Correlating Market Movements With Sentiments: A Longitudinal Study

Nicholas Daly[1], Khurshid Ahmad[1], and Colm Kearney[2]

[1] School of Computer Science and Statistics,
Trinity College Dublin, Ireland,
{dalyni, kahmad}@cs.tcd.ie
[2] School of Business,
Trinity College Dublin, Ireland

Abstract. An adaptation of two well-established measures of changes in financial markets - *return* and *volatility* - is presented for the analysis of changes: (a) in consumer confidence about national economies, and (b) in sentiments articulated in financial news stories text and extracted automatically by counting the frequency of polarity words. This analysis is then compared and contrasted with the changes in key stock market indices - by looking at the actual distribution of returns of the sentiment, consumer confidence and stock market indices. We have looked at three stock market indices (S&P 500, ISEQ, and Nikkei) for periods of two years or more. For consumer confidence we have analysed the Michigan Consumer Confidence Index and the Irish Economic and Social Research Institute's Consumer Confidence survey. Sentiment analysis was carried out on a ten-year archive of stories published in the Irish Times (1995-2005); The distributions of each of the three returns, all the stock market indices, sentiment time series, and to a lesser extent the two consumer confidence surveys, appear not to follow the normal distribution of a random variable.

Key words: Financial News Sentiment, Stock Market Indices, Consumer Confidence

1 Introduction

There has been much research on the impact of information or 'news' on financial markets. News was traditionally interpreted as a causal variable in financial market models, and much of the early work used readily quantifiable aspects of news as a proxy for the news itself. Examples of this work include the timing of news arrival, the volume of news items and the type of news (periodically expected announcements and general publicly available news). Amongst the contributions to this literature are Mitchell and Mulherin (1994), Coval and Shumway (2001) and Antweiler and Frank (2004), who have used public information, broadcast news and internet news.

Economists and financial market experts have been dealing with the impact of 'good' and 'bad' news on market prices. This kind of sentiment analysis is

almost always conducted post-hoc wherein typically the effect of the scheduled news announcements from enterprises - about profits/losses for instance or initial public announcements-is related to the prices and traded volumes of the shares in a given enterprise (see, for examples, Bauwens, Omrane and Giot (2005), Chang and Taylor (2003), and references therein). There have been some attempts to use such an impact analysis for predictive purposes (Koppel and Shrtimberg (2004)).

Research has emerged that deals with the role of information and news in determining financial asset prices have focussed on the use of linguistic analysis. Busse and Green (2002) investigate the effects on financial markets of a second-by-second record of market analysts' opinion taken from the CNBC financial news channel. They take a recorded corpus of 20 weeks of CNBC Call programs, amounting to 322 stock reports on 84 days and subjectively assess the positive or negative sentiment of each report. They find that positive reports have a statistically and economically significant price impact for one minute following the report, while negative reports have a more gradual, longer-lasting impact over 15 minutes. In both cases, trading intensity doubles subsequent to the broadcast reports, and the price effects are discernible up to four weeks.

With the development of more robust language processing software, the latest studies have used a computational measure of sentiment in text in their analyses of news impact on the markets. Ahmad et al. (2004) have attempted to correlated the behaviour of FTSE-100 index with streaming financial news from Reuters. Davis et al. (2006) investigate the effects of optimistic or pessimistic language in firm earnings press releases on future firm performance. Davis et al. use DICTION, a content analysis system due to Hart (2000), to analyse the press release and find that optimistic or pessimistic language in earnings announcements can predict future firm earnings, with optimistic language preceding earnings increases, and pessimistic language preceding earnings decreases. Tetlock (2007) and Tetlock et al. (2008) also investigate how the affective component of text, in this case news text, can forecast market activity. Both articles analyse the relationship between a 'pessimism' factor automatically generated from the news text and market activity. Tetlock (2007) has used the General Inquirer dictionary of affect which has words tagged as positive, negative, weak, strong etc, to derive a vector of category counts for each text in their corpora. He then applies principal components analysis to reduce the dimensions of this vector to a single 'pessimism' factor which can be readily generated for any text. He takes 16 years of the daily 'Abreast of the Market' column from the *Wall Street Journal* and generates the pessimism factor for each text, creating a pessimism time-series and relates this to market pricing and trading volumes for large firms and for smaller companies represented by the *Dow Jones Index* and the *Fama-French* SMB (small-minus-big) index respectively. He reports that high pessimism values are followed by low returns for one or two days before returning to fundamentals over the following week. More generally, impact of the first three moments of news on the moments of trading volumes and returns

has been investigated by Engle and Ng (1993), Harris and Raviv (1993) and by Hutson, Kearney and Lynch (2008).

The structure of the paper is as follows: In section 2 we briefly describe the method used to determine the properties of an affect time series and how the this series can be compared and contrasted with the consumer confidence and stock market series. Data used in this paper is described in Section 3. We then show that aspects of the behaviour of an affect time series are the same as encountered in the behaviour of the other two series - in particular the moments of affect series deviate from that of a normal distribution (Section 4). Section 5 concludes the paper.

2 Method

The change in the *return* of a financial instrument is the defined as the natural logarithm of its price or volume changes on two consecutive intervals when the trading was carried; for stock prices and volumes one usually deals with closing prices on two consecutive days. The volatility is defined for the instruments as the standard deviation of returns over a reporting period- months or years for example. Return and volatility of stock market indices are used to compute risk and potential profit/loss related to the performance of enterprises included in the indices. We have extended the definition of return and volatility for two indirect and perhaps contested sources that impact on the market: sentiment and confidence in the markets.

The changes in 'market sentiment' as articulated in financial news stories is computed through a count of polarity laden words - up/down, rise/fall, boom/bust- individual news stories are normalized for the length of the stories. These words are counted independent of the context of usage: this may appear counterintuitive especially context appears paramount in language use, and in any case language use is reputedly imbued with uncertainty. However, there is considerable evidence from the analysis of political news dating back to the 1930's (Lasswell (1948), Namenwirth and Lasswell (1970)), and work in related areas of content analysis (Philip et al (1966)), indicates that a context independent frequency count does yield an insight into the polarity content of texts. More recently, the frequency count approach has been used to analyze the behaviour of financial markets (Tetlock (2007)), and to compare the treatment of different religious groups in newspaper articles (2008). The return and volatility are defined as the natural logarithm of the normalized frequency counts on two consecutive intervals of time - these may be physically contiguous times (yesterday and today for instance) or may be separated by weekends, or in some cases whenever the sentiment was published from a previous reporting. The volatility of sentiments is the standard deviation of returns.

Consumer Confidence Surveys for finding out the views of individuals on current economic conditions and future events through questionnaire surveys; the individuals may be resident of a country/region. While the survey methods differ from country to country the overall process is the same with individu-

als selected at random from varying social economic background to provide an accurate representation of the population. Individuals are asked a series of questions and then asked to response positively or negatively. These responses are aggregated to form a consumer confidence index. The index contains two subsets regarding current and future economic performances. The reporting frequency for consumer confidence indices is typically monthly (or quarterly for older data, c. 1980's). We define the return of a confidence index as the natural logarithm of the value of the index in two consecutive moths. The volatility is the standard deviation of the confidence returns.

There have been attempts to show a causal relationship between consumer sentiment and household (Carroll, Fuhrer and Wilcox, (1994)), and attempts to correlate consumer sentiment with 'recession headline news', the state of the economy and (US) Presidential popularity (Blood and Phillips (1995)).

3 Data Sources

3.1 News Sources

Irish News News sources for consideration where collected from the "Irish Times" between the years 1996-2005. Article retrieval was accomplished through the use of LexisNexis a popular search archive for various global newspapers. The "Irish Times" was selected due to its position as the leading newspaper within Ireland and as such would provide an excellent source of sentiment and commentary regarding the overall health and perception of the Irish economy. Articles concerning Ireland with the words *Economy* or *Economic* being of major mention and being relevant to Ireland. From the criteria set out a total of 9,439 articles where collected over the ten year period. While there are several other sources which provide excellent economic commentary few provide extensive coverage of the Irish economy and would provide a sparse amount of data for analysis.

US News Data for examination of the US financial sentiment was collected from the New York Times with the same criteria as that set out above. Again the New York Times was selected due to it's standing as a major US newspaper. The above criteria yielded 12,439 articles over the ten year period.

3.2 Consumer Confidence

USA: Michigan Consumer Confidence Index The Conference Board, a nonprofit business-improvement organization founded in 1916, started surveying consumer confidence in 1967. Each month out of a pool of 120,000 pre-selected, representative families, 5,000 are sent a questionnaire. About 3,500 families return the five-question survey, which asks them to rate as positive, negative, or neutral current business conditions, their expectations of conditions six months from now, their current employment situation, employment expectations in six

months, and expectations in six months for total family income. The board then comes up with a number representing consumer confidence indexed to a baseline of 100 used in 1985. The questions comprise a 'preamble' and an interrogative clause.

Ireland: ESRI's Consumer Confidence Index Initially the EU Commission Harmonised Consumer Survey launched in 1972 concentrated on only five member states with surveys carried out on a quarterly basis. The survey is now carried out in all member states on a monthly basis.

The ESRI carries out a nationally representative survey of Ireland released on a monthly basis with a minimum of 1,000 completed questionnaires. The sample design is such that the survey represents the views of people from a variety of regions of the country and also including all different labor force and educational background levels. The survey's questionnaire is designed to provide a reflection of public perception regarding the overall health of the economy, future performance and recent trends in unemployment. The ESRI CSI is carried out similarly to the Michigan survey with only some minor differences.

All-Japan Consumer Confidence Index The Monthly All-Japan Consumer Confidence Survey is conducted under the authority of the Prime Minister of Japan with an ostensible objective of understanding of three variables affecting consumer confidence: "shifts in consumer perception[s]", "price expectations", and "state of the household". Typically 6720 households are sent a questionnaire for the survey to be conducted on the 15th of every month. These households are a random sample from the universe of 47.8 households within Japan. People residing in the sampled households respond via questionnaires or face-to-face interview (CCI (2008)).

The respondents are expected to answer questions in four perception indices: two relate to variables that can be easily quantified or rendered numeric, income growth and employment; the other two categories include questions related to consumer sentiment - "overall livelihood" and willingness to buy durable goods". The respondents give their view on a graded scale (1 to 5) about the prospect of the economy over the next six months. The average of the four indices gives the All-Japan Consumer Confidence Index.

3.3 Stock Market Indices

ISEQ We use ISEQ 20 - a benchmark stock market index composed of 20 companies which trade on theIrish Stock Exchange (ISE) (Stocmhalartán na hÉireann). The 20 have the highest trading volume and market capitalization. The ISEQ Price Indices are calculated and published once per minute during the trading day (08:00 -17:15). ISEQ Total Return Indices are calculated and published at the end of the day (17.15). We use the closing prices.

S & P 500 The S&P 500, comprising the share prices of 500 of "large cap common stocks" traded in two US stock exchanges (New York and NASDAQ), is an index whose components are weighted according to the total market value of their outstanding shares. S&P 500 is regarded as a 'bellwether' for the US economy and is one of the leading indicator of the US economy. The inclusion in the index is determined by the market capitalization and 'financial viability' amongst other terms. Financial viability involves the company reporting four consecutive periods of profitable growth. The Index is produced by a committee - economists and others employees of the Standards and Poor company -(Details are available on http://www2.standardandpoors.com/spf/pdf/index/SP_500_Factsheet.pdf and other pages pointed therefrom).

Nikkei Nikkei heikin kabuki (stock prices) is a stock market index for the Tokyo Stock Exchange (TSE) published by the Tokyo newspaper Nihon Keizai Shimbun (Nikkei). It is a bellwether index for Japanese and related Asian economies. Nikkei 225 is an in index of 225 major Japanese companies. The methodology for computing Nikkei 225 is similar to that of other stock market indices mentioned above. Nikkei 225 traces the rise/fall/rise/fall of the Japanese economy since the 1970's when its value was 5000 and then leading upto an eight-fold increase to 40,000 in the 1990's and now hovers around 12,000. In the last boom period leading to Jan 2007, the index showed an appreciation of 60%, whilst FTSE 100 and Dow Jones Index only appreciated by 20 odd points; all three have shown a fall of over 20% in them last year.

3.4 A Summary of Our Data Repository

A total of nine series were used in our analysis. Varying between 2500 and 2800 data points or so for stock market indices covering (almost) daily trading between 1996-2005. We have collected over 9000 news items for sentiment analysis covering the same period economy/economic. The consumer sentiment series are usually produced monthly, and in one case (Japan CCI) the series was produced on a quarterly basis; for the monthly Michigan and ESRI consumer confidence we have over 110 points and for the Japanese CCI we have a total of 57 points. Details are given in Table 1 below.

4 Stylised Facts for Returns for Stock Market, Consumer Confidence and Sentiment Indices

A set of stylized facts are observations that are repeated in so many contexts that these observations are usually and widely 'accepted as empirical truths and set boundaries to which new hypotheses must conform' . There is one important property that appears in all our times series - the returns are not distributed according to the normal (or Gaussian) distribution. We also look at a number of statistical characteristics of the five daily time series (3 stock market indices,

Table 1. Description of time series used in this paper

	Region	Start date	End Date	Frequency	No. of returns
Stock Market Indices					
ISEQ 25	Ireland	02/01/1996	31/12/2005	Daily	2517
Nikkei 225	Japan	04-Jan-95	31-Dec-05	Daily	2770
S&P 500	USA	03-Jan-95	30-Dec-05	Daily	2770
Consumer Confidence Indices					
ESRI CSI	Ireland	Feb-96	Dec-05	Monthly	119
Michigan	USA	Jan-95	Dec-05	Monthly	131
CCI	Japan	Mar-95	Dec-05	Monthly*	57
Affect word frequency in Financial News					
Irish Times	Ireland	02/01/1996	31/12/2005	Daily	2052
New York Times	USA	02/01/1996	31/12/2005	Daily	2052

ISEQ, Nikkei, and S&P 500, and four sentiment indices extracted from articles in Irish Times and New York Times containing the key words Economic/Economy) and that of 3 monthly series of consumer confidence indicators (ESRI's Consumer Sentiment Index, Michigan's Consumer Confidence Index, and Japanese Ministry of Trade & Industry's Confidence Index).The methods used in this analysis follow closely that of Stephen Taylor's (2005) analysis of asset price dynamics and volatility. The analysis was carried on the standardized series of returns. The distribution of frequencies computed in five intervals: 0 to 0.25, 0.25 to 0.5, 0.5 to 1, 1 to 1.5, 2 to 3, and standard deviations greater than three. We note that all our series appear not to behave like they are normally distributed.

4.1 Summary Statistics

The summary of the eight returns can be summarized through the use of four moments: mean ($\hat{r}$), standard deviation (σ), skewness (skew) , and kurtosis (kurt) (Table 2).

Table 2. Summary statistics for time series of financial, consumer confidence, and sentiment returns between 1996-2005

Name	Region	$10^4 * \hat{r}$(a)	100 * σ(b)	Skew(c)	Kurt(d)	Total(n)
		Financial Returns				
S&P 500	USA	3.61	1.11	-0.11	6.81	2770
ISEQ 25	Ireland	1.8	0.48	-0.48	4.96	2517
Nikkei 225	Japan	-0.74	1.45	-0.01	6.34	2770
		Consumer Confidence Returns				
CCI	Japan	8.9	6.12	-0.6	0.42	57
Michigan	USA	-4.93	4.17	-0.05	4.45	131
ESRI CSI	Ireland	-5.61	4.92	-0.22	3.48	119
		Sentiment Returns				
Irish Times	Ireland	Return*10000	Stdev*100	Skew	Kurt	No
	Positive	-1.92	43.92	0.07	3.41	2734
	Negative	1.63	52.06	0.02	1.76	2734
New York Times	US					
	Positive	-0.79	35.89	-0.23	4.92	3376
	Negative	0.169	39.70	0.17	4.92	3376

4.2 The Shape of the Returns' Distribution

The distribution of all our returns appears not to follow the Gaussian distribution, to a greater or lesser extent, on the basis of some of the stylized facts. The work reported below was carried out on standardized daily returns, $r'=(r-\hat{r})/\sigma$, rendering a new series of zero mean and unit standard deviation.

Table 3. Average frequencies for standardized daily financial returns ($\hat{r}$)

Range	Normal (a)	ISEQ (b)	S&P (c)	Nikk (d)	Average (e)	Diff. (e-a)
0 to 0.25	19.74%	27.29%	26.43%	25.10%	**26.62%**	**6.88%**
0.25 to 0.5	18.55%	23.80%	21.23%	19.60%	**21.63%**	**3.08%**
0.5 to 1	29.98%	28.09%	28.05%	28.30%	**27.59%**	**-2.38%**
1 to 1.5	18.37%	11.56%	12.78%	15.00%	**13.08%**	**-5.29%**
1.5 to 2	8.81%	4.61%	6.17%	6.90%	**5.77%**	**-3.04%**
2 to 3	4.28%	3.34%	3.94%	4.10%	**3.95%**	**-0.33%**
3+	0.27%	1.31%	1.41%	1.00%	**1.43%**	**1.16%**

The first two rows of Table 3 show that there are more observations in the range between $\hat{r}\pm0.5\sigma$ than are expected from a normal distribution, and these results are in a very high peak. The final row ($\hat{r}\pm3\sigma$) that there are more 'extreme observations above and below three standard deviations' (Taylor (2005:71)). The kurtosis for the financial returns is in excess of three - which is the value of the normal or Gaussian distribution (Table 2).

The distribution of returns for the consumer sentiment indices also show departure from normal distribution (Table 4): There are 1% more observations

Table 4. Average frequencies for standardized monthly consumer confidence returns ($\hat{r}$)

Range			Normal(a)	CSI(b)	Uni. Mich(c)	CCI(d)	Average(e)	Diff.(e-a)
0	to	0.25	19.74%	26.89%	18.32%	15.79%	**17.06%**	**-2.69%**
0.25	to	0.5	18.55%	18.49%	25.19%	19.30%	**22.24%**	**3.69%**
0.5	to	1	29.98%	28.57%	32.06%	35.09%	**33.57%**	**3.60%**
1	to	1.5	18.37%	12.61%	14.50%	22.81%	**18.66%**	**0.29%**
1.5	to	2	8.81%	5.88%	3.05%	1.75%	**2.40%**	**-6.41%**
2	to	3	4.28%	7.56%	6.11%	5.26%	**5.69%**	**1.41%**
3+			0.27%	0.00%	0.76%	0.00%	**0.38%**	**0.11%**

that lie between $\hat{r}\pm0.5\sigma$ than are expected from a normal distribution, but when we look between one standard deviation $\hat{r}\pm\sigma$, then there are just under 5% more observations in this range than if the consumer sentiment returns were to follow normal distribution (The figure is 7% for financial returns, see Table 3 above). The final two rows ($\hat{r}\pm2\sigma$, $\hat{r}\pm3\sigma$) the presence of extreme observations here as well.

The distribution of returns for phrases carrying positive and negative sentiments (polar affect words) suggests that there more observations than expected in the range $\hat{r}\pm0.5\sigma$, and even more observations in the $\hat{r}\pm\sigma$, than are expected from the normal distribution (Table 5). There are more extreme observations when compared with a normal distribution ($\hat{r}\pm3\sigma$).It is worth noting that the frequency distributions across the New York Times and Irish Times are comparable and highly correlated.

Table 5. Average frequencies for standardised daily polar affect returns ($\hat{r}$)

Range			Normal(a)	% Pos(b)	% Neg(c)	Average(d)	Diff.(d-a)
Irish Times							
0	to	0.25	19.74%	24.07%	23.12%	**23.59%**	**3.85%**
0.25	to	0.5	18.55%	21.69%	22.13%	**21.91%**	**3.36%**
0.5	to	1	29.98%	30.32%	29.74%	**30.03%**	**0.05%**
1	to	1.5	18.37%	13.31%	13.79%	**13.55%**	**-4.82%**
1.5	to	2	8.81%	5.27%	5.30%	**5.29%**	**-3.52%**
2	to	3	4.28%	4.02%	4.65%	**4.33%**	**0.05%**
	3+		0.27%	1.32%	1.28%	**1.30%**	**1.03%**
New York Times							
0	to	0.25	19.74%	23.20%	23.79%	**23.50%**	**3.76%**
0.25	to	0.5	18.55%	21.07%	21.19%	**21.13%**	**2.58%**
0.5	to	1	29.98%	30.43%	29.54%	**29.99%**	**0.01%**
1	to	1.5	18.37%	14.37%	14.64%	**14.50%**	**-3.87%**
1.5	to	2	8.81%	5.84%	6.46%	**6.15%**	**-2.66%**
2	to	3	4.28%	4.12%	3.50%	**3.81%**	**-0.47%**
3	Plus		0.27%	0.98%	0.89%	**0.93%**	**0.66%**

4.3 Correlation Between Financial, Consumer and Sentiment series

It will be interesting to see to the extent the time series of financial returns correlate with the time series of consumer confidence returns - the correlation between the changes in the value of financial instruments and that of consumer confidence. The correlation between changes in the value of financial instruments and sentiment in newspapers is, of course, focus of our study. However, we have only conducted a longitudinal analysis of news sources , so in this respect the sentiment/financial and sentiment/consumer confidence our observations will be more tentative than is the case for financial/consumer confidence (*return*) correlation.

The consumer confidence index has much higher average yearly returns than is the case for the corresponding average returns from the stock market indices in all the three regions we have sampled - Ireland, Japan and the USA (Table 6).

Table 6. Yearly Averages of Consumer Confidence and Stock Markets Indices

	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
CSI	59	79	-22	54	78	-231	-266	172	235	-46
ISEQ	3	7	3	0.5	2	-0.1	-6	4	4	3
Irish Times Pos	-11.29	7.97	-9.26	2.76	1.00	4.33	-9.74	-23.30	12.66	-6.19
Irish Times Neg	4.32	-12.27	8.50	-13.44	5.64	-3.40	18.60	-20.88	5.68	74.12
Michigan	52	44	-13	40	-57	-86	-20	55	40	-50
S&P	7	11	9	7	-4	-6	-11	9	3	1
NY Times Pos	-14.77	-0.22	-3.31	3.22	7.25	-10.97	6.40	1.68	-8.02	0.61
NY Times Neg	1.90	15.14	-16.27	18.97	-6.47	-12.30	7.20	0.67	-7.56	-0.30

We have also analysed the volatility associated with the financial and consumer confidence returns and found the correlation between averages and volatility (the second moment or standard deviation) are different: the correlation between the averages for the financial return and consumer confidence is quite high (>50%) for all the three regions. But the correlation between volatility is less stronger (ranging between ±40% (Table 7)

Table 7. Correlation between Consumer Confidence and Market returns and volatility

Stock Market Index	Consumer Confidence	Average ($\hat{r}$)	Volatility (σ)
ISEQ	ESRI CCI	74%	-3%
S&P 500	Michigan	73%	-39%
Nikkei	Primie Minsters Office CCI	57%	41%

5 Conclusion

We believe that ours is a first of the combined studies of changes in different indices that may have an impact on the financial markets and our initial results are encouraging. We have used the computation of returns and volatilities to create dimensionless variables in three different areas - prices of financial instruments, and crafted indices of consumer confidence and that of 'market' sentiment.

There are missing values in all the series; most are daily time series, but some by necessity and design are monthly series. We have not excluded weekends and the term 'daily' is not quite accurate as there are days missing in the news and for holidays in stock market data and in the news. The narrative of the contribution of sentiments to the financial instruments and to consumer confidence (and vice versa) is even less complete than our narrative of the relationship between financial returns and that of consumer confidence. We are in the process of expanding our news sentiment corpus of texts, to include all G7 member countries and to see whether or not the sentiments' stylized facts have the same properties as the Irish and US sentiment series have.

More work will be reported and pooling our results with some of our colleagues working exclusively on s similar topic, but focusing on German markets, consumers, and sentiment articulated in German, will expand the scope of our studies and will help us in making a more convincing narrative.

References

Ahmad, K. Cheng, D., Taskaya,T., Ahmad,S., Gillam,L., Manomaisupat,P., Traboulsi, H., Hippisley,A.: 'The mood of the (financial) markets: In a corpus of words and of pictures.' In (Eds). Andrew Wilson, Dawn Archer and Paul Rayson. Corpus linguistics around the world. Amsterdam/New York: Rodopi, pp 18-32.(2006)

Ahmad, K., Gillam, L., Cheng, D,: Society Grids In (Eds.) Simon Cox and David Walker. & Proceedings of the UK e-Science All Hands Meeting 2005. 18-21 September, Nottingham UK. Swindon: EPSRC. pp 923-930. (2005)

Ahmad, K. News media and the use of polar words in emotive contexts. Edderkoppspinn eller nettverk:, Synaps, 21, 2008, p20 - 36

Antweiler, W., Frank.M.Z.: Is all that talk just noise? The information content of internet stock message boards. Journal of Finance, 59 (3), 1259-1294, (2004)

Bauwens, L., Omrane, W.B., Giot.P.: News Announcements, Market Activity and Volatility in the Euro/Dollar Foreign Exchange Market. Journal of International Money and Finance. 24, 1108-1125 (2005)

Blood, D.J., Phillips, P.C.B.: "Recession headline news, consumer sentiment, the state of the economy and presidential popularity: A time series analysis 1989-1993. Int. J. Public Opinion Research 7(1) 2-22 (1995).

Busse, J.A., Green, T.C.: Market efficiency in real-time. Journal of Financial Economics, 65:-437, (2002).

Carroll, C. D., Fuhrer,J.C., Wilcox, D.W.: "Does Consumer Sentiment Forecast Household Spending? If So, Why?". The American Economic Review. 84(5), 1397-1408. (1994)

CCI http://www.esri.cao.go.jp/en/stat/shouhi/zgshohi_kaisetu-e.html (site visited 30 Dec 2008).

Chang, Y., Taylor, S.J.: Information Arrivals and Intraday Exchange Rate Volatility. Financial Markets, Institutions and Money, 13, 85-112 (2003)

Coval, J.D., and Shumway, T.: Is sound just noise? Journal of Finance, 56(5) 1887-1910, (2001)

Davis,A.K., Piger,J.M., Sedor, L.M.: Beyond the numbers: An analysis of optimistic and pessimistic language in earnings press releases. Technical Report, Federal Reserve Bank of St Louis, (2006)

Engle, R.F., Ng, V.K.: Measuring and testing the impact of news on volatility. Journal of Finance, 48(5), 1749-1778 (1993)

Harris, M., Raviv.A.: Differences of opinion make a horse race. The Review of Financial Studies, 6(3), 473-506 (1993)

Hart. Roderick P.: DICTION 5.0: The Text-Analysis Program. Sage, Thousand Oaks, Ca., (2000)

Hutson, E.,Kearney, C.,Lynch, M.: 'Volume and skewness in international equity markets'. The Journal of Banking and Finance 32(7), 1255-1268. (2008)

Koppel, M., Shtrimberg, I.: "Good News or Bad News? Let the Market Decide" In AAAI Spring Symposium on Exploring Attitude and Affect in Text , Palo Alto CA, pp. 86-88. (Seen on http://www.cs.biu.ac.il/k̃oppel/papers/sentimentchapter-koppel.pdf 28 Dec 2008). (2004)

Lasswell, H. D. (1948). Power and personality. London:Chapman & Hall.

Mitchell, M.L., J. Harold Mulherin.J.H.: The impact of public information on the stock market. Journal of Finance, 49(3):-950, (1994)

Namenwirth, Z., and Lasswell, H. D: The changing language of American values : a computer study of selected party platforms. Beverly Hills (Calif.) : Sage Publications. (1970)

Stone, P J., Dunphy, D.C., Smith,M.S., Ogilvie, D.M., et al.: The General Inquirer: A Computer Approach to Content Analysis. Boston:The MIT Press. (For downloading the GI Lexicon, please go to http://www.wjh.harvard.edu/ĩnquirer/ ,Accessed 17th April 2008). (1966)

Taylor, S J.: Asset Price Dynamics, Volatility, and Prediction. Princeton, NJ: Princeton University Press. (2005)

Tetlock, P. C.: Giving content to investor sentiment: The role of media in the stock market. Journal of Finance, 62 (3), 1139-1168 (2007)

Tetlock, P. C., Saar-Tsechansky, M., Macskassy, S.: More than words: Quantifying language to measure firms' fundamentals. Journal of Finance 63(3), 1437-1467 (2008)