Quantify the Effect of Artificial Intelligence News on Chinese Stock Movements

Jinghua Tan[1,2,3,*], Jun Wang[1,2,3], Menglan Zeng[1,2,3], Chenhao He[1,2,3], Yan Chen[1,2,3] , Qing Li[1,2,3]

1. School of economics and information engineering, Southwestern University of Finance and Economics, Chengdu, China; 2. Sichuan Provincial Key Laboratory of financial intelligence and financial engineering, Southwestern University of Finance and Economics, Chengdu, China; 3. Fintech Innovation Center, Southwestern University of Finance and Economics, Chengdu, China; *Corresponding Author：Jinghua Tan

Abstract

Both traditional finance and modern behavioral finance believe that the volatility of the stock market comes from the release, dissemination and absorption of information from different views. To study the mechanism, in this study, putting our attention on the high-profile Chinese Artificial Intelligence (AI) sector, we firstly quantify the news articles with natural language processing techniques, with indicators obtained, and then investigate the impact of firm-specific news on AI stock price. Following the above method, our three main findings are as follows:(1) Quantitative measure with NLP techniques of news sentiment is significant to be adopted to study the media-aware stock movements in AI sector of Chinese stock market.(2) For news-sensitive companies which share similar characteristic of relatively low share price, news release are more likely to result in abnormal returns. For news-pre-sensitive companies, news leak ahead of time can cause stock abnormal returns beforehand. As a result, when the news is truly appeared before the public, it only causes a slight stock price fluctuation. (3) The whole sentiment in the AI sector is mainly optimistic, and typically, investors blindly pursue and absorb good information in the market.

Introduction

Supported by a series of national policies such as “Internet Plus strategy”and "Made in China 2025 strategy", the artificial intelligence industry has achieved all-round development in China. From 2012 to 2018, the number of investment institutions and investment amount in the field of artificial intelligence in China have realized a leap-forward growth. Specifically, the number of AI-related enterprises founded between 2012 and 2018 reached 78% of the total number of such enterprises. The investment amount in 2018 alone reached a staggering number of 2349 billion RMB. Moreover, the Artificial Intelligence Index from the day of its establishment on March 21, 2016 to July 15, 2016, has risen from 1019 to 1097. Such a 6% increase in AI sector is sharply contrasted with a 16% decrease of Shanghai Composite index at the same time. The difference shows enough evidence that the AI sector has attracted much attention from investors in the stock market.

This abnormal fluctuations in stock movements is essentially caused by the excessive speculation operations of investors, when they receive the news information related with artificial intelligence and have a positive expectation on the stock in this sector. In fact, such media-aware stock movements are frequently observed. For example, with a piece of news article reporting WISEOFT's national major project passed the acceptance check, stock price of this company has achieved a staggering 5% raise, also accomplished with a 1.6% raise in the AI sector within just 1 day from January 7,2019 to January 8, 2019.

Theoretically, the efficient market hypothesis (EMH) states that a stock price is driven by 'unemotional' investors, who make rational decisions based on the information available. When they receive new information, they are constantly affected by the professional and expert opinions within the news articles and updating their beliefs about the directions of markets, which cause stock prices to fluctuate (Fama 1965). News articles cover relevant new information about firms' fundamentals, and therefore, cause the stock movements. In addition, according to behavioral finance, abnormal fluctuations of stocks are caused by the emotional impulses of irrational investors (De Long et al. 1990; Shleifer and Vishny 1997). That is, investors may be affected by peer opinions from social media or personal attitudes from news articles. For example, Tetlock (2007) found that high news pessimism indicated downward pressure on market prices and that high or low pessimism tended to increase market trading volume.

Although traditional finance and modern behavioral finance have different perspectives of how information shapes stock movements, both agree that the volatility of stock markets are inevitably swayed by the release, dissemination and absorption of information (Li et al. 2016). Understanding the mechanism of the impact from news articles' sentiment on AI sector will meaningfully help investors sense the risk in stock market when news articles are released. In addition, investigating how firm-specific AI news articles can affect relevant firms will also benefit the government taking reasonable actions to regulate the markets.

In this study, we first quantify the news articles into sentiment factor with natural language processing techniques. This is achieved by adopting and extending Tetlock's (2007, 2008) method which quantifies the news articles in terms of the proportion of negative and positive sentiment words. Second, we use the event study methodology and basic vector autoregressions (VARs) to reveal the deep relationships between news and stock index movements. We discover that news may leak ahead of time, so that stock market fluctuated only slightly and briefly after the news release. Third, ANOVA analysis is applied to explore the difference of sensitive and pre-sensitive firms, and find that the stocks with relatively higher stock price are more pre-sensitive to the AI news.

Related Work

The healthy development of Chinese stock market is closely related to the stability of each single sector. As the increasingly important elementary part of Chinese stock market, AI sector with a slight move may affect the stock market as a whole. With the constantly improving AI techniques and the abnormal fluctuations in stock prices related with this sector, more and more news articles were published nationwide in 2017. However, the effectiveness and functionality of these articles are yet to be studied.

Meanwhile, these firm-specific news articles convey the professional or peer opinions. Such opinions may affect the irrational investors, and then cause abnormal fluctuations of stocks by the emotional impulses. Thus, the sentiment indicators in the news allow us to explore the news effect on relevant stock movements. Our goal is to fully understand the mechanisms of AI news percolation and its impact on stock markets. More importantly, exploring such impact, are significant to help reform government regulation and strengthen the supervision of information release and dissemination.

Researchers have explored the power of verbal information on stock markets due to the observation of stock price fluctuations with news feeds. However, there is few studies on the effect of AI news on stock movements. It is a substantial challenge to understand the mechanism of news articles percolation and its degree of impact on listed firms. In this section, we evaluate the existing relevant research. We first introduce the related work focusing on the media-aware stock movements, and then present the approaches to quantify the effect of news articles on stock movements.

Media-aware Stock Movements

In traditional finance, the efficient market hypothesis states that “rational” investors are constantly updating their beliefs about the directions of markets as they receive new information about firm fundamentals, which enables stock prices to reflect their intrinsic values. For example, Fama and French (1993) identified three risk factors affecting returns on stocks—the overall market, the firm size, and the book-to-market equity ratio (BE/ME). Dechow (1994) demonstrated that both accounting earnings and cash flows can be used to measure a firm’s performance as reflected in stock returns.

However, modern behavioral finance has discovered that, with emotional impulses of investors or irrational investors, it is common to observe various financial anomalies. Gilbert and Karahalios (2010) reported that increased levels of anxiety, worry, and fear produced downward pressure on the Standard and Poors (S&P) 500 index. J. Bollen, H. Mao, and X. Zeng, found that the emotions of tweets affected stock trends for a brief period after the release of the tweets, which reveals that news information has influences on stock markets. Therefore, the study of media-aware stock movements began by investigating the influence of breaking news on stock fluctuations. For example, P. C. Tetlock, M. Saar-Tsechansky, and S. Macskassy found that high news pessimism suggested downward pressure on market prices and that high or low pessimism indicated the increasing of market trading volume. L. Fang and J. Peress used the number of newspaper articles about a stock as a proxy for the stocks overall media exposure and found that mass media could alleviate informational frictions and affect security pricing even if it did not supply genuine news. J. E. Engelberg and C. A. Parsons revealed that local press coverage increased the daily trading volume of local retail investors, from 8% to nearly 50% depending on the specification.

Nonetheless, these studies focus on general stock markets instead of concentrating on specific sector to investigate the mechanism of the media-aware influence. In fact, according to Schumaker and Chen, the experiments on each sector could offer a superior performance. Meanwhile, the present work mainly examined the media-aware stock movements on special sectors, such as the real estate and the finance, few of them explore the impact of firm-specific news articles on the burgeoning AI sector in Chinese stock markets. Thus, in this study, we try to figure out the mechanism of how firm-specific news affecting the abnormal fluctuation of the AI sector and the relevant stock. As a result, it can allow us to better understand the pattern of such media-aware stock movements, which will contribute to the regulation of the stock market and strengthen the supervision of information release and dissemination to reduce abnormal fluctuations.

Representation of News Articles

To fully understand the mechanisms of AI news percolation and its impact on stock markets, it is essential for us to analyze the dataset. In this study, the dataset consists of two sub-datasets: news information and stock transaction data. In terms of the stock transaction data, there have been numerous theories applying to identify the influence of the stock information, like the three risk factors for returns on stocks from Fama and French (1993) and the research of Dechow (1994). By contrast, processing news articles remains a considerable challenge of combining with the stock information. Therefore, it is critical to extract valuable sentiment information from textual articles to explore the effect of news on stock movements.

Due to limitations of text mining techniques in the early stages of this research, the number of news articles has been widely used as an indicator of the influence of news. This numerical value is generally treated as one of the independent variables in multiple linear regression models to capture the relationship between news and stock market dependent variable, such as the stock price, trading volume, or abnormal return. For example, Chan (2003) built an econometric model using the number of news articles as the dependent variable and abnormal returns as the independent variable and found that investors tended to react slowly to bad news. However, quantifying the influence of news using news counts is too simple because the influence of news comes from its content, which includes firm fundamentals, macroeconomic conditions, and professional or peer opinions.

Realizing such limitations, researchers have studied various text mining techniques, i.e., term vector, and sentiment analysis, to extract valuable information from news articles. In natural language processing, the basic approach for representing an article in a machine-friendly form is to transform it into a term vector, where each entry is a weighted term in the article. The weight of a term can be calculated as a Boolean value, where the weight of a word is 1 if the word exists, 0 otherwise. Such a textual representation is called a bag-of-words model. In this way, Schumaker and Chen (2008) represented a quarterly financial report as a term vector using full words and studied its influence on stock prices. Realizing that some words are irrelevant to the main topic and that using full words may not scale very well, researchers have resorted to various types of word-based and sentence-based sentiment analysis techniques. For example, H. Chen, E. C.-N. Huang, H.-M. Lu, and S.-H. Li utilized Opinion Finder, a document-level sentiment analyser, to calculate the sentiment index of each news article and found that this index obviously improved predictive precision. These methods are summarized in Table 1.

Table 1: Literature comparison of the news representation.

Category	title	Focus			Analysis Model		
		Information sources	Markets	Period	Model	Response	Performance Measure/s
News counts	chan (2003)	Down Jones News	NYSE, AMEX, S&P500	Month	Linear model	return	RMSE
	Mitchell and Mulherin(1994)	Down Jones News	NYSE, AMEX, OTC	Day	Linear model	Return, Volume	RMSE
Term vector	Schumacher	WSJ Yahoo!Finance	S&P500	2005.10.26—2005.11.28(Minute)	SVM	Prices	MSE, Return
	One-day ahead	Google New, Wikipedia pages	S&P500	Day	SVM	Direction	AUC
	Barak	Tehran Stock Exchange (TSE)	Iran	2002-2012(Day)	DT,ANN,SVM	Return, Risk	Accuracy
	eMAQT	Web	CSI100	2011.1.1-2011.12.31(Day)	SVR	Prices	Accuracy, RMSE
sentiment analysis	TeSIA	Discussion board, Web	CSI100	2011.1.1-2011.12.31(Day)	STR	Direction	Accuracy，RMSE
	Bollen	Twitter	S&P500	2018.2.28-2018.12.20(Day)	OpinionFinder and GPOMS mood time series	direction, prices	MAPE，accuracy
	Fangbinxin	web media, social media	CSI100	2015.1.1-2015.12.31(Day)	TeSIA	Direction	Accuracy, MCC
	Chen et al.	Yahoo!Finance	S&P500	Day	SVR	Return	RMSE

In fact, in behavioural finance, abnormal fluctuations of stocks are caused by the emotional impulses of irrational investors (Shleifer and Vishny 1997). In general, investors' emotion may be affected by professional or peer opinions from news articles. To analyse such opinions, researchers have resorted to various types of word-based and sentence-based sentiment analysis techniques (Pang and Lee 2008). Moreover, Li et al. has proved that a finance-oriented dictionary method is a simple and efficient way to explore the media-aware stock movements.

Therefore, in this study, we follow the approach of Tetlock (2007), and measure the contents of news articles by standardizing fraction of emotion words in each news story. Specially, we use three different perspectives of quantifying news emotions, namely pessimism, optimism, and emotion divergence. In Tetlock's study, he measures the contents of news articles in terms of general-domain emotional word dictionaries. However, Loughran and McDonald (2012) found that approximately three-quarters of all negative words in a general emotion word dictionary are not considered negative in a financial context. The word "bear" originally refers to a carnivorous mammal; it also indicates widespread pessimism in the finance domain, such as a "bear market". Some typical emotion words, such as "crude", "tire", or "capital" are more likely to identify a specific industry segment in financial events than expressing a negative sentiment as suggested by general sentiment words. Therefore, in this study, to improve the precision of sentiment analysis, we use the finance-specific sentiment word list (Li et al. 2014b) rather than general sentiment words.

Experiment Design

Experimental Data

In this study, the dataset consists of two sub-datasets: stock transaction data and news articles. The transaction data are related to Artificial Intelligence Index (AII) with 105 constituent AI stocks. Specifically, stock transaction data include turnover, current ratio, stock price, stock return, market value and market return, which is obtained from the China Securities Market Research (CSMAR) database (www.gtarsc.com), the largest and most accurate financial and economic database in China.

The firm-specific news refers to the news articles which cover fundamentals and management information on a certain company. To get these news articles, we first applied a firms-focused Web crawler to download news content from 8 mainstream financial websites in China, including *eastmoney.com*, *tengxun.com*, etc. during the period between 2016 and 2018. Specifically, titles, publishers, release time, and the content of news reports are extracted from the acquired news web page. Finally, 57,399 firm-specific news are obtained to the selected 105 listed companies. Table 1 presents a summary of the experimental data.

Table 2. Data statistics.

News Category	Firms Category	Number of firms	Number of news
	Small	39	13, 105
Firm-specific news	Medium	40	19, 082
	Large	26	25, 212

News Quantification

Tetlock (2008) presented a simple and effective approach to quantify the news articles to study the media-aware stock movements. It measures the contents of news articles by standardizing fraction of emotion words in each news article. Formally, following Tetlock (2007), we extract the following characteristics: pessimism (negative media sentiment P_t^-), optimism (positive media sentiment P_t^+), and emotion divergence (D_t). These characteristics are calculated as follows:

$$P_t^- = \frac{N_t^-}{N_t^+ + N_t^-}, P_t^+ = \frac{N_t^+}{N_t^+ + N_t^-}, D_t = \frac{N_t^+}{N_t^+ + N_t^-} \tag{1}$$

where, $N_t^+ (N_t^-)$ is the number of positive (negative) sentiment words found in the media on the t^{th} day. Note that, we combine all qualifying news articles for each firm on a given trading day into a single composite article.

Tetlock (2008) analyse the sentiment of news articles in terms of general-domain emotional word dictionaries to capture the sentiment in firm-specific news. However, 73.8% of the negative sentiment words in such general-domain emotional word dictionaries are not considered negative in a financial context. For instance, the word "bear" originally referred to an ursine animal but indicates poor earnings in the financial markets, e.g., "a bear stock". Therefore, to improve the precision of sentiment analysis, we resort to a finance-specific sentiment word list created in previous study (Li et al., 2014b) rather than general sentiment words.

Firm-specific News Articles and Relevant Stock Returns

We use the event study methodology to reveal the deep relationships between news and stock index movements. And then, the ANOVA analysis is applied to explore the difference of sensitive and pre-sensitive firms.

Regression Model for Influence Analysis

We apply an event study, an important empirical mean in financial study method, to investigate the effects of sentiment on stock movements in the field of AI. Essentially, an event study is a statistical method to explain the impact of specific events on stock abnormal returns. The event study period is divided into a pre-event estimation period and an event period, as shown in Figure 2. The pre-event estimation refers to a period of time before an event occurs, which is applied to train the model parameters. The event period is the influential period of the event (Ahern 2009).

Typically, the pre-event estimation period is longer than the event period. According to Brown and Warner (1980), event information can affect stock prices for only a short period of time, and they have almost no long-term effects on investors' goals. Therefore, in this study, we use the event study method to determine whether there is an information leakage of policy news or whether investors are slow to react to policy news by observing the abnormal return of AI for each of the 10 days before and after the release of policy news, but also performed robustness checks with different windows. Specifically, the event study period is divided into a pre-event estimation period and an event period.

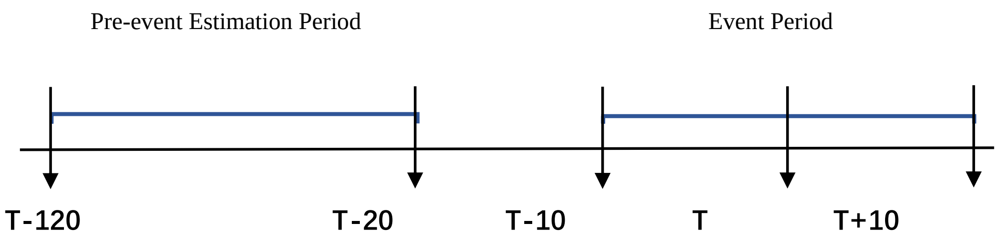

Fig 1. Time Line for an Event Day.

We regard the issue of policy news as an event, and the day of policy news release as an "event day". The event day is T. 10 days before the event day are used as the pre-event period, namely $T-10$. 10 days after the event day are used as the post-event period, namely $T+10$. The 120 days before the pre-event period are used as the pre-event estimation period, i.e. the $(T-120)-(\mathrm{T}-20)$ interval. The estimation period is used to measure the normal return before the event occurs. The stage of $(T-$

$10) - (\mathrm{T} + 10)$ is the stage of event occurrence which is the influential period of the event (Ahern 2009). We adopt stock price return to evaluate the risk-return trade-off. The estimated stock returns are calculated as follows:

$$\mathrm{R}_{i,t}^{\mathrm{Est}} = \alpha_i + \beta_i R_t^{m,f} + \delta_i SMB_t + h_i HML_t + \varepsilon_{i,t} \quad (1)$$

where $\mathrm{R}_{i,t}^{\mathrm{Est}}$ is the return of stock i on the t^{th} day during the pre-event estimation period, $R_t^{m,f}$ denotes the market profitability on the t^{th} day during the pre-event estimation period, SMB_t is the scale factor that stands for small market capitalization minus big market capitalization on the t^{th} day during the pre-event estimation period, HML_t is the net asset market value ratio factor that stands for high book-to-market ratio minus low book-to-market ratio on the t^{th} day during the pre-event estimation period, and t is the day calculated during the pre-event estimation period. SMB_t and HML_t are used to measure the historic excess returns of small capital over big capital and of value stocks over growth stocks. The abnormal return ($AR_{i,t}$) of stock i on the t^{th} day during event period is defined as:

$$AR_{i,t} = \mathrm{R}_{i,t}^{\text{Eve-a}} - \mathrm{R}_{i,t}^{\text{Eve-p}} \quad (2)$$

where t is the day measured relative to the event, t = 0, 1,. . ., 9, 10. Here, day 0 is the day on which the news event is released, $\mathrm{R}_{i,t}^{\text{Eve-a}}$ is actual return on the t^{th} day during the event period, and $\mathrm{R}_{i,t}^{\text{Eve-p}}$ is the estimated return on the t^{th} day of the event period in terms of returns in the pre-event estimation period. Therefore, the abnormal return of t^{th} in event period, $\overline{AR_t}$ is calculated by

$$\overline{AR_t} = \frac{1}{n_i} \sum_{i-1}^{n} AR_{i,t} \quad (3)$$

We apply t-tests to examine whether the abnormal return and the cumulative abnormal return are significant. Specifically, $T(t,t)$ and $T(t_1,t_2)$ are t-statistics for $\overline{AR_t}$ and defined as

$$T(t,t) = SAR \bullet \sqrt{n_i} \tag{4}$$

SAR is calculated as follows:

$$SAR = \frac{1}{n_i} \sum_{i=1}^{i=n_i} \frac{AR_{i,t}}{S_i} \tag{5}$$

where S_i is the residual standard deviation of stock i obtained from our regression model, as shown in formula (1). SAR is the standardized abnormal return.

To deeply understand the mechanism on how the digitalized news information affects stock markets, we adopt an ordinary least squares (OLS) equation to explore the relationship between abnormal return and its determinants. We are particularly interested in the sentiment of news of listed firms. Specifically, we consider the following determinants.

Circulation market value (MV) is calculated by the number of tradable shares multiplied by the stock price at a certain time. Turnover rate (TO) is stock trading frequency, which reflects the strength of stock liquidity. Current ratio (CR) is a liquidity ratio that measures whether a firm has enough resources to meet its short-term obligations. MV, TO, and CR have been shown to be important factors affecting securities markets (Bhide 1993; Datar, Naik, and Radcliffe 1998; Mulyono and Khairurizka 2009). Therefore, we selected these three variables to form the basic model in order to examine the impact of news information on abnormal returns.

Moreover, Recent studies have found that the variance in the information disclosure behaviours of listed companies may lead to different investor responses and thereby affect their market performance. In particular, Buskirk (2012) found that more frequent disclosure does accelerate the rate at which information is impounded into price. In fact, the potential role of investor sentiment in financial markets has received considerable attention. As Baker and Wurgler (2006) argue, “Now, the question is no longer, as it was a few decades ago, whether investor sentiment affects stock prices, but rather how to measure investor sentiment and quantify its effects.” Therefore, we added P_t^- , P_t^+ and D_t as measurements of emotion to evaluate the effect of textual information on stocks, as suggested by Tetlock (2007).

According to the above discussion, we adopted the abnormal return of news information(C_{AR}) as the explanatory variable, and the OLS equation describing the relationship between C_{AR} and its determinants is

$$C_{AR} = \alpha + \beta_1 MV + \beta_2 TO + \beta_3 CR + \delta_1 P_t^- + \delta_2 P_t^+ + \delta_3 D_t + \varepsilon \qquad (6)$$

Multi-way Analysis of Variance (ANOVA)

In statistics, One-way ANOVA is a technique that can be used to compare means of two or more groups (using the F distribution). This technique can be used only for numerical response data, the “Y”, usually one variable, and numerical or (usually) categorical input data, the “X”, always one variable, hence “one-way”. As an extended version of the One-Way ANOVA, Multi-Factorial ANOVA is a statistical technique which can be used to investigate whether a dependent variable is affected by multiple independent variables (factors). Like before, the “Y”, usually one variable, and the "X_i" represent three different factors and their interactions. In this study, Multi-Factorial ANOVA model is employed to ascertain whether the three factors are influential in determining if a company is news-sensitive or not, which include company market value, trading volumes and the amount of company-related news. This ANOVA test is the F-test statistic. The formula for the F-test is as follows:

$$F_A = \frac{MS_A}{MS_e} \tag{7}$$

$$F_B = \frac{MS_B}{MS_e} \tag{8}$$

$$F_{A*B} = \frac{MS_{A*B}}{MS_e} \tag{9}$$

Where, MS_A and MS_B are the Mean Square of factor A and factor B between groups, respectively. MS_{A*B} is Mean Square of the interaction between factor A and B between groups; MS_e is the Mean Square within groups. In this paper, we use regression model to classify firms into two types: news-sensitive and news-pre-sensitive. Then, the Multi-Factorial ANOVA is used to find out the influence of various factors on the company's category (news-sensitive or news-pre-sensitive).

Empirical Results

In this paper, we first investigate the effect of firm-specific news articles related with artificial intelligence sector on stock markets. This is achieved by quantifying the news articles with natural language processing techniques and studying its impact with classical event study methodology. Then, we utilize the regression model to explore the effect of media sentiment on relevant stock abnormal returns in terms of news context. Finally, we use analysis of variance (ANOVA) to study the commonalities of two subgroups firms which are with different performance affected by news.

Event Study on the Firm-specific Real Estate News

In this section, we study the effect of news articles related with the AI stocks, and estimate the abnormal return of these stocks for each of the 10 days before and after relevant news released to explore such effectiveness. To estimate the abnormal returns in our event study methodology, we adapt the Fama-French model three-factor model. The Fama-French method can significantly explain price outperformance for the majority of stocks (Fama and French, 1993). Therefore, we can treat them as the risk factors that are used to estimate abnormal returns. We set a pre-event estimation period of [−120, −20] trading days to find the regression coefficients.

In this study, we focus on the short-term effect of news with different emotions (positive and negative) and examine their effect on AI sector in Chinese stock markets for each of the 10 days before and after information released. In Table 3, our empirical results are in accordance with our expectations to some extent. That is, optimistic news tends to have positive impact on stock returns while pessimistic news has downward on stock returns. There are several more interesting findings: (1) The abnormal return (AR) of AI stocks is statistically significant within the [-1, 6] time windows before news released. (2) The AR is negative before the news released ([-10, -2] days) and positive after the news released ([-1, 1] days). (3) On the other periods before and after the release of news, the stock market experienced random fluctuations.

Table 3. Abnormal return and cumulative abnormal return.

	Positive news		**Negative news**		**All news**	
Day	**AR (1)**	**CAR (2)**	**AR (3)**	**CAR (4)**	**AR (5)**	**CAR (6)**
-10	0.39	4.44	0.42	0.83	-0.51	-0.51
-9	0.12	4.56	-3.48	-2.65	-0.26*	-0.77*
-8	0.43	4.99	-2.04	-4.69	-0.14	-0.40

-7	0.63*	5.62*	-1.28**	-5.97	-0.78**	-0.91*
-6	0.41	6.03	4.46	-1.51	-0.42*	-1.20**
-5	0.65*	6.68**	2.46	0.94	-0.38*	-0.80*
-4	-0.60*	6.08**	1.37	2.31	0.36	-0.02
-3	0.20	6.28	0.42	2.74	0.40	0.76
-2	-0.03	6.25	3.34	6.08*	-0.19	0.21
-1	0.15**	6.40**	-1.44**	4.63*	1.47***	1.29**
0	0.37**	6.77***	-3.65**	3.98**	2.34***	3.82***
1	0.80**	7.57***	1.79*	5.77*	0.38**	2.72***
2	0.44*	8.01*	4.72	10.49	-0.57***	-0.19***
3	-0.17*	7.84***	-1.23	9.26	-0.15**	-0.72***
4	0.31	8.15	1.45*	10.71**	-1.22***	-1.37***
5	-0.05	8.09	-3.29**	7.42	-0.81**	-2.03**
6	0.09	8.19	-3.58	3.84	-0.94*	-1.76***
7	0.33	8.52	0.17	4.01	-0.42	-1.37
8	1.34*	9.86**	-3.65	0.36	-0.55*	-0.97*
9	0.89	10.75	0.61*	0.97*	-0.62	-1.17
10	0.68	11.43	-3.17	-4.20	-1.13	-1.76

Notes: This table shows the results of the abnormal return (one basis point equals a daily return of 0.1%) within a certain time before and after the release of news. We estimate abnormal returns and cumulative abnormal returns for each of the 10 days before and after policy news are released based on the Fama-French 3 factors in our event study methodology. We set the pre-event estimation window to [-135, -16] trading days to fully conduct our regression analysis.

*, **, *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

These findings provide concrete support for the effectiveness of news information and information leakage of Chinese stock market (Gao and Tse, 2001; Ying-Peng, 2013). Specially, the significant AR within the [-1, 6] shows that the released news articles affect investors' behaviours and lead to the abnormal return in the stock market. One good explanation is that new information released increases the information transparency of the markets and affect the investors' expectation on the market. The AR before the $(t-1)^{th}$ is positive and then becomes negative ($p < 0.01$), which indicated news has been leaked in the first 1 day ahead of news released. For the periods of abnormal fluctuations, one possible explanation is that as the stock market absorbs information, the emotional investors tend to be calmed, and even no longer cares about the news information. Another explanation is that investors who know the leaked news may buy the stocks before the news released and sell them after the news released to make profits. After news released, the investments' behaviours of other unknown investors lead to temporary abnormal fluctuations in the stock market. After 6 days of news released, the articles gradually have slight effect on the stock market.

Due to more vividly display our research results, Figure 2, 3 and 4 were plotted based on columns (1), (3) and (5) in Table 3. The figure shows the abnormal return for each of the 10 trading days before and after news releasing. We find that the abnormal returns of positive news present an increasing trend. Positive news brings about sustained growth in abnormal return, a significant increase does occur within the [-10, +10] time windows. This is partially because investors are more likely to blindly pursue and absorb good information. However, negative news has caused greater abnormal return volatility in the stock market (reported in column (3), Table 3). Relatively, the released bad news has a sharp influence on the markets. That is, investors change their expectations fast when they receive bad articles, and such effectiveness is soon be diluted by the o … news affects the markets and ga … s released. This proves that new … hen change their behaviours (rep

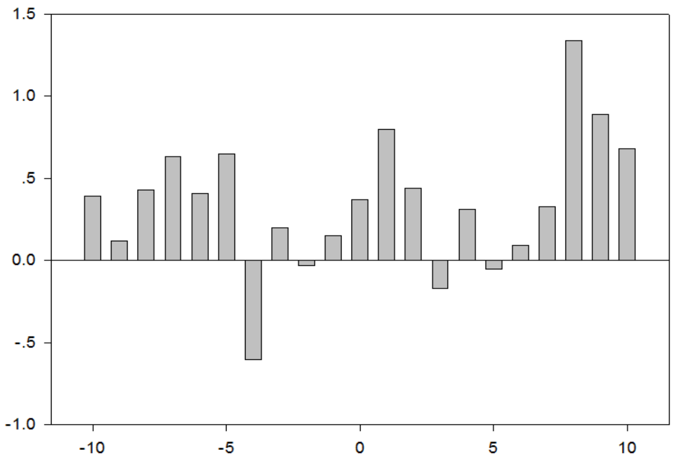

Fig 2. Abnormal return of positive news

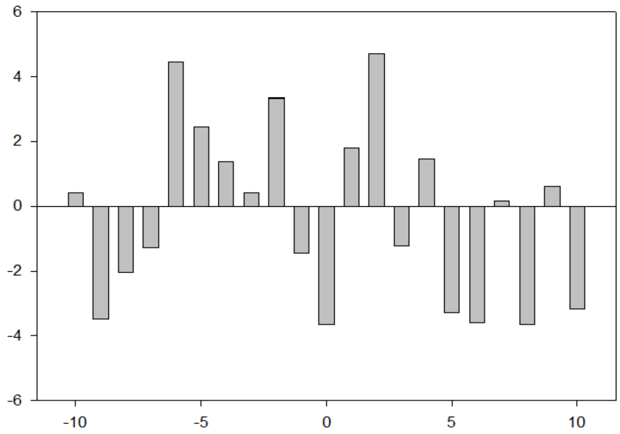

Fig 3. Abnormal return of negative news

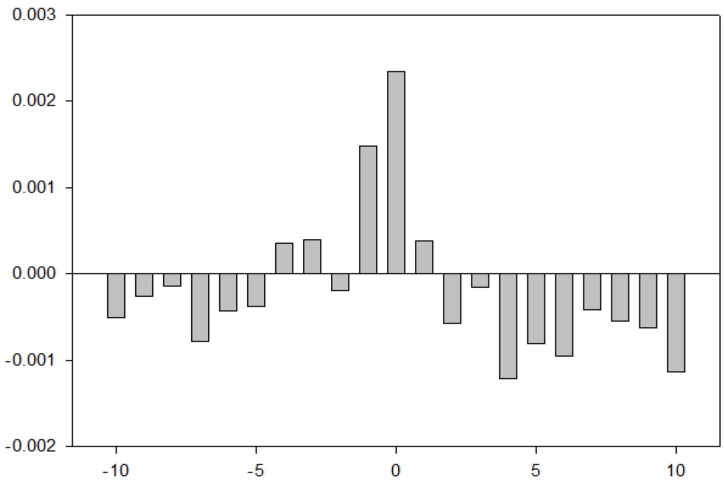

Fig 4. Abnormal return of all news

Effectiveness of News on Abnormal Returns in Terms of Emotion

In this section, we utilize OLS to further explore the relationship between C_{AR} and its determinants. We are particularly interested in the sentiment of news information of listed firms. In the regression, we explore several control variables, which include P_t^- (negative index), P_t^+ (positive index) and D_t (bias index), to assess the sentiment of news information.

The measure of P_t^- (negative index), P_t^+ (positive index) and D_t (bias index) is the sentiment penchant of an event of news. Moreover, the regression includes three basic control variables for numerous firm characteristics: market value (MV), turnover rate (TO), and current ratio (CR). In this study, to extract the sentiment penchant of a textual message more accurately, we adapt the financial-oriented approach. This provides consolidated support for further analysis on the effect of news event on stock markets in terms of textual sentiment.

Table 4. Effect of emotion on abnormal return

	C_{AR} (-1)	C_{AR} (0)
MV	0.2011318*** (0.0019025)	0.1400263*** (0.0017017)
TO	0.0004047*** (0.0000784)	0.0005181*** (0.0000701)
CR	-0.0024132*** (0.0004562)	-0.001252*** (0.000408)
P_t^-	-0.0012598** (0.0005147)	-0.0065333** (0.0031127)
P_t^+	0.0025084** (0.001201)	0.00885** (0.0042531)
D_t	0.0139358*** (0.0021878)	0.009241*** (0.0019569)

Notes: This table reports the results from OLS regression (C_{AR}) regressed on circulation P_t^- (negative index), P_t^+ (positive index) and D_t (bias index) to assess the effect of the sentiment of news information of listed firms. The C_{AR} is the abnormal return within a certain time after the release of

news. The measure of P_t^- (negative index), P_t^+ (positive index) and D_t (bias index) is the sentiment penchant of an event of news. Moreover, the regression includes three basic control variables for numerous firm characteristics: market value (MV), turnover rate (TO), and current ratio (CR).
The p-values are reported in parentheses. *p < 0.10; **p < 0.05; ***p < 0.01.

The regressions include three basic control variables for numerous firm characteristics, such as MV, TO, and CR. Specifically, the circulation MV is calculated by the number of tradable shares multiplied by stock price at a certain time. TO is the stock trading frequency, which reflects the strength of stock liquidity. CR is a liquidity ratio that measures whether a firm has enough resources to meet its short-term obligations. MV, TO, and CR have been shown to be important factors that affect securities markets (Bhide 1993; Datar, Naik, and Radcliffe 1998; Mulyono and Khairurizka 2009).

Similar to the work of Tetlock (2008), we run a set of regression to ensure that the covariance on stock C_{AR} with three basic factors in the basic model. In the set of regressions, we use each C_{AR} at a certain point of time after news event as the dependent variable, where the basic model is a benchmark for expected volatility.

We analyse the C_{AR} regression that include news sentiment information of listed firms in greater detail. As shown in Table 4, the regression results of MV, TO, and CR are statistically significant ($p < 0.01$) at two different points of time.

Besides, the regression results for news sentiment information are also statistically significant ($p < 0.05$) at two different points of time. Therefore, this result is consistent with our previous experiments: sentiment of news enlarges the changing trend of abnormal returns. Therefore, there are two interesting findings: (1) The effectiveness of positive or negative emotional indicators are both statistically significant. Similar with the findings of Das and Chen (2007) and Schumaker et al. (2012), optimistic news tends to have positive impact on stock returns while pessimistic news has downward on stock returns. (2) Divergence reflects the difference between optimistic emotion and pessimistic emotion. In Table 4, the effect of divergence on stock returns are significantly positive.

In addition, in order to prevent the influence between emotional variables, this study conducts a robust test on each emotional variable to ensure the accuracy of the research results. In Table 5, 6 and 7, the results of this study show that the three emotional indexes were statistically significant and passed the robust test.

Table 5. Effect of negative index on abnormal return

	C_{AR} (-1)	C_{AR} (0)
MV	0.3580344*** (0.002509)	0.3568864*** (0.0025065)
TO	0.0027592*** (0.0001616)	0.0027023*** (0.0001614)
CR	-0.0115813*** (0.0014614)	-0.0010759*** (0.0002739)
P_t^-	-0.000582* (0.0003472)	-0.000312186* (0.0006509)
P_t^+		
D_t		

Table 6. Effect of positive index on abnormal return

	C_{AR} (-1)	C_{AR} (0)
MV	0.4096103*** (0.0031039)	0.4092781*** (0.0031003)
TO	0.0043062***	0.0042721***

	(0.0001944)	(0.0001942)
CR	-0.0175347*** (0.0002816)	-0.0174813*** (0.0002814)
P_t^-		
P_t^+	0.0006186* (0.0006509)	0.012537** (0.0031149)
D_t		

Table 7. Effect of bias index on abnormal return

	C_{AR} (-1)	C_{AR} (0)
MV	0.5490827*** (0.0030329)	0.4708836*** (0.0024372)
TO	0.0134097*** (0.0024873)	0.0134526*** (0.0024816)
CR	-0.0100514*** (0.0002316)	-0.0099477*** (0.0002315)
P_t^-		
P_t^+		
D_t	0.0022812*** (0.0004562)	0.0013625*** (0.000078)

ANOVA Analysis Between Two Subcategorized Firms

In order to study the significant differences between news-sensitive stocks and news-pre-sensitive stocks. We first estimate the coefficients of the factor "*emotion*" for each stock, and investigate their significance to obtain the two groups of firms. The firms in the news-pre-sensitive group are significantly affected by three different emotion factors ahead of the news released. Then, we find that the top 9 news-pre-sensitive stocks are pre-sensitive to all three emotion factors and 52 stocks are pre-sensitive to one or two emotion factors. Finally, we recognize 61 AI listed stocks as pre-sensitive firms.

In this study, we first investigate what kind of firms are tending to leak the information before the news released. The ANOVA analysis is used to find out the significant difference characteristics between the two groups, including company market value (MV), trading volumes (TV) and the amount of company-related news (AN). It would be helpful for the investors to make decisions when receive new information in the stock markets.

Table 8. ANOVA analysis in different groups.

Source	SS	df	MS	F	P-value	PES
MV	1788.781	1	1788.781	37.213**	0.000	0.847
TV	88.784	2	44.392	1.889	0.165	0.090
AN	358.920	1	358.920	4.415*	0.009	0.242
MV x TV	307.177	2	153.589	2.251	0.055	0.076
TV x AN	162.884	2	81.442	2.136	0.127	0.082
MV x AN	325.972	1	325.972	4.219*	0.009	0.237

The p-values are reported in parentheses. *p < 0.10; **p < 0.05; ***p < 0.01.

From Table 8, the F-value is statistically significant (F-value > the critical value of F-test = 3.923) in terms of MV, AN and MV x AN. The result is significant at $p < 0.05$. This indicates that the company market value (MV) is different between the two groups. Specifically, we find that the news-pre-sensitive firms has a big MV, and generally such firms with a higher stock price. One good example is that companies with relatively high stock prices have more chance for arbitrage and thus are more likely to receive continuous attention from investors. As a result of that, news leak is more likely to happen. That is, the firms with higher stock prices generally attract much attention in the markets and its new information reflects fast in the market.

Conclusion:

To our best knowledge, this is the first work to explore the impact of firm-specific news articles on the AI sector in Chinese stock markets. In fact, media-aware stock movements have attracted much attention from both academia and industry (Li et al. 2018). Experiments on sectors conducted by (Schumacher and Chen, 2009) have achieved superior performance. Therefore, putting our focus on the overheated AI industry with a great amount of news coverage, we investigated how AI news affects stock movements.

There exist several interesting findings in our research. First, with the adoption of NLP techniques, we found that positive news articles basically result in net gains in abnormal daily returns, however, impact of negative news is somehow uncertain for it sometimes bring net losses while sometimes net gains. One explanation is that investors' optimistic moods make themselves hold a wait-and-see attitude toward the negative news. Therefore, it can be inferred that investors' investment behavior is mainly optimistic and thus leads to the continuous rise of the stock price of Chinese AI companies in recent years. Second, news leak ahead of time may cause stock price abnormal fluctuation beforehand. As a result, stock price fluctuation caused by the news is rather slight when it is truly appeared before the public. We define such kind of companies as news-pre-sensitive. Contrastively, for those companies without news leak, news release will normally cause significant abnormal return. Thus, we define such kind of companies as news-sensitive. Third, conclusions based on the ANOVA investigation of the two kind companies are as follows: companies with comparably low share price are more prone to be news-sensitive, while with relatively high share price to be news-pre-sensitive. This is probably because of the fact that companies with relatively high share prices have more chance for arbitrage and thus are more likely to receive continuous attention from investors. As a result of that, news leak is more likely to happen.

In addition, unlike traditional case studies, we extend Fama-French three-factor model by adding the news sentiment factors to build a simple but effective framework which can be generalized to study the effect of news articles related with other specific sectors. Meanwhile, it would be interesting to explore more information sources including micro-blog, discussion board and so on.

Reference

[1] Ahern, K. R. 2009. Sample selection and event study estimation. Journal of Empirical Finance 16 (3):466–82.

[2] Baker, M., and J. Wurgler. 2006. Investor sentiment and the cross-section of stock returns. Journal of Finance 61 (4):1645–80.

[3] Bhide, A. 1993. The hidden costs of stock market liquidity. Journal of Financial Economics 34 (1):31–51.

[4] Brown, S. J., and J. B. Warner. 1980. Measuring security price performance. Journal of Financial Economics 8 (3):205–58.

[5] Chen Y, Das M. An automated technique for image noise identification using a simple pattern classification approach. In 50th Midwest Symposium on Circuits and Systems. IEEE, 2007: 819-822.

[6] Datar, V. T., N. Y. Naik, and R. Radcliffe. 1998. Liquidity and stock returns: An alternative test. Journal of Financial Markets 1 (2):203–219.

[7] De Long, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann. 1990. Noise trader risk in financial markets. Journal of Political Economy 98 (4):703–38.

[8] E. Engelberg and C. A. Parsons, “The causal impact of media in financial markets,” J. Finance, vol. 66, no. 1, pp. 67–97, 2011.

[9] Gao, Y., & Tse, Y. K. (2001). Capital control, market segmentation and cross-border flow of information: some empirical evidence from the Chinese stock market. SSRN Electronic Journal.

[10] Gilbert, E., Karahalios, K., & Sandvig, C. (2010). The network in the garden: designing social media for rural life. American Behavioural Scientist, 53(9), 1367-1388.

[11] Hahn, R. W., & Tetlock, P. C. (2008). Has economic analysis improved regulatory decisions? Journal of Economic Perspectives, 22(1), 67-84.

[12] H. Chen, E. C.-N. Huang, H.-M. Lu, and S.-H. Li, “AZ Smart- Stock: Stock prediction with targeted sentiment and life support,” IEEE Intell. Syst, vol. 26, no. 6, pp. 84–88, 2011.

[13] J. Bollen, H. Mao, and X. Zeng, “Twitter mood predicts the stock market,” J. Comput. Sci., [3] vol. 2, no. 1, pp. 1–8, 2011.

[14] Loughran, T., & Mcdonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-Ks. Journal of Finance, 66(1), 35-65.

[15] Li, Q., T. J. Wang, P. Li, L. Liu, Q. X. Gong, and Y. Z. Chen. 2014b. The effect of news and public mood on stock movements. Information Sciences 278 (10):826–40.

[16] Li, Q., Chen, Y., Jiang, L. L., Li, P., & Chen, H. (2016). A tensor-based information framework for predicting the stock market. ACM Transactions on Information Systems, 34(2), 1-30.

[17] Mulyono, D. M., and R. Khairurizka. 2009. The effect of financial ratios, firm size, and cash flow from operating activities in the interim report to the stock return. Chinese Business Review 8 (6):44.

[18] Pang, H., & Tan, K. L. (2008). Verifying completeness of relational query answers from online servers. ACM Transactions on Information and System Security, 11(2), 1-50.

[19] P. C. Tetlock, M. Saar-Tsechansky, and S. Macskassy, “More than words: Quantifying language to measure firms’ fundamentals,” J. Finance, vol. 63, no. 3, pp. 1437–1467, 2008.

[20] P. C. Tetlock, “Giving content to investor sentiment: The role of media in the stock market,” J. Finance, vol. 62, no. 3, pp. 1139– 1168, 2007.

[21] Peress, J., & Fang, L. H. (2008). Media coverage and the cross-section of stock returns. The Journal of Finance, 64(5), 2023-2052.

[22] R. P. Schumaker and H. Chen, “Textual analysis of stock market prediction using breaking financial news: The AZFin text system,” ACM Trans. Inf. Syst., vol. 27, no. 2, pp. 12:1–12:19, 2009.

[23] Shleifer, A., & Vishny, R. W. (1997). A survey of corporate governance. Journal of Finance, 52.

[24] Shleifer, A., and R. W. Vishny. 1997. A survey of corporate governance. Journal of Finance 52 (2):737–83.

[25] Sharpe, S. A. (1994). Financial market imperfections, firm leverage, and the cyclicality of employment. American Economic Review, 84.

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

[27] Throckmorton, G. S., & Dechow, P. C. (1994). In vitro strain measurements in the condylar process of the human mandible. Archives of Oral Biology, 39(10), 853-67.

[28] W. S. Chan, "Stock price reaction to news and no-news: Drift and reversal after headlines," J. Financ. Econ., vol. 70, no. 2, pp. 223–260, 2003.

[29] Xu, S. X., and X. Q. Zhang. 2013. Impact of Wikipedia on market information environment: Evidence on management disclosure and investor reaction. Management Information Systems Quarterly (MISQ) 37 (4):1043–68.