Price–Volume Relationship in Polish Stock Market

K. Karpio[a], P. Łukasiewicz[a] and A. Orłowski[a,b]

[a]Katedra Informatyki SGGW, Nowoursynowska 159, 02-776 Warszawa, Poland

[b]Instytut Fizyki PAN, al. Lotników 32/46, 02-668 Warszawa, Poland

A relationship between daily prices of Polish WIG index and trading volumes is investigated. By introducing variables related to a number of last prices and volumes, a history of values in a certain period of time (which could be regarded as an investor memory) is taken into account. Different characteristics of autocorrelations for prices and trading volumes are observed. By studying mutual correlations between the variables, a local maximum at about 100 trading days is discovered. The Granger causality test is performed, indicating very strong influence of prices on volumes. This property can be considered as a sign of markets maturity.

PACS: 89.65.-s, 89.65.Gh, 02.30.Nw

1. Introduction

The levels of prices and trading volumes are the base quantities describing current situation on a stock exchange (stock market). Those quantities are observed by investors with a big concern. They are the one of the base criteria during investment decision making but the relations between them have much deeper meaning. Among others, they allow for overall performance evaluation of stock markets, give insight about spreading new information, and influence the operation of futures markets [1, 2]. For these reasons prices, volumes and their mutual relations are the subject of numerous scientific papers [1–6]. In a majority of works daily changes of prices and volumes – daily return rates are analyzed. However investors look further in the past and compare the current prices and volumes to their level in shorter or longer time horizon. The length of the horizon is of course the matter of an investor's decision. It may be equal to dozens or even hundreds of days.

In this paper we would like to study the problem of relations between prices and volumes from an investor's point of view. We propose new variables which express current prices and volumes in relation to their past values observed in a certain time horizon (investor memory). We study the distributions of the variables and a degree of the autocorrelations as a function of the investment horizon. During next step we analyze mutual correlations between the variables as well as we investigate causality of those correlations with the means of the Granger test.

2. Data selection and variables

We analyze the main Polish Stock Exchange index WIG which reflects the behavior of the stock market. Data consist of daily closure prices and trading volumes from 02-01-1995 to 21-10-2010, what corresponds to 3961

days. We excluded from analysis the first few years related to the preliminary stage of the operation of Polish market, when the number of assets was very small and they weren't even priced every day.
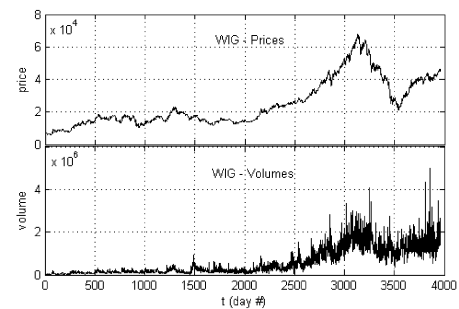

Fig. 1. WIG daily prices and trading volumes for analyzed period of time. Time is expressed as an ordinal number of trading day.

First of all we would like to draw attention to the fact that prices as well as volumes increased by a few orders of magnitude in the last 15 years (Fig.1). So the term "big" or "small" price (volume) cannot be applied directly to the raw data. In order to be able to compare the changes of data within so wide period of time we define the new time series. We compare the current price of the WIG index to the average closure price of the last k days. The k will be called window and will reflect the memory of an investor. Volume will be treated the same way – dividing daily trading volume by the average of the last k volumes. Thus values of the new time series could be expressed as:

$$P_{k\,t} = \frac{p_t}{\frac{1}{k}(p_{t-1} + p_{t-2} + \ldots + p_{t-k})} \tag{1a}$$

$$V_{k\,t} = \frac{v_t}{\frac{1}{k}\big(v_{t-1} + v_{t-2} + \ldots + v_{t-k}\big)}, \qquad (1b)$$

where p_t and v_t- daily closure price and trading volume respectively, at time t, $(t = k+1, \ldots, 3961)$. Time t will be the ordinal number of session and k – the size of the window. We analyzed windows sizes starting from 1 to 400. The latter corresponds to almost two-year-window, which is relatively long memory horizon. For the window of 1 makes $P_{1\,t}$, $V_{1\,t}$ become daily price and volume return rates. We will further denote by P_k, V_k the variables with values of P_{kt} and V_{kt} respectively.

3. Behavior of price variable

The sample time series of the P_{201} and P_{401} are presented in Fig. 2, together with the WIG index and running averages of the 201 and 401 prices respectively. On the horizontal axes time is represented by the ordinal number of trading days.
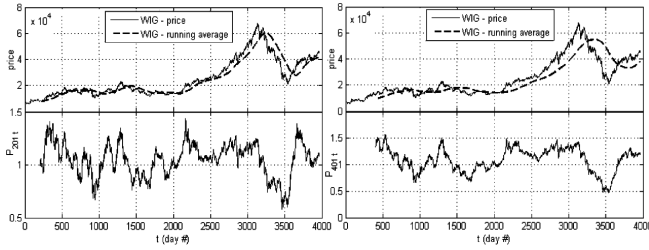

Fig. 2. WIG daily prices as well as 201 and 401 sessions running averages (upper plots) for analyzed period of time. Daily values of P_{201} and P_{401} are also shown (bottom plots).

There are no visible trends for the both variables. Similar behavior was observed for other analyzed values of k. On the other hand the increase of dispersion of P_k with k could be observed (please note different vertical scale on the bottom plots). The behavior of the dispersion is obvious because our variable depends on the greater number of historical prices. We present distributions of variable P_k for selected values of k in Fig. 3.
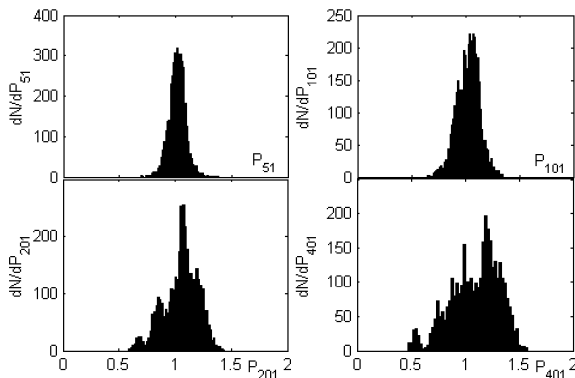

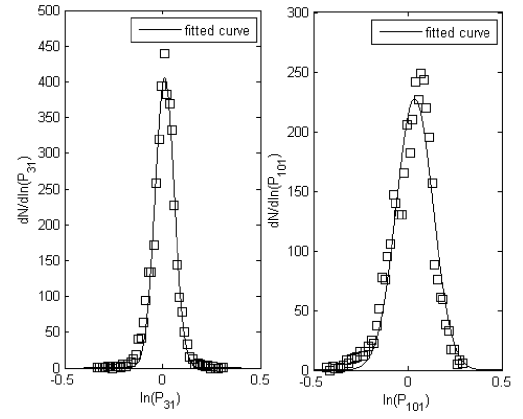Fig. 3. Distributions of P_k for selected values of k.

Fig. 4. The distributions of $\tilde{P}_k = \ln(P_k)$ for selected values of k.

Distributions of P_k could be approximated by normal distribution. However, by definition, values of P_k are within the range 0 to infinity what is inconsistent with the range of normal distribution. Such a discrepancy became more and more significant when analyzing bigger windows because of the increase of P_k dispersion. In order to have the consistent analysis in full range of windows we will use variable $\tilde{P}_k = \ln(P_k)$. The distributions of $\tilde{P}_k$ together with the fits of normal distributions for selected values of k are presented in Fig. 4.

We utilized a normal distribution described by the function given by

$$f(x) = a \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \qquad (2)$$

where a is a normalization factor, μ denotes an average value, and σ is a standard deviation.

We applied a normal fits to $\tilde{P}_k$ for the full range of windows analyzed. Their qualities have been measured by R^2 and were within the range of 0.9 to 0.99, decreasing with th window size. In order to find possible autocorrelations we investigated the width of the distributions as a function of the window size. Of course in the case of fit and data discrepancy the width of the normal distribution fitted could be not reliable. So to measure the width independently we used interquantile range to measure the dispersion. Of course we are aware of other methods of evaluating the dispersion but the detailed analysis of this aspect is beyond the scope of this paper. However, in further work we plan to perform comprehensive fluctuation analysis based on the well-known Hurst exponent.

In the case of the absence of autocorrelations the width of the distribution scales as a power function of number of periods with the exponent equal to 0.5. The quality of the power fit in both cases is very good (Fig. 5). Obtained result indicates on a weak positive autocorrelations – exponent being 0.55 ± 0.01. That means the dispersions of the distributions are somewhat greater that in the case of the lack of autocorrelations and indicate on keeping trend.
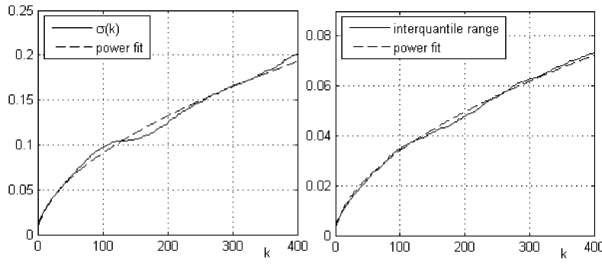
Fig. 5. Width of the Gaussian fit (left plot) and interquantile range (right plot) as a function of the window size.

4. Behavior of volume variable

By analogy to prices, in Fig. 6 we present WIG trading volumes together with moving average (top plots) and time series V_k (bottom plots) for $k = 201$ and 401. It can be noticed that dispersion of trading volumes is much bigger than it is for prices. There are also visible huge spikes for some isolated days. Time series for trading volumes qualitatively resemble price time series (see Fig. 1), where big prices come together with big volumes.
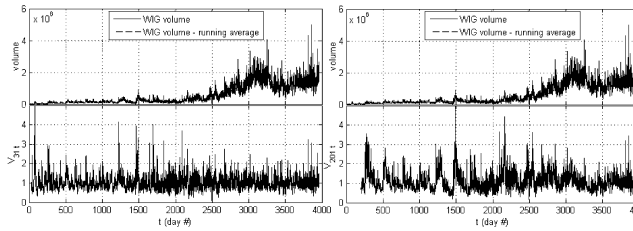

Fig. 6. WIG daily volumes as well as 31 and 201 sessions running averages (upper plots) for analyzed period of time. Daily values of V_{31} and V_{201} are also given (bottom plots).

The distributions of V_k are asymmetric toward the region of high volumes and are very well described by the log-normal function. To show this fact we present the distributions of V_k for $k = 51$, 101, 201, and 401 together with the log-normal fits given by the Eq. (2).

The very good fit quality is characteristic for all the windows analyzed: R^2 being of about 0.99 (Fig. 7). Hence distributions of $\tilde{V}_k = \ln(V_k)$ are Gaussians. Because of the very good agreement of data with fitted function there is no need to use positional measurements of the dispersion like interquantile range, what was the case for prices. We present the width of the distributions as the function of number of periods – trading days in Fig. 8. The width scales as a power law however the exponent is far lower than in the absence of autocorrelations. The fit exhibits the very large negative autocorrelations (0.101 ± 0.003). When increasing the horizon the increase of the width of the volume distribution keeps to be relatively small. One should keep in mind that a number of assets is limited and usually large demand causes rather
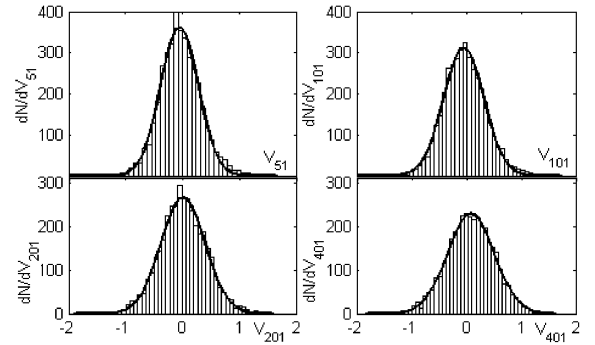

Fig. 7. The distributions of $\tilde{V}_k = \ln(V_k)$ for, $k = 51$, 101, 201, and 401, together with the normal fits given by the Eq. (2).

price to increase instead of a big volume. When a price increases some people change their mind and do not buy assets. And thus the volume does not increases as was intended. Similar behavior takes place in the case of big supply which causes price to drop and thus some people refuse to sell the assets they possess limiting trading volume. On the other hand the prices are not limited, they can be very big.
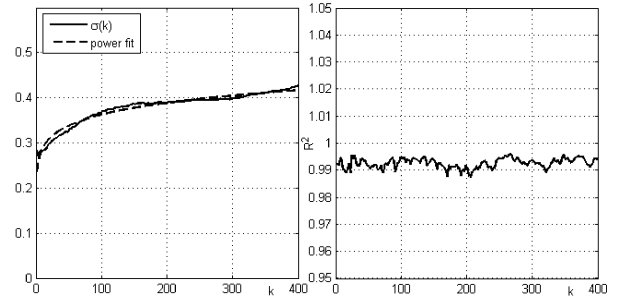

Fig. 8. Width of the distributions as a function of the window size together with the power fit (left plot). Gaussian fit quality expressed by R^2 as a function of the window size (right plot).

5. Price–volume relationship

We calculated the correlation coefficients r between $\tilde{P}_k$ and $\tilde{V}_k$ for the analyzed windows sizes, which are presented in Fig. 9.

As can be seen the correlation coefficient r has a local maximum for an investor memory of about 75 trading days, what corresponds to roughly four months. The maximum is followed by the relatively flat minimum for 100–150 trading days. After that the correlation increases linearly. We investigate the relationships between both variables in more details in the following part of this paper.

We further investigate the relationship between prices and volumes by studying the 2-dimensional correlation
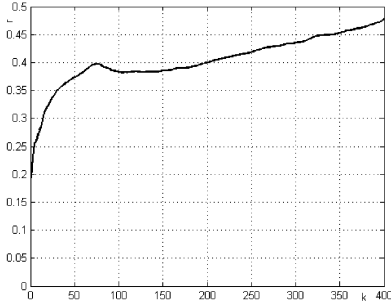
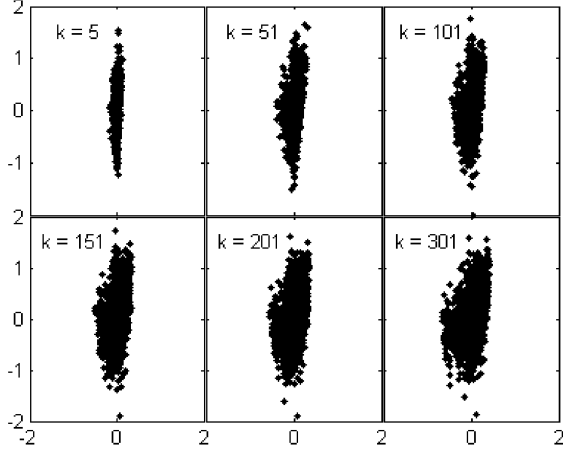Fig. 9. Correlation coefficient r between $\tilde{P}_k$ and $\tilde{V}_k$.

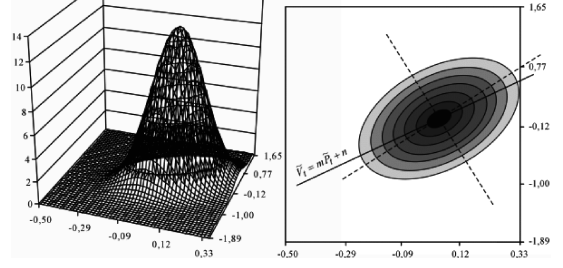Fig. 10. Correlation plots for $k = 5, 51, 101, 151, 201,$ and 301.

Fig. 11. Joint distribution of $\tilde{P}_k$ and $\tilde{V}_k$ for the window of 151 days.
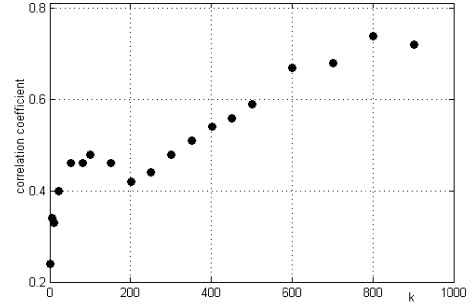
Fig. 12. Correlation coefficient *versus* the window size.

plots. The plots for selected values of k are given in Fig. 10. In order to quantitatively describe the relationship we applied the following procedure. For each of the investment horizon we calculated the histogram representing correlation plot. To each of the histogram we fit the 2-dimensional normal distribution given by the:

$$f(p,v) = \frac{A}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(\frac{-1}{2(1-\rho^2)}\left[\frac{(p-\mu_1)^2}{\sigma_1^2}\right.\right.$$

$$\left.\left. -\frac{2\rho(p-\mu_1)(v-\mu_2)}{\sigma_1\sigma_2} + \frac{(v-\mu_2)^2}{\sigma_2^2}\right]\right), \qquad (3)$$

where parameters μ_1, μ_2, σ_1, σ_2 are two average values and two standard deviations of marginal distributions of $\tilde{P}_k$ and $\tilde{V}_k$. Parameter ρ is the correlation coefficient ($|\rho| \leq 1$).

Parameters of the function (3) have been evaluated by minimizing χ^2 weighted by the inversed error of data. For every histogram bin we calculated the error as a squared root of the bin content. As a fit quality we used weighted χ^2 divided by the number of degrees of freedom. For all the investigated windows we obtained a good fit quality.

We investigated the fit parameters describing the shape of the 2-dimensional distributions as a function of the investor memory. For the next part of this paper we describe the behavior of the correlation coefficient and the slope of the regression model given by the $\tilde{V}_t = m\tilde{P}_t + n + \varepsilon_t$ (Fig. 11).

The correlation coefficient obtained using the function (4) exhibits previously noticed local maximum (see Fig. 12). We obtained slightly different position of the local maximum because of the implementation of different procedure (compare to Fig. 9). Previously we simply calculated the correlation coefficient between two time series, where each data point has been taken into account with equal weight. Now we fit 2-dim normal distribution using the weighted χ^2 method, where every point on the distribution has been weighted based on its inversed error. The latter method favors points with small errors while points with large errors have minor impact on the fitted function parameters.

As can be seen in Fig. 10 the distribution of points in the correlation plots tends to rotate in the direction of P axis. In order to investigate this behaviors we plot the slope of the regression line as a function of window. The slope seems to stabilize on the level slightly above 1 at the window of about 250 trading days (Fig. 13). This point is close to the local minimum in Fig. 11 where the correlation coefficient assumes local minimum value. Those points correspond to the investor memory of about 200–250 days what correspond to roughly one year. It is close to the value of local autocorrelations described in the frames of the local Hurst exponent [7-9]. The stabilization means that investors do not look beyond the horizon of about one year. In order to present the result of coefficient stabilization we plot its values for windows up to 1000 of trading days.
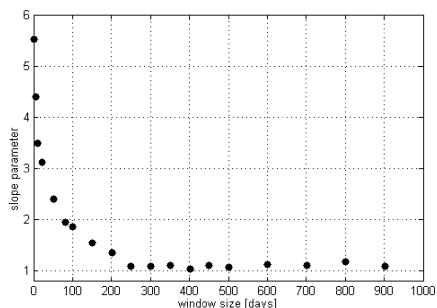
Fig. 13. Slope parameter as a function of the window size.

6. Price–volume causality

The last part of this paper is devoted to the study of relationship between prices and volumes in frames of the Granger causality. We would like to obtain the answer for the question: does $\tilde{P}_k$ cause $\tilde{V}_k$ or vice-versa. One can say that X variable Granger causes Y ($X \to Y$), if the current values of Y can be predicted with the greater accuracy then X taking into account past values of both X and Y [10, 11]. The starting point in the Granger test is the vector autoregressive model given by

$$\tilde{V}_t = B_0 + \sum_{i=1}^{l} \alpha_i \tilde{V}_{t-i} + \sum_{i=1}^{l} \beta_i \tilde{P}_{t-i} + \varepsilon_t, \qquad (4)$$

where $\tilde{V}_t$ and $\tilde{P}_t$ are values of the variables $\tilde{V}_k$ and $\tilde{P}_k$ respectively, for a given window size k.

The null hypothesis is $\beta_1 = \beta_2 = \ldots = \beta_l = 0$, what means that $\tilde{P}_k$ does not Granger causes $\tilde{V}_k$. With the assumption that null hypothesis is true the statistics SG of the Granger test has F distribution (see [11, 12] for details). If the SG/F^*, (where F^* is a critical value) exceeds 1, we can conclude that null hypothesis is false. Using the Granger test we study causality in both directions: $\tilde{P}_k \to \tilde{V}_k$ and $\tilde{V}_k \to \tilde{P}_k$. In Fig. 14 we plot the SG/F^* for $k = 55$ and 201 as a function of the number of past values taken into account by the test, see Eq. (4).
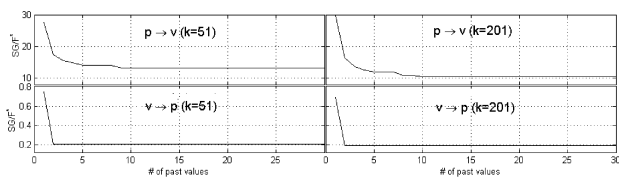

Fig. 14. The SG/F^* as a function of the number of past values taken into account during prediction using Granger test for two size of the window: 51 (left plot) and 201 (right plot).

As can be seen in Fig. 14 the plots stabilize at a certain level. This behavior is characteristic for all the windows investigated. During the next step we assume common

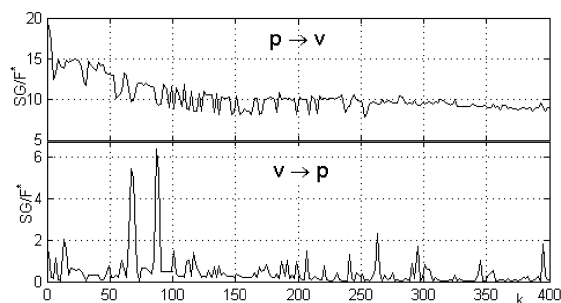number of past values in the Granger test equal to 12, for all the windows studied.

Fig. 15. The SG/F^* as a function of the window size.

We plot value of SG/F^* as a function of window for both scenarios: $\tilde{P}_k \to \tilde{V}_k$ and $\tilde{V}_k \to \tilde{P}_k$ in the Fig. 15. Thus we are able to study both the presence the directional causality as well as to compare the force of the effect for various sizes of windows.

The values of SG/F^* in Fig. 15 unambiguously prove that $\tilde{P}_k$ causes $\tilde{V}_k$ ($\tilde{P}_k \to \tilde{V}_k$). This effect is very clear because for all the windows the Granger statistics exceed the critical values for at least 8 times. On the other hand, in the case of the inverse relation $\tilde{V}_k \to \tilde{P}_k$ we observe low values of SG/F^*, thus we face the one-directional causality. For the majority of windows the ratio does not exceed unity, thus for those instatnces volumes cannot be regarded as a cause of prices. However for selected values of k one can speak about two-directional relation. Spikes visible in Fig. 15 (upper plot) indicate that for some window sizes, mainly of about 65 and 85, volumes could also cause prices.

In the case of Polish market we observe similar behavior to the one characteristic for US market, while being different from the one observed for emerging markets in Latin America. For the latter one can conclude that volumes influence prices [6]. In the majority of analyses regarding mature financial markets the influence of prices on volumes with no reverse relation has been observed (see for example [13, 14]). More complex dependency has been observed for emerging markets. For some of them the mutual dependency exists. The behavior similar to the one observed for Polish market (prices causing volumes) has been also observed for Mexican market, which is under great influence of the US mature one.

7. Summary and conclusions

The relationship between prices and volumes has been studied using the new variables P_k and V_k as well as $\tilde{P}_k = \ln P_k$ and $\tilde{V}_k = \ln V_k$, which take into account historical horizon of prices and volumes respectively. The conclusions can be summarized in the following items.

1. Distributions of P_k and V_k are log-normal, thus $\tilde{P}_k$ and $\tilde{V}_k$ are normal. Moreover the compatibility of

V_k with log-normal is very high, regardless of the window size k.

2. The $\tilde{P}_k$ exhibits weak positive autocorrelations indicating on small tendency to keep trend of changes. On the other hand the $\tilde{V}_k$ has very strong negative autocorrelations.

3. Linear correlation coefficient between $\tilde{P}_k$ and $\tilde{V}_k$ has a local maximum of 0.5 at a window of about 100 trading days. The maximum is then followed by the relatively flat minimum around $k = 200$. After the minimum correlation coefficient increases monotonically with the window size.

4. The 2-dimensional correlation plots are well described by the 2-dimensional Gaussian functions. Distributions rotate with the window size toward price axis. The rotation of the distributions was described by the slope of the regression model $\tilde{V}_t = m\tilde{P}_t + n + \varepsilon_t$. The value of m decreases for $k \in [1; 250]$ and then stabilizes on the level of about 1. So the increasing window size (investor memory) does not change the slope. This means that investors do not look beyond the horizon of about one year.

5. For each k price $\tilde{P}_k$ causes $\tilde{V}_k$. This effect is very strong and is characteristic for mature markets. For some values of k the volumes also causes the price, so the relation is then two-directional.

References

[1] S.C. Gold, *J. Accounting and Finance Research* **2**, 85 (2004).

[2] J.M. Karpoff, *J. Financial and Quantitative Analysis* **22**, 109 (1987).

[3] A.R. Gallant, P.E. Rossi, G. Tauchen, *The Review of Financial Studies* **5**, 199 (1992).

[4] K. Saatcioglu, L.T. Starks, *International J. Forecasting* **14**, 215 (1998).

[5] C.K.Y. Leung, G.C.K. Lau, Y.C.F. Leong, *J. Real Estate Research* **23**, 253 (2002).

[6] R.R. Kamath, *International Business and Economics Research J.* **7**, 7 (2008).

[7] D. Grech, Z. Mazur, *Physica A* **336**, 133 (2004).

[8] K. Karpio, A.J. Orłowski, P. Łukasiewicz, *Acta Physica Polonica A* **117**, 619 (2010).

[9] C.K. Peng *et al.*, *Physial Review E* **47**, 3730 (1993).

[10] C.W.J. Granger, *Econometrica* **37**, 424 (1969).

[11] W. Charemza, D. Deadman, *New directions in econometric practice: General to specific modelling, cointegration, and vector autoregression*, Edward Elgar Pub., Cheltenham 1997.

[12] T.C. Mills, R.N. Markellos, *Econometric analysis of financial time series*, Cambridge University Press, Cambridge 2008.

[13] G. Chen, M. Firth, O.M. Rui, *Financial Review* **38**, 153 (2001).

[14] C. Lee, B. Swaminathan, *J. Finance* **55**, 2017 (2000).