Mémoire de Fin d'Études

Sentiment Analysis of Twitter Data

and the

Efficient Market Hypothesis

Présenté et soutenu par: M^{r} Alexandre Bénétreau

Numéros étudiant: Panthéon-Sorbonne n°: 11340541, Paris Descartes n°: 21103850

Date de soutenance: 18 juin 2015

Directeur de soutenance: Prof. Angelo Secchi

Membres du jury: M^{r} Michał Krawczyk et M^{r} Julien Nelson

Master: M2R Économie et Psychologie

« L'Université de Paris 1 Panthéon Sorbonne n'entend donner aucune approbation ni désapprobation aux opinions émises dans ce mémoire ; elles doivent être considérées comme propres à leur auteur. »

Acknowledgements

I would like to address my particular thanks to Prof. Angelo Secchi for his unwavering support in statistical and coding matters. The WeFeel project also deserves a special mention for graciously accepting to share its data. Finally, I would also like to extend my thanks to the people on the web forums tex.stackexchange and stackoverflow where I could always turn when in doubt regarding LaTeXformatting and statistical computing with gretl.

Abstract

This thesis discusses the claim of 16 computational finance articles according to which it is possible to predict the stock market using sentiment analysis of social media data. The purpose of this paper is to investigate whether this is indeed true or not. In economic theory, the efficient market hypothesis states that markets are not predictable, that they follow a random walk and that irrational behaviour cancels out in the aggregate. However, behavioural economics research shows that investors are in fact subject to predictable biases which affect the markets. This study uses data from the WeFeel project that analyses tweets in English to infer social mood on a world scale. It also uses data from the Wilshire 5000 index from June 2014 to March 2015. The hypothesis is that changes in aggregate mood arousal mediate stock market fluctuations. Yet linear regression shows that there is no relation between emotional arousal and the stock market, nor between primary emotions and the stock market. Hence, the conclusion is that global social sentiment as derived from social media has no relation with stock market fluctuations. Further research may better focus on social media specialised in the stock markets, such a finance micro-blogging data.

Keywords: sentiment analysis, efficient market hypothesis, social networks, computational finance, behavioural finance, stock market, emotion recognition, stock market prediction, social sentiment, behavioural economics

Contents

Introduction and Theoretical Background

The desire to know the future is old. In ancient civilisations and tribes, it was one of the functions of priests, oracles and Sybyls to attempt to see beyond the present. Today, science seems to have taken over this function. By the power of statistics, reliable probabilities of future outcomes can be constructed.

Four years ago, a study attracted a lot of attention, claiming that "collective mood states derived from large-scale Twitter feeds are correlated to the value of the Dow Jones Industrial Average (DJIA) over time. [...] We find an accuracy of 87.6% in predicting the daily up and down changes in the closing values of the DJIA ..." (Bollen, Mao, & Zeng, 2011). This result was featured in articles by CNN, Wired, and Time, amongst others (Quest, 2010; McMillan, 2011; Grossman, 2010). In the following years, a flurry of studies appeared which to our knowledge all contended that there is a significant relationship between the results of textual analysis of news or microblogging with various financial indices or stocks (Rao & Srivastava, 2012, 2014; Zhang, Fuehres, & Gloor, 2012; Oliveira, Cortez, & Areal, 2013; Dey, 2014; Kaminski & Gloor, 2014; Levenberg, Pulman, Moilanen, Simpson, & Roberts, 2014; Li et al., 2014; Ohmura, Kakusho, & Okadome, 2014; Shirawadekar, 2014; Si et al., 2014; Siganos, Vagenas-Nanos, & Verwijmeren, 2014; Sprenger, Tumasjan, Sandner, & Welpe, 2014; Sul, Dennis, & Yuan, 2014; Khadjeh Nassirtoussi, Aghabozorgi, Ying Wah, & Ngo, 2015). How do these studies come to these conclusions? Firstly, they gather data from microblogging or news websites. Secondly, they use computerised methods of text analysis to infer the emotions and moods of the writers. This is called sentiment analysis. Thirdly, they compare the social moods to the market movements and even go as far as to attempt to predict the markets.

Psychological Background

Before performing sentiment analysis, researchers have to decide on a framework in order to classify emotions. Numerous models of emotions have been developed over the years, dating as far back as ancient Greek philosophers. There are three main modern models: the basic-emotions, the multicomponential, and the hierarchical model (Frijda, 2008). The basic-emotions model posits that there are basic components of emotions which have neural correlates. Specific brain regions and/or circuits would be activated for each different emotion. Yet the correlations between the neural activation and the felt emotions range from weak to moderate, thus only partially validating this approach.

The multicomponential model views emotions as bundles of component processes. At the extreme, a single component could correspond to an emotion. This leads to the abandonment of a strict classification of emotions, as each emotion would be unique and only loosely linked to other emotions. Language would only tentatively describe the various emotions and their links. One could label this model as bottom-up in that it starts by analysing the components as well as their possible links and only subsequently creates categories.

The hierarchical model describes components as differing in their organizational power. For instance, Parrott's hierarchy of emotions differentiates between primary, secondary, and tertiary emotions (2001). According to him, there are six primary emotions: love, joy, surprise, anger, sadness, and fear. Each of these contains a set of secondary emotions; the primary emotion of joy contains secondary emotions such as affection, cheerfulness, pride, etc. Following the same principle, the secondary emotions contain a set of tertiary emotions. For instance, the secondary emotion "cheerfulness" includes bliss, satisfaction, and amusement (see appendix A for the complete chart). Contrary to the previous model, this one is top-down, as the categories are determined from the start.

The hierarchical model was found to outperform the competing multicomponential approach in emotion classification in tweets (Esmin, de Oliveira, & Matwin, 2012). Furthermore, Parrott's hierarchy of emotions takes into account the pleasure-dominance-arousal theory of emotion. This theory states that emotions can be further characterised along three different bipolar axes: pleasure-displeasure, dominance-submissiveness, and arousal-nonarousal. For instance, emotions such as "joy" and "love" would rank high on the pleasure scale; "boldness" and "power" would rank high in dominance while "fear" and "loneliness" would rank low on the same scale; "boredom" and "concentration" would be opposites along the arousal-nonarousal axis (Mehrabian, 1996). This addition allows for a more fine-grained sentiment analysis of textual data.

Economic Background

Traditionally, economists have considered that the financial markets follow a *random walk* (Sewell, 2011). The expression means that for an external observer, the financial markets are unpredictable, they move in indiscernible ways. This is not to say that there is no reason for these movements, but in theory, the variations of the market are not predictable.

Therefore, neither technical nor fundamental analysis (two investment strategies) should allow for correct predictions of the financial markets. The question then is how investors such as Mr Buffet manage to make profits if no patterns can be discovered in the financial markets. According to Keynes (1923), this would be simply due to risk-seeking investing behaviours that turn out well .

It may appear counter-intuitive that markets cannot be forecast since economics views the individual as rational, therefore predictable. What is in fact considered unpredictable are the new informations which determine the rational agent's investment. From there, it is only a short step to the *efficient market hypothesis* (EMH), which firstly states that investors are rational. Secondly, whenever they do not behave rationally, they act randomly. This random behaviour cancels out in the aggregate; prices are thus not influenced. Thirdly, in the case that investors are consistently irrational over time, there are rational economic agents (arbitragers) who counter their influence on the prices (Yen & Lee, 2008).

Apart from the rationality of economic agents, the EMH also assumes that information is accessible and taken into account by the investors. Of course, perfect information for all financial investors at the same time does not seem plausible. Even the most zealous investors need sleep, which means that their colleagues who are awake may have a few hours of advance in terms of new information. But there are several forms of the EMH which reflect that information may not be available instantaneously to all: In the weak form, market efficiency is defined as past information that is fully and swiftly reflected in the stock price; semi-strong form market efficiency means that public information is be fully and quickly reflected in its stock price; and in the strong form, all information is fully and quickly reflected in the stock price (Fama, 1970).

Due to the random walk and the EMH, the field of economics has regarded the impact of investors' emotions on the stock prices to be negligible. Indeed, according to the random walk, emotional states of economic agents do not make the financial markets predictable. Moreover, the EMH states that any irrational behaviour caused by emotional influences is brought into balance by arbitragers.

However, the rise of behavioural economics led to strong criticism of both the random walk and the EMH. It is now recognised in economics that investors are subject to biases. Three main groups of biases can be marked out: heuristics, emotions and framing (Baker & Nofsinger, 2010).

Regarding emotional biases, research has shown effects of daily sunshine on market returns. Seasonal variations in the duration of the day, daylight saving changes, and special events such as sporting contests as well as religious holidays appear to impact returns (Shumway, 2010). Although these studies are in accordance with the psychological literature, they contradict the efficient market hypothesis. The EMH can however not be rejected based on these behavioural results since the effects of mood states on investors appear to be modest. More specifically, overconfidence is a sentiment which has been linked to higher trading volume, higher prices, and bubbles. This sentiment is fostered by positive affects such as love and joy (Ifcher & Zarghamee, 2014). Furthermore, high arousal messages/stimuli and agreeable messages/stimuli are better recalled and are allocated more attentional resources (Vogt, De Houwer, Koster, Van Damme, & Crombez, 2008; Lang, Dhillon, & Dong, 1995). Therefore, pleasurable, high arousal social moods could very well impact the markets in a predictable fashion.

Caveats

Before the hypothetical effects of population sentiment on the markets be investigated, a few considerations are necessary. Oddly enough, it seems to have escaped all previous studies that there may be a substantial hurdle to take into account: algorithmic trading. Indeed, what if the trades were mostly performed by algorithms? Would that not mean that the influence of social moods on the markets is irrelevant since algorithms are not capable of human emotion? According to a British governmental report, 30% of the UK's equity trading volume and over 60% for the USA's may now be generated through high frequency trading (HFT) (Beddington, 2012). These figures are corroborated by Foucault (2012), who indicates in a review of the literature that algorithmic trading accounts for at least 30% of the trading volumes, while in some markets the figure could be as high as 80%. However, to conclude that these algorithms are insensitive to social mood is fallacious. Algorithms can be enhanced to include semantic complex event processing (SCEP), such that the web, social media and media outlets are monitored so as to react as fast as possible to the news or mood swings of the population (Teymourian, 2014; G. Mitra & Mitra, 2011). For instance, a fake Associated Press tweet claimed on April 23rd 2013 at 1:07 pm that there had been an explosion at the White House and that Obama had been injured. Within three minutes, the DJIA (Dow Jones Industrial Average) had plunged by more than a 100 points, which is not explainable without algorithmic trading (Matthews, 2013). Therefore, although algorithms cannot feel emotion, they do react to tweets. What is more, hedge funds such as Derwent Capital Markets and Cayman Atlantic explicitly include sentiment analysis in their algorithms.

Of course, if trade orders are now initiated more often by algorithms than by humans, it is sensible to ask whether social media and press activity are not independently operated by algorithms too. There are clear steps in that direction; for instance, some Twitter accounts are operated by bots and recently a news article was written by a program ("Robot writes LA Times earthquake breaking news article", 2014). However, we can safely assume that the overwhelming majority of tweets and journalistic texts are from human origin (Linshi, 2014). That does not mean that social media production is an exact reflection of peoples' minds; a happy tweet does not necessarily equal a happy person. Textual analysis is also limited when it comes to differentiating subtle content such as irony from literal statements. In spite of these limitations, we can assume that social media analysis delivers a reliable general tendency of the social mood.

Hypotheses and Overview of the Study

The major difficulty of studies like this one is knowing whether variations in the social mood cause fluctuations of the market, or whether fluctuations in the market or some related variable cause variations in the social mood. For instance, showing a correlation between positive social moods and market signals such as increases in trading volume and stock prices does not mean that sentiment or mood influences the markets. A good research should show that social mood either clearly causes the markets to fluctuate or that it does not.

The social media data was generously shared by the WeFeel project. It explores whether Twitter can provide an accurate, real-time signal of the world's emotional state. The endeavour is a joint venture by computer scientists at CSIRO and mental health researchers at The Black Dog Institute. The tweets collected are all in English and are analysed according to Parrott's hierarchy of emotions, allowing for a measurement of daily mood shifts in the English-speaking Twitter community.

Regarding the choice of the stock market to analyse, this must depend on the countries where English-speaking Twitter users come from. As of October 2013, roughly 25% of active Twitter users were from the US, the next biggest English-speaking country being the UK with 5% (Schoonderwoerd, 2013). These numbers suggest that the majority of tweets in English come from the US. English-speaking sentiment is therefore expected to impact more the American market, as captured by the Wilshire 5000 index. Both datasets cover the period from June 2014 to March 2015, nine months in total. The econometric analyses mainly consist in OLS regressions and stationarity testing with the goal of running a Granger causality analysis. It is expected that the analyses will yield significant results indicating a small to intermediate effect of population sentiment on the markets.

Based on the literature, we can state the following hypothesis:
H: Changes in aggregate mood arousal mediate stock market fluctuations.

Method

Data

In contrast to many other studies which favour the DJIA (Dow Jones Industrial Average) or the Standard & Poor's 500 as a market index, we opt for the Wilshire 5000 Full Cap Price Index. Indeed, this index has the advantage of providing a larger aggregate of the US stock market and is float adjusted to include shares of stock not considered available to "ordinary" investors. The index is not seasonally adjusted and provides us with the daily market close.

The tweets of the WeFeel Project come from a random 10% sample from Gnip (a social media aggregation company), and another source that specifically monitors the public Twitter API (application programming interface) for a large vocabulary of emotion terms. The emotions are identified by means of a database of emotion terms that was compiled from multiple sources, including the ANEW and LIWC corpora, and a list of moods from LiveJournal. A crowdsourcing task was used in order to classify these words according to Parrott's hierarchy of emotions. In this paper, we collected semantically pre-processed data from 8 873 773 550 tweets in English published over the course of nine months. The emotional categorisation of the tweets leads to table 1.

Not all tweets can be categorised, some contain words such as "indifferent" or "serious" whose position in the Parrot hierarchy is unclear (Other). Furthermore, some tweets do not contain emotional words (No emotion).

The markets file displays instances of missing data due to the stock markets closing during the weekends and public holidays (91 cases). Furthermore, several cases of

Table 1

Total number of tweets per primary emotion

Surprise	501 721 446
Joy	1 762 160 521
Love	823 829 871
Sadness	1 732 492 465
Anger	533 844 790
Fear	312 558 854
Other	1 878 141 034
No emotion	1 329 024 569

ineffective tweet collection were detected and removed from the dataset (8 cases).

Arousal

We can approximate arousal in different ways. The first and easiest one would be to just look at the total number of tweets per day, irrespective of the emotional components:

$$Arousal = t \tag{1}$$

Conversely, we focus on emotions in the second method. Primary emotions' arousal was assessed by the WeFeel Project with a scale ranging from 1 (not arousing) to 9 (very arousing): Surprise (6,57), Joy (5,55), Love (5,36), Sadness (2,81), Anger (5,93), and Fear (6,14). We use this formula to estimate the total level of arousal for each day:

$$Arousal = \frac{\sum_{i=1}^{n} e_i * a_i}{10000} \tag{2}$$

where e stands for the daily number of tweets per emotion and a for the arousal weight corresponding to each emotion. The sum of these products is divided by 10000 for readability's sake.

Analyses and Results

In model 1, we linearly regress the market values against population arousal (computed using the second equation). As can be seen from table 2, the OLS coefficient of total arousal is -0.14, which means that a one percent increase in total arousal is associated with a decrease of the dependent variable by 14 percent. The p-value of the OLS coefficient and of the F-test are both significant at the 0.001 threshold, which means that we can reject the null hypotheses and assume that the regressor (arousal) has indeed a non-zero coefficient. The R^2 shows that 29% of the variation in the market (Wilshire 5000) is determined by the linear relationship between the market and arousal. This means that we obtain a negative correlation of -0.53 between the market and arousal. The various criteria for model selection indicate that the model does fit the data well. Indeed, log-likelihood is positive, while the Schwartz, Akaike, and Hannan-Quinn criteria are negative. Testing for heteroscedasticity with White's test reveals that we cannot reject the assumption of homoscedasticity. However, the χ^2 test also reveals that the residuals do not follow a normal distribution. The output of the Durbin-Watson test and the first-order autocorrelation coefficient $\hat{\rho}$ denote a strong positive autocorrelation, meaning that the residual error terms are, on average, close in value to one another. This could mean that the model is functionally misspecified or that an important variable has been omitted.

We detect an outlier for the data of the 12^{th} of August 2014, as can be seen on the bottom right of figure 1. This is presumably caused by the announcement of the death of the actor Robin Williams, as indeed the sadness levels reach their all time peak on that day (more than 13 Mio tweets), while no major catastrophe occurred worldwide. When using an outlier-robust linear method (least absolute deviation), the arousal coefficient moves merely from -0.14 to -0.17, which indicates that we can keep the outlier in our dataset.

Table 2
Model 1, (T = 205)

Dependent variable: logWILSHIRE5000

	Coefficient	Std. Error	t-ratio	p-value
const	11,3918	0,157041	72,5403	0,0000
logAROUSAL	−0,146571	0,0160856	−9,1119	0,0000

Mean dependent var	9.960921	S.D. dependent var	0.028117
Sum squared resid	0.114460	S.E. of regression	0.023745
R^2	0.290276	Adjusted R^2	0.286779
$F(1, 203)$	83.02654	P-value(F)	7.89e–17
Log-likelihood	476.8983	Akaike criterion	−949.7965
Schwarz criterion	−943.1505	Hannan–Quinn	−947.1084
$\hat{\rho}$	0.829063	Durbin–Watson	0.337207

Test for normality of residual –
Null hypothesis: error is normally distributed
Test statistic: $\chi^2(2) = 25.4273$
with p-value = 3.00982e-06

White's test for heteroskedasticity –
Null hypothesis: heteroskedasticity not present
Test statistic: LM = 0.683387
with p-value = $P(\chi^2(2) > 0.683387) = 0.710566$

In order to verify the statistical soundness of the OLS, we need to test whether the regression is spurious or not. A spurious correlation is characterised by two properties: all the variables in the regression are non-stationary and not cointegrated. Before we examine the issue of cointegration, we must ascertain that the variables are stationary. We test the variables outside of the regression for stationarity, always using a six lag order: The arousal variable with constant and trend appears to be stationary at the 0.01 significance level (p-value: 0.001498), while we cannot reject the H0 of non-stationarity for the market variable with constant and trend (p-value: 0.1357).

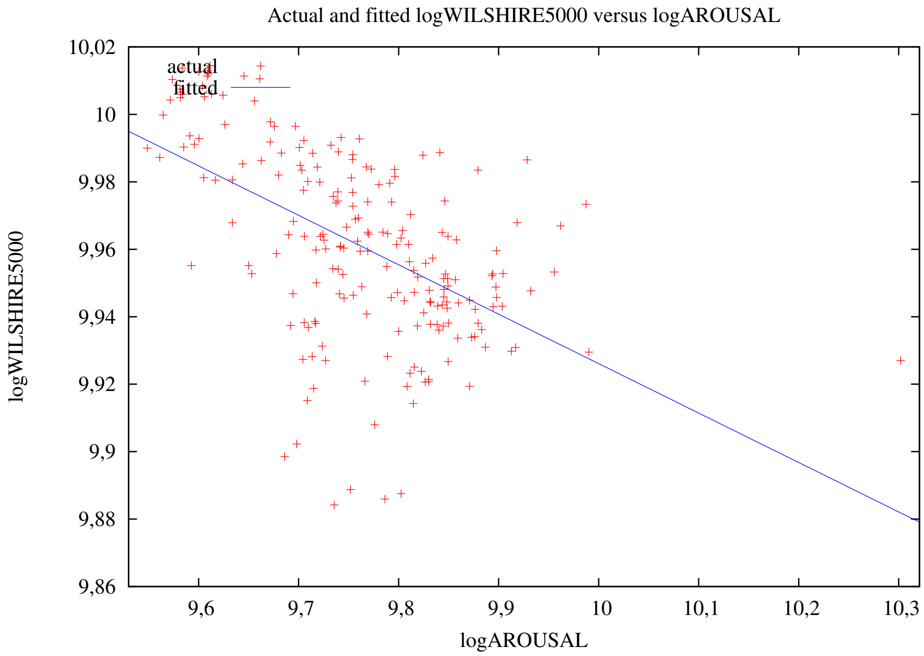

Figure 1. Scatterplot with fitted regression line of the log of Wishire 5000 against the log of Arousal.

The KPSS test is run separately on the two variables of interest and confirms these results. The null hypothesis is reversed compared to the ADF test, such that significant results lead us to reject the hypothesis of stationarity. With trend and constant, we can assume stationarity for the arousal variable (test statistic: 0.0970596). However, we reject the stationarity hypothesis at the five percent threshold for the Wilshire 5000 at lag order six with constant and trend (test statistic: 0.166264). Therefore, we consider the arousal variable stationary and the market variable non-stationary.

We run a Dickey-Fuller test on the residuals of the regression. Indeed, cointegration is defined as the occurrence of stationarity in the error term of the regression. The residuals are stationary at the 0.01 significance level without constant (p-value: 0.009575). However, checking this result with the KPSS and the AFD-GLS tests contradicts the preceding test: The KPSS without trend leads to rejection of the stationarity hypothesis at threshold level 0.01 (test statistic: 0.802903); the ADF-GLS with constant and no trend cannot reject the non-stationarity hypothesis (p-value: 0.1317). We conclude that the residuals of the OLS are non-stationary, which indicates that the variables are not cointegrated. Hence, we cannot exclude a spurious correlation with this model.

Table 3
Model 2, (T = 204)

Dependent variable: d_logWILSHIRE5000

	Coefficient	Std. Error	t-ratio	p-value
const	0.0108791	0.0520040	0.2092	0.8345
logAROUSAL	−0.00107676	0.00532694	−0.2021	0.8400

Mean dependent var	0.000368	S.D. dependent var	0.007837
Sum squared resid	0.012464	S.E. of regression	0.007855
R^2	0.000202	Adjusted R^2	-0.004747
$F(1, 202)$	0.040858	P-value(F)	0.840015
Log-likelihood	700.2435	Akaike criterion	−1396.487
Schwarz criterion	−1389.851	Hannan–Quinn	−1393.802
$\hat{\rho}$	−0.023407	Durbin–Watson	2.041331

Due to the non-stationarity of the market variable, we are forced to work with differentiation. However, a new OLS shows that there is virtually no relationship between arousal (equation two) and the market returns, which are the first difference of the market values. Indeed, the p-values are non-significant, the regression coefficients very close to zero. And as can be seen in figure 2, the regression line is nearly flat. No relationship is found either when both variables are differenced. Taking the first equation for arousal, we are confronted with the same outcomes, namely an insignificant OLS, unless when the market variable remains undifferentiated (see appendix B).

This could mean that emotional arousal does not matter, but only the type of emotion and its associated valence does. To test that assumption, we run various OLS

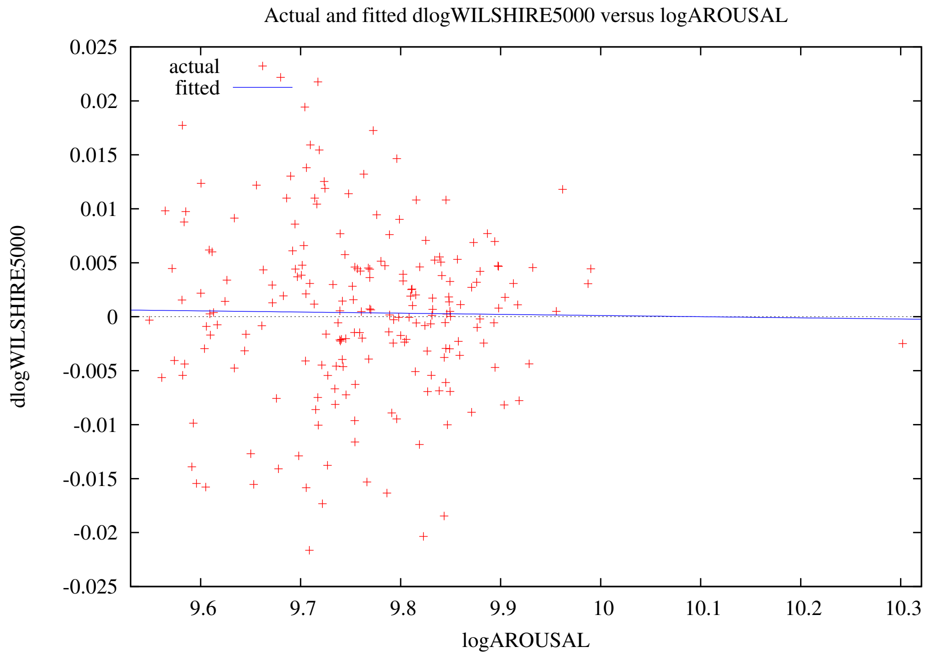

Figure 2. Scatterplot of the first order difference of the log of Wishire 5000 against the log of Arousal.

models that include the daily number of tweets for each primary emotion (surprise, joy, love, sadness, anger, and fear). This results in non-significant output with correlation zero (see appendix C). Trying with lagged regressor variables does not change anything (see appendix D). The lags allow to see if an independent variable can affect the dependent variable over the course of a few days instead of immediately.

Hence, we must reject our main hypothesis and state that according to our data, emotional arousal does not influence the stock markets, and neither do primary emotions.

Discussion

Contrary to our expectations, we do not find any trustworthy correlation between emotional arousal and the market returns, nor between primary emotions and market returns. Our results stand in stark contrast to Bollen et al.'s (2011). A possible explanation for this divergence may lie in the statistical methodology of their study. For instance, they find significant correlations at the 0.10 and 0.05 threshold. Needless to say, this is not exceptional. But the real problem here is that they test 49 multiple hypotheses in the process and therefore run into the multiple comparisons problem. This should be accounted for by performing a Bonferroni correction for multiple tests. Once this is done, one realises that *none* of the correlations are significant at the 0.10 threshold (Lawly Wurm, 2014). Hence, the subsequent Granger analysis does not make any sense, because there cannot be causation without correlation (as long as there is no selection bias).

However, we cannot dismiss all other studies similarly. The main difference between these studies and our paper is that they analyse tweets which contain words related to the stock market or a company, such as "dollar", "money", "Microsoft", etc. For instance, Zhang et al. (2012) show that market time series correlate with time series of market-related tweets. Rao and Srivastava (2012, 2014) use tweets concerning specific companies; Sprenger et al. (2014) use targeted data from investment microblogs; and Sul et al. (2014) collected tweets regarding specific corporations. Our purpose was more general, as we collected data from tweets that were not necessarily related to the market, just as Bollen et al. did.

From a theoretical perspective, this means that the EMH (efficient market hypothesis) retains validity as long as general, market-unspecific emotional data is used to predict the markets. We believe however that the EMH may lose its validity when using media content and messages on the topic of the markets. It is conceivable that a real-time technology of media collection and analysis predicts the markets accurately, as the aforementioned papers suggest. In practice, this would mean that an entity with that technology would have access to a far greater amount of information than other investors. Online messages do not only have a potential emotional content, but an informational one. The ability to capture data and information immediately and to make sense of it would certainly allow for at least short-term predictions.

If there is no correlation between population sentiment and the stock markets, then this means that they should not be able to influence each other. However, this is not correct, as times of economic crisis such as the stock market crash of 2007 deeply affected the economy, causing unemployment and a measurable decrease in happiness (Graham, Chattopadhyay, & Picon, 2010). Therefore, whether the economy is suffering a crisis or thriving may be a factor to take into account. This is a weakness that could be addressed by time series over several years. In practice, this is easily achievable for financial data, but hardly feasible for microblogging data. One does not simply download eight billion tweets and conduct textual analysis on them with an IPad. As long as Twitter does not donate part of its treasure trove of tweets to the Library of Congress, access over several years of data will remain a researcher's dream. That being said, one can still rule out the small sample bias with data covering nine months.

Another possibility can be raised to explain the results, namely that the primary emotions are too unspecific. Running regressions with secondary or even tertiary emotions such as hope, confidence, anxiety, and greed may be more conclusive. Indeed, greed and fear are commonly thought to be the decisive emotions in the markets.

Conclusion

Even though initial correlations look promising, neither emotional arousal nor primary emotions are correlated with the stock markets. Hence, our results imply that there cannot be any causal link between the two. We believe that this is due to the unspecific nature of the arousal and of the primary emotions data. It seems only natural after all that tweets concerning the performance of the Paris Symphonic Orchestra or aunt Annie's cookies have little relevance for the markets. There is evidence, however, that fine-grained microblogging data pertaining to the stock markets is relevant and can even be predictive of the stock returns. For future studies in this field, this means that researchers should avoid unspecific data and rather collect material that directly concerns the market. Another suggestion for future work would be to consider secondary and tertiary emotions. And a last suggestion would be to take into account the different components of the emotion (valence, arousal, and dominance) in the analyses.

References

Baker, H. K. & Nofsinger, J. R. (2010). *Behavioral finance: Investors, corporations, and markets*. Hoboken, N.J.: Wiley.

Beddington, J. (2012). *The future of computer trading in financial markets: An international perspective*. The Government Office for Science. London.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8.

Dey, S. K. (2014). Stock market prediction using Twitter mood. *International Journal of Scientific & Engineering Research*, *5*(5), 44–47.

Esmin, A. A. A., de Oliveira, R. L., & Matwin, S. (2012). Hierarchical classification approach to emotion recognition in Twitter. In *11th International Conference on Machine Learning and Applications (ICMLA)* (pp. 381–385). doi:10.1109/ICMLA.2012.195

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383–417.

Foucault, T. (2012). Algorithmic trading: Issues and preliminary evidence. In *Market microstructure: Confronting many viewpoints* (pp. 1–40). John Wiley & Sons.

Frijda, N. H. (2008). The psychologist's point of view. In *Handbook of emotions* (3rd ed., pp. 68–87). New York: Guilford Press.

Graham, C., Chattopadhyay, S., & Picon, M. (2010). Adapting to adversity: Happiness and the 2009 economic crisis in the United States. *Social Research*, 715–748.

Grossman, L. (2010, October). Twitter can predict the stock market. *Wired*.

Ifcher, J. & Zarghamee, H. (2014). Affect and overconfidence: A laboratory investigation. *Journal of Neuroscience, Psychology, and Economics*, *7*(3), 125–150. doi:10.1037/npe0000022

Kaminski, J. & Gloor, P. (2014, June). Nowcasting the Bitcoin market with Twitter signals. Cambridge, USA.

Keynes, J. M. (1923). Some aspects of commodity markets. *Manchester Guardian Commercial: European Reconstruction Series*, *13*, 784–786.

Khadjeh Nassirtoussi, A., Aghabozorgi, S., Ying Wah, T., & Ngo, D. C. L. (2015, January). Text mining of news-headlines for FOREX market prediction: A multi-layer dimension reduction algorithm with semantics and sentiment. *Expert Systems with Applications*, *42*(1), 306–324. doi:10.1016/j.eswa.2014.08.004

Lang, A., Dhillon, K., & Dong, Q. (1995, June). The effects of emotional arousal and valence on television viewers' cognitive capacity and memory. *Journal of Broadcasting & Electronic Media*, *39*(3), 313–327. doi:10.1080/08838159509364309

Lawly Wurm. (2014, April). The junk science behind the 'Twitter hedge fund' [Buy the hype]. Retrieved June 9, 2015, from http://sellthenews.tumblr.com/post/21067996377/noitdoesnot

Levenberg, A., Pulman, S., Moilanen, K., Simpson, E., & Roberts, S. (2014). Predicting economic indicators from web text using sentiment composition. *International Journal of Computer and Communication Engineering*, *3*(2), 109–115. doi:10.7763/IJCCE.2014.V3.302

Li, Q., Wang, T., Li, P., Liu, L., Gong, Q., & Chen, Y. (2014, September). The effect of news and public mood on stock movements. *Information Sciences*, *278*, 826–840. doi:10.1016/j.ins.2014.03.096

Linshi, J. (2014, August). Twitter: No, 23 million of our active users are not actually bots. *Time*.

Matthews, C. (2013). How does one fake tweet cause a stock market crash? *Time*.

McMillan, G. (2011, March). Let Twitter tell you where to invest your money. *Time*.

Mehrabian, A. (1996). Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, *14*(4), 261–292.

Mitra, G. & Mitra, L. (2011, July). *The handbook of news analytics in finance*. John Wiley & Sons.

Ohmura, M., Kakusho, K., & Okadome, T. (2014). Stock market prediction by regression model with social moods. *International Journal of Social, Management, Economics and Business Engineering*, *8*(10), 2942–2945.

Oliveira, N., Cortez, P., & Areal, N. (2013). Some experiments on modeling stock market behavior using investor sentiment analysis and posting volume from Twitter. In *Proceedings of the 3rd International Conference on Web Intelligence, Mining and Semantics* (p. 31). ACM.

Robot writes LA Times earthquake breaking news article. (2014, March). *BBC News Online*.

Parrott, W. G. (2001). *Emotions in social psychology: Essential readings*. Key Readings in Social Psychology. Psychology Press.

Quest, R. (2010, October). *China raises interest rates; protests, strikes in France*. CNN.

Rao, T. & Srivastava, S. (2012). Analyzing stock market movements using Twitter sentiment analysis. In *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)* (pp. 119–123). IEEE Computer Society.

Rao, T. & Srivastava, S. (2014). Twitter sentiment analysis: How to hedge your bets in the stock markets. In *State of the art applications of social network analysis* (pp. 227–247). Springer.

Schoonderwoerd, N. (2013). *4 ways how Twitter can keep growing - PeerReach blog.*

Sewell, M. (2011). History of the efficient market hypothesis. *University College London - Research Note*, *11*(4), 04.

Shirawadekar, A. (2014). *Stock market movement analysis using social feeds* (Doctoral dissertation, Rochester Institute of Technology).

Shumway, T. (2010). Mood. In *Behavioral finance: Investors, corporations, and markets* (pp. 671–679). Hoboken, N.J.: Wiley.

Si, J., Mukherjee, A., Liu, B., Pan, S. J., Li, Q., & Li, H. (2014). Exploiting social relations and sentiment for stock prediction. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1139–1145).

Siganos, A., Vagenas-Nanos, E., & Verwijmeren, P. (2014, July). *Divergence of sentiment and stock market trading* (SSRN Scholarly Paper No. ID 2462140). Social Science Research Network. Rochester, NY.

Sprenger, T. O., Tumasjan, A., Sandner, P. G., & Welpe, I. M. (2014, November). Tweets and trades: The information content of stock microblogs. *European Financial Management*, *20*(5), 926–957. doi:10.1111/j.1468-036X.2013.12007.x

Sul, H., Dennis, A. R., & Yuan, L. I. (2014). Trading on Twitter: The financial information content of emotion in social media. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on System Sciences* (pp. 806–815). IEEE.

Teymourian, K. (2014, November). *Introduction to semantic complex event processing.* Freie Universität Berlin.

Vogt, J., De Houwer, J., Koster, E. H. W., Van Damme, S., & Crombez, G. (2008, January). Allocation of spatial attention to emotional stimuli depends upon arousal and not valence. *Emotion*, *8*(6), 880–885.

Yen, G. & Lee, C. (2008). Efficient market hypothesis (EMH): Past, present and future. *Review of Pacific Basin Financial Markets and Policies*, *11*(2), 305–329.

Zhang, X., Fuehres, H., & Gloor, P. A. (2012). Predicting asset value through Twitter buzz. In *Advances in collective intelligence 2011* (pp. 23–34). Springer.

Appendix A

Table A1

Parrot's chart of emotions

Primary emotion	Secondary emotion	Tertiary emotions
Love	Affection	Adoration, affection, love, fondness, liking, attraction, caring, tenderness, compassion, sentimentality
	Lust	Arousal, desire, lust, passion, infatuation
	Longing	Longing
Joy	Cheerfulness	Amusement, bliss, cheerfulness, gaiety, glee, jolliness, joviality, joy, delight, enjoyment, gladness, happiness, jubilation, elation, satisfaction, ecstasy, euphoria
	Zest	Enthusiasm, zeal, zest, excitement, thrill, exhilaration
	Contentment	Contentment, pleasure
	Pride	Pride, triumph
	Optimism	Eagerness, hope, optimism
	Enthrallment	Enthrallment, rapture
	Relief	Relief
Surprise	Surprise	Amazement, surprise, astonishment
Anger	Irritation	Aggravation, irritation, agitation, annoyance, grouchiness, grumpiness
	Exasperation	Exasperation, frustration
	Rage	Anger, rage, outrage, fury, wrath, hostility, ferocity, bitterness, hate, loathing, scorn, spite, vengefulness, dislike, resentment
	Disgust	Disgust, revulsion, contempt
	Envy	Envy, jealousy
	Torment	Torment
Sadness	Suffering	Agony, suffering, hurt, anguish
	Sadness	Depression, despair, hopelessness, gloom, glumness, sadness, unhappiness, grief, sorrow, woe, misery, melancholy
	Disappointment	Dismay, disappointment, displeasure
	Shame	Guilt, shame, regret, remorse
	Neglect	Alienation, isolation, neglect, loneliness, rejection, homesickness, defeat, dejection, insecurity, embarrassment, humiliation, insult
	Sympathy	Pity, sympathy
Fear	Horror	Alarm, shock, fear, fright, horror, terror, panic, hysteria, mortification
	Nervousness	Anxiety, nervousness, tenseness, uneasiness, apprehension, worry, distress, dread

Appendix B

Table B1

Model 3, (T = 204)

Dependent variable: dlogWILSHIRE5000

	Coefficient	Std. Error	t-ratio	p-value
const	0.0307961	0.0981006	0.3139	0.7539
logTWEETSDAY	−0.00167968	0.00541520	−0.3102	0.7567

Mean dependent var	0.000368	S.D. dependent var	0.007837
Sum squared resid	0.012461	S.E. of regression	0.007854
R^2	0.000476	Adjusted R^2	-0.004472
$F(1, 202)$	0.096211	P-value(F)	0.756745
Log-likelihood	700.2714	Akaike criterion	−1396.543
Schwarz criterion	−1389.907	Hannan–Quinn	−1393.858
$\hat{\rho}$	−0.023338	Durbin–Watson	2.041175

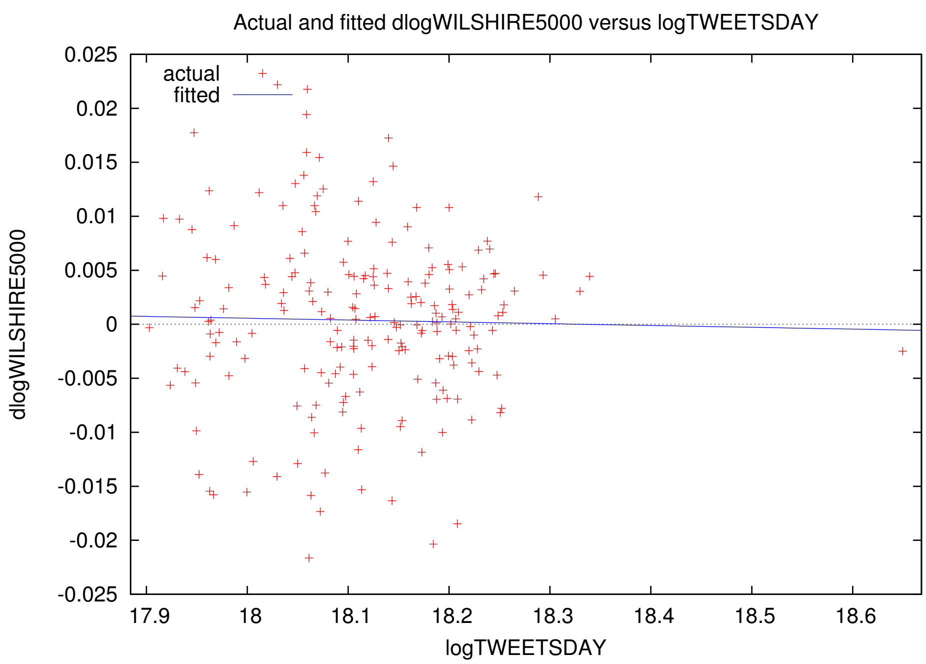

Figure B1. Scatterplot of the first order difference of the log of Wishire 5000 against the log of Arousal (computed only with the total number of tweets per day).

Appendix C

Table C1

Model 4, (T = 204)

Dependent variable: dlogWILSHIRE5000

	Coefficient	Std. Error	t-ratio	p-value
const	0.0608088	0.152552	0.3986	0.6906
l_SURPRISE	0.00472125	0.0104067	0.4537	0.6506
l_JOY	0.0158747	0.0639742	0.2481	0.8043
l_LOVE	−0.0159689	0.0134270	−1.1893	0.2358
l_SADNESS	0.00604487	0.0190355	0.3176	0.7512
l_ANGER	0.0140351	0.00925615	1.5163	0.1311
l_FEAR	0.00507350	0.00924583	0.5487	0.5838
l_OTHER	−0.0432042	0.0252193	−1.7131	0.0883
l_NOEMOTION	0.00872600	0.102708	0.0850	0.9324

Mean dependent var	0.000368	S.D. dependent var	0.007837
Sum squared resid	0.011993	S.E. of regression	0.007842
R^2	0.038032	Adjusted R^2	-0.001433
$F(8, 195)$	0.963688	P-value(F)	0.465695
Log-likelihood	704.1778	Akaike criterion	−1390.356
Schwarz criterion	−1360.493	Hannan–Quinn	−1378.276
$\hat{\rho}$	−0.037012	Durbin–Watson	2.073441

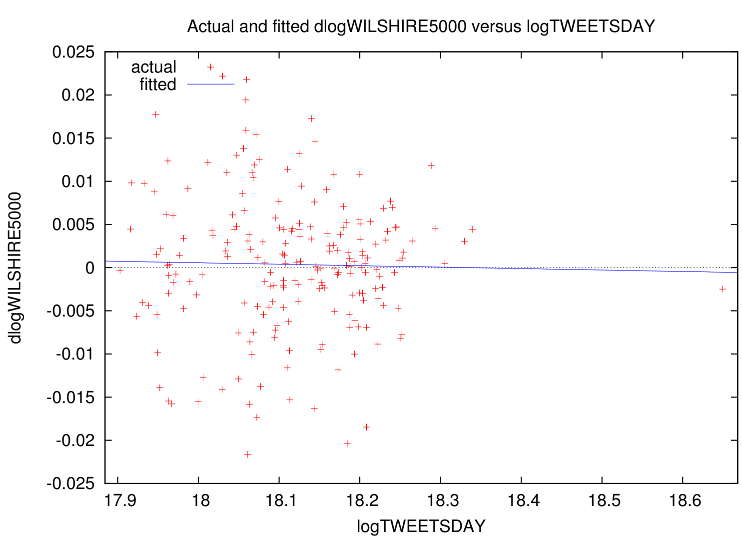

Figure C1. Scatterplot of the first order difference of the log of Wishire 5000 against the log of emotions.

Appendix D

Table D1

Model 5, (T = 200)

Dependent variable: dlogWILSHIRE5000

	Coefficient	Std. Error	t-ratio	p-value
const	0.000287219	0.000566919	0.5066	0.6130
dlogAROUSAL	0.0553817	0.0739208	0.7492	0.4547
dlogAROUSAL_1	0.0258612	0.0875416	0.2954	0.7680
dlogAROUSAL_2	0.0209823	0.0879244	0.2386	0.8116
dlogAROUSAL_3	−0.0304015	0.0743110	−0.4091	0.6829
d_logTWEETSDAY	−0.0625415	0.0772126	−0.8100	0.4190
d_logTWEETSDAY_1	−0.0316824	0.0908109	−0.3489	0.7276
d_logTWEETSDAY_2	−0.0299038	0.0914550	−0.3270	0.7440
d_logTWEETSDAY_3	0.0281902	0.0777453	0.3626	0.7173

Mean dependent var	0.000302	S.D. dependent var	0.007889
Sum squared resid	0.012261	S.E. of regression	0.008012
R^2	0.009939	Adjusted R^2	-0.031529
$F(8, 191)$	0.239682	P-value(F)	0.982862
Log-likelihood	686.1767	Akaike criterion	−1354.353
Schwarz criterion	−1324.669	Hannan–Quinn	−1342.340
$\hat{\rho}$	−0.023156	Durbin–Watson	2.041611

Table D2
Model 6, (T = 201)

Dependent variable: dlogWILSHIRE5000

	Coefficient	Std. Error	t-ratio	p-value
const	0.193822	0.236626	0.8191	0.4139
l_SURPRISE_1	−0.00387262	0.0125700	−0.3081	0.7584
l_SURPRISE_2	−0.00124344	0.0125267	−0.0993	0.9210
l_SURPRISE_3	0.00714000	0.0122411	0.5833	0.5605
l_JOY_1	0.00812509	0.0850710	0.0955	0.9240
l_JOY_2	0.110957	0.0856704	1.2952	0.1970
l_JOY_3	−0.0540705	0.0821280	−0.6584	0.5112
l_LOVE_1	0.000358636	0.0173344	0.0207	0.9835
l_LOVE_2	0.0196426	0.0173382	1.1329	0.2589
l_LOVE_3	−0.0501688	0.0166459	−3.0139	0.0030
l_SADNESS_1	0.0127177	0.0263394	0.4828	0.6298
l_SADNESS_2	0.0340494	0.0266874	1.2759	0.2038
l_SADNESS_3	−0.0182308	0.0256357	−0.7111	0.4780
l_ANGER_1	0.00284563	0.0134583	0.2114	0.8328
l_ANGER_2	0.00569893	0.0135584	0.4203	0.6748
l_ANGER_3	−0.0188304	0.0123716	−1.5221	0.1299
l_FEAR_1	0.00229064	0.0109675	0.2089	0.8348
l_FEAR_2	0.0128126	0.0112610	1.1378	0.2568
l_FEAR_3	0.00359261	0.0110045	0.3265	0.7445
l_OTHER_1	−0.0272957	0.0326076	−0.8371	0.4037
l_OTHER_2	0.0237766	0.0327915	0.7251	0.4694
l_OTHER_3	−0.0223413	0.0300908	−0.7425	0.4588
l_NOEMOTION_1	0.0121285	0.139345	0.0870	0.9307
l_NOEMOTION_2	−0.221981	0.140327	−1.5819	0.1156
l_NOEMOTION_3	0.166666	0.133421	1.2492	0.2133

Mean dependent var	0.000326	S.D. dependent var	0.007876
Sum squared resid	0.010570	S.E. of regression	0.007932
R^2	0.148004	Adjusted R^2	-0.014280
$F(32, 168)$	0.912004	P-value(F)	0.606824
Log-likelihood	705.0196	Akaike criterion	−1344.039
Schwarz criterion	−1235.030	Hannan–Quinn	−1299.929
$\hat{\rho}$	−0.057329	Durbin–Watson	2.111853