Modified detrended fluctuation analysis based on empirical mode decomposition for the characterization of anti-persistent processes

Xi-Yuan Qian\textsuperscript{a,b}, Gao-Feng Gu\textsuperscript{b,c}, Wei-Xing Zhou\textsuperscript{a,b,c,*}

\textsuperscript{a} School of Science, East China University of Science and Technology, Shanghai 200237, China
\textsuperscript{b} Research Center for Econophysics, East China University of Science and Technology, Shanghai 200237, China
\textsuperscript{c} School of Business, East China University of Science and Technology, Shanghai 200237, China

\textbf{A B S T R A C T}

Detrended fluctuation analysis (DFA) is a simple but very efficient method for investigating the power-law long-term correlations of non-stationary time series, in which a detrending step is necessary to obtain the local fluctuations at different timescales. We propose to determine the local trends through empirical mode decomposition (EMD) and perform the detrending operation by removing the EMD-based local trends, which gives an EMD-based DFA method. Similarly, we also propose a modified multifractal DFA algorithm, called an EMD-based MFDFA. The performance of the EMD-based DFA and MFDFA methods is assessed with extensive numerical experiments based on fractional Brownian motion and multiplicative cascading process. We find that the EMD-based DFA method performs better than the classic DFA method in the determination of the Hurst index when the time series is strongly anticorrelated and the EMD-based MFDFA method outperforms the traditional MFDFA method when the moment order \( q \) of the detrended fluctuations is positive. We apply the EMD-based MFDFA to the 1 min data of Shanghai Stock Exchange Composite index, and the presence of multifractality is confirmed. We also analyze the daily Austrian electricity prices and confirm its anti-persistence.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

The dynamics of an evolving complex system can be usually recorded as time series, whose temporal correlation structure embeds much information about the interactions among the microscopic constituents of the system. There are a wealth of approaches proposed to determine the correlation strength of a time series \cite{1,2}. In recent years, the detrended fluctuation analysis (DFA) has become one of the most extensively adopted methods. The idea of DFA was originally invented to investigate the long-range dependence in coding and noncoding DNA nucleotide sequences \cite{3,4}, and the presence of trends and non-stationarity is found to have significant impact on the analysis \cite{5–7}. An extension of the DFA method can be used to unveil the multifractal nature hidden in time series, termed the multifractal DFA (MFDFA) \cite{8}. The (MF)DFA method can also be generalized to investigate higher-dimensional fractal and multifractal measures \cite{9}.

There are also other applications of the DFA method. One example is the investigation of lagged correlations of nonstationary signals, in which the so-called lagged DFA can determine the largest correlation uncovering the existence of underlying delays in the evolution of real time series \cite{10}. On the other hand, there are many situations where several...
variables are simultaneously recorded that exhibit long-range dependence or multifractal nature. A detrended cross-correlation analysis (DXA) was proposed to investigate the long-range cross-correlations between two nonstationary time series, which is a generalization of the DFA method [11–15]. These methods can be easily generalized to study the power-law cross-correlations in higher-dimensional fractal or multifractal signals.

A common step of all the aforementioned methods is to remove the local trends at different timescales. There is an extensive study comparing the performance of several detrending methods in the DFA methods [16]. Numerical experiments show that the classic DFA algorithm outperforms the detrending moving average algorithm [17–19] for short time series with weak trends and Fourier-based DFA [20], whereas the DMA method works slightly better than DFA for long time series with weak trends. However, for a real time series, it is usually not known if there is a trend and, if any, what the functional form of this trend is. A conventional strategy is to assume that the local trends are in the form of polynomial, which works quite well. Recently, a model-free timescale-adaptive detrending approach has been proposed, which is based on the empirical mode decomposition (EMD) approach [21]. For a given time series, EMD is designed to decompose it into a limited number of intrinsic mode functions (IMFs) and a residue component. The EMD-based trend can be considered as the combination of the monotonic residue and IMFs which are significantly distinguishable from the pure white noise, and the timescale of an EMD-based trend is the averaged timescale of the decomposed IMFs. In this work, we employ the EMD as a detrending tool to modify the DFA algorithm, as well as the MFDFA algorithm.

There are several relevant studies aiming at applying the EMD to study the correlations in time series. Numerical experiments based on fractional Gaussian noise with Hurst index $H_0$ show that the EMD acts as a dyadic filter bank which is able to extract the Hurst index $H$ in good agreement with the "true value" $H_0$ when $H_0 > 0.5$ but shows a clear deviation from $H_0$ when $H_0 < 0.5$ [22,23]. A close scrutiny unveils that the extracted $H$ is systematically greater than $H_0$[22,23]. An alternative method applies the EMD-based Hilbert spectral analysis to determined the Hurst index and numerical experiments using fractional Brownian motions show that the extracted Hurst indexes are consistent with the preset indexes for $H_0 \in (0, 1)$ [24]. This method can also be used to calculate the multifractal scaling exponents [24]. There are also studies combining the EMD and the DFA method. One idea is to perform EMD on the original time series to remove the trend or seasonality and a DFA is followed, which was applied to study the correlation properties of long daily ozone records [25]. A slight different method is to conduct the EMD detrending process and the DFA directly on the original time series and a two-parameter scale of randomness for the DFA is proposed to replace the DFA scaling exponent, which was applied to identify characteristics and complexity of heartbeat interval caused by the effects of aging and of illness [26]. Indeed, there is no power-law dependence of the detrended fluctuation function on the timescale.

As we will show below, our EMD-based DFA and MFDFA algorithms are different from the previous efforts in the combination of EMD and DFA. The performance of the EMD-based DFA and MFDFA methods is assessed with numerical experiments based on fractional Brownian motion and multiplicative cascading process. We find that the EMD-based DFA method performs better than the classic DFA method in the determination of the Hurst index when the time series is strongly anticorrelated, and the EMD-based MFDFA method outperforms the traditional MFDFA method when the moment order $q$ of the detrended fluctuations is positive.

The paper is organized as follows. In Section 2, we review the algorithms of DFA, MFDFA and EMD and propose the EMD-based DFA and MFDFA algorithms. In Section 3, we test the performance of the EMD-based DFA and MFDFA algorithms with extensive numerical experiments. The EMD-based DFA and MFDFA methods are applied in Section 4 to investigate the multifractal nature of the return time series of the Shanghai Stock Exchange Composite index and the anti-persistence of electricity prices. Section 5 gives a brief summary.

2. The EMD-based DFA and MFDFA algorithms

2.1. The DFA algorithm

The original DFA algorithm contains the following five steps [3,4].

Step 1. Consider a time series $x(t)$, $t = 1, 2, \ldots, N$. First construct the cumulative sum

$$u(t) = \sum_{i=1}^{t} x(i), \quad t = 1, 2, \ldots, N.$$  \hfill (1)

Step 2. The new series $u(t)$ is partitioned into $N_s$ disjoint segments of the same size $s$, where $N_s = \lfloor N/s \rfloor$. Each segment can be denoted by $u_s$, such that $u_s(i) = u(i + l) - \tilde{u}_s(i)$ for $1 \leq i \leq s$, where $l = (v - 1)s$.

Step 3. In each segment $u_s$, we determine the local trend $\tilde{u}_s$, with the method of polynomial fitting. When a polynomial of order $\ell$ is adopted in this step, the DFA method is called DFA-$\ell$ (DFA-1 if $\ell = 1$, DFA-2 if $\ell = 2$, DFA-3 if $\ell = 3$, and so on). We can then obtain the residual sequence

$$e_s(i) = u_s(i) - \tilde{u}_s(i), \quad 1 \leq i \leq s.$$  \hfill (2)
Step 4. The detrended fluctuation function $F(v, s)$ of the segment $u_i$ is defined as the root of the mean squares of the sample residuals $\epsilon_v(i)$

$$[F(v, s)]^2 = \frac{1}{s} \sum_{i=1}^{s} [\epsilon_v(i)]^2. \quad (3)$$

The overall detrended fluctuation is calculated by averaging over all the segments, that is,

$$[F(s)]^2 = \frac{1}{N_s} \sum_{i=1}^{N_s} [F(v, s)]^2. \quad (4)$$

Step 5. Varying $s$, we can determine the power-law relation between the detrended fluctuation function $F(s)$ and the timescale $s$,

$$F(s) \sim s^H, \quad (5)$$

where $H$ is the DFA scaling exponent. In many cases including the fractional Brownian motions, the DFA scaling exponent $H$ is identical to the Hurst index $\eta$, which is related to the power spectrum exponent $\eta$ by $\eta = 2H - 1$ [27, 28] and thus to the autocorrelation exponent $\gamma$ by $\gamma = 2 - 2H$.

### 2.2. The MFDFA algorithm

The MFDFA method has the same first three steps as the DFA method, and we need only to revise the last two steps [8].

Step 4. The $q$th order overall detrended fluctuation is calculated as follows,

$$F_q(s) = \left\{ \frac{1}{N_s} \sum_{i=1}^{N_s} [F(v, s)]^q \right\}^{1/q}, \quad (6)$$

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

$$F_0(s) = \exp \left\{ \frac{1}{N_s} \sum_{i=1}^{N_s} \ln[F(v, s)] \right\}, \quad (7)$$

according to L'Hôpital's rule.

Step 5. Varying the value of $s$, we can determine the power-law dependence of the detrended fluctuation function $F_q(s)$ on the size scale $s$, which reads

$$F_q(s) \sim s^{h(q)}, \quad (8)$$

where $h(q)$ is the generalized Hurst index.

It is obvious that the DFA is a special case of the MFDFA when $q = 2$. In the standard multifractal formalism based on partition function, the multifractal nature is characterized by a spectrum of scaling exponents $\tau(q)$, which is a nonlinear function of $q$ [29]. For each $q$, we can obtain the corresponding traditional $\tau(q)$ function through

$$\tau(q) = qh(q) - D_f, \quad (9)$$

where $D_f$ is the fractal dimension of the geometric support of the multifractal measure.

### 2.3. The EMD algorithm

Empirical mode decomposition is an innovative data processing algorithm for nonlinear and non-stationary time series [21]. It decomposes the time series $x(t)$ into a finite number of intrinsic mode functions, which satisfy the following two conditions: (1) in the whole set of data, the numbers of local extrema and the numbers of zero crossings must be equal or differ by 1 at most; and (2) at any time point, the mean value of the “upper envelope” (defined by the local maxima) and the “lower envelope” (defined by the local minima) must be zero.

The decomposing process is called a sifting process, which can be described with the following six steps [21]: (1) Identify all extrema of $x(t)$; (2) Interpolate the local maxima to form an upper envelope $U(x)$; (3) Interpolate the local minima to form a lower envelope $L(x)$; (4) Calculate the mean envelope:

$$\mu(t) = \frac{[U(x) + L(x)]}{2}; \quad (10)$$

(5) Extract the mean from the signal

$$g(t) = x(t) - \mu(t); \quad (11)$$
Fig. 1. (a) Log–log plots of the detrended fluctuation function $F(s)$ with respect to the timescale $s$ for two randomly selected FBM time series with $H_0 = 0.2$ and $H_0 = 0.8$ using the EMD-based DFA algorithm. The solid lines are the least-squares power-law fits to the data. (b) Assessing the performance of the EMD-based DFA method through extensive numerical experiments with fractional Brownian motions. A comparison of the estimated Hurst index $H$ with the true value $H_0$ is illustrated. The error bar is determined by the standard deviation of the 100 estimated $H$ values for each $H_0$.

and (6) Check whether $g(t)$ satisfies the IMF conditions. If YES, $g(t)$ is an IMF, stop sifting; If NO, let $x(t) = g(t)$ and keep sifting. Finally, we obtain

$$r_n(t) = x(t) - \sum_{i=1}^{n} g_i(t),$$

where $r_n$ is a residue representing the trend of the time series.

2.4. The EMD-based DFA and MFDFA algorithms

We can now embed the EMD algorithm into the DFA and MFDFA to modify the third step of the algorithms, while keeping all other steps unchanged.

Step 3. For each segment $u_i$, we obtain the EMD-based local trend $\bar{u}_i = r_n(i)$ with the sifting process. We can then obtain the residuals

$$e_i(i) = u_i(i) - r_n(i), \quad 1 \leq i \leq s.$$  

Note that, the trend $r_n(i)$ should be determined for each segment separately at each timescale.

This gives the EMD-based DFA and MF DFA methods. Recall that the authors of [25] remove the trend from the time series according to the EMD method and perform DFA on the resultant series. Alternatively, the modified DFA algorithm in [26] compares the detrended fluctuations between the stationary time series and its shuffled surrogate to defined two parameters of complexity. One can see that these methods differ essentially from the ones proposed in the previous works [25,26].

3. Validating the methods through numerical experiments

3.1. EMD-based DFA of fractional Brownian motions

We test the EMD-based DFA with synthetic fractional Brownian motions (FBMs). In this paper, we use the free MATLAB software FracLab 2.03 developed by INRIA to synthesize fractional Brownian motions with Hurst index $H_0$. In our test, we investigate fractional Brownian motions with different Hurst indices $H$ ranging from 0.05 to 0.95 with an increment of 0.05. The size of each time series is $2^{16} = 65536$. For each $H_0$, we generate 100 FBM time series. Each time series is analyzed by the EMD-based DFA algorithm.

In Fig. 1(a), we show the log–log plot of the detrended fluctuation $F(s)$ as a function of the timescale $s$ for two randomly selected synthetic fractional Brownian motions with $H_0 = 0.2$ (strongly anticorrelated) and $H_0 = 0.8$ (strongly correlated), respectively. There is no doubt that the power-law scaling between $F(s)$ and $s$ is very evident and sound, and the scaling range spans more than two orders of magnitude. The estimates of the Hurst indices are $H = 0.197$ and $H = 0.796$ for the anticorrelated and correlated time series, which are very close to the corresponding $H_0$ values. The EMD-based DFA algorithm is able to well capture the self-similar (or self-affine) nature of the fractional Brownian motions and results in precise estimation of the Hurst index.

We confirm that there is also a power-law dependence of $F(s)$ on $s$ for other synthetic FBMs with different Hurst index $H_0$. For each $H_0$, we determine the Hurst index $H$ for each FBM time series as done in Fig. 1(a) and obtain 100 $H$ values. The mean of the 100 $H$ values is calculated for each $H_0$. The resultant mean Hurst indices $H$ are plotted against $H_0$ in Fig. 1(b). We can see that the estimated Hurst indices $H$ are very close to the preset values $H_0$. The deviation of the estimated Hurst index $H$ from $H_0$ becomes larger for larger values of $H_0$. When comparing with the EMD variance method which fails to give
Assessing the performance of the EMD-based DFA method with reference to the original DFA methods using numerical experiments on fractional Brownian motions. The relative ratio $H/H_0$ of the estimated Hurst index $H$ over the true value $H_0$ is plotted as a function of $H_0$ for the EMD-based DFA and the original DFA algorithms.

Since in practice empirical time series might be shorter, we perform the same analysis for different time series sizes from $2^{10}$ to $2^{15}$. The results are quantitatively very similar to Fig. 1(b), except that the error bar decreases with increasing time series size.

We now turn to assess the performance of the EMD-based MFDFA algorithm with synthetic multifractal signals. There are many methods proposed to generate multifractal signals, such as the multiplicative cascading method \cite{30–32}, the fractionally integrated singular cascade method \cite{33–35}, the random $W$ cascades method \cite{36,35}, and so on. The multiplicative cascading process is widely used to model multifractal measures in many complex systems. As the simplest one, the $p$ model was originally invented to simulate the energy-dissipation field in turbulent flows \cite{31}. In this work, we adopt the textbook $p$ model to generate multifractal signals.

We start from a line and partition it into two segments of the same length and assign two given proportions of measure $p_1 = 0.3$ and $p_2 = 1 - p_1$ to them. Then each segment is divided into two smaller segments and the measure is redistributed in the same multiplicative way. This procedure is repeated for 16 times and at last we generate a multifractal signal of size $2^{16} = 65536$. If the multiplicative cascade process goes to infinity, the mass exponent function has an analytic expression as follows

$$\tau(q) = -\ln(p_1^q + p_2^q)/\ln 2.$$  

The empirical mass exponent function of the constructed multifractal signal should be well approximated by Eq. (14). We perform the EMD-based MFDFA on the binomial measure and determine the empirical mass exponent function $\tau(q)$ and the singularity spectrum $f(\alpha)$, which are illustrated in Fig. 3. We also draw the theoretical lines in Fig. 3 for comparison. It is clear that the two curves overlap with each other. In addition, we also perform the MFDFA on the same multifractal signal and obtain the empirical $\tau(q)$ and $f(\alpha)$ functions, which are plotted in Fig. 3 as well. In general, the two approaches have comparable performance. Both approaches can provide nice results for negative $q$’s and deviate for positive $q$’s. When $q \geq 2$, the EMD-based MFDFA overestimates $\tau(q)$, while the classical MFDFA underestimates $\tau(q)$, which is also observed
4. Applications

In this section, we apply the EMD-based MF DFA method to study the multifractal nature of the high-frequency return time series of the Shanghai Stock Exchange Composite (SSEC) index. The multifractal properties of financial returns have been investigated extensively [37,38]. Concerning the Chinese stock market, there are also numerous multifractal analyses based on the multiplier method [39], the MF DFA method [40–42], and the partition function approach [43,40,44–47]. The presence of multifractality in the Chinese stock market is well documented.

We have performed the EMD-based MF DFA on the 1 min high-frequency data of the SSEC index from 4 January 2000 to 18 April 2008. There are 240 min in the double continuous auction on each trading days [48,49], and the size of the data is 471202. For comparison, we have also conducted the classical MF DFA with the polynomial order \( \ell = 1, 2, 3 \). The resulting \( \tau(q) \) functions for \(-5 \leq q \leq 5\) are illustrated in Fig. 4(a). It is evident that the four algorithms give consistent results since the four \( \tau(q) \) curves overlap. There are two intriguing characteristic points in Fig. 4(a). When \( q = 2 \), \( \tau(q) = 0 \) for all the four curves. It follows that the Hurst index of the SSEC returns is \( H = h(2) = [\tau(2) + 1] / 2 = 0.5 \), which is in agreement with the well-known fact that stock returns are uncorrelated. When \( q = 0 \), Fig. 4 gives \( \tau(q) = -1 \), which is in line with Eq. (9). In certain sense, these two points verify the correctness of the four algorithms.

Given that EMD-based DFA seems to perform better for anti-persistent processes, it is interesting to analyze anti-persistent time series. Since electricity prices are known to be anti-persistent [50,51], we chose the daily Austrian electricity prices from 2002 to 2010 for analysis. The weekly seasonality was removed before analysis. The results are shown in Fig. 4(b). The Hurst indices are \( 0.32 \pm 0.01 \) for the EMD-based DFA, \( 0.33 \pm 0.01 \) for DFA-1, \( 0.33 \pm 0.01 \) for DFA-2, and \( 0.32 \pm 0.01 \) for DFA-3.

5. Summary

In summary, we have proposed a modified detrended fluctuation analysis for fractal and multifractal signals based on the empirical mode composition. The polynomial local trend in the classic DFA algorithm is replaced by an EMD-based local trend. The modified (MF) DFA is called the EMD-based (MF) DFA.

The performance of the EMD-based DFA and MF DFA methods is assessed with extensive numerical experiments based on fractional Brownian motion and multiplicative cascading process. For the EMD-based DFA method, we investigated different multifractal structures.
Hurst index $0.05 < H_0 < 0.95$ and generated 100 fractional Brownian motions for each $H_0$. The accuracy of the estimated Hurst indices $H$ obtained from the EMD-based DFA with reference to $H_0$ is compared with those from the classical DFA. We found that the EMD-based DFA performs better than the classical DFA method in the determination of the Hurst index when the time series is strongly anticorrelated, especially when $H_0 < 0.3$, while the classical DFA outperforms when $H_0 > 0.4$. In all cases, the EMD-based DFA is able to determine the Hurst index $H$ with a deviation less than 3% from the true value $H_0$, that is $(H - H_0)/H_0 < 3%$.

For the EMD-based MFDFA method, we constructed a multifractal signal based on the $p$ model. The empirical mass exponent functions $\tau(q)$ of the EMD-based MFDFA and classical MFDFA were compared with the analytical expression. We found that the EMD-based MFDFA outperforms the traditional MFDFA method when the moment order $q$ of the detrended fluctuations is positive. The usefulness of the EMD-based MFDFA in the multifractal analysis is thus validated.

We have applied the EMD-based MFDFA to the 1 min return data of Shanghai Stock Exchange Composite index. The EMD-based MFDFA gives very similar results as the classical MFDFA methods with different detrending polynomials. The presence of multifractality is confirmed. In addition, the EMD-based DFA and the classical DFA methods give similar Hurst indices for anti-persistent Austrian electricity prices.

We conclude that the EM-based DFA and MFDFA methods have comparable performance as the classical DFA and MFDFA methods in the analysis of fractal and multifractal time series. In certain cases, the EMD-based methods can give better results. The only shortcoming of the EMD-based DFA and MFDFA algorithms is that the detrending process based on EMD is more time-consuming.

Acknowledgments

We are grateful to professor Zhaohua Wu for providing the Matlab codes for EMD. We acknowledge financial support from the National Natural Science Foundation of China under grant 11075054 and the Fundamental Research Funds for the Central Universities.

References


