Ekonomia Menedżerska
2007, nr 1, s. 143–163

Henryk Gurgul*, Roland Mestel**, Tomasz Wójtowicz*

Distribution of Volume on the American Stock Market

1. Introduction

It is well known that trade on stock market can take place, when market participants assign different values to equity. Relations on stock market become more complex, when there is a number of assets and a number of diverse traders. It is necessary to take into account not only heterogeneity of traders but also changing environment. A good understanding of trading volume is very important for both traders and researchers. Usually at least 3 reasons are cited for the importance of trading volume. The first one is that there is a contradiction between the homogenous trader assumption and positive trading volume. In this model it is assumed that trade without new information is possible. The only reason for trading is the motive to speculate. It is very common that volume data are reported together with financial data, especially with price data. In the literature it is not clear what kind of information is reflected in volume data. Some researchers express doubt about any link between prices, trading volume and volatility. The reason may be the fact that volume increase can be caused by different interpretations of the same information by investors, or even by identical interpretation by investors who nevertheless behave differently because of differing initial expectations.

A third reason is the impact of market imperfection on trading volume. In this context the effect of the institutional form of the market on trading volume may be not correctly understood. The widespread view is that trading volume on imperfect market is lower that in a perfect market. This view is confirmed by

* Wydział Zarządzania, Akademia Górniczo-Hutnicza w Krakowie
** Institut für Banken und Finanzierung, University of Graz

results of empirical investigations. These results imply that new information does not always cause immediately clearing of the markets. The second explanation is that investors frequently make mistakes, which can be corrected in following subperiods. This increases trading activity in the whole period under consideration.

Since investors are heterogeneous in their interpretations of new information, prices may remain unchanged even though new information is revealed to the market. This will be the case if some investors interpret the news as good whereas others find it to be bad. Another situation in which relevant new information may leave stock prices unchanged can be when, although investors interpret the information identically, they start with diverse prior expectations. From this it follows that changes in stock prices reflect an aggregation or averaging of investors' adapted beliefs.

On the other hand, stock prices may only change if there is positive trading volume. One important question arising from this is whether volume data are simply a descriptive parameter of the trading process, or whether they contain unique information that can be exploited for modelling stock returns or return volatilities. As with prices, trading volume and volume changes mainly reflect the available set of relevant information on the market. Unlike stock prices, however, a revision in investors' expectations always leads to an increase in trading volume, which therefore reflects the sum of investors' reactions to news. Because this process leads to higher trading volume, it preserves the differences between investors' reactions to the arrival of new information, differences that may get lost in the averaging process that fixes prices. Studying the joint dynamics of stock prices and trading volume therefore improves our understanding of the dynamic properties of stock markets.

Trading volume is not examined in the literature as often as prices behaviour over time. Although there is no consistent theory of volume, there are some important contributions to this subject, which will be reviewed in next section.

The rest of this paper proceeds as follows. The main ideas in the context of important volume contribution are reviewed in the next section. The third section describes the basic properties of the distributions examined. A short description of the data and their statistical properties are given in the fourth and fifth sections, respectively. The sixth section summarizes the empirical results of distribution fitting and goodness-of-fit tests. A final section concludes the paper.

2. Previous contributions to trading volume

While investigations of the distributions of stock return data started very early [23, 35], the first contributions to volume distribution appeared considerably later

[1]. Trading volume appears in the literature in at least three settings: its relation to the bid-ask spread, its relation to price changes, and its relation to information. Previous contributions uniformly indicate that volume is negatively related to the bid-ask spread. This finding is consistent with some theoretical models [17, 21].

Several theoretical models consider the relation of trading volume to price changes. T.W. Epps [20] formulated a model where the volume on transactions in which the price ticks up is greater than the volume on downticks.

T. Copeland [16] derived a model of the sequential arrival of information to investors. By mean of simulation he demonstrated that after new information, volume is positively related to price changes. Copeland's theory suggests that where new information is disseminated sequentially, rather than simultaneously, to investors, there is a sequence of transitional price equilibria that are accompanied by persisting high trading volume. One important implication of this model is the existence of positive contemporaneous as well as causal relations between price volatilities and trading activities.

The widely reported fat-tailed return distributions are mainly explained by the Clark hypothesis [14]. According to this hypothesis the distribution of prices and volume is jointly exposed to an unobservable directing process. Clark's model was extended by T.W. Epps and M.L. Epps [22], G. Tauchen and M. Pitts [44] and T.G. Andersen [2]. The underlying hypothesis of Clark's model is known in the literature as the mixture of distribution hypothesis (MDH). The MDH implies strong positive contemporaneous linkages, but no causal ones between volume and return volatility data, whereas return levels and volume data display no interactions [30]. C.G. Lamoureux and W.D. Lastrapes [31] examined Clark's hypothesis using the GARCH model extended by a volume variable. The authors by means of the time series of single U.S. stocks demonstrated that volume explains GARCH effect. T.W. Epps and M.L. Epps [22] have suggested that volume moves with measures of within-day price variability because the distribution of transaction price change is a function of volume. Andersen [2] uses in his contribution a rational expectation model, and assumes non-informational trading and common information arrivals. He demonstrates that this improves the empirical fit of the implemented moment restrictions. The dynamic version of his model reduces the estimated volatility persistence of U.S. stocks.

In a framework where stock prices are assumed to be noisy, some recent studies by L. Blume et al. [9] and M. Suominen [42] argue that trading volume conveys unique information to the market, which is not contained in prices. The former model assumes that informed traders reveal their private information to the market through trades, and that uninformed traders learn from volume data about the precision and dispersion of an informational signal. This model implies that return volatility and trading volume exhibit time persistence even when information arrivals do not.

R. Gallant et al. [24] in their extensive study estimated joint conditional return-volume density for the U.S. market using time series over 57 years. To make the time series stationary they detrended return series and volume series as well. Moreover they apply seasonal adjustments of both time series. The contributors found a positive (but non-linear) dependence between return volatility and volume. This relation is almost constant for negative standardized volume data, and has a tendency to increase for positive standardized volume realizations. C.G. Lamoureux and W.D. Lastrapes [32], by means of *the bivariate mixture model*, find no support for the idea that volume data has explanatory power for volatility persistence.

In the light of the reviewed contributions, large volumes and large price changes (positive or negative) can be attributed to information flows (the sequential information arrival model), or to a common directing process that can be interpreted as the flow of information (MDH). The large cost of taking a short position gives an explanation for the observation that, in stock markets, the volume connected with a price increase usually exceeds that connected with an equal price decrease, since costly short sales restrict some investors to trading on the basis of new information.

A common conclusion from these models is that trading volume not only describes market behaviour but actually affects it since it directly enters into the decision process of market participants. In this sense a strong relationship (contemporaneous as well as causal) between volume and return volatility is suggested.

In addition to the theoretical approaches described above, some empirical studies have been performed which deal with volume-price relations on capital markets. The relationship between trading volume and price changes per se was investigated by C. Hiemstra and J.D. Jones [29], T.J. Brailsford [10] and B.S Lee and O.M. Rui [33], mainly using index data. The results of these studies differ in detail, but on the whole they deliver evidence to support a positive volume-price relationship. A few studies also use the returns and volume of individual stocks e.g. R.L. Antoniewicz [3] who finds that the returns of individual stocks on high-volume days are more sustainable than are returns on low-volume days. S.E. Stickel and R.E. Verrechia [43] find that when earnings announcements are accompanied by higher volume, returns are more sustainable over the following days. In the model of G.O. Orosel [39], high stock returns lead investors who do not participate in the stock market to increase their estimate of the profitability of stock market participation. Results by T. Odean and S. Gervais [37] imply that a positive market return should lead to greater volume. T. Chordia and B. Swaminathan [12] investigate the role of trading volume in the cross-autocorrelation patterns observed in stock returns. The authors observe that returns of stocks with high trading volume precede returns of stocks with lower trading volume. This is seen to confirm the speed of adjustment hypothesis, which suggests that high volume stocks adjust quicker to new information than low volume stocks. T. Chordia *et al.* [13] report a negative cross-sectional relationship between expected stock returns and both the level and the changes of trading volume.

In a more recent contribution by M.D. McKenzie and R.W. Faff [36] the authors analyze the relations between trading volume and the autocorrelation properties of daily stock returns. They notice the significant impact of trading volume on time varying autocorrelations in stock returns. In this study the authors report that higher trading volume is mostly accompanied by a drop in return autocorrelation. Recently, R. Connolly and C. Stivers [15] investigated the autocorrelation properties of stock returns in relation to abnormal turnover on a weekly basis. They found momentum in stock returns in the case of contemporaneous, abnormally high trading volume, and averse in returns in the case of abnormally low trading volume.

The relation between stock return volatility and trading volume has been analyzed in a more recent contribution by H. Bessembinder and P.J. Seguin [7], W.A. Brock and B.D. LeBaron [11], S. Avouyi-Dovi and E. Jondeau [4], B.S. Lee and O.M. Rui [33], H. Gurgul *et al.* [27] and [28]. These contributors uniformly confirm a strong relationship (contemporaneous as well as dynamic) between return volatility and trading volume. However, the investigations conducted by A.F. Darrat *et al.* [18], based on intraday data from DJIA stocks, only provide evidence of significant lead/lag relations, but do not confirm a contemporaneous correlation between return volatility and trading volume.

C.G. Lamoureux and W.D. Lastrapes [31], mentioned above, were the first to apply stochastic time series models of conditional heteroscedasticity (GARCH-type) to explore the contemporaneous relationship between volatility and volume data. The authors found that persistence in stock return variance vanishes for the most part when trading volume is included in the conditional variance equation. If trading volume is considered to be an appropriate measure for the flow of information into the market, then this finding is consistent with the MDH. However, one has to realize that the observation by C.G. Lamoureux and W.D. Lastrapes [31] is mainly proof of the fact that trading volume and return volatility are driven by identical factors, leaving the question of the source of the joint process largely unresolved. This GARCH cum volume approach has been applied and extended in several subsequent studies, such as [26, 32, 2, 10, 38]. M. Glaser and M. Weber [25] find that investors who think that they have above average investment skills (but who do not have above average returns) trade significantly more.

An extensive study of stock return distribution for some stock markets is presented by A. Peiro [40] and E. Eberlein and U. Keller [19].

This contribution investigates the statistical properties of trading volume, especially distribution fit. Unlike most other studies on this issue, we use individual stock data instead of index data. Empirical distributions of common stock returns and abnormal returns have been studied extensively. Analogous information on trading volume is very sparse. Specification of empirical distributions is important because they enable researchers to:

– identify measures for which the distributions of abnormal trading volume approximate normality,
– define the level of misstatements in achieved significance levels when performing typical statistical tests based on the t-distribution in trading volume studies,
– find out how statistical tests are affected by the length of the estimation period, the length of prediction interval, clustering of events and the size of the firm.

In practice [34] there are some several measures of trading activities for individual stocks:

1) number of trades per period,
2) share volume $X_{j,t}$,
3) dollar volume $P_{j,t}X_{j,t}$,
4) relative dollar volume $P_{j,t}X_{j,t}\Big/\sum_j P_{j,t}X_{j,t}$,
5) share turnover (turnover ratio) $\tau_{j,t} = X_{j,t}/N_{j,t}$,
6) dollar turnover $\nu_{j,t} = P_{j,t}X_{j,t}/P_{j,t}N_{j,t} = \tau_{j,t}$.

One can see that the last two measures are equal. The most common measures used in empirical investigations are given by 2, 3 and 5. To measure aggregate trading activity similar measures can be defined. Although the definition of dollar turnover may seem redundant since it is equivalent to share turnover, it becomes more relevant in the portfolio case.

It is well known in the literature that return data and especially volume data exhibit clustering, autocorrelation (long memory), fat tails, leptokurtosis and are positively skewed; i.e. they are highly non-normal.

According to A. Timmermann and C.W.J. Granger [45] the most promising direction for investigations of stock market quality in the context of the efficient market hypothesis (EMH) is not just modelling the first conditional moment of returns or volume data, but an estimation of the full conditional distribution of returns or volume dependent on the given information in the last time period.

According to the authors, the EMH does not imply that all changes in this density are unpredictable. It does, however, require that certain functions of the probability distribution are not predictable. As a concrete example, there is now substantial evidence that volatility of asset returns varies over time in a way that can be partially predicted. For this reason there has been considerable interest in improved volatility forecasting models in the context of option pricing. Does this violate market efficiency?

Clearly the answer is no unless a trading strategy could be designed that would use this information in the options market to identify under- and over-valued options. If options markets are efficient, option prices should incorporate the best volatility forecasts at all points in time. To the knowledge of A. Timmermann and C.W.J. Granger [45] no similar results exist yet for the full predictive density of asset returns and volume. However according to those authors it is likely that methods now being developed for predicting the conditional skew, kurtosis and higher order moments of asset returns will also find some use in tests of market efficiency. From this point of view a first, very important step towards developing such a theory is proper distribution fitting to returns and volume data.

While B. Ajinkya and P. Jain [1] found non-normality in raw volume index data, according to these authors the log-volume of index data converges to a normal distribution as the number of securities per portfolio increases. They suggest log-volume as the most proper measure of trading activity also for individual stocks. Therefore, the main task of our paper is to perform a distribution fit to the return and log-volume data of companies listed in DJIA as well. We extract 23 companies and test the log-volume distribution of each of them in the whole period from August 1997 to October 2004, and in two subperiods (August 1997–February2001; March 2001–October 2004).

In the next section we review briefly the properties of potential distributions and their suitability to fit empirical return and volume distribution

3. Distributions

In order to find the best fit several distributions are considered: scaled Student t distribution, exponential power distribution, logistic distribution, hyperbolic distribution, normal inverse Gaussian distribution and α-stable distribution. These distributions have been chosen because they are extensively used in the analysis of the unconditional distribution of stock returns.

Scaled Student *t* distribution has a density function:

$$f(x)=\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu\sigma^2}}\left[1+\frac{(x-\mu)^2}{\nu\sigma^2}\right]^{-\frac{\nu+1}{2}},$$

where Γ denotes the gamma function, μ is a location parameter, $\sigma > 0$ is a scale parameter, and $\nu > 0$ is a shape parameter usually referred to as degrees of freedom. A scaled Student *t* distribution is a generalisation of Student *t* distribution in the sense that if *X* follows a scaled Student *t* distribution, then μ and σ are the mean and standard deviation respectively, and (*X*–μ)/σ follows a Student *t* distribution with ν degrees of freedom. Scaled Student *t* distribution was proposed in the analysis of stock returns by P.D. Praetz [41] and R. Blattberg and N. Gonedes [8].

The exponential power distribution has a density function:

$$f(x)=\frac{\exp\left[-\frac{1}{2}\left|\frac{x-\mu}{\alpha}\right|^{\frac{2}{1+\beta}}\right]}{2^{\frac{3+\beta}{2}}\alpha\Gamma\left(\frac{3+\beta}{2}\right)},$$

where μ, α > 0 and –1 < β ≤ 1 are location, dispersion and shape parameters, respectively. The latter can be seen as a measure of "non-normality" distribution. If β > 0, the distribution is leptokurtic and if β < 0, it is platykurtic. When β = 0 distribution is equal to normal and when β tends to –1 the distribution converges to uniform distribution. If *X* follows an exponential distribution, then $E(X)=\mu$ and $\mathrm{var}\,(X)=\alpha^2\,\Gamma\left(\frac{3(1+\beta)}{2}\right)\Big/\Gamma\left(\frac{1+\beta}{2}\right)$.

Logistic distribution has a density function

$$f(x)=\frac{\exp\left(\frac{x-\mu}{\alpha}\right)}{\alpha\left(1+\exp\left(\frac{x-\mu}{\alpha}\right)\right)^2},$$

where μ is mean and α > 0 is a scale parameter. If *X* follows a logistic distribution, then $E(X)=\mu$ and $\mathrm{var}(X)=\frac{\pi^2}{3}\alpha^2$.

One possible parameterisation of the PDF of **the hyperbolic distribution** can be written in the form:

$$f(x)=\frac{\sqrt{\alpha^2-\beta^2}}{2\alpha\delta K_1\left(\delta\sqrt{\alpha^2-\beta^2}\right)}\exp\left\{\beta(x-\mu)-\alpha\sqrt{\delta^2+(x-\mu)^2}\right\},$$

where K_1 is a modified Bessel function of the third kind with a parameter 1, μ is a location parameter, δ > 0 is a scale parameter and 0 ≤ |β| < α are shape parameters. α is responsible for the steepness, and β for skewness. The hyperbolic distribution was introduced by Barndorff-Nielsen [5]. In [19] this distribution fits well to German return data.

The density of NIG (**the Normal Inverse Gaussian)** is given by:

$$f(x)=\frac{\alpha\delta K_1\left(\alpha\sqrt{\delta^2+\left(x-\mu\right)^2}\right)}{\pi\sqrt{\delta^2+\left(x-\mu\right)^2}}\exp\left\{\beta\left(x-\mu\right)+\delta\sqrt{\alpha^2-\delta^2}\right\}$$

The interpretation of the parameters of the NIG is identical with those of the hyperbolic distribution. The NIG distribution is able to model symmetric and asymmetric distribution with possibly heavy tails. Its tails are often classified as "semi-heavy". If α tends to zero, the NIG distribution converges to a heavy-tailed Cauchy distribution with location parameter μ and scale parameter δ. Normal inverse Gaussian distributions were introduced by Barndorff-Nielsen [6] as a subclass of generalised hyperbolic laws with parameter $\lambda = -1/2$.

The class of **α-stable distributions** introduced by Lévy (1924) has no closed formula of density for all but three values of the parameter α: α = 1/2 (Lévy distribution), α = 1 (Cauchy distribution) and α = 2 (Gaussian distribution). Hence, stable distributions are defined instead in terms of their characteristic functions. The most often used parametrisation of the characteristic function of an α-stable distribution is:

$$\Phi(t)=\begin{cases}\exp\left\{-\gamma^\alpha\left|t\right|^\alpha\left(1-i\beta\,\mathrm{sgn}(t)tg\dfrac{\pi\alpha}{2}\right)+i\delta_1 t\right\} & \alpha\neq 1\\ \exp\left\{-\gamma^\alpha\left|t\right|\left(1+i\beta\dfrac{2}{\pi}\mathrm{sgn}(t)\ln(t)\right)+i\delta_1 t\right\} & \alpha=1\end{cases}$$

where $0 < \alpha \leq 2$ is an index of stability, $-1 \leq \beta \leq 1$ is a skewness parameter, $\gamma > 0$ is a scale parameter and δ_1 is a location parameter. As mentioned above, for $\alpha = 2$, the distribution is Gaussian. While if $0 < \alpha < 2$, the distribution has a fatter tail than in the Gaussian. If X follows an α-stable distribution with $\alpha > 1$, then $E\left(X\right)=\delta_1$.

A parameterisation of the characteristic function which is more useful in applications, is given by:

$$\Phi(t)=\begin{cases}\exp\left\{-\gamma^\alpha\left|t\right|^\alpha\left(1+i\beta\,\mathrm{sgn}(t)tg\dfrac{\pi\alpha}{2}\left((\gamma\left|t\right|)^{1-\alpha}-1\right)\right)+i\delta_0 t\right\} & \alpha\neq 1\\ \exp\left\{-\gamma\left|t\right|\left(1+i\beta\dfrac{2}{\pi}\mathrm{sgn}(t)\left(\ln\left|t\right|+\ln\gamma\right)\right)+i\delta_0 t\right\} & \alpha=1\end{cases}.$$

The value of this representation is that characteristics (and densities) are jointly continuous in all four parameters. The parameters α, β and γ have the same meaning for the two parameterisations, while the location parameters are related by:

$$\delta_1 = \delta_0 - \beta\gamma tg\frac{\pi\alpha}{2},$$

if $\alpha \neq 1$
and

$$\delta_1 = \delta_0 - \beta\gamma\frac{2}{\pi},$$

when α = 1.

It is worth noting that unlike all the distributions described above, members of the class of **α-stable distributions,** except the Gaussian distribution, have infinite variance. From the empirical investigations of E. Eberlein and U. Keller [19] it follows that this kind of distribution applies to returns on the American stock market.

4. Data

Our data set (supplied by *Reuters Austria* and *Deutsche Börse*, controlled by us for mistakes and adjusted for dividends and stock splits) comprises the daily percentage rates of return and the natural logarithms of trading volume series for 23 companies listed in the DJIA in the whole period from August 1997 to October 2004. For each company calculations exclusively concentrate on the period of its DJIA membership. Therefore it was possible to extract 23 companies over the whole above-mentioned period. Continuously compounded stock returns are calculated from daily stock prices at close, adjusted for dividend payouts and stock splits. In order to check the robustness of our results in respect to sample size, the whole period has been divided into two sub periods: from August 1997 to February 2001 and from March 2001 to October 2004.

5. Descriptive statistics

We start our investigations with some basic descriptive analysis of the time series under study. As can be seen from Panel A of Table 1 the mean of daily percentage stock returns over the whole period ranges from −0.035% (HP) to +0.058% (WalMart) with a median of 0.016% (Procter & Gamble).

Table 1

Aggregated summary statistics for stock market data of DIJA companies

Panel A: Daily percentage stock returns				
	Mean	**Std.dev.**	**Skewness**	**Kurtosis**
Min	−0.034549	1.6457	−3.8872	5.0711
1. Quartile	−0.015194	2.0112	−0.36585	5.6813
Median	0.01629	2.2308	−0.12103	7.4464
3. Quartile	0.028528	2.4299	0.068617	10.591
Max	0.057856	3.119	0.2241	74.44
Panel B: Daily log-volume				
	Mean	**Std.dev.**	**Skewness**	**Kurtosis**
Min	14.2	0.36	−0.28	3.56
1. Quartile	14.8	0.41	0.09	3.91
Median	15.4	0.44	0.25	4.18
3. Quartile	15.6	0.46	0.37	4.64
Max	16.5	0.63	0.77	6.26

The well-known fact of a “fat-tailed and highly-peaked” distribution of return series is mostly present in our data. The smallest kurtosis of stock returns is equal to 5.07 (Caterpillar) while the largest is 74.44 (Procter & Gamble). Furthermore, a negative median of skewness (−0.12 for McDonald’s) means that majority of the examined stock returns are left-skewed.

Despite lower values of kurtosis, log-volume series as well as returns show fat tails. On the other hand, skewness of log-volume shows the opposite pattern –the majority of samples are right-skewed (lower quartile equals 0.09).

Observed values of skewness and kurtosis are the main reason for the non-normal distribution of both returns and log-volume series. In fact, the Jarque-Bera test rejects normality in all cases except one (log-volume of AmEx from August 1997 to February 2001).

6. Empirical results

The parameters of the considered distributions have been estimated by the maximum likelihood method.

All necessary computations have been carried out in Matlab, however the parameters of α-stable distribution have been estimated following the method of Nolan (1997) in his program STABLE [46]. The results of these estimations are summarised in table 2, which contains the main descriptive statistics for all estimated distribution parameters (including cases where the distribution was rejected by goodness-of-fit tests). (Detailed results are not presented here because of the lack of space. They are available upon request).

With regard to the estimation of the parameters, some facts should be pointed out. The computed estimates of shape parameter β in the power exponential distribution are greater than zero. In the case of returns series they are, in fact, in the whole period greater than 0.5. This means highly leptokurtic distributions. In comparison to the return series, estimates of β for log-volume series are smaller but still positive. In the whole period they range from 0.14 to 0.57, which indicates semi-heavy tails.

A similar conclusion can be deduced from the estimated values of shape parameter α of hyperbolic and NIG distributions: for returns they are close to zero, whereas for log-volumes they are more distant. This once again confirms the fat-tailednes of stock returns, and the semi-heavy tails of log-volume distributions. Moreover, estimated degrees of freedom in Student t distribution, ranging in the main period from 3.28 to 6.14, and from 5.44 to 16.3 for return and log-volume series respectively confirm nonnormality of examined data.

In order to clarify the properties of data series over time, we compare the values of estimated parameters in each of the subperiods. If we consider daily returns, visible differences between descriptive statistics can be observed for Student's *t*, exponential and hyperbolic distribution. The values of degrees of freedom in Student *t* distribution are significantly greater in the first subperiod (August 1997–February 2001) than in the other considered period. Moreover, in the same period, shape parameter β in the power exponential distribution has its lowest values. On the other hand, hyperbolic distribution has significantly greater estimated values from March 2001 to October 2004.

If we take into account the interpretation and role of individual distribution parameters, it will be clear that all the differences mentioned above indicate that daily returns in the subperiod from August 1997 to February 2001 are closer to normality, than returns in the whole period or in the second subperiod. An analogous conclusion can be drawn from estimated parameter values of log-volume distributions. The values of degrees of freedom in Student *t* distribution, and of β of power exponential distribution in the second subperiod in connection with higher values of α in NIG distribution in the first one suggest that the log-volume data are much more leptokurtic in the second subperiod from March 2001 to October 2004 than in the earlier one.

Table 2

Aggregated summary statistics for estimated distributions parameters.

Panel A: Daily percentage stock returns													
		03.01–10.04				08.97–02.01				08.97–10.04			
		min	median	mean	max	min	median	mean	max	min	median	mean	max
logistic	μ	−0.06	−0.01	0.00	0.06	−0.10	0.00	0.01	0.09	−0.07	0.00	0.00	0.06
	α	0.66	1.05	1.03	1.49	0.99	1.30	1.29	1.74	0.89	1.18	1.16	1.62
Student's t	μ	−0.07	−0.01	0.00	0.06	−0.10	0.00	0.01	0.09	−0.08	0.00	0.00	0.06
	σ	0.92	1.43	1.43	1.99	1.55	1.89	1.93	2.49	1.23	1.66	1.66	2.22
	υ	2.78	4.02	4.36	7.39	4.06	6.00	6.22	10.17	3.28	4.57	4.68	6.14
exponential	μ	−0.09	−0.01	−0.01	0.07	−0.14	0.00	0.00	0.1	−0.11	0.00	−0.01	0.05
	α	0.59	0.87	0.91	1.53	0.97	1.49	1.44	1.87	0.64	1.08	1.10	1.50
	β	0.38	0.83	0.76	1.00	0.27	0.49	0.52	0.86	0.47	0.69	0.70	1.00
hyperbolic	α	0.53	0.80	0.85	1.26	0.49	0.72	0.79	1.32	0.50	0.76	0.76	1.11
	β	−0.11	0.00	0.00	0.09	−0.05	0.03	0.04	0.15	−0.04	0.02	0.02	0.08
	δ	0.00	0.73	0.85	2.31	1.09	1.95	2.02	3.21	0.43	1.22	1.14	1.94
	μ	−0.25	−0.02	−0.02	0.24	−0.64	−0.11	−0.16	0.25	−0.41	−0.05	−0.7	0.13
NIG	α	0.28	0.49	0.53	0.92	0.33	0.54	0.58	1.12	0.30	0.51	0.50	0.82
	β	−0.09	0.00	0.00	0.09	−0.04	0.02	0.03	0.14	−0.04	0.02	0.01	0.07
	δ	1.23	1.81	1.93	3.47	2.26	3.08	3.12	4.25	1.46	2.38	2.30	3.18
	μ	−0.24	−0.02	−0.02	0.21	−0.59	−0.09	−0.14	0.20	−0.38	−0.06	−0.06	0.13
α-stable	α	1.55	1.71	1.73	1.87	1.73	1.85	1.83	1.91	1.63	1.76	1.77	1.86
	β	−0.29	0.07	0.05	0.35	−0.15	0.11	0.14	0.52	−0.16	0.09	0.10	0.34
	γ	0.73	1.15	1.14	1.61	1.19	1.47	1.49	1.97	1.00	1.32	1.31	1.78
	δ	−0.11	−0.01	−0.01	0.09	−0.16	0.00	−0.01	0.08	−0.14	−0.02	−0.02	0.07
Panel B: Daily log-volume													
		03.01–10.04				08.97–02.01				08.97–10.04			
		min	median	mean	max	min	median	mean	max	min	median	mean	max
logistic	μ	14.4	15.4	15.4	16.6	14.0	15.3	15.2	16.3	14.2	15.4	15.3	16.5
	α	0.14	0.21	0.21	0.26	0.20	0.24	0.25	0.36	0.20	0.24	0.24	0.35
Student's t	μ	14.4	15.4	15.4	16.6	14.0	15.3	15.2	16.3	14.2	15.4	15.3	16.5
	σ	0.21	0.32	0.32	0.43	0.31	0.38	0.39	0.59	0.31	0.38	0.38	0.55
	υ	3.68	6.00	6.77	12.8	4.82	9.35	10.1	20.0	5.44	8.61	8.73	16.3
exponential	μ	14.4	15.4	15.4	16.6	14.0	15.3	15.2	16.3	14.2	15.4	15.3	16.5
	α	0.16	0.25	0.25	0.38	0.25	0.34	0.34	0.55	0.25	0.30	0.32	0.46
	β	0.21	0.48	0.46	0.84	0.01	0.28	0.30	0.59	0.14	0.33	0.34	0.57
hyperbolic	α	4.07	5.24	5.23	6.71	3.91	5.31	5.65	10.9	3.43	5.03	5.04	6.34
	β	0.18	0.93	0.97	2.08	−1.01	1.04	1.12	3.72	−0.97	0.71	0.63	1.84
	δ	0.19	0.36	0.38	0.83	0.30	0.59	0.70	2.12	0.29	0.54	0.59	1.22
	μ	14.4	15.2	15.2	16.4	13.4	15.0	15.0	16.4	14.3	15.2	15.2	16.3

Table 2 cont.

NIG	α	2.79	3.98	4.08	5.29	2.82	4.60	4.90	10.7	2.52	4.24	4.19	5.82
	β	0.17	0.88	0.92	2.00	−1.03	0.97	1.11	4.25	−0.95	0.69	0.62	1.88
	δ	0.31	0.53	0.55	1.00	0.52	0.76	0.88	2.21	0.53	0.71	0.76	1.34
	μ	14.4	15.2	15.2	16.4	13.3	15.0	15.0	16.4	14.3	15.2	15.2	16.3
α-stable	α	1.62	1.86	1.85	2.00	1.68	1.86	1.86	2.00	1.84	1.87	1.88	1.99
	β	−1.00	0.54	0.33	0.86	−0.93	0.61	0.50	1.00	−0.82	0.44	0.36	1.00
	γ	0.16	0.25	0.25	0.32	0.23	0.28	0.29	0.42	0.23	0.29	0.29	0.44
	δ	14.4	15.4	15.4	16.5	14.0	15.2	15.1	16.3	14.2	15.4	15.3	16.4

The adequacy of the estimated distribution is examined with Kolmogorov–Smirnov (KS) and chi-square goodness-of-fit tests. These tests were chosen because they are commonly used in distribution fitting and their critical values are known and tabulated. Values of the cumulative distribution functions, needed in testing procedures, are estimated via numerical integration. In order to compute chi-square goodness-of-fit statistics, the real line has been partitioned into disjointed intervals of equal probability. The numbers of intervals were proportional to the sample length and were equal to 38 for whole period, and 28 for sub periods. In both tests a 5% significance level was considered. The detailed results of goodness-of-fit tests are not presented here.

However, some remarks about them should be noted. All considered distributions were rejected by one of the goodness-of-fit tests in four return samples of returns (Alcoa and IBM in period 08.97–10.04, Honeywell and Johnson&Johnson in 08.97–02.01), and in the case of five samples of log-volume (Honeywell in period 08.97–10.04; Honeywell and HP in 08.97–02.01; Citigroup and GE in 03.01–10.04).

For the majority of samples, more than one distribution is fitted. In such a situation a criterion for choosing the best distribution is needed. As is well known, the Kolmogorov–Smirnov statistic is equal to the greatest distance between the empirical and the theoretical cumulative distribution functions.

Thus, it could be said that the distribution that has the smallest value of KS statistics describes the sample in the best way. This can be true only in the central part of the sample. But if the tails are of interest, the Anderson–Darling statistic should be considered instead of the Kolmogorov–Smirnov, because it better describes behaviour of the distribution on tails. Hence the Anderson–Darling statistic should be also taken into account in describing the best fit.

Table 3 reports the best fits based on Kolmogorov–Smirnov and Anderson–Darling statistics respectively. Dashes indicate rejection of all examined distributions.

For the majority of return and log-volume samples, the hyperbolic or NIG distribution have the smallest values of both the KS and AD statistics. It should be noted that in majority of samples the values of Kolmogorov–Smirnov statistics for hyperbolic and NIG distribution were very close. The same is true of Anderson–Darling statistics. If we take this into account, it turns out that NIG and hyperbolic distributions describe the log-volume and stock returns in the best way, in the case of stock returns, this is in accordance with mentioned results from the literature.

Table 3

The best fits based on Kolmogorov–Smirnov and Anderson–Darling statistics.

Panel A: Daily percentage stock returns						
	KS statistic			**AD statistic**		
	08.97– –10.04	**08.97– –02.01**	**03.01– –10.04**	**08.97– –10.04**	**08.97– –02.01**	**03.01– –10.04**
Alcoa	–	NIG	Student's *t*	–	hyperbolic	NIG
Altria	stable	hyperbolic	Student's *t*	stable	hyperbolic	Student's *t*
AmEx	hyperbolic	Student's *t*	hyperbolic	hyperbolic	hyperbolic	NIG
Boeing	Student's *t*	hyperbolic	stable	Student's *t*	stable	Student's *t*
Caterpillar	hyperbolic	hyperbolic	NIG	hyperbolic	hyperbolic	NIG
Citigroup	Student's *t*	stable	stable	Student's *t*	stable	Student's *t*
CocaCola	hyperbolic	hyperbolic	NIG	NIG	stable	NIG
Disney	Student's *t*	hyperbolic	Student's *t*	Student's *t*	stable	Student's *t*
DuPont	hyperbolic	hyperbolic	hyperbolic	hyperbolic	hyperbolic	hyperbolic
Exxon	hyperbolic	stable	stable	Student's *t*	stable	stable
GE	Student's *t*	stable	NIG	NIG	stable	NIG
GM	NIG	hyperbolic	NIG	NIG	stable	NIG
Honeywell	NIG	–	NIG	Student's *t*	–	Student's *t*
HP	NIG	Student's *t*	Student's *t*	Student's *t*	Student's *t*	Student's *t*
IBM	–	stable	NIG	–	stable	NIG
Johnson & Johnson	hyperbolic	–	hyperbolic	hyperbolic	–	hyperbolic
JPMorgan chase	exponen- tial	hyperbolic	exponen- tial	hyperbolic	hyperbolic	exponen- tial
McDo- nald's	NIG	NIG	hyperbolic	NIG	NIG	Student's *t*
3M	NIG	hyperbolic	Student's *t*	NIG	NIG	stable
Merck	Student's *t*	stable	Student's *t*	Student's *t*	stable	Student's *t*

Table 3 cont.

Procter & Gamble	Student's *t*	stable	Student's *t*	Student's *t*	stable	Student's *t*
UnitTech	hyperbolic	logistic	hyperbolic	hyperbolic	exponen-tial	Student's *t*
WalMart	NIG	Student's *t*	NIG	NIG	NIG	stable

Panel B: Daily log-volume						
	KS statistic			**AD statistic**		
	08.97– –10.04	**08.97– –02.01**	**03.01– –10.04**	**08.97– –10.04**	**08.97– –02.01**	**03.01– –10.04**
Alcoa	NIG	NIG	NIG	hyperbolic	hyperbolic	NIG
Altria	hyperbolic	hyperbolic	NIG	hyperbolic	hyperbolic	NIG
AmEx	hyperbolic	hyperbolic	hyperbolic	hyperbolic	hyperbolic	stable
Boeing	hyperbolic	NIG	hyperbolic	NIG	NIG	hyperbolic
Caterpillar	hyperbolic	Student's *t*	stable	hyperbolic	NIG	stable
Citigroup	hyperbolic	NIG	–	hyperbolic	NIG	–
CocaCola	hyperbolic	hyperbolic	hyperbolic	hyperbolic	hyperbolic	hyperbolic
Disney	stable	hyperbolic	NIG	stable	hyperbolic	NIG
DuPont	hyperbolic	hyperbolic	exponen-tial	hyperbolic	hyperbolic	hyperbolic
Exxon	hyperbolic	hyperbolic	stable	hyperbolic	hyperbolic	stable
GE	hyperbolic	hyperbolic	–	hyperbolic	hyperbolic	–
GM	hyperbolic	hyperbolic	hyperbolic	hyperbolic	stable	hyperbolic
Honeywell	–	–	stable	–	–	stable
HP	hyperbolic	–	NIG	hyperbolic	–	NIG
IBM	NIG	NIG	hyperbolic	NIG	NIG	hyperbolic
Johnson & Johnson	NIG	NIG	hyperbolic	NIG	hyperbolic	hyperbolic
JPMorgan chase	stable	hyperbolic	hyperbolic	hyperbolic	hyperbolic	hyperbolic
McDo-nald's	hyperbolic	hyperbolic	stable	stable	hyperbolic	stable
3M	hyperbolic	NIG	hyperbolic	NIG	hyperbolic	hyperbolic
Merck	stable	stable	stable	NIG	hyperbolic	NIG
Procter & Gamble	stable	NIG	NIG	stable	NIG	hyperbolic
UnitTech	Student's *t*	hyperbolic	stable	Student's *t*	NIG	stable
WalMart	NIG	NIG	hyperbolic	hyperbolic	hyperbolic	hyperbolic

7. Conclusions

While distributions of stock returns on an index and company basis are well documented in the literature, there are only a few contributions that relate to volume distribution fit. Although there is up to now no general theory of trading activity, especially with respect to returns and volatility, the most frequently used volume proxy is the logarithm of the number of shares traded (log-volume). Therefore this proxy of trading activity was the subject of our investigations. We performed our computations using the relatively short time series of volume, which were available to us after removing errors, stock split and dividend adjustments. For the sake of comparability we chose return data from the same time period.

In the light of our computations, in most cases the best fit to returns can be reached by normal inverse and hyperbolic distributions. This is in line with the results from the literature presented above. Volume data exhibit more autocorrelation than return data, especially long memory property. Therefore, a normal distribution (except in the case of index data) does not fit volume data. Most distributions that fit pretty well to return data fit much worse to volume data. However it follows from our investigations that those distributions which fit relatively well to return data are also likely to fit to the log-volume data of the considered companies, although the quality of fit is in the latter case worse.

Literature

[1] Ajinkya B., Jain P. 1998: *The behavior of daily stock market trading volume*, Journal of Accounting and Economics 11, pp. 331–359.

[2] Andersen T.G. 1996: *Return volatility and trading volume: An information flow interpretation of stochastic volatility*, Journal of Finance 51, pp. 169–204.

[3] Antoniewicz R.L. 1993: *Relative Volume and Subsequent Stock Price Movements*, working paper, Governors of the Federal Reserve System.

[4] Avouyi-Dovi S., Jondeau E. 2000: *International transmission and volume effects in G5 stock market returns and volatility*, BIS Conference Papers No. 8, pp. 159–174.

[5] Barndorff-Nielsen O. 1977: *Exponentially decreasing distributions for the logarithm of particle size*, Proceedings of the Royal Society of London A353, pp. 401–419.

[6] Barndorff-Nielsen O. 1995: *Normal inverse Gaussian processes and the modeling of stock returns,* Research Report # 300, Aarhus University.

[7] Bessembinder H., Seguin P.J. 1993: *Price volatility, trading volume and market depth: evidence from futures markets*, Journal of Financial and Quantitative Analysis 28, pp. 21–39.

[8] Blattberg R., Gonedes N. 1974: *A Comparison of the Stable and Student Distributions as Statistical Models for Stock Prices*, Journal of Business 47, pp. 244–280.

[9] Blume L., Easley D., O'Hara M. 1994: *Market statistics and technical analysis: The role of volume*, Journal of Finance 49, pp. 153–181.

[10] Brailsford T.J. 1996: *The empirical relationship between trading volume, returns and volatility*, Accounting and Finance 35, pp. 89–111.

[11] Brock W.A., LeBaron B.D. 1996: *A dynamic structural model for stock return volatility and trading volume*, The Review of Economics and Statistics 78, pp. 94–110.

[12] Chordia T., Swaminathan B. 2000: *Trading volume and cross-autocorrelations in stock returns*, Journal of Finance 55, pp. 913–935.

[13] Chordia T., Subrahmanyam A., Anshuman V.R. 2001: *Trading activity and expected stock returns*, Journal of Financial Economics 59, pp. 3–32.

[14] Clark P.K. 1973: *A subordinated stochastic process model with finite variance for speculative prices*, Econometrica 41, pp. 135–155.

[15] Connolly R., Stivers C. 2003: *Momentum and reversals in equity-index returns during periods of abnormal turnover and return dispersion*, Journal of Finance 58, pp. 1521–1555.

[16] Copeland T. 1976: *A model of asset trading under the assumption of sequential information arrival*, Journal of Finance 31, pp. 135–155.

[17] Copeland T., Galai D. 1983: *Information effects on the bid-ask spread*, Journal of Finance 31, pp. 1457–69.

[18] Darrat A.F., Rahman S., Zhong M. 2003: *Intraday trading volume and return volatility of the DJIA stocks: A note*, Journal of Banking and Finance 27, pp. 2035–2043.

[19] Eberlein E., Keller U. 1995: *Hyperbolic distributions in finance*, Bernoulli 1, pp. 281–299.

[20] Epps T.W. 1975: *Security price changes and transaction volumes: theory and evidence*, American Economic Review 65, pp. 586–597.

[21] Epps T.W 1976: *The demand for brokers' services: the relation between security trading volume and transaction cost*, Bell Journal of Economics, pp. 163–194.

[22] Epps T.W., Epps M.L. 1976: *The stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distribution hypothesis*, Econometrica 44, pp. 305–321.

[23] Fama E.F. 1965: *The behaviour of stock market prices*, Journal of Business 38, pp. 34–105.

[24] Gallant R., Rossi P., Tauchen G. 1992: *Stock Prices and Volume*, Review of Financial Studies, pp. 199–242.

[25] Glaser M., Weber M. 2004: *Which past returns affect trading volume?*, Working Paper, University of Mannheim.
[26] Glosten L.R., Jagannathan R., Runkle D.E. 1993: *On the relation between the expected value and the volatility of the nominal excess return on stocks*, Journal of Finance 48, pp. 1779–1801.
[27] Gurgul H., Majdosz P., Mestel R. 2004: *Trading volume and stock prices on the Austrian stock market*, Proceedings of the 22nd International Conference "Mathematical Methods in Economics", Czech Society for Operations Research, Czech Econometric Society, Masaryk University Brno, Faculty of Economics and Administration, pp. 195–200.
[28] Gurgul H., Mestel R., Schleicher C. 2003: *Stock market reactions to dividend announcements: empirical evidence from the Austrian stock market*, Financial Markets and Portfolio Management, vol. 17, No 3, pp. 332–350
[29] Hiemstra C., Jones J.D. 1994: *Testing for linear and nonlinear Granger causality in the stock price - volume relation*, Journal of Finance 49, pp. 1639–1664.
[30] Karpoff J.M. 1987: *The relation between price changes and trading volume: A survey*, Journal of Financial and Quantitative Analysis 22, pp. 109–126.
[31] Lamoureux C.G., Lastrapes W.D. 1990*: Heteroscedasticity in stock return data: Volume versus GARCH effects*, Journal of Finance 45, pp. 221–229.
[32] Lamoureux C.G., Lastrapes W.D. 1994*: Endogenous trading volume and momentum in stock-return volatility*, Journal of Business and Economic Statistics 12, pp. 253–260.
[33] Lee B.S., Rui O.M. 2002: *The dynamic relationship between stock returns and trading volume: Domestic and cross-country evidence*, Journal of Banking and Finance 26, pp. 51–78.
[34] Lo A.W., Wang J. 2000: *Trading volume: Definitions, data analysis, and implications of portfolio theory,* Review of Financial Studies 13, pp. 257–300.
[35] Mandelbrot B. 1963: *The variation of certain speculative prices,* Journal of Business 36, pp. 394–419.
[36] McKenzie M.D., Faff R.W. 2003: *The determinants of conditional autocorrelation in stock returns*, Journal of Financial Research 26, 2003, pp. 259–274.
[37] Odean T., Gervais S. 2001: *Learning to be overconfident*, Review of Financial Studies 14, pp.1–27.
[38] Omran M.F., McKenzie, E. 2000: *Heteroscedasticity in stock returns data revisited: Volume versus GARCH effects*, Applied Financial Economics 10, pp. 553–560.
[39] Orosel G.O. 1998: *Participation costs, trend chasing and volatility of stock prices*, Review of Financial Studies 11, pp. 521–557.

[40] Peiro A. 1994: *International evidence on the distribution of stock returns*, Applied Financial Economics 4, pp. 431–439.
[41] Praetz P.D. 1972: *The distribution of share price changes*, Journal of Business 45, pp. 49–55.
[42] Suominen M. 2001: *Trading volume and information revelation in stock markets*, Journal of Financial and Quantitative Analysis 36, pp. 545–565.
[43] Stickel S.E., Verrechia R.E. 1994: *Evidence that Volume Sustains Price Changes*, Financial Analyst Journal (November-December), pp. 57–67.
[44] Tauchen G., Pitts M. 1983*: The price variability-volume relationship on speculative markets*, Econometrica 51, pp. 485–505.
[45] Timmermann A., Granger C.W.J. 2004: *Efficient Market Hypothesis and Forecasting*, International Journal of Forecasting, Elsevier, vol. 20 (1), pp. 15–27.
[46] www.cas.american.edu/~jpnolan

Appendix

List of companies included in the sample

Company	Name	ISIN code
Alcoa	Alcoa Inc.	US0138171014
Altria	Altria Group Inc.	US02209S1033
AmEx	American Express Co.	US0258161092
Boeing	Boeing Co.	US0970231058
Caterpillar	Caterpillar Inc.	US1491231015
Citigroup	Citigroup Inc.	US1729671016
CocaCola	Coca-Cola Co.	US1912161007
Disney	Walt Disney Co.	US2546871060
DuPont	E.I. DuPont de Nemours & Co.	US2635341090
Exxon	Exxon Mobil Corp.	US30231G1022
GE	General Electric Co.	US3696041033
GM	General Motors Corp.	US3704421052
Honeywell	Honeywell International Inc.	US4385161066
HP	Hewlett-Packard Co.	US4282361033
IBM	International Business Machines Corp.	US4592001014
Johnson&Johnson	Johnson & Johnson	US4781601046
JPMorganChase	JPMorgan Chase & Co.	US46625H1005
McDonald’s	McDonald’s Corp.	US5801351017
3M	3M Co.	US88579Y1010
Merck	Merck & Co. Inc.	US5893311077
Pro&Gamble	Procter & Gamble Co.	US7427181091
UnitTech	United Technologies Corp.	US9130171096
WalMart	Wal-Mart Stores Inc.	US9311421039