been pondered – because they represent the almost totality of the tweets universe – and secondly, in order to reduce the high volatility deriving from unrelated or misunderstood tweets, it has been decided to include only the messages strictly connected to the stock valuation. In other words, the financial literacy of the bloggers has been proxied through the selection of the tweets where the company's ticker was mentioned. Totally, almost 88,000 thousand of tweets has been grouped for the Apple stock, about 44,000 for Facebook, and less than 32,000 for Google.

Regarding instead one of the key variable the tweet sentiment a scoring algorithm assigned a value ranging from -20 to +20, respectively to extremely negative or positive tweets. In addition, a group of other variables has been ideated to take into account variations and nuances of the sentiment score above-mentioned. Thus, in a similar fashion as in Oliveira et

al. [27], it has been defined a simple sentiment mean (**SM**), and its ratio with the one-lagged value (**SR**). Furthermore, a set of indicators aimed to capture the bullishness (or bearishness) of the market has been suggested: the simple mean has been computed first individually for positive (**BBSp**) and negative tweets (**BBSn**), and then a ratio between the two has been proposed (**BBSR**). Finally, it has been added the klout score, and a 5-moving averages for the sentiment (**SMMA**).
It has been then used a simple ordinary least squares (OLS) regression model and a linear probability model (LPM) to assess the type of impact the general sentiment had on the option prices:

$$\boldsymbol{y}_t = \boldsymbol{x}_t\beta + \epsilon_t \qquad (5)$$
$$\boldsymbol{y}_t^* = \boldsymbol{x}_t\beta + \epsilon_t \qquad (6)$$

Where $\boldsymbol{y}_t^*$ is a latent variable observable only in terms of his sign. In other words:

$$\boldsymbol{y}_t^* = \begin{cases} 0, & \left(\frac{P_t}{P_{t-1}}\right) \leq 1 \\ 1, & \left(\frac{P_t}{P_{t-1}}\right) > 1 \end{cases} \qquad (5)$$

According to [12] and [13] this variable has been called *Trend*, and it has been constructed as a dummy variable with a value of 1 that indicates an *up*-movement, while 0 a *down*-movement. Furthermore, instead of selecting by hand, which of those variables to be included in the model or testing different models, it has been decided to use a selection model that automatically inserts or excludes a certain variable on the base of a threshold significance level. In this case, the value of a variable to be part of the model is 0.05, while 0.1 for being removed. There are different types of *stepwise regression* model, and here the backward version has been implemented.

The backward stepwise regression assumes to estimate the full model with all the explanatory variables in a first place. Then, if the least-significant term is statistically insignificant, it removes that variable and re-estimates the model (otherwise it stops). The process is then reiterated.

At the same time, for each step, if the most-significant excluded term is statistically significant, it adds that variable back and reestimates the model (otherwise it stops). The algorithm is thus alternatively choosing the least significant variable to drop and to be reintroduced in the model. It is a particularly smart and convenient way to select the statistical meaningful variables on the base of pre-fixed significance threshold values without having to deal with each one by hand.

Concerning the Nasdaq estimation instead, another set of indicators has been integrated, with the aim of replicating synthetically the index-as already proposed in Corea (2015). The main instruments embedded in the analysis have been therefore obtained as the simple average of the three stock's sentiment (SIT) and the weighted variation in their respective tweets volumes (SITw). The two relative moving-average versions have been also incorporated (SITma and SITwma).

A different procedure has also been implemented. In order to predict the Nasdaq-100 option oscillations, three of the major technology companies of the index itself (i.e., Google, Apple, and Facebook) have been selected ex-ante because they are expected to have a stronger weight within the stocks bundle belonging to the Nasdaq index. Hence, it has been done a kind of qualitative principal component analysis, in order to take into account from the beginning only the stocks with a higher explanatory power for the index.

Afterwards, the following models have been tested:

$$M1: \quad P_t = \alpha + \varphi P_{t-1} + \epsilon_t \qquad (6)$$

$$M2: \quad P_t = \alpha + \varphi P_{t-1} + \beta_1 SM_{Apple} + \beta_2 SM_{Google} + \beta_3 SM_{Facebook} + \epsilon_t \qquad (7)$$

$$M3: \quad P_t = \alpha + \varphi P_{t-1} + \beta_1 SMMA_{Apple} + \beta_2 SMMA_{Google} + \beta_3 SMMA_{Facebook} + \epsilon_t \qquad (8)$$

$$M4: \quad P_t = \alpha + \varphi P_{t-1} + \beta SIT_{t-1} + \epsilon_t \qquad (9)$$

The same has been studied for the weighted version, the moving average one, and finally the weighted moving average, respectively M5, M6, and M7.

3.Empirical analysis and discussion

As already previously explained, two regressions have been run for each company's stock, one for the price (the OLS), and one for trend (the LPM). The Nasdaq is going to be considered first, and then the single companies' forecasts. However, only the results relative to the call options will be showed - for the sake of completeness, the analysis has been implemented also on put options, and the conclusions remain almost the same ones. The results have been though taken out of the current work because they did not add any further value neither generate new insights for the investigation.

The results for the Nasdaq data are then shown in the *Tables 1 – 4. Tables 1 and 3* have the same meaning, and they show the results of the OLS and LPM regressions. *Tables 2 and 4* show instead the adjusted R^2 and root mean squared error for all the models considered, in order to provide a fast way to assess whether the augmented models performed better and were more accurate with respect to the benchmark.

Table 1 OLS Regression results for the different models (2-7) with respect to the benchmark (1).

Model	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
	$Price_t$	**$Price_t$**	**$Price_t$**	**$Price_t$**	**$Price_t$**	**$Price_t$**	**$Price_t$**
$Price_{t-1}$	0.998***	0.998***	0.997***	0.998***	0.998***	0.997***	0.997***
	(2328.36)	(923.86)	(655.98)	(923.50)	(923.14)	(656.77)	(656.59)
Apple SM_{t-1}		0.149					
		(0.73)					
Google SM_{t-1}		-0.382**					
		(-2.15)					
Facebook SM_{t-1}		0.288*					
		(1.75)					
Apple $SMMA_{t-1}$			-0.256				
			(-0.45)				
Google $SMMA_{t-1}$			0.214				
			(0.42)				
Facebook $SMMA_{t-1}$			0.138				
			(0.31)				
SIT_{t-1}				0.0132			
				(0.12)			
$SITw_{t-1}$					0.0425		
					(0.16)		
$SITma_{t-1}$						0.0690	
						(0.19)	
$SITwma_{t-1}$							0.103
							(0.13)

T-statistics in parentheses. *p<0.1,** p<0.01,*** p<0.001.

Table 2 Adjusted R^2 and root mean squared error for all the models.

Model	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**
Adj-R^2	0.9965	0.9963	0.9953	0.9963	0.9963	0.9953	0.9953
RMSE	23.357	24.231	27.733	24.249	24.249	27.722	27.722

Table 3 LPM Regression results for the different models (2-7) with respect to the benchmark (1).

Model	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**	**(13)**	**(14)**
	$Trend_t$	$Trend_t$	$Trend_t$	$Trend_t$	$Trend_t$	$Trend_t$	$Trend_t$
Apple SM_{t-1}		0.00217					
		(0.72)					

Google SM $_{t-1}$		0.00315					
		(1.21)					
Facebook SM $_{t-1}$		0.00315					
		(1.31)					
Apple SMMA $_{t-1}$			-0.00548				
			(-0.76)				
Google SMMA $_{t-1}$			0.0152**				
			(2.35)				
Facebook SMMA$_{t-1}$			-0.000516				
			(-0.09)				
SIT $_{t-1}$				0.00276*			
				(1.67)			
SITw$_{t-1}$					0.00877**		
					(2.19)		
SITma$_{t-1}$						0.00171	
						(0.38)	
SITwma$_{t-1}$							0.00790
							(0.81)
Trend $_{t-1}$	0.679***	0.703***	0.713***	0.702***	0.703***	0.713***	0.713***
	(127.66)	(55.21)	(45.79)	(55.21)	(55.24)	(45.77)	(45.76)

Table 4 Adjusted R^2 and root mean squared error for all the models.

Model	**M8**	**M9**	**M10**	**M11**	**M12**	**M13**	**M14**
Adj-R^2	0.4611	0.4933	0.5091	0.4932	0.4935	0.5083	0.5084
RMSE	0.3666	0.3558	0.3504	0.3558	0.3557	0.3506	0.3506

Differently with respect to previous results in the literature [12, 13], it seems that the Twitter explanatory power is fairly low concerning the price estimation. Indeed, only the sentiment tracking index variable seems to be slightly statistically significant, and the benchmark autoregressive model performs better than any other more complex variations. The opposite is true though in the trending case, in which the microblogging proves once again to be relevant for increasing the forecasting ability of the statistical models. The situation drastically changes when intraday data are instead considered for the single companies' option forecasting. Indeed, as it is provided in *Table 5*, the predictions are more complex and heterogeneous, and some of the forecasts are more accurate and complete for the directional models than with respect to the price estimations, while some others the other way round.

Table 5 Stepwise variable selection for the high-frequency prices and trends.

	APPLE		**FACEBOOK**		**GOOGLE**	
	Price	***Trend***	***Price***	***Trend***	***Price***	***Trend***
Price$_{t-1}$	1.000***		0.443***		0.900***	
	(32313.78)		(7.11)		(27.97)	
Trend $_{t-1}$		0.403***		0.624***		0.459***

	APPLE	FACEBOOK		GOOGLE	
	(17.79)		(17.50)		(5.93)
Klout $_{t-1}$	0.00859***	0.711***	0.00713***	0.762*	0.00532*
	(11.35)	(9.10)	(8.67)	(2.63)	(1.94)
SR $_{t-1}$	0.00910*			-7.883**	
	(2.11)			(-2.92)	
BBSp $_{t-1}$	0.0166***			-3.206*	
	(3.81)			(-1.97)	
BBSn $_{t-1}$	-0.00778*	-4.147***		-5.151*	-0.0378*
	(-1.74)	(-8.59)		(-2.31)	(-1.95)
SM $_{t-1}$		-1.970*			
		(-2.56)			
BBSR $_{t-1}$		-7.655***	-0.0207*	-13.20*	-0.0455*
		(-8.31)	(-1.85)	(-2.38)	(-1.75)

T-statistics in parentheses. *p<0.1,** p<0.01,*** p<0.001.

It can be noticed that, no matter the stock taken into account, the Klout score has a significant impact on the option price: influencing traders or investors who release their opinions on the web may actually affect the market's evolution. A second interesting fact is that simple indicator such as the sentiment mean has a low, meaning for these forecasting models, while more refined variables (e.g., the bullishness-bearishness ratio) are more valuable to the analysis. In particular, it seems also that negative news influence more the option prices than positive ones.

4.Conclusions

A vast literature is exploring the implications of new data sources for different field applications, and relevant progresses have been made especially in financial markets. Twitter and social media data may represent a new frontier of quantitative financial modeling, and in this work it has been given a contribution to this area. Two months of tweets have been collected, with a specific focus on three big technology companies – Apple, Google, and Facebook – and used them for refining option pricing forecasting models. It has been tried first of all to predict the Nasdaq-100 variations, with poor results concerning the price forecasting and slightly more encouraging ones regarding the directional changes. Afterwards, single companies' options have been considered, and a set of indicators has been built in order to augment simple autoregressive models. The achievements of the models in this case are relevant - even if on a modest scale, due mainly to the length of the time series - and further works will investigate for sure longer time series, different and multiple stocks, and different sectors. If will prove to be consistent, these further adjustments would increase the model and techniques standardization, making the analysis generall y applicable and transferable to different environments and maybe additional speculative instruments.

Acknowledgment

None

Conflicts of interest

The author has no conflicts of interest to declare.

References

[1] Agarwal A, Xie B, Vovsha I, Rambow O, Passonneau R. Sentiment analysis of twitter data. In proceedings of the workshop on languages in social media 2011 (pp. 30-8). Association for Computational Linguistics.

[2] Antweiler W, Frank MZ. Is all that talk just noise? The information content of internet stock message boards. The Journal of Finance. 2004; 59(3):1259-94.

[3] Asur S, Huberman B. Predicting the future with social media. International conference in web intelligence and intelligent agent technology (WI-IAT), IEEE/WIC/ACM 2010 (pp.492-99).

[4] Baker M, Wurgler J. Investor sentiment and the cross-section of stock returns. The Journal of Finance. 2006; 61(4):1645-80.

[5] Baker M, Wurgler J. Investor sentiment in the stock market. Journal of Economics Perspectives. 2007; 21(2): 129-52.

[6] Barber B, Lehavy R, McNichols M, Trueman B. Can investors profit from the prophets? Security analyst recommendations and stock returns. The Journal of Finance. 2001;56(2):531-63.

[7] Bollen J, Mao H. Twitter mood as a stock market predictor. Computer. 2011; 44(10):91-4.
[8] Bollen J, Mao H, Zeng X. Twitter mood predicts the stock market. Journal of Computational Science. 2011; 2(1):1-8.
[9] Bordino I, Battiston S, Caldarelli G, Cristelli M, Ukkonen A, Weber I. Web search queries can predict stock market volumes. PloS one. 2012; 7(7):e40014.
[10] Brown ED. Will twitter make you a better investor? A look at sentiment, user reputation and their effect on the stock market. In proceedings of the southern association for information systems conference 2012 (pp. 36-42).
[11] Choi H, Varian H. Predicting the present with Google trends. Economic Record. 2012; 88(s1):2-9.
[12] Corea F. Why social media matters: the use of twitter in portfolio strategies. International Journal of Computer Applications. 2015; 128(6):25-30.
[13] Corea F, Cervellati EM. The power of micro-blogging: how to use twitter for predicting the stock market. Eurasian Journal of Economics and Finance. 2015; 3(4):1-7.
[14] Culotta A. Towards detecting influenza epidemics by analyzing twitter messages. In proceedings of the first workshop on social media analytics 2010 (pp. 115-22). ACM.
[15] Da Z, Engelberg J, Gao P. In search of attention. The Journal of Finance. 2011; 66(5):1461-99.
[16] Da Z, Engelberg J, Gao P. The sum of all FEARS investor sentiment and asset prices. Review of Financial Studies. 2015; 28(1):1-32.
[17] De Choudhury M, Sundaram H, John A, Seligmann DD. Can blog communication dynamics be correlated with stock market activity? In proceedings of the nineteenth ACM conference on hypertext and hypermedia 2008 (pp. 55-60). ACM.
[18] Dhar V, Chang EA. Does chatter matter? The impact of user-generated content on music sales. Journal of Interactive Marketing. 2009; 23(4):300-7.
[19] Fisher KL, Statman M. Investor sentiment and stock returns. Financial Analysts Journal. 2000; 56(2):16-23.
[20] Koski JL, Rice EM, Tarhouni A. Day trading and volatility: evidence from message board postings in 2002 vs. 1999. University of Washington. 2008.
[21] Lavrenko V, Schmill M, Lawrie D, Ogilvie P, Jensen D, Allan J. Language models for financial news recommendation. In proceedings of the ninth international conference on information and knowledge management 2000 (pp. 389-96). ACM.
[22] Mao H, Counts S, Bollen J. Predicting financial markets: Comparing survey, news, twitter and search engine data. arXiv preprint arXiv:1112.1051. 2011.
[23] Mao H, Counts S, Bollen J. Quantifying the effects of online bullishness on international financial markets. In ECB workshop on using Big Data for forecasting and statistics, Frankfurt, Germany 2015.
[24] Mishne G, Glance NS. Predicting movie sales from blogger sentiment. In AAAI spring symposium: computational approaches to analyzing weblogs 2006 (pp. 155-8).
[25] Mittal A, Goel A. Stock prediction using twitter sentiment analysis. Standford University, CS229 (2011. http://cs229.stanford. edu/proj2011/GoelMittal-Stock Market Prediction UsingTwitter SentimentAnalysis. pdf). Accessed 10 October 2015.
[26] Oh C, Sheng O. Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement. ICIS 2011 proceedings 2011 (pp. 1-19).
[27] Oliveira N, Cortez P, Areal N. On the predictability of stock market behavior using stocktwits sentiment and posting volume. In progress in artificial intelligence 2013 (pp. 355-65). Springer Berlin Heidelberg.
[28] Ruiz EJ, Hristidis V, Castillo C, Gionis A, Jaimes A. Correlating financial time series with micro-blogging activity. In proceedings of the fifth ACM international conference on web search and data mining 2012 (pp. 513-22). ACM.
[29] Schumaker RP, Chen H. Textual analysis of stock market prediction using breaking financial news: the AZFin text system. ACM Transactions on Information Systems (TOIS). 2009; 27(2):12.
[30] SINTEF. Big Data, for better or worse: 90% of world's data generated over last two years. ScienceDaily. Accesed 22 May 2013.
[31] Sprenger TO, Welpe IM. Tweets and Trades: The Information Content of Stock Microblogs. 2010. Available at SSRN 1702854.
[32] Tetlock PC. Giving content to investor sentiment: the role of media in the stock market. The Journal of Finance. 2007; 62(3):1139-68.
[33] Tetlock PC, Saar-Tsechansky MA, Macskassy S. More than words: quantifying language to measure firms' fundamentals. The Journal of Finance. 2008; 63(3):1437-67.
[34] Tumasjan A, Sprenger TO, Sandner PG, Welpe IM. Predicting elections with twitter: what 140 characters reveal about political sentiment. ICWSM 2010 (pp. 178-85).

Francesco Corea is PhD student in Economics at LUISS Guido Carli University. He received his Master's Degree in Finance from Nova School of Business and Economics in 2013, and a Master's Degree in Economics (2013) and a Bachelor in Business Administration (2011) from LUISS. His research interests are econometrics and big data applications to finance and economics, as well as behavioural finance, venture capital, and insurance analytics. He is also an Anthemis Fellow, and he spent some years in the industry working for different players in the big data space.
Email: fcorea@luiss.it