Are the Trading Volume and the Number of Trades Distributions Universal?

Vikram S. Vijayaraghavan and Sitabhra Sinha

Abstract. Analysis of dynamical phenomena in financial markets have revealed the existence of several features that appear to be invariant with respect to details of the specific markets being considered. While some of these "stylized facts", such as the inverse cubic law distribution of price returns indeed seem to be universal, there is less consensus about other phenomena. In particular, there has been a long-running debate in the literature about whether the distributions of trading volume $V_{\Delta t}$ and the number of trades $N_{\Delta t}$ occurring in a given time interval Δt, are universal, and whether the volume distribution is Levy-stable. In this article, we analyse data from the National Stock Exchange of India, both daily and high frequency tick-by-tick, to answer the above questions. We observe that it is difficult to fit the $V_{\Delta t}$ and $N_{\Delta t}$ distributions for all stocks using the same theoretical curve, e.g., one having a power-law form. Instead, we use the concept of the stability of a distribution under temporal aggregation of data to show that both these distributions converge towards a Gaussian when considered at a time-scale of $\Delta t = 10$ days. This appears to rule out the possibility that either of these distributions could be Levy-stable and at least for the Indian market, the claim for universality of the volume distribution does not hold.

1 Introduction

A financial market comprising a large number of interacting components, viz., agents involved in trading assets whose prices fluctuate with time as a result of

Vikram S. Vijayaraghavan
Department of Physics, Indian Institute of Technology Madras, Chennai, India,
e-mail: vikramsv@smail.iitm.ac.in

Sitabhra Sinha
The Institute of Mathematical Sciences, C.I.T. Campus, Chennai, India,
e-mail: sitabhra@imsc.res.in

F. Abergel, B.K. Chakrabarti, A. Chakraborti, M. Mitra (Eds.)
Econophysics of Order-driven Markets.

the constant stream of external news and other information affecting the actions of the traders, is a paradigmatic example of a complex system. However, despite such inherent complexity, markets appear to exhibit several statistically regular features which make them amenable to a rigorous analysis by techniques based on the statistical mechanics of physical systems, a discipline that is often referred to as econophysics [1–3]. Indeed, many of the empirical relations obtained by such analysis appear to be statistically invariant or *universal* with respect to different markets, periods of observations and the type of assets being considered. These *stylized facts* of the market (as they are often referred to in the economic literature) include the celebrated *inverse cubic law* for the distribution of price (or index) fluctuations as measured by their logarithmic returns [4]. First observed in the developed markets of advanced economies [5], it has later been reported also in emerging markets at all stages of their development [6, 7]. Another robust feature characterizing financial markets is *volatility clustering*, i.e., the occurrence of long-range temporal correlations in the magnitude of price fluctuations [8]. Thus, periods marked by large fluctuations (i.e., high volatility) often tend to be persistent, as is seen across many different markets.

There have also been claims that other quantifiers of market activity, such as the distributions for order size q (i.e., the number of shares traded in a particular transaction), trading volume V_t (i.e., the total number of shares traded in a given period) and the number of trades N_t over a specific time interval, possess universal forms [9, 10]. However, the evidence for the invariance of these distributions seems less unequivocal. Note that the three distributions are not completely independent of each other, as the volume $V_{t,\Delta t}$ over a particular time interval $[t, t + \Delta t]$ is related to the number of trades $N_{t,\Delta t}$ and the sizes of each trade q_i that takes place in the interval as

$$V_{t,\Delta t} = \sum_{i=1}^{N_{t,\Delta t}} q_i \,. \tag{1}$$

For US markets, the N_t cumulative distribution appears to follow an approximately "inverse cubic" form, i.e., $P(N_t > x) \sim x^{-\beta}$ with $\beta \simeq 3.4$ [9]. Both the trade size and volume cumulative distributions have been claimed to be Levy-stable with exponents $\zeta_q \simeq 1.53$ (the so-called "inverse half-cubic law") and $\zeta_V \simeq 1.7$, respectively [11]. However, not only has the universality of these exponents been challenged, even the power-law form of the distributions appear to be dependent on the type of stock and the market being considered. For example, an early study of the volume distribution of several stocks in the London Stock Exchange (LSE) did not show any evidence of power-law scaling [12], but it was pointed out later that this depended on whether one was considering the downstairs or upstairs market in LSE. As splitting a large order into several smaller parts is regularly practised in the downstairs market (but rare in the upstairs market) it is probably not surprising that long tails can only be seen when the trades in the upstairs market are included in the volume data [13]. A re-analysis of the US stock data complicated the issue further by showing that the cumulative distribution of trading volume over 15-minute intervals has a tail exponent of around 2.2, i.e., outside the Levy-stable regime [14].

More recent work on emerging markets such as the Korean [15] and Chinese [16] exchanges have also revealed significant deviations from the Levy-stable power-law tails of volume distribution reported for the developed markets of US, London and Paris [17] (see also [18, 19]). There have also been related studies that try to fit the entire distribution of trading volume rather than focusing only on the tail, e.g., by using the q-Gamma distribution [20, 21].

The reason that the universality (or otherwise) of the distributions for q, V_t and N_t is of interest to the econophysics community is because this may provide insights towards understanding the statistical relationship between price returns and market activity. It is frequently said that it takes volume to move prices, implying that the dynamics of price fluctuations (measured by the log-returns) can be understood in terms of the distributions of trade size, number of trades and trading volume. Indeed, the price impact function, that measures how the volume of shares traded affects the price movement, tries to quantify such a relation. By assuming a square-root functional form for the impact (based on empirical analysis of US markets), Gabaix *et al.* [22] have developed a theory of market movements where the long-tailed return distribution arises as a consequence of the long-tailed volume distribution. The square-root relation between price and volume leads to the result that the price return distribution exponent ($\simeq 3$) is twice the volume distribution exponent (~ 1.5), thereby connecting the inverse cubic and half-cubic laws. However, we have recently shown that the occurrence of power-law tailed distributions for price and volume with their characteristic exponents do not critically depend on the assumption of a square-root price impact function [23], nor does the existence of the inverse cubic law for returns necessarily imply an exponent of around $3/2$, or even a power-law nature, for the distribution of trading volume [24][1].

It is in this context that we report our analysis of the data for market activity in the National Stock Exchange (NSE) of India in this article. As this market has already been shown to exhibit the inverse cubic law of returns [6, 7], the absence of a Levy-stable nature for the volume distribution would appear to argue against the theoretical work relating the return and volume distribution exponents on the basis of a square-root form for the price impact function. While our earlier work on the trade and volume distributions in this market had also shown the absence of a clear power-law functional form for either [25], here we use an alternative procedure to show that the two distributions do not have the same behavior as that reported for the developed markets. In particular, we use the concept of stability of a distribution under temporal aggregation of data to show that both the quantities converge to a Gaussian distribution at a time-scale of $\Delta t = 10$ days. This evidence against the Levy-stable nature of the volume distribution (even though the return distribution follows the inverse cubic law) suggests that the theoretical framework of [23, 24] can better explain the market dynamics than arguments based on square-root price impact function whose predictions about the relations between return and volume is not matched by the empirical data. It is of course possible that the deviation from the

[1] In fact, our numerical results show that even a log-normal distribution of trading volume can result in a power-law tailed return distribution.

Levy-stable nature for volume and trade size distributions is a result of the emerging nature of the market which is yet to evolve into a completely developed form. Just as the network representing the relations between price movements of different stocks (measured by the cross-correlation between returns) has been suggested to change over time from being homogeneous to one having a clustered organization as the market matures [26], the volume distribution could, in principle, become more and more heavy tailed as market activity increases, eventually becoming Levy-stable at a certain stage of market development.

2 The Indian Financial Market

There are 23 different stock markets in India. The largest of these is the National Stock Exchange (NSE) which accounted for more than half of the entire combined turnover for all Indian financial markets in 2003–04 [27], although its market capitalization was comparable to that of the second largest market, the Bombay Stock Exchange. The NSE is considerably younger than most other Indian markets, having commenced operations in the capital (equities) market from Nov 1994. However, by as early as 2004 it had become the world's third largest stock exchange (after NASDAQ and NYSE) in terms of transactions [27]. It is thus an excellent source of data for studying the trading frequency and volume statistics in an emerging market.

Description of the data set. The low-frequency data that we analyze consists of the daily volume and number of trades for the entire NSE market, as well as, for individual stocks, available from the exchange web-site [28]. The period we have considered begins at March 1994 (for the entire market) or the date from which data for a particular stock has been recorded in the NSE database (for individual stocks) and ends at May 2010. For the market data, this corresponds to 3910 working days. We also consider high-frequency tick-by-tick data containing information of all transactions carried out in the NSE between Jan 1, 2003 and Mar 31, 2004. This information includes the date and time of trade, the price of the stock during transaction and the number of shares traded. This database is available in the form of CDs published by NSE.

3 Results

To investigate the nature of the volume and number of trades distribution in detail, we first consider the high-frequency tick-by-tick data. To calculate these quantities we use a time-interval $\Delta t = 5$ minutes and normalize the resulting variables by subtracting the mean and dividing by their standard deviation. The resulting distributions of normalized trading volume $v = \frac{V_{t,\Delta t} - \langle V \rangle}{\sqrt{\langle V^2 \rangle - \langle V \rangle^2}}$ and number of trades $n = \frac{N_{t,\Delta t} - \langle N \rangle}{\sqrt{\langle N^2 \rangle - \langle N \rangle^2}}$, where $\langle \ldots \rangle$ represents time average, for all stocks that are traded

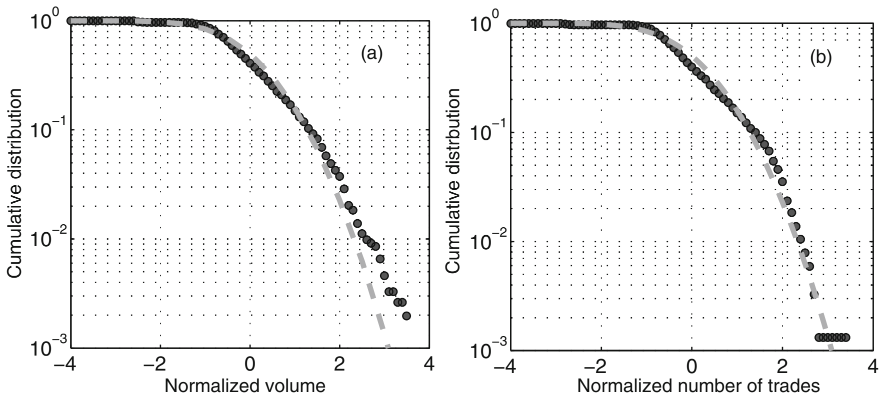

Fig. 1 Cumulative distribution of (a) normalized trading volume and (b) normalized number of trades in $\Delta t = 5$-minute time intervals for all stocks traded in NSE in December 2003. The cumulative standard normal distribution, i.e., $\mathcal{N}(0, 1)$, is shown for comparison (*broken line*)

in NSE are shown in Fig. 1. Direct comparison with the standard normal distribution $\mathcal{N}(0, 1)$ shows that both of these quantities are distributed differently from a Gaussian.

As the exact nature of the distributions for the entire market is difficult to characterize, we now consider the volume and number of trades data for *individual stocks*. Fig. 2 shows the corresponding distributions for a particular stock which appear to possess tails described by a power-law decay. Using the Clauset-Shalizi-Newman (CSN) estimator based on maximum likelihood and Kolmogorov-Smirnov statistic [29], we obtain exponents of -2.87 and -3.11 for the volume and number of trades respectively, both of which lie outside the Levy-stable regime. However, the values of these exponents differ from stock to stock. More importantly, the power-law nature of the decay itself is not entirely representative of the ensemble of stocks. The deviation of the distributions from a power-law is quite apparent visually for several frequently traded stocks (e.g., the volume distribution of SBI).

As the best-fit distributions for the high-frequency volume and number of trades statistics of the NSE do not appear to have a form that is common to all stocks, we cannot readily use this data to decide whether these distributions are Levy-stable or not. Instead, we shall use an indirect approach based on the idea of the stability of a distribution under time-aggregation of the corresponding random variables[2]. A distribution is said to be *stable*, when a linear combination of random variables independently chosen from the distribution has the same distribution, up to a trans-

[2] It should be noted here that the convergence to a stable form, which follows from the Central Limit Theorem, is strictly valid only when the variables being aggregated are statistically independent. However, if correlations do exist between the variables, then provided that these correlations decay sufficiently fast, the theorem still holds and the convergence result can be applied. We have explicitly verified that the auto-correlation function for trading volume shows an exponential decay with time-lag.

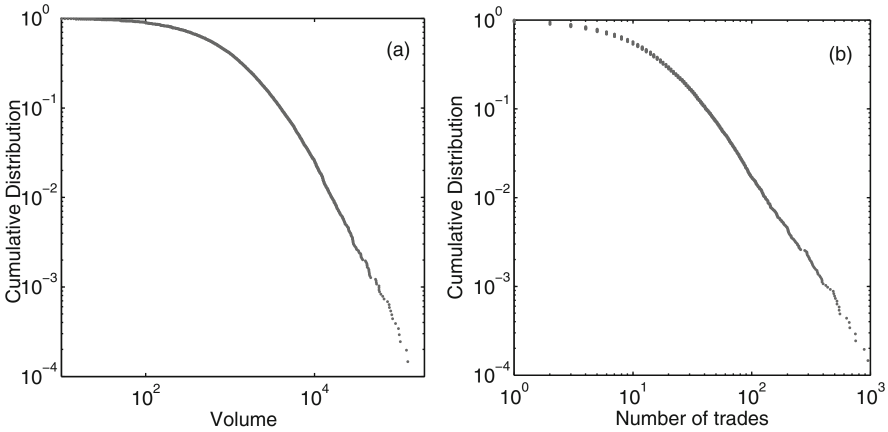

Fig. 2 Cumulative distribution of (a) volume and (b) number of trades in $\Delta t = 5$-minute time intervals for a particular stock (Colgate) traded at NSE between Jan 2003 and March 2004

lation in the mean and a scale factor in the variance. Thus, a sum of independent, identically distributed random variables always converge to a stable distribution. In terms of symbols, if x_1 and x_2 are random variables chosen from a stable distribution $P_{\text{stable}}(x)$, then for any pair of positive constants a, b, the composite variable $a\ x_1 + bx_2$ has the same distribution, but possibly with a different mean and variance. If the mean is identical to the original distribution, then it is said to be *strictly stable* (or stable in the narrow sense) [30]. This is a generalization of the classical *Central Limit Theorem*, according to which, a variable generated by adding a large number of random numbers from arbitrary distributions having finite variance will eventually be seen to follow a Gaussian distribution. Removing the restriction of finite variance results in other possible stable distributions, including the Cauchy and Levy distributions. In particular, a cumulative probability distribution having a power-law tail exponent $\alpha > -2$ has an unbounded second moment. It is, thus, Levy-stable and will not converge to a Gaussian even if we consider an aggregate quantity generated by summing together many random variables generated using such a distribution. Here, we shall use the fact that if the volume or the number of trades, when aggregated over long time periods, converges to a Gaussian distribution, then the original distribution of $V_{t,\Delta t}$ or $N_{t,\Delta t}$ (respectively) could not have been Levy-stable.

Fig. 3 shows the time-series of daily trading volume and number of trades for all stocks traded at NSE. To address the non-stationary nature of the variation in both the quantities, we calculate the mean (μ_t) and standard deviation (σ_t) over a moving window. The data is then de-trended by subtracting the mean and normalized by dividing by the standard deviation, i.e., $x_{t,\text{daily}} = (X_{t,\text{daily}} - \mu_t)/\sigma_t$, where $X_{t,\text{daily}}$ can represent either the daily volume or number of trades. The window used in Fig. 3 has a width of 10 days but small variations in the window size do not critically affect the results. One can also check whether the fluctuations from the mean values

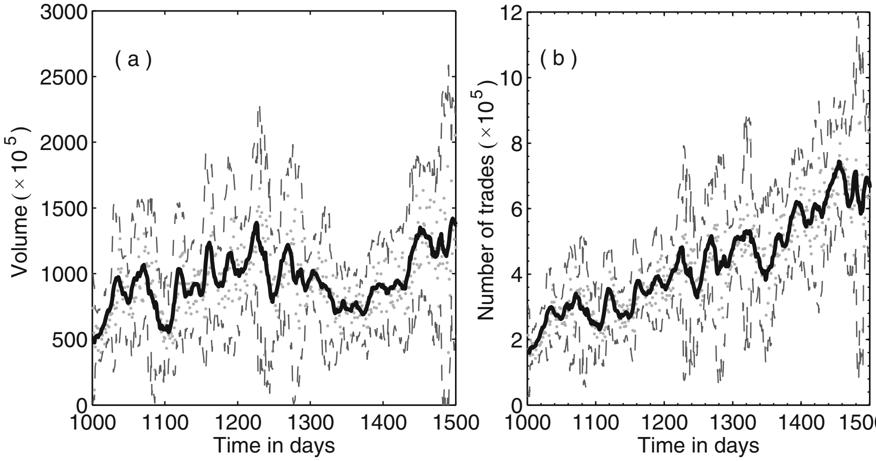

Fig. 3 The time-series of (a) the total volume of all stocks traded $V_{t,\mathrm{daily}}$ and (b) the total number of trades in the market $N_{t,\mathrm{daily}}$ (*dots*), shown for the interval $T = 1000$–1500 days in the daily NSE data. The continuous curves represent the moving average(μ_t) of the corresponding quantities calculated over a moving window (having a width of 10 days and shifted in steps of 1 day). The standard deviation (σ_t) calculated over the window is used to show the range of fluctuations (*dotted lines*) in the quantities $V_{t,\mathrm{daily}}$ and $N_{t,\mathrm{daily}}$ expected from a Gaussian distribution (i.e., $\mu \pm 3\sigma$)

observed in these quantities agree with those expected from a Gaussian distribution by verifying if most data points lie within the bounds representing three standard deviations above and below the mean (which account for about 99.7% of all data points if they are normally distributed). As seen from Fig. 3, this indeed appears to be the case.

To obtain a more reliable comparison between the empirical and normal distributions, we next use a graphical method, specifically the Quantile-Quantile or Q-Q plots [31], for comparing the de-trended, normalized volume and number of trades data with the standard normal distribution. The abscissa and ordinate of any point in such a Q-Q plot correspond to the quantiles (i.e., points taken at regular intervals from the cumulative distribution function) of the theoretical and empirical distributions being compared, respectively. Linearity of the resulting curve implies that the empirical distribution is indeed similar to the theoretical distribution, in this case, the standard normal distribution. While the daily data (Fig. 4a,d) shows deviation from linearity at the ends, the agreement between the two distributions become better when the data is aggregated over several days. Indeed, when we consider the volume and number of trades over a 10-day period, the corresponding distributions appear to match a normal distribution fairly well as indicated by the linearity of the Q-Q plots (Fig. 4c,f). This is also shown by direct graphical comparison of the distributions of these quantities aggregated over 10 days with the normal distribution shown in Fig. 5.

For a more rigorous determination of the nature of the distributions for the temporally aggregated volume and number of trades data, we turn to statistical tests for

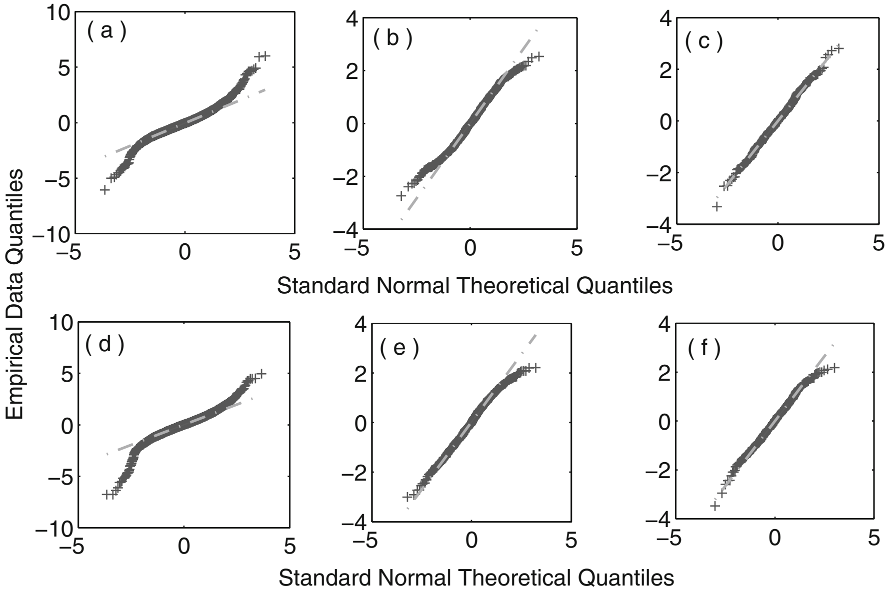

Fig. 4 Q-Q plots comparing the distributions of normalized and de-trended volume (a-c) and number of trades (d-f) for the entire market to a standard normal distribution at different scales of temporal aggregation. The aggregation is over 1 day for (a,d), 5 days for (b,e) and 10 days for (c,f). The *broken line* used for evaluating linearity of the plots is obtained by extrapolating the line joining the the first and third quartiles of each distribution. The linear nature of the plots for both volume and number of trades aggregated over 10 days suggest that the quantities converge to a Gaussian distribution at this level of temporal aggregation

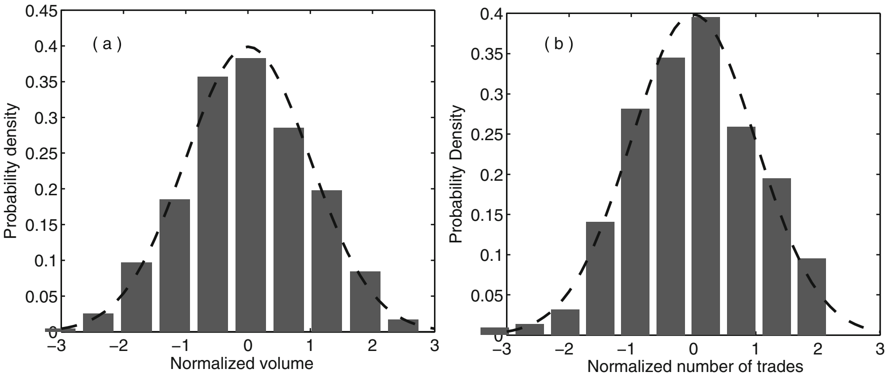

Fig. 5 Probability distribution of the de-trended and normalized trading volume (a) and number of trades (b) for all stocks in NSE aggregated over 10 days. For comparison, the standard normal distribution is shown (*broken curve*)

normality. Such tests go beyond simple regression-based best-fit of an empirical distribution by a theoretical curve and provides measures for the goodness of fit of the theoretical distribution to the data. Here, we use the Lilliefors test and the Anderson–

Darling test for testing whether the distribution of $v_{t,\text{daily}}$ and $n_{t,\text{daily}}$ approaches the Gaussian form as the data aggregation is done over longer and longer time-scales. For both these tests the null hypothesis (H_0) considered is that the empirical data is described by a Gaussian distribution. The Lilliefors test begins by estimating the mean and standard deviation of the underlying distribution from the data. It then calculates the test statistic, which is the maximum deviation of the empirical distribution from a normal distribution with the estimated mean and standard deviation. The null hypothesis is rejected when the maximum deviation becomes statistically significant. For all results reported here, we have fixed the level of significance at 5%. The p-value for the test indicates the probability of obtaining the observed maximum deviation assuming H_0 to be true, and a small value indicates that it is very unlikely that the empirical data follows a Gaussian distribution. The Anderson–Darling test is a non-parametric method for determining whether the empirical data is generated by a specific probability distribution and is considered to be one of the most powerful statistical tests for identifying deviations from normality [32]. It estimates the value of a test statistic, A^2, which is then compared with standard critical values of the theoretical distribution against which the empirical data is being tested. For example, the null hypothesis that a Gaussian distribution explains the empirical data can be rejected if the estimated test statistic A^2 exceeds 0.751.

The results of both statistical tests for the volume and number of trades data for the entire market is shown in Table 1. While the daily data clearly does not fit a Gaussian distribution, when aggregated over 10 days the trading volume does appear to be normally distributed, as the null hypothesis cannot be rejected for either of the tests we have used. Similarly, as the temporal aggregation is increased to 10 days for the number of trades data, the resulting distribution does appear to converge to a Gaussian form according to both the tests. As the time-period over which the daily data has been collected is relatively large (~ 16 years) we also checked whether the convergence to a Gaussian with increasing temporal aggregation also holds for subsets of the entire data-set. We have verified that even when the data is split into three approximately equal parts, with each sub-set corresponding to a period of about 5 years, the time-aggregated volume and number of trades distributions approach a Gaussian distribution according to the statistical tests.

Table 1 Normality tests for trading volume and number of trades for the entire NSE market at different scales of temporal aggregation

	Temporal Aggregation	Anderson–Darling test Reject H_0?	Anderson–Darling test Statistic (A^2)	Lilliefors test Reject H_0?	Lilliefors test p-value
	1 day	Y	26.631	Y	0
Volume	5 day	Y	2.1738	Y	0.0097
	10 day	N	0.2110	N	0.7100
	1 day	Y	28.694	Y	0
Trades	5 day	Y	1.4519	Y	0.0050
	10 day	N	0.3764	N	0.7950

Thus far we have been considering together all stocks that are traded in the NSE. In order to verify if the convergence to Gaussian distribution is also seen when the trading volume data for *individual stocks* is aggregated over longer periods, we shall now look at a few representative stocks from different market sectors. The cumulative distributions of the volume traded over the course of a single day for two stocks (Colgate and SBIN) are shown in Fig. 6. Both appear approximately linear in a semi-logarithmic graph, suggesting that the distribution may be fit by

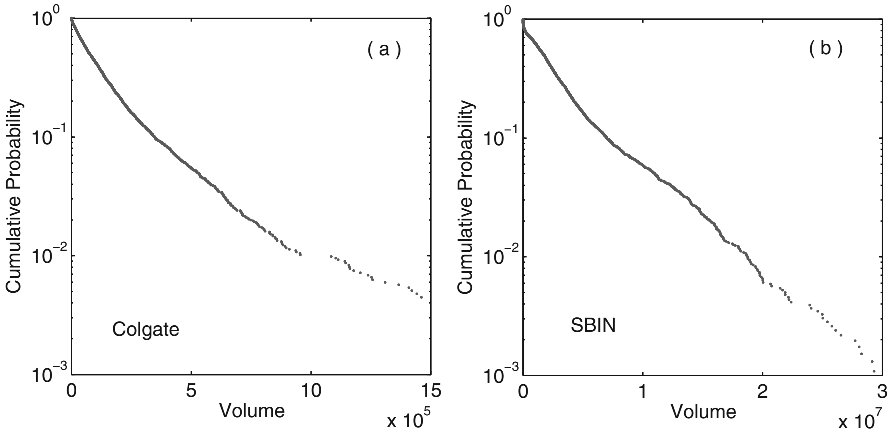

Fig. 6 Cumulative distribution of the total daily volume for two representative stocks: (a) Colgate and (b) SBIN, during the period March 1994 to May 2010. Note that the ordinate has a logarithmic scale. Thus, the linear nature of the distributions suggest that they are approximately exponentially decaying

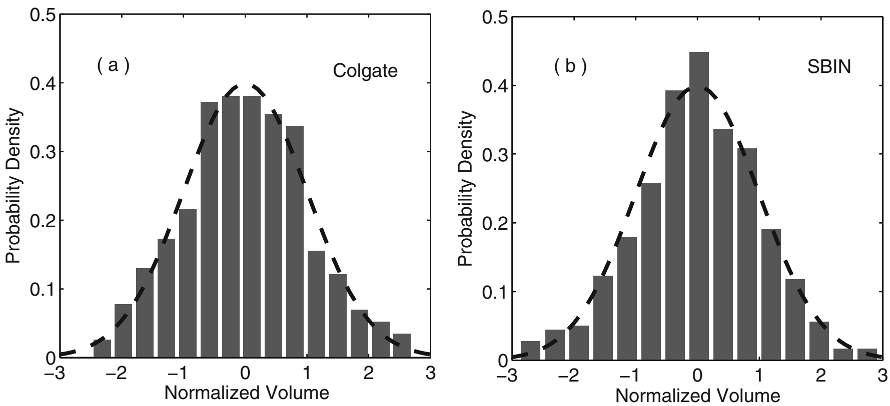

Fig. 7 Probability distribution of the de-trended and normalized trading volume aggregated over 10 days for two stocks: (a) Colgate and (b) SBIN. For comparison, the standard normal distribution is shown (*broken curve*). The time-series was de-trended by subtracting the mean calculated over a moving window (having a width 10 days and shifted in steps of 1 day) and normalized by dividing with the standard deviation calculated over the same window

an exponential form. However, when we look at the volume traded over 10 days, the corresponding de-trended and normalized distributions appear to be reasonably well-fit by the standard normal distribution (Fig. 7).

Table 2 Lilliefors test for normality of trading volume distribution for representative individual stocks in NSE

Stock	5-day aggregate		10-day aggregate	
	Reject H_0?	p-value	Reject H_0?	p-value
ABANLLOYD	Y	0.0	N	0.127
ACC	Y	0.0	N	0.727
COLGATE	Y	0.0	N	0.508
DABUR	Y	0.0	Y	0.002
DRREDDY	Y	0.001	N	0.002
GAIL	Y	0.0	Y	0.017
GLAXO	Y	0.0	Y	0.010
GODREJIND	Y	0.0	N	0.636
HCLTECH	Y	0.0	Y	0.014
HDFCBANK	Y	0.0	N	0.899
ICICIBANK	Y	0.0	N	0.558
INFOSYSTCH	Y	0.0	Y	0.0
IOC	Y	0.0	Y	0.0
RELCAPITAL	Y	0.0	Y	0.029
RELIANCE	Y	0.0	Y	0.0
SATYAMCOMP	Y	0.003	N	0.104
SBIN	Y	0.0	N	0.502
TCS	Y	0.036	N	0.787

Table 3 Anderson–Darling test for normality of trading volume distribution for representative individual stocks in NSE

Stock	5-day aggregate		10-day aggregate	
	Reject H_0?	Statistic (A^2)	Reject H_0?	Statistic (A^2)
ABANLLOYD	Y	6.696	N	0.546
ACC	Y	2.199	N	0.326
COLGATE	Y	6.691	N	0.229
DABUR	Y	7.013	Y	0.876
DRREDDY	Y	2.9960	N	0.389
GAIL	Y	4.084	Y	1.212
GLAXO	Y	6.833	Y	1.281
GODREJIND	Y	4.757	N	0.270
HCLTECH	Y	1.891	Y	1.099
HDFCBANK	Y	3.954	N	0.298
ICICIBANK	Y	3.504	N	0.611
IOC	Y	3.585	N	0.491
RELCAPITAL	Y	4.135	Y	1.113
RELIANCE	Y	7.806	Y	14.07
SATYAMCOMP	Y	2.001	N	0.521
SBIN	Y	2.537	N	0.339
TCS	Y	1.300	N	0.439

As in the case of the data for the entire market, we have carried out the Lilliefors test (Table 2) and the Anderson–Darling test (Table 3) for the volume data at different levels of aggregation. As is seen from the test results, while the volume traded over 5 days cannot be described by a Gaussian distribution for any of the stocks, when we consider the volume traded over 10 days, the Gaussian distribution appears to be a reasonable fit for many of the stocks considered.

Thus, our results indicate that, at least for the Indian market, the proposed invariant forms for the volume and number of trade distributions that have been observed in the developed markets of USA, London and Paris [17] do not hold true. In particular, the trading volume distribution does not follow a Levy-stable form. It has been suggested that, in the developed markets, the Levy-stability of the V_t distribution is a consequence of the Levy-stable trade size (q) distribution. Thus, a reason for the deviation of the volume distribution from Levy-stability could be inferred by looking at Eq. (1). If the distribution of q_i is Levy-stable but not that of N_t, the heavier tail of the former distribution would appear to dominate the nature of the tail of the V_t distribution. Presumably, this is what is happening in developed markets where we note that ζ_q and ζ_V are almost same (within error bars) [11]. However, in the Indian market, the distribution of q_i, even though it appears to fit a power-law, is clearly outside the Levy-stable regime. For instance, the exponent obtained by the CSN estimator for all trades carried out in December 2003 at NSE is $\zeta_q \simeq 2.63$ (Fig. 8). Thus, in the Indian financial market, the nature of the distribution for V_t may be dominated by that of N_t instead of the q distribution. Indeed, our earlier analysis had shown that there is a strong (almost linear) correlation between N_t and V_t [25], which would appear to support this hypothesis. It suggests that, for emerging markets where the trade size distribution has not yet become Levy-stable, the volume distribution would closely follow the distribution of the number of trades which is outside the Levy-stable region (as seen also for developed markets).

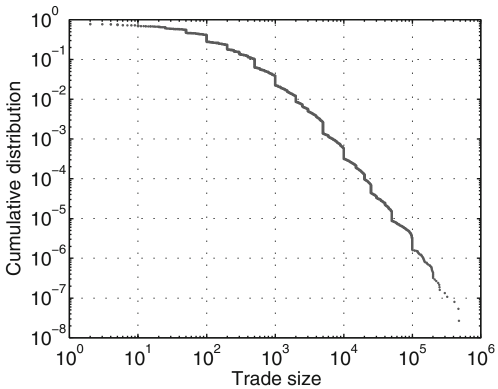

Fig. 8 Cumulative distribution of trade sizes for all transactions carried out at NSE during December 2003

4 Conclusions

In this article, we have examined the statistical properties of the distributions of the trading volume and the number of trades in the National Stock Exchange, the largest Indian financial market. Using both low-frequency (daily) and high-frequency (tick-by-tick) data, we have tried to characterize the nature of these distributions. In particular, we have sought to establish whether or not the distributions are Levy-stable by examining their stability under temporal aggregation. Our results show that although from the tick-by-tick or daily data it is difficult to exactly characterize the nature of the distribution of volume and number of trades, when we consider these quantities aggregated over a period of several days (e.g., 10 days), the resulting distribution approaches a Gaussian form. This has been verified both graphically using Q-Q plots and plots of the probability distribution functions, as well as, with statistical tests of normality, such as the Lilliefors test and the Anderson–Darling test. This suggests that the distributions of volume and number of trades are not Levy-stable, as otherwise they could not have converged to a Gaussian distribution when aggregated over a long period. Our results are significant in the context of the ongoing debate about the universality of the nature of the volume and number of trades distributions. Unlike the Levy-stable nature of the volume and trade size distributions seen in developed markets, the emerging financial market of India appears to show a very different form for these distributions, thereby undermining the claim for their universality.

Acknowledgements We are grateful to M. Krishna for invaluable assistance in obtaining and analyzing the high-frequency NSE data. We gratefully acknowledge the assistance of R. K. Pan in processing both the daily and high-frequency data.

References

1. Mantegna RN, Stanley HE (2000) An Introduction to Econophysics. Cambridge University Press, Cambridge
2. Farmer JD, Smith DE, Shubik M (2005) Is economics the next physical science? Physics Today 58(9):37–42
3. Sinha S, Chakrabarti BK (2009) Towards a physics of economics, Physics News 39 (2):33–46
4. Gopikrishnan P, Meyer M, Amaral LAN, Stanley HE (1998) Inverse cubic law for the distribution of stock price variations, Eur Phys J B 3:139–140
5. Lux T (1996) The stable paretian hypothesis and the frequency of large returns: An examination of major German stocks, Applied Financial Economics 6:463–475
6. Pan RK, Sinha S (2007) Self-organization of price fluctuation distribution in evolving markets, Europhys Lett 77:58004
7. Pan RK, Sinha S (2008) Inverse-cubic law of index fluctuation distribution in Indian markets, Physica A 387:2055–2065
8. Bouchaud JP (2005) The subtle nature of financial random walks, Chaos 15:026104
9. Plerou V, Gopikrishnan P, Amaral LAN, Gabaix X, Stanley HE (2000) Economic fluctuations and anomalous diffusion, Phys Rev E 62: R3023–R3026

10. Gopikrishnan P, Plerou V, Gabaix X, Stanley HE (2000) Statistical properties of share volume traded in financial markets, Phys Rev E 62:R4493–R4496
11. Plerou V, Gopikrishnan P, Gabaix X, Amaral LAN, Stanley HE (2001) Price fluctuations, market activity and trading volume, Quant Finance 1: 262–269
12. Farmer JD, Lillo F (2004) On the origin of power-law tails in price fluctuations, Quant Finance 4:C7–C11
13. Plerou V, Gopikrishnan P, Gabaix X, Stanley HE (2004) On the origin of power-law fluctuations in stock prices, Quant Finance 4:C11–C15
14. Eisler Z, Kertesz J (2006) Size matters: Some stylized facts of the stock market revisited, Eur Phys J B 51:145–154
15. Hong BH, Lee KE, Hwang JK, Lee JW (2009) Fluctuations of trading volume in a stock market, Physica A 388:863–868
16. Qiu T, Zhong LX, Chen G, Wu XR (2009) Statistical properties of trading volume of Chinese stocks, Physica A 388:2427–2434
17. Plerou V, Stanley HE (2007) Tests of scaling and universality of the distribution of trade size and share volume: Evidence from three distinct markets, Phys Rev E 73:046109
18. Racz E, Eisler Z, Kertesz J (2009) Comment on "Tests of scaling and universality of the distributions of trade size and share volume: Evidence from three distinct markets", Phys Rev E 79:068101
19. Plerou V, Stanley HE (2007) Reply to "Comment on 'Tests of scaling and universality of the distribution of trade size and share volume: Evidence from three distinct markets' ", Phys Rev E 79:068102
20. Queiros SMD (2005) On the emergence of a generalised Gamma distribution: Application to traded volume in financial markets, Europhys Lett 71:339–345
21. Mu G-H, Chen W, Kertesz J, Zhou W-X (2009) Preferred numbers and the distribution of trade sizes and trading volumes in the Chinese stock market, Eur Phys J B 68:145–152
22. Gabaix X, Gopikrishnan P, Plerou V, Stanley HE (2003) A theory of power-law distributions in financial market fluctuations, Nature 423:267–270
23. Vikram SV, Sinha S (2010) A mean-field model of financial markets: Reproducing long-tailed distributions and volatility correlations. In: Basu B *et al* (eds) Econophysics and Economics of Games, Social Choices and Quantitative Techniques. Springer, Milan
24. Vikram SV, Sinha S (2010) Emergence of universal scaling in financial markets from mean-field dynamics, arXiv:1006.0628
25. Sinha S, Pan R K (2006) The power (law) of Indian markets: Analysing NSE and BSE trading statistics. In: Chatterjee A, Chakrabarti B K (eds) Econophysics of Stock and Other Markets. Springer, Milan
26. Pan RK, Sinha S (2007) Collective behavior of stock price movements in an emerging market, Phys Rev E 76:046116
27. National Stock Exchange (2004) Indian securities market: A review. (http://www.nseindia.com/content/us/ismr2005.zip)
28. http://www.nseindia.com/
29. Clauset A, Shalizi CR, Newman MEJ (2009) Power-law distributions in empirical data, SIAM Review 51:661-703
30. Nolan JP (2010) Stable Distributions: Models for heavy tailed data. Birkhauser, Boston
31. Thode H (2002) Testing for Normality. Marcel Dekker, New York
32. Kvam PH, Vidakovic B (2007) Nonparametric Statistics with Applications to Science and Engineering. Wiley, Hoboken