Good news and bad news: Do online investor sentiments reaction to return news asymmetric?

ALYA AL NASSERI[1], FAEK MENLA ALI[2], and ALLAN TUCKER[3]

[1] Brunel Business School, Brunel University London, UK
[2] Department of Economics and Finance, Brunel University London, UK
[3] Department of Information Systems, Brunel University London, UK

Abstract
There is growing evidence to suggest that the impact of positive and negative news are asymmetric- that the impact of bad news has a much greater effect on investors' sentiment than positive news does; investors react more harshly when bad news is disseminated. Using daily data from 30 companies listed on the DJIA index over the period April 3, 2012 to April 5, 2013, we analyse 289,443 online tweets, from the so-called StockTwits, and construct a measure of online investor sentiment. The aim of this paper is to explore the asymmetric responses of online investor sentiments to different news in different market conditions (i.e. bull versus bear market). Applying data mining techniques for sentiment detection coupled with a nonlinear econometric model, we find strong evidence of asymmetry. The result provides evidence that investor sentiments exhibit different sensitivity to news in the bull and bear market. In particular, return effects on sentiment are positive (negative) in the bull and (bear) market where this effect is more pronounced in the bull market.

1 Introduction

Over the past few decades, a growing body of literature has focused the debate on how fast information is incorporated into security prices (Fama, 1965). Most of these studies are based on the assumptions underlying the two leading theories in financial economics: Efficient Market Hypothesis (EMH) and Random Walk (RW) theory. However, the largest criticism of these theories is that both do not incorporate the behavioural component in their models and that news is treated as neutral information. Psychological researchers however argue that the effect of different types of news (good or bad) certainly affect individual sentiments of that news item. A large body of recent finance literature however, recognize the various affects of news on investor sentiments and have provided empirical evidence that while news undoubtedly influences security prices in the stock market, its impact on public mood and emotions (sentiments) may play an equally important role (i.e. Baker and Wurgler, 2007; Brown and Cilff 2004; Verma Verma, 2007). Behavioural finance has provided evidence that noise investors' emotions, preferences and mistaken beliefs can affect the decisions of other investors in the market and may result in shifting the asset's value from its fundamental level. The extreme deviations from fundamentals may be a result of noise traders overreacting or under-reacting to good and bad news, causing price levels and risk to deviate far more drastically from expected levels than would have been actually required by the news. Despite the well-recognized literature on the impact of news on sentiments, some previous studies treat news information as neutral and rarely differentiate between good and bad news (Bowman, 1983, among others). It is unlikely that investors' responses to positive and negative information are symmetric. It has been argued that news of all type (i.e., positive vs. negative) should have different impact on investor sentiments. Therefore, it is unaspiringly that investors shows different reactions depending on the types of news released in the market. Empirical evidence in the context of macro economic news, firm-specific news argues that investors' responses to positive and negative information are asymmetric and that negative news has a more significant and harsh affect than does positive news (for a summary, see Soroka, (2006) and Pritamani and Singal, (2001)). Using news releases on traditional sources of information such as those in the *Wall Street Journal* and *Newswire* (e.g. Tetlock, 2007) show that news sentiment has an effect on market reactions and that news of negative nature has a very significant influence on some market indicators such as market liquidity.

The contribution of the present study to the existing literature is threefold. First we examine the impact of different news (good vs. bad news) on investor sentiments by using relatively new data from an online stock forum (StockTwits). The high volume of message posts, the real-time message streams and the efficient diffusion mechanism of information are the three distinct features of the stock micro-blogging forum. Therefore, our sentiment measure reflects the natural market conversations while concurrently distinguishing between good and bad news information. Second, Our empirical methodology accounts for possible asymmetries in the effects of market news on investor sentiments in different states of the market by employing the nonlinear model to examine the effect of news returns on investor sentiments in two different regimes; bull and bear market periods. Third, this paper seeks to draw research in behavioural finance together with research from data mining and build a more thorough account of the impact and magnitude of effects of asymmetric responses of investor's sentiments to different types of news. In particular, this paper combines data mining techniques with financial econometric modelling to investigate the asymmetries evident in online investors' opinion to different types of market news.

The paper is organised as follows. Section 2 reviews the related literature on online investing forums and different classification algorithms. Section 3 presents the data, the classification method employed, and our investor sentiment measure. Section 4 describes the simple Markov- switching model of returns to estimate the bull and bear market regimes. Section 5 presents and discusses the empirical results. Finally, Section 6 provides some concluding remarks.

2 Related Work

2.1 Online Investment Forum and Stock Market

A growing body of empirical research has been undertaken to investigate the predictive power of online investing forums in predicting various financial market indicators; all of these papers have focused on message boards, financial news articles and recently on micro-blogging forums. *Internet message board* is one of the most popular investment forums that provides an effective means for investors to communicate, disseminate and discover information (Delort et al., 2012) Previous research studies have begun to explore the impact of stock message boards on financial markets and stock price behaviour (Wysocki, 1998; Tumarkin and Whitelaw, 2001; Antweiler and Frank, 2004). The two initial papers to analytically investigate Internet posting were Wysocki (1998) and Tumarkin and Whitelaw (2001) who measured the correlation between the message volume and the next day trading volume and returns. Their findings reveal that firms with high volume postings characterized as high market valuation with high return and accounting performance, high volatility and trading volume. One major criticism of the above-mentioned studies is that they rely too heavily on quantitative data of internet message boards such as message volume and users' ratings. Unlike previous works, the most complete study of Internet message boards is by Antweiler and Frank (2004), who focused on qualitative as well as quantitative data analysis of the internet messages posted on Yahoo Finance and Raging Bull. They determined the correlation between activity on Internet message boards and stock volatility and trading volume. They found that positive shocks to message board posting levels do predict negative stock returns the following day. The literature related that the impact of *financial news articles* and investment stories on stock markets are vast. Gidofalvi (2001) presents an approach for investigating the relationship between the financial news articles and short-term price movement. Similar patterns adopted by Schumaker and Chen (2009), who provided evidence that validate the importance of news on the performance of stock prices. Recently, with the pragmatic innovation of *stock micro-blogging forums* around the world, platforms like StockTwits and TweetTrader have spread widely as an online discussion forum among investors and traders. A study by Sprenger and Welpe (2014) investigate the relationship of market prices of publicly traded companies with the StockTwits sentiment. They show that sentiment of StockTwits (i.e. bullishness) is significantly associated with abnormal stock returns and message volume, while that sentiment has power in predicting the next day trading volume.

3 Methods & Data

3.1 Classifier Algorithms

Three different classifier algorithms are applied to the sentiment detection process in this paper, namely: Naïve Bayes (NB), Decision Tree (DT) and Support Vector Machine (SVM). The following subsection will elaborate in more detail the three models of machine-learning classifiers

- **Naive Bayes Classifier**

A Naive Bayes classifier is a simple classifier technique based on Bayes' Theorem. It is based on the naive assumption, which states that a given attribute is independent of the other attributes contained in a given sample, and it considers each of these attributes discretely when classifying a new incoming instance. The Naive Bayes algorithm is based on the joined probabilities of words or a document belonging to a class in a given text (Witten et al., 1999).

- **Decision Tree Classifier**

The decision tree method is one of the most frequently used techniques for classification problems. It exploits a tree structure consisting of nodes, leaves and branches. Decision trees used for classification problems are often called classification trees where each node represents the predicted class of a given feature. It applies the concept of information gain or entropy reduction, which is based on the selection of a decision node and further splitting the nodes into sub-nodes. The normalised information gain is an impurity-based criterion that uses the entropy measure (Rokach and Maimon, 2005) to evaluate the effectiveness of an attribute for splitting the data. These criteria state that the attribute with the greatest normalised information gain is chosen to make the decision.

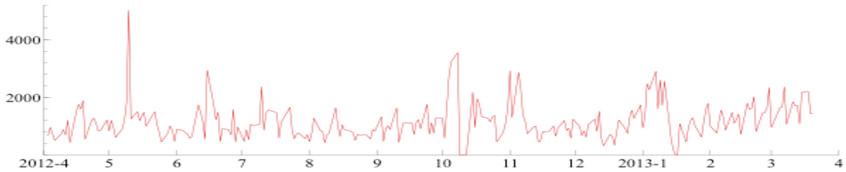
- **Support Vector Machines (SVMs)**.

Support vector machines (SVM) are the most widely used techniques for textual analysis applications; they have proven excellent empirical success with strong theoretical foundations (Tong and Koller, 2002). The primary aim of SVM is to find a maximum hyperplane, which clearly separates the instances and non-instances of a given class relative to the target variables (Barakat and Bradley, 2007). This common approach is generally used when the instances of the target variables are described as linearly separable whereby the target variable should have only two class values. On the other hand, there are some cases where the target variable may have more than two class values where, in this case, described as nonlinearly separable. For a non-linearly separable data, SVM makes use of Kernel methods to transform the data from an input space or parametric space into a high-dimensional feature space. SVM attempts to maximise the margin space between the separating hyperplane and the data instances.

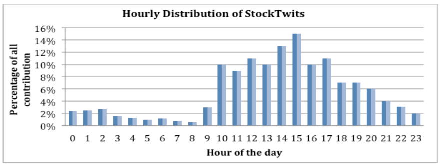
3.2 StockTwits Data

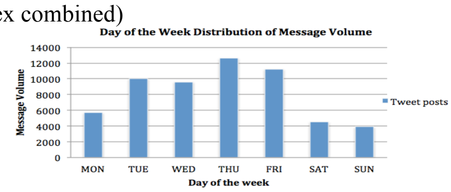
One year of StockTwits data about the companies listed on the DJIA Index are collected for the period April 3rd 2012 – 5thApril 2013.The sample period consists of 252 days only because the U.S. stock market is idle on weekends and national holidays. In line with Antweiler and Frank (2004), messages are aligned with US market hours; messages posted after 4:00 pm (market closing) are combined together with pre-market messages up to 9:30 am (market opening) on the following trading day.

Figure 1 shows the distribution of tweet messages, where panels A, B and C display such message postings respectively over the sample period of one year, over the days of the week, and over the hours of the day. A graphical inspection suggests that the StockTwits postings are reasonably stable over the considered period of study. Nonetheless, some increase in the volume of postings is observed during the early summer, the autumn months (i.e., Halloween), Christmas, and New Year's Eve (see panel A), suggesting that people tend to post more actively during these special occasions. Moreover, consistent with previous studies (e.g., Oh and Sheng, 2001), the volume of tweets posted during working days is high (i.e., reaching a peak on Thursdays), as opposed to the low volume of postings observed during the weekend and on public holidays (see panel B). It is also evident that message postings are concentrated between 10:00am and 5:00pm (see panel C), which suggests the high activity of day traders; hence, more sentiment is developed during the market hours.



Panel A: The evolution of daily StockTwits messages (posting activity of 30 companies of the DJIA index combined)





Panel B: The distribution of StockTwits posts throughout the week (average postings of all companies in the sample are considered across days of the week)

Panel C: The percentage of postings of the 30 companies of the DJIA index during the daytime

Fig. 1**.** The distribution of StockTwits postings

3.3 StockTwits Messages' Manual Labelling

In order to manage the huge amount of StockTwits messages collected for this study, a random selection of a representative sample of 2,892 tweets on all 30 stocks on the Dow Jones Index are hand-labelled as either buy, hold or sell signals based on a redefined dictionary (Harvard-IV-4 classification dictionary). These hand-labelled messages constitute the training set, which is then used as an input for what will be used as training set for different machine learning models. The results of the percentage allocation of the manual classifications of tweet messages into the three distinct classes are shown in Table 1. Table 1 shows that roughly half of these messages were considered to be "buy" signals (47.06%). The remaining messages for "sell" signals were (32.54%) which roughly constitute three quarters of “buy” signals whereas the “hold” signals were (20.40%).

Table 1. Number and percentage distribution oft he manual classification of posting

Table 1: Number and percentage distribution of the manual classification of postings				
Class	**Buy**	**Hold**	**Sell**	**Total**
Number	1,361	590	941	2,892
Percentage	47.06%	20.40%	32.54%	100%

The results of the study indicate that the stock micro-blogging forum seems to be more balanced in terms of the distributions of buy vs. sell messages than internet message boards where the ratio of buy vs. sell signals appears to be unbalanced, ranging from 7:1 (Dewally, 2003) to 5:1 (Antweiler and Frank, 2004). The finding that “hold” messages constitute a relatively small percentage of 20.40%, does not confirm that of Sprenger et al. (2014), who found that almost half of the messages manually classified were considered to be “hold” signals. It follows that the higher distribution of buy and sell messages may provide evidence that there is more relevant financial information present in such forums.

3.4 Automated Classification

To extract the bullishness measure that serves as a proxy for investor sentiment, each message has to be classified into one of the three distinct classes {sell, buy or hold}. Our StockTwits data contain nearly 300,000 text messages - far too many to classify manually. To manage the message classification task, we employ well-recognised methods from computational linguistics. Unlike previous studies (i.e. Antweiler and Frank (2004) and Sprenger et al. (2014)) that use Naïve Bayes Classifiers for classifying messages, our study takes a different approach by comparing the classification performance of three different machine algorithms: (NB), (DT) and (SVM).[1] Training the selected sample of StockTwits messages in Weka using the three machine learning algorithms (NB, DT and SVM) reveals that the (Random Forest) Decision Trees classifier results in a higher accuracy rate compared to the other two classifier algorithms. Table 2 presents a consolidated summary of all performance metrics of the three classifiers. It is evident that there is no clear winning classifier in terms of the performance evaluation method used, yet the Decision Tree classifier is possibly the best classifier in terms of almost all the metrics. The 10-fold cross-validation experiments achieved accuracy rates of 66.70%, 62.80% and 65.20%, where 1,929, 1,815 and 1,887 instances were correctly classified out of 2,892 for RandF, NB and SMO, respectively. It follows that the RandF decision tree classifier outperforms the NB and SMO counterparts in predicting investor sentiment class (i.e., buy, hold and sell) of StockTwits postings. The weighted averages of the three classes of RandF classifier are also reported in Table 2, achieving 65.50%, 66.70% and 66.20% for precision, recall and F-measures, respectively. Figure 2 shows the graph-

[1]In normal settings, machine learning algorithms are designed for the purpose of maximising the classification accuracy and minimising the error rate as far as possible (Kukar and Kononenko, 1998).

ical representation of the comparative performance of the three discussed classifiers using some of the important measures given in Table 2.

Table 2. Summary Resuls of the classification performance evaluation of NB, RandF and SMO

Weighted Average Metrics for (Buy, Hold and Sell) Class	Classifiers		
	Naive Bayes	**Random Forest**	**SVM**
Accuracy Rate	62.80%	**66.70%**	65.25%
Correctly Classified Instances	1,815	**1,929**	1,887
Incorrectly Classified Instances	1,077	**963**	1,055
TP Rate	62.80%	**66.70%**	65.20%
FP Rate	21.80%	**20.80%**	24.60%
Precision	62.90%	**66.50%**	65.90%
Recall	62.80%	**66.70%**	65.20%
F-Measure	62.60%	**66.20%**	64.00%
ROC Area	77.60%	**79.80%**	73.60%

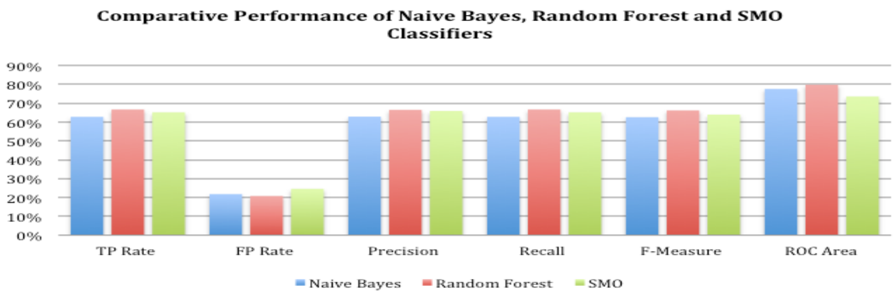


Fig. 2. Comparative Performance of NB, RandF and SMO classifiers

Further, to make sure that our classification accuracy is good enough, we perform an out-of-sample testing. In Weka, training on the first ten months of the year and testing on the remaining two months. Table 3 provides a comparison of the manual classification of hold-out messages and the automated classification of the Random Forest algorithm. The results suggest that the Random Forest algorithm preforms reasonably well, as indicated by the relatively small numbers of misclassifications in each sentiment class. Finally, Table 4 shows the assigned labels for the entire set of StockTwits postings. The reported postings are 140,350, 26,157, and 122,517 for those of 'buy', 'hold' and 'sell', respectively[2].

Table 3. Overall classification distribution of Random Forest (supplied test)

	Classified by Algorithm			
Class	**Buy**	**Hold**	**Sell**	**Manual Classification**
Buy	315	26	158	**499**
Hold	47	48	35	**130**
Sell	94	11	205	**310**
Total Classified by Algorithm	**456**	**85**	**398**	**939**
% Classification by Algorithm as per Class	48.56%	9.05%	42.39%	100%

Table 4. The Overall distribution of the total StockTwits postings of 30 companies of the DJIA index

Table 5: The Overall distribution of the total StockTwits postings of 30 companies of the DJIA index

Class	Manual Classification (in %)	Automatic Classification (in %)	Total Tweets per Class
Buy	47.60	48.56	140,350
Hold	20.40	9.05	26,157
Sell	32.54	42.39	122,517
Total	100%	100%	**289,024**

[2] Following Antweiler and Frank (2004), the 'hold' postings are removed from the analysis as they are considered noise and convey neutral opinions.

3.5 Investor Sentiment Measure

In this research paper bullishness measure is extracted from StockTwits data which then be used as a proxy for investor sentiment. In the stock market, bullishness can be defined as optimism that a particular investment is potentially profitable. The classification algorithm classified all the tweet messages into three distinct classes M^c where $c \in \{Buy, Hold, Sell\}$. The bullishness of messages is an important tweet feature that determines the proportion of buy and sell signals on a particular day t. It is used to aggregate the three different message classes M_t^{Buy} and M_t^{Sell} and M_t^{Hold} in a given time interval. This research study has carried forward the work of Antweiler and Frank, (2004b) by defining bullishness B_t using three different measures as follows:

$$B_t = \left[\frac{M_t^{Buy} - M_t^{Sell}}{M_t^{Buy} + M_t^{Sell}}\right] \quad (1), \quad B_t^{*} = ln\left[\frac{1+M_t^{Buy}}{1+M_t^{Sell}}\right] = \ln\left[\frac{2+M_t(1+B_t)}{2+M_t(1-B_t)}\right] B_t \ln(1+M_t) \quad (1)$$

$$B_t^{**} = \left(M_t^{Buy} - M_t^{Sell}\right) = B_t M_t \quad (3)$$

where M_t indicates the overall number of messages and M_t^{Buy} and M_t^{Sell} indicate the total number of traders' messages conveying buy and sell signals on day *t* respectively. The first bullishness measure is an essential component for obtaining results of the two other measures while these last two measures are more comprehensive measures as both take into account the number of messages M_t as well as the ratio of bullish to bearish messages. The measure B_t^{**} appears to outperform both alternatives; hence, this measure is used to measure bullishness, which is used as a proxy for investor sentiment in this research study. Because a markedly large number of messages are tweeted on a daily basis, normalisation is therefore needed for these messages as this will assist the model's estimation. More specifically, as B_t^{**} may contain negative values and in order to take into account such values, the following formula of normalisation is considered:

$$\overline{B}_{it}^{**} = \frac{(B_{it}^{**} - min\, B_i^{**})}{(min\, B_i^{**} - max\, B_i^{**})} \quad (4)$$

where $\overline{B}_{it}^{**}$ is the normalised value of bullishness B^{**} of company i at time t, and *max* B_i^{**} and *min* B_i^{**} indicate respectively the maximum and minimum value of the bullishness measures of company i over the sample period. This measure represents the number of investors' messages expressing a particular sentiment (buy or sell), giving more weight to a larger number of messages in a specific sentiment.

3.6 Stock Return Data

The financial data are obtained from Bloomberg for the actively traded blue chip stocks of the 30 companies making up the DJIA index for the period between April 3rd 2012 and April 5th 2013. No extraordinary market conditions were reported during this period, so it represents a good base test for the evaluation. The return series of the DJIA index stocks are computed by taking the first differences of the logarithm of the daily closing prices, multiplied by 100. There are several reasons why DJIA index is being focused to adequately reflect the US stock market. One of theses reason is that DJIA is a price-weighted average of 30 largest market capitalisations of the industrial companies in the US equity market traded on the NYSE and the NASDAQ. For more justified reasons see Al-Nasseri et al. (2016).

4 Market Regimes

Following Chen (2007) and Kurov (2010), we estimate the regime switching probabilities using a simple Markov-switching model of stock returns to identifies the periods of the two different market regimes namely; bull and bear market as follow;

$$R_t = \mu_{S_t} + \varepsilon_t, \qquad \varepsilon_t \sim i.i.d.N\left(0, {\sigma^2}_{S_t}\right) \tag{5}$$

where R_t is the daily returns on the DJIA index and S_t is an unobserved dummy variable that indicates the two different states of market regimes; bull or bear market. Therefore, $\mu_{S_t}$and ${\sigma^2}_{S_i}$ are the state- dependent mean and variance of returns, respectively. The model in Eq. (5) is used to statistically identify two regimes classifications based on smoothed probability. These two regimes are; regime 0 with a lower variance of returns and higher returns so called (bull market) and regime 1 with a higher variance and lower returns (bear market) as indicated in Figure 3. The mean variances of the model are estimated jointly with maximum likelihood. Once the model is estimated, regime classification based on smoothed probabilities of bull and bear market at different point in time are computed to identifies a periods of each regimes states separately.

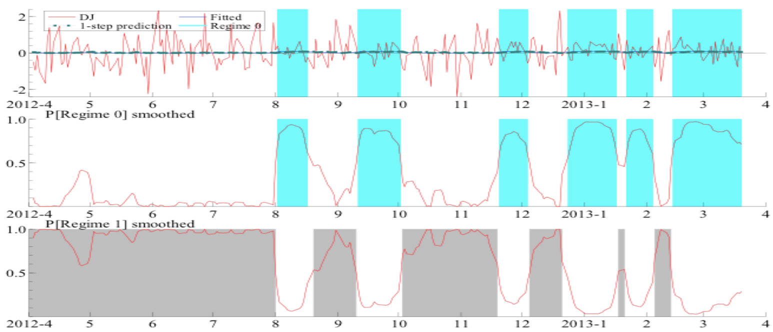


Fig. 3. Regime classification based on smoothed probability

We create two indicator variables labeled I^{bull} and $I^{bear}$for the bull and bear market respectively as shown in Eqs. (6a) and (6b):

$$I_{it}^{bull} = \begin{cases} 1 & if\ r_{it} > 0 \\ 0 & otherwise \end{cases} \quad (6a) \qquad I_{it}^{bear} = \begin{cases} 1 & if\ r_{it} \leq 0 \\ 0 & otherwise \end{cases} \quad (6b)$$

5 Empirical Results

In order to empirically investigate the asymmetrical response of investor sentiments to returns in different states of the market, the bullishness equation is estimated by including two market regimes (bull and bear market) to show whether the returns help to explain investor sentiment. Both interaction terms as defined previously in Eqs. (19a and 19b) are added in the bullishness equations as follows:

$$\overline{B}_{it}^{**} = \alpha_1 + \phi\overline{B}_{it-1}^{**} + \gamma_1 I_{it}^{bull}\, r_{it} + \gamma_2 I_{it}^{bear}\, r_{it} + \beta_1 NWK_t + \beta_2 MKT_t + \varepsilon_{it} \quad (7)$$

The results presented in Table 5 indicate that the model specification in Eq. (7) suffers from serial autocorrelation in the data series. Therefore, the model is modified to include lagged bullishness to assist in removing the serial correlations (Dickey and Fuller, 1979). The panel regression with company fixed effects is used where the market index and first-day-of-the week dummy were added to the regression to control for the market-wide effect and the negative return on the first trading day of the week, respectively. The results reported in Table 5 show that bullishness tends to respond to stock returns positively in the bull market and negatively in the bear market. The significant positive coefficient of γ_1= +0.0157 indicates that positive returns trigger an increase in investor bullishness in the bull market by 1.57 %, while the negative coefficient of γ_2= -0.0122 implies therefore that a negative return triggers a reduction in investor bullishness by 1.22% in the bear market. A possible explanation for this is that, when the stock return r_t is positive (negative), investor bullishness exhibits a pronounced increase (decrease), which implies that in the bull market an investor becomes more bullish whereas in the bear market investor is likely become more bearish. These findings are in line with the subjective evidence: When the market is on a bull run as it was in the late 1990s, investors appear to become more bullish. This finding is consistent with (De Bondt, 1993) who found that increased bullishness could be expected after a market rise and increased bearishness after a market fall. This evidence is also in line with the existence of bandwagon effect (Brown and Cliff, 2004), which states that good returns in a given period drive optimism and they found that stock returns predict sentiments. The magnitude effects of the impact of returns on bullishness appear to be greater in the bull market compared to bear market. This finding is in line with Verma and Verma (2007) who found a stronger effect on bullish sentiments during the period of positive return (growth) than the effects on bearishness during the period of negative return (decline).

Table 5. The asymmetric response of the investor sentiment to the change in stock returns in the bull and bear markets.

$\overline{B}_{it}^{**}$	Coefficient	S.E
α_1	0.1885***	(0.0038)
ϕ	0.3592***	(0.0107)
γ_1	0.0157***	(0.0019)
γ_2	-0.0122***	(0.0019)
β_1	-0.0164***	(0.0027)
β_2	-0.0037*	(0.0020)
R^2	0.6290	
N-observations	7,530	
Durbin-Watson statistics	2.0845	

Note (*), (**), and (***) denote significance levels at 10%, 5%, and 1%, respectively. Standard errors are shown in parentheses

6 Conclusion

In this paper we proposed a novel approach by combining various text-mining techniques and financial econometric modelling to investigate the asymmetric behaviour of investor on news sentiments in two different state of the market. This paper seeks to answer three interrelated questions. Can text-mining techniques accurately predict sentiment on StockTwits? Do investor reactions to good and bad news are asymmetric? Do investor react more harshly when bad news is disseminated? Our findings provide significant evidence of the effectiveness of different classifiers algorithms in predicting online investor sentiments in financial market. Despite the well performance of all of our three classifiers algorithms (NB, RandF and SMO), the Random Forest classifier however achieved the best results and proves capable in predicting sentiments of online financial text. Our findings show that investor behave asymmetrically to the good and bad news in the bull and bear market. We find that investor sentiment show positive (negative) impact to news in the bull and bear market respectively. Furthermore, our result indicates that investors are more sensitive to positive news and react much more positively in the bull market than otherwise do in the bear market. Overall our results are consistent with the conjecture of Kurov (2010) that investor shows asymmetric response to information news in the market and that effect are very dependent on the market regime since there is a greater positive impact on bullishness during the period of growth in the bull market than the negative impact on bullishness during the period of decline in the bear market.

References

1. Al-Nasseri, A., Menla Ali, F. and Tucker, A. (2016) Dispersion of stock returns and investor sentiment: StockTwits evidence, Unpublished Paper.
2. Antweiler, W. and Frank, M.Z. 2004b. Is all that talk just noise? The information content of Internet stock message boards. *The Journal of Finance,* **59**(3), pp. 1259-1294.
3. Baker, M. and Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, **21**(2): 129–151.
4. Barakat, N.H. and Bradley, A.P. 2007. Rule extraction from support vector machines: a sequential covering approach. *Knowledge and Data Engineering, IEEE Transactions on,* **19**(6), pp. 729-741.
5. Bennett, K.P. and Mangasarian, O.L., 1992. Robust linear programming discrimination of two linearly inseparable sets. *Optimization methods and software*, *1*(1), pp.23-34.
6. Dickey, D.A. and Fuller, W.A. 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association,* **74**(366a), pp. 427-431.
7. Bowman, R. G. (1983). Understanding and Conducting Event Studies. *Journal of Business Finance and Accounting*, **10** (4), pp. 561–84.
8. Brown, G.W. and Cliff, M.T. 2004. Investor sentiment and the near-term stock market. *Journal of Empirical Finance,* **11**(1), pp. 1-27.
9. Chen, S.S., 2007. Does monetary policy have asymmetric effects on stock returns? *Journal of Money, Credit and Banking* **39**, 667–688.

10. Cortes, C. and Vapnik, V. 1995. Support vector machine. *Machine Learning,* **20**(3), pp. 273-297.
11. De Bondt, W.P. 1993. Betting on trends: Intuitive forecasts of financial risk and return. *International Journal of Forecasting,* **9**(3), pp. 355-371.
12. Delort, J., Arunasalam, B., Leung, H. and Milosavljevic, M. 2012. The impact of manipulation in Internet stock message boards. *International Journal of Banking and Finance,* **8**(4), pp. 1.
13. Dewally, M. 2003. Internet investment advice: Investing with a rock of salt. *Financial Analysts Journal*, *59*(4), pp.65-77.
14. Fama, E.F. 1965. The behavior of stock-market prices. *The journal of Business,* **38**(1), pp. 34-105.
15. Gidofalvi, G. and Elkan, C. 2001. Using news articles to predict stock price movements. *Department of Computer Science and Engineering, University of California, San Diego*.
16. Kukar, M. and Kononenko, I. 1998. Cost-Sensitive Learning with Neural Networks. *ECAI* 1998, pp. 445-449.
17. Kurov, A. 2010. Investor sentiment and the stock market's reaction to monetary policy. *Journal of Banking and Finance,* **34**(1), pp. 139-149.
18. Oh, C. and Sheng, O. (2011), "Investigating Predictive Power of Stock Micro Blog Sentiment in Forecasting Future Stock Price Directional Movement", in 32nd *International Conference on Information Systems*, AIS, p.17.
19. Pritamani, M. and V. Singal (2001). Return Predictability Following Large Price Changes and Information Releases. *Journal of Banking and Finance*, **25**, pp. 631–51.
20. Rokach, L. and Maimon, O. 2005. Decision trees. *Data Mining and Knowledge Discovery Handbook.* Springer, pp. 165-192.
21. Schumaker, R.P. and Chen, H. 2009. Textual analysis of stock market prediction using breaking financial news: The AZFin text system. *ACM Transactions on Information Systems (TOIS),* **27**(2), pp. 12.
22. Soroka, S. N. (2006). Good News and Bad News: Asymmetric Responses to Economic Information. *Journal of Politics*, **68**, pp. 372–85.
23. Sprenger, T.O., Tumasjan, A., Sandner, P.G. and Welpe, I.M. 2014. Tweets and trades: The information content of stock microblogs. *European Financial Management,* **20**(5), pp. 926-957.
24. Tetlock, P. C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance*, 62, pp. 1,139–1,168
25. Tong, S. and Koller, D. 2002. Support vector machine active learning with applications to text classification. *The Journal of Machine Learning Research,* **2**, pp. 45-66.
26. Tumarkin, R. and Whitelaw, R.F., 2001. News or noise? Internet postings and stock prices. *Financial Analysts Journal*, pp. 41-51.
27. Witten, I.H., Frank, E., Trigg, L.E., Hall, M.A., Holmes, G. and Cunningham, S.J., 1999. Weka: Practical machine learning tools and techniques with Java implementations. Morgan Kaufmann, San Francisco.
28. Wysocki, P. 1998. Cheap talk on the web: The determinants of postings on stock message boards. *University of Michigan Business School Working Paper.*
29. Verma, R. and Verma, P. 2007. Noise trading and stock market volatility. *Journal of Multinational Financial Management,* **17**(3), pp. 231-243.