Managerial Economics
2015, vol. 16, no. 1, pp. 7–37
http://dx.doi.org/10.7494/manage.2015.16.1.7

Krzysztof Brania*, Henryk Gurgul**

The impact of estimation methods and data frequency on the results of long memory assessment***

1. Introduction

Long memory describes the high order correlation structure of a series. If a time series exhibits long memory, there is a persistent temporal dependence between observations, even when considerably separated in time. The autocorrelation function (ACF) of series with long memory tails off hyperbolically. These series have low-frequency spectral distributions. In contrast to long memory, short memory is characterized by the low order correlation structure of a series. The presence of long memory means that the market does not immediately respond to information upcoming in the financial market. The market reacts to it gradually over a period of time. This is why past price changes can be used as a significant basis for the prediction of future price changes. The main implication of long memory is that shocks to the volatility process tend to have long-lasting effects. Such persistence is a crucial component of risk management, investment portfolios, and derivative pricing.

There has been a large amount of research on long memory in economic and financial time series. The presence of long memory in asset returns has important implications for many of the models used in modern financial economics. For example, the pricing of derivative securities with martingale methods is no longer valid, since most of the stochastic calculus employed in martingale analysis is inconsistent with long memory. Long memory is also inconsistent with

* Jagiellonian University in Cracow, Faculty of Mathematics and Computer Science, e-mail: krzysztofbrania@gmail.com
** AGH University of Science and Technology, Faculty of Management, Department of Application of Mathematics in Economics, e-mail: henryk.gurgul@gmail.com
*** Financial support for this paper from the National Science Centre of Poland (Research Grant DEC--2012/05/B/HS4/00810) is gratefully acknowledged.

the usual statistical inference methods that are employed to estimate and conduct hypothesis testing in the CAPM model.

The present study – in contrast with most papers on this subject – concentrates on a comparison of long memory estimates as calculated by different methods.

We have used the most recent data (both daily and hourly) for the indices and companies chosen. The period under study begins on 31 December, 2013, and ends on 19 December, 2014. This time period is sufficient for high frequency data and even daily data.

The outline of this paper is as follows: the most important contributions involved with long memory in returns, return volatility, and trading are reviewed in the next section. The third section provides a definition of long memory and outlines the most important traditional estimation methods of the long memory parameter. This section also briefly characterizes the data basis. The empirical results are presented in the fourth section. In the last section, the results are summarized.

2. Returns and trading volume

It is widely accepted that stock prices reflect investor expectations about the future development of a firm. Upcoming information is the main factor that changes investor beliefs and, therefore, is the main reason for changes. There are situations when prices do not move in spite of new, important upcoming information. This is possible when particular groups of investors interpret the same new information differently. Sometimes, they interpret new information identically but start from different initial expectations. From a mathematical point of view, changes in stock prices reflect the sum or average of investor behavior in reaction to the news. It is clear that stock price changes can be noticed if there is a positive trading volume.

Together with price data, volume data is also reported. As in the case of prices, trading volume changes depend on market information. In contrast to stock prices, a change of intraday expectations is a source of a rise in trading volume. Trading volume reflects the sum of investor decisions in reaction to the news. Differences between investors in their interpretation of new information are not cancelled out, as in the case where an averaging process determines prices. Stock price behavior and trading volume changes are helpful in determining the dynamic properties of stock markets. The observation of both variables allows for a better understanding of the importance of upcoming information on the market. It is worth noting that speculation motivates investments, even in the absence of new information.

Clark (1973) formulated the *Mixture of Distribution Hypothesis* (MDH), which states that stock returns and trading volume are jointly dependent on a latent information flow variable. This hypothesis posits trading volume as a proxy for the upcoming information stochastic process, which MDH implies as a positive contemporaneous relationship between volume and return volatility data.

An alternative hypothesis formulated by Copeland (1976) is known as the *sequential information flow model* (SIAH). According to this conjecture, news is disseminated sequentially rather than simultaneously to market participants. Sequential information flow is the source of a sequence of transitional price equilibriums. They are accompanied by persistent high trading volume. The most important implication of Copeland's hypothesis is the existence of positive contemporaneous as well as causal relations between price volatilities and trading volume.

Darrat et al. (2003), using intraday trading data for 30 stocks in the DJIA, also found that high trading volume causes high return volatility (which accords with the SIAH but not the MDH).

In many financial time series, there is a persistence in autocorrelation. This property is called long memory. The notion of long memory was formulated by British hydrologist Harold Edwin Hurst (1951). The earliest contributions to the subject of long memory in time series are those by Mandelbrot and Van Ness (1968) and Mandelbrot (1971). They formalized Hurst's empirical findings using cumulative river flow data, see Geweke and Porter-Hudak (1983), Hosking (1981). Granger and Joyeux (1980) introduced fractionally-integrated ARMA models, which were discussed by Sowell (1992), Beran (1992), and Baillie (1996), among others.

Finance researchers in both theoretical and empirical studies have focused on long memory (persistence) in financial asset returns. The finding of long memory in financial data would contradict the Efficient Markets Hypothesis of Fama (1970). The EMH is based on the assumption of the martingale behavior of market prices that rules out long memory. The first application of the persistence concept to finance is that of Greene and Fielitz (1977). They used the rescaled – range (R/S) method of Hurst; in this way, they confirmed the existence of long memory in daily equity returns. However, this result was rejected by Lo (1991), who also used the R/S method. Neither did investigations by Aydogan and Booth (1988), Crato (1994), Cheung et al. (1993), and Cheung and Lai (1995), Barkoulas and Baum (1996), Hiemstra and Jones (1997) detected the presence of long memory in finance data. Beveridge and Oickle (1997) investigated long memory dependence in Canadian daily stock returns by ARIMA models and found long memory mean reversion.

Contributions by Booth et al. (1982), Helms et al. (1984), Cheung and Lai (1993), Fang et al. (1994), and Barkoulas et al. (1997) detected long memory in some kinds of foreign currency rates.

In the following years, researchers came back to stock markets and started to investigate not only returns but also return volatility (absolute values of returns or squared returns) and trading volume, also with respect to long memory and bivariate long memory.

Estimation results by Bollerslev and Mikkelsen (1996) provided new evidence that the apparent long-run dependence in US stock market volatility is best described by a mean-reverting fractionally integrated process, so that a shock to the optimal forecast of future conditional variance dissipates at a low hyperbolic rate.

Granger and Zhuanxin (1996) justified the relevance of long memory by using returns from a daily stock market index. The authors found that a number of other processes can be long memory, such as generalized fractionally integrated models resulting from aggregation, time-changing coefficient models, and possibly nonlinear models.

Bollerslev and Jubinski (1999) checked the behavior of stock trading volume and volatility for individual firms from the Standard & Poor 100 composite index. They found evidence for the MDH. The long-run hyperbolic decay rates appeared to be common across each volume-volatility pair. Moreover, fractionally-integrated processes best describe long-run temporal dependencies in volume and volatility series.

Koop et al. (1997) conducted a Bayesian analysis of ARFIMA models and described the testing of ARFIMA against ARIMA alternatives.

Lobato and Velasco (2000) checked the properties of 30 equities in the DJIA with respect to long memory. According to this study, trading volume exhibited long memory. In addition, volatility and volume shared the same degree of long memory for most stocks. However, the authors did not detect a common long memory component for both processes.

Like both of the above studies by Koop et al. (1997) and Lobato and Velasco (2000), we use individual stock data instead of index data in our contribution.

In the next section, we explain the notion of long memory in detail.

3. Long memory

The necessary and sufficient condition that a covariance stationary stochastic process exhibits long memory with memory parameter d is that its spectral density function $f(\lambda)$ satisfies:

$$f(\lambda) \sim c\lambda^{-2d} \text{ as } \lambda \to 0^{+} \tag{1}$$

Here, c is a finite positive constant, and symbol "~" means that the ratio of the left- and right-hand sides tends to one at the limit. According to the literature (Granger and Joyeux (1980), Hosking (1981), Beran (1994)), when the process satisfies condition (1) and $d > 0$ its autocorrelation function dies out at a hyperbolic rate; i.e.:

$$\rho_k \sim c_\rho k^{2d-1} \text{ as } k \to \infty$$

Parameter d determines the nature of the memory of the process. If $d > 0$, the spectral density is unbounded near the origin, and the process exhibits long memory. If $d = 0$, the spectral density is bounded at 0, and the process is called short memory. When $d < 0$, the spectral density is zero at the origin, and the process is labelled antipersistent and displays negative memory.

Typical of long memory processes, satisfying (1) is the class of autoregressive fractionally integrated moving average (ARFIMA) processes introduced into econometrics by Granger and Joyeux (1980).

X_t is called an ARFIMA(p, d, q) process if:

$$\Phi(B)(1-B)^d(x_t-\mu)=\Theta(B)\varepsilon_t$$

where $\Phi(z)=1-\varphi_1 z-\ldots-\varphi_p z^p$ and $\Theta(z)=1-\theta_1 z-\ldots-\theta_q z^q$ are lag polynomials of order p and q respectively in the backshift operator B with roots outside the unit circle, ε_t is iid (0, σ^2), and $(1-B)^d$ is defined by binomial expansion:

$$(1-B)^d=\sum_{j=0}^{\infty}\frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)}L^j$$

In addition to the previously mentioned properties of memory, if $d > -0.5$, the ARFIMA process is invertible and possesses a linear Wold representation, and if $d < 0.5$, it is covariance stationary. Thus, if $0 < d < 0.5$ the process is stationary and exhibits long memory. Many non-stationary series can be transformed by integers integrating into stationary ones with a spectral density satisfying (1).

There are several methods for the estimation of long memory parameter d. We will review the main methods briefly in the next subsections. Major methods will be reviewed in subsections 3.1–3.4, while other methods will be shortly outlined in part 3.5.

3.1. Maximum likelihood estimator

Maximum likelihood estimation (MLE) in the time domain needs an assumption about the exact form of the estimated ARFIMA model.

Then the exact Gaussian likelihood function for given sample $\{x_t\}_{t=1\ldots T}$ is:

$$L\left(d,\varphi,\theta,\sigma^2,\mu\right)=-\frac{T}{2}\ln|\Sigma|-\frac{1}{2}\left(X-\mu l\right)^T\Sigma^{-1}\left(X-\mu l\right) \quad (2)$$

where $\mathbf{l} = (1,\ldots,1)^T$, $X = (x_1,\ldots, x_T)^T$, φ and θ are the parameters of autoregression and moving average polynomials respectively, μ is the mean of the process, and Σ is its covariance matrix. Sowell (1992) proved that the exact maximum likelihood estimator (EML) obtained by maximizing the likelihood function (2) is consistent and asymptotically normal; i.e.:

$$\hat{d}_{EML} \sim N\left(d,(\pi^2 T/6-c)^{-1}\right)$$

where $c = 0$ when $p = q = 0$ and $c > 0$ otherwise.

Other properties of MLE and methods of solving some computational problems are discussed in Sowell (1992) and Doornik and Ooms (2003). There are several modifications of exact maximum likelihood estimation; e.g. modified profile likelihood (see Cox and Reid (1987) and An and Bloomfield (1993)) or conditional maximum likelihood (see Tanaka (1999) and Nielsen (2004)). The main drawback of such maximum likelihood estimators is their sensitivity to any model misspecification, so they can easily be influenced by any short-run dynamics.

3.2. GPH estimator

Another class of estimators of long memory parameter d are semiparametric estimators based on the approximation (1) of the spectral density function near the origin. Among them, the most popular is the log-periodogram regression method originally developed by Geweke and Porter-Hudak (1983) and analyzed in detail by Robinson (1995a). Semiparametric estimators use only information from the periodogram for very low frequencies; thus, they are robust to short-run dynamics. Based on condition (1), after taking the logarithms and inserting sample quantities, the long memory estimator is computed from the approximate regression relationship:

$$\ln\left(I(\lambda_j)\right) \approx \text{const} - 2d\ln(\lambda_j)$$

where $\lambda_j = \frac{2\pi j}{T}$ are the Fourier frequencies and $I(\lambda) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T} x_t e^{it\lambda}\right|^2$ is the periodogram of the given sample $x_1,\ldots,x_T$. The Geweke and Porter-Hudak (GPH) estimator is then defined as the OLS estimator in the above regression using only $j = 1,\ldots,m$ its first values, where $m=m(T)$ is a bandwidth parameter satisfying condition:

$$\frac{1}{m}+\frac{m}{T}\to 0 \text{ as } T\to\infty$$

Geweke and Porter-Hudak originally suggested choosing m equal $\sqrt{T}$. For further considerations about the optimal bandwidth, see Hurvich et al. (1998) and Henry and Robinson (1996). The asymptotical normality of the GPH estimator was initially proved by Robinson (1995a) for $d \in (-1/2,\ 1/2)$. But recently, Kim and Philips (1999) and Velasco (1999a) showed that it is consistent for $d \in (-1/2,\ 1)$ and has an asymptotically-normal limit distribution for $d \in (-1/2,\ 3/4)$:

$$\hat{d}_{GPH} \sim N\left(d, \frac{\pi^2}{24m}\right)$$

There are several modifications of the GPH estimator. For example, Agiakloglou et al. (1993) suggested replacing the constant in the regression by the polynomial in order to reduce bias (see also Andrews and Guggenberger (2003)). Similarly, an estimator that allows a short-run component was proposed by Shimotsu and Phillips (2002a).

The univariate GPH estimator described above can be generalized for the multivariate case. Consider $x_t = \left(x_{1,t},\ldots,x_{N,t}\right)$ a covariance stationary N-dimensional vector process with mean vector μ and covariance matrix Γ_j at lag j and a fractional integration vector $(d_1,\ldots,d_N)$, i.e. each $x_{i,t}$ is integrated of order d_i. For any a, $b=1,\ldots,N$ and $\lambda_j = \frac{2\pi j}{T}$ define the crossperiodogram of the process x_t:

$$I_{ab}(\lambda) = \left(\frac{1}{(2\pi T)^{1/2}}\sum_{t=1}^{T} x_{a,t}e^{it\lambda}\right)\left(\frac{1}{(2\pi T)^{1/2}}\sum_{t=1}^{T} x_{b,t}e^{it\lambda}\right)^{*}$$

where the asterisk means complex conjugation. For a bandwidth parameter m define $Y_{kj} = \ln\left(I_{kk}(\lambda_j)\right)$, $k=1,\ldots,N$, $j=1,\ldots,m$. Then the multivariate GPH estimator of fractional integration d_k is given by:

$$\hat{d}_k = -\frac{\sum_{j=1}^{m} \nu_j Y_{kj}}{2\sum_{j=1}^{m} \nu_j^2} \text{ where } \nu_j = \ln\lambda_j - \frac{1}{m}\sum_{j=1}^{m}\ln\lambda_j \qquad (3)$$

For individual series $\{x_{j,t}\}_{t=1,\ldots,T}$ this estimator is equivalent to the univariate GPH estimator previously described but based on its asymptotic normality (Robinson (1995a)). A Wald-type test for null hypothesis:

$$H_0: \quad Pd = \rho$$

for a $u \times N$ matrix P and $N \times 1$ vector ρ can be constructed. Test statistic:

$$4m\left(P\hat{d}-\rho\right)^{T}\left(P\hat{\Omega}P^{T}\right)^{-1}\left(P\hat{d}-\rho\right)$$

has the limiting χ_u^2 distribution, where $\hat{d}=\left(\hat{d}_1,\ldots,\hat{d}_N\right)$ and $\hat{\Omega}$ is a consistent estimate of the limiting variance of $2\sqrt{m}\left(\hat{d}-d\right)$ (see Robinson (1995a)). In the case of testing for a common long memory parameter of the process, ρ is a vector of zeroes and $P=\left(I_{N-1}\vdots 0\right)-\left(0\vdots I_{N-1}\right)$ is a $(N-1)\times N$ matrix, where I_{N-1} is the identity matrix of dimension $N-1$.

When the existence of a common order of integration d is assumed, the restricted least square estimator is given by:

$$\hat{d}=-\frac{1}{2}\frac{\sum_{j=1}^{m}1_N^T\hat{\Omega}^{-1}Y_j\nu_j}{1_N^T\hat{\Omega}^{-1}1_N\sum_{j=1}^{m}\nu_j^2} \tag{4}$$

where $Y_j=\left(Y_{1j},\ldots,Y_{Nj}\right)^T$ and 1_N is a $N \times 1$ vector of ones. Like the unrestricted estimates, the $\hat{d}$ is asymptotically normally distributed.

3.3. Whittle estimator

Another class of semiparametric estimators includes the narrow-band Gaussian or local Whittle estimators introduced by Künsch (1987) and developed by Robinson (1995b), Lobato (1999). In the univariate case, it is defined as a maximizer of the likelihood function:

$$Q(g,d)=-\frac{1}{m}\sum_{j=1}^{m}\left[\ln\left(g\lambda_j^{-2d}\right)+\frac{I(\lambda_j)}{g\lambda_j^{-2d}}\right] \tag{5}$$

The ranges of consistency and asymptotic normality of the local Whittle estimator are the same as those for the HPG estimator (see Velasco [1999b] and Phillips and Shimotsu (2004)), but the Whittle estimator is more efficient because asymptotically:

$$\hat{d}_{LW}\sim N\left(d,\frac{1}{4m}\right)$$

For further modifications of the local Whittle estimator, see for example Shimotsu and Phillips (2002b) or Andrews and Sun (2004).

As with the GPH estimator, the local Whittle estimator can be defined in the multivariate case. The corresponding (concentrating) likelihood function is

$$Q(d) = \frac{2}{m}\sum_{i=1}^{N} d_i \sum_{j=1}^{m} \ln\lambda_j - \ln\left|\hat{R}(d)\right| \tag{6}$$

where:

$$\hat{R}(d) = \frac{1}{m}\sum_{j=1}^{m} \Lambda_j \operatorname{Re}\left\{I(\lambda_j)\right\}\Lambda_j \tag{7}$$

with $\Lambda_j = diag\left(\lambda_j^{d_1},\ldots,\lambda_j^{d_N}\right)$ and a crossperiodogram matrix $I(\lambda)$. The estimator $\hat{d} = \left(\hat{d}_1,\ldots,\hat{d}_N\right)$ is defined as a maximizer of the concentrating likelihood function (6). It can be computed in two ways: by a numerical maximizing of (6) or using the two-step procedure proposed by Lobato (1999). The first step is to compute the univariate QMLE for every series (denote that vector by $\hat{d}^{(1)}$) and the second is to compute the following expression:

$$\hat{d}^{(2)} = \hat{d}^{(1)} - \left(\frac{\partial^2 Q(d)}{\partial d \partial d^T}\bigg|_{\hat{d}^{(1)}}\right)^{-1}\left(\frac{\partial Q(d)}{\partial d}\bigg|_{\hat{d}^{(1)}}\right)$$

As shown by Lobato (1999), the above two-step estimator has the same asymptotic distribution as the QMLE based on equation (6), but it is straightforward to calculate. Under the reasonable assumption:

$$\hat{d}^{(2)} \sim N\left(d, \frac{1}{\sqrt{m}}E^{-1}\right)$$

where $E = 2\left(I_N + R \circ R^{-1}\right)$ and $\circ$ denotes the Hadamard product of two matrices.

Based on these asymptotic properties, a test for the null hypothesis of a linear set of restrictions on d is available. Consider P which is $q \times N$ matrix, $N \times 1$ vector ρ and the null hypothesis:

$$H_0: \quad Pd = \rho$$

Then the test statistic:

$$m\left(P\hat{d}^{(2)} - \rho\right)^T \left(P\hat{E}^{-1}P^T\right)^{-1}\left(P\hat{d}^{(2)} - \rho\right)$$

is asymptotically χ_q^2 distributed under the null hypothesis. It allows testing for a common long memory parameter. In this case, ρ is a vector of zeroes and $P = \left(I_{N-1} \vdots 0\right) - \left(0 \vdots I_{N-1}\right)$ is a $(N-1) \times N$ matrix. On the other hand, it allows a test of whether the vector process is $I(0)$ or $I(1)$. In this case, $P = I_N$ and ρ is $N \times 1$ vector of zeroes or ones, respectively.

If the existence of a common order of integration is assumed, the estimator of d_* can be computed by maximizing the likelihood function:

$$Q_*(d) = \frac{2Nd}{m}\sum_{j=1}^{m}\ln\lambda_j \tag{8}$$

The resulting QMLE d_* is asymptotically normally distributed:

$$\hat{d}_* \sim N\left(d_*, \frac{1}{4Nm}\right)$$

3.4. Long memory versus Hurst exponent

The Hurst exponent is applied as a measure of the long memory of a time series. It reflects the autocorrelations of the time series. In addition, it is related to the rate at which these fall as the lag between pairs of values increases. The first definition of the Hurst exponent was originally formulated in hydrology. The aim of the study conducted by Hurst was to determine the optimum dam size for the Nile River. The volatile rain and drought conditions for this river were observed over a long period of time.

The Hurst exponent is known as the "index of dependence" or "index of long-range dependence." It measures the relative tendency of a time series either to regress strongly to the mean or to cluster in a direction. A value H in the range 0.5–1 indicates a time series with long-term positive autocorrelation. This means that large values in the time series will probably be followed by other large values. A Hurst exponent between 0 and 0.5 describes a time series with long-term switching between high and low values in neighboring pairs. This means that a large value will probably be followed by a low value, and that the value after that will tend to be large again. The case where $H = 0.5$ can indicate a completely uncorrelated series. This value characterizes those series for which the autocorrelations at small time lags can be positive or negative. However, the absolute values of the autocorrelations vanish exponentially to zero rather quickly. For $0.5 < H < 1$ and $0 < H < 0.5$, the autocorrelations fall according to a power law. The Hurst exponent and long memory parameter are related according to the following equation:

$$d = H - 0.5$$

The Hurst exponent, H, is defined in terms of the asymptotic behavior of the rescaled range as a function of the time span of a time series as follows:

$$\mathrm{E}\left[\frac{R(n)}{S(\mathrm{n})}\right] = Cn^H \quad \text{as} \quad n \to \infty$$

where $R(n)$ is the range of the first n values and $S(n)$ is their standard deviation, $\mathrm{E}(X)$ is the expected value, n is the time span of the observation (number of data points in a time series), C is a constant.

3.5. Applied methods

Now, we will list the methods used in our computations.

The Geweke and Porter-Hudak method. There is a well-known method, introduced in the work of Geweke and Porter-Hudak (1983), described in section 3.2.

A modified Geweke and Porter-Hudak method. This is a modification of the periodogram method. The algorithm divides the frequency axis into logarithmically-equally spaced boxes and then averages the periodogram values connected with frequencies inside the box.

The Whittle estimator. This also performs the periodogram analysis as described in section 3.3.

The remaining computations were conducted based on the Hurst exponent by using R software with the additional package "fArma." We used nine different functions to estimate the self-similarity parameter or long range dependence in a time series, as described by Taqqu, Teverovsky, Willinger (1995) and by Palma (2007). Now, we give a short overview of the methods used:

The R/S Rescaled Range Statistic method. A well-known method originally developed by H.E. Hurst (1951).

The aggregated variance method. The original time series is divided into blocks. Afterwards, the sample variance within each block is calculated. The slope $\beta = 2H - 2$ from the least square fit of the logarithm of the sample variances versus the logarithm of the block sizes provides an estimated value of the Hurst exponent H.

The aggregated variance method is based on the observation that the variance of the sample mean of a long memory process of m observations behaves like $Var(\bar{y}_m) \sim cm^{2d-1}$, for large m, where c is a positive constant. Next, by dividing a sample of size n, $\{y_1,\ldots,y_n\}$ into k blocks of size m, we have $\log[Var(\bar{y}_j)] \sim c + (2d-1)\log(j)$, for j = 1,2,…,k, where $\bar{y}_j$ is the average of the j-th block. Consequently, we see that the heuristic least squares estimator of d is:

$$\hat{d} = \frac{1}{2} - \frac{\sum_{j=1}^{k}\left[\log(j) - a\right]\{\log[Var(\bar{y}_j)] - b\}}{2\sum_{j=1}^{k}[\log(j) - a]^2} \tag{9}$$

where $a = \frac{1}{k}\sum_{j=1}^{k}\log(j)$ and $b = \frac{1}{k}\sum_{j=1}^{k}\log[Var(\bar{y}_j)]$. Thus, for a short-memory process, $d = 0$ and the slope of the line described by formula (9) should be –1. On the other side, for a long memory process with parameter d, the slope is $\beta = 2d - 1 = 2H - 2$.

The differenced aggregated variance method. In order to differentiate jumps and slowly decaying trends, this method differences the sample variances of successive blocks. The slope $\beta = 2H - 2$ from the least squares fit of the logarithm of the differenced sample variances versus the logarithm of the block sizes ensures an estimate of the Hurst exponent H.

The aggregated absolute value method. This computes the Hurst exponent from the absolute values of an integrated time series process. Again, the slope $\beta = H - 1$ of the regression line of the logarithms of the statistic versus the logarithm of block sizes gives an estimate for the Hurst exponent H.

Higuchi's fractal dimension method. This implements a technique similar to the absolute value method. Instead of blocks it uses a sliding window to compute the aggregated series. The function includes the calculating of the length of a path and finding its fractal dimension D. The slope $D = 2 - H$ from the least squares regression of logarithm of the expected path lengths on the logarithm of the window sizes provides an estimate of Hurst exponent H.

Peng's variance of residuals method. The series is divided into blocks of size m. Within each block the cumulated sums are computed up to t and the least squares line $y = \alpha x + \beta$ is fitted to the cumulated sums. Next, the sample variance of the residuals is calculated proportional to m^{2H}. Let σ_k^2 be the estimated residual variance from the regression model within block k, $\sigma_k^2 = \frac{1}{m}\sum_{t=1}^{m}(x_t - \widehat{\alpha_k} - \widehat{\beta_k} t)^2$, where $\widehat{\alpha_k}$ and $\widehat{\beta_k}$ are the least squares estimators of the intercept and the slope of the regression line. Let $F^2(k)$ be the average of these variances, $F^2(k) = \frac{1}{k}\sum_{j=1}^{k}\sigma_j^2$. As described by Peng et al. (1994), for a random walk it behaves like $F(k) \sim ck^{\frac{1}{2}}$, while for a long-range sequence, $F(k) \sim ck^{d+\frac{1}{2}}$. Then, by taking logarithms we have $\log F(k) \sim \log c + \left(d + \frac{1}{2}\right)\log k$. Thus, by fitting the least squares regression we get $\log F(k) = \alpha + \beta \log k + \varepsilon_k$. Next we may obtain an estimate of d as $\hat{d} = \hat{\beta} - \frac{1}{2}$, where $\hat{\beta}$ is the least squares estimator of parameter β. Therefore, if the result is plotted in a log-log plot on m, we will get a straight line with a slope $2H$.

3.6. Data description

In our computation, we picked a broad spectrum of most important indices and companies from leading stock markets over the world. The data consists of the log-returns, squared returns which are the measure of volatility and the natural logarithms of trading volume series for 8 indices (Australia – 200 largest Australian companies (AUS.IDX), France – 40 largest French companies (FRA.IDX), Germany-30 major German companies (DEU.IDX), Great Britain – 100 top UK companies as per capitalization (GBR.IDX), Japan – over 200 of Japanese leading companies (JPN.IDX), Netherlands (NLD.IDX), Switzerland – 20 Swiss blue-chips (CHE.IDX), United States of America – 500 major American companies (USA.IDX)) and 43 companies (including 5 British, 4 French, 4 German, 5 Swiss, and 25 American). The price data is at close. We consider both daily and hourly intervals. The period of time begins on 31 December, 2013, and ends on 19 December, 2014. Each hourly time series includes nearly 8,000 observations. We used the most-recent available financial data. Naturally, zero values were omitted in all considered time series.

Price and volume data comes from contracts for difference (CFDs). CFD is an agreement between two parties – "buyer" and "seller" – to exchange the difference between the current value of an underlying asset and its value at contract time. Therefore, CFDs are financial derivatives that allow traders to take advantage of price movements on underlying financial instruments and are very often used to speculate on such markets. Considering equities, CFD is an equity derivative that allows investors to speculate on share price movements without ownership of the underlying shares. The main advantages of CFDs, compared to futures contracts (agreements between two parties to buy or sell an asset for a price agreed upon today with payment and delivery occurring at the future point – delivery date), is that the sizes of contracts are smaller, making them more available for small traders and pricing much more transparent. A CFD never expires and is a simple mirror of the underlying financial instrument, while a futures contract tends to only converge near the expiry date compared to the price of the underlying instrument.

4. Empirical results

The Hurst exponents were calculated using R software with the additional package "fArma". We used nine different functions (referring to the methods outlined in [3.5.]) to estimate the self-similarity parameter or long range dependence in a time series, as described by Taqqu, Teverovsky, Willinger (1995).

The descriptive statistics of the estimated Hurst exponents are shown in tables 1–6 for, respectively: hourly log-returns, hourly squared log-returns, hourly log-volume, daily log-returns, daily squared log-returns, daily log-volume (the calculations were rounded to the fourth decimal place).

The ranges, standard deviations, skewness, and kurtosis of the Hurst exponents for these time series depend on the method applied.

Taking into account the mean and median of the Hurst exponent for the methods used, one can see that, for most of the methods applied, there is no indication of the existence of long memory. In most cases, we see an anti-persistence (median and mean are below 0.5, see Table 1). In these time series (called a mean-reverting series), an increase will most likely be followed by a decrease or vice-versa (i.e., values will tend to revert to a mean).

Table 1

Descriptive statistics of estimated Hurst exponents for hourly log-returns

	Aggregated variance method	Differenced aggregated variance method	Aggregated absolute value method	Higuchi's fractal dimension method	Peng's variance of residuals method	R/S Rescaled Range Statistic method	Geweke and Porter-Hudak method	Modified Geweke and Porter-Hudak method	Whittle estimator
Minimum	0.132	0.3427	0.2427	0.3273	0.2783	0.3655	–0.0094	0.1824	0.3853
First quartile	0.3757	0.5381	0.4859	0.4237	0.3922	0.5051	0.3745	0.3001	0.4482
Median	0.424	0.6413	0.5276	0.4792	0.4333	0.5377	0.4884	0.3497	0.4863
Mean	0.4186	0.6344	0.53	0.4756	0.4308	0.5553	0.455	0.341	0.4776
Third quartile	0.4962	0.7344	0.5863	0.5259	0.4739	0.6149	0.5433	0.3828	0.5058
Maximum	0.6202	0.9596	0.7441	0.6118	0.5436	0.7557	0.7201	0.4775	0.5963
Standard deviation	0.0982	0.1351	0.0977	0.0667	0.0576	0.0803	0.1445	0.0654	0.0433
Skewness	–0.6068	0.0318	–0.3795	0.0266	–0.4197	0.0835	–0.9016	-0.1955	–0.2408
Kurtosis	0.1073	–0.3758	0.3048	-0.7185	–0.2414	–0.3187	0.7081	-0.4962	–0.0149

Table 2

Descriptive statistics of estimated Hurst exponents for hourly squared log-returns

	Aggregated variance method	Differenced aggregated variance method	Aggregated absolute value method	Higuchi's fractal dimension method	Peng's variance of residuals method	R/S Rescaled Range Statistic method	Geweke and Porter-Hudak method	Modified Geweke and Porter-Hudak method	Whittle estimator
Minimum	–0.0168	0.0044	0.3601	0.7138	–0.0598	0.4143	0.3889	0.3164	0.4836
First quartile	0.2263	0.3998	0.5363	0.8776	0.3099	0.5286	0.5108	0.3935	0.5083
Median	0.3623	0.4976	0.6199	0.9196	0.4121	0.5657	0.5624	0.4382	0.5331
Mean	0.3572	0.4754	0.6299	0.9117	0.3874	0.5699	0.5755	0.4465	0.5587
Third quartile	0.4721	0.5756	0.733	0.9525	0.4799	0.6168	0.6062	0.499	0.5902
Maximum	0.7471	0.8505	0.9071	1.0155	0.7709	0.7345	0.9504	0.6975	0.7916
Standard deviation	0.1757	0.1806	0.1459	0.0605	0.1519	0.0645	0.1075	0.0786	0.0668
Skewness	–0.0104	–0.5201	0.1691	–0.7224	–0.4959	0.1013	1.068	0.706	1.324
Kurtosis	–0.644	0.318	–0.8625	0.6699	0.6398	–0.0279	1.6027	0.5491	1.4392

Table 3

Descriptive statistics of estimated Hurst exponents for hourly log-volume

	Aggregated variance method	Differenced aggregated variance method	Aggregated absolute value method	Higuchi's fractal dimension method	Peng's variance of residuals method	R/S Rescaled Range Statistic method	Geweke and Porter-Hudak method	Modified Geweke and Porter-Hudak method	Whittle estimator
Minimum	0.6661	0.5986	0.7652	0.8653	0.6052	0.393	0.713	0.5665	0.6508
First quartile	0.8154	0.9879	0.9342	0.9232	0.7539	0.7238	0.9614	0.7186	0.7141
Median	0.8667	1.1156	0.9915	0.966	0.7933	0.8224	1.0441	0.7701	0.7525
Mean	0.8687	1.0824	0.978	0.9509	0.8076	0.7948	1.0462	0.789	0.79
Third quartile	0.9396	1.1629	1.0413	0.9765	0.8613	0.8857	1.1372	0.8423	0.841
Maximum	1.0155	1.6047	1.1209	1.0244	1.4806	1.0489	1.5199	1.449	0.99
Standard deviation	0.0838	0.1823	0.0761	0.0357	0.1194	0.1268	0.1484	0.1252	0.1019
Skewness	–0.2211	–0.278	–0.822	–0.6678	3.3043	–0.773	0.2092	2.638	0.8139
Kurtosis	–0.8085	1.0608	0.1311	–0.3222	17.2019	0.828	0.769	12.329	–0.5253

Table 4

Descriptive statistics of estimated Hurst exponents for daily log-returns

	Aggregated variance method	Differenced aggregated variance method	Aggregated absolute value method	Higuchi's fractal dimension method	Peng's variance of residuals method	R/S Rescaled Range Statistic method	Geweke and Porter-Hudak method	Modified Geweke and Porter-Hudak method	Whittle estimator
Minimum	–0.2245	–0.4231	–0.0441	0.1214	0.1804	0.3224	–0.9465	0.0092	0.3189
First quartile	0.1464	0.3114	0.3009	0.2607	0.3774	0.4764	0.3057	0.2401	0.3921
Median	0.3337	0.5244	0.5199	0.3492	0.4346	0.5585	0.3901	0.335	0.4612
Mean	0.3197	0.5354	0.475	0.3648	0.4393	0.5888	0.3967	0.3289	0.453
Third quartile	0.4587	0.7164	0.61	0.4603	0.4899	0.6841	0.5714	0.4389	0.4915
Maximum	0.8123	1.4306	0.9931	0.6067	0.6702	0.8977	1.4013	0.7174	0.6283
Standard deviation	0.2354	0.329	0.2411	0.1342	0.0903	0.1419	0.3589	0.169	0.0739
Skewness	–0.1033	0.19	0.0118	0.3082	–0.0199	0.3216	–0.8187	0.087	0.3278
Kurtosis	–0.4201	0.9644	–0.5192	–1.037	0.228	–0.9212	3.6982	–0.4338	–0.5242

Table 5

Descriptive statistics of estimated Hurst exponents for daily squared log-returns

	Aggregated variance method	Differenced aggregated variance method	Aggregated absolute value method	Higuchi's fractal dimension method	Peng's variance of residuals method	R/S Rescaled Range Statistic method	Geweke and Porter-Hudak method	Modified Geweke and Porter-Hudak method	Whittle estimator
Minimum	–0.4449	-0.4979	–0.1711	0.7299	0.0841	0.3816	0.0086	0.2487	0.4281
First quartile	0.2148	0.1874	0.4627	0.8495	0.3464	0.5181	0.3417	0.396	0.5115
Median	0.3744	0.4516	0.6286	0.8973	0.4123	0.5812	0.5012	0.5315	0.5709
Mean	0.3614	0.4113	0.6032	0.9081	0.4203	0.5912	0.4893	0.5297	0.5814
Third quartile	0.5419	0.6176	0.8038	0.9728	0.5309	0.6686	0.6125	0.6378	0.6381
Maximum	0.8223	1.2728	1.0637	1.0629	0.7053	0.9243	1.059	0.898	0.7848
Standard deviation	0.2619	0.3694	0.2456	0.0781	0.1628	0.1186	0.2263	0.1568	0.0974
Skewness	–0.605	–0.0933	–0.6185	0.1325	–0.2627	0.4111	0.1673	0.3328	0.3488
Kurtosis	0.2836	–0.0233	0.3139	–0.7625	–0.4714	–0.0634	–0.2997	–0.8239	–0.5454

Table 6

Descriptive statistics of estimated Hurst exponents for daily log-volume

	Aggregated variance method	Differenced aggregated variance method	Aggregated absolute value method	Higuchi's fractal dimension method	Peng's variance of residuals method	R/S Rescaled Range Statistic method	Geweke and Porter-Hudak method	Modified Geweke and Porter-Hudak method	Whittle estimator
Minimum	0.3462	0.0239	0.5697	0.638	0.6165	0.6444	0.132	0.3007	0.6155
First quartile	0.5924	0.6343	0.8073	0.8681	0.738	0.7764	0.9851	0.5881	0.6834
Median	0.7021	0.9828	0.9001	0.9493	0.8371	0.8778	1.1313	0.6885	0.725
Mean	0.7064	0.9407	0.9094	0.9138	0.8149	0.9093	1.1273	0.6982	0.7358
Third quartile	0.8213	1.2306	1.0306	0.9757	0.8843	1.0433	1.336	0.8246	0.7778
Maximum	0.9522	0.7685	1.1703	1.0657	1.0611	1.3322	1.7634	1.067	0.9228
Standard deviation	0.1473	0.3627	0.1562	0.0966	0.1052	0.1759	0.3099	0.1615	0.0758
Skewness	–0.1872	–0.3411	–0.0523	–1.1565	–0.2103	0.4625	–0.6907	0.0063	0.5943
Kurtosis	–0.4756	–0.3537	–0.8755	0.5435	–0.6171	–0.7609	0.9337	–0.2941	–0.3566

This means that future values have a tendency to return to a long-term mean. The definition of long-range dependence assumes a stationary time series. In the case of non-stationarity, the estimators of the Hurst exponent can be greater than one (which happened in several computations included in our contribution).

Table 2 shows quite similar picture to Table 1. However, the values of mean and median are for hourly squared log-returns higher than for hourly log-returns. Some time series may reflect long memory.

In contrast to those from Table 1 and Table 2 for returns and squared returns means and medians of the Hurst exponent for log-volume (see Table 3) calculated by all these methods indicate the existence of long memory. Both skewness and kurtosis are considerable in size and depend on the estimation method used.

Generally, the means and medians (see tables 4, 5 and 6) are slightly lower than the respective values for hourly log-returns. We observed this in the case of short memory for the majority of the time series tested. Moreover, the skewness and kurtosis have the same order of magnitude.

The above distributions are similar to the respective distributions for hourly squared log-returns, both in terms of mean/median and skewness/kurtosis.

In the case of trading volume, there are the largest differences between distributions in corresponding time intervals. For hourly data, there were exceptionally high values of excess kurtosis (17.2019 and 12.329), for daily data no value is greater than 1. These values, therefore, could be seen as outliers. The means, medians, and skewness for both frequencies of data seem to be consistent.

As follows from tables 1–6, the distributions of empirical results are rather regular. There are absolute values of skewness greater than 1 in a few cases only. The same statement holds true with respect to absolute values of kurtosis greater than 1 (with the exception of the above-mentioned outliers). The hypothesis about the normal distribution of estimated values should be checked by normality tests.

In the next step, we investigate this property in a formal way using the six most popular normality tests (the Anderson-Darling, Cramer von Mises, Kolmogorov-Smirnov, Pearson chi-square, Shapiro-Francia, and Shapiro-Wilk tests). The calculations were conducted in R software with the additional package "nortest."

Table 7 shows the results of normality tests for Hurst exponents estimated by each method in hourly log-returns. Similar computations were conducted for hourly squared log-returns, hourly log-volume, daily log-returns, daily squared log-returns, and daily log-volume (to save the space, they are available to the reader upon request).

These normality tests confirmed previous assumptions about the normal distributions of Hurst exponent estimates. Table 8 summarizes the results of these investigations – in each cell is the number of normality tests which rejected the hypothesis about normal distribution at significance level $\alpha = 0.05$.

Table 7

Normality tests for estimated Hurst exponents for hourly log-returns (p-values)

	Aggregated variance method	Differenced aggregated variance method	Aggregated absolute value method	Higuchi's fractal dimension method	Peng's variance of residuals method	R/S Rescaled Range Statistic method	Geweke and Porter-Hudak method	Modified Geweke and Porter-Hudak method	Whittle estimator
Anderson-Darling	0.113	0.9447	0.6626	0.9329	0.6623	0.2671	0.0046	0.7538	0.0018
Cramer von Mises	0.144	0.9303	0.6058	0.9467	0.6754	0.176	0.0024	0.6517	0.0168
Kolmogorov-Smirnov	0.2371	0.9567	0.4726	0.9848	0.656	0.1856	0.0014	0.6798	0.1246
Pearson chi-square	0.0693	0.4882	0.1684	0.9519	0.6743	0.4454	0.0901	0.813	0.053
Shapiro-Francia	0.141	0.9668	0.7066	0.8731	0.6011	0.6787	0.0094	0.8678	0.0546
Shapiro-Wilk	0.1026	0.9633	0.4611	0.9479	0.5459	0.5354	0.0092	0.8727	0.0463

Table 8

Summary of results of normality tests

	Aggregated variance method	Differenced aggregated variance method	Aggregated absolute value method	Higuchi's fractal dimension method	Peng's variance of residuals method	R/S Rescaled Range Statistic method	Geweke and Porter-Hudak method	Modified Geweke and Porter-Hudak method	Whittle estimator
Hourly log-returns	0	0	0	0	0	0	5	0	3
Hourly squared log-returns	0	0	0	1	1	0	5	0	6
Hourly log-volume	0	1	6	6	6	1	1	5	6
Daily log-returns	0	0	0	1	0	0	6	0	0
Daily squared log-returns	0	0	0	0	0	0	0	0	0
Daily log-volume	0	0	0	6	2	0	0	0	3

As follows from the table shown above, the hypothesis about normality – in general – could not be rejected. The biggest deviation from a normal distribution is represented by hourly log-volume. The list of estimation methods that generate results which are far away from a normal distribution includes: the Whittle estimator, the Geweke and Porter-Hudak method, and Higuchi's fractal dimension method.

Example tabulation schemes of the estimated Hurst exponents for hourly log-returns and hourly log-volume are presented in Table 9. Similar schemes of the estimated Hurst exponents (to save the space they are available to the reader upon request) were done for hourly squared log-returns, daily log-returns, daily squared log-returns, and daily log-volume. These calculations gave us an insight into the long memory property of the time series under study.

Table 9

Tabulation schemes for hourly log-returns and hourly log-volume *

	H < 0	**[0; 0.49]**	**(0.49; 0.51)**	**[0.51; 1]**	**H > 1**
AUS.IDX	0 (0)	5 (0)	1 (0)	3 (8)	0 (1)
CHE.IDX	0 (0)	1 (0)	3 (0)	5 (9)	0 (0)
DEU.IDX	0 (0)	2 (0)	2 (0)	5 (9)	0 (0)
FRA.IDX	0 (0)	5 (1)	1 (0)	3 (4)	0 (4)
GBR.IDX	0 (0)	3 (0)	2 (0)	4 (9)	0 (0)
JPN.IDX	0 (0)	1 (0)	3 (0)	5 (9)	0 (0)
NLD.IDX	0 (0)	5 (0)	2 (0)	2 (8)	0 (1)
USA.IDX	0 (0)	4 (1)	1 (0)	4 (8)	0 (0)
BMW.DEU	0 (0)	6 (0)	0 (0)	3 (6)	0 (3)
COMM.DEU	0 (0)	4 (0)	1 (0)	4 (7)	0 (2)
DEBK.DEU	0 (0)	3 (0)	0 (0)	6 (6)	0 (3)
EON.DEU	0 (0)	4 (0)	2 (0)	3 (6)	0 (3)
BNPP.FRA	0 (0)	7 (0)	0 (0)	2 (6)	0 (3)
LVHM.FRA	0 (0)	3 (0)	2 (0)	4 (6)	0 (3)
SANO.FRA	0 (0)	4 (0)	1 (0)	4 (6)	0 (3)

Table 9 cont.

	H < 0	[0; 0.49]	(0.49; 0.51)	[0.51; 1]	H > 1
TOTAFRA	0 (0)	5 (0)	2 (0)	2 (6)	0 (3)
CSG.CHE	0 (0)	6 (0)	1 (0)	2 (6)	0 (3)
NEST.CHE	0 (0)	7 (0)	0 (0)	2 (6)	0 (3)
NOVA.CHE	0 (0)	7 (0)	0 (0)	2 (6)	0 (3)
ROCH.CHE	0 (0)	7 (0)	0 (0)	2 (6)	0 (3)
UBS.CHE	0 (0)	8 (0)	0 (0)	1 (7)	0 (2)
BHPP.GBR	0 (0)	4 (0)	1 (0)	4 (7)	0 (2)
BP.GBR	0 (0)	5 (0)	2 (0)	2 (7)	0 (2)
HSBC.GBR	0 (0)	2 (0)	1 (0)	6 (7)	0 (2)
RIO.GBR	0 (0)	6 (0)	0 (0)	3 (7)	0 (2)
VOD.GBR	0 (0)	7 (0)	0 (0)	2 (9)	0 (0)
AMAZ.USA	0 (0)	4 (0)	1 (0)	4 (7)	0 (2)
ATT.USA	0 (0)	4 (0)	0 (0)	5 (8)	0 (1)
BOA.USA	0 (0)	4 (0)	2 (0)	3 (8)	0 (1)
CHEV.USA	0 (0)	4 (0)	3 (0)	2 (7)	0 (2)
CISC.USA	0 (0)	5 (0)	2 (0)	2 (9)	0 (0)
COPA.USA	0 (0)	5 (0)	1 (0)	3 (7)	0 (2)
DISN.USA	0 (0)	6 (0)	0 (0)	3 (7)	0 (2)
EBAY.USA	0 (0)	6 (0)	0 (0)	3 (6)	0 (3)
EXXO.USA	0 (0)	4 (0)	2 (0)	3 (8)	0 (1)
GEEL.USA	0 (0)	5 (0)	1 (0)	3 (8)	0 (1)
GEMO.USA	0 (0)	4 (0)	1 (0)	4 (7)	0 (2)
HEPA.USA	1 (0)	5 (0)	2 (0)	1 (6)	0 (3)
HOME.USA	0 (0)	4 (0)	2 (0)	3 (5)	0 (4)
IBM.USA	0 (0)	5 (0)	0 (0)	4 (7)	0 (2)

Table 9 cont.

INTC.USA	0 (0)	3 (0)	0 (0)	6 (9)	0 (0)
JOJO.USA	0 (0)	5 (0)	0 (0)	4 (7)	0 (2)
JPMC.USA	0 (0)	5 (0)	1 (0)	3 (9)	0 (0)
MCDN.USA	0 (0)	4 (0)	1 (0)	4 (2)	0 (7)
MSFT.USA	0 (0)	5 (0)	2 (0)	2 (8)	0 (1)
ORCL.USA	0 (0)	7 (0)	1 (0)	1 (6)	0 (3)
PHMO.USA	0 (0)	3 (0)	1 (0)	5 (6)	0 (3)
PRGA.USA	0 (0)	5 (0)	1 (0)	3 (7)	0 (2)
STAR.USA	0 (0)	3 (0)	1 (0)	5 (7)	0 (2)
WMS.USA	0 (0)	6 (0)	0 (0)	3 (7)	0 (2)
YHOO.USA	0 (0)	6 (0)	0 (0)	3 (6)	0 (3)

* The numbers concerning log-volume are given in parentheses.

A summary of the tabulation schemes is shown in Table 10. In each cell, there is the number of time series (from the 51 time series used in this study) with long memory as determined by a particular number of methods (which are at the head of each column).

Table 10

Long memory property in considered time series

Number of methods	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Hourly log-returns	0	3	12	16	11	6	3	0	0	0
Hourly squared log-returns	0	0	3	3	9	18	12	1	2	3
Hourly log-volume	0	0	1	0	1	1	16	16	8	8
Daily log-returns	5	7	11	8	13	5	1	1	0	0
Daily squared log-returns	0	1	4	1	14	15	9	3	4	0
Daily log-volume	0	0	0	0	2	9	10	17	10	3

As can be seen from Table10, long memory seems to be invariant with respect to the frequency of data – hourly and daily series have similar properties. Certainly, the strongest long memory is seen in hourly and daily log-volume. Hourly and daily squared returns exhibit weaker long memory. Finally, hourly and daily log-returns do not exhibit long memory at all. By and large, these observations are in line with previous results reported in other contributions.

As a last step, we investigated whether the frequency of data affects the probability distributions of the estimated Hurst exponents. Table 11 presents the p-values of a two-sample Kolmogorow-Smirnov test that compares the distributions of the estimated Hurst exponents for, respectively: hourly log-returns versus daily log-returns, hourly squared log-returns versus daily squared log-returns, and hourly log-volume versus daily log-volume.

Table 11

A comparison of distributions using a two-sample Kolmogorov-Smirnov test (*p*-values)

	Aggregated variance method	**Differenced aggregated variance method**	**Aggregated absolute value method**	**Higuchi's fractal dimension method**	**Peng's variance of residuals method**	**R/S Rescaled Range Statistic method**	**Geweke and Porter-Hudak method**	**Modified Geweke and Porter-Hudak method**	**Whittle estimator**
Log-returns	0.0035	0.0032	0.0069	0	0.4051	0.0243	0.0238	0.119	0.0728
Squared log-returns	0.5616	0.0725	0.557	0.1863	0.4084	0.1863	0.0017	0.0004	0.0429
Log-volume	0	0.0069	0.0017	0.0728	0.1873	0.0069	0.0035	0.0035	0.0728

At a significance level $\alpha = 0.05$, the distributions of the estimated Hurst exponents for hourly and daily log-returns are statistically different for six of the nine methods of the estimation of the Hurst exponent. Analogously, for squared log-returns this property is satisfied by three methods. Finally, for log-volume, this property is satisfied by six methods. We can claim, therefore, that the probability distributions of the estimated Hurst exponent depend – to some extent – on the time frequency chosen.

5. Conclusions

The main goal of this study was to detect the long memory in the financial time series (returns, volatility, and trading volume) and to identify the impact of data frequency on the long memory properties of the financial time series. In addition, we tried to compare "consistency" of results derived by nine methods for the estimation of Hurst exponents (long memory).

The computations were conducted for 51 time series of log-returns, squared log-returns, and log-volume by nine methods with respect to hourly and daily data.

The bases of this comparison were the descriptive statistics of the estimated (by nine methods) Hurst exponents, probability distributions, and the impact of data frequency on the results.

Irrespective of data frequency and company or market index, long memory could not be detected in the log-returns. However, the size of the Hurst exponents of log-returns seems to rise with data frequency. In addition, the probability distributions of the Hurst exponents for hourly and daily data are mostly significantly different. In squared log-returns (a proxy for volatility), most of the methods used indicate the existence of long memory. By and large, in the case of squared log-returns, the frequency of data with respect to long memory does not matter. Also, differences in distributions for hourly and daily data are less pronounced than in the case of log-returns.

All of the methods used indicate significant long memory in log-volume data irrespective of its frequency. In addition, the distributions of the estimates of Hurst exponents depend to a large extent on the frequency of the time series data.

References

[1] Agiakloglou, C., Newbold, P. and Wohar, M. (1993) 'Bias in an estimator of the fractional difference parameter', *Journal of Time Series Analysis,* vol. 14, pp. 235–246.

[2] An, S. and Bloomfield, P. (1993) Cox and Reid's modification in regression models with correlated errors, Technical Report, North Carolina State University.

[3] Andrews, D.W.K. and Guggenberger, P. (2003) 'A bias-reduced log-periodogram regression estimator for the long memory parameter', *Econometrica,* vol. 71, pp. 675–712.

[4] Andrews, D.W.K. and Sun, Y. (2004) 'Adaptive local polynomial Whittle estimation of long-range dependence', *Econometrica,* vol. 72, pp. 569–614.
[5] Aydogan, K. and Booth, G.G. (1988) 'Are there long cycles in common stock returns?', *Southern Economic Journal*, vol. 55, pp. 141–149.
[6] Baillie, R.T. (1996) 'Long memory processes and fractional integration in econometrics', *Journal of Econometrics*, vol. 73, pp. 5–59.
[7] Barkoulas, J.T. and Baum, C.F. (1996) 'Long term dependence in stock returns', *Economics Letters*, vol. 53, pp. 253–259.
[8] Barkoulas, J.T, Labys, W.C. and Onochie, J. (1997) 'Fractional dynamics in international commodity prices', *Journal of Futures Markets*, vol. 17, pp. 161–189.
[9] Beran, J.A. (1992) 'Statistical methods for data with long-range dependence', *Statistical Science*, vol. 7, pp. 404–427.
[10] Beran, J.A. (1994) Statistics for Long memory Processes, London: Chapman and Hall.
[11] Beveridge, S. and Oickle, C. (1997) 'Long memory in the Canadian stock market', *Applied Financial Economics*, pp. 667–672.
[12] Bollerslev, T. and Jubinski, D. (1999) 'Equity trading volume volatility: latent information arrivals and common long-run dependencies', *Journal of Business & Economic Statistics,* vol. 17, pp. 9–21.
[13] Bollerslev, T. and Mikkelsen, H.O. (1996) 'Modeling and pricing long memory in stock market volatility', *Journal of Econometrics,* vol. 73, pp. 151–184.
[14] Booth, G.G., Kaen, F.R. and Koveos, P.E. (1982) 'R/S analysis of foreign exchange markets under two international monetary regimes', *Journal of Monetary Economics,* vol. 10, pp. 407–415.
[15] Cheung, Y.W. and Lai, K.S. (1993) 'Do gold market returns have long memory?', *Financial Review*, vol. 28, pp. 181–202.
[16] Cheung, Y.W. and Lai, K.S. (1995) 'A search for long memory in international stock market returns', *Journal of International Money and Finance*, vol. 14, pp. 597–615.
[17] Cheung, Y.W., Lai, K. and Lai, M. (1993) 'Are there long cycles in foreign stock returns?', *Journal of International Financial Markets, Institutions and Money*, vol. 3, pp. 33–47.
[18] Clark, P.K. (1973) 'A subordinated stochastic process model with finite variance for speculative prices', *Econometrica*, vol. 41, pp. 135–155.
[19] Copeland, T. (1976) 'A model of asset trading under the assumption of sequential information arrival', *Journal of Finance*, vol. 31, pp. 135–155.
[20] Cox, D.R. and Reid, N. (1987) 'Parameter orthogonality and approximate conditional inference (with discussion)', *Journal of the Royal Statistical Society Series B*, vol. 49, pp. 1–39.

[21] Crato, N. (1994) 'Some international evidence regarding the stochastic memory of stock returns', *Applied Financial Economics*, vol. 4, pp. 33–39.
[22] Darrat, A.F., Rahman, S. and Zhong, M. (2003) 'Intraday trading volume and return volatility of the DJIA stocks: A note', *Journal of Banking and Finance*, vol. 27, pp. 2035–2043.
[23] Doornik, J.A. and Ooms, M. (2003) 'Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models', *Computational Statistics and Data Analysis*, vol. 42, pp. 333–348.
[24] Fama, E.F. (1970) 'Efficient capital markets: A review of theory and empirical work', *Journal of Finance,* vol. 25, pp. 383–417.
[25] Fang, H., Lai, K. and Lai, M. (1994) 'Fractal structure in currency futures prices', *Journal of Futures Markets*, vol. 14, pp. 169–181.
[26] Geweke, J. and Porter-Hudak, S. (1983) 'The estimation and application of long memory time series models', *Journal of Time Series Analysis*, vol. 4, pp. 221–238.
[27] Granger, C.W.J. and Joyeux, R. (1980) 'An introduction to long memory time series models and fractional differencing', *Journal of Time Series Analysis*, vol. 1, pp. 15–29.
[28] Granger, C.W.J. and Zhuanxin, D. (1996) 'Varieties of long memory models', *Journal of Econometrics*, vol. 73, pp. 61–67.
[29] Greene, M.T. and Fielitz, B.D. (1977) 'Long-term dependence in common stock returns', *Journal of Financial Economics*, vol. 5, pp. 339–349.
[30] Helms, B.P., Kaen, F.R. and Rosenman, R.E. (1984) 'Memory in commodity futures contracts', *Journal of Futures Market*. vol. 4, pp. 559–567.
[31] Henry, M. and Robinson, P.M. (1996) 'Bandwidth choice in Gaussian semiparametric estimation of long range dependence', in Robinson, P.M. and Rosenblatt, M. (eds) Athens Conference on Applied Probability and Time Series Analysis, Volume II: Time Series Analysis, In Memory of E. J. Hannan, New York, Springer, pp. 220–232.
[32] Hiemstra, C. and Jones, J.D. (1997) 'Another look at long memory in common stock returns', *Journal of Empirical Finance*, vol. 4, pp. 373–401.
[33] Hosking, J.R.M. (1981) 'Fractional differencing', *Biometrika*, vol. 68, pp. 165–176.
[34] Hurst, H.R. (1951) 'Long-term storage capacity of reservoirs', *Transactions of the American Society of Civil Engineers*, vol. 1, pp. 519–543.
[35] Hurvich, C.M., Deo, R.S. and Brodsky, J. (1998) 'The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long memory time series', *Journal of Time Series Analysis,* vol. 19, pp. 19–46.
[36] Kim, C. S. and Philips, P.C.B. (1999) Log periodogram regression in the nonstationary case, *Mimeo,* Yale University.

[37] Koop, G., Ley, E., Osiewalski, J. and Steel, M.F.J. (1997) 'Bayesian analysis of long memory and persistence using ARFIMA models', *Journal of Econometrics*, vol. 76, pp. 149–169.

[38] Künsch, H.R. (1987) 'Statistical aspects of self-similar processes', in Prokhorov, Y. and Sazanov, V.V. (eds), Proceedings of the First World Congress of the Bernoulli Society. Utrecht, VNU Science Press, pp. 67–74.

[39] Lo, A.W. (1991) 'Long-term memory in stock market prices', *Econometrica*, vol. 59, pp. 1279–1313.

[40] Lobato, I.N. (1999) 'A semiparametric two-step estimator in a multivariate long memory model', *Journal of Econometrics,* vol. 90, pp. 129–153.

[41] Lobato, I.N. and Velasco, C. (2000) 'Long memory in stock-market trading volume', *Journal of Business & Economic Statistics*, vol. 18 (4), pp. 410–427.

[42] Mandelbrot, B.B. (1971) 'When can a price be arbitraged efficiently? A limit to the validity of the random walk and martingale models', *Review of Economics and Statistics*, vol. 53, pp. 225–236.

[43] Mandelbrot, B.B. and Van Ness J.W. (1968) 'Fractional Brownian Motion, Fractional Noises and Applications', *SIAM Review*, vol. 10 (4), pp. 422–437.

[44] Nielsen, M.Ø. (2004) 'Efficient likelihood inference in nonstationary univariate models, *Econometric Theory,* vol. 20, pp. 116–146.

[45] Palma W. (2007) Long memory time series. Theory and methods, New York: Wiley Interscience.

[46] Peng C.K., Buldyrev S.V., Simons M., Stanley H.E. and Goldberger A.I. (1994) 'Mosaic organization of DNA nucleotides', *Physical Review E*, vol. 49, pp. 1685–1698.

[47] Phillips, P.C.B. and Shimotsu, K. (2004) 'Local Whittle estimation in nonstationary and unit root cases', *Annals of Statistics,* vol. 34 (2), pp. 656–692.

[48] Robinson, P.M. (1995a) 'Log-periodogram regression of time series with long range dependence', *Annals of Statistics,* vol. 23, pp. 1048–1072.

[49] Robinson, P.M. (1995b) 'Gaussian semiparametric estimation of long range dependence', *Annals of Statistics,* vol. 23, pp. 1630–1661.

[50] Shimotsu, K. and Phillips, P.C.B. (2002a) 'Pooled log periodogram regression', *Journal of Time Series Analysis,* vol. 23, pp. 57–93.

[51] Shimotsu, K. and Phillips, P.C.B. (2002b) 'Exact local Whittle estimation of fractional integration', *Cowles Foundation Discussion Paper* 1367.

[52] Sowell, F.B. (1992) 'Maximum likelihood estimation of stationary univariate fractionally integrated time series models', *Journal of Econometrics,* vol. 53, pp. 165–188.

[53] Tanaka, K. (1999) 'The nonstationary fractional unit root', *Econometric Theory,* vol. 15, pp. 549–582.

[54] Taqqu M.S., Teverovsky V. and Willinger W. (1995) 'Estimators for long-range dependence: an empirical study' *Fractals*, vol. 3(4), pp. 785–798.
[55] Velasco, C. (1999a) 'Non-stationary log-periodogram regression', *Journal of Econometrics,* vol. 91, pp. 325–371.
[56] Velasco, C. (1999b) 'Gaussian semiparametric estimation of non-stationary time series', *Journal of Time Series Analysis,* vol. 20, pp. 87–127.