The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific and
Cultural Organization

IAEA
International Atomic Energy Agency

2229-4

School and Workshop on Market Microstructure: Design, Efficiency and Statistical Regularities

21 - 25 March 2011

Financial Market Micostructure: Empirical Studies

Fabrizio LILLO
Scuola Normale Superiore
Piazza dei Cavalieri
PISA
ITALY

Strada Costiera 11, 34151 Trieste, Italy - Tel.+39 040 2240 111; Fax +39 040 224 163 - sci_info@ictp.it

Observatory of
Complex Systems

SCUOLA NORMALE SUPERIORE PISA

SANTA FE INSTITUTE
Celebrating 20 years of Complexity Science

(Financial) Market microstructure: empirical studies

Fabrizio Lillo

Scuola Normale Superiore di Pisa and University of Palermo (Italy)

Santa Fe Institute (USA)

ICTP, Trieste, March 21-22, 2011

Outline

- Financial markets as model complex systems
- Efficient Market Hypothesis
- Financial data
- Stylized facts of aggregate financial variables
- Market microstructure
- Market impact
- Order flow
- Heterogeneity of investors in a financial market

Financial markets as complex systems

A financial market can be described as a `model' complex system.




In a financial market there are many agents interacting to perform the collective task of finding the best price for a financial asset.

There are many different types of financial markets

- Stock exchanges (New York, London, Tokyo)
- Foreign exchange markets (Global market)
- Derivative markets (Chicago, New York, Paris)
- Bond markets (London)
- Commodities
- …..

Financial market as a model complex system

- The study of financial markets has an obvious importance on his own.
- However I believe that financial markets are an ideal model system to study the interaction of many individuals taking decisions under risk. The system is ideal because
    - It is an extremely competitive environment where the fitness of an investor can often be identified with her ability of generating profit
    - The interaction mechanism is clearly defined
    - The availability of very detailed and large datasets (down to individual behavior) allows to perform careful empirical analyses
    - In some cases the flow of external information can be identified and monitored (news stream, financial analyst's forecasts, etc)

Quantitative approach to financial markets

Roughly speaking two types of approaches are possible in the study of financial markets, and, more generally, of social systems.

- Assume that agents in the system have a given amount (homogeneous or heterogeneous) of rationality. The process of price formation is based on the decision making of agents.
- Make use of a more pragmatic approach consisting in analyzing the dynamical properties of financial variables looking for statistical regularities

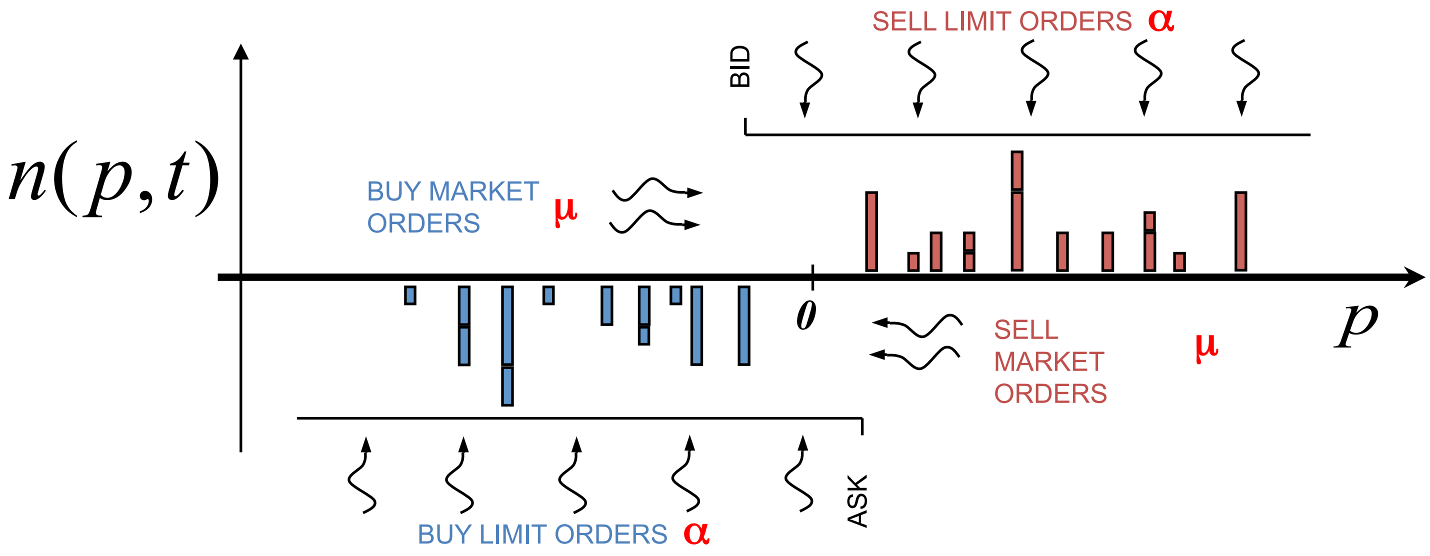
Perfect rationality

- The standard approach in financial economics consists in assuming that agents in the market have perfect rationality and perfect knowledge of other agents' preferences
- In this idealized market it is possible to investigate the conditions allowing an equilibrium between supply and demand
- Moreover the model is (sometimes) able to make falsifiable prediction on the behavior of aggregate quantities, such as price (e.g. Capital Asset Pricing Model)

Bounded rationality

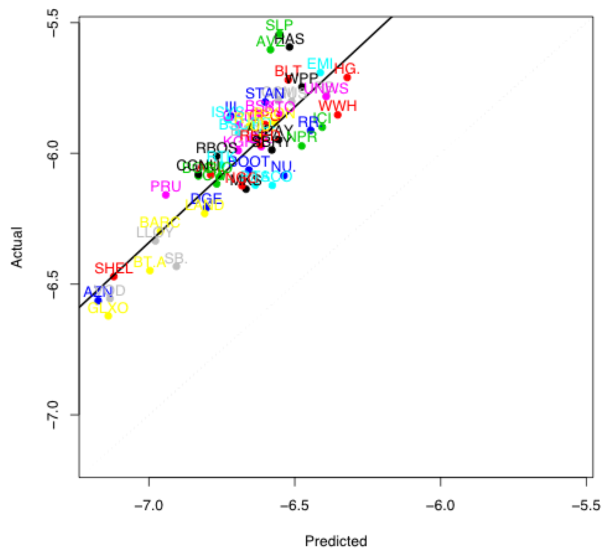
- In recent years there has been an increasing interest of scientific community on models of bounded rationality (H. Simon), i.e. models where agents have only limited cognitive and computational abilities.
- An extreme approach consists in assuming that agents have zero intelligence, i.e. they act randomly. Surprisingly some empirical facts can be explained by this kind of model, proving the importance of interaction rules

Zero-intelligence model



Volatility = $$\frac{\mu^{5/2}\delta^{1/2}\sigma^{-1/2}}{\alpha^2}$$

Strategy or structure?



Farmer et al., PNAS 2005

Efficient Market Hypothesis

Samuelson (1965) stated that in an informationally efficient market, price changes must be unforecastable if they fully incorporate the expectations and information of all market participants

"The market is said to be efficient with respect to some information set if prices would be unaffected by revealing that information to all participants. Moreover, efficiency with respect to an information set implies that it is impossible to make economic profits by trading on the basis of that information set"

(Malkiel, 1992)

Efficient Market Hypothesis

- Therefore, under the efficient market hypothesis, **price changes must be unforecastable**
- The more efficient the market, the more random is the sequence of price changes
- A widespread model of price that incorporates the efficiency of the market is the **Random Walk Hypothesis**

Random Walk Hypothesis

- The Efficient Market Hypothesis suggests that a good framework to describe price dynamics is continuous or discrete time stochastic processes
- The price must be described by a martingale

$$E(P_{t+1}|P_t, P_{t-1}, ...) = P_t$$

i.e. the best forecast of tomorrow's price is simply today's price

The attempts to model the price of a financial asset as a stochastic process go back to the 1900 pioneering work of Louis Bachelier

The simplest model for price dynamics in discrete time is

$$p_t = \mu + p_{t-1} + \epsilon_t$$

where $p_t = \log P_t$, μ is a constant, and ϵ_t is a noise term consistent with the Efficient Market Hypothesis

Random walk hypothesis

Depending on the properties of ϵ_t, we distinguish

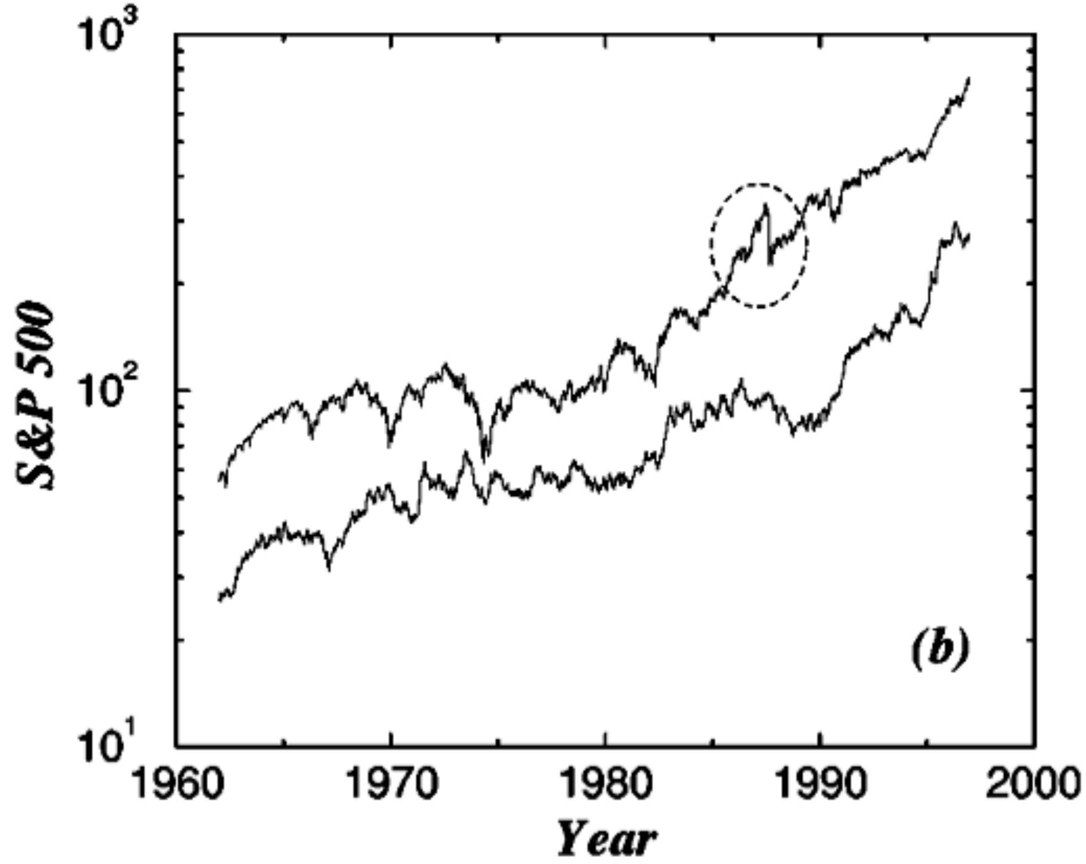
- Independent Identically Distributed increments: for example Gaussian distributed (stable laws)
- Independent increments
- Uncorrelated increments: the weaker form implies the vanishing of the linear autocorrelation

In continuous time the **geometric Brownian motion** is considered the simplest random process describing the price dynamics of a financial asset.

$$dP(t) = \mu P(t)dt + \sigma P(t)dW$$

This equation is used as one of the fundamental assumptions of the so-called Black and Scholes (B&S) model. The B&S
model allows to obtain the rational price for a simple financial contract (an European option) issued on an underlying fluctuating financial asset.

Random walk hypothesis



Ideal vs real

An idealized model of stock market where the stock price dynamics is described by a geometric Brownian motion exists and provides the theoretical foundation for quantitative finance.

What do real data say?

Financial data

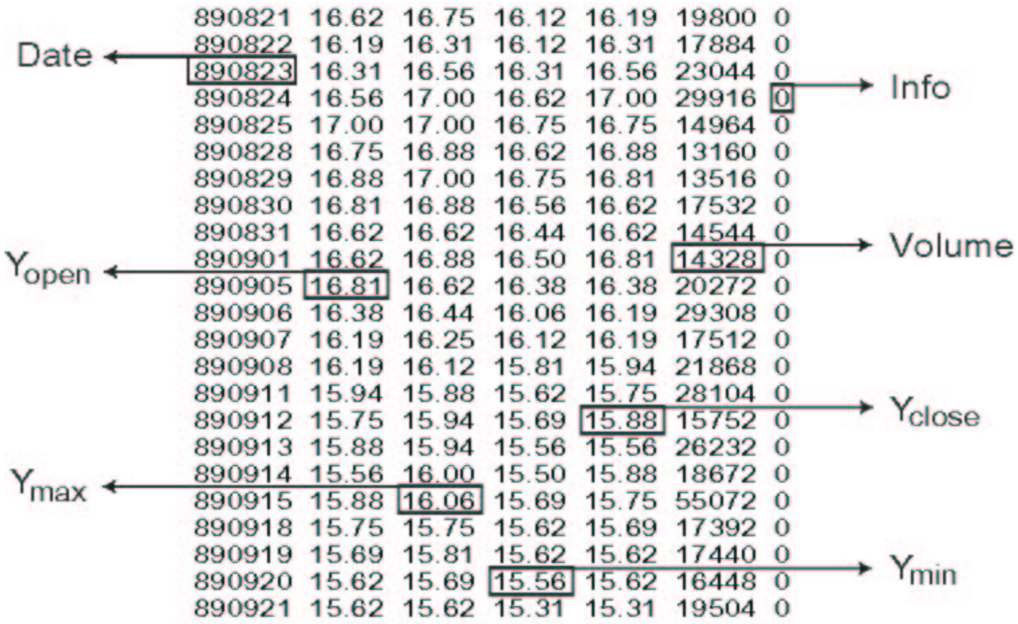
- Data are essential for the development and testing of scientific theories
- In the last years social sciences have experienced a transition from a low rate of data production to an high rate of data production
- This is due to the availability of datasets combining an high resolution and a large size
- New levels of resolution raise new issues in data handling, visualization, and analysis

Financial data

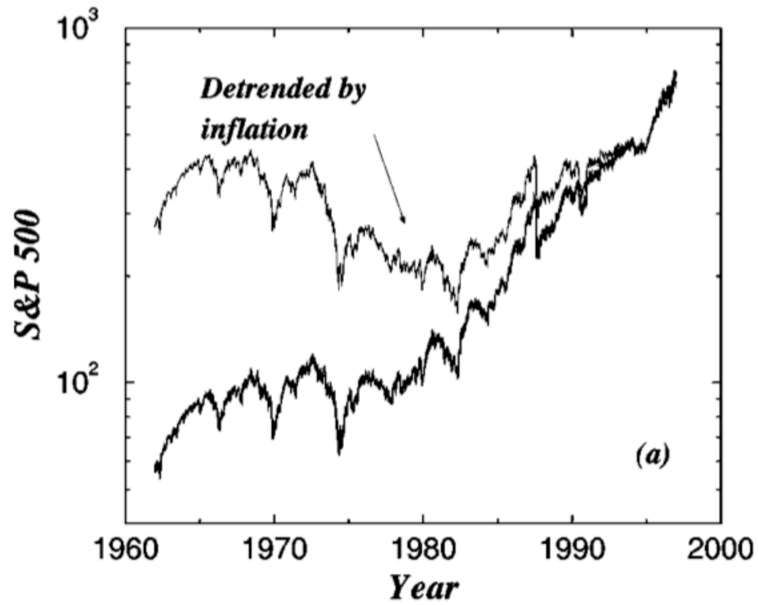
- In the last thirty years the degree of resolution of financial data has increased
  - Daily data
  - Tick by tick data
  - Order book data
  - “Agent” resolved data

Daily data

- Daily financial data are available at least since nineteenth century
- Usually these data contains opening, closing, high, and low price in the day together with the daily volume
- Standard time series methods to investigate these data



890821	16.62	16.75	16.12	16.19	19800	0
890822	16.19	16.31	16.12	16.31	17884	0
890823	16.31	16.56	16.31	16.56	23044	0
890824	16.56	17.00	16.62	17.00	29916	0
890825	17.00	17.00	16.75	16.75	14964	0
890828	16.75	16.88	16.62	16.88	13160	0
890829	16.88	17.00	16.75	16.81	13516	0
890830	16.81	16.88	16.56	16.62	17532	0
890831	16.62	16.62	16.44	16.62	14544	0
890901	16.62	16.88	16.50	16.81	14328	0
890905	16.81	16.62	16.38	16.38	20272	0
890906	16.38	16.44	16.06	16.19	29308	0
890907	16.19	16.25	16.12	16.19	17512	0
890908	16.19	16.12	15.81	15.94	21868	0
890911	15.94	15.88	15.62	15.75	28104	0
890912	15.75	15.94	15.69	15.88	15752	0
890913	15.88	15.94	15.56	15.56	26232	0
890914	15.56	16.00	15.50	15.88	18672	0
890915	15.88	16.06	15.69	15.75	55072	0
890918	15.75	15.75	15.62	15.69	17392	0
890919	15.69	15.81	15.62	15.62	17440	0
890920	15.62	15.69	15.56	15.62	16448	0
890921	15.62	15.62	15.31	15.31	19504	0



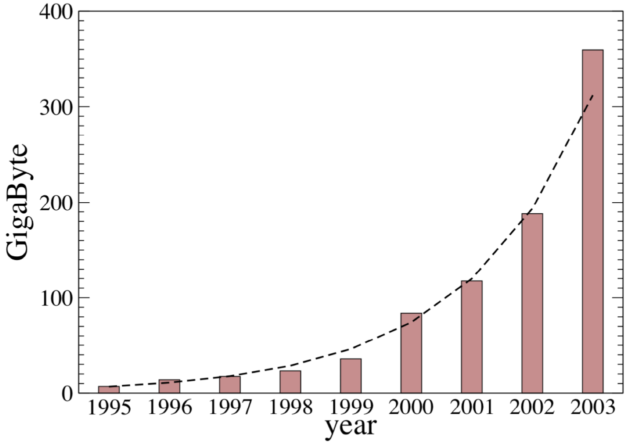
(from Gopikrishnan et al 1999)

Tick by tick data

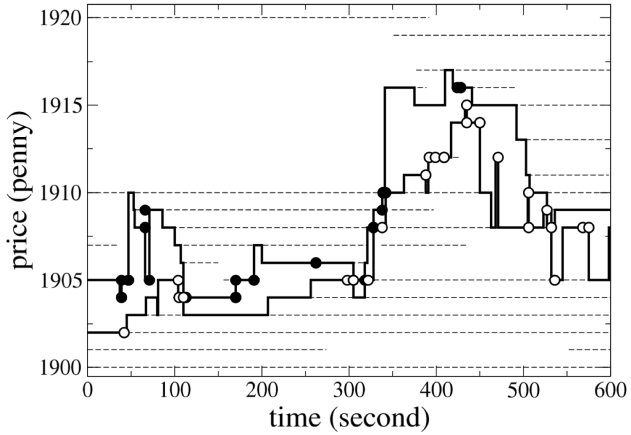
- Financial high frequency data usually refer to data sampled at a time horizon smaller than the trading day
- The usage of such data in finance dates back to the eighties of the last century
  - Berkeley Option Data (CBOE)
  - TORQ database (NYSE)
  - HFDF93 by Olsen and Associates (FX)
  - CFTC (Futures)

- Higher resolution means new problems
  - data size: example of a year of a LSE stock
    - 12kB (daily data)
    - 15MB (tick by tick data)
    - 100MB (order book data)
  - irregular temporal spacing of events
  - the discreteness of the financial variables under investigation
  - problems related to proper definition of financial variables
  - intraday patterns
  - strong temporal correlations
  - specificity of the market structure and trading rules.

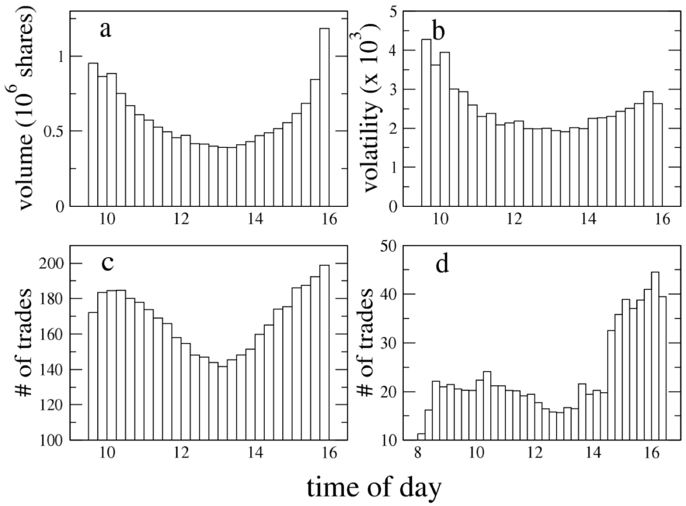
Data size



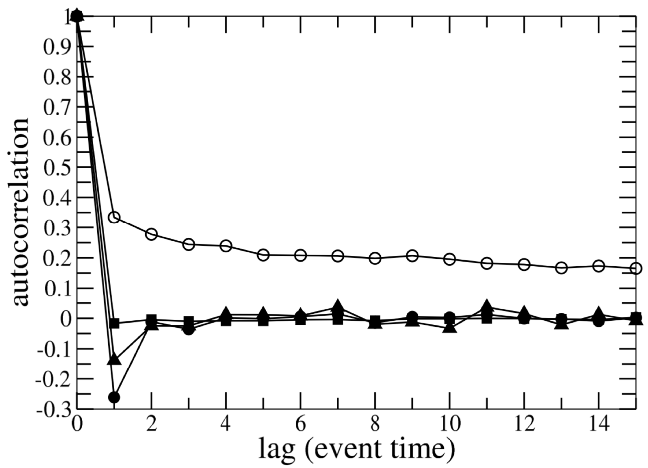
Irregular temporal spacing



Periodicities



Temporal correlations



Lillo and Miccichè, Encyclopedia of Quantitative Finance, 2010

- More structured data require more sophisticated statistical tools
  - data size:
    - more computational power
    - better filtering procedures
  - irregular temporal spacing of events
    - point processes, ACD model, CTRW model…
  - the discreteness of the financial variables under investigation
    - discrete variable processes
  - problems related to proper definition of financial variables
  - intraday patterns
  - strong temporal correlations
    - market microstructure
  - specificity of the market structure and trading rules.
    - better understanding of the trading process

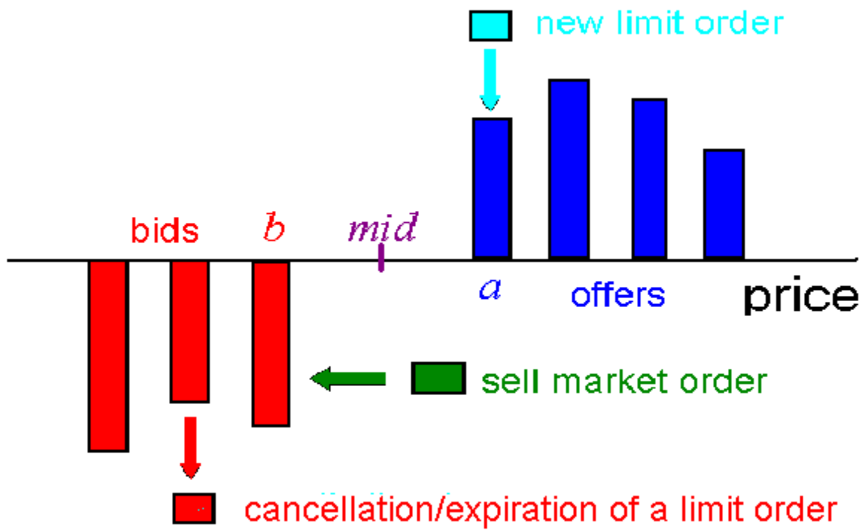
Order book data

- The next resolution of financial data contains data on all the orders placed or canceled in the market
- Many stock exchanges (NYSE, LSE, Paris) works through a double auction mechanism
- Order book data are

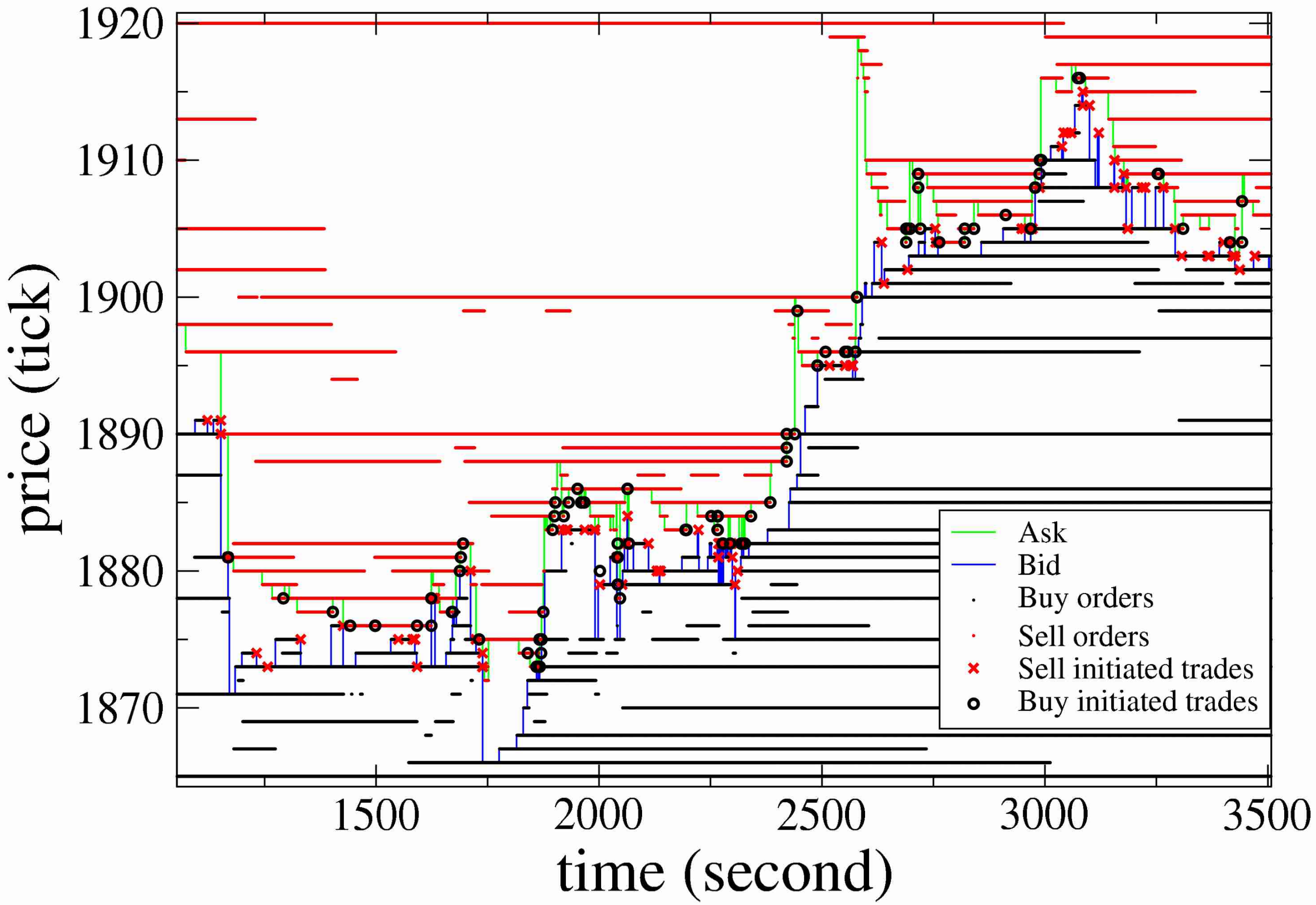
fundamental to

investigate the price

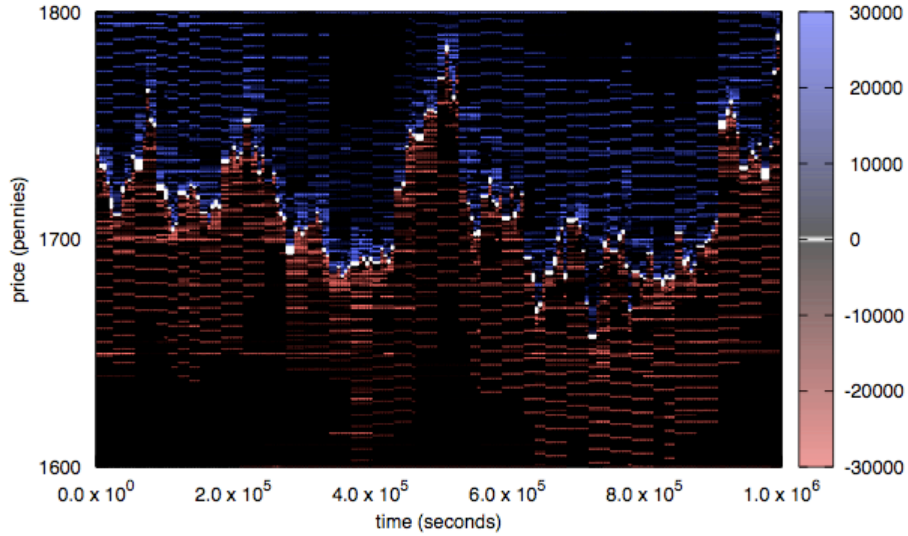
formation mechanisms



Representation of limit order book dynamics



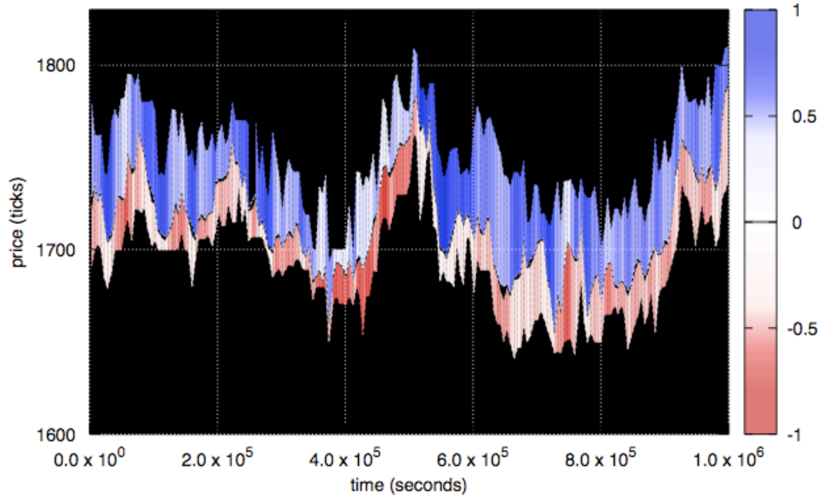
(Ponzi, Lillo, Mantegna 2007)



More structured data require more sophisticated visualization tools

Figure 2. The order book of the stock GSK during the first 32 trading days of 2002. For all days during the period we made five snapshots of the order book at the 5000'th, 10000'th, ..., 25000'th second of the day. The lack of orders is indicated by black color, buy/sell orders by red/blue, and the bid-ask spread by white. For every price level we indicated the total volume of limit orders by coloring as indicated on the right. The ends of the scale correspond to orders for 30000 shares or more.

$$I^{\text{buy}} = \frac{V_t^{\text{buy}}}{V_t^{\text{buy}} + V_t^{\text{sell}}}$$



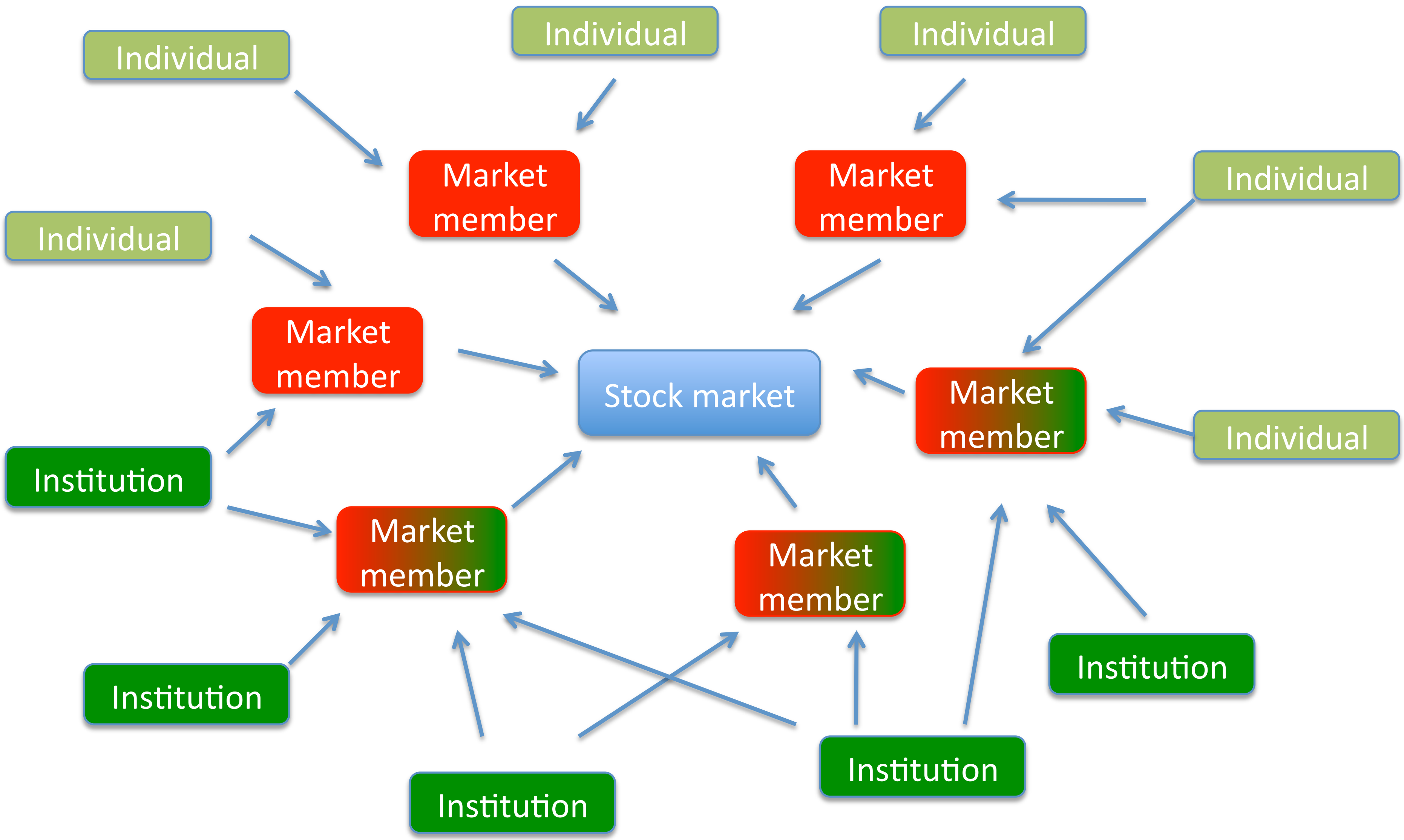
Eisler, Kertesz, Lillo, 2007

Figure 5. The order imbalance and the median-to-median width of the order book of GSK calculated during the first 10 trading days of the year 2002. We took six snapshots of the book during every trading day of the period from its 3000'th to its 28000'th second every 5000 seconds. Red/blue color corresponds to buy/sell orders, and the black line between them to the bid-ask spread. The coloring also indicates the value of I^{buy} and I^{sell} (see right for values), while the width of the graph shows the median of the book in terms of its total volume.

Agent resolved data

- In the recent years there has been an increasing availability and interest toward databases allowing to distinguish, at least partly, the trading activity of “agents” or “classes of agents”.
- In principle, this type of databases allows to investigate empirically the agent’s behavior and strategies, and to study the interaction between agents.

Structure of a financial market



Example: Momentum and contrarian strategies

Momentum investors are buying stocks that were past winners.

A contrarian strategy consists of buying stocks that have been losers (or selling short stocks that have been winners).

The contrarian strategy is formulated on the assumption that the stock market overreacts and a contrarian investor can exploit the inefficiency related to market overreaction by reverting stock prices to fundamental values.

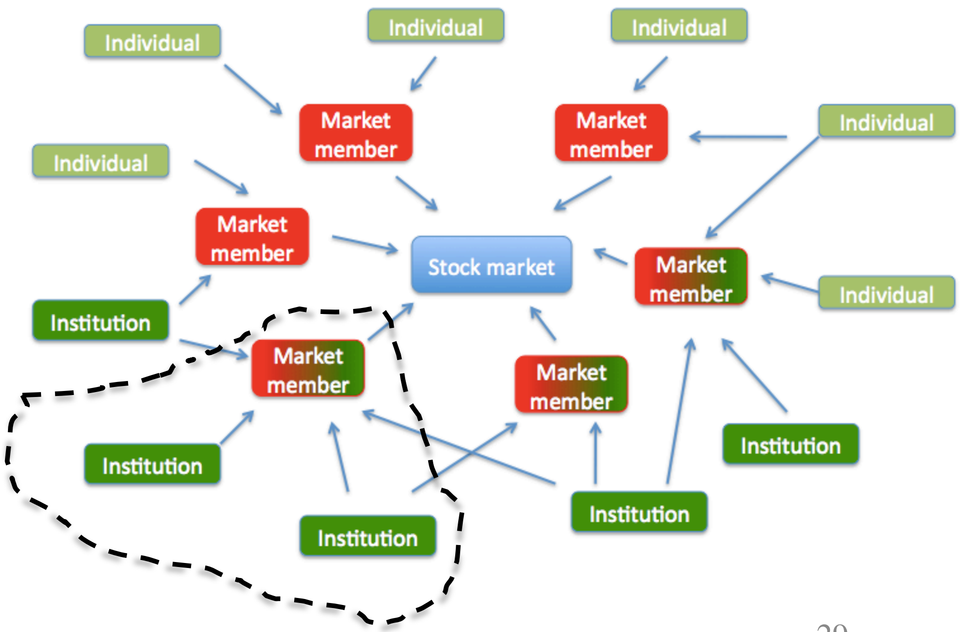
- Is it possible to detect empirically such strategies?
- Are there classes of agents using preferentially these strategies?

Grinblatt, Titman and Wermers (1995)

They investigated the trading pattern of fund managers by examining the quarterly holdings of 155 mutual funds (information from CDA Investment Technologies and CRSP data) over the 1975-1984 period.

- The large majority of funds (77%) had a momentum investment profile.
- Authors found relatively weak evidence that funds tended to buy and sell the same stocks at the same time (herding).

¶M.Grinblatt et al, American Economic Review 85, 1088-1105 (1995)

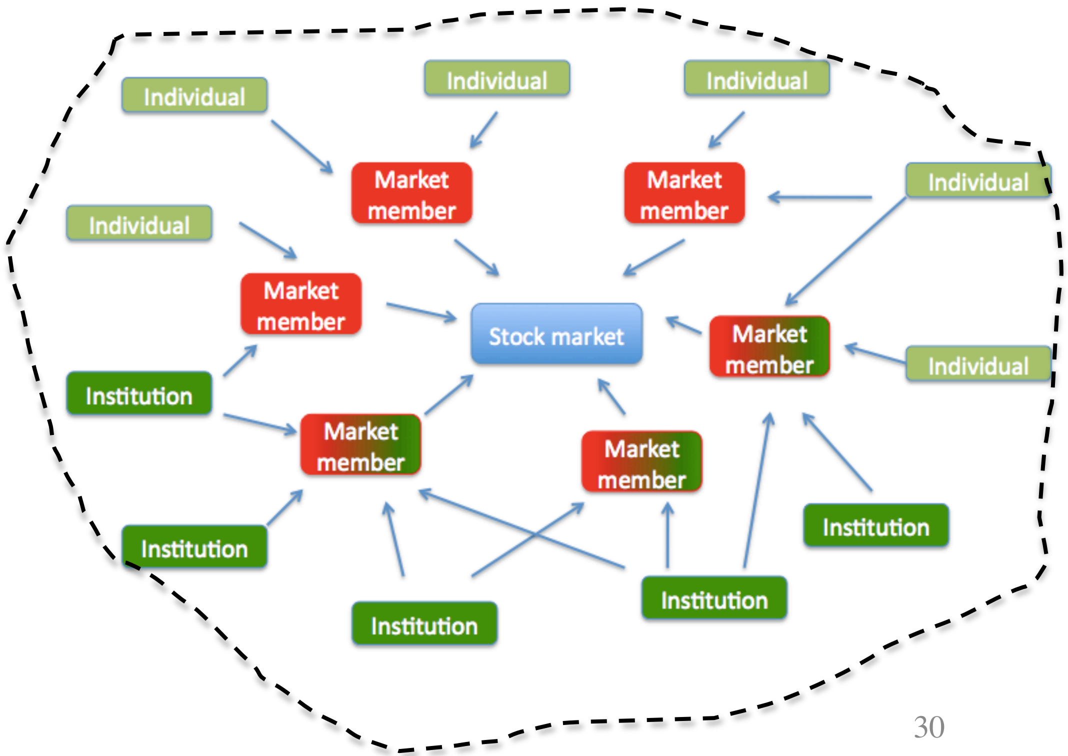


Grinblatt and Keloharju (2000)

Grinblatt and Keloharju[¶] investigated the central register of shareholdings for Finnish Central Securities Depository, a comprehensive data source. This data set reports individual and institutional holdings and stock trades on a daily basis. Data consists of each owner's stock exchange trades from Dec 27, 1994 through Dec 30, 1996.

- Foreign investors tend to be momentum investors
- Individual investors tend to be contrarian
- Domestic institutional investor tend to present a mixed behavior.

¶M.Grinblatt and M.Keloharju, J. of Financial Economics 55, 43-67 (2000)

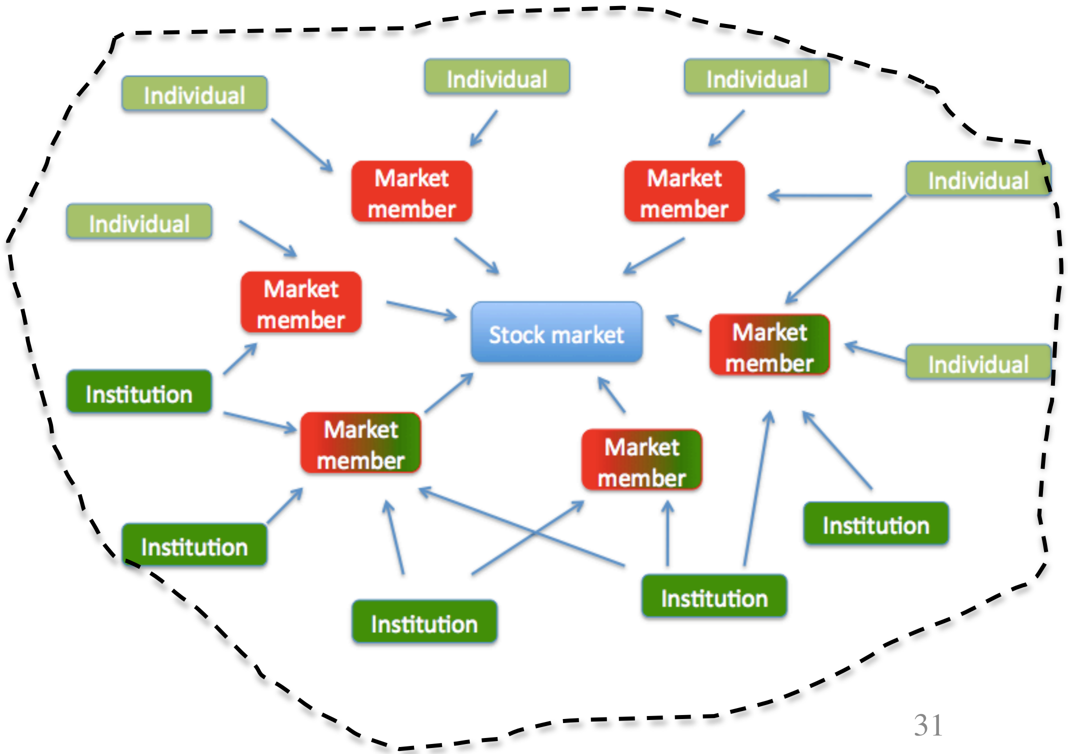


Example: profitability of classes of agents

Studies performed by Barber, Lee, Liu and Odean¶ have the performance of individual and institutional investors at the Taiwan Stock Exchange. Data allow authors to identify trades made by individuals and by institutions, which fall into one of four categories (corporations, dealers, foreigners, or mutual funds).

- Individual investor trading results in systematic and, more importantly, economically large losses
- In contrast, institutions enjoy an annual performance boost of 1.5 percentage points

¶ B.M.Barber, Y.-T.Lee, Y.-J.Liu and T.Odean, Do Individual Day Traders Make Money?
Evidence from Taiwan (2004).
¶ B.M.Barber, Y.-T.Lee, Y.-J.Liu and T.Odean, Just How Much Do Individual Investors Lose by Trading? (2005).



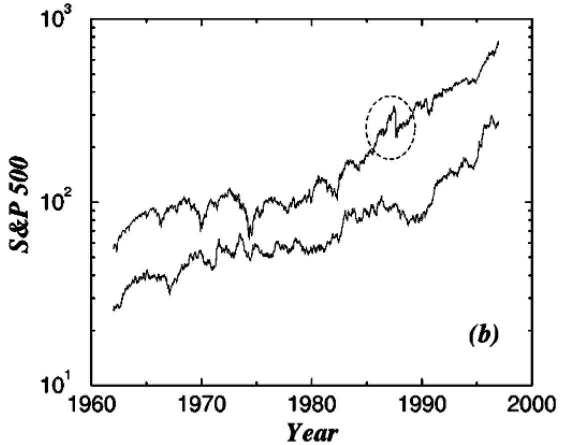
What do the data say?

Statistical regularities or stylized facts

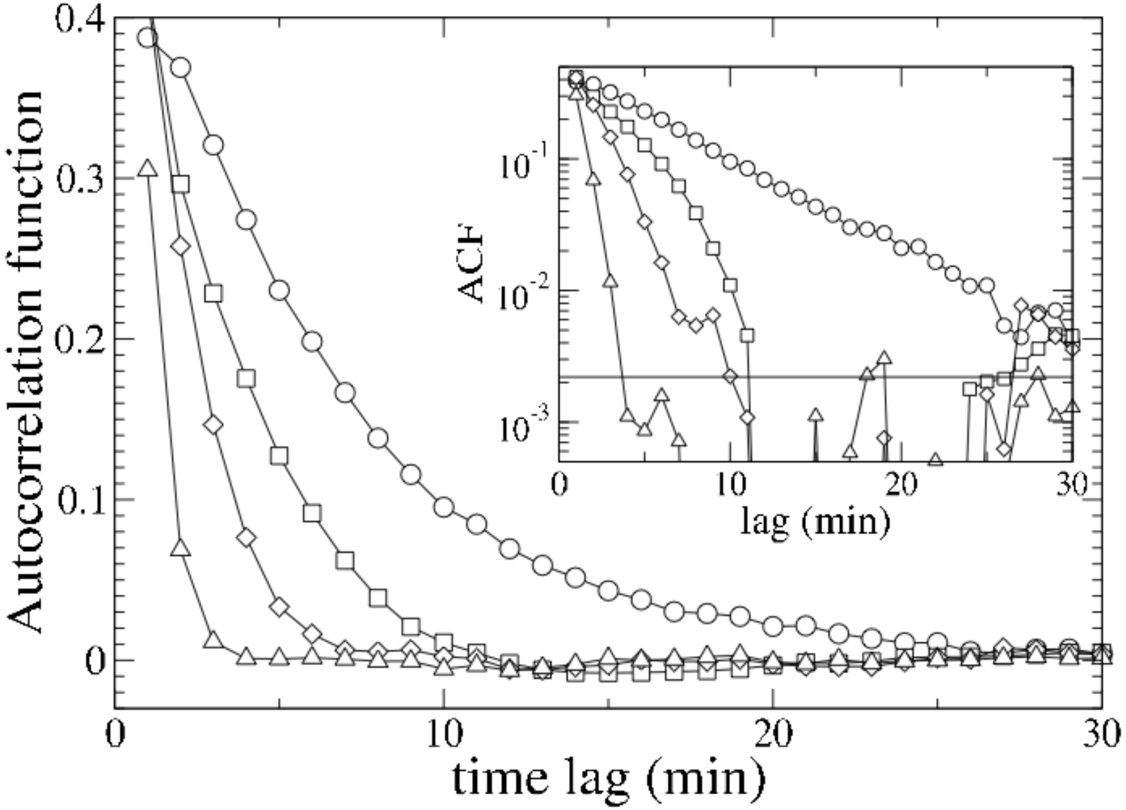
Random walk hypothesis

Depending on the properties of ϵ_t , we distinguish

- Independent Identically Distributed increments: for example Gaussian distributed (stable laws)
- Independent increments
- Uncorrelated increments: the weaker form implies the vanishing of the linear autocorrelation



Linear efficiency

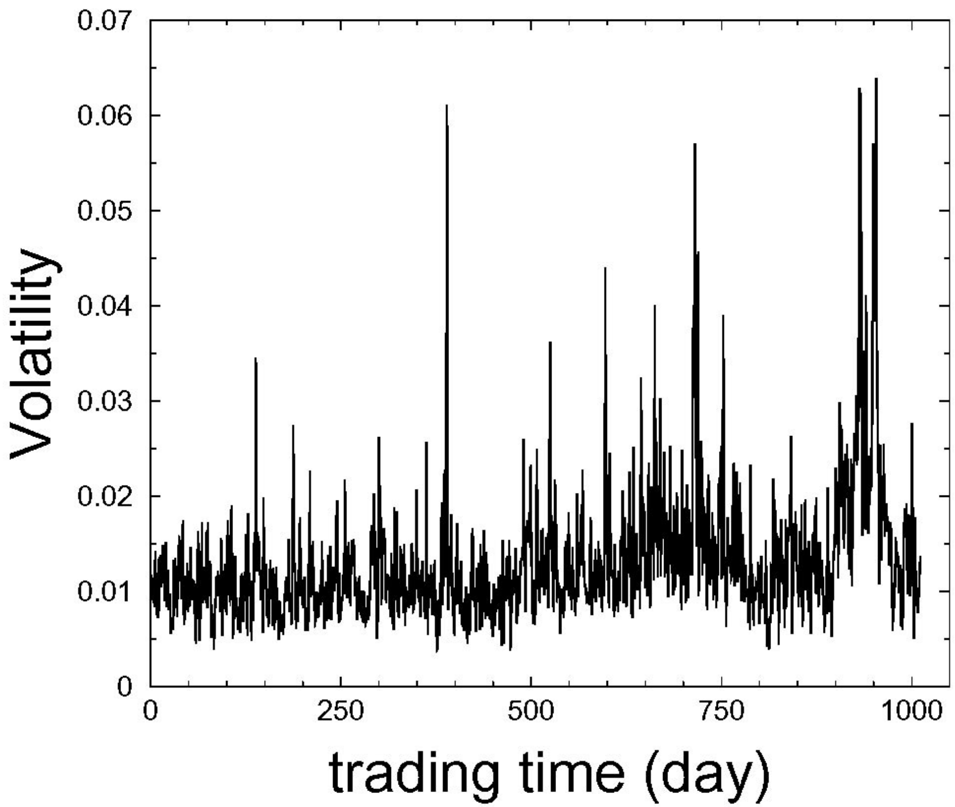


Example: S&P 500
sampled at 1 min
time horizon
1983 - 2004

Characteristic decay time:
378s (1983-1988)
144s (1988-1993)
96s (1994-1999)
36s (1999-2004)

Higher-order correlations

Are higher order correlation present?

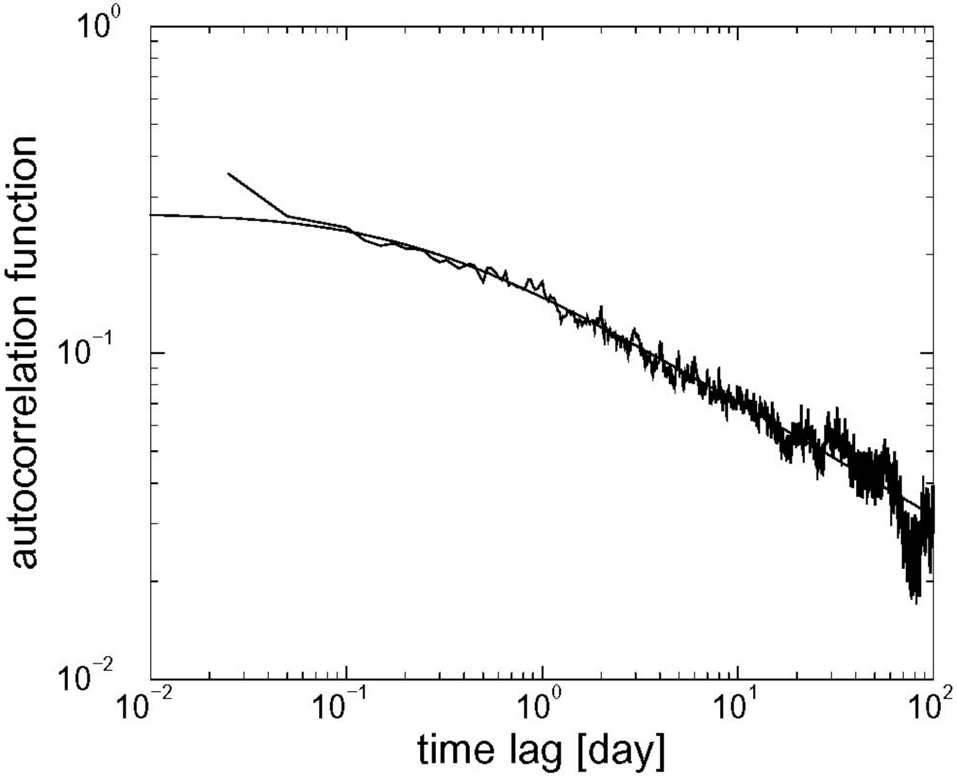


Example: General Electric Co. 1995-1998

The volatility, i.e. the standard deviation of returns, is itself a stochastic process.

Volatility autocorrelation

The volatility autocorrelation is a slow-decaying function



The decay is *compatible* with a power-law decay

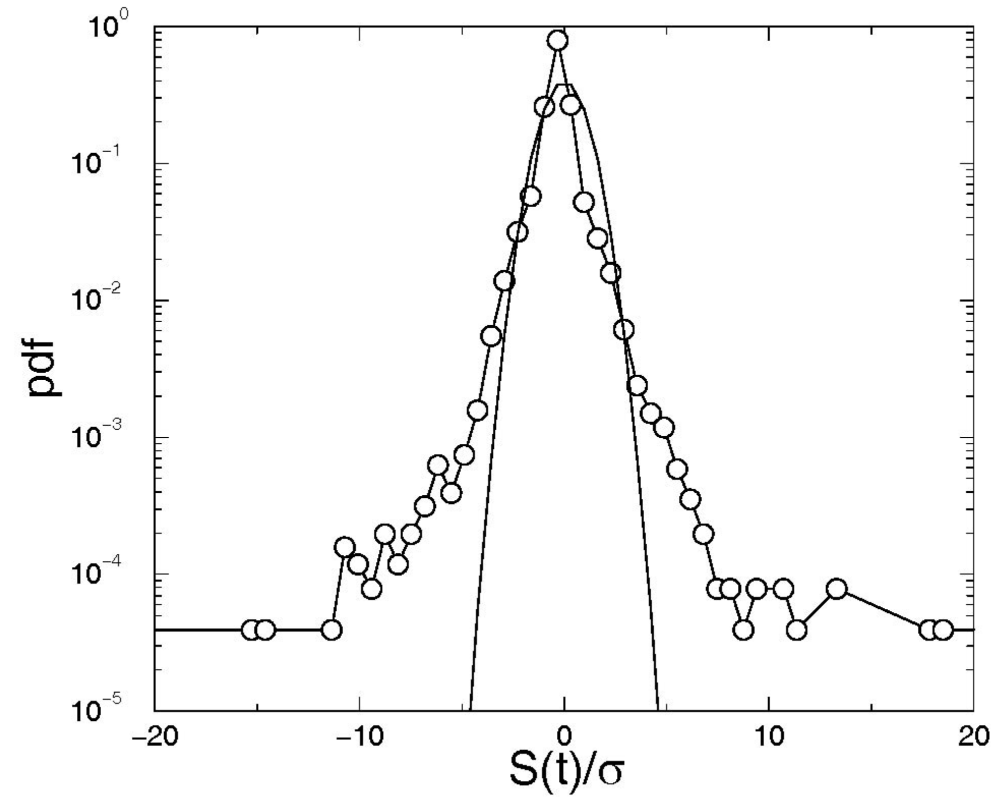
$$\mathrm{ACF}(\sigma(\tau)) \propto \tau^{-\eta}$$

$$\eta \approx 0.3$$

S&P 500 sampled at
1 min time horizon
1984 - 1996

Empirical properties of return pdf

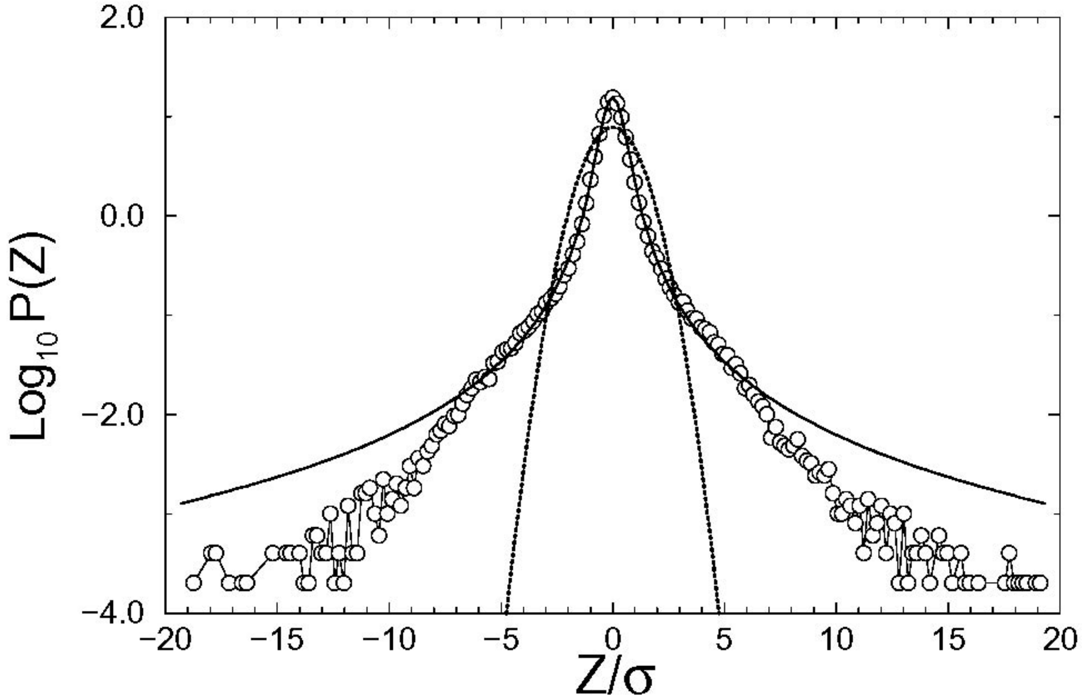
Leptokurtosis increases at short time horizons



Example: Xerox Co.
10 minutes data
1994-1995

The return variance is finite

Several empirical studies have shown that the variance of return pdfs is *finite*

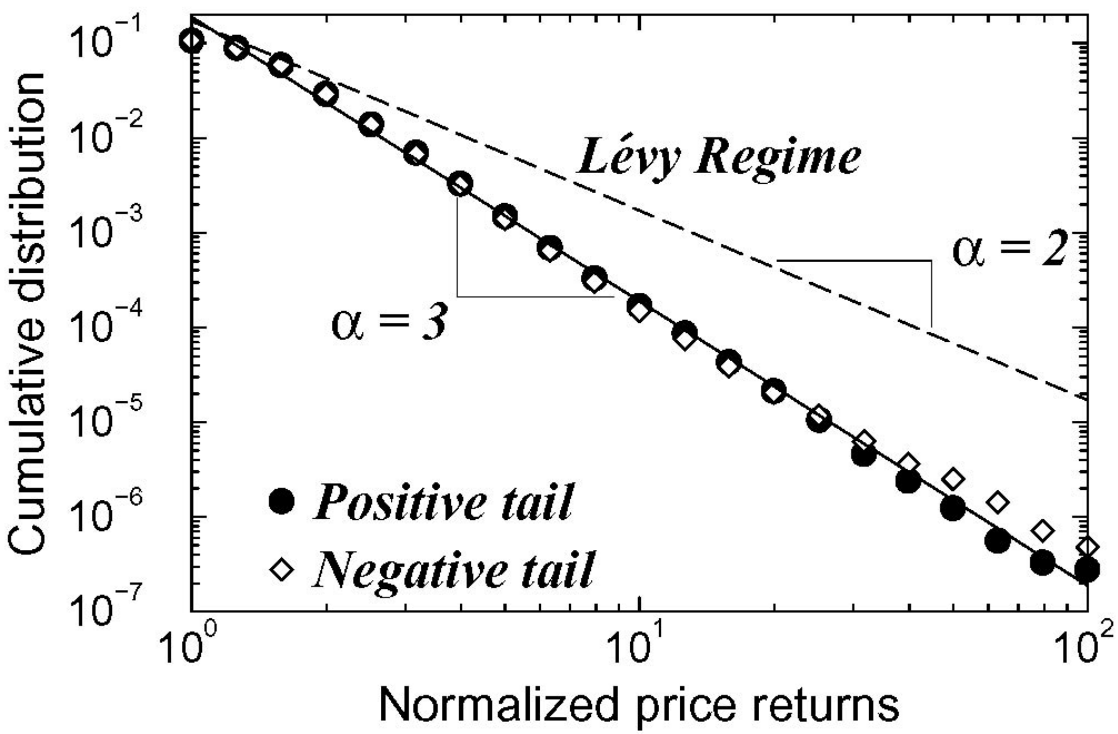


High-frequency investigation of the S&P 500 index (1 minute time horizon 1984-1989)

(Mantegna and Stanley 1995)

Rare events

Rare events are described by a power-law tail with an exponent $\alpha \approx 3$ in the cumulative probability F_p=1-F(x)



ΔT=20 min

(Gopikrishnan et al 2000)

Can we understand the microscopic origin of these stylized facts?

Market microstructure

- Market microstructure "is devoted to theoretical, empirical, and experimental research on the economics of securities markets, including the role of information in the price discovery process, the definition, measurement, control, and determinants of liquidity and transactions costs, and their implications for the efficiency, welfare, and regulation of alternative trading mechanisms and market structures" (NBER Working Group)

Strategic model: Kyle (1985)

- The model describes a case of information asymmetry and the way in which information is incorporated into price.
- It is an equilibrium model
- There are several variants: single period, multiple period, continuous time
- The model postulates three (types of) agents: an informed trader, a noise trader, and a market maker (MM)

- The terminal (liquidation) value of the asset is v, normally distributed with mean p_0 and variance Σ_0.
- The informed trader knows v and enters a demand x
- Noise traders submit a net order flow u, normally distributed with mean 0 and variance σ^2_u.
- The MM observes the total demand y=x+u and then sets a price p. All the trades are cleared at p, any imbalance is exchanged by the MM.

- The informed trader wants to trade as much as possible to exploit her informational advantage
- However the MM knows that there is an informed trader and if the total demand is large (in absolute value) she is likely to incur in a loss. Thus the MM protects herself by setting a price that is increasing in the net order flow.
- The solution to the model is an expression of this trade-off

Informed trader

- The informed trader conjectures that the MM uses a linear price adjustment rule $p=\lambda y+\mu$, where λ is inversely related to liquidity.
- The informed trader's profit is

$$\pi=(v-p)x=x[v-\lambda(u+x)-\mu]$$

and the expected profit is

$$E[\pi]=x(v-\lambda x-\mu)$$

- The informed traders maximizes the expected profit, i.e.

$$x=(v-\mu)/2\lambda$$

- In Kyle's model the informed trader can loose money, but on average she makes a profit

Market maker

- The MM conjectures that the informed trader's demand is linear in v, i.e. $x=\alpha+\beta v$
- Knowing the optimization process of the informed trader, the MM solves

$$(v-\mu)/2\lambda=\alpha+\beta v$$

$$\alpha=-\mu/2\lambda \qquad \beta=1/2\lambda$$

- As liquidity drops the informed agent trades less
- The MM observes y and sets

$$p=E[v|y]$$

Solution

- If X and Y are bivariate normal variables, it is

$$E[Y|X=x]=\mu_Y+(\sigma_{XY}/\sigma_x^2)(x-\mu_X)$$

- This can be used to find

$$E[v|y]=E[v|u+\alpha+\beta v]$$

- The solution is

$$\alpha = -p_0\sqrt{\frac{\sigma_u^2}{\Sigma_0}}; \quad \mu = p_0; \quad \lambda = \frac{1}{2}\sqrt{\frac{\Sigma_0}{\sigma_u^2}}; \quad \beta = \sqrt{\frac{\sigma_u^2}{\Sigma_0}};$$

Solution (II)

- The impact is linear and the liquidity increases with the amount of noise traders

$$p = p_0 + \frac{1}{2}\sqrt{\frac{\Sigma_0}{\sigma_u^2}}\,y$$

- The informed agent trades more when she can hide her demand in the noise traders demand

$$x = (v - p_0)\sqrt{\frac{\sigma_u^2}{\Sigma_0}}$$

- The expected profit of the informed agent depends on the amount of noise traders

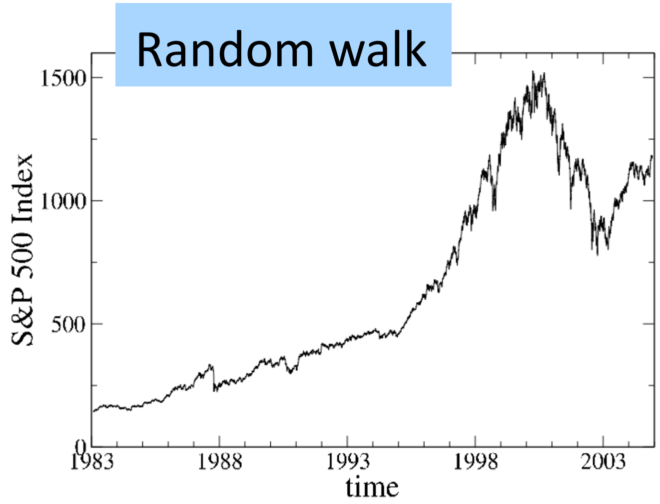
$$E[\pi] = \frac{(v - p_0)^2}{2}\sqrt{\frac{\sigma_u^2}{\Sigma_0}}$$

- The noise traders loose money and the MM breaks even (on average)

Kyle model - summary

- The model can be extended to multiple periods in discrete or in continuous time
- The main predictions of the model are
  - The informed agent "slices and dices" her order flow in order to hide in the noise trader order flow
  - Linear price impact
  - Uncorrelated total order flow
  - Permanent and fixed impact

Price formation and random walk



- Price dynamics is often modeled in terms of a random walk
- This process is mechanically determined by the interplay between order flow and price response
- Specifically, from a statistical point of view, price dynamics is determined by three components
  - The market structure
  - The (unconditional) price response to individual transactions (or events) -> Price (or market) impact as a function of volume
  - The statistical properties of the flow of orders initiating transactions

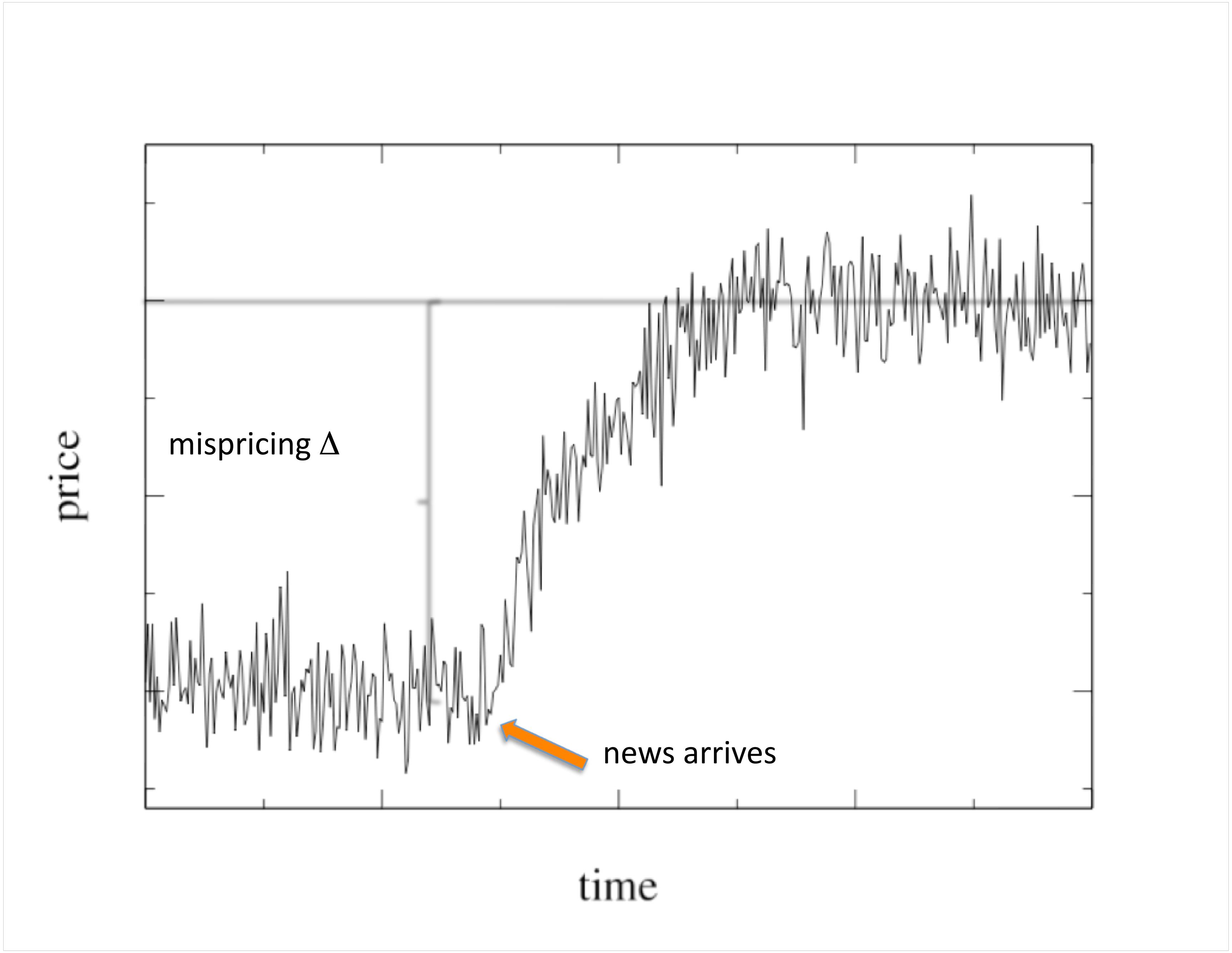
Current paradigm

- There are two types of traders: informed and uninformed
- Informed traders have access to valuable information about the future price of the asset (fundamental value)
- Informed traders sell (buy) over- (under-)priced stocks making profit AND, through their own impact, drive quickly back the price toward its fundamental value

- In this way information is incorporated into prices,

markets are efficient, and prices are unpredictable

Current paradigm



Is this the right explanation? Orders of magnitude

- Information
  - How large is the relative uncertainty on the fundamental value? 10^{-3} or 1 (Black 1986)
  - Financial experts are on the whole pretty bad in forecasting earnings and target prices
- Time
  - Time scale for news: 1 hour-1day (?)
  - Time scale for trading: 10^{-1}s:10^{0}s
  - Time scale for market events: 10^{-2}:10^{-1}s
  - Time scale for "large" price fluctuations: 10 per day
- Volume
  - Daily volume: 10^{-3}:10^{-2} of the market capitalization of a stock
  - Volume available in the book at a given time: 10^{-4}:10^{-5} of the market capitalization
  - Volume investment funds want to buy: up to 1% of a company

Consequences

- 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

Price (or market) impact

- Price impact is the correlation between an incoming order and the subsequent price change
- For traders impact is a cost -> Controlling impact
- Volume vs temporal dependence of the impact

Why price impact?

- Given that in any transaction there is a buyer and a seller, why is there price impact?
  – Agents successfully forecast short term price movements and trade accordingly (i.e. trade has no effect on price and noise trades have no impact)
  – The impact of trades reveals some private information (but if trades are anonymous, how is it possible to distinguish informed trades?)
  – Impact is a statistical effect due to order flow fluctuations (zero-intelligence models, self-fulfilling prophecy)

"Orders do not impact prices. It is more accurate to say that orders forecast prices" (Hasbrouck 2007)

Market impact

- Market impact is the price reaction to trades
- However it may indicate many different quantities
  - Price reaction to individual trades
  - Price reaction to an aggregate number of trades
  - Price reaction to a set of orders of the same sign placed consecutively by the same trader (hidden order)
  - Price reaction in a market to a trade in another market (e.g. electronic market and block market)

Volume and temporal component of market impact of individual trades

- Market impact is the expected price change due to a transaction of a given volume. The response function is the expected price change at a future time

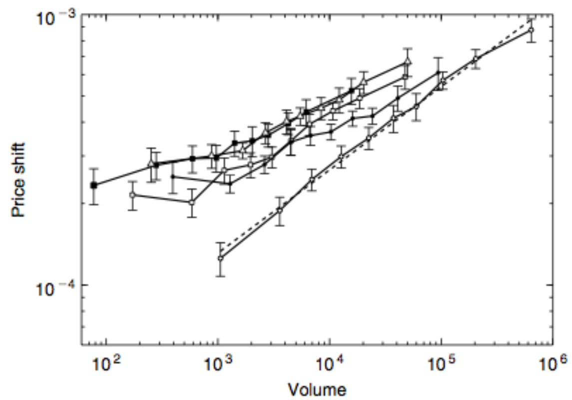


FIGURE 2.4 Market impact function of buy market orders for a set of five highly capitalized stocks traded in the LSE, specifically AZN (*filled squares*), DGE (*empty squares*), LLOY (*triangles*), SHEL (*filled circles*), and VOD (*empty circles*). Trades of different sizes are binned together, and the logarithmic price change's average size for each bin is shown on the vertical axis. The *dashed line* is the best fit of the market impact of VOD with a functional form as described in Eq. 2.8. The value of the fitted exponent for VOD is $\psi = 0.3$.

$$f(V) \equiv E[(p_{t+1} - p_t)\varepsilon_t | v_t = V]$$

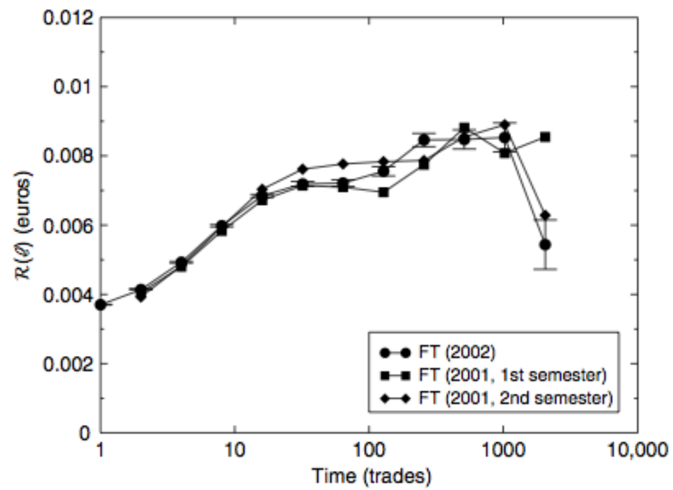
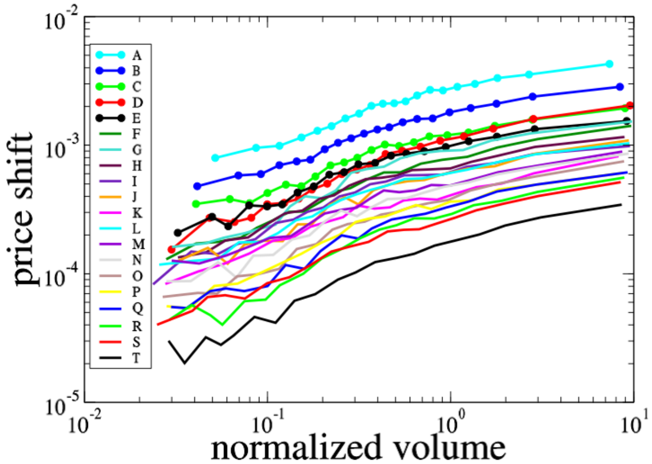


FIGURE 2.8 Average empirical response function $\mathcal{R}_\ell$ for FT during three different periods (1st and 2nd semester of 2001 and 2002); error bars are shown for the 2002 data. For the 2001 data, the y axis has been rescaled such that $\mathcal{R}_1$ coincides with the 2002 result. $\mathcal{R}_\ell$ is seen to increase by a factor ~ 2 between $\ell = 1$ and $\ell = 100$.

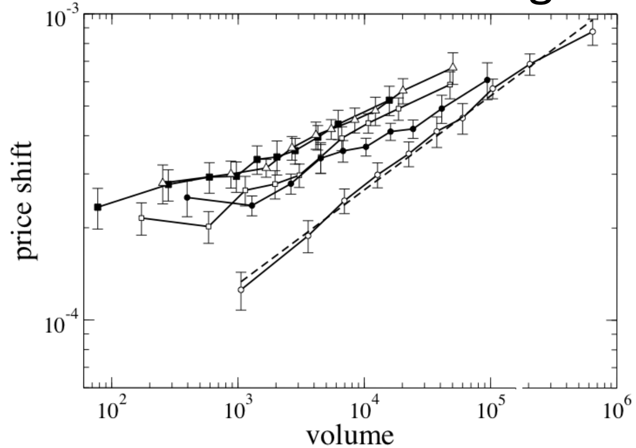
$$R(\ell) = E[(p_{t+\ell} - p_t)\varepsilon_t]$$

Master curve for individual impact

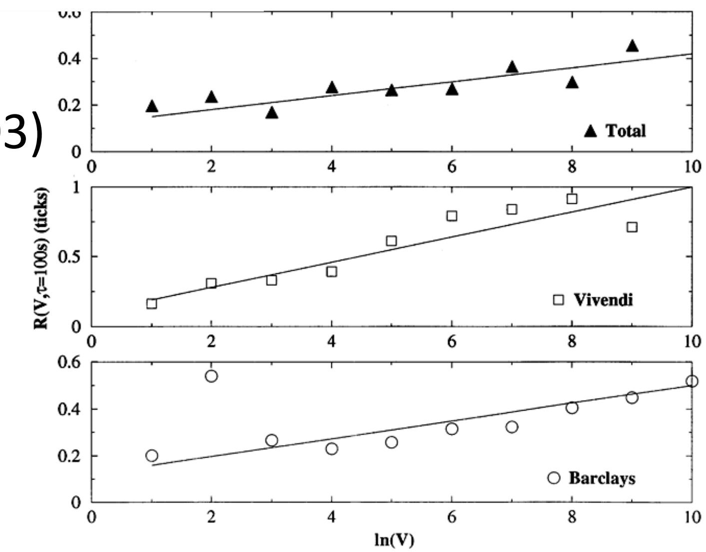
NYSE



London Stock Exchange



Paris Bourse
(Potters et al. 2003)



Impact of individual transaction is NOT universal

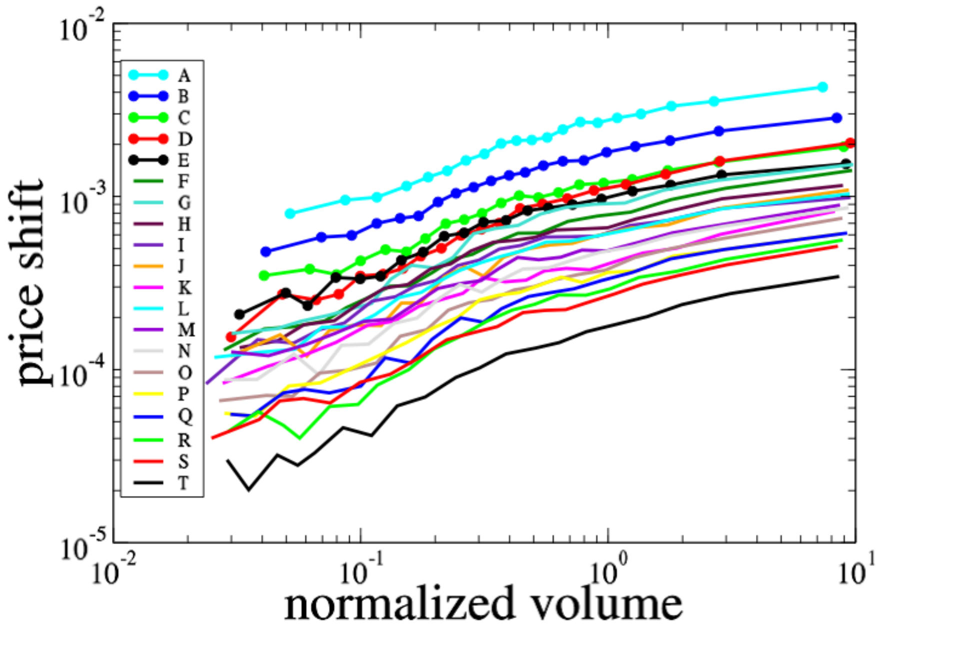
Individual market impact is a concave function of the volume

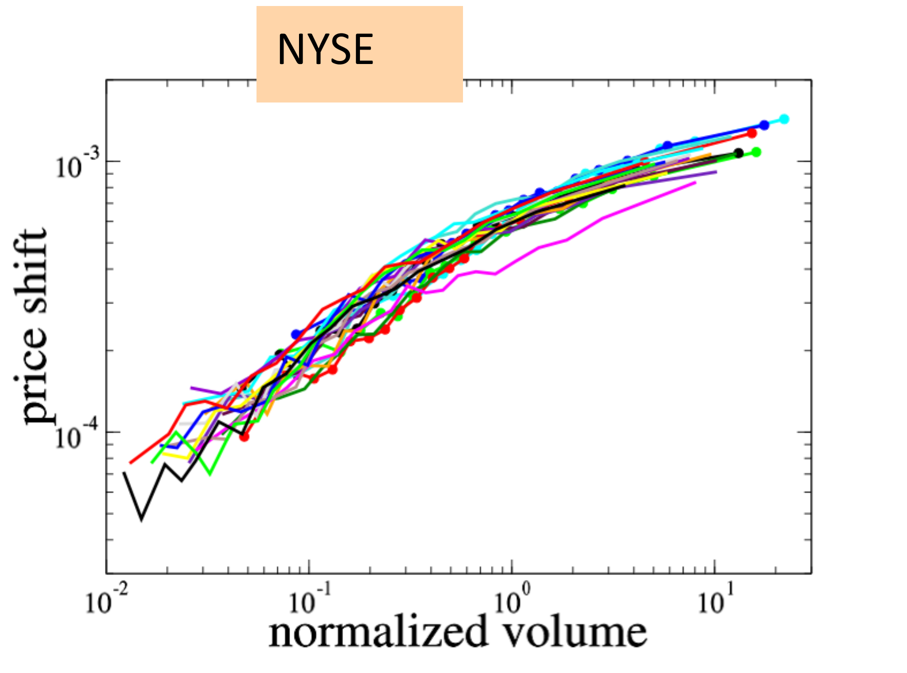
Master curve for individual impact

(Lillo et al. Nature 2003)

GROUP A -> least capitalized group

GROUP T -> most capitalized group





$$r = C^{-\gamma} f(VC^{\delta})$$

Fluctuations of the impact

Let us decompose the conditional probability of a return r conditioned to an order of volume V as

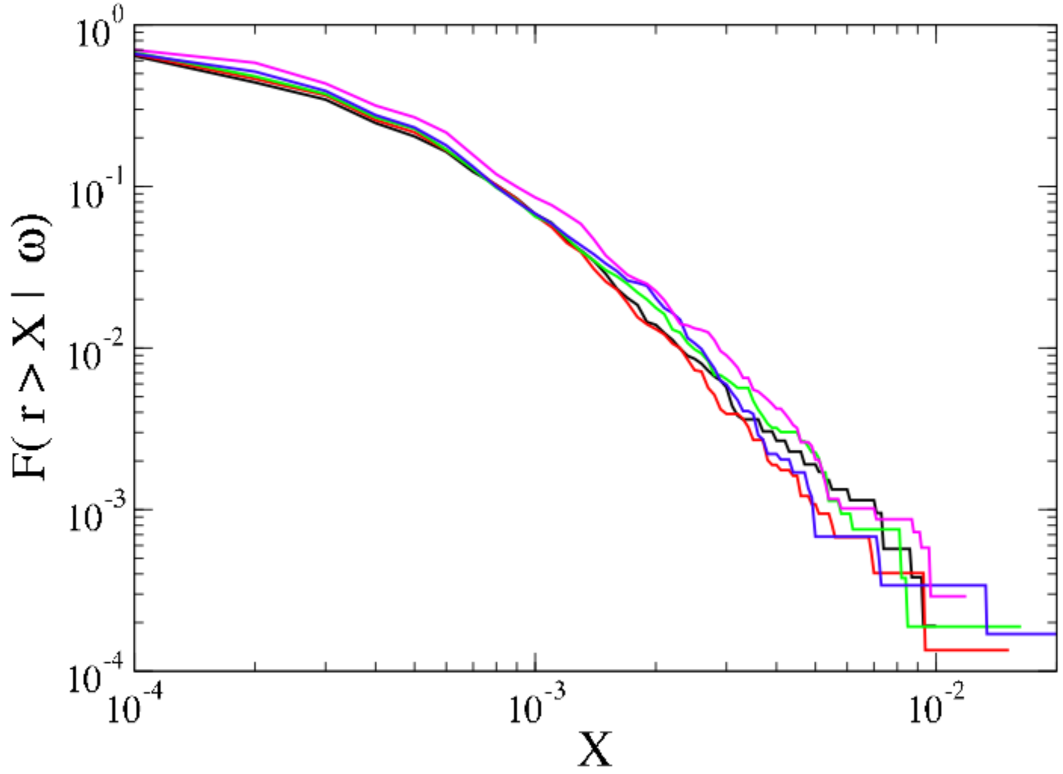
$$p(r|V) = (1 - g(V))\delta(r) + g(V)f(r|V)$$

and we investigate the cumulative probability

$$F(r > X|V) = \int_X^\infty f(r|V)\, dr$$

for several different value of V.

This is the cumulative probability of a price return r conditioned to the volume and to the fact that price moves



Different curves correspond to different trade volume

Independent from the volume !!

- The role of the transaction volume is negligible. The volume is important in determining whether the price moves or not

- **The fluctuations in market impact are important**

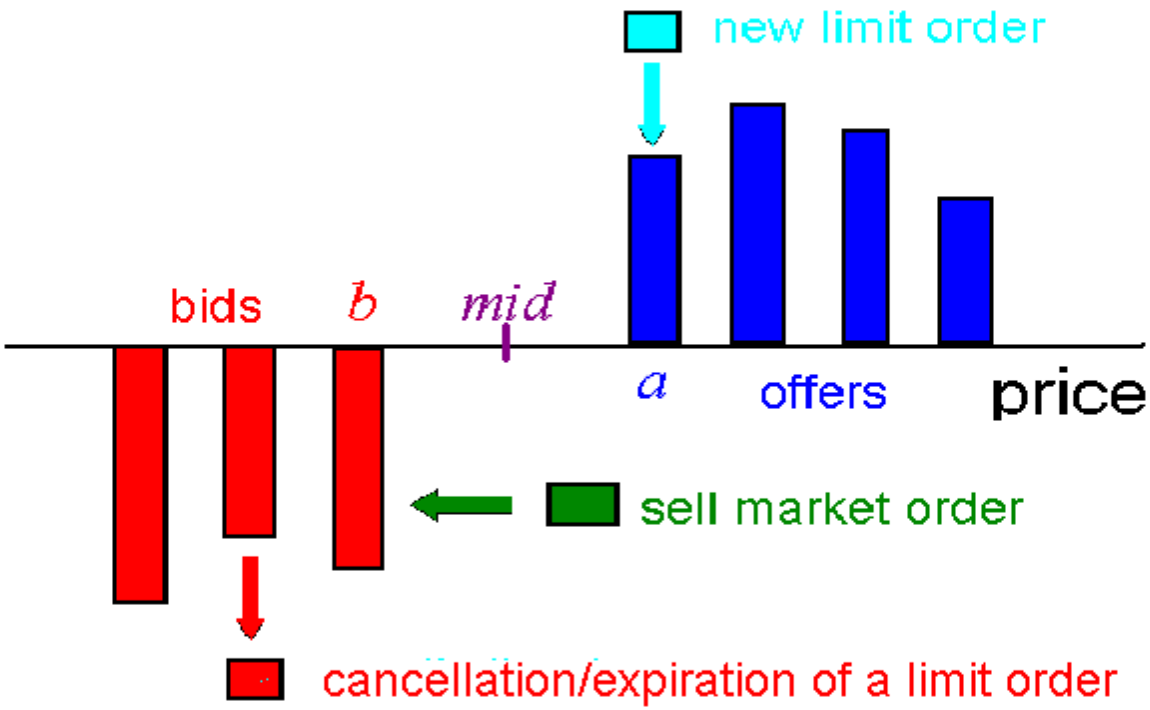
- The impact function is NOT deterministic and the fluctuations of price impact are very large.
- These results show that the picture of the book as an approximately constant object is substantially **wrong**

- Central role of **fluctuations** in the state of the book

- How can small volume transactions create large price changes ?

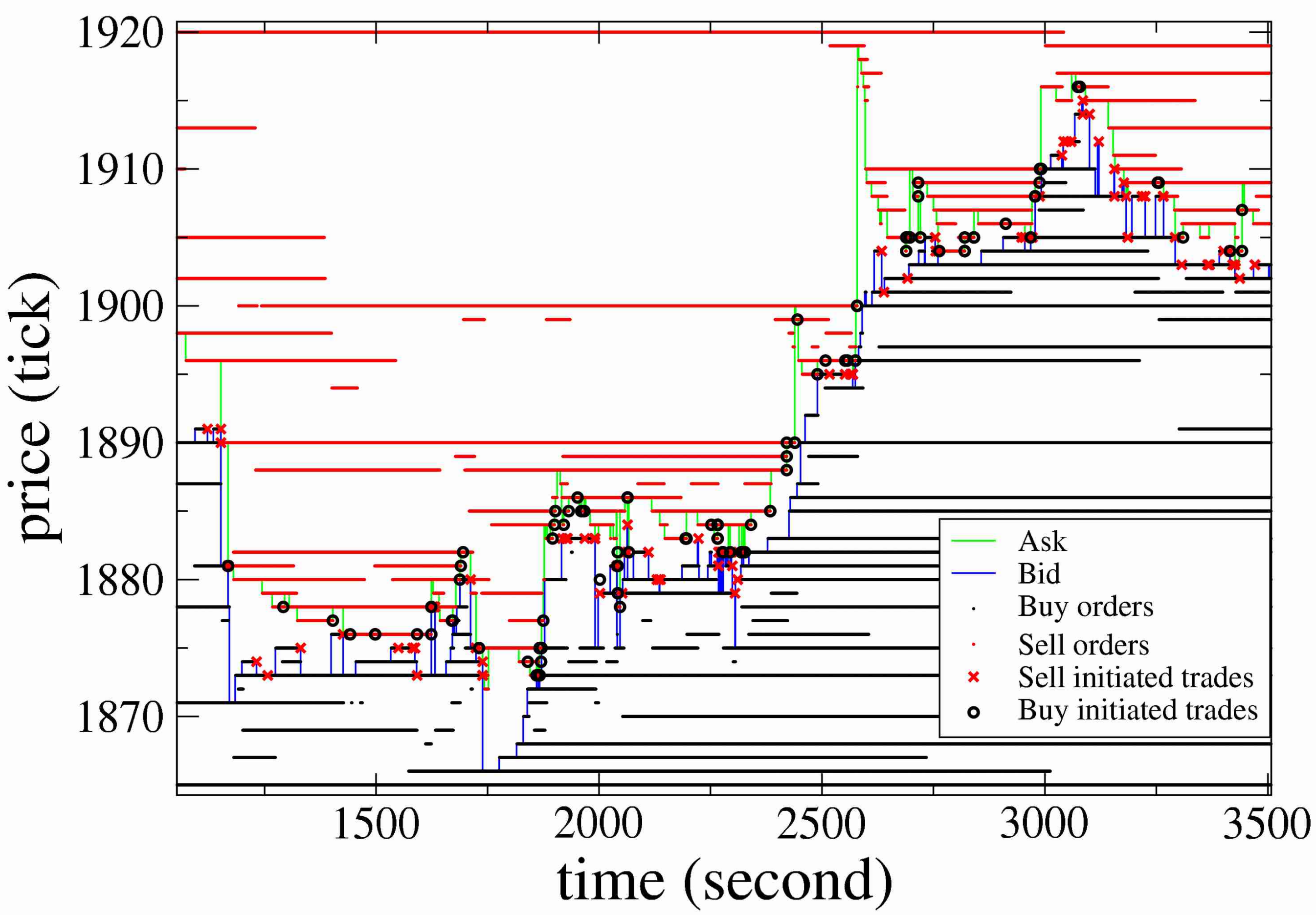
Continuous double auction

- Many stock exchanges (NYSE, LSE, Paris) works through a double auction mechanism



Limit order book

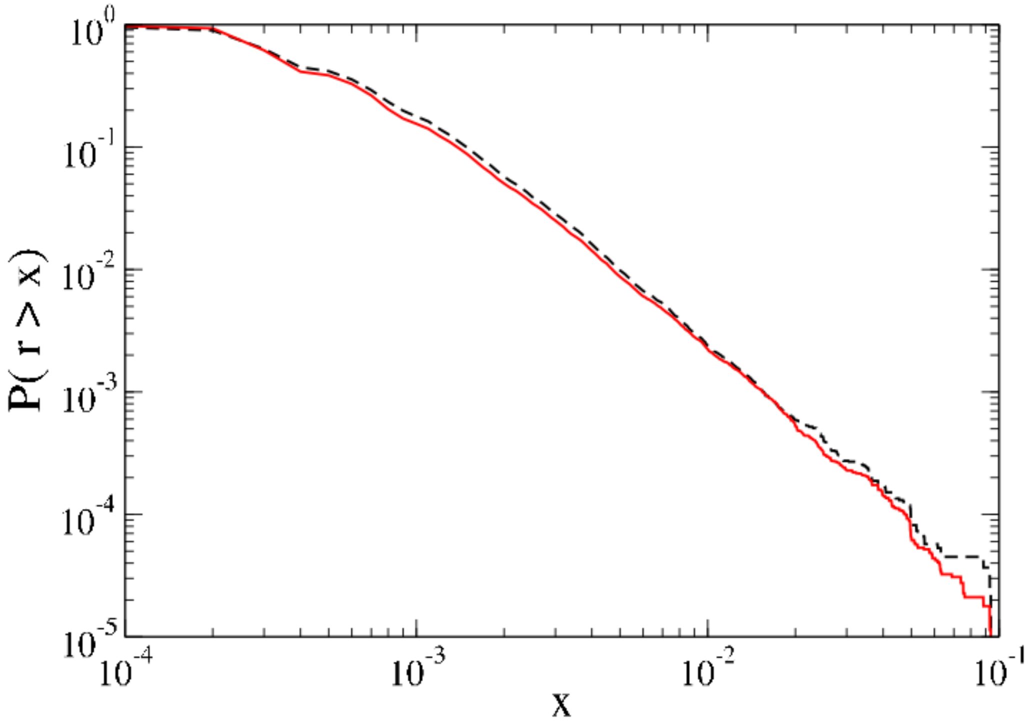
Representation of limit order book dynamics



Case study

- Large price changes are due to the granularity of supply and demand
- The granularity is quantified by the size of gaps in the Limit Order Book

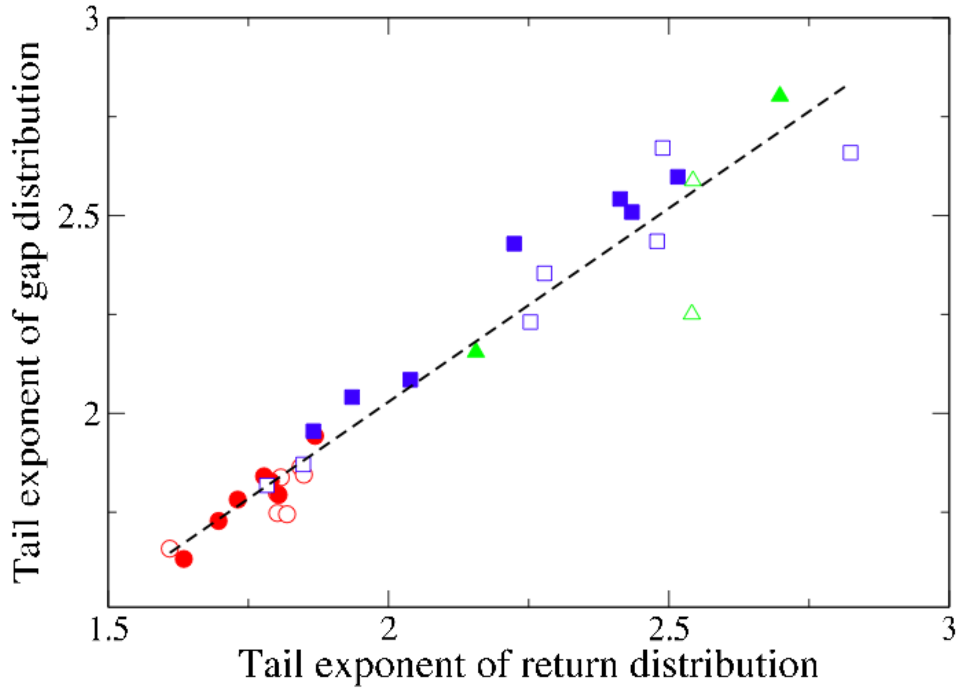
Origin of large price returns



- First gap distribution (red) and return distribution (black)

Large price returns are caused by the presence of large gaps in the order book

Tail exponents (Farmer et al 2004)



Low liquidity (red), medium liquidity (blue), high liquidity (green)

A similar exponent describes also the probability density of the successive gaps

Walking up the book?

• The analysis of transactions in both large and small tick size LSE stocks reveal that the “walking up” of the book, i.e. a trades that involves more than one price level in the limit order book, is an extremely rare event

# of gaps	0	1	2	3	4	5	>5
AZN	44%	49%	5.8%	0.80%	0.15%	0.026%	0.22%
VOD	64%	34%	1.7%	0.094%	0.010%	0.0002%	0.19%

• This again strengthens the idea that market order traders strongly condition their order size to the best available volume

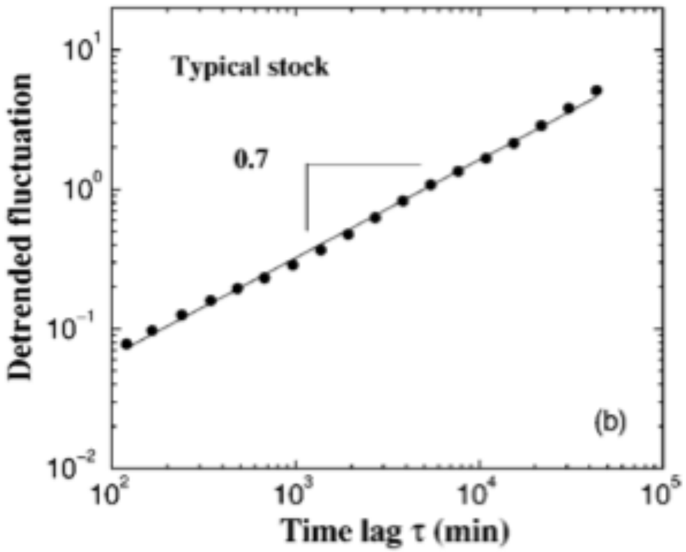
• Thus the use of the instantaneous shape of the limit order book for computing the market liquidity risk can be very misleading

Financial markets are sometimes found in a state of temporary liquidity crisis, given by a sparse state of the book. Even small transactions can trigger large spread and market instabilities.

- Are these crises persistent?
- How does the market react to these crises?
- What is the permanent effect of the crises on prices?

Persistence of spread and gap size

DFA of spread (Plerou et al 2005)



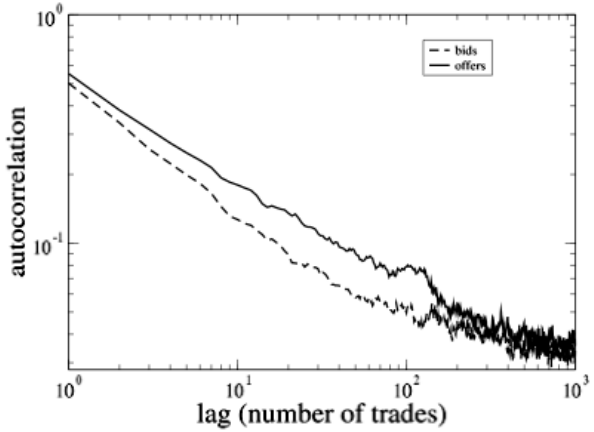


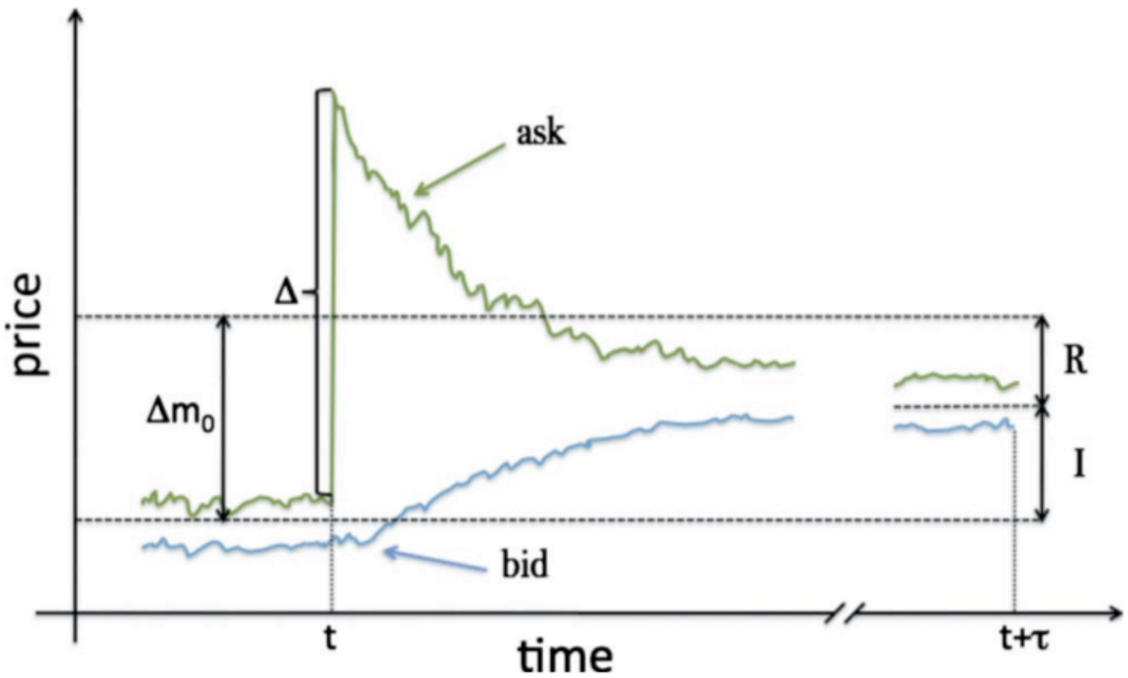
Fig. 4. Autocorrelation function of the first gap size for bids ($g_{-1}(t)$) and offers ($g_1(t)$) in a log-log plot. The data refer to Astrazeneca.

Table 1. Correlation coefficient of the first three gap size g_i on the buy side ($i = -3, -2, -1$) and on the sell side ($i = 1, 2, 3$) of the limit order book. The data shown refer to the stock Astrazeneca.

ρ	g_{-3}	g_{-2}	g_{-1}	g_1	g_2	g_3
g_{-3}	1.00	0.35	0.24	0.10	0.08	0.08
g_{-2}	0.35	1.00	0.27	0.11	0.08	0.08
g_{-1}	0.24	0.27	1.00	0.15	0.15	0.13
g_1	0.10	0.11	0.15	1.00	0.33	0.30
g_2	0.08	0.08	0.15	0.33	1.00	0.41
g_3	0.08	0.08	0.13	0.30	0.41	1.00

Market reaction to temporary liquidity crises

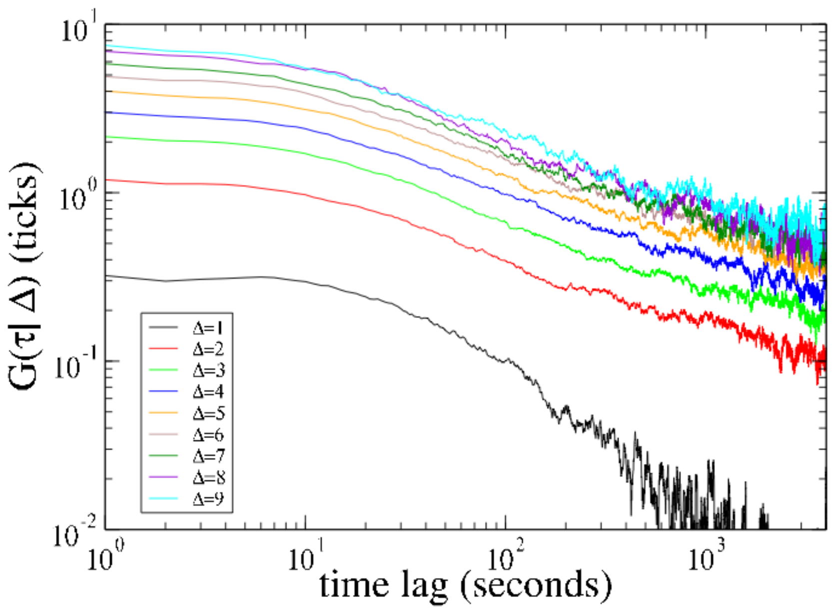
- We quantify the market reaction to large spread changes.
- The presence of large spread poses challenging questions to the traders on the optimal way to trade.
- Liquidity takers have a strong disincentive for submitting market orders given that the cost, the bid-ask spread, is large
- Liquidity provider can profit of a large spread by placing limit orders and obtaining the best position. However the optimal order placement inside the sprea[d] a nontrivial problem.

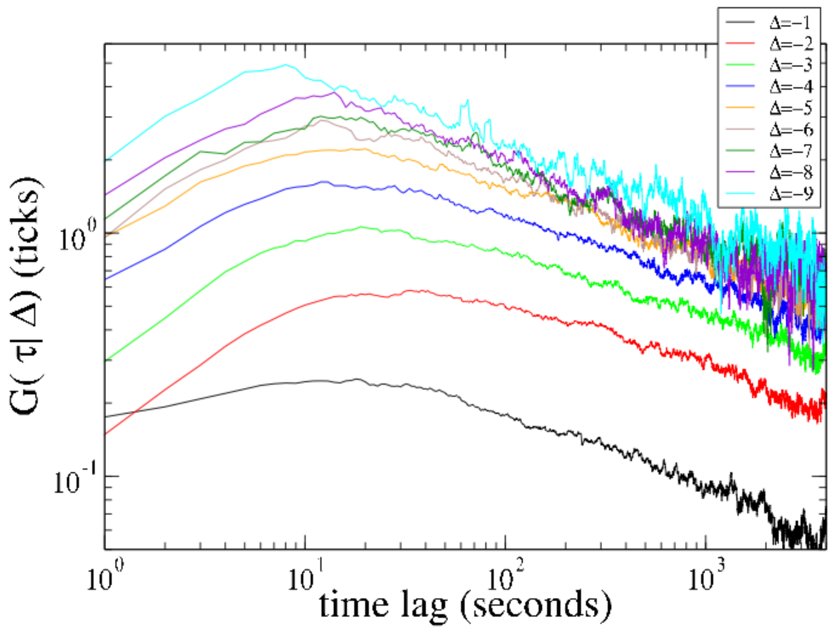


  - Rapidly closing the spread -> priority but risk of informed traders
  - Slowly closing the spread -> "testing" the informed traders but risk of losing priority

We wish to answer the question: **how does the spread s(t) return to a normal value after a spread variation?**
To this end we introduce the quantity

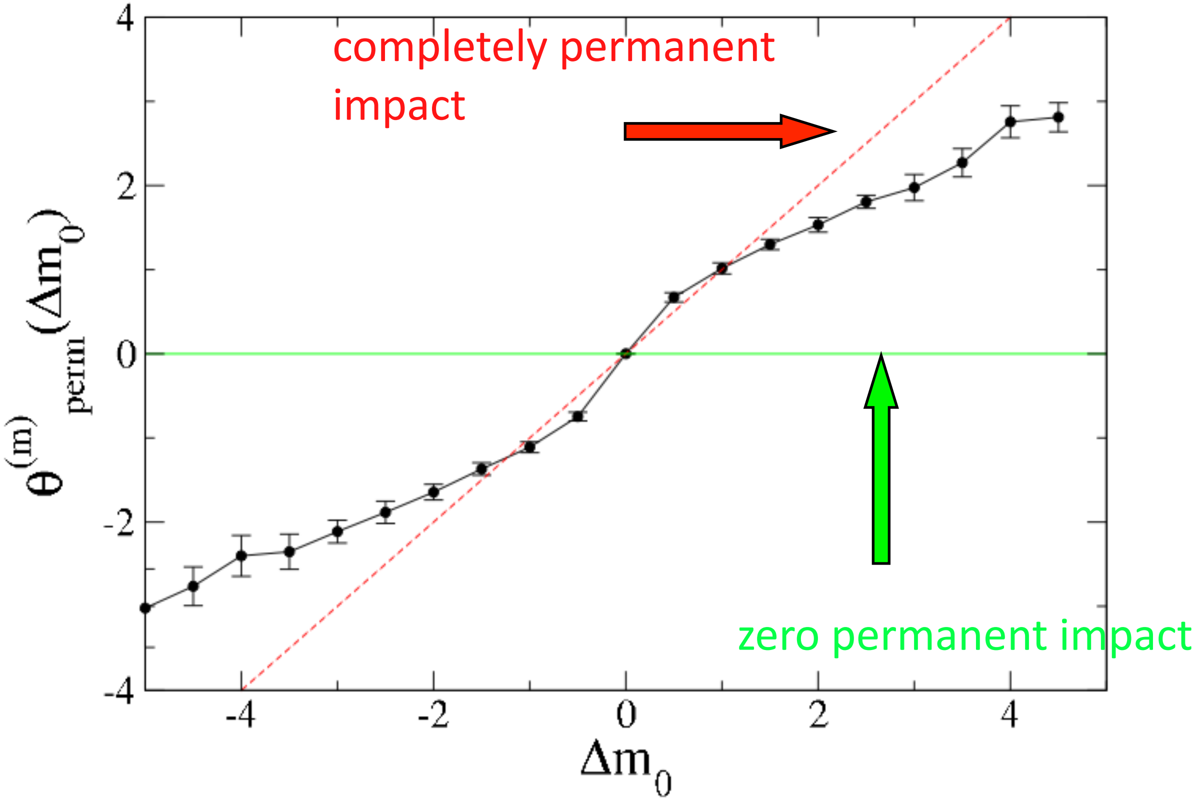
$$G(\tau|\Delta) = E(s(t+\tau)|s(t) - s(t-1) = \Delta) - E(s(t))$$





$$G(\tau|\Delta) \sim \tau^{-\delta} \qquad \delta \simeq 0.4 - 0.5$$

Permanent impact

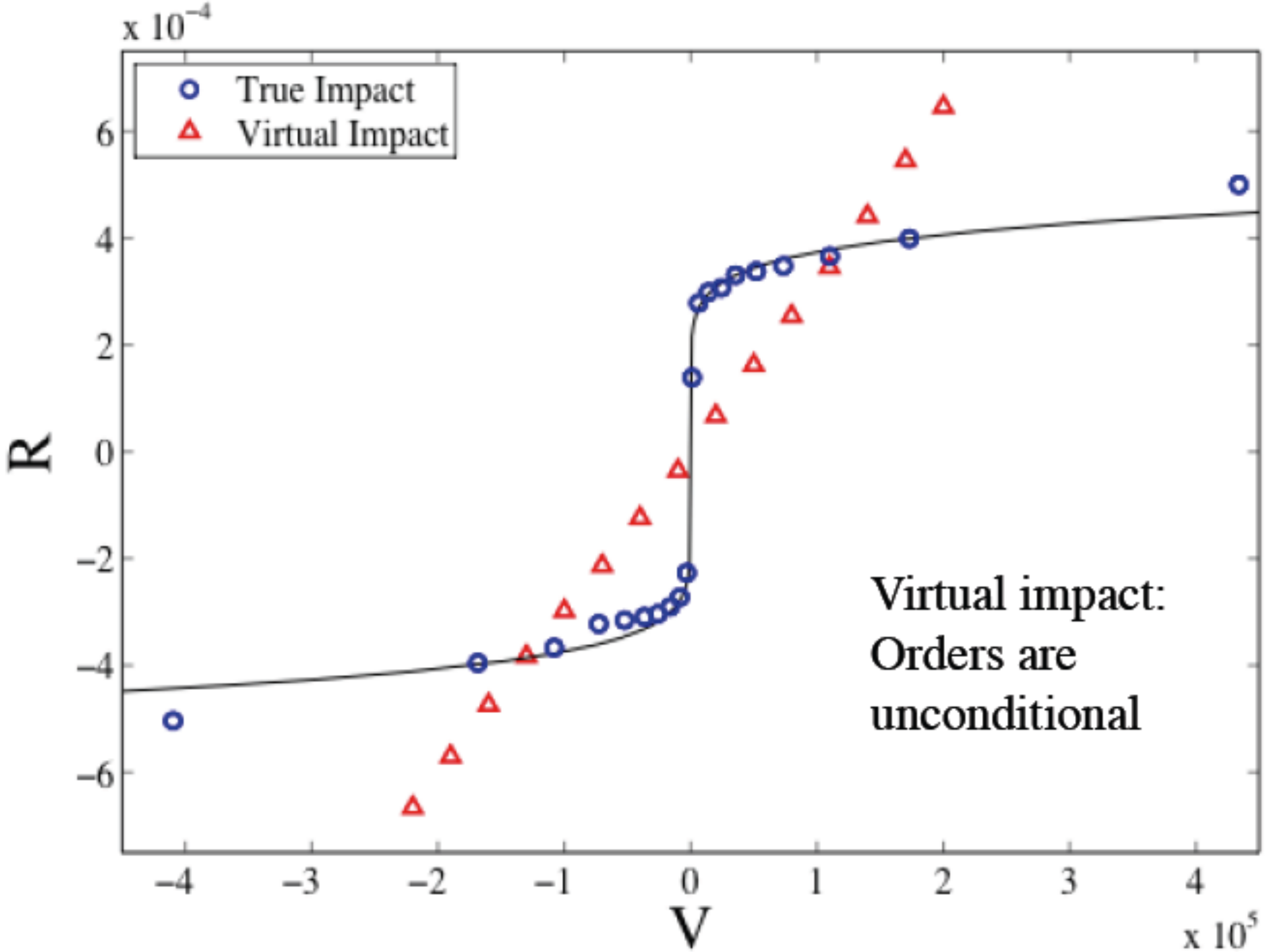


Permanent impact is roughly proportional to immediate impact

Why is single trade impact concave?

- Liquidity takers condition their orders on what is offered.
  - When offer is deep, they submit large orders
  - When offer is shallow, they submit small orders
  - Result is that observed impact grows slowly with size

Market impact f(V) for single trade



Farmer, Gillemot, Lillo, Mike, Sen (Quantitative Finance, 2004)
Weber and Rosenow (Quantitative Finance, 2006), Gerig (2007)

What is the origin of fat tails and clustered volatility?

- A common belief is that large part of the story is explained by the inhomogeneous rate of trading
- Theories making use of this point of view are for example:

–Subordinated processes (Mandelbrot and Taylor 1967, Clark 1973, Ane and Geman 2000)

- price shift due to individual transactions are Gaussian (or thin tailed), but when many trades are aggregated in a time interval, the return distribution can be fat tailed. This is due to the fluctuation of number of trades or volume in the time interval

– Volume fluctuations (Gabaix et al. 2003)

Alternative time clocks

We define transaction time as

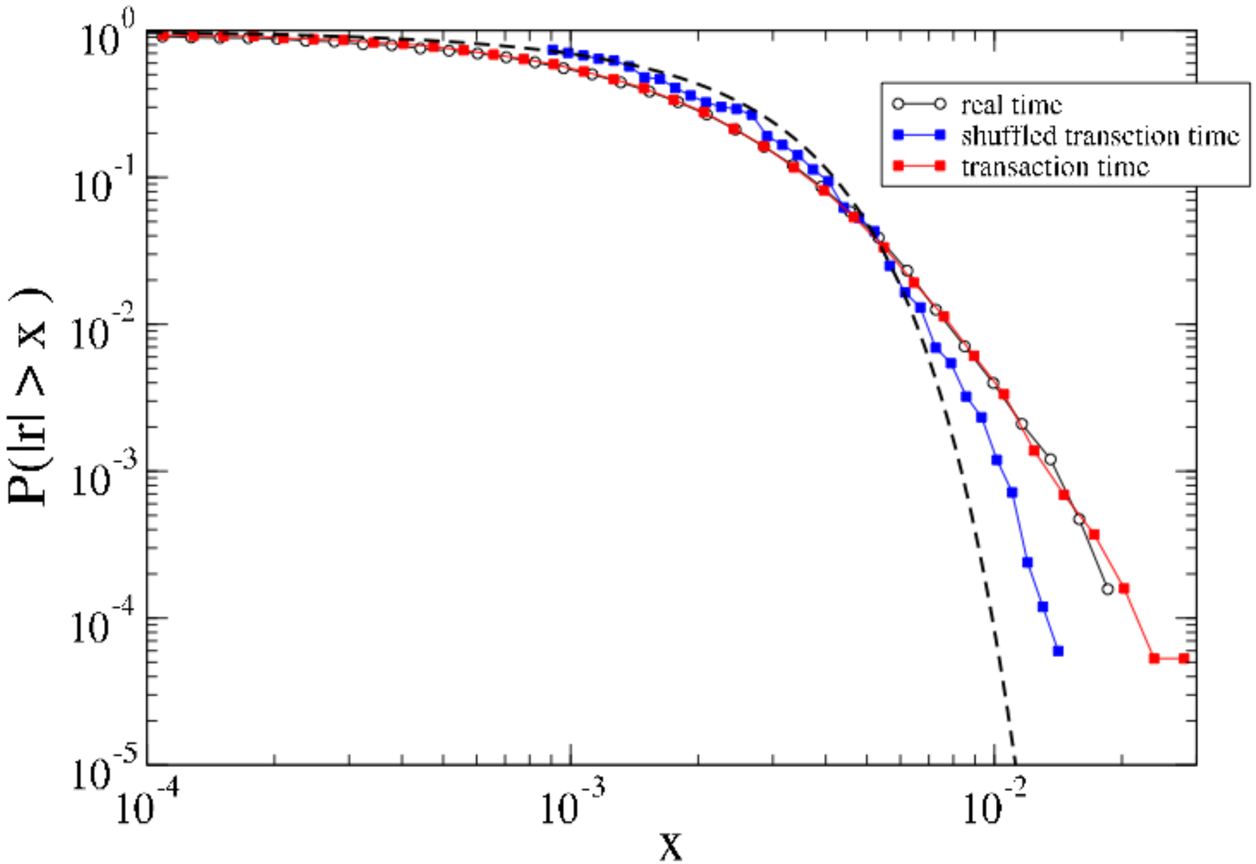
$$\tau_\theta(t_i) = \tau_\theta(t_{i-1}) + 1$$

where t_i is the time when transaction i occurs.

- We define **shuffled transaction time** as follows:
  - We associate to each trade the corresponding price change.
  - Then we randomly shuffle transactions. We do this so that we match the number of transactions in each real time interval, while preserving the unconditional distribution of returns but destroying any temporal correlations.

Which time reproduces better the real time volatility?
Transaction time or shuffled transaction time ?

Shuffling experiments: return distribution

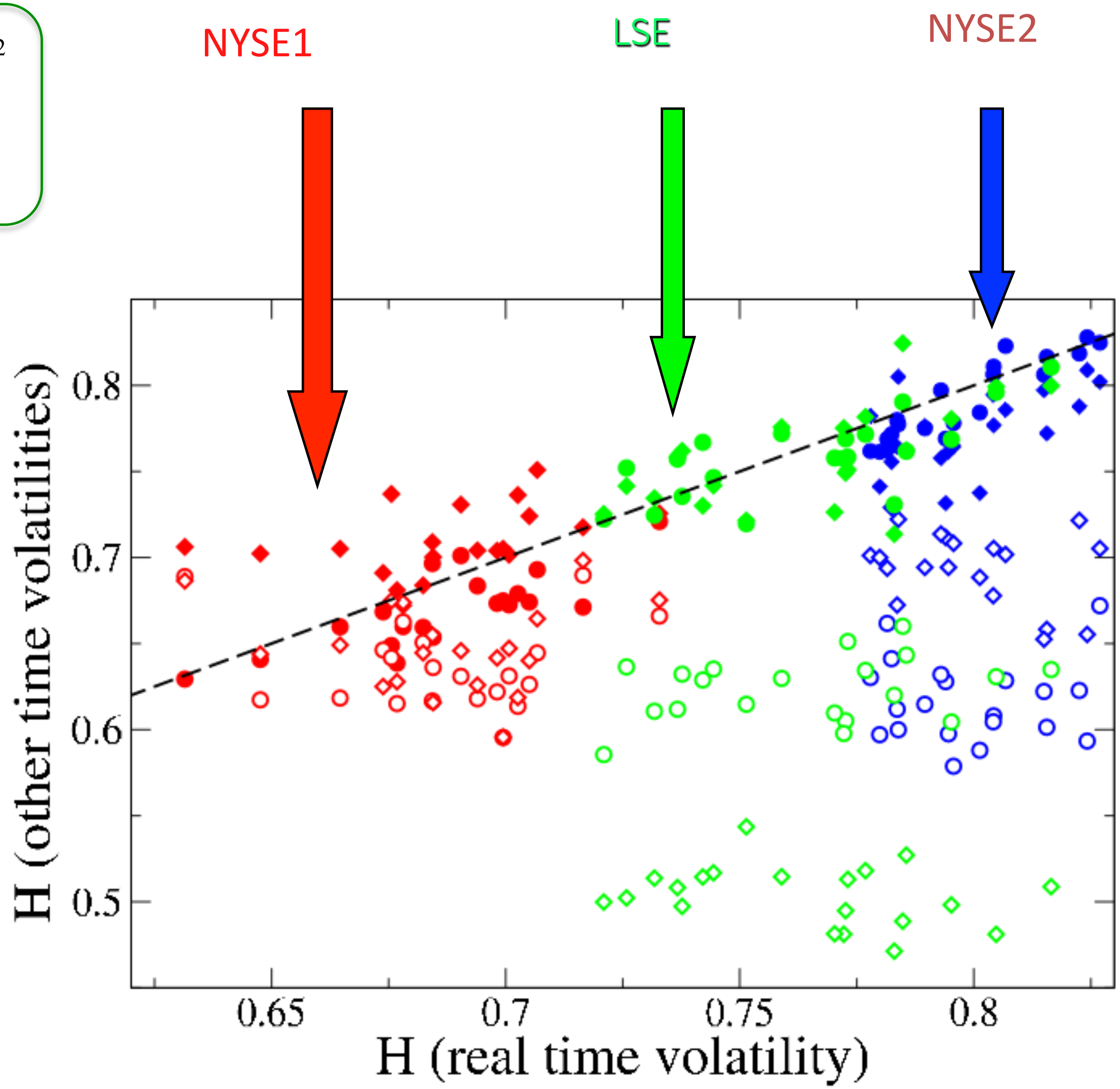


Clustered volatility

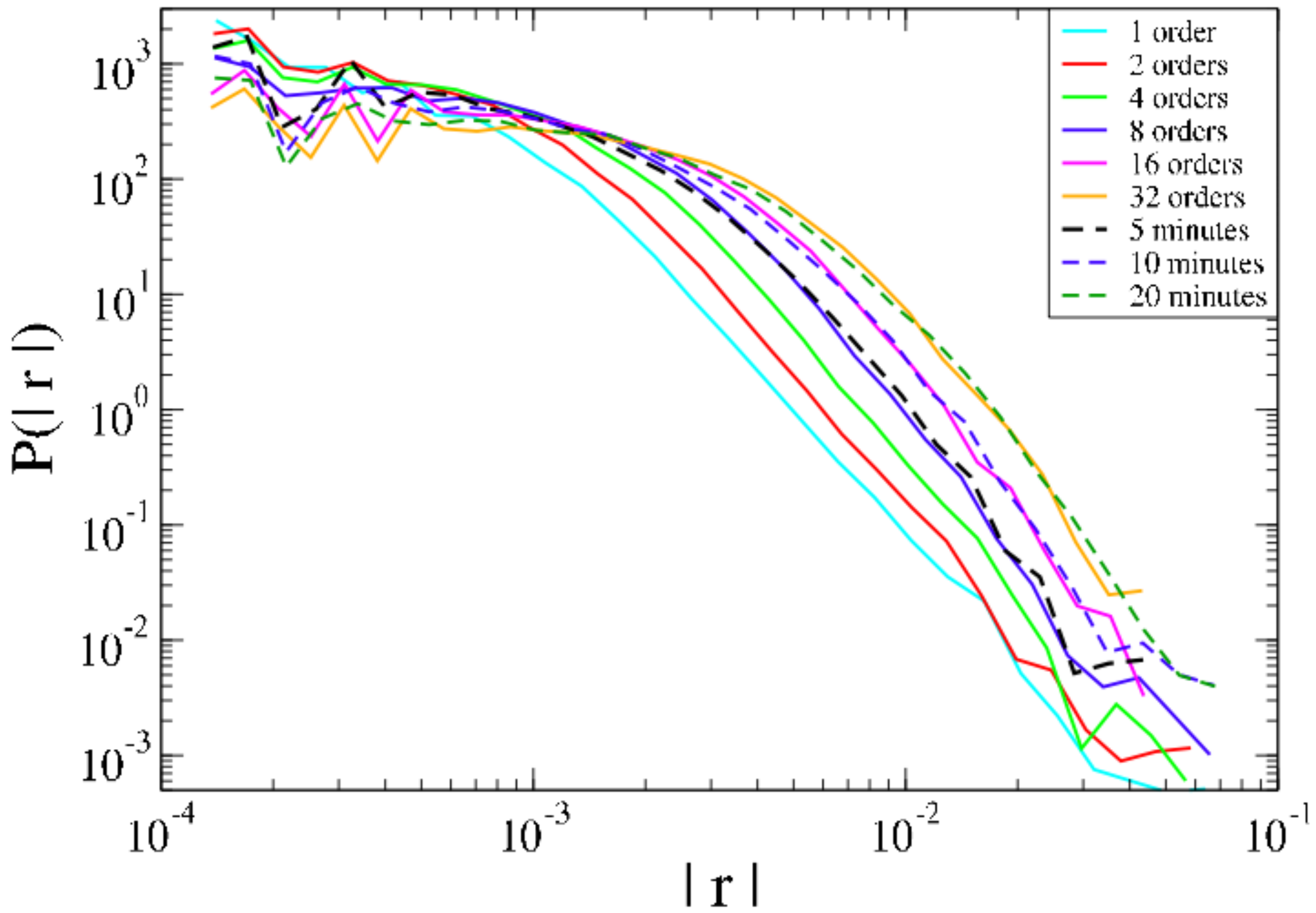
$$Corr[\sigma(t)\sigma(t+\tau)] \approx \tau^{2H-2}$$

0<H<1 Hurst exponent

- solid circles = trans. time
- solid diamond= volume time
- open circles= shuffled trans.time
- open diamond= shuffled volume time



Return distribution for fixed number of transactions



The type of aggregation (time or number of trades) does not matter

- Fat tails of return distribution and clustered volatility are closer to the real one in transaction (volume) time rather than in shuffled transaction (volume) time.
- These results indicate that the main drivers of heavy tails are the fluctuations of the price reaction to individual transactions.
- Tick size is also important as emphasized by the comparison of NYSE1 and NYSE2 dataset.
- Our analysis suggests that fluctuations in trading rate are not the most important determinant of return's fat tails and clustered volatility

Order flow

- One of the key problem in the analysis of markets is the understanding of the relation between the order flow and the process of price formation.
- The order flow is strongly depending on

  - the decisions and strategies of traders

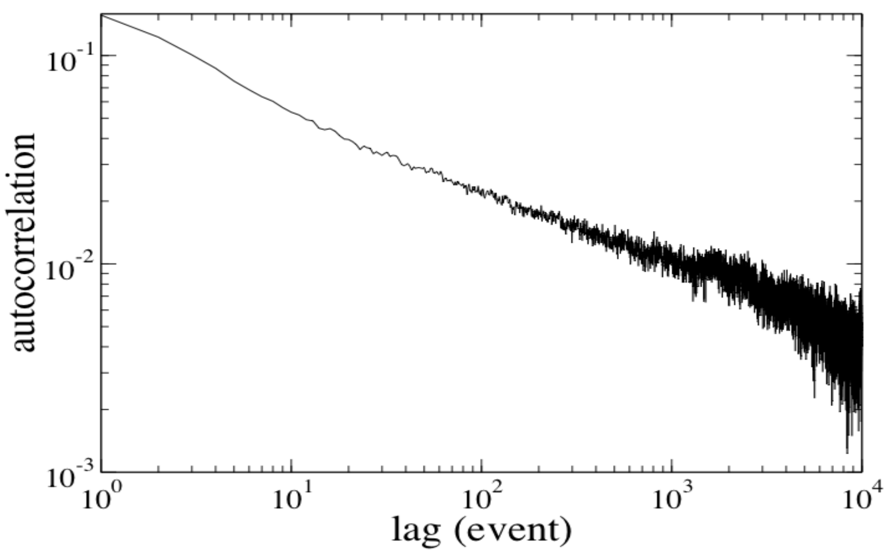
  - the exploiting of arbitrage opportunities
- This problem is thus related to the balance between supply and demand and to the origin of the market efficiency.

Market order flow

- We investigate the London Stock Exchange in the period 1999-2002
- We consider market orders, i.e. orders to buy at the best available price triggering a trade
- We consider the symbolic time series obtained by replacing buy orders with +1 and sell orders with -1
- The order flow is studied mainly in event time

....+1,+1,-1,-1,-1,+1,-1,+1,+1,+1,-1,-1,+1,...

Order flow dynamics

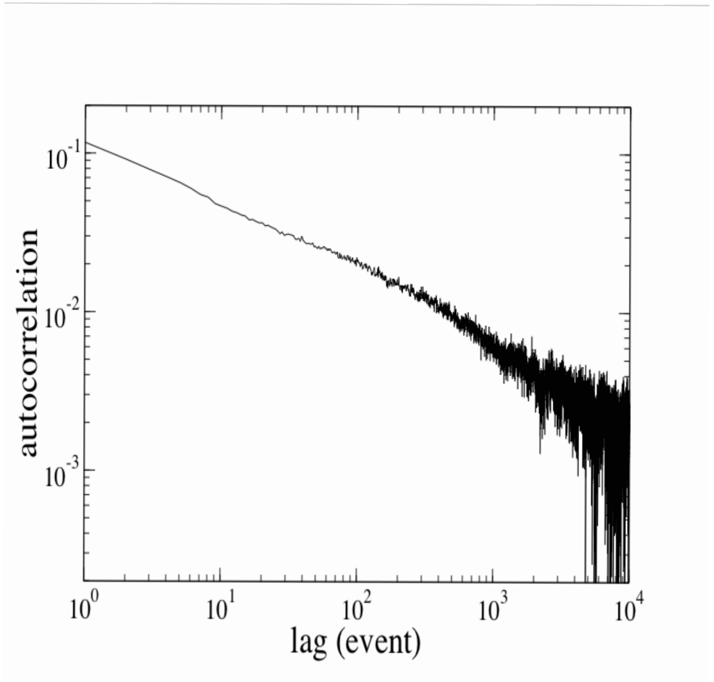


Time series of signs of market orders is a long memory process (Lillo and Farmer 2004, Bouchaud et al 2004)

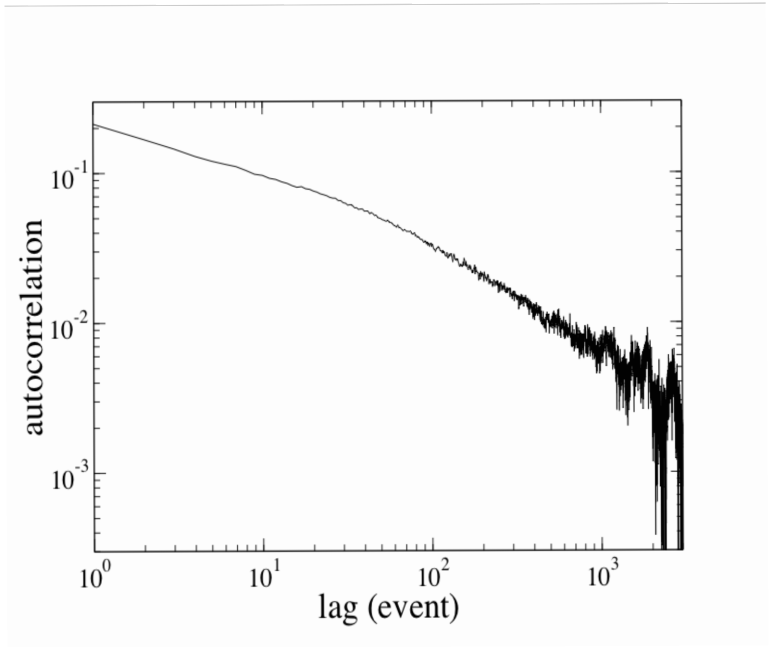
$$C(\tau) \approx \tau^{-\gamma} \approx \tau^{-0.5}$$

- Why is the order flow a long-memory process ?
- We show (empirically) that:
  - It is likely due to splitting
  - It requires a huge heterogeneity in agent size

Limit Orders



Cancellations



The sign time series of the three types of orders is a long-memory process

Hurst exponent →

$$\begin{cases} H_{mo} = 0.695 \pm 0.039 \\ H_{lo} = 0.716 \pm 0.054 \\ H_{ca} = 0.768 \pm 0.059 \end{cases}$$

What is the origin of long-memory in order flow?

Two explanations has been proposed

- Herding among market participants (LeBaron and Yamamoto 2007). Agents herd either because they follow the same signal(s) or because they copy each other trading strategies. Direct vs indirect interaction
- Order splitting (Lillo, Mike, and Farmer 2005). To avoid revealing true intentions, large investors break their trades up into small pieces and trade incrementally (Kyle, 1985). Convert heavy tail of large orders volume distributions in correlated order flow.

Is it possible to quantify **empirically** the contribution of herding and order splitting to the autocorrelation of order flow?
Note that this is part of the question on the origin of *diagonal effect* raised in Biais, Hillion and Spatt (1995).

Decomposing the autocorrelation function

Assume we know the identity of the investor placing any market order.

- For each investor i we define a time series of market order signs ϵ_t^i which is equal to zero if the market order at time t was not placed by investor i and equal to the market order sign otherwise
- The autocorrelation function can be rewritten as

$$C(\tau) = \frac{1}{N}\sum_t \sum_{i,j} \epsilon_t^i \epsilon_{t+\tau}^j - \left(\frac{1}{N}\sum_t \sum_i \epsilon_t^i\right)^2$$

Decomposing the autocorrelation function

We rewrite the acf as $C(\tau) = C_{split}(\tau) + C_{herd}(\tau)$ where

$$C_{split}(\tau) = \sum_i \left(P^{ii}(\tau) \left[\frac{1}{N^{ii}(\tau)} \sum_t \epsilon_t^i \epsilon_{t+\tau}^i \right] - \left[P^i \frac{1}{N^i} \sum_t \epsilon_t^i \right]^2 \right)$$

$$C_{herd}(\tau) = \sum_{i \neq j} \left(P^{ij}(\tau) \left[\frac{1}{N^{ij}(\tau)} \sum_t \epsilon_t^i \epsilon_{t+\tau}^j \right] - P^i P^j \left[\frac{1}{N^i} \sum_t \epsilon_t^i \right] \left[\frac{1}{N^j} \sum_t \epsilon_t^j \right] \right)$$

N^i is the number of market orders placed by agent i, $P^i = N^i/N$, $N^{ij}(\tau)$ is the number of the number of times that an order from investor i at time t is followed by an order from investor j at time $t+\tau$, and $P^{ij}(\tau) = N^{ij}(\tau)/N$

Market members

- At LSE there are typically 250-300 market members trading a stock. Of those roughly 80 are significantly active in a six month period.
- There is a huge heterogeneity in market member activity at LSE. The 15 most active ones are responsible for 80-90% of transactions.
- The activity of market members (independently from their trading direction) is characterized by the persistence

$$\tilde{P}^{ii}(\tau) = P^{ii}(\tau) - (P^i)^2$$

Market members persistence

Market member activity is highly clustered in (transaction) time. I.e. there is some degree of predictability that a member active now will be active in the near future.

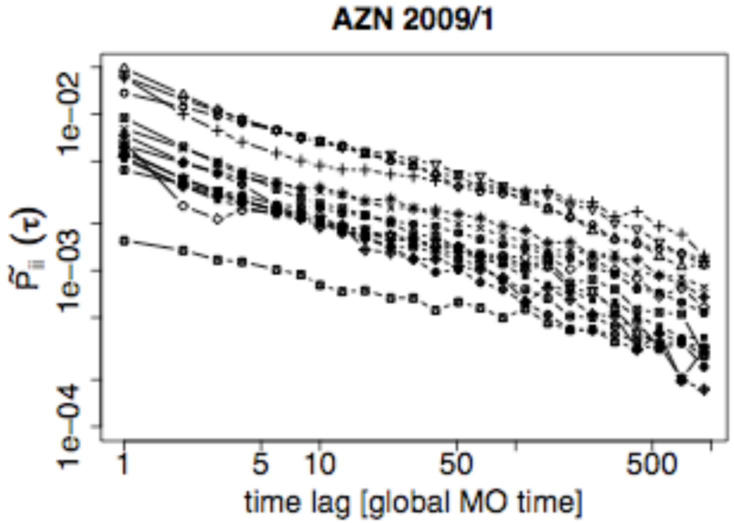


Figure: The diagonal terms of persistence in activity, i.e., $P^{ii}(\tau) - [P^i]^2$ of MO placement for the 15 most active participant codes, the first half of 2009 for AZN. For many codes this quantity is consistently positive, indicating a significant clustering in their activity.

Herding or splitting?

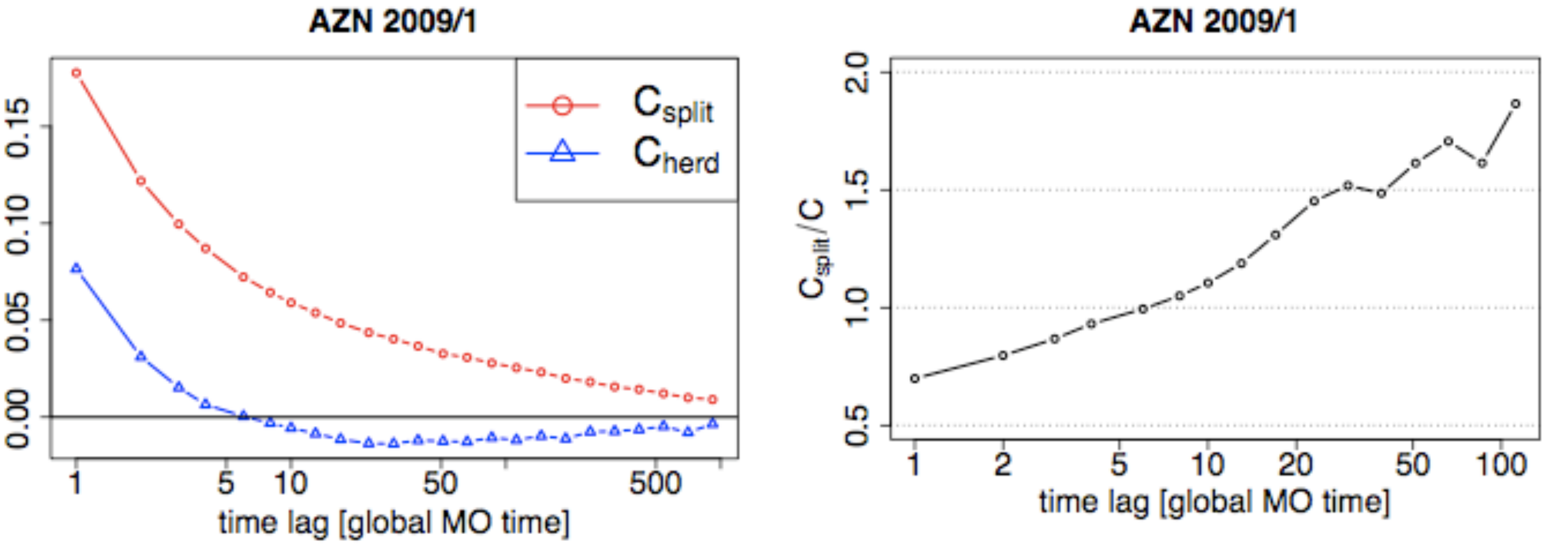


Figure: Left panel. The splitting and the herding term of the correlation of MO signs (the two terms sum up to $C(\tau)$) for the first half year of 2009 for AZN. Right panel. The splitting ratio of MO signs (defined as the ratio of the splitting term in the correlations and the entire correlation) for the first half year of 2009 for AZN.

Splitting dominates herding (especially for large lags)

Hidden orders

- In financial markets large investors usually need to trade large quantities that can significantly affect prices. The associated cost is called market impact
- 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 hidden orders 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

Model of order splitting

- There are N hidden orders (traders).
- An hidden order of size L is composed by L revealed orders
- The initial size L of each hidden order is taken by a given probability distribution P(L). The sign s_i (buy or sell) of the hidden order is initially set to +1 or -1 in a random way.
- At each time step an hidden order i is picked randomly and a revealed order of sign s_i is placed in the market. The size of the hidden order is decreased by one unit..
- When an hidden order is completely executed, a new hidden order is created with a new size and a new sign.

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

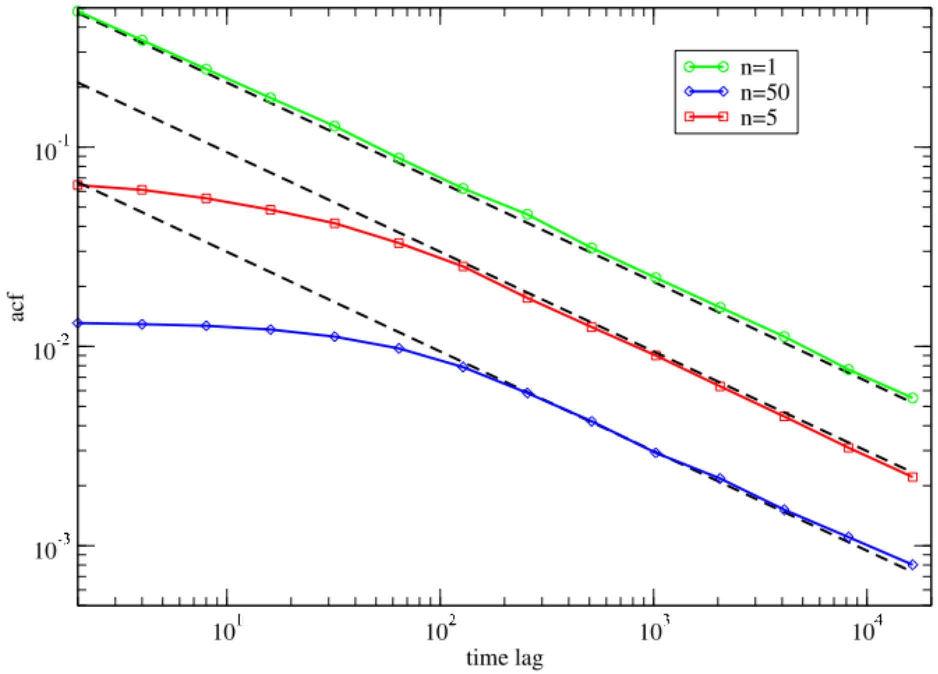
$$p(L) = \frac{\alpha}{L^{\alpha+1}} \qquad L \geq 1 \qquad \alpha > 1$$

The rationale behind this assumption is that

1. It is known that the market value of mutual funds is distributed as a Pareto distribution (Gabaix *et al.,* 2003)
2. It is likely that the size of an hidden order is proportional to the firm placing the order

We prove that the time series of the signs of the revealed order has an autocorrelation function decaying asymptotically a

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



Testing the models

• It is very difficult to test the model because it is difficult to have information on the size and number of hidden orders present at a given time.

• We try to cope with this problem by taking advantage of the structure of financial markets such as London Stock Exchange (LSE).

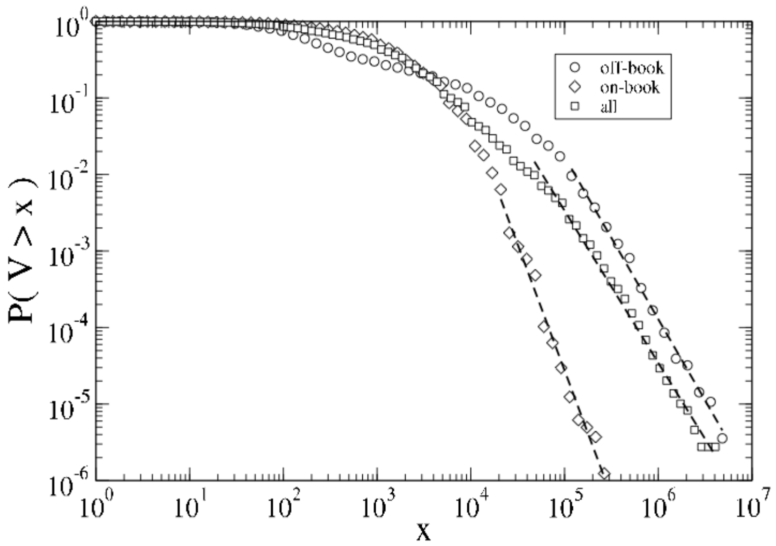
• At LSE there are two alternative methods of trading

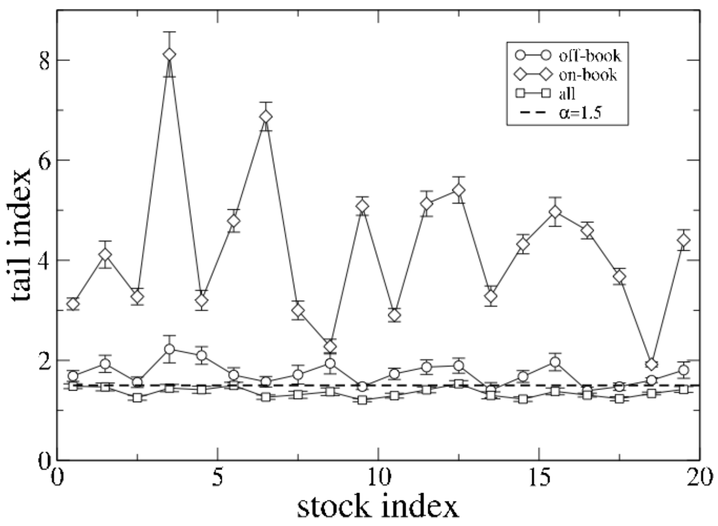
- The on-book (or downstairs) market is public and execution is completely automated (Limit Order Book)
- The off-book (or upstairs) market is based on personal bilateral exchange of information and trading.

We assume that revealed orders are placed in the on-book market, whereas off-book orders are proxies of hidden orders

Volume distribution

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 $\gamma=\alpha-1$.

Is it possible to identify directly hidden orders?

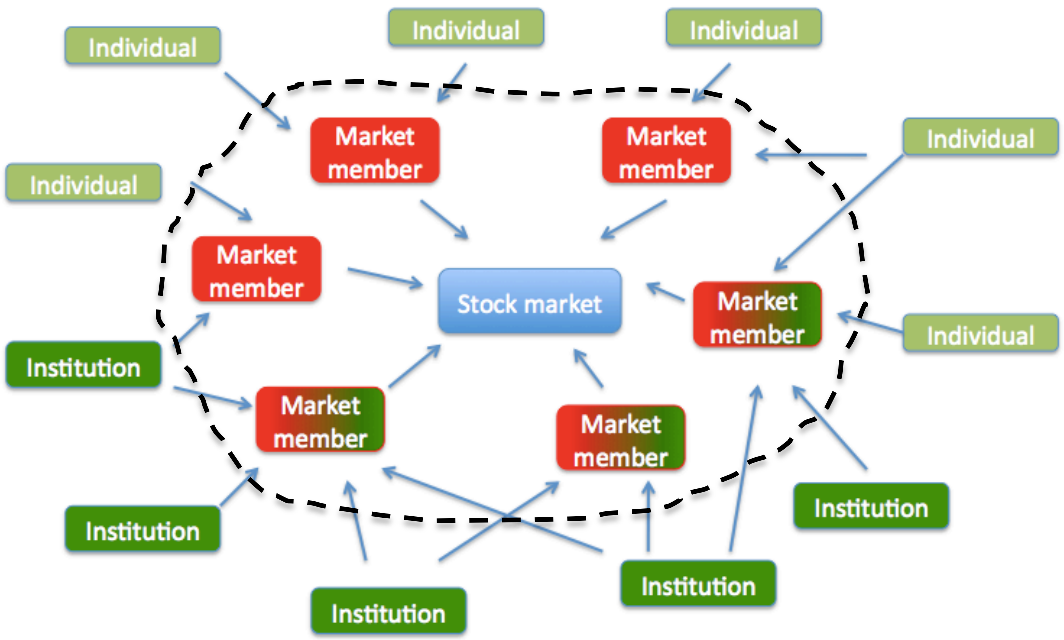
Our investigation

- The investigated market is the Spanish Stock Exchange (BME)

❑ Firms are credit entities and investment firms which are members of the stock exchange and are entitled to trade in the market.

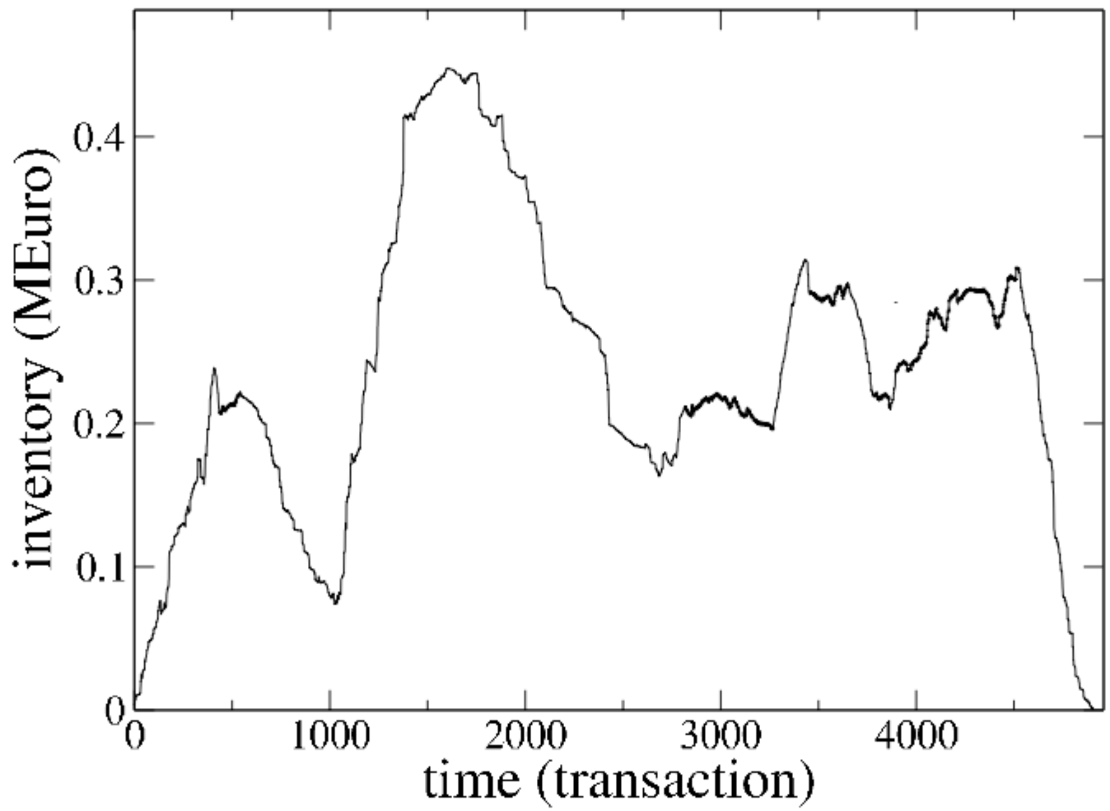
❑ 200 Firms in the BME (350/250 in the NYSE)

VALOR	VOLUMEN	PRECIO	SOCCOM	SOCVEN	HORA	FECHA
TEF	236	2187	9405	9858	90108	01/06/2000
TEF	1764	2187	9405	9487	90108	01/06/2000
ANA	110	3800	9839	9855	90109	01/06/2000
CAN	37	2194	9839	9578	90109	01/06/2000
CAN	151	2200	9839	9412	90109	01/06/2000
VIS	214	700	9821	9561	90109	01/06/2000
SOL	286	1299	9839	9838	90110	01/06/2000
ALB	104	2710	9839	9843	90110	01/06/2000
ALB	29	2719	9839	9419	90110	01/06/2000
ACX	97	3689	9839	9843	90111	01/06/2000
AGS	120	1445	9839	9487	90111	01/06/2000
AGS	110	1448	9839	9485	90111	01/06/2000
ACS	107	2930	9839	9863	90111	01/06/2000
SCH	11226	1045	9858	9880	90112	01/06/2000
CTE	96	1935	9839	9832	90112	01/06/2000
CTE	50	1955	9839	9872	90112	01/06/2000
CTE	14	1958	9839	9426	90112	01/06/2000
FER	237	1296	9839	9560	90112	01/06/2000
SGC	50	3980	9820	9560	90113	01/06/2000
ACR	161	1139	9839	9487	90113	01/06/2000
ACR	47	1140	9839	9845	90113	01/06/2000
DRC	20	803	9839	9573	90114	01/06/2000
DRC	267	805	9839	9484	90114	01/06/2000
AUM	111	1649	9839	9474	90114	01/06/2000



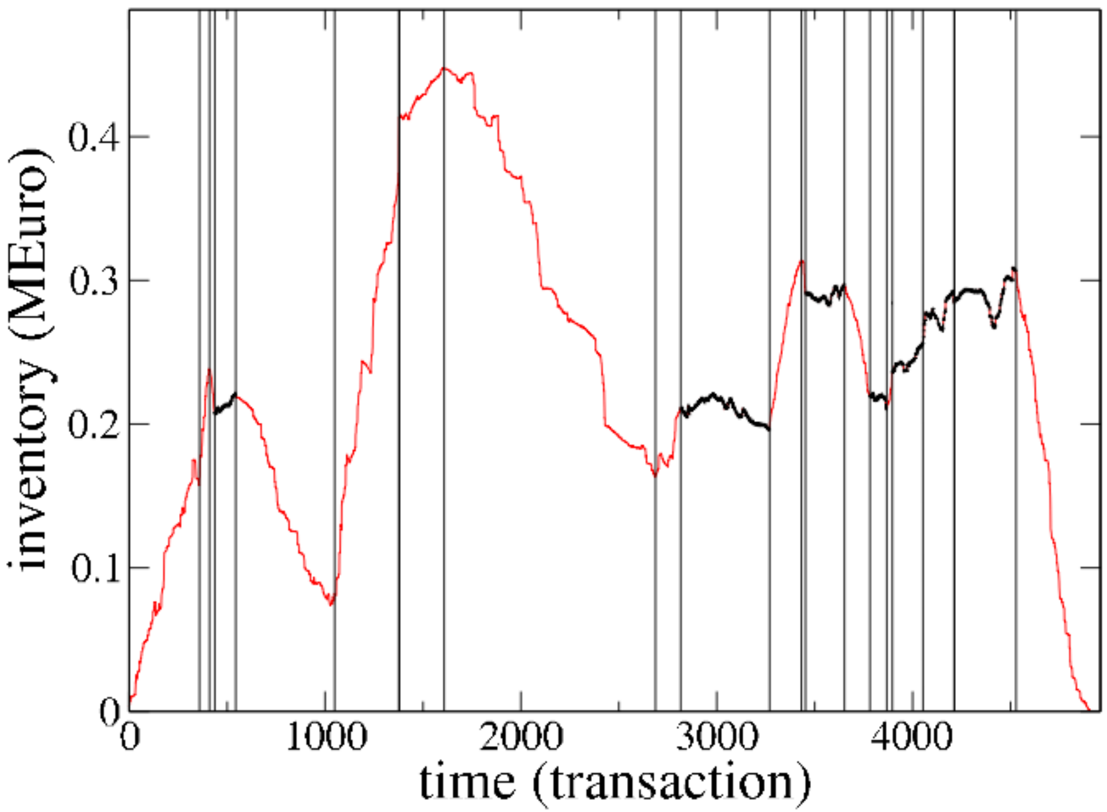
- Investigation at the level of market members and not of the agents (individuals and institutions
- The dataset covers the whole market
- The resolution is at the level of individual trade (no temporal aggregation)

A typical inventory profile



Credit Agricole trading Santander

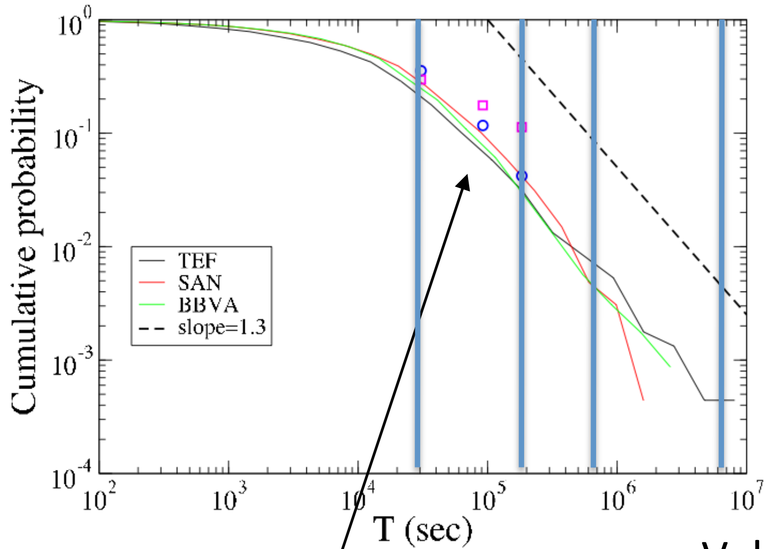
Detecting hidden orders



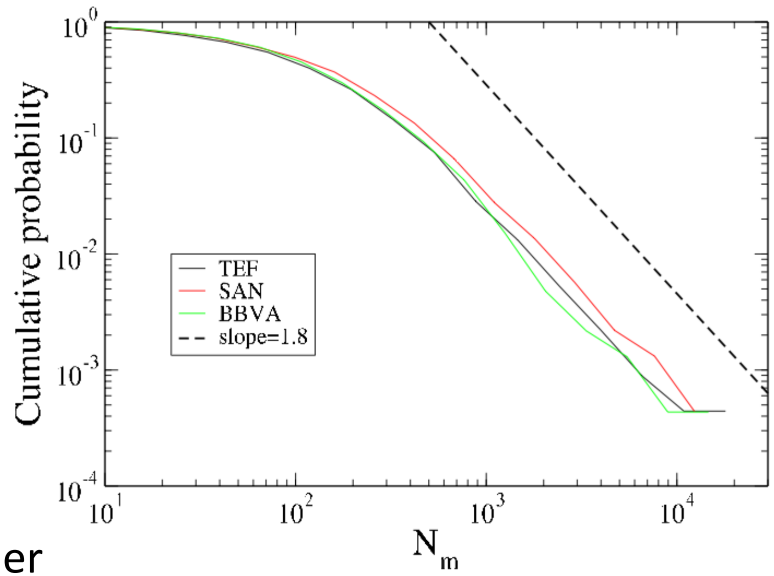
We developed a statistical method to identify periods of time when an investor was consistently (buying or selling) at a constant rate -> Hidden orders

Distributional properties of hidden orders

Investment horizon

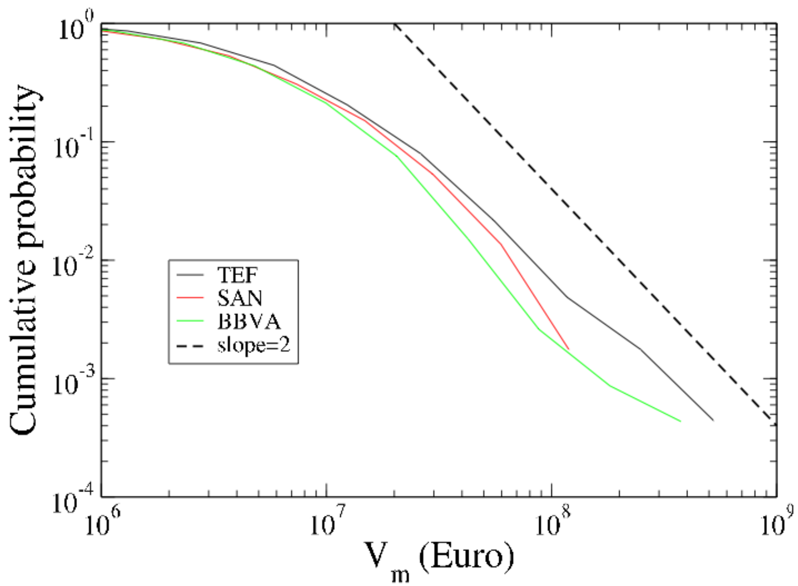


Number of transactions



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

Volume of the order



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

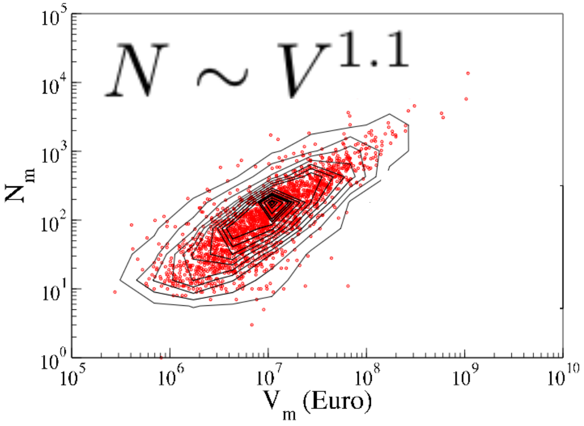
$$P(T > x) \sim \frac{1}{x^{1.3}} \quad P(N_m > x) \sim \frac{1}{x^{1.8}} \quad P(V_m > x) \sim \frac{1}{x^2}$$

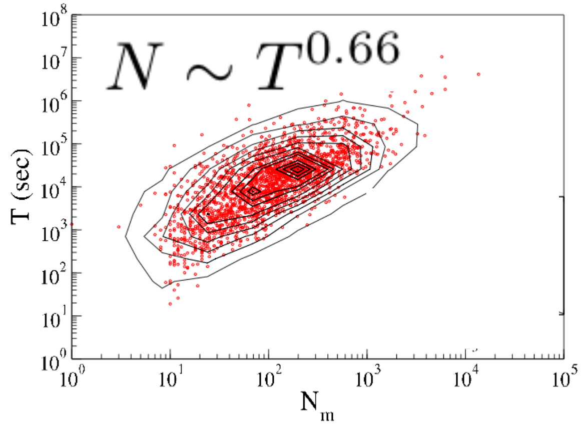
These results are not consistent with the theory of Gabaix et al. Nature 2003)

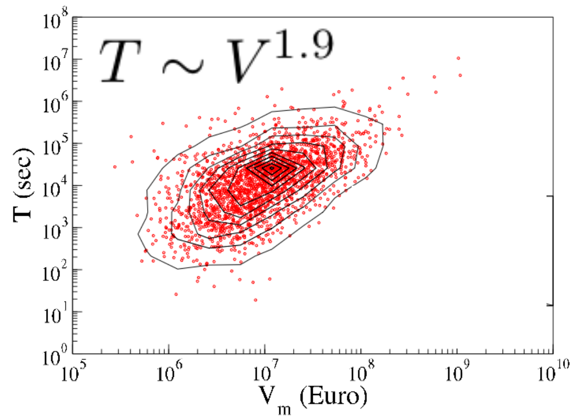
$$P(T > x) \sim \frac{1}{x^3} \qquad P(N_m > x) \sim \frac{1}{x^3} \qquad P(V_m > x) \sim \frac{1}{x^{3/2}}$$

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)

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

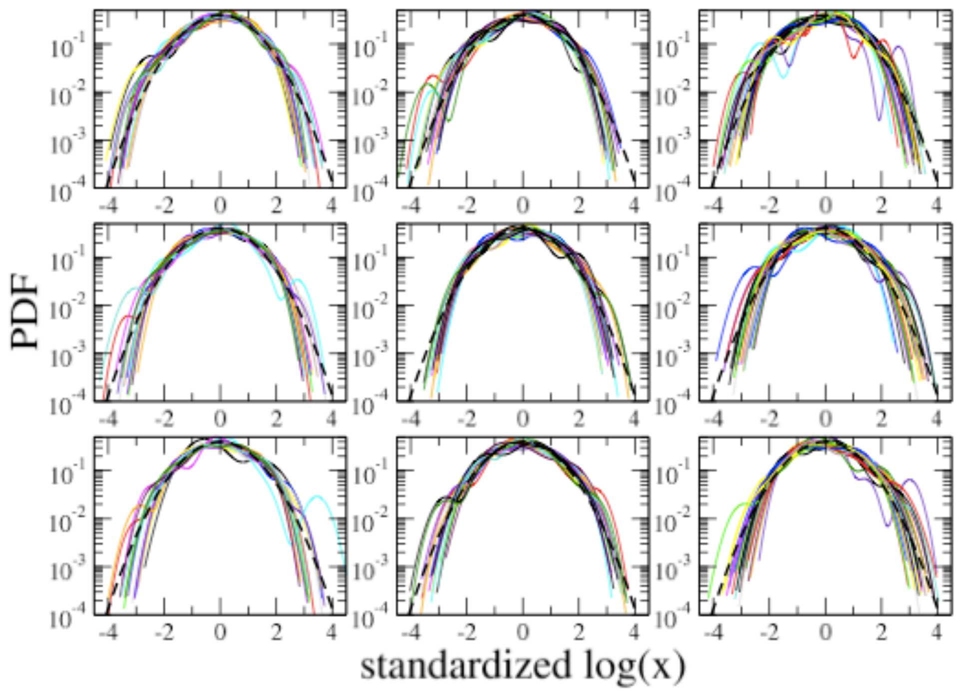


Figure 4.5: Probability density function of the standardized logarithm of the variables T, N_{maj} and V_{maj} of the firms for which the Jarque-Bera test of lognormality cannot be rejected. Specifically, for each stock and each variable we consider the firms for which the lognormal hypothesis cannot 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}).

- 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

Can we use the detected hidden orders to compute the market impact of hidden orders?

Market impact of hidden orders

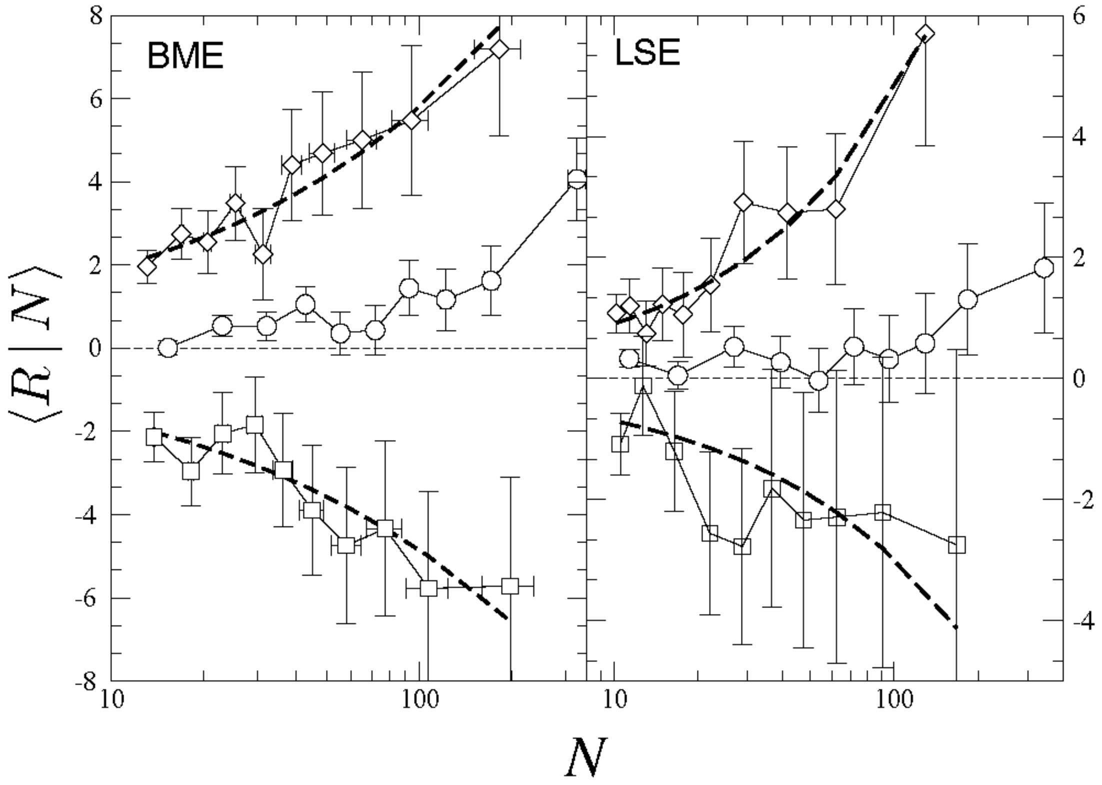


Figure 4: Average rescaled market impact R for hidden orders shorter than 1 day as a function of N for the BME (left) and LSE (right). Circles are the results for all hidden orders, while squares are the results when there is a low fraction of market orders ($f_{mo} < 0.2$) and diamonds are for when there is a large fraction of market orders ($f_{mo} > 0.8$). Dashed lines are power law fits $R \sim N^{\gamma}$. Values of γ are reported in Table II.

Moro, Vicente, Moyano, Gerig, Farmer, Vaglica, Lillo, Mantegna, Physical Review E 2009

Impact vs *N*

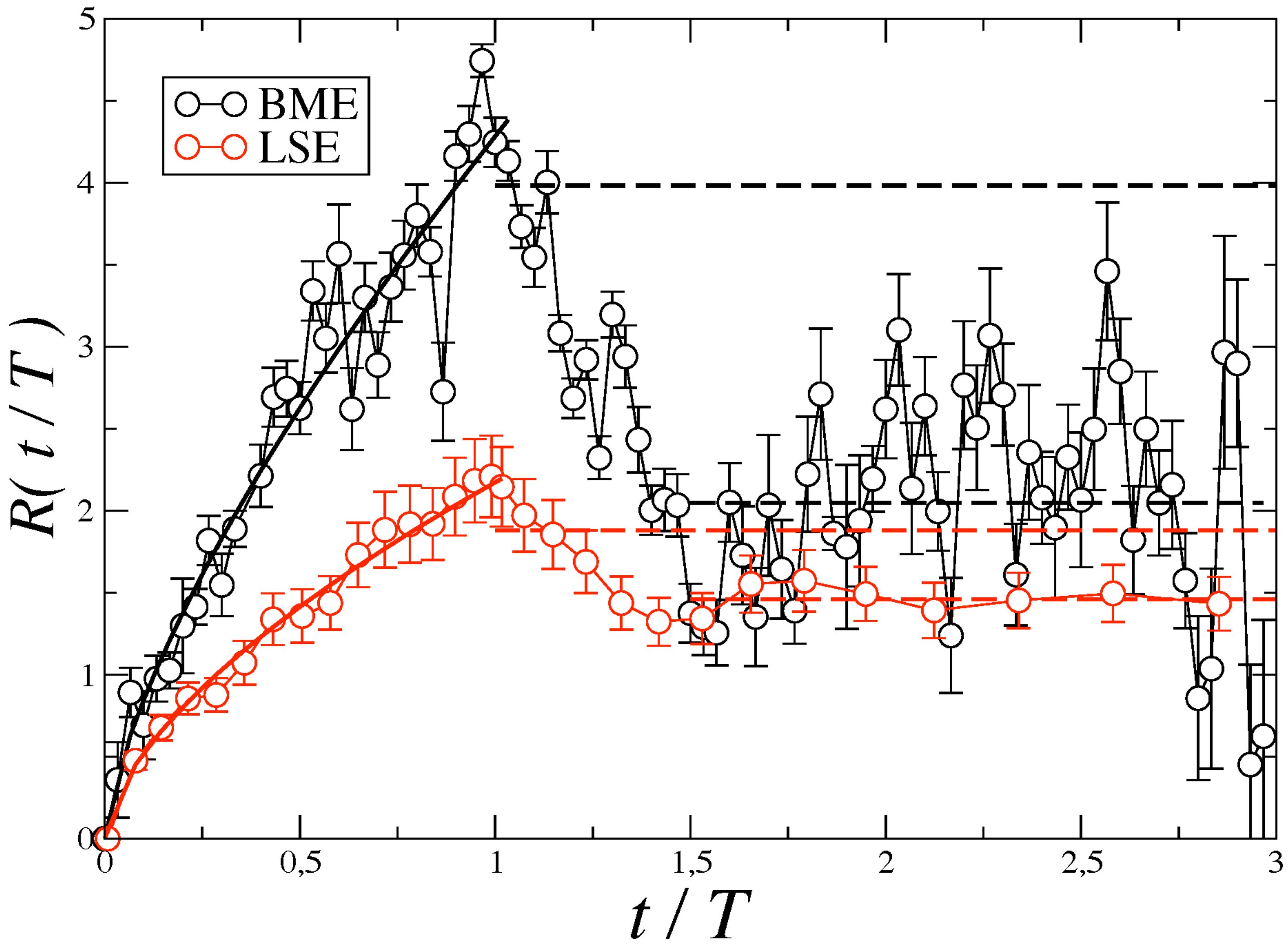
We find that for both groups the relation <R|N> is well described by:

$$|\langle R|N\rangle| = A\ N^{\gamma}$$

Table II: Parameters of the fitting of the market impact with Eq. 15.

Market	$A_{f_{mo}>0.8}$	$\gamma_{f_{mo}>0.8}$	$A_{f_{mo}<0.2}$	$\gamma_{f_{mo}<0.2}$
BME	0.63 ± 0.17	0.48 ± 0.07	-0.63 ± 0.22	0.44 ± 0.09
LSE	0.17 ± 0.05	0.72 ± 0.10	-0.16 ± 0.14	0.64 ± 0.30

Market impact versus time



Solid lines are power-law fits while dashed lines correspond to temporary (upper) and permanent (lower) market impact. Temporary impact R_{temp} is measured at the end of the hidden order t/T=1 while permanent impact R_{perm} is obtained through an average of *R(t/T)* with 1.5<t/T<3. Data are only for f_{mo}> 0.8.

Moro, Vicente, Moyano, Gerig, Farmer, Vaglica, Lillo, Mantegna, Physical Review E 2009

R_{perm} *and* R_{temp}

The power low fits give:

$$R \sim (4.28 \pm 0.21) \times \left(\frac{t}{T}\right)^{0.71 \pm 0.03} \qquad (BME)$$
$$R \sim (2.13 \pm 0.05) \times \left(\frac{t}{T}\right)^{0.62 \pm 0.02} \qquad (LSE)$$

The drop in impact is:

R_{perm}/R_{temp}= 0.51 ± 0.22 for BME

R_{perm}/R_{temp}= 0.73 ± 0.18 for LSE

Fair pricing condition

Suppose that the price after reversion is equal to the average price paid during execution.

If during execution price impact grows like $A\,X\,(t/\tau)^{\beta}$ then the average price paid by the agent who executes the order is:

$$\langle p \rangle = p_t + A \int_0^1 (t/T)^{\beta} d(t/T) = p_t + \frac{A}{1+\beta},$$

i.e. the permanent impact is $1/(1+\beta)$ of the peak impact

In our case by using the values of β obtained in the previous figure we get $1/(1+\beta) \approx 0.58 \pm 0.01$ for the BME and $1/(1+\beta) \approx 0.62 \pm 0.02$ for the LSE which are statistically similar to the ratios R_{perm}/R_{temp} for each market .

Long memory and efficiency

- **How can the long memory of order flow be compatible with market efficiency?**

- In the previous slides we have shown two empirical facts
  - Single transaction impact is on average non zero and given by

$$E\left[r|v\right] = sign(v)f(v) = \varepsilon f(v)$$

  - The sign time series is a long memory process

$$E\left[\varepsilon_t \varepsilon_{t+\tau}\right] \approx \tau^{-\gamma}$$

Naïve model

- Consider a naïve random walk model of price dynamics

$$p_{t+1} - p_t \equiv r_t = \varepsilon_t f(v_t) + \eta_t$$

- It follows that

$$E[r_t r_{t+\tau}] \propto E[\varepsilon_t \varepsilon_{t+\tau}] \approx \tau^{-\gamma}$$

- If market order signs ε_t are strongly correlated in time, price returns are also strongly correlated, prices are easily predictable, and the market inefficient.

- It is not possible to have an impact model where the impact is both fixed and permanent
- There are two possible modifications
  - A fixed but transient impact model (Bouchaud et al. 2004)
  - A permanent but variable (history dependent) impact model (Lillo and Farmer 2004, Gerig 2007, Farmer, Gerig, Lillo, Waelbroeck)

Fixed but transient impact model (Bouchaud et al 2004)

The model assumes that the price just after the (t-1)-th transaction is

$$p_t = p_{-\infty} + \sum_{k=1}^{\infty} G_0(k)\varepsilon_{t-k} f(v_{t-k}) + noise$$

and return is

$$r_t = p_{t+1} - p_t = G_0(1)\varepsilon_t f(v_t) + \sum_{k=1}^{\infty}[G_0(k+1) - G_0(k)]\varepsilon_{t-k} f(v_{t-k}) + noise$$

where the propagator $G_0(k)$ is a decreasing function.

The propagator can be chosen such as to make the market exactly efficient. This can be done by imposing that the volatility diffuses normally. The volatility at scale ℓ is

$$V_\ell \equiv E[(p_{n+\ell} - p_n)^2] = \sum_{j=0}^{\ell} G_0^2(\ell - j) + \sum_{j>0}[G_0(\ell + j) - G_0(j)]^2 + 2\Delta(\ell) + \Sigma^2 \ell$$

where Δ is a correlation-induced contribution

The correlation in the order flow decays as a power law with exponent γ

Assume that $G_0(\ell)$ itself decays at large ℓ as a power law, $\Gamma_0 \ell^{-\beta}$. When $\beta, \gamma < 1$, the asymptotic analysis of $\Delta(\ell)$ yields:

$$\Delta(\ell) \approx \Gamma_0^2 c_0 I(\gamma, \beta) \ell^{2-2\beta-\gamma} \quad \textbf{(2.27)}$$

where $I > 0$ is a certain numerical integral. If the single trade impact does not decay ($\beta = 0$), we recover the above superdiffusive result. But as the impact decays faster, superdiffusion is reduced, until $\beta = \beta_c = (1 - \gamma)/2$, for which $\Delta(\ell)$ grows exactly linearly with ℓ and contributes to the long-term value of the volatility. However, as soon as β exceeds β_c, $\Delta(\ell)$ grows sublinearly with ℓ, and impact only enhances the high-frequency value of the volatility compared to its long-term value Σ^2, dominated by "news." We therefore reach the conclusion that the long-range correlation in order flow does not induce long-term correlations nor anticorrelations in the price returns if and only if the impact of single trades is transient ($\beta > 0$) but itself nonsummable ($\beta < 1$). This is a rather odd situation in which the impact is not permanent (since the long-time limit of G_0 is zero) but is not transient either because the decay is extremely slow. The convolution of this semipermanent impact with the slow decay of trade correlations gives only a finite contribution to the long-term volatility. The mathematical constraint $\beta = \beta_c$ will be given more financial flesh later.

The model is able to make predictions on the response function defined as

$$R_\ell \equiv E[\varepsilon_n (p_{n+\ell} - p_n)]$$

which can be re-expressed in terms of the propagator and of the order sign correlation C_j

$$\mathcal{R}_\ell = G_0(\ell) + \sum_{0<j<\ell} G_0(\ell - j)C_j + \sum_{j>0} [G_0(\ell + j) - G_0(j)]\, C_j$$

From a mathematical point of view, the asymptotic analysis can again be done when $G_0(\ell)$ decays as $\Gamma_0 \ell^{-\beta}$. When $\beta + \gamma < 1$, one finds:

$$\mathcal{R}_\ell \approx_{\ell \gg 1} \Gamma_0 c_0 \frac{\Gamma(1-\gamma)}{\Gamma(\beta)\Gamma(2-\beta-\gamma)} \left[\frac{\pi}{\sin \pi\beta} - \frac{\pi}{\sin \pi(1-\beta-\gamma)}\right] \ell^{1-\beta-\gamma} \qquad \textbf{(2.29)}$$

where we have explicitly given the numerical prefactor to show that it exactly vanishes when $\beta = \beta_c$, which means that in this particular case one cannot satisfy oneself with the leading term. When $\beta < \beta_c$, one finds that $\mathcal{R}_\ell$ diverges to $+\infty$ for large ℓ, whereas for $\beta > \beta_c$, $\mathcal{R}_\ell$ diverges to $-\infty$, which is perhaps counterintuitive but means that when the decay of single trade impact is too fast, the accumulation of mean reverting effects leads to a negative long-term average impact—see Figure 2.7. When β is precisely equal to β_c, $\mathcal{R}_\ell$ tends to a finite positive value $\mathcal{R}_\infty$: The decay of single trade impact precisely offsets the positive correlation of the trades.

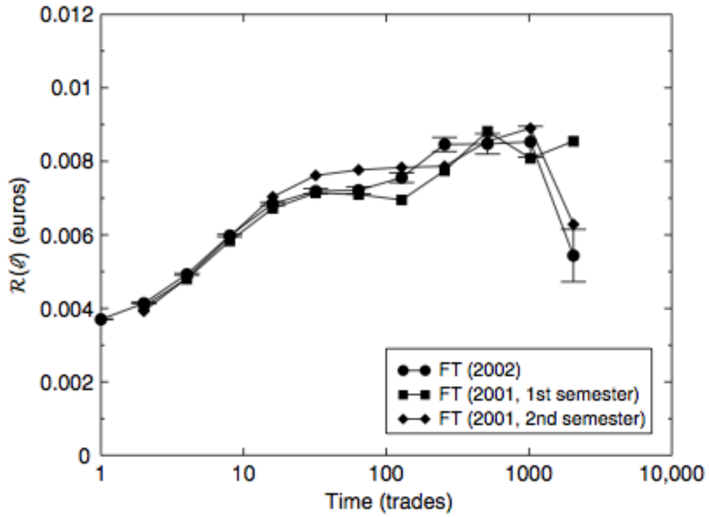


FIGURE 2.8 Average empirical response function $\mathcal{R}_\ell$ for FT during three different periods (1st and 2nd semester of 2001 and 2002); error bars are shown for the 2002 data. For the 2001 data, the y axis has been rescaled such that $\mathcal{R}_1$ coincides with the 2002 result. $\mathcal{R}_\ell$ is seen to increase by a factor ~2 between $\ell = 1$ and $\ell = 100$.

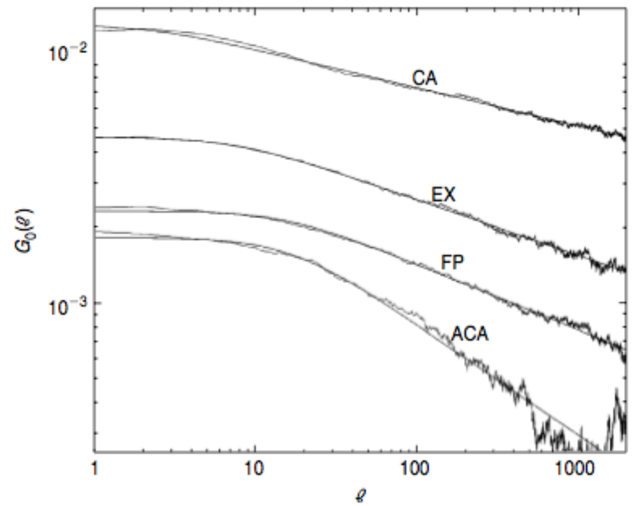


FIGURE 2.9 Comparison between the empirically determined $G_0(\ell)$, extracted from $\mathcal{R}$ and $\mathcal{C}$ using Eq. 2.28, and the power-law fit $G_0^f(\ell) = \Gamma_0/(\ell_0^2 + \ell^2)^{\beta/2}$ for a selection of four stocks: ACA, CA, EX, and FP.

History dependent, permanent impact model

- We assume that agents can be divided in three classes
  - Directional traders (liquidity takers) which have large hidden orders to unload and create a correlated order flow
  - Liquidity providers, who post bid and offer and attempt to earn the spread
  - Noise traders
- The strategies of the first two types of agents will adjust to remove the predictability of price changes

Model for price diffusion

We neglect volume fluctuations and we assume that the naïve model is modified as

$$p_{t+1} - p_t \equiv r_t = \theta(\varepsilon_t - \hat{\varepsilon}_t) + \eta_t \qquad \hat{\varepsilon}_t = E_{t-1}[\varepsilon_t | \Omega]$$

where Ω is the information set of the liquidity provider.

Ex post there are two possibilities, either the predictor was right or wrong
Let p^+_t (p^-_t) be the probability that the next order has the same (opposite) sign of the predictor and r^+_t (r^-_t) are the corresponding price change

- The efficiency condition $E_{t-1}[r_t|\Omega]=0$ can be rewritten as

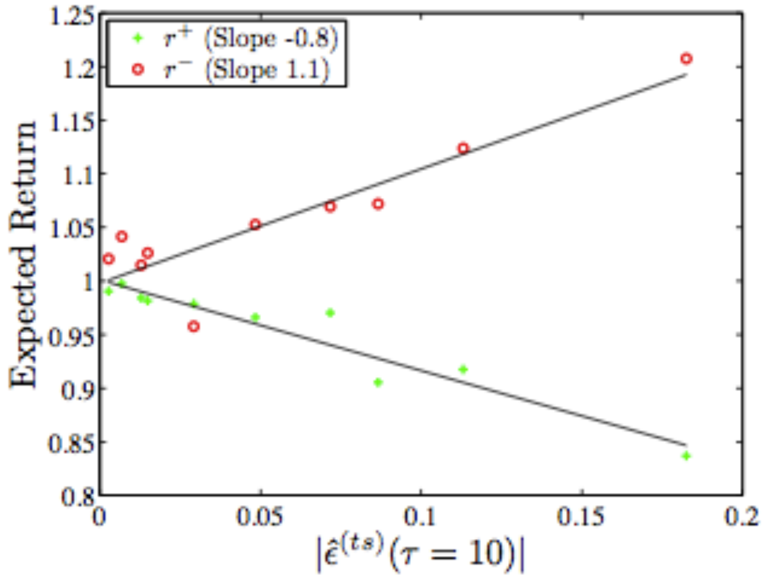
$$p_t^+ r_t^+ - p_t^- r_t^- = 0$$

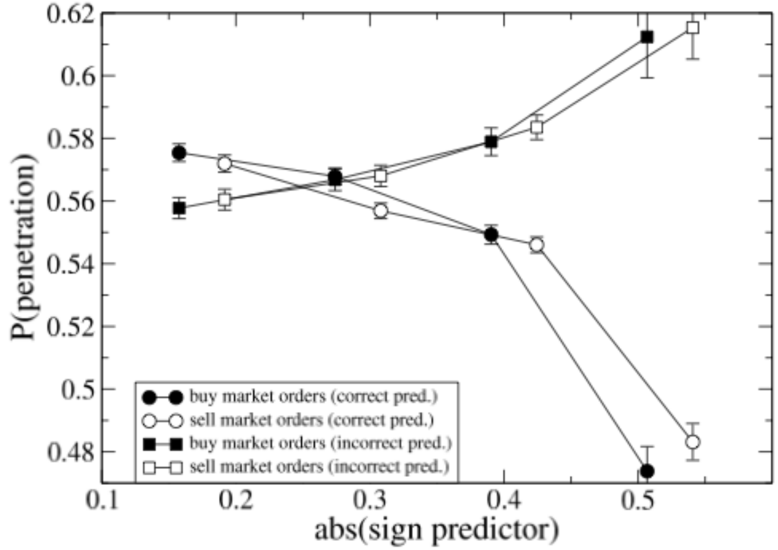
- The market maker has expectations on p^+_t and p^-_t given her information set Ω and adjusts r^+_t and r^-_t in order to make the market efficient

-----> MARKET EFFICIENCY

ASYMMETRIC LIQUIDITY MODEL

Empirical evidence of asymmetric liquidity





A linear model

The history dependent, permanent model is completely defined when one fixes

- the information set Ω of the liquidity provider
- the model used by the liquidity provider to build her forecast $\hat{\varepsilon}_t$

As an important example we consider the case in which

- the information set is made only by the past order flow
- the liquidity provider uses a finite or infinite order autoregressive model to forecast order flow

$$r_t = \theta\left(\varepsilon_t - \sum_{i=1}^{K} a_i \varepsilon_{t-i}\right) + \eta_t$$

If the order flow is long memory, i.e. $E[\varepsilon_t \varepsilon_{t+\tau}] \approx \tau^{-\gamma}$ the optimal parameters of the autoregressive model are $a_k \approx k^{\frac{\gamma-3}{2}} \equiv k^{-\beta-1}$ and the number of lags K in the model should be infinite.

If, more realistically, K is finite the optimal parameters of the autoregressive model follows the same scaling behavior with k

Under these assumptions and if K is infinite the linear model becomes mathematically equivalent to the fixed-temporary model (or propagator) model by Bouchaud et al. with

$$\theta a_i = G(i+1) - G(i) \quad \text{or} \quad G(i) = \theta[1 - \sum_{j=1}^{i-1} a_j]$$

Impact of hidden orders

The above model allows to make quantitative prediction on the impact of an hidden order

Assume an hidden order of length N is placed by a liquidity taker by using a slice and dice strategy which mixes the trades with the flow of noise traders with a constant participation rate π

The impact of the hidden order is

$$E\left[p_N\right] - p_0 \approx \varepsilon\theta\left(1 + \frac{2^{\beta-1}\pi^{\beta}}{1-\beta}\left[(2N-1)^{1-\beta} - 1\right]\right) \approx \pi^{\beta} N^{1-\beta}$$

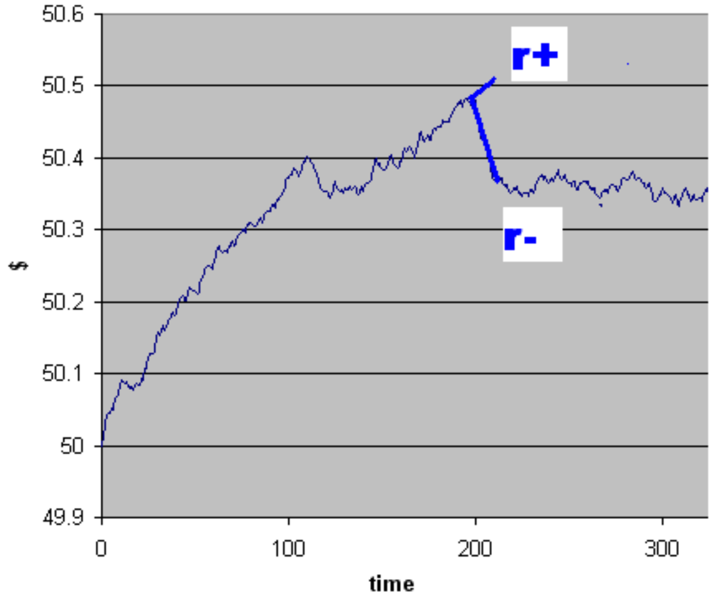
An empirical value γ=0.5, gives β=0.25,which in turn implies that the impact of an hidden order should grow as the ¾ power of its size. Moreover, as expected, the impact is smaller for slower execution (i.e. smaller π)

Permanent impact

The model allows to compute the permanent impact, i.e. the price change after the price has relaxed back to its long term value

If π=1 then

$$E\left[p_{\infty}\right]-p_{0}=\varepsilon\theta N\frac{4^{H-1}\sqrt{\pi}\Gamma(H)\sec\left[(K-H)\pi\right]}{\Gamma(3/2+K-H)\Gamma(2H-1-K)}\approx\varepsilon\theta\frac{N}{K^{\beta}}$$



The permanent impact is linear and vanishes only if K is infinite, recovering the Bouchaud et al idea of a completely temporary market impact

- Asymmetric liquidity depends on the information set Ω.
- This model predicts the existence of two classes of traders that are natural counterparties in many transactions
  - Large institutions creates predictable component of order flow by splitting their large hidden orders
  - Hedge funds and high frequency traders removes this predictability by adjusting liquidity (and making profit)
- This ecology of market participants is empirically detectable?
- What is the interaction pattern between market participants?

Coping with heterogeneity

- One of the distinguishing features of physics is that it has to do with collections of entities which are identical one to the other. There is no way of distinguishing an electron from another, given that they are identical in essence.
- The representative agent paradigm assumes "that the choices of all the diverse agents in one sector can be considered as the choices of one 'representative' standard utility maximizing individual whose choices coincide with the aggregate choices of the heterogeneous individuals" (Kirman 1992). This paradigm has been severely criticized.
- In my opinion, heterogeneity is an open problem and the challenge is to find classes of models of an economy with heterogeneous interacting agents.

A possible approach

- A different approach is to investigate real systems for which detailed data on the behavior of agents is recorded -> agent based empirical modeling
- This approach has not been followed very often due to the lack of data
- One of the difficulties is that one has to infer strategies, preferences and sometimes payoffs from data on agents' action.
- We use this approach for the investigation of financial markets

Limit order placement

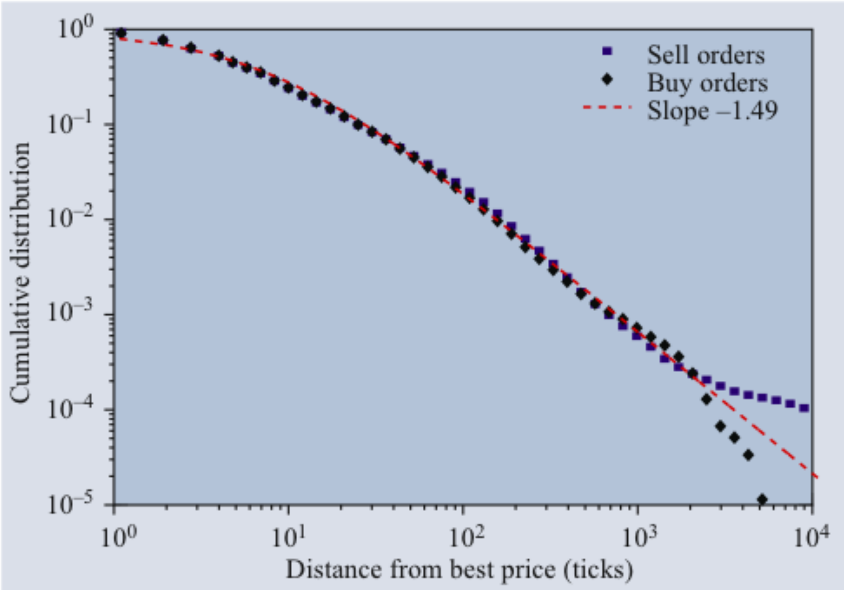


Figure 2. An estimate of the cumulative probability distribution based on a merged data set, containing the relative limit-order sizes $\delta(t)$ for all 50 stocks across the entire sample. The solid curve is a nonlinear least squares fit to the logarithmic form of equation (1).

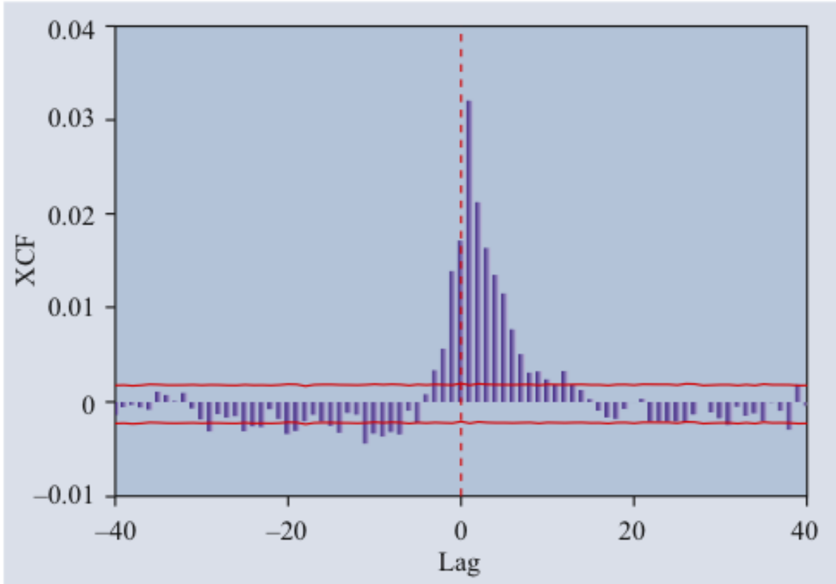


Figure 5. The cross autocorrelation of the time series of relative limit prices $\delta(t)$ and volatilities $v(t-\tau)$, averaged across all 50 stocks in the sample.

(from Zovko and Farmer 2002)

Limit order price is power law distributed with a tail exponent in the range 1.5-2.5
Moreover limit order price is correlated with volatility

An utility maximization argument

For a given limit price Δ, volatility σ, and investment horizon T the investor is faced with the lottery

$$\begin{cases} \Delta & \text{with probability} \quad \operatorname{erfc}\left[\frac{\Delta}{\sqrt{2\sigma^2 T}}\right] \\ 0 & \text{with probability } 1 - \operatorname{erfc}\left[\frac{\Delta}{\sqrt{2\sigma^2 T}}\right] \end{cases}$$

From first passage time of the price random walk

So the expected utility is $U_{T,\sigma}(\Delta) = \operatorname{erfc}\left[\frac{\Delta}{\sqrt{2\sigma^2 T}}\right] u(\Delta)$ where u(Δ) is the utility function.

Agent optimizes her limit order placement by choosing the lottery (i.e. the value of Δ) which maximizes the expected utility.

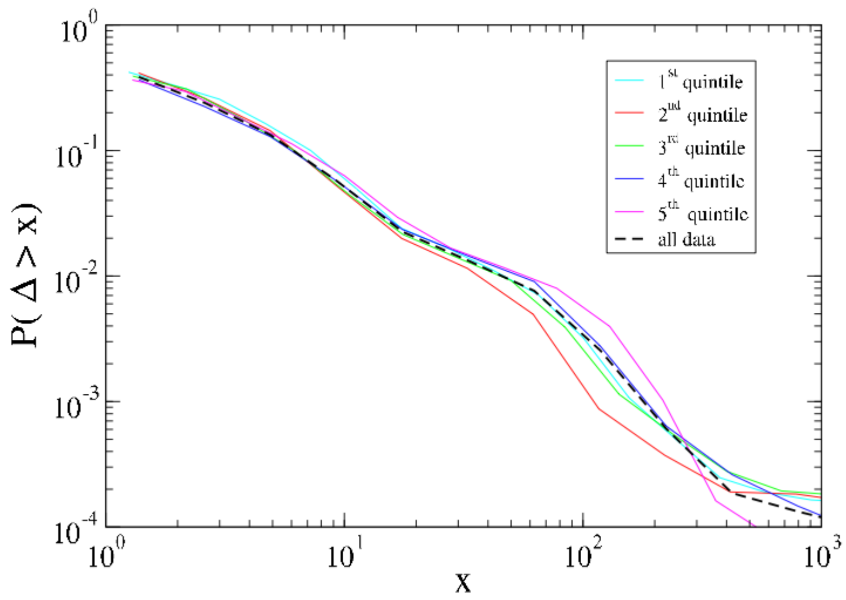
For several choices of u(Δ) it is possible to solve the problem analytically. For example for power utility function $u(x)=C\, x^{\alpha}$

$$\Delta^* = \sqrt{2} g^{-1}(\alpha)\sigma T^{1/2}$$

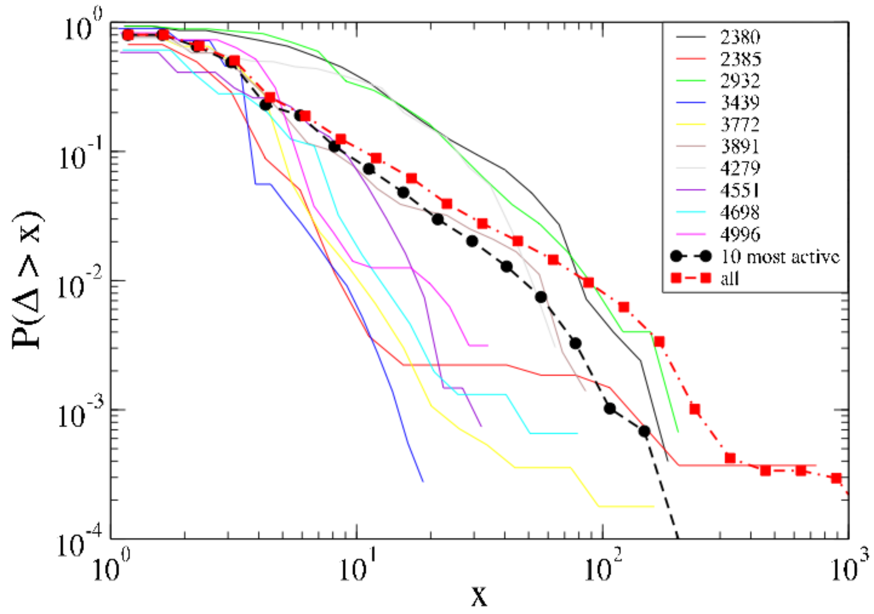
Some variable must be highly fluctuating: market (σ) or agent (α or T)?

It is possible to show that heterogeneity in volatility or utility function cannot explain the value of the exponent empirically observed.
The only possibility is a strong heterogeneity in investment horizon T

$$P_T(T) \sim \frac{1}{T^{\zeta_T}} \qquad \zeta_T \simeq 1.3 \div 1.7$$



Heterogeneity in volatility



Heterogeneity in traders preferences

Curiously, the same exponent of the time scale distribution is obtained by considering the time to fill of limit orders, metaorder splitting, and a modified GARCH (Borland and Bouchaud 2005)

How the market adapt to your trading?

- We decompose the total impact of a given type of order book event into a contribution from the same broker and a contribution from all other brokers.

Response function -> $$\mathcal{R}_{\pi_1}(\ell) = \frac{\langle (p_{t+\ell} - p_t) I(\pi_t = \pi_1) \epsilon_t \rangle}{P(\pi_1)}.$$

$$\mathcal{R}^{\text{same}}_{\pi_1}(\ell) = \frac{\left\langle \sum_{t'=t}^{t+\ell-1} (p_{t'+1} - p_{t'}) I(b_{t'} = b_t) I(\pi_t = \pi_1) \epsilon_t \right\rangle}{P(\pi_1)}.$$

$$\mathcal{R}^{\text{same}}_{\pi_1}(\ell) + \mathcal{R}^{\text{diff}}_{\pi_1}(\ell) = \mathcal{R}_{\pi_1}(\ell).$$

$$\mathcal{R}^{\text{diff}}_{\pi_1}(\ell) = \frac{\left\langle \sum_{t'=t}^{t+\ell-1} (p_{t'+1} - p_{t'}) I(b_{t'} \neq b_t) I(\pi_t = \pi_1) \epsilon_t \right\rangle}{P(\pi_1)}.$$

(with B. Toth, Z. Eisler, J. Kockelkoren, J.-P. Bouchaud, and J.D. Farmer)

Response function is a delicate balance

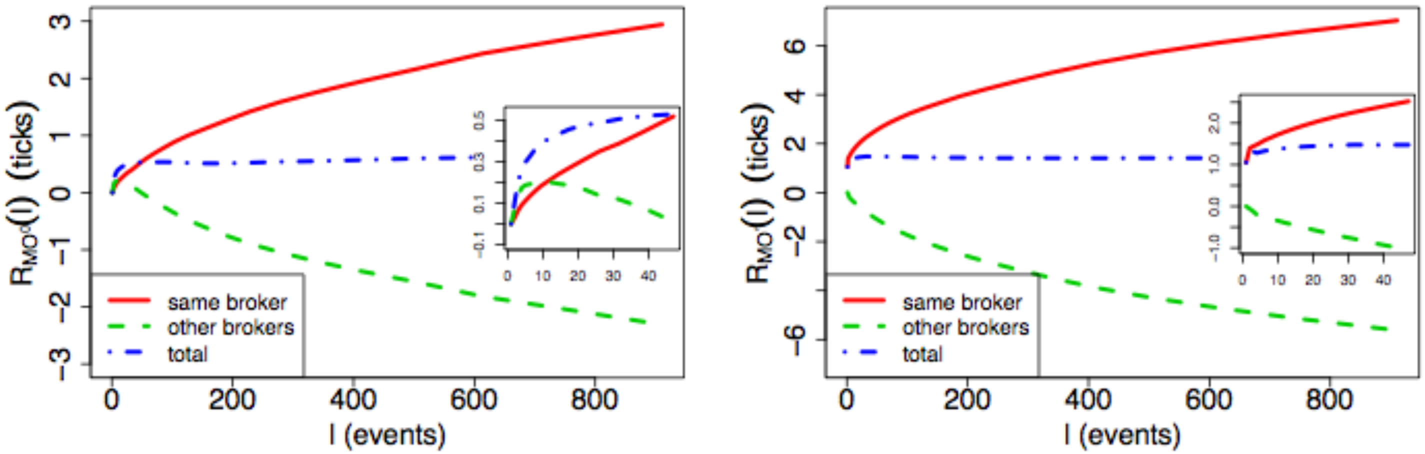


Figure 1: The response function $\mathcal{R}_{\pi_1}(\ell)$ and its contributions coming from the orders of the same broker ($\mathcal{R}^{\text{same}}_{\pi_1}(\ell)$) and of different brokers ($\mathcal{R}^{\text{diff}}_{\pi_1}(\ell)$). (left) The case of $\pi_1 = \text{MO}^0$. (right) The case of $\pi_1 = \text{MO}'$. The insets show a zoom for small ℓ.

These two contributions very nearly offset each other, leading to a total impact that is nearly constant in time and much smaller than both these contributions.

Dynamical liquidity picture -> the highly persistent sign of market orders must be buffered by a fine-tuned counteracting limit order flow in order to maintain statistical efficiency (i.e. that the price changes are close to unpredictable, in spite of the long-ranged correlation of the order flow).

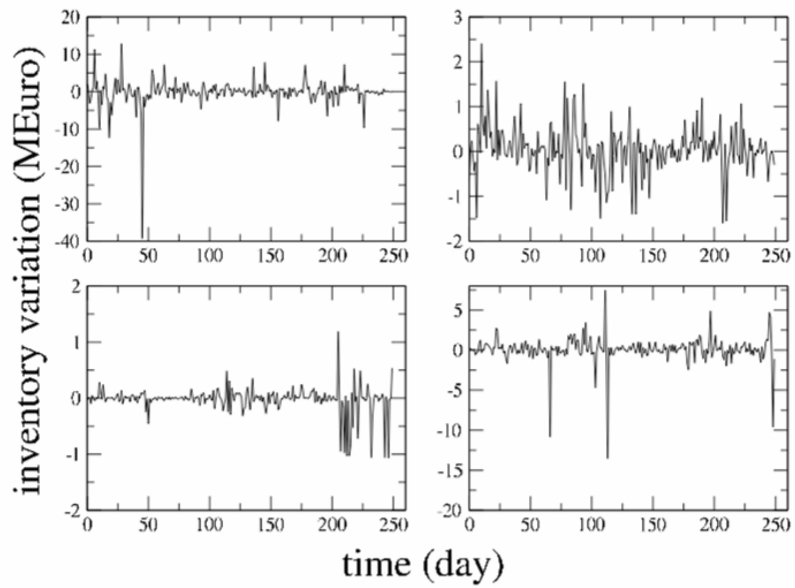
Daily inventory variation time series

We quantify the trading activity of a firm in a given time period τ by introducing the inventory variation

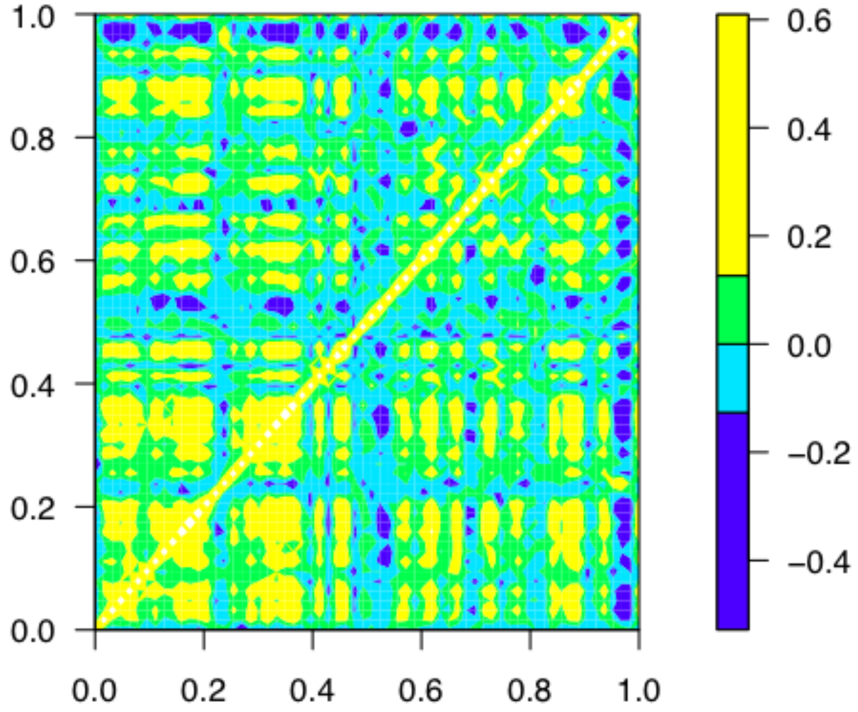
$$v_i(t) \equiv \sum_{s=t}^{t+\tau} \epsilon_i(s) p_i(s) V_i(s)$$

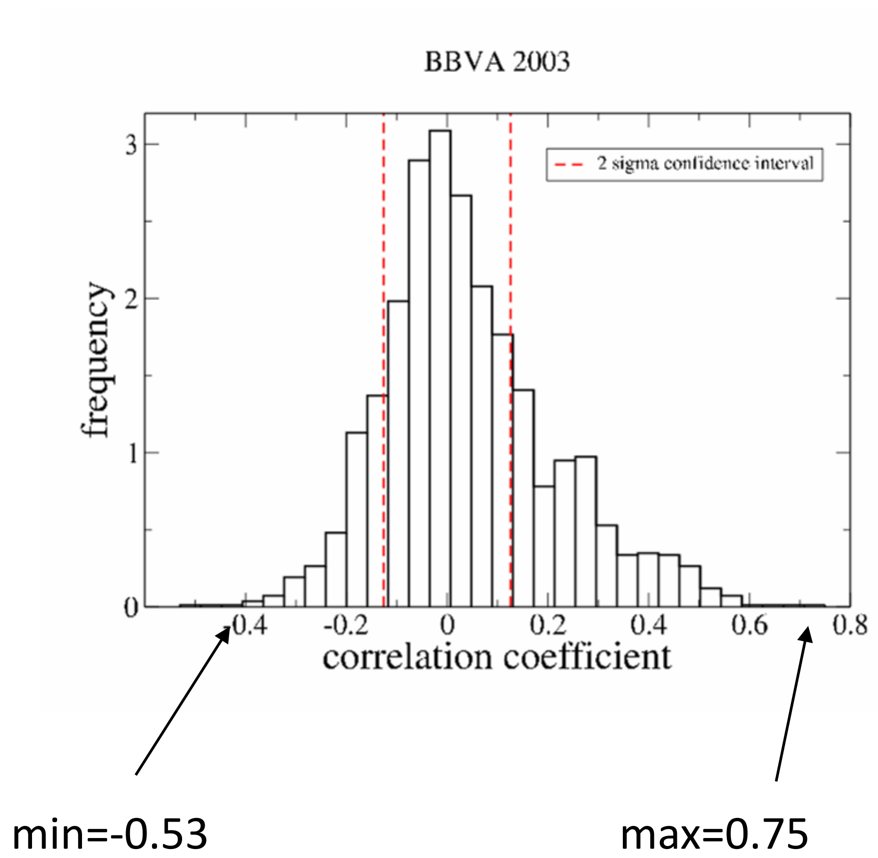
sign price volume

• Inventory variation is a measure of the net buy/sell position of agent i



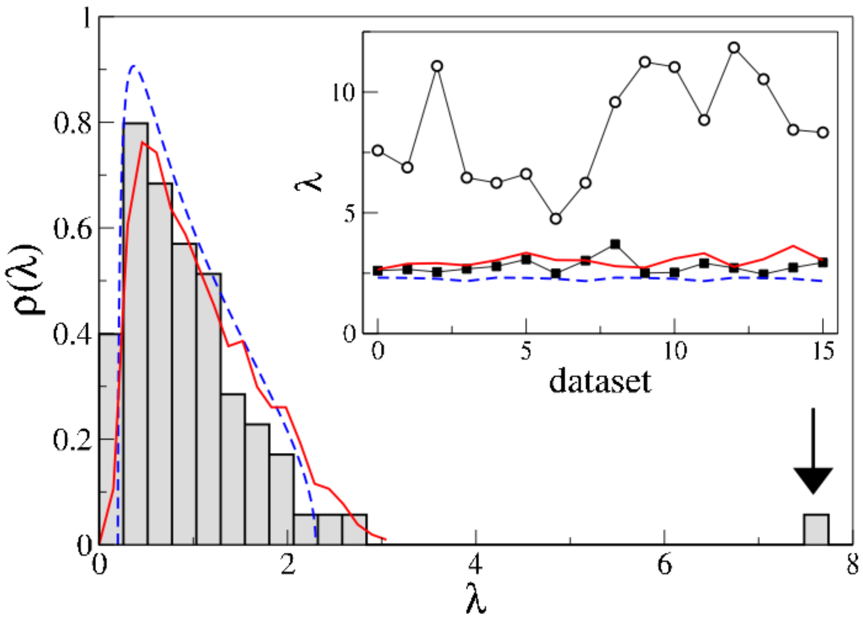
Cross correlation matrix of inventory variation

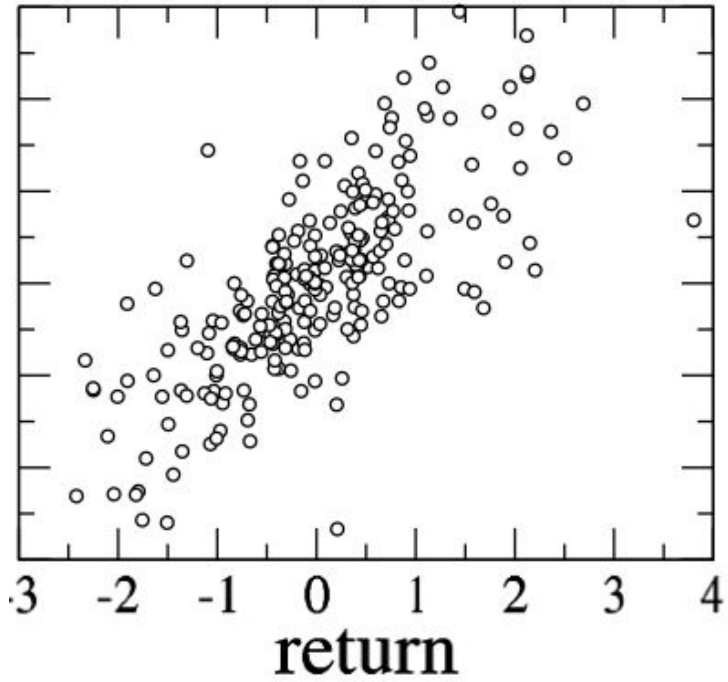




Trading activity is significantly cross correlated among firms

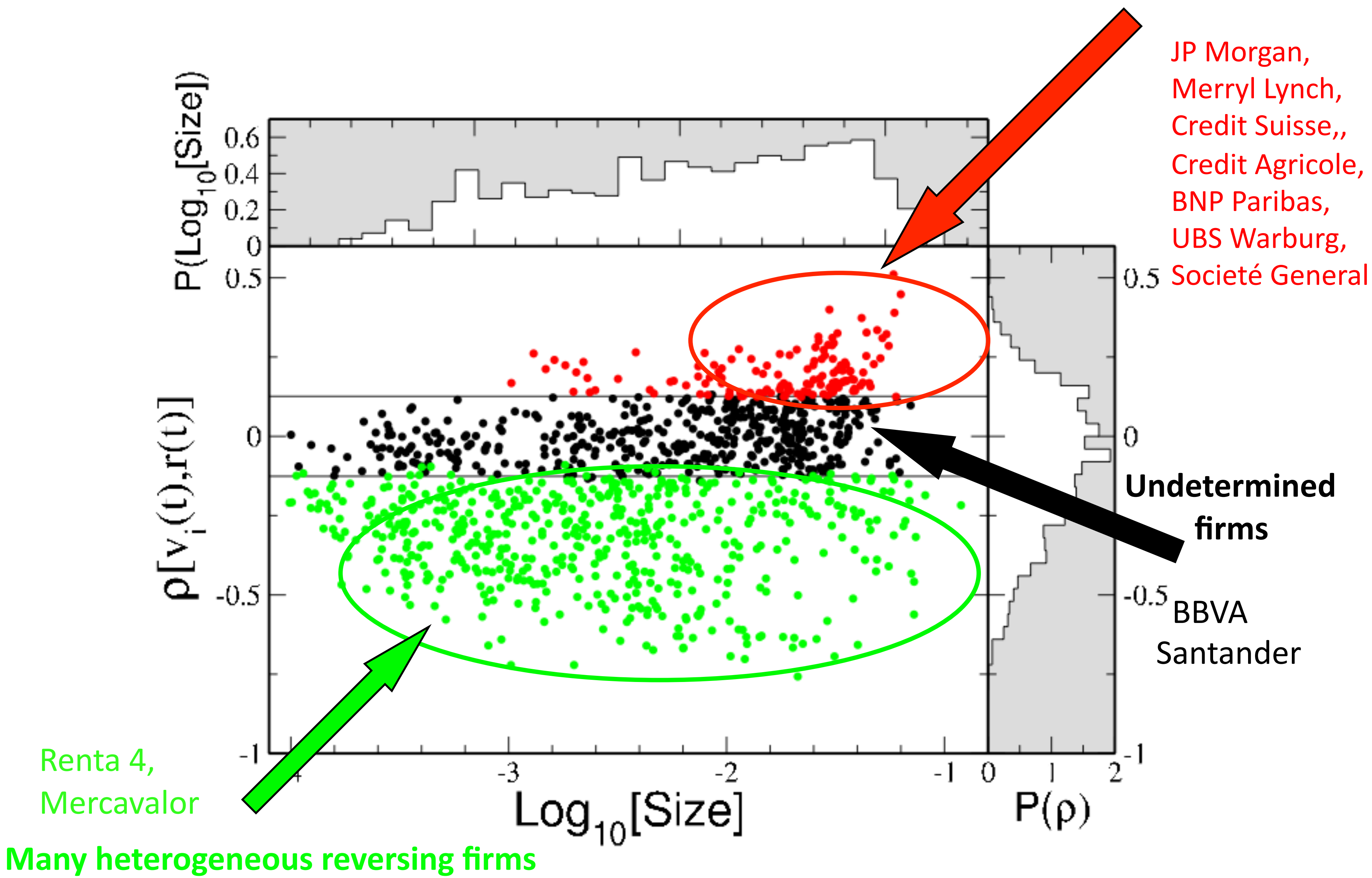
Origin of collective behavior





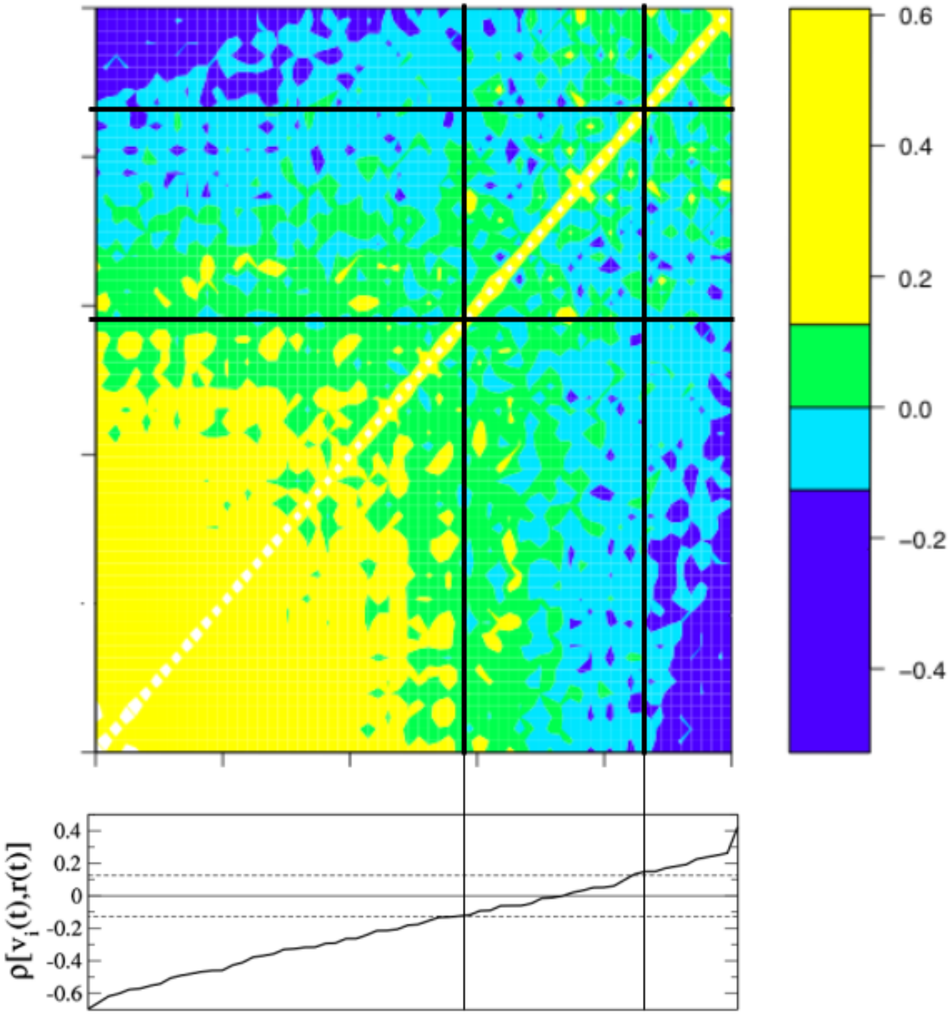
- The first eigenvalue is not compatible with random trading and is therefore carrying information about the collective dynamics of firms.
- The corresponding factor is significantly correlated with price return.
- There are groups of firms having systematically the same position (buy/sell) as the other members of the group they belong to.

The role of size



Size = average daily fraction of volume

Inventory variation correlation matrix obtained by sorting the firms in the rows and columns according to their correlation of inventory variation with price return



Correlated order flow

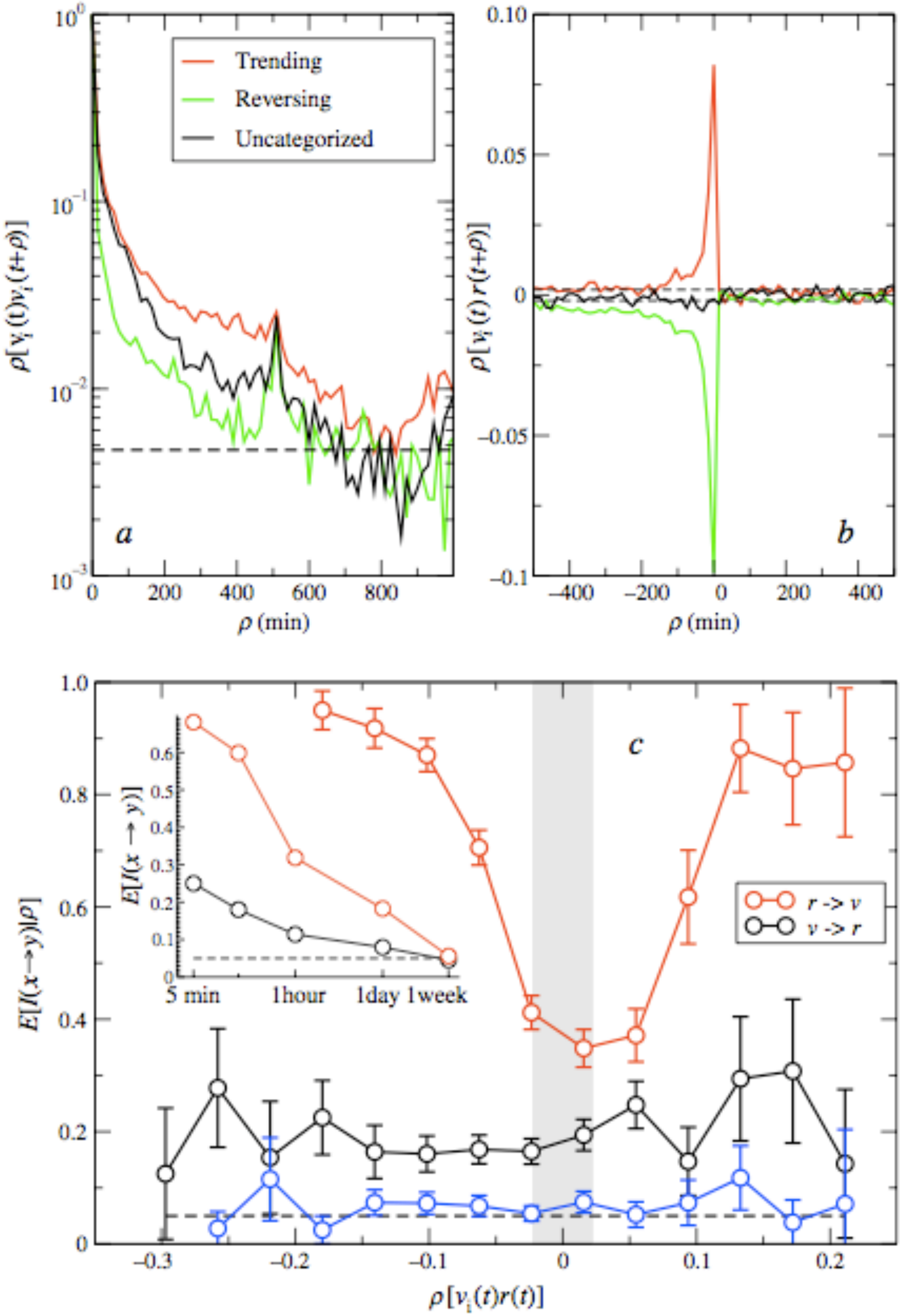
Inventory variation is long range correlated

$$\langle v_i(s+\tau)\, v_i(s)\rangle \approx \tau^{-\alpha}$$

Granger causality

$$I(X \to Y) = \begin{cases} 1 & \text{if X Granger} - \text{causes Y} \\ 0 & \text{if X does not Granger} - \text{causes Y} \end{cases}$$

- Returns cause inventory variations
- Inventory variations does not cause returns

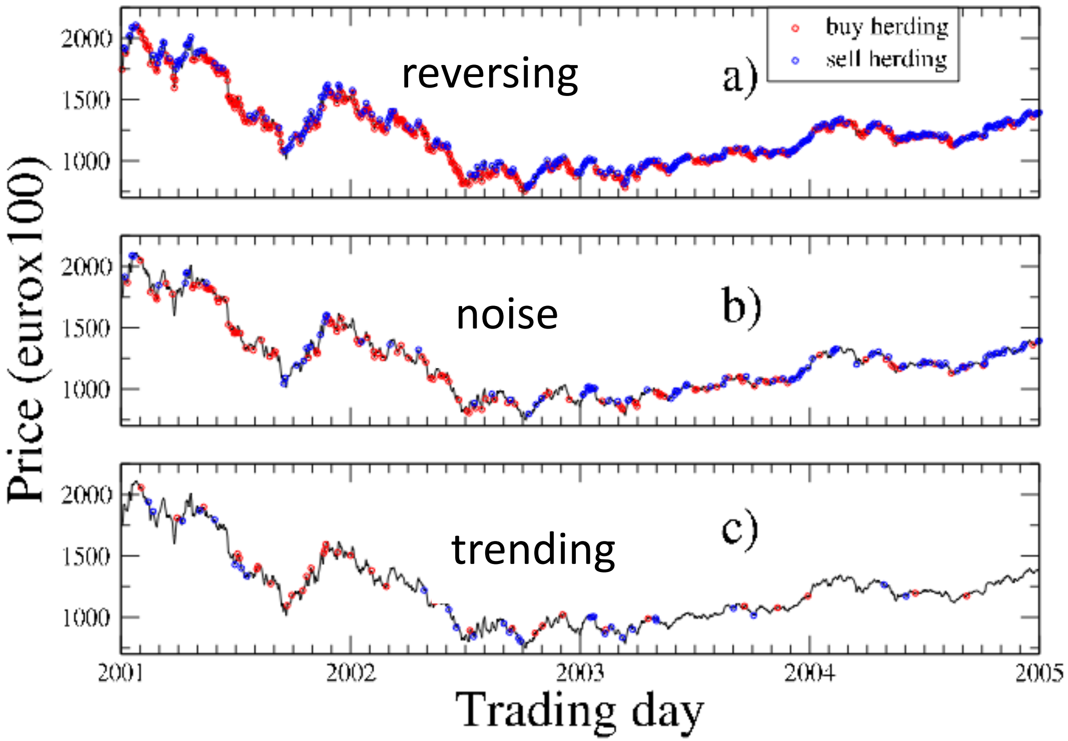


Herding

Herding indicator (see also Lakonishok et al, 1992)

$$h = \frac{\#\ of\ buy\ firms}{\#\ of\ buy\ firms\ +\ \#\ of\ sell\ firms}$$

We infer that herding is present in a given group when the probability of the observed number of buying or selling firms is smaller than 5% under the binomial null hypothesis.



	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

PhD opportunity

The **CALL FOR APPLICATIONS FOR ADMISSION TO PhD PROGRAM in Mathematics for Finance at the Scuola Normale Superiore di Pisa is open**

(see http://www.sns.it/en/scuola/ammissione/corsodiperfezionamento/scienze/).

The deadline for application is Thursday March 31, 2011 (Note that there is another available deadline on Monday September 12, 2011).

Co-authors

- Rosario Mantegna, Gabriella Vaglica, Adam Ponzi, Michele Tumminello (Univ. of Palermo)
- Doyne Farmer, Austin Gerig, Szabolsc Mike, Laszlo Gillemot, Bence Toth (Santa Fe Institute, USA)
- Esteban Moro, Javier Vicente, Luis Moyano (Univ. Carlos III of Madrid)
- Jean-Philippe Bouchaud (CFM, Paris)

Bibliography

- J.P. Bouchaud, J. D. Farmer, and F. Lillo, "How markets slowly digest changes in supply and demand", in T. Hens and K.R. Schenk-Hoppè (editors), Handbook of Financial Markets: Dynamics and Evolution, Handbook of Finance, 57-160 (2009).
- E. Moro, J. Vicente, L.G. Moyano, A. Gerig, J.D. Farmer, G. Vaglica, F. Lillo, R. N. Mantegna, Market impact and trading profile of hidden orders in stock markets. Physical Review E 80, 066102 (2009).
- A. Ponzi, F. Lillo, R. N. Mantegna, Market reaction to a bid-ask spread change: A power-law relaxation dynamics. Physical Review E 80, 016112 (2009).
- G. Vaglica, F. Lillo, E. Moro, and R.N. Mantegna, Scaling laws of strategic behavior and size heterogeneity in agent dynamics. Physical Review E, 77 036110 (2008).
- F. Lillo, E. Moro, G. Vaglica, and R.N. Mantegna, Specialization and herding behavior of trading firms in a financial market. New Journal of Physics, 10, 043019 (2008).
- L. Gillemot, J. D. Farmer, F. Lillo, There's more to volatility than volume. Quant. Fin. 6, 371-384 (2006).
- J. D. Farmer, A. Gerig, F. Lillo, S. Mike, Market efficiency and the long-memory of supply and demand: Is price impact variable and permanent or fixed and temporary? Quantitative Finance 6, 107-112 (2006).
- F. Lillo, S. Mike, J. D. Farmer, Theory for long memory in supply and demand. Physical Review E 71, 066122 (2005).
- F. Lillo, J.D. Farmer, The key role of liquidity fluctuations in determining large price fluctuations. Fluctuations and Noise Letters 5, L209 (2005).
- F.Lillo and J.D. Farmer, The long memory of efficient market. Studies in Nonlinear Dynamics and Econometrics 8, 1 (2004).
- J.D. Farmer, L. Gillemot, F. Lillo, S. Mike and A. Sen, What really causes large price changes? Quantitative Finance 4 383–397 (2004).
- Fabrizio Lillo J. Doyne Farmer, Rosario N. Mantegna, Master curve for the price-impact function, Nature 421, 129-130 (2003).