Amount of news before stock market fluctuations

Yoshifumi Tahira
Takayuki Mizuno

March, 2016

Amount of news before stock market fluctuations

Yoshifumi Tahira · Takayuki Mizuno

Abstract The views of the Wikipedia pages of companies listed in the Dow Jones Industrial Average (DJIA) were correlated with the future DJIA changes. Such an increase in views suggest that new information is circulating about these companies. We elucidate that such fluctuations in the number of news articles about a stock market are correlated with the future changes of its index by investigating 9,150,000 news articles distributed by Thomson Reuters and the stock market indexes between December 2007 and April 2012. When the number of news articles about companies listed in NYSE and NASDAQ increase/decrease in one week, the Standard & Poor's 500 index (S&P500 index) tends to fall/rise in the next week. On the other hand, the fluctuation in other stock market indexes, which are rarely correlated with NYSE and NASDAQ, are basically random. Markets notably react to the fluctuation in the number of business sector news articles. These characteristics are observed every year.

Keywords Econophysics · Stock market · Business news · Exogenous shock

1 Introduction

Violent price fluctuations, which don't follow normal distributions, often happen in financial markets and disturb market economics. The mechanisms of such violent price fluctuations are researched in continental finance and econophysics.

Y. Tahira
Chuo University Graduate School of Science and Engineering, 1-13-27 Kasuga, Bunkyo-ku, Tokyo, Japan
E-mail: ytahira@phys.chuo-u.ac.jp

T. Mizuno
National Institute of Informatics, SOKENDAI (The Graduate University for Advanced Studies) School of Multidisciplinary Sciences, PRESTO Japan Science and Technology Agency, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, Japan

Asset prices are moved by an endogenous mechanism that follows the past price changes made by trend followers and by an exogenous mechanism of the reactions to sudden news in financial markets (Jiang, Guo & Zhou, 2007; Sornette, 2006). It is proved that endogenous mechanisms follow a multiplicative process by investigating the historical tick-by-tick data of prices. This multiplicative process generates a fat tail of price changes (Newman, 2005). Exogenous mechanisms have also been investigated using news archives and historical tick-by-tick price data (Mizuno, Ohnishi & Watanabe, 2015; Hisano et al., 2013; Alanyali, Moat & Preis, 2013; Mizuno et al., 2012; Petersen et al., 2010; Rangel, 2011). Natural language processing for news articles identified a relationship between price changes and text words in news articles (Bollen, Mao & Zeng, 2011; Schumaker & Chen, 2009; Thelwall et al., 2010). Furthermore, investigations have addressed the relationship between people's reactions to news and price changes. In prospect theory in behavioral economics, people are more concerned about avoiding loss than earning profit in stock trading (Tversky & Kahneman, 1991). When risk exists that a stock might fall, people desperately gather information about it from the web in to avoid that risk. Therefore, we can assume that the search volume and topic views about a stock increase on the web before it falls. T. Preis et al. showed that the fluctuations of the views of Wikipedia pages of 30 companies listed in the Dow Jones Industrial Average (DJIA) are correlated with future DJIA changes (Preis, Moat & Stanley, 2013; Moat et al., 2013; Curme et al., 2014). The amount of the views of these pages increases before DJIA falls.

In this study, we argue that the fluctuations in the number of news articles about a stock market are related to the changes of its stock market index. Since news media concentrate on the topics that attract people's attention, we assume that the number of news articles on a certain topic will increase before a stock market index falls, as in the relationship between Wikipedia views and DJIA. We establish this assumption by investigating the news articles distributed by Thomson Reuters and the time series of stock market indexes.

The remainder of this article is organized as follows. In Section 2, we explain the news data of Thomson Reuters. In Section 3, we explain our method that investigates the relationship between the fluctuation in the number of news articles and the changes of stock market indexes. In Section 4, we apply this method to the S&P500 index and news articles about NYSE and NASDAQ. In Sections 5 and 6, we investigate this relationship with respect to a few topics and compare years. Section 7 is our conclusion.

2 Business news data

We used the 9,150,000 news articles between December 2007 and April 2012 that were distributed for industrial investors by Thomson Reuters, one of the world's largest financial information providers (Thomson Reuters, 2008). These news articles were classified into alert, headline and story take overwrite. Alert is the urgent news and only provides a title. Headline is other news that

also only provides a title. Story take overwrite is both alert and headline texts that are subsequently distributed. We used only the 9,150,000 news articles classified into alert and headline to count a substantial number of news articles.

Each news article contains keywords appended by Thomson Reuters called "topic code". Each topic code is categorized into the following 22 classifications: "Arts/Culture/Entertainment," "Asset Class/Property," "Business Sector," "Commodity," "Crime," "Currency," "Disaster/Accident," "Event Type," "Genre," "Geography," "Health/Medicine" "Indicator Type," "Intellectual Property," "Language," "Legacy News Topic," "News Flag/Status," "Organization," "Physical Asset Type," "Religion," "Sport," "Sports combined with Geography," and "Sporting Competitions" by Thomson Reuters. Each news article also contains codes that represent the stock name called the stock name code.

News articles contain the stock name codes and the topic codes. In this study, we use only news articles containing code reported more than 130 weeks out of 230 weeks between December 2007 and April 2012. For example, news articles containing "IBM" code were reported 230 weeks. Due to this rule, the numbers of stock name codes and topic codes were 500 and 676.

3 Method

We introduce the trading strategy of Preis, Moat & Stanley (2013); Moat et al. (2013); Curme et al. (2014) based on Wikipedia views and Google's search volume to clarify the correlation between the fluctuation in the number of news articles about a certain stock market and its future changes.

We defined the number of news articles about listed company i in week t as $n_i(t)$ and calculated the average number of news articles for previous week Δt: $N_i(t-1, \Delta t) = \sum_{\tau=1}^{\Delta t} n_i(t-\tau)/\Delta t$. If the number of news articles increased in week t such that $n_i(t) > N_i(t-1, \Delta t)$, we virtually sell a unit of the stock market index at closing price $p(t+1)$ on the first trading day of week $t+1$, and virtually buy a unit of the stock market index at closing price $p(t+2)$ on the first trading day of week $t+2$. On the other hand, if the number of news articles decreased in week t such that $n_i(t) \leq N_i(t-1, \Delta t)$, we virtually buy a unit of the stock market index at closing price $p(t+1)$ on the first trading day of week $t+1$ and virtually sell a unit of the stock market index at closing price $p(t+2)$ on the first trading day of week $t+2$. We estimated cumulative returns R_i of this trading strategy between December 2007 and April 2012. To confirm this performance, we introduced a trading strategy that randomly buys or sells. Cumulative returns R_i are normalized by the standard deviation from 100,000 random trading strategies. We focus on the distribution of cumulative returns R_i in a stock market.

4 Relationship between fluctuations in number of news articles and future stock market index changes

We show the relationship between the fluctuations of the number of news articles about a certain stock market and future stock market index changes using the trading strategy introduced in Section 3. We use the S&P500 index, the Nikkei 225 index, and news articles about NYSE, NASDAQ, the Tokyo Stock Exchange, the Hong Kong Exchanges and Clearing, the Shanghai Stock Exchange and the Korea Stock Exchange.

Fig. 1(a) shows the probability density distribution of cumulative returns R_i when the S&P500 index is traded based on the fluctuations of the number of news articles about the listed companies in NYSE and NASDAQ for previous week $\Delta t = 7$ weeks. The mean of the cumulative returns of these trading strategies $\langle R \rangle = 0.40$ is significantly higher than that of random trading strategies $\langle R \rangle = 1.3 \times 10^{-3}$. Since the p value of the Kolmogorov-Smirnov test is 1.58×10^{-11}, the following null hypothesis is rejected "Distribution of these trading strategies and that of random trading strategies are the same.". Fig. 1(b) is the probability density distribution of cumulative returns R_i when the Nikkei 225 index is traded based on the fluctuations of the number of news articles about listed companies in NYSE and NASDAQ. The mean of the cumulative returns from these trading strategies $\langle R \rangle$ is 0.17, and the p value of the Kolmogorov-Smirnov test is 0.041. The null hypothesis is rejected at a significance level of 5%, but the p value of Fig. 1(b) exceeds Fig. 1(a), and the distribution of such trading strategies more closely resembles the random trading strategies.

We also investigated the Δt dependence of the mean of cumulative return $\langle R \rangle$. Fig. 2 shows the means of cumulative returns $\langle R \rangle$ when the S&P500 and Nikkei 225 indexes are traded based on the fluctuations of the number of news articles about the listed companies in NYSE and NASDAQ for previous Δt weeks. The mean of cumulative returns $\langle R \rangle$ of the S&P500 index exceeds that of the Nikkei 225 index for every Δt.
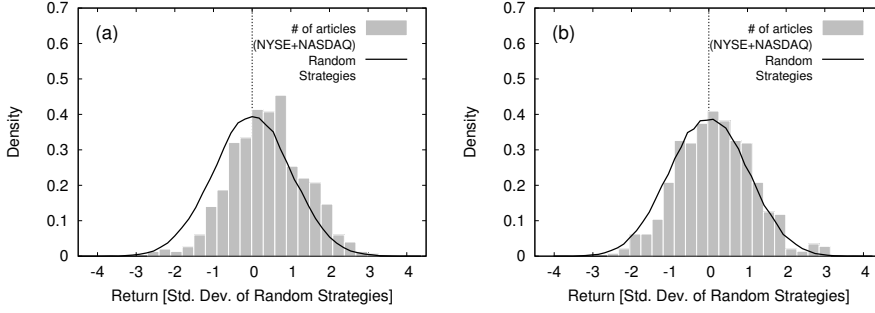
Fig. 1 Probability density distributions of cumulative returns R_i when (a)S&P500 index and (b)Nikkei 225 index are traded based on fluctuations in number of news articles about NYSE and NASDAQ. The curves represent distribution of cumulative returns from random trading strategies. Δt is 7 weeks (see Section 3).
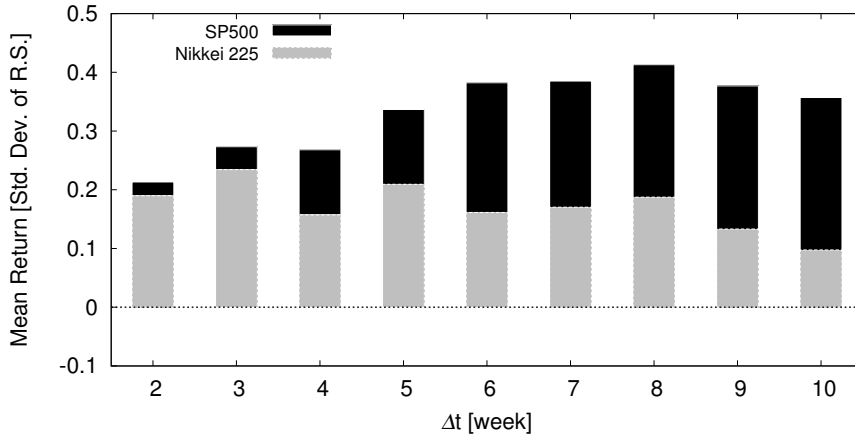

Fig. 2 Means of cumulative returns $\langle R \rangle$ based on fluctuations of number of news articles about listed companies in NYSE and NASDAQ for previous Δt weeks. Black and gray are means of cumulative returns $\langle R \rangle$ when S&P500 index and trading Nikkei 225 index are traded, respectively.

5 Cumulative return for each topic

Next, we show that the news topics are related to the future changes of stock market indexes. We estimate the means of cumulative returns $\langle R \rangle$ when the S&P500 index is traded based on the fluctuations in the number of news

articles for previous $\Delta t = 7$ weeks for each classification of topics[1](see Section 2).

Table 1 expresses the number of topics and the mean of cumulative returns $\langle R \rangle$ for each topic classification. The mean of cumulative returns $\langle R \rangle$ of the business sector class is the highest, $\langle R \rangle = 1.10$, and that of the sport&sporting competition class is the lowest, $\langle R \rangle = 0.30$. The density of the cumulative returns of the business sector class is significantly different from that of random trading strategies, because the p value of the Kolmogorov-Smirnov test is 1.57×10^{-33}. On the other hand, the density of the cumulative returns of the sport&sporting competition class resembles that of the random trading strategies, because the p value of the Kolmogorov-Smirnov test is 0.25. Stock markets notably react to fluctuations in the number of business sector class news articles. Table 2 expresses the top 10 topics for cumulative returns in the business sector class. The "Metal/Mining" topic is especially related with future changes of the S&P500 index.

Table 1 Number of topics and mean of cumulative returns $\langle R \rangle$ for each topic classification (see Section 2). Δt is 7weeks (see Section 3)

Classification	Number of Topics	Mean of Return
Business Sector	199	1.10
Event Type	24	1.08
Commodity	53	0.93
Organization	16	0.46
Asset Class / Property	85	0.37
Geography	207	0.33
Sport & Sporting Competition	25	0.30

Table 2 Top 10 topics for cumulative return in business sector class

Ranking	Topic Name	Cumulative Return
1	Metal/Mining	3.58
2	Entertainment Production	3.46
3	Food Processing	3.17
4	Food & Drug Retailing	3.01
5	Leisure & Recreation	2.97
6	Aluminum	2.94
7	Real Estate	2.84
8	Restaurants	2.78
9	Utilities	2.74
10	Media/Publishing	2.50

[1] We don't use the following classifications that number of topics is less than 10. "Arts/Culture/Entertainment," "Crime," "Currency," "Disaster/Accident," "Genre," "Health/Medicine," "Indicator Type," "Intellectual Property," "Language," "Legacy News Topic," "News Flag/Status," "Physical Asset Type," "Religion," "Sport combined with Geography")

6 Cumulative returns for every year.

We observed the density of the cumulative return for every year from 2008 to 2011. Fig. 3(a) is a time series of the S&P500 index. Falls occur between September 2008 and March 2009 and between August 2011 and October 2011. Fig. 3(b) is the probability density distribution of the annual cumulative returns of the S&P500 index based on the fluctuations in the number of news articles of business sector topics. Although the cumulative returns are different every year, the performance of the trading strategies is always good. The means of cumulative returns $\langle R \rangle$ are positive every year. In the stock falls of 2009 and 2011, the mean of cumulative returns $\langle R \rangle$ is especially high. In the Kolmogorov-Smirnov test, the null hypothesis is rejected at a significance level of 1% every year.
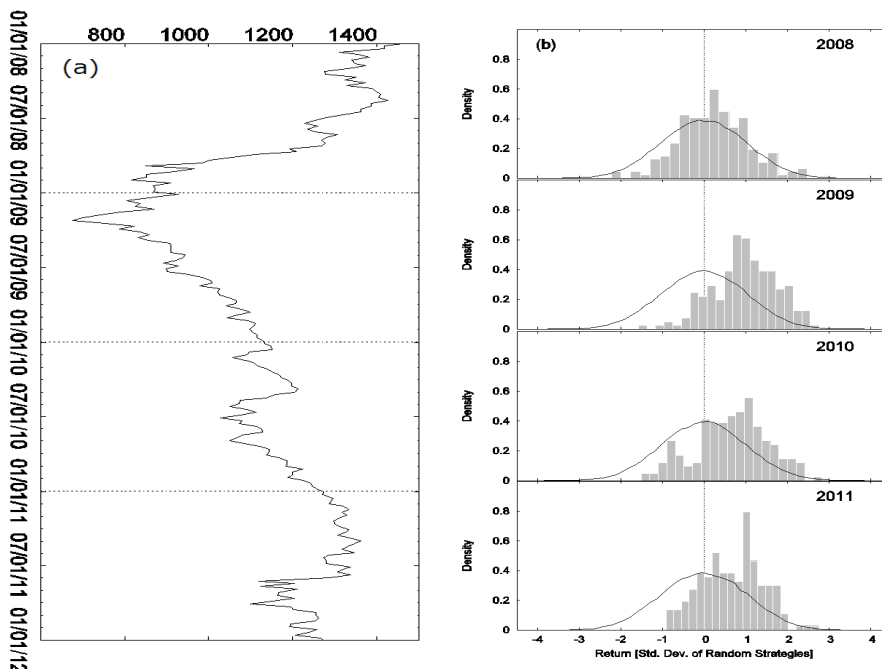

Fig. 3 (a)Time series of S&P500 index from 2008 to 2011. (b)Probability density distributions of annual cumulative returns of S&P500 index based on fluctuations in number of news articles of business sector topics. The curves represent distributions of annual cumulative returns of random trading strategies. Δt is 7 weeks (see Section 3). Means of cumulative returns are 0.22, 0.97, 0.66 and 0.67 from 2008 to 2011. p values of Kolmogorov-Smirnov test are 1.40×10^{-3}, 6.03×10^{-35}, 2.94×10^{-15} and 5.33×10^{-16}, respectively.

7 Conclusion

We showed the correlation between the fluctuations in the number of news articles about a stock market and the future changes of its stock market index using 9,150,000 news articles distributed by Thomson Reuters and stock market indexes between December 2007 and April 2012. When the number of news articles about the listed companies in NYSE and NASDAQ increase/decrease in a week, the S&P500 index tends to fall/rise in the next week. On the other hand, the fluctuations in the number of these news articles are rarely correlated with the changes of other stock market indexes. Stock markets notably react to the fluctuation in the number of business sector news articles. These characteristics are observed every year. Prospect theory in behavioral economics argue that investors are more concerned about avoiding loss. Therefore, when a stock might fall, they need more information about it, and news media notably report such information. The number of news articles often increases before the stock falls.

A trading strategy based on the number of news articles can get returns as high as trading strategies based on Wikipedia views. For example, Moat et al. (2013) virtually traded DJIA using the views of 285 pages listed in the "General economic concept" subsection on Wikipedia Moat et al. (2013). The means of the cumulative returns ranged from 0.19 to 0.89 per year. On the other hand, when the S&P500 index is virtually traded using business sector news articles, the means of the cumulative returns were 0.22, 0.97, 0.66 and 0.67 from 2008 to 2011. These tendencies suggest differences in the quality between the news from financial information providers and web sources. Future work will investigate this speculation.

We can also predict the changes of other asset price. For example, when the Japanese government's ten-year bonds are traded using business sector news in Japanese, we get the mean of cumulative returns 0.35 for the same period. Another future work will identify the relationship between various financial markets and the number of news articles.

References

Jiang ZQ, Guo L & Zhou WX (2007) Endogenous and exogenous dynamics in the fluctuations of capital fluxes. The European Physical Journal B 57(3):347-355.

Sornette D (2006) Endogenous versus Exogenous Origins of Crises. In: S. Alberio, V. Jentsch & H. Kantz (ed) Extreme events in nature and society, Springer Berlin Heidelberg, pp 95-119

Newman ME (2005) Power laws, Pareto distributions and Zipf's law. Contemporary Physics 46(5):323-351.

Mizuno T, Ohnishi T & Watanabe T (2015) Novel and topical business news and their impact on stock market activities. arXiv preprint arXiv:1507.06477

Hisano R, Sornette D, Mizuno T, Ohnishi T & Watanabe T (2013) High quality topic extraction from business news explains abnormal financial market volatility. PLoS ONE 8(6):e64846.

Alanyali M, Moat HS & Preis T (2013) Quantifying the relationship between financial news and the stock market. Scientific Reports 3:3578.

Mizuno T, Takei K, Ohnishi T & Watanabe T (2012) Temporal and cross correlations in business news. Progress of Theoretical Physics Supplement 194:181-192.

Petersen AM, Wang F, Havlin S & Stanley HE (2010) Market dynamics immediately before and after financial shocks: Quantifying the Omori, productivity, and Bath laws. Physical Review E 82(3):036114.

Rangel JG (2011) Macroeconomic news, announcements, and stock market jump intensity dynamics. Journal of Banking & Finance 35(5):1263-1276.

Bollen J, Mao H & Zeng X (2011) Twitter mood predicts the stock market. Journal of Computational Science 2(1):1-8.

Schumaker RP & 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.

Thelwall M, Buckley K, Paltoglou G, Cai D & Kappas A (2010) Sentiment strength detection in short informal text. Journal of the American Society for Information Science and Technology 61(12):2544-2558.

Tversky A & Kahneman D (1991) Loss aversion in riskless choice: A reference-dependent model. The Quarterly Journal of Economics 106(4):1039-1061.

Preis T, Moat HS & Stanley HE (2013) Quantifying trading behavior in financial markets using Google Trends. Scientific Reports 3:1684.

Moat HS, Curme C, Avakian A, Kenett DY, Stanley HE & Preis T (2013) Quantifying Wikipedia usage patterns before stock market moves. Scientific Reports 3:1801.

Curme C, Preis T, Stanley HE & Moat HS (2014) Quantifying the semantics of search behavior before stock market moves. Proceedings of the National Academy of Sciences 111(32):11600-11605.

Thomson Reuters (2008) Reuters Newsscope Sentiment Engine:Guide to sample data and system overview. Thomson Reuters White Paper

Thomson Reuters (2008) Reuters Newsscope Sentiment Engine:Output image and file format. Thomson Reuters White Paper