

# Economic News and Stock Market Correlation: A Study of the UK Market

Lee Gillam, Khurshid Ahmad, Saif Ahmad, Matthew Casey, David Cheng, Tugba Taskaya, Paulo C F de Oliveira and Pensiri Manomaisupat

Department of Computing, University of Surrey, Guildford, Surrey, GU2 7XH, UK  
{l.gillam, k.ahmad, s.ahmad, m.casey, d.cheng, t.taskaya, p.oliveira, p.manomaisupat}@surrey.ac.uk

## Abstract

In this paper, we examine the potential correlation between news and the stock market. Our goal is to match the movement of the market to events that cause the movements. The method employed attempts to produce a time-series of so-called ‘market sentiment’ from news stories that can be compared directly with time-series of stock markets and, in future, individual stocks. The results gained from a prototype based on this method indicate that news stories relate to movements in the financial market and so such a time-series is possible. We conclude that with systematic improvement of present techniques these results could be used in combination with existing financial prediction techniques, possibly improving the accuracy of such predictions.

## Background

News about financial markets has been treated in a variety of ways. Almeida, Goodhart and Payne [1] have studied the effects on the Dollar/Deutschmark exchange rate of macroeconomic announcements looking specifically at the 15 minutes following the announcement. Andersen and Bollerslev [2] and an earlier paper by Goodhart [7] have also studied these macroeconomic effects. Trading strategies based on taking strategic positions based on rumour and result have been treated by Brunnermeier [5]. Chang and Taylor [6] deal with news as ‘information arrivals’ which “are measured by the numbers of news items that appeared in the Reuters News Service”. Their approach is based on counting the number of news items in specific categories to find the most significant impacts on volatility. Chang and Taylor also provide a large literature review on the subject of information arrivals. Paradoxically, Roley and Selon [8] discuss the effects of market reactions when expected monetary policy actions do not occur – the news is that there is no news.

According to the efficient market hypothesis, the effect of news is rapidly discounted in the market. Indeed, the movement of the market according to news is ascribed, according to Almeida, Goodhart and Payne [1] as “the unexpected part of an announcement”. If, for example, a company’s profits are in line with expectations, the share price will already to some extent contain the expectation. Higher or lower profit announcements will move the price up or down as this is the reaction to the unexpected part.

Financial traders working on an intraday basis react to news events in certain ways and make an investment decision that eventually either appears consensual or contrarian. The value of the market at some future point indicates whether their approach was a success or not. Part of the challenge therefore is to determine what the ‘expected’ part of the news is and determine the

extent of the possibly unexpected part. This requires some understanding of what is termed ‘market sentiment’ – whether the market feels ‘good’ or ‘bad’ about certain things.

The goal therefore is to quantify market sentiment that is expressed in financial news. Traders talk of ‘bull’ and ‘bear’ markets; markets that are rising and will continue to rise or falling and will continue to fall. Financial reports will to some extent reflect the feeling within and about a particular market, aspects of that market such as industry sectors or specific companies, or in relation to macroeconomic factors at a given time. We believe that if this feeling can be expressed numerically, the news itself can be treated as a time-series and compared with the time series of actual market values. The conversion of textual data to numeric data enables news to be manipulated using existing time-series mechanisms; one could consider the use of Bollinger Bands [4] representing positive and negative sentiment.

Financial news may explain current economic performance or predict the state of the future economy. As such, there is a ‘past’ or ‘future’ element within such stories. The stories may contain a report of how the market performed, in which case the market affects the news or it may contain elements predicting the expected performance of a specific company in which case the market is affected by the news. Either way, there will be some correlation between news and the market. If we can determine such a correlation, further interesting research questions regarding the use of language will certainly be generated.

In this paper, we investigate the possibility of extracting ‘market sentiment’ from financial news announcements and whether market sentiment can be seen to have an effect on market movement. We evaluate this approach in terms of exceptions and the need for inferencing over

news events to determine a fuller picture which provides the basis for future work. This work is the focus of an EU co-funded project called Generic Information Decision Assistant (GIDA – IST-2000-31123).

### Market Movements

At time of writing, the financial markets have dropped about one-fifth (Dow Jones) to one-third (FTSE) of their value since peaks around December 1999. Certain drops can be seen around the time of catastrophic events such as those of September 11, which caused a lowest point around September 21 (see Figure 1) followed by a subsequent recovery to levels that were indicative of the slowly falling market.

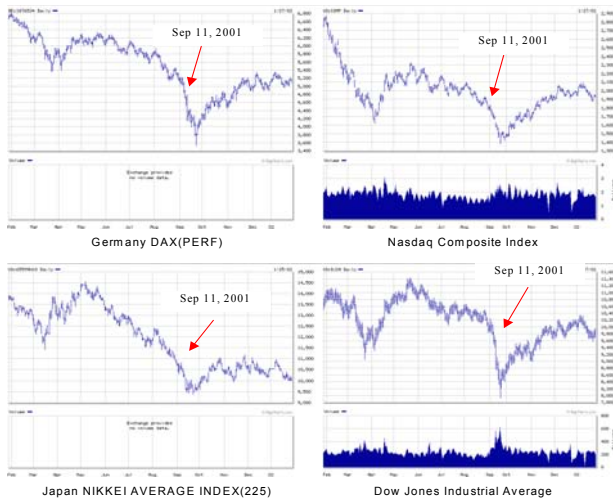


Figure 1: Movement of major market indices from Feb 2001 to Jan 2002 showing major drop following September 11

Movements of share prices for individual companies can be even more meteoric in nature. In September 1989, a Microsoft share was just under \$1, at the peak of December 1999 it was \$119.94 and trades currently around \$51. A previous favourite of the UK market, Marconi, traded at £12.76 at its peak in August 2000 and yet in July 2002 it has traded as low as 3.1 pence per share, having lost over 99% of its value. Those with impressive hindsight would suggest that it was their strategy of focussing on telecommunications and the negative sentiment regarding telecommunications at the time. Potentially, this sentiment would be evident in news regarding slowing markets and the like. An historical study would be required to confirm this.

Different ‘types’ of traders apply different strategies in selecting appropriate instruments to buy and sell. Intraday traders look for technical patterns on charts such as *Head and Shoulders*, *Double Tops* and a variety of Japanese Candlestick patterns. These traders are generally referred to as *technicalists*. On the other hand, *fundamentalists* look for the value of the company they are investing in by considering items such as the *price / earnings* ratio or the *growth* of the company.

Some traders may use various combinations of these two distinctions to evoke a trading strategy. For technicalists, the volume of arriving news makes its appropriate treatment difficult. For fundamentalists, missing an important news item is the major worry. By focussing on market sentiment, we hope to provide some indication of feeling regarding, initially, a particular market.

### Measuring Market Sentiment

As reported in, we have previously modelled market sentiment using news announcements (financial / industrial / political) in the UK along with the daily closing value of the FTSE 100. For a specific experiment, the data set was taken from November 2001 and consists of a corpus of 30 news items per day published on the Reuters News site - [www.reuters.co.uk](http://www.reuters.co.uk) between 9:00 AM and 5:30 PM and the corresponding closing value of the FTSE 100 for the day. The corpus contains time stamped business news from Britain, the United States and Europe.

We have identified over 70 terms each for conveying ‘good’ and ‘bad’ news (see Ahmad et al [1] for further details). The news stories were analysed using System Quirk (<http://www.computing.surrey.ac.uk/SystemQ>) to find the frequency of occurrences of these terms and Table 1 shows a snapshot of the results for each week under consideration.

Week	Good Word Frequency	Bad Word Frequency
1	58	40
2	71	75
3	77	66
4	73	59
5	72	28
<b>Total</b>	<b>351</b>	<b>268</b>

Table 1: Frequency of ‘Good’ and ‘Bad’ words in November 2001 news stories

For display purposes, the frequencies for ‘good’ and ‘bad’ words along with those of the FTSE 100 are then scaled according to Equation 1 below:

$$v_i = \frac{f_i}{f_{\max}}$$

Equation 1: Scaling values within data sets

The scaled values can then be plotted alongside each other using, for example, Microsoft Excel as in Figure 2. This particular plot shows some interesting initial features, however the major result is that we have a method by which we can begin to experiment with texts as time-series. The method may initially appear naïve, however we believe there is value in continuing in this vein. Certainly this approach shows no initial reasons

why continuing its development would not be worthwhile.

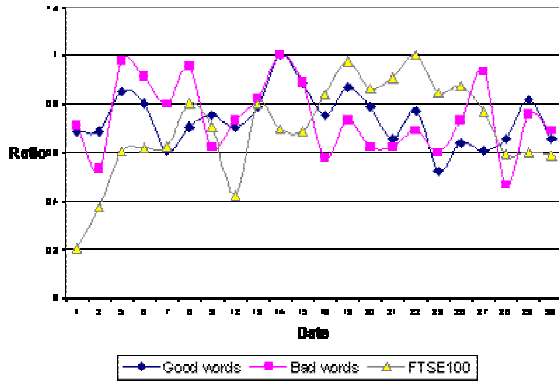


Figure 2: 'Good' and 'bad' words plotted alongside FTSE 100 values (all scaled)

Figure 3 below shows the method used to combine these data. Applying this method to the data enables potential correlations to be made between the index and the good or bad news. Furthermore, well-developed time-series analysis (TSA) algorithms can be applied over these data and may potentially be used to make predictions regarding the market itself, although this is considered future work.

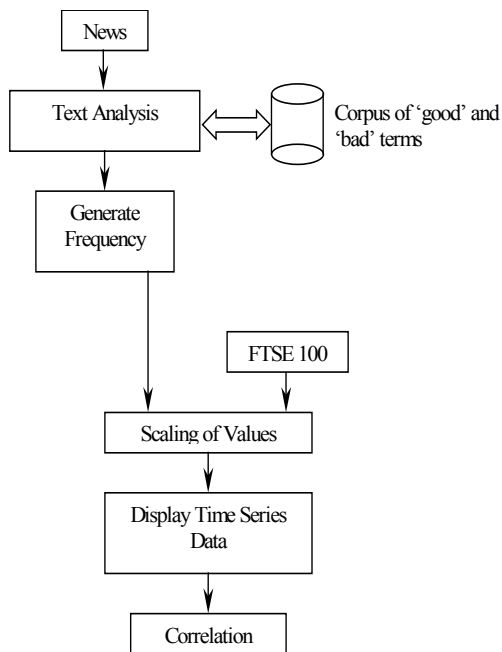


Figure 3: Method for processing news and financial time series

### Prototype of a News Time Series System

Having developed what appears on the surface to be a simplistic mechanism for converting text to a time-series of market sentiment, the next phase is the development of a prototype environment in which both the treatment of the news and the potential for combination with TSA algorithms are possible. From specific news feeds, Reuters news is published in NewsML, an application of XML. For web-based

news, HTML is common which can be converted to XHTML, another application of XML. Our prototype Sentiment and Time Series: Financial analysis system (SATISFI) therefore handles XML-based news and combines this handling with terminology extraction techniques to provide absolute frequencies for the sentiment.

SATISFI was validated using the same data set as for the previous experiment. A correlation routine has been implemented to provide an initial determination of whether the results of this analysis have some significance that is worth investigating further. Also, the capability to 'shift' a particular pattern (for example to determine whether the market is affected on a particular day by the news from the previous day) has been implemented.

Figures 4 and 5 show screenshots of SATISFI plotting the 'good' and 'bad' sentiments alongside the FTSE 100. The system can display all three results on a single chart and can provide correlations between any two.

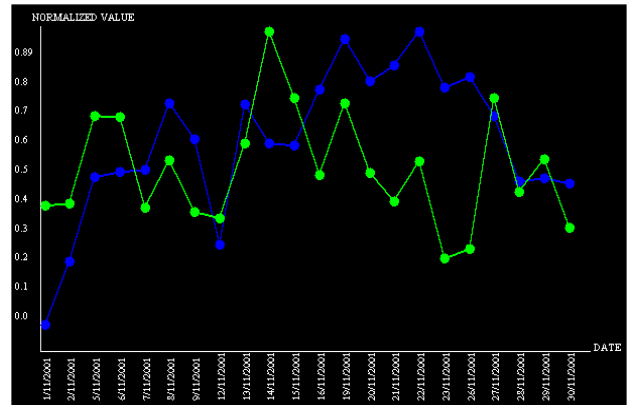


Figure 4: SATISFI Screenshot: 'Good' (Green – starts as upper line) Vs FTSE 100 (Blue – starts as lower line) for November 2001.

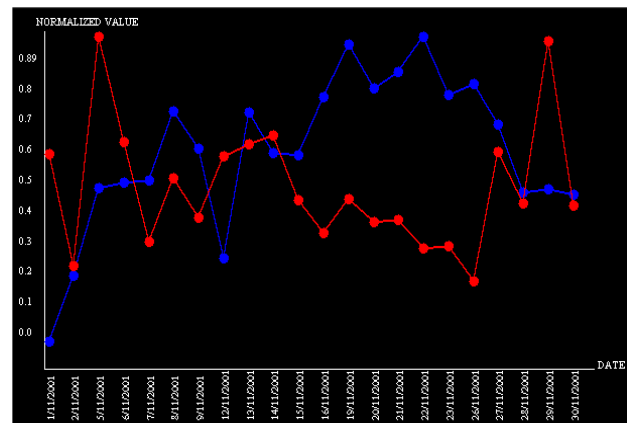


Figure 5: SATISFI Screenshot: 'Bad' (Red – starts as upper line) Vs FTSE 100 (Blue – starts as lower line) for November 2001.

Initial experiments with SATISFI appear to show that to some extent the FTSE 100 reacts to financial news.

Table 2 shows correlation results for shifted and unshifted versions of FTSE 100 with upward and downward movement indicators for November 2001. It appears that bad sentiment on a particular day anti-correlates with the FTSE, that is, 2 days after bad sentiment, the FTSE rises. This is certainly an interesting result.

	'Good'	'Bad'
FTSE	0.15	-0.31
FTSE Shifted One Day Ahead	-0.08	-0.34
FTSE Shifted Two Days Ahead	-0.48	-0.58

Table 2: Results of correlations between FTSE 100 and sentiments

### Discussion and Future Work

In this paper we examined the effects of news on the FTSE 100. To examine this further, a prototype has been developed that can produce some form of market sentiment. Although the approach appears naïve from a linguistic point of view, it does suggest initial promise for the treatment of news data as time series.

To evaluate this approach, 172 Reuters news stories for the week 4-8 February 2002 were analysed by 6 native English speakers to determine whether a particular news item was good or bad for the stock market. The same analysis was also carried out using SATISFI and the resulting correlations are shown in Table 3 below.

	FTSE Vs 'Good'	FTSE Vs 'Bad'
Manual	0.21	-0.67
SATISFI	-0.15	-0.25

Table 3: Comparative Correlation Results of News Analysed Manually and by SATISFI for February 4, 2002 to February 8, 2002.

We observed that the results from correlations produced by human analysis were better than those produced automatically. We have examined the reasons for this difference and propose to make a number of improvements in the treatment of sentiment. It is important to know the *entity* involved in the 'good' or 'bad' evaluation and how that may affect other entities – *rising inflation* is considered good using this method, however this is bad for stocks, and hence rules regarding the knock-on effects of certain events need to be considered. Information Extraction techniques may provide a valuable basis for the generation of such information. Negation and the consideration of reportage versus prediction are also issues – the phrase "not seen rising" is predictive and negates the rise. However, if this were about inflation, this would be 'good' for stocks. Double negation introduces a further complication – the phrase "wouldn't be unhappy with" exists in our corpus. We also intend to determine the 'goodness' or 'badness' of particular words and phrases, for example *surge* or *bankruptcy* are more evocative than *slight rise* or *declining*. Finally, it will be beneficial to determine events that are considered

positive or negative – despite its connotations, *restructuring* seems to have some positive aspects to it, whereas *accounting fraud* represents an extremely negative sentiment.

These improvements are the focus of ongoing work of the authors within the GIDA project, in combination with other project partners.

### References

- [1] Khurshid Ahmad, Paulo C F de Oliveira, Pensiri Manomaisupat, Matthew Casey and Tugba Taskaya "Description of Events: An Analysis of Keywords and Indexical Names" in Proceedings of the LREC2002 workshop on Event Modelling for Multilingual Document Linking
- [2] Alvaro Almeida, Charles Goodhart, and Richard Payne (1998) "The Effects of Macroeconomic News on High Frequency Exchange Rate Behaviour" in Journal of Financial and Quantitative Analysis, Vol. 33, No. 3, September 1998
- [3] Torben G Anderson and Tim Bollerslev (1996) "DM-Dollar Volatility: Intraday Activity Patterns, Macroeconomic announcements and longer run dependencies" in Journal of Finance, Vol. 53: 219-265
- [4] John Bollinger (2001) "Bollinger on Bollinger Bands" McGraw-Hill Education - Europe; ISBN: 0071373683
- [5] Markus K. Brunnermeier (1998) "Buy on Rumours - Sell on News: A Manipulative Trading Strategy" in FMG Discussion Papers from Financial Markets Group and ESRC
- [6] Yuanchen Chang and Stephen J. Taylor (2002) "Information Arrivals and Intraday Exchange Rate Volatility", Journal of International Financial Markets, Institutions and Money, forthcoming.
- [7] Charles Goodhart (1990) "'News' and the Foreign Exchange Market". LSE Financial Markets Group Discussion Paper 071
- [8] V.Vance Roley and Gordon H. Sellon Jr. (1998) "Market Reaction to Monetary Policy Nonannouncements." Federal Reserve Bank of Kansas City.