

## THE IMPACT OF NEWS ON STOCK MARKET: QUANTIFYING THE CONTENT OF INTERNET-BASED FINANCIAL NEWS

Xiqian Zhao, Juan Yang, Lili Zhao and Qing Li  
Southwestern University of Finance & Economics, Chengdu, P.R. China  
liq\_t@swufe.edu.cn

### ABSTRACT

In this article, we propose an approach to quantify the content of financial news and investigate its impact on the Chinese stock market. Our three main findings are: (1) Internet-based financial news tends to have a negative impact on the movements of Chinese stock market; (2) Firms with Shenzhen Stock Exchange are more affected by the Internet-based financial news than those of Shanghai Stock Exchange; (3) The impact of news on stock market is essentially determined by their contents. With these findings, we believe that Internet-based financial news contains hard-to-quantify information of firms, which investors could incorporate into stock prices timely.

**Keywords:** financial news, stock market, quantified language, text mining.

### INTRODUCTION

The movements of stock market are determined by enormous factors including overall economy, inflation, trading strategies, return on equity (ROE), market sentiment, and firm itself. Intuitively, news articles reporting the fundamentals of firms or economics should affect the stock market to some degree. The impact of quantitative information on stock market prices has been studied by a voluminous literature. Although quantitative data has been successfully applied to stock market forecast, there are evidences that substantial stock market movements do not seem to correspond to the changes in quantitative measures of firms' fundamentals [1]. This indeed suggests qualitative variables could be considered to complement quantitative measures to fully assess firm valuation. In particular, the linguistic descriptions of firms, especially online news articles, capture some aspects of firms' fundamentals. Such linguistic information could be utilized to overcome the incompleteness or biased measures of firms' fundamentals of analyst and publicly disclosed accounting variables.

With the advent of Web techniques, the Internet-based news media becomes a major vehicle to discover and disseminate the information. The timely and broad dissemination mechanism essentially makes it a potentially important source of information about firms' fundamental values. However, few researchers have studied the impact of news articles on stock prices, let alone the Internet-based news. This is partially because of the difficulty of transferring qualitative information into numerical data for further econometric analysis. When studying the connection between news and stock market, previous studies limit on the number or title of news articles. Some researchers go even further by utilizing the number of positive and negative words

in the news. We argue that such crude quantitative measures could not fully reflect the linguistic power on firms' fundamental values. In this article, we propose a sophisticated measure, i.e. support vector regression (SVR), to quantify the Internet-based News, and explore its impact on the daily returns of the relevant stocks. Here, we are particularly interested in 1) the concrete evidence to demonstrate the interaction between the Internet-based financial news and their relevant firms in Chinese stock market; 2) which kind of firm is more sensitive to the online news; 3) which type of financial news is more influential.

The rest of this work is organized as follows. We first briefly describe relevant studies and background techniques in the next section. Then we present the design details for analyzing the impact of the Internet-based news articles in Chinese stock market. In the following part, we conduct experimental studies. At last, this article is concluded with speculation on how the current work can be further improved.

### LITERATURE REVIEW

Since Keynes brought up the concept of "animal spirits" in 1930's, a large number of researchers are trying to understand the determinants of wild movements in stock market prices. Observing the fluctuation of stock prices accompanied by the news publishing, researchers have devoted substantial attention to explore the power of verbal information on stock market. The earliest research can be traced back to the work of Niederhoffer [2]. It was the first report that the movements of stock market are influenced by the news related with world events in the New York Times. Keown and Pinkerton [3] study the change of stock returns caused by the release of news on firm acquisition. Mitchell and Mulherin [4] point out that the number of Dow Jones announcements has a direct impact on several kinds of securities market activities including trading volume and market returns. These earlier findings reveal the connection of news and movements of stock markets.

Later on, more endeavors are devoted to finding out the influential patterns of news on stock market. For example, Cutler et al. [5] find out that one-third of variance in stock returns is brought by the news articles. Chan [6] examines the monthly stock returns of firms affected by the headlines of news containing these firms. This study shows that bad news reports result in a negative drift of the stock market up to 12 months. Fang and Peress [7] study the connections between the average stock return of certain firms and their media coverage in a period, and find out there is a significant return premium on stock with no media coverage. Tetlock [1] [8] use news articles from the Wall Street Journal to measure the interaction between the media and the stock market. He finds the high media pessimism predicts downward pressure on market prices followed by a reversion to fundamentals, and

unusually high or low pessimism predicts high market trading volume. Takeda and Yamazaki [9] take a further step by studying the relations between the stock price of a firm and the exposure of a firm in a NHK TV program. One interesting finding of this study is that a simple report of a firm's old story still has a potentially positive impact on its stock price.

All of these studies reveal that the release of news reports on a firm tends to affect its stock market price to a certain degree. These researches focus on understanding the power of linguistic information on stock market. However, the way to quantifying the news articles in these studies is rather crude. Existing techniques can be categorized into two sub-groups. One is counting the times that a firm is mentioned in the headline or news body. The other is counting the words carrying the tone information (bad or good words) in news articles related with a particular firm. Such simple methods only partially capture firm fundamentals revealed in the news, and consequentially weaken or distort the impact of news on stock market analysis.

To overcome this problem, computer scientists apply text mining and machine learning techniques to quantify verbal information and predict its impact on stock prices. Since Fledman and Dagan [10] combine the Knowledge Discovery in Databases (KDD) and text categorization paradigms, a large number of economists have devoted substantial attention to exploring the power of textual information on stock market. For instance, Wiithrich et al. [11] predict stock movement using textual information contained in news articles published on the Web. Several techniques, such as rule-based, k-NN algorithm and neural network, have been employed to produce the forecasts. Antweiler and Frank [12] extract messages from the Yahoo! Finance (YF) and Raging Bull (RB). They utilize Naïve Bayes and Support Vector Machine algorithms to code the individual message as bullish, bearish, or neither, and then find that stock messages can help predict market volatility. Schumaker and Chen [13] examine the role of financial news articles on three different textual representations: Bag of Words, Noun Phrases, and Named Entities and analyze their ability to predict stock prices twenty minutes after an article release. Lagreid and Sandal [14] employ text mining techniques to create a framework that enables extraction of entities (companies, people and other concepts) from Web-based business news.

However, they ignore the influence of the other key stock performance indicators, such as earning-to-price, share turnover, and price-to-equity on the stock market. These quantitative variables can complement qualitative measures to fully assess firm valuation. And it is rare to find the research about the relationship between news articles and stock market in China.

In this paper, we are dedicated to investigating the relation between Internet financial news articles and the stock returns in Chinese stock market, which considers both the linguistic information and quantitative variables. However, the detail of the proposed methodology is described in the next section.

#### DETAILS OF METHODOLOGY DESIGN

In this section, we describe the details of our methodology design to quantify the content of Internet-based financial

news and the method to analyze its impact on Chinese stock market. First, we briefly introduce the data sets for our analysis. We then present the way to quantify the content of online news. Last, we describe our econometric analysis method for studying the impact of news on stock market.

#### Data

Chinese stock market consists of two parts, Shanghai and Shenzhen Stock Exchange. We select 85 firms related with information technology from both stock exchanges. The timeframe to analyze the movements of stock market is between August 1, 2010 and September 30, 2010. The volatility of the market index within this period is relative stable so that we can avoid the wild influence of economic fluctuation. To explore the impact of online financial news on stock market, we construct two data sets. One is the stock trading data, and the other is a synthetic data set containing the Internet-based financial news articles related with the above 85 selected firms. Details of the data sets are:

(1) Stock trading data set. This data set contains the financial information of the selected firms, which is collected from RESSET Financial Research Database ([www.resset.cn](http://www.resset.cn)). We use the Shanghai Composite Index and the Shenzhen Component Index as the market index. It is noted that the stock market index does not refer to the bonus reinvestment, so we only adjust the trading data without considering it. In this way, the trading data can be in accordance with the market index [15].

(2) Internet-based news data set. We use a focused crawler to automatically collect a bundle of news articles related with the selected firms within the study period. Specifically, 4448 news articles are collected from several important financial news portal including Hexun.com, JRJ.com and Finance.Sina.com. Then we exclude stories in the first week after a firm has been newly added to the index to prevent the well-known price increase associated with a firm's inclusion in the Shanghai Composite Index or Shenzhen Component Index [16]. News articles which are released during the twenty minutes before the closure of the market are regarded as the news in the next day [13]. Similarly, if the news articles are released in the non-market days, they are account to the next day's data. For example, the news on weekends is treated as the data of next Monday.

#### Quantifying the Content of News

Financial news contains lots of information which can affect economic activities. To investigate the impact of financial news on stock market, we first use text mining technique to convert the unstructured news document to structured vector. Each document (news) is treated as a term vector where term weight is valued by the standard TF\*IDF definition [17]. With such transformation, we can apply SVR (Support Vector Regression) technique to analyze the relation between the news and stock market.

Most traditional machine learning techniques, such as Neural Network, embody Empirical Risk Minimization (ERM) principle which minimizes the error in the training data. However, these traditional approaches can result in overfitting to the training data. According to Gunn [18], SVR is based on Structure Risk Minimization (SRM) principle, which minimizes an upper bound on the expected risk. It can

build a much more accurate and robust regression model to solve overfitting perfectly. That's the reason why we choose SVR to measure the impact of news articles on stock market. In applying the SVR technology, we consider the news documents vectors as the input of SVR, and the corresponding stocks returns as the output. Meanwhile, we divide the news dataset into training data set (2921 news articles) and testing data set (1527 news articles). With the training data set, we build our regression model. And then we use testing data set to evaluate the SVR performance. Finally, we use the final SVR model to quantify the content of Internet-based news articles and get the news indicator (*news*), so as to analyze the impact of news stories on stock market.

### Econometric Analysis Method

Here, we introduce two ways—event study methodology and cross-sectional regression method to analyze the impact of Internet-based news articles on stock returns respectively. (1) Event study methodology. In this paper, we use the standard event study methodology as described by MacKinlay [19]. We define the date when a piece of news is released in the Internet media as the event day. We then choose the event windows that are the period over which stock return react to the event. We define *t* as the event day. We employ a 10-day event window, denoted by [*t*-2, *t*+7]. *t*-2 means two days before the event day. *t*+7 means the seventh day after the event day. We set the estimation window at 180 transaction days prior to the event window. The timing sequence is illustrated with a time line in Figure 1.

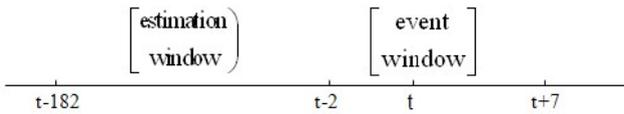


Figure 1 Time line for an event study

We adopt the Capital Asset Pricing Model (CAPM) as our market model. The CAPM due to Sharpe [20] and Lintner [21] is an equilibrium theory where the expected return of a given asset is determined by its covariance with the market portfolio. The following CAPM market model was estimated for each news-related firm.

$$r_{i,t} = r_{f,t} + \beta_i(r_{M,t} - r_{f,t}) + \varepsilon_{i,t} \quad (1)$$

Where  $\varepsilon_{i,t}$  is the zero mean disturbance term.  $r_{i,t}$  is the anticipated normal return for the stock of firm *i* in period *t*.  $r_{f,t}$  is non-risk interest rate.  $r_{M,t}$  is the return of portfolio, which we respectively use Shanghai comprehensive index and Shenzhen component index to present. We use the Ordinary Least Square (OLS) method to estimate parameter  $\beta_i$ . Then we calculate the abnormal return for the stock of firm *i* in period *t*.

$$ar_{i,t} = r_{i,t} - \hat{r}_{i,t} \quad (2)$$

$ar_{i,t}$  is the abnormal return for the stock of firm *i* in period *t*.

$\hat{r}_{i,t}$  is the actual return for the stock of firm *i* in period *t*.

Then we calculate the average abnormal return and the cumulative average abnormal return in equation (3) and (4), which are respectively denoted by  $AAR_t$  and  $CAR_t$ .

$$AAR_t = \frac{1}{n} \sum_{i=1}^n ar_{i,t} \quad (3)$$

$$CAR_t = \sum_{\tau=1}^n AAR_{\tau} \quad (4)$$

(2) Cross-sectional analysis of abnormal returns. In this section we provide a multivariate analysis of abnormal returns. Specifically, we explore whether the observed abnormal returns are affected by various factors that describe the firm characteristics as well as overall market conditions. The following is the multivariate regression model.

$$Car_i = a + b_1 news_i + b_2 \ln Size_i + b_3 Trnv_i + b_4 EP_i + b_5 PB_i + \varepsilon \quad (5)$$

Where  $Car_i$  is the cumulative abnormal return of the news-related firm *i* from *t*-2 to *t*+*j*. *news* is the qualified indicator of news articles calculated by SVR, which is used to explore the relation between news and stock market. *lnSize* is the nature log of the firm scale which is presented by the firm's total share capital. *Trnv* is the daily turnover rate, which is a very important indicator of liquidity. *EP* is the earnings-price ratio. And *PB* is the price-equity ratio.

## EXPERIMENTAL STUDY

### Event study of the stock market

In order to study if the news events have an impact on the stock market, we test the simple relationship using the event study method. In Table 1, we discuss the interaction between the news events and the whole stock market that includes the Shanghai and Shenzhen stock market. To well understand the trends of stock price, here is a graphical description about the AAR & CAR for various windows surrounding the event day in Figure 2. In a nutshell, it is observed that stock price decreases as time goes on until the fifth day after the news events. This is quite understandable that the effect of a news stories on stock price is negative and strong during the first several days but to be gradually attenuated with the time passing by. However, there are two peaks in the curve of CAR.

- The first peak occurs on day *t*, the event day, when the news is released. Here, the AAR is -0.0048 and the CAR reaches -0.0089. They are both very significant at the statistical level of 1%, which indicates a strong and immediate impact on stock price. We can easily understand that on the news event day, all of the investors have acquired the information, so it results in an immediate change of the stock price.
- The second peak occurs on day *t*+5. On that day the stock price drops to the lowest level. Meanwhile, we can observe that there is a relatively continuous and particular change from day *t*+2 to day *t*+5, and the CAR changes from -0.0106 to a minimum of -0.0162. Since it is a long time after the news, the investors might have just been brought eyeballs to the other issues so that the influence is almost to the ends. Therefore, the stock price is gradually going back to its normal status.

So, it actually reveals that most firms' stock performance is affected by the news reports from the internet media. In addition, we can easily discover that AAR & CAR are lower

than zero, it reflect the news stories has a negative impact on the stock price, no matter how positive and negative the news articles are. This is different from the study of Tetlock [1].

TABLE I. THE SIMPLE INTERACTION BETWEEN THE NEWS AND STOCK MARKET

	t-2	t-1	t	t+1	t+2
AAR	-0.0004	-0.0036***	-0.0048***	0.0000	-0.0017***
CAR	-0.0004	-0.0040**	-0.0089***	-0.0089	-0.0106
	t+3	t+4	t+5	t+6	t+7
AAR	0.0019***	-0.0022***	-0.0015**	0.0014	-0.0012**
CAR	-0.0125***	-0.0147***	-0.0162***	-0.0147	-0.0159***

Notes: The different asterisk (\*, \*\*, \*\*\*) denotes stationarity on the basis of each test at the 10%, 5%, and 1% statistical level respectively, similarly hereinafter.

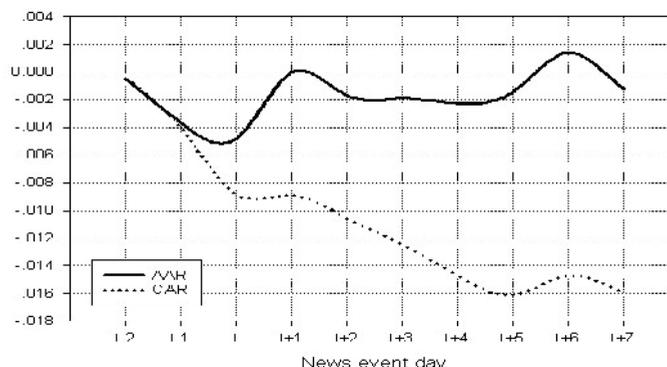


FIGURE II. TRENDS OF THE STOCK PRICE WITH NEWS EVENTS

TABLE II. THE SIMPLE INTERACTION BETWEEN THE NEWS AND SHANGHAI STOCK MARKET

	t-2	t-1	t	t+1	t+2
AAR	0.0007	-0.0034***	-0.0034**	0.0008	-0.0007
CAR	0.0007	-0.0027	-0.0061**	-0.0069**	-0.0076*
	t+3	t+4	t+5	t+6	t+7
AAR	0.0045***	-0.0031***	-0.0011	0.0010	-0.0022**
CAR	-0.0121***	-0.0152***	-0.0162***	-0.0152***	-0.0174***

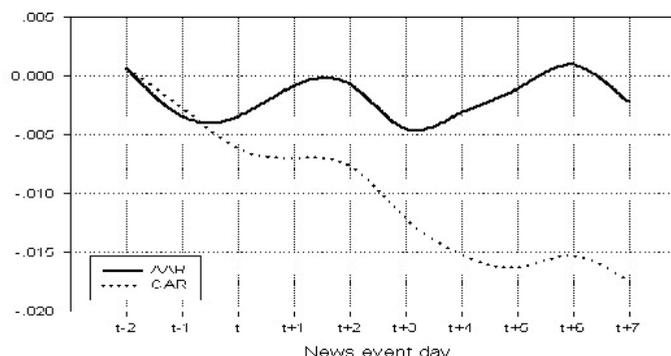


FIGURE III. TRENDS OF THE STOCK PRICE WITH NEWS EVENTS ON SHANGHAI STOCK MARKET

It is known that the firms which are listed in Shanghai Stock Exchange are mostly the state-owned big enterprises, while the other firms listed in Shenzhen Stock Exchange are private-owned and small. Due to the difference about the firms in each stock market, we explore the impact of news articles on different stock market respectively in table 2 (figure 3) and table 3 (figure 4).

● Event study of Shanghai stock market.

When the news stories are reported, the stock has a negative movement on day t quickly. Then it lasts several days because that the t-statistical test of AAR & CAR are of great

significance in three days: day t, day t+3, and day t+4. The CAR reaches a maximum of -0.0152 on the fourth day after the news event day.

● Event study of Shenzhen stock market.

We can see a different exploration of the interaction in Shenzhen stock market from table 3 and figure 4 using the standard event study method. The influential duration is also about five days with a negative impact. On the event day, the AAR is -0.0062 and CAR is -0.115, both having a significance on 1% level. Furthermore, there is another three responses to the news articles after the event which almost as

the same as that in Shanghai stock market. But the impact lasts only five days due to the non-significance of AAR &

CAR in the following days (day t+6, day t+7) in table 3.

TABLE III. THE SIMPLE INTERACTION BETWEEN THE NEWS AND SHENZHEN STOCK MARKET

	t-2	t-1	t	t+1	t+2
AAR	-0.0015	-0.0038	-0.0062***	0.0007	-0.0026***
CAR	-0.0015	-0.0053	-0.0115***	-0.0108	-0.0134***
	t+3	t+4	t+5	t+6	t+7
AAR	0.0005	-0.0015***	-0.0018**	0.0019	-0.0001
CAR	-0.0129	-0.0143**	-0.0161***	-0.0143	-0.0145

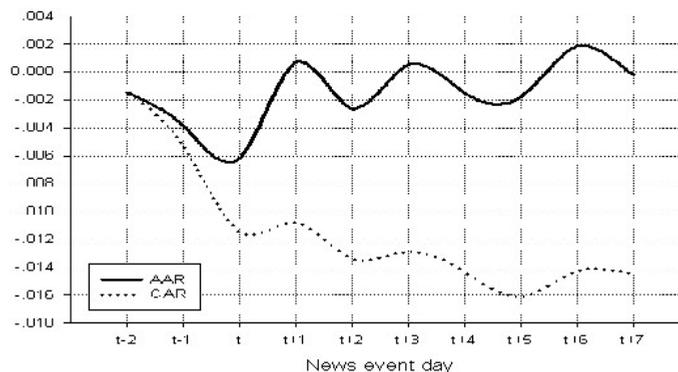


FIGURE IV. TRENDS OF THE STOCK PRICE WITH NEWS EVENTS ON SHENZHEN STOCK MARKET

### Cross-sectional analysis of Shanghai Stock Market

The results of the cross-sectional analysis in Shanghai stock market are presented in table 4. We examine the heteroskedasticity using the White's test [22]. The robust t-statistics of the regression are displayed in parentheses. The coefficients of the *news* are significant, particularly from t+1 to t+4, which can indicate that the news articles have a great influence on the stock returns. However, the impact is gradually attenuated as time goes on because of the reduction of t-statistical level. Also, it is obvious that the indicators about firm scale (*lnSize*) and turnover rate (*Trnv*) have high

level of significance. It reveals that the investors have an enthusiastic response when the news is released. To do the further analysis, we can find the model performs excellence on t+1, for the adjusted R-squared reaches up to 0.441. In addition, the firm scale (*lnSize*) is less significant from t to t+2, but turnover rate (*Trnv*) is significant in these days. We can conclude that whatever the firm scale is, news articles always induce strong influence on stock. In the following days, the firm scale (*lnSize*) becomes more and more significant. It means there is different impact on the firms, whose scales are different.

TABLE IV. THE RESULTS OF MULTIVARIATE REGRESSION MODEL FOR SHANGHAI STOCK MARKET

	a	b1(news)	b2(lnSize)	b3(Trnv)	b4(EP)	b5(PB)	adj.R <sup>2</sup>
t	0.065 (-1.038)	-0.091* (-1.743)	-0.002 (-0.549)	-0.009*** (5.761)	-0.948** (-2.536)	0.0009 (-0.643)	0.329
t+1	-0.098 (-1.038)	-0.090*** (-3.524)	0.005 (1.350)	-0.007*** (-4.371)	-1.388*** (-3.340)	-0.0005 (-0.205)	0.441
t+2	-0.047 (-0.390)	-0.085** (-2.423)	0.003 (0.683)	-0.009*** (-4.0002)	-1.579*** (-2.893)	-0.0001 (-0.450)	0.297
t+3	-0.198 (-1.472)	-0.109** (-2.456)	0.009* (1.675)	-0.008*** (-3.223)	-1.978*** (-3.191)	0.0022 (0.928)	0.282
t+4	-0.293 (-1.994)	-0.103** (-2.215)	0.014** (2.366)	-0.009*** (-3.32)	-2.649*** (-3.991)	-0.0011 (-0.359)	0.363
t+5	-0.396 (-2.639)	-0.093* (-1.960)	0.018*** (2.990)	-0.006** (-2.424)	-2.841*** (-4.184)	0.0002 (0.057)	0.336
t+6	-0.520 (-3.245)	-0.108** (-2.125)	0.023*** (3.485)	-0.003 (-1.264)	-3.047*** (-4.122)	0.0025 (0.895)	0.288
t+7	-0.589 (-3.688)	-0.090* (-1.775)	0.026*** (3.980)	-0.004 (-1.563)	-3.411*** (-4.588)	0.0022 (0.796)	0.338

### Cross-sectional analysis of Shenzhen Stock Market

Table 5 shows the result of cross-sectional analysis in Shenzhen Stock Market. From the table, we can see the coefficients of the *news* are significant and negative as well. The conclusion is the same as that of Shanghai Stock Market. Then, we can discover the absolute coefficient of the *news* is 0.235 on the day *t*, which is the largest absolute value among all those days. And it is decreasing as time goes by. So, we can draw a conclusion that the strongest impact of news articles on stock happens on the date of *t*, on which day news

articles are released. The influence lasts from *t* to *t*+7, which is longer than that of Shanghai stock market.

At the same time, we find the coefficients of turnover rate (*Trnv*) are also very significant during the event window. However, the significance of firm scale (*lnSize*) is not so obvious. Thus, we can draw the conclusion that no matter what scale the firm may be, there is always strong influence of news articles on stock. The relationship between the impact and firm scale (*lnSize*) is not obvious, even though we add the news indicator. This is similar to the study of Wei Wang [23].

TABLE V. THE RESULTS OF MULTIVARIATE REGRESSION MODEL FOR SHENZHEN STOCK MARKET

	a	b1( <i>news</i> )	b2( <i>lnSize</i> )	b3( <i>Trnv</i> )	b4( <i>EP</i> )	b5( <i>PB</i> )	adj.R <sup>2</sup>
t	0.029 (1.236)	-0.235*** (-4.222)	0.001 (1.267)	-0.011*** (-6.173)	-0.474* (-1.854)	-0.004** (-2.252)	0.415
t+1	0.037 (0.123)	-0.204** (-6.207)	0.001 (0.782)	-0.01*** (-5.428)	-0.656** (-2.452)	-0.003** (-2.249)	0.447
t+2	0.007 (0.227)	-0.161*** (-3.964)	0.002 (1.622)	-0.010*** (-4.303)	-0.832** (-2.576)	-0.005** (-2.260)	0.289
t+3	0.038 (1.092)	-0.183*** (-3.951)	0.001 (0.730)	-0.010*** (-4.013)	-1.038*** (-2.982)	-0.005** (-2.433)	0.273
t+4	0.044 (1.139)	-0.219*** (-4.319)	0.001 (0.053)	-0.009*** (-3.574)	-1.086*** (-2.849)	-0.006** (-2.285)	0.268
t+5	0.050 (1.291)	-0.195*** (-3.803)	0.000 (0.067)	-0.008*** (-3.110)	-0.172*** (-3.015)	-0.004* (-1.673)	0.225
t+6	0.008 (0.177)	-0.163*** (-3.014)	0.002 (0.801)	-0.006** (-2.053)	-1.27* (-2.990)	-0.003 (-1.090)	0.151
t+7	0.029 (0.539)	-0.120** (-2.017)	0.001 (0.398)	-0.006* (-1.972)	-1.409*** (-3.017)	-0.003 (-1.017)	0.104

Comparing the results of Shanghai stock market and Shenzhen stock market, we can discover some findings as follows.

First, the news articles have impact on both of the stock markets. Especially, the indicators about news articles (*news*), turnover rate (*Trnv*) of the two stock markets are all significant.

Second, the influence of news articles on Shenzhen stock market is stronger than that of Shanghai stock market. From table 4 and table 5, we can note that the maximum absolute coefficient of *news* in Shenzhen stock market is 0.235, but that is 0.108 in Shanghai stock market. The minimum value of *news* in Shenzhen stock market is 0.12, while that is 0.049 in Shanghai stock market.

Third, news articles have different impact on the firms, which has different scales. The coefficient of firm scale (*lnSize*) is significant to a certain degree in Shanghai stock market. However, the result of firm scale (*lnSize*) is not noticeable in Shenzhen stock market.

Finally, the duration of the news articles' influence on stock is different in the two stock markets. The impact strength and duration of Shenzhen stock market are stronger and longer

than that of Shanghai stock market.

### The impact of different types of news articles on stock market

By using the text mining technique, we automatically divide the news articles into four categories: (1) the capacity of profitability, (2) personnel reshuffle and lawsuit, (3) new product offerings and strategic cooperation, and (4) mergers and financing. Table 6 shows the news variable (*news*) of cross-sectional regression models which are based on different types of news articles about the whole stock market. We can find the first three categories have significant impact on the stock market. When the news stories are released, there is an immediate negative impact on the stock, especially the news articles about the capacity of profitability. However, the stock market is not sensitive to the news articles about mergers and financing. We find this kind of news articles are not reported very frequently but last long, which is different from the other types of news articles. Thus, we can conclude that news with different contents have different impact on the stock market.

TABLE VI. THE RESULTS OF MULTIVARIATE REGRESSION MODEL WITH DIFFERENT TYPES OF NEWS ARTICLES

	capacity of profitability (36.3%)	change of personnel and lawsuit(33.6%)	new product offerings and strategic cooperation (21.6%)	mergers and financing (8.8%)
t	-0.199*** (-3.652)	0.138* (1.763)	-0.049 (-0.609)	-0.138 (-1.599)
t+1	-0.132*** (-3.940)	-0.182** (-2.413)	-0.059 (-1.290)	-0.240* (-2.036)
t+2	-0.107*** (-2.858)	-0.172* (-1.774)	-0.151** (-2.182)	-0.087 (-0.693)
t+3	-0.131*** (-2.837)	-0.330** (-2.579)	-0.111 (-1.244)	-0.119 (-0.794)
t+4	-0.0.139** (-2.805)	-0.358** (-2.584)	-0.170*** (-1.758)	-0.039 (-0.230)
t+5	-0.130*** (-2.637)	-0.312** (-2.139)	-0.171*** (-1.705)	-0.060 (-0.343)
t+6	-0.113** (-2.165)	-0.385** (-2.340)	-0.165 (-1.571)	-0.176 (-1.097)
t+7	-0.092* (-1.752)	-0.410** (-2.365)	-0.087** (-0.728)	-0.205 (-1.217)

## CONCLUSION

This research is dedicated to investigating the influence of news articles on stock market in China by quantifying the Internet-based news articles using support vector regression technology. Our intension is to assist the government in supervising the securities market and help the individual investor make intelligent decisions. In particular, we explore the impact of different types of news articles on stock market, and compare the stocks' performance in two markets including Shanghai and Shenzhen stock market. In this study, there are several interesting findings summarized as follows:

- (1) Internet-based financial news tends to have a negative impact on the movements of Chinese stock market, no matter how positive and negative the news articles are.
- (2) The firms listed in Shenzhen Stock Exchange react much stronger and the impact lasts a longer period than that in Shanghai Stock Exchange. Note that the characteristics of firms in two markets are different, we conclude that the smaller the firms are, the more information can be noticed.
- (3) News articles with different contents have different impact on the stock price. When the news stories are released, there is an immediate negative impact on the stock, especially the news articles about the capacity of profitability.

This study can be extended in a few interesting ways. First, more online information sources can be considered. For example, we can enrich the information sources for the public by mining messages from Internet-based forum, Blog and twitter. In addition, we do not yet take much numeric time series data into account. A more sophisticated econometric model can be considered for more accurate analysis so as to achieve higher quality.

## REFERENCES

[1] Tetlock, P. C. , Saar Tsechansky, M. and Macskassy, S. (2008) 'More than words: quantifying language to measure firms' fundamentals', *The Journal of Finance*, Vol. 63, No.3,

pp.1437-1467.

[2] Niederhoffer, V. (1971) 'The analysis of world events and stock prices', *The Journal of Finance*, Vol. 44, No. 2, pp. 193-219.

[3] Keown, A. J. and Pinkerton, J. M. (1981) 'Merger announcements and insider trading activity: An empirical investigation', *The Journal of Finance*, Vol. 36, No. 4, pp. 855-869.

[4] Mitchell, M. L. and Mulherin, J. H. (1994) 'The impact of public information on the stock market', *The Journal of Finance*, Vol. 49, No. 3, pp. 923-950.

[5] Cutler, D. M., Poterba, J. M. and Summers, L. H. (1998) 'What moves stock prices?', *The Journal of Portfolio Management*, pp.56.

[6] Chan, W. S. (2003) 'Stock price reaction to news and no-news: drift and reversal after headlines', *Journal of Financial Economics*, Vol. 70, No. 2, pp. 223-260.

[7] Fang, L. and Peress, J. (2009) 'Media coverage and the cross-section of stock returns', *The Journal of Finance*, Vol. 64, No. 5, pp. 2023-2052.

[8] Tetlock, P. C. (2007) 'Giving content to investor sentiment: The role of media in the stock market', *The Journal of Finance*, Vol. 62, No. 3, pp. 1139-1168.

[9] Takeda, F., and Yamazaki, H. (2006) 'Stock Price Reactions to Public TV Programs on Listed Japanese Companies', *Economics Bulletin*, Vol. 13, No. 7, pp.1-7.

[10] Feldman, R. and Dagan, I. (1995) 'Knowledge discovery in textual databases (KDT)', *Proceedings of the first international conference on knowledge discovery and data mining (KDD-95)*, pp. 112-117.

[11] Wiithrich, B., Permunetilleke, D., Leung, S., Cho, V., Zhang, J. and Lam, W. (1998) 'Daily prediction of major stock indices from textual www data', *Proceedings of the 4th international conference on knowledge discovery and data mining(KDD-98)*.

[12] Antweiler, W. and Frank, M. Z. (2004) 'Is all that talk just noise? The information content of internet stock message

Xiqian Zhao, Juan Yang, Lili Zhao, and Qing Li

boards', *The Journal of Finance*, Vol. 59, No. 3, pp. 1259-1294.

[13] Schumaker, R. P. and Chen, H. (2006) 'Textual Analysis of Stock Market Prediction Using Financial News', *Americas Conference on Information Systems*.

[14] Lareid, T. and Sandal, P. C. (2006) 'Financial News Mining: Extracting useful information from continuous stream of text'.

[15] Jingmei Zhao and Fengyun Wu. (2009) 'Lucy Number Worship and Asset Price Anomaly', *Economic Research Journal*, Vol. 6, pp. 129-141.

[16] Shleifer, A. (1986) 'Do demand curves for stocks slope down?', *The Journal of Finance*, Vol. 41, No. 3, pp. 579-590.

[17] Ricardo, B. and Berthier, R. (1999) 'Modern information retrieval', *ACM*, Vol. 463.

[18] Gunn, S. R. (1998) 'Support vector machines for classification and regression', *ISIS technical report*, Vol. 14.

[19] MacKinlay, A. C. (1997) 'Event studies in economics and finance', *Journal of economic literature*, Vol. 35, No. 1, pp.13-39.

[20] Sharpe, W. F. (1964) 'Capital asset prices: A theory of market equilibrium under conditions of risk', *The Journal of Finance*, Vol. 19, No. 3, pp. 425-442.

[21] Lintner, J. (1965) 'The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets', *The Review of Economics and Statistics*, Vol. 47, No. 1, pp. 13-37.

[22] White, H. (1980) 'A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity', *Journal of the Econometric Society*, pp. 817-838.

[23] Wei Wang and Yu Zhou. (2002) 'An Empirical Study on the Size Effect and Time Effect of the Stock Market in China —Based on Shanghai Stock Market', *Economic Research Journal*, Vol. 10, No. 10.