Physica A 390 (2011) 3136–3146

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Stability and lack of memory of the returns of the Hang Seng index

Krzysztof Burnecki *, Janusz Gajda, Grzegorz Sikora

Hugo Steinhaus Center, Institute of Mathematics and Computer Science, Wroclaw University of Technology, Wyspianskiego 27, 50-370 Wroclaw, Poland

ARTICLE INFO

Article history:
Received 2 December 2010
Received in revised form 16 April 2011
Available online 2 May 2011

Keywords:
Hang Seng index
Long memory
FARIMA
Lévy stable distribution
Mean-squared displacement

ABSTRACT

In this paper we show that the logarithmic returns of the Hang Seng index from January 2, 1987 to November 14, 2005 statistically resemble a sequence of independent identically distributed Lévy stable random variables. This is in stark contrast to Xiu and Jin (2007) [39], where long-memory FARIMA processes with Gaussian noise were suggested as well fitted to the data. The lack of memory is checked by using Lo's modified R/S statistic and a new method of estimation of the memory parameter d which applies the notion of empirical mean-squared displacement. In order to test stability of the data we employ several statistical tests based on the empirical distribution function. Finally, we also show that the returns possess no conditional heteroscedasticity property thus excluding the ARCH/GARCH family of processes as possible underlying models.

1. Introduction

The long memory (or long-range dependence or the Joseph effect first used by Mandelbrot and Wallis [1]) is a peculiarity of many sundry phenomena, which concerns events that are arbitrarily distant but still influence each other exceptionally strongly. This effect can be empirically observed, e.g., by a slowly decaying autocovariance function [2,3]. Such phenomena can be modeled by different stationary stochastic processes like, e.g., fractional Brownian noise, fractional Lévy stable noise, and the fractional ARIMA (FARIMA) processes with either light-tailed or heavy-tailed noise [2,4,5]. In recent years, scientists have investigated the long memory property in different fields such as finance, econometrics, network modeling, hydrology, climate studies, astronomy, linguistics or DNA sequencing [2,4,6–10]. Besides in the above single systems, the long memory property has been found on cross-correlations between different systems, ranging from finance to geophysics [11–13].

The property of long-range dependence is strictly related to increments of self-similar processes with stationary increments (SSSI) [14]. Fractional Brownian motion (FBM), which is a Gaussian SSSI process, was introduced a few decades ago, see Ref. [15] and references therein. It appears that the increments of FBM, called fractional Brownian noise, for the Hurst exponent greater than 0.5, have such a slowly decaying autocovariance function that it is not absolutely summable. This behavior serves as a classical definition of the long-range dependence [2]. The study of non-Gaussian SSSI processes was initiated more recently [2,4,5,14]. Ever since the pioneering work by Mandelbrot [16], and Montroll and Scher [17], Lévy stable processes have enjoyed great popularity as flexible modeling tools in economics and natural sciences. The importance of Lévy stable distributions or processes in physics, astronomy and related areas has long been widespread [3,18–28]. Fractional Lévy stable noise, which is an increment process of the fractional Lévy stable motion, for the Hurst parameter greater than $1/\alpha$, where α is the index of stability, serves as an example of an infinite variance process which exhibits long-range dependence. In the infinite variance case there is no standard definition of long-range dependence as the autocovariance function does not exist (is infinite). Therefore, definitions used in the literature incorporate other measures of dependence, like, e.g., codifference, or different functionals like partial sums and maxima [4,29]. We also note that truncated

* Corresponding author. Tel.: +48 713203530; fax: +48 713202654.
E-mail address: krzysztof.burnecki@pwr.wroc.pl (K. Burnecki).

0378-4371/$ – see front matter
doi:10.1016/j.physa.2011.04.025

Lévy flights and processes have become of interest to researchers [30–32]. They are modifications of the Lévy stable motion with either abrupt or smooth truncation to achieve finite second moments. Another modification, called a tempered Lévy motion, takes a different approach, exponentially tempering the probability of large jumps [33,34].

Fractional Brownian and Lévy stable noises are continuous-time stationary processes. The corresponding (in the limit sense) discrete-time processes are fractional autoregressive integrated moving-average (FARIMA) time series with Gaussian and Lévy stable noises. For a certain choice of their parameters they become long-range dependent [2,35]. For example, in the Gaussian case, the standard definition of long memory implies that the memory parameter d of FARIMA time series has to be greater than zero. Stationary discrete-time processes are attractive from the modeling point of view as the time series theory is well developed and it provides different forecasting tools [36,37].

In the literature different methods of assessing long-range dependence and estimating the memory parameter d have been developed [2,4,35]. It is important to realize what are the assumptions and limitations of various tools and what is the exact output of different estimators [38]. This also refers to self-similarity index estimators. For example, a very well-know R/S method, in the general Lévy stable case, does not return H, which is true only in the Gaussian case, but the value $d+1/2$, where $d = H - 1/\alpha$ and α is the index of stability [35,9].

In this paper we show that the procedure presented in Ref. [39] for the financial data is performed incorrectly which makes their conclusions irrelevant. Furthermore, we specify and justify statistically the model underlying the data which has completely different properties than the model constructed in Ref. [39], namely the log-return data are neither dependent nor Gaussian. This paper is structured as follows: in Section 2 we recall basic facts about the most prominent examples of long-range dependent processes, namely fractional Brownian and Lévy stable motions, and FARIMA processes. In Section 3 we study returns of the weekly close prices of the Hang Seng index. We show that the data are stationary and justify statistically that the underlying distribution is Lévy stable with $\alpha < 2$, and definitely not Gaussian which is assumed in Ref. [39]. To this end we employ well-known and not so well-known statistical tests for Gaussian and Lévy stable distributions. Moreover, we also analyze an 'in-between' distribution often used for the financial data modeling, i.e. normal–inverse Gaussian. It appears that this distribution is also no match for the Lévy stable law. Furthermore, we illustrate graphically Gaussian, Lévy stable, and normal–inverse Gaussian fits. In Section 4 we prove that the data exhibit no long memory by applying such tools like Lo's modified R/S statistic and a new method of estimating the memory parameter d applying the notion of sample mean-squared displacement. As a consequence, the data cannot be modeled by FARIMA processes, which was suggested in Ref. [39]. This finding coincides with different studies on long-range dependence in the literature where strong dependence has been observed almost only in the stock indices volatility, e.g. in absolute values of returns [10,40]. In Section 5 we show that the analyzed data display no conditional heteroscedasticity, thus excluding GARCH processes as appropriate models. In Section 6 a summary of the results is presented.

2. Fractional motions and FARIMA processes

Fractional Brownian motion (FBM) is a generalization of the classical Brownian motion (BM). Most of its statistical properties are characterized by the Hurst exponent $0 < H < 1$. For further properties of FBM and its applications to physics see Refs. [7,41–44]. For any $0 < H < 1$, FBM of index H (Hurst exponent) is a mean-zero Gaussian process $B_H(t)$ with the following integral representation [7,45]:

$$B_H(t) = \int_{-\infty}^{\infty} \left\{ (t-u)_+^{H-\frac{1}{2}} - (-u)_+^{H-\frac{1}{2}} \right\} \mathrm{d}B(u), \tag{1}$$

where $B(t)$ is a standard Brownian motion and $(x)_+ = \max(x, 0)$.

FBM is H-self-similar, namely for every $c > 0$ we have $B_H(ct) = c^H B_H(t)$ in distribution, and has stationary increments. It is the only Gaussian process satisfying these properties. For $H > 1/2$, the increments of the process are positively correlated and exhibit long-range dependence (long memory, persistence), whereas for $H < 1/2$, the increments of the process are negatively correlated and exhibit short-range dependence (short memory, antipersistence) [7].

It can be generalized to a fractional Lévy stable motion (FLSM) [6,7,29,41,46]:

$$Z_H^{\alpha}(t) = \int_{-\infty}^{\infty} \left\{ (t-u)_+^{d} - (-u)_+^{d} \right\} \mathrm{d}L_\alpha(u), \tag{2}$$

where $L_\alpha(t)$ is a Lévy α-stable motion (LSM), $0 < \alpha \le 2$, $0 < H < 1$, and $d = H - 1/\alpha$. The process is α-stable (for $\alpha = 2$ it becomes an FBM), H-self-similar and has stationary increments. Analogously to the FBM case, we say the increments of the process exhibit positive (long-range) dependence if $d > 0$ ($H > 1/\alpha$), and negative dependence when $d < 0$ ($H < 1/\alpha$) [29]. This is due to the behavior of the integrand in (2) [35]. When $d < 0$ the integrand has singularities at $u = 0$ and $u = t$. These singularities magnify the fluctuations of LSM, and cause large spikes in the paths of the FLSM process. These paths become very irregular. Their dependence structure resembles that of a negatively correlated process, namely negative (positive) values are likely to follow positive (negative) values of the increments of the process. Thus, we shall refer to this case as to the negative dependence scenario. In the case when $d > 0$, the integrand is bounded and positive, for all $t > 0$. Thus the jumps in the paths of FLSM, due to the fluctuations of LSM, are not as magnified as in the case $d < 0$. In this case, especially for large values of H, the integrand decays slowly. This implies that the past fluctuations of the process $L_\alpha(t)$

influence significantly the present values of the process $Z_H^\alpha(t)$. This resembles the behavior of positively correlated process, namely positive (negative) values are likely to follow positive (negative) values of the increments of the process. This case is referred to as the long-range dependence scenario. Therefore, as in the Gaussian case, the parameter d controls the sign of dependence.

The classic and abundant class of processes with long memory property is a class of models known as fractionally integrated processes. They are attractive because their long-range dependence structure can be modeled by the memory parameter d. One of the most popular examples of such process is the FARIMA (also known as ARFIMA) model, originally introduced by Granger and Joyeux [47] and Hosking [48]. The FARIMA(p, d, q) time series is defined as the solution of equations

$$\Phi(\nabla)\Delta^d X_n = \Theta(\nabla)\epsilon_n, \quad n \in \mathbb{Z},$$

where ∇ is the backward shift operator, i.e. $\nabla(X_n) = X_{n-1}$, and Δ is the difference operator, i.e. $\Delta X_n = X_n - X_{n-1}$. The polynomials Φ and Θ have the form

$$\Phi(\nabla) = 1 - a_1\nabla - a_2\nabla^2 - \cdots - a_p\nabla^p,$$
$$\Theta(\nabla) = 1 - b_1\nabla - b_2\nabla^2 - \cdots - b_q\nabla^q. \tag{3}$$

They correspond to autoregressive (AR) and moving average (MA) parts, respectively. ϵ_j's, which stand for the noise, are independent and identically distributed (i.i.d.) random variables. They may be either Gaussian, non-Gaussian with finite variance or they may have infinite variance (for example symmetric and skewed non-Gaussian Lévy stable distributions).

The name fractional corresponds to the memory parameter d (or fractional differencing exponent). The fractional difference operator Δ^d has the binomial expansion

$$\Delta^d = (1-B)^d = \sum_{j=0}^{\infty} \pi_j(d)B^j, \tag{4}$$

where the coefficients $\pi_j(d)$'s look like in the Taylor expansion of the function $f(z) = (1-z)^d$ and can be represented by the gamma function Γ:

$$\pi_j(d) = \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)}, \quad j = 0, 1, \ldots. \tag{5}$$

The linear solution of FARIMA processes takes the form

$$X_n = \sum_{j=1}^{\infty} c_j(d)\epsilon_{n-j}, \tag{6}$$

where $c_j(d)$'s are defined by the equation

$$\frac{\Theta(z)(1-z)^{-d}}{\Phi(z)} = \sum_{j=0}^{\infty} c_j(d)z^j, \quad |z| < 1, \tag{7}$$

for details see Ref. [37]. In the case of the Lévy stable noise for $0 < \alpha < 2$ the process X_n is a well-defined stationary moving average provided that $-\infty < d < 1 - 1/\alpha$. For more information about FARIMA processes with Lévy stable noise we refer the reader to Ref. [37].

Finally, observe that the FARIMA time series is asymptotically self-similar with the Hurst parameter $H = d+1/\alpha$, namely, as $N \to \infty$,

$$N^{-H}\sum_{n=1}^{[Nt]} X_n \xrightarrow{D} CZ_H^\alpha(t), \tag{8}$$

where $\xrightarrow{D}$ means the convergence in the distribution and C is a constant [35]. For the relation between FARIMA and FLSM processes see also Ref. [49].

3. Distribution of the returns

The analyzed time series consists of the weekly close price data of the Hong Kong Hang Seng index since January 2, 1987, until November 14, 2005 presented in the left panel of Fig. 1. This gives 985 observations, denoted by $\{Y_n : n = 1, 2, \ldots, 985\}$, which are freely available on the Internet (from http://finance.yahoo.com). The same set of data was studied in Ref. [39]. Since the data clearly have non-stationary structure and fitting the distribution requires stationarity, our investigations are conducted on the logarithmic returns of the Hang Seng index, see the right panel in Fig. 1. The stationarity of the returns is studied in more detail in Section 5 and also in Ref. [39], when first-order differenced logarithms of the weekly close prices are considered.

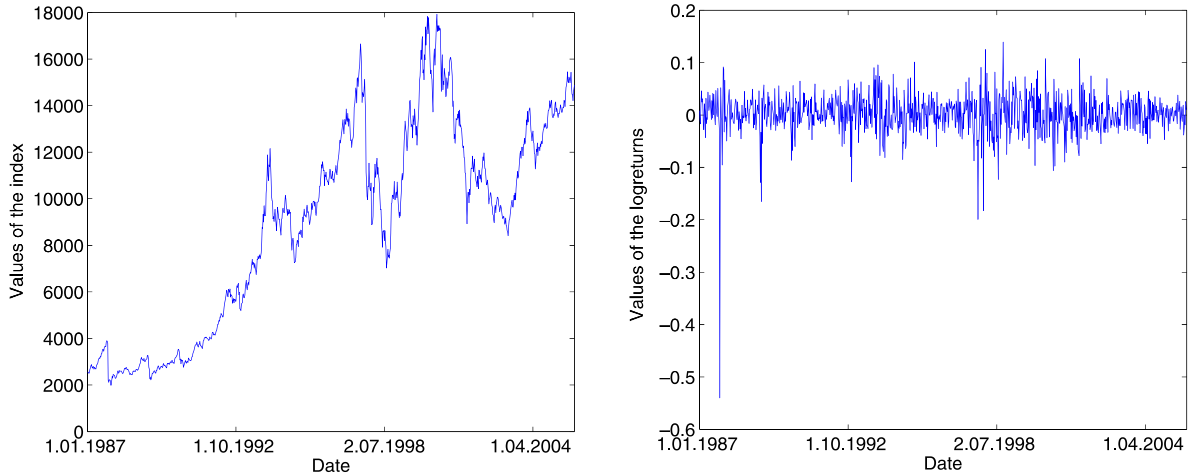

Fig. 1. Hang Seng index weekly prices (left panel) and their logarithmic returns (right panel).

All further analysis is based on logarithmic returns $\{X_n = \ln(Y_{n+1}/Y_n) : n = 1, 2, \ldots, N = 984\}$. We consider here three possible probability laws underlying the series: Gaussian, Lévy stable and normal-inverse Gaussian (NIG). The NIG distribution, introduced in Ref. [50], was chosen as it is able to model symmetric and asymmetric distributions with possibly long tails in both directions. Its tail behavior is often classified as "semi-heavy", i.e. the tails are lighter than those of Lévy stable laws, but much heavier than the Gaussian. Moreover, empirical experience shows an excellent fit of the NIG law to different financial data [51].

To check if the data come from a population with a specific distribution, we performed various statistical tests based on empirical distribution function. The tests are related to different statistics which measure the vertical distance between the empirical $F_N(x)$ and theoretical $F(x)$ cumulative distribution function (CDF). The values of the statistics calculated for different distributions serve as a first insight into the goodness of fit. Statistical tests provide a mechanism for making quantitative decisions about a distribution. The intent is to determine whether there is enough evidence to reject a conjecture or hypothesis about the distribution. We calculated so-called p-values which are probabilities of obtaining a test statistic at least as extreme as the one that was actually observed for the data, assuming that the hypothesis is true. The lower the p-value, the less likely the result is if the hypothesis is true. One often rejects the hypothesis about the distribution when the p-value is less than 0.05 or 0.01 [52].

We used two classes of such measures, namely the Kolmogorov–Smirnov and Cramér–von Mises [52–54]. The Kolmogorov–Smirnov statistic is given by

$$D = \sup_x |F_N(x) - F(x)|. \tag{9}$$

This statistic can be written as a maximum of two non-negative supremum statistics:

$$D^+ = \sup_x\{F_N(x) - F(x)\} \quad \text{and} \quad D^- = \sup_x\{F(x) - F_N(x)\}. \tag{10}$$

Belonging to the same class is the Kuiper statistic given by

$$V = D^+ + D^-. \tag{11}$$

The second class forms the Cramér–von Mises family

$$CM = N \int_{-\infty}^{\infty} (F_N(x) - F(x))^2 \psi(x) \mathrm{d}F(x). \tag{12}$$

$\psi(x)$ is a special weight function, for $\psi(x) = 1$ we obtain W^2 Cramér–von Mises statistic and for $\psi(x) = [F(x)(1 - F(x))]^{-1}$ we arrive at A^2 Anderson and Darling one.

The values of the statistics for the three considered distributions are depicted in Table 1. Comparing values of the statistics for the Gaussian, Lévy stable and NIG distributions we clearly see that values for the Gaussian distribution are the largest, and for the stable case they are much lower than in the NIG case. The maximum likelihood method was used to estimate parameters of the Gaussian distribution. We obtained that $\mu_G = 0.0018$ and $\sigma_G = 0.0385$. To estimate the parameters of the Lévy stable distribution, we employed a regression-type estimator, which is regarded as both accurate and fast [51,55]. The estimated parameters are: $\alpha_S = 1.7896$, $\sigma_S = 0.0207$, $\beta_S = -0.1750$, and $\mu_S = 0.0026$. We note that the negative skewness parameter β_S suggests the data to be left-skewed (the skewness, negative or positive, is often observed for the return data). In order to estimate NIG parameters we used the maximum likelihood method, see Ref. [51]. The estimated parameters are: $\alpha_{NIG} = 30.1870$, $\beta_{NIG} = -5.1916$, $\delta_{NIG} = 0.0364$, and $\mu_{NIG} = 0.0081$.

Table 1
Values of the test statistics and corresponding p-values (in parentheses) under the assumption of the Gaussian, Lévy stable and NIG laws.

D	V	W^2	A^2
Gaussian			
0.0745 (<0.001)	0.1425 (<0.001)	1.8979 (<0.001)	12.1704 (<0.001)
Lévy stable			
0.0173 (0.5120)	0.0332 (0.3180)	0.0468 (0.2880)	0.3266 (0.2980)
NIG			
0.0253 (0.2750)	0.0444 (0.2060)	0.1006 (0.2760)	0.6489 (0.2620)

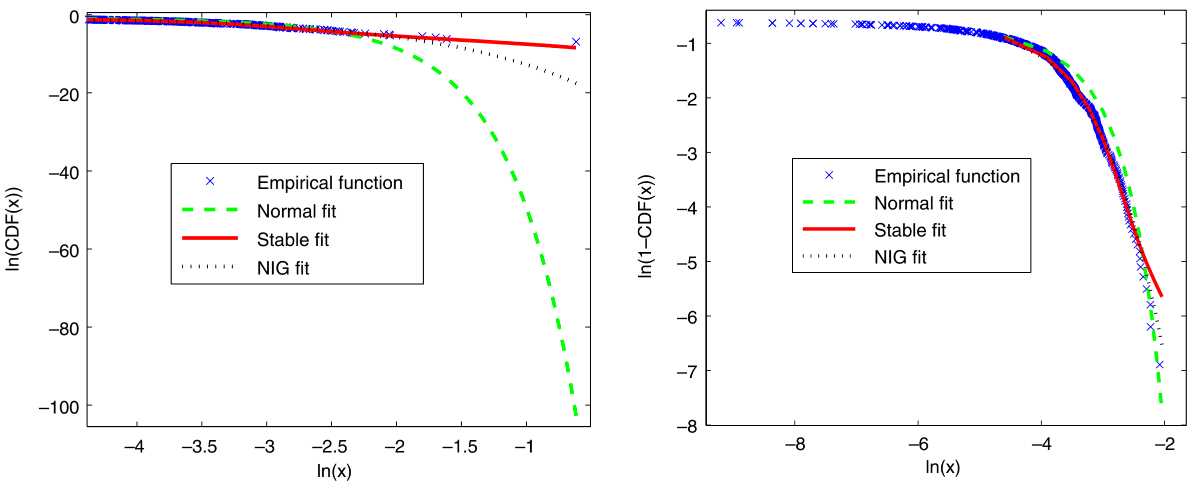

Fig. 2. Left (left panel) and right (right panel) tails of the empirical (blue crosses), fitted Gaussian (green dashed), Lévy stable (red solid line), and normal inverse Gaussian (black dotted) distribution functions on a log–log scale. We can observe that the Lévy stable distribution is clearly superior in the left tail area. In the right tail, the normal distribution is definitely not acceptable in the middle region. The other distributions look pretty much alike.

To calculate p-values, we used a procedure based on Monte Carlo simulations [54]. First, for the given sample size we estimated a vector of parameters $\hat{\theta}$ under the assumption that the sample is distributed according to one of the considered distributions. Next, we calculated t values for different statistics. Next, we generated 1000 samples of the same length as our data, given the vector $\hat{\theta}$. For each sample we estimated the parameters and calculated the statistics. The p-values were calculated as a proportion of times that the test statistic quantity is at least as large as t.

As we can see in Table 1, for the Gaussian distribution, the calculated p-values for all considered tests were lower than 0.001, showing that a Gaussian distribution hypothesis should be definitely rejected for the data. Moreover, p-values for the Lévy stable distribution were quite high and always greater than those calculated under the NIG assumption. This analysis indicates stability of the data. Finally, we illustrated Gaussian, Lévy stable and NIG fits in Fig. 2, where tails of the empirical and fitted distribution functions are presented on a double logarithmic scale. We can see that the Lévy stable distribution is fitted well, especially in the left tail.

4. Long memory of the data

Now, for the logarithmic returns of the price data, we check the hypothesis of the lack of long memory in the Hang Seng time series. To this end we apply three methods of estimation of the memory parameter d, namely the R/S statistic, Lo's modified R/S statistic and a new method based on the notion of mean-squared displacement. We note that the same Hang Seng time series was studied in Ref. [39]. However, this empirical research was mainly based on the analysis of the log-prices not logarithmic returns. Such data obviously have non-stationary structure, hence the discussion on the long memory of the data presented in Ref. [39] cannot be justified.

Our empirical study begins with the R/S method [56,57]. It is the most classical and well-known method for estimation of the self-similarity parameter H. In the case of Gaussian (i.e., finite variance) time series the result of the method is indeed an estimate of H. The relationship with memory parameter d is $H = d + 1/2$. For the Lévy α-stable (infinite variance) time

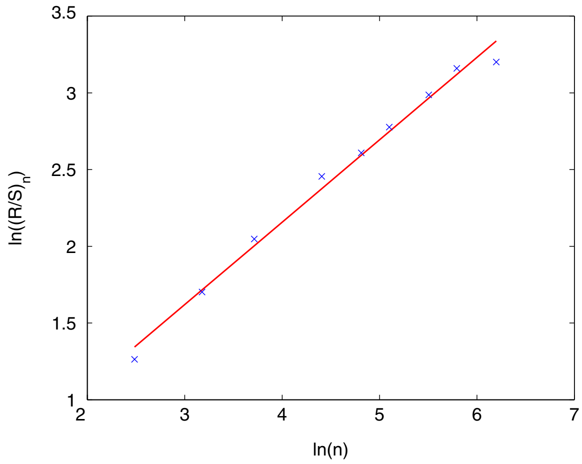

Fig. 3. R/S statistic (blue crosses). The estimated slope of the fitted line (red solid line) equals 0.537, which corresponds to $d = 0.037$.

series the method does not return H, only $d + 1/2$, where $d = H - 1/\alpha$ [58]. Naturally, if $\alpha = 2$, this reduces to H. We begin with dividing a series (of logarithmic returns) of length $N = 984$ into blocks of equal length n. Then for each subseries ($m = 1, 2, \ldots, [N/n]$) the $(R/S)_m$ statistic is computed according to the formula:

$$(R/S)_m = \frac{\max\limits_{1\le k\le n} \sum\limits_{i=1}^{k}(X_i - \overline{X}_m) - \min\limits_{1\le k\le n} \sum\limits_{i=1}^{k}(X_i - \overline{X}_m)}{S_m}, \tag{13}$$

where $\overline{X}_m$ and S_m are sample mean and sample standard deviation of the mth subseries respectively. Finally, the sample mean of the $(R/S)_m$ statistics over blocks is calculated and denoted by $(R/S)_n$. After such a procedure, we plot the $(R/S)_n$ statistic against n on a log–log scale with the fitted least squares line, see Fig. 3.

The slope should be equal to $d + 1/2$. We obtain that $d = 0.037$. In our analysis we used only divisors of N greater than 10, i.e. $n \in \{12, 24, 41, 82, 123, 164, 246, 329, 492\}$ [57].

Next, we consider a Lo's modified R/S statistic (MRS) V_q [59] defined by

$$V_q = \sqrt{N}(R/S_q)_N, \tag{14}$$

which is a version of the formula (13) for $n = N$ with the same numerator and modified denominator

$$S_q = \sqrt{S^2 + 2\sum_{i=1}^{q}\omega_i(q)\hat{\gamma}_i}, \tag{15}$$

where $\hat{\gamma}_i$ are the sample autocovariances and weights $\omega_i(q)$ are given by

$$\omega_i(q) = 1 - \frac{i}{q+1}, \quad q < N. \tag{16}$$

Plotting V_q with respect to q on a log–log scale and fitting the least squares line, leads, for the finite variance case, to an estimate of H, namely the slope of the line equals $1/2 - H$. One may check that, as it was the case for R/S statistic, that the formula for the Lévy α-stable distribution can be generalized to $-d$, where $d = H - 1/\alpha$. In Fig. 4 we depict a plot of the V_q statistic against q on a double logarithmic scale with the fitted least squares line. The slope is equal to $-d = 0.101$, hence $d = -0.101$.

In the case of absence of long memory, Lo showed that for the finite variance case, under the right choice of q,

$$V_q \xrightarrow{D} W_1 = \max_{0\le t\le 1} W_0(t) - \min_{0\le t\le 1} W_0(t), \tag{17}$$

where W_0 is a standard Brownian bridge. Since the distribution of the random variable W_1 is known, Lo proposed testing the null hypothesis

$$\mathcal{H}_0 = \{\text{no long memory, i.e. } d = 0\} \tag{18}$$

with a 95% (asymptotic) acceptance region [0.809, 1.862]. Therefore, Lo's modified R/S statistic can serve as a tool for the validation procedure. We note here that the right choice of q is an open problem and the test statistic V_q has a strong bias toward accepting the null hypothesis [60].

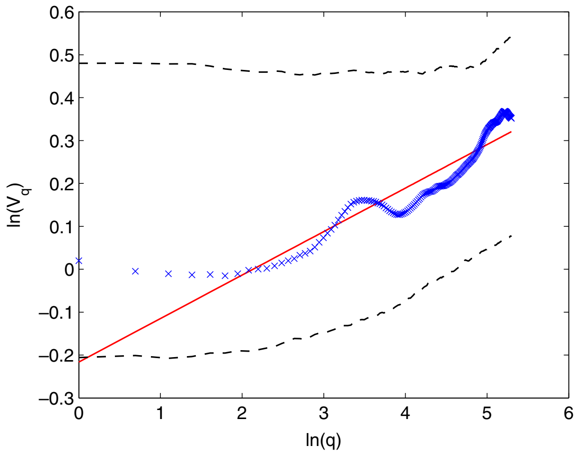

Fig. 4. Modified R/S statistic V_q with respect to q (blue crosses). The slope of the fitted line (red solid line) equals $-d = 0.101$, hence $d = -0.101$. We also plot the 95% acceptance region of $\mathcal{H}_0$ (black dashed lines) obtained via Monte Carlo simulation under the assumption of i.i.d. Lévy stable random variables. The values of the statistic lie within the confidence interval supporting the absence of long memory.

However, for the infinite variance case, there are yet no analytical results for the asymptotics of V_q. Hence, under the assumption the analyzed time series is a sequence of i.i.d. Lévy stable random variables, we calculated the confidence interval applying Monte Carlo simulations with 1000 samples (for the idea of calculating confidence intervals, see, e.g. Ref. [40]). The results are presented in Fig. 4. We can see that the values of the statistic lie within the confidence region for all q's. This justifies the claim of absence of long memory.

In order to estimate the memory parameter d we also used a sample mean-squared displacement (sample MSD) [61]. Let $\{Y_n,\ n = 1, \ldots, N\}$ be a sample of length N. Then, the sample MSD is defined by

$$M_N(\tau) = \frac{1}{N-\tau}\sum_{k=1}^{N-\tau}(Y_{k+\tau} - Y_k)^2. \tag{19}$$

The sample MSD is a time average MSD on a finite sample regarded as a function of difference τ between observations. It is a random variable in contrast to the ensemble average which is deterministic and always infinite for the Lévy stable case with $\alpha < 2$.

If N becomes large and τ small, the sample comes from an FLSM with $\alpha \le 2$ (for $\alpha = 2$ it becomes an FBM), then

$$M_N(\tau) \xrightarrow{D} S\tau^{2d+1}, \tag{20}$$

where $d = H - 1/\alpha$ and S is a Lévy stable random variable which does not depend on τ [61]. Therefore, the memory parameter d characterizes the stochastic process in terms of the speed of transport, and, for $d \ne 0$ we obtain so-called anomalous dynamics, either superdiffusive (for $d > 0$) or subdiffusive (for $d < 0$). In particular, for $d = 0$, so an LSM, we obtain that $M_N(\tau)$ behaves like τ exactly as for a BM.

As a by-product, the sample MSD can serve as a method of estimating d. The method is well defined for the general Lévy stable case and the estimator has a very small variance, which will be discussed in another paper. In order to apply it, first, we calculated the sample MSD for the partial sum process $\{Y_n = \sum_{i=1}^{n} X_i\colon n = 1, 2, \ldots, 984\}$:

$$M_N(\tau) = \frac{1}{984-\tau}\sum_{k=1}^{984-\tau}(Y_{k+\tau} - Y_k)^2\,.$$

Next, applying (20), we fitted linear regression line according to

$$\ln(M_N(\tau)) = \ln(C) + (2d+1)\ln(\tau), \quad \tau = 1, 2, \ldots 10,$$

where C is assumed to be constant. Our analysis yields $2d + 1 = 1.0022$ and therefore $d = 0.0011$. Finally, with the help of Monte Carlo simulations, we calculated the confidence interval for the value of sample MSD under the hypothesis of i.i.d. Lévy stable random variables (the partial sum process is then an LSM). The results are presented in Fig. 5. We can see that the values lie with in the confidence interval supporting the hypothesis of absence of long memory.

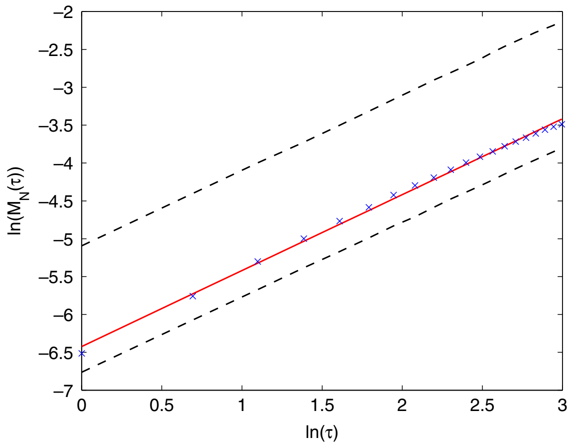

Fig. 5. Sample MSD (blue crosses) with fitted line (red solid line) and estimated 95% confidence interval (black dashed lines) obtained via Monte Carlo simulation under the assumption of i.i.d. Lévy stable random variables. The slope of the line equals 1.0022 leading to $d = 0.0011$. The values of the statistic lie within the confidence interval supporting the absence of long memory.

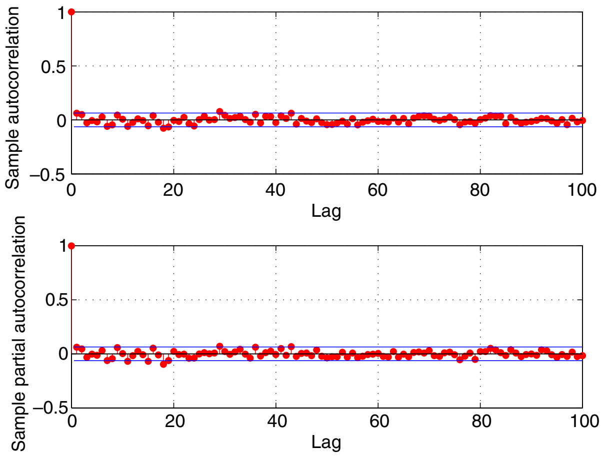

Fig. 6. Autocorrelation (top panel) and partial autocorrelation (bottom panel) functions of the Hang Seng index logarithmic returns. Blue horizontal lines represent 95% confidence intervals calculated under the i.i.d. assumption.

5. Randomness of the data

We now examine the data by plotting the autocorrelation and partial autocorrelation functions. The sample autocorrelation function at lag k is defined as

$$r_k = \frac{\sum_{t=1}^{N-k}(X_t - \overline{X})(X_{t+k} - \overline{X})}{\sum_{t=1}^{N}(X_t - \overline{X})}, \quad \text{for } k = 0, 1, \ldots \tag{21}$$

and stands for a measure of the linear dependence between observations with time lag k [3]. For i.i.d. data $r_k = 0$ for all $k \neq 0$. For moving average processes of order $q(\mathrm{MA}(q))$, the autocorrelation function is zero for lags beyond q, hence the function is also useful for estimating the q parameter.

For assessing the order p of an autoregressive process $(\mathrm{AR}(p))$ we use the partial autocorrelation function, which, at lag k, is defined as a correlation between the predictor errors of values X_n and X_{n+k} represented in terms of $X_{n+1}, \ldots, X_{n+k-1}$. The partial autocorrelation function for AR(p) processes behaves much like the autocorrelation function for MA(q) processes, see Ref. [62] and references therein. For i.i.d. series the function should be equal to zero except for the lag $k = 0$.

Plots of autocorrelation and partial autocorrelation functions for the analyzed data are presented in Fig. 6. We can see that the values fall within the 95% confidence intervals calculated under the i.i.d. assumption, thus we can infer that the data do not exhibit serial correlation. In order to support this claim we perform further analyses. If the data exhibit no serial

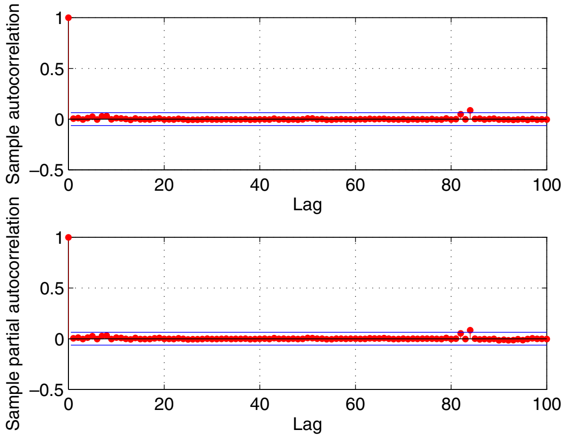

Fig. 7. Autocorrelation (top panel) and partial autocorrelation (bottom panel) functions of the Hang Seng index squared logarithmic returns. Blue horizontal lines represent 95% confidence intervals calculated under the i.i.d. assumption.

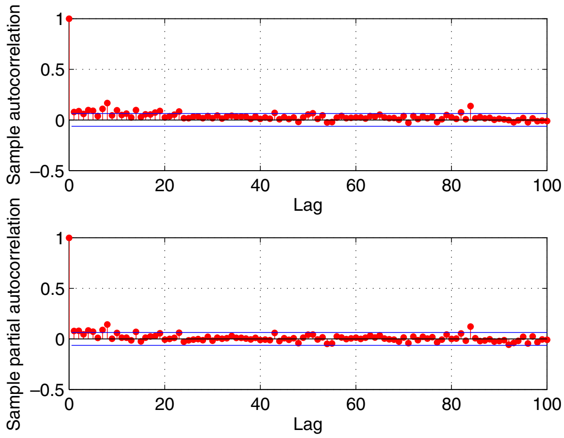

Fig. 8. Autocorrelation (top panel) and partial autocorrelation (bottom panel) functions of the Hang Seng index absolute logarithmic returns. Blue horizontal lines represent 95% confidence intervals calculated under the i.i.d. assumption.

dependence, one should check possible heteroscedasticity and volatility clustering nature of time series. To this end we consider a generalized autoregressive conditional heteroscedasticity (GARCH(p, q)) model [63]:

$$\epsilon_t = \sqrt{h_t}\eta_t, \quad \eta_t \sim N(0, 1), \tag{22}$$

$$h_t = a_0 + \sum_{i=1}^{q} a_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} b_j h_{t-j}. \tag{23}$$

Taking into account that the correlation is only a measure of linear dependence, one can should explore higher order serial dependence. We note that if the data is truly independent, then nonlinear transformations like squaring or taking absolute values should preserve independence. In our case, we can observe in Fig. 7 there is also no serial correlation in squared returns. A similar procedure was performed for absolute values of the returns, which is presented in Fig. 8. To give a better insight into the structure of our series we also performed the ARCH test [64] for the presence of ARCH effects in our series. First, for a given p we estimated the ARCH(p) model and then we computed the Lagrange multiplier statistic $N \times R^2$, where N is the sample size and R^2 is the coefficient of determination. Under the null hypothesis, the asymptotic distribution of the test statistic is χ^2 with p degrees of freedom. We note that the ARCH($p + q$) process is locally equivalent to the GARCH(p, q) process [65]. At the significance level $\alpha = 0.05$, at lags 5, 10, and 50, we obtained the following p-values: 0.98, 0.99, and 0.40 correspondingly. This confirms the claim that the series does not follow a GARCH model.

In addition, we performed the Ljung–Box test which takes into account magnitude of correlation as a group. The Ljung–Box statistic is given by [66]:

$$Q = N(N+2)\left(\frac{\rho_1^2}{N-1} + \frac{\rho_2^2}{N-2} + \cdots + \frac{\rho_K^2}{N-K}\right), \tag{24}$$

where ρ_i is a squared autocorrelation at lag i and K is a number of autocorrelation lags included in the statistic.

The Q statistic is χ^2 distributed as $N \to \infty$. This allows one to test the null hypothesis that there is no serial correlation in the data. We calculated the p-values for $K = 1, \ldots, 20$. The mean of the values was 0.1279 and only four of them were lower than 0.05 (but always larger than 0.01). From these results we can expect that there is no serial correlation in the data.

6. Conclusions

We have analyzed a time series consisting of the weekly close price data of the Hong Kong Hang Seng index since January 2, 1987, until November 14, 2005. The same set of data was studied by Xiu and Jin in Ref. [39]. Their empirical research was mainly based on the analysis of the log-prices. Such data obviously have a non-stationary structure, hence the discussion on the long memory of the data, in particular applying the modified R/S statistic, presented in Ref. [39] cannot be justified. Moreover, the authors neglect non-Gaussianity of data, which has an impact on the interpretation of the results, see Sections 3–4. Fitting the FARIMA model, which is discussed in Ref. [39] can be viable only for stationary data which is definitely not the case for the log-prices. Although in the second part of Ref. [39] the authors perform first-order differencing which leads to logarithmic returns of the index, hence proper stationary data, the discussion about that data is brief and it is only suggested that the time series can be expressed as an ARMA(2, 1) model.

Our investigations have been conducted on the sequence of logarithmic returns of the weekly close price data. The first aim of our analysis was to substantiate the hypothesis of the stability of the data. The estimated stability index $\alpha = 1.7896$. To check the hypothesis of stability, we have implemented the following statistical tests: Kolmogorov–Smirnov, Kuiper, Cramér–von Mises, and Anderson–Darling [52,54]. The calculated p-values for the considered tests were high indicating the goodness of the Lévy stable fit. We have also compared the Lévy stable distribution with normal-inverse Gaussian which is often considered for the return data in the literature and it appears that the Lévy stable fit is superior.

In order to examine the memory of the Hang Seng time series, we have applied different methods of estimation of the memory parameter d. Among estimation techniques used were R/S and MRS statistics and a new method of estimation d based on the mean-squared displacement (MSD). To check the hypothesis of absence of long memory, relevant confidence intervals were calculated with the help of Monte Carlo simulations under the i.i.d. assumption. Finally, we have studied the dependence structure by analysis of autocorrelation functions and performed the ARCH and Ljung–Box tests to exclude possible heteroscedasticity and volatility clustering nature of the time series.

The procedure proposed in this paper can be applied to different real-world time series, when one needs to verify long-range dependence property, check whether the underlying distribution is either light- or heavy-tailed, and to construct and validate a stochastic model which fits the data well.

Acknowledgment

The second and third authors would like to acknowledge a partial support of the Fellowship co-financed by European Union within European Social Fund.

References

[1] B. Mandelbrot, J. Wallis, Water Resour. Res. 4 (1968) 909.
[2] J. Beran, Statistics for Long-Memory Processes, Chapman & Hall, NY, 1994.
[3] R.N. Mantegna, H.E. Stanley, An Introduction to Econophysics: Correlations and Complexity in Finance, Cambridge University Press, 2000.
[4] P. Doukham, G. Oppenheim, M.S. Taqqu (Eds.), Theory and Applications of Long Range Dependence, Birkhüser, Boston, 2003.
[5] E. Eberlein, M. Taqqu (Eds.), Dependence in Probability and Statistics, Birkhäuser, Boston, 1986.
[6] A. Janicki, A. Weron, A Simulation and Chaotic Behaviour of α-Stable Stochastic Processes, Marcel Dekker Inc., New York, 1994.
[7] A. Weron, K. Burnecki, Sz. Mercik, K. Weron, Phys. Rev. E 71 (2005) 016113.
[8] P.A. Varotsos, N.V. Sarlis, E.S. Skordas, H.K. Tanaka, M.S. Lazaridou, Phys. Rev. E 74 (2006) 021123.
[9] K. Burnecki, J. Klafter, M. Magdziarz, A. Weron, Physica A 387 (2008) 1077;
A. Stanislavsky, K. Burnecki, M. Magdziarz, A. Weron, Astrophys. J. 693 (2009) 1877.
[10] S.H. Kang, C. Cheong, S.-M. Yoon, Physica A 389 (2010) 4844.
[11] B. Podobnik, H.E. Stanley, Phys. Rev. Lett. 100 (2008) 084102.
[12] B. Podobnik, D. Horvatic, A.M. Petersen, H.E. Stanley, Proc. Natl. Acad. Sci. USA 106 (2009) 22079.
[13] S. Arianos, A. Carbone, J. Stat. Mech. (2009) P03037.
[14] P. Embrechts, M. Maejima, Internat. J. Modern Phys. B 14 (2000) 1399; Selfsimilar Processes, Princeton University Press, Princeton, 2002.
[15] B. Mandelbrot, J.W. Van Ness, SIAM Rev. 10 (1968) 422;
B. Mandelbrot, Fractals and Scaling in Finance, Springer, Berlin, 1997.
[16] B. Mandelbrot, J. Business 36 (1963) 394.
[17] E.W. Montroll, H. Scher, J. Stat. Phys. 10 (1973) 101.
[18] R. Metzler, J. Klafter, Phys. Rep. 339 (2000) 1.
[19] A.V. Chechkin, J. Klafter, V.Y. Gonchar, R. Metzler, L.V. Tanatarov, Phys. Rev. E 67 (2003) 010102(R).
[20] I. Eliazar, J. Klafter, J. Stat. Phys. 111 (2003) 739.
[21] A.V. Chechkin, J. Klafter, I.M. Sokolov, Europhys. Lett. 63 (2003) 326.
[22] B. Dybiec, E. Gudowska-Nowak, Phys. Rev. E 69 (2004) 016105; Fluct. Noise Lett. 4 (2004) L273.
[23] B. Dybiec, E. Gudowska-Nowak, P. Hänggi, Phys. Rev. E 73 (2006) 046104.
[24] A.V. Chechkin, J. Klafter, V.Yu. Gonchar, R. Metzler, L.V. Tanatarov, Chem. Phys. 284 (2002) 233.
[25] I.M. Sokolov, J. Klafter, A. Blumen, Phys. Today 55 (2002) 48.
[26] I.M. Sokolov, Phys. Rev. E 63 (2000) 011104; 63 (2001) 056111.
[27] A.A. Stanislavsky, Phys. Rev. E 61 (2000) 4752; Theoret. and Math. Phys. 138 (2004) 418.
[28] A. Piryatinska, A.I. Saichev, W.A. Woyczynski, Physica A 349 (2005) 375.

[29] G. Samorodnitsky, M.S. Taqqu, Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance, Chapman & Hall, London, 1994.
[30] R.N. Mantegna, H.E. Stanley, Phys. Rev. Lett. 73 (1994) 2946.
[31] I. Koponen, Phys. Rev. E 52 (1995) 1197.
[32] B. Podobnik, P.Ch. Ivanov, Y. Lee, H.E. Stanley, Europhys. Lett. 52 (2000) 491.
[33] J. Rosinski, Stoch. Proc. Appl. 117 (2007) 677.
[34] B. Baeumera, M.M. Meerschaert, J. Comput. Appl. Math. 233 (2010) 2438.
[35] S. Stoev, M.S. Taqqu, Fractals 12 (2004) 95.
[36] P.J. Brockwell, R.A. Davis, Introduction to Time Series and Forecasting, Springer, New York, 2002.
[37] P.S. Kokoszka, M.S. Taqqu, Stochastic Process. Appl. 60 (1995) 19.
[38] L. Giraitis, P. Kokoszka, R. Leipus, J. Appl. Prob. 38 (2001) 1033.
[39] J. Xiu, Y. Jin, Physica A 377 (2007) 138.
[40] R. Weron, Physica A 312 (2002) 285.
[41] E. Lutz, Phys. Rev. E 64 (2001) 051106.
[42] S.C. Kou, X.S. Xie, Phys. Rev. Lett. 93 (2004) 180603.
[43] W. Deng, E. Barkai, Phys. Rev. E 79 (2009) 011112.
[44] J.H. Jeon, R. Metzler, Phys. Rev. E 81 (2010) 021103.
[45] B. Mandelbrot, C. R. Acad. Sci. Paris 240 (1965) 3274.
[46] N.W. Watkins, D. Credgington, R. Sanchez, S.J. Rosenberg, S.C. Chapman, Phys. Rev. E 79 (2009) 041124.
[47] C.W.J. Granger, R. Joyeux, J. Time Series Anal. 1 (1980) 15.
[48] J.R.M. Hosking, Biometrika 68 (1981) 165.
[49] M. Magdziarz, A. Weron, Studia Math. 181 (2007) 44.
[50] O.E. Barndorff-Nielsen, Normal//Inverse Gaussian Processes and the Modelling of Stock Returns, Research Report 300, Department of Theoretical Statistics, University of Aarhus, 1995.
[51] R. Weron, Computationally intensive value at risk calculations, in: J.E. Gentle, W. Haerdle, Y. Mori (Eds.), Handbook of Computational Statistics: Concepts and Methods, Springer, Berlin, p. 911.
[52] R.B. D'Agostino, M.A. Stephens, Goodness-of-Fit Techniques, Marcel Dekker, New York, 1986.
[53] K. Burnecki, G. Kukla, R. Weron, Physica A 287 (2000) 269.
[54] K. Burnecki, J. Janczura, R. Weron, Building loss models, in: P. Cizek, W Härdle, R. Weron (Eds.), Statistical Tools for Finance and Insurance, second ed., Springer, Berlin, 2011, p. 293.
[55] I.A. Koutrouvelis, JASA 75 (1980) 918.
[56] H.E. Hurst, Trans. Am. Soc. Civ. Eng. 116 (1951) 770.
[57] R. Weron, B. Przybylowicz, Physica A 283 (2000) 462.
[58] M.S. Taqqu, V. Teverovsky, in: R.J. Adler, R. Feldman, M.S. Taqqu (Eds.), A Practical Guide To Heavy Tails: Statistical Techniques and Applications, Birkhäuser, Boston, 1998, p. 177.
[59] A.W. Lo, Econometrica 59 (1991) 1279.
[60] V. Teverovsky, M.S. Taqqu, W. Willinger, J. Statist. Plann. Inference 80 (1999) 211.
[61] K. Burnecki, A. Weron, Phys. Rev. E 82 (2010) 021130.
[62] J.D. Cryer, K.-S. Chan, Time Series Analysis: With Applications in R, second ed., Springer, New York, 2008.
[63] T. Bollerslev, J. Econometrics 31 (1986) 307.
[64] R.F. Engle, Econometrica 50 (1982) 987.
[65] J.H.H. Lee, Econom. Lett. 37 (1991) 265.
[66] G.M. Ljung, G.E.P. Box, Biometrika 65 (1978) 297.