
A Computational Account of Investor Behaviour in Chinese and US Market

Zeyan Zhao^{*}, Khurshid Ahmad

School of Computer Science and Statistics, Trinity College, the University of Dublin, Dublin, Ireland

Email address:

zhaoz@scss.tcd.ie (Z. Zhao), kahmad@scss.tcd.ie (K. Ahmad)

To cite this article:

Zeyan Zhao, Khurshid Ahmad. A Computational Account of Investor Behaviour in Chinese and US Market. *International Journal of Economic Behavior and Organization*. Vol. 3, No. 6, 2015, pp. 78-84. doi: 10.11648/j.ijjebo.20150306.11

Abstract: Using vector autoregressive models (VAR) and Granger causality tests, we have looked at the impact of news sentiment on Shanghai Stock Exchange Composite (SSEC) returns based on negative sentiment (words) in newspaper texts about the Chinese economy for a period of 15 years (2000-2014, 22000 news items comprising 15 million tokens). Negative sentiment words were extracted using a well-known sentiment lexicon and a computer program based on a bag-of-words model. In addition to the negative sentiment, we have analysed the impact of traded volume and S&P 500 index: S&P (lagged) returns and negative sentiment appear to have an impact on the SSEC index.

Keywords: Time Series Analysis, GARCH(1,1), Vector Autoregressive, Granger Causality, Sentiment Analysis

1. Introduction

Investor behaviour is expected to be that of a rational person – neither unduly risk averse nor excessively risk seeking. However, it has been argued that during times of market volatility, the investor behaves irrationally and is driven by sentiment, expressed in terms of unjustified fear or blatant optimism. Sentiment analysis relies on text mining where typically a psycho-social dictionary (Stone et al, 1966[1]), and variations thereof as reported in Loughran and McDonald, 2011[2]. The dictionary is used to identify sentiment bearing phrases in a wide variety of texts – ranging from conventional media using digital channels (newspapers and wire services) to text that is available on micro-blogging sites or on social media networks, and texts available from state agencies like revenue and taxation agencies. There are systems that do not use a prescribed sentiment lexicon but rather use programs that are trained to learn that certain single and compound words appearing in situations of negative, positive or neutral evaluation (see Généreux, Poibeau, & Koppel, 2011[3]). In finance literature the estimation of the impact of sentiment, extracted from social and legacy media on daily and intra-day market returns or stock price returns, shows that the sentiment can add to the volatility of the indices and stocks (see for example Groß-Klußmann and Hautsch (2011)[4]). The texts are made available at different times and the frequency distribution of

sentiment bearing phrases over time (diachronic variation of sentiments) is rendered as a time series. The sentiment time series and stock price (returns) time series are used together in econometric models to ascertain the impact of sentiment on the economic variables. Typically one or two regression models are used in key research papers, but increasingly one sees the use of machine learning techniques. Table 1 below presents exemplar studies in sentiment analysis classified according to textual sources and the econometric models used.

A multitude of text mining systems have been employed in numerous domains to extract meaning, sentiment, and information from text. These systems typically use dictionary based lookup of terms, classification of topics, concept and phrases. Such systems that have been used previously in financial literature include the General Inquirer (Stone et al, 1966[1]), Linguistic inquiry and word count (Pennebaker et al, 2001[14]), Newscats (Mittermayer, 2004[15]), OpinionFinder (Wilson et al, 2005[16]), and AZFin text system (Schumaker and Chen, 2009[17]). These systems and other text analysis resources have been used in conjunction with quantitative and econometric models to determine causal impact of news on financial markets (see Table 1).

We have chosen to study equity market in a large economy that is representative of economies in the world, and may have been affected by the USA. China is the second largest economy, with close trading relationship with other Asian

countries, and has been impacted by the US (economy) which is its largest trading partner. We have explored Shanghai Stock Exchange Composite Index (SSEC) behaviour and the impact of news from Xinhua news agency about China in Chinese news sources for generating the proxy of sentiment. In this paper, the interaction between return and investor sentiments is estimated using vector autoregressive model (Sims, 1980[18]) and the causality relationship between markets and

investor sentiments is measured using Granger causality tests (Granger and Newbold, 1977[19]). (We find that SSEC return does not have serial correlation if we use Newey and West (1987)[20] standard errors.) The major world index, S&P 500, has significant negative leading impact on SSEC and the Chinese negative investor sentiment influences SSEC in an opposite way.

Table 1. Summary of sentiment analysis classified according to textual sources and the econometric models used – content analysis is carried out with bag-of-words (BoW) model.

Text type	Source	Econometric Model	Reference
Online messages	Message Boards	Naïve Bayes, Support Vector Machine	Antweiler and Frank (2004)[5]
		Classifier ensemble	Das and Chen (2007) [6]
		Panel Regression	Henry (2005) [7]
Corporate releases	EDGAR, Compustat	Naïve Bayes,	Li (2010) [8]
		OLS & Fama-Macbeth regression	Loughran and McDonald (2011)[2]
		Multivariate regression,	Jegadeesh and Wu (2013)[9]
Financial News	Wall Street Journal, NY Times,	VAR & OLS regression	Tetlock (2007)[10], Tetlock et al (2008)[11]
	Dow Jones News Service, News	Panel & Fama-Macbeth regression	Engelberg et al (2012)[12]
	Wires	Support Vector Regression	Schumaker et al (2012)[13]

2. Returns

China has been developing rapidly during the last three decades. With a population of 1.3 billion, Chinese economy is the 2nd largest in the world by nominal GDP and the largest by purchasing power parity according to the IMF in 2014. We have used Shanghai Stock Exchange Composite (SSEC) and this market shows a degree of volatility during

the last 15 years.

The SSEC index is a traded stock market index based on the returns of the top 50 listed companies in China by market cap. We use the daily time series of SSEC index prices over the period through 01/01/2000 to 31/12/2014. The return time series is calculated using the natural log of the ratio of price at time t and at $t - 1$ (Figure 1).

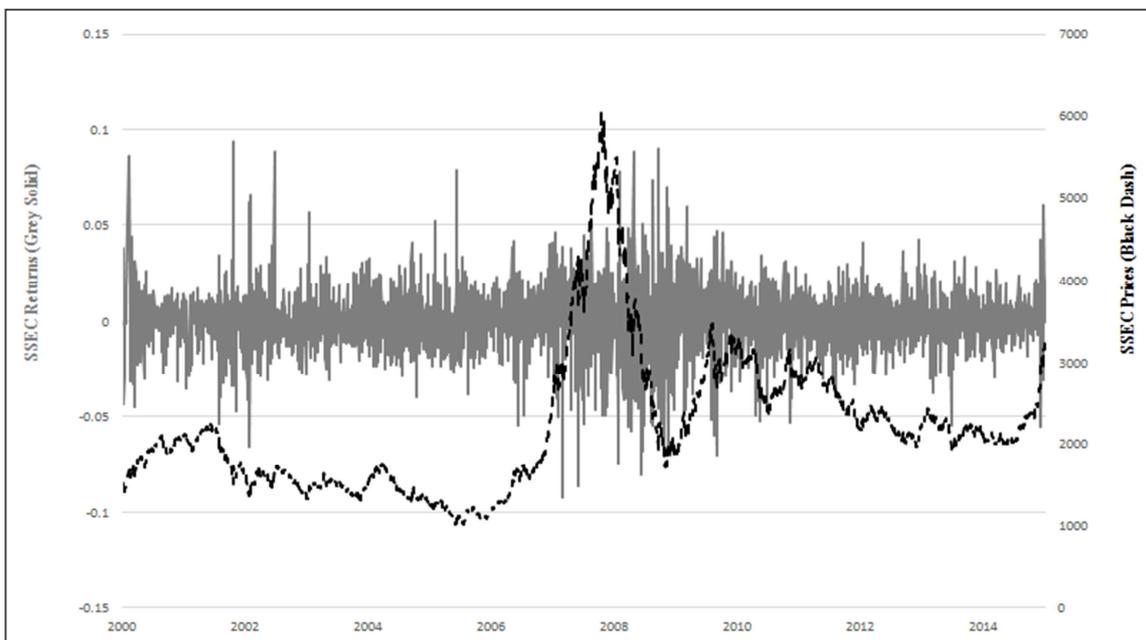


Figure 1. SSEC price and return time series. (Note the period of high volatility, 2007 – 2009).

Like most financial asset prices, SSEC returns series has a fat tail, asymmetric and aggregated normality, no serial correlation, volatility clustering and time-varying cross-correlation.

We establish the stylised facts, 5 indices and 4 firms, comprising various moments like mean, standard deviation, skewness, kurtosis, and z-statistics. The mean of the return series for S&P 500 and SSEC are close to zero (0.89×10^{-4} and 2.38×10^{-4}) and it is true of all the indices and firms in

Table 2. We also look at some representative firms' stock return series for comparison. All the firm level stock returns have slightly positive mean, higher standard deviations and non-normal distributions (Table 2).

Table 2. Summary statistics for return time series for indices: observations, mean values (10^0), standard deviations (10^2), skewness, kurtosis and z-statistics for each series.

Series	Obs.	$10^4\bar{r}$	10^2sd	Skew.	Kurt.	z
Indices						
SP500	3773	0.92	1.28	-0.18	8.02	0.44
DAX	3823	0.97	1.55	0.02	4.27	0.38
FTSE	3906	-0.15	1.22	-0.16	6.48	-0.07
NIKKEI	3693	-0.23	1.56	-0.41	6.19	-0.09
SSEC	3625	2.30	1.58	-0.09	4.35	0.88
Firms						
Alibaba	71	14.12	2.19	-0.21	-0.61	0.54
Baidu	2367	12.35	3.52	-0.15	11.61	1.71
China Telecom	3052	4.72	2.62	0.38	6.79	1.00
Sinopec	3571	6.12	2.61	0.22	5.74	1.40

The Pearson's correlation coefficients¹ measuring the degree of linear association between markets show that S&P 500, DAX and FTSE are in excess of 50% (Table 3). Nikkei and SSEC have low correlation to other indices but show relatively high correlation among each other. Commodity index and Purchasing Managers' Index (PMI) are positively correlated with market indices but the Dollar index is correlated negatively. Almost all firm level stock² returns are positively correlated with market indices (except Alibaba has negative correlation with Nikkei).

Table 3. Serial correlation amongst key market indices and with macroeconomic indicators and US currency indices (RICI – Euronext Rogers International Commodity Index, DTWEXM – Federal Reserve Trade Weighted U.S. Dollar Index, PMI – Purchasing Managers' Index).

Series	SP500	DAX	FTSE	NIKKEI	SSEC
DAX	62%				
FTSE	54%	81%			
NIKKEI	12%	26%	30%		
SSEC	4%	9%	10%	22%	
RICI	28%	28%	34%	16%	11%
DTWEXM	-12%	-12%	-17%	-10%	-9%
PMI	3%	2%	2%	3%	1%
Alibaba	42%	7%	18%	-5%	3%
Baidu	47%	33%	30%	12%	10%
China Telecom	62%	40%	40%	12%	22%
Sinopec	54%	35%	34%	15%	23%

The fact that individual return series are not

1 $\rho[X, Y] = Cov[X, Y] / \sqrt{V[X]V[Y]}$

2 Note that these Chinese companies are listed on NYSE or NASDAQ and data are also obtained there.

auto-correlated, means that return series values are independent of each other and mean reversing over time; a correlation between different indices indicate that return series are not entirely independent.

Table 4. Estimation of a GARCH(1,1) model for daily log-returns.

Series	SP500	DAX	FTSE	NIKKEI	SSEC
ω	0.000002 (0.083)	0.000002 (0.043)	0.000001 (0.251)	0.000004 (0.009)	0.000003 (0.023)
α'	0.093 (0.000)	0.091 (0.000)	0.102 (0.000)	0.102 (0.000)	0.076 (0.000)
β'	0.896 (0.000)	0.899 (0.000)	0.889 (0.000)	0.883 (0.000)	0.913 (0.000)
$\alpha' + \beta'$	0.989	0.990	0.991	0.985	0.989
log-lik.	11964.17	11232.22	12511.88	10587.07	10309.24

In simple regression models, we assume that variance of error term remains constant over time. It has been suggested that most of the key indices, firm stocks and commodities show a degree of volatility clustering: high variance during extreme periods and low variance during normal periods are observed. In order to understand the extent to which the returns are impacted by volatility, typically GARCH models are used to compute the conditional variance (h_t). We have used the so-called GARCH models for h_t (Bollerslev (1986) [21] and Taylor(1986) [22]):

$$h_t = \omega + \alpha'(r_{t-1} - \mu)^2 + \beta'h_{t-1} \quad (1)$$

The model produces h_t , the conditional variance, a one-period ahead estimate based on the past standardised return. This equation explains high volatility begetting high volatility and vice versa: α' contains the information about asset risk during the previous period and β' interprets dependency on variance during the previous period for the daily log returns. For all the five market indices in our case, α' is around 0.1 and β' close to 0.9. The persistence ($\alpha' + \beta'$) is close to, and less than, unity in all cases as return series of these indices revert to the mean value. A higher α' (or β') normally indicates the variance of returns series is dependent on its past squared returns (or conditionally dependent on its past variance). SSEC has the highest β' of the five indices, with relatively lower α' (and less significant ω).

The results of GARCH(1,1) model for the indices show that S&P 500 has the largest α' , although the persistence of S&P 500 and SSEC are same, indicating that S&P 500 is perhaps a more efficient market than the SSEC. We look at S&P 500 and SSEC, in conjunction with negative sentiment, to test the relationship between these variables.

Volatility clustering perhaps is one manifestation of irrational behaviour. And, as this behaviour is motivated by sentiment, then the lack of auto-correlation and the presence of volatility clusters provides prima facie evidence of the key role played by sentiment and justifies our investigation. Note that we are going to use auto-regressive techniques, which rely on volatility clustering and constant variance, our claims

relating to the impact of sentiment maybe weakened.

3. Sentiment Analysis

The investor sentiment is articulated through news and views that are made available to other investors, either as the report of events related to unexpected market downturn or upturn or the expectation of such events. The former is typically published in newspaper or online media as reportage and expectation is usually in opinion columns.

The more widely source of business information in China is the Xinhua news agency, which is government owned. Xinhua's output is translated into English. This translation is available daily from the news aggregator LexisNexis with frequent additions are updated during a day. We assume that Xinhua's output is an authentic account of business and finance in China.

3.1. Corpus Collection

We downloaded news articles from a business news provider "LexisNexis News and Business". "LexisNexis News and Business" annotates each news item with a large set of keywords that fall in categories like politics, technology, and business and finance. Within the business and finance category, we have searched for a cluster of terms, internal to LexisNexis but under two broad headings "Banking & Finance" and "Economy & Economic Indicators". Moreover, we set a search criteria that tokens³ "China" or "Chinese" has to occur 3 or more times in each news item in order to make sure that output news articles are primarily dealing with a Chinese event. The result of the downloading is a corpus of texts from Jan 2000 to Dec 2014 and contains over 22,000 articles with around 16 million terms (or tokens). Our corpus, on average, has 1,500 articles per annum that are published on 327 days of the year (Table 5).

Table 5. Summary of Chinese corpus from 01/01/2000 to 31/12/2014.

	Chinese Corpus	(St. dev)
Coverage (days)	4,898	
Number of articles	22,169	
Total number of tokens	15,703,055	
Number of tokens per year	1,046,801	(±491,960)
Number of tokens per day	3,206	(±2,739)

3.2. Psychological Dictionary and Investor Sentiment

We have used a bag-of-words (BoW) model for extracting investor sentiment from our text corpus. Essentially a BoW model assumes that words are distributed independently of each other in a text. A computer programme is used based on BoW model to identify individual words. These words are matched against a pre-existing list of words, e.g. in a

sentiment 'dictionary'. For analysing investor sentiment, the words in the dictionary are further classified into sentiment categories – e.g. categories of positive or negative. Every time the programme finds a word belonging to a given category, the category count is increased correspondingly. We have used a BoW programme called *Rocksteady*⁴ together with a well-established and updated affect dictionary General Inquirer developed initially by Stone et al (1966) [1].

3.3. Time Stamps in News

Xinhua General News Service is one of the major news wire service providers in China. Newspaper organisation subscribes news reports from news wire providers and re-write and publish the stories. Most informed traders do have access to news wire services however public generally do not have the same access and they normally read economic and financial section in the daily newspapers. As a consequence there is a significant gap (we take into account in the calculations) between the time the news being available from wire services and some audiences actually reading it.

We notice that the date stamp on the news from "Xinhua news agency" uploaded by LexisNexis was in New York local time⁵ however the original news was in fact published by Xinhua in Beijing local time⁶ according to Xinhua website (Figure 2). A conversion programme for time zone adaption between New York and Beijing is used in our research.

3.4. Estimation of Models

Basically, we are looking at the relationship between S&P500 and SSEC returns and their past historical values on a daily basis over a 15 year period (Jan 2000 – Dec 2014) together with the investor sentiment (proxy) during the same period. A vector autoregressive or VAR model is conducted to test the relationship between different inputs and their historical measures. The assumption of our model is that the expectation of regression residuals is independently and identically distributed (*i.i.d*). The regression of the return variable is conducted using a number of endogenous variables, including five lags of SSEC returns, traded volumes, S&P 500 returns and negative sentiments, and exogenous variables, including dummy variables of day-of-the-week and month-of-the-year effects as well as five lags of conditional volatility measure⁷. We use detrended log volume (Campbell et al, 1992 [23]) of SSEC index as the measure of traded volume. The Newey and West (1987) [20] robust standard errors are used to reduce the heteroskedasticity of residuals.

4 A sentiment analysis system developed at Trinity College Dublin.

5 <http://www.nexis.com/auth/checkbrowser.do?ipcounter=1&cookieState=0&rand=0.7635313126958804&bhcp=1> (access required)

6 http://news.xinhuanet.com/english/2015-10/28/c_134757333.htm

7 1st difference of conditional variances of GARCH(1,1) models

3 A token is an individual occurrence of a linguistic unit in speech or writing.

Multivariate time series model, namely Vector Autoregressive model, is used as our estimation technique. The errors are fitted using ordinary least squares. We also

conduct Chi-square tests and test the causality relationship between endogenous variables.

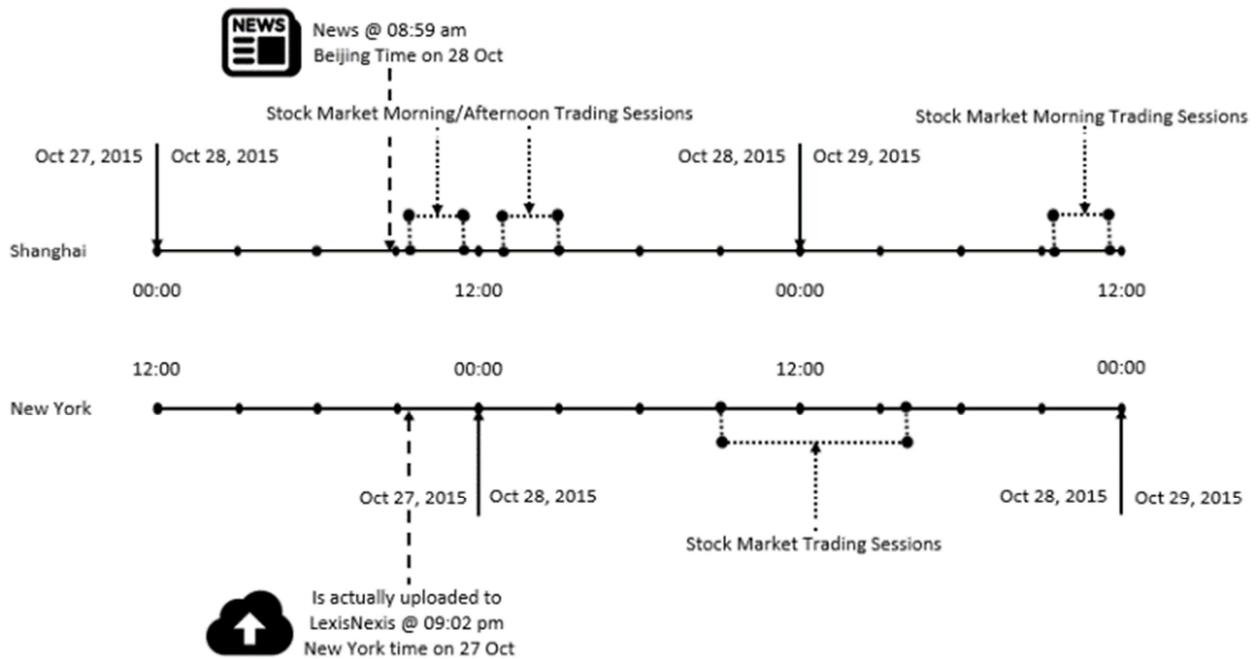


Figure 2. News timing.

The hypotheses we will test in this paper are listed below using 5th order Vector Autoregressive model with error term ε_t (L_n is the order operator, $L_1 r_t = r_{t-1}$):

Hypothesis I: whether detrended log volume has significant impact on return (2).

$$r_t^{SSEC} = \alpha + L_5 \beta r_t^{SSEC} + L_5 \gamma V_t^{SSEC} + Exog_t + \varepsilon_t \quad (2)$$

Hypothesis II: whether US index return has significant impact on return when taking into account the effect in hypothesis I (3).

$$r_t^{SSEC} = \alpha + L_5 \beta r_t^{SSEC} + L_5 \gamma V_t^{SSEC} + L_5 \delta r_t^{SP500} + Exog_t + \varepsilon_t \quad (3)$$

Hypothesis III: whether investor sentiment has significant impact on returns when taking into account the effect in hypothesis I (4).

$$r_t^{SSEC} = \alpha + L_5 \beta r_t^{SSEC} + L_5 \gamma V_t^{SSEC} + L_5 \theta Neg_t^{SSEC} + Exog_t + \varepsilon_t \quad (4)$$

Hypothesis IV: whether investor sentiment has significant impact on returns when taking into account the effect in hypothesis I and II (5).

$$r_t^{SSEC} = \alpha + L_5 \beta r_t^{SSEC} + L_5 \gamma V_t^{SSEC} + L_5 \delta r_t^{SP500} + L_5 \theta Neg_t^{SSEC} + Exog_t + \varepsilon_t \quad (5)$$

The coefficients (α , β , γ , δ , and θ) measure the sign and magnitude of impact of past values of the dependent and independent variables.

3.5. Impact of Detrended Volume, S&P 500 Index and Sentiment

We tested the hypotheses on a 15 year daily data set of the endogenous and exogenous variables associated with SSEC. We have noted that the dependence of the current value of SSEC return on its past values is statistically significant, this is because the series values are not auto-correlated (see Table 6).

Hypothesis I was not rejected as SSEC is impacted by 1st and 3rd day lag of detrended volume.

Hypothesis II was not rejected as SSEC is impacted considerably by first day lag of S&P 500.

Hypothesis III was not rejected as SSEC is impacted by 1st and 2nd day lag of sentiment.

Hypothesis IV was not rejected as SSEC is impacted by first day lag of S&P 500 and 1st and 2nd day lag of sentiment.

The results are summarised in Table 6.

We have tested the interdependence between different endogenous variables using Granger causality tests (χ^2). S&P 500 returns and SSEC negative sentiment Granger cause SSEC returns at 0.01 and 0.05 levels respectively.

4. Conclusion

This paper used SSEC and S&P 500 indices and the proxy of Chinese investor sentiment based on Xinhua General news corpus from LexisNexis and tested the impact of past investor sentiment on stock returns. We confirmed the SSEC index returns follows the general properties of return series. The 1st

day lag of US index (S&P 500) returns positively impacts SSEC returns and influences the 1st and 2nd day lags of detrended log traded volume switching their roles. Investor negative sentiment takes the responsibility of reducing the SSEC returns in one day and recovering two thirds of the

reduction in two days. If the S&P 500 returns impact is taken into account when we look at the investor sentiment impact, the impact of latter tends to be weaker. The Granger causality tests show that S&P 500 returns and Chinese negative sentiment (though weaker) Granger cause SSEC.

Table 6. Hypothesis tests of impact of S&P 500 and negative sentiment on SSEC returns and volumes: *, ** and *** denote values of coefficients' (α , β , γ , δ , and θ) statistical significance at 0.1, 0.05, and 0.01 levels respectively. All coefficients are in basis points.

		Dependent variable: r_t^{SSEC}							
	Tests	H (I)	H (II)	H (III)	H (IV)				
βr_t^{SSEC}	L_1	-194	-315	-180	-301				
	L_2	-42	31	-37	33				
	L_3	231	220	239	225				
	L_4	327	312	318	303				
	L_5	-145	-114	-154	-124				
γV_t^{SSEC}	L_1	21	**	17	*	**	17	*	
	L_2	7		11	**		10	*	
	L_3	-16	***	-17	***	-16	***	-16	***
	L_4	-6		-7		-5		-6	
	L_5	-1		0		-1		-0	
δr_t^{SP500}	L_1		1980	***			1954	***	
	L_2		296				295		
	L_3		-346				-321		
	L_4		384				363		
	L_5		51				67		
θNeg_t^{SSEC}	L_1				-9.0	**	-7.9	**	
	L_2				6.8	**	6.6	**	
	L_3				0.2		-0.2		
	L_4				1.2		0.5		
	L_5				-4.7		-3.7		
V_t^{SSEC}	$\chi^2(5)$	8.4	7.8		8.2		7.4		
r_t^{SP500}	$\chi^2(5)$		69.7	***			67.2	***	
Neg_t^{SSEC}	$\chi^2(5)$				12.1	**	9.8	*	

References

- [1] Stone, P. J., Dunphy, D. C., & Smith, M. S. with associates (1966). The General Inquirer: A Computer Approach to Content Analysis. The MIT Press, Cambridge.
- [2] Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of Finance, 66(1), 35-65.
- [3] Génereux, M., Poibea, T., & Koppel, M. (2011). Sentiment Analysis Using Automatically Labelled Financial News Items. In (Ed.) K. Ahmad. Affective Computing and Sentiment Analysis (pp. 101-114). Springer Netherlands.
- [4] Groß-Klußmann, A., & Hautsch, N. (2011). When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions. Journal of Empirical Finance, 18(2), 321-340.
- [5] Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. The Journal of Finance, 59(3), 1259-1294.
- [6] Das, S. R., & Chen, M. Y. (2007). Yahoo! for Amazon: Sentiment extraction from small talk on the web. Management Science, 53(9), 1375-1388.
- [7] Henry, P. (2005). Is the internet empowering consumers to make better decisions, or strengthening marketers' potential to persuade. Online consumer psychology: Understanding and influencing consumer behavior in the virtual world, 345-360.
- [8] Li, F. (2010). The information content of forward-looking statements in corporate filings—A naïve Bayesian machine learning approach. Journal of Accounting Research, 48(5), 1049-1102.
- [9] Jegadeesh, N., & Wu, D. (2013). Word power: A new approach for content analysis. Journal of Financial Economics, 110(3), 712-729.
- [10] Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. The Journal of Finance, 62(3), 1139-1168.
- [11] Tetlock, P. C., SAAR-TSECHANSKY, M. A. Y. T. A. L., & MacKassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. The Journal of Finance, 63(3), 1437-1467.
- [12] Engelberg, J. E., Reed, A. V., & Ringgenberg, M. C. (2012). How are shorts informed? Short sellers, news, and information processing. Journal of Financial Economics, 105(2), 260-278.
- [13] Schumaker, R. P., Zhang, Y., Huang, C. N., & Chen, H. (2012). Evaluating sentiment in financial news articles. Decision Support Systems, 53(3), 458-464.

- [14] Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count: LIWC 2001*. Mahway: Lawrence Erlbaum Associates, 71, 2001.
- [15] Mittermayer, M. A. (2004, January). Forecasting intraday stock price trends with text mining techniques. In *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on* (pp. 10-pp). IEEE.
- [16] Wilson, T., Hoffmann, P., Somasundaran, S., Kessler, J., Wiebe, J., Choi, Y., & Patwardhan, S. (2005, October). OpinionFinder: A system for subjectivity analysis. In *Proceedings of hlt/emnlp on interactive demonstrations* (pp. 34-35). Association for Computational Linguistics.
- [17] Schumaker, R. P., & Chen, H. (2009). Textual analysis of stock market prediction using breaking financial news: The AZFin text system. *ACM Transactions on Information Systems (TOIS)*, 27(2), 12.
- [18] Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, 1-48.
- [19] Granger, C. W. J., & Newbold, P. (1977). *Forecasting economic time series*. Academic Press.
- [20] Newey, W. K., & West, K. D. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, 777-787.
- [21] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3), 307-327.
- [22] Taylor, S. J. (1986). *Modelling financial time series*. Wiley, New York.
- [23] Campbell, J. Y., Grossman, S. J., & Wang, J. (1992). Trading volume and serial correlation in stock returns (No. w4193). National Bureau of Economic Research.