



This is an electronic reprint of the original article. This reprint may differ from the original in pagination and typographic detail.

Zhou, Wei Xing; Mu, Guo Hua; Kertész, János

Random matrix approach to the dynamics of stock inventory variations

Published in: New Journal of Physics

DOI: 10.1088/1367-2630/14/9/093025

Published: 01/09/2012

Document Version Publisher's PDF, also known as Version of record

Published under the following license: CC BY-NC-SA

Please cite the original version:

Zhou, W. X., Mu, Ğ. H., & Kertész, J. (2012). Random matrix approach to the dynamics of stock inventory variations. *New Journal of Physics*, *14*, Article 093025. https://doi.org/10.1088/1367-2630/14/9/093025

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

PAPER • OPEN ACCESS

Random matrix approach to the dynamics of stock inventory variations

To cite this article: Wei-Xing Zhou et al 2012 New J. Phys. 14 093025

View the article online for updates and enhancements.

Related content

- Specialization and herding behavior of trading firms in a financial market Fabrizio Lillo, Esteban Moro, Gabriella Vaglica et al.
- Statistical identification with hidden Markov models of large order splitting strategies in an equity market Gabriella Vaglica, Fabrizio Lillo and Rosario N Mantegna
- Recurrence interval analysis of highfrequency financial returns and its application to risk estimation Fei Ren and Wei-Xing Zhou

Recent citations

- Correlation Structure and Evolution of World Stock Markets: Evidence from Pearson and Partial Correlation-Based **Networks** Gang-Jin Wang et al
- The q-dependent detrended crosscorrelation analysis of stock market Longfeng Zhao et al
- Market Correlation Structure Changes Around the Great Crash: A Random Matrix Theory Analysis of the Chinese Stock Market Rui-Qi Han et al



IOP ebooks[™]

to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

New Journal of Physics

The open-access journal for physics

Random matrix approach to the dynamics of stock inventory variations

Wei-Xing Zhou 1,2,3,6 , Guo-Hua Mu 1,2,3 and János Kertész 4,5,6

¹ School of Business, East China University of Science and Technology, Shanghai 200237, People's Republic of China
² School of Science, East China University of Science and Technology, Shanghai 200237, People's Republic of China
³ Research Center for Econophysics, East China University of Science and Technology, Shanghai 200237, People's Republic of China
⁴ Institute of Physics, Budapest University of Technology and Economics, Budapest H-1111, Hungary
⁵ Department of Biomedical Engineering and Computational Science, Aalto University, Espoo FI-00076, Finland E-mail: wxzhou@ecust.edu.cn and kertesz@phy.bme.hu

New Journal of Physics **14** (2012) 093025 (19pp) Received 11 June 2012 Published 14 September 2012 Online at http://www.njp.org/ doi:10.1088/1367-2630/14/9/093025

Abstract. It is well accepted that investors can be classified into groups owing to distinct trading strategies, which forms the basic assumption of many agent-based models for financial markets when agents are not zero-intelligent. However, empirical tests of these assumptions are still very rare due to the lack of order flow data. Here we adopt the order flow data of Chinese stocks to tackle this problem by investigating the dynamics of inventory variations for individual and institutional investors that contain rich information about the trading behavior of investors and have a crucial influence on price fluctuations. We find that the distributions of cross-correlation coefficient C_{ij} have power-law forms in the bulk that are followed by exponential tails, and there are more positive coefficients than negative ones. In addition, it is more likely that two individuals or two institutions have a stronger inventory variation correlation than one

⁶ Authors to whom any correspondence should be addressed.

Content from this work may be used under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

New Journal of Physics **14** (2012) 093025 1367-2630/12/093025+19\$33.00

individual and one institution. We find that the largest and the second largest eigenvalues (λ_1 and λ_2) of the correlation matrix cannot be explained by random matrix theory and the projections of investors' inventory variations on the first eigenvector $u(\lambda_1)$ are linearly correlated with stock returns, where individual investors play a dominating role. The investors are classified into three categories based on the cross-correlation coefficients C_{VR} between inventory variations and stock returns. A strong Granger causality is unveiled from stock returns to inventory variations, which means that a large proportion of individuals hold the reversing trading strategy and a small part of individuals hold the trending strategy. Our empirical findings have scientific significance in the understanding of investors' trading behavior and in the construction of agent-based models for emerging stock markets.

Contents

1.	Introduction	2						
2.	The data	3						
3.	Statistics of the correlation matrix between two time series of inventory variations	s 4						
	3.1. Distributions of cross-correlation coefficients	4						
	3.2. The eigenvalue spectrum	7						
	3.3. Distribution of eigenvector components	9						
	3.4. Information in eigenvectors for deviating eigenvalues	10						
4.	Inventory variation and stock return	11						
	4.1. Categorization of investors	11						
	4.2. Causality	13						
	4.3. Herding behavior	16						
5.	Summary	17						
Acknowledgments								
1. Information 2. The data 3. Statistics of the correlation matrix between two time series of inventory variations 3.1. Distributions of cross-correlation coefficients 3.2. The eigenvalue spectrum 3.3. Distribution of eigenvector components 3.4. Information in eigenvectors for deviating eigenvalues 1 4. Inventory variation and stock return 4.1. Categorization of investors 4.2. Causality 4.3. Herding behavior 1 4.3. Herding behavior 1 4. Summary 1 Acknowledgments 1 References								

1. Introduction

Stock markets are complex systems, whose elements are heterogenous individual and institutional investors interacting with each other by stock exchanges [1-4]. Stock price fluctuates due to investors' trading activities, and the cross-sectional relation between investors' stock inventory variations and stock returns has attracted much attention [5]. The large body of literature falls into three groups: to study the relation between past returns and inventory variations, to investigate the contemporaneous relation between inventory variations and stock returns and to analyse the return predictability of inventory variations [6]. The main findings are that institutions are trending investors adopting the momentum trading strategy [6–8], while individuals are reversing investors who buy previous losers and sell previous winners [6, 8–11] and stock returns lead inventory variations but not vice versa [5–8].

However, there is evidence showing different trading patterns. Lillo *et al* [5] investigated the trading behavior of about 80 firms that were members of the Spanish Stock Exchange and found that there were more reversing firms than trending firms. They also found that the largest

eigenvalue of the correlation matrix of inventory variations cannot be explained by random matrix theory (RMT) and its eigenvector contains information on stock price fluctuations. Both buying and selling herding behavior has been observed for trending and reversing firms. In a recent paper, an alternative identification approach of investor clusters has been proposed from their real trading activity [12], which is based on the statistical validation of links in complex networks [13–15].

In this work, we perform a similar analysis to that in [5] based on the trading records of Chinese investors in the Shenzhen Stock Exchange. As officially documented by the Spanish Stock Exchange, 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; that is, a firm is a market member that usually acts as a brokerage house for individual and institutional investors⁷. Therefore, in the Spanish case [5], the inventory variation over a given time interval is a sum of the inventory variations of many individuals and institutions. Different from the Spanish case, our data set contains both individual and institutional investors, which allows us to observe interesting investor behavior. Our analysis starts from the perspective of random matrix theory, which has been extensively used to investigate the cross-correlations of financial returns in different stock markets [16-18]. However, very few studies have been conducted on the Chinese stocks [19] and, to our knowledge, there is no research reported on the dynamics of inventory variations of Chinese investors. Alternatively, there are studies on Chinese equities at the transaction and trader level from the complex network perspective [20-23]. We note that random matrix theory has very wide applications in physics from high-energy physics to statistical physics; see [24] for a brief review and see the papers in this issue.

This paper is organized as follows. Section 2 describes the data and the method for constructing the time series of investors' inventory variations. Section 3 studies the statistical properties of the elements, eigenvalues and eigenvectors of the correlation matrix of inventory variations. Section 4 investigates the contemporaneous and lagged cross-correlation between inventors' inventory variations and stock returns to divide investors into three categories and their herding behavior. Section 5 summarizes our findings.

2. The data

We analyze 39 stocks actively traded on the Shenzhen Stock Exchange in 2003. The database contains all the information needed for the analysis in this work. For each transaction *i* of a given stock, the data record the identities of the buyer and seller, the types (individual or institution) of the two traders, the price p_i and the size q_i of the trade, and the time stamp. Therefore, the trading history of each investor is known. For each stock, we identify active traders who had more than 150 transactions, amounting to about three transactions per week. If the number of active traders of a stock is less than 120, we exclude it from analysis. In this way, we have 15 stocks for analysis.

Following Lillo *et al* [5], we investigate the dynamics of the inventory variation of the most active investors who executed more than 120 transactions for each stock. Although the trading period of each day consists of call auction and continuous auction, their behavior is different in many aspects and is usually studied separately [25, 26]. We stress that all the transactions in both call auction and continuous auction are included in our investigation. The daily inventory

⁷ Private communication with R N Mantegna.

Table 1. Basic statistics of the investigated stocks. The first column is the stock code, which is the unique identity of each stock. The second and third columns present investor-averaged total inventory variation $\langle \sum_t v \rangle_i$ and average absolute variation $\langle \langle |v| \rangle_t \rangle_i$. The fourth to eighth columns give the number of investors N, the number of trending investors N^{tr} , the number of reversing investors N^{re} , the number of uncategorized investors N^{un} and the slope of the factor versus stock return k. The variables in the ninth to thirteenth columns are the same as those in the five 'all investors' columns but for individual investors, and the last five columns are for institutional investors. Each value in the last row gives the sum of the numbers in the same column.

			All investors					Individuals					Institutions				
Code	$\langle \sum_t v \rangle_i$	$\langle \langle v \rangle_t \rangle_i$	N	N _{tr}	N _{re}	N _{un}	k	Nind	$N_{\rm ind}^{\rm tr}$	$N_{\rm ind}^{\rm re}$	$N_{\rm ind}^{\rm un}$	k _{ind}	Nins	$N_{\rm ins}^{\rm tr}$	$N_{\rm ins}^{\rm re}$	$N_{ m ins}^{ m un}$	kins
000001	-2.99×10^{6}	1.12×10^5	80	7	41	32	0.83	61	4	39	18	0.83	19	3	2	14	0.04
000002	1.01×10^{6}	1.46×10^5	80	6	29	45	0.49	42	2	26	14	0.52	38	4	3	31	0.06
000012	-1.64×10^{6}	7.88×10^4	81	5	20	56	0.19	78	5	20	53	0.18	3	0	0	3	0.06
000021	$5.09 imes 10^5$	4.35×10^4	81	2	43	36	0.78	64	1	43	20	0.78	17	1	0	16	0.04
000063	1.46×10^7	2.46×10^5	80	9	20	51	0.13	20	2	13	5	0.71	60	7	7	46	0.05
000488	$1.35 imes 10^6$	8.92×10^4	80	5	5	70	0.08	69	1	5	63	0.06	11	4	0	7	0.06
000550	4.20×10^{6}	7.53×10^4	81	2	37	42	0.18	45	0	35	10	0.18	36	2	2	32	0.06
000625	$1.87 imes 10^6$	1.22×10^5	80	10	26	44	0.71	62	8	24	30	0.69	18	2	2	14	0.05
000800	2.53×10^6	2.56×10^5	80	6	19	55	0.30	31	1	17	13	0.31	49	5	2	42	0.06
000825	6.33×10^6	9.38×10^4	80	5	38	37	0.69	50	3	37	10	0.70	30	2	1	27	0.05
000839	8.13×10^{5}	6.67×10^4	80	2	40	38	0.84	60	2	38	20	0.84	20	0	2	18	0.04
000858	$1.75 imes 10^6$	1.66×10^5	82	4	21	57	0.33	31	0	19	12	0.67	51	4	2	45	0.05
000898	$6.26 imes 10^6$	1.20×10^5	83	6	32	45	0.79	46	0	31	15	0.80	37	6	1	30	0.04
200488	3.59×10^{6}	$5.05 imes 10^4$	80	9	30	41	0.65	53	7	25	21	0.66	27	2	5	20	0.05
200625	$6.27 imes 10^6$	1.26×10^5	83	9	14	60	0.50	46	5	9	32	0.58	37	4	5	28	0.05
\sum	-	_	1211	87	415	709	-	758	41	381	336	-	453	46	34	373	-

variation of an investor *i* trading a given stock on day *t* is defined as follows:

$$v_i(t) = \sum^{+} p_i(t)q_i(t) - \sum^{-} p_i(t)q_i(t), \qquad (1)$$

where $\sum_{i=1}^{i} p_i(t)q_i(t)$ is the total buy quantity on trading day t and $\sum_{i=1}^{i} p_i(t)q_i(t)$ is the total sell quantity on the same day. The basic statistics of the 80 most active traders and the resultant inventory variations are given in table 1. The time series of the daily inventory variations of three investors trading stock 000001 are shown in figure 1.

3. Statistics of the correlation matrix between two time series of inventory variations

3.1. Distributions of cross-correlation coefficients

The empirical correlation matrix C is constructed from the time series of inventory variation $v_i(t)$ of the investigated stock, defined as

$$C_{ij} = \frac{\langle (v_i - \langle v_i \rangle)(v_j - \langle v_j \rangle) \rangle}{\sigma_i \sigma_j}.$$
(2)



Figure 1. Typical time series of the daily inventory variations of three randomly selected investors trading stock 000001.

Since the results for individual stocks are quantitatively similar, we put the cross-correlation coefficients of the 15 stocks into one sample. We find that the mean value is $\langle C_{ij} \rangle = 0.02$ for the real data and 0 for the shuffled data. When the types of investors are taken into account, the mean value of the cross-correlation coefficients is $\langle C_{ij} \rangle = 0.048$ (shuffled: -0.001) for both *i* and *j* being individuals, $\langle C_{ij} \rangle = 0.014$ (shuffled: 0) for both *i* and *j* being institutions and $\langle C_{ij} \rangle = -0.008$ (shuffled: -0.001) for *i* being individual and *j* being institution.

Figure 2 plots the daily returns of stock 000001 and the sliding average values of the correlation coefficients $\langle C_{ij} \rangle$ for comparison. We observe that large values $\langle C_{ij} \rangle$ appear during periods of large price fluctuations by and large, which is reminiscent of a similar result for cross-correlations of financial returns [18]. However, the short time period of our data sample does not allow us to reach a decisive conclusion. There are also less volatile time periods with large $\langle C_{ij} \rangle$. The situation is quite similar for other stocks.

Figure 3(a) shows the empirical probability distributions of C_{ij} , which is calculated using daily inventory variation. The four curves with different markers correspond to C_{ij} , $C_{ind,ind}$, $C_{ind,ins}$ and $C_{ins,ins}$, respectively. It is found that most coefficients are small and the tails are exponentials:

$$P(C) \propto \begin{cases} e^{-\lambda_{-}C}, & -0.6 < C \leq -0.1, \\ e^{-\lambda_{+}C}, & 0.1 < C \leq 0.6, \end{cases}$$
(3)

where $\lambda_{+} = 8.8 \pm 0.2$ and $\lambda_{-} = 11.1 \pm 0.3$ for C_{ij} , $\lambda_{+} = 8.9 \pm 0.2$ and $\lambda_{-} = 12.5 \pm 0.4$ for $C_{\text{ind,ind}}$, $\lambda_{+} = 10.8 \pm 0.4$ and $\lambda_{-} = 10.5 \pm 0.4$ for $C_{\text{ind,ins}}$ and $\lambda_{+} = 6.6 \pm 0.5$ and $\lambda_{-} = 8.7 \pm 0.5$ for $C_{\text{ins,ins}}$, respectively. We find that there are more positive cross-correlation coefficients $(\lambda_{+} < \lambda_{-})$ when both investors are individuals or institutions. In contrast, the distribution is symmetric $(\lambda_{+} \approx \lambda_{-})$ when one investor is an individual while the other is an institution. This finding implies that herding behavior is more likely to occur among the same type of investors, and individuals have larger probability to herd than institutions. We shuffle the original time



Figure 2. Evolution of the five-day average cross-correlation coefficient $\langle C_{ij} \rangle$ for inventory variations and the daily return *R*.



Figure 3. Empirical distributions of the cross-correlation coefficients for all 15 stocks. (a) Log–linear plot of $P(C_{ij})$ for cross-correlations between any two investors, any two individuals, any one individual and one institution, and any two institutions, which shows exponential forms when $0.1 < |C| \leq 0.6$. The dashed line corresponds to the result of shuffled data. (b) Log–log plot of $P(C_{ij})$, which shows power-law forms when $10^{-5} < |C| \leq 0.01$. (c) Comparison of the occurrence numbers of positive and negative cross-correlations. The ordinate gives $\log_{10}[1 + N(C_{ij})]$ rather than $\log_{10}[N(C_{ij})]$ for better presentation.

series and perform the same analysis. The resulting distributions collapse onto a single curve, which has an exponential form

$$P(C) = \lambda_{\text{shuf}} e^{-\lambda_{\text{shuf}}C},\tag{4}$$

where the parameter $\lambda_{shuf} = 23.3$ is determined using robust regression [27, 28]. It is not surprising that real data have higher cross-correlations than shuffled data, which is confirmed by $\lambda_{\pm} < \lambda_{shuf}$. This exponential distribution is different from the Gaussian distribution for the shuffled data of financial returns [18].

Figure 3(b) plots the distributions in double logarithmic coordinates where the negative parts are reflected to the right with respect to $C_{ij} = 0$. Nice power laws are observed spanning

over three orders of magnitude:

$$P(C) \propto \begin{cases} (-C)^{-\gamma_{-}}, & -0.01 < C \leqslant -10^{-5}, \\ C^{-\gamma_{+}}, & 10^{-5} < C \leqslant 0.01, \end{cases}$$
(5)

where $\gamma_{+} = 0.69 \pm 0.01$ and $\gamma_{-} = 0.69 \pm 0.01$ for C_{ij} , $\gamma_{+} = 0.67 \pm 0.02$ and $\gamma_{-} = 0.62 \pm 0.02$ for $C_{\text{ind,ind}}$, $\gamma_{+} = 0.67 \pm 0.02$ and $\gamma_{-} = 0.70 \pm 0.02$ for $C_{\text{ind,ins}}$ and $\gamma_{+} = 0.72 \pm 0.02$ and $\gamma_{-} = 0.73 \pm 0.02$ for $C_{\text{ins,ins}}$, respectively. It is found that $\gamma_{-} \approx \gamma_{+}$ and all the power-law exponents are close to each other. An intriguing feature is that the distributions of C_{ij} , $C_{\text{ind,ind}}$ and $C_{\text{ind,ins}}$ exhibit evident bimodal behavior, which is reminiscent of the distributions of waiting times and inter-event times of human short message communication [29]. Certainly, the underlying mechanisms are different and the factors causing the bimodal distribution of the cross-correlations are unclear.

We have tried to fit the data with the stretched exponential distribution [30, 31]. However, we do not observe significant improvement, especially in the tails. The worst fit is obtained for the 'ind-ind' case.

It is natural that we are interested more in large cross-correlations. The preceding discussions focus on cross-correlations not larger than 0.6. As shown in figure 3(a), there are pairs of inventory variation time series that have very large cross-correlations that look like outliers. For better visibility, we plot in figure 3(c) the numbers of occurrences of positive and negative cross-correlations in ten nonoverlapping intervals for the four types of pairs. It is shown that N(C > 0) > N(C < 0) in all intervals for $C = C_{ij}$, $C_{ind,ind}$ and $C_{ind,ins}$. In contrast, $N(C_{ind,ins} > 0) < N(C_{ind,ins} < 0)$ when $C_{ind,ins} < 0.6$ and $N(C_{ind,ins} > 0) > N(C_{ind,ins} < 0)$ when $C_{ind,ins} > 0.6$. Hence, for larger cross-correlations (|C| > 0.6), there are many more occurrences of positive cross-correlations than negative ones for all four types of pairs. This striking feature can be attributed to two reasons. The first is that a large proportion of investors react to the same external news in the same direction [5]: they buy following good news and sell following bad news. The second is that investors imitate the trading behavior of others of the same type and rarely imitate other investors of different types. The second reason is rational because the friends of individual (or institutional) investors are more likely individual (or institutional) investors.

3.2. The eigenvalue spectrum

For the correlation matrix C of each stock, we can calculate its eigenvalues whose density $f_c(\lambda)$ is defined as [16]

$$f_{\rm c}(\lambda) = \frac{1}{N} \frac{{\rm d}n(\lambda)}{{\rm d}\lambda},\tag{6}$$

where $n(\lambda)$ is the number of eigenvalues of *C* less than λ . If *M* is a $T \times N$ random matrix with zero mean and unit variance, $f_c(\lambda)$ is self-averaging. In particular, in the limit $N \to \infty$, $T \to \infty$ and $Q = T/N \ge 1$ fixed, the probability density function $f_c(\lambda)$ of eigenvalues λ of the random correlation matrix *M* can be described as [16, 18, 32]

$$f_c(\lambda) = \frac{Q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_{\max} - \lambda)(\lambda - \lambda_{\min})}}{\lambda},\tag{7}$$



Figure 4. Eigenvalue spectrum of the correlation matrix of inventory variation of investors trading stock 000001 within 1 day time horizon in 2003. The solid line is the spectral density obtained by shuffling independently the buyers and the sellers in such a way as to maintain the same number of purchases and sales for each investor as that in the real data. The dashed blue line shows the spectral density predicted by random matrix theory using equation (7) with Q = 237/80 = 2.96. The inset shows the largest eigenvalue λ_1 (\bigcirc) and the second largest eigenvalue λ_2 (\Box) of all 15 investigated data sets from 15 stocks. The solid line indicates the upper thresholds by shuffling experiments, and the dashed line presents the threshold predicted by random matrix theory.

with $\lambda \in [\lambda_{\min}, \lambda_{\max}]$, where λ_{\min}^{\max} is given by

$$\lambda_{\min}^{\max} = \sigma^2 (1 + 1/Q \pm 2\sqrt{1/Q}), \tag{8}$$

and σ^2 is equal to the variance of the elements of *M* [16, 32]. The variance σ^2 is equal to 1 in our normalized data.

Figure 4 illustrates the probability distribution $f_{\rm c}(\lambda)$ of the correlation matrix of inventory variation of investors trading stock 000001. The solid line is the spectral density obtained by shuffling independently the buyers and the sellers in such a way as to maintain the same number of purchases and sales for each investor as that in the real data, while the dashed blue line shows the spectral density predicted by random matrix theory using equation (7) with Q = 237/80 = 2.96. We find the largest eigenvalue is well outside the bulk and the second largest eigenvalue also escapes the bulk. The results for the other 14 stocks are quite similar. In the inset of figure 4, we plot the largest eigenvalues λ_1 and the second largest eigenvalues λ_2 for all 15 stocks. We find that all the largest eigenvalues are well above the upper thresholds determined from shuffling experiments and the thresholds λ_{max} in equation (8) predicted by random matrix theory. Moreover, all the second largest eigenvalues are above the two threshold lines and some of them are well above the thresholds. These findings indicate that both the largest and the second largest eigenvalues carry information about the investors, which is different from the results of the Spanish stock market, where only the largest eigenvalue is larger than the up thresholds while the second largest eigenvalue is within the bulk [5]. This discrepancy can be attributed to the difference in the two markets and the fact that our analysis



Figure 5. Distribution of eigenvector components *u*: (a) all the eigenvectors associated with the eigenvalues in the bulk $\lambda_{\min} < \lambda < \lambda_{\max}$ after normalization for each eigenvector for stock 000001; (b) the same as (a) but for stock 200625; (c) all the eigenvectors associated with the largest eigenvalues λ_1 after normalization for all the stocks; and (d) all the eigenvectors associated with the second largest eigenvalues λ_2 after normalization for all the stocks. The solid lines show the Gaussian distribution predicted by random matrix theory.

contains both individuals and institutions, while Lillo *et al* studied only firms. It probably implies that an RMT analysis of the Spanish data is not capable of distinguishing firms' classes. In contrast, at least two classes of Chinese investors can be identified based on λ_2 . However, it is hard to understand what the differences between the two classes of Chinese investors are.

3.3. Distribution of eigenvector components

If there is no information contained in an eigenvalue, the normalized components of its associated eigenvector should conform to a Gaussian distribution [16-18]:

$$f(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right).$$
(9)

Since the empirical eigenvalue distribution $f_c(\lambda)$ deviates from the theoretic expression (6) from random matrix theory with two large eigenvalues outside the bulk of the distribution, it is expected that the associated eigenvectors also contain certain information.

For correlation matrices of financial returns, the components of an eigenvector with the eigenvalue λ in the bulk of its distribution ($\lambda_{\min} < \lambda < \lambda_{\max}$) are distributed according to equation (9) [16–18]. Panels (a) and (b) of figure 5 show the empirical distributions of the eigenvector components *u* with the eigenvalues in the bulk for two typical stocks. Rather than analyzing the vector for one eigenvector, we normalized the components of each eigenvector and put all the eigenvectors together to gain better statistics, since each eigenvector has only 80 components. We find that the distributions of ten stocks are well consistent with the Gaussian, while the other five stocks exhibit high peaks at the center around u = 0.

IOP Institute of Physics DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

For deviating eigenvalues λ_1 and λ_2 , the distribution for each stock is very noisy and deviates from Gaussian. We treat the components of the 15 eigenvectors as a sample to have better statistics. The two distributions obtained are illustrated in figures 5(c) and (d). It is evident that both deviate from the Gaussian distribution and the distribution for λ_1 is more skewed. We have plotted the components of each eigenvector and found that they are remarkably asymmetric with respect to u = 0. The quantile–quantile plots in reference to a normal distribution further confirm that the eigenvectors are not Gaussian. We do not show those figures here.

3.4. Information in eigenvectors for deviating eigenvalues

We have shown that the largest and the second largest eigenvalues deviate from the RMT prediction and the distributions of their eigenvector components are not Gaussian. It implies that these eigenvectors carry some information. For $u(\lambda_2)$, it is not clear what kind of information they carry. We find no evident dependence of the magnitude of $u_i(\lambda_2)$ on the average absolution inventory variation $\langle |v_i| \rangle$, the total variation $\sum v_i$, or the maximal absolute variation $\max\{|v_i|\}$. The same conclusion is obtained for $u_i(\lambda_1)$, which differs from the conclusion that the eigenvector components of the return correlation matrix depend on the market capitalization in a logarithmic form [18]. For the correlation matrix of stock returns, the deviating eigenvalues except λ_1 contain information on industrial sectors [18] or stock categories [19]. It is natural to conjecture that λ_2 for inventory variations contains information on investor classification based on their trading behavior. Unfortunately, we are not able to identify clear differences in the trading behavior of investors, although we will show in the next section that the investors can be categorized into three trading types. We find no relation between the trading strategy category and the magnitude of the vector component $u(\lambda_2)$. We thus focus on extracting the information from $u(\lambda_1)$.

For the correlation matrix whose elements are the correlation coefficients of price fluctuations of two stocks, the eigenvector of the largest eigenvalue contains market information [16, 17]. The market information indicates the collective behavior of stock price movements [33], which can be unveiled by the projection of the time series on the eigenvector [18]. We follow this approach and calculate the projection G(t) of the time series $V_i(t) = [v_i(t) - \langle v_i(t) \rangle]/\sigma_i$ on the eigenvector $u(\lambda_1)$ corresponding to the first eigenvalue [5]:

$$G(t) = \sum_{i} V_i(t) \times u_i(\lambda_1)(t).$$
(10)

The projection G can be called the factor associated with the largest eigenvalue [5]. We plot the factor G(t) against the normalized return R(t) for stock 000001 in figure 6(a). There is a nice linear dependence between the two variables and a linear regression gives the slope $k = 0.83 \pm 0.04$. It indicates that these most active investors have a dominating influence on the price fluctuations. This is not surprising because the majority of the investors had very few trading activities.

Panels (b) and (c) of figure 6 illustrate the relation between the factor for individuals and institutions and the return of the stock. Linear regression gives $k_{ind} = 0.83 \pm 0.04$ and $k_{ins} = 0.41 \pm 0.06$. Comparing panels (b) and (c) with (a), we find that the influence of individuals matches excellently with the whole sample, which can be quantified by the facts that $k_{ind} = k$ and $k_{ind} < k$. The results are similar for other stocks. The resulting k_{ind} and k_{ins} are plotted in figure 6 against k for all 15 stocks. For individual investors, we find that $k_{ind} = k$ for 13 stocks and $k_{ind} > k$ for 2 stocks. In contrast, we find that $k_{ins} < k$ except for one stock.



Figure 6. Influence of the most active investors trading stock 000001 on the price fluctuations for all the investigated investors (a), for individual investors (b) and for institutional investors (c), where the slopes are $k = 0.83 \pm 0.04$, $k_{\text{ind}} = 0.83 \pm 0.04$ and $k_{\text{ins}} = 0.41 \pm 0.06$, respectively. Panel (d) plots k_{ind} and k_{ins} against k, where each symbol corresponds to a stock.

An interesting question is to ask the lagged correlation between returns R and dominant mode G. In particular, one would be interested in the potential predictability of G on the returns. The factor G(t) has predictive power if it is significantly correlated with R(t+1). This test cannot be done using the data in figure 6 because the G(t) time series contains information on the time period investigated. Indeed, the analysis performed above is in a descriptive manner, as done in factor analysis in econometrics [34]. A crude analysis by plotting R(t+1) against G(t)does not uncover any significant correlations. A proper way is to determine the G(t) time series in moving windows. However, we do not proceed due to the fact that we have only the data over one year. Nevertheless, we argue that this can be done when longer data sets are available and this idea can be adopted in other cases when the records are sufficiently long.

4. Inventory variation and stock return

4.1. Categorization of investors

Following Lillo *et al* [5], we divide the investors into three categories according to the crosscorrelation coefficient C_{V_iR} between the inventory variation V_i and the stock return R. The investor *i* belongs to the trending or reversing category if its inventory variation is positively or negatively correlated with the return. We use a wieldy significant threshold to categorize the investors:

$$\pm 2\sigma = \pm 2/\sqrt{N_T},\tag{11}$$

where N_T is the number of time records for each time series [5]. We also verify the robustness of equation (11) by comparing the experimental results with the results of a null hypothesis



12

Figure 7. Panels (a)–(c) show the scatter plots of C_{VR} versus a proxy of the size of the investor. For each stock, the proxy is the ratio of the value exchanged by the investor (scaled by a factor of 10⁴) to the capitalization of the stock. Each marker refers to an investor trading a specific stock. The three kinds of markers refer to investors whose inventory variations are positively correlated (\bigcirc), negatively correlated (\square) or uncorrelated (\triangle) with returns according to the block bootstrap analysis. The two dashed lines indicate the 2σ threshold calculated using equation (11). Panels (d)–(f) are contour plots of the correlation matrix of daily inventory variation of investors trading the stock 000001. We have sorted the investors into rows and columns according to their cross-correlation coefficients of inventory variation with its price return C_{VR} . The evolution of C_{VR} in the same order as in the matrix is shown in the bottom panel, where the dashed lines bound the $\pm 2\sigma$ significance intervals.

based on a block bootstrap of both R and V. In this regard, 1000 block bootstrap replicas with a block length of 20 are performed. For each investor, we have checked whether the estimated correlation with return exceeds the 0.977 25 quantile or is smaller than the 0.022 75 quantile of the correlation distribution obtained from bootstrap replicas. The results are shown in figures 7(a)–(c). There are 1211 investors in the whole sample in figure 7(a), including 453 institutional investors in figure 7(b) and 758 individual investors in figure 7(c).

As shown in the last row of table 1, the numbers of the three kinds of investors (trending, reversing and uncategorized) are 46, 34 and 373 for institutional investors and 41, 381 and 336 for individual investors, respectively. We find that most institutional investors are uncategorized and there are more trending investors than reversing investors. These results are different from the Spanish case, where only one-third investors are uncategorized and the number of reversing firm investors is about thrice as much as the number of trending firm investors [5]. In contrast, about half of the individuals are reversing investors and only 6% of the individuals are trending investors. The observation that most investors are uncategorized is probably due to the fact that the Chinese market was emerging and its investors are not experienced. Comparing individuals

and institutions, we find a larger proportion of individuals exhibiting reversing behavior. This indicates that these individuals buy when the price drops and sell when the price rises on the same day. This finding is very interesting since it explains the worse performance in stock markets [10, 11].

The empirical evidence for the significant cross-correlation between inventory variation $V_i(t)$ of trending and reversing investors and stock return R(t) leads us to adopt a linear model for the dynamics of inventory variation as a first approximation [5]:

$$V_i(t) = \gamma_i R(t) + \epsilon_i, \tag{12}$$

where γ_i is proportional to the cross-correlation coefficient $C_{V_i,R}$. It follows immediately that the cross-correlation coefficient between the inventory variations of two investors is

$$C_{ij} = C_{V_i, V_i} = \gamma_i \gamma_j. \tag{13}$$

If two investors belong to the same category, either trending ($\gamma_i > 2\sigma$ and $\gamma_j > 2\sigma$, significantly) or reversing ($\gamma_i < -2\sigma$ and $\gamma_j < -2\sigma$, significantly), the value of C_{ij} is expected to be significantly positive. In contrast, if two investors belong, respectively, to the trending and reversing categories, the value of C_{ij} is expected to be significantly negative. To show the performance of the model, we plot the contours of the correlation matrix of inventory variation for all investors, for individuals and for institutions, where the investors are sorted according to their cross-correlation coefficients C_{VR} of the inventory variation with the price return. Figure 7(d) shows that the left-top corner gives large positive C_{ij} values and the left-bottom and right-top corners give large negative C_{ij} values, as expected. Figure 7(e) gives better results for individual investors, validating the linear model (12). The results in figure 7(f) are worse for institutional investors, which is due to the fact that most C_{VR} values are small for institutions, as illustrated in figure 7(c). However, figure 7(f) does not invalidate the linear model, since there are only three trending institutions and two reversing institutions. Indeed, the situation is quite similar for other individual stocks with very few investors of the same category, as shown in table 1.

4.2. Causality

In section 4.1, we have shown that the inventory variation $V_i(t)$ and the stock return R(t) have significant positive or negative correlation for part of the investors. It is interesting to investigate the lead-lag structure between these two variables. For the largest majority of reversing and trending firms in the Spanish stock market, it is found that returns Granger cause inventory variation but not vice versa at the day or intraday level, and the Granger causality disappears over longer time intervals [5]. Here, we aim to study the same topic for both individual and institutional investors.

We first investigate the autocorrelation function $C_{V(t)V(t+\tau)}$ of the inventory variation time series sampled in 15 min time intervals. Figure 8(a) shows the three autocorrelation functions for all the trending, reversing and uncategorized investors. Each autocorrelation function is obtained by averaging the autocorrelation functions of the investors in the same category to have better statistics. It is found that the inventory variation is long-term correlated and the correlation is significant over dozens of minutes and the correlation is stronger for trending investors than reversing investors, which can be partly explained by the order splitting behavior of large investors [5, 35–38]. We also find that uncategorized investors have stronger autocorrelations than trending and reversing investors, which is different from the Spanish case [5]. However, it



Figure 8. The first column (a, d, g, j) shows the results for all investors. The second column (b, e, h, k) shows the results for all individual investors. The third column (c, f, i, l) shows the results for all institutional investors. (a–c) Averaged autocorrelation functions $C_{V(t)V(t+\tau)}$ of the 15 min inventory variation V for trending, reversing and uncategorized investors. The dashed lines give the 5% significance level. (d)–(f) Averaged lagged cross-correlation functions $C_{V(t)R(t+\tau)}$ of the 15 min inventory variation for trending, reversing and uncategorized investors. The dashed lines bound the $\pm 2\sigma$ significance interval. (g)–(i) Conditional expected value of the indicator $I(x \rightarrow y)$ of the rejection of the null hypothesis of non-Granger causality between x and y with 95% confidence as a function of time horizons ΔT . The dashed lines show the 5% significance level. (j)–(l) Conditional expected value of the indicator $I(x \rightarrow y)$ as a function of the simultaneous cross-correlation $C[V_i(t), R(t)]$. The black symbols refer to the Granger test on shuffled data and the dashed lines bound $\pm 2\sigma$ significance interval.

is not clear why uncategorized investors have stronger autocorrelations than trending investors. Figures 8(b) and (c) illustrate the results for individuals and institutions. We observe that institutions have stronger long memory than individuals. This implies that institutions are more persistent in adopting their trading strategies while individuals are more likely to change their strategies.

Panels (d)–(f) of figure 8 illustrate the averaged lagged cross-correlation functions $C_{V(t)R(t+\tau)}$ between inventory variations and returns. The results in the three panels are qualitatively the same. For uncategorized investors, no significant cross-correlations are found between inventory variations and returns, which is trivial due to the 'definition' of this category, as shown in figures 7(a)–(c). For trending and reversing investors, it is evident that the returns lead the inventory variations by dozens of minutes ($\tau < 0$), where the cross-correlation $C_{V(t)R(t+\tau)}$ is significantly nonzero. When the price drops, trending investors will sell stock shares to reduce their inventory in a few minutes, while reversing investors will buy shares to increase their inventory. When the price rises, trending investors will buy shares to increase their inventory. In the meanwhile, we also observe nonzero cross-correlations for $\tau > 0$ in shorter time periods, which means that the inventory variations lead returns.

To further explore the lead-lag structure between inventory variations and returns, we perform Granger causality analysis. We define an indicator $I(X \to Y)$, whose value is 1 if X Granger causes Y and 0 otherwise [5]. In our analysis, the time resolution of the two time series is 15 min. The values of $I(V \to R)$ and $I(R \to V)$ for all investors are determined at different time scales ΔT . The average indicator values $E[I(V \to R)]$ and $E[I(R \to V)]$ are plotted in figures 8(g)-(i) with respect to ΔT for all investors, individual investors and institutional investors. Both $I(V \to R)$ and $I(R \to V)$ are decreasing functions of ΔT . We note that $E[I(X \to Y)]$ is the percentage of investors with $I(X \to Y) = 1$. Figure 8 shows that there are more investors with $I(R \to V) = 1$ than investors with $I(V \to R) = 1$. On average, bidirectional Granger causality is observed at the intraday time scales and the Granger causality disappears at the weekly level or longer. Moreover, individual investors are more probable to be influenced by the intraday price fluctuations than institutions, because the $I(R \to V)$ values of individuals are greater than those of institutions at the same time scale level.

We then investigate the impact of investor category on the causality indicator. The results for $\Delta T = 4$ (i.e. 1 h) are depicted in figures 8(j)–(l). The middle parts bounded by two vertical lines at $C_{VR} = \pm \sigma$ correspond to uncategorized investors. The left parts $(C_{VR} < -\sigma)$ correspond to reversing investors and the right parts $C_{VR} > \pm \sigma$ correspond to trending investors. It is found that an investor adjusts his inventory following price fluctuations with very large probabilities when his $|C_{VR}|$ value is large. This conclusion holds for both individual and institutional investors. The strong Granger causality from inventory variations cannot be attributed to the non-Gaussianity in the distributions of the variables, as verified by bootstrapping analysis. The effect that inventory variations Granger cause return is stronger for institutional investors. Qualitatively similar results are obtained for other ΔT values. The observation that inventory variations seem to also Granger causing inventory variations have temporary or permanent price impacts [39–41]. We note that the Granger causality from inventory variations have temporary or permanent price impacts [39–41].

ons,	resp	ecuv	ery. I	i ne i	ime i	IONZ	on is	one	uay.			
Reversing					Trer	nding		Uncategorized				
$n_{\rm d}^+$	$n_{\rm d}^-$	$n_{\rm s}^+$	$n_{\rm s}^-$	$n_{\rm d}^+$	$n_{\rm d}^-$	$n_{\rm s}^+$	$n_{\rm s}^-$	$n_{\rm d}^+$	$n_{\rm d}^-$	$n_{\rm s}^+$	$n_{\rm s}^-$	
60	63	0	0	0	0	0	0	6	8	0	0	
23	19	0	0	0	0	0	0	3	6	0	0	
26	37	0	0	0	0	0	0	2	1	0	0	
54	62	0	0	0	0	0	0	4	3	1	0	
34	34	2	1	0	0	0	0	0	0	1	1	
0	0	0	0	0	0	0	0	2	1	0	0	
17	34	0	0	0	0	0	0	0	0	2	1	
21	15	0	0	3	7	0	0	5	2	0	0	
29	25	0	0	0	0	0	0	7	7	0	0	
34	31	0	0	0	0	0	0	0	0	0	0	
61	58	0	0	0	0	0	0	7	7	0	0	
38	36	0	0	0	0	0	0	0	1	0	0	
55	40	0	0	0	0	0	0	3	4	0	0	
31	24	0	0	0	0	0	0	2	7	0	3	
6	6	0	0	0	0	0	0	3	6	0	3	
	$ \begin{array}{r} \overline{n_d^+} \\ \overline{n_d^+} \\ \overline{60} \\ 23 \\ 26 \\ 54 \\ 326 \\ 54 \\ 34 \\ 0 \\ 17 \\ 21 \\ 29 \\ 34 \\ 61 \\ 38 \\ 55 \\ 31 \\ 6 $	$\begin{tabular}{ c c c c c } \hline & Revel \\ \hline n_d^+ & n_d^- \\ \hline n_d^- & n_d^- \\ \hline$	$\begin{tabular}{ c c c c c c } \hline Reversing \\ \hline $n_{\rm d}^+$ $n_{\rm d}^-$ $n_{\rm s}^+$ \\ \hline $n_{\rm d}^-$ $n_{\rm s}^-$ \\ \hline $n_{\rm d}^-$ \\ \hline $n_{\rm d}^-$ $n_{\rm d}^-$ \\ \hline $n_{\rm d}^$	$\begin{tabular}{ c c c c c c c c c c c c c c c c c c c$	Reversing n_{d}^{+} n_{d}^{-} n_{s}^{+} n_{s}^{-} n_{d}^{+} 60 63 0 0 0 23 19 0 0 23 19 0 0 0 26 37 0 0 26 37 0	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	Reversing Trending n_{d}^{+} n_{d}^{-} n_{s}^{+} n_{s}^{-} n_{d}^{+} n_{d}^{-} n_{s}^{+} n_{s}^{-} n_{d}^{+} n_{d}^{-} n_{s}^{+} n_{s}^{-} n_{d}^{+} n_{d}^{-} n_{s}^{+} n_{s}^{-} n_{s}^{+} $n_{s}^$	Reversing Trending n_d^+ $n_d^ n_s^+$ $n_s^ n_d^+$ $n_d^ n_s^+$ n_s^- 60 63 0 0 0 0 0 0 23 19 0 0 0 0 0 0 0 26 37 0 0 0 0 0 0 0 26 37 0	Reversing Trending U n_d^+ $n_d^ n_s^+$ $n_s^ n_d^+$ $n_d^ n_s^+$ $n_s^ u_d^+$ 60 63 0 0 0 0 0 0 6 23 19 0 0 0 0 0 0 3 26 37 0 0 0 0 0 2 54 62 0 0 0 0 0 2 17 34 0 0 0 0 0 0 2 15 0 3 7 0 0 0 0 29 25 0 0 0 0 0 0 31 0 0 0 0 0 0 0 55 40 0 0 0 0 0 3 31 24 0 0 <	Reversing Trending Uncates n_d^+ $n_d^ n_s^+$ $n_s^ n_d^+$ $n_d^ n_s^+$ $n_s^ u_{n_s^+}$ $n_{n_s^-}^ u_{n_s^+}$ $u_{n_s^-}^ u_{n_s^+}^ u_{n_s^+}^ u_{n_s^+}^ u_{n_s^+}^ u_{n_s^+}^ u_{n_s^+}$	Reversing Trending Uncategorize n_d^+ $n_d^ n_s^+$ $n_d^ n_s^+$ $n_s^ u_{0}^+$ $u_{0}^ u_{0}^+$ $u_{0}^ u_{0}^+$ u_{0}^+ $u_{0}^ u_{0}^+$ u_{0}^+	

Table 2. The number of herding days for different groups of investors. The total number of trading days is 237. The superscripts '+' and '-' indicate buy herding and sell herding, respectively. The subscripts 'd' and 's' indicate individuals and institutions, respectively. The time horizon is one day.

by the fact that the Chinese stock market was most inefficient as a young emerging market and the investors were less skilled and more irrational.

4.3. Herding behavior

Herding and positive feedbacks are essential for the boom of bubbles [42, 43]. These topics have been studied extensively to understand the price formation process [44–47]. Herding is a phenomenon where a group of investors are trading in the same direction over a period of time. Here, we try to investigate possible herding behavior in different groups of investors.

We study possible buy and sell herding behavior among groups of the same investors. Investors are classified into different groups based on their types (individual or institution) and their categories (reversing, trending or uncategorized). We define a herding index as follows [5]:

$$h = \frac{N^+}{N^+ + N^-},$$
(14)

where N^+ is the number of buying investors and N^- is the number of selling investors in the same group over a given time horizon. When the herding index *h* is smaller than 5% under a binomial null hypothesis, we assess that herding is present. In our analysis, we fix the time horizon to one day and determine the number of days that herding was present for different groups of investors. The results are depicted in table 2.

According to table 2, there are no buying and selling herding days observed for trending institutions. For trending individuals and reversing institutions, herding is observed in only one

stock on very few days. For categorized investors, we see slightly more herding days in a few stocks. For reversing investors, the number of herding days is greater than that for other investors and we observe comparable buying and selling herding days. Our findings are consistent with those for the Spanish stock market, especially in the sense that reversing investors are more likely to herd [5]. Our analysis also allows us to conclude that individuals are more likely to herd than institutions in 2003.

5. Summary

In summary, we have studied the dynamics of investors' inventory variations. Our data set contains 15 stocks actively traded on the Shenzhen Stock Exchange in 2003 and the investors can be identified as either individuals or institutions.

We studied the cross-correlation matrix C_{ij} of inventory variations of the most active individual and institutional investors. It is found that the distribution of cross-correlation coefficient C_{ij} is asymmetric and has a power-law form in the bulk and exponential tails. The inventory variations exhibit stronger correlation when both investors are either individuals or institutions, which indicates that the trading behavior is more similar within investors of the same type. The eigenvalue spectrum shows that the largest and the second largest eigenvalues of the correlation matrix cannot be explained by the random matrix theory and the components of the first eigenvector $u(\lambda_1)$ carry information about stock price fluctuation. In this respect, the behavior differs for individual and institutional investors.

Based on the contemporaneous cross-correlation coefficients C_{VR} between inventory variations and stock returns, we classified investors into three categories: trending investors who buy (sell) when stock price rises (falls), reversing investors who sell (buy) when stock price rises (falls), and uncategorized investors. We also observed that stock returns predict inventory variations. It is interesting to find that about 56% of individuals hold trending or reversing strategies and only 18% of institutions hold strategies. Moreover, there are far more reversing individuals (50%) than trending individuals (6%). In contrast, there are slightly more trending institutions (10%) than reversing institutions (8%). Hence, Chinese individual investors are prone to selling winning stocks and buying losing stocks, which provides supporting evidence that trading is hazardous to the wealth of individuals [9, 10].

In this work, we have focused on the zero-lag cross-correlations among inventory variations. It is possible to perform further investigations. One possibility is to study the correlation matrix constructed from time-lagged cross-correlations of the inventory variations as in [48–50]. In the Granger causality analysis, the lagged cross-correlations between inventory variation and return have already been considered. For inventory variation and return and their magnitudes, when longer time series become available, we can perform detrended cross-correlation analysis [51] and multifractal cross-correlation analysis [52, 53] and determine their statistical significance [54]. Another possibility is to research the absolute values of the inventory variations, which is a natural idea as the pair of return and volatility.

Acknowledgments

We are grateful to Fabrizio Lillo and Rosario N Mantegna for fruitful discussion and anonymous referees for insightful suggestions. This work was partially supported by the National Natural Science Foundation of China (grant numbers 10905023, 11075054 and 71131007), the Shanghai Rising Star Program (grant number 11QH1400800), the Shanghai Chenguang Program (grant number 2010CG32) and the Fundamental Research Funds for the Central Universities.

References

- [1] Bouchaud J P 2008 Nature 455 1181
- [2] Lux T and Westerhoff F 2009 Nature Phys. 5 2-3
- [3] Farmer J D and Foley D 2009 Nature 460 685-6
- [4] Schweitzer F, Fagiolo G, Sornette D, Vega-Redondo F, Vespignani A and White D R 2009 Science 325 422-5
- [5] Lillo F, Moro E, Vaglica G and Mantegna R N 2008 New J. Phys. 10 043019
- [6] Griffin J, Harris J H and Topaloglu S 2003 J. Finance 58 2285-320
- [7] Choe H, Kho B C and Stulz R M 1999 J. Financial Econ. 54 227-64
- [8] Grinblatt M and Keloharju M 2000 J. Financial Econ. 55 43-67
- [9] Odean T 1998 J. Financ. 53 1775–98
- [10] Barber B M and Odean T 2000 J. Finance 55 773-806
- [11] Barber B M, Lee Y, Liu Y and Odean T 2009 *Rev. Finance Stud.* 22 609–32
- [12] Tumminello M, Lillo F, Piilo J and Mantegna R N 2012 New J. Phys. 14 013041
- [13] Tumminello M, Miccichè S, Lillo F, Varho J, Piilo J and Mantegna R N 2011 J. Stat. Mech. 2011 P01019
- [14] Tumminello M, Miccichè S, Lillo F, Piilo J and Mantegna R N 2011 PLoS One 6 e17994
- [15] Tumminello M, Miccichè S, Dominguez L J, Lamura G, Melchiorre M G, Barbagallo M and Mantegna R N 2011 PLoS One 6 e23377
- [16] Laloux L, Cizeau P, Bouchaud J P and Potters M 1999 Phys. Rev. Lett. 83 1467-70
- [17] Plerou V, Gopikrishnan P, Rosenow B, Amaral L A N and Stanley H E 1999 Phys. Rev. Lett. 83 1471-4
- [18] Plerou V, Gopikrishnan P, Rosenow B, Amaral L A N, Guhr T and Stanley H E 2002 Phys. Rev. E 65 066126
- [19] Shen J and Zheng B 2009 Europhys. Lett. 86 48005
- [20] Jiang Z Q and Zhou W X 2010 Physica A 389 4929-41
- [21] Wang J J and Zhou S G 2011 Physica A 390 398-409
- [22] Sun X Q, Cheng X Q, Shen H W and Wang Z Y 2011 Physica A 390 3427-34
- [23] Wang J J, Zhou S G and Guan J H 2012 Neurocomputing 92 44-53
- [24] Forrester P J, Snaith N C and Verbaarschot J J M 2011 J. Phys. A: Math. Theor. 36 R1-10
- [25] Gu G F, Chen W and Zhou W X 2008 Physica A 387 3173-82
- [26] Gu G F, Ren F, Ni X H, Chen W and Zhou W X 2010 Physica A 389 278-86
- [27] Holland P W and Welsch R E 1977 Commun. Stat. Theory Methods 9 813-27
- [28] Street J O, Carroll R J and Ruppert D 1988 Ann. Stat. 42 152-4
- [29] Wu Y, Zhou C S, Xiao J H, Kurths J and Schellnhuber H J 2010 Proc. Natl Acad. Sci. USA 107 18803-8
- [30] Laherrere J and Sornette D 1998 Eur. Phys. J. B 2 525-39
- [31] Ivanov P C, Yuen A, Podobnik B and Lee Y K 2004 Phys. Rev. E 69 056107
- [32] Sengupta A M and Mitra P P 1999 Phys. Rev. E 60 3389–92
- [33] Plerou V, Gopikrishnan P, Rosenow B, Amaral L A N and Stanley H E 2001 Physica A 299 175-80
- [34] Jolliffe I T 2002 Principal Component Analysis (New York: Springer)
- [35] Vaglica G, Lillo F, Moro E and Mantegna R N 2008 Phys. Rev. E 77 036110
- [36] Moro E, Vicente J, Moyano L G, Gerig A, Farmer J D, Vaglica G, Lillo F and Mantegna R N 2009 Phys. Rev. E 80 066102
- [37] Vaglica G, Lillo F and Mantegna R N 2010 New J. Phys. 12 075031
- [38] Toth B, Lemperiere Y, Deremble C, de Lataillade J, Kockelkoren J and Bouchaud J P 2011 *Phys. Rev.* X 1 021006
- [39] Bouchaud J P, Gefen Y, Potters M and Wyart M 2004 Quant. Finance 4 176-90
- [40] Farmer J D, Gerig A, Lillo F and Mike S 2006 Quant. Finance 6 107-12

New Journal of Physics 14 (2012) 093025 (http://www.njp.org/)

18

IOP Institute of Physics **D**EUTSCHE PHYSIKALISCHE GESELLSCHAFT

- [41] Bouchaud J P, Kockelkoren J and Potters M 2006 Quant. Finance 6 115-23
- [42] Sornette D 2003 Why Stock Markets Crash (Princeton, NJ: Princeton University Press)
- [43] Sornette D 2003 Phys. Rep. 378 1–98
- [44] Lakonishok J, Shleifer A and Vishny R W 1992 J. Financial Econ. 32 23-43
- [45] Wermers R 1999 J. Finance 54 581-622
- [46] Nofsinger J R and Sias R W 1999 J. Finance 54 2263–95
- [47] Sias R W 2004 Rev. Financial Studies 17 165–206
- [48] Potters M, Bouchaud J P and Laloux L 2005 Acta Phys. Pol. B 36 2767-84
- [49] Podobnik B, Wang D, Horvatic D, Grosse I and Stanley H E 2010 Europhys. Lett. 90 68001
- [50] Wang D, Podobnik B, Horvatic D and Stanley H E 2011 Phys. Rev. E 83 046121
- [51] Podobnik B and Stanley H E 2008 Phys. Rev. Lett. 100 084102
- [52] Zhou W X 2008 Phys. Rev. E 77 066211
- [53] Jiang Z Q and Zhou W X 2011 Phys. Rev. E 84 016106
- [54] Podobnik B, Jiang Z Q, Zhou W X and Stanley H E 2011 Phys. Rev. E 84 066118

19