TRACKING SCALING EFFECTS IN MUTUAL FUNDS RETURN TIME SERIES

Arkadiusz Biernacki

Silesian University of Technology, Institute of Computer Science
Akademicka 16, 44-100 Gliwice, Poland
arkadiusz.biernacki@polsl.pl

(Received October 24, 2012; revised version received November 28, 2012; final version received December 3, 2012)

Data coming from many fields of science and technology, ranging from hydrology through network traffic to economics, show long range dependence and self-similarity. These properties result in significant consequences and usually require a redefinition of well grounded assumptions and theories. In the case of financial markets, the classical models which often assume that the dynamics of economic time series is described by the random walk, may incorrectly evaluate the investment risk. Therefore, it is important to understand the dynamics of returns generated by different financial instruments. In this work, we tested fifteen different mutual funds investing in stocks through a stock exchange. We found that the distribution of funds daily returns cannot be described by the random walk. Furthermore, using several different method, we provide empirical evidence, that the daily returns of the analysed funds may exhibit long-range correlations and fractal behaviour.

DOI:10.5506/APhysPolB.43.2103
PACS numbers: 89.65.Gh

1. Introduction

Lots of research have been made over the years in order to explain how information diffuses amongst investors and how it is then influencing on the prices of financial instruments. Since financial markets are regarded as complex systems, therefore, the models that have been created to describe and explain their behaviour usually have comprised too many simplified rules. One of the important findings in the statistical characteristics of financial markets were long-range correlations and self-similarity in the return series, produced by different kinds of financial indicators or traded securities. However, it is not always clear what are the causes of these statistical properties, and to what extent they are present in a different kind of financial instruments like *e.g.* mutual funds.

Mutual funds are investment vehicles that pools funds collected from many investors for the purpose of investing in securities such as stocks, bonds money market instruments and similar assets. They are operated by money managers, who invest the collected money from investors and attempt to produce capital gains and income for the investors. Mutual funds may be interesting investment opportunity for small investors, being able to access professionally managed, diversified portfolios of equities, bonds and other securities, which would be difficult to create with a limited amount of capital. Based on the investment objective and the degree of investment risk, the funds can be classified into stock funds, bond funds, active configuration fund, *etc.* For example, a stock fund is one that invests mainly in stocks, and a bond fund adopts a collective investment scheme that invests mostly in bonds and other debt securities. In this work, we focus on the stock funds.

An assessment of a mutual fund gain *versus* its risk is of paramount importance for policy makers and regulators since it provides some guidance on appropriate classification strategies. Furthermore, it is crucial for the investment funds industry, portfolio and risk management purposes. One of the key factor when assessing the investment risk is the volatility of the fund price. The term volatility represents a generic measure of the magnitude of market fluctuations. It is also directly related to the amount of information arriving in the financial market at a given time. In the case of a mutual fund, the volatility is a measure of how much the fund price is liable to fluctuate. As the volatility is of great interest to investors, because it quantifies the risk, understanding its statistical properties has important practical implications. Nonetheless, financial markets, as it was shown in numerous works [1–3], are non-linear dynamic systems producing non-random, non-linear time series, which often exhibit long-term correlations, self-similarity or trends. Therefore, the employment of standard statistical analysis methods to price volatility can give misleading results. Thus, taking into account the above considerations, we focus on several basic statistical properties of time series composed of mutual fund prices: the probability distribution, autocorrelation, long range dependence and self-similarity. The main objective of this paper is to explore if the "new" analytical techniques, which form the bedrock of quantitative capital market theory: a concepts of chaos, fractals, long-range correlations, concern also the mutual fund market. The analysis focus on fifteen mutual funds which operate on the Polish stock market.

2. Long range dependence and self-similarity

Long range dependence (LRD) is a property of certain stationary time series. In the time domain, the LRD exhibits itself as a high degree of correlation between distantly separated data points, while in the frequency domain, a significant level of power at frequencies near zero can be observed. The LRD is inherently defined over a range of scales, and in many ways, may be a difficult statistical property to estimate. In the time-domain, it is measured only at high lags (strictly at infinite lags) of the autocorrelation function: those very lags where only a few samples are available, and where the measurement errors are the largest. In the frequency domain, it is measured at the frequencies near zero, again where it is hardest to make measurements.

There are several way of characterising LRD processes. A widespread definition takes into account the autocorrelation function $\gamma(k)$, where we define a process as a long-memory process if for $k \to \infty$ there exist the relation

$$\gamma(k) \sim k^{-\alpha} L(k)\,, \tag{1}$$

where $0 < \alpha < 1$ and $L(k)$ is a slowly varying function at infinity. The degree of the LRD, or a process memory, is given by the exponent α; the smaller α, the longer the process memory.

In many publications, the terms LRD and self-similarity are used interchangeably, which may lead to confusion. A self-similar process behaves the same when viewed at different degrees of magnification, or different time scales. Self-similar processes can sometimes be described using heavy-tailed distributions, what in the case of financial time series translates to heavy-tailed distribution of financial instrument returns. The fractal characteristics is a result of the LRD of the time series and fat-tail distributions of its returns. Some self-similar processes may exhibit the LRD, but not all processes having the LRD are self-similar.

The LRD and self-similarity are also discussed in terms of the Hurst exponent H, which is simply related to α from (1). For a stochastic process there exists a relation

$$H = 1 - \alpha/2\,. \tag{2}$$

When $H \in (0.5, 1]$ the process is positively correlated, which implies that the time series is persistent. The persistence is characterised by the LRD effects on all time scales, *i.e.* if the series has been up or down in the last period, then the chances are that it will continue to be up or down, respectively, in the next period. Consequently, all daily price changes are correlated with all future daily price changes; all weekly price changes are correlated with all future weekly price changes and so on. Short-memory processes have $H = 1/2$, and its autocorrelation function decays faster than the autocorrelation of the LRD process. On the other hand, when $H \in [0, 0.5)$ we have anti-persistence. This means that whenever the time series has been up in the last period, it is more likely that it will be down in the next period.

Thus, an anti-persistent time series will be more ragged than a pure random walk with $H = 0.5$. While the Hurst parameter is perfectly well-defined mathematically, it may be a difficult property to measure in real life.

3. Previous works

In an economics and capital market theory, it had been long assumed that there exists a natural balance between a supply and demand of stocks or currencies. This balance should have been preserved as long as a significant event would have changed the supply or demand. After such event occurred, the market was supposed to find a new equilibrium and stabilised itself until a next event would throw it off balance again. These assumptions led to a market efficiency hypothesis. According to [4], in an efficient capital market, all information available to investors is already reflected and discounted in the securities prices. Investors react to information instantaneously after the information has been received, and not in a cumulative fashion to a series of events. The security prices only move when new information arrives in the market. This means that today change in the security price is caused only by today unexpected new information. Yesterday news is no longer important, and today returns are unrelated to yesterday returns, *i.e.* the security returns are independent. Additionally, if to collect enough historical data, the distribution of the security returns should, according to the theory of large numbers, approach the Gaussian distribution, *i.e.* the security price returns behave like a sequence of i.i.d. Gaussian random variables. However, it was shown that the financial data are usually characterised by a probability distribution that exhibits power-law behaviour and cannot be fitted using the Gaussian distribution [5]. Some prominent approaches, including the Levy stable distribution [6], leptokurtic distribution [7], q-Gaussians [8, 9] were reported.

Therefore, taking into account the above cited works, the current researches clearly show that capital markets are complex and interdependent systems, where the state of the system is continuously fluctuating, with no natural equilibrium state. The financial time series generated by markets have long-term correlations and trends, which are a result of feedback effects, and can be more or less chaotic. Furthermore, researches show that historical time series derived from financial markets typically exhibit distinct non-periodic cyclical patterns that indicate the presence of significant power at low frequencies manifested as the LRD [10, 11]. The above mentioned findings have often become a source of controversies, primarily because of the wide range of implications that the presence of the LRD in financial series returns has on many of the paradigms used in economics. Above all, such findings are inconsistent with the efficient market hypothesis and require redefining statistical and stochastic analysis techniques which has been used in finance theory and its applications.

In economics and finance, the LRD has a long history and has remained a topic of active research is the studies of economic and financial time series. The range of applications of the LRD processes spans from macroeconomics to finance. In the macroeconomics, for example, the authors of [12] found evidence of long-memory in the gross national product (GNP) data of the USA. Although several criticisms have been raised to this work, further studies confirmed the evidence of LRD properties in the GNP data. In [13], it was shown that the monthly US Consumer Price Index (CPI) had LRD properties. A related study by [14] confirmed the presence of the LRD in inflation time series. The LRD and self-similarity were also observed in stock indices [1, 2], foreign exchange markets [3, 15, 16], traded volumes [17] and interest rates [18].

In addition to finance and economics, LRD processes have been observed in different natural and human phenomena ranging from hydrology [19], through meteorology [20] to geophysics, *e.g.* the temperature of the Earth [21]. The LRD has also started to play an important role in the analysis and performance modelling of traffic volume in communications networks. The LRD and self-similar traffic was observed in local networks [22] as well as in the Internet [23]. The latest comprehensive review of LRD presence in different domains may be found, among others, in [24].

Motivated by the aspiration to reduce the literature gap, in this paper, we investigate non-linear properties embedded in the return series of several mutual funds which invest at least 75% of their assets in stocks. Therefore, we analyse a return distribution, autocorrelation and we estimate the Hurst exponent of the funds daily returns. As the result of our analysis, we found that the distribution of daily returns does not follow Gaussian distribution, which is often applied to modelling of financial instruments returns. Furthermore, we found that the examined mutual fund market exhibits the LRD and self-similar properties, and the intensity of these phenomena depend on a particular mutual fund.

4. Measurements

While the Hurst parameter is mathematically well defined, the empirical determination of the LRD property of a time series usually is not trivial. The basic reason lays in a strong autocorrelation of LRD processes, which makes statistical fluctuations quite large. Thus, tests for the LRD usually require a considerable amount of data because the measurement should be done at tails of a distribution, where not so much data are available. Furthermore, different methods of the Hurst parameter estimation often give inconclusive or even contradictory results. The assessment results may be biased by trends, periodicity and corruptions in the data. Therefore, some authors suggested the use of a "portfolio" of estimators instead of relying on a single estimator, which could give a misleading assessment caused by properties of the time series under investigation [25]. Thus, in this paper, we use three widespread, well-known techniques, which have been used for some time, to estimate the Hurst exponent: R/S, Aggregated Variance (AV) and Differenced Aggregated Variance (DAV). All the chosen techniques have freely available code and are implemented, amongst others, in Rmetrics software [26], which is a part of the Cran R environment. We shortly describe the above mentioned methods in the next sections.

4.1. R/S

The main concept of the R/S analysis is to calculate the relation between the range of the values exhibited in a portion of the time series and the standard deviation of the values over the same portion of the time series. Thus, for a given set of observations $\{X_i\}, i \in \{1 \dots N\}$, with the partial sum

$$Y_n = \sum_{i=1}^{n} X_i \,, \qquad n \in \{1 \dots N\} \,,$$

and sample variance

$$S_n^2 = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \frac{1}{n} Y_n \right)^2 \,, \qquad n \in \{1 \dots N\} \,,$$

the rescaled adjusted range statistic or R/S-statistic is defined by

$$(\mathrm{R/S})_n = \frac{1}{S_n} \left[\max_{0 \le t \le n} \left(Y_t - \frac{t}{n} Y_n \right) - \min_{0 \le t \le n} \left(Y_t - \frac{t}{n} Y_n \right) \right] \,, \tag{3}$$

where we compute the difference between rescaled cumulative deviations adjusted to their mean. As we see from (3), to determine H parameter, a given sample of N observations is subdivided into blocks, each of size n. Then, for each lag n, $n \le N$, one estimates $(\mathrm{R/S})_n$. The graphical R/S approach consists then of plotting the estimates $(\mathrm{R/S})_n$ *versus* logarithmically spaced values of n. The parameter H can be estimated by fitting a line to the points in the plot.

4.2. Aggregated Variance

This method estimates the Hurst parameter from the variance of an aggregated (averaged) time series. The original time series $\{X_i\}$ is divided into blocks of size m, and for each block an average is computed

$$X_k^{(m)} = 1/m \sum_{i=(k-1)m+1}^{km} X_i \,, \qquad k = 1, 2, \ldots, [N/m] \,.$$

Then the sample variance within each block is computed according to the formula

$$\mathrm{Var} X^{(m)} = \frac{1}{N/m} \sum_{k=1}^{N/m} \left(X_k^{(m)} - \hat{X} \right)^2 \,,$$

where $\hat{X}$ is the mean of the time series $\{X_i\}$. If the series is Gaussian or at least has finite variance, $X^{(m)}$ scales proportionally to m^{2H-2} for large N/m and m. When this procedure is repeated for different values of m, one can plot the logarithm of the sample variances *versus* the logarithm of the block sizes. The resulting points should form a straight line with the slope $2H - 2$.

4.3. Differenced Aggregated Variance

This method calculates the differences between sample variances of successive blocks of a time series. The method tries to distinguish non-stationary elements of a time series: jumps and slowly decaying trends, from the LRD. The calculations are performed according to the formula

$$\mathrm{Var} X^{m_i+1} - \mathrm{Var} X^{m_i} \,, \tag{4}$$

where m_i are the successive values of m as defined in the AV method. The slope $2H - 2$ from the least square fit of the logarithm of the differenced sample variances (4) *versus* the logarithm of the block sizes provides an estimate for the Hurst exponent H.

5. Dataset

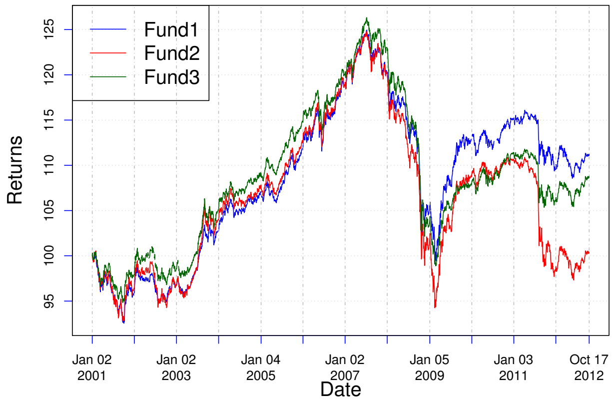
We took data from fifteen mutual funds which invest majority of their capital in a stock exchange purchasing stocks offered to the public. The selected funds publish every trading day their net asset value (NAV). The NAV is the price at which open-end funds stand ready to issue new shares or redeem existing shares, and is computed each trading day after the market close. The NAVs are believed to reflect the "fair value pricing" what is practice that gives fund companies discretion in reporting their best estimate of fund share value taking into account the estimated market value of infrequently traded securities or securities listed on foreign exchanges. The NAVs data were downloaded from popular Polish stock data provider bossa.pl and have the following symbols in the provider database: PIO3, PKCA, SEB3, AASW, AIFA, AIP7, AIPA, ARNE, INGA, DWA+, UNIA, AFAZ, AIP6, AIP8, ARDS. In our work, we identify the above listed funds by giving them consecutive numbers from 1 to 15, and we refer to them using the convention: *Fund number*, *e.g.*: Fund 1 refers to the PIO3 fund, Fund 4 refers to the AASW fund, *etc.* For our analysis, we took the NAVs from January 2001 to October 2012. The data set contains only trading days, thus excluding holidays and weekends. This means that five-day returns are the returns of five consecutive trading days, which may not necessarily need to be the days from Monday to Friday. A plot of daily closing prices for the three selected funds, analysed in this work, is shown in Fig. 1. In order to improve the comparison, the fund returns were adjusted to a common starting value of 100 for each presented time period. As we can observe, there is a certain degree of correlation between the fund NAVs. However, the investing strategies for the presented funds are different what can be noticed when to rescale the NAV plot in Fig. 1 (b) and 1 (c).

As the real life data is susceptible to trends and periodicity resulting from daily usage patterns, thus, running it through an off-the-shelf estimation method may give an inaccurate assessment of the Hurst parameter. Financial time series generally exhibit significant autoregressive tendencies, therefore, an empirical investigation of LRD properties in financial assets returns should take into account the presence of trends and oscillations which are the result of high-frequency autocorrelations [10]. During the Hurst parameter estimation, it is important to minimize the above mentioned disturbances in a time series. Otherwise, the estimation algorithm may classify the time series as having a long-term "memory" when it has in reality a short-term "memory" property. In order to avoid misleading results, we employ a method called de-trending, which was used amongst others in [10].

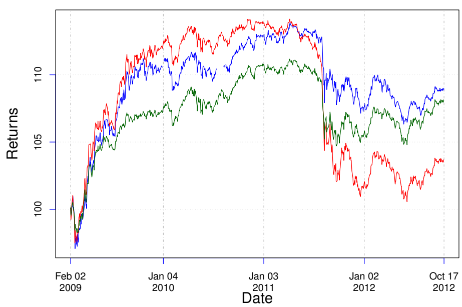
To apply de-trending procedure, let us assume that X_t is the NAV of a mutual fund on a time t. As a result of the procedure, we obtain

$$R_t = \ln(X_{t+1}/X_t)\,. \tag{5}$$

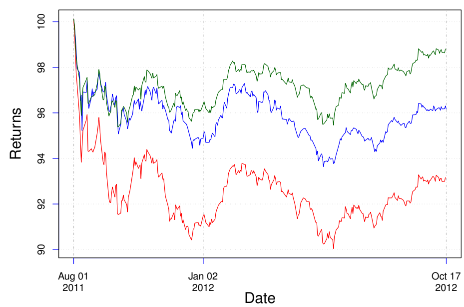
After applying (5) to the mutual fund NAV series, the obtained results R_t were further processed by the Hurst estimation procedures, described in Sec. 4.



(a) From January 2001 to October 2012.



(b) From February 2009 to October 2012.



(c) From August 2011 to October 2012.

Fig. 1. Evolution of mutual fund returns.

6. Results

Firstly, we look at the returns distribution of three randomly selected mutual funds covering the period from 2001 to 2012, Fig. 2. We plotted the empirical distributions of NAVs (fund returns) daily changes and compared them to the normal distribution with the same mean and variance. The figure reveals that the normal distribution is not suitable for the description of the NAVs distribution of the examined mutual funds. The return distribution of the examined mutual funds is clearly narrower and has fatter tails compared to the normal distribution. This is an indication that mutual funds returns do not follow a random walk. In this case, the investment decisions based on the assumption of the Gaussian price distribution may over- or underestimate risk and return potential on all trading horizons.

Secondly, we calculated the autocorrelation function C_τ for the de-trended data series R_t, computed in (5), of the three selected funds, where $\tau \in \{1, 2 \dots 64\}$ denotes consecutive trading days. As it was already mentioned

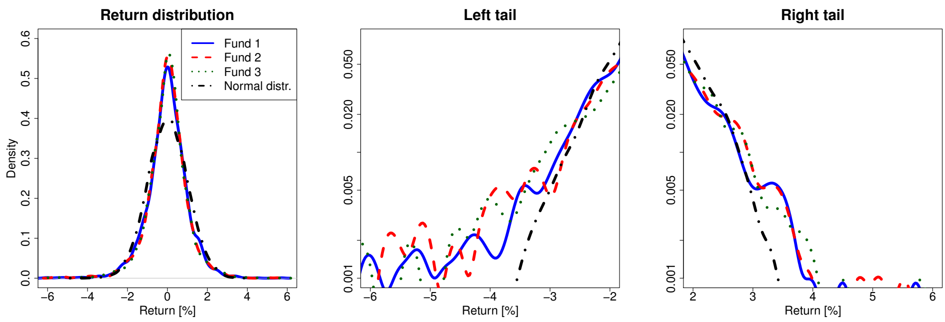


Fig. 2. Distribution of the mutual funds daily returns.

in Sec. 2, the autocorrelation of a LRD process is likely to decrease to zero in a manner described by (1). As shown in Fig. 3, the autocorrelation for the funds 2 and 3 decreases relatively slowly, clearly different from an exponential decay, although with some breaks in the slope. In the case of Fund 1, when we increase the autocorrelation lag, we can observe fluctuations of the autocorrelation coefficient.

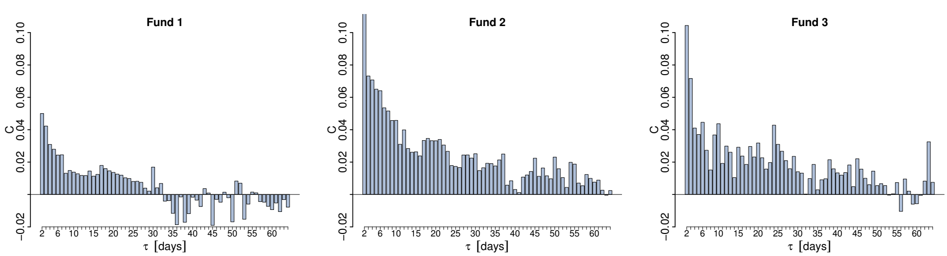


Fig. 3. Autocorrelation function C_τ for the de-trended NAV series R_t of mutual funds.

In order to further explore the autocorrelation function, we plotted its decay in a logarithmic scale in Fig. 4. Employing regression analysis, we found that the autocorrelation function for the examined funds decreases proportionally to $k^{-\alpha}$ with $\alpha \approx 1.01$ for about 21 days for Fund 1, $\alpha \approx 0.58$ for about 55 days for Fund 2, and $\alpha \approx 0.71$ for Fund 3 for about 50 days. These findings indicate that the empirical autocorrelation function may be reasonably approximated by the power-law relation (1) at least for funds 2 and 3.

The logarithmic plot of R/S analysis, described in Sec. 4.1, for the de-trended NAVs of Fund 1 is presented in Fig. 5 (a). As may be seen, the results are well described by the linear regression with the Hurst exponent $H \approx 0.51$, which indicates a short-memory property of the fund NAVs.

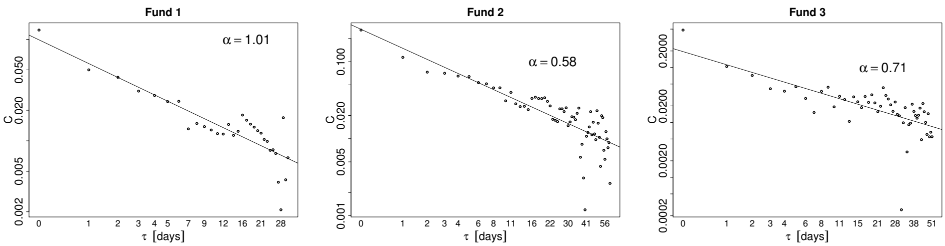
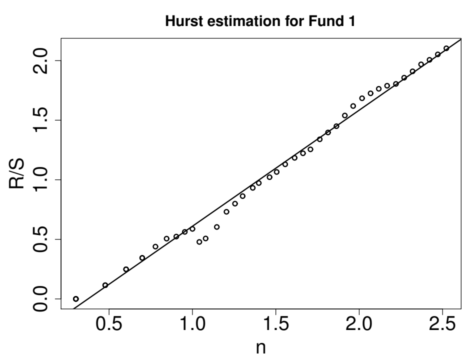
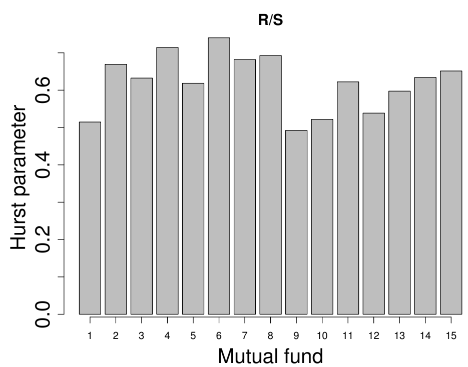


Fig. 4. Log-scaled autocorrelation function C_τ for the de-trended NAV series R_t of mutual funds.

The comparative analysis of all fifteen funds, which we take into account in our work, is shown in Fig. 5 (b). As we can see, the value of the Hurst parameter ranges from about 0.48 to 0.73. For most of the analysed funds, with the exception of Fund 9, the Hurst parameter is higher than 0.5. When to use (2) to compute the relation between H and α parameters presented in Fig. 4, the results for the funds 1, 2 and 3 are generally in an agreement with the results obtained from the autocorrelation function analysis.



(a) R/S method, single case estimation

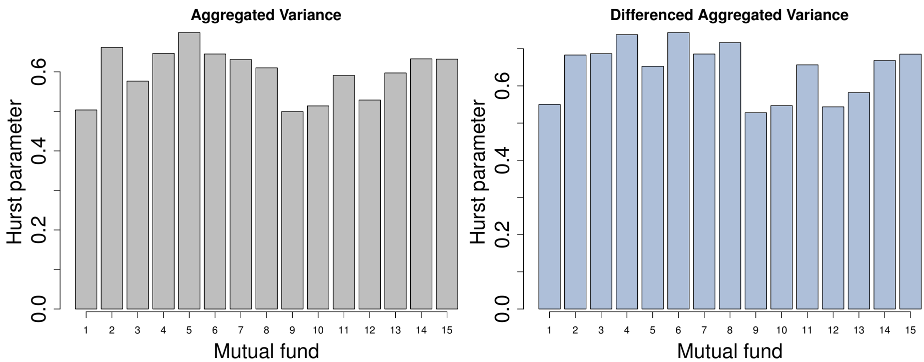


(b) R/S method

Fig. 5. Estimation of the Hurst parameter using R/S method for the de-trended NAV series R_t of mutual funds.

The results of the Hurst parameter assessment using the AV method, described in Sec. 4.2, are presented in Fig. 6 (a). In this analysis, the Hurst parameter ranges from about 0.52 for Fund 9 to 0.73 for Fund 5, which in most cases can be interpreted as a weak form of persistence in returns of the funds daily prices. Furthermore, usually the values of the Hurst parameters are higher compared to those obtained using the R/S method in Fig. 5 (b).

The DAV method, described in Sec. 4.3, gives correlated results to the AV method, Fig. 6 (b). The Hurst parameter ranges from about 0.52 for Fund 9 to 0.73 for Fund 6. This time, the analysis also suggests that most of the examined funds may possess LRD property. The estimations for funds 3, 5 and 6 are noticeable different compared to their the estimations using the AV method presented in Fig. 6 (a).



(a) Aggregated Variance method (b) Differenced Aggregated Variance method

Fig. 6. Estimation of the Hurst parameter using aggregate variance methods for the de-trended NAV series R_t of mutual funds.

7. Conclusions

In this paper, we investigated the statistical properties of several mutual funds returns investing in a stock exchange. We analysed their return distribution, autocorrelation and long range dependence. We provided empirical evidences that the mutual fund return distribution does not match the Gaussian distribution. Furthermore, most of the investigated funds show a long-range dependence behaviour which is significantly different from what a random walk would produce. Using R/S and aggregated variance methods, we estimated the Hurst parameter, which is a measure of the LRD, for every analysed fund. The value of the Hurst parameter ranges for the examined funds from about 0.48 to 0.73 depending on the estimation method used which indicates a certain degree of self-similarity in fund returns. The differences in the assessment may be a result of different investing strategies employed by the funds managers *e.g.*: some of the funds prefer investing rather in large liquid companies (blue-chips funds), others funds may specialise in small and medium size companies (small-caps funds); some of the funds may be interested in dividend stocks (dividend funds), while the other prefer locating their capital in growth stocks.

As a result of our study, we can state that methods based on the assumption of random walk of mutual fund prices may not correctly estimate the investment risk. Our results may contribute to development of new financial models which will better describe the price dynamics of this kind of financial instruments. A natural continuation of this work would be to extend the performed analysis on funds which undertake a wider range of investment and trading activities than mutual funds like *e.g.* hedge funds.

REFERENCES

[1] D.S. Ho, C.K. Lee, C.C. Wang, M. Chuang, *Physica A* **332**, 448 (2004).

[2] L. Zunino *et al.*, *Chaos, Solitons Fractals* **41**, 2331 (2009).

[3] S. Galluccio, G. Caldarelli, M. Marsili, Y.C. Zhang, *Physica A* **245**, 423 (1997).

[4] E.F. Fama, *J. Finance* **25**, 383 (1970).

[5] R.N. Mantegna, H.E. Stanley, *Nature* **376**, 46 (1995).

[6] M. Levy, S. Solomon, G. Ram, *Int. J. Mod. Phys.* **C7**, 65 (1996).

[7] S. Solomon, M. Levy, `arXiv:cond-mat/0005416 [cond-mat.stat-mech]`.

[8] S. Drożdż *et al.*, *Acta Phys. Pol. B* **34**, 4293 (2003).

[9] R. Rak, S. Drożdż, J. Kwapień, *Physica A* **374**, 315 (2007).

[10] A.W. Lo, *Econometrica* **59**, 1279 (1991).

[11] Z. Ding, C.W.J. Granger, R.F. Engle, *Journal of Empirical Finance* **1**, 83 (1993).

[12] F.X. Diebold, G.D. Rudebusch, *Journal of Monetary Economics* **24**, 189 (1989).

[13] R.T. Baillie, C.F. Chung, M.A. Tieslau, *Journal of Aapplied Econometrics* **11**, 23 (1996).

[14] U. Hassler, J. Wolters, *Journal of Business & Economic Statistics* **13**, 37 (1995).

[15] N. Vandewalle, M. Ausloos, *Eur. Phys. J.* **B4**, 257 (1998).

[16] J.T. Barkoulas, W.C. Labys, J.I. Onochie, *Financial Review* **34**, 91 (1999).

[17] J. De Souza, L.G. Moyano, S.M. Duarte Queirós, *Eur. Phys. J.* **B50**, 165 (2006).

[18] D.O. Cajueiro, B.M. Tabak, *Phys. Lett.* **A372**, 181 (2008).

[19] B.B. Mandelbrot, J.R. Wallis, *Water Resources Research* **4**, 909 (1968).

[20] B.B. Mandelbrot, J.R. Wallis, *Water Resources Research* **5**, 321 (1969).

[21] D.L. Turcotte, *Fractals and Chaos in Geology and Geophysics*, Cambridge University Press, July 1997, p. 412.

[22] W.E. Leland, M.S. Taqqu, W. Willinger, D.V. Wilson, *Comput. Commun. Rev.* **23**, 183 (1993).

[23] M.E. Crovella, A. Bestavros, *IEEE/ACM Transactions on Networking* **5**, 835 (1997).

[24] J. Kwapień, S. Drożdż, *Phys. Rep.* **515**, 115 (2012).

[25] S. Murad, V. Teverovsky, W. Willinger, *Fractals* **3**, 785 (1995).

[26] D. Wuertz *et al.*, Rmetrics, 2012.