Journal of the Korean Physical Society (2022) 80:526–532
https://doi.org/10.1007/s40042-022-00448-4

Online ISSN 1976-8524
Print ISSN 0374-4884

ORIGINAL PAPER - CROSS-DISCIPLINARY PHYSICS AND RELATED AREAS OF SCIENCE AND TECHNOLOGY

Multifractal analysis of social media use in financial markets

Gabjin Oh[1]

Received: 13 December 2021 / Revised: 27 January 2022 / Accepted: 27 January 2022 / Published online: 25 February 2022

Abstract
We analyze the nonlinear properties of social media activity(SMA) using the multifractal detrended fluctuation analysis (MF-DFA) method. Social media data related to the stock market are gathered from social media platforms. Using data on over 2000 firms in the Korean stock market for 2018–2020, we study social media activity and its differences to evaluate associated nonlinear and statistical properties. We find that the cumulative distribution function of SMA follows a stretched exponential distribution with $\beta = 0.85$. The Hurst exponent of SMA for three datasets (2018, 2019, 2020 year) is larger than 0.9, whereas the Hurst exponents of shuffled time series have values of approximately 0.5. In particular, we find a multifractal structure in both SMA and SMA difference results irrespective of the period and degree of multifractality defined as $\alpha_{\max} - \alpha_{\min}$, which reaches a maximum value during the COVID-19 pandemic as a financial crisis.

Keywords Econophysics · Multifractal · Social media

1 Introduction

In determining which stocks to select over others, investors consider the information generated from social media [1–4]. Investors' attention is a significant resource among many alternatives [5]. We establish social media activity as an attention source of heterogeneous investors. Economic systems have been described as some of most complex systems with nontrivial interactions among investors. In particular, investors who want to make profits in the market with valuable information regarding financial markets are always looking for new and meaningful information with which to help them invest. Hirshleifer et al. [6] asserted that the information generated from social media can play a crucial role in the discovery of asset prices. In other words, understanding social media dynamics has become very important to modern empirical asset pricing. The econophysics and finance fields have posed crucial research questions regarding which information from social media systems is valuable in financial markets [1–5]. Although various characteristics of the financial market have been revealed from information generated in the financial market"called stylized factors"deviating from the random walk process of the efficiency market hypothesis (EMH) [7–14], studies that attempt to explain the financial market using information generated from social media are still insufficient. In this paper, we test the nonlinear characteristics of social media activity for firms listed on the Korean stock market. We examine the statistical and nonlinear features that have already been found in economic systems, including fat tails in risk asset prices [7–10], long-range correlation [11–14], and multifractality [15–25]. The dataset used, which includes the aggregated social media activity of firms in both the KOSPI and KOSDAQ markets, was constructed from Naver Finance (http://finance.naver.com/), which contains messages written by investors at the firm level. For each firm at a given time, we find many messages created by numerous writers. We test whether valuable information from social media activity relates to multifractality, which can be seen as a feature of complex systems and which is observed in several fields [15–25]. In particular, multifractality is measured from nonlinear time series using several methods, including the partition function method [26], wavelet transform method [27], generalized Hurst exponent method [28], and multifractal detrended fluctuation analysis [MF-DFA] method [29]. Here, we employ the MF-DFA method to quantify multifractality observed in the behavior of writers engaged in the

✉ Gabjin Oh
phecogjoh@chosun.ac.kr

[1] College of Business, Chosun University, Gwangju 61452, South Korea

social media system on the stock market because as a stable means of calculating the scaling relationships in nonlinear time series. We investigate social media activity on the stock market in terms of the multifractal structure of SMA using datasets for 2018, 2019, and 2020. The main contribution of this paper is to analyze the nonlinear properties of SMA for firms listed on the Korean stock market and to find the relationship between complexity and market stability. We calculate the multifractality for the SMA and its differences and the associated shuffled time series and find that a multifractal structure exists regardless of the subperiod considered. The degree of multifractality, defined as the difference the maximun and minimun values of local Hurst exponent, reaches a maximum value during the COVID-19 pandemic as a financial crisis. The next section describes the dataset and methodology used. Section 4 presents the empirical results. Finally, conclusions are reported in Sect. 5.

2 Data

We employ message data from the Naver Finance website to measure social media activity. We introduce our sample data gathered from social media sites. As mentioned in the introduction, social media activity (SMA), defined by the number of postings for a given day, reflects investor sentiment in the Korean stock market. SMA is measured from the number of messages listed on the Naver Finance website. To minimize stylistic noise in messages, we sum all messages created in a given time period so that their activity values are systematically similar across different firms and times. This ensures that we do not select social media features that capture variation in the Naver Finance data for different firms and days. We employ several datasets and different time scales, including periods of hours and days. Here, we consider the aggregated messages created on all firms listed on the Korean stock market with an hourly frequency to analyze the nonlinear features of social media activity on the Korean stock market for January 2018 to December 2020. The data used were gathered from Naver Finance. Social media activity was studied for the Korean stock market with the total messages of firms defined as follows:

$$\mathrm{SMA}(t) = \sum_{i=1}^{n} \mathrm{SMA}_i(t), \qquad \mathrm{SMA}_i(t) = \sum_{t=t}^{\delta t} \mathrm{SMA}_i(t) \quad (1)$$

where $\mathrm{SMA}_i(t)$ and δt denote the social media activity of a potential investor of a certain firm i at hour t and $\mathrm{SMA}(t)$ corresponds to the total number of investor messages. We use the difference value of SMA(t) to analyze the varied effects of social media activity in terms of complexity, which are defined as follows.

$$\mathrm{DSMA}(t) = \mathrm{SMA}(t) - \mathrm{SMA}(t-1) \quad (2)$$

3 Methodology

To examine whether time variation in social media activity (SMA) on the Korean stock market is related to nonlinearity and whether it corresponds to market stability, we investigate the nonlinear dynamics of SMA and the difference in SMA using the multifractal detrended fluctuation analysis (MF-DFA) method [28]. The MF-DFA method is defined by the three steps. In the first step, we calculate the accumulated data from $\mathrm{SMA}(t)$ with a zero mean and are defined as follows.

$$y(t) = \sum_{k=1}^{t} [\mathrm{SMA}(k) - \overline{(\mathrm{SMA})}], \quad \overline{(\mathrm{SMA})} = \frac{1}{t}\sum_{k=1}^{t} \mathrm{SMA}(k) \quad (3)$$

where $\mathrm{SMA}(k)$ is the total social media activity at time k and $\overline{(\mathrm{SMA})}$ is the mean value of $\mathrm{SMA}(k)$. Here, $y(t)$ defined in Eq. (3) can be represented as the accumulated data.

Next, we divide $y(t)$ into N_s boxes of size s. To calculate the fluctuations in each box, we estimate the trend in given box v by an m-order polynomial using the least-square method, which is subtracted from the original data in box v, defined as follows:

$$F_2(s,v) \equiv \frac{1}{s}\sum_{i=1}^{s} (y((v-1)s+i) - y_v(i))^2 \quad (4)$$

for each box v, $v = 1, ..., N_s$ and

$$F_2(s,v) \equiv \frac{1}{s}\sum_{i=1}^{s} (y(N-(v-N_s)s+i) - y_v(i))^2 \quad (5)$$

for each box v, $v = N_s+1, ..., 2N_s$. $y_v(i)$ is the local trend of box v estimated by the m-order polynomial fitting function. Here, we establish a quardratic polynomial fitting function (DFA-2). Finally, to estimate the nonlinearity or complexity of SMA, we employ scaling function $F_q(s)$ and is calculated with its q-order moment by the mean of the appropriate function of F_q for all boxes. Scaling function $F_q(s)$ with q is defined as follows:

$$F_q(s) \equiv \frac{1}{N_s}\sum_{v=1}^{N_s} (F_2(s,v)^{q/2})^{(1/q)} \sim s^h(q) \quad (6)$$

To examine whether the time evolution in SMA is nonlinear and whether it is related to market stability, we consider whether scaling function $F_q(s)$ with q depends on scale s. Then, to determine whether the SMA data are nonlinear, we estimate scaling function $\tau(q)$ using the following equation:

$$\tau(q) = qh(q) - D_{\mathrm{f}}, \quad (7)$$

where D_f is the fractal dimension. Here, D_f is equal to one because of the 1-dimensional time series. We analyze the temporal structure of SMA using multifractal exponent $\tau(q)$. $\tau(q)$ can reflect the various fluctuations from negative $(-q)$ to positive (q). We analyze the relationship between $\tau(q)$ and q to test nonlinearity. We can consider multifractality in social media activity if there is a nonlinear relationship, while it is considered monofractal when $\tau(q) = \alpha q$. Next, we consider the dimension and local Holder exponent using the multifractal spectrum, which can be estimated by a Legendre transform of $\tau(q)$ defined as follows.

$$f(\alpha) \equiv \alpha q - \tau(q), \quad \alpha \equiv \frac{\mathrm{d}\tau(q)}{\mathrm{d}q}, \tag{8}$$

where α is the Holder exponent, which can characterize singularity in a given dataset, and $f(\alpha)$ is the dimension of the data. In examining social media activity on the financial market, it is important that complexity in the social media system be organized by the heterogeneity of writers and nontrivial interactions among them. Rather, the writing process, usually on social media, should be affected by financial market complexity. Therefore, if there are many different writers in terms of investment strategy, α has a broad range, while α of SMA has one value among writers with a homogeneous investment strategy. We test the complexity of social media activity and whether it is related to financial market stability. To do so, we consider an alternative estimator to measure SMA complexity using the range of the local Holder exponent defined by $\Delta\alpha$, $\alpha_{\text{maximum}} - \alpha_{\text{minimum}}$.

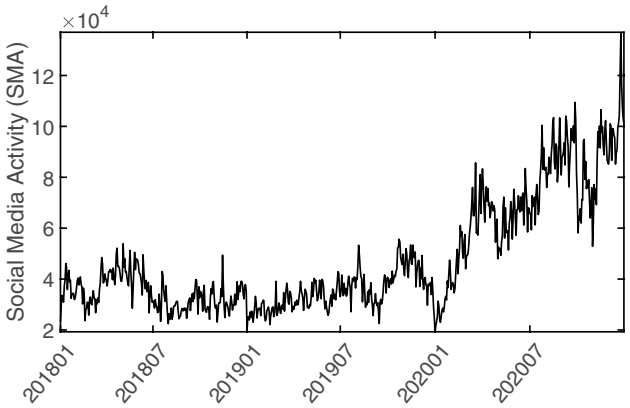

Fig. 1 The evolution of the total number of messages from Naver Finance on the Korean stock market

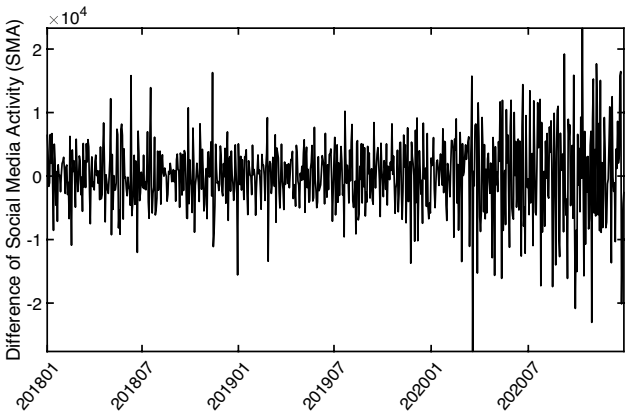

Fig. 2 The evolution of the difference in the total number of messages from Naver Finance on the Korean stock market

4 Empirical results

In a complete market in which investors have rational expectations, the information created in a financial market explains risky assets and will follow a random walk. In other words, financial theory suggests that stock prices should following a random walk with constant drift. However, we can read numerous texts from news, social media, and other professional channels. We find that the temporal dynamics of news media content are characterized in nonlinearity theory and are related to the broad risk of the stock market. We analyze the nonlinear features of both social media activity and its differences as alternative information on the financial market using the multifractal detrended fluctuation analysis (MF-DFA) method for Naver Finance content. We test whether information from social media on the Korean stock market is related to nonlinearity and market stability. First, to investigate the statistical characteristics of social media activity on the stock market, we estimate the probability density function of the difference in SMA. Figure 1 shows the aggregated messages posted on Naver Finance from January 2018 to December 2020, and Fig. 2 displays the difference in SMA (DSMA) defined by Eq. (2). We find that the aggregated number of messages follows a mean-reverting process before the COVID-19 pandemic, while SMA has increased monotonically since 2020, as shown in Fig. 1. Figure 3 shows the cumulative distribution function (CDF) of the DSMA on the Korean stock market obtained using both log-log(a) and linear-log(b) plots. The red, green, yellow, and black lines indicate the empirical data, power-law, stretched exponential, and Gaussian distribution, respectively. In Fig. 3, we find that the cumulative distribution function of the DSMA follows a stretched exponential distribution with $\beta = 0.85$, which deviated from the normal distribution in the efficiency market hypothesis. To further verify the heavy tail feature of the probability distribution function (PDF) of social media activity, we calculate the kurtosis of PDF and find kurtosis of 7.3. In other words, there is a large change over time in social media activity. As a measure of temporal correlation, we employ a detrended fluctuation analysis (DFA-2) and

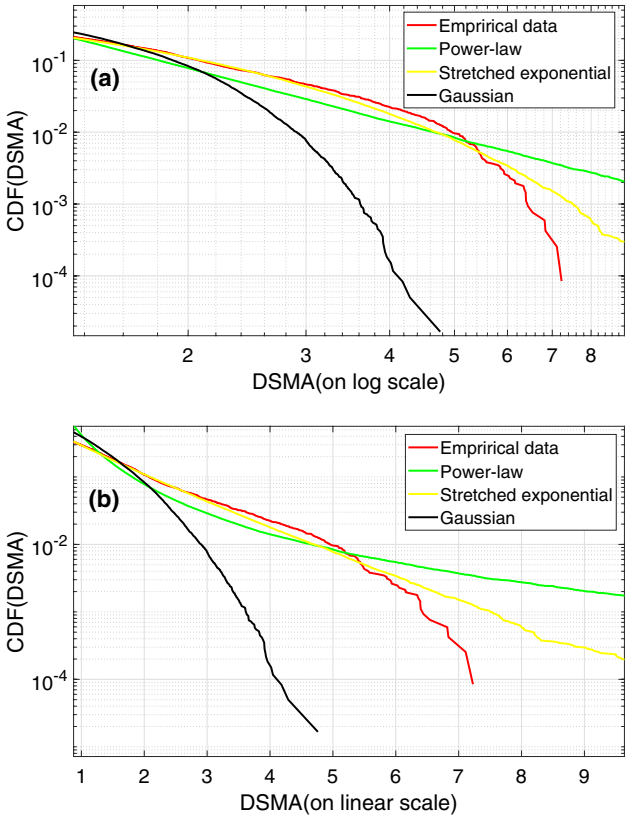

Fig. 3 The cumulative distribution function (CDF) of the difference in the total number of messages on the Korean stock market obtained using both log-log (**a**) and linear-log (**b**) plots. The red, green, yellow, and black lines indicate the empirical data, power-law, stretched exponential, and Gaussian distribution, respectively. The CDF of the empirical data follows a stretched exponential distribution with $\beta = 0.85$

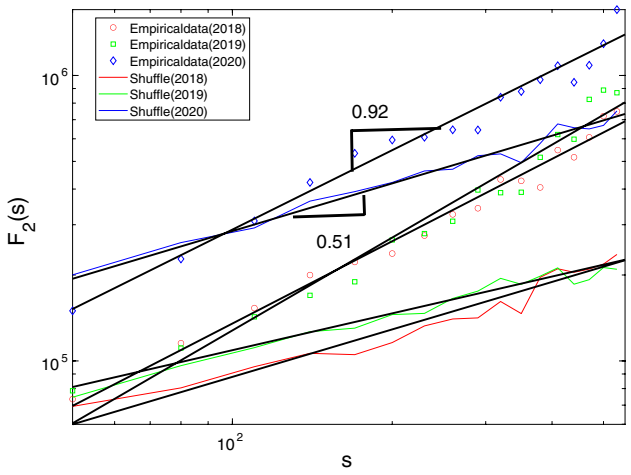

Fig. 4 The figure shows the relationship between the scaling function $F_2(s)$ and s for the total number of messages for three periods, including 2018, 2019, and 2020. The red circles, green squares, and blue diamonds indicate the total number of messages for the three periods, respectively. The Hurst exponent estimated by the linear relationship between the scale and variation function shows a long-range correlation regardless of the subperiod considered, while the shuffle dataset for the three datasets shows a random walk process with an exponent of 0.5

test whether the temporal correlation properties of social media activity on the financial market deviate from a random walk process described by the efficiency market hypothesis (EMH). The Hurst exponents calculated by the DFA-2 method can be interpreted as the existence of a memory in a given dataset and ranged from 0 to 1. The Hurst exponent is expected to be larger (smaller) than 0.5 if SMA has a long (short)-range correlation. We also employ the shuffling method to test whether the observed temporal correlation results are calculated by chance. We estimate the relationship between $F_2(s)$ and s, and the Hurst exponents can be calculated from the linear relationships in the log-log plots, which are shown in Fig. 4. In Fig. 4, the Hurst exponents for the three datasets (2018, 2019, and 2020) are greater than 0.5, while the Hurst exponent shuffled datasets show a value of approximately 0.5. This means that the opinions of writers using Naver Finance show long-term correlations regardless of the dataset considered.

To better understand the nonlinear properties of social media activity on the Korean stock market, we analyze the multifractality of SMA and its differences across the three datasets so that economic status is equal across different years. This ensures that we do not choose economic status based on the variation in social media activity in different states. We use a yearly dataset to measure multifractality to test for the possibility that SMA differs across different years. Figure 5 displays the relationship between $F_q(s)$ and s through q in the range of $-5 \leq q \leq 5$ and confirms the multifractality of SMA. We also created surrogate data from the original data, which eliminated the temporal correlations through the shuffling method. Figure 6a, c show multifractality scaling functions $\tau(q)$ and q of the SMA and its shuffling time series for the three different years, and (b) and (d) demonstrate multifractal dimensions $f(\alpha)$ and α calculated via Legendre transformation with $\tau(q)$ and q, respectively. In Fig. 6, the multifractality distribution of SMA for 2020, including the COVID-19 pandemic period, is much greater than the multifractality of both 2018 and 2019. The degree of multifractality estimated from the shuffled data, after eliminating temporal correlation and preserving the probability density function, is diminished but remains. The heavy tail distribution of SMA should contribute to the multifractality and temporal correlation. We observe similar results for the difference in SMA.

To assess whether the complexity defined from the multifractality relates to market status, we attempt to calculate multifractality in subperiods including 60 days of approximately one quarter. These tests may determine the specific market status underlying the multifractality results.

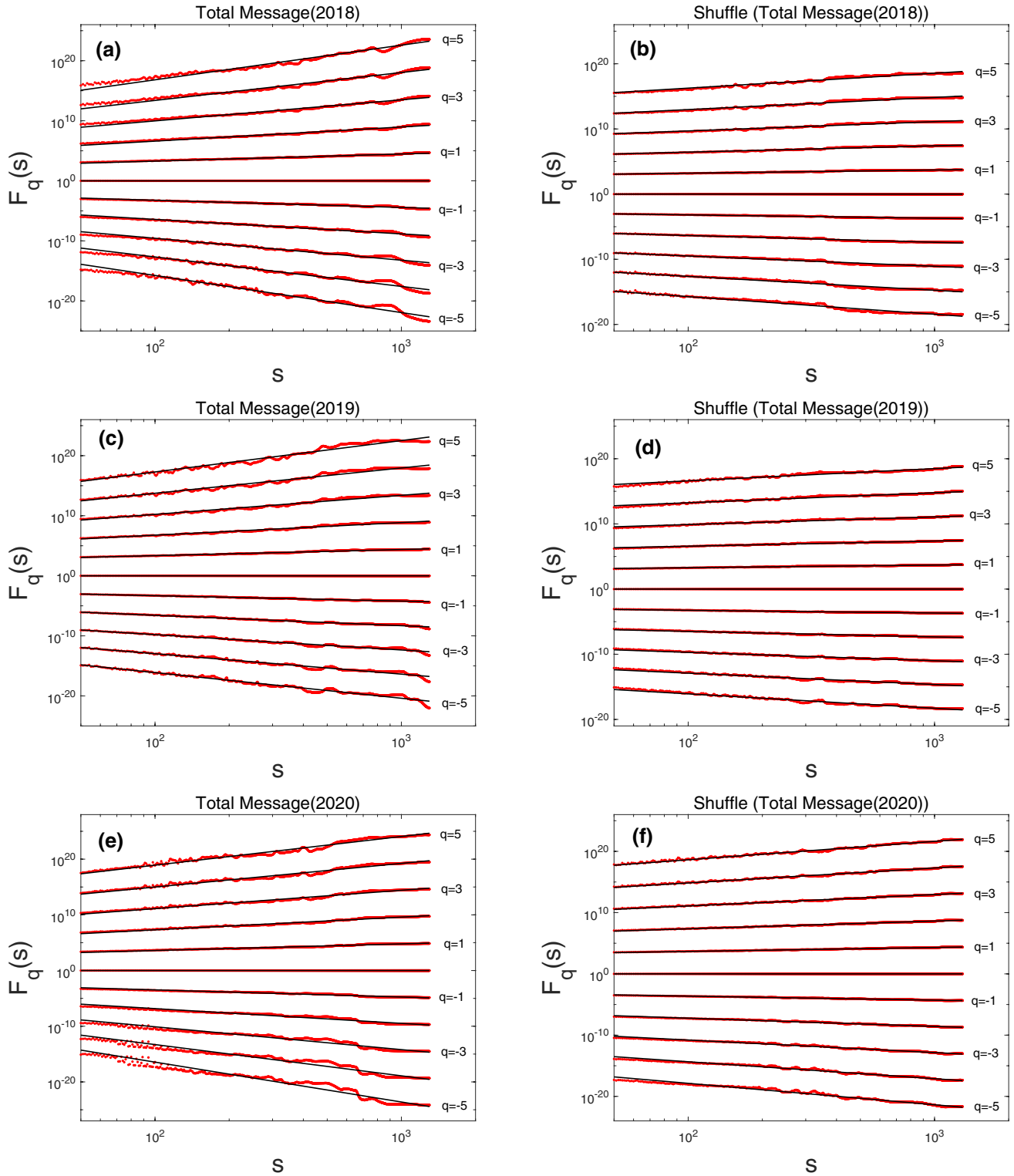

Fig. 5 The relationships between $F_q(s)$ and s with the moment q ranges from -5 to 5 for six data set: **a** total number of message (TNM (2018)), **b** the shuffled data of total number of message (TNM (2018)), **c** total number of message (TNM (2019)), **d** the shuffled data of total number of message (TNM (2019)), **e** total number of message (TNM (2020)), **f** the shuffled data of total number of message (TNM (2020))

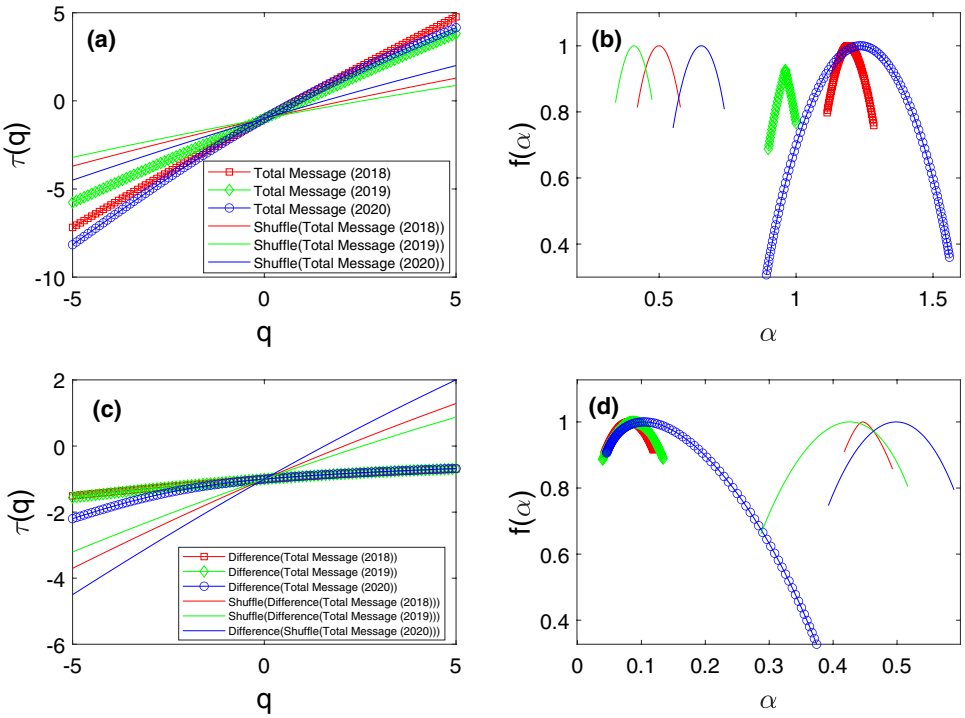

Fig. 6 Panels **a** and **c** show the multifractal scaling functions $\tau(q)$ with q of the empirical data and shuffled time series for three subperiods. Panels **b** and **d** show the multifractal dimension $f(\alpha)$ with α calculated by a Legendre transform from both $\tau(q)$ and τ of **a** and **c**, respectively. The red squares, the green diamons, and blue circles correspond to the total number of messsage (2018), (2019), and (2020)

A decomposition of complexity into investor activity may suggest candidate economical sources of the complexity of SMA that could be tested in further research. We thus test whether the complexity of SMA relates to occurrence of the COVID-19 pandemic as an economic crisis. Figure 7 shows the evolution of complexity estimated by $\alpha_{\max} - \alpha_{\min}$ and the stock market volatility estimated by the KOSPI index return time series, and finds that both complexity and volatility has a maximum value during the COVID-19 pandemic period. The degree of multifractality, $\Delta\alpha$, also increases in the COVID-19 period and sharply decreases after the financial crisis and the correlation between complexity and volatility as market risk over the sample period is also statistically significant at 0.39.

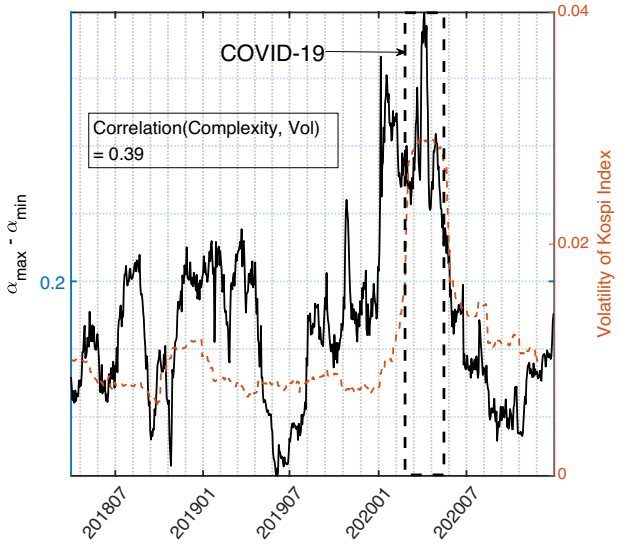

Fig. 7 The evolution of complexity estimated by $\alpha_{\max} - \alpha_{\min}$ and volatility calculated by standard deviation of KOSPI index return time series. The both complexity and volatility reaches a maximum value during the COVID-19 pandemic period

5 Conclusion

The decision-making of investors faced with many alternatives should consider primarily alternatives obtained from valuable information provided through social media. Heterogeneous investors may benefit from the numerous pieces of information generated by social media systems if the attention features of social media activity coincide with characteristics that increase wealth. Recently, with social media influencing numerous retail investors, the utility of a social media activity should be affected by how retail investors select information posted on social media. We argue that many investors want to solve the problem of how to find valuable information from the social media system. This paper investigates the nonlinear properties of social media activity (SMA) on the Korean stock market. We construct a measure of social media content that appears to correspond to investor activity. The multifractality of SMA is consistent with the results of the financial market. The market volatility of the financial market places upward pressure on the complexity of social media activity; high values of SMA

lead to temporarily low stability in the financial market. Furthermore, the SMA impact of the COVID-19 pandemic as a financial factor appears to be especially complex. This result is consistent with the sentiment theories of investors under the assumption that social media content is related to the behavior of individual investors. In this paper, we address how social media activity on different economic statuses relates to the complexity and stability of the financial market. To do so, we analyzed the statistical and nonlinear features of social media activity on the Korean stock market to find the underlying dynamics involved. We used the aggregated social media activity of firms listed on both the KOSPI and KOSDAQ markets and the associated differences in SMA to estimate the cumulative density function and Hurst exponents. We analyzed the multifractal structure via multifractal detrended fluctuation analysis (MF-DFA). We find that the cumulative distribution function of the difference in SMA follows a stretched exponential distribution with $\beta = 0.85$, which deviated from the normal distribution in the efficiency market hypothesis. The Hurst exponent of SMA for three datasets (2018, 2019, 2020) is higher than 0.9, whereas the Hurst exponents of shuffled time series have values of approximately 0.5. Finally, we find a multifractal structure in both SMA and SMA difference results irrespective of the period and degree of multifractality, where $\alpha_{\max} - \alpha_{\min}$ reaches a maximum value during the COVID-19 pandemic as a financial crisis. Our results imply that the underlying dynamics of SMA should be characterized by nonlinear dynamics or complex theory.

Acknowledgements This study was supported by research fund from Chosun University (K206026015-1, 2018).

References

1. M. Baker, J. Wurgler, J. Financ. **61**, 1645–1680 (2006)
2. P.C. Tetlock, J. Financ. **62**, 1139–1168 (2007)
3. Z. T. Ke, B. T. Kelly, D. Xiu, Natl. Bur. Econ. Res. w26186 (2019)
4. Z. Da, J. Engelberg, P. Gao, Rev. Financ. Stud. **28**, 1–32 (2015)
5. J.A. Cookson, M. Niessner, J. Financ. **7**, 173–228 (2020)
6. D. Hirshleifer, S.S. Lim, S.H. Teoh, Rev. Asset Pricing Stud. **1**, 35–73 (2011)
7. M. Gentzkow, B. Kelly, M. Taddy, J. Econ. Lit. **57**, 535–574 (2019)
8. R.N. Mantegna, H.E. Stanley, Nature **376**, 46 (1995)
9. R.N. Mantegna, H.E. Stanley, Nature **383**, 587 (1996)
10. V. Plerou, P. Gopikrishnan, H.E. Stanley, Nature **421**, 130 (2003)
11. X. Gabaix, P. Gopikrishnan, V. Plerou, H.E. Stanley, Nature **423**, 267 (2003)
12. Y. Liu, P. Gopikrishnan, Cizeau, Meyer, Peng, H.. E. Stanley, Phys. Rev. E **60**, 1390 (1999)
13. K. Yamasaki, L. Muchnik, S. Havlin, A. Bunde, H.E. Stanley, Proc. Natl. Acad. Sci. USA **102**, 9424 (2005)
14. G. Oh, S. Kim, C. Eom, J. Korean Phys. Soc. **48**, 197 (2006)
15. W.C. Jun, G. Oh, S. Kim, Phys. Rev. E **73**, 066128 (2006)
16. P. Norouzzadeh, B. Rahmani, Phys. A **3**(67), 328–336 (2006)
17. H. Kim, G. Oh, S. Kim, Physica A **30**, 4286–4292 (2011)
18. G. Oh, J. Korean Phys. Soc. **64**, 1751–1757 (2014)
19. C.K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, A.L. Goldberger, Phys. Rev. E **49**, 1685–1689 (1994)
20. S.V. Buldyrev, A.L. Goldberger, S. Havlin, R.N. Mantegna, M.E. Matsa, C.-K. Peng, M. Simons, H.E. Stanley, Phys. Rev. E **24**, 5084–5091 (1995)
21. A. Bunde, S. Havlin, J.W. Kantelhardt, T. Penzel, J.-H. Pete, K. Voigt, Phys. Rev. Lett. **85**, 3736–3739 (2000)
22. P. Talkner, R.O. Weber, Phys. Rev. E **62**, 150–160 (2000)
23. PCh. Ivanov, L.A.N. Amaral, A.L. Goldberger, S. Havlin, M.G. Rosenblum, Z.R. Struzik, H.E. Stanley, Nature **399**, 461–465 (1999)
24. G. Oh, J. Korean Phys. Soc. **71**, 19–27 (2017)
25. M.U. Rehman, N. Ahmad, X.V. Vo, Phys. A **587**, 126489 (2022)
26. Z.Q. Jiang, W.X. Zhou, Physica A **387**, 1585–1592 (2008)
27. Z.R. Struzik, A.P.J.M. Siebes, Phys. A **309**, 388–402 (2002)
28. D. Grech, Z. Mazur, Phys. A **336**, 133–145 (2004)
29. J.W. Kantelhardt, S.A. Zschiegner, E.K. Bunde, S. Havlin, A. Bunde, H.E. Stanley, Phys. A **316**, 87 (2002)