News Diversity and Financial Market: Causality and Structural Breaks

Lajos Horváth[1], Zhenya Liu[2,3], Ling Zhai[2]
[1]Department of Mathematics, University of Utah, United States
[2]School of Finance, Renmin University of China, China
[3]CERGAM, Aix-Marseille University, France

January 15, 2020

Abstract

Recently, more and more news media textual data with high-dimensional and unstructured features has been used in the economic and financial research, and has become a hot topic in the field of Media, Economics and Finance, and Computer Science. However, generally studies fail to make full use of big data to predict the financial market crash effectively, especially the prices movements. This paper investigates the relationship between financial news diversity and financial crash applying the change-point detection. Empirical analysis shows that (1) there does exist a realtionship between news diversity and financial market, (2) the news diversity tends to decline when the market falls and volatility soars, and goes back to higher level when the market upward and recovers, (3) the multiple structural breaks detected do help the forecast of price movements of stock market. Therefore, the change-point detection of financial news diversity can be a helpful earning warning of financial market falls or recoveries.

1 Introduction

With the rapid development of the Internet and the further improvement of computer technology, more and more textual data has been applied to the research in economy and finance. Different from traditional data, textual data is natrually high-dimensional and unstructured, which can be handled by the using of topic modelling and other machine learning methods. In recent years, textual big data has gradually become active in the field of economic finance (Baker et al., 2016; Shapiro et al., 2018; Tetlock, 2007; Kim and Kim, 2014; Jegadeesh and Wu, 2013; Fang and Peress, 2009).

Among textual data, the more commonly used information is usually derived from daily financial news (Thorsrud, 2018). The financial transaction data reflects the trader's decision-making (Simon, 1955), buying or selling, which affects the price movements and volatility of the financial market. These decisions may be affected by various types of information in trader's environment. In modern society, our interaction with the Internet is becoming a new source of information (Vespignani, 2009; King, 2011; Vosen and Schmidt, 2011). Additionally, Alanyali et al. (2013) pointed out that traders can not only actively receive information by searching for information online, but also receive news by passively or actively receiving news from influential financial media.

The performance of financial markets accumulates for some time. Unlike the effective market hypothesis, securities prices have fully responded to public and private information (Malkiel and Fama, 1970), behavioral finance has proven that news plays an important role in pricing mechanisms (Nofsinger, 2005). For instance, the occurrence of a financial crisis is not a one-day event. It takes a certain amount of time to ferment, resulting in a concentrated outbreak at a certain point in time. These subtle changes in fundamentals cannot be absorbed by all investors to the financial market, instead, the financial media are served as the carriers of these tiny information. In addition, when financial markets are experiencing violent fluctuations, the financial media will disseminate timely information to the public and track reports in a more professional way.

Financial media have a certain degree of interaction with the dynamics of financial markets. Alanyali et al. (2013) found that there was a positive correlation between the number mentioned by media and the daily trading volume of the stocks on the day before and the same day as the news release. That is, financial markets and financial news are intrinsically interrelated. Further, Curme et al. (2017) noted that there were interesting features of news diversity. For instance, during the "silly seasons" in summer the media would mainly focus on frivolous topics and the news topics were more concerntrated on fewer issues during a war or natural disasters or financial crisis. Therefore, Curme et al. (2017) developed a news diversity index based on *Financial Times*, one of the major financial media, and revealed that there was forecast effect of news diversity on the trading volume and the news diversity tended to drop when the stock market failed. However, they did not find a relation between news diversity and the subsequent price movements.

In this context, the objective of our paper is to examine the relationship between the financial news diversity and the performance of financial market furtherly and then discover the prediction effect of financial diversity on the dynamics of financial market, especially the prices and returns of stock market index. We select the financial news in *Financial Times* during the period of 2007 to 2009 from the ProQuest Database. To investigate the news diversity and its prediction effection, we construct news diversity index based on Curme et al. (2017) and apply change-point detection in the dependent case. We present evidence that the news diversity can grasp and forecast the performance of stock market well in general by the change-point analysis.

This paper may contribute to the academic discussion in several ways. First, instead of studying the forecast of trading volume (Curme et al., 2017), we focus on the forecast of the performance of financial market, and reveal an early warning effect of news diversity on the prices and returns of stock index, which is more crucial to the market and investors. Second, we use the method of change-point detection, a mathematical technique, to explore the relationship, rather than continuing previous studies applying conventional empirical methods (Birz and Lott, 2013; Zhang et al., 2017; Curme et al., 2017), which can get around endogenous problems. The remainder of this paper is organised as follows. Section 2 is literature review, Section 3 introduces the methodology and data, Section 4 discusses the empirical findings, Section 5 presents our conclusion.

2 Literature Review

There are increasing and innovative studies on the application of textual analysis in economic and financial filed of late years.

In economic studies, textual analysis is widely used for constructing index derived from business newspapers; see e.g. Baker et al.(2016) for index of economic policy uncertainty (EPU), and related works on EPU (Lee, 2018; Zalla, 2017; Colombo, 2013; Gulen and Ion, 2015), and Thorsrud(2018), Shapiro et al.(2018) for daily economic cycle index and economic sentiment. Not only newspapers, digging up and exploiting textual information nested in annual reports is also an expedient way for research in economics, for instance, Hoberg and Phillips (2016) measured product similarity for a new way of industry classification based on textual information analysing product descriptions in the annual reports (10-K).

In financial field, textual analysis is more devoted to uncovering relationship between mass media and the financial market. Firstly, novel proxy variables for financial market is an initial and fruitful effort with textual information. From earlier advertising costs (Lou, 2014; Chemmanur, 2019), the number of news (Barber and Odean, 2007) and even the forum speech (Antweiler and Frank, 2004) to latest studies on search volume indicator (SVI, Engelberg and Gao, 2011) characterizing investor attention measured by Google Trends, textual analysis does make an impressive achievement in the financial market research. Besides, media coverage derived from textual information can be served as a direct connection between media and the financial market, which draws a great number of studies. Fang and Peress (2009) believed that by reaching out to investors, media can alleviate information friction and affect safe pricing, even if it does not provide intrinsic news. Moreover, more socially responsible companies (Cahan et al., 2015), higher returns (Fang and Peress, 2009) and stronger momentum (Hillert et al., 2014) are positive correlated with higher media coverage.

Recently, textual data originating from well-known and credible financial newspapers, such as Wall Street Journal and Financial Times, has been employed to probe financial market movements. News Implied Volatility (NVIX) is a trendy keyword. Manela and Moreira (2017) used the front page of the Wall Street Journal to build text-based uncertainty metrics and the NVIX peaked during stock market crashes. Moreover, there are growing investigations about the interaction between financial market movements and financial news changes. Alanyali et al. (2013) used the corpus of Financial Times and found that there is a positive correlation between the number mentioned by the Financial Times and the daily trading volume of the company's stock on the day before and the same day as the press release. Curme et al. (2017) constructed a news diversity indicator and showed that the diversity of financial news can improve the trading volume forecast of the stock market. Evidence suggested that while financial news concerns tend to focus on a fewer topics after the stock market declines, news diversity will be healthier after the market begins to move up.

In sum, this section deals with literature from two perspectives: the application of textual analysis in economic and financial researches, and the relationship between financial news and financial market explored from textual data. Firstly, there are numerous papers studying the application of textual analysis in nowadays economic and financial researches, which is a popular research interests. Secondly, the relationship between financial news and financial market has attracted major attention, which focus on the traditional and simple econometric model to solve the economic and financial problems based on some indicators derived from textual data. To the best of our knowledge, few researches concentrated on the application of textual analysis in the prediction of financial market and the research methods are relatively simple in most studies. Therefore, based on previous research, our paper introduces more logical mathematical theories to explore the relationship between economic and financial news and financial market events, in order to do some prediction for the financial market.

3 Methodology and Data

This section introduces the Latent Dirichlet allocation(LDA; Blei et al.(2003)). Furthermore, the nonparametric method of change in the mean(Aue and Horváth, 2013) and a binary segmentation method(Vostrikova, 1981), which are applied to detect multiple change-points in the series of *News Diversity* we computed using topic modelling method, will be discussed. And then we will present the textual data, data resource and some data preprocess.

3.1 Topic Modelling

Since the text naturally has the characteristics of high dimensionality, the information contained in the text is difficult to quantify. The first step in making these textual data easy to analyse is usually to reduce its dimensions by collecting words into groups. A common tool is the topic modelling. In this way, the text corpus is divided into documents, each of which is usually treated as a collection of unordered words or as a "word bag", that is, it's a composition of words that have high or low probability in various topics. Latent Dirichlet allocation(LDA; Blei et al.(2003)) is one of the most popular methods in modelling hidden topics in text.

First, we introduce some terms in the LDA model:

1. A *word* w is defined to be an item from a vocabulary indexed by $\{1, \dots, V\}$, the distiburion is Multinominal φ_k, V is the number of *words*, and it's 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 and it's 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's a fixed number;

4. *Topic* z and its distribution is Multinominal θ_m, K is the number of topics, and it's given;

5. *Distibution* φ_k and θ_m have conjugate distributions *Dirichlet*(α) and *Dirichlet*(β).

The generative process under the LDA model are as follows:

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

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

Following the idea of Curmer et al.(2017), we can construct diversity of news. Each document m, it represents each paragraph is a mixture of K topics ($K = 50$). According to the LDA model, we can get the $k-$dimensional topic vector $\theta_m = (\theta_{m,1}, \theta_{m,2}, \ldots, \theta_{m,K})$, 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 can reach a topic matrix $\rho_{k,t}$.

The **News Diversity** then follows:

$$H_t = -\sum_{k=1}^{K} \rho_{k,t} \log\left(\rho_{k,t}\right) \tag{1}$$

3.2 Nonparametric method: Change in the mean

The used model in our paper is the signal-plus-noise model, $\mathbb{N}$ denotes the set of observations

$$X_t = \mu_t + \varepsilon_t, \quad t \in \mathbb{N} \tag{2}$$

where $(\mu_t : t \in \mathbb{N})$ is the signal and $(\varepsilon_t : t \in \mathbb{N})$ is the noise part, with $E\left[\varepsilon_t\right] = 0$ and $E\left[\varepsilon_t^2\right] = \sigma^2$.

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

$$H_0 : \mu_1 = \cdots = \mu_n \equiv \mu \tag{3}$$

against the "one change in the mean" alternative

$$\begin{aligned} H_A &: \text{there is an integer } k^*, 1 \leq k^* < n, \\ &\quad \text{such that } \mu_1 = \cdots = \mu_{k^*} \neq \mu_{k^*+1} = \cdots = \mu_n \end{aligned} \tag{4}$$

We are still interested in the CUSUM procedures, $Z_n = (Z_n(x) : x \in [0,1])$, which is also a basic object of research in the independent case studying in Csörgo and Horváth (1997)

$$Z_n(x) = \frac{1}{\sqrt{n}} \left(\sum_{t=1}^{\lfloor nx \rfloor} X_t - \frac{\lfloor nx \rfloor}{n} \sum_{t=1}^{n} X_t \right), \quad x \in [0,1] \tag{5}$$

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

$$Z_n(x) = \frac{1}{\sqrt{n}}\left(\sum_{t=1}^{\lfloor nx\rfloor}\varepsilon_t - \frac{\lfloor nx\rfloor}{n}\sum_{t=1}^{n}\varepsilon_t\right), \quad x \in [0,1]. \tag{6}$$

On the basis of (6), the CUSUM process under the null hypothsis is dependent with the unknown mean μ, and is decided by the large-sample behavior of the partial sums of $(\varepsilon_t : t \in \mathbb{N})$. In our settings, the noise part does not an independent case, they are possibly correlated.

We define the standardized partial sum process $S_n = (S_n(x) : x \in [0,1])$ by

$$S_n(x) = \frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor x\rfloor}\varepsilon_t, \quad x \in [0,1]. \tag{7}$$

Next, we consider the asymptotic behavior of the partial sums of $(\varepsilon_t : t \in \mathbb{N})$. According to Aue and Horváth (2013), partial sums satisfies

$$S_n \Rightarrow \omega W \tag{8}$$

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

If (8) holds, it can be easliy shown that

$$Z_n \Rightarrow \omega B \tag{9}$$

as $n \to \infty$, where $B = (B(t) : t \in [0,1])$, with $B(t) = W(t) - tW(1)$, is a Brownian bridge. As Aue and Horváth (2013) pointed that ω^2 is typically referred to as the long-run variance. If the estimator $\hat{\omega}_n^2$ from the given data is consistent with the long-run variance, (9) implies

$$\frac{1}{\hat{\omega}_n} Z_n \Rightarrow B \tag{10}$$

as $n \to \infty$.

A large volume of literature has researched the optimal esitimator for ω^2. For instance, Hannan (1957), Berk (1974), Newey and West (1987), Andrews (1991) and Robinson and Velasco (1997) have made great contributions. In the serial correlation case, we refer to the recent work by Müller (2007) for a more efficient estimator of the long-run variance of small sample serial correlation models, such as first order autoregressive and moving average processes.

Müller (2007) gave an estimator, $\hat{\omega}^2_{\mathrm{UA}}(p)$, which has competitive performance compared to previously studied long-run variance estimators. Let φ_l and r_l with $r_1, r_2, \ldots$ and $l = 1, 2, \ldots$ be a set of continuous eigenfunctions and eigenvalues of $k(r,s) = \mathrm{E}[W(r)W(s)]$. Phillips (1998) worked out the eigenvalues and eigenfunctions of $k(r,s) = \mathrm{E}[W(r)W(s)]$ that $r_l = \pi^{-2}\left(l - \frac{1}{2}\right)^{-2}$ and $\varphi_l(s) = \sqrt{2}\sin\left(\pi\left(l - \frac{1}{2}\right)s\right)$.

The recommend esrimator by Müller (2007) is modified

$$\hat{\omega}^2_{\mathrm{UA}}(p) = p^{-1} \sum_{l=1}^{p} \hat{\xi}_l^2 \tag{11}$$

where $\hat{\xi}_l = r_l^{-1/2} T^{-1} \sum_{t=1}^{T} \varphi_l(t/T) u_{T,t}$, and the choice of parameter p is a trade-off between the robustness and efficiency, which we decide $p = 2$ accodring to the simulations and analysis promoted by Müller (2007).

Following Bazarova et al. (2011, 2014), our testing procedures are based on

$$\frac{1}{\hat{\omega}_n} \sup_{0 \le x \le 1} |Z_n(x)| \tag{12}$$

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

$$\frac{1}{\hat{\omega}_n} \sup_{0 \le x \le 1} |Z_n(x)| \xrightarrow{D} \sup_{0 \le x \le 1} |B(x)|. \tag{13}$$

3.3 Multiple changes

We would like to detect multiple change-points in the news diversity. Assume that $X_i = \mu_i + \varepsilon_i$, where $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are dependent random variables. The multiple changes in the mean means that

$$\mu_i = \begin{cases} \Delta_1, & 1 \le i \le k_1^* \\ \Delta_2, & k_1^* < i \le k_2^* \\ \vdots & \\ \Delta_{m+1}, & k_m^* < i \le n \end{cases} \tag{14}$$

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

$$H_0' : \Delta_1 = \Delta_2 = \cdots = \Delta_{m+1} \tag{15}$$

We introduce a binary segmentation method, and firstly test the null hypothsis in Section 3.2. If H_0 is rejected, implying that we could locate the first changepoint $\hat{k}(1)$. Next, divide the fixed sample into two subsamples $\left\{\mathbf{X}_i, 1 \le i \le \hat{k}(1)\right\}$ and $\left\{\mathbf{X}_i, \hat{k}(1) < i \le n\right\}$, and then test both subsamples for further changes.

3.4 Textual data

In order to focus on investigating the relationship between the diversity of topics appearing in financial news, represented by daily issues of *Financial Times*, and financial market. We use a corpus of daily financial news issued by *Financial Times*, which is accessible in the ProQuest Dataset and the daily closing prices of FTSE 100 and S&P 500. The sample period is from January 2, 2007 to December 31, 2009. According to the data requirement for the LDA model, we do some data preprocessing before applying the textual data to the LDA model, including lowercase all words, change hyphens to whitespace, only letters “a” to “z” are left, clean single-letter words and stemmed stopwords, and other data preprocess according to Curme et al.(2017).

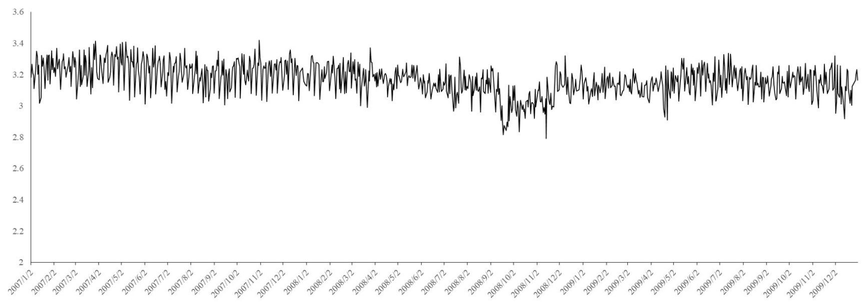

Figure 1: News diversity from 2007/01/02 to 2009/12/31.

4 Empirical Analysis

In this section, we will discuss about the news diversity over our sample period, multiple change-point detection of news diversity, and survey the performance of financial market at news diversity's structural breaks, especially the prices and returns, to investigate whether the news diversity is a good prediciton of financial market movements by applying change-point analysis.

4.1 News diversity over 2007-2009

Once the LDA is trained, each document m, namely each paragraph, in our whole corpus containing daily news in *Financial Times* from January 2, 2007 to December 31, 2009, can be presented by the K-dimensional topic vector $\theta_m = (\theta_{m,1}, \theta_{m,2}, \ldots, \theta_{m,K})$, and then we can calculate the news diversity index following Curmer et al. (2017).

As we trained and calculated, Figure 1 shows the news diversity index (including the weekend) and Figure 2 shows the box plot of news diversity, aggregated by weekdays. It is evident that news diversity in weekend is generally below that of weekday's for the special columns in weekend. Therefore, in order to get a more smooth sequence and the CUSUM method free from the data feature making the change-point detection more convincing, we drop the data of Saturdays. And Figure 3 seems that news diversity getting rid of Saturdays is more stationary than before. Additonally, the topics news focuses are naturally associated. Once a news breaks out, the media will continue to follow it. As a result, the diversity of daily news will not be independent[1].

4.2 News diversity and financial market

4.2.1 Granger causality test

In order to examine the relationship between news diversity and financial market, especially to see whether the news diversity could help predict the performance of the financial market, we do the Granger causality test to depict the general relation in advance.

Before the Granger causality test, we firstly test for unit roots for the sequence of news diversity, FTSE 100 and SP 500 daily closing price sequences (ADF, PP, DF-GLS), and the

[1]The independece of news diversity can be rejected by white noise test.

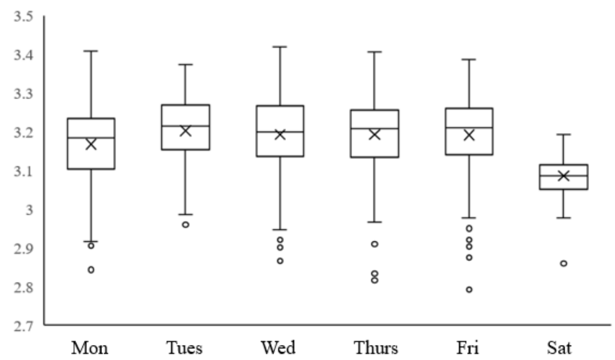

Figure 2: Box plots of the news diversity, aggregated by weekday.

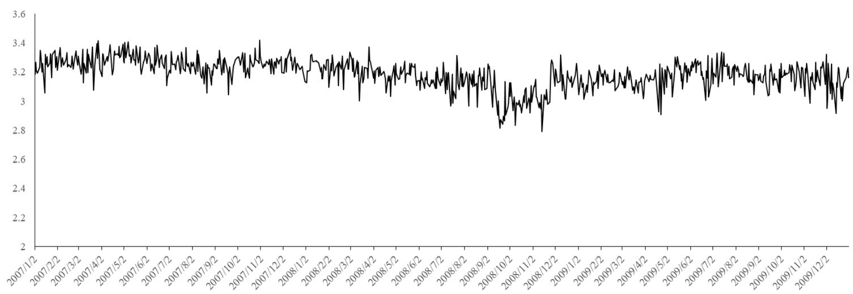

Figure 3: News diversity from 2007/01/02 to 2009/12/31 dropping weekend.

closing price series are expressed in natural logarithms. Table 1 shows that the closing price series of FTSE 100 and SP 500 are $I(1)$, and the news diversity is a stationary series.

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
SP_500(Level)	-1.093	-1.106	-1.234
SP_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

Next following Granger (1969), we fit a VAR and do the Granger causality tests to look into whether there is some relationship between news diversity and the first difference of stock market index.

The null hypothesis of Granger causality Wald test is each of the other endogenous variables does not Granger-cause the dependent variable in that equation. Therefore, Table 2 and Table 3 present that news diversity "Granger causes" SP 500 and FTSE 100 index

Table 2: Granger causality test for diversity and FTSE 100

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

Table 3: Granger causality test for diversity and SP 500

Equation	Excluded	chi2	Prob ≥ chi2
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

and these two index also "Granger cause" news diversity. Therefore, till now, it shows that there does exit relationship between news diversity and stock market, and what's important, past values of news diversity are useful for predicting the stock market. As a result, the following change-point detection of news diversity can be used to explain the performance of stock market in some way.

4.2.2 Change-point detection of news diversity

By applying the change in mean test and the binary segmentation method in the fixed sample of the news diversity series which is except for Saturdays, we can find 6 change points over the whole sample period, which lie at 2007/03/23, 2007/06/27, 2008/04/16, 2008/09/05, 2008/11/24, 2009/04/27, respectively[2]. According to the multiple changepoint detection, 2008/04/16 is the first changepoint deteceted over the whole sample (*Period 1*), which can be seperated into two periods with the point of 2008/04/16. 2007/06/27 is the changepoint for the first half (*Period 2*) and 2009/04/27 is for the second half (*Period 3*). Next, 2007/03/23 is detected over 2007/01/02 to 2007/06/27 (*Period 4*). Futher, we detect 2008/11/24 over the period from 2008/04/16 to 2009/04/27 (*Period 5*) and 2008/09/05 over 2008/04/16 to 2008/11/24 (*Period 6*).

Combining the statistics we calculated in "Change in mean" and the topic diversity's trend shown in Figure 4, we can focus on the sign of $S(k) - \frac{k}{n}S(n)$, that is to say, if $S(k) - \frac{k}{n}S(n)$ is positive at the change-point, the change-point detection method has grasped the maximum decline of the news diversity and the topic diversity is more likely to decrease compared to the prior period and vice versa. Futhermore, with the help of the figure, we can verify whether the diversity is going to drop or rise after the change-point over its respective periods.

[2]Because of the long-run variance estimators are closely bound up with the sample sizes, the change-point detection will stop when the period is less than 100 trading days.

Table 4: Tests for structural breaks during the whole sample

Period	Statistics	Location of structural breaks
2007/01/02-2009/12/31	10.660***	2008/04/16
2007/01/02-2008/04/16	3.132***	2007/06/27
2008/04/17-2009/12/31	3.287***	2009/04/27
2007/01/02-2007/06/27	4.160***	2007/03/23
2008/04/17-2009/04/27	4.044***	2008/09/05
2008/09/05-2009/04/27	5.025***	2009/11/24

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. *** Significant at 1% significance level.

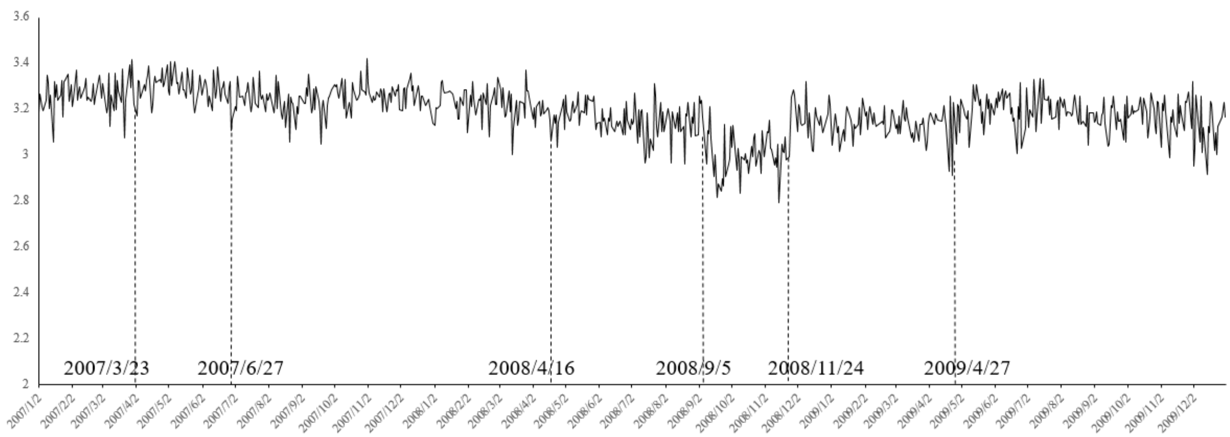

Figure 4: Change-points in news diversity from 2007/01/02 to 2009/12/31.

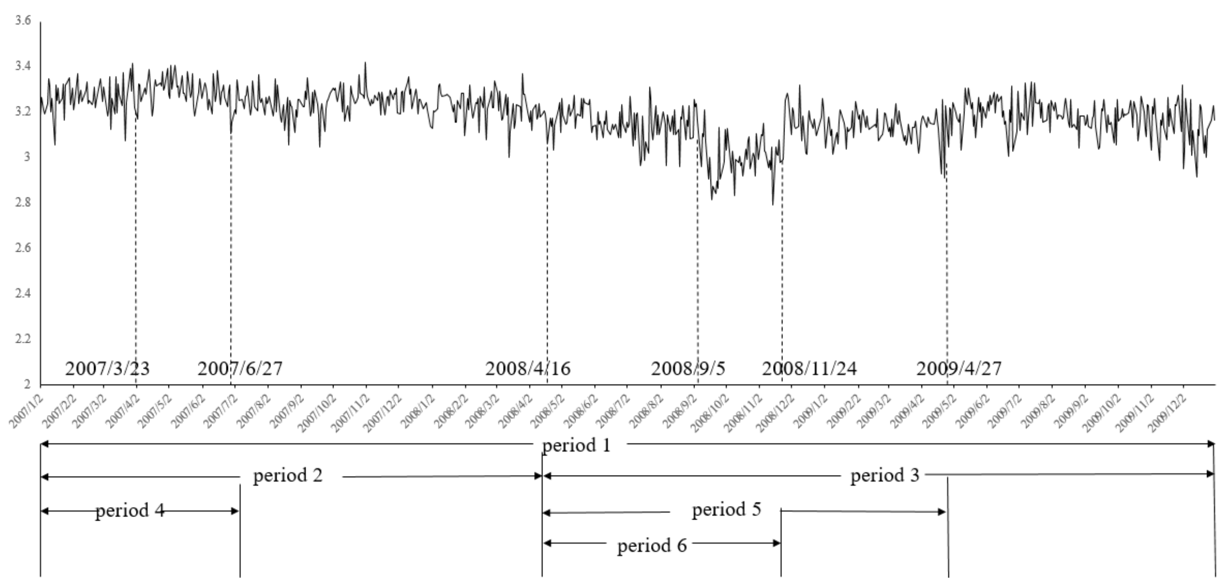

Figure 5: Change-points in news diversity from 2007/01/02 to 2009/12/31 with periods division.

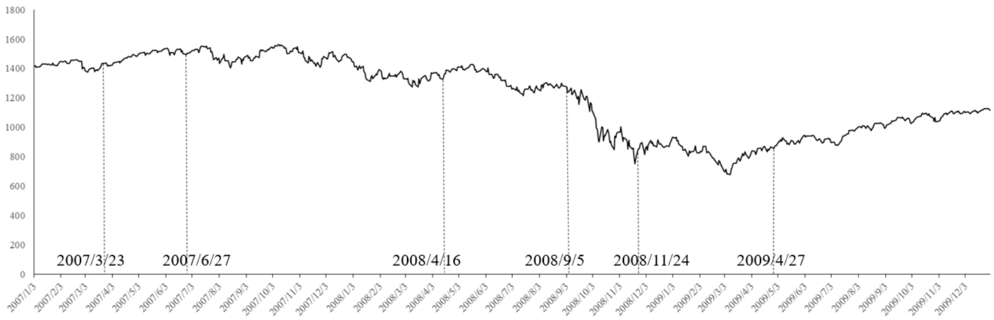

Figure 6: Prices of SP500 at the time of news diversity's change-points.

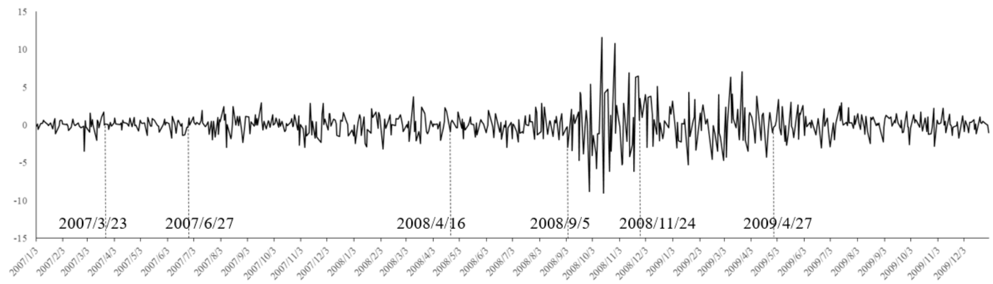

Figure 7: Returns of SP500 at the time of news diversity's change-points.

After calculation, we find that at 2008/04/16, the first change-point, indicates that the topic diversity is about to decline over the second half of *period 1*. Change-point at 2007/06/27 shows that the decrease of news diversity later over *period 2* and 2009/04/27 shows that the news diversity is to increase over the second half of *period 3*. 2007/03/23 presents that news diversity tends to go up later in the *period 4*, and in *period 3* the two more change-points 2008/09/05 and 2008/11/24 furtherly indicate that the news diversity is going to decrease and rebound later in *period 5* and *period 6*.

4.2.3 Performance of stock market at time of news diversity's changepoint

It should be noted that we are not to detect the change-point exactly the time when the topic diversity started to increase or decrease, but to grasp the time when the cumulated change is statistically significant. Therefore, in this part, we will discuss about the performance of the stock market (prices and returns), during the period of 2007 to 2009, focusing on the dynamics around the timing of the news diversity's changepoints particularly. Furthermore, we would like to investigate what financial events have happened to affect the stock market before the change-point and how the stock market behaved when the topic tended to be more concentrated or not.

Firstly, let's look into the performance of the stock market, especially the prices and returns of SP 500 index and FTSE 100 index at the time of news diversity's change-points.

Combine the analysis on whether the news topic is to more concentrated or not according to CUSUM and the figures of the performance of SP 500 index and FTSE 100 index. At 2007/06/27, when the news diversity is to drop later over *period 2*, it is clear that the price

Figure 8: Prices of FTSE100 at the time of news diversity's change-points.

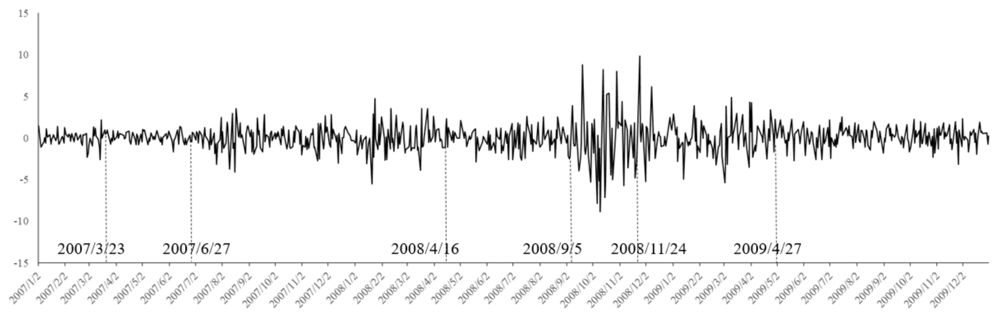

Figure 9: Returns of FTSE100 at the time of news diversity's change-points.

begun to fall and the returns were becoming more and more volatile. 2008/04/16 is the change-point of the whole period which indicated that the news topic was to concentrated on fewer topics later and it seems to be a good warning for the dramatic decline of prices and extreme volatility during the latter half of year 2008. Based on the CUSUM, 2008/09/05 also indicated a descent of news diversity over *6*, and it is evident that the price experienced a rapid drop and the returns sunk into a maelstrom during these periods.

On the other hand, could the news diversity give us an early prompt or hint that the stock market is beginning to recover? It should be noted that we also detect two change-points, 2008/11/24 and 2009/04/27, which indicated an increase of topic diversity over the corresponding period. Therefore, let's focus on the performance of stock market after these two change-points. 2008/11/24 is the change-point over the period from 2008/04/16 to 2009/04/27, and the returns were less volatile from 2008/11/24 compared to the hitting turbulence from 2008/09/06 to 2008/11/24. And coincidentally, there were also rallying equities in the stock market, which can be reflected on the prices of SP 500 and FTSE 100 index. Even more interesting, the first change-point, 2007/03/23, showed an increasing trend of topic diversity before the big chaos. In turn, the prices went up and the volatility was moderate compared to the early stage.

The connection between the ascent of news diversity and the rebound of stock market reflects more vividly at the change-point of 2009/04/27 over 2008/04/16 to 2009/12/31.The price of SP 500 index increased steadily after April 2009 and the returns volatility started to recover to normal and almost original levels.

Furthermore, in order to explore the relation between news diversity and stock market performance and the predict effect of financial news on the stock market, we review the financial events happened around the change-points to study what have attributed the changes of news diversity and whether they have reflected on stock market immediately.

According to the St.Louis Federal Reserve Bank, the Financial Crisis of 2007-2009 may start on February 27, 2007 when the Federal Home Loan Mortgage Corporation (Freddie Mac) made an announcement that it will no longer buy the most risky subprime mortgages and mortgage-related securities[3]. However, the risk early-warning was not appreciated by the investors, especially the European financial institutions in truth who continued purchasing mortgage-related securities. And after the announcement, the market even went up for some time.

At April 2, 2007, the leading subprime mortgage lender New Century Financial Corporation called for Chapter 11 bankruptcy protection. The financial news had begun to cover the bad financial situation, such as Standard and Poor's and Moody's Investor Services downgrade over 100 bonds backed by second-lien subprime mortgages and Bear Stearns informs investors that it is suspending redemptions in June 2007. In other words, the financial news had begun to focus on fewer topics related to "credit", "loan", "bank", etc., and the change-point at 2007/06/27 can be served as a starting point of the Financial Crisis of 2007-2009, which is consistent with the increasingly worse financial situation and French et al. (2010) considered the first symptoms of the World Financial Crisis appeared in the summer of 2007.

The situation began to deteriorate in the summer of year 2007. Standard and Poor's places 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+. And the bad situation began to the European countries, BNP Paribas, France's largest bank, halted redemptions on three investment funds and the Chancellor of the Exchequer authorizes the Bank of England to provide liquidity support for Northern Rock, the United Kingdom's fifth-largest mortgage lender. What's worse, financial market pressures intensified, which can be reflected in diminished liquidity in interbank funding markets.

Situations went worse in the beginning of year 2008, Countrywide Financial was purchased by Bank of America in January and Bear Stearns was merged by JP Morgan in March. The situation in Europe is not optimistic neither. The Treasury of the United Kingdom took Northern Rock into state ownership in February. Therefore, the first detected change-point, 2008/04/16 shows that the financial news has concentrated on the bad situation of financial market for a long time, not very short at least. And very little helps a mickle, it has become a very precise early warning of the turmoil on Wall Street and the following global crisis, with the help of the change-point at April 2008.

In summer of year 2008, more and more financial institutions failed, such as Wachovia in June and Indy Mac in July. There will be a melodramatic drop in the price and an immense fluctuation in return of SP 500 in September, marked with the failure of Fannie & Freddie, Merrill Lynch, Lehman and Washington Mutual. The bankruptcy of Lehman took a hard bit

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

on the short-termdebtmarket. There has been a significant decrease of news diversity before the quickened pace of merger focusing on related topics and is grasped by the change-point at 2008/09/05.

After the chaos in Autumn in 2008 and under the immense pressure, the US government started a bailout. At October 3, 2008, Congress passed the law of Emergency Economic Stabilization Act of 2008 and established the $700 billion Troubled Asset Relief Program (TARP). The US government also tried hard to inject liquidity into the precarious financial institutions in October and November in 2008. As a result, the stock market began to break free from the most sever situation, and the financial news started to turn back to its "healthy state" and focused more general topics which can be shown at the changepoint of 2008/11/24.

The most remarkable turnaround showed after Mr.Obama was sworn in January 2009. The Obama administration had over $1.1 trillion to continue the rescue for the financial market and the economy after the sign of American Reinvestment and Recovery Act of 2009 on February 17. The stock market began to warm up and the financial news finally diversified back to its original level, marked with the change-point at 2009/04/27.

Therefore, we may come to a conclusion that there does exist an intrinsic relation between news diversity and stock market performance. To be specific, the concentration of financial news topics is more focused when the situation of financial market gets worse. Considering that there will always be some negligible things which the financial market can not reflect on the financial market instantaneously, such as failure of small business and banks, the mortgage of some people cannot be repaid and the beginning deterioration of the property market indicators, however, all of these can be covered by financial news. Furthermore, when the news diversity goes back to healthier and its original level, which means the public no longer pays such close attention to the poor market, it can be a sign of the rebound of financial situation.

5 Conclusion

This study breaks new ground and is the first paper to investigate the relationship between financial news diversity and the performance of fiancial market by using the change-point detection. Overall, we reveal an intrinsic relation between news diversity and stock market performance, for the reason that almost every financial events can be grasped by the major financial media and may not be reflected on the market in a timely, effective and complete manner. Specifically, the financial news topics will be more focused when the situation of financial market gets worse, and when the market warms up, the diversity will go back to higher level. And the change-points of these different stages can make a good prediction of the financial market.

References

[1] Alanyali, M., Moat, H. S., and Preis, T. Quantifying the relationship between financial news and the stock market. *Scientific reports 3* (2013), 3578.

[2] Andrews, D. W. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica: Journal of the Econometric Society* (1991), 817–858.

[3] Antweiler, W., and Frank, M. Z. Is all that talk just noise? the information content of internet stock message boards. *The Journal of finance 59*, 3 (2004), 1259–1294.

[4] Aue, A., and Horváth, L. Structural breaks in time series. *Journal of Time Series Analysis 34*, 1 (2013), 1–16.

[5] Baker, S. R., Bloom, N., and Davis, S. J. Measuring economic policy uncertainty. *The quarterly journal of economics 131*, 4 (2016), 1593–1636.

[6] Barber, B. M., and Odean, T. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The review of financial studies 21*, 2 (2007), 785–818.

[7] Bazarova, A., Berkes, I., and Horváth, L. Trimmed stable ar (1) processes. *Stochastic processes and their applications 124*, 10 (2014), 3441–3462.

[8] Bazarova, A., Berkes, I., and Horváth, L. Change point detection with stable ar (1) errors. In *Asymptotic Laws and Methods in Stochastics*. Springer, 2015, pp. 179–193.

[9] Berk, K. N. Consistent autoregressive spectral estimates. *The Annals of Statistics* (1974), 489–502.

[10] Birz, G., and Lott Jr, J. R. The effect of macroeconomic news on stock returns: New evidence from newspaper coverage. *Journal of Banking & Finance 35*, 11 (2011), 2791–2800.

[11] Blei, D. M., Ng, A. Y., and Jordan, M. I. Latent dirichlet allocation. *Journal of machine Learning research 3*, Jan (2003), 993–1022.

[12] Cahan, S. F., Chen, C., Chen, L., and Nguyen, N. H. Corporate social responsibility and media coverage. *Journal of Banking & Finance 59* (2015), 409–422.

[13] Chemmanur, T. J., and Yan, A. Advertising, attention, and stock returns. *Quarterly Journal of Finance* (2019), 1950009.

[14] Colombo, V. Economic policy uncertainty in the us: Does it matter for the euro area? *Economics Letters 121*, 1 (2013), 39–42.

[15] Csorgo, M., and Horváth, L. *Limit theorems in change-point analysis*. John Wiley & Sons Chichester, 1997.

[16] Curme, C., Zhuo, Y., Moat, H. S., and Preis, T. Quantifying the diversity of news around stock market moves. *Chester Curme, Ying Daisy Zhuo, Helen Susannah Moat and Tobias Preis, Quantifying the diversity of news around stock market moves, Journal of Network Theory in Finance 3*, 1 (2017), 1–20.

[17] Engelberg, J., and Gao, P. In search of attention. *The Journal of Finance 66*, 5 (2011), 1461–1499.

[18] Fang, L., and Peress, J. Media coverage and the cross-section of stock returns. *The Journal of Finance 64*, 5 (2009), 2023–2052.

[19] Gulen, H., and Ion, M. Policy uncertainty and corporate investment. *The Review of Financial Studies 29*, 3 (2015), 523–564.

[20] Hannan, E. The variance of the mean of a stationary process. *Journal of the Royal Statistical Society: Series B (Methodological) 19*, 2 (1957), 282–285.

[21] Hillert, A., Jacobs, H., and Müller, S. Media makes momentum. *The Review of Financial Studies 27*, 12 (2014), 3467–3501.

[22] Hoberg, G., and Phillips, G. Text-based network industries and endogenous product differentiation. *Journal of Political Economy 124*, 5 (2016), 1423–1465.

[23] Jegadeesh, N., and Wu, D. Word power: A new approach for content analysis. *Journal of Financial Economics 110*, 3 (2013), 712–729.

[24] Kim, S.-H., and Kim, D. Investor sentiment from internet message postings and the predictability of stock returns. *Journal of Economic Behavior & Organization 107* (2014), 708–729.

[25] King, G. Ensuring the data-rich future of the social sciences. *science 331*, 6018 (2011), 719–721.

[26] Lee, S. Economic policy uncertainty in the us: Does it matter for korea? *East Asian Economic Review 22*, 1 (2018), 29–54.

[27] Lou, D. Attracting investor attention through advertising. *The Review of Financial Studies 27*, 6 (2014), 1797–1829.

[28] Malkiel, B. G., and Fama, E. F. Efficient capital markets: A review of theory and empirical work. *The journal of Finance 25*, 2 (1970), 383–417.

[29] Manela, A., and Moreira, A. News implied volatility and disaster concerns. *Journal of Financial Economics 123*, 1 (2017), 137–162.

[30] Müller, U. K. A theory of robust long-run variance estimation. *Journal of Econometrics 141*, 2 (2007), 1331–1352.

[31] Newey, W. K., West, K. D., et al. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 3 (1987), 703–708.

[32] Nofsinger, J. R. Social mood and financial economics. *The Journal of Behavioral Finance 6*, 3 (2005), 144–160.

[33] Phillips, P. C. New tools for understanding spurious regressions. *Econometrica* (1998), 1299–1325.

[34] Robinson, P. M., and Velasco, C. 11 autocorrelation-robust inference. *Handbook of Statistics 15* (1997), 267–298.

[35] Shapiro, A. H., Sudhof, M., and Wilson, D. Measuring news sentiment. Federal Reserve Bank of San Francisco.

[36] Simon, H. A. A behavioral model of rational choice. *The quarterly journal of economics 69*, 1 (1955), 99–118.

[37] Tetlock, P. C. Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance 62*, 3 (2007), 1139–1168.

[38] Thorsrud, L. A. Words are the new numbers: A newsy coincident index of the business cycle. *Journal of Business & Economic Statistics* (2018), 1–17.

[39] Vespignani, A. Predicting the behavior of techno-social systems. *Science 325*, 5939 (2009), 425–428.

[40] Vosen, S., and Schmidt, T. Forecasting private consumption: survey-based indicators vs. google trends. *Journal of Forecasting 30*, 6 (2011), 565–578.

[41] Vostrikova, L. Y. Detecting ¡disorder¡ in multidimensional random processes. In *Doklady Akademii Nauk* (1981), vol. 259, Russian Academy of Sciences, pp. 270–274.

[42] Zalla, R. Economic policy uncertainty in ireland. *Atlantic Economic Journal 45*, 2 (2017), 269–271.

[43] Zhang, Y., Zhang, Z., Liu, L., and Shen, D. The interaction of financial news between mass media and new media: Evidence from news on chinese stock market. *Physica A: Statistical Mechanics and its Applications 486* (2017), 535–541.