STOCK RETURNS VERSUS TRADING VOLUME: IS THE CORRESPONDENCE MORE GENERAL?

Rafal Rak\textsuperscript{a}, Stanislaw Drozd\textsuperscript{b,c}, Jaroslaw Kwapien\textsuperscript{b} 

Pawel Osiewimka\textsuperscript{b}

\textsuperscript{a}Faculty of Mathematics and Natural Sciences, University of Rzeszow 
Rejtana 16a, 35-959 Rzeszow, Poland

\textsuperscript{b}Complex Systems Theory Department
The H. Niewodnicznski Institute of Nuclear Physics, Polish Academy of Sciences
Radzikowskiego 152, 31-342 Krakow, Poland

\textsuperscript{c}Faculty of Physics, Mathematics and Computer Science
Cracow University of Technology
Warszawska 24, 31-155 Krakow, Poland

(Received September 18, 2013; revised version received October 14, 2013)

This paper presents a quantitative analysis of the relationship between the stock market returns and corresponding trading volumes using high-frequency data from the Polish stock market. First, for stocks that were traded for sufficiently long period of time, we study the return and volume distributions and identify their consistency with the power-law functions. We find that, for majority of stocks, the scaling exponents of both distributions are systematically related by about a factor of 2 with the ones for the returns being larger. Second, we study the empirical price impact of trades of a given volume and find that this impact can be well described by a square-root dependence: \( r(V) \sim V^{1/2} \). We conclude that the properties of data from the Polish market resemble those reported in literature concerning certain mature markets.

DOI:10.5506/APhysPolB.44.2035 
PACS numbers: 89.20.–a, 89.65.Gh, 89.75.–k

1. Introduction

Every day millions of individuals all over the world make billions of orders to buy or sell stocks according to their own investment strategies and in reaction to huge amount of arriving information. These individual decisions taken together define very complex behaviour of the financial markets [1–5] and lead to such characteristics of the financial data as multifractality [6–11], long memory, nonlinear correlations [9, 12, 13], the leverage effect [14, 15], fat tails of financial data fluctuations [16–30], known together as the financial stylized facts.
The fat tails of pdfs mean that, for example, the nature of logarithmic price fluctuations (returns) differs from the Gaussian noise model as the former can be much larger than the latter expects. The relative magnitude of the most spectacular events observed in the financial markets is a good example of this property. Let us assume that we have a bi-variate time series of price and volume recordings collected over some period of time. In general, the returns can be defined as

$$\begin{align*}
  r(\Delta t_k) &= \frac{R(\Delta t_k) - \langle R(\Delta t_k) \rangle}{\sigma}, \\
  R(\Delta t_k) &= \ln Q(t_k + \Delta t_k) - \ln Q(t_k),
\end{align*}$$

(1)

where $\Delta t_k$ is length of the $k$th interval of time, $Q(t)$ is stock price, $\sigma$ is standard deviation, and $\langle \cdot \rangle$ is mean over all the intervals $k = 1, \ldots, T$. For each return $r(\Delta t_k)$, there exists an appropriate trading volume $V(\Delta t_k)$. The intervals $\Delta t_k$ can be defined in various ways: they can either cover a constant number of consecutive transactions $n_T$, be defined by constant trading volume, or be equal to each other with $\Delta t_k = \Delta t$ for all $k$s. The latter definition is the most common one.

It was documented in literature that the non-Gaussian pdfs of the return can be described in most cases by power-law tails of the form of $p(x) \sim x^{-(1+\alpha)}$ with $\alpha > 0$. Exact values of the scaling index $\alpha$ depend on the return definition and vary from market to market [22, 24, 31–33], and from past to present [20, 34, 35]. The scaling index $\alpha_r$ for the returns depends also strongly on either $\Delta t$ or $n_T$. Typically, one observes a gradual increase of $\alpha_r$ with $\Delta t$ and $n_T$ from an initial value of $3 \leq \alpha_r \leq 4.5$ (see Ref. [17, 18, 34]), which is usually maintained over a range of the shortest intervals, towards higher values for longer intervals or larger number of aggregated transactions. This increase corresponds to convergence of the distribution of the returns towards a stable, Gaussian distribution.

From the pure mathematical perspective, speaking about a convergence is a delicate issue in this context as the convergence towards a Gaussian distribution induced by the Central Limit Theorem does not alter the power-law tail slopes. However, the following two remarks have to be made. First, any empirical data is finite and its pdf/cdf tails do not cover the distant regions which may reveal the actual power-law tails that are being repelled towards infinity while CLT exerts its influence. Thus, in econophysics literature it became commonplace that under the notion of “power-law slope”, one considers the effective slope of the power function that was best-fitted to empirical distribution in its non-central region. In this study, we just follow this approach. Second, one has to keep in mind that financial data is usually non-linearly correlated and cannot be represented by i.i.d. processes [4].
This implies that the standard form of CLT may not be completely adequate in this case, leaving space for its non-extensive generalizations (see Sec. 3). The exact form of convergence might thus differ from the classic CLT picture.

On the other hand, tail shape of the aggregated trading volume or trade size distributions was a matter of some dispute in literature [36], there is nevertheless, substantial evidence that, for at least the American and Chinese stock markets, the trade size and aggregated volume pdfs can be described by a scale-free tails with the exponent \(1.5 \leq \alpha_V \leq 2.8\) [22, 24, 31–33], i.e., around the Lévy limit of \(\alpha = 2\). This implies that the volume pdf convergence towards a Gaussian distribution is rather slow [32]. One of interesting empirical findings is that, for a given market, sometimes a simple relation between both scaling indices exists [18, 24, 31, 32]

\[
\frac{\alpha_r}{\alpha_V} \simeq \xi, \tag{2}
\]

where specific value of \(\xi\) depends on a study. For example, Ref. [24–26] and [24] reported \(\xi \approx 2\). This relation was found empirically for those data, for which the effective power-law tails were observed. How it behaves for large time scales for which the distributions converge closer to a Gaussian distribution, has not been studied since the data samples are in such cases rather small. The above relation between the returns and trade sizes or volume (depending on a definition of the returns) can be related to the empirical price impact function \(r = f(V)\) describing how a trade or aggregated volume of a given size modifies the price. This function is known to be concave, but its exact form is debatable with possible logarithmic [39] or (wider documented) power-law relation [25, 26, 28, 36, 40]. The latter one can be written as

\[
r(V) = cV^\beta, \quad 0 < \beta < 1, \tag{3}
\]

where \(c\) is some positive constant. For example, Gabaix et al. [25, 26] who considered the American stock market argued that \(\beta \approx 1/2\), while Farmer and Lillo [28, 36] reported \(\beta \approx 0.3\) for the London market. It was also seen that this function changes its form under a change of the aggregation time scale, going from strongly nonlinear for small \(\Delta t\) or \(n_T\) to rather linear for larger scales [41].

Based on results of their empirical study, Gabaix et al. [25, 27] formulated a simple theory that was able to explain why both the returns and volume are power-law distributed. They assumed that large market participants — mutual funds, whose activity effectively govern the prices, are power-law distributed and that those funds optimize their trading strategies. Their approach allowed them to derive Eq. (3). This theory opened a debate, in which both the model assumptions and the empirical evidence, which it was grounded on, was criticised [28, 36, 42, 43] and an alternative approach
was proposed pointing out to fluctuations of liquidity, and not the volume
distribution as a principal cause of the form of the return pdf [41]. However,
since the authors of the original model were able to rebut the main objections
against it [26], it now seems that the two approaches may be complementary
and point out to phenomena that in fact coexist.

Here, we do not want to consider empirical data in the light of either
theory, but we rather aim at considering statistical distributions of the
returns and trading volume for a market which was not analysed before
in this context and compare results with those known from literature for
other markets. The Warsaw Stock Exchange (WSE) has total capitalization
of about 180 billion USD (August 2013) and, as such, is still counted
among the small markets. However, many properties of data from the WSE
are similar to the respective properties of data from the mature markets
[44–46]. Therefore, taking all this into consideration, it is interesting to
verify whether any relation similar to Eq. (2) can also be identified on the
WSE.

2. Results

We study the distribution of trading volumes and returns based on high-
frequency data from the largest 14 companies listed in the WSE. Our data
cover a time interval starting on November 17, 2000 and ending on March 6,
2008. The data comprises basically information on all trades which took
place in this period but, unfortunately, it does not offer data from the order
book. Therefore, we cannot distinguish trades which were buyer- or seller-
initiated and have to treat all trades together, which is similar to an approach
of Ref. [31, 32], but different from those in Ref. [24, 36].

We start our analysis with creating the cumulative distribution functions
of the returns and volumes, separately for each company. In this case, we
consider evenly sampled data with $\Delta t = 1$ min. Figure 1 exhibits typical
results, which indicate that all the return distributions possess power-law
tails, while this property is also observed for volume distributions, even
though sometimes in a less clear form (like in the case of Pekao, see the
bottom right graph).

In order to compare the scaling exponents of both distributions, we esti-
imated their values by two independent methods: least square fits of power-

\[ f(x) = ax^{-\alpha} \]

law functions \( f(x) = ax^{-\alpha} \) and the Hill estimators [47]. For a given time
series \( X_1 \leq X_2 \leq \ldots \leq X_N \), the Hill estimator is defined by

\[ \text{HE}_\alpha = \left( \frac{1}{k} \sum_{i=1}^{k} \log(X_i) - \log(X_{k+1}) \right)^{-1}. \]  

(4)
Using both methods for each company, we obtained results that are collected in Table I. For the returns, the exponents are in the range of $3.4 \leq \alpha_r \leq 5.8$, while for the volumes, the exponents are significantly smaller: $1.6 \leq \alpha_V \leq 2.4$. It is important that both methods give comparable results. (We also employed the third method of calculating the scaling exponents, namely the Meerschaert–Scheffler estimator [48], but it provided us with rather unreasonable estimates, similar for both quantities, so we excluded it from our analysis.)
TABLE I

The tail empirical scaling exponents of the cumulative distributions of absolute returns $|r(\Delta t)|$ (second column) and normalized trading volume $V(\Delta t)$ (third column) for 14 Polish companies. The fourth column represents relation (2). The remaining three columns represent the tail exponents and their ratio as given by the Hill estimator [47].

<table>
<thead>
<tr>
<th>Company name</th>
<th>$\alpha_r$</th>
<th>$\alpha_V$</th>
<th>$\alpha_r/\alpha_V$</th>
<th>HE$_{\alpha_r}$</th>
<th>HE$_{\alpha_V}$</th>
<th>HE$<em>{\alpha_r}/$HE$</em>{\alpha_V}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agora</td>
<td>4.3</td>
<td>2</td>
<td>2.15</td>
<td>4.04±0.01</td>
<td>1.64±0.02</td>
<td>2.46</td>
</tr>
<tr>
<td>BRE</td>
<td>4.1</td>
<td>1.9</td>
<td>2.16</td>
<td>4.17±0.015</td>
<td>1.74±0.02</td>
<td>2.40</td>
</tr>
<tr>
<td>Comarch</td>
<td>4.6</td>
<td>2</td>
<td>2.3</td>
<td>5.00±0.06</td>
<td>1.85±0.05</td>
<td>2.70</td>
</tr>
<tr>
<td>Kęty</td>
<td>5.5</td>
<td>2.4</td>
<td>2.3</td>
<td>5.77±0.04</td>
<td>2.09±0.03</td>
<td>2.76</td>
</tr>
<tr>
<td>KGHM</td>
<td>4.6</td>
<td>2.4</td>
<td>1.92</td>
<td>5.15±0.02</td>
<td>2.23±0.05</td>
<td>2.31</td>
</tr>
<tr>
<td>Mostostal Exp.</td>
<td>4.6</td>
<td>2.4</td>
<td>1.92</td>
<td>4.64±0.01</td>
<td>2.22±0.01</td>
<td>2.09</td>
</tr>
<tr>
<td>Netia</td>
<td>3.4</td>
<td>1.7</td>
<td>2</td>
<td>3.55±0.02</td>
<td>1.66±0.01</td>
<td>2.14</td>
</tr>
<tr>
<td>Orbis</td>
<td>3.8</td>
<td>1.8</td>
<td>2.11</td>
<td>4.05±0.01</td>
<td>1.59±0.03</td>
<td>2.35</td>
</tr>
<tr>
<td>Pekao</td>
<td>4.2</td>
<td>2.1</td>
<td>2</td>
<td>5.05±0.03</td>
<td>1.86±0.01</td>
<td>2.72</td>
</tr>
<tr>
<td>PKN Orlen</td>
<td>4.7</td>
<td>2.4</td>
<td>1.96</td>
<td>5.11±0.04</td>
<td>2.25±0.02</td>
<td>2.27</td>
</tr>
<tr>
<td>Prokom</td>
<td>4.2</td>
<td>1.9</td>
<td>2.21</td>
<td>4.45±0.04</td>
<td>1.77±0.03</td>
<td>2.51</td>
</tr>
<tr>
<td>Stalexport</td>
<td>4.7</td>
<td>2.4</td>
<td>1.96</td>
<td>5.12±0.02</td>
<td>2.16±0.02</td>
<td>2.37</td>
</tr>
<tr>
<td>Softbank</td>
<td>3.9</td>
<td>1.8</td>
<td>2.17</td>
<td>3.73±0.01</td>
<td>1.76±0.01</td>
<td>2.12</td>
</tr>
<tr>
<td>TP S.A.</td>
<td>4.6</td>
<td>2.1</td>
<td>2.19</td>
<td>5.13±0.04</td>
<td>1.68±0.03</td>
<td>3.05</td>
</tr>
<tr>
<td>⟨·⟩</td>
<td>4.37</td>
<td>2.1</td>
<td>2.09</td>
<td>4.64</td>
<td>1.89</td>
<td>2.45</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.5</td>
<td>0.26</td>
<td>0.14</td>
<td>0.66</td>
<td>0.25</td>
<td>0.27</td>
</tr>
</tbody>
</table>

Interestingly, although the power law relation for the returns and volumes is different than in Ref. [18, 31, 32], the approximate dependence (2) seems to be similar: $\alpha_r/\alpha_V = 2.09±0.14$ (see the third column in Table I). The Hill estimator on average gives numbers that are larger than those of the power-law fitting ($\text{HE}_{\alpha_r}/\text{HE}_{\alpha_V} = 2.45±0.27$) which is rather a typical relation of these two methods (see also [22]).

To check whether similar value of $\xi$ holds for different time scales, we constructed analogous time series for longer lags. Figure 2 shows the cumulative distributions for two lags: $\Delta t = 1$ min. and $\Delta t = 120$ min. for all the companies taken together. It is now clear that, for the Polish stock market, $\xi \approx 2$ holds also for larger values of scaling exponents than $\alpha_r = 3$ and $\alpha_V = 1.5$ discussed in Ref. [25, 26] in the context of the American markets.

Next, we compare our data with those studied by Zhou [24] by fixing $n_T = 1$. We thus look at time series constructed from the returns caused by individual transactions and the corresponding time series of trade volumes. As we can observe in Fig. 3, the relation between $\alpha_r$ and $\alpha_V$ is still preserved with $\xi \approx 2$ once again. This outcome is different than that for the Chinese stock market reported in Ref. [24], where the ratio was $\approx 1.5$, but it is again close to the result for the American stock market [18, 31, 32].
Fig. 2. Log–log plot of the cumulative distribution of absolute returns $|r(\Delta t)|$ (left) and of normalized trading volume $V(\Delta t)$ (right) for $\Delta t = 1$ and 120 min. over the period November 17, 2000–March 6, 2008. The distributions were averaged over the largest 14 companies.

Fig. 3. Log–log plot of the cumulative distribution of absolute transaction returns for $n_T = 1$ (left) and the corresponding normalized transaction volume (right) for the tick-by-tick data over the period November 17, 2000–March 6, 2008, averaged over the largest 14 companies.

It can be easily shown that relation (2) is valid always if both the return and the volume distributions have power-law tails and the price impact function $r(V)$ is deterministic with a power-law form as in Eq. (3). Let $r \sim V^\beta$, then we get

$$x^{-\alpha r} \sim P(|r| > x) \simeq P\left(cV^\beta > x\right) = P\left(V > (xc^{-1})^{1/\beta}\right) \sim x^{-(1/\beta)\alpha_V},$$

(5)
\[ \xi \equiv 1/\beta = \alpha_r/\alpha_V. \] (6)

To make our results more complete, we have to mention that in the period considered there was a company (Internet Group) with apparently different dynamics. It developed scale-free tails of the pdfs neither for the returns nor the volume (Fig. 4). We suspect that this was because it was a highly speculative stock that went through a phase of bankruptcy, followed

Fig. 4. (Top) Log–log plots of the cumulative distributions of absolute returns \( r(\Delta t = 1 \text{ min.}) \) (left) and normalized trading volume \( V(\Delta t = 1 \text{ min.}) \) (right) for atypical Polish company (IG). (Middle) Locally defined slopes of the distributions shown above. (Bottom) Empirical ratio \( \xi = \alpha_r/\alpha_V \) calculated for the corresponding parts of the distributions. Horizontal axis shows the rightmost 25 histogram bins.
by a spectacular recovery and then again collapsed towards bankruptcy with no serious institutional investors involved. It is curious, however, that the locally defined slopes of the pdfs are also approximately equal to 2 in this case. Whether this is purely coincident or rather indicates that the power-law tails are not a necessary condition for the constant ratio of local slopes, we cannot determine at present, but this phenomenon is doubtlessly worth further studies.

Fig. 5. (Left) Absolute returns as a function of volume for three different companies: PKN Orlen, Prokom, Pekao. The vertical dashed lines denote the quartiles of the total number of data points. (Right) The cumulative distributions of absolute returns corresponding to the three volume intervals: circles, triangles, and squares.
Having shown that, typically, both the return and the volume distribution has scale-free tails, now let us look more closely at the price impact given by Eq. (3). We address a question if existence of such a price impact can be postulated based on empirical data at our disposal. First, we seek any dependence of the return distribution on the volume size. We create a scatter plot $r-V$ for three exemplary companies from different stock sectors (left panels of Fig. 5). By increasing the trading volume, we observe the increasing variance of the absolute returns. To better visualize this, we divide the whole volume range into three parts: $V \leq 100$, $100 < V \leq 1000$ and $V > 1000$, and calculate the return cdfs conditioned on $V$ for each part separately. The resulting distributions are shown in right panels of Fig. 5. For majority of stocks, $\alpha_r$ is noticeably dependent on $V$ in such a way that the larger trading volumes are considered, the gentler is the slope.

Now, we can consider the expectation $E(r^2|V)$ which was proposed in Ref. [25, 26] as an indication of possible square-root dependence $r(V)$

$$E \left( r^2 | V \right) = a + bV \Rightarrow r(V) \sim V^{1/2}. \quad (7)$$

Since $\xi \approx 2$ suggests that the relation described by Eq. (5) holds for the WSE data, we are motivated to look at the empirical form of the price impact function. We compare the expectation $E(r^2|V)$ calculated for real (Fig. 6) and surrogate data (Fig. 7). By following Ref. [28], the latter are constructed by assuming the existence of exact price impact function with some exponent $0 < \beta < 1$ in Eq. (3) and by calculating the artificial returns $r(V(t))$ from

![Graphs](image-url)
the real time series of volumes $V(t)$. If the empirical price impact $r(V)$ is square-root indeed, we expect that $E(r^2|V)$ will qualitatively be similar for the actual and the artificial data if we take $\beta = 1/2$. On the other hand, after taking $\beta \neq 1/2$, the results in each case have to be different.

Fig. 7. Conditional expectation $E[r^2|V]$ for time series of artificial absolute returns created from the original time series of volumes (averaged over all 14 companies) by assuming an exact price impact function $r(V) = cV^\beta$ with $c$ being a constant for $\beta = 0.3, 0.5, 0.7$, respectively. (Left) Volumes $V$ are aggregated over $\Delta t = 1$ min. consecutive returns. (Right) The same as on left, but here volumes $V$ are aggregated over $n_T$ consecutive transactions.
In Fig. 7, we show the results obtained for different choices of $\beta$. In agreement with the above arguments, for all the considered $\Delta t$ or for large enough $n_T$, the expectation $E(r^2|V)$ shows linear behaviour for $\beta = 1/2$, while for $\beta \neq 1/2$ no obvious linear dependence is seen. These results convince us that the square-root price impact function takes place for Polish stocks, indeed.

3. $q$-Gaussian fits to volume distributions

Among the distributions considered in the context of financial fluctuations, there are lognormal distributions, stretched exponentials, truncated Lévy distributions, and the $q$-Gaussian distributions [45, 49–53]. Our previous research has proved that $q$-Gaussians are a good theoretical representation of the empirical return distributions (not only for the stock markets but also for the currency ones) [54, 55]. Based on this experience, we prefer to fit the financial data with the $q$-Gaussians. There are also valid theoretical arguments supporting the use of these distributions in this case [4].

The $q$-Gaussians were discovered in the field of nonextensive statistical mechanics which is a candidate theory to generalize the traditional Maxwell–Boltzmann–Gibbs statistical mechanics to nonequilibrium systems with long-range power-law correlations [56, 57]. Since from a physicist’s point of view it seems that the financial markets can be thought of as systems of this kind, applying the distributions associated with the nonextensive statistics looks natural in this case. The $q$-Gaussians are a family of distributions that maximize the nonextensive Tsallis entropy [56] given by

$$S_q = k_B \frac{1 - \int [p(x)]^q dx}{q - 1},$$

where $p(x)$ is a probability distribution and $k_B$ is the Boltzmann constant, under the conditions

$$\int x [p(x)]^q dx = \mu_q, \quad \int (x - \mu_q)^2 [p(x)]^q dx = \sigma_q^2.$$

Up to a normalization constant, the formula for $q$-Gaussians reads [57, 58]

$$G_q(x) \sim \exp_q \left[ -B_q (x - \mu_q)^2 \right],$$

where $\exp_q x = [1 + (1-q)x]^{1/q}$ and $B_q = [(3-q)\sigma_q^2]^{-1}$. The $q$-Gaussians are defined for $0 < q < 3$. Unlike e.g. the Pareto and other already-listed distributions, the $q$-Gaussians can consistently fit the whole range of fluctuations, not only the tails. The asymptotic behaviour of $q$-Gaussians is of
the power-law type with the scaling exponent $\alpha \equiv \alpha_{qG}$ uniquely determined by $q$ according to the relation

$$\alpha_{qG} = \frac{3 - q}{q - 1}. \quad (11)$$

In particular, $q \geq 5/3$ corresponds to $\alpha_{qG} \leq 2$, i.e. to the Lévy-stable regime.

We consider trading volume distributions for different stocks listed in the WSE. For each stock, we compare the cumulative distributions obtained for different time scales $\Delta t$ or different number of aggregated trades $n_T$. All the cumulative distributions are fitted by appropriate $q$-Gaussian cdfs [45]. Exemplary results are shown in Fig. 8. The agreement between the empirical data and the fits is encouraging: for all the time lags considered in our study (shown and not shown), one obtains a good theoretical representation of the data over the whole range of values. In each case, the inaccuracy does not exceed a few largest events. The $q$-Gaussian fits to the tick-by-tick data are also quite good. Moreover, all values of the empirical scaling exponents ($\alpha_V$) are similar to values of $\alpha_{VqG}$ for appropriate $qs$.

Fig. 8. Log-log plot of the cumulative distributions of the aggregated trading volume for different time scales $\Delta t = 1, 15$, and 120 min. (left), as well as for different number of consecutive transactions $n_T = 1, 15$, and 30 (right). The solid lines represent the best theoretical fits of the cumulative $q$-Gaussians [45] with appropriate $qs$. 
4. Summary

In the present contribution, we focused our attention on a relationship between large returns and trading volume based on data from the Warsaw Stock Exchange, which is an emerging market. Our intention was to investigate whether any systematic relation between the distributions of returns and volumes exists in this case, similar to outcomes of earlier works which were focused on larger markets like the American, Chinese, London, and Paris ones.

We have shown that relation (2) with $\xi \approx 2$ (or $\approx 2.5$ based on the Hill estimator) holds for majority of the analysed the WSE stocks. We also observed that the price impact of trading volume can be modelled by a square-root function (as described by Eq. (3) with $\beta = 1/2$). These outcomes go in parallel with some earlier studies (see especially Ref. [18, 25, 26, 31, 32]). Another curious result of our study worth further investigation is that one can observe a relation between the local slopes of both types of the distributions even if they do not display any power-law tails. However, this result was obtained only for a single company and, therefore, we do not want to draw any decisive conclusions based on it yet.

An additional interesting observation done in our work is that not only the distributions of returns but also the distributions of trading volumes can be well described by the $q$-Gaussian functions.

REFERENCES