Big data, news diversity and financial market crash

Sabri Boubaker, Zhenya Liu, Ling Zhai

Version of Record: https://www.sciencedirect.com/science/article/pii/S0040162521001876
Manuscript_e436649e174fda46ae3e6b8f4be591e5

Big Data, News Diversity and Financial Market Crash

Sabri Boubaker[a,b], Zhenya Liu[c,d*], Ling Zhai[c]

[a] EM Normandie Business School, Métis Lab, France
[b] International School, Vietnam National University, Hanoi, Vietnam
[c] School of Finance, Renmin University of China, China
[d] CERGAM, Aix-Marseille University, France

Abstract

A vast quantity of high-dimensional, unstructured textual news data is produced every day, more than two decades after the launch of the global Internet. These big data have a significant influence on the way that decisions are made in business and finance, due to the cost, scalability, and transparency benefits that they bring. However, limited studies have fully exploited big data to analyze changes in news diversity or to predict financial market movements, specifically stock market crashes. Based on modern methods of textual analysis, this paper investigates the relationship between news diversity and financial market crashes by applying the change-point detection approach. The empirical analysis shows that (1) big data is a relatively new and useful tool for assessing financial market movements, (2) there is a relationship between news diversity and financial market movements. News diversity tends to decline when the market falls and volatility soars, and increases when the market is on an upward trend and in recovery, and (3) the multiple structural breaks detected improve the ability to forecast stock price movements. Therefore, changes to news diversity, embedded in big data, can be a useful indicator of financial market crashes and recoveries.

Keywords: Big data; News diversity; Textual analysis; Change-point; Financial Crisis
JEL codes: C81, G01, G14, O16

* Corresponding author. Add: No. 59 Zhongguancun Street, Haidian District Beijing, 100872, P.R. China
Email address: sabri.boubaker@gmail.com (Sabri Boubaker), zhenya.liu@ruc.edu.cn (Zhenya Liu), lingzhai@ruc.edu.cn (Ling Zhai)

1. Introduction

Big textual data has gradually developed into an active research topic in economics and finance following the rapid, unprecedented development of the Internet and computer technology (e.g., Jegadeesh and Wu, 2013; Burnap et al., 2015; Baker et al., 2016; Zhang et al., 2016; Kayser and Blind, 2017; Haroon and Rizvi, 2020). The information most commonly used in research is usually derived from daily news (Thorsrud, 2020). For instance, Liao et al. (2019) examine a novel dataset of news events and find that media sentiment plays an important role in mergers and acquisitions. Moreover, firms associated with more negative media sentiment, stemming from news outlets, are more likely to be targeted by hedge funds (Wang and Wu, 2020).

The financial transaction data reflects traders' decision-making (buying or selling), which affects the price movements and volatility of the financial market (Simon, 1955). These decisions may be affected by various types of information, which stem from the traders' environments. In modern society, interaction with the Internet is becoming our main method of sourcing information (Vespignani, 2009; King, 2011; Vosen and Schmidt, 2011). Alanyali et al. (2013) point out that traders can not only actively obtain information by searching for it online, but also receive information by actively, or even passively, receiving news from influential financial media.

The financial media have a certain degree of interaction with the dynamics of the financial markets. Authoritative financial media plays a crucial role in disseminating information to investors. News can have a significant effect upon investor sentiment and this is reflected in their decisions. Moreover, when financial markets experience violent fluctuations, the financial media is expected to disseminate information promptly to the public, to track the situation as it unfolds, and to provide regular, professional reports. For instance, Alanyali et al. (2013) find a positive correlation between the frequency of instances that the Financial Times mentions a company and the daily trading volume of the stocks on the previous day and the day of the news release, suggesting that financial markets and news are intrinsically interrelated. Further, Curme et al. (2017) note that interesting examples of news diversity can be found within news from the Financial Times, one of the major financial media outlets. Curme et al. (2017) reveal that there is a forecast effect of news diversity on the trading volume, and news diversity tends to drop when the stock market falls. However, they do not find a relation between news diversity and subsequent stock price movements.

The idea of news diversity as being a measure of news importance is simple. If the media is reporting similar events across multiple outlets, topics covered by the news will be concentrated. This means that such events are highly influential and can have a significant impact on movements within the financial market. Therefore, the level of news diversity is closely correlated with the financial market movements. The more concentrated the news diversity index, the more accurately it will reflect market movements.

In this context, we emphasize that there is a natural connection between news diversity and the financial markets and they can influence each other. The objective of our paper is to further examine the relationship between news diversity and the movements of financial markets, and to assess the prediction effect of news diversity on market prices and volatility of the financial market. We examine the news published by the Financial Times during the period of 2007 to 2009 and 2018 to 2019 from the online database ProQuest. We use it to construct a news diversity index based on Curme et al. (2017) and apply change-point detection methodology.

This paper analyzes whether news topics are significantly concentrated prior to stock market crashes and whether the media covers more concentrated topics prior to market recoveries (these being marked by a continuous increase in stock prices and a decline in volatility). It is precisely because of the mutually influential relationship between news diversity and financial markets that we use the method of change-point detection, based on the cumulative sum, to identify significant changes in news diversity. This paper provides evidence that news diversity can be used to accurately forecast the movements of the stock market, by using change-point analysis.

This paper contributes to the literature in several ways. First, it combines news media and financial markets data and uses a natural language processing technology to innovatively extract information from the news. It then uses this to improve the ability to predict financial market movements. Second, instead of studying the forecast of trading volume (Curme et al., 2017), it focuses on the forecast of financial market movements and documents an early warning effect of news diversity on market prices and volatility, which is more important to the market and investors. Third, it uses the change-point detection method, a mathematical technique, to explore this relationship, rather than applying conventional empirical methods, as has been done within previous studies (Birz and Lott, 2013; Zhang et al., 2017; Curme et al.,

2017)[1].

The remainder of the paper is organized as follows. Section 2 presents the theoretical background and hypotheses, Section 3 introduces the methodology and data, Section 4 discusses the empirical findings, and Section 5 presents several robustness tests. Section 6 concludes.

[1] This method is supplemented by financial time series models for robustness.

2. Theoretical background and hypotheses

Quantified by Shannon's Entropy (Shannon, 1948), news diversity measures the news content's cohesiveness across media outlets, which assesses whether the news concerns few topics or is dispersed among a high number of topics (Curme et al., 2017). More precisely, if a high number of "words" focus on the same topics, a low diversity news index should reflect the significance of the news topics and this should correlate with the main trends in the financial markets. This section discusses the theoretical background of how news diversity affects financial market movements.

2.1. Information in news

Authoritative financial media, such as the Financial Times, the Wall Street Journal, and the New York Times, is a major mechanism for disseminating information to investors (Fang and Peress, 2009; Griffin et al., 2011). Business news plays an important role in the way investors value stocks. Financial media can provide the market with fundamental information and continuously influences the decision-making of investors. The efficient market hypothesis (EMH) suggests that stock prices are a timely and adequate reflection of information (Malkiel and Fama, 1970). Therefore, no matter what type of news, whether it be political, environmental, social, financial, or economic, it will always cause investors to continuously adjust their expectation of the future returns that a firm's investment projects can achieve. The adjustment of such expectations can be reflected in the adjusted demand and supply functions (Hisano et al., 2013; Hagenau et al., 2013). Through these two functions, stock prices can rise and fall. Therefore, both the news and investor expectations of future cash flows brought by the media, play a vital role in stock pricing.

2.2. Investor sentiment and news

A considerable amount of research has been devoted to linking news with financial markets by using text mining to predict market dynamics, such as closing prices, trading volume, and volatility (e.g., Tetlock, 2007; Engelberg and Parsons, 2011; Kleinnijenhuis et al., 2013; Atkins et al., 2018). Contrary to the efficient market hypothesis, behavioral finance emphasizes the role of behavioral and emotional factors in investors' financial decision making. Recent studies on investor psychology and social sentiment show that financial markets respond to news asymmetrically. Financial media can influence the stock market considerably and can magnify the animal instincts of investors. Investors are more sensitive to negative news when the

market is held hostage by uncertainty, and expectations for future earnings become increasingly uncertain. According to the “framework effect”, investors’ reactions to negative news are asymmetrical (Tan and Chua, 2004; Antweiler and Frank, 2004). News, which acts as the most authoritative and extensive method of media messaging, has been shown to play an important role in shaping investor sentiment. For instance, Tetlock (2007) finds that the high level of media pessimism, stemming from newspapers, puts downward pressure on market prices, and unusually high or low pessimism leads to high levels of the market trading volume. Schumaker and Chen (2009) show that the model containing both article terms and the stock price has the best performance in terms of proximity to the future stock price. According to the “framework effect”, investor sentiment can provide a better perspective when studying the correlation between news and market movements, particularly during periods of financial crashes. Many research studies have focused on similar periods of crisis, examining whether the authoritative and influential newspapers may have a greater impact on investor sentiment in instances of financial turmoil (Preis et al., 2010; Bollen et al., 2011).

2.3. Hypotheses

News diversity derived from daily news measures the degree of concentration of news topics and categorizes the most useful information for investors. According to EMH, investors are exposed to centralized information and adjust their expectations promptly, leading to immediate changes in the market’s supply-demand curve, thus driving stock price adjustments. Therefore, in terms of the information embedded within the news, it is to be expected that news diversity correlates with market movements and should indicate whether the financial market pays more attention to certain types of events.

This paper’s primary focus is on the correlation between news diversity and financial market movements during financial crashes, marked by extremely high volatility, which can be a relatively effective indicator of financial market movements.

Behavioral finance theory provides a possible explanation for the correlation between decreasing news diversity and increasing market volatility. Specifically, when influential financial media outlets begin reporting negative news, the negative news becomes increasingly concentrated, and investors tend to overreact, making more extreme financial decisions than they usually would, owing to their pessimistic mood and sentiment. Thus, decreasing levels of diversity may be associated with increasing financial market volatility. When the news begins reporting once again on

more diverse topics, the financial market gradually stabilizes, investor confidence starts to recover, financial decisions are swiftly adjusted, and financial market volatility decreases accordingly[2]. Therefore, our first hypothesis is as follows:

H1: There is a negative correlation between news diversity and financial market volatility.

All market investors carry out financial adjustments, as necessary, in a timely manner after they receive various pieces of news. Consequently, the movements of the financial market should follow the release of information. In reality, before the market collapses, the media has already begun to report on the bear market, referring to alarming cases, worsening market fundamentals, and the concerns of economists. Moreover, investors' pessimistic moods can aggravate the situation. As a result, the news becomes increasingly concentrated, leading to a decreasing diversity index before reaching extreme levels of volatility and the market collapses. Contrastingly, when the news focuses on more diverse topics and disseminates increasingly encouraging news, investors adjust their financial decisions correspondingly and the financial market begins to recover. In line with the above reasoning, our second hypothesis is as follows:

H2: News diversity can be treated as a signal for predicting financial crashes and recoveries.

[2] This paper focus on the relationship between news topic diversity and financial market movements. Information in news and investor sentiment discussed here are only possible effects of this mechanism.

3. Methodology and data

This section first introduces the method for modeling "hidden" topics in texts, this being the Latent Dirichlet Allocation (LDA). It also discusses the nonparametric method of change in the mean and the binary segmentation methods that are applied to detect multiple change-points in the series of news diversity. Moreover, it presents the textual data, data resources, and data preprocessing.

3.1. Methodology

Since texts naturally exhibit high dimensionality, the information they contain is difficult to quantify. The first step in making textual data easy to analyze consists of reducing its dimensions by clustering words into groups. Topic modeling is a commonly used tool for reducing dimensions. The text corpus is divided into documents, each of which is treated as a collection of unordered words or as a "word bag". LDA is one of the most popular methods for modeling "hidden" topics in texts (LDA; Blei et al., 2003). Following Curme et al. (2017), we construct a diversity index of news topics to gauge the concentration of news topics.

Based on the news diversity index[3], we use the signal-plus-noise model to detect change-points. N denotes the set of observations

$$X_t = \mu_t + \varepsilon_t, \qquad t \in N \tag{1}$$

where X_t $(t \in N)$ denotes the time series of news topic diversity index, μ_t $(t \in N)$ is the signal and ε_t $(t \in N)$ is the noise part, with $E[\varepsilon_t] = 0$ and $E[{\varepsilon_t}^2] = \sigma^2$.

We assume that the news diversity index is a series of dependent observations. $X_1, X_2, \ldots, X_n$, and test the null hypothesis that there is no-change in the mean of the news diversity H_t

$$H_0: \mu_1 = \cdots = \mu_n \equiv \mu$$

against the "one change in the mean" alternative

$$H_A: there\ \ is\ \ an\ \ integer\ \ k^*, 1 \le k^* < n, such\ \ that\ \ \mu_1 = \cdots = \mu_{k^*} \ne \mu_{k^*+1} = \cdots = \mu_n.$$

We are interested in the CUSUM procedures, $Z_n = (Z_n(x): x \in [0,1])$,

[3] Details are given in Appendix A.

$$Z_n(x) = \frac{1}{\sqrt{n}}\left(\sum_{t=1}^{\lfloor nx \rfloor} X_t - \frac{\lfloor nx \rfloor}{n}\sum_{t=1}^{n} X_t\right), x \in [0,1] \quad (2)$$

where $\lfloor \cdot \rfloor$ denotes the integer part.

As Aue and Horváth (2013) point out

$$Z_n \Rightarrow \omega B \quad (3)$$

as $n \to \infty$, where $B = (B(t): t \in [0,1])$, with $B(t) = W(t) - tW(1)$, is a Brownian bridge and the long-run variance ω^2 is unknown and has to be estimated with an estimator $\widehat{\omega}_n^2$, which comes to

$$\frac{1}{\widehat{\omega}_n} Z_n \Rightarrow B \quad (4)$$

as $n \to \infty$. Following Berkes et al. (2011), we apply the Bartlett estimator to compute the long-run variance.

We also detect multiple change-points in the news diversity by introducing a binary segmentation method, following Horváth et al. (2017).

3.2. Textual data

This study uses daily news from the Financial Times (accessible through the ProQuest Database) and the daily closing prices of the FTSE 100 and S&P 500 to investigate the relationship between the diversity of topics appearing in the news and the financial market. The sample period is from 01/02/2007 to 12/31/2009. We preprocess data before analyzing the textual data using the LDA model, including changing all words to lowercase, changing hyphens to whitespace, leaving only letters "a" to "z", removing all single-letter words, stemming stopwords, and undertaking other data preprocesses, as in Curme et al. (2017).

4. Empirical analysis

This section discusses news diversity over the sample period, the multiple change-point detection of news diversity, the movements of the financial market at news diversity's structural breaks, and the financial events which occur around these structural breaks. We do this in order to analyze whether news diversity is an effective predictor of financial market movements using change-point analysis.

4.1. News diversity over 2007-2009

Once the LDA is trained, each document (paragraph) m, in the whole corpus of daily news in the Financial Times from 01/02/2007 to 12/31/2009, can be represented by the K-dimensional topic vector $\theta_m = (\theta_{m,1}, \theta_{m,2}, \dots, \theta_{m,K})$. The news diversity index is then calculated following the methods of Curme et al. (2017).

As proposed in Curme et al. (2017), news diversity during the weekend is generally lower than it is during the working week. We therefore choose to drop the data on Saturdays to obtain a smoother sequence. Additionally, the topics covered by the news are naturally associated with each other. When the news breaks, the media will follow it. As a result, the diversity of daily news is not expected to be independent[4].

4.2. News diversity and the financial market

4.2.1. Granger causality test

We perform a Granger causality test to examine the relationship between news diversity and financial markets, specifically to see whether there is a connection between levels of news diversity and the movements of the financial market in general. Prior to the Granger causality test, we first test for unit roots of the sequence of news diversity, FTSE 100, and S&P 500 daily closing price sequences (ADF, PP, DF-GLS). The closing price series are expressed in natural logarithms. Table 1 shows that the closing price series of the FTSE 100 and S&P 500 are $I(1)$, and the news diversity is a stationary series.

[4] The independence of news diversity can be rejected by the white noise test.

Table. 1

Test for unit roots.

Variable	ADF	DF_GLS	PP
Diversity (Level)	-5.373	-4.617	-25.295
FTSE 100 (Level)	-1.257	-1.411	-1.251
FTSE 100 (1st difference)	-13.889	-13.691	-30.686
S&P 500 (Level)	-1.093	-1.106	-1.234
S&P 500 (1st difference)	-23.239	-22.944	-32.373
CV 1%	-3.970	-3.480	-3.970
CV 5%	-3.416	-2.890	-3.416
CV 10%	-3.130	-2.570	-3.130

Table. 2

Granger causality test for diversity and FTSE 100 (and S&P 500).

Equation	Excluded	chi2	Prob≥chi2
L.FTSE	Diversity	12.905	0.000
L.FTSE	ALL	12.905	0.000
Diversity	L.FTSE	77.576	0.000
Diversity	ALL	77.576	0.000
L.SP	Diversity	14.964	0.000
L.SP	ALL	14.964	0.000
Diversity	L.SP	53.509	0.000
Diversity	ALL	53.509	0.000

Next, following Granger (1969), we carry out the Granger causality tests to test whether there is a relationship between news diversity and the first difference of stock market index.

The null hypothesis of the Granger causality test is that each of the other endogenous variables does not Granger-cause the dependent variable in the equation. Therefore, Table 2 and Table 3 show that news diversity “Granger causes” S&P 500 and FTSE 100 index and these two indices also “Granger cause” news diversity. This, therefore, demonstrates that a relationship exists between news diversity and the stock market. More importantly, it shows that past values of news diversity are useful for predicting the stock markets.

4.2.2. Change-point detection of news diversity

We use the change-point detection of news diversity to explain the movements of stock markets. In particular, we apply the change in the mean test and the binary segmentation method in the fixed sample of the news diversity and find three change points over the whole sample period, namely, on 06/27/2007, 04/16/2008, and 04/27/2009. According to the multiple change-point detection, 04/16/2008 is the first detected change-point over the whole sample (Period 1) that can be divided into two periods; before and after 04/16/2008. The change-point of the first half is 06/27/2007 (Period 2) and that of the second half is 04/27/2009 (Period 3).

Table. 3

Tests for structural breaks during the whole sample.

Period	Statistics	Location of structural breaks
01/02/2007-12/31/2009	1.875***	04/16/2008
01/02/2007-04/16/2008	1.403**	06/27/2007
04/17/2008-12/31/2009	1.249**	04/27/2009

Notes: This table shows the results of change-points detections over the whole sample. The corresponding critical values in the table are obtained from Monte-Carlo simulations with 10,000 replications for Brownian Bridge simulations. *** (**) denotes statistical significance at 1% (5%) threshold level.

Next, we focus on the sign of CUSUM, $S(k) - \frac{k}{n}S(n)$. The positive sign of $S(k) - \frac{k}{n}S(n)$ means that the change-point captures the decline of the news diversity and it is more likely to decrease compared with the previous period and vice versa.

Based on the calculation, we find that on 04/16/2008, the first change-point indicates that the topic diversity is about to decline over the second half of Period 1. The change-point at 06/27/2007 shows a decrease in news diversity over the later Period 2 and 04/27/2009 shows that the news diversity level is to increase over the second half of Period 3.

4.2.3. Movements of the stock market and changes in news diversity

It should be noted that the objective is not to detect the specific change-point and time when the topic diversity level started to increase or decrease, but to detect the time when the cumulated change becomes statistically significant. Therefore, in this section, we discuss stock market movements (in relation to prices and returns), during the period 2007-2009, focusing on the dynamics around the timing of the news diversity's change-points. Moreover, we investigate the nature of the financial events

reported by the news media, which affected the stock market before the change-points. We examine how the stock market reacted when the topic became more concentrated (or not).

We first examine the movements of the stock market (prices and returns of S&P 500 index and FTSE 100 index) at the time of the news diversity's change-points and combine this with our analysis of whether the news topics are more concentrated or not according to CUSUM and the movements of S&P 500 and FTSE 100 indices. On 06/27/2007, when the news diversity level dropped over the later Period 2, it is clear that prices began to fall and the returns were becoming increasingly volatile. The change-point of the whole period was on 04/16/2008, which indicates that the news was about to become concentrated on fewer topics. The change-point, 04/16/2008, appears to be a clear indicator of the dramatic decline of prices and the extreme fluctuation of returns to come during the trading days that followed.

We also consider whether the news diversity level could provide us with an early prompt or hint that the stock market is beginning to recover. In this respect, it is worth noting that we detect an additional change-point on 04/27/2009, which indicates an increase in topic diversity over the corresponding period from 04/27/2009 to 12/31/2009. It is, therefore, interesting to focus on the stock market movements after this change-point date. The change-point during Period 3 is 04/27/2009, where returns were less volatile compared to the significant turbulence from 04/16/2008. The S&P 500 index increased steadily after April 2009 and the levels of return volatility began to revert, almost back to their original levels.

To explore the relationship between news diversity and stock market movements, and to assess the prediction effect of the news on the stock market, we review the financial events that occurred around the time of the change-points. We do this in order to study what may have been attributed to the changes in news diversity levels and whether they were fully and immediately reflected on the stock market.

According to the St. Louis Federal Reserve Bank, the financial crisis of 2007-2009 may have started on 02/27/2007 when the Federal Home Loan Mortgage Corporation (Freddie Mac) made the announcement that it will no longer buy the riskiest subprime mortgages and mortgage-related securities[5]. However, this early-warning indicator of risk was not fully appreciated by the investors, particularly by the European financial institutions, who continued to purchase mortgage-related

[5] See https://fraser.stlouisfed.org/timeline/financial-crisis.

securities.

On 04/02/2007, the leading subprime mortgage lender, New Century Financial Corporation, called for Chapter 11 bankruptcy protection. The news outlets had begun to report on the negative financial situation[6]. The news was focusing on fewer topics relating to "credit", "loan", "bank", etc., and the change-point on 06/27/2007 can be considered as the starting point of the financial crisis of 2007-2009, which aligns with the worsening financial situation. French et al. (2010) outline that the first signs of the global financial crisis were also beginning to appear during the summer of 2007.

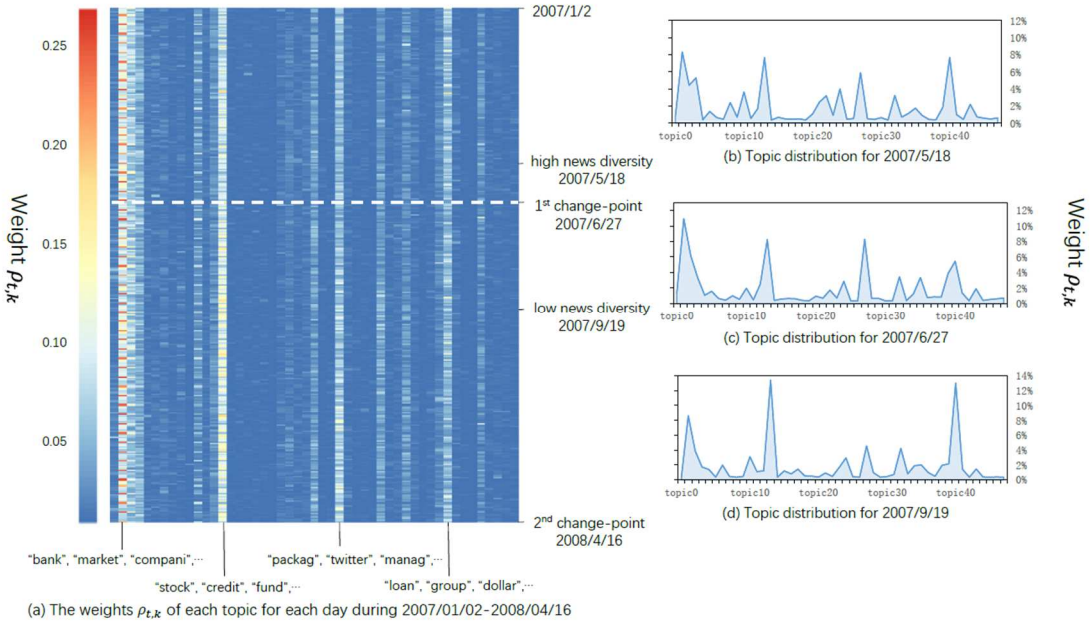

Fig. 1. The prominence of topics in the *Financial Times*.

Notes: (a) The weights $\rho_{k,t}$ of each topic k for each day t in Period 2. Highly weighted topics are annotated with three typical words. (b) The distribution of topics for the day 05/18/2007, which is earlier than the first change-point, exhibits a high news diversity index. (c) The distribution of topics for the day 06/27/2007, which is the first change-point, exhibits a relatively low news diversity index. (because the news diversity index in 06/27/2007 is lower than 05/18/2007, but higher than 09/19/2007) (d) The distribution of topics for the day 09/19/2007, which is later than the first change-point, exhibits a low news diversity index. The white dotted line indicates the time of the first change-point.

In order to further understand the levels of news diversity around the time of the first change-point, 06/27/2007, Fig. 1 shows the weight of news topics in Period 2 and the distribution of news topics at a selection of specific time points. It is clear in (a) that the topic containing words such as "stock", "credit", "fund", "package", "loan" is

[6] For instance, Standard and Poor's and Moody's Investor Services downgraded over 100 bonds backed by second-lien subprime mortgages and Bear Stearns informed investors that it is suspending redemptions in June 2007.

becoming more and more present in the news being reported, resulting in an increasingly concentrated news diversity index. Before the first change-point, this being 06/27/2007, the distribution of topics is more uniform, and each topic accounts for a relatively small proportion of the news reported (here we take the date 05/18/2007 as an example). Around the first change-point, the proportion of individual topics began to increase, with topics containing words such as "bank", "credit" and "package" standing out significantly. As discussed, after the first change-point, the level of news diversity declines, and the proportion of individual topics that contain words related to the financial crash far exceeds other common topics (here we take the date 09/19/2007 as an example).

The situation began to deteriorate in the summer of 2007 and worsened in early 2008[7]. The first detected change-point, 04/16/2008, shows that the news had focused on the negative situation within the financial market for an extended period of time. This has come to be recognized as an accurate early warning sign of the impending turmoil on Wall Street and the global financial crisis, indicated by the change-point in April 2008.

In the summer of 2008, increasing numbers of financial institutions were failing, such as Wachovia in June and IndyMac in July. This led to a dramatic drop in prices and an immense fluctuation in the returns of the S&P 500 in September, marked by the failure of Fannie & Freddie, Merrill Lynch, Lehman, and Washington Mutual. The bankruptcy of Lehman was particularly damaging for the short-term debt market. The Financial Times had been focusing on related topics and there was a significant decrease in news diversity before the pace of financial institution failures began to quicken.

Following the chaos of autumn 2008, and being under immense pressure, the US government established a bailout[8] scheme. As a result, the stock market began to free itself from the most severe impacts of the situation, and the news started to return to

---

[7] Standard and Poor's placed 612 securities backed by subprime residential mortgages on a credit watch, Countrywide Financial Corporation was warned of "difficult conditions", Bear Stearns liquidated two hedge funds that invested in various types of mortgage-backed securities, American Home Mortgage Investment Corporation filed for Chapter 11 bankruptcy protection and Fitch Ratings downgrades Countrywide Financial Corporation to BBB+. The worsening financial situation began to affect the European countries, BNP Paribas, France's largest bank, halted redemptions on three investment funds and the Chancellor of the Exchequer authorized the Bank of England to provide liquidity support for Northern Rock, the United Kingdom's fifth-largest mortgage lender. To add to this, financial market pressures intensified, which was reflected in diminished liquidity in interbank funding markets. In 2008, Countrywide Financial was purchased by Bank of America in January and Bear Stearns merged with JP Morgan in March. The situation in Europe was also worsening. The Treasury of the United Kingdom took Northern Rock into state ownership in February.

[8] On October 3rd, 2008, the Congress passed the law of Emergency Economic Stabilization Act and established the $700 billion Troubled Asset Relief Program (TARP). The US government also worked to inject liquidity into the precarious financial institutions in October and November 2008.

its “healthy state”. The most remarkable turnaround was seen after President Obama was sworn in in January 2009[9]. The stock market began to pick up and the news returned to its original diversity levels, marked by the change-point on 04/27/2009.

We may therefore conclude that there is an intrinsic relationship between news diversity levels and stock market movements. To be specific, the concentration of news topics is more focused when the condition of the financial market worsens. There will always remain negligible elements and events which cannot be reflected on the financial market instantaneously, such as the failure of small businesses and banks, people defaulting on their mortgages, and the initial deterioration of the property market indicators. However, all of these elements can be covered by news outlets. Furthermore, when news diversity returns to its original, healthier level (meaning the public is no longer paying such close attention to the poor market condition) it can be a sign of the financial markets rebounding.

[9] The Obama administration had over $1.1 trillion to rescue the financial market and the economy after the signing of the American Reinvestment and Recovery Act on February 17th 2009.

5. Robustness checks

This section conducts a series of robustness checks to further confirm the negative correlation between news diversity and financial market volatility, in order to assess the predictive power of news diversity. Moreover, we introduce a "distraction event" (Peress and Schmidt, 2020) to strengthen the results in Section 4.

First, we conduct a classical empirical test. The Financial Times is released daily at around 05:00 London time, which is significantly earlier than the opening time of continuous trading for the S&P 500 index. Moreover, Atkins et al. (2018) claim that the behavior of time series data from financial markets is influenced by information from the system, capturing its past behavior, and information regarding the underlying fundamentals embedded in news feeds. Based on the time difference and the mechanism of the financial time series, we investigate the relationship between news diversity and the subsequent movements of the financial market by training a popular and classical Auto-Regressive and Moving Average (ARMA) model, where we choose CBOE's volatility index (VIX) to measure market volatility.

We introduce the difference of VIX index ($v_t \equiv VIX_t - VIX_{t-1}$) as it is required for stationary time series to train the ARMA model and it denotes the change of daily volatility. According to the Auto Correlation Function (ACF), Partial Auto Correlation Function (PACF) and information criterion, we opt for the ARMA(2,1) model. Furthermore, to establish the connection between news diversity and market volatility, we add changes in news diversity ΔH_t as a regressor to our model of volatility changes, and find a significantly negative coefficient (t=-2.31, N=751, $p<0.05$). The negative coefficient is consistent with our research hypothesis, indicating that there is a negative correlation between news diversity and market volatility. In other words, a decrease in news diversity H_t tends to precede increased market volatility and increases in news diversity tend to precede a drop in market volatility on trading days. Consequently, when we combine this with the change-point detection in news diversity, we conclude that news diversity is negatively correlated with financial market volatility and it can be regarded as a signal for impending financial crashes and recoveries.

Second, we extend the data beyond the 2008 financial crisis and add two control time periods to further verify the validation of change-point detection in different market conditions. We choose a different time period, from April 2018 to March

2019[10], where there is severe stock market volatility, to check the robustness of the method. Since the recession in the autumn of 2018, the US stock market has been under various pressures, such as the increase in the expected yield of US Treasury bonds, the Fed's interest rate hike, and the Sino-US trade war, and its condition subsequently began to deteriorate. Market panic intensified and volatility increased, and the stock market gradually slowed down until the beginning of 2019. The news diversity has declined significantly since October 2018. We detect a change-point during this period, which is on 10/19/2018 (significant at the 1% significance level). Based on this change-point, we focus on US stock market movements and market volatility around this time. We find a negative correlation between news diversity and market volatility, suggesting that when the news diversity begins to decline, market volatility tends to increase, and the change-point signals the ensuing sharp drop in the market.

We then carry out a robustness test that finds no change-point in news diversity during 2019, where there are no major financial crashes. The news diversity index is relatively stable and there is no upward or downward trend. Furthermore, since recovering from the crash of October 2018, the stock market has been relatively stable. It has been on an upward trend and has not experienced sharp fluctuations during this period.

Third, we show that exogenous shocks may influence news diversity and its effect on the stock market. Following in the footsteps of Peress and Schmidt (2020), we introduce a "distraction event", one which captured worldwide headlines but had an arguably negligible effect on the economy and financial markets. One of the advantages of the LDA model is that we can extract topics from the daily news (e.g., Fig. 1). Among these 50 topics, we notice that there is one focusing on the US government and president containing keywords such as "parti" ("party"), "minist" ("ministry"), "elect" ("election"), "govern" ("government"). We label this as an "election topic". Taking into account the nature of the news reported by the Financial Times, and the typical "distraction event" mentioned in Peress and Schmidt (2020), the Democratic National Convention (08/25/2008 - 08/28/2008) is considered to be a suitable exogenous shock to news diversity.

With the help of the topic matrix $\rho_{k,t}$ (see, Appendix A), we verify that news concerning the election topic is more concentrated during the Democratic National

[10] After weighing the length of the sample and the time of the financial crash in the autumn of 2018, we choose the sample from April 2018 to March 2019. As long as the sample length exceeds about half a year and includes the autumn of 2018, different sample periods have little effect on the change-point detection.

Convention than during other weeks. In other words, more news focused on this topic, and it is a topic that is not directly associated with the economy and the financial markets. However, although the news diversity value is lower than the average (which is 3.11) during the period between the second (04/16/2008) and third (04/27/2009) change-points, there was no significant drop in the news diversity index during the convention, compared to the substantial decline of the index in late September 2008. Moreover, taking advantage of the intraday data of the S&P 500 index, and the price pattern from 04/16/2008 to 04/27/2009, we find that during the convention, the stock market was relatively stable and there were no violent fluctuations.

As a result, when there are exogenous shocks to news diversity, and these shocks are not directly related to the economy and financial markets, news topics tend to become more concentrated, and therefore news diversity decreases slightly. However, the concentration of news topics is temporary, and the news diversity soon returns to its previous level. Moreover, because the exogenous shock that causes the concentration of news topics is not directly related to the economy, the short-term decline in news diversity will not be reflected in the overall stock market in the long term. Therefore, even if these exogenous shocks from the news with little connection to the economy have a temporary, rather than a cumulative effect, on news concentration, they will not qualitatively affect our conclusions.

Incorporating exogenous shocks to our news diversity and change-point tests strengthens the results of our paper in two respects. First, exogenous shocks often manifest as a concentrated discussion on a particular topic (e.g., election topic). However, to build news diversity, multiple topics are required (50 topics in this article), meaning that the concentrated discussion on a specific topic caused by exogenous shocks will not bring about a significant decline in news diversity. Second, the change-point detection applied in this paper is based on a cumulative sum. In other words, for a change-point to be detected, the news diversity index must be continuously declining. However, it is clear that exogenous shocks do not meet this condition, while the research object discussed in this paper, i.e., an extreme event such as the financial crisis, does. Therefore, there would be no detected change-point in the news diversity due to exogenous shocks in the news that would thus cause an inaccurate prediction of financial crashes.

6. Conclusion

Technological forecasting is, now more than ever before, at the heart of the field of finance, due to the global and powerful e-market, new technologies and recent useful applications, such as big data and cloud computing. This study breaks new ground by investigating the relationship between news diversity and the movements of the financial market using change-point detection. Based on multiple change-point detections, this paper reveals two types of change-points that exist in the time series of news topic diversity. One type demonstrates that news diversity tends to decrease compared with the prior period, and the other type confirms that the diversity increases relative to the first half of the period. The first type of change-point demonstrates that the financial media is focusing on intensive negative news and the market will become more volatile and is at risk of collapse in the following trading days. The second type of change-point shows that news diversity has returned to a more diversified level and the stock market is going to recover. The empirical analysis reveals that there exists an intrinsic relationship between news diversity and stock market performance. Almost every financial event can be covered by the major financial media outlets, however, it may not be reflected on the market in a timely and complete manner. More specifically, news topics are more concentrated when the condition of the financial market worsens. However, when the markets warm up, the news diversity returns to a higher level. Therefore, the change-points of these different stages are useful for predicting the financial market.

Several implications flow from these observations. Firstly, with the further development of big data, relevant research methods and data sources suitable for financial market research should be more diverse and technical. This study begins by using news from the financial media and provides an initial study between news information and financial markets during financial crashes using a mathematical model. Therefore, significant benefits can be derived when news media, computer science, and mathematical methods are combined when studying and forecasting the financial market. Moreover, modern-day investments do not need to only rely on traditional data sources, such as market fundamentals and financial statements. Innovative data sources such as the news from the financial media can be incorporated as an effective supplement to existing research.

Appendix A. Topic Modeling

First, we introduce a selection of terms used in the LDA model:

1) A *word* w is defined to be an item from a vocabulary indexed by $\{1, \dots, V\}$, the distribution is Multinominal φ_k, V is the number of *words*. It is a fixed number;
2) A *document* m is a sequence of N words denoted by $d = (w_1, w_2, \dots, w_N)$, where N_m is the number of words in document m. It is a random variable;
3) A *corpus* D is a collection of M documents denoted by $D = \{d_1, d_2, \dots, d_M\}$, where M is the number of documents in the corpus. It is a fixed number;
4) *Topic* z and its distribution is Multinominal θ_m, K is the number of topics. It is a given number;
5) *Distribution* φ_k and θ_m have conjugate distributions $Dirichlet(\alpha)$ and $Dirichlet(\beta)$.

The generative process under the LDA model are as follows:

1) Sampling topic distribution θ_m of document m, $\theta \sim Dirichlet(\alpha)$;
2) Sampling word distribution φ_k of topic k, $\varphi \sim Dirichlet(\beta)$;
3) Sampling topic $Z_{m,n}$ of word $w_{m,n}$ in document m, $z \sim Multi(\theta)$;
4) Sampling word $w_{m,n}$ in topic $Z_{m,n}$, $w \sim Multi(\varphi)$;
5) Repeat N_m times.

Finally, we can obtain the topic distribution θ_m of document m, and the corresponding word distribution φ_k.

Following Curme et al. (2017), we construct diversity of news. Each document m represents each paragraph as a mixture of K topics ($K = 50$). According to the LDA model, we compute the k-dimensional topic vector $\theta_m = (\theta_{m,1}, \theta_{m,2}, \dots, \theta_{m,K})$, where D_t is the documents in the Financial Times issue on day t, and $n(D_t)$ is the number of documents in the set D_t. Thus, we have a topic matrix $\rho_{k,t}$.

The **News Diversity** is then as follows:

$$H_t = -\sum_{k=1}^{K} \rho_{k,t} \log\left(\rho_{k,t}\right). \qquad \text{Eq. (A. 1)}$$

Appendix B. Nonparametric method: Change in the mean

The model used in our paper is the signal-plus-noise model, N denotes the set of observations

$$X_t = \mu_t + \varepsilon_t, \qquad t \in N \qquad \text{Eq. (B. 1)}$$

where $X_t(t \in N)$ denotes the time series of news topic diversity index, $\mu_t(t \in N)$ is the signal and $\varepsilon_t(t \in N)$ is the noise part, with $E[\varepsilon_t] = 0$ and $E[{\varepsilon_t}^2] = 0$.

Considering the characters and practical significance of the news diversity index, we focus on dependent observations, X_1, X_2, …, X_n, to test the null hypothesis that there is no-change in the mean of the news diversity H_t

$$H_0\colon\ \mu_1 = \mu_2 = \cdots = \mu_n \equiv \mu$$

against the "one change in the mean" alternative

$$H_A\colon \text{there is an integer } k^*, 1 \le k^* < n, \text{such that } \mu_1 = \cdots = \mu_{k^*} \neq \mu_{k^*+1} = \cdots = \mu_n.$$

The CUSUM procedures, $Z_n = (Z_n(x)\colon x \in [0,1])$, is interested.

$$Z_n(x) = \frac{1}{\sqrt{n}}\left(\sum_{t=1}^{\lfloor nx \rfloor} X_t - \frac{\lfloor nx \rfloor}{n}\sum_{t=1}^{n} X_t\right), \qquad x \in [0,1] \qquad \text{Eq. (B. 2)}$$

where $\lfloor \cdot \rfloor$ denotes the integer part. Clearly, under the null hypothesis,

$$Z_n(x) = \frac{1}{\sqrt{n}}\left(\sum_{t=1}^{\lfloor nx \rfloor} \varepsilon_t - \frac{\lfloor nx \rfloor}{n}\sum_{t=1}^{n} \varepsilon_t\right), \qquad x \in [0,1]. \qquad \text{Eq. (B. 3)}$$

On the basis of (B.3), the CUSUM process under the null hypothesis is independent of the unknown mean μ, and is decided by the large-sample behavior of the partial sums of $(\varepsilon_t\colon\ t \in N)$. The noise part is not independent. It is possibly correlated.

Define the standardized partial sum process $S_n = (S_n(x)\colon x \in [0,1])$ by

$$S_n(x) = \frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor x \rfloor} \varepsilon_t\,, \qquad x \in [0,1]. \qquad \text{Eq. (B. 4)}$$

According to Aue and Horváth (2013), partial sum satisfies

$$S_n \Rightarrow \omega W \qquad \text{Eq. (B. 5)}$$

in the Skorohod space [0,1], where '⇒' denotes weak convergence as $n \to \infty$, and $W = (W(t)\colon t \in [0,1])$ is a standard Brownian motion, with known continuous and non-degenerate covariance kernel $k(r,s) = E[W(r)W(s)]$.

If (B.5) holds, it can be easily shown that

$$Z_n \Rightarrow \omega B \qquad \text{Eq. (B. 6)}$$

as $n \to \infty$, where $B = (B(t): t \in [0,1])$, with $B(t) = W(t) - tW(1)$, is a Brownian bridge. ω^2 is typically referred to as the long-run variance, which can be estimated by $\widehat{\omega}_n^2$. (B.6) implies

$$\frac{1}{\widehat{\omega}_n} Z_n \Rightarrow B \qquad \text{Eq. (B. 7)}$$

as $n \to \infty$.

A large volume of literature has researched the optimal estimator for ω^2. In the serial correlation case, we follow the suggestion of Berkes et al. (2011) by applying the Bartlett estimator for a consistent estimator of the long-run variance of small sample serial correlation models, such as first order autoregressive and moving average processes.

The Bartlett estimator is given by

$$\hat{s}_n^2 = \hat{\gamma}_0 + 2\sum_{j=1}^{n-1} \omega\left(\frac{j}{h(n)}\right)\hat{\gamma}_j \qquad \text{Eq. (B. 8)}$$

where

$$\hat{\gamma}_j = \frac{1}{n}\sum_{i=1}^{n-j}(X_i - \bar{X}_n)\left(X_{i+j} - \bar{X}_n\right) \qquad \text{Eq. (B. 9)}$$

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i\,. \qquad \text{Eq. (B. 10)}$$

$\omega(\cdot)$ is the kernel and $h(\cdot)$ is the length of the window. Assume that $\omega(\cdot)$ and $h(\cdot)$ satisfy the following standard assumptions:

$$\omega(0) = 1,$$

$$\omega(t) = 0 \;\; if \;\; t > a \;\; \text{with some} \;\; a > 0,$$

$$\omega(\cdot) \;\; \text{is a Lipschitz function,}$$

$\widehat{\omega}(\cdot)$, the Fourier transform of $\omega(\cdot)$, is also Lopschitz and integrable and

$$h(n) \to \infty \;\; \text{and} \;\; \frac{h(n)}{n} \to \infty \qquad \text{as} \;\; n \to \infty.$$

We use $h(n) = n^{1/2}$ as the window and the flat top kernel

$$\omega(t) = \begin{cases} 1, & 0 \le t \le 0.1 \\ 1.1 - |t|, & 0.1 \le t \le 1.1 \\ 0, & t \ge 1.1 \end{cases} . \qquad \text{Eq. (B. 11)}$$

Following Bazarova et al. (2014), our testing procedures are based on

$$\frac{1}{\widehat{\omega}_n} \sup_{0 \le x \le 1} |Z_n(x)| \qquad \text{Eq. (B. 12)}$$

and it follows immediately from (B.10) that under the null hypothesis

$$\frac{1}{\widehat{\omega}_n} \sup_{0 \le x \le 1} |Z_n(x)| \to^D \sup_{0 \le x \le 1} |B_n(x)| . \qquad \text{Eq. (B. 13)}$$

Appendix C. Multiple changes

This part follows Horváth et al. (2017). Assume that $X_i = \mu_i + \varepsilon_i$, where ε_1, ε_2,…,ε_n are dependent random variables. The multiple changes in the mean means that

$$\mu_i = \begin{cases} \Delta_1, & 1 \le i \le k_1^* \\ \Delta_2, & k_1^* \le i \le k_2^* \\ & \vdots \\ \Delta_{m+1}, & k_m^* \le i \le n \end{cases}, \qquad \text{Eq. (C. 1)}$$

where m denotes m changes in the mean, which is a given integer, and $\Delta_1, \dots, \Delta_{m+1}, k_1^*, \dots, k_m^*$ are unknown parameters. We test the null hypothesis that

$$H_0'\colon\ \Delta_1 = \Delta_2 = \cdots = \Delta_{m+1}.$$

We introduce a binary segmentation method, and test the null hypothesis in Section 3.1. If H_0 is rejected, this implies that we could locate the first change-point $\hat{k}_1$. Next, we divide the fixed sample into two subsamples $\{X_i,\ 1 \le i \le \hat{k}_1\}$ and $\{X_i,\ \hat{k}_1 < i \le n\}$, and then test both subsamples for further changes.

Acknowledgements

The paper has benefited from the useful comments and constructive suggestions by anonymous referees. The usual disclaimer applies.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Reference

Alanyali, M., Moat, H. S., Preis, T., 2013. Quantifying the relationship between financial news and the stock market. Sci. Rep. 3(1), 1-6. https://doi.org/10.1038/srep03578

Antweiler, W., Frank, M. Z., 2004. Is all that talk just noise? The information content of internet stock message boards. J. Financ. 59(3), 1259-1294. https://doi.org/10.1111/j.1540-6261.2004.00662.x

Atkins, A., Niranjan, M., Gerding, E., 2018. Financial news predicts stock market volatility better than close price. J. Financ. Data Sci. 4(2), 120-137. https://doi.org/10.1016/j.jfds.2018.02.002

Aue, A., Horváth, L., 2013. Structural breaks in time series. J. Time Ser. Anal. 34(1), 1-16. https://doi.org/10.1111/j.1467-9892.2012.00819.x

Baker, S. R., Bloom, N., Davis, S. J., 2016. Measuring economic policy uncertainty. Quart. J. Econ. 131(4), 1593-1636. https://doi.org/10.1093/qje/qjw024

Bazarova, A., Berkes, I., Horváth, L., 2014. Trimmed stable AR (1) processes. Stoch. Process. Their Appl. 124(10), 3441-3462. https://doi.org/10.1016/j.spa.2014.05.001

Berkes, I., Horváth, L., Schauer, J., 2011. Asymptotics of trimmed CUSUM statistics. Bernoulli 17(4), 1344-1367. https://doi.org/10.3150/10-BEJ318

Birz, G., Lott, J. R., 2011. The effect of macroeconomic news on stock returns: New evidence from newspaper coverage. J. Bank. Financ. 35(11), 2791-2800. https://doi.org/10.1016/j.jbankfin.2011.03.006

Blei, D. M., Ng, A. Y., Jordan, M. I., 2003. Latent dirichlet allocation. J. Mach. Learn. Res. 3, 993-1022. https://dl.acm.org/doi/10.5555/944919.944937

Bollen, J., Mao, H., Zeng, X., 2011. Twitter mood predicts the stock market. J. Comput. Sci. 2(1), 1-8. https://doi.org/10.1016/j.jocs.2010.12.007

Burnap, P., Rana, O. F., Avis, N., Williams, M., Housley, W., Edwards, A., ... Sloan, L., 2015. Detecting tension in online communities with computational Twitter analysis. Technol. Forecast. Soc. Change 95, 96-108. https://doi.org/10.1016/j.techfore.2013.04.013

Curme, C., Zhuo, Y., Moat, H. S., Preis, T., 2017. Quantifying the diversity of news around stock market moves. J. Network Theory Financ. 3(1), 1-20. http://doi.org/10.21314/JNTF.2017.027

Engelberg, J. E., Parsons, C. A., 2011. The causal impact of media in financial markets. J. Financ. 66(1), 67-97. https://doi.org/10.1111/j.1540-6261.2010.01626.x

Fang, L., Peress, J., 2009. Media coverage and the cross-section of stock returns. J. Financ. 64(5), 2023-2052. https://doi.org/10.1111/j.1540-6261.2009.01493.x

French, K., Baily, M., Campbell, J., Cochrane, J., Diamond, D., Duffie, D., ... Stulz, R., 2010. The Squam Lake report: Fixing the financial system. Princeton University Press, Princeton, NJ. https://www.jstor.org/stable/j.ctt7sjcm

Granger, C. W., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37(3), 424-438. https://doi.org/10.2307/1912791

Griffin, J. M., Hirschey, N. H., Kelly, P. J., 2011. How important is the financial media in global markets?. Rev. Financ. Stud. 24(12), 3941-3992. https://doi.org/10.1093/rfs/hhr099

Hagenau, M., Liebmann, M., Neumann, D., 2013. Automated news reading: Stock price prediction based on financial news using context-capturing features. Decis. Support Syst. 55(3), 685-697. https://doi.org/10.1016/j.dss.2013.02.006

Haroon, O., Rizvi, S. A. R., 2020. COVID-19: Media coverage and financial markets behavior—A sectoral inquiry. J. Behav. Exp. Econ. 27, Article 100343. https://doi.org/10.1016/j.jbef.2020.100343

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), Article e64846. https://doi.org/10.1371/journal.pone.0064846

Horváth, L., Pouliot, W., Wang, S., 2017. Detecting at-most-m changes in linear regression models. J. Time Ser. Anal. 38(4), 552-590. https://doi.org/10.1111/jtsa.12228

Jegadeesh, N., Wu, D., 2013. Word power: A new approach for content analysis. J. Financ. Econ. 110(3), 712-729. https://doi.org/10.1016/j.jfineco.2013.08.018

Kayser, V., Blind, K., 2017. Extending the knowledge base of foresight: The contribution of text mining. Technol. Forecast. Soc. Change 116, 208-215. https://doi.org/10.1016/j.techfore.2016.10.017

King, G., 2011. Ensuring the data-rich future of the social sciences. Science 331(6018), 719-721. https://doi.org/10.1126/science.1197872

Liao, R. C., Wang, X., Wu, G., 2021. The Role of Media in Mergers and Acquisitions. J. Int. Financial Mark. Inst. Money in press. https://doi.org/10.1016/j.intfin.2021.101299

Malkiel, B. G., Fama, E. F., 1970. Efficient capital markets: A review of theory and empirical work. J. Financ. 25(2), 383-417. https://doi.org/10.2307/2325486

Peress, J., Schmidt, D., 2020. Glued to the TV: Distracted noise traders and stock market liquidity. J. Financ. 75(2), 1083-1133. https://doi.org/10.1111/jofi.12863

Preis, T., Reith, D., Stanley, H. E., 2010. Complex dynamics of our economic life on different scales: insights from search engine query data. Philos. Trans. R. Soc.

A-Math. Phys. Eng. Sci. 368(1933), 5707-5719. https://doi.org/10.1098/rsta.2010.0284

Schumaker, R. P., Chen, H., 2009. Textual analysis of stock market prediction using breaking financial news: The AZFin text system. ACM Trans. Inf. Syst. 27(2), 1-19. https://doi.org/10.1145/1462198.1462204

Shannon, C. E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27(3), 379-423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Simon, H. A., 1955. A behavioral model of rational choice. Quart. J. Econ. 69(1), 99-118. https://doi.org/10.2307/1884852

Tan, S. J., Chua, S. H., 2004. "While stocks last!" Impact of framing on consumers' perception of sales promotions. J. Consum. Mark. 21(5), 343-355. https://doi.org/10.1108/07363760410549168

Tetlock, P. C., 2007. Giving content to investor sentiment: The role of media in the stock market. J. Financ. 62(3), 1139-1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x

Thorsrud, L. A., 2020. Words are the new numbers: A newsy coincident index of the business cycle. J. Bus. Econ. Stat. 38(2), 393-409. https://doi.org/10.1080/07350015.2018.1506344

Vespignani, A., 2009. Predicting the behavior of techno-social systems. Science 325(5939), 425-428. https://doi.org/10.1126/science.1171990

Vosen, S., Schmidt, T., 2011. Forecasting private consumption: Survey-based indicators vs. Google trends. J. Forecast. 30(6), 565-578. https://doi.org/10.1002/for.1213

Wang, X., Wu, G., 2020. The Role of the Media in Hedge Fund Activism. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3584071.

Zhang, Y., Zhang, G., Chen, H., Porter, A. L., Zhu, D., Lu, J., 2016. Topic analysis and forecasting for science, technology and innovation: Methodology with a case study focusing on big data research. Technol. Forecast. Soc. Change 105, 179-191. https://doi.org/10.1016/j.techfore.2016.01.015

Zhang, Y., Zhang, Z., Liu, L., Shen, D., 2017. The interaction of financial news between mass media and new media: Evidence from news on Chinese stock market. Physica A 486, 535-541. https://doi.org/10.1016/j.physa.2017.05.051