Specialization and herding behavior of trading firms in a financial market

To cite this article: Fabrizio Lillo *et al* 2008 *New J. Phys.* **10** 043019

View the article online for updates and enhancements.

You may also like

New Journal of Physics

Specialization and herding behavior of trading firms in a financial market

Fabrizio Lillo[1,2,3]**, Esteban Moro**[4]**, Gabriella Vaglica**[1,3] **and Rosario N Mantegna**[1,3]

[1] Dipartimento di Fisica e Tecnologie Relative, Università di Palermo, Viale delle Scienze, I-90128 Palermo, Italy
[2] Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA
[3] CNR-INFM, Unità Operativa di Roma, Centro di Ricerca e Sviluppo SOFT, Roma, Italy
[4] Grupo Interdisciplinar de Sistemas Complejos (GISC), Departamento de Matemáticas, Universidad Carlos III de Madrid, Avenida de la Universidad 30, E-28911 Leganés, Spain

Abstract. Agent-based models of financial markets usually make assumptions about agent's preferred stylized strategies. Empirical validations of these assumptions have not been performed so far on a full-market scale. Here we present a comprehensive study of the resulting strategies followed by the firms which are members of the Spanish Stock Exchange. We are able to show that they can be characterized by a resulting strategy and classified in three well-defined groups of firms. Firms of the first group have a change of inventory of the traded stock which is positively correlated with the synchronous stock return whereas firms of the second group show a negative correlation. Firms of the third group have an inventory variation uncorrelated with stock return. Firms tend to stay in the same group over the years indicating a long term specialization in the strategies controlling their inventory variation. We detect a clear asymmetry in the Granger causality between inventory variation of firms and stock return. We also detect herding in the buying and selling activity of firms. The herding properties of the two groups are markedly different and consistently observed over a four-year period of trading. Firms of the second group herd much more frequently than the ones of the first group. Our results can be used as an empirical basis for agent-based models of financial markets.

IOP Institute of Physics ⟨Φ⟩ DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

Contents

1. Introduction

The modeling of complex systems [1, 2] benefits from the study of agent-based models. A particularly interesting complex system is that of financial markets. Despite many agent-based models of financial markets having been investigated [3]–[10], only in a few cases [11]–[17], has an empirical investigation of agent strategies been possible due to the lack of accessible data. There are institutional and individual investors taking investment decisions and their decisions are in most cases executed by financial and brokerage firms which are allowed to trade in a specific market. Recent studies have empirically shown that the dynamics of institutional and individual trading can show detectable statistical regularities in investment decisions down to a daily or intradaily time horizon [16]. Proprietary trading data obtained from NYSE [13], the Korean Stock Exchange [14] and from NASDAQ [16] have shown that stock returns have some ability to forecast inventory variation of groups of investors whereas the evidence of return predictability on the basis of investor inventory variation is negligible both at a daily and intradaily time horizon.

In the present study, we empirically investigate the presence of detectable resulting investment strategies of the firms entitled to trade in the Spanish Stock Market. In this market, firms are local and foreign credit entities and investment firms which are members of the stock exchange and are the only firms entitled to trade. Approximately, 75% of them are major financial institutions and 25% are established securities dealers. Both of these groups may trade on their own behalf and also on behalf of other individuals and/or institutions which are not members of the market. We empirically show that, although a firm may act on behalf of many individuals and institutions having different strategies, firms self-organize in groups to the extent that in most cases it is possible to characterize a firm with a specific resulting strategy. Our results are the first empirical detection of the existence of a resulting strategy and herding of firms performed over an entire market and are consistently observed over a four-year time period. These results can be used as an empirical basis for agent-based models of financial markets.

## 2. Specialization

The activity of a market participant with respect to a given asset is well represented by the inventory variation which is the value exchanged as a buyer minus the value exchanged as a seller in a given time interval. In this paper, we investigate the inventory variation of financial firms exchanging a financial asset at the Spanish Stock Exchange during the years 2001 through

to 2004. In 2004 this market was eighth in the world in market capitalization. Our database corresponds to the electronic open market, Sistema de Interconexión Bursátil Electrónico (SIBE), and allows us to follow each transaction performed by all the registered firms. We focus our investigation on Telefónica (TEF), Banco Bilbao Vizcaya Argentaria (BBVA), Banco Santander Central Hispano (SAN) and Repsol (REP) stocks, which are four highly capitalized stocks and on the most active firms which have traded at least 200 trading days with at least 1000 transactions per year during the period 2001–2004. We investigate the market dynamics by focusing on the trading of each selected stock separately for each available calendar year. By doing so we have 4×4 distinct sets of results. The number of active firms is around 70 with a minimum and a maximum value equal to 54 and 82, respectively. The homogeneity of obtained results for these sets provides us with an indication about the general validity of them.

We first consider daily inventory variation of the investigated stock. Let $v_i(t)$ indicate the inventory variation of firm i during the day t. We first investigate the statistical properties of this multivariate set by considering its correlation coefficient matrix. This matrix has both positive and negative statistically significant correlation coefficients $\rho[v_i(t), v_j(t)]$. To estimate if the detected correlations are carrying information about the market dynamics we perform a principal component analysis in which the factor selection is done with the help of methods based on the random matrix theory [18]–[20]. Particular attention has to be paid to spurious correlation that might be due to the buy–sell counterparts present in each transaction. By performing numerical experiments where we shuffle independently the buyers and the sellers, in such a way to maintain the same number of purchases and sales for each firm as in the real data, we have verified that the first eigenvalue is not consistent with the null hypothesis of random trading among firms (see figure 1) and is therefore carrying information about the collective dynamics of firms. The same cannot be said for the second eigenvalue and therefore we limit our investigation to the first one.

To elucidate the nature of this information we investigate the time profile of the factor associated with the first eigenvalue: we find that there is a statistically significant correlation of this factor with the price return time series (see figures 2(a) and (b)). The factor price return correlation ranges from 0.47 to 0.74. This empirical evidence leads us to hypothesize that the dynamics of the inventory variation can be described as a first approximation by the linear relation $v_i(t) = \gamma_i r(t) + \varepsilon_i(t)$ where γ_i is proportional to the correlation between price return and inventory variation $\rho[v_i(t), r(t)]$ and $\varepsilon_i(t)$ is a zero mean white noise term describing the idiosyncratic part of the strategy of the firm.

We categorize the firms characterized by a significantly positive value of $\rho[v_i(t), r(t)]$ as trending firms and those characterized by a significantly negative value of $\rho[v_i(t), r(t)]$ as reversing firms. We address the remaining firms as the uncategorized ones. If inventory variations and returns were independently identically distributed Gaussian variables, we could use a significant threshold such as, for example, $\pm 2\sigma = \pm 2/\sqrt{N_T}$, where N_T is the number of time records for each time series. These assumptions are certainly not realistic for real market data and therefore we verify the robustness of this choice by comparing the experimental results with the results of a null hypothesis based on a block bootstrap of both $r(t)$ and $v_i(t)$. Specifically, for each firm i and for each of the 16 investigated data sets we have performed 10 000 block bootstrap replicas of our $r(t)$ and $v_i(t)$ time series by using a block length of 20 trading days. For each firm we have checked whether the estimated correlation with return exceeds the 0.97725 quantile or is smaller than the 0.02275 quantile (corresponding to 2σ for Gaussian variables) of the correlation distribution obtained from the bootstrap replicas.
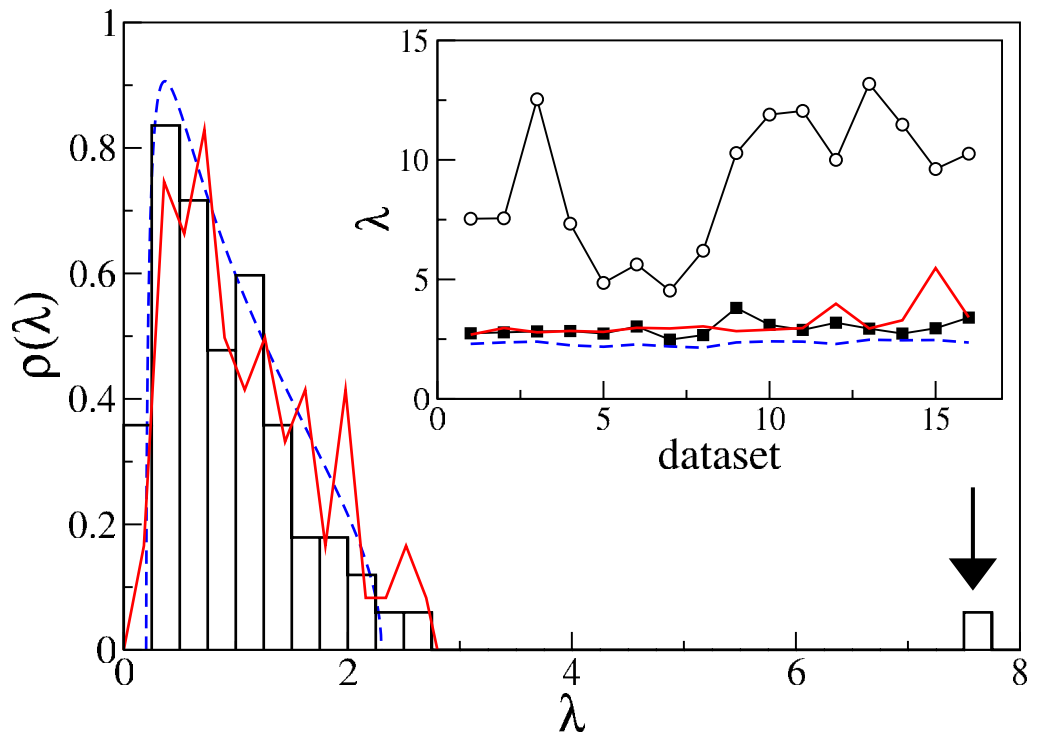
Figure 1. Histogram (rectangles) of the eigenvalue spectrum of the correlation matrix of inventory variation of firms trading the stock BBVA in 2001. The black arrow indicates the first eigenvalue. The blue dashed line is the spectral density expected by the random matrix theory [18] where each time series is replaced by a uncorrelated random inventory time series. The solid red line is the averaged spectral density obtained by shuffling independently the buyers and the sellers, in such a way to maintain the same number of purchases and sales for each firm as in the real data. Other shuffling experiments give similar results. In the inset we show the first (empty circles) and second (filled squares) eigenvalues of the 4×4 investigated sets. The dashed blue line again indicates the threshold expected by the random matrix theory, and the solid red line is the upper threshold expected by the shuffling experiment. The first eigenvalue is well above the thresholds obtained with the random matrix theory and shuffling methods for all the investigated sets.

In figure 2(c) red (green) circles are firms whose inventory variation is significantly positively (negatively) correlated with price return according to our bootstrap test. Black circles indicate firms not significantly correlated with returns. In the same figure we also draw the lines $\pm 2\sigma = \pm 2/\sqrt{N_T}$ thresholds. The figure shows that the categorization in three groups of firms selected according to the $\pm 2\sigma$ thresholds is validated by the block bootstrap procedure for the largest majority of firms with only a very limited number of exceptions. This result also holds for smaller block sizes. Having verified with a block bootstrap procedure that a $\pm 2\sigma$ threshold is quite a reliable indicator of the quantile associated with the detected $\rho[v_i(t), r(t)]$ values, for the sake of simplicity, in the rest of this paper we will use a categorizing procedure based on the $\pm 2\sigma$ threshold.
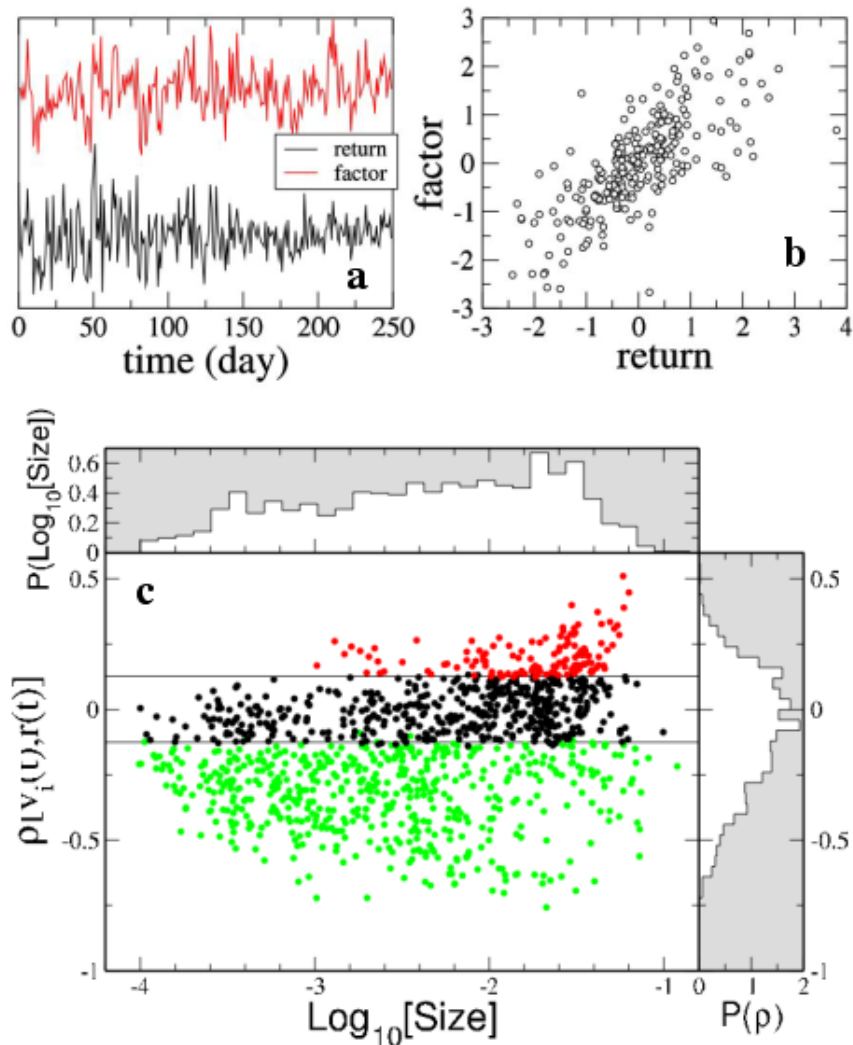
Figure 2. Panel (a) shows the time evolution of the first factor (red line) of the correlation coefficient matrix of daily inventory variation of firms trading the stock BBVA in 2003 and the daily stock return of the same stock (black line). Panel (b) shows the scatter plot of these quantities. A high degree of correlation (in the figure equals to 0.72) is observed. Panel (c) shows the scatter plot of $\rho[v_i(t), r(t)]$ versus a proxy of the size of the firm. For each stock and each year this proxy is the ratio between the value exchanged by the firm and twice the total value transacted in the market. Each of the 1115 circles refers to a firm trading a specific stock in a specific year. Red (green) circles refer to firms whose inventory variation is positively (negatively) significantly correlated with returns according to the block bootstrap analysis whereas black circles refer to firms whose inventory variation is not significantly correlated with returns. The horizontal lines indicate the 2σ threshold. In the side panels of the figure we show the marginal probability density function of the correlation coefficient $\rho[v_i(t), r(t)]$ and of the firm size.

Table 1. Number of active firms trading the Telefónica stock belonging to the groups of reversing, uncategorized and trending firms for the calendar years of the period 2001–2004.

	2001	2002	2003	2004
Reversing	43	39	42	37
Uncategorized	28	31	31	29
Trending	11	10	8	6
Total	82	80	81	72

Table 1 indicates that about 50% of the firms are reversing whereas firms with trending strategy are observed in approximately 10% of the cases. The rest remain uncategorized. Finally, figure 2(c) also indicates a significant correlation between the strategy and size of the firm: specifically, we find that the Spanish market is composed of a few mostly large trending firms and many reversing firms with a very heterogeneous size. As table 1 indicates, the percent categorization over the four years is rather stable but what is the probability that a firm categorized in a given group will remain in the same group or will move to another group next year? We have computed the probability $P(Y|X)$ of a firm being in group X in a given year and moving to group Y during the next year. We have averaged these probabilities over the three changes of year present in our database. For the group of reversing firms ($X = R$), these probabilities are $P(R|R) = 71\%$, $P(U|R) = 16\%$, $P(T|R) = 2\%$ and $P(E|R) = 11\%$, where T indicates trending firms, U uncategorized ones and E indicates that the firm has exited from the set of active firms. For the trending group we analogously obtain $P(R|T) = 3\%$, $P(U|T) = 35\%$, $P(T|T) = 44\%$ and $P(E|T) = 18\%$ whereas for the uncategorized firms we estimate $P(R|U) = 19\%$, $P(U|U) = 62\%$, $P(T|U) = 7\%$ and $P(E|U) = 12\%$. These probabilities show that a firm usually tends to preferentially stay in the same group over the years indicating a long term specialization. This behavior is more pronounced for reversing firms ($P(R|R) = 71\%$) rather than for trending firms ($P(T|T) = 44\%$). Uncategorized firms show an intermediate behavior. The probability to move from reversing to trending firms, or *vice versa*, is rather low. These results, obtained for Telefónica, are representative of the other stocks of our sample.

In summary, our results are consistent with the hypothesis that the price return of the traded stock acts as the major common factor for all the firms. This one-factor model also predicts that the cross-correlation between the inventory variation of two firms is proportional to $\rho[v_i(t), v_j(t)] \simeq \gamma_i \gamma_j$ and thus the correlation of inventory variation of two firms is positive when both firms are either trending or reversing while the cross-correlation of inventory variation of a trending and a reversing firm is negative. In order to illustrate how well the model reproduces the empirical data, we show in figure 3 the contour plot of the correlation matrix of daily inventory variation plotted by sorting the firms into rows and columns according to their value of correlation γ_i. The approximately blocked structure of the matrix indicates that the proposed model gives a good basic description of the data. Moreover it can be shown [21], that the correlation matrix of the model is composed by a large eigenvalue $\lambda_1 \sim \sum \gamma_i^2$ and $N - 1$ small eigenvalues, similar to what is seen for empirical data. These properties are indeed statistically detected in real data consistent with our observation that the price variation is an important common factor characterizing the dynamics of the inventory return.
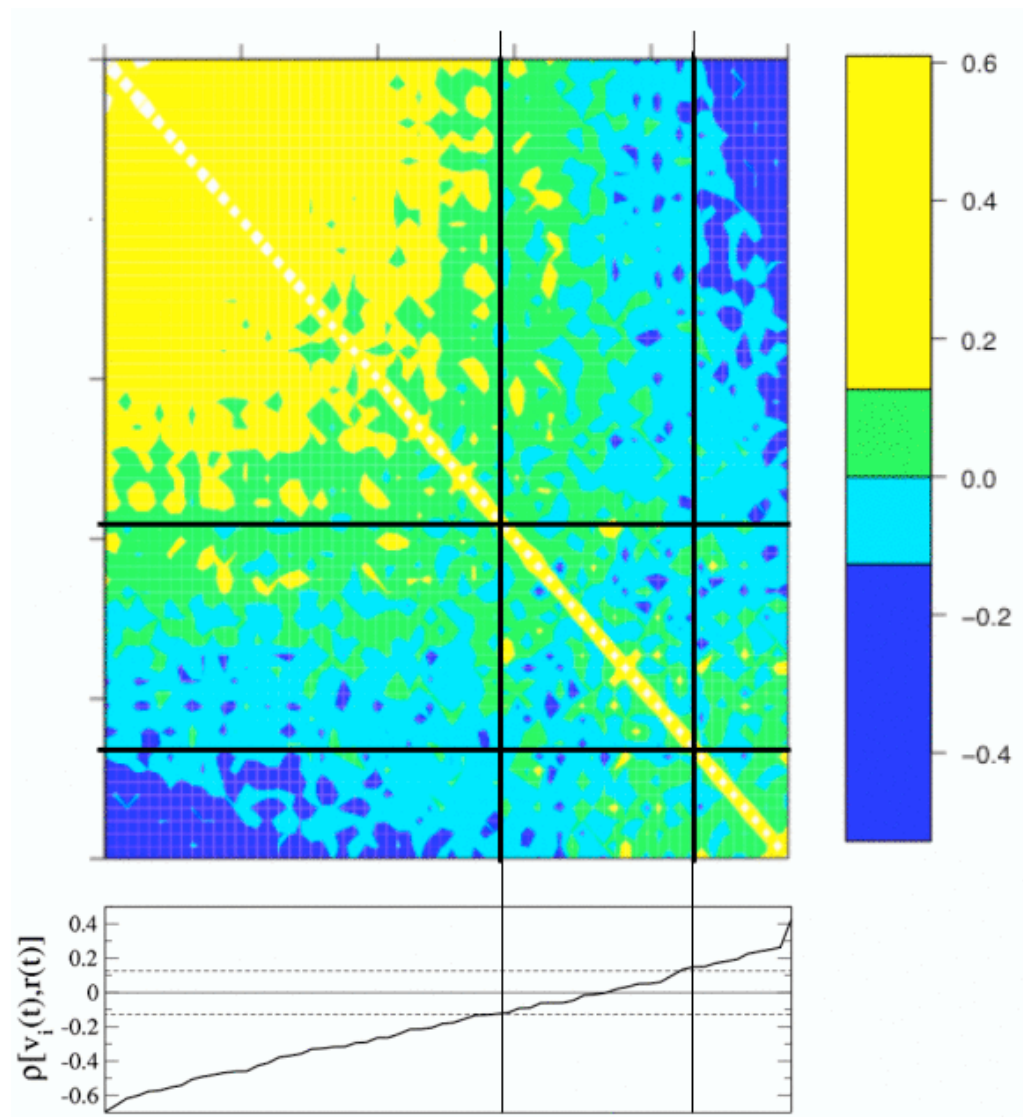
Figure 3. Contour plot of the correlation matrix of daily inventory variation of firms trading the stock BBVA in 2003 plotted by sorting the firms into rows and columns according to their value of correlation of inventory variation with BBVA price return $\rho[v_i(t), r(t)]$. The bottom panel shows the value of $\rho[v_i(t), r(t)]$ of the firms in the same order as in the matrix. The dashed lines in the bottom panel bound the 2σ significance interval. The colors of the matrix are chosen to highlight positive and negative firm daily inventory variation cross correlation values $\rho[v_i(t), v_j(t)]$ which are according to a given significance level. Specifically, yellow (blue) indicates positive (negative) cross correlation with a significance of 2σ, whereas green (cyan) indicates positive (negative) cross correlation within the 2σ interval. Two groups of firms are seen, one on the top left corner and the other on the bottom right corner. These two groups present a significant level of anticorrelation of their inventory variation profile.

3. Causality

The correlation between inventory variation and price return raises the question of the causality relation between these two variables [13, 16] and of the time correlation properties of inventory variations. While efficiency requires that price returns are uncorrelated, little is known about the time correlation of inventory variation. Here we investigate the causality problem for the firms of our database starting from a 15 min time horizon. As figure 4(a) shows, there is an average significant autocorrelation of $v_i(t)$ for more than one trading day. However, while a large majority of the trending firms are characterized by an autocorrelation of $v_i(t)$ which is significantly positively autocorrelated, reversing firms show a more heterogeneous behavior. For a few of them $v_i(t)$ is negatively autocorrelated on a timescale of approximately 15 min whereas the large majority show a positive autocorrelation with quite a heterogeneous timescale. Thus few large trending firms act on a long timescale most probably in order to build a position by splitting large orders [22]–[25] to minimize their market impact [26], whereas many reversing firms of different size act both on a short and a long timescale.

The analysis of lagged cross correlation between $v_i(t)$ and $r(t)$ indicates a clear asymmetry. As figure 4(b) shows, inventory variation $v_i(t)$ is correlated with price return in the near past $r(t + \tau)$ ($\tau < 0$), whereas $v_i(t)$ is not significantly correlated with $r(t + \tau)$ in the near future ($\tau > 0$) for trending and reversing firms. This observation suggests an asymmetric causal relation between v_i and r. To make this observation more rigorous we perform a series of Granger causality tests [27]. A Granger causality test investigates whether a quantity X can help to linearly forecast another quantity Y. It is said that X fails to Granger-cause Y if for all $s > 0$ the mean squared error of a forecast Y_{t+s} based on $(Y_t, Y_{t-1}, \ldots)$ is statistically indistinguishable from the mean squared error of a forecast of Y_{t+s} that uses both $(Y_t, Y_{t-1}, \ldots)$ and $(X_t, X_{t-1}, \ldots)$. The null hypothesis states that X does not Granger-cause Y, and is statistically assessed under the hypothesis of Gaussian disturbances with respect to a threshold value often chosen equal to 95%. The detection of Granger-causality therefore indicates that a time series X can be seen as a useful predictor of the time series Y. It is worth noting that there is no guarantee that Granger-causality directly implies true causation. In a typical use of a bivariate Granger-causality test one investigates both whether X helps forecast Y and whether Y helps forecast X [28]. Here we apply the Granger causality test to detect the ability of inventory variation to predict future stock return and *vice versa*. The null hypothesis is that the past p values of X do not help in predicting the value of Y. In our analysis we choose $p = 10$ corresponding to 150 trading minutes for the shortest time horizon. We consider both $X = r(t)$, $Y = v_i(t)$ and $X = v_i(t)$, $Y = r(t)$. For each firm and for each investigated set we construct an indicator $I(X \rightarrow Y)$ assuming the value 1 when the null hypothesis is rejected therefore implying that X Granger-causes Y and 0 when the null hypothesis of absence of causality cannot be rejected at a 95% confidence level. The results of the 1115 tests are summarized in figure 4(c) where we show the conditional average value of $I(X \rightarrow Y)$ both for $X = r(t)$, $Y = v_i(t)$ and for $X = v_i(t)$, $Y = r(t)$ as a function of the $\rho[v_i(t), r(t)]$ of the considered firm i. Figure 4(c) shows that for the large majority of firms which are categorized as trending or reversing (i.e. with $|\rho[v_i(t), r(t)]| > 2\sigma$) the stock return $r(t)$ Granger-causes $v_i(t)$ but not *vice versa*. In fact the average value of $I(r(t) \rightarrow v_i(t))$ (red symbols) is close to one for reversing and trending firms whereas it tends to a minimal value of approximately 0.3 when $\rho[v_i(t), r(t)]$ is close to zero (mostly uncategorized firms). On the other hand, the average value of $I(v_i(t) \rightarrow r(t))$ (black symbols) is close to 0.2 and approximately
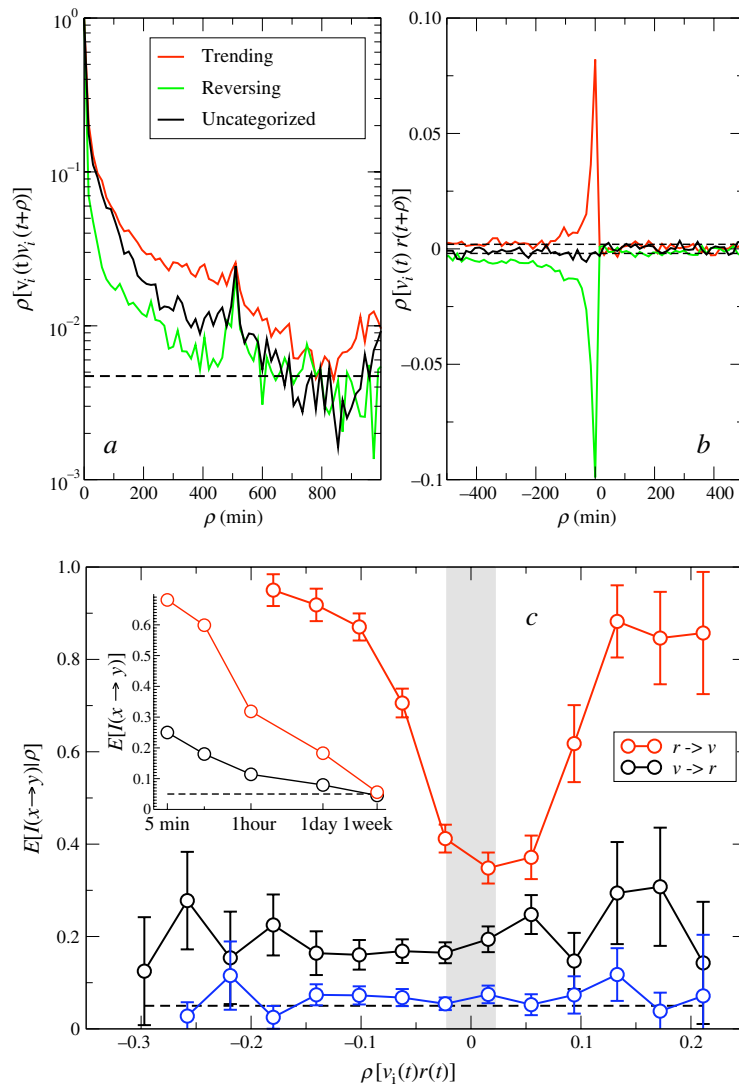
Figure 4. Panels (a) and (b) show the averaged autocorrelation of $v_i(t)$ and averaged lagged cross-correlation $\rho[v_i(t), r(t+\tau)]$ (respectively) for the different firm groups. The dashed lines bound the 2σ significance interval. In panel (c), we show the conditional expected value of the indicator $I(X \to Y)$ of the rejection of the null hypothesis of non-Granger causality between X and Y with 95% confidence as a function of the simultaneous cross correlation $\rho[v_i(t), r(t)]$. Red symbols refer to the test that $r(t)$ is Granger-causing $v_i(t)$ whereas black symbols refer to the test that $v_i(t)$ is Granger-causing $r(t)$. Granger-causality is tested on the time series of $v_i(t)$ and $r(t)$ obtained at each $\Delta t = 15$ min and over $p = 10$ lags, corresponding to 150 min. The blue line is a result of the Granger test on shuffled data. The shaded area corresponds to values of ρ within the $\pm 2\sigma$ significance level. The inset shows the average value of the indicator for other time horizons Δt with $p = 10$ and the same color code as before. The dashed line is the value 0.05 expected for a test of a null hypothesis performed by using a 95% confidence level.

constant as a function of $\rho[v_i(t), r(t)]$. The value of 0.2 is smaller than any value observed in the previous test but significantly different from the value of 0.05 expected for a test of a null hypothesis performed by using a 95% confidence level. This discrepancy might be due to a certain degree of Granger causality also for the case $v_i(t) \rightarrow r(t)$ for a minority of the firms or could alternatively just be an effect of the non Gaussianity of the considered time series. In the attempt to discriminate the role of non Gaussianity we have performed the same $I(v_i(t) \rightarrow r(t))$ test on a shuffled version of the time series. The result of this investigation is shown in figure 4(c) as a blue line. In this last case the average value of $I(v_i(t) \rightarrow r(t))$ is close to 0.07, a value not too different from the expected value of 0.05. The result of the shuffling therefore suggests that the observed values of the average of $I(v_i(t) \rightarrow r(t))$ in real data cannot be ascribed to the distributional properties of $r(t)$ and $v_i(t)$ and might indicate that $v_i(t)$ Granger-causes $r(t)$ only for a small number of firms. Finally, we notice that Granger-causality disappears as the time horizon increases. For example, in figure 4(c) we find that no significant Granger-causality is found between $r(t)$ and $v_i(t)$ at the weekly level. In summary our analysis indicates that for the largest majority of reversing and trending firms returns are Granger causing inventory variation but not *vice versa* at the day or intraday level.

4. Herding and net flow measure

Are firms belonging to the same group behaving in a similar way at specific time intervals? To answer this question we use an indicator based on the inventory variation of each firm. The herding indicator

$$h = \frac{\# \text{ of buying firms}}{\# \text{ of buying firms} + \# \text{ of selling firms}} \tag{1}$$

of the group is the number of buying firms divided by the number of firms of the group which are active in the specific time interval (buying or selling). This herding indicator is a simplified version of the herding measure introduced in [11] to quantify the herding of institutional investors in selecting a basket of stocks. Differently than in [11], here we limit our investigation to the univariate case of the investment in a single stock. We infer that herding is associated to the observation of a high value (buy herding) or low value (sell herding) of h by evaluating the probability to observe a number of buying (selling) firms equal or larger than the empirically detected one under a binomial null hypothesis. Specifically, we infer that herding is present when the probability of the observed number of buying or selling firms is smaller than 5% under a binomial null hypothesis where the probability of a firm of a group to be a buyer or a seller is estimated from data for each investigated year. We estimate h for the three groups of firms at the 15 min and 1 day time horizon. Table 2 shows that firms characterized by a reversing resulting strategy present herding in a significant fraction of time intervals. Specifically, the percentage of herding intervals averaged over four years ranges from 31.3% for the 15 min time horizon to 64.1% for the 1 day time horizon. The percentage of herding is much less pronounced for firms with a trending resulting strategy. For this group we observe a percentage of herding intervals of a few percent for the time horizon both of 15 min (4.4%) and 1 trading day (6.3%). The uncategorized firms present a behavior which is intermediate between those observed for reversing and trending firms. It is worth noting that for the selected threshold of 5% used both for buying and selling herding intervals the expected percentage of buying and selling herding time intervals being consistent with the null hypothesis is 10% when the number of firms of

Table 2. Percentage of herding intervals observed for the groups of reversing, uncategorized and trending firms actively trading the Telefónica stock during the period 2001–2004. The percentage of herding intervals is also provided separately for buying (BH) and selling (SH) herding. For each block of results, the first three rows refer to the 1 day time horizon (250 trading days) and the last three rows refer to the 15 min intraday time horizon (8500 trading intervals).

	2001			2002		
	ALL	BH	SH	ALL	BH	SH
Reversing (1 day)	66.8	34.8	32.0	65.2	34.8	30.4
Uncategorized (1 day)	22.4	11.2	11.2	16.4	7.2	9.2
Trending (1 day)	10.4	7.2	3.2	6.4	2.4	4.0
Reversing (15 min)	35.1	17.4	17.7	34.5	17.3	17.2
Uncategorized (15 min)	10.1	5.3	4.8	11.6	5.7	5.9
Trending (15 min)	3.7	2.1	1.6	6.7	3.4	3.3
	2003			2004		
	ALL	BH	SH	ALL	BH	SH
Reversing (1 day)	64.8	31.2	33.6	59.6	27.2	32.4
Uncategorized (1 day)	21.2	10.8	10.4	19.2	10.4	8.8
Trending (1 day)	6.0	2.0	4.0	2.4	1.2	1.2
Reversing (15 min)	29.2	14.7	14.5	26.6	13.3	13.3
Uncategorized (15 min)	10.2	5.3	4.9	11.5	6.3	5.2
Trending (15 min)	3.9	1.7	2.2	3.3	1.7	1.6

the group is sufficiently large. When the number of firms is about 10 or less (as in our case for the trending firms) due to the discretization of the binomial distribution the 10% value is the upper boundary expected according to the null hypothesis. At daily time horizon reversing and uncategorized firms are herding for a percentage of time intervals which is much larger than 10%. Conversely, trending firms are herding rarely or not herding at all by considering that their percentage of herding is compatible with the one expected for a group of elements of their size on a pure chance basis.

An illustration of the occurrence of the herding time intervals estimated for the one day time horizon is provided in figure 5 for the Telefónica stock. Each panel refers to a different group of firms. The herding days of reversing firms are highly frequent and approximately uniformly distributed over the investigated time period. The prevalence over long periods of time of the kind of observed herding (buying or selling) is related to the prevalence of a bull or bear market phase.

The herding measure of equation (1) is quite effective in detecting herding for firms of approximately the same size and frequency of the trading activity. However the herding indicator h does not give information on the net flow of value of a given group of firms. A firm acts often as an intermediary for many different clients and in a given day the total purchased value can be close to the total sold value. In these cases, the sign of the inventory variation is just the effect of random fluctuations and the net flow of the firm is small. Here we want to verify that when a given group of firms herds (according to h) the net flow of value of the group is consistent with the direction of herding indicated by h. To achieve this goal we adapt to our investigations the
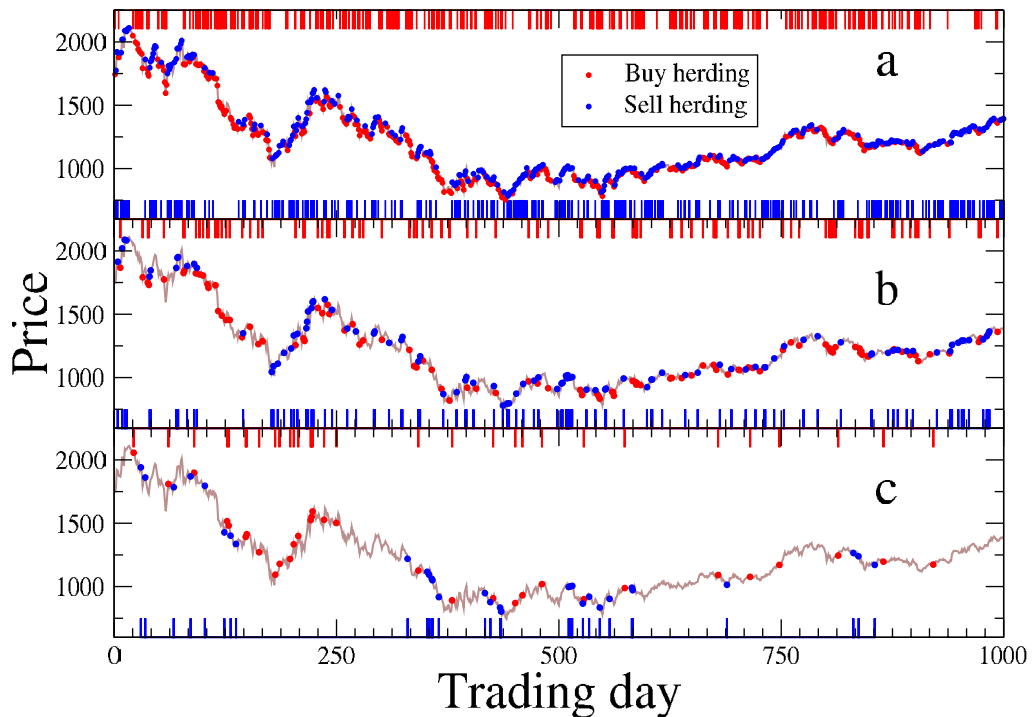
Figure 5. The thin brown line is the daily closure price of Telefónica stock for the January 2001–December 2004 time period. The three panels refer to reversing (a), uncategorized (b) and trending (c) firms. Red circles indicate buying herding days whereas blue circles indicate selling herding days. The herding days are also indicated by the red (buy herding) and blue (sell herding) segments drawn on the top and bottom time axes of each panel.

buy ratio b used in [15]. Specifically, for each time interval and for each group we compute

$$b = \frac{\sum_{i \in \text{buying firms}} v_i}{\sum_{i \in \text{all firms}} |v_i|}. \tag{2}$$

The buy ratio b varies between zero and one. Low values of b indicate time intervals when firms of the considered group are mostly selling whereas high values close to one indicate that firms are mostly buying. In table 3, we show the mean value of b for all the groups and for the same values of time horizon used in table 2. The results refer to the active firms trading the Telefónica stock during 2001 and a similar behavior is observed for the other investigated years. The values of b are computed both unconditionally on all the investigated time intervals and conditioning on the intervals characterized as buying or selling herding intervals by the herding indicator h with the associated binomial test.

The expected value of b under a null hypothesis of no herding cannot be estimated *a priori* especially if we constrain the null hypothesis by asking it to reflect the size heterogeneity of the considered set and the total exchanged value of each firm. For this reason in table 3 we also show the values of $\langle b \rangle_{\text{Shuf}}$ estimated for artificial time series of the inventory variation obtained by randomly shuffling the time of all the transactions occurring between the considered firms. In each shuffled series each firm maintains the number of transactions and the amount of exchanged value (hence, the firm's yearly net flow) as in the real data. Table 3 shows that

Table 3. Mean value of the buy ratio b of firms which are active in a given time interval. The mean values are computed for real data both unconditional and conditional on the buying (BH) and selling (SH) herding intervals for the reversing, uncategorized and trending groups trading the Telefónica stock in 2001. In the second column, we also provide the mean value of b obtained from shuffled time series. In our shufflings we use 25 000 daily intervals and 425 000 15 min intervals in the computation of $\langle b \rangle_{Shuf}$. The first three rows refer to the 1 day time horizon whereas the last three rows refer to the 15 min time horizon. The dispersion of mean values is one standard deviation. The number in parenthesis is the number of records of the considered set. The presence of symbols △ and □ selects the pairs of mean values which pass a 99% confidence level t-test that the two mean values are equal (see text for details).

	$\langle b \rangle_{Shuf}$	$\langle b \rangle$	$\langle b \rangle_{BH}$	$\langle b \rangle_{SH}$
Reversing	0.538	0.52 ± 0.28	0.77 ± 0.15	0.22 ± 0.16
(1 day)		(250)	(87)	(80)
Uncategorized	0.418	0.48 ± 0.16	0.55 ± 0.16	0.43 ± 0.15
(1 day)	△	□ (250)	□ (28)	△ □ (28)
Trending	0.556	0.51 ± 0.25	0.81 ± 0.20	0.22 ± 0.19
(1 day)		(250)	(18)	(8)
Reversing	0.5095	0.50 ± 0.22	0.69 ± 0.18	0.32 ± 0.18
(15 min)		(8500)	(1480)	(1500)
Uncategorized	0.4783	0.49 ± 0.23	0.70 ± 0.21	0.30 ± 0.20
(15 min)		(8500)	(452)	(406)
Trending	0.5105	0.51 ± 0.27	0.88 ± 0.19	0.17 ± 0.20
(15 min)		(8500)	(181)	(137)

the mean value of b is close to but not exactly equal to 1/2 both for the shuffled time series and for the real data, when the average is performed unconditionally over all intervals. The observed values are closer to 0.5 for the 15 min time horizon. Differently, when the average is computed conditioning on buying or selling herding intervals one obtains the mean values of b deviating from both the values obtained from the unconditional analysis of real data and from the shuffled time series. To estimate the statistical reliability of the differences observed between the unconditional mean values, the mean values obtained from shuffled data and the conditional ones we perform two different t-tests at a 99% confidence threshold of the hypothesis that the mean values of each pair of the considered samples are the same. Specifically we compare (i) the unconditional mean value $\langle b \rangle$ with each of the mean values estimated conditioning on the buying $\langle b \rangle_{BH}$ and selling $\langle b \rangle_{SH}$ herding intervals and (ii) the mean value obtained by shuffled time series $\langle b \rangle_{Shuf}$ with each corresponding $\langle b \rangle_{BH}$ and $\langle b \rangle_{SH}$.

The results summarized in table 3 show that the differences between the unconditional mean values or the mean values obtained from shuffled time series of the buy ratio b and the conditional ones are not consistent with the null hypothesis that they belong to the same distributions in almost all cases with the exception for the 1 day time horizon of uncategorized firms. Therefore, with only this exception, the investigation of b indicates that during herding periods there is a consistent net flow of exchanged value by the considered group which is in agreement with the herding action of the firms.

5. Conclusions

Our results show that a large number of firms trading a financial asset in a financial market are characterized by well-defined and detectable trending or reversing resulting strategies. We also show that trending and reversing firms present a characteristic pattern of herding behavior both at daily and at intradaily time horizons. Reversing firms are herding quite frequently and uniformly in time whereas trending firms are herding more rarely.

Market dynamics can therefore be seen as the interplay of at least two classes of traders, different with respect to their size heterogeneity and responding to the price changes in different ways. It is possible that the fluctuation of price returns, i.e. the market volatility, is significantly affected by the fluctuations in the relative trading intensity of the two groups. Our results open up the possibility of setting up agent-based models of financial firms trading in a financial market. These models can now be empirically grounded in the type of resulting strategies characterizing the dynamics of real firms.

Acknowledgments

References

[1] Simon H A 1962 *Proc. Am. Phil. Soc.* **106** 467
[2] Anderson P W 1972 *Science* **177** 393
[3] Gode D K and Sunder S 1993 *J. Political Econ.* **101** 119
[4] Arthur W B 1994 *Am. Econ. Rev.* **84** 406
[5] Levy M, Levy H and Solomon S 1994 *Econ. Lett.* **45** 103
[6] Lux T and Marchesi M 1999 *Nature* **397** 498
[7] Le Baron B, Arthur W B and Palmer R 1999 *J. Econ. Dyn. Control* **23** 1487
[8] Le Baron B 2000 *J. Econ. Dyn. Control* **24** 679
[9] Tesfatsion L 2001 *J. Econ. Dyn. Control* **25** 281
[10] Challet D, Marsili M and Zhang Y-C 2005 *Minority Games: Interacting Agents in Financial Markets* (Oxford: Oxford University Press)
[11] Lakonishok J, Shleifer A and Vishny R W 1992 *J. Finance Econ.* **32** 23
[12] Härdle W and Kirman A 1995 *J. Econ.* **67** 227
[13] Nofsinger J R and Sias R W 1999 *J. Finance* **54** 2263
[14] Choe H, Kho B-C and Stulz R M 1999 *J. Finance Econ.* **54** 227
[15] Grinblatt M and Keloharju M 2000 *J. Finance Econ.* **55** 43
[16] Griffin J M, Harris J H and Topaloglu S 2003 *J. Finance* **58** 2285
[17] Kossinets G and Watts D J 2006 *Science* **311** 88
[18] Laloux L, Cizeau P, Bouchaud J P and Potters M 1999 *Phys. Rev. Lett.* **83** 1467
[19] Plerou V, Gopikrishnan P, Rosenow B, Amaral L A N and Stanley H E 1999 *Phys. Rev. Lett.* **83** 1471
[20] Plerou V, Gopikrishnan P, Rosenow B, Amaral L A N, Guhr T and Stanley H E 2002 *Phys. Rev.* E **65** 066126

[21] Lillo F and Mantegna R N 2005 *Phys. Rev.* E **72** 016219
[22] Chan L K C and Lakonishok J 1995 *J. Finance* **50** 1147
[23] Gabaix X, Gopikrishnan P, Plerou V and Stanley H E 2003 *Nature* **423** 267
[24] Lillo F and Farmer J D 2004 *Stud. Nonlinear Dyn. Econometrics* **8** 1
[25] Vaglica G, Lillo F, Moro E and Mantegna R N 2008 *Phys. Rev.* E **77** 036110
[26] Hasbrouck J 1991 *J. Finance* **46** 179
[27] Granger C W J 1969 *Econometrica* **37** 424
[28] Hamilton J D 1994 *Time Series Analysis* (Princeton, NJ: Princeton University Press)