

https://dspace.org/

Critical Insights Into the Design of Big Data Analytics Research: How Twitter “Moods” Predict Stock Exchange Index Movement

Maree, Stiaan; Johnston, Kevin

2015-12-15
LINK Centre, University of the Witwatersrand (Wits), Johannesburg
http://hdl.handle.net/10539/20330
Maree, S., & Johnston, K. (2015). Critical insights into the design of big data analytics research: How Twitter “moods” predict stock exchange index movement. The African Journal of Information and Communication (AJIC), 15, 53-67. https://doi.org/10.23962/10539/20330

Downloaded from a DSpace repository. https://dspace.org/

CRITICAL INSIGHTS INTO THE DESIGN OF BIG DATA ANALYTICS RESEARCH: HOW TWITTER "MOODS" PREDICT STOCK EXCHANGE INDEX MOVEMENT

Stiaan Maree
Department of Information Systems, University of Cape Town, South Africa

Kevin Johnston
Associate Professor, Department of Information Systems, University of Cape Town, South Africa

ABSTRACT

The research explored whether one or more of the South African Twitter moods could be used to predict the movement of the Johannesburg Stock Exchange (JSE) All Share Index (ALSI). This is a proof of principle study in the field of big data analytic research in South Africa, which is at a relatively early stage of development. The research methods used secondary data from Twitter's application programming interfaces (APIs), and formulated a model to extract public mood data and search for a causal effect of the mood on the closing values of the JSE ALSI. Over three million tweets were gathered and analysed over a 55-day period, with data collected from the JSE for 39 weekdays, from which only one variable (mood states) was considered. Four of the South African Twitter mood states did not produce any correlation with the movement of the JSE ALSI. The mood Depression had a significant negative correlation with the same day's JSE ALSI values. The major finding was that there was a highly significant positive correlation between the Fatigue mood and the next day's closing value of the JSE ALSI, and a significant causality correlation from the Fatigue mood to the JSE ALSI values. The findings support the behavioural finance theory (Wang, Lin & Lin, 2012), which states that public mood can influence the stock market. Organisations and governments could use Twitter data to gauge public mood and to ascertain the influence of public mood on particular issues. However, very large data sets are required for analytical purposes, possibly five to ten years of data, without which predictability is likely to be low.

KEYWORDS

twitter moods, predict, stock exchange index movement, big data analytics, Africa and developing countries.

INTRODUCTION: A PROOF OF PRINCIPLE STUDY

Predicting stock market movements has motivated researchers and practitioners to formulate new models and methodologies (Atsalakis, Dimitrakakis, & Zopounidis, 2011). The daily fluctuations of the Dow Jones Industrial Average (DJIA) was predicted with an 86,7% accuracy using Twitter public moods by Bollen, Mao and Zeng (2011). Research into public moods has been of interest to academics from several disciplines, such as psychology (Stolarski, Matthews, Postek, Zimbardo, & Bitner, 2014), social science (Johnston & Newman, 2014), politics (Ellis & Faricy, 2011), and information sciences (Li, Wang, Li, Liu, Gong, & Chen, 2014). Practitioners from government (Hakhverdian, 2012) and industry (Bollen, Mao, & Zeng, 2011) appear to be interested in public moods, and their effects.

The aim of this proof of principle study was to understand the complexities involved in designing a study to explore whether analysis of one or more of the South African Twitter public moods could be used to predict the movement of the Johannesburg Stock Exchange (JSE) All Share Index (ALSI). The study engages with big data research approaches and methodologies, which are at a relatively early stage of development in South Africa (Kahn, Higgs, Davidson, & Jones, 2014). This preliminary study could be utilised as the foundation for designing studies for other African stock exchanges. Premier stock exchanges such as the London Stock Exchange (LSE) may be interested in the study, as it has relevance to their potential partnerships with African stock exchanges, including Egypt, Kenya, Morocco and Nigeria, as reported by Blas (2014). Forgha (2012, p. 176) stated that there are "twenty nine stock exchange markets in Africa, twenty one are members of the African Securities Exchanges Association (ASEA)". It is unclear if several have closed down as, according to the World Stock Exchanges (2014), there are 16 stock exchanges in Africa, four each in North and East Africa, three in West Africa, and five in Southern Africa. Irrespective of the number of stock exchanges in Africa, this study could be replicated for these countries and their stock exchanges.

The research was conducted in South Africa, using 3,104,364 tweets from within South Africa, and the daily closing prices of the Johannesburg Stock Exchange (JSE) All Share Index (ALSI), over a 55-day period. The tweets were tweeted by 282,211 unique users. Of the collected tweets, 2,305,063 tweeted by 259,671 unique users had "feeling" scores. The study analysed publicly available data downloaded from Twitter, using two of Twitter's application programming interfaces (APIs), and the closing values of the JSE ALSI. A model (XPOMS) was developed based on the Profile of Mood States (POMS) to extract the mood from Twitter data and quantitative analysis was conducted to extract the moods from these South African tweets. XPOMS generated a score for 2,3 million tweets which had "feeling" scores. The moods as defined by the XPOMS model were then mapped against the JSE ALSI closing values to see whether it was possible to use one or more of the moods to predict the movement of the ALSI.

The research focused on a microstructure of the market, rather than the market as whole. Market microstructure has been defined by O'Hara (2003) as the "study of the process and outcomes of exchanging assets under a specific set of rules". More commonly, this refers to the trading mechanisms used for real financial assets and the effect on price information and discovery, transactions costs, market structure and investor behaviour. The relationship between the transparency of such information and the process by which the price is determined was focused on. The value of publishing this work now is that it is one of the building blocks towards the future full-scale analysis of stock exchange index movements.

In 2014, South Africa was estimated to have a population of 54 million (StatsSA, 2014) and an annual gross domestic product of ZAR3,008 trillion (StatsSA, 2015). The World Bank ranks South Africa as an upper-middle income economy. In April 2011 it joined BRICS (Brazil, Russia, India, China, South Africa), a forum for economically significant emerging economies (Khan, 2011). South Africa has 11 official languages and is considered a country with a diverse range of cultures (Mwakikagile, 2010).

The Johannesburg Stock Exchange (JSE) Limited, established in 1887, is South Africa's stock market and the biggest in Africa (Eita, 2012). The market capitalisation grew from USD151 million in 1998 (Eita, 2012) to nearly USD1 billion in 2015 (ZAR12 368 billion on 20 July) with approximately 400 listed companies, making it the 19th largest in the world (JSE, 2015). Initiatives to improve the JSE's functioning were introduced in the late 1990s, including the Stock Exchanges Control Amendment Act 54 of 1995, an electronic clearing and settlement system, and a real-time stock exchange news service (Eita, 2012). Further legislative change was introduced in 2005 with the passage of the Securities Services Act No 36 of 2004, which aims to align with international regulatory practice.

The research focused on one particular index of the FTSE/JSE Africa Index Series, namely the All Share Index (ALSI). "The JSE All Share Index is calculated based on component share prices that are averaged according to specific rules which are impacted by stock splits and dividends" (Campbell, 2011, p. 4). Research on predictability has previously been conducted with respect to the JSE, in which it was found that past and present values of inflation can help predict the JSE returns (Eita, 2012). Evidence has been found for the predictive ability of simple technical trading rules on the JSE; however, the results lack statistical significance (Campbell, 2011). The yield curve has also been found to be a powerful tool for predicting down-swings on the JSE (Clay & Keeton, 2011).

LITERATURE REVIEW: MOODS AND PREDICTIVE CAPABILITY

The literature search commenced with the article "Twitter mood predicts stock market" by Bollen et al. (2011), which called for future studies to factor in location and language. This study attempted to replicate the Bollen et al. (2011) work in a developing world context, using a refined analytical tool. Articles from the reference section of the Bollen et al. (2011) article, as well as articles that referenced the Bollen et al. (2011) article were reviewed. Google Scholar was used to find articles on relevant topics including Twitter, mood states and stock exchange prediction, and the JSE. Twitter's website was also searched for useful information. The literature review investigated the current state of knowledge on using Twitter moods to predict stock market movements, on its limitations in making predictions, and on how the topic fits into the wider context, as suggested by Saunders, Lewis and Thornhill (2009).

TWITTER

Since Twitter's inception in 2006, it has seen tremendous growth, with 140 million active users posting 340 million tweets daily (Golbeck, Grimes, & Rogers, 2010; Rios & Lin, 2012). "Twitter is an Internet social-network and micro-blogging platform with both mass and interpersonal communication features for sharing 140-character messages, called tweets, with other people, called followers" (Chen, 2011, p. 755). Users tweet on everything, including the weather, news, sports results, and their feelings and moods. South Africa has over 1,1 million registered Twitter users (Vermeulen, 2012).

PROFILE OF MOOD STATES (POMS)

The Profile of Mood States (POMS) is a simple low-cost, user friendly instrument whose factor-analytical structure has been validated numerous times, which has been used in hundreds of research studies, and has been normed for various populations (Pepe & Bollen, 2008). The POMS questionnaire (McNair, Lorr & Droppelman, 1971) measures six dimensions of mood, namely Tension-anxiety, Depression-dejection, Anger-hostility, Vigour-activity, Fatigue-inertia and Confusion-bewilderment (Pepe & Bollen, 2008). POMS and derivations of POMS are simple instruments, which are not machine-learning algorithms.

Public moods can be tracked with the use of large-scale surveys, but the accuracy is limited by the degree to which the indicators correlate with public mood (Bollen et al., 2011). Great improvements have been made in the use of social media, such as Twitter, to track public mood (Bollen et al., 2011).

TWITTER MOOD CLASSIFICATION

Although many Fortune 500 companies are using social media platforms such as Twitter to interact with customers (Culnan, Mchugh, & Zubillaga, 2010), few are mining customers' tweets or moods. Bollen et al. (2011) classified Twitter moods using two tools, firstly OpinionFinder, which analysed tweets on a positive versus negative mood scale, and secondly Google-Profile of Mood states (GPOMS) derived from POMS. GPOMS classifies public mood according to six dimensions, namely Calm, Happy, Kind, Alert, Vital and Sure (Bollen et al., 2011). In this 2012 study, a new tool called Extended Profile of Mood States (XPOMS) was developed which included Afrikaans terms to classify the Twitter stream and search data. XPOMS had the same six moods as the original POMS, namely Depression, Tension, Anger, Vigour, Fatigue and Confusion.

STOCK EXCHANGE PREDICTION

A generation ago, academic financial economists accepted the efficient market hypothesis (EMH) as the leading theory for stock market prediction (Malkiel, 2003). "This hypothesis is associated with the view that stock market

price movements approximate those of a random walk. If new information develops randomly, then so will market prices, making the stock market unpredictable apart from its long-run uptrend" (Malkiel, 2005, p. 1). The implication is that stock market trends can only be predicted with 50% accuracy, according to EMH (Bollen et al., 2011). Lately, however, economists doubt the efficiency of EMH (Malkiel, 2005), and Bollen et al. (2011) identified two issues with EMH. The first problem is that several studies have concluded that stock prices can, to a certain degree, be predicted and do not follow a random walk (Bollen et al., 2011; Malkiel, 2003). The second problem is that, although news is unpredictable, early indicators can be extracted from social media to predict economic indicators (Bollen et al., 2011).

Other theories of stock market prediction have since emerged, such as behavioural finance (Bollen et al., 2011). "Behavioural finance combines behavioural and financial theory with the aim of analyzing the psychology, behavior and mood involved in financial decision-making, meaning the results of such research fall within the realms of both psychology and finance" (Wang, Lin & Lin, 2012, p. 696). Behavioural finance has proved that financial decisions are driven by mood and emotions (Bollen et al., 2011; Subrahmanyam, 2007). "Results indicate that the accuracy of Dow Jones Industrial Average (DJIA) predictions can be significantly improved by the inclusion of specific public mood dimensions but not others" (Bollen, Mao, & Zeng, 2011, p. 1). Subrahmanyam (2007) tied mood to stock market changes through behavioural finance, and Edmans, Garcia and Norli (2007) noticed that stock market changes can be influenced by sporting events, which affect the country as a whole. "This suggests that investor mood (ostensibly negative on cloudy days) affects the stock market" (Subrahmanyam, 2007, p. 17).

Event study analysis is concerned with measuring the impact of events on the value of companies and thus stock markets, using specific mathematical formulae (Dimpfl, 2011; Hart, 2006). Globally important news is processed quickly, and affects the volatility of stock markets (Dimpfl, 2011). Event study analysis could measure the validity of the XPOMS tool developed to see whether news events were picked up by Twitter moods, and to see whether events picked up and reflected by Twitter moods affected the JSE ALSI movement.

TWITTER MOOD PREDICTS THE STOCK MARKET

The correlation between public moods as gathered from Twitter with the DJIA was researched by Bollen et al. (2011). Between February 28 and December 19, 2008, almost 10 million tweets (9 853 498) were analysed to see if Twitter moods could predict the DJIA. Bollen et al. (2011) found that one of the moods (Calm) could indeed predict the DJIA with 86,7% accuracy. This review of the relevant literature informed the framing of the research question: Could one or more of the South African Twitter moods be used to predict the movement of the JSE ALSI?

RESEARCH METHODOLOGY

The research filled a gap revealed by Bollen et al. (2011) for further research on using Twitter moods to predict stock market movements in specific geographical areas. The purpose of this research was threefold, firstly as a further check on Bollen et al.'s (2011) methods and findings, secondly to attempt to predict stock market movements in South Africa, and thirdly to see if it would hold true for markets in a developing country such as South Africa.

From the research problem, the following hypothesis (and null hypothesis) was developed:

H1. One or more of the South African Twitter moods can be used to predict the movement of the JSE ALSI.
H10: One or more of the South African Twitter moods cannot be used to predict the movement of the JSE ALSI.
The movement of the ALSI of the JSE is the dependent variable, changing according to the independent variables, the South African Twitter moods.

The following six sub-hypotheses spring from the main hypothesis:

H2. Depression South African Twitter mood classified according to XPOMS can be used to predict the movement of the JSE ALSI.
H3. Tension South African Twitter mood classified according to XPOMS can be used to predict the movement of the JSE ALSI.
H4. Anger South African Twitter mood classified according to XPOMS can be used to predict the movement of the JSE ALSI.
H5. Vigour South African Twitter mood classified according to XPOMS can be used to predict the movement of the JSE ALSI.
H6. Fatigue South African Twitter mood classified according to XPOMS can be used to predict the movement of the JSE ALSI.
H7. Confusion South African Twitter mood classified according to XPOMS can be used to predict the movement of the JSE ALSI.

The study will benefit many interest groups, including investors who are trying to make better decisions about stock market movements, as well as academics who study the predictability of the stock market or who study Twitter as a gauge for public mood. Stock market prediction has attracted research, with none being able to fully predict the market (Schumaker & Chen, 2009). Various studies have been conducted on predicting stock markets using a variety of methods (Atsalakis, et al., 2011; Bollen et al., 2011; Zhang & Wu, 2009).

The research was guided by a positivist philosophy, which entails "working with an observable social reality and that the end product of such research can be law-like generalizations similar to those produced by the physical and natural scientists" (Remenyi, Williams, Money, & Swartz, 1998, p. 32). The phenomena of Twitter moods can be observed and hypotheses developed to test whether the Twitter moods can be used to predict the movements of the JSE ALSI. A hallmark of the positivist approach is that the research is undertaken value-free, and data is collected unobtrusively (Saunders, et al., 2009). The data resources (Twitter moods and JSE ALSI closing values) were downloaded from the Internet, and the researchers were unable to influence either of these data sets. Complete freedom of the researchers' values is impossible and, in this case, Twitter search terms, geographical co-ordinates, XPOMS classification terms and time intervals of downloads could have been influenced by the researchers values (Saunders et al., 2009). Quantitative analysis (such as Spearman rank correlation and Granger causality) are closely associated with positivism (Saunders et al., 2009).

The approach to the research was deductive, in that the theory that Twitter moods can be used to predict stock market movement was tested. The following steps, as listed by Saunders et al. (2009), were followed in the research process. Firstly, theories from existing literature by Bollen et al. (2011) led to the deduction of seven hypotheses. These hypotheses were expressed proposing a relationship between two variables (South African Twitter moods and JSE ALSI closing values). The hypotheses were then tested, after which the outcome of the inquiry was examined and the theory revisited. Important characteristics of a deductive approach include the "search to explain causal relationships between two variables" (Saunders et al., 2009, p. 125). The independent variables were *Twitter moods* causing movements on the dependent variable, the JSE ALSI. "Studies that establish causal relationships between variables may be termed explanatory research: (Saunders et al., 2009, p. 140). The study tried to establish a causal relationship between two variables, hypothesising that South African Twitter moods can be used to predict the JSE ALSI movement.

Although data was collected over a two month period, the study was cross-sectional, because it focused on Twitter mood's ability to predict the JSE ALSI movement at a specific point in time, and not on how Twitter mood's ability to predict the JSE ALSI movement changed over time (Saunders et al., 2009). The population consisted of all tweets from within the borders of South Africa for a period of 55 days, and the daily closing price of the JSE ALSI over the same period, plus an additional five days to test for the effect of lag. Twitter exposes random tweets to the Streaming and Search APIs (GET statuses/sample, 2012), so the sample can be considered as a probability sample (Saunders et al., 2009). A total of 3,104,364 tweets were collected, tweeted by 282,211 unique users. In terms of language sampling, English and Afrikaans tweets were used, as these were the only two mentioned in a study done on Twitter language usage by Fischer (2011), and English and Afrikaans are the only two South African languages that form part of Twitter's translation project (Twitter Translation Center, n.d.).

The research tool developed and used to classify the Twitter stream and search data was called the XPOMS, based on POMS. The original POMS consist of six moods, namely Depression, Tension, Anger, Vigour, Fatigue and Confusion, represented by 65 words, which were retained in XPOMS (McNair et al., 1971). This research extended POMS, using Princeton University's WordNet application (WordNet, n.d.). A lexicon of additional terms that link back to the original POMS terms was created for XPOMS, and the terms were translated into Afrikaans using Google Translate. The extended and translated POMS terms were then used to identify one of the six POMS moods per tweet.

Figure 1 details the various applications and databases created to download and analyse South African tweets and the JSE ALSI closing values. tweets containing the following terms were downloaded: "I feel", "I am feeling", "I'm feeling", "I dont feel", "I don't feel", "I'm", "Im", "I am", "makes me". The English terms chosen were exactly the same as the ones used by Bollen et al. (2011), and these terms were also translated into Afrikaans. The end result (Table 1) showed the mood score value per day, calculated by adding up all the tweets falling into a specific mood, dividing that total by the amount of 'feeling' tweets for the day, and multiplying by a factor of 100. Table 1 also contains the JSE ALSI closing value plus lag of four days. APIs and program code are available from the authors on request.

FIGURE 1:THE APPLICATIONS AND DATABASES CREATED TO DOWNLOAD AND ANALYSE SOUTH AFRICAN TWEETS AND THE JSE ALSI CLOSING VALUES.

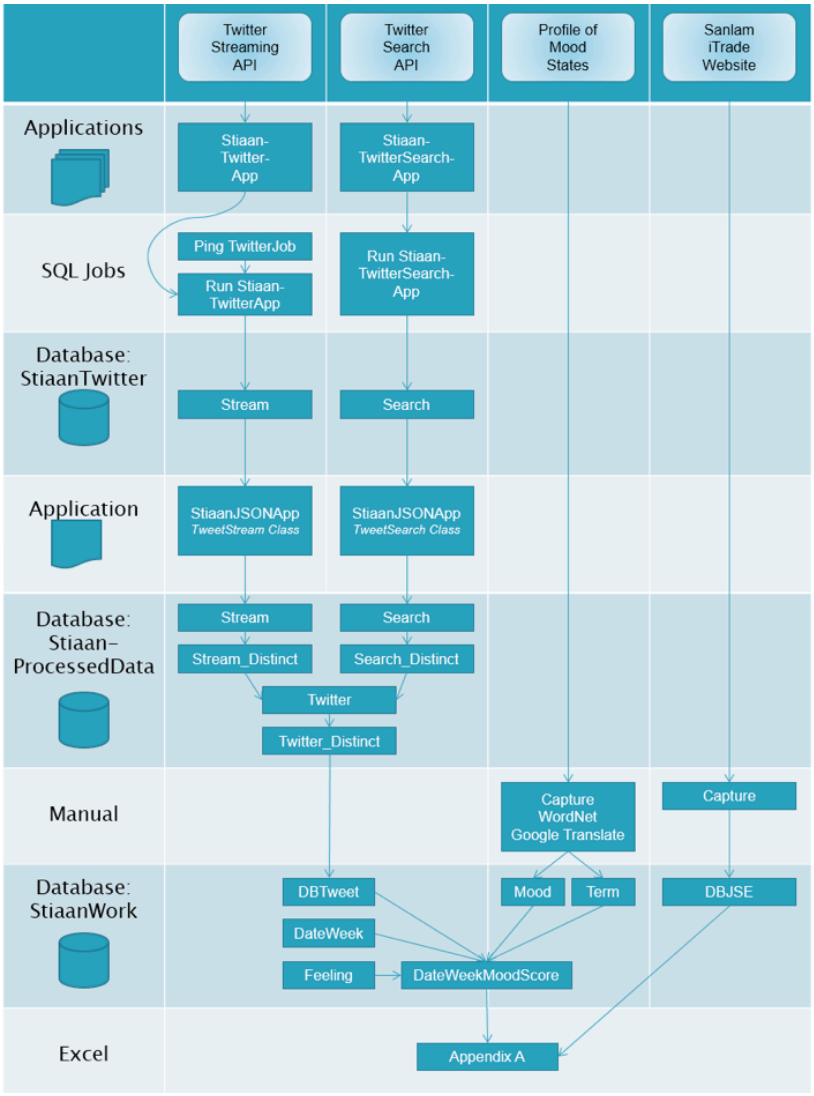


TABLE 1: SUMMARY OF MOOD SCORE VALUES AND ALSI CLOSING VALUES PLUS LAG OVER ONE, TWO, THREE AND FOUR DAYS.

End date	Total	Mood	Alsi	1-day lag	2-day lag	3-day lag	4-day lag
2012-06-11	0.397728673	Tension	33,617	33,855	34,038	33,826	33,960
2012-06-12	0.383339311	Tension	34,038	34,038	33,960	33,960	34,439
2012-06-13	0.372300819	Tension	33,826	33,826	34,439	34,439	34,714

RESEARCH FINDINGS AND ANALYSIS

The lag was tested for one to four days, as can be seen in Table 1. The research methodology was discussed and summarised in Figure 1; however the order of the analysis of the themes differed in that hypotheses H2 to H7 were discussed first, as the main hypothesis (H1) was dependent on these sub-hypotheses.

Of the data collected from 9 June to 2 August 2012, 2,305,063 tweets by 259,671 unique users had "feeling" scores and thus could be used for analysis. When examining the geographical data as entered by the users, places from across South Africa were represented. JSE ALSI data was collected for the period from 9 June to 8 August 2012 to include a four-weekday lag.

Applications used for analysis included: Microsoft Excel, Statistica 10, MATLAB, the Granger Causal Connectivity Analysis MATLAB Toolbox (Seth, 2010), R version 2.15 with MSBVAR (Brandt & Davis, 2012) and lmtest (Hothorn, Zeileis, Millo & Mitchell, 2012).

Data analytics according to XPOMS (H2 to H7)

The six weighted mood scores for the 39 days, along with the JSE ALSI closing values and values for one to four days lag, were used for the analysis as shown in the research process model (Figure 1). The six mood hypotheses were analysed against four rounds of tests, where each hypothesis could be accepted or rejected at the end of each round. The first round tested reliability and validity of the data, using basic statistical and event study analysis. The second round investigated Spearman rank correlations between the moods and JSE ALSI data with lags. The third round of tests involved testing causality of the relationships using the Granger causality test. Finally, in the fourth round, a neural network was set up to test actual prediction, as shown in Table 2.

TABLE 2: ANALYSIS OF SIX MOOD HYPOTHESES THROUGH FOUR TESTS

	Reliability & validity	Spearman	Granger	Neural network
Depression	?	?	?	
Tension	?	?	?	
Anger	?	?	?	
Vigour	?	?	?	
Fatigue	?	?	?	
Confusion	?	?	?	

The data from the DateWeekMoodScore and DBJSE was combined into one spreadsheet for all six moods, which formed the base on which all analysis was done. The basic statistics of the data were examined for reliability and validity by copying the data into Statistica. Table 3 details the results. The JSE data was not analysed as the JSE data was merely a listing of the day's closing ALSI values.

TABLE 3: DESCRIPTIVE STATISTICS OF THE SIX MOODS, AND THE JSE ALSI VALUES

	Valid N	Mean	Median	Minimum	Maximum	Lower-quartile	Upper-quartile	Varience	Std. dev.	Coef. var.	Skewness
Depression	39	6.38	6.36	5.77	6.87	6.18	6.57	0.1	0.2489	3.89952	-0.01194
Tension	39	0.36	0.35	0.25	0.5	0.32	0.39	0	0.0601	16.734	0.51985
Anger	39	2.26	2.23	1.97	3.08	2.17	2.34	0	0.2012	8.89979	2.39165
Vigour	39	3.73	3.66	3.37	4.48	3.55	3.92	0.1	0.2578	6.90752	0.88703
Fatigue	39	3.19	3.17	2.88	3.55	3.03	3.34	0	0.1915	5.99873	0.37921
Confusion	39	0.94	0.95	0.77	1.19	0.9	0.99	0	0.0822	8.7565	0.30856
JSE 0-day lag	39	34,085.9	33,999	33,253	35,071	33,793	34,289	184,411	429.43H1	1.25985	0.67083
JSE 1-day lag	39	34,127.6	34,035	33,253	35,244	33,815	34,292	212,138	460.584	1.3496	0.72676
JSE 2-day lag	39	34,169	34,038	33,253	35,469	33,815	34,439	255,779	505.746	1.48013	0.81152
JSE 3-day lag	39	34,206.1	34,041	33,253	35,487	33,815	34,534	299,628	547.383	1.60025	0.81273
JSE 4-day lag	39	34,248.9	34,070	33,253	35,495	33,815	34,597	337,664	581.089	1.69666	0.7565

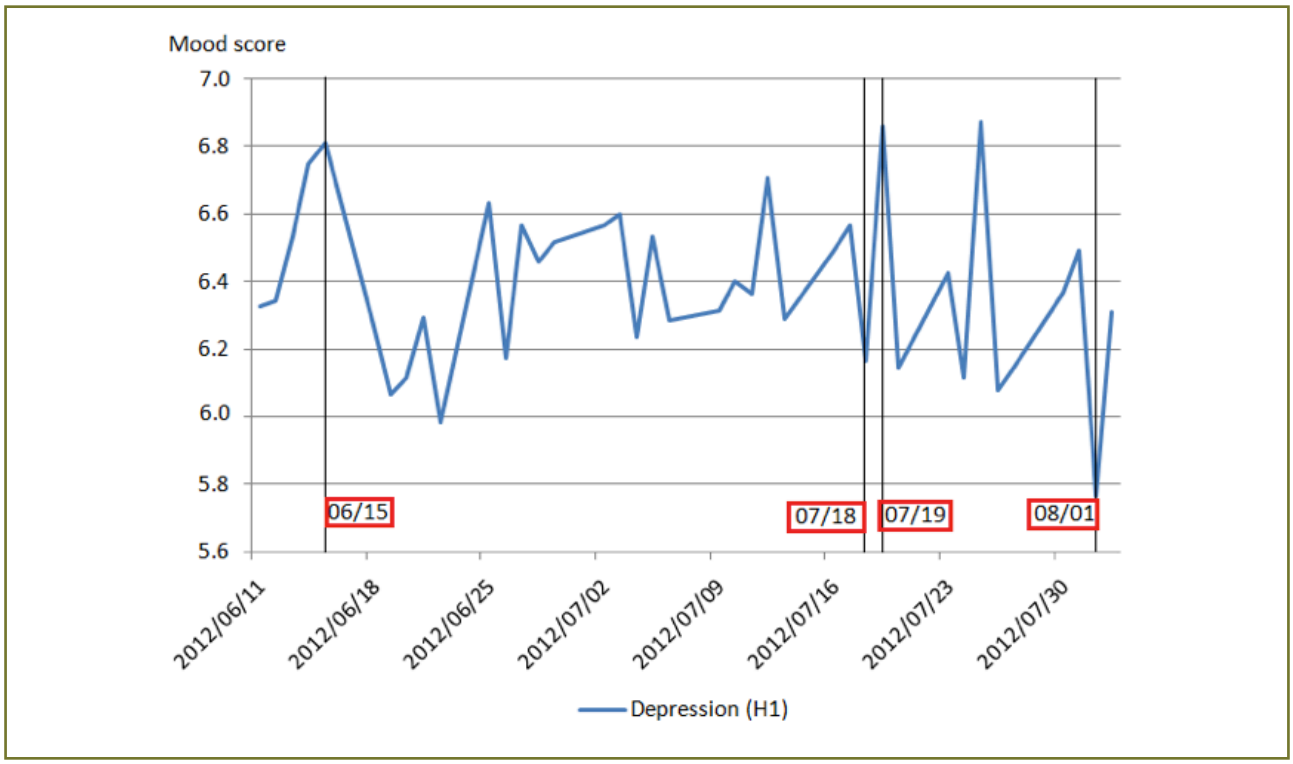
As can be seen in Table 3, the means and medians are close to each other, meaning that the average scores are close to the score in the middle (when arranged from biggest to smallest). In terms of skewness, only the Anger mood scores above 2; this means that it is probably skewed to a significant degree (Brown, 1997); all the other moods have low skewness scores. Reliability of the data examines whether the measuring tool consistently measures the data (Salkind, 2004), in this case whether the XPOMS tool and Twitter programs yield the same results consistently. Normally statistical reliability tests include test-retest reliability, parallel forms reliability and internal consistency reliability (Salkind, 2004).

A variation of test-retest was to split the 39 days' worth of data into two groups, then to compare means and standard deviations to see if the results of the two groups were more or less the same. When comparing the results for the first 19 days and the last 20 days, the major differences appear in the Anger row, where the maximum and standard deviations in the two tables differed. Compared to the basic statistics, the correlations between the means and standard deviation figures were significant ($p<0.5$), meaning that the XPOMS model and the Twitter programs measure results consistently.

Cronbach's Alpha (a coefficient of reliability) which measures internal consistency of a test or scale was not applicable, as it measures reliability of the same construct or concept, and the six moods did not fall into one construct to measure (Salkind, 2004; Tavakol & Dennick, 2011).

Validity is closely related to reliability and is "the property of an assessment tool that indicates that the tool does what it says it does" (Salkind, 2004, p. 289). This research used construct validity to "correlate the set of test scores with some theorized outcome that reflects the construct for which the test is being designed" (Salkind, 2004, p. 289), using event study analysis. Results of the moods were mapped to newsworthy events that might have influenced the scores of the moods. Results of all six moods were analysed and found to be impacted by news. As seen in Figure 2, a high Depression score occurred on June 15, 2012, due to news relating to the unavailability of school textbooks in many South African schools. Depression dipped on Nelson Mandela's birthday on July 18, 2012. The Depression mood in South Africa was very low when Chad le Clos beat Michael Phelps to Olympic gold on August 1, 2012.

FIGURE 2: NEWS EVENTS MAPPED ON DEPRESSION MOOD



The next step was to determine if any evidence of linear relationships existed between any of the moods and the JSE ALSI, including the four-day lag. This was done by running a Spearman rank correlation test on the data. When both variables are intervals, as was the case in the research, then the Spearman rank correlation coefficient is chosen to work out the correlation coefficient (Keller, 2012). Statistica was used where $p<0,01$ (Table 4).

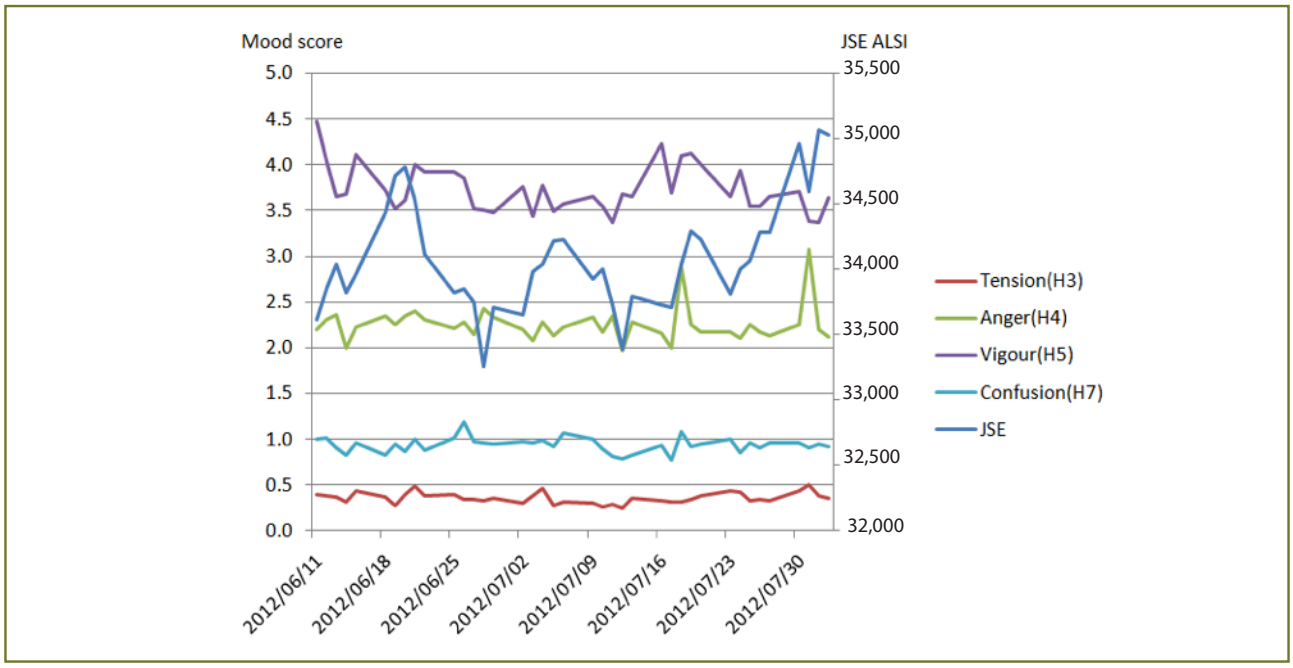
TABLE 4: DATASET SPEARMAN RANK CORRELATION (P < 0,01 IN RED)

	Depression	Tension	Anger	Vigour	Fatigue	Confu-sion	JSE 0-day lag	JSE 1-day lag	JSE 2-day lag	JSE 3-day lag	JSE 4-day lag
Depression	1.0000										
Tension	-0.1415	1.0000									
Anger	-0.2433	0.2490	1.0000								
Vigour	0.0059	0.2759	-0.0342	1.0000							
Fatigue	0.2466	-0.0506	0.1237	-0.1002	1.0000						
Confusion	-0.0164	0.1943	0.2140	0.2411	- 0.0753	1.0000					
JSE 0-day lag	-0.4641	0.2609	0.1806	-0.1293	0.2070	0.0325	1.0000				
JSE 1-day lag	-0.2544	0.2536	0.0480	-0.0389	0.4177	0.0355	0.7638	1.0000			
JSE 2-day lag	-0.1971	0.2684	-0.1645	-0.0671	0.3238	0.1420	0.5361	0.7917	1.0000		
JSE 3-day lag	0.0214	0.2382	-0.1918	-0.1282	0.3800	0.2659	0.4024	0.5897	0.8071	1.0000	
JSE 4-day lag	0.0252	0.2906	-0.1382	-0.0185	0.2957	0.1779	0.2562	0.4380	0.6213	0.8330	1.0000

When analysing the findings of the Spearman rank correlation coefficient, the significant values to note were marked in red. Table 4 has a high level of confidence (99%) and results were highly significant (Keller, 2012).

The hypotheses involved the Twitter mood as the independent variable and the JSE ALSI and lag values as the dependent variables, so correlations between these variables were examined first. It was interesting to note that no significant correlation exists between Tension, Anger, Vigour, Confusion and any JSE ALSI and lag values. As can be seen in Table 4, there were positive correlations between, for example, Anger and JSE four-day lag variables, but these correlations are not significant, because p>0,05. In Figure 3, the JSE ALSI values were mapped against the four moods, showing no noticeable correlation between the values.

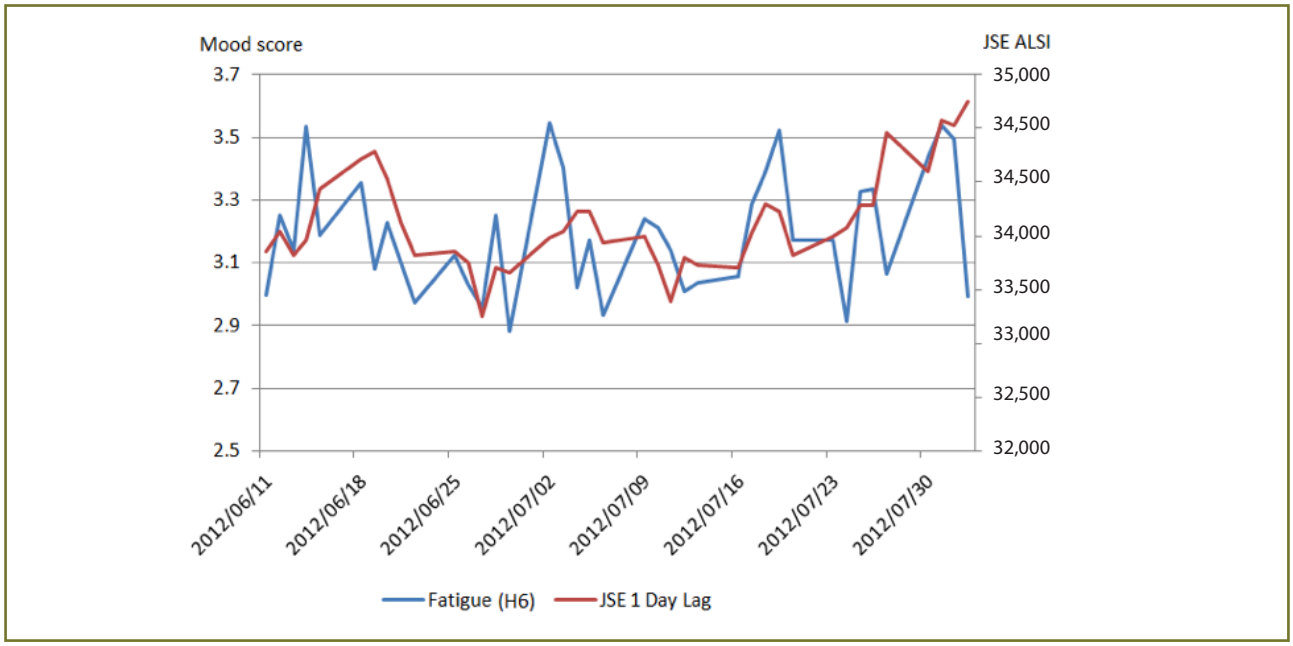
FIGURE 3 : COMPARISON OF FOUR MOOD SCORES WITH THE JSE ALSI



The interpretation of Figure 3 was that there were no relationships between Tension, Anger, Vigour, Confusion and any JSE ALSI and lag values. The independent variables had no bearing on the dependent variable; in other words, observing the Tension, Anger, Vigour, Confusion Twitter mood cannot be useful in predicting the JSE ALSI movement. This led to the acceptance of the null hypotheses for the four variables (H30, H40, H50 and H70).

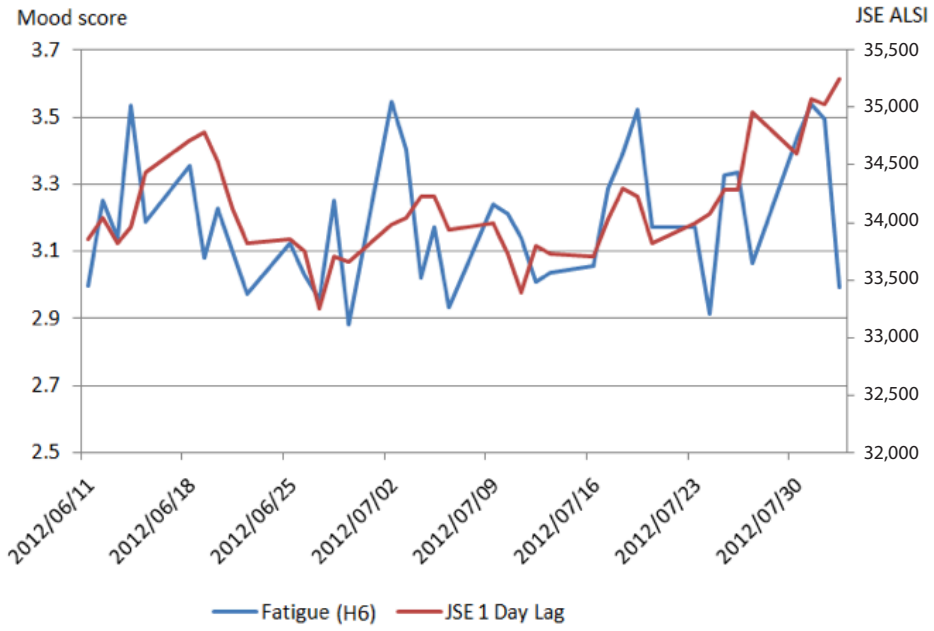
The first significant correlation between Twitter mood and the JSE ALSI was between Depression and the JSE same day ALSI values. The correlation coefficient was at -0,4641, where $p < 0,01$, thus highly significant. A negative correlation coefficient means that there is an inverse relationship between the two variables (Keller, 2012). When the Depression mood score went down, on the same day the JSE ALSI went up, and vice versa – when the Depression mood went up, on the same day the JSE ALSI went down. This relationship could be observed in Figure 4 between the Depression mood score indicated in blue and the JSE ALSI values in red. From 31 July to 1 August 2012 the Depression mood score fell from 6,49 to 5,77; while the JSE ALSI climbed from 34,597 to 35,071.

FIGURE 4: RELATIONSHIP BETWEEN THE DEPRESSION MOOD AND THE JSE ALSI



More significant correlations existed between Fatigue mood score and JSE ALSI values of one, two and three days' lag. Table 4 showed that positive correlations existed between Fatigue mood and JSE one-day lag (coefficient 0,4177), JSE two-day lag (coefficient 0,3238) and JSE three-day lag (coefficient 0,3800). The coefficients were significant ($p<0,05$) whereas the JSE one-day lag coefficient (0,4177) was highly significant with $p<0,01$. The research only focused on the highly significant correlations. The correlation coefficient was a positive one, thus the interpretation of the numbers was that when Fatigue mood score goes up, one day later the JSE ALSI goes up, and when Fatigue mood score goes down, one day later the JSE ALSI goes down. This effect can be seen in Figure 5. For example, from July 16 to July 17 2012 the Fatigue mood score climbed from 3,06 to 3,29 and the JSE ALSI with a one-day lag climbed from 33,707 to 34,035.

FIGURE 5: RELATIONSHIP BETWEEN THE FATIGUE MOOD AND A ONE-DAY LAG OF THE JSE ALSI



The Spearman rank correlation coefficients indicated that there were two moods that had highly significant correlations to the JSE ALSI values with lag. The question whether the correlations were enough to prove the hypothesis that the moods could predict the JSE ALSI movement was still unanswered. Thus, neither hypotheses nor null hypotheses for Depression or Fatigue could be accepted or rejected at this stage. Bollen et al. (2011, p. 4) used more than only Spearman rank correlation coefficients to prove prediction; they used the Granger causality test in order to test "whether one time series has predictive information about the other or not".

The focus of the research was to find out whether the JSE ALSI can indeed be predicted by South African moods on Twitter. Four null hypotheses were accepted, but significant correlations were found for two hypotheses (H2 and H6), which required further analysis. A significant correlation was found between Depression mood (H2) and the same day JSE ALSI and Fatigue mood (H6) and the next day's JSE ALSI. Thus the next round of analysis was to test the Granger causality correlation of these two hypotheses, as was done by Bollen et al. (2011). "According to G-causality, a variable X1 'Granger causes' a variable X2 if information in the past of X1 helps predict the future of X2 with better accuracy than is possible when considering only information in the past of X2 itself" (Seth, 2010, p. 262). For the analysis, three different algorithm packages in two different applications were used. The first software package, MATLAB, implemented the Granger Causal Connectivity Analysis MATLAB Toolbox, developed by Seth (2010).

Depression mood scores and JSE data for the same day were entered into MATLAB in a matrix. The Granger causality correlation, probability and thus significance were then worked out for a one-day lag, using MATLAB. Table 5 can be read as follows: according to the Granger causality test, the variable at the top (either 1a or 2b) causes the variable on the left (either 1A or 2B), with the significance as in the table. The Granger causality test significance with which Depression (1a) caused JSE values (2B) was 0,4634, which was insignificant, because p>0,05. What was interesting to note however, was that the Granger causality correlation from 2b to 1A was indeed significant, meaning that, with a one-day lag, JSE values caused a highly significant Depression mood on Twitter, according to the Granger causality test.

TABLE 5: MATLAB DEPRESSION JSE GRANGER CAUSALITY P-VALUE

	Variable 1a (depression)	Variable 2b (JSE)
Variable 1A (depression)	NaN	0.0013
Variable 2B (JSE)	0.4634	NaN

The findings were also inspected using R version 2.15, importing two packages, namely MSBVAR (Brandt & Davis, 2012) and lmtest (Hothorn et al., 2012). The Depression mood score and JSE data were imported into R, using a comma separated value (CSV) list. After running both Granger causality tests, similar results were observed to the MATLAB tests. Although the figures were slightly different from the MATLAB tests, the interpretations were exactly the same.

For both R tests, the probabilities that Depression mood caused JSE ALSI data using the Granger causality test were insignificant, as p = 0,456139621 > 0,05. Yet again, using the Granger causality test, the results showed that there was a highly significant (p = 0,001547018) correlation, which indicated that JSE ALSI values caused Depression

mood on Twitter. The findings resulted in the null hypothesis for Depression being accepted (H20). Depression mood cannot be used to predict the JSE ALSI movement, but JSE ALSI values can cause Depression mood.

The last XPOMS hypothesis that needed to be tested was the Fatigue mood score. The matrix was entered into MATLAB, similar to the Depression matrix and the Granger causality tests were run. Table 6 shows that there was a significant Granger causality correlation coefficient between Variable 1a (Fatigue) and Variable 2B (JSE). The coefficient meant that Fatigue mood caused JSE values, using a one-day lag, according to the Granger causality test.

TABLE 6: MATLAB FATIGUE JSE GRANGER CAUSALITY P-VALUE

	Variable 1a (fatigue)	Variable 2b (JSE)
Variable 1A (fatigue)	NaN	0.6313
Variable 2B (JSE)	0.0221	NaN

The same values were again tested in R using both the MSBVAR (Brandt & Davis, 2012) and lmtest (Hothorn et al., 2012) packages. The matrix was loaded into R using a CSV containing the Fatigue mood scores and the JSE ALSI closing values. As was observed when using the Granger causality test for the Depression mood, the scores from the R software looked slightly different from the MATLAB results, but the findings implied the same results.

Using the Granger causality test, the results indicated that there was a significant ($p = 0{,}0253 < 0{,}05$) correlation between Fatigue mood and JSE ALSI values with one day's lag. The result meant that Fatigue can indeed be used to predict the following day's JSE ALSI movement. The null hypothesis (H60) for Fatigue was thus rejected and the hypothesis (H6) accepted.

The last step of the research entailed implementing a neural network for prediction and seeing whether the addition of Fatigue mood scores made predictions more accurate. A table was populated in Microsoft SQL Server 2008 R2 with the following information: (i) the Fatigue mood score for the day, (ii) the JSE ALSI value for the day, along with (iii) the previous six days' closing values, and (iv) the effect that these would have on the next day's JSE movement. For the sake of programming, the up movement effect was marked as a 1 (or true) and the down as a 0 (or false). The table was then used to train a neural network. A Microsoft SQL Server Analysis Services project was created in Visual Studio 2008, from which a Microsoft Neural Network mining structure was created. The model used the Fatigue mood score and JSE ALSI values as input columns and was trained to predict the effect column. The neural network needed to use input data to make predictions. The input data was generated using a case table, which had the same columns as the training table, except for the effect, which the mining model would predict.

The same two-table structure was created for data that did not contain the Fatigue score, to see whether the addition of the Fatigue score made a better prediction possible. Another mining model was created using the same software. The result was that the historical JSE ALSI data only (no added Fatigue mood score) could predict 23 out of the 39 movements correctly. The addition of Fatigue mood score improved the prediction to 24 correctly predicted movements out of the 39. As the results indicated, the prediction was slightly more accurate with the addition of Fatigue mood scores, albeit very slightly. The availability of more days' worth of data would improve training of the neural network, and thus the accuracy of the prediction. When the details of the neural network were inspected, Fatigue mood score did not play a major role in the algorithm; in fact a Fatigue mood score of 3,320 to 3,547 was only the fourteenth most important determinant of movement.

The researcher was aware that using the same data (except for the Effect column) for training and cases was not ideal in implementing the predictive power of the Fatigue mood score. However, only 39 days were available for training and testing, and the alternative option (which was to break this data up into training and testing groups) would render inadequate results. The use of the neural network was simply a quick test to implement Fatigue mood score and see whether the use of this data would indeed improve predictions, which it did. Future research could include testing more days of Fatigue mood scores and JSE ALSI values, which would yield more options for prediction of JSE ALSI movement.

ANALYSIS OF TWITTER MOOD (H1)

The main theme of the research was to find out whether one or more of the South African Twitter moods could be used to predict the movement of the JSE ALSI. The main theme was broken into six sub-themes, each investigating a different mood. All the null hypotheses were accepted, except for Fatigue (H6), where the hypothesis was accepted. This meant that the Twitter mood, Fatigue, could be used to predict the movement of the JSE ALSI. The results were then tested using a neural network and, although only a small improvement in accuracy of prediction was detected, the addition of the Twitter mood Fatigue did make a positive contribution towards the results. The acceptance of the Fatigue hypothesis (H6) led the researcher to also accept the main theme hypothesis, H1.

OTHER ANALYSIS

Other important analysis conducted on the data included ignoring weekend tweets between Fridays 17:00 and Sundays 17:00. The reason for this was to have the same number of valid tweets per day so that the effect of this on lag could be tested. The same significant and highly significant results were shown by the Spearman rank

correlations; however, the magnitude of the coefficient was smaller in each case. The test did not have an improved effect on lag correlations, prompting the researcher to include weekend tweets.

SUMMARY OF FINDINGS

The data passed reliability and validity tests. Four null hypotheses were accepted after applying the Spearman correlation coefficient to the data, while H20 and H6 was accepted after the Granger test as summarised in Table 7.

TABLE 7: ACCEPTANCE OF SIX MOOD HYPOTHESES THROUGH STATISTICAL TESTS

	Reliability & validity	Spearman	Granger	Neural network
Depression			$H2_0$	
Tension		$H3_0$		
Anger		$H4_0$		
Vigour		$H5_0$		
Fatigue			H6	
Confusion		$H7_0$		

The acceptance of the Fatigue hypothesis (H6) led to the acceptance of the main theme hypothesis (H1), meaning that one of the South African Twitter moods, Fatigue, can be used to predict the JSE ALSI movement.

DISCUSSION: INTERPRETING SIGNIFICANT CORRELATION

One of the major findings of the research was that there was a highly significant negative correlation between Depression mood and the same day's JSE ALSI. When the Granger causality tests were run, it was discovered that there was, in fact, a significant causality relationship between the two variables, but exactly in the opposite direction, as hypothesised. The JSE ALSI influences the Depression mood on Twitter; when the JSE ALSI goes down, Depression mood on Twitter goes up. When generalising from data to an observation, "an increase in the sample size can increase the generalizability of the sample points to a sample estimate, but does not increase the generalizability of the sample estimate to the corresponding population characteristic" (Lee & Baskerville, 2003, p. 234). Thus, although the sample size was extensive, one could not generalise that all African stock market indices would influence the Depression moods in those areas.

The results of the research were similar to the study done by Bollen et al. (2011). The mood which could best predict the stock exchange, according to Bollen et al. (2011), was Calm (which transposes to Tension from POMS), whereas the research found that in South Africa, Fatigue was a better predictor. The biggest finding of the research was the positive correlation between Fatigue mood and JSE ALSI values with a one-day lag. No literature has been found that would explain this phenomenon, but the researcher argued as follows: The Fatigue mood score indicates that many people are feeling exhausted, tired, weary or lethargic. A person generally does not feel constantly fatigued and after appropriate remedy such as sleep and/or exercise (Berger & Motl, 2000), the day after the high Fatigue score, a positive effect is noticed on the JSE ALSI value.

Hart and Webber (2005) investigated the effect of information technology (IT) infrastructure investment on the value of the firms in South Africa, where no significant market reactions were identified. Bhattacharya, Daouk, Jorgenson and Kehr (2000) conducted a similar study in Mexico. One of the possible reasons for the lack of market reactions could be that "investors do not regard news announcements as value-relevant" (Hart & Webber, 2005, p. 50). The results of this study suggest that investors do regard news as value-relevant, when looking at the reaction of Le Clos's Olympic gold medal on both the Depression mood on Twitter and on the ALSI on the JSE.

Lastly, it is important to note that stock market prediction is a research field that has to be seen in context (Malkiel, 2003, p. 72):

> Given enough time and massaging of data series it is possible to tease almost any pattern out of every data set. Moreover, the published literature is likely to be biased in favour of reporting such results. Significant effects are likely to be published in professional journals while negative results, or boring confirmations of previous findings, are relegated to the file drawer or discarded.

As advised in the comment on the limitations of the study, more than 39 days' worth of data needs to be analysed in order to consistently and accurately predict JSE ALSI movements.

CONCLUSION: CAUSAL RELATIONSHIP BETWEEN MOOD AND INDEX

MOVEMENT OBSERVED

The research question was "Could one or more of the South African Twitter moods be used to predict the movement of the JSE ALSI?" A model (XPOMS) built on an existing psychological model (POMS) was developed to extract moods from Twitter. Seven hypotheses were developed to test each mood and the main theme of the research. Four of the six moods did not have any correlations between Twitter mood and JSE ALSI values, thus the null hypotheses for these four were accepted after doing a Spearman rank correlation test on the data.

A significant negative correlation between the Depression mood and the same day JSE ALSI values were found, but after a Granger causality test, the null hypothesis was accepted, because it seemed that the JSE ALSI caused the Depression mood, and not vice versa. The main finding of the research was that there existed a highly significant positive Spearman rank correlation between Fatigue mood and JSE ALSI with one day's lag. The Granger causality test revealed that there was a significant causal relationship between Fatigue mood and JSE ALSI, meaning that Fatigue mood causes JSE ALSI movements. The neural network used for prediction showed that the prediction results were slightly better with the addition of Fatigue Twitter mood.

Researchers in developing countries, particularly those researching African stock exchanges could find this research useful, as "spillovers to individual African countries evolve" (Sugimoto, Matsuki & Yoshida, 2013, p. 1). Researchers could investigate if the findings spill over to other African countries and stock exchanges.

Methodologically, the research could be improved by the collection of five to ten years' data to use for analysis. Reflecting on the research from a substantive point of view, the results were very similar to the findings of Bollen et al. (2011). The results did indeed strengthen the claim that Twitter mood can be used to predict stock exchange movements. The results also confirm behavioural finance theory, which states that public mood can influence stock markets (Subrahmanyam, 2007). The research adds to the scientific body of knowledge by confirming previous research, the development of the XPOMS model and the development of software, which can be used by future researchers.

A recommendation for future research would be to obtain more than 39 days' worth of Twitter data to match against the JSE ALSI, as well as other African stock exchanges. Other methodologies for working out the day's mood score could be applied to the data. For example, the use of a few days before and after the measured day's mean and standard deviation of the mood scores could be used in addition to the day's own mood score. Other correlations that might yield interesting results would be to investigate the correlations between moods, giving lag between days; this could answer questions such as does Anger Twitter mood cause Depression Twitter mood, and what influence does this have on stock market predictions? Event study could also be applied with more detail, using mathematical formulae (Hart, 2006). Given Twitter's immediate nature, intraday correlations between Twitter mood and stock exchange data could also yield interesting results.

To test the researcher's XPOMS model on Bollen et al.'s (2011) data, and Bollen et al.'s (2011) GPOMS model on the researcher's data, could have added another dimension of reliability and validity to the research.

The complexity of doing big data research and the scarcity of data centres which can collect, store, analyse, and manage the data resources are an issue which needs investigation (Kahn, Higgs, Davidson, & Jones, 2014).

ACKNOWLEDGEMENT

This work is based on the research supported in part by the National Research Foundation of South Africa (Grant Number 91022).

REFERENCES

Atsalakis, G. S., Dimitrakakis, E. M., & Zopounidis, C. D. (2011). Elliott wave theory and neuro-fuzzy systems, in stock market prediction: The WASP system. *Expert Systems with Applications, 38*(8), 9196-9206.

Berger, B. G., & Motl, R. W. (2000). Exercise and mood: A selective review and synthesis of research employing the profile of mood states. *Journal of Applied Sport Psychology*, *12*(1), 69-92.

Bhattacharya, U., Daouk, H., Jorgenson, B., & Kehr, C. (2000). When an event is not an event: The curious case of an emerging market. *Journal of Financial Economics, 55*(1), 69-101.

Blas, J. (2014, 28 September). London Stock Exchange to pursue African company listings. *Financial Times*. Available from http://www.ft.com/cms/s/0/2247f5f0-4313-11e4-8a43-00144feabdc0.html#axzz3F4dXyfAv

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

Brandt, P. T., & Davis, W. R. (2012). *MSBVAR: Markov-Switching, Bayesian, vector autoregression models [online]*. Available from http://cran.r-project.org/web/packages/MSBVAR/index.html

Brown, J. D. (1997). *Skewness and kurtosis [online]*. Available from http://www.jalt.org/test/bro_1.htm

Campbell, M. H. (2011). Simple technical trading rules on the JSE Securities Exchange of South Africa, Part 2. *Proceedings of the World Congress on Engineering*. July 6-8. London, U.K.

Chen, G. M. (2011). Tweet this: A uses and gratifications perspective on how active Twitter use gratifies a need to connect with others. *Computers in Human Behaviour, 27*(2), 755-762.

Clay, R., & Keeton, G. (2011). The South African yield curve as a predictor of economic downturns: An update. *African Review of Economics and Finance, 2*(2), 167-193.

Culnan, M. J., Mchugh, P. J., & Zubillaga, J. I. (2010). How large U.S. companies can use Twitter and other social media to gain business value. *MIS Quarterly Executive, 9*(4), 243-260.

Dimpfl, T. (2011). The impact of US news on the German stock market – An event study analysis. Quarterly Review of Economics and Finance, 51(4), 389-398.

Edmans, A., Garcia, D., & Norli, O. (2007). Sports sentiment and stock returns. *Journal of Finance 62*(4), 1967-1998.

Eita, J. H. (2012). Inflation and stock market returns in South Africa. *International Business & Economics Research Journal, 11*(6), 677-686.

Ellis, C., & Faricy, C. (2011). Social policy and public opinion: How the ideological direction of spending influences public mood. *The Journal of Politics, 73*, 1095-1110.

Forgha, N.G. (2012). An investigation into the volatility and stock returns efficiency in African stock exchange markets. I*nternational Review of Business Research Papers, 8*(5), 176-190.

Fischer, E. (2011). *Language communities of Twitter [online]*. Available from http://www.flickr.com/photos/walkingsf/6277163176/in/photostream#

GET statuses/sample [online]. (2012). Available from https://dev.twitter.com/docs/api/1/get/statuses/sample

GET statuses/show/:id [online] (2012). Available from https://dev.twitter.com/docs/api/1/get/statuses/show/%3Aid

Golbeck, J., Grimes, J.M., & Rogers, A. (2010). Twitter use by the US. Congress. *Journal of the American Society for Information Science and Technology, 61*(8), 1612-1621.

Hakhverdian, A. (2012). The causal flow between public opinion and policy: Government responsiveness, leadership, or counter movement? *West European Politics, 35* (6), 1386-1406.

Hart, M. L., & Webber, R. (2005). The shareholder-wealth effects of information technology infrastructure investments in South Africa. *Investment Analysts Journal, 62*, 42-53.

Hart, M. L. (2006). Using event study methodology to assess financial impact of media announcements on companies. In D. Remenyi (Ed.): *Proceedings of the 5th European Conference on Research Methods in Business and Management*. Dublin, Ireland: Trinity College, 161-169.

Hothorn, T., Zeileis, A., Millo, G., & Mitchell, D. (2012). *lmtest: Testing linear regression models [online]*. Available from http://cran.r-project.org/web/packages/lmtest/index.html

Johnston, C. D., & Newman, B. J. (2014). Economic inequality and U.S. public policy mood across space and time (February 22, 2014). Available from *Social Science Research Network* http://dx.doi.org/10.2139/ssrn.2399927

JSE. (2015). Data from the Johannesburg Stock Exchange [JSE]. Available from https://www.jse.co.za

Kahn, M., Higgs, R., Davidson, J., & Jones, S. (2014). Research data management in South Africa: How we shape up. *Australian Academic & Research Libraries, 45*(4), 296-308.

Keller, G. (2012). *Managerial statistics*. 9th ed. China: South-Western.

Khan, M. J. (2011). The BRICS and South Africa as the gateway to Africa. *The Journal of the Southern African Institute of Mining and Metallurgy, 111*(7), 493-496.

Lee, A.S., & Baskerville, R.L. (2003). Generalizing generalizability in information systems research. *Information Systems Research, 14*(3), 221-243.

Li, Q., Wang, T., Li, P., Liu, L., Gong, Q., & Chen, Y. (2014). The effect of news and public mood on stock movements. *Information Sciences, 278,* 826-840.

Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *Journal of Economic Perspective, 17*(1), 59-82.

Malkiel, B. G. (2005). Reflections on the efficient market hypothesis: 30 years later. *The Financial Review, 40*(1), 1-9.

McNair, D. M., Lorr, M. & Droppleman, L. E. (1971). *Manual for the profile of mood states (POMS)*. San Diego, CA: Educational and Industrial Testing Service.

Mwakikagile, G. 2010. *South Africa as a multi-ethnic society*. United Kingdom: Continental Press. *Partner providers of Twitter data [online]*. Available from https://dev.twitter.com/docs/twitter-data-providers

O'Hara, M. (2003). Presidential address: Liquidity and price discovery. *The Journal of Finance, 58*(4), 1335-1354.

Pepe, A., & Bollen, J. (2008). Between conjecture and memento: Shaping a collective emotional perception of the future. *Proceedings of the Association for the Advancement of Artificial Intelligence Spring Symposium on Emotion, Personality, and Social Behavior*. March 26-28, Palo Alto, California.

Remenyi, D., Williams, B., Money, A., & Swartz, E. (1998). *Doing research in business and management: An introduction to process and method*. London: Sage.

Rios, M., & Lin, J. (2012). Distilling massive amounts of data into simple visualizations: Twitter case studies. *Proceedings of the Association for the Advancement of Artificial Intelligence Conference on Weblogs and Social Media*, June 4-7, Dublin, Ireland.

Salkind, N. J. (2004). *Statistics for people who think they hate statistics*. Thousand Oaks, USA: Sage.

Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research methods for business students*. Harlow: Prentice Hall.

Schumaker, R. P., & Chen, H. (2009). Textual analysis of stock market prediction using breaking financial news: The AZFinText system. *ACM Transactions on Information Systems, 27*(2), 12:1-12:19.

Seth, A. K. (2010). A MATLAB toolbox for Granger causal connectivity analysis. *Journal of Neuroscience Methods, 186*(2010), 262-273.

StatsSA. (2014, 31 July). Statistical release P0302: Mid-year population estimates 2014. Pretoria: South Africa: Author.

StatsSA. (2015, 24 February). Statistical release P0441: Gross domestic product, Fourth quarter 2014. Pretoria, South Africa: Author.

Stolarski, M., Matthews, G., Postek, S., Zimbardo, P.G. & Bitner, J. (2014). How we feel is a matter of time: Relationships between time perspectives and mood. *Journal of Happiness Studies, 15*(4), 809-827.

Subrahmanyam, A. (2007). Behavioural finance: A review and synthesis. *European Financial Management, 14*(1), 12-29.

Sugimoto, K., Matsuki, T., & Yoshida, Y. (2013). The spillover analysis on the African stock markets in the crises period. Available from http://www.apeaweb.orgwww.apeaweb.org/confer/osaka13/papers/Sugimoto_Kimiko.pdf

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2*, 53-55.

Twitter Translation Center [online]. Available from http://translate.twttr.com/welcome

Vermeulen, J. (2012). *Biggest social networks in South Africa [online]*. Available from http://mybroadband.co.za/news/internet/44061-biggest-social-networks-in-south-africa.html

Wang, Y., Lin, C., & Lin, J. (2012). Does weather impact the stock market? Empirical evidence in Taiwan. *Quality & Quantity, 46*(2), 695-703.

World Stock Exchanges. (2014). Stock exchanges in Africa. Available from http://www.world-stock-exchanges.net/africa.html

WordNet *[online]*. Available from http://wordnet.princeton.edu/

Zhang, Y., & Wu, L. (2009). Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network. *Expert Systems with Application, 36*(5), 8849-8854.