Modelling the short term herding behaviour of stock markets

Yoash Shapira, Yonatan Berman and Eshel Ben-Jacob
School of Physics and Astronomy, The Raymond and Beverly Sackler Faculty of Exact Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel
E-mail: eshelbj@gmail.com

Received 1 November 2013, revised 6 April 2014
Accepted for publication 8 April 2014
Published 21 May 2014

Abstract
Modelling the behaviour of stock markets has been of major interest in the past century. The market can be treated as a network of many investors reacting in accordance to their group behaviour, as manifested by the index and effected by the flow of external information into the system. Here we devise a model that encapsulates the behaviour of stock markets. The model consists of two terms, demonstrating quantitatively the effect of the individual tendency to follow the group and the effect of the individual reaction to the available information. Using the above factors we were able to explain several key features of the stock market: the high correlations between the individual stocks and the index; the Epps effect; the high fluctuating nature of the market, which is similar to real market behaviour. Furthermore, intricate long term phenomena are also described by this model, such as bursts of synchronized average correlation and the dominance of the index as demonstrated through partial correlation.

Keywords: econophysics, market herd behavior, Epps effect, stock market modelling

1. Introduction
One of the central ideas in economics is that investors take rational decisions in accordance to information in the form of news, sometimes referred to as Homo economicus [1, 2]. This has led
to the notion of market efficiency and the hidden assumption that capital markets are dominated by exogenous variables as external pressure and changes in the environment and not by endogenous variables such as human behaviour [3, 4]. However, recent advances in economics and behavioural finance have led to the understanding that in practice, financial markets are not perfectly efficient and the effects of human behaviour on financial markets are gaining more and more attention [3–6].

The importance of human behaviour and its effect on the stock market were recently supported by a novel approach of analysing large scale data regrading financial topics, appearing in the media. These analyses include mood analysis on Twitter [7], search volume analysis on Google [8], article viewing analysis on Wikipedia [9] and even the analysis of financial news [10]. These examples demonstrate the correlation between stock market moves and behavioural trends reflected in various media channels.

A prominent human social behaviour that might have effect on financial decisions made by investors, is the human tendency to belong to a group. A common expression of that behaviour is herding, which is common to many natural systems: ant colonies, fish schooling, bird flocking, animal herding and even bacterial growth [11, 12]. In financial markets this is especially true for analysts and consultants, who trade their customers money. Investors are influenced by the decisions of other investors and most investment managers prefer not to invest contrary to the group [13, 14]. New information is very noisy and not accurate, whereas the group decisions are perceived as better than the individual ones [13–19].

Within this framework, two major features of financial markets are very well known, yet still noteworthy:

- Market prices are highly fluctuating and unpredictable. The short time fluctuations are an order of magnitude larger than the effect of long range drifts.
- Strong collective behaviour in the stock market is identified [13–19]. This collective behaviour is observed in the correlations between the prices of stocks, and in the pronounced correlation between different indexes in different markets [20, 21].

These features are hard to explain as the result of new information. The flow of new information, by itself, is not enough to explain the highly fluctuating behaviour of financial markets [22, 23]. The rate of new substantial information is very low, compared to the market fluctuations rate. In addition, the information is often not clear enough to cause a correlated behaviour.

Following the observations above, a combination of the unpredictable reaction to information and the collective behaviour in markets is necessary when trying to model the market behaviour. Therefore, we propose a simple model that balance between these two characteristics.

The suggested stock market model consists of the coupled dynamics of the market index and several stocks. In the stock market, the index is a good representation of the group activity [24, 25]. Every investor is effected by the actions of other investors, and when making a decision regarding a single transaction, the group behaviour is taken into account. Consequently, the transaction return should be strongly correlated to the change in the index, at the time of the transaction.

In this paper, we present an analysis of the proposed model used to describe the observed short time scale behaviours of financial markets. In section 2, we describe our model and the
method used for its calculation and analysis. In section 3 the lagged autocorrelation of the stock returns with different time lags is analysed and compared to empirical data. The Epps effect apparent in the correlations between the modelled stocks is presented in section 4. Section 5 features the analysis of daily results using the suggested model focusing in the fluctuative nature of the daily returns of the index and the stocks. In section 6 long term behaviour of the model is described. Finally, a discussion of the results and a summary of the presented work are brought on section 7.

2. Method

We devise a new stock market model that produces time series of stock returns and their weighted average, the index return.

Let us start with \( n \) time series, each representing the return of a certain stock (for example, \( n = 30 \) is the number of stocks in the Dow-Jones index). In each time interval the return of each stock is denoted by \( C_{ij} \), where \( i \) and \( j \) are the stock and time indexes respectively. The change of the index, \( IC_j \), for the same interval, is defined as:

\[
\frac{1}{n} \sum_{i=1}^{n} C_{ij}.
\]  

(1)

Given \( C_{ij} \) for all \( i \) and \( IC_j \) it is possible to calculate \( C_{ij+1} \) using the correlation between stock \( i \) and the index and a random term, based on the rationale presented in section 1:

\[
C_{ij+1} = R_i \cdot IC_j + K \cdot \text{sgn} \left( \mathcal{N}(0, 1) \right) \cdot r.
\]  

(2)

\( R_i \) stands for the initial correlation between the change in the index and that specific stock. For simplicity it is assumed that \( R_i \) is constant for all stocks. \( K \) is an arbitrary number (of none real qualitative significance. The parameters \( R \) and \( K \) have a major quantitative significance, though) and \( r \) is a random number, distributed so that it represents the reaction of investors to information and their general behaviour. This was done by producing a random number generating function with a tail that decays according to a power law. This function should be constant up to a certain value and decays like \( x^\alpha \) above that value. Following the procedure presented in [26], the random distribution was chosen to be:

\[
p(x) = \begin{cases} C & x \leq 0.2 \\ A \cdot x^{-2.8} & x > 0.2. \end{cases}
\]  

(3)

After normalization (i.e. \( \int_0^\infty p(x) \, dx = 1 \)) \( A = (1.8/2.8) \cdot 0.2^{1.8} \) and \( C = A \cdot 0.2^{-2.8} \).

The term \( \text{sgn} \left( \mathcal{N}(0, 1) \right) \) was used in equation (2) in order to randomly produce the numbers 1 and \(-1\) in equal probability multiplying \( r \).

Using the described process, \( IC_j \) is calculated and then used to calculate \( C_{ij+1} \) and so on for the different stocks. This way, long sequences of \( C_{ij} \) and \( IC_j \) are produced.

Remembering that each day is composed of many time intervals, the daily change of the index, \( D_x \), is calculated as the sum of some consecutive values of \( IC_j \), where for the next day, \( D_{k+1} \), it is the sum of the next consecutive time intervals and so on. The term ‘daily’ is used to describe that each ‘day’ consists of a specific number of time intervals. The calculated daily
return is dimensionless, as it is synthetically constructed. In order to compare it to real market data, gauging or normalization is essential. This will be discussed in section 5. In addition, since stock trading is continuous, the time steps, on which the stock returns are set, are rather artificial. However, the model update rate can be simply seen as a sampling frequency. This way, gauging will also enable us to overcome this artificiality.

3. Lagged autocorrelation

It is reasonable to expect that such an iterative process as presented in the previous section, will produce ‘autocorrelated’ sequences. An analysis of the correlation of a modelled stock to itself with different time lags was done. Given a time series $S(t)$, describing the stock return in time, the term $\text{corr} \left( S(t), S(t + \tau) \right)$ was calculated (where $\text{corr}$ is defined as: $\text{corr} \left( X, Y \right) = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}$) for different values of $\tau$, measured in transactions (or time intervals). We will refer this term as the lagged autocorrelation of $S(t)$ for time lag $\tau$. Figure 1 presents results obtained for the lagged autocorrelation of an arbitrary stock in a model of 30 stocks, using the parameters $R = 0.8$ and $K = 0.6$, running $10^6$ transactions. The choice of parameters was such that $R$, representing the average correlation between the index and any of the stocks, is similar to a nominal value in real markets ranging usually from 0.5 to 0.9 [27]. Naturally, this value changes in time and is also different between different markets and between different stocks within the same market. However, for simplicity, a constant value is used through the whole analysis presented in this paper. The parameter $K$ was chosen so that $R^2 + K^2 = 1$. As previously stated, the choice of parameters has an insignificant qualitative effect, while a thorough investigation of the quantitative significance of these parameters exceeds the scope of this paper.

The qualitative picture illustrated in figure 1 is consistent with real market data. The correlation decreases exponentially in time up to some point ($\sim$15–20 transactions) at which the

![Figure 1. The lagged autocorrelation of a modelled stock. A semi-log plot of the time-lagged correlation of one stock, chosen arbitrarily out of 30 modelled stocks, as a function of the time lag, using $R = 0.8$ and $K = 0.6$ for $10^6$ transactions. The dashed black line is a linear fit of the results for $1 \leq \tau \leq 20$.](image-url)
correlation essentially disappears (it is purely statistical). These results should be compared to those presented in [21, 28, 29], where the same qualitative results are obtained. In addition, by comparing the results to the empirical data, it is possible to add a time scale to the artificial time intervals used. Since in real market data such correlations exist for time lags of up to a few minutes, we can evaluate that for the chosen set of parameters in the model, each time interval or transaction is equivalent to a few seconds. For example, the actual time interval between consecutive announcements of the Dow-Jones index is 2 s [30], 1 s for the DAX index [31] and 15 s for S&P500 [32] and FTSE100 [33].

4. The Epps effect

The Epps effect is a phenomenon described by the decrease of empirical correlation between the returns of two different stocks with the increase of the sampling frequency of the data [34]. Since the description of the effect [34–37], many efforts were made in order to shed more light on the effect and its sources.

Two major factors leading to the effect have been revealed up to now [37–41]:

- Possible lead-lag effect between stock returns, which can appear mainly between stocks of different capitalization or for some functional dependencies between them.
- Asynchronicity of ticks in case of different stocks.

The analysis of empirical data showed that the explanation of the effect solely by any of the factors above is not satisfactory [37, 38]. Using the suggested model, it can be deduced that the explanation for the Epps effect might be different than previously suggested. The proposed factors for the Epps effect are irrelevant in the model, since no direct correlation or functional dependencies exist between the modelled stocks and the possibility of ticks asynchronicity was not taken into account. The correlation between the stocks in the model is only due to their correlation to the index, and the modelled herd behaviour concept.

In figure 2 (a) results of the average of the correlation between the modelled stocks to one specific arbitrary modelled stock are presented.

The calculation was conducted as follows: a modelled stock market with index and \( n \) stocks was produced for \( 10^6 \) transactions and the parameters used were \( R = 0.8 \) and \( K = 0.6 \). The resulted time series can be numbered from 0 to \( n \), where \( S_i(t) \) is the \( i \)th stock return time series (the stocks were arbitrarily numbered) and \( S_0(t) \) is the modelled index return time series. For a given integer \( m \), it is possible to calculate the average values of sliding windows of width \( m \) (measured in transactions) of \( S_i(t) \), denoted as \( \bar{S}_{i,m}(t) \). Using these calculations, the following term can be derived:

\[
C_{m,1} = \frac{1}{n-1} \sum_{i=2}^{n} corr (\bar{S}_{i,m}(t), \bar{S}_{i,m}(t)).
\]

\( C_{m,1} \) is a measure of the average correlation between the first stock return and the return of the other stocks (stock–stock correlation) for a given sampling frequency. Obviously, increasing the sampling frequency of the data is equivalent to decreasing the sliding window width, \( m \). The results presented in figure 2(a) for \( C_{m,1} \) are clearly consistent with the Epps effect.
The effect is also apparent when taking into account the average correlation between the stock returns and the index return (index–stock correlation), calculated as follows and presented in figure 2(a):

\[
C_{m,0} = \frac{1}{n} \sum_{i=1}^{n} corr \left( \bar{S}_{0,m}(t), \bar{S}_{i,m}(t) \right).
\]  

(5)
Also calculated was the average correlation between the index and the stocks return for increasing sampling frequency of the data (without averaging over sliding windows of changing widths), as presented in figure 2(b).

From figure 2 it is possible to observe that although the $R$ parameter in the model was 0.8, the average correlation was 0.3, when looking at each transaction separately (large sampling frequency). As the sampling frequency decreases, the average correlation goes closer to 0.8. These results, in which the correlation between pairs of stocks is lower than the correlation between stocks and the index (see figure 2(a)), are consistent with actual observations [24].

Identifying the Epps effect in the model results suggests that the underlying assumptions of the model presented previously might be additional factors to the Epps effect. Thus, it demands further work to be carried out in order to test this conclusion thoroughly.

5. Daily results

The fluctuative nature of the market is a key feature of the market one would expect the suggested model to possess. In order to identify this feature in the model, the distribution of the market return should be analysed.

First, the index return time series ($IC_i$) is used to produce the ‘daily’ return time series of the index. Suppose each day is defined to be consisted of $d$ transactions (in our case $d = 780$ transactions/day is used, as transaction every 30 s is assumed), it is possible to define the daily return of the index simply by summing over each ‘day’. This time series is defined as follows:

$$D_k = IC_{d(k-1)+1} + IC_{d(k-1)+2} + ... + IC_{d,k},$$  \hspace{1cm} (6)

where $k$ is the day number.

The modelled index and stocks return series are synthetically created and therefore have no natural scale and can be seen as dimensionless. $D$, therefore, also lacks a scale and should be normalized in order to be quantitatively compared to any data representing real markets. The choice of normalization was such that the standard deviation of the normalized modelled daily change time series will match the typical standard deviation of daily change time series of leading indexes. It might be seen as a method of adding units to the modelled daily return, as it is naturally dimensionless, while the real index daily change time series are measured in percent.

This normalization was done the following way: the mean value of $D$ ($\mu_D$) that is relatively close to 0 and its standard deviation ($\sigma_D$) were calculated. In addition, the daily change time series of each major index taken into account (Dow-Jones, S&P500, NASDAQ, DAX, FTSE100, CAC40, NIKKEI225, Hang Seng, Shanghai stock exchange and Bombay stock exchange) was calculated as well. This was done by taking $T_k' = 100 \cdot (I_{k+1} - I_k)/I_k$ for the daily change in day $k$ of the index $I$, where $I_k$ is the closing price of $I$ in day $k$ measured in points. Each of these time series has a mean value (relatively close to 0) and a standard deviation. When averaging over all ten indexes taken into account the average standard deviation was $\bar{\sigma} \approx 1.5\%$. Therefore, the normalized modelled daily change time series was calculated simply by taking $\hat{D} = \frac{\hat{\sigma}}{\sigma_D} \cdot D$, so that $\sigma_{\hat{D}} = \hat{\sigma}$. Due to the additivity of the daily change (see equation (6)), this implicates that the actual index and stocks returns can be simply
normalized by taking $\hat{C}_{ij} = \frac{\sigma_i}{\sigma_j} \cdot C_{ij}$, where $i$ and $j$ are the stock and time (or transaction) indexes respectively.

Figure 3 presents a typical daily change plot in time. The results were obtained for a modelled market with 30 stocks, where $R = 0.8$ and $K = 0.6$ and for 780 transactions a day, reflecting the number of 30 s time intervals in a 6.5 h trading day. The presented results should be compared to those presented in figure 1 in [26].

Figure 3 shows typical results of the daily return time series for a modelled index. These results resemble results obtained in practice for actual indexes.

Yet another measure of how fluctuative is a modelled index, which can be used to test its resemblance to actual indexes, is the generation of a probability distribution function (pdf) of the daily change (the absolute value of the daily change was taken into account, as the distribution is approximately symmetric) and its comparison to real data. Such a comparison to the daily change probability distribution function of some of the leading indexes was made, and is presented in figure 4. These results were obtained, as previously presented, for a modelled market with 30 stocks, where $R = 0.8$ and $K = 0.6$ and for 780 transactions a day, this time taking account of 70 000 days in order to improve the statistical significance of the results. The probability distribution function of the modelled index taken into account is the normalized time series $\hat{D}$.

Both qualitatively and quantitatively, a similarity between the modelled index and the actual data is obtained. This resemblance is a good measure for similar fluctuative nature, and the modelled index is therefore close in these characteristics to actual indexes. Also depicted in figure 4 is that the distribution of the daily return both in real markets and in the modelled index approximately follows a Lévy stable distribution [42] (see figure 4(a)), which is known for characterizing many processes in economics and finance in general [43, 44]. While the model provide no explanation for this specific functional dependence of the distribution, the assumptions on which it stands are enough in order to achieve this dependence. Furthermore, the quantitative similarity demonstrates that the nominal values of $R$ and $K$ used, as well as the value of the power law exponent, make sense.
Figure 4. The probability distribution function of the absolute daily change. (a) The probability distribution function of the absolute daily change of leading indexes—Dow-Jones (blue), S&P500 (green), NASDAQ (red) and FTSE100 (cyan). Also presented is the probability distribution function of a Lévy stable distribution [42] (dashed black curve) with the parameters $\alpha = 1.6$, $\beta = 0$, $\gamma = 0.55$ and $\mu = 0$; (b) the probability distribution function of the absolute daily change for the modelled index with 30 stocks, where $R = 0.8$, $K = 0.6$ and $d = 780$ transactions/day, taking account of 70,000 days (magenta curve) and the probability distribution function of a Lévy stable distribution (dashed black curve, the parameters taken were the same as presented in (a)). The light green curve shows the results for the modelled index with 30 stocks where $R = 0.8$, $K = 0.6$ and $d = 780$ transactions/day, taking account of 70,000 days (the same parameters as the magenta curve), where the modelled index was calculated for a normally distributed random number and not power law distributed, meaning equation (2) was taken as $C_{i,j+1} = R_i \cdot I C_j + K \cdot r$, where $r \sim N(0, 1)$. 
While figure 4(b) demonstrates the high similarity between the model results and real market data, it is clearly seen that as expected, a different choice for the random distribution describing the non-correlated term in the model (see section 2), such as a normal distribution, will fail in describing both qualitatively and quantitatively the fluctuation nature of the market. The normal distribution fails to capture extreme changes, as the distribution tail is short, leading to the rapid decay of the daily change pdf depicted in figure 4(b) (the light green curve). In that sense, the results presented in figure 4 demonstrate that the choice in a power law distribution function in equation (2) makes sense.

Also to be noted is that regardless of the observed similarities between the volatility distribution of the model results and of actual indexes, no temporal order can be obtained using the model, in the sense that exists in indexes [26]. This is, of course, due to the purely randomized method of constructing the modelled index.

6. Long term behaviour

The results presented in the previous sections show how different known behaviours of stock markets can be obtained using the suggested simple model. This success shows that the herding perspective of the driving forces of stock markets is significant and relevant. However, while relatively short term characteristics were obtained, no long term characteristics were analysed using the suggested model in the previous sections.

6.1. Partial correlations

Another market characteristic that be obtained using the model is the high partial correlations with respect to some stock comparing to the relatively low partial correlations with respect to the index [24]. In order to test this behaviour, a stock market containing 30 stocks was modelled, using the parameters $R = 0.8$ and $K = 0.6$. Daily returns were obtained using 780 transactions a day. Given a sliding window of 22 days, the partial correlation of each pair of stocks was calculated, with respect to the index and with respect to two given stocks (arbitrarily chosen) for each window, for consecutive 2000 windows. The results of these calculations are presented in figure 5.

As expected, a significant difference was obtained between the average partial correlation of the stocks with respect to an arbitrary stock comparing to the average partial correlation with respect to the index [26]. In accordance with the previous results, the suggested model is found successful in qualitatively and even quantitatively describing known behaviours of the stock market. The results presented in figure 5 of the average partial correlation with respect to the index are very similar to those obtained from real data [45]. However, the small difference that still exists between the modelled and the empirical results might be attributed to the effect of external events that was not taken into account in the model.

6.2. Synchronized correlation

In the previous sections we used the suggested model in order to show that using the underlying assumptions described in section 1 known behaviours of stock markets can be achieved.

We are also able to demonstrate more complex behaviours of the market using our model. One of these behaviours is the synchronized high correlation between stocks or correlation...
bursts appearing in major markets [25]. This feature is achievable through random calculation using the suggested model. This is done by analysing the daily return of \( n \) stocks market (we used \( n = 30 \)), taking into account 780 transactions a day, using \( R = 0.8 \) and \( K = 0.6 \). For each stock \( i \), we calculated the correlation between its daily return and any other stock \( j \) daily return for a sliding window of 22 days, thus producing a time series of correlations, denoted as \( X_{ij} \).

Following this calculation it is possible to obtain the average correlations time series of \( i \):

\[
X_i = \frac{1}{n-1} \sum_{j \neq i} X_{ij}.
\]

Typical results for \( X_i \) for different stocks in 2000 days are presented in figure 6 (a). The most prominent behaviour depicted in figure 6(a) is the ‘synchronized’ spikes in the averaged correlation for all stocks, meaning that all stocks show sudden extreme correlation ‘bursts’ to all other stocks. This result is a known phenomenon [25, 45, 46], and the results presented in figure 6(a) should be compared to those presented in figure 3 in [45], which demonstrate the same behaviour for real leading indexes. Apart from presenting this well known behaviour of the market using the suggested model, it also demonstrates that this result is a partly statistical feature, and not the result of an ordered behaviour of the market as might be hypothesized. As previously stated, no feature or behaviour of the market that involves some sort of temporal order be the result of the suggested model due to its randomized nature.

A related behaviour is obtained when calculating the average partial correlation using the index as a mediating variable between pairs of stocks (simply replace the correlation between \( i \) and \( j \) with partial correlation between \( i, j \) and the index) and will be denoted as \( \tilde{X}_i \). These results are presented in figure 6(b). One can easily see that figures 6(a) and (b) complete together the empirical results depicted in figure 1 on [25]. The average correlations and the average partial correlations can be divided for the calculation of the index cohesive force (ICF) [24, 25].

![Figure 5. The average partial correlation between pairs of modelled stocks. The calculation was performed with sliding window of 22 days, with 780 transactions a day, 30 stocks, \( R = 0.8 \) and \( K = 0.6 \). The blue curve was obtained for the modelled index as a mediating variable and the red and green curves were obtained by taking two arbitrary stocks as the mediating variables. Comparison should be made to figure 6 in [26].](image-url)
ICF is a quantitative measure of the cohesive effect the index has on the dynamics of the stock correlations. This refers to the observed effect the index has on stock correlations, where large changes of the index result in higher index–stock correlations [25]. The value of the ICF is the quotient of the averaged average correlations and the averaged average partial correlations, or \( \frac{\sum \bar{X}_i}{\sum \bar{X}_i} \). The model results of the ICF are presented in figures 6(c) and (d).

Figure 6. Bursts of synchronized correlation in the modelled index. (a) \( X_i \) for all \( n = 30 \) stocks, calculated using a sliding window of 22 days, 780 transactions per day, \( R = 0.8 \) and \( K = 0.6 \). Each horizontal line represents the average correlation of one stock, where the left y-axis displays the number of the stock. The black curve is the averaged average correlation for all stocks (\( \frac{1}{n} \sum_{i=1}^{n} X_i \)); (b) \( \bar{X}_i \) for all \( n = 30 \) stocks, calculated using a sliding window of 22 days, 780 transactions per day, \( R = 0.8 \) and \( K = 0.6 \). Each horizontal line represents the average partial correlation of one stock, where the left y-axis displays the number of the stock. The black curve is the averaged average partial correlation for all stocks (\( \frac{1}{n} \sum_{i=1}^{n} \bar{X}_i \)); (c) the index cohesive force, calculated using a sliding window of 22 days, 780 transactions per day, \( R = 0.8 \) and \( K = 0.6 \). The result is the quotient of the averaged average correlation and the averaged average partial correlation with respect to the index; (d) the index cohesive force dependence on the averaged average partial correlation (as calculated in (b) and (c), respectively).
As previously mentioned, one can easily see that the model results illustrated in figure 6 are very similar to such obtained empirically for different markets [24, 25, 45, 46]. However, while the results featured in the previous sections show similarities between the model results and known empirical results for short term market phenomena, these results capture long term behaviours, while the underlying assumptions are exactly the same. This may lead to the idea that the index effect on the market is a actually a relatively long lasting effect.

7. Summary and discussion

As the 2008 financial crisis brought more attention to the validity of the current economic theories, it should be taken into account that social tendencies of investors might have effect on their financial decisions. Here we devised a new simple model to quantitatively describe some of the short and long term behaviours of real stock markets taking into account this effect.

The model is consisted of numerically constructing time series for the stock market index return and the stock returns. The time series building process involves two terms: the first is a random one, which represents the uncoordinated influence of many economic variables. The complexity of those variables makes them unpredictable and thus expressed in the process by a random number power-law distributed. The second term is a correlation coefficient multiplied by the index return at the time of the transaction, which represents the collective group behaviour of investors.

The use of constant correlation values between the index and the stocks is a simplifying assumption. The probability density function of stock returns is not stationary and correlations between stocks and between stocks and the index change over time. There are findings of an increase of the mean correlation between stocks with increasing absolute returns over time [47]. However, this dependence was not taken into account in the presented model, as it was kept simplified, at this point, based on a few assumptions and a small number of free parameters. The dependence of correlations between stocks and between the index and the stocks on time and return will be treated in future work. In order to take account of this dependence, it is first essential to analyse the dependence of the mean correlation between stocks, used in [47] and the correlation between index return and a specific stock transaction, used in the presented model.

The model was able to capture several key short time features of the market: the lagged autocorrelation in the return for short times (see section 3); the Epps effect of the change in correlations when enlarging the window size (see section 4); the transition to daily time scale and the fluctuative nature of the daily return (see section 5). In addition, more complex behaviours related to longer time scales were also described by the model, such as the averaged partial correlations of stocks with respect to other stocks and the index, and the synchronized high correlation bursts between stocks (see section 6).

These results demonstrate that some aspects of the human behaviour complexity reflected in the stock market can be described using models such as the suggested one. Moreover, it is the herding notion set in the core of the model that allows for these results. The short term lagged autocorrelation, for example, that can be found in stocks and is described by the model cannot be explained through flow or aggregation of information (due to the short time lag between consecutive transactions), but by some kind of herd behaviour. The term herding is associated with emotional behaviour, where the herd follows some leader blindly, without rational
criticism and can lead to extreme catastrophes. This notion is in accordance with actual observations that relate many catastrophes in financial markets to endogenous processes in addition to rational processing of new information [48–53]. In addition, the Epps effect, extensively researched and thought of as originated in different factors related to technical aspects of trading and to inherent lagged correlation between stocks, is shown as possibly originated in investors herd behaviour.

The model describes relatively well short term behaviours, but only some long term behaviours can be described. Other behaviours, such as those involving temporal order [26], are naturally impossible to be achieved using this simple model. Nevertheless, following the model results, it can be deduced that such behaviours, as temporal order or other similar long term effects, cannot be explained solely through the postulated assumptions of the suggested model. More specifically, the correlation between stock and index returns, and the fluctuative nature of the return are not sufficient while trying to capture more complex behaviours. It is however demonstrated that any long range analysis, should take into account that short range behaviour can have a long range influence. The suggested model demonstrates such an influence by the qualitative description of the average correlation bursts observed in real markets (see section 6.2). In that sense, this feature can also be attributed to the market herd behaviour. However, these correlation bursts, not being a short term phenomenon, may broaden the sense in which herd behaviour is perceived. While for the short term behaviours, the index can be thought of as reflecting a ‘safe choice’ for investors, and therefore is responsible for the stocks lagged autocorrelation, such an explanation cannot be given for the long term behaviours.

Establishing the above, more research should be carried out: the introduction of different weights and correlations for different stocks; analysing the significance of parameter choice; testing the model reaction to perturbation in one or more stocks and in the index; taking into account the existence of clusters of stocks; taking into account information that has a specific and direct effect on the market; calculation of the long term influence of the short term behaviour, analysed in this work and so on.

Acknowledgments

We thank Dror Y Kenett for fruitful discussions and for his comments regarding the results. This research has been supported in part by the Tauber Family Foundation and the Maguy-Glass Chair in Physics of Complex System at the Tel-Aviv University.

References

[1] Henrich J, Boyd R, Bowles S, Camerer C, Fehr E, Gintis H and McElreath R 2001 Am. Econ. Rev. 91 73
[2] Thaler R H 2000 J. Econ. Perspect. 14 133
[3] Kahneman D and Tversky A 1979 Econometrika 47 263
[4] Bondt W F and Thaler R H 1987 J. Finance 42 557
[5] Fama E F 1998 J. Financ. Econ. 49 283
[6] Kumar A 2009 J. Finance 64 1889
[7] Bollen J, Mao H and Zeng X 2011 J. Comput. Sci. 2 1
[8] Preis T, Moat H S and Stanley H E 2013 Sci. Rep. 3 1684
[9] Moat H S, Curme C, Avakian A, Kenett D Y, Stanely H E and Preis T 2013 Sci. Rep. 3 1801
[10] Alanyali M, Moat H S and Preis T 2013 Sci. Rep. 3 3578
[11] Yang X S 2009 Stochastic Algorithms: Foundations and Applications (Berlin: Springer) p169
[12] Schultz D, Lu M, Stavropoulos T and Ben-Jacob E 2013 Sci. Rep. 3 3578
[13] Gopikrishnan P, Rosenow B, Plerou V and Stanley H E 2001 Phys. Rev. E 64 035106
[14] Harrison J M and Kreps D M 1978 Q. J. Econ. 92 323
[15] Pan R K and Sinha S 2007 Phys. Rev. E 76 046116
[16] Harmon D, de Aguiar M, Chinellato D, Braha D, Epstein I and Bar-Yam Y 2011 Predicting Economic Market Crises Using Measures of Collective Panic
[17] Forbes K and Rigobon R 2002 J. Finance 57 2223
[18] Beine M and Candelon B 2011 Quant. Finance 11 299
[19] Allez R and Bouchaud J P 2011 New J. Phys. 13 025010
[20] Goswami B, Ambika G, Marwan N and Kurths J 2012 Physica A 391 4364
[21] Song D M, Tumminello M, Zhou W X and Mantegna R N 2011 Phys. Rev. E 84 026108
[22] Nakamura T and Small M 2007 Physica A 383 96
[23] Gopikrishnan P, Plerou V, Gabaix X, Amaral L and Stanley H E 2001 Physica A. 299 137
[24] Shapira Y, Kenett D Y and Ben-Jacob E 2009 Eur. Phys. J. B 72 657
[25] Kenett D Y, Shapira Y, Madi A, Bransburg-Zabary S, Gur-Gershgoren G and Ben-Jacob E 2011 PLoS One 6 e19378
[26] Shapira Y, Kenett D Y, Raviv O and Ben-Jacob E 2011 AIP Adv. 1 022127
[27] Kenett D Y, Shapira Y, Gur-Gershgoren G and Ben-Jacob E 2011 J. Eng. Sci. Technol. Rev. 4 218 (www.jestr.org/downloads/Issue3SI/SI3.pdf)
[28] Mantegna R N and Stanley H E 2000 Introduction to Econophysics: Correlations and Complexity in Finance (Cambridge: Cambridge University Press)
[29] Sewell M 2011 UCL RN/11/01 www.cs.ucl.ac.uk/fileadmin/UCL-CS/images/Research_Student_Information/RN_11_01.pdf
[30] Dow Jones Industrial Average Fact Sheet 2013 http://www.djindexes.com/mdsidx/downloads/fact_info/Dow_Jones_Industrial_Average_Fact_Sheet.pdf
[31] DAX Fact Sheet 2013 http://www.dax-indices.com/EN/MediaLibrary/Document/Factsheet_DAX_en.pdf
[32] S&P500 Fact Sheet 2012 http://us.spindices.com/documents/additional-material/sp-500-trifold.pdf
[33] FTSE100 Fact Sheet 2011 http://www.ftse.com/Indices/UK_Indices/Downloads/FTSE_100_Index_Fact-sheet.pdf
[34] Epps T W 1979 J. Am. Stat. Assoc. 74 291
[35] Reno R 2003 Int. J. Theor. Phys. Appl. Finance 6 87
[36] Tóth B and Kertész J 2007 Physica A 383 54
[37] Tóth B and Kertész J 2009 Quant. Finance 9 793
[38] Zhang L 2011 J. Econ. 160 33
[39] Lundin M C, Dacorogna M M and Müller U A 1998 Correlation of high frequency financial time series Financial Markets Tick by Tick: Insights in Financial Markets Microstructure ed P Lequeux(New York: Wiley)
[40] Lo A W and Craig MacKinlay A 1990 J. Econ. 45 181
[41] Tóth B and Kertész J 2009 Physica A 388 1696
[42] Nolan J P 2003 Stable Distributions: Models for Heavy-Tailed Data (Boston, MA: Birkhauser)
[43] Mandelbrot B 1960 Int. Econ. Rev. 1 79
[44] Mantegna R N and Stanley H E 1995 Nature 376 46
[45] Kenett D Y, Raddant M, Lux T and Ben-Jacob E 2012 PLoS One 7 e31144
[46] Kenett D Y, Raddant M, Zatlavi L, Lux T and Ben-Jacob E 2012 Int. J. Mod. Phys.: Conf. Series 16 13
[47] Preis T, Kenett D Y, Stanley H E, Helbing D and Ben-Jacob E 2012 Sci. Rep. 2 752
[48] Kenett D Y, Ben-Jacob E, Stanely H E and Gur-Gershgoren G 2013 Sci. Rep. 3 2110
[49] Smith V L, Suchanek G L and Williams A W 1988 Econometrica 56 1119
[50] Johansen A and Sornette D 2002 arXiv:cond-mat/0210509
[51] Johansen A and Sornette D 2010 Brussels Economic Review (Cahiers Economiques de Bruxelles) 53 201
[52] Cont R and Wagalah L 2013 Math. Finance 23 718
[53] Filimonov V and Sornette D 2012 Phys. Rev. E 85 056108