

Self-similarity and non-Markovian behavior in traded stock volumes

Frank R. Brown¹, David Pravica¹, and Martin Bier^{2,3,a}

¹ Departement of Mathematics, East Carolina University, Greenville, NC 27858, USA

² Departement of Physics, East Carolina University, Greenville, NC 27858, USA

³ M. Smoluchowski Institute of Physics, Jagiellonian University, ul. Lojasiewicza 11, 30-348 Kraków, Poland

Received 23 August 2015 / Received in final form 13 September 2015

Published online 18 November 2015

© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract. The volume traded daily for 17 stocks is followed over a period of about half a century. We look at the volume of stocks traded in a certain time interval (day, week, month) and analyze how long that traded volume keeps monotonically increasing or decreasing. On all three times scales we find that the sequence of traded volumes behaves neither like a sequence of independent and identically distributed variables, nor like a Markov sequence. A compressed exponential survival function with the same parameters at all timescales is firmly established. A day with an increase (decrease) of traded volume is most likely followed by a day with a decrease (increase) of traded volume. We show how the apparent self-similarity results because the small day-to-day anticorrelation carries over when larger time intervals are considered. The observed small anticorrelation can be explained as a consequence of market forces and trader reactions.

1 Introduction

For the obvious monetary reasons, stock movements and stock prices have been subject to very intense analyses. However, there is also a good scientific rationale behind such studies. The stock market is an evolving complex system, but, more so than for biochemical or ecological systems, there is the availability of an abundance of data. Publicly available numbers actually constitute the input and the output of the stock market.

The focus of this article will be on volumes of traded stock. The relation between price and volume is still the subject of much research and speculation [1,2]. Traded volume can be taken as a measure of liquidity and investor interest and a price change of a stock is often considered more significant if at the same time a large volume of that stock is being traded. Getting a better understanding of volume dynamics is the goal of our analysis.

Subsequent price changes of a stock turn out to be uncorrelated when time intervals of more than about 15 minutes are taken [3,4]. There are many independent inputs that determine the price change and, by the Central Limit Theorem, we expect a Gaussian distribution for price changes of stocks. The expected Gaussian distribution of price changes, however, turns out not to emerge when the prices of actual stocks are followed to sufficient accuracy and statistical significance. It is found that the distribution of price changes of actual stocks is too leptokurtic, i.e., the tails are too fat to be successfully fit

with a Gaussian. About half a century ago Mandelbrot found good fits with power laws [5,6]. Much of the modeling and fitting since has centered around power law formalisms [3,7]. Mandelbrot emphasized that these power laws are scale free, i.e., they do not have a characteristic timescale and the same law is found for different timesteps. The resulting self-similar or fractal structures are central in his major text on the subject [6].

Below we take the volume that is traded during a day and compare it to the volume that is traded during the next day. We will not concern ourselves with the exact distribution of the increments. For our analysis we take only the *sign* of the change and disregard the magnitude. From a modeling perspective, this approach is justified by the fact that traders commonly react to “decrease” or “increase” before they react to actual numerical values. By looking at just the sign, we simplify the analysis. To ensure that we have sufficient data for statistically significant results we go back about half century. We also follow the volumes traded during time intervals other than a day (week, month). The question we next wish to answer is: does an increase (decrease) of a stock’s traded volume make an increase (decrease) in the next timestep more likely? When the context is the price of the stock, this is called “momentum” and traders are familiar with how a price increase (decrease) can trigger subsequent price increases (decreases). The memory and momentum in the market is what is supposed to give rise to bubbles and crashes [8]. In what follows we will use the notion of “momentum” also when analyzing subsequent increments of traded volume.

^a e-mail: bierm@ecu.edu

In the simplest possible case, without any memory, a sequence of traded stock volumes would be a sequence of independent and identically distributed (IID) numbers. The volume V of traded stock is bounded from below by zero and bounded from above by a limit that is set by how fast stocks can change hands. Assume that for the volume of traded stock during Δt , there is a constant probability distribution on this interval. Suppose next that during every time interval, the traded volume is a random draw from this distribution. This would imply an IID sequence.

Imagine doing three draws from an IID sequence and let these draws represent traded volume. Next randomly label these draws as 1, 2, and 3. There are six possible ways, i.e. permutations, to do this labeling. We now have V_1 , V_2 , and V_3 . The case that $V_3 > V_2 > V_1$, i.e. two subsequent increases, is just one of the six possible permutations (ignoring equality) and has a probability of 1/6. Thus, the probability of two subsequent increments that are both positive is 1/6.

Extending the reasoning of the previous paragraph, we find that the probability of t or more subsequent increments of the same sign equals

$$S(t) = 1/(t + 1)! \quad (1)$$

Equation (1) is valid for any distribution! So by looking at just the sign of the increments and not at the exact values, we circumvent a major source of computational difficulty. We will include the IID sequence in our graphs as a point of reference. There is no memory as we move from one volume in the IID sequence, V_{j-1} , to a next volume V_j . However, it is important to realize that there is an anti-momentum effect associated with the IID sequence when we consider subsequent increments $\Delta V_j = (V_j - V_{j-1})$ and $\Delta V_{j+1} = (V_{j+1} - V_j)$. This can be easily intuited by imagining the distribution to be a Gaussian with average V_{avg} . Take three subsequent volumes V_{j-1} , V_j , and V_{j+1} that are drawn from that distribution. For V_{j-1} the average outcome is V_{avg} . An increase, i.e. $\Delta V_j = (V_j - V_{j-1}) > 0$, implies that $V_j > V_{avg}$. Now the next draw, V_{j+1} , is more likely going to be to the left than to the right of V_j . This means that the positive increment, $\Delta V_j > 0$, will with more than 50% probability be followed by a negative increment, $\Delta V_{j+1} = (V_{j+1} - V_j) < 0$. This effect generalizes to any distribution and is actually implicit in equation (1).

We also include in our analysis a model with a one-step-deep memory. To do so we take the sign of the difference $\Delta V_j = (V_j - V_{j-1})$ and next formulate a Markov model for the sign of ΔV_{j+1} . Suppose that a stock's traded volume has increased from day (week, month) $j - 1$ to day (week, month) j , i.e. $\Delta V_j > 0$. In this simple approach, the probability that it will move to the decreasing state in the next timestep, i.e. $V_{j+1} < V_j$, can be represented by a transition rate. In other words, we consider a Markov system:

$$\mathbf{I} \Rightarrow \mathbf{D}, \quad (2)$$

where \mathbf{I} represents the state in which the stock's traded volume is increasing and \mathbf{D} represents the state in which the stock's traded volume is decreasing. The transition

rate from \mathbf{I} to \mathbf{D} is α_{ID} . The transition rate from \mathbf{D} to \mathbf{I} is α_{DI} .

With the mechanism of Schematic (2) it is important to realize that, in terms of the V_j 's, the "memory of the market" goes back one step, as both V_j and V_{j-1} are necessary to establish the sign of $\Delta V_j = V_j - V_{j-1}$. In terms of the ΔV_j 's, however, there is no memory. The sign of ΔV_{j+1} depends only on the sign of ΔV_j and on a transition rate α , i.e. we have a two-state Markov process.

The Markov setup (2) allows for the description of momentum effects. When the transition rates, α_{ID} and α_{DI} , are small, it is hard to get out of the state that you are in. That means that an increase (decrease) of traded volume is more likely if traded volume is already increasing (decreasing), i.e., it means momentum.

With this Markov model the probability $S(t)$ that a stock survives in the same state for t days or longer is

$$S(t) = \exp[-t/\tau], \quad (3)$$

where $\tau = 1/\alpha_{ID}$ for the increasing state and $\tau = 1/\alpha_{DI}$ for the decreasing state. If increase or decrease of traded volume were determined by a coin toss with an unbiased coin, we would have $\alpha_{ID} = \alpha_{DI} = \ln 2$ and $S(t) = 2^{-t}$ for both states. For $\tau \neq 1/\ln 2$ we have a biased coin toss, i.e. the survival probability is different from 1/2 at $t = 1$. Also the biased-coin-toss Markov sequence will be included in our graphs as a point of reference.

An awareness of the past that goes deeper than just one step can be implemented through time-dependent transition rates in Schematic (2). In that case we no longer have a Markov process. A common way to implement such a memory is by taking power law behavior, i.e. $\alpha \propto t^{k-1}$, for both the involved rates. For future convenience we adopt the form

$$\alpha(t) = \frac{k}{\tau} \left(\frac{t}{\tau}\right)^{k-1}. \quad (4)$$

Here t is the time that the stock has been in its present state and $\alpha(t)$ is the transition rate out of that state at time t . For the decrease of $S(t)$ over the next time interval dt we then have $dS(t) = -\alpha(t)S(t)dt$. We now find for the probability $S(t)$ that a stock survives in the same state for t days or longer:

$$S(t) = \exp\left[-\left(\frac{t}{\tau}\right)^k\right]. \quad (5)$$

This is the survival function for the well-known Weibull distribution [9]. We can conceive of k as a *memory parameter*. For $k = 1$ the Markovian setup of Schematic (2) and the accompanying exponential survival function (cf. Eq. (3)) are retrieved. For $k > 1$ we have a transition rate out of a state that increases as the system spends more time in that state. In our case this corresponds to traders increasingly holding on to a stock as more volume of that stock is traded. It also implies an increasing urge to start trading a stock if less and less volume of that stock is traded. $S(t)$ then takes the form of a

compressed exponential. For $k < 1$, a stretched exponential, which extends increasing or decreasing trends, results. Survival curves like equation (5) have been commonly found in the analysis of the failure of devices like light bulbs. This is because the failure rate of a device often depends on the age of the device [10]. Time-dependent transition rates as in equation (4) and the resulting compressed and stretched exponentials as in equation (5) have furthermore been employed in the analysis of the kinetic behavior of proteins [11,12].

With equations (4) and (5), the memory of the system goes back to the moment that the system came into the state (I or D) that it is in. So it is a short-term memory. The time varying transition rates, furthermore, imply that the aforementioned momentum can grow or decay as the time in a state evolves.

In the next section we will show how equation (5), with $k > 1$, leads to a good fit of the data. Sections 3 and 4 contain theoretical analysis of the results. Section 5 is the Discussion section.

2 Data analysis

For the seventeen stocks listed in Table 1, we take data going back four to five decades. The day-to-day volumes are freely available on the world wide web¹. For each sequence of subsequent increases (decreases) we record how long it lasts and how many times a sequence of that length occurs in the entire datafile. We organize the data in a survival histogram as explained in Figure 1. Ultimately, we have 18 sets of more than 10^4 data points each. Our analysis of the two-state system mimics an analysis that is commonly performed on ion channels that flip between an open and a closed state [12]. Having more than 10^4 sampling points in the original record has generally been found to be sufficient to get good estimates of transition rates or to establish a self-similar nature.

Equation (5) can be represented as a linear function by taking the logarithm of the negative logarithm of $S(t)$ for the vertical coordinate and taking $\ln t$ for the horizontal coordinate:

$$\ln[-\ln S(t)] = k \ln t - k \ln \tau. \quad (6)$$

The slope can be identified with the memory parameter k . The rate parameter τ can next be determined from the intercept with the vertical ($\ln t = 0$) axis. Figure 2 shows fits of equation (6) to the cumulative record of “All” stocks. It is apparent that the Weibull exponential, cf. equation (5), with $k \approx 1.2$ is a good fit. In all six graphs neither the IID sequence (cf. Eq. (1)), nor the Markovian coin toss (cf. Eq. (3)) can account for the data.

For the three different timescales (daily, weekly, and monthly) Figure 3 depicts k and τ values for our seventeen stocks. At all three timescales and for both the

¹ Yahoo! Finance, <http://finance.yahoo.com>. After entering the symbol for a stock, go to “Historical Prices” via the menu on the left. Next, at the bottom of the page, there is a link to a CSV file with the pertinent data

Table 1. The seventeen stocks selected for analysis. We also include the sum of all stock data.

Symbol	Stock	Starting Date
BA	Boeing	January 2, 1962
CAT	Caterpillar	January 2, 1962
CVX	Chevron	January 2, 1970
DD	DuPont	January 2, 1962
DIS	Disney	January 2, 1962
GE	General Electric	January 2, 1962
HON	Honeywell	January 2, 1970
HPQ	Hewlett-Packard	January 2, 1962
IBM	IBM	January 2, 1962
JNJ	Johnson and Johnson	January 2, 1962
KO	Coca Cola	January 2, 1962
MCD	McDonald’s	January 2, 1970
MMM	3M Corp	January 2, 1970
MRK	Merck	January 2, 1970
PG	Procter and Gamble	January 2, 1970
UTX	United Technologies	January 2, 1970
XOM	Exxon-Mobil	January 2, 1970
(All)	All stocks	January 2, 1962

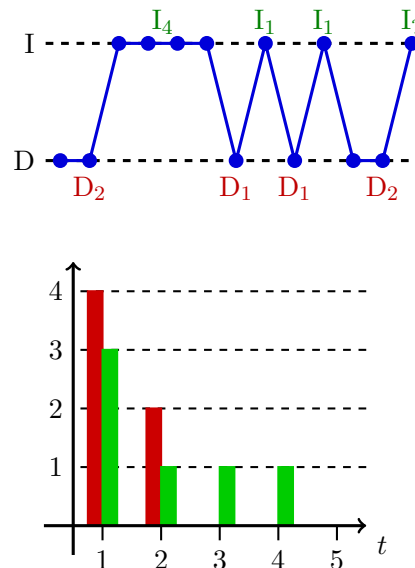


Fig. 1. An increasing/decreasing record and the corresponding histogram of survival times. The green bars correspond to the “increasing times,” where we have one interval of four units and two intervals of one unit. The “decreasing times” are represented by the red bars. There we have two intervals of two units and two intervals of one unit.

increasing and the decreasing state, we find k to be significantly different from unity, i.e. there is a notable deviation from Markov behavior. Merging the seventeen data files for the stocks, we find the results for the cumulative “All” in Table 1. These estimated k and τ for the “All” file are tabulated in Table 2. Figure 3 and Table 2 lead to the conclusion that we have a remarkable self-similarity. Over a

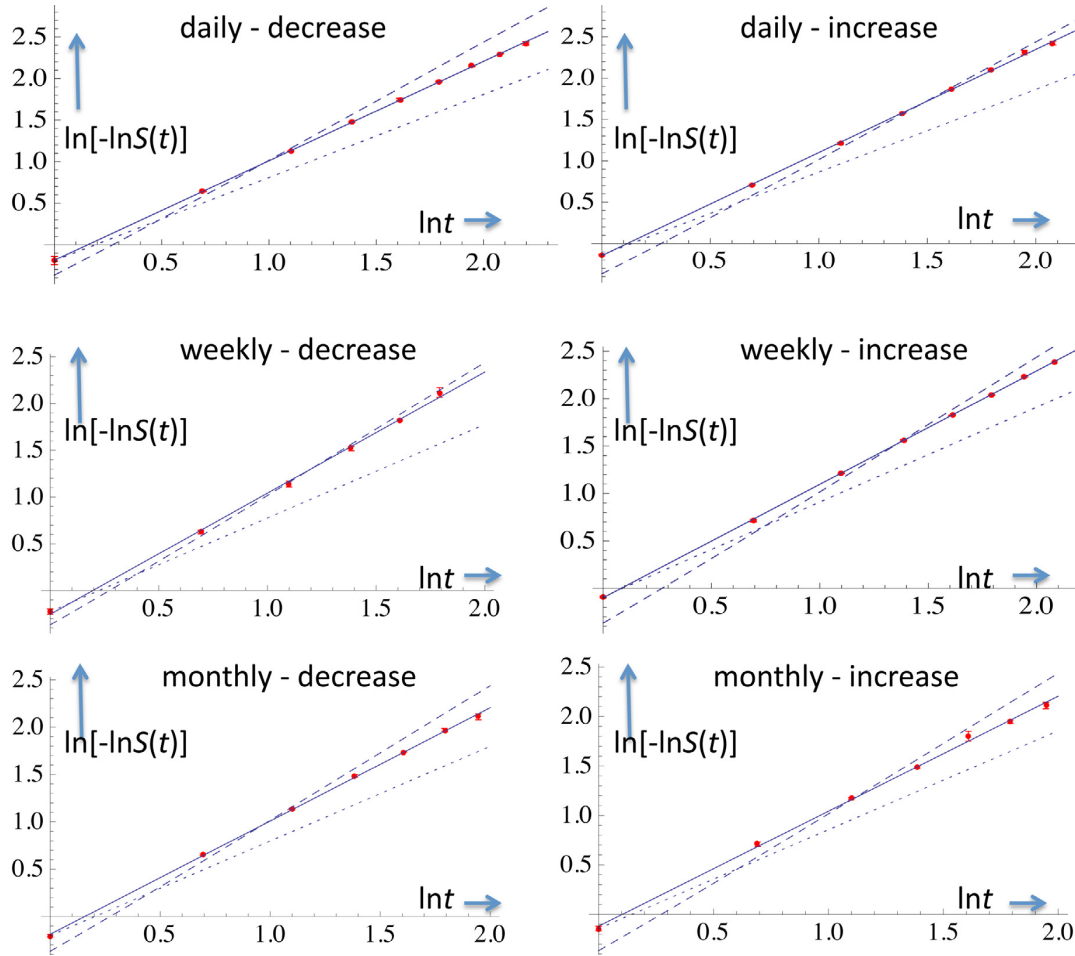


Fig. 2. For the traded volume per day, per week, and per month of all of the stocks together (last entry in Tab. 1), we followed the procedure explained in Figure 1 and recorded and tabulated how long the runs of subsequent increases and decreases lasted. Shown in red are the datapoints $(\ln(-\ln S(t)), \ln t)$, where $S(t)$ is the survival function for $t = 1, 2, 3, \dots$ days. The 95% confidence interval are indicated and are generally smaller than the size of the point. The result appears significantly different from what an IID sequence (dashed lines, cf. Eq. (1)) and a biased Markovian coin toss (dotted lines, cf. Eq. (3)) would give. The solid line represents the fit of a compressed exponential (cf. Eq. (5)). The values of k and τ that give the best fits are indicated in Table 2.

range of timescales and for both increasing and decreasing runs, a compressed exponential with $k = 1.215 (\pm 0.058)$ is observed.

As a further test of self-similarity, we evaluate our data sets using the three-sample Anderson-Darling test [13] on the data sets daily-weekly-monthly for decreasing and increasing sequences for each stock and for the combination of data for all stocks. The test determines to what extent two or more distributions are identical. All data set triplets (daily-weekly-monthly) for individual stocks for both increasing and decreasing states passed the Anderson-Darling test at the 0.01 level. The combined data for all stocks passed for the decreasing state but failed for increasing. Collectively, these results provide strong support for self-similar behavior for all the individual stocks and partial support for the combined data set.

The numbers that are collected in Table 2 are indicative of a significant anti-momentum effect. A Markovian coin toss would give rise to $(k, \tau) = (1, 1/\ln 2) = (1.00, 1.44)$. This would obviously lead to $S(1) = 0.50$, $S(2) = 0.25$, $S(3) = 0.125$, etc. The $(k, \tau) = (1.2, 1.16)$ that we find for the daily data, leads to $S(1) = 0.43$, $S(2) = 0.15$, and $S(3) = 0.044$ for the first three days. So there is an apparent anticorrelation, i.e. when the system is in one state, the probability to be in the other state at the next timestep is always larger than $1/2$. This is suggestive of a restoring force that drives the traded volume per day (week, month) back to an average. It is interesting to note that, where we find an anti-momentum for traded volumes, momentum and “superdiffusion” have been observed for the movement of stock prices (in particular, the S&P 500 index) [14].

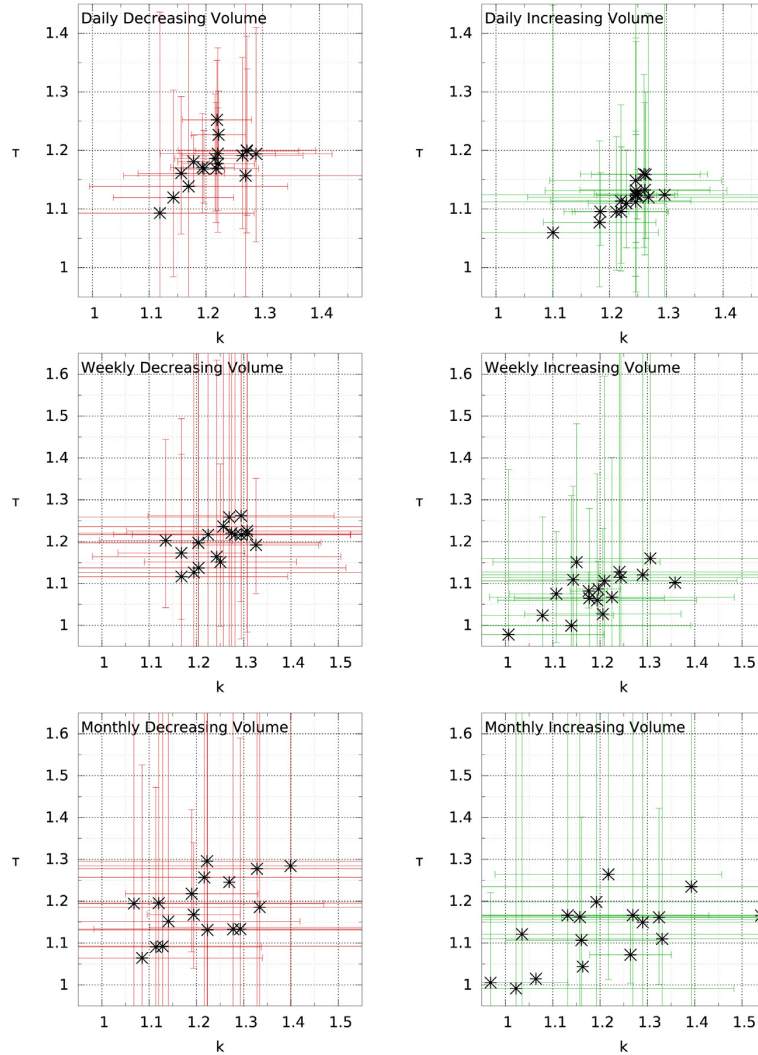


Fig. 3. The results of fits of the Weibull parameters k and τ (cf. Eq. (5)) for the volume of traded stock for three different timescales: daily (top), weekly (middle), and monthly (bottom). We find a compressed exponential at all timescales for both the increasing and the decreasing state. With the average k of 1.2 on all timescales we find the process to be self-similar. Error bars indicate 95% confidence intervals.

Table 2. The values for k and τ when the volume data for all stocks together are fit to equation (6). To obtain these numbers a linear regression was performed. The 95% confidence intervals for τ were calculated. We took the further of the two $\ln \tau$ end points to determine the confidence intervals.

k values	Increasing	Decreasing
Daily	1.246 (± 0.045)	1.194 (± 0.025)
Weekly	1.197 (± 0.030)	1.295 (± 0.139)
Monthly	1.163 (± 0.099)	1.196 (± 0.059)
τ values	Increasing	Decreasing
Daily	1.124 (± 0.070)	1.169 (± 0.043)
Weekly	1.087 (± 0.041)	1.216 (± 0.231)
Monthly	1.108 (± 0.123)	1.170 (± 0.097)

3 Self-similarity

Self-similarity, i.e. repetition of the same pattern on different timescales, appears to be a key feature of our data.

The basis of equation (1) is an IID sequence. Of course, if the traded volume is an IID sequence on the day-to-day timescale, then it will also be an IID sequence on the week-to-week and on the month-to-month timescale. After all, in all cases you are doing independent draws from a distribution. Because we take only the signs and not the exact values of the day-to-day increments, the sequence for the day-to-day timescale does not determine the sequences for week-to-week and month-to-month.

For the case of a Markov sequence, it is obvious that the day-to-day unbiased coin toss, $S(t) = 2^{-t}$, leads to

the same $S(t) = 2^{-t}$ on the week-to-week and month-to-month timescales. In that case the probability distribution associated with $t = 1$ is always 50-50. However, our data show a small anticorrelation from one point in time to the next. Within the accuracy achievable with our datasets, the anticorrelation appears to have the same value at all of the three considered timescales.

Below we show how an anticorrelation for short time intervals carries over to the longer time intervals. With $S(t) = \exp[-t/\tau]$ and $\tau = 1/\ln 2$, we have a 50-50 distribution at time $t = 1$. We describe a small anticorrelation from one timestep to the next by taking $\tau = (1 - \varepsilon)/\ln 2$, where ε is small. A first order Taylor approximation gives $S(t) \approx \exp[-(\ln 2)t] (1 - \varepsilon(\ln 2)t)$ for small deviations from the unbiased case. With $\delta = \varepsilon(\ln 2)$ and considering one step ($t = 1$), we have for the probability to stay in the same state and for the probability to switch to the other state, respectively,

$$P_1(\text{same}) = \frac{1}{2}(1 - \delta) \quad \text{and} \quad P_1(\text{other}) = \frac{1}{2}(1 + \delta). \quad (7)$$

We let $P_n(\text{same})$ be the probability that the system is in the same state at time n as that it was at time $n = 0$. Likewise, $P_n(\text{other})$ is the probability that the system is in the other state at time n . To derive the probabilities $P_n(\text{same})$ and $P_n(\text{other})$ we use

$$\begin{aligned} 1 &= (P_1(\text{same}) + P_1(\text{other}))^n \\ &= \sum_{j=0}^n \binom{n}{j} P_1^j(\text{same}) P_1^{n-j}(\text{other}) \end{aligned} \quad (8)$$

and realize that we go to the other state in the course of n steps if $(n - j)$ is odd. After taking $q = P_1(\text{other})/P_1(\text{same})$ and rewriting equation (8) as $1 = f_n(q) \equiv P_1^n(\text{same}) (1 + q)^n$, we can identify $P_n(\text{same})$ with the even part of $f_n(q)$, i.e. $P_n(\text{same}) = (f_n(q) + f_n(-q))/2$. After some algebra we then derive

$$\begin{aligned} P_n(\text{same}) &= \frac{1}{2} (1 + (-\delta)^n) \quad \text{and} \\ P_n(\text{other}) &= \frac{1}{2} (1 - (-\delta)^n). \end{aligned} \quad (9)$$

If the increases and decreases are always by the same amount, then we have for the average net change after N steps

$$\begin{aligned} \Delta V &\propto \sum_{n=1}^N [P_n(\text{same}) - P_n(\text{other})] \\ &= [-\delta + \delta^2 - \delta^3 \dots + (-\delta)^N] = -\frac{\delta(1 + (-\delta)^N)}{1 + \delta}. \end{aligned} \quad (10)$$

For large N this sum yields $\Delta V \propto -\delta/(1 + \delta)$. For small δ we can just take the first-order approximation, which leads to $\Delta V \propto -\delta$.

This last result is what contains the scale invariance. If, for instance, there was an increase on day t , then a decrease is not just more likely on day $t + 1$. Also over

a subsequent block of days, $t + 1, t + 2, \dots, t + N$, a net decrease is more likely than a net increase. For small δ the anticorrelation between a day and a subsequent block of days is effectively the same as between a day and the next day. Of course, the same reasoning applies if we go backward in time (from t to $t - 1, t - 2, \dots, t - N$) and it is equally valid when going from a block of subsequent days to the one day before or after the block. Ultimately, this means that the same anticorrelation that exists between a day and the next day also exists between a block of days of length n_1 and a subsequent or previous block of days of length n_2 .

The scale invariance happens because the exact form of the survival curve $S(t)$ is not important when, from one day to the next, survival-or-not is close to a 50-50 coin toss. For the (anti)correlation between two subsequent blocks it is ultimately only the last day of block 1 and the first day of block 2 that matter. Between these two days the (anti)correlation is of order δ . All the correlations between all the other days together sum up to order δ^2 .

The reasoning of the previous paragraph still applies for a compressed or stretched exponential (cf. Eq. (5) with $k \neq 1$) as long as the (anti)correlation from one day to the next is small. If we take $S(t) = \exp[-(t/\tau)^k]$ with $\tau = (1 - \varepsilon)/\ln 2$ and $k = 1 + \kappa$, then ε and κ describe the deviation from the unbiased coin toss. Expanding at $t = 1$ and only keeping the first order terms in ε and κ , it is found that:

$$S(1) \approx \frac{1}{2} [1 - \ln 2 (\kappa \ln(\ln 2) + \varepsilon)]. \quad (11)$$

The δ that denotes the anticorrelation in equation (7) is now given by $\delta = \ln 2 (\kappa \ln(\ln 2) + \varepsilon)$.

The difference between the biased coin-toss and the compressed exponential only becomes significant when considering sequences of three or more days, i.e. when considering $S(2), S(3), \dots$. Ultimately, it is only at order δ^2 that non-Markovian behavior and ‘‘memory’’ can be identified.

Summarizing, the scale invariance can be understood in the following way. For the unbiased Markovian coin-toss it is obvious that going to the other state vs. staying in the same state is 50-50, whether it is day-to-day, week-to-week, or month-to-month. For a small deviation δ from the unbiased case, the breaking of the scale invariance occurs only at order δ^2 . For sufficiently small δ this is outside the limits of our accuracy. So though τ in equations (3) and (5) has the dimension of time, it is ultimately not a characteristic timescale of the system. The (anti)correlation between neighboring days works out to also be the (anti)correlation between neighboring blocks of days.

4 Deriving an underlying map

We model our day-to-day dynamics with an iterative map

$$x_{t+1} = T(x_t) = x_t + f(x_t) \quad \text{on } x \in (0, c). \quad (12)$$

The $(0, c)$ interval corresponds to a state X (I or D). We start with a random seed x_0 from a uniform distribution over this interval. We compute the next iterates using equation (12). The first iterate x_n outside the $(0, c)$ interval signifies a jump to the other state. In that case we have an n -step stay in state X before “escape” occurs. For the unbiased Markovian coin-toss case, equation (12) would simply be the well-known Bernoulli Map [15], i.e. $f(x) = x$ (so $x_{t+1} = 2x_t$) and $c \in \mathcal{R}^+$.

A coin toss with small bias δ corresponds to $x_{t+1} = 2(1 + \delta)x_t$ and again $c \in \mathcal{R}^+$. In that case a fraction $1/(2(1 + \delta)) \approx \frac{1}{2}(1 - \delta)$ stays inside the interval upon an iteration. The escaping fraction then corresponds to $P = 1 - 1/(2(1 + \delta)) \approx \frac{1}{2}(1 + \delta)$, as in equation (7).

Giving $f(x)$ a concavity results in deviation from the well-understood behavior of the Bernoulli map. To derive the $f(x_t)$ that corresponds to our compressed exponential we follow a standard method [16].

Writing equation (12) as $x_{t+1} - x_t = f(x_t)$, we can identify the left-hand-side with a derivative as long as $(x_{t+1} - x_t)$ is sufficiently small. From $dx_t/dt = f(x_t)$, we then get t_{esc} , which is an approximation for the total time in state X before the escape:

$$\int_{x_0}^c \frac{dx}{f(x)} = t_{esc}. \tag{13}$$

The probability density function $\Psi(t_{esc})$ of the waiting times is related to the distribution $p_{in}(x_0)$ of injection points. We assume this latter distribution to be uniform, i.e. $p_{in}(x_0) = 1/c$. From $\Psi(t_{esc})dt_{esc} = p_{in}(x_0)dx_0$, we then have

$$\frac{1}{c} \int_{x_0}^c dx = \int_0^{t_{esc}} \Psi(\theta) d\theta = 1 - S(t_{esc}), \tag{14}$$

where $S(t)$ is the survival function (cf. Eq. (5)). This equation leads us to $x_0 = cS(t_{esc})$. Substituting this result in equation (13), we establish a relation between $f(x)$, which characterizes the map, and the measured survival function $S(t)$:

$$\int_{x_0}^c \frac{dx}{f(x)} = S^{-1}(x_0/c). \tag{15}$$

This gives us

$$f(x) = - \left[\frac{dS^{-1}(x/c)}{dx} \right]^{-1}. \tag{16}$$

It is easy to check that our compressed exponential $S(t) = \exp[-(t/\tau)^k]$ leads to

$$f(x) = \frac{kx}{\tau} (-\ln[x/c])^{(1-1/k)}. \tag{17}$$

The iteration map now takes the form

$$x_{t+1} = T(x_t) = x_t \left(1 + \frac{k}{\tau} (-\ln[x_t/c])^{(1-1/k)} \right). \tag{18}$$

We first observe here that c cancels out if we rescale: $x' = cx$. So c is ultimately just a scaling constant that we may

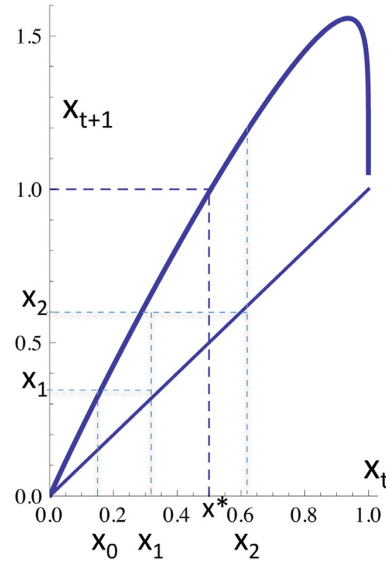


Fig. 4. The map $x_{t+1} = T(x_t)$ (cf. Eq. (18)) for $c = 1$, $k = 1.2$, and $\tau = 1.15$. It is shown how a run starts with a random number x_0 on the unit interval and how the next iterates are generated. A run with n iterates comes to an end when $x_n = T(x_{n-1}) > c$. The distribution of run lengths mimics the distribution corresponding to the survival function $S(t)$ (cf. Eq. (5)).

set equal to unity without loss of generality. Figure 4 shows $x_{t+1} = T(x_t)$ for $c = 1$, $k = 1.2$, and $\tau = 1.15$.

The dashed line in Figure 4 indicates the value for which $T(x^*) = 1$. Any initial value $x_0 \in (x^*, 1)$ is immediately mapped out of the unit interval. The segment $(0, x^*)$ is mapped onto the unit interval and x^* should be approximately equal to $S(1)$ (cf. Eq. (11)). The values x^* and $S(1)$ differ by a small amount because equation (13) is an approximation.

Figure 4 shows that, for our parameter values, the curvature of $T(x_t)$ is small on $0 < x_t < x^*$. Nevertheless, it is this curvature that is associated with the non-Markovian behavior and the “memory” of the system. For $k = 1$, equation (18) reduces to $x_{t+1} = (1 + 1/\tau)x_t$. In case $k \neq 1$, the term that multiplies x_t on the right hand side of equation (18) depends on x_t itself and it changes as x_t grows in the course of subsequent iterations.

The “memory” can also be understood in the following way. Start with the unit interval $(0, 1)$. If $k = 1$, then the same fraction $(1 - x^*)$ is mapped out of the unit interval at every subsequent iteration. If $k \neq 1$, i.e. if there is curvature, then the fraction that is mapped out of the unit interval changes from one iteration to the next. This is easily understood after realizing that the curvature on $(0, x^*)$ makes a uniform distribution on $(0, x^*)$ map onto a nonuniform distribution on $(0, 1)$. With a nonuniform distribution on $(0, 1)$, $(1 - x^*)$ is no longer the fraction that is mapped out of the interval at the next iteration. The relation between the iteration number and

the “mapped-out fraction” is what ultimately constitutes the memory of the system.

5 Discussion

There are several versions of the Efficient Market Hypothesis. In one of its weakest formulations, the hypothesis states that prices follow a random walk with no time correlations. For stock prices such a random walk is indeed what is found. If the time interval Δt is taken to be sufficiently large (in practice $\Delta t \approx 15$ minutes), then price changes follow a white noise pattern, i.e. a flat power spectrum and no long or short term correlations. However, there are market variables that do exhibit a memory. When the volatility of a stock price (the standard deviation around an average price over a fixed period) is followed for a long time, long term correlations are observed as well as spectral densities that follow a $S(\omega) \propto 1/\omega^\gamma$ pattern, i.e. a power law, where γ is close to unity [3].

The IID sequence (Eq. (1) and the dashed line in Fig. 2) makes for a random walk where today’s traded volume is not related to tomorrow’s traded volume, except that they are draws from the same distribution. In the case of the biased coin-toss (Eq. (3) and the dotted line Fig. 2) there is a very shallow memory. Tomorrow’s ΔV_{j+1} depends on today’s ΔV_j and on a probability distribution. As $\Delta V_j = V_j - V_{j-1}$, the memory is effectively one step deep.

The fact that the compressed exponential with $k = 1.2$ is an almost perfect fit in Figure 2 means that the market generally exhibits a memory that goes deeper than one step. Today’s increment in traded volume depends on the increments of days past! However, it is only at second order in the small day-to-day anticorrelation δ that the memory is apparent.

For the traded volume the same compressed exponential is found in a range of timescales varying from days to months. We explained this apparent self-similarity as a consequence of the fact that the day-to-day anticorrelation δ is very small. Correlations that are not between nearest neighbors are of order δ^2 and higher and thus are negligible. The correlation of the traded volume of two subsequent weeks, week 1 and week 2, ultimately boils down to the correlation between the Friday of week 1 and the Monday of week 2.

Stretched exponentials and a transition rate (cf. Eqs. (4) and (5)) that varies as a negative power of t have been commonly found in the analysis of ion channels and other proteins. In many cases, underlying mechanisms and elegant models that give rise to such decreasing transition rates have been formulated [11,12,17]. For a compressed exponential no such simple model suggests itself. However, we have formulated here a “first order” explanation for why a compressed exponential with the same characteristics is observed at different timescales.

We repeated our analysis for stock prices. More than with volume, long term trends, like inflation, may be an issue with price and such long term trends may interfere

with good statistics. However, we look at the lengths of sequences of subsequent increases (or decreases). Sequences of ten or more subsequent increases (or decreases) are already extremely rare and ten timesteps is a sufficiently short span to ignore the long term trends. For both the increasing and the decreasing state we found $k \approx 1.05$ for the day-to-day changes of the closing prices. For the weekly and monthly closing prices the deviation from the Markovian $k = 1$ was not statistically significant. However, when performing the analysis for the price, a serious ambiguity occurs. Stock prices stay the same on subsequent days in about 5% of all cases. Added to this ambiguity is the fact that, prior to decimalization in 2000 and 2001, stock prices were given to an accuracy of 1/8 of a dollar. So there are dilemmas when stock prices remain the same with more such dilemmas occurring before 2000. Such dilemmas are not a major issue when volume is considered. For each of the stocks in Table 1 the daily traded volume is of the order of 10^5 . That the same volume is traded on two subsequent days happens only in 0.20% of all cases. For subsequent weeks and months it is 0.034% and 0.011%, respectively. In these rare cases where there was no change in volume, we included that iteration in the sequence as if it was an increase.

For stock prices, $k = 1$ and unbiased-coin-toss statistics are the expected outcome. Such an outcome is consistent with the weakest formulation of the Efficient Market Hypothesis. If the k for a stock price were significantly different from one, it would imply a pattern and a predictability of the price. The idea of the Efficient Market Hypothesis is that if a stock price were to go through predictable highs and lows, traders would start buying when low and selling when high. Through a simple supply-and-demand mechanism these traders’ actions would then eliminate the pattern.

Other researchers have studied traded volumes with other methods. In references [18,19] long term correlations for traded volumes are reported. It is still a puzzle as to what the underlying mechanism is for such a long term memory. In this article we have been looking at the *sign of the increments* of traded volume. We, effectively, took the derivative of the time sequence and just followed for how long it stays positive or negative. We cannot identify long term correlations as our analysis erases all memory when the state of the system switches from **I** to **D** or **D** to **I** (cf. Fig. 1). However, we do find a short term correlation.

We find an anti-momentum ($S(1) < 1/2$) that even grows as the time in a state gets longer ($\alpha_{DI} = \alpha_{ID} \propto t^{0.2}$). Such anti-momentum is readily explained in terms of trader behavior. In reference [20] a distinction is made between buyer-initiated traded volume, V_b , and seller-initiated traded volume, V_s . Next, it is empirically found that the price change, Δp , is a monotonically increasing function of $\Delta V = V_b - V_s$. In reference [21] a mechanism is formulated that leads to $\Delta p \propto \text{sgn}(\Delta V) \sqrt{|\Delta V|}$. The picture that thus emerges is the following. A trader that wishes to buy a certain stock drives up the price significantly when he buys too much volume of that stock in too short a time. Likewise, a trader that wants to sell a

certain stock decreases the price significantly by putting too much volume of that stock on the market all at once. In wanting to not affect a stock's price through their own actions, traders will thus start trading less if too much volume is traded. We thus have a negative feedback loop, i.e. an anti-momentum, through which the speed of trading will converge to an optimum. This optimum is the compromise where traders move towards their ideal portfolios at a speed that keeps price movement at a manageable minimum. How anxiety can drive action-versus-inaction dynamics in general has actually been modelled with Weibull distributions [22].

For the dichotomous signal $f(t)$ of Figure 1 with residence times in both states given by the same compressed exponential (cf. Eq. (5)), we find for the autocorrelation function [23,24]:

$$g(\Delta t) = \langle f(t)f(t + \Delta t) \rangle = \exp \left[-2 \left(\frac{\Delta t}{\tau} \right)^k \right]. \quad (19)$$

The Wiener-Khinchin Theorem tells us that the associated spectral density, $S(\omega)$, is the Fourier transform of the autocorrelation function $g(\Delta t)$, i.e.

$$S(\omega) = \int_{-\infty}^{\infty} g(t)e^{i\omega t} dt = 2 \int_0^{\infty} e^{-2(t/\tau)^k} \cos(\omega t) dt. \quad (20)$$

This spectrum does not resemble a power law and no significant part of it is well approximated by a power law. As was mentioned before, spectral densities of the power-law-type, $S(\omega) \propto 1/\omega^\alpha$, are commonly found in econophysics [3,19,25]. The spectrum of equation (20) actually does not look very different from the well-known Lorentzian spectrum ($S(\omega) \propto 1/(\omega_0^2 + \omega^2)$, where ω_0 is the constant characteristic frequency) that is associated with Markovian transitions [24,26]. In our case the characteristic frequency ω_0 is very close to $1/\tau$. We worked with timesteps of length one. This gives us a sampling rate of one. For $1/\tau$ we have values close to unity (cf. Tab. 2). So the region $\omega < 1/\tau$ constitutes the part of our spectrum that we have data for and this part is flat.

Much of the scientific analysis of stock prices and stock traffic has concentrated on long term correlations and on the large fluctuations in the tails of the distributions. This is sensible as enigmatic market crashes are such large and rare fluctuations. Our focus, however, has been on how trade and traders operate on the short term. We follow the sign of the day-to-day change in traded volume. We have shown how the day-to-day pattern carries over when larger time intervals are considered. The pattern we have identified can be qualitatively understood as a consequence of market forces and trader behavior. It is not yet clear how the numerical values that we found for k and τ (cf. Eq. (5) and Tab. 2) exactly emerge. We speculate that an agent-based model could lead to an understanding of the origin of these numbers. But this would be a subject for future research.

Author contribution statement

Authors contributed equally to this paper.

References

1. T. Chordia, B. Swaminathan, J. Financ. **55**, 913 (2000)
2. A. Joulin, A. Lefevre, D. Grunberg, J.P. Bouchaud, Wilmott Magazine **37**, 1 (2008)
3. R.N. Mantegna, H.E. Stanley, *An Introduction to Econophysics; Correlations and Complexity in Finance* (Cambridge University Press, Cambridge, 2000)
4. R. Cont, Quant. Financ. **1**, 223 (2001)
5. B. Mandelbrot, J. Business **36**, 394 (1963)
6. B. Mandelbrot, *The Fractal Geometry of Nature* (W.H. Freeman and Company, New York, 1983)
7. Y. Liu, P. Gopikrishnan, P. Cizeau, M. Meyer, C.-K. Peng, H.E. Stanley, Phys. Rev. E **60**, 1390 (1999)
8. K. Watanabe, H. Takayasu, M. Takayasu, Physica A **383**, 120 (2007)
9. W. Weibull, J. Appl. Mech. Trans. ASME **18**, 293 (1951)
10. D.R. Cox, *Renewal Theory* (Wiley, New York, 1962)
11. H. Frauenfelder, S.G. Sligar, P.G. Wolynes, Science **254**, 1598 (1991)
12. J.B. Bassingthwaite, L.S. Liebovitch, B.J. West, *Fractal Physiology* (Oxford University Press, New York, 1994)
13. F.W. Scholz, M.A. Stephens, J. Am. Statist. Assoc. **82**, 918 (1987)
14. M.H. Cohen, P. Venkatesh, in *Practical Fruits of Econophysics*, edited by H. Takayasu (Springer, Tokyo, 2005), p. 147
15. R. Klages, in *Reviews of Nonlinear Dynamics and Complexity*, edited by H.G. Schuster (Wiley-VCH, Weinheim, 2010), Vol. 3, p. 169
16. N. Korabel, E. Barkai, Phys. Rev. E **82**, 016209 (2010)
17. J. Gallaher, K. Wodzińska, T. Heimburg, M. Bier, Phys. Rev. E **81**, 061925 (2010)
18. I.N. Lobato, C. Velasco, J. Bus. Econ. Stat. **18**, 410 (2000)
19. P. Gopikrishnan, V. Plerou, X. Gabaix, H.E. Stanley, Phys. Rev. E **62**, R4493 (2000)
20. V. Plerou, P. Gopikrishnan, X. Gabaix, H.E. Stanley, Phys. Rev. E **66**, 027104 (2002)
21. X. Gabaix, P. Gopikrishnan, V. Plerou, H.E. Stanley, Nature **423**, 267 (2003)
22. U. Khan, M.B. Stinchcombe, Am. Econ. Rev. **105**, 1147 (2015)
23. L.S. Liebovitch, J. Fischbarg, J.P. Koniarek, Math. Biosci. **78**, 203 (1986)
24. L.S. Liebovitch, J. Fischbarg, J.P. Koniarek, Math. Biosci. **84**, 37 (1987)
25. G. Bonanno, F. Lillo, R.N. Mantegna, Physica A **280**, 136 (2000)
26. J. Wuttke, Algorithms **5**, 604 (2012)

Open Access This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.