Physica A 525 (2019) 1343–1350

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

New dynamics between volume and volatility

Zeyu Zheng [a,b,c,1], Jun Gui [a,b,c,d,e,*,1], Zhi Qiao [f,g,1], Yang Fu [a,b,c], H.Eugene Stanley [h], Baowen Li [i]

[a] *Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China*
[b] *Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110016, China*
[c] *Key Laboratory of Network Control System, Chinese Academy of Sciences, Shenyang 110016, China*
[d] *University of Chinese Academy of Sciences, Beijing 100049, China*
[e] *School of Information Science and Engineering, Shenyang University of Technology, Shenyang 110870, China*
[f] *Department of Physics and Centre for Computational Science and Engineering, National University of Singapore, Singapore 117542, Republic of Singapore*
[g] *NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, Singapore 117456, Republic of Singapore*
[h] *Center for Polymer Studies, Boston University, MA 02215, USA*
[i] *Department of Mechanical Engineering, University of Colorado, Boulder, CO 80309, USA*

HIGHLIGHTS

- Multiscale volume-conditional volatility distributions are unified.
- Volume-conditional volatility distributions are well described by power-laws with exponential cutoffs.
- The highest volatility in a given range of volume is strongly correlated with previous-day volume.
- Both volume and volatility have the predicting power for the volatility.

ARTICLE INFO

Article history:
Received 2 January 2019
Received in revised form 12 February 2019
Available online 27 March 2019

Keywords:
Dynamics
Volume
Volatility
Local maximum volatility
Scaling

ABSTRACT

Understanding, quantifying and predicting market fluctuation has become increasingly important in recent decades. Volatility and volume are the two commonly used quantities to study the market dynamics and the relationship between these two has been modeled and debated for years with several hypothesis been put forward. Using empirical data, we investigate the causality and correlation between volume and volatility and find new ways in which they interact, particularly when the levels of both are high. We find that the volume-conditional volatility distribution scales with volume as a power-law function with an exponential cutoff. We exploit the characteristics of a volume-volatility scatterplot and find a strong correlation between logarithmic volume and a quantity we define as local maximum volatility (LMV), the highest volatility observed in a given range of volume. This supports our empirical analysis, showing that volume is an effective parameter for prediction of the maximum value of volatility for both same-day and near-future time periods. The joint conditional probability of volume and volatility also indicates if we invoke both quantities, the prediction of the largest next-day volatility will be better than invoking either one alone. This approach is thus a greatly improved method of risk assessment.

* Correspondence to: Shenyang Institute of Automation, Chinese Academy of Sciences, No.114 Nanta Street, Shenyang 110016, People's Republic of China.
E-mail address: guijun@sia.cn (J. Gui).

1 These authors contributed equally to this work.

https://doi.org/10.1016/j.physa.2019.03.100
0378-4371/

1. Introduction

After recent financial crisis, more and more people are starting to realize that financial market is complex [1,2] bearing lots of instability and risk as volatility changes widely across time. The core part of this field has been focusing on the relationship between the price volatility and trading volume. There have been studies reporting that absolute price change (volatility) and trading volume are positively correlated [3], while others indicate that the correlation is weak [4] and their analyses of time-lag correlations produce a variety of contradictory results [5–12]. The subtleties of the relationship between volume and volatility remain unclear [13] and disagreement persists. For example, Zhou provides some evidence supporting the possibility that volume might play a minor role in extreme price fluctuations at transaction level [14]. Brailsford et al. report a significant cross-correlation between overnight return and trading volume [15]. Brooks indicates that including lagged volume may lead to modest improvements in forecasting performance [16] while Clark shows a nearly parabolic functional relationship between volume and volatility [17], and a popular model developed by Clark holds that volatility could be modeled as a subordinated random process, in which volume, insofar as it affects trading times, accounts for the majority of observed volatility clustering and leptokurtosis (i.e., heavy tails). On the other hand, several studies report that volume is only nominally useful in predicting volatility. Koulakiotis et al. report a negative relationship between volatility and trading volume [18]. Lamoureux and Lastrapes show that ARCH effects tend to disappear (i.e., volatility persistence is lost) when volume is included in the variance equation [6]. Sharma et al. even suggest that price returns of the NYSE are best described by the GARCH model in the absence of volume as a mixing variable [19]. Recently, Gillemot et al. demonstrated that the subordinated random process developed by Clark accounts for, at most, only a small fraction of observed volatility clustering and leptokurtosis [20].

Besides the statistical and empirical studies, several information-based theories have been developed. Clark [17] first developed the mixture of distributions hypothesis (MDH). However, different types of traders and the lagged effect between volume and volatility are not considered in MDH which leads to several new theories including the sequential arrival of information hypothesis (SAIH) [21], the dispersion of belief hypothesis [22], the noise trader hypothesis [23] and etc.

There are two key problems to the above studies: one is that they are more focus on the linear relationship between volume and volatility, lacking of more detailed non-linear relationship despite the heteroskedasticity among the volatilities; the other is that the empirical studies are not well supported by the theoretical work with gaps remaining except for the mixture of distributions hypothesis (MDH) developed by Clark [17]. Other notable theoretical information models including the sequential arrival of information hypothesis (SAIH) [21], the dispersion of belief hypothesis [22] and the noise trader hypothesis [23], but they are having some contradictory results.In order to clearly uncover the underlying non-linear relationship between volume and volatility (absolute price change [24]) and the behavior theories behind, we focus on the most fundamental features of these two quantities, starting by examining the probability density function (pdf) of each, as well as the linear casual relationship between them. Then we go one step further by investigating volume-conditional pdf of volatility in our dataset. Based on these elementary analyses, we show that the pdf for volume-conditional volatility is actually invariant under volume change when the units of volatility are scaled appropriately. This scaling property is useful because it is in line with a similar scaling law found in other complex systems, e.g., atmospheric and biological systems [25,26], giving us a better understanding of the underlying dynamics, which allows us to extrapolate large volume values that correspond to large market fluctuations from fluctuations at small volume values. Here we propose a new pdf that links volatility and volume, investigate the highest volatility distribution value in specified volume regimes, and propose the quantity "local maximum volatility" (LMV). We demonstrate that this quantity is strongly correlated with both the trading volume on a given day and the trading volume on the previous day compared to the normal volatility we used. Also we have found behavior explanations behind the LMV which is the overconfidence, which proves to be a perfect representative of market participant's overreaction on the market fluctuation. Finally we combine volume and volatility and find that the two taken together can be used as a much improved predictor of risk.

2. Materials and methods

We analyze the 30 stocks comprising the Dow Jones Industrial Average, using daily values from the 17-year time period from April 1990 to June 2007, for a total of 130,410 data points. We avoid data after June 2007 due to the potential for high non-stationarity in the volume time series associated with the world financial crisis, although further analysis indicates that our results do not change when post-2007 data is included.

For each of the 30 stocks i, we calculate the daily logarithmic change, commonly referred to as the *return*, of price $p_i(t)$

$$R_i(t) \equiv lnp_i(t) - lnp_i(t-1), \tag{1}$$

and also the daily normalized logarithmic trading volume $\tilde{Q}_i(t)$, calculated from the trading volume $Q_i(t)$ as

$$\tilde{Q}_i(t) \equiv lnQ_i(t) - Y_i, \tag{2}$$

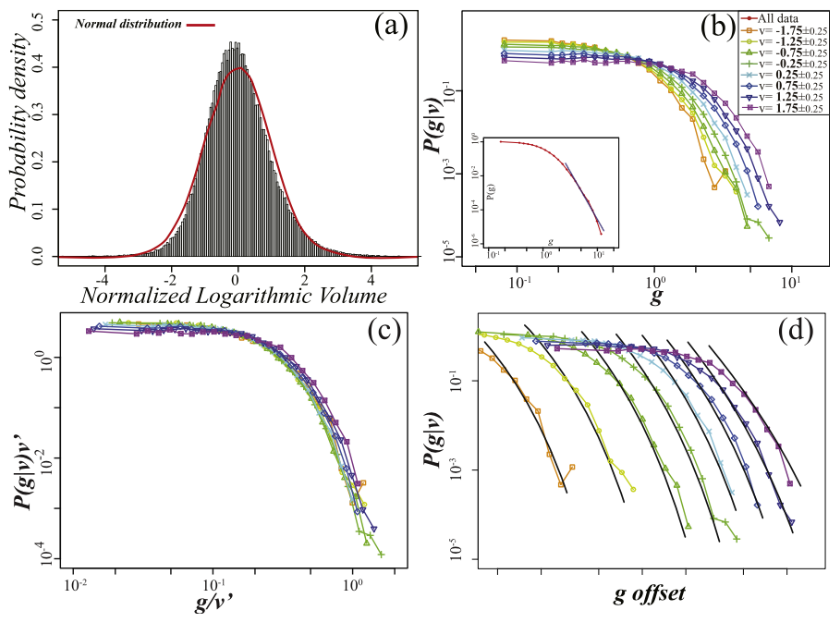

Fig. 1. Range of volumes in the dataset, as well as probability density function (pdf). $P(g)$ and conditional pdf $P(g|v)$, demonstrating that all conditional volatility distributions behave the same in scaled units. (a) Pdf of normalized logarithmic volume v Eq. (4) for the 30 Dow Jones Industrial Average stocks for the period from April 1990 to June 2007, well fit by a normal curve (solid line). (b) Conditional pdf showing the distribution of volatility given a specified volume range. Inset: unconditioned pdf of volatility, roughly fit by a power-law distribution. (c) When the curves in (b) are plotted in scaled parameter of v', where $v' \equiv v + 4.5$, data for all volumes collapse onto the same curve. (d) Same as (b), but shifted horizontally for better visibility with tails given power-law fits with exponential cut offs described by Eqs. (5) and (6).

for a given stock i, where Y_i represents a least-squares linear fit of $\ln Q_i$ [27], which removes the global trend over the entire 17-year period. For each different stock, we define the normalized volatility $g_i(t)$ and normalized logarithmic volume $v_i(t)$ from the raw returns and raw logarithmic volume by

$$g_i(t) \equiv \frac{R_i(t) - \langle R_i(t) \rangle}{\sigma_R} \tag{3}$$

and

$$v_i(t) \equiv \frac{\tilde{Q}_i(t) - \langle \tilde{Q}_i(t) \rangle}{\sigma_{\tilde{Q}}}, \tag{4}$$

where $\langle \cdots \rangle$ denotes a time average over the period studied. Here $\sigma_R = \sqrt{\langle R^2 \rangle - \langle R \rangle^2}$ and $\sigma_{\tilde{Q}} = \sqrt{\langle \tilde{Q}^2 \rangle - \langle \tilde{Q} \rangle^2}$ are the standard deviations of $R(t)$ and $\tilde{Q}(t)$, respectively. Note that the volatility is expressed in terms of absolute value while the logarithmic volume can be both positive and negative. In this paper, the volume indicates the normalized logarithmic volume $v_i(t)$, and volatility indicates the normalized volatility $g_i(t)$.

3. Results

3.1. Dynamics between volume and volatility

We begin by examining the probability density function (pdf) of the normalized logarithmic trading volume, which we find in Fig. 1(a) to be in excellent agreement with a unit Gaussian. The normal curve is often a null model for various econometric quantities. For example, Wang et al. [28] have shown that a normal curve is also a good fit for trading values. However, the pdf of volatility is widely known to be more leptokurtic (i.e., fat-tailed) than a normal fit, which we show in the inset picture in Fig. 1(b) as a log–log plot. The solid red line is the pdf of volatility, the tail of which we observe roughly matches a power-law distribution, as was pointed out in Ref. [29]. The scaling exponents of distributions are systematically related which was also found in Ref. [30]. We also find the distribution of returns to be leptokurtic as well, being better fit by a Laplace distribution than a Gaussian, in agreement with work by Podobonik et al. on NYSE stocks [31].

The tendency of trading volume and price change to move together has important implications in the prediction of financial risk. Recent studies have revealed the long-term cross-correlation of volume changes with price changes [25], power-law cross-correlations between trading activity and volume traded [32], and also the positive correlation of price changes with volume [4,33]. As the absolute value of return, volatility should be a better indicator for market fluctuation

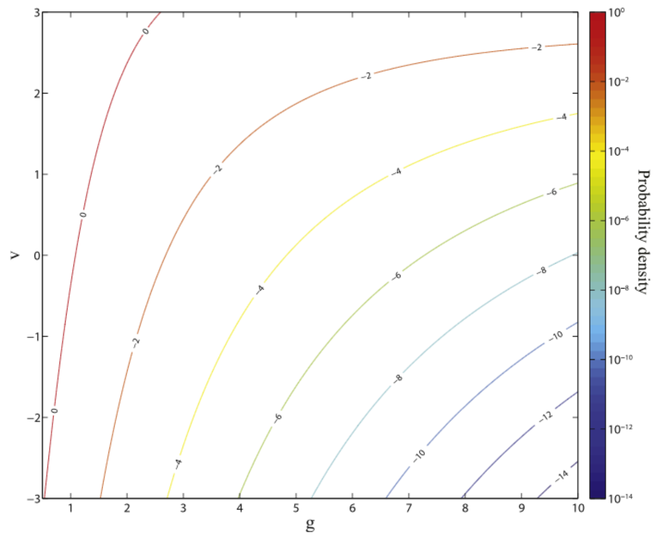

Fig. 2. Fit of the joint distribution of volatility and volume. Using fitted values for α, β, a, b, we show the contour plot for the probability density function of Eq. (6) to fit the data seen in Fig. 1(b). We show that either decreasing g (volatility) or increasing v (volume) results in monotonic increases in the probability density. Higher volatilities are far more localized in their range of volumes than low volatilities, leading to the possibility that higher volatilities may be predicted from volume.

and so we investigate the pdf of volatility given a specified volume. As shown in Fig. 1(b), the conditional volatility distributions for various volumes seem very similar, which leads us to search for scaling features that unify these distributions. We draw inspiration from the work of Yamasaki et al., who analyzed the distribution of return intervals τ between volatilities larger than a specified threshold q [26]. They found that the distributions for different q across seven stocks and currencies all collapsed to a single curve when plotted in units scaled by the mean return interval, dependent on q. We investigate here whether a similar scaling parameter exists that could unify these distributions. This scaling parameter should incorporate volume dependence the same way τ incorporates q dependence in Yamasaki's work.

Redrawing the conditional volatility distributions using the scale parameter v', where $v' \equiv v + 4.5$, results in all conditional distributions collapsing onto the same curve, regardless of the value of volume, as shown in Fig. 1(c), meaning that all conditional volatility distributions are unified, differing only by a factor of the volume chosen, very similar to Yamasaki's findings on volatility return intervals. We have chosen the offset in a volume of 4.5 to avoid singularities and unphysical values, since normalized volume as defined in Eq. (4) can be a non-positive quantity.

We next investigate what unified pdf these distributions follow. In Fig. 1(d) the volume-conditional pdfs are offset for better visibility. We notice that the tails of these distributions are too curved to fit power-laws. After investigating such distributions as log-normal and stretched exponential, we find the best fit using power-law distributions with exponential cutoffs. Thus the distribution of volatility given a certain value of volume should be

$$P(g|v) \sim g^{-\xi} e^{-\varsigma g}. \tag{5}$$

However, as Fig. 1(c) shows, the above pdf can be scaled in $v'(v' \equiv v + 4.5)$, which leads us to add volume as a variable of the conditional volatility distribution function. Thus we assume $\xi = \alpha v + a$ and $\varsigma = \beta v + b$, making Eq. (5)

$$P(g, v|v) \sim g^{-(\alpha v + a)} e^{-(\beta v + b) g}. \tag{6}$$

Using a maximum likelihood estimation for the data shown in Fig. 1(b), we find $\alpha = 0.4$, $\beta = -1.23$, $a = 2.5$, and $b = 3$. We draw a contour plot using these parameters with Eq. (6) in Fig. 2, showing that g (volatility) and v (volume) increase concurrently given a certain probability density value. Specifically, we note that while low volatilities can occur over the entire range of volumes fairly regularly, higher volatilities have a strong tendency to occur only with larger volumes, meaning that high volatilities may be predictable from volume, although low volatilities cannot.

3.2. Local maximum volatility (LMV) and correlation with volume

As a consequence, we restrict our analysis to the days comprising the largest portion of volatilities—which is appropriate, given that days of high volatility are the ones of greatest interest to traders and market researchers. To do this we introduce the quantity "local maximum volatility" (LMV), which, because it is closely related both to a given day's volume and the volume of previous days, allows the possibility of making predictive statements.

We define the LMV parameter, denoted by g_{LM}, by partitioning the values of observed trading volumes into bins $u_1, u_2, u_3, \ldots, u_n$. Then

$$g^j_{LM} \equiv \max(\{g_t\}) \forall t \mid v_t \in u_j. \tag{7}$$

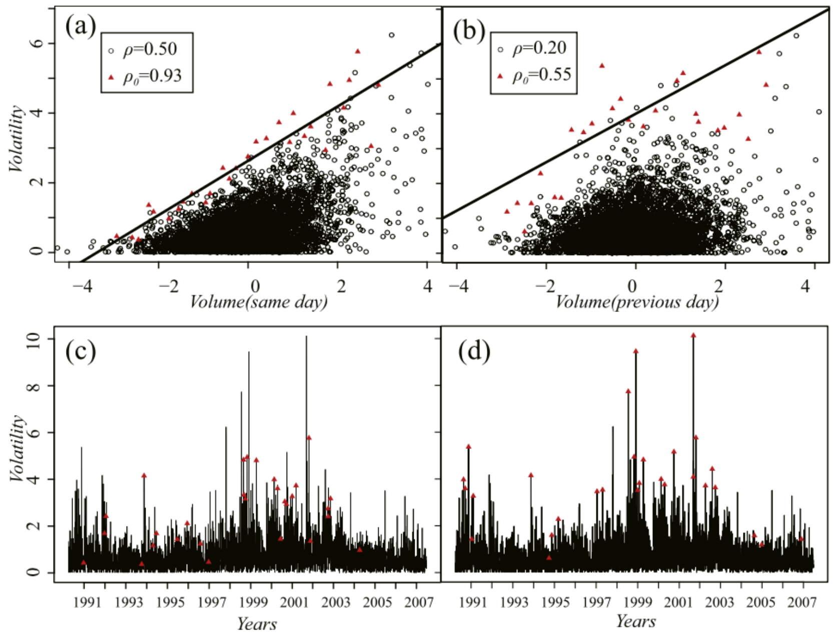

Fig. 3. While volatility is not highly correlated to volume, LMV is highly correlated to both today's volume and to yesterday's volume (linear fits for LMV shown). LMV days occur throughout the period which we study. Shown is a scatter plot of volatility vs. volume for the example of The Boeing Company (BA): (a) Volatility $g(t)$ vs. normalized logarithmic volume $v(t)$, (b) volatility $g(t)$ vs. normalized logarithmic volume the day before, $v(t-1)$. The red solid triangles depict the largest values in each bin of g (from −3 to 3 we delineate 30 bins evenly). ρ_0 is the correlation coefficient between logarithmic volume and LMV (Eq. (7)), while ρ represents correlation coefficient between logarithmic volume and volatility. The volatility time series and LMV (red triangle) are shown for LMV based on (c) concurrent volume and (d) previous day's volume.

LMV is the maximum volatility observed in a given range of trading volumes, i.e., the volatility of the most volatile day a given trading volume has co-occurred with. Although correlation between volatility and logarithmic volume is weak, we find that, in general, LMV and logarithmic volume are highly correlated. We demonstrate this in Fig. 3(a) using the example of the Boeing Company (BA). For BA, we observe that while the correlation coefficient between same-day volume and volatility is only 0.5, the correlation coefficient between volume and LMV is 0.93. We further investigate the correlation between volatility and volume using the scatter plot of volatility against volume in Fig. 3. A characteristic triangular shape can be seen in both the scatter plot of (a) volatility vs. the same-day volume and (b) volatility vs. the previous day's volume. The volume ranges used to define LMV are delimited by defined bins as is shown in Fig. 1(a) (30 bins evenly divided from −3 to 3). As defined in Eq. (7), we use the highest volatility in each given bin. In both cases, the maximum volatility matching a given volume is shown in red triangles and a linear regression fit is shown in solid black, visually confirming the calculated correlation. Because it is possible that the volatilities used in LMV could originate in a narrow, unusually volatile time window (e.g., one week), and thus be giving spurious results, we investigate the timing of the high volatility days used. Figs. 3(c) and 3(d) show that these high volatility days do indeed occur throughout the span of the time period under consideration, which ensures the universal representativeness of LMV.

We now generalize the analysis shown in Fig. 3 for same-day and one-day offsets to variable time offsets up to 16 days. Our results are shown in Fig. 4. In the figure, we show the mean correlation coefficient against time-lag for the 30 DJIA stocks. The figure shows that while the correlation between volume and volatility quickly drops to zero for almost any nonzero time-lag, the correlation between volume and LMV retains significant value (≈0.4) at a one-day lag and remains noticeable (≈0.2) even with a 4-day time-lag, indicating significant potential for predicting days of potential largest volatility, and therefore largest risk, which is extremely important in protecting investments during a financial crisis [34].

3.3. Predicting power of volume and volatility

The fact that the possible volatility is closely tied to the same-day trading volume is intuitive, as the extent to which the price can change is a function of quantity of trading that has transpired in a given day. The connection between volume and future volatility is more interesting. Because volatility is already widely known to correlate with its own values in the immediate future, this result may seem trivial. We later present evidence (see Fig. 5) that our findings go beyond this obvious result, that the inclusion of volume really does add non-redundant information into the prediction scheme. Additionally, as has been shown by Gillemot et al. the tendency for volatility to cluster is not a simple volume effect resulting from reductions in average trading time [20].

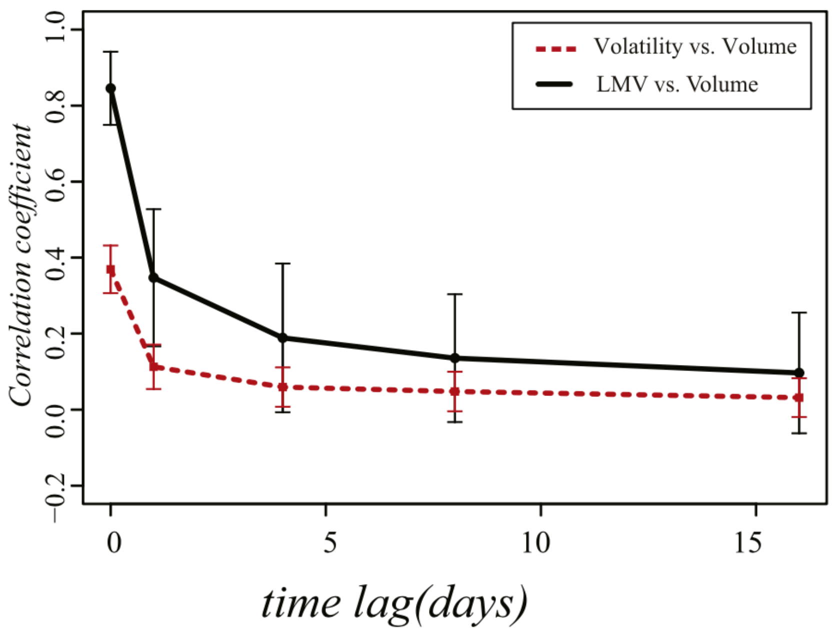

Fig. 4. There is a weak correlation between volatility and volume (red squares), though this effect quickly drops off with time lag. LMV has a stronger correlation with volume (black circles) throughout the range of time lags. Shown is the mean correlation coefficient vs. time lag for the 30 DJIA stocks. The error bars depict ± standard deviation. Note that the mean correlation coefficients for time-lag = 0 days, 1 day are very similar to those found in Fig. 4, which depicts results for only The Boeing Company (BA).

Fig. 5(a) and (b) show the conditional distribution of $P(v(t)|g_0(t+1))$ and $P(g(t)|g_0(t+1))$. Here $g_0(t+1)$ represents the subset that contains the highest 1% or lowest 1% of volatilities. Fig. 5(a)shows the quintile distribution of the volume today, given a specified volatility tomorrow while Fig. 5(b) shows the volatility today, given a specified volatility tomorrow. In the absence of memory effects, Figs. 5(a) and 5(b) would be completely flat distributions, in both highest and lowest volatility cases. Instead we clearly see memory effect in highest cases. 20% of volumes account for roughly 40% of the days preceding the highest (top 1%) volatilities. This effect is monotone across the quintiles and the most extreme example of underrepresentation being that the lowest 20% of volumes account for approximately 10% of the highest volatilities. For the lowest 1% of volatilities we find no such effect. The distribution of days preceding low volatility is statistically the same as a flat distribution across volume. That low volatility days do not have a statistically different distribution in volumes agrees with earlier observations seen in Fig. 2 that low volatilities have a broad range across volumes, and hence are not predictable from volumes. We observe similar results when considering the distribution a day's volatility, knowing that the next day will have a particular high or particularly low volatility. Again, days prior to high volatility are overrepresented in the highest quintile of volatility, but days prior to low volatility have a distribution that is essentially flat across quintiles in volatility.

In summary, Fig. 5(a) and (b) show not only that high volatility tends to follow high volatility, but also high volatility tends to follow high volume. No such significant effects can be observed for low volatility.

Extending this analysis, we include both volume and volatility in order to better predict next-day volatility. Fig. 5(c) and (d) give $P(g(t+1) = A|v_n(t), g_n(t))$, the distribution of the days preceding the highest or lowest 1% of volatilities according to preceding volatilities and volumes broken up into quintiles ($n = 1 \dots 5$). The probabilities are given in units of $P(g(t+1) = A)$, the unconditioned probability of a defined volatility (top or bottom 1%) day, which is equal to 1%. Fig. 5(c) and (d) therefore divide the 1304 data points (1% of all data points) into $5 \times 5 = 25$ equal-sized sets of approximately 52 points each. Fig. 5(c) shows the relative probability of that one particular set of data points to precede a high volatility day with probability proportional to circle radius. Essentially, Fig. 5(c) and (d) are heat maps with bubble size being used in lieu of color intensity.

Were there no next-day memory effect, all bubbles would be of equal size. However, in Fig. 5(d) we find that the joint conditional probability for the top quintiles $P[g(t+1) = p|v_5(t), g_5(t)]$ is approximately three times the size of the unconditioned probability $P(g(t+1) = A)$, indicating that days with the top quintiles of both volatility and volume are overrepresented in the days preceding high volatility by a factor of three. In contrast, the probability for the bottom quintiles $P[g(t+1) = p|v_1(t), g_1(t)]$ is only half that of the unconditioned probability, meaning that days with the bottom quintiles in both volatility and volume are underrepresented in the days preceding high volatility by a factor of two. We compare this to the results yielded from the investigation in Fig. 5(a) and (b), where the greatest overrepresentation by quintile is approximately only a factor of two. This indicates that the volume and volatility combined are a more powerful predictor of upcoming high volatility than either volume or volatility alone. The variation of results by both row and column also indicates that there is information potentially important for volatility prediction embedded into

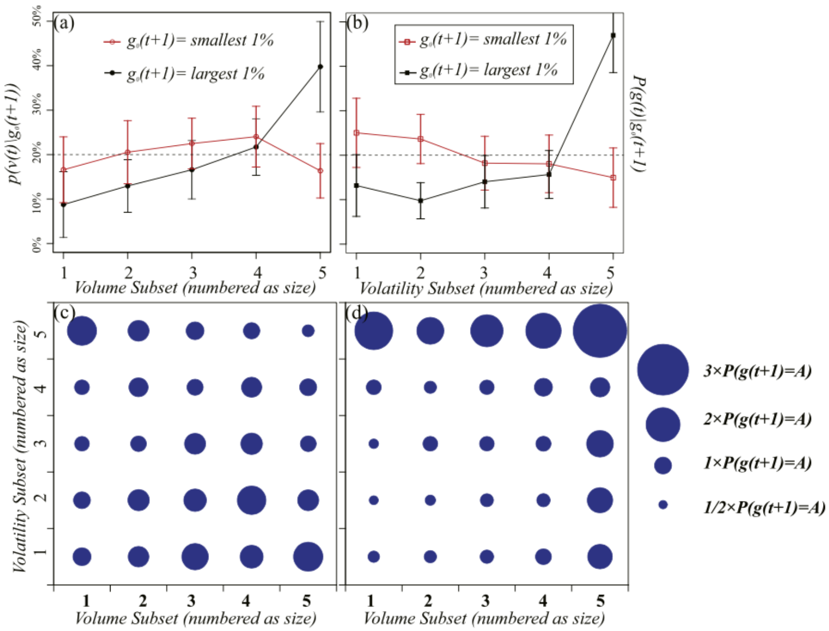

Fig. 5. Predicting power of volume and volatility. Days preceding high volatility are characterized by higher volume and higher volatility than usual. Figures (a) and (b) show the conditional distribution of $P(v(t)|g_0(t+1))$ and $P(g(t)|g_0(t+1))$, averaged over all 30 stocks. Here $g_0(t+1)$ represents the subset containing largest (closed symbols) or smallest (open symbols) 1% of volatility observed. The volatility $g(t)$ and normalized logarithmic volume $v(t)$ are divided into quantiles. The error bars show $\pm$ standard deviation. Figures (c) and (d) show $P(g(t+1) \in g_0|v_n(t), g_n(t))$, the distribution by quintile (n = 1…5) for v and g values of today given that tomorrow will have the (c) smallest or (d) largest 1% of volatility, where n is increasing with size and g_0 denotes the set of the (c) smallest 1% of day volatilities or (d) the largest 1% of day volatilities. The radius of the circle is proportional to $P(g(t+1) \in g_0|v_n(t), g_n(t)) = 1\%$, the unconditioned probability that the next day's volatility will be in the smallest or largest 1% of volatility.

both quantities. We confirm this by applying a simple multiregression model predicting next-day volatility from either volatility alone or volatility and volume together. We find an average 6% increase in the R^2 value when volume is included.

By contrast, Fig. 5(c) shows the distribution by quintile of volume and volatility for days preceding the very lowest volatility days. The variation in bubble size is considerably reduced compared to that of Fig. 5(d), showing that days preceding low volatility are far more evenly distributed in volatility and volume. Additionally, there are no clear pronounced trends across row or column that would indicate a clear effect of either volume or volatility on the next day's volatility value.

Discussion and conclusion

In this paper, we have examined the relationship between trading volume, volatility, and LMV and have uncovered the scaling laws and memory effect between them. We have combined volume and volatility and found that while the same-day correlation between the logarithmic volume and volatility is fairly weak, the same day and time-lagged correlation between logarithmic volume and a quantity we introduce as "local maximum volatility" (LMV) are both very strong. This finding may help explain the inconsistency between investors intuition about market stability during high volume days and the empirical fact that the relationship is not strong. Although it is essential that a trader understands the effects of trading volume, the weak correlation coefficient (≈0.2) is unable to explain the importance of trading volume. While humans often interpret correlations to be stronger than they are (i.e., illusory correlation [35]), in the case of volume-volatility correlations there are obvious mechanisms indicating their reality. Thus, we further investigate and find out that through the strong correlation between volume and LMV, a trader's interpretation may be justified. We believe LMV to be a more accurate representation of an investor's memory than the actual volatilities themselves. The cognitive bias in which humans disproportionately focus their attention on negative experiences and threats over positive experiences and aid is well-documented in cognitive psychology and termed the "negativity bias" [36], summarized by Baumeister et al. [37] as "bad is stronger than good". The manifestation of negativity bias in trading in the form of volatility asymmetry – wherein negative price changes cause a market to become more volatile than positive price changes – has been observed in many different countries [38–42]. Thus our findings using LMV match the behavior of investors because LMV is a more important quantity when it comes to human perception. An investor may thus be justified in having an negative attentional bias because (s)he does not know the next-day volatility level in advance and must treat the "risk of risk" as the relevant quantity. Our findings also indicate that high volatility tends to follow high trading volume, although low volatility is largely unaffected by volume. Because we observe that high volatility strongly affects trading volume, we posit that volume can be used to predict future highest volatility. Based on the new dynamics we provided and the empirical analysis, we find the good use of volume in predicting near-future high volatility. Our analysis shows that using both volume and volatility in the prediction is better than using either of them alone. Further, we have introduced the functional form that gives the tail of the volume-conditional volatility distribution and shown that the distribution is unified across wide ranges of volumes when viewed in scaled units making the abscissa the volatility divided by the volume. Thus, we are able to explain not only why high volatility tends to occur with large volume, but also to what extent the latter effects the former.

Acknowledgments

This research is supported by the Program for National Natural Science Foundation (71671182). And the Boston University Center for Polymer Studies is supported by NSF Grant PHY-1505000 and by DTRA Grant HDTRA1-14-1-0017. We also thank F.Ling for the constructive suggestions. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

[1] W.B. Arthur, Complexity economics: A different framework for economic thought, Cfa Dig. 13 (2) (2015) 15–37.
[2] J. Kwapień, S. Drożdż, Physical approach to complex systems, Phys. Rep. 515 (3–4) (2012) 115–226.
[3] T. Ané, H. Geman, Order flow, transaction clock, and normality of asset returns, J. Financ. 55 (5) (2010) 2259–2284.
[4] J.M. Karpoff, The relation between price changes and trading volume: A survey, J. Financ. Quant. Anal. 22 (1) (1987) 109–126.
[5] R.L. Crouch, A nonlinear test of the random-walk hypothesis, Amer. Econ. Rev. 60 (1) (1970) 199–202.
[6] C.G. Lamoureux, W.D. Lastrapes, Heteroskedasticity in stock return data: volume vs garch effects, J. Financ. 45 (1990) 487–498.
[7] V. Plerou, P. Gopikrishnan, X. Gabaix, L.A.N. Amaral, H.E. Stanley, Price fluctuations, market activity and trading volume, Quant. Finance 1 (2001) 262–269.
[8] C.C. Ying, Stock market prices and volumes of sales, Econometrica 34 (3) (1966) 676–685.
[9] R.J. Rogalski, The dependence of prices and volume, Rev. Econ. Stat. 60 (2) (1978) 268–274.
[10] G.E. Tauchen, M. Pitts, The price variability-volume relationship on speculative markets, Econometrica 51 (2) (1983) 485–505.
[11] J.A. Stephan, R.E. Whaley, Intraday price change and trading volume relations in the stock and stock option markets, J. Financ. 45 (1) (1990) 191–220.
[12] X. Gabaix, P. Gopikrishnan, V. Plerou, H.E. Stanley, A theory of power-law distributions in financial market fluctuations, Nature 423 (6937) (2003) 267–270.
[13] I.N. Lobato, C. Velasco, Long memory in stock-market trading volume, Bus. Econ. Stat. 18 (2000) 410–427.
[14] W.X. Zhou, Universal price impact functions of individual trades in an order-driven market, Quant. Finance 12 (8) (2012) 1253–1263.
[15] T.J. Brailsford, The empirical relationship between trading volume return and volatility, Account. Financ. 36 (1) (2014) 89–111.
[16] C. Brooks, Predicting stock index volatility: can market volume help?, J. Forecast. 17 (1) (1998) 59–80.
[17] P.K. Clark, A subordinated stochastic process model with finite variance for speculative prices, Econometrica 41 (1) (1973) 135–155.
[18] A. Koulakiotis, A. Dasilas, P. Molyneux, Does trading volume influence garch effects? - some evidence from the greek market with special reference to banking sector, Invest. Manag. Financ. Innov. 4 (3) (2007) 33–38.
[19] J.L. Sharma, M. Mougoue, R. Kamath, Heteroscedasticity in stock market indicator return data: volume versus garch effects, Appl. Financ. Econ. 6 (4) (1996) 337–342.
[20] L. Gillemot, J.D. Farmer, F. Lillo, There's more to volatility than volume, Quant. Finance 6 (5) (2006) 371–384.
[21] T.E. Copeland, A model of asset trading under the assumption of sequential arrival, J. Finance 31 (1976) 1149–1168.
[22] M. Harris, A. Raviv, Differences of opinion make a horse race, Rev. Financ. Stud. 6 (3) (1993) 473–506.
[23] J.B. De Long, A. Shleifer, L.H. Summers, R.J. Waldmann, Noise trader risk in financial markets, J. Political Econ. 98 (4) (1990) 703–738.
[24] R.N. Mantegna, H.E. Stanley, An Introduction to Econophysics: Correlations and Complexity in Finance, Cambridge University Press, 1999.
[25] B. Podobnik, D. Horvatic, A.M. Petersen, H.E. Stanley, Cross-correlations between volume change and price change, Proc. Natl. Acad. Sci. USA 106 (52) (2009) 22079.
[26] K. Yamasaki, L. Muchnik, S. Havlin, A. Bunde, H.E. Stanley, Scaling and memory in volatility return intervals in financial markets, Proc. Natl. Acad. Sci. USA 102 (26) (2005) 9424–9428.
[27] J.M. Chambers, T.J. Hastie, Statistical Models in S. Wadsworth and Brooks/Cole Computer ScienceSeries, Brooks/Cole. CRC Press, 1992.
[28] F. Wang, K. Yamasaki, S. Havlin, H.E. Stanley, Statistical regularities of equity market activity, Quant. Finance (2009).
[29] F. Wang, S.J. Shieh, S. Havlin, H.E. Stanley, Statistical analysis of the overnight and daytime return, Phys. Rev. E 79 (2) (2009) 056109.
[30] R. Rak, S. Drozdz, J. Kwapien, P. Oswiecimka, Stock returns versus trading volume: is the correspondence more general?, Acta Phys. Polon. 44 (10) (2013) 2035.
[31] B. Podobnik, D. Horvatic, A.M. Petersen, H.E. Stanley, Quantitative relations between riskreturn and firm size, Europhys. Lett. 85 (5) (2009) 50003–50008.
[32] R. Rak, S. Drozdz, J. Kwapien, P. Oswiecimka, Detrended cross-correlations between returns, volatility, trading activity, and volume traded for the stock market companies, Quant. Finance 112 (4) (2015) 48001.
[33] A.R. Gallant, P.E. Rossi, G. Tauchen, Stock prices and volume, Rev. Financ. Stud. 5 (2) (1992) 199–242.
[34] Z. Zheng, B. Podobnik, L. Feng, B. Li, Changes in cross-correlations as an indicator for systemic risk, Sci. Rep. 2 (11) (2012) 888.
[35] L.J. Chapman, Illusory correlation in observational report, J. Verbal Learn. Verbal Behav. 6 (1) (1967) 151–155.
[36] P.C. Wason, The processing of positive and negative information, Q. J. Exp. Psychol. 11 (2) (1959) 92–107.
[37] R.F. Baumeister, E. Bratslavsky, C. Finkenauer, K.D. Vohs, Bad is stronger than good, Rev. Gen. Psychol. 5 (4) (2001) 477–509.
[38] R.F. Engle, N.G. Victor K., Measuring and testing the impact of news on volatility, J. Financ. 48 (5) (1993) 1749–1778.
[39] J.M. Zakoian, Threshold heteroskedastic models, J. Econom. Dynam. Control 18 (5) (1994) 931–955.
[40] E. Sentana, Quadratic arch models, Rev. Econom. Stud. 62 (4) (1995) 639–661.
[41] R. Jayasuriya, Shamila A. Rossiter, Asymmetric volatility in asian equity markets, J. Int. Financ. Econ. (2008).
[42] J. Tenenbaum, D. Horvatić, S.C. Bajić, B. Pehlivanović, B. Podobnik, H.E. Stanley, Comparison between response dynamics in transition economies and developed economies, Phys. Rev. E 82 (2) (2010) 527–537.