Methods of Non-Extensive Statistical Physics in Analysis of Price Returns on Polish Stock Market

Ł. Bil[a,b], D. Grech[b] and E. Podhajska[c]

[a]Faculty of Physics, University of Warsaw, L. Pasteura 5, 02-093 Warsaw, Poland

[b]Institute of Theoretical Physics, University of Wrocław, pl. M. Borna 9, 50-204 Wrocław, Poland

[c]Faculty of Life Sciences and Technology, Wrocław University of Environmental and Life Sciences,
pl. Grunwaldzki 24A, 50-363 Wrocław, Poland

We use methods of non-extensive statistical physics to describe quantitatively the memory effect involved in returns of companies from WIG 30 index on the Warsaw Stock Exchange. The entropic approach based on the generalization of the Boltzmann–Gibbs entropy to non-additive Tsallis q-entropy is applied to fit fat tailed distribution of returns to q-normal (Tsallis) distribution. The existence of long term memory effects in price returns generated by two-point autocorrelations are checked via calculation of the Hurst exponent within detrended fluctuation analysis approach. The results are collected for diversified frequency of data sampling. We confirm the perfect inverse cubic power law for low time-lags (≈ 1 min) of returns for the main WIG 30 index as well as for the most of separate stocks, however this relationship does not hold for longer time-lags. The particular emphasis is given to a study of an independent fit of probability distribution of positive and negative returns to q-normal distribution. We discuss in this context the asymmetry between tails in terms of the Tsallis parameters $q^{\pm}$. A qualitative and quantitative relationship between the frequency of data sampling, the parameters q and $q^{\pm}$, and the corresponding main Hurst exponent H is provided to analyze the effect of memory in data caused by linear and nonlinear autocorrelations. A new quantifier based on asymmetry of the Tsallis index instead of skewness of distribution is proposed which we believe is able to describe the stage of market development and its robustness to speculation.

1. Introduction

Among variety of problems involved in study of financial markets, the presence of fat tails in distributions of returns and their power law features [1–7] seem to be crucial. The well known century-old Gaussian model [8] simply underestimates the probability of large events and therefore cannot be used as a proper model in financial risk analysis. Events such as the 1987 stock market crash where the Standard&Poor 500 index — the leading on US stock market — dropped by a magnitude of over 20 standard deviations and many other crashes clearly observed in long stock market history all over the world, would have never occurred if probability distribution of financial returns was of the Gaussian nature. Although the origin of fat tails is not entirely known, it is agreed that their source must lie in memory effects in data but of more complicated origin than non-vanishing simple two-point autocorrelation function. It is well established fact that nonlinear correlations between returns visible as volatility clustering (i.e., volatility correlations seen in the simplest case as autocorrelations between absolute returns) generate fat distribution tails [9–11]. Also multifractal properties of data in series are likely to produce similar results [12]. Hence, the quantitative description of shape of tails in distribution of financial data may serve as important global quantifier identifying much more complex nonlinear phenomena which are difficult for simultaneous detection and identification in complex financial system with other tools. This is why it is interesting to look for the statistical features of financial system in order to find the reflection of much more complicated intrinsic varied organization of financial complexity in these statistical features.

2. Data analysis

The well-known fact connected with fat tails of return distribution in finance is the so-called inverse cubic power law [13–15]. It concerns the cumulated probability that for given time interval (time-lag) Δt the probability $P(|r| > x)$ that a financial return r is in absolute value larger than x has the power law form

$$P(|r| > x) \sim x^{-\zeta} \tag{1}$$

with $\zeta \approx 3$. This exponent is claimed [14, 15] to be almost independent of Δt, size of stock and period of measurement. It is shown also to hold for different stock market indices [7, 13–17] across different sizes of stocks, in different time periods and for time-lags in returns ranging from 1 min to even one month. Its accuracy has been reported over as many as 80 standard deviations for some stock markets and applies as well to commodity market [18] and to traded currency exchange rates [19, 20]. Note that within this paper we will use the standard definition of return $R(t, \Delta t)$ calculated for the time-lag Δt:

$$R(t, \Delta t) = \frac{p(t) - p(t - \Delta t)}{p(t - \Delta t)}. \tag{2}$$

The statistically different stocks can be compared together only if the normalization procedure is applied to returns, i.e., the normalized and centered returns $r(t, \Delta t)$ are considered

$$r(t, \Delta t) = \frac{R(t, \Delta t) - \langle R(t, \Delta t) \rangle_T}{\sigma_T}, \quad (3)$$

where $\langle R(t, \Delta t) \rangle_T$ is the mean value of R over the time window T and σ_T is its standard deviation ($\sigma_T^2 = \langle R^2 \rangle_T - \langle R \rangle_T^2$).

A formalism grounded on statistical physics that appears to be attractively compact and economic for describing the inverse cubic law, other cases of fat tails of probability distribution as well as the Gaussian distributions, is the one based on the generalized non-extensive entropy proposed by Tsallis [21, 22]. This formalism leads to equilibrium distributions of *power law* type instead of *exponential* ones typical for extensive systems and governed by normal distribution. Indeed, the optimization of non-additive Tsallis entropy yields, for the stationary state, the following distribution of probabilities called *q*-normal or Tsallis distribution [22]:

$$P(x) = N_q [1 + B_q x^2 (q-1)]^{\frac{1}{(1-q)}}, \quad (4)$$

where

$$B_q = [(5 - 3q)\sigma^2]^{-1} \quad (5)$$

and σ^2 is the variance of data. The normalization constant N_q for $1 < q < 3$ (this case is the most interesting since it develops for large $|x|$ the power law form of distribution tails) is expressed by the Gamma–Euler function Γ as

$$N_q = \frac{\Gamma\left(\frac{1}{q-1}\right)}{\Gamma\left(\frac{3-q}{2q-2}\right)} \sqrt{\frac{q-1}{\pi} B_q}. \quad (6)$$

Note that Eq. (4) develops for $|x| \to \infty$ the power law form $P(x) \sim x^{2/(1-q)}$ or equivalently at the level of cumulative distribution leads to

$$P(|r| > x) \sim x^{-\frac{3-q}{q-1}}, \quad (7)$$

which coincides with Eq. (1) and recovers the inverse cubic law for $q = 3/2$. On the other hand, the limit $q \to 1$ refers to the additive entropy case (the Boltzmann–Gibbs entropy) which manifests as the well known normal distribution case [22] at the level of optimization of this entropy. The *q*-normal distribution and in particular the value of non-extensive q parameter will be therefore in focus of our interest while making a fit to real financial data. It should be stressed at this point that the meaning, value and behavior of the non-extensive exponent, or the Tsallis parameter q is still an open question in non-extensive thermostatistics studies. In particular its relation with memory effects in signal still requires explanation. When moved into finance, this problem seems to be even more interesting and intriguing.

We examined normalized and centered returns (see Eq. (3)) of most developed companies in Poland quoted on the Warsaw Stock Exchange which make the content of WIG 30 index. The statistics of these returns was found from the very recent two years intraday and interday quotations* in the period March 27, 2013–March 31, 2015. First, all obvious artifacts in data connected with periods of no quotes have been carefully removed. The examined period counted finally $T \approx (7.6-18) \times 10^4$ data points (depending on particular stock) for the time-lag $\Delta t = 1$ min. Then we constructed, separately for each company, the normalized and centered returns $r(t)$ calculated for all other discussed time-lags Δt, i.e., the intraday returns for $\Delta t = 5$–60 min and the interday returns for $\Delta t = 1-4$ days, all in moving time window every 1 min. The final results for various companies have been found qualitatively identical and quantitatively very close together, therefore we present them here only for chosen companies. We focused on good representative examples of different economic branches in Poland: banking (PKO BP), telecommunication (Orange), fuel and energy sector (PKN Orlen) and insurance sector (PZU). Our findings were then compared with the results of capital weighted stock index WIG 30 which can be considered as the economic "reference frame", since it accommodates full variety of stocks and therefore should be more robust to speculation made by investors within just one branch.

3. Results and concluding remarks

Figures 1–4 show the quality of fit to the Tsallis distribution for returns of companies from WIG 30 for variety of time-lags ranging from 1 min to 4 days. The log-linear scale is applied to reveal all data details more exactly.
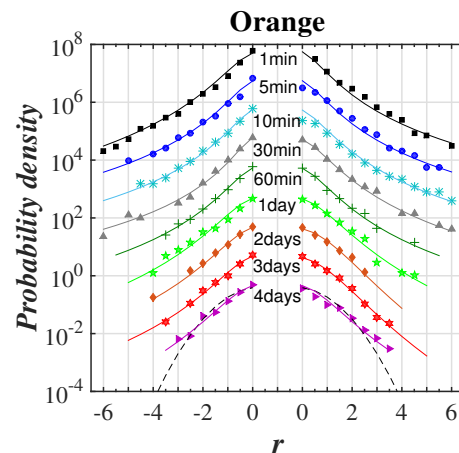

Fig. 1. Probability density function in log–linear scale of normalized and centered returns for Orange stock from WIG 30 index calculated for various time-lags Δt from the period: March 2013–March 2015. The plots for different Δt were drawn as dots and are relatively artificially rescaled vertically 10 times for better display. The dashed line corresponds to Gaussian distribution while solid lines present the best fit of *q*-normal Tsallis distribution done separately to positive and negative returns.
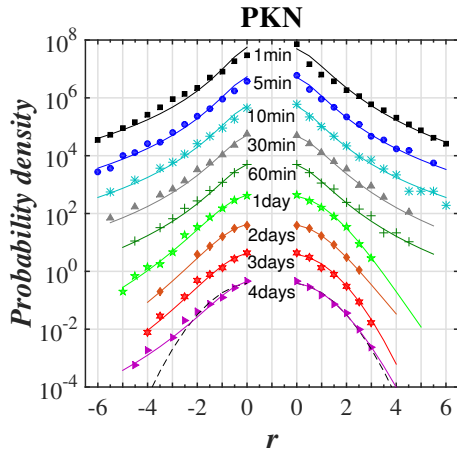

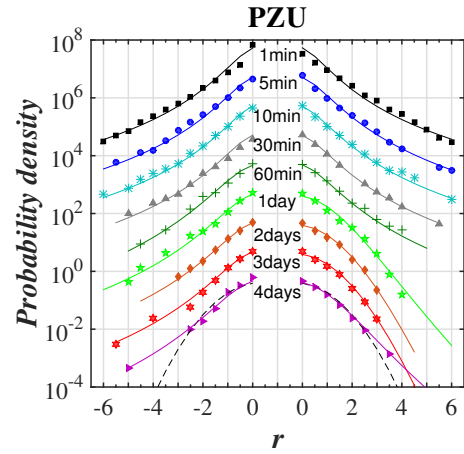Fig. 2.   The same as in Fig. 1 for PKN Orlen stock.

Fig. 4.   The same as in Fig. 1 for PZU stock.
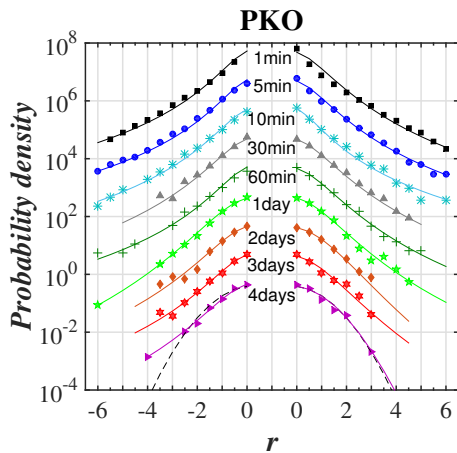

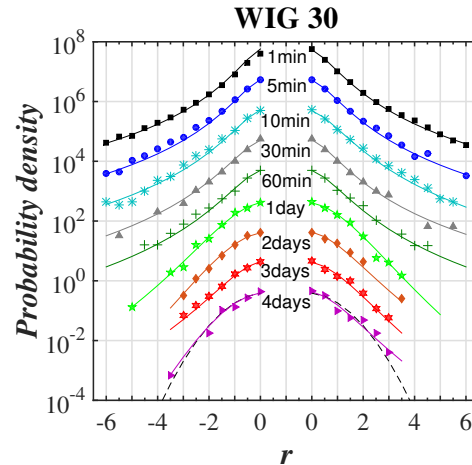Fig. 3.   The same as in Fig. 1 for PKO BP stock.

Fig. 5.   The same as in Fig. 1 for the capital weighted WIG 30 index.

Figure 5 shows also for comparison how such a fit works for returns of WIG 30 index. The best fit was obtained here as the minimal squared distance of the q-normal distribution profile and the real data distribution (in log scale). However we did not stop fitting data to symmetric Tsallis distribution. All fits have been done twice. First we made them for all returns which gave us the best fit q value for all returns independently of their sign. Then, the same was performed separately for positive and negative returns. We have introduced then the respective q^- and q^+ values of the Tsallis index for the left and right distribution tail. The latter case is actually shown in plots in Figs. 1–5. This enables to register in quantitative way the presence of asymmetry between heavy tails. Our approach differs from the one based simply on skewness of probability distribution, since we consider our scheme as more indicative for the different form of far ($|x| \to \infty$) left and right tails. This asymmetry was not so far a subject of deep study in econophysics (see, however [20]).

All the fitting results have been summarized for comparison in Table I. The Tsallis index has also been shown once more for better visualization as a function of the

time-lag Δt — in Fig. 6 for separate companies and in Fig. 7 for WIG 30 data.

The above findings have been confronted with the results of two-point linear autocorrelation function of returns $C_2(s) = \langle r_i r_{i+s} \rangle$ with time-lag s in order to reveal any similarities if present. The average is taken over points in considered time window. Recall that in a case of long term memory caused by linear correlation in data, $C_2(s)$ is mainly shaped according to the analytically proven power law: $C_2(s) \sim H(2H-1)s^{2H-2}$ [23] by the Hurst exponent $0 \le H \le 1$ [24]. The provided power law form explains that for $H = 1/2$ returns are uncorrelated, once other values of H correspond respectively to linearly anticorrelated ($0 < H < 1/2$) or linearly autocorrelated ($1/2 < H < 1$) data. We used DFA technique [25] to calculate values of H exponent. This method seems to be most reliable at present comparing with other available in literature. The H values in Figs. 7, 8 have been calculated with high precision $\delta H \sim 0.01$. The outcomes of this calculation are shown as a function of time-lags in Fig. 7 (bottom part) for WIG 30 data and in the following Fig. 8 for ingredient examined stocks.

TABLE I

Results of q-normal distribution fit to statistics of returns for chosen stocks within WIG 30 index. Shown are also results of independent fit of Tsallis parameter $q^{\pm}$ to the right and left tails of probabilty distribution.

Δt	Tsallis index	OrangePL	PKN Orlen	PKO BP	PZU	WIG 30
1 min	q^-	1.42	1.50	1.46	1.46	1.51
	q	1.45	1.45	1.41	1.46	1.49
	q^+	1.49	1.41	1.38	1.47	1.48
5 min	q^-	1.48	1.47	1.47	1.45	1.48
	q	1.48	1.44	1.43	1.44	1.47
	q^+	1.49	1.44	1.40	1.43	1.46
10 min	q^-	1.50	1.46	1.44	1.46	1.46
	q	1.50	1.46	1.46	1.46	1.44
	q^+	1.47	1.46	1.49	1.44	1.41
30 min	q^-	1.54	1.44	1.39	1.44	1.43
	q	1.47	1.38	1.37	1.43	1.45
	q^+	1.45	1.38	1.36	1.43	1.46
60 min	q^-	1.47	1.42	1.44	1.35	1.41
	q	1.47	1.42	1.42	1.40	1.41
	q^+	1.45	1.39	1.34	1.40	1.41
1 day	q^-	1.38	1.26	1.26	1.37	1.19
	q	1.35	1.25	1.27	1.35	1.18
	q^+	1.33	1.07	1.28	1.10	1.15
2 days	q^-	1.37	1.22	1.33	1.38	1.13
	q	1.32	1.20	1.24	1.33	1.20
	q^+	1.20	1.11	1.17	1.03	1.23
3 days	q^-	1.38	1.23	1.38	1.38	1.28
	q	1.27	1.20	1.26	1.33	1.25
	q^+	1.21	1.01	1.23	1.03	1.20
4 days	q^-	1.33	1.30	1.33	1.33	1.06
	q	1.25	1.29	1.28	1.29	1.20
	q^+	1.16	1.03	1.02	1.16	1.24

TABLE II

Relative asymmetry ratio $\frac{|\delta q|}{q}$ of Tsallis distribution to intraday stock data on Polish market calculated from fitting results shown in Table I. The averaged relative asymmetry ratio $\langle \frac{|\delta q|}{q} \rangle$ taken over all intraday time-lags from 1 min to 60 min is also shown for individual representative components of WIG 30.

	Δt	OrangePL	PKN Orlen	PKO BP	PZU	WIG 30
$\frac{|\delta q|}{q}$	1 min	4.8%	6.2%	5.7%	0.7%	2.0%
	5 min	0.7%	2.1%	4.9%	1.4%	1.4%
	10 min	2.0%	0.0%	3.4%	1.4%	3.5%
	30 min	6.1%	4.3%	2.2%	0.7%	2.1%
	60 min	1.4%	2.1%	7.0%	3.6%	0.0%
$\langle \frac{|\delta q|}{q} \rangle$		3.0%	3.0%	4.6%	1.5%	1.8%

TABLE III

Same as in Table I but for interday time-lags from 1 day to 4 days.

	Δt	OrangePL	PKN Orlen	PKO BP	PZU	WIG 30
$\frac{|\delta q|}{q}$	1 day	3.7%	15.2%	1.6%	20.0%	3.4%
	2 days	12.9%	9.2%	12.9%	26.3%	8.3%
	3 days	13.4%	18.3%	11.9%	26.3%	6.4%
	4 days	13.6%	20.9%	24.2%	13.2%	15.0%
$\langle \frac{|\delta q|}{q} \rangle$		10.9%	15.9%	12.7%	21.5%	8.3%

Our first observation is that the cubic power law ($q = 3/2$) can be very well recognized for WIG 30 index for $\Delta t = 1$ min (see Table I and Figs. 6, 7). The separate stocks obey this law for $1 < \Delta t < 30$ min, usually with minor time delay comparing with the main index. However, already for time-lags $\Delta t \approx 60$ min we observe remarkable departure from the cubic law with $q \approx 1.4$ corresponding to $\zeta = 4$ (see Eq. (1)). We checked also that the value $q = 1.2$ suggested as asymptotic one for developed markets (see, e.g., [26]) is reached for Polish market for time-lags ten times shorter than reported for mature stocks as long as $\Delta t \approx 40$ days. As a result distribution of intraday returns for $\Delta t \geq 1$ day starts in case of Polish stocks also from the lower value ($q = 1.2$–1.3) than reported for the established US market ($q = 1.4$) in [26].

Another observation is that the Tsallis parameter q and the Hurst exponent H are shown to be in principle not quantitatively connected for real data. This situation is explicitly revealed in Fig. 7 where intraday returns exhibit fat tails with varying $q \approx 1.2$–1.25, while $H = 0.5$ indicates simultaneously that no linear autocorrelation

exist in the corresponding series of WIG 30. This result clearly points that higher order correlations and nonlinear correlations play fundamental role in genesis of fat tails in distribution of returns on stock market, and that expectation that $q = 1$ should correspond to $H = 1/2$ is not fundamentally confirmed therefore. Note, however, that due to accuracy in H exponent calculation within DFA, the evolution of H exponent with time-lag Δt reveals the presence of some residual linear autocorrelations for returns in this complex stock market system.

In particular, the negative linear autocorrelation ($H \approx 0.45$) can be seen for all ingredient stocks for $\Delta t \approx 1$ min which agrees with results calculated directly for financial data (see, e.g. [20]). For longer time-lags stocks become slightly persistent with maximum of $H \approx 0.55$ around $\Delta t \approx 20$–40 min. The diversified level of $0.4 \leq H \leq 0.55$ is seen for interday data of separate stocks. One can observe also that the WIG 30 index is remarkably more stable for interday data since it shows no linear autocorrelations present for such quotations in the considered large 2 years window, while it does reveal (like its components) some residual positive persistence for interday returns (see the bottom part in Fig. 7).

The deeper analysis of memory effects in stock data can be done looking at asymmetry between left and right tail of probability distribution. We considered an independent fit of q-normal Tsallis distribution done separately for both distribution tails, instead of the standard skew-

ness. This way two more Tsallis indices q^+ and q^- are calculated as result of fit to positive and negative returns correspondingly. These values can be used as an alternative description of asymmetry between right and left tails — particularly for large absolute returns. The found q^+ and q^- values are collected for main stocks on Polish market and for WIG 30 in Table I. this point it is ob-

vious to introduce the asymmetry index $\delta q \equiv q^+ - q^-$. Such asymmetry index does not show any uniform behavior along Δt — neither its absolute value $|\delta q|$ does — contrary to behavior of the main Tsallis index q which decreases for WIG 30 with growing Δt to the asymptotic value $q \simeq 1.2$ at $\Delta t \gtrsim 1$ day. This is seen clearly in right parts of Fig. 6.
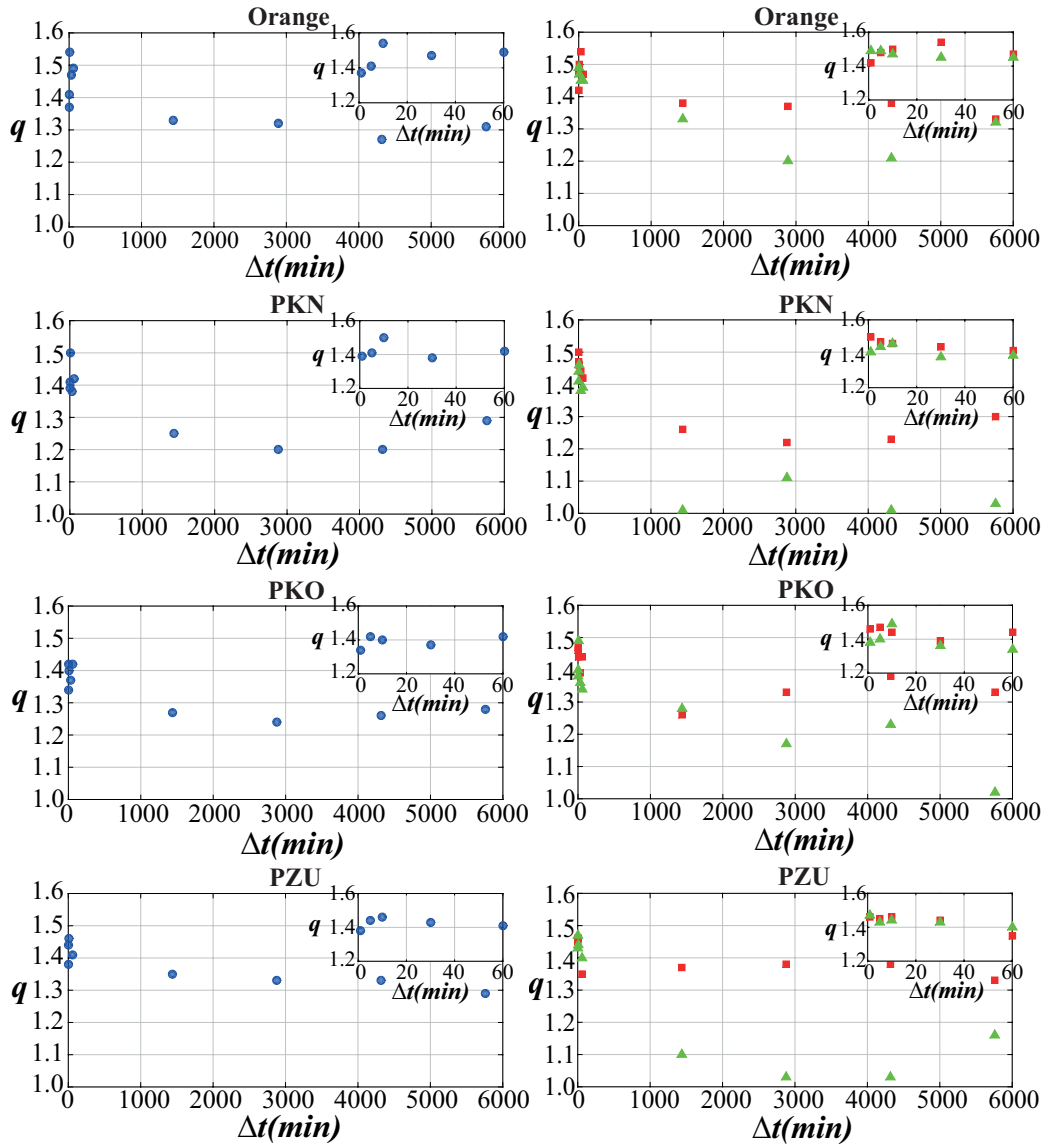

Fig. 6. Dependence between Tsallis parameters: q (left) (● all returns) and $q^\pm$ (right) (▲ positive returns, ■ negative returns) versus the time-lag used for calculation of returns. Inbox in the right top corner shows magnification of this dependence for higher frequency data $\Delta t \leq 60$ min. Note that the horizontal axis in these plots show the clock time which is equivalent to stock trading time only for time-lags $\Delta t < 480$ min. For other Δt values the proper adjustment between clock and trading time is necessary — the clock time is roughly three times larger than trading time because of amount of trading hours in a day.

However, let us have a look at the relative absolute asymmetry between tails of positive and negative returns. It can be defined in straightforward way as:

$$\frac{|\delta q|}{q} = \frac{|q^+ - q^-|}{q}. \tag{8}$$

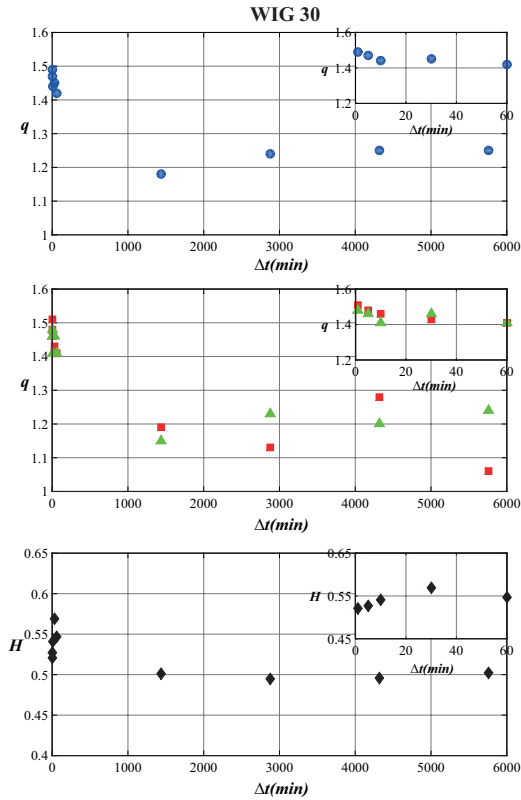Its values are collected in Tables II, III. We propose

Fig. 7. Same as in Fig. 6 but for WIG 30 signal.(● all returns, ▲ positive returns, ■ negative returns). The bottom plot shows additionally the Hurst exponent H values calculated using DFA technique for integrated time series of returns generated at different time-lags Δt.
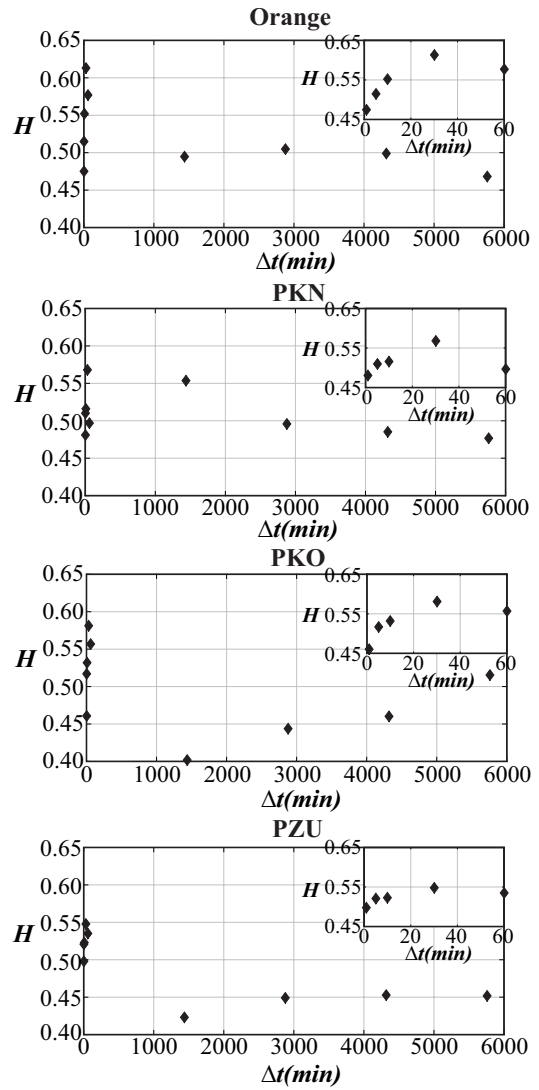

Fig. 8. Dependence between Hurst exponent and the time-lag Δt for discussed Polish stocks from WIG 30. The inboxes in the right top corners reveal this dependence for better display for higher frequencies of data collection $\Delta t \leq 60$ min.

to use it as a new quantifier measuring the normalized asymmetry between temptation among investors (letting a stock to go up) and the level of fear among traders (pushing the market prices down). The value of $|\delta q|/q$ should be therefore bigger for younger and more speculative markets than for well established ones. In the case of developing stock markets this relative asymmetry index should manifest the presence of more unbalanced emotions among traders. Its value should presumably be less remarkable for the whole market index — as the capital-weighted average measure of stock prices — than for returns of particular stock. This property is really seen in Tables II, III for real data.

Some regularity can be noticed for $|\delta q|/q$ index from Tables II, III. It exhibits much different values for interday and intraday data. Indeed, if one takes the average over various considered time-lags, the value $\langle|\delta q|/q\rangle$ for distribution of interday returns on Polish stock exchange market is $\langle|\delta q|/q\rangle_{\text{inter}} \lesssim 5\%$ while the corresponding average for intraday returns is remarkably higher $(\langle|\delta q|/q\rangle_{\text{intra}} \gtrsim 10\%)$. This happens for the most of stocks. Thus we may conclude that the average relative asymmetry index is about 3–15 times larger for intraday returns than for interday data. This noticeable fact requires some explanation.

We hope to have the one. It is likely that the mechanism standing behind is a reflection of different tools basically used by investors in interday and intraday trading. Presumably, in the first case stochastic econometric models are mostly used by institutional investors who trade in short time scale while in the second case the traditional technical analysis is applied by individual traders who trade more often in longer time horizon.

Anyway these results lead to interesting observation that trading in larger time horizon involves more asymmetric price speculation, i.e., it has features of less balanced temptation over fears among investors. Thus it seems interesting to use this new quantifier and investigate further its properties also for various foreign stocks and time-lags. The regularity of $|\delta q|/q$ and its remarkable difference between intraday and interday trading may be

useful for independent detection of speculation level on stock market, particularly because we observe also that positive returns tend to the random case ($q \approx 1$) usually much faster than negative returns do (see Fig. 6). This asymmetry strongly depends on the frequency of data sampling and trials to describe left and right tail of return distribution where just one value of q makes a clear simplification of the problem. Thus the memory effect in returns is evidently asymmetric for positive and negative returns and requires further detailed study not only in case of Polish stock market.

References

[1] E.F. Fama, *J. Business* **38**, 34 (1965).

[2] R.N. Mantegna, H.E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance*, Cambridge Univ. Press, Cambridge 2000.

[3] R.N. Mantegna, H.E. Stanley, *Nature* **383**, 587 (383).

[4] P. Gopikrishnan, V. Plerou, X. Gabaix, L.A.N. Amaral, H.E. Stanley, *Physica A* **299**, 137 (2001).

[5] P. Gopikrishnan, V. Plerou, Y. Liu, X. Gabaix, L.A.N. Amaral, H.E. Stanley, *Physica A* **287**, 362 (2000).

[6] K. Kiyono, Z.R. Struzik, Y. Yamamoto, *Phys. Rev. Lett.* **96**, 068701 (2006).

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

[8] L. Bachelier, *Ann. Sci. Ecole Norm. Sup.* **17**, 21 (1900).

[9] B. Podobnik, K. Matia, A. Chessa, P.Ch. Ivanov, Y. Lee, H.E. Stanley, *Physica A* **300**, 300 (2001).

[10] B. Podobnik, P.Ch. Ivanov, Y. Lee, A. Chessa, H.E. Stanley, *Europhys. Lett.* **50**, 711 (2000).

[11] Y. Liu, P. Gopikrishnan, P. Cizeau, M. Meyer, C.-K. Peng, H.E. Stanley, *Phys. Rev. E* **60**, 1390 (1999).

[12] M.I. Bogachev, J.F. Eichner, A. Bunde, *Phys. Rev. Lett.* **99**, 240601 (2007).

[13] T. Lux, *Appl. Fin. Econ.* **6**, 463 (1996).

[14] P. Gopikrishnan, V. Plerou, L.A.N. Amaral, M. Meyer, H.E. Stanley, *Phys. Rev. E* **60**, 5305 (1999).

[15] X. Gabaix, P. Gopikrishnan, V. Plerou, H.E. Stanley, *Nature* **423**, 267 (2003).

[16] V. Plerou, P. Gopikrishnan, L.A.N. Amaral, M. Meyer, H.E. Stanley, *Phys. Rev. E* **60**, 6519 (1999).

[17] S. Drożdż, J. Kwapień, F. Grümmer, F. Ruf, J. Speth, *Acta Phys. Pol. B* **34**, 4293 (2003).

[18] K. Matia, L.A.N. Amaral, S. Goodwin, H.E. Stanley, *Phys. Rev. E* **66**, 045103 (2002).

[19] U.A. Müller, M.M. Dacorogna, R.B. Olsen, O.V. Pictet, M. Schwarz, C. Morgenegg, *J. Bank. Fin.* **14**, 1189 (1995).

[20] S. Drożdż, J. Kwapień, P. Oświęcimka, R. Rak, *New J. Phys.* **12**, 105003 (2010).

[21] C. Tsallis, *J. Stat. Phys.* **52**, 479 (1988).

[22] C. Tsallis, S.V.F. Levy, A.M.C. Souza, R. Maynard, *Phys. Rev. Lett.* **75**, 3589 (1995).

[23] M.S. Taqqu, V. Teverovsky, W. Willinger, *Fractals* **3**, 785 (1995).

[24] H.E. Hurst, *Trans. Am. Soc. Civil Eng.* **116**, 770 (1951).

[25] C.K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, A.L. Goldberger, *Phys. Rev. E* **49**, 1685 (1994).

[26] M. Ausloos, K. Ivanova, *Phys. Rev. E* **68**, 046122 (2003).