Fractality and Multifractality Analysis of News Sentiments Time Series

Sergei Sidorov, *Member, IAENG,* Alexey Faizliev and Vladimir Balash

***Abstract*—Using artificial intelligence algorithms, providers of news analytics calculate the sentiment score of almost every economic and financial news in real time. The sentiment score of negative, neutral, positive news are assigned to be -1, 0, 1, respectively. We constructed time series of news sentiments as follows: a nine-month period of 2015 was divided into non-overlapping consecutive intervals of equal length, and then we calculated the sum of sentiment scores of all news within each time interval. In this paper we examine long-range dependance and self-similarity of time series of sentiments of economic and financial news using the Fluctuation Analysis (FA), the Detrended Fluctuation Analysis of order 1 (DFA), Rescaled Range Analysis (R/S), Average Wavelet Coefficient Method (AWC) and Fourier Transform Method (FTM). Moreover, in this paper we use Multifractal Detrended Fluctuation Analysis (MDFA) to examine multifractality of time series of news sentiments. In addition, we analyzed three time series of news sentiments for companies traded on the London Stock Exchange (LSE), the New York Stock Exchange (NYSE) and The Stock Exchange of Hong Kong (SEHK). Empirical results obtained by this methods show that time series of news sentiments exhibit self-similarity (as well as a long memory property). The Hurst exponent (as well as the long-range correlation exponent) is greater than 0.55 over four orders of magnitude in time ranging from several minutes to dozen of days. Moreover, we show that both NYSE and LSE time series of news sentiments exhibit multifractality. On the other hand, we demonstrate that the SEHK time series have monofractal behavior.**

***Index Terms*—long-range correlation, time series, detrended fluctuation analysis, Hurst exponent.**

I. Introduction

A multifractal system could be considered as a generalization of a fractal system. Multifractal is an extension of a fractal in which a single Hurst exponent is not sufficient to characterize its behavior, and therefore we need a continuous spectrum of Hurst exponents. Many papers defined a multifractal structure as a superposition of homogeneous monofractal structures.

Multifractal analysis was initially introduced for the study of turbulence and was applied for the measurement of the turbulent flow velocity in the study of energy dissipation. Over the last thirty years, multifractal as well as fractal analysis have been applied extensively for analysis of medical signals [1], [2], [3], fluid flow [4], time series [5], [6], [7], [8], traffic signals [9], images [10], earthquake signals [11], [12], and in many other areas. The survey of multifractal spectrum computing methods can be found in [13].

Long range dependance and self-similarity of financial time series have been the focus of attention for many researchers in the last several years [14]–[25]. The recent years have seen interest in multifractality analysis of financial time series [19], [20], [22]–[30]. However, there are very few works devoted to the self-similarity or multifractality analysis of the statistical properties of the news flow. Following papers [31], [32], [33] and [34], in this work we analyze time series of news sentiments.

News analytics is an unconventional approach to the analysis of news flow based on artificial intelligence techniques. Introduction to the news analytics tools and techniques can be found in [35] and [36]. For each piece of news, news analytic providers find its sentiment score using artificial intelligence algorithms [35]. The sentiment score of a neutral news is assumed to be 0. Every positive piece of news has sentiment score 1. If news is negative then its sentiment score is assigned to be -1. Then we constructed time series of news sentiments as follows. We use the final data sample and divide the whole period into non-overlapping consecutive intervals of equal length $\delta = 1$ minute and calculated the sum of all news sentiments within each time interval. Note that the sentiment extraction from news is a difficult task and in many real applications it is necessary to use sophisticated technics [37] and big data approaches [38].

This work uses five estimators

- Fluctuation Analysis (FA),
- Detrended Fluctuation Analysis of orders 1 (DFA),
- Rescaled Range Analysis (R/S),
- Average Wavelet Coefficient Method (AWC),
- Fourier transform method

to detect long-range auto-correlation and self-similarity of time series of news sentiments. In addition to the work [31], in this paper we use Multifractal Detrended Fluctuation Analysis (MDFA) to examine multifractality of time series of news sentiments.

Our analysis is based on a 189 trading days period from news analytic data taken from January 1, 2015 to September 22, 2015. In addition to analyzing time series of intensity for all available news from all stock exchanges, the following research question was of our interest: is the multifractal structure of news sentiments similar for different stock exchanges? We examine time series of news sentiments related to different time zones. We consider news affecting companies trading on the European, American and Asian stock exchanges separately. More precisely, we analyzed companies trading on the London Stock Exchange (LSE), the New York Stock Exchange (NYSE) and The Stock Exchange of Hong Kong (SEHK). As the paper [39] we excluded from our analysis periods of low news intensity, e.g. weekends and holidays as well as night hours at local time.

In different time zones, hours of high and low news activity do not coincide, although overlap partially. It may be explained by the partial intersection of the lists of trading companies and agents that generate news. We assume that the fractal structure of time series of news sentiments is

Manuscript received January 08, 2018; revised March 08, 2017. This work is supported by RFBR (grant 16-01-00507).

V. A. Balash, S. P. Sidorov, A. R. Faizliev are with Saratov State University, Saratov, 410012 RUSSIA e-mail: sidorovsp@info.sgu.ru.

more complex for bigger stock exchanges and for larger differentiation of the agents generated news.

The results of this work show that the long-range power-law correlation takes place in time series of news sentiments. All methods showed the presence of a long-range correlation in the time series of news sentiments (the Hurst exponent is greater than 0.5). Moreover, this paper shows that time series of news sentiments exhibit multifractality.

II. Methods for detecting auto-correlation

A. Auto-correlation types

Let $X = (x_t)_{t=1}^n$ be a time series with large n and let $s \in \mathbb{N}$, $s \ll n$. Let $C(s)$ denote the (auto) correlation between $X_1 = (x_t)_{t=1}^{n-s}$ and $X_2 = (x_{t+s})_{t=1}^{n-s}$. The correlation can be of several types:

1) x_t are *uncorrelated*; it is obvious that if X_1 and X_2 are uncorrelated then $C(s)$ must be zero, $C(s) = 0$;
2) *short-range correlations* of $(x_t)_{t=1}^n$ results in exponentially declining of $C(s)$, i.e. $C(s) \sim e^{-s/s_0}$, where s_0 is a parameter of decay;
3) long-range correlation of the $(x_t)_{t=1}^n$, $C(s)$ should follow a power-law dependence: $C(s) \sim s^{-\gamma}$, $0 < \gamma < 1$.

We use the following methods for estimation of the correlation.

B. Fluctuation analysis (FA)

FA algorithm consists of two steps:

1) *Integration.* We calculate the cumulative deviate series as follows
$$y_k = \sum_{t=1}^{k} (x_t - \overline{x}),\ k = 1, 2, \ldots, n, \tag{1}$$
where $\overline{x} = \sum_{t=1}^{n} x_t$.
2) The *FA fluctuation function* is defined by
$$F(s) = F_{\text{FA}}(s) = \left(\frac{1}{n-s} \sum_{k=1}^{n-s} (y_{k+s} - y_k)^2 \right)^{1/2}. \tag{2}$$

Note that (2) is actually a special case of the structure function in turbulence [40].

C. DFA method

DFA was proposed in the papers [41], [42]. This method is used for studying the indirect scaling of the long-range dependence. DFA method was effectively applied for analysis of economical and finance time series in papers [14]–[18]. While DFA has some drawbacks [43], the work [44] remarks that DMA and DFA stay the options of choice for evaluating the long-range correlation of time series. DFA algorithm includes five stages:

1) *Integration.* We calculate y_k, $k = 1, 2, \ldots, n$, using (1).
2) *Cutting.* We divide the $(y_k)_{k=1}^n$ into $n_s = [n/s]$ non-crossing intervals, each of length s, starting with y_1.
3) *Fitting.* For each interval $l = 1, \ldots, n_s$ we construct a fitting linear function P_l (trend) by means of least-square fit of the data $(y_k)_{k=(l-1)s+1}^{ls}$. Denote $y_k^* = P_l(k)$.
4) *Detrending.* The detranded time series are obtained by $\epsilon_k(s) = y_k - y_k^*$.
5) We calculate variance of residuals ϵ_i:
$$F_l^2(s) = \frac{1}{s} \sum_{i=1}^{s} \epsilon_{(l-1)s+i}^2(s),\ l = 1, \ldots, n_s.$$
6) *DFA fluctuation function* is defined by
$$F(s) = F_{\text{DFA}}^{[m]}(s) = \left(\frac{1}{n_s} \sum_{l=1}^{n_s} F_l^2(s) \right)^{1/2}.$$

Derivation of DFA can be found in the work [45].

D. The auto-correlation parameter

The fluctuation functions $F(s)$ deduced in DFA let us examine the s-dependance of F. In the case of long-range power-law correlation of x_t, the F must follow a power-law
$$F(s) \sim s^{\alpha}$$
for large enough s. The fluctuation parameter α is connected with the value of correlation exponent γ in the following way (see [46]) $\alpha = 1 - \gamma/2,\ 0 < \gamma < 1$. The correlation parameter α reflects the auto-correlation properties of time series as follows:

1) $\alpha = 1/2$, the time series uncorrelated (white noise) or short-range correlated;
2) $\alpha < 1/2$, it is anti-correlated;
3) $\alpha > 1/2$, it is long-range power-law correlated;
4) $\alpha = 1$, pink noise ($1/f$ noise).

E. Rescaled Range Analysis (R/S)

The R/S algorithm is one of the most widely used methods for scale exponent estimation. R/S algorithm estimates the value of the Hurst exponent based on an empirical data set for the long-range dependent process that generated the data set. R/S-analysis uses a heuristic approach developed in [47], [48], [49], [50], [51].

R/S algorithm consists of the following steps:

1) *Cutting.* We divide the $(x_k)_{k=1}^n$ into $n_s = [n/s]$ non-overlapping intervals $X^{(i)}(s) := \{x_{(i-1)s+1}, \ldots, x_{is}\}$, $i = 1, \ldots, n_s$, each of length s, starting with x_1.
2) *Accumulation.* We calculate the accumulated series for each window $X^{(i)}(s)$, $i = 1, \ldots, n_s$, as follows
$$y_j^{(i)}(s) = \sum_{t=(i-1)s+1}^{j} (x_t - \overline{x}(s)), \tag{3}$$
where $\overline{x}(s) = 1/s \sum_{t=(i-1)s+1}^{is} x_t$ is the mean of ith window of size s, and $j = (i-1)s+1, \ldots, is$.
3) *Range calculation.* We compute the range of deviation within each window for the accumulated series as follows $R^{(i)}(s) = \max_{(i-1)s+1 \leq j \leq is}\{y_j^{(i)}(s)\} - \min_{(i-1)s+1 \leq j \leq is}\{y_j^{(i)}(s)\}$.
4) *Standard deviation calculation.* For each window $i = 1, \ldots, n_s$ we find the standard deviation:
$$S^{(i)}(s) = \sqrt{\frac{1}{s} \sum_{t=(i-1)s+1}^{is} (x_t - \overline{x}(s))^2}.$$

5) *Computation of R/S statistics:*

$$\langle R(s)/S(s)\rangle := \frac{1}{n_s}\sum_{i=1}^{n_s}\frac{R^{(i)}(s)}{S^{(i)}(s)},\ s = s_1, \ldots, s_L.$$

6) *Estimation of Hurst exponent.* We solve the following least mean square problem

$$\sum_{j=1}^{L}\left(\log\left(\langle R(s_j)/S(s_j)\rangle\right) - H\log s_j - b\right)^2 \to \min_{H,b}$$

R/S algorithm computes the R/S-statistic for different s and plots the resulting estimates versus the s on loglog scale. Then the Hurst parameter can be estimated via the slope of the resulting log-log plot. As it pointed out in [52], classical R/S-analysis is not suitable for small samples, but can be highly effective for quite large samples and it often provides a rather accurate picture of the presence or absence of long-range dependence in a given empirical data sets. Moreover, R/S-analysis demonstrates relative robustness under heavy tails with infinite variance in the marginal distribution of the data [47]–[50].

On the other hand, R/S-analysis is quite sensitive relative to the presence of explicit short-range dependence structures and its bias. Therefore, these shortcomings of R/S analysis lead to the fact that many researchers do not consider this algorithm as a rigorous statistical method.

III. Wavelet-based methods

A. The Hurst exponent

The self-similarity parameter $0 < H < 1$ of self-affine processes is also called the Hurst (or roughness) exponent [53]. The Hurst exponent is commonly used for measuring the duration of long-range dependence of a stochastic process.

There are three possibilities:

- If $H = 0.5$ then $C(s) = 0$ which means that past and future increments are uncorrelated (Brownian motion);
- In case $H > 0.5$ we have $C(s) > 0$ and the increments are positively correlated (the process $\{X(t)\}_t$ is called persistent).
- If $H < 0.5$ then $C(s) < 0$ and increments are negatively correlated (the process $\{X(t)\}_t$ is called anti-persistent or anti-correlated).

B. The Average Wavelet Coefficient Method

The Average Wavelet Coefficient Method (AWC) was proposed in papers [54] and [55].

The method is used for measuring the temporal self-affine correlations of a time series by estimating its Hurst exponent. The AWC method is based on the wavelet transform (good review of the wavelet transform can be found in books [56] and [57]).

The strategy for the data-analysis by the AWC method consists of three main steps:

1) Wavelet transformation of the data $X(t)$ into the wavelet domain, $\mathcal{W}[X](a, b)$, where a, b are scale and location parameters, respectively [56], [57].
2) Then for a given scale a we can find a representative wavelet amplitude for that particular scale, and to study its scaling. To do so we calculate the averaged wavelet coefficient $W[X](a)$ according to the equation
$$W[X](a) = \langle|\mathcal{W}[X](a, b)|\rangle_b,$$
where $\langle\cdot\rangle_b$ denotes the standard arithmetic mean value operator with respect to the b.
3) For a self-affine process $X(t)$ with exponent H, the spectrum $W[X](a)$ should scale as $a^{H+0.5}$ [55]. Therefore, to find $H + 0.5$ we should plot $W[h](a)$ against scale a in a log-log plot. As it is pointed out in [55], a scaling regime consisting of a straight line in this plot implies a self-affine behavior of the data.

C. Fourier transform method

Let us recall that the Fourier transform of the function f and inverse transform are defined by

$$\widehat{f}(\zeta) = \sum_t e^{-i\zeta t} f(t), \tag{4}$$

$$f(t) = \sum_\zeta e^{i\zeta t}\widehat{f}(\zeta). \tag{5}$$

Fourier inversion theorem states that f can be reconstructed from $\widehat{f}$.

If $\rho_x(s)$ is autocorrelation function of $x(t)$, $\rho_x(s) = \sum_t x(t)x(t+s)$, then the Fourier transform of $\rho_x(s)$ is

$$\widehat{\rho}(\zeta) = |\widehat{x}(\zeta)|^2. \tag{6}$$

Therefore, to find the value of $\rho_x(s)$ one should make the following steps

1) Find $\widehat{x}(\zeta)$, the Fourier transform of $x(t)$, using (4).
2) Find $\widehat{\rho}(\zeta)$, the Fourier transform of $\rho_x(s)$, using (6)
3) Compute $\rho_x(s)$ as the inverse transform of $\widehat{\rho}_x(\zeta)$ using (5).

Fourier transform is a classical way of correlation exponent estimation γ, but it is often not appropriate due to noisy nature, non-stationarity and imperfect measurement of data x_t.

$\widehat{\rho}(\zeta)$ is a power spectrum $P(\omega)$ of the Fourier transform. $\omega = 0 : Fs/n : Fs/2$ is a frequency increment (angular frequency), where Fs is sampling frequency, and n - number of samples.

IV. Multifractal Detrended Fluctuation Analysis (MDFA)

The multifractal generalization of DFA [30] can be briefly sketched as follows.

It is well-known that the self-similarity of fractal structures can be described by the global exponent $H\ddot{o}lder$, or the local Hurst exponent α. If the fractal is monofractal, then it can be described by only one $H\ddot{o}lder$ exponent, and in the case of a multifractal, different parts of the structure are characterized by different values of α, which means the existence of a spectrum $f(\alpha)$.

MDFA algorithm includes seven stages:

1) *Integration.* We calculate $y_k = \sum_{t=1}^{k}(x_t - \overline{x})$, $k = 1, 2, \ldots, n$, where $\overline{x} = \sum_{t=1}^{n} x_t$.
2) *Cutting.* We divide the $(y_k)_{k=1}^n$ into $n_s = [n/s]$ non-crossing intervals, each of length s, starting with y_1.

3) *Fitting.* For each interval $l = 1, \ldots, n_s$ we construct a fitting linear function P_l (trend) by means of least-square fit of the data $(y_k)_{k=(l-1)s+1}^{ls}$. Denote $y_k^* = P_l(k)$.
4) *Detrending.* The detranded time series are obtained by $\epsilon_k(s) = y_k - y_k^*$.
5) We calculate variance of residuals ϵ_i:

$$F_l^2(s) = \frac{1}{s} \sum_{i=1}^{s} \epsilon_{(l-1)s+i}^2(s), \ l = 1, \ldots, n_s.$$

6) *MDFA fluctuation function* is defined by

$$F_q(s) = \left(\frac{1}{n_s} \sum_{l=1}^{n_s} F_l^q(s) \right)^{1/q},$$

where q can take any real value except for $q = 0$. When $q = 0$, we have

$$F_0(s) = \exp\left(\frac{1}{4n_s} \sum_{l=1}^{n_s} \ln[F_l^2(s)] \right) \approx s^{h_0},$$

according to L'Hospital's rule.
7) Changing the size of the segment s, we can find a power relation between the function $F_q(s)$ and the size s:

$$F_q(s) \sim s^h(q).$$

Then we calculate the multifractal scale factor $\tau(q)$ which can be considered as a characteristic of multifractality:

$$\tau(q) = qh(q) - D_f$$

where D_f denotes the fractal dimension of the geometric support of the multifractal measure [30], $h(q)$ is generalized Hurst exponent or Holder exponent, which coincides with Hurst exponent for $q = 2$, $h(q)$. When analyzing time series, we have $D_f = 1$. If $\tau(q)$ is a nonlinear function of q then the signal is multifractal.

Another characteristic of multifractality for time series may be the singularity spectrum $f(\alpha)$, which is connected with $\alpha(q)$ through the Legendre transformation as follows [58]:

$$\begin{cases} \alpha(q) = d\tau(q)/dq \\ f(q) = q\alpha - \tau(q) \end{cases}$$

where $\alpha(q)$ is Holder singular index which is also called singular strength and $f(\alpha)$ is a multifractal spectrum which determines dimensions of subsets of the time series that are characterized by $\alpha(q)$.

The curve $\alpha \sim f(\alpha)$ is a single-humped function for a multifractal and is a single point for a mono-fractal. The shape and length of the curve $f(\alpha)$ give us an important information on characteristics of the distribution of the data set. First, we can determine the width of the spectrum $\Delta\alpha = \alpha_{\text{max}} - \alpha_{\text{min}}$. The parameter $\Delta\alpha$ describes the inhomogeneity of the probability distribution measured by the total fractal structure and determines the degree of multifractality. The function $f(\alpha)$ has the maximum value f_{max} at some point α_0, $f_{\text{max}} = f(\alpha_0)$, which corresponds to the peak of the multifractal spectrum [12]. Larger values of $\Delta\alpha$ and f_{max} correspond to a stronger degree of multifractality.

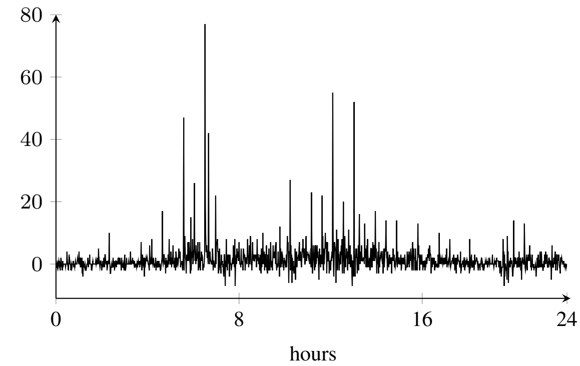

Fig. 1. Dynamics of sentiments (February 2, 2015)

V. Empirical Results

A. *Time series of news sentiments*

Providers of news analytics obtain and aggregate data from different sources (including news agencies and business reports) and social media. Our data cover the period from January 1, 2015 to September 22, 2015 (i.e. 189 trading days). We consider all the news released during this period. Initially we performed data selection and cleaning process as described in [32] or [34].

For each piece of news, news analytic providers find its sentiment score using artificial intelligence algorithms. The sentiment score of neutral news is assumed to be 0. Every positive piece of news has sentiment score 1. If news is negative then its sentiment score is assigned to be -1.

Then we constructed time series of news sentiments as follows. We divided 189-day period Δ into n non-crossing consecutive segments $\Delta_1, \ldots, \Delta_n$ of the same longevity δ minutes, $\Delta = \Delta_1 \cup \ldots \cup \Delta_n$. We found x_t, the sum of all sentiment scores of all economical and finance news reported in the world during each interval Δ_t, $t = 1, 2, \ldots, n$. The sequence $x_1, x_2, \ldots, x_n$ is the time series of news sentiments with the δ minutes window. The overall sentiment of news for the 189-day period is 2011463. Table I shows the summary statistics. We used the time window $\delta = 1$.

TABLE I
Summary statistics of time series, $\delta = 1$ minutes

	All news	LSE	NYSE	SEHK
n	$2.7 \cdot 10^5$	$1.3 \cdot 10^4$	$1.5 \cdot 10^4$	$1.5 \cdot 10^4$
Sum sentiment	279750	90253	32127	8940
Mean	1.03	0.69	0.22	0.06
Minimum	-45	-214	-254	-25
Maximum	237	97	138	24
St. deviation	4.03	3.00	3.18	1.08
Median	0	0	0	0
Skewness	11.54	6.25	-8.31	0.37
Kurtosis	282.18	426.22	601.95	29.67

Fig. 1 plots an example of time series of news sentiments with 1-min window. Time series are quite volatile and demonstrate a non-stationary behavior. The amount of positive news is much greater than the amount of negative news during the period for all stock exchanges.

B. Self-Similarity Analysis

In this subsection we demostrate the auto-correlation and self-similarity analysis with the use of FA, DFA, R/S analysis, AWC and Fourier transform method.

First we use FA and DFA (of order 1) methods to evaluate the correlation and Hurst exponent. Then, we obtained values of $F(s_i)$ for different interval lengths $s_i \in [10^1, 10^{4.5}]$ using FA and DFA. The task of selecting a needed length for scaling range has been examined in the papers [59], [60], [61]. The influence of scaling range on the efficiency of some detranding techniques has been investigated in the same works. We used two ranges $[10^1, 10^{4.5}]$ (for 1-min data) and $[10^1, 10^4]$ (for 5-min data) in our analysis. This values belong to the scaling range suggested by these papers.

Figures 2, 3, 4 and 5 plot the results of FA and DFA application to all news, NYSE news, LSE news and SEHK time-series of news sentiments respectively (with $\delta = 1$). Results of DFA show that the estimates of the Hurst exponent are equal to 0.69 (all news), 0.64 (NYSE), 0.55 (LSE), 0.69 (SEHK) which indicates the presence of long-range correlation in the corresponding time series. Linear regressions are highly significant, since the determination coefficients R^2 are close to 1. It must be pointed out that Detranded Fluctuation Analysis let us discover the crossover effect shown in the study [46].

AWC method gives similar results with DFA method in identical estimates of the Hurst exponent on the whole scale of s both for all time series and NYSE news, LSE news and SEHK time-series of news sentiments (Table II). The differences between estimates of the Hurst exponent obtained by means of different methods are less than 0.1. It should be noted that regression errors are very small and the value of R^2 is close to 1 for FA, DFA and AWC methods. We noted that AWC method also allowed to detect the so called crossover effect (Fig. 7). It should be mentioned that the crossover effect matches with 1 day period (or 24 hours). Estimates of the Hurst exponent, for small (about one day) as well as large scales (more than a day) using DFA and AWC methods yield similar results, the difference being less than 0.1, on the whole scale of for all news, NYSE news, LSE news and SEHK time-series of news sentiments.

Fig. 6 demonstrates the log-log plot of dependance R/S statistics of s for all news time series with $\delta = 1$ obtained by R/S analysis. In this case, though the value of the Hurst exponent is less than $H = 0.63$, it still demonstrates that there are positive long-range correlations in time series. Linear regressions are also highly significant. In contrast to the DFA, R/S analysis as well as FA failed to reveal the presence of the crossover effect.

The Fourier transform analysis of the all news time series of news sentiments (with $\delta = 1$) are presented in Fig. 8. The figure shows the raw power spectrum, $P(w)$ vs. angular frequency w for the news sentiments data. We can see that the raw power spectrum is too noisy, which does not allow us to accurately estimate the Hurst exponent and certainly prevents us from detecting the effect of the crossover. We can apply the log-binning technique [62] to reduce its noise. The result of the log-binning smoothing is shown in Fig. 9. The dashed lines of figures correspond to the slope of $(-2H-1)$ with the value of $H = 0.61$ for all news time series with time window $\delta = 1$. It should be noted that while the regression is significant, it is considerably worse than regressions obtained by FA, DFA, AWC and R/S analysis. Moreover, it is not possible to identify the effect of the crossover even after the smoothing of power spectrum.

Table II presents the estimates of the Hurst exponent obtained by FA, DFA, R/S, AWC and Fourier transform methods for our time series.

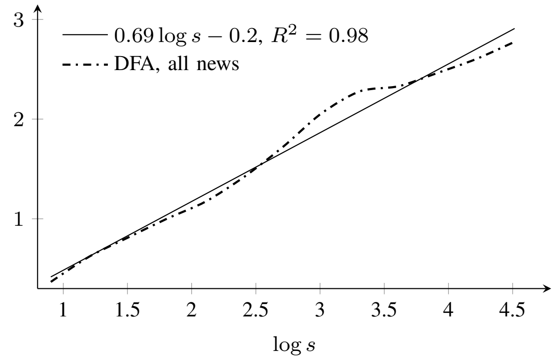

Fig. 2. $\log F(s)$ versus $\log s$ for the DFA estimation method, all news

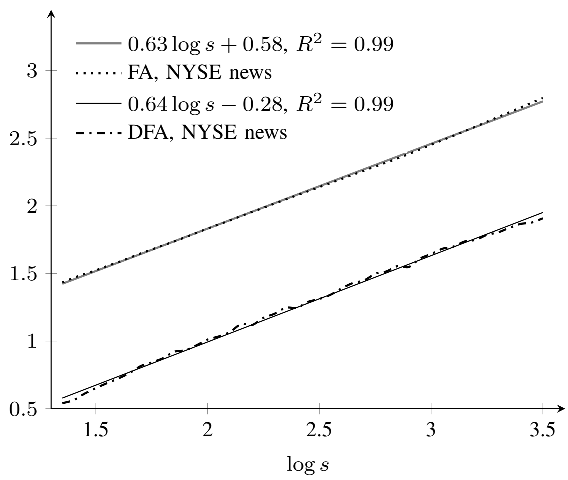

Fig. 3. $\log F(s)$ versus $\log s$ for the FA and the DFA estimation methods, NYSE

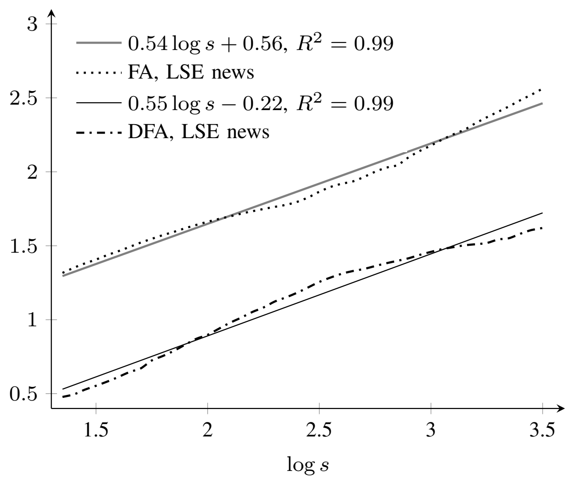

Fig. 4. $\log F(s)$ versus $\log s$ for the FA and the DFA estimation methods, LSE

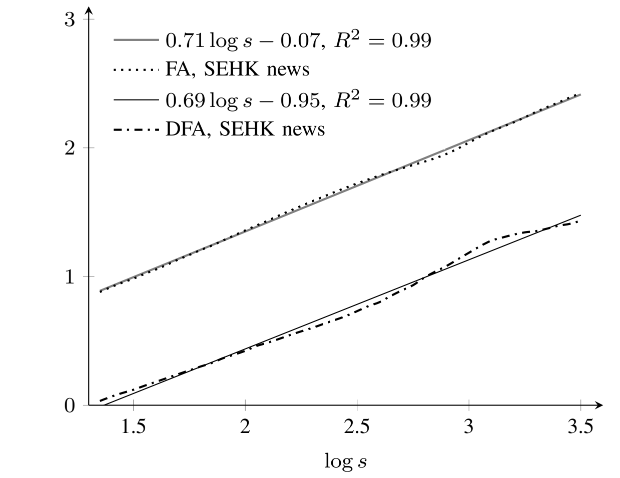

Fig. 5. $\log F(s)$ versus $\log s$ for the FA and the DFA estimation methods, SEHK

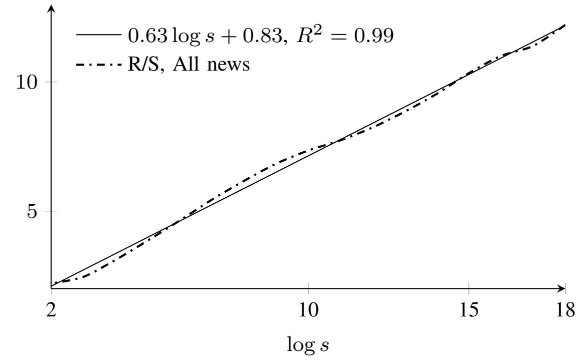

Fig. 6. $\log R/S$ versus $\log s$ for the R/S method, all news

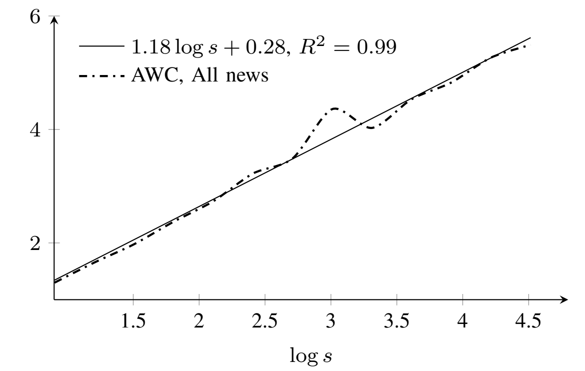

Fig. 7. $\log W[X](a)$ versus $\log a$ for AWC estimation method, all news

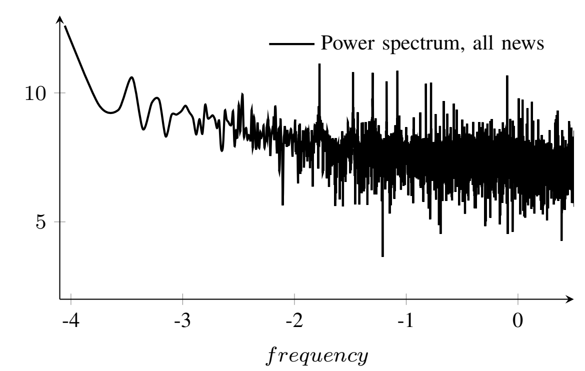

Fig. 8. Results for the Fourier transform method, all news

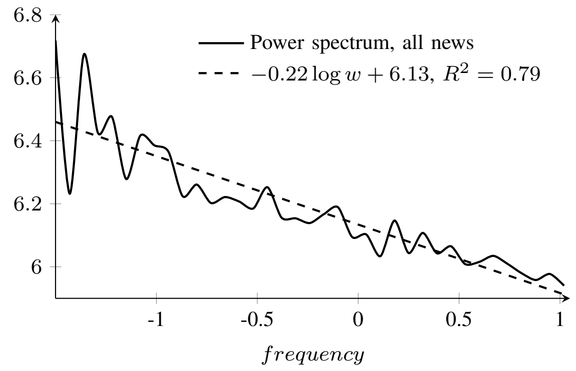

Fig. 9. Results for the Fourier transform method, all news

TABLE II
ESTIMATES OF THE HURST EXPONENT

Method	All	NYSE	LSE	SEHK
FA	0.711	0.627	0.543	0.710
DFA	0.690	0.638	0.554	0.693
R/S analysis	0.631	0.618	0.521	0.710
AWC	0.683	0.722	0.648	0.687
Fourier transform	0.609	0.643	0.588	0.611

C. Multifractality Analysis

In this subsection we examine the multifractality of time series of news sentiments using the MDFA algorithm. Figure 10 shows the dependence of log-values of fluctuation function on the length of non-crossing intervals for different values of $q := -5 : 0.5 : 5$. One can see that slopes are different at distinct points q, which is evidence of multifractality. For periods longer than one day, the alignment of points and slopes can be observed.

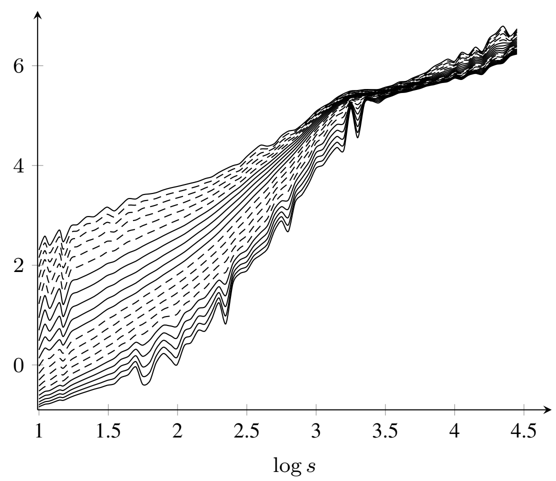

Fig. 10. $\log F(s)$ versus $\log s$ for the MDFA estimation method, $\delta = 1$

In addition, Fig. 11 shows the multifractal spectrum of news sentiments time series. It should be noted that $f(\alpha)$ is in a wide range ($\alpha = 0.82$), and curves are concentrated over a sufficiently large range, which is the indication of multifractality. NYSE and LSE time series exhibit multifractality as well, however their degree of multifractality is less than for the time series generated by all news. At the same time $f(\alpha)$ for SEHK time series is in a narrow range ($\alpha = 0.38$) and the curve is concentrated on a small range, which indicates mono-fractality.

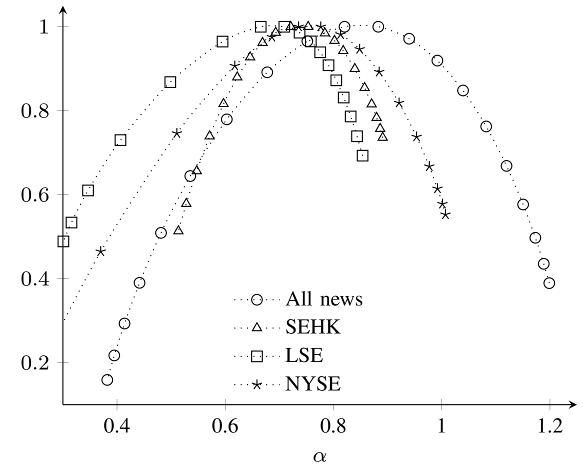

Fig. 11. Multifractal spectrum of $f(\alpha) - \alpha$

VI. Conclusion

In this paper we use the FA, DFA, R/S analysis, AWC method and Fourier transform technique to examine the presence of long-range correlation of financial and economic time series of news sentiments. The results of this paper show that the long-range power-law correlation take place in time series of news sentiments. The paper shows that the behavior of long range dependence for time series of sentiment intensity most similar to the news flow intensity [32]. The results show that the self-similarity property is a stable characteristic of the sentiment of news information flow which serves the financial industry and stock markets. All methods showed the presence of a long-range correlation in the time series of news sentiments (the Hurst exponent is greater than 0.5) both for time series of all news sentiments and for three time series of news sentiments for three major stock exchanges (NYSE, LSE, SEHK). Results obtained by the DFA and AWC for sufficiently large scale have revealed the effect of crossover (corresponding to one day). Shortcomings of the classic FA method, R/S analysis and Fourier technique do not allow us to determine the effect of crossover.

Moreover, we show that the time series of news sentiments exhibit multifractality. This statement can also be attributed to the time series of news sentiments related to individual stock exchanges (NYSE and LSE). On the other hand, we demonstrate that the SEHK time series have monofractal behavior. We can conclude that the fractal structures of time series of sentiments are more complex for large stock exchanges with a wide differentiation of operating agents. At the same time, the presence of multifractality in the time series of news sentiments can be explained by the fact that news is formed in the process of information interaction within a complex system with a large number of subsystems and participants. The elements of the system are more than 34000 largest companies located around the world, financial market participants, investors, traders, analytical and news agencies, media, etc. In our opinion, these are the main reasons why the time series of news sentiments are multifractal and have long-range memory. This paper also revealed a long-range power dependence in high-frequency data for relatively limited time periods (up to 1 week). In our opinion, at such relatively short intervals of time, the correlation between news events is caused more by the delay in the reaction of different agents to events (the time sequence of interactions between agents) than the temporal correlation between news events.

References

[1] A. Ouahabi, Ed., *Signal and Image Multiresolution Analysis*. John Wiley & Sons, Inc., 2012.

[2] E. Gerasimova, B. Audit, S.-G. Roux, A. Khalil, O. Gileva, F. Argoul, O. Naimark, and A. Arneodo, *A Wavelet-Based Method for Multifractal Analysis of Medical Signals: Application to Dynamic Infrared Thermograms of Breast Cancer*. Cham: Springer International Publishing, 2014, pp. 288–300.

[3] S. M. Shekatkar, Y. Kotriwar, K. P. Harikrishnan, and G. Ambika, "Detecting abnormality in heart dynamics from multifractal analysis of ECG signals," *Scientific Reports*, vol. 7, no. 1, p. 15127, 2017.

[4] D. Veneziano and A. K. Essiam, "Flow through porous media with multifractal hydraulic conductivity," *Water Resources Research*, vol. 39, no. 6, p. 1166, 2003.

[5] K. Xu, P. Shang, and G. Feng, "Multifractal time series analysis using the improved 0-1 test model," *Chaos, Solitons & Fractals*, vol. 70, no. Supplement C, pp. 134–143, 2015.

[6] E. Green, W. Hanan, and D. Heffernan, "The origins of multifractality in financial time series and the effect of extreme events," *The European Physical Journal B*, vol. 87, no. 6, p. 129, Jun 2014.

[7] A. L. Karperien, H. F. Jelinek, and H. Ahammer, "Multifractal formalism in image and time series analysis," *Banach Center Publications*, vol. 109, pp. 23–45, 2016.

[8] H. Xiong and P. Shang, "Weighted multifractal analysis of financial time series," *Nonlinear Dynamics*, vol. 87, no. 4, pp. 2251–2266, 2017.

[9] J. Wang, P. Shang, and X. Cui, "Multiscale multifractal analysis of traffic signals to uncover richer structures," *Phys. Rev. E*, vol. 89, no. 3, p. 032916, Mar. 2014.

[10] N. M. G. Al-Saidi, A. J. Mohammed, E. A. Ogada, and A. M. Ahmed, *Fractal Fourier Coefficients With Application To Identification Protocols*. World Scientific, pp. 67–80.

[11] D. Sornette and G. Ouillon, "Multifractal scaling of thermally activated rupture processes," *Phys. Rev. Lett.*, vol. 94, p. 038501, Jan 2005.

[12] L. Telesca, V. Cuomo, V. Lapenna, and M. Macchiato, "Identifying space-time clustering properties of the 1983-1997 Irpinia-Basilicata (southern Italy) seismicity," *Tectonophysics*, vol. 330, no. 1, pp. 93–102, 2001.

[13] R. Lopes and N. Betrouni, "Fractal and multifractal analysis: A review," *Medical Image Analysis*, vol. 13, no. 4, pp. 634–649, 2009.

[14] Y. Liu, P. Cizeau, M. Meyer, C.-K. Peng, and H. E. Stanley, "Correlations in economic time series," *Physica A: Statistical Mechanics and its Applications*, vol. 245, pp. 437–440, 1997.

[15] P. Cizeau, Y. Liu, M. Meyer, C.-K. Peng, and H. E. Stanley, "Volatility distribution in the S&P500 stock index," *Physica A: Statistical Mechanics and its Applications*, vol. 245, pp. 441–445, 1997.

[16] M. Ausloos, N. Vandewalle, P. Boveroux, A. Minguet, and K. Ivanova, "Applications of statistical physics to economic and financial topics," *Physica A: Statistical Mechanics and its Applications*, vol. 274, pp. 229–240, 1999.

[17] W. Willinger, M. S. Taqqu, and V. Teverovsky, "Stock market prices and long-range dependence," *Finance Stochast*, vol. 3, pp. 1–139, 1999.

[18] M. Ausloos and K. Ivanova, "Introducing false EUR and false EUR exchange rates," *Physica A: Statistical Mechanics and its Applications*, vol. 286, pp. 353–366, 2000.

[19] X. Lu, X. Sun, and J. Ge, "Dynamic relationship between japanese yen exchange rates and market anxiety: A new perspective based on MF-DCCA," *Physica A: Statistical Mechanics and its Applications*, vol. 474, pp. 144–161, 2017.

[20] X. Zhang and W. Xiao, "Arbitrage with fractional Gaussian processes," *Physica A: Statistical Mechanics and its Applications*, vol. 471, pp. 620–628, 2017.

[21] J. Kwapień and S. Drożdż, "Physical approach to complex systems," *Physics Reports*, vol. 515, no. 3–4, pp. 115–226, 2012.

[22] D. Stošić, D. Stošić, T. Stošić, and H. E. Stanley, "Multifractal analysis of managed and independent float exchange rates," *Physica A: Statistical Mechanics and its Applications*, vol. 428, pp. 13–18, 2015.

[23] P. Suárez-García and D. Gómez-Ullate, "Multifractality and long memory of a financial index," *Physica A: Statistical Mechanics and its Applications*, vol. 394, pp. 226–234, 2014.

[24] X. Lu, J. Tian, Y. Zhou, and Z. Li, "Multifractal detrended fluctuation analysis of the Chinese stock index futures market," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 6, pp. 1452–1458, 2013.

[25] W. Wang, K. Liu, and Z. Qin, *Multifractal Analysis on the Return Series of Stock Markets Using MF-DFA Method*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 107–115.

[26] ——, "Multifractal analysis on the return series of stock markets using MF-DFA method," in *International Conference on Informatics and Semiotics in Organisations*. Springer, 2014, pp. 107–115.

[27] C. Chen and Y. Wang, "Understanding the multifractality in portfolio excess returns," *Physica A: Statistical Mechanics and its Applications*, vol. 466, pp. 346–355, 2017.

[28] P. Mali and A. Mukhopadhyay, "Multifractal characterization of gold market: A multifractal detrended fluctuation analysis," *Physica A: Statistical Mechanics and its Applications*, vol. 413, pp. 361–372, 2014.

[29] J. R. Thompson and J. R. Wilson, "Multifractal detrended fluctuation analysis: Practical applications to financial time series," *Mathematics and Computers in Simulation*, vol. 126, pp. 63–88, 2016.

[30] J. W. Kantelhardt, S. A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, and H. E. Stanley, "Multifractal detrended fluctuation analysis of nonstationary time series," *Physica A: Statistical Mechanics and its Applications*, vol. 316, no. 1, pp. 87–114, 2002.

[31] S. Sidorov, A. Faizliev, and V. Balash, "Long range correlation in time series of news sentiments," in *Lecture Notes in Engineering and Computer Science: Proceedings of The World Congress on Engineering 2017, 5-7 July, 2017, London, U.K.*, vol. II, 2017, pp. 549–554.

[32] S. Sidorov, A. Faizliev, V. Balash, and E. Korobov, "Long-range correlation analysis of economic news flow intensity," *Physica A: Statistical Mechanics and its Applications*, vol. 444, pp. 205–212, 2016.

[33] S. Sidorov, A. Faizliev, and V. Balash, "Measuring long-range correlations in news flow intensity time series," *International Journal of Modern Physics C*, vol. 28, no. 08, p. 1750103, 2017.

[34] ——, "Scale invariance of news flow intensity time series," *Nonlinear Phenomena in Complex Systems*, vol. 19, no. 4, pp. 368–377, 2016.

[35] G. Mitra and Y. Xiang, Eds., *The Handbook of Sentiment Analysis in Finance*. Albury Books, 2016.

[36] G. Mitra and L. Mitra, Eds., *The Handbook of News Analytics in Finance*. John Wiley & Sons, 2011.

[37] X. Wei, H. Lin, and L. Yang, "Cross-domain sentiment classification via constructing semantic correlation," *IAENG International Journal of Computer Science*, vol. 44, no. 2, pp. 172–179, 2017.

[38] C. Costa and M. Y. Santos, "Big data: State-of-the-art concepts, techniques, technologies, modeling approaches and research challenges," *IAENG International Journal of Computer Science*, vol. 44, no. 3, pp. 285–301, 2017.

[39] S. P. Sidorov, A. Revutskiy, A. Faizliev, E. Korobov, and V. Balash, "Stock volatility modelling with augmented GARCH model with jumps," *IAENG International Journal of Applied Mathematics*, vol. 44, pp. 212–220, 2014.

[40] A. N. Kolmogorov, "A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds number," *J. Fluid Mech.*, vol. 13, pp. 82–85, 1962.

[41] C. Peng, S. Buldyrev, S. Havlin, M. Simons, H. Stanley, and A. Goldberger, "Mosaic organization of DNA nucleotides," *Phys. Rev. E*, vol. 49, pp. 1685–1689, 1994.

[42] C. Peng, S. Buldyrev, A. Goldberger, S. Havlin, F. Sciortino, M. Simons, and H. Stanley, "Long-range correlations in nucleotide sequences," *Nature*, vol. 356, pp. 168–170, 1992.

[43] R. Bryce and K. Sprague, "Revisiting detrended fluctuation analysis," *Scientific Reports*, vol. 2:315, 2012.

[44] Y.-H. Shao, G.-F. Gu, Z.-Q. Jiang, W.-X. Zhou, and D. Sornette, "Comparing the performance of FA, DFA and DMA using different synthetic long-range correlated time series," *Scientific Reports*, vol. 2:835, 2012.

[45] M. S. Taqqu, V. Teverovsky, and W. Willinger, "Estimators for long-range dependence: An empirical study," *Fractals*, vol. 3, pp. 785–798, 1995.

[46] J. W. Kantelhardt, E. Koscielny-Bunde, H. H. Rego, S. Havlin, and A. Bunde, "Detecting long-range correlations with detrended fluctuation analysis," *Physica A: Statistical Mechanics and its Applications*, vol. 295, pp. 441–454, 2001.

[47] B. Mandelbrot and J. Wallis, "Computer experiments with fractional gaussian noises. parts 1,2,3," *Water Resour. Res.*, vol. 5, pp. 228–267, 1969.

[48] ——, "Robustness of the rescaled range R/S in the measurement of noncyclic long-run statistical dependence," *Water Resour. Res.*, vol. 5, pp. 967–988, 1969.

[49] ——, "Some long-run properties of geophysical records," *Water Resour. Res.*, vol. 5, pp. 321–340, 1969.

[50] B. Mandelbrot and M. Taqqu, "Robust R/S analysis of long-run serial correlation," in *Proc. of the 42nd Session of the Int. Statistical Institute*, vol. Bull. Int. Statist. Inst. 48 (Book 2), Manila, 1979, pp. 69–104.

[51] J. Bassingthwaighte and G. Raymond, "Evaluating rescaled range analysis for time series," *Ann. Biomed. Eng.*, vol. 22, pp. 432–444, 1994.

[52] V. Teverovsky, "A critical look at lo's modified R/S statistic," *Journal of Statistical Planning and Inference*, vol. 80, no. 1–2, pp. 211–227, 1999.

[53] J. Feder, *Fractals*. New York: Plenum Press, 1988.

[54] A. R. Mehrabi, H. Rassamdana, and M. Sahimi, "Characterization of long-range correlations in complex distributions and profiles," *Phys. Rev. E*, vol. 56, pp. 712–722, 1997.

[55] I. Simonsen, A. Hansen, and O. M. Nes, "Determination of the hurst exponent by use of wavelet transforms," *Phys. Rev. E*, vol. 58, pp. 2779–2787, Sep 1998.

[56] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia: SIAM, 1992.

[57] D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis*. Cambridge: Cambridge University Press, 2000.

[58] T. C. Halsey, M. H. Jensen, L. Kadanoff, I. I. Procaccia, and B. I. Shraiman, "Fractal measures and their singularities: The characterization of strange sets," *Phys. Rev. A Gen. Phys.*, vol. 33, no. 2, pp. 1141–1151, 1986.

[59] D. Grech and Z. Mazur, "On the scaling ranges of detrended fluctuation analysis for long-term memory correlated short series of data," *Physica A*, vol. 392, pp. 2384–2397, 2013.

[60] ——, "Scaling range of power laws that originate from fluctuation analysis," *Phys. Rev. E*, vol. 87, p. 052809, May 2013.

[61] ——, "Impact of scaling range on the effectiveness of detrending methods," *Acta Physica Polonica A*, vol. 127, no. 3-A, pp. 59–65, 2015.

[62] I. Simonsen, "Measuring anti-correlations in the Nordic electricity spot market by wavelets," *Physica A: Statistical Mechanics and its Applications*, vol. 322, pp. 597–606, 2003.