

# Trading strategy of a stock index based on the frequency of news releases for listed companies

Yoshifumi Tahira<sup>1</sup> · Takayuki Mizuno<sup>2</sup>

Published online: 10 September 2016

© The Author(s) 2016. This article is published with open access at Springerlink.com

**Abstract** The browsing frequency of Wikipedia pages for companies listed in the Dow Jones Industrial Average (DJIA) has been shown to be related to future DJIA changes. The number of Wikipedia page views increases after new information for these companies is released. Therefore, the frequency of listed companies' news releases often reflects future stock market conditions. We show that the trading strategy performance of a stock index based on the frequency of news releases is better than that of a trading strategy that randomly buys or sells its stock index. When the number of news articles for companies listed on the NYSE and NASDAQ increases/decreases in a week, the Standard & Poor's 500 index (S&P500 index) tends to fall/rise the following week. In particular, the trading strategy performance using news articles for the business sector is good. We confirmed these characteristics for the period of December 2007 to April 2012.

**Keywords** Econophysics · Stock market · Business news · Exogenous shock

**JEL Classification** G11 · G14 · G17

---

✉ Takayuki Mizuno  
mizuno@nii.ac.jp

Yoshifumi Tahira  
ytahira@phys.chuo-u.ac.jp

<sup>1</sup> Chuo University Graduate School of Science and Engineering, 1-13-27 Kasuga, Bunkyo-ku, Tokyo, Japan

<sup>2</sup> 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

## 1 Introduction

Asset prices are moved by an endogenous mechanism that follows past price changes created by trend followers and by an exogenous mechanism of reactions to sudden financial market news (Jiang et al. 2007; Sornette 2006). Investigation of historical tick-by-tick data of prices has shown that endogenous mechanisms follow a multiplicative process. 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 et al. 2015; Hisano et al. 2013; Alanyali et al. 2013; Mizuno et al. 2012; Petersen et al. 2010; Rangel 2011). Natural language processing for news articles has identified a relationship between price changes and text words in news articles (Bollen et al. 2011; Schumaker and 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 and Kahneman 1991). When there is a risk of a stock falling, people desperately gather information on this stock to avoid that risk. Therefore, the number of internet hits for the stock increases before its price falls. Preis et al. showed that the browsing frequency of Wikipedia pages for 30 companies listed in the Dow Jones Industrial Average (DJIA) is related to future DJIA changes (Preis et al. 2013; Moat et al. 2013; Curme et al. 2014). The number of Wikipedia page views increases before the DJIA falls. However, we cannot determine why people look at a Wikipedia page from only the number of views.

Since news media concentrates on topics that attract attention, the number of news articles on a certain topic might increase before a stock market index falls, as in the relationship between the browsing frequency of Wikipedia pages and DJIA. In this paper, we show that the trading strategy performance of a stock index based on the frequency of news releases for the stock market is better than that of a trading strategy that randomly buys or sells the stock index. This result suggests that the frequency of news releases is related to future stock market conditions. We can understand why news attracts attention because news articles contain detailed information of events.

The remainder of this article is organized as follows: In Sect. 2, we explain the Thomson Reuters' news data. In Sect. 3, we introduce a stock index trading strategy based on the frequency of news releases for listed companies. In Sect. 4, we apply this trading strategy to the S&P500 index. In Sect. 5, we review typical news topics where the trading strategy performance is good. In Sect. 6, we list the annual performance from 2008 to 2011 and we discuss our conclusions in Sect. 7.

## 2 Business news data

We used news articles from December 2007 to April 2012 that were distributed for institutional investors by Thomson Reuters, one of the world's largest financial information providers (Thomson Reuters 2008a, b). These news articles were

classified into alert, headline, and story take overwrite categories. Alert is urgent news that 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 focused on only the 9,150,000 news articles classified as alerts and headlines to amass a substantial number of news articles.

Thomson Reuters appends topic codes to the news articles. These topic codes are classified into 22 groups: “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”. Each group is composed of related topic codes. News articles also have a stock name code if related to a stock.

In this study, we used only news articles containing codes reported in more than 130 weeks out of 230 weeks between December 2007 and April 2012. For example, news articles containing the “IBM” code were reported in 230 weeks. Using this rule, the number of stock name codes and topic codes were 500 and 676, respectively.

### 3 Trading strategy based on the frequency of news releases

We introduce a stock market index trading strategy based on the frequency of news releases. We represent the number of news articles for listed company  $i$  in week  $t$  as  $n_i(t)$  and calculate the average number of news articles for the previous  $\Delta t$  weeks:  $N_i(t-1, \Delta t) = \sum_{\tau=1}^{\Delta t} n_i(t-\tau)/\Delta t$ . If the number of news articles increases in week  $t$  such that  $n_i(t) > N_i(t-1, \Delta t)$ , we sell a unit of the stock market index at closing price  $p(t+1)$  on the first trading day of week  $t+1$  and 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 decreases in week  $t$  such that  $n_i(t) \leq N_i(t-1, \Delta t)$ , we buy a unit of the stock market index at closing price  $p(t+1)$  on the first trading day of week  $t+1$  and 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 the cumulative returns  $R_i$  of this trading strategy between December 2007 and April 2012. To confirm the performance of this trading strategy, we also introduce a trading strategy that randomly buys or sells the stock market index. The cumulative returns  $R_i$  are normalized by the standard deviation from 100,000 random trading strategies. We investigated the distribution of cumulative returns  $R_i$  for listed companies in each stock market.

### 4 Statistical test of trading strategy performance

We show the performance of the trading strategy introduced in Sect. 3 using the S&P500 index, the Nikkei 225 index, and news articles on companies listed on the NYSE and NASDAQ.

Figure 1a shows the probability density distribution of cumulative returns  $R_i$  when the S&P500 index is traded based on the frequency of news releases for companies listed on the NYSE and NASDAQ for the previous  $\Delta t = 7$  weeks. The mean of the cumulative returns  $\langle R \rangle = 0.40$  is significantly higher than the mean using random trading strategies  $\langle R \rangle = 1.3 \times 10^{-3}$ . Since the  $p$  value of the Kolmogorov–Smirnov test is  $1.58 \times 10^{-11}$ , the following null hypothesis is rejected: “The distribution of trading strategies using the frequency of news releases is the same as that of random trading strategies.”

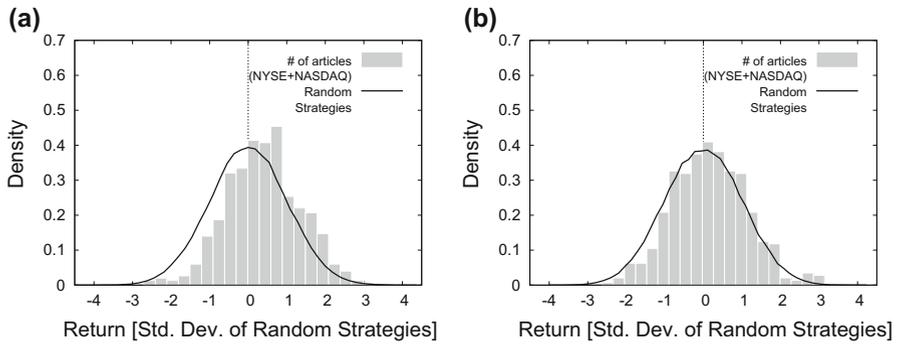
Figure 1b displays the probability density distribution of cumulative returns  $R_i$  when the Nikkei 225 index is traded based on the frequency of news releases for companies listed on the NYSE and NASDAQ for the previous  $\Delta t = 7$  weeks. In this trading strategy, the mean of the cumulative returns  $\langle R \rangle$  is 0.17 and the  $p$  value of the Kolmogorov–Smirnov test is 0.041. The null hypothesis is also rejected at a significance level of 5%. However, the distribution of cumulative returns of this trading strategy is similar to that of the random trading strategies because the  $p$  value of Fig. 1b is significantly larger than that of Fig. 1a.

We investigated the  $\Delta t$  dependence of the cumulative returns. Figure 2 shows the relationship between the means of cumulative returns  $\langle R \rangle$  and  $\Delta t$  weeks when the S&P500 and Nikkei 225 indexes are traded based on the frequency of news releases for companies listed on the NYSE and NASDAQ for the previous  $\Delta t$  weeks. The mean of the cumulative returns  $\langle R \rangle$  of the S&P500 index is larger than that of the Nikkei 225 index for every  $\Delta t$ .

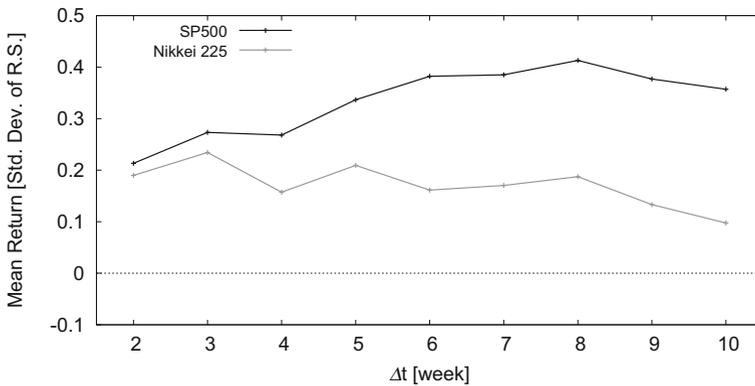
## 5 Trading strategy performance for each news topic

Thomson Reuters’ news articles have topic codes, as shown in Sect. 2. We measured the trading strategy performance using the S&P500 index and the news articles with each topic code.<sup>1</sup> For each group of topic codes, we show the number of topic codes and the mean of the cumulative returns  $\langle R \rangle$  in Table 1. The mean of cumulative returns  $\langle R \rangle$  using news articles for the business sector is the highest,  $\langle R \rangle = 1.10$ . The distribution of the cumulative returns using the business sector’s news is significantly different from that of the random trading strategies because the  $p$  value of the Kolmogorov–Smirnov test is  $1.57 \times 10^{-33}$ . On the other hand, the mean using the sport&sporting competition’s news is the lowest at  $\langle R \rangle = 0.30$ . The distribution of the cumulative returns using the sport&sporting competition’s news is closely similar to that of the random trading strategies because the  $p$  value of the Kolmogorov–Smirnov test is 0.25. Stock markets notably react to the frequency of news releases for the business sector. Table 2 displays the top 10 topics for cumulative returns in the business sector group. The “Metal/Mining” topic is especially related to future changes of the S&P500 index.

<sup>1</sup> We do not use the following 14 topic groups for which the number of topic codes 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”.



**Fig. 1** Probability density distributions of cumulative returns  $R_t$  when **a** S&P500 index and **b** Nikkei 225 index are traded based on the frequency of news releases for companies listed on the NYSE and NASDAQ. The *curves* represent the distribution of cumulative returns from random trading strategies.  $\Delta t$  is 7 weeks (see Sect. 3)



**Fig. 2** Means of cumulative returns  $\langle R \rangle$  based on the frequency of news releases for companies listed on the NYSE and NASDAQ for previous  $\Delta t$  weeks. *Black* and *gray* indicate the means of cumulative returns  $\langle R \rangle$  when the S&P500 index and Nikkei 225 index are traded, respectively

**Table 1** Number of topic codes and mean of cumulative returns  $\langle R \rangle$  for each group of topic codes (see Sect. 2)

Topic group	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 and sporting competition	25	0.30

$\Delta t$  is 7 weeks (see Sect. 3)

## 6 Annual trading strategy performance

Figure 3a expresses a time series of the S&P500 index. The index fell between September 2008 and March 2009 and between August 2011 and October 2011. Figure 3b shows the probability density distribution of the annual cumulative returns for 2008 to 2011 when the S&P500 index was traded using the business sector's news. 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. The mean of cumulative returns  $\langle R \rangle$  is especially high during the 2009 and 2011 stock market crises. In the Kolmogorov–Smirnov test, the null hypothesis is always rejected at a significance level of 1 % every year.

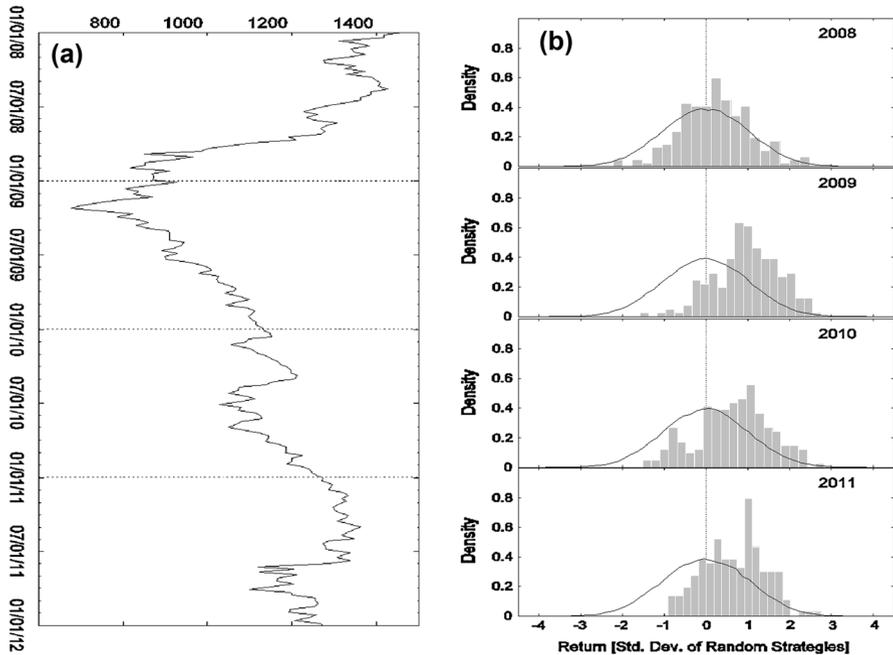
## 7 Conclusion

We showed that the trading strategy performance of a stock index based on the frequency of news releases is better than that of a trading strategy that randomly buys or sells its stock index. When the number of news articles for companies listed on the NYSE and NASDAQ increases/decreases in a week, the S&P500 index tends to fall/rise the following week. On the other hand, the trading strategy performance of the Nikkei 225 index using news articles for companies listed on the NYSE and NASDAQ closely resembles the performance of the random trading strategy. Therefore, the frequency of news releases for companies listed in a stock market reflects the future stock market conditions. We also showed that the trading strategy performance using news articles for the business sector is especially good. These characteristics were confirmed for each year from 2008 to 2011.

We set weekly trading intervals in this paper. If we trade a stock market index at longer intervals, the performance of our trading strategy will decline because the stock market is affected by various economic factors and the social environment. If we trade a stock market index at shorter intervals, the performance of our trading strategy will also decline because the stock market is affected by market microstructure noise. We will determine the best trading intervals in the future.

**Table 2** Top 10 topics for cumulative returns in the business sector group

Ranking	Topic name	Cumulative return
1	Metal/mining	3.58
2	Entertainment production	3.46
3	Food processing	3.17
4	Food and drug retailing	3.01
5	Leisure and 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



**Fig. 3** **a** Time series of the S&P500 index from 2008 to 2011. **b** Probability density distributions of annual cumulative returns when the S&P500 index is traded using the business sector’s news. The curves represent the distributions of annual cumulative returns of random trading strategies.  $\Delta t$  is 7 weeks (see Sect. 3). The means of the cumulative returns are 0.22, 0.97, 0.66, and 0.67 from 2008 to 2011. The  $p$  values of the Kolmogorov–Smirnov test are  $1.40 \times 10^{-3}$ ,  $6.03 \times 10^{-35}$ ,  $2.94 \times 10^{-15}$ , and  $5.33 \times 10^{-16}$ , respectively

Moreover, in our future work, we aim to clarify the trading strategy performance for various financial markets. For example, we can confirm that the mean of cumulative returns is positive when the Japanese government’s ten-year bonds are traded using Japanese business sector news.

**Acknowledgments** T.M. was subsidized by the Ishii Memorial Securities Research Promotion Foundation and JSPS KAKENHI Grant Numbers 24710156 and 25220502. The authors thank the Yukawa Institute for Theoretical Physics at Kyoto University. Discussions during the YITP workshop YITP-W-15-15 on “Econophysics 2015” were useful to complete this work.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

Alanyali M, Moat HS, Preis T (2013) Quantifying the relationship between financial news and the stock market. *Sci Rep* 3:3578  
 Bollen J, Mao H, Zeng X (2011) Twitter mood predicts the stock market. *J Comput Sci* 2(1):1–8

- Curme C, Preis T, Stanley HE, Moat HS (2014) Quantifying the semantics of search behavior before stock market moves. *Proc Natl Acad Sci* 111(32):11600–11605
- 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
- Jiang ZQ, Guo L, Zhou WX (2007) Endogenous and exogenous dynamics in the fluctuations of capital fluxes. *Eur Phys J B* 57(3):347–355
- Mizuno T, Takei K, Ohnishi T, Watanabe T (2012) Temporal and cross correlations in business news. *Prog Theor Phys Suppl* 194:181–192
- Mizuno T, Ohnishi T, Watanabe T (2015) Novel and topical business news and their impact on stock market activities. *arXiv preprint arXiv:1507.06477*
- Moat HS, Curme C, Avakian A, Kenett DY, Stanley HE, Preis T (2013) Quantifying Wikipedia usage patterns before stock market moves. *Sci Rep* 3:1801
- Newman ME (2005) Power laws, Pareto distributions and Zipf's law. *Contemp Phys* 46(5):323–351
- Petersen AM, Wang F, Havlin S, Stanley HE (2010) Market dynamics immediately before and after financial shocks: quantifying the Omori, productivity, and Bath laws. *Phys Rev E* 82(3):036114
- Preis T, Moat HS, Stanley HE (2013) Quantifying trading behavior in financial markets using Google Trends. *Sci Rep* 3:1684
- Rangel JG (2011) Macroeconomic news, announcements, and stock market jump intensity dynamics. *J Bank Finance* 35(5):1263–1276
- Schumaker RP, Chen H (2009) Textual analysis of stock market prediction using breaking financial news: the AZFin text system. *ACM Trans Inf Syst (TOIS)* 27(2):12
- Sornette D (2006) Endogenous versus exogenous origins of crises. In: Albeverio S, Jentsch V, Kantz H (eds) *Extreme events in nature and society*. Springer, Berlin, pp 95–119
- Thelwall M, Buckley K, Paltoglou G, Cai D, Kappas A (2010) Sentiment strength detection in short informal text. *J Am Soc Inf Sci Technol* 61(12):2544–2558
- 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
- Tversky A, Kahneman D (1991) Loss aversion in riskless choice: a reference-dependent model. *Q J Econ* 106(4):1039–1061