Observatory of
Complex Systems

SANTA FE INSTITUTE
Celebrating 20 years of Complexity Science

Empirical market ecology studies II

Fabrizio Lillo

University of Palermo (Italy) and Santa Fe Institute (USA)

in collaboration with
Rosario Mantegna, Gabriella Vaglica (Univ. of Palermo)
Esteban Moro (Univ. Carlos III of Madrid)

Santa Fe, July 29, 2008

- The dynamics of many socioeconomic systems is determined by the decision making process of agents.

- The decision process depends on agent’s characteristics, such as preferences, risk aversion, behavioral biases, etc.

- In addition, in some systems the size of agents can be highly heterogeneous leading to very different impacts of agents on the system dynamics.

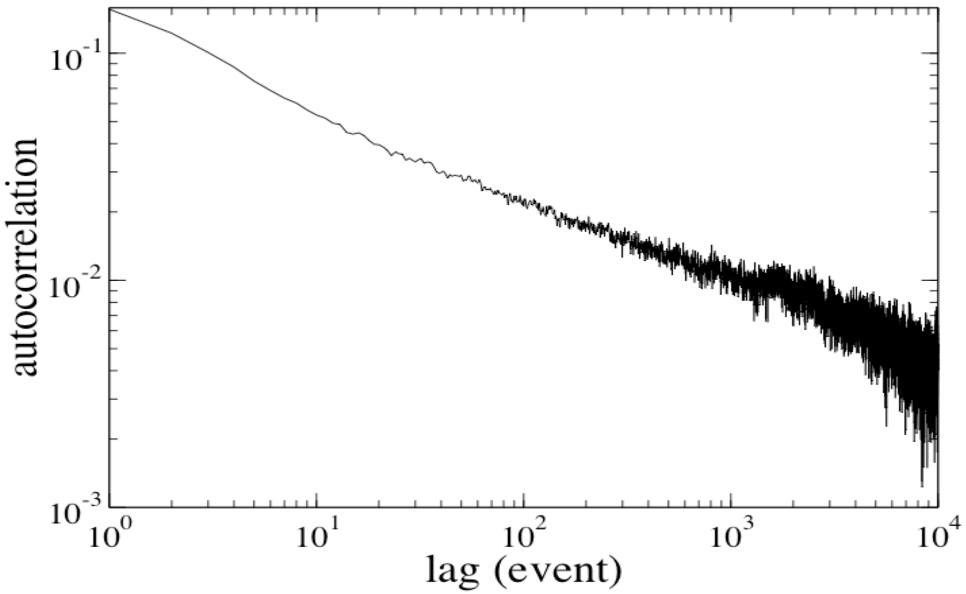
- The large size of some agents poses challenging problems to agents who want to control their impact, either by forcing the system in a given direction or by hiding their intentionality.

- In financial markets many participants (e.g. mutual funds) have a large size
- When these participants decide to trade, for example to rebalance their portfolio, the volume they want to trade is usually very large.
- These volumes are usually several orders of magnitude larger than the volume available at the best or even in the whole book at any time
- Financial markets are in a state of latent liquidity, meaning that the displayed liquidity is a tiny fraction of the true (hidden) liquidity supplied/demanded
- Delayed market clearing: traders are forced to split their large orders in pieces traded incrementally as the liquidity becomes available
- Market participants forms a kind of ecology, where different species act on different sides of liquidity demand/supply and trade with very different time scales

We want to find empirical evidence of these long-lasting orders at a market level and we want to characterize their statistical properties

Evidence for long-lasting orders

- We consider the symbolic time series obtained by replacing buyer initiated trades with +1 and seller initiated trades with -1



$$\rho(\tau) \sim \tau^{-0.5}$$

Market order sign is a long memory process

Lillo and Farmer 2004, Bouchaud et al. 2004

The long memory of order sign is statistically robust and persists across several days

This fact raises two questions:

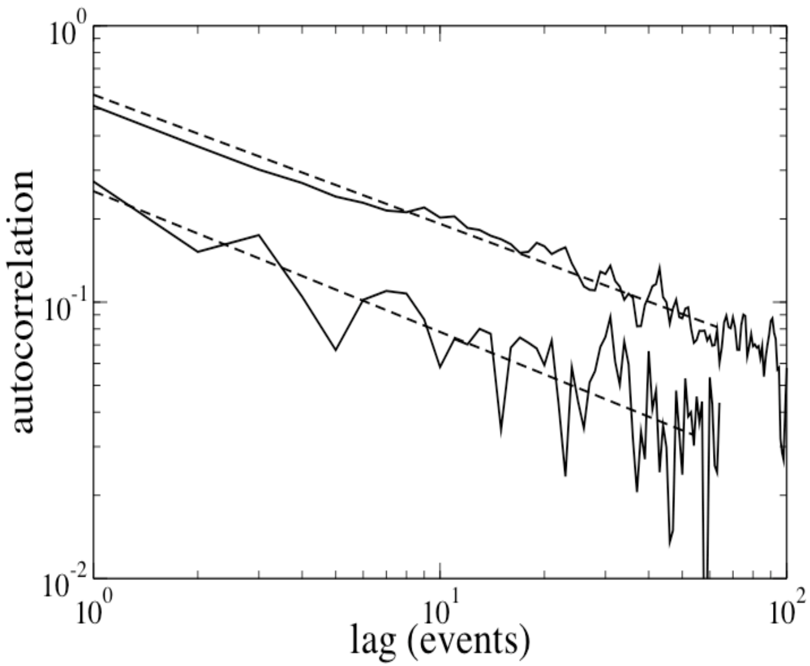
- What is origin of the long-memory property of order flow ?
- How can the market be efficient in the presence of long memory of order sign ?
  – See Austin's talk

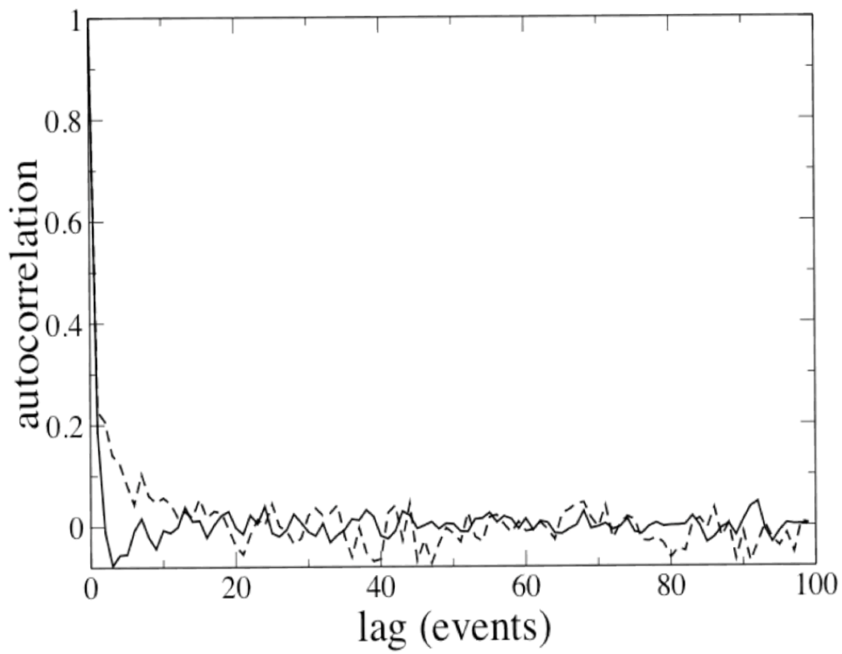
What is the origin of the long memory?

- Two alternative explanations
  - Long memory is a property of the order flow of each investor, independent of the behavior of other investors (Lillo, Mike, and Farmer, 2005)
  - Long memory arises because investors herd in their trading though an imitation process that involves an interaction between them (LeBaron and Yamamoto, 2008)

Individual institutions

- We analyze the behavior of individual broker codes trading a stock

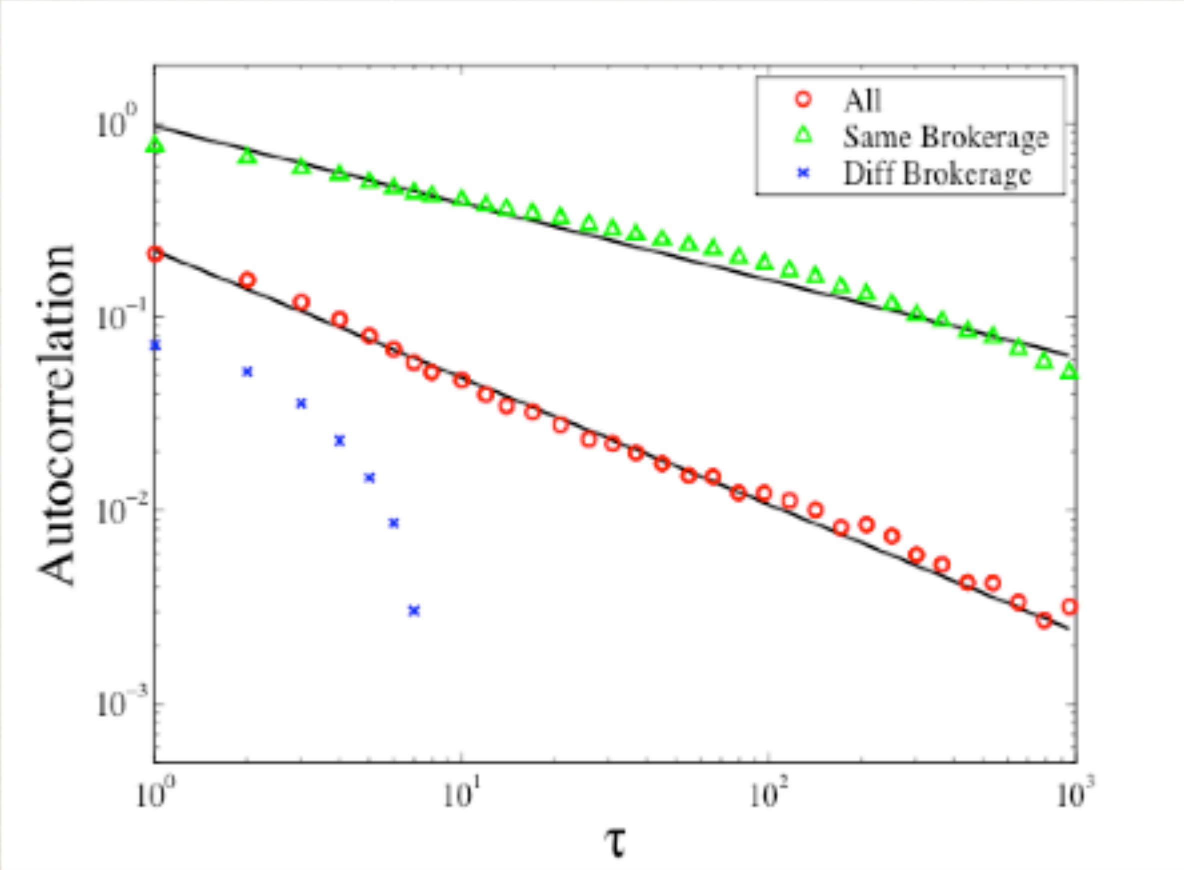




Heterogeneity probably related to different trading strategies

Autocorrelation of volume sign (vs. transaction number)

Autocorrelation of trade signs
for same vs. different broker codes



A theory for the origin of long memory

- We hypothesize that the cause of long memory in order signs is delay in market clearing
- Suppose that a large investor decides to buy 10,000,000 shares. It is likely that she decides to split the order and to trade the order incrementally over an extended period of time.
- We postulate that unrevealed hidden orders are broken up into pieces, which we call revealed orders

Model of order splitting

- We assume that the distribution of initial hidden order size is a Pareto distribution

$$p(L) = \frac{\alpha}{L^{\alpha+1}} \qquad L \geq 1 \qquad \alpha > 1$$

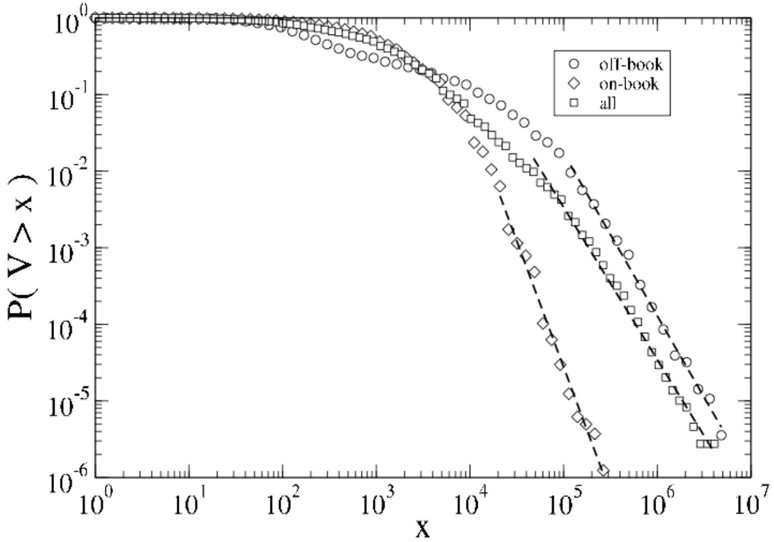
(Pareto distribution of mutual fund size? See Yoni's talk)

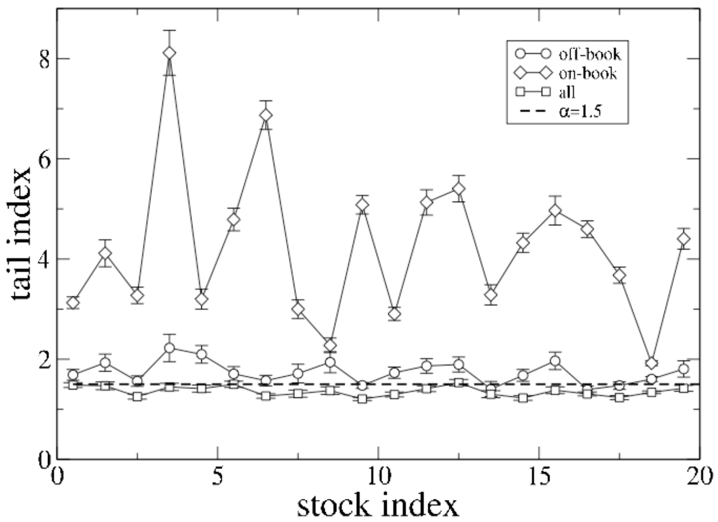
We prove that for our model the time series of the signs of the revealed order has an autocorrelation function decaying asymptotically as

$$\rho(\tau) \sim \frac{N^{\alpha-2}}{\alpha} \frac{1}{\tau^{\alpha-1}} \longrightarrow \gamma = \alpha - 1$$

Lillo, Mike, and Farmer 2005

Testing the model





- The volume of on-book and off-book trades have different statistical properties.
- The exponent α=1.5 for the hidden order size and the market order sign autocorrelation exponent γ are consistent with the order splitting model which predicts the relation $\gamma=\alpha-1$.

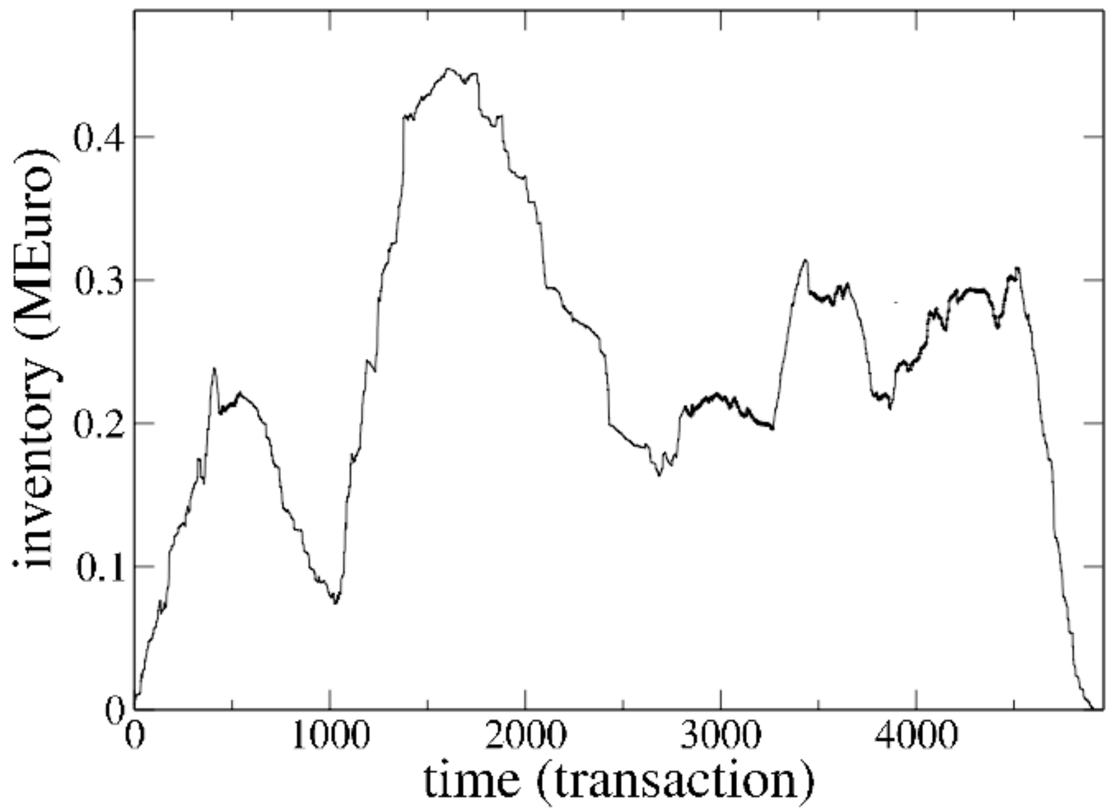
Lillo, Mike, and Farmer 2005

Detecting hidden orders

- In financial markets large investors usually need to trade large quantities that can significantly affect prices. The associated cost is called <u>market impact</u>
- For this reason large investors refrain from revealing their demand or supply and they typically trade their large orders incrementally over an extended period of time.
- These large orders are called packages or <u>hidden orders</u> and are split in smaller trades as the result of a complex optimization procedure which takes into account the investor's preference, risk aversion, investment horizon, etc..

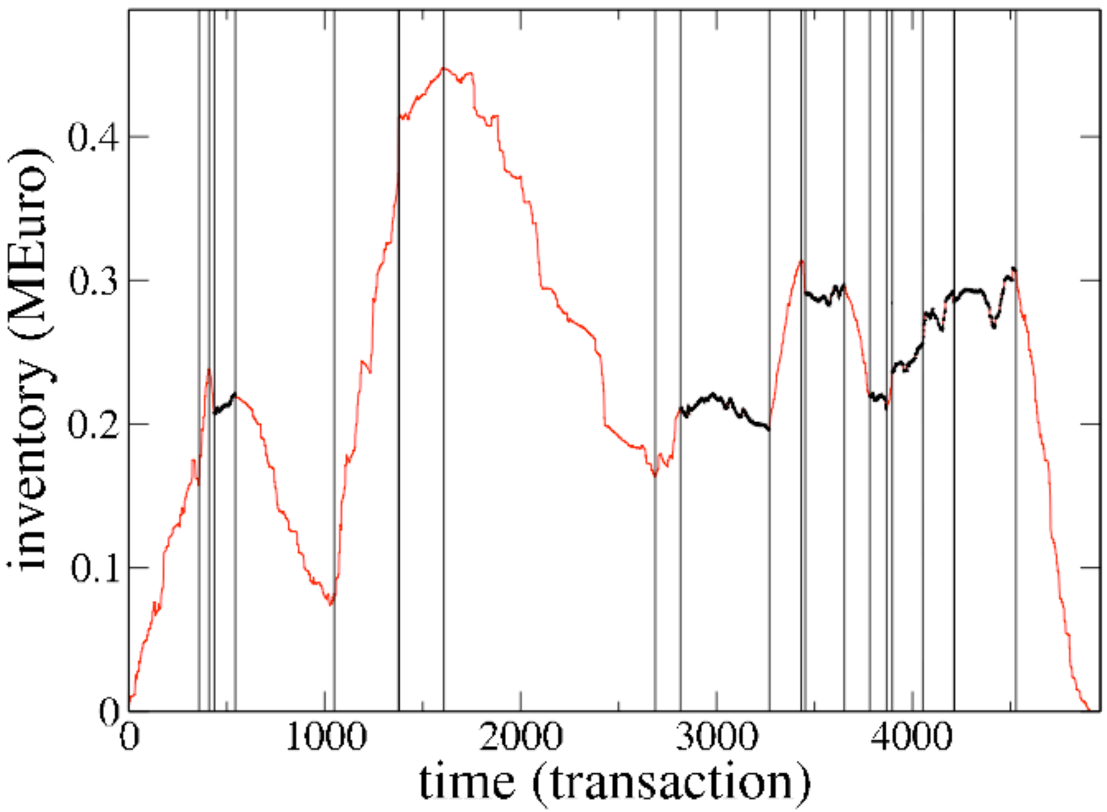
- We want to detect empirically the presence of hidden orders from the trading profile of the investors
- We make use of the market member data in the Spanish Stock Market
- Since market members are not individual investors, we have to use statistical detection methods

A typical inventory profile



Credit Agricole trading Santander

Detecting hidden orders



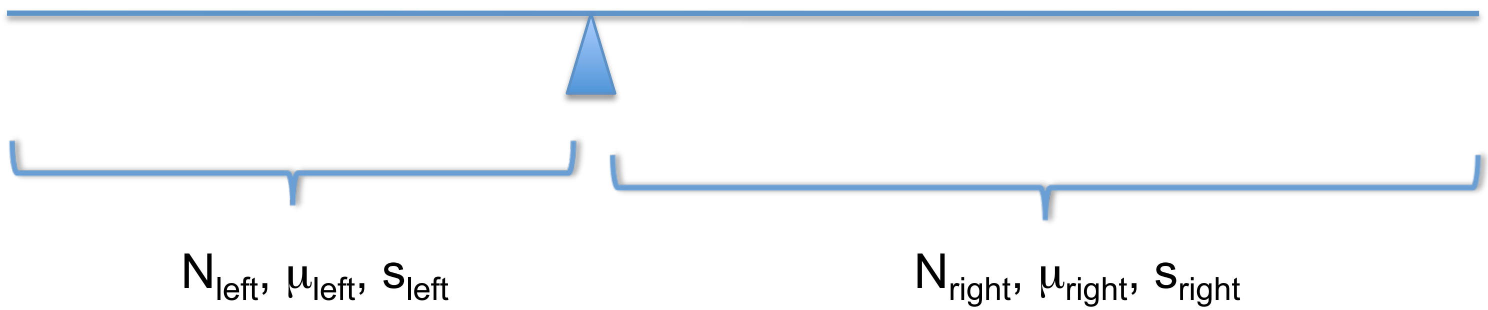
We developed a statistical method to identify periods of time when an investor was consistently (buying or selling) at a constant rate

Working hypothesis: the detected patches are hidden orders

Vaglica, Lillo, Moro, and Mantegna 2008

Segmentation algorithm

If the series is composed of many segments with different mean values, the segmentation maximizes the difference in the mean values between adjacent segments (adapted from Bernaola-Galvan et al 2001).



- The algorithm search for the position where the t statistics $(\mu_{left}-\mu_{right})/s_D$ is maximal
-A t (modified) test is performed and and the sequence is cut if the probability is higher than a predefined threshold (in our case 99%)
- If the cut is accepted the procedure continues recursively on the left and right subsets created by each cut.
- Before a new cut is accepted one also computes t between the right-hand new segment and its right neighbor (created by a previous cut) and t between the left-hand new segment and its left neighbor (created by a previous cut) and one checks if both values of t continue to be statistically signficant according to the selected threshold.

Directional hidden orders

- We are interested in <u>directional hidden orders</u>, i.e. hidden orders with a well defined direction to buy or to sell.
- We call buy hidden orders when $V_{buy}/V_{tot} > 0.75$ and sell hidden order when $V_{sell}/V_{tot} > 0.75$. For buy hidden orders $V_m = V_{buy}$ and $N_m = N_{buy}$, and analogously for sell hidden orders.

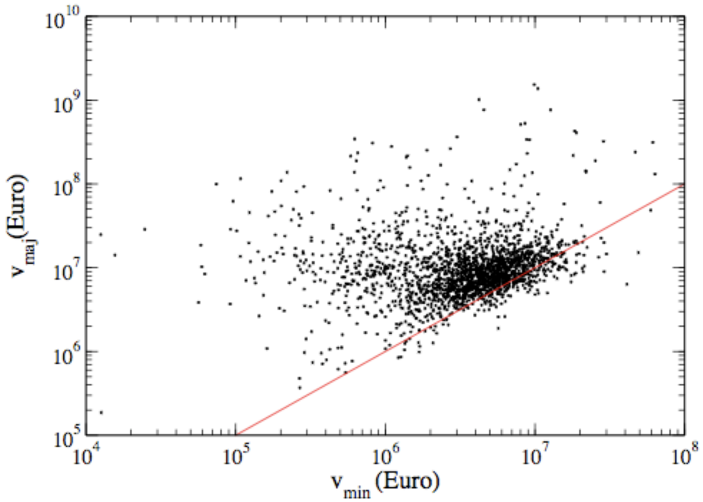


Figure 3.9: Scatter plot of the mean exchanged value per transaction considering transaction in the same direction as the total trend of the patch (v_{maj}) and those in the opposite direction (v_{min}). The figure is relative just to those directional patches with $\theta = 0.75$ for which v_{min} is defined, i.e. patches which $N_{min} \neq 0$. Patches are detected through Segmentation algorithm. The figure refers to the stock Telefónica during 2001-2004. Red line. Curve of equation $Y = X$. Almost all empirical points stay above the red line, indicating that on average the value v_{maj} is greater than N_{min} for directional patches.

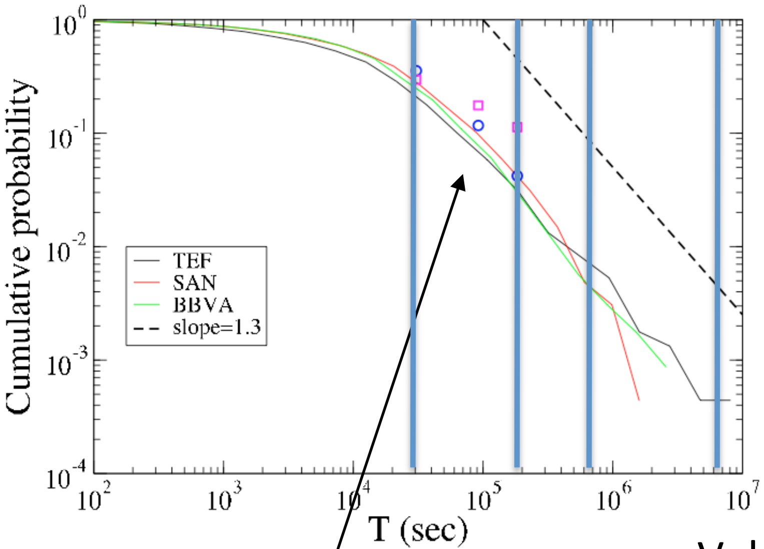
Investigated variables

- Each hidden order is characterized by three quantities
  1) The time duration T of the hidden order (in seconds)
  2) The number of transactions N_m characterizing the direction of the order (in number of trades)
  3) The volume V_m characterizing the direction of the hidden order (in Euros)

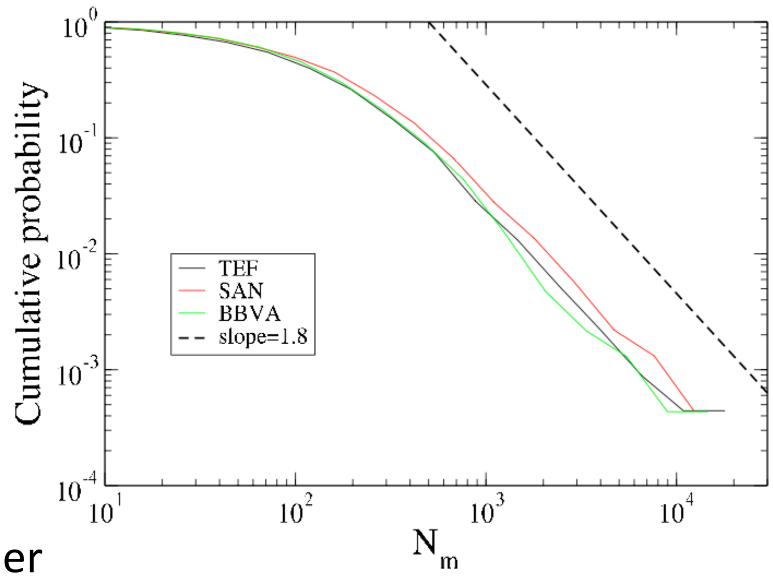
- We measure the distributional properties of these variables and their mutual dependencies

Distributional properties of hidden orders

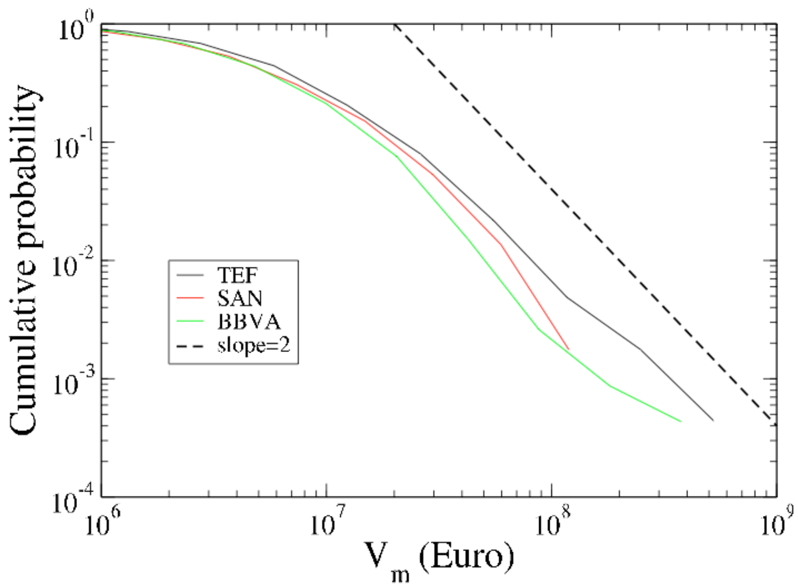
Investment horizon



Number of transactions



Volume of the order



Circles and squares are data taken from Chan and Lakonishok at NYSE (1995) and Gallagher and Looi at Australian Stock Exchange (2006)

Large hidden orders

The distributions of large hidden orders sizes are characterized by power law tails.

	BBVA	(2104)	SAN	(2086)	TEF	(2062)
$\zeta_{V_{maj}}$	2.3	(1.9; 2.7)	2.0	(1.7; 2.3)	1.9	(1.6; 2.2)
$\zeta_{N_{maj}}$	2.0	(1.7; 2.3)	1.7	(1.4; 2.0)	1.7	(1.4; 2.0)
ζ_T	1.5	(1.3; 1.7)	1.5	(1.3; 1.7)	1.2	(1.0; 1.4)

Table 4.1: Tail exponents of the distribution of T, N_{maj}, and V_{maj} estimated with the Hill estimator (or Maximum Likelihood Estimator). In parenthesis we report the 95% confidence interval. The number in parenthesis nearby the tick symbol is the number of patches detected for the considered stock.

Power law heterogeneity of investor typical (time or volume) scale

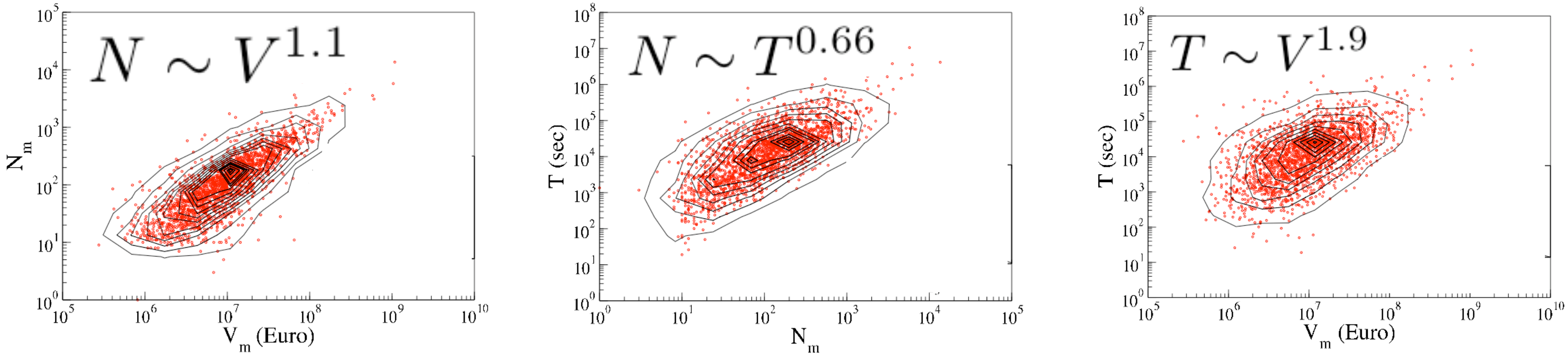
$$P(T > x) \sim \frac{1}{x^{1.3}} \quad P(N_m > x) \sim \frac{1}{x^{1.8}} \quad P(V_m > x) \sim \frac{1}{x^2}$$

These results are not consistent with the theory of Gabaix et al. (Nature 2003)

$$P(T > x) \sim \frac{1}{x^3} \qquad P(N_m > x) \sim \frac{1}{x^3} \qquad P(V_m > x) \sim \frac{1}{x^{3/2}}$$

Allometric relations of hidden orders

We measure the relation between the variables characterizing hidden orders by performing a Principal Component Analysis to the logarithm of variables.



	BBVA (2104)		SAN (2086)		TEF (2062)	
g_1	1.08	(1.05 ; 1.12)	1.06	(1.01 ; 1.10)	1.07	(1.04 ; 1.11)
g_2	1.81	(1.69 ; 1.93)	1.81	(1.68 ; 1.94)	2.00	(1.88 ; 2.14)
g_3	0.68	(0.65 ; 0.71)	0.68	(0.65 ; 0.70)	0.62	(0.59 ; 0.64)

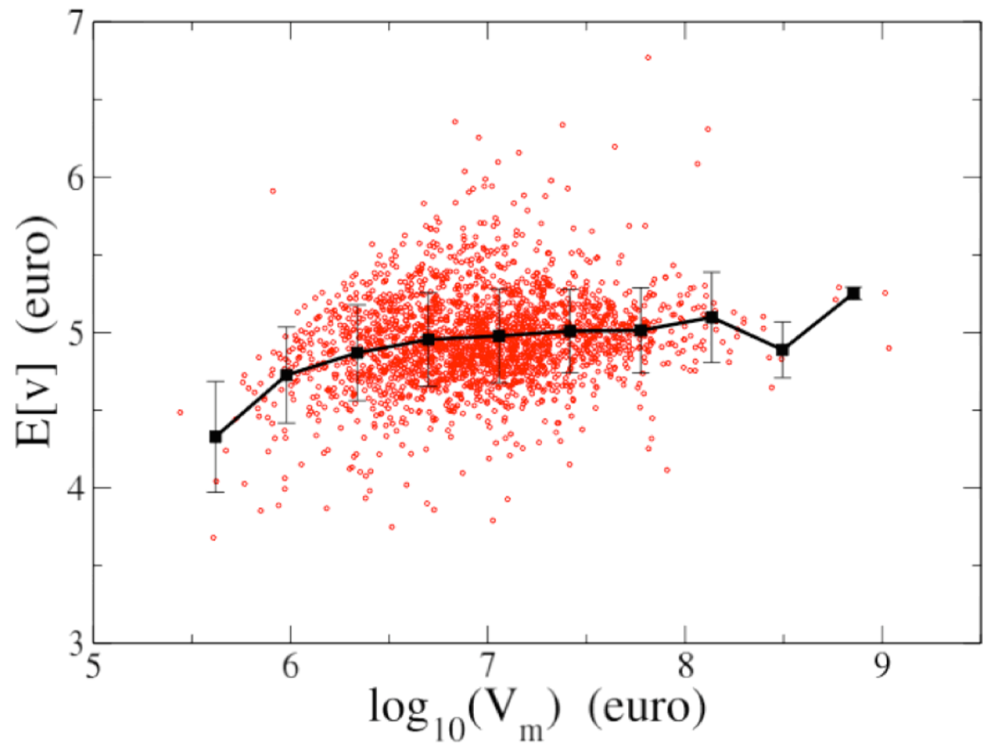
Table 4.3: Exponents of the allometric relations defined in Eq. 4.7. The exponents are estimated with PCA and the errors are estimated with bootstrap. In parenthesis we report the 95% confidence interval. The number in parenthesis nearby the tick symbol is the number of patches detected for the considered stock.

Comments

- The almost linear relation between N and V indicates that traders do not increase the transaction size above the available liquidity at the best (see also Farmer et al 2004)
- For the N_m-V_m and the T-N_m relations the fraction of variance explained by the first principal value is pretty high
- For the T-V_m relation the fraction of variance explained by the first principal value is smaller, probably indicating an heterogeneity in the level of aggressiveness of the firm.
- Also in this case our exponents (1.9, 0.66, 1.1) are quite different from the one predicted by Gabaix et al theory (1/2, 1, 1/2)

On the relation between N and V

The mean transaction volume weakly depends on the hidden order volume



This is consistent with the “selective liquidity taking” mechanism proposed in Farmer and al (2004) to explain the shape of the one trade impact. Traders condition the market order volume to be smaller or equal to the volume at the opposite best.

Role of agents heterogeneity

- We have obtained the distributional properties and the allometric relations of the variables characterizing hidden orders by pooling together all the investigated firms
- Are these results an effect of the aggregation of firms or do they hold also at the level of individual firm?

Heterogeneity and power law tails

- For each firm with at least 10 detected hidden orders we performed a Jarque-Bera test of the lognormality of the distribution of T, N_m, and V_m

	BBVA	SAN	TEF
T	75 (15/20)	63 (17/27)	77 (24/31)
N_m	90 (18/20)	100 (27/27)	100 (31/31)
V_m	90 (18/20)	100 (27/27)	94 (29/31)

- For the vast majority of the firms we cannot reject the hypothesis of lognormality
- The power law tails of hidden order distributions is mainly due to firms (size?) heterogeneity

Individual firms

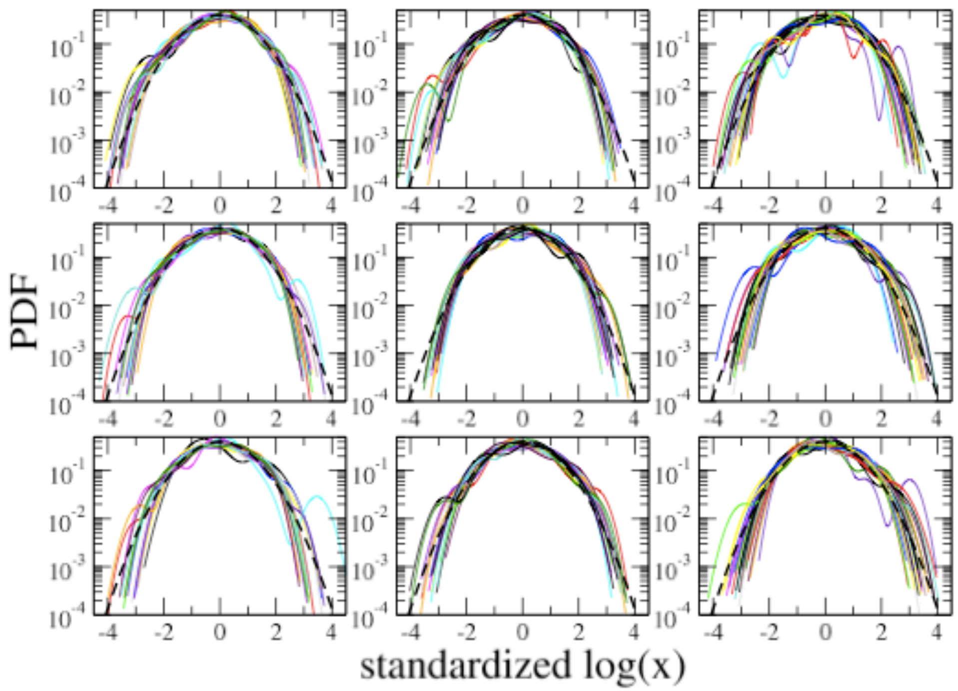


Figure 4.5: Probability density function of the standardized logarithm of the variables T, N_{maj} and V_{maj} of the firms for which the Jarque-Bera test of lognormality cannot be rejected. Specifically, for each stock and each variable we consider the firms for which the lognormal hypothesis cannot be rejected (see Table 4.2). For each of these firms we compute the logarithm of the variable, we subtract the mean value and divide by the standard deviation. According to the null hypothesis these normalized variables should be Gaussian distributed. In the figure we plot in a semi-log scale the probability density functions for each firm (continuous lines) and we compare them with the Gaussian probability density function (dashed line). Each column refers to a firm (from left to right, BBVA, SAN, TEF) and each row refers to a variable (from top to bottom T, N_{maj} and V_{maj}).

Heterogeneity and allometric relations

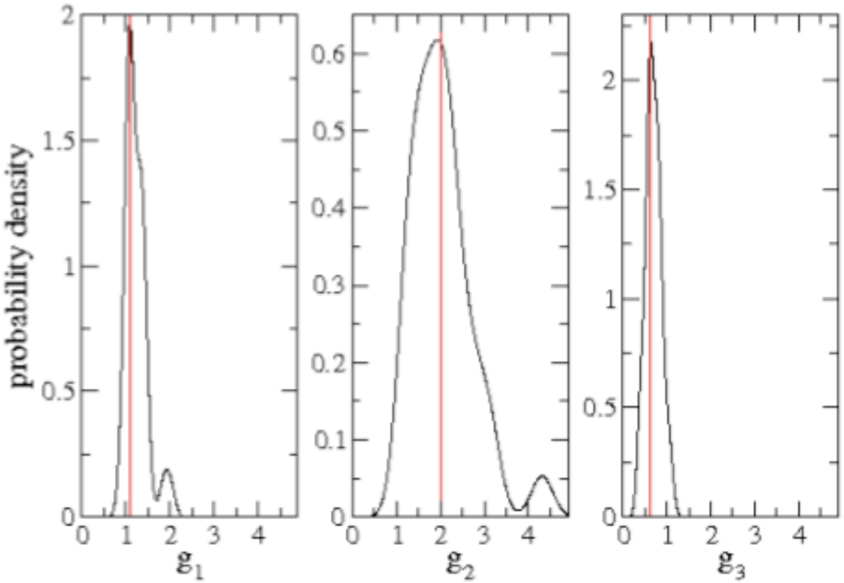


Figure 4.9: Probability density function of the three exponents g_1, g_2, and g_3 describing the allometric relations of Eq. 4.7 computed on the patches of individual firms with at least 10 patches. The stock investigated is Telefónica. The red vertical lines indicates the values of the scaling exponents computed in the pool of all firms and reported in Table 4.3. It is worth noting that the dispersion of g_2 is significantly larger than the one for the other two exponents.

- The scaling exponent between variables are quite consistent across different firms
- The exponent with the broadest distribution is the one relating T and V_m
- The allometric relations are not the effect of traders heterogeneity

3D Principal Component Analysis

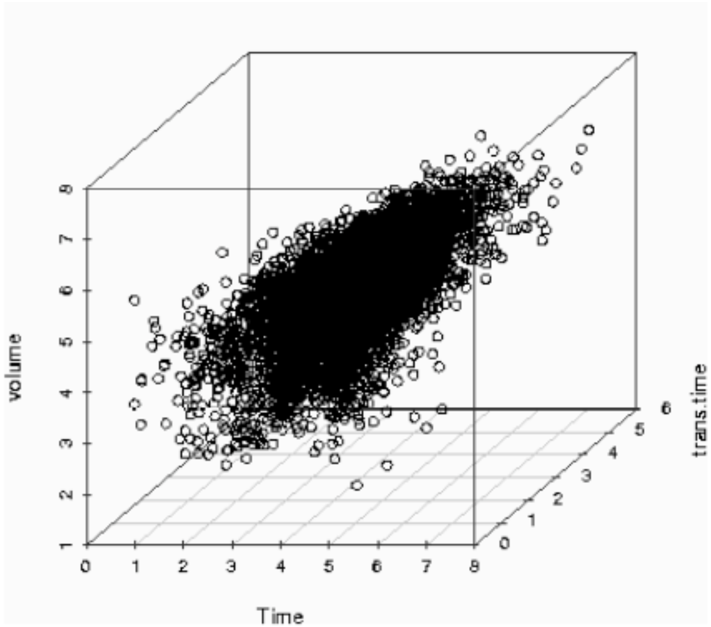


Figure 4.8: Scatter plot of the three variables ($\log T, \log N_{maj}, \log V_{maj}$). The considered stock is Telefónica and the period is 2001-2004.

- The eigenvalues explain 81%, 15%, and 4% of the variance
- The first direction is a size variable
- The points lie on a 2D manifold
- The scaling exponents are consistent (but different) from the ones obtained with the 2D PCA

Table 4.4: Approximate value of percentage of firms with at least 10 patches for which the 3D PCA gives the first (λ_1) or second (λ_2) eigenvalue larger than the corresponding first or second eigenvalue obtained from the PCA of the pool of all the patches. The numbers in parenthesis are the absolute numbers of firms.

	BBVA (2104)	SAN (2086)	TEF (2062)
λ_1	90 (18/20)	85.2 (23/27)	87.1 (27/31)
λ_2	15 (3/20)	18.5 (5/27)	22.6 (7/31)

- Hidden orders of an individual firm are essentially explained by one size variable.
- The second eigenvalue of the pool, is mainly due to the heterogeneity between firms.

Conclusions

- Order flow is a long memory process
- The origin is delayed market clearing and hidden orders
- Hidden order size is very broadly distributed
- Heterogeneity of market participants plays a key role in explaining fat tails of hidden order size
- Scaling relations between the variables characterizing the hidden orders are more universal and are probably related to some universal optimization scheme

Open questions

- Are these statistical regularities observed also in other markets? (Javier and the LSE)
- How much are our results dependent on the specific algorithm for identifying hidden orders ?
- Is there any difference between hidden orders made mainly by market orders and hidden orders made mainly of limit orders?
- What is the market impact function of hidden orders?
- What is the origin of broadly distributed hidden order sizes? How is it related to the distribution of the size of investors?
- What is the optimization which is able to account for the allometric relations between the variables characterizing the hidden orders?