Observatory of
Complex Systems

SCUOLA
NORMALE
SUPERIORE
PISA

SANTA FE
INSTITUTE
Celebrating 20 years of Complexity Science

How market microstructure affects liquidity and market risks

Fabrizio Lillo

Scuola Normale Superiore di Pisa and University of Palermo (Italy)
Santa Fe Institute (USA)

Perm (Russia), February 2, 2012

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)

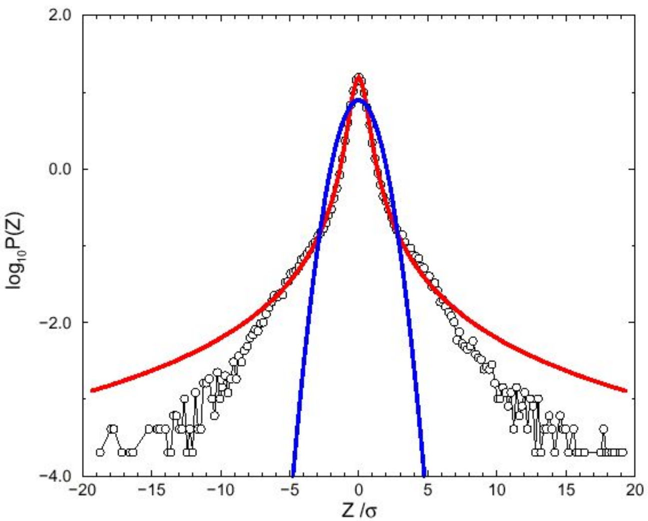
Why should market microstructure be relevant for risk and stress testing?

- Connection between microscopic and macroscopic time scales
- Temporary liquidity crises
- Price impact
- Liquidity risk
- Fire sales

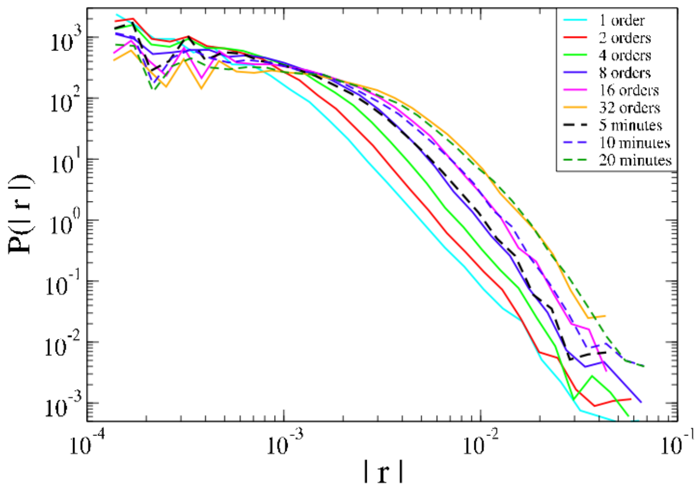
- Institutional design makes the system more robust: e.g. a clearinghouse for CDS (Duffie and Zhu, Cont)

Micro-macro connection

- Financial markets are intrinsically unstable and display large price fluctuations (Mandelbrot, Fama, Mantegna and Stanley, etc)
- The origin of these short time scales large price fluctuations is weakly related to news (see, e.g. Bouchaud et al 2009)
- There is an intriguing evidence that individual trade price returns have the same properties as returns on longer time scales
- Is microstructure important to explain and model stylized facts? (fat tails, clustered volatility, multifractality, etc)

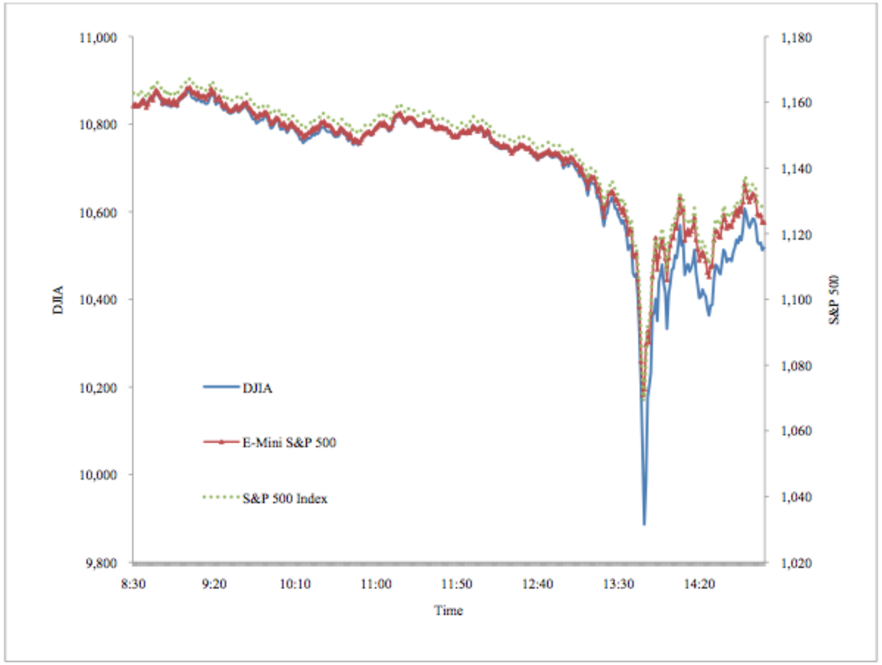


Return distributions



An extreme example: the flash crash

- May 6, 2010
- Initiated at E-mini S&P 500 futures
- Price drop of 1% per minute
- Contagion to other assets: ETF, Indices, and then stocks: **the 20 millisecond cascade**
- Over 20,000 trades across more than 300 securities were executed at prices more than 60% away from their values just moments before. Many at a penny or less, or as high as $100,000, before prices of those securities returned to their "pre-crash" levels.
- By the end of the day, major futures and equities indices "recovered" to close at losses of about 3% from the prior day.



This figure presents end-of-minute transaction prices of the Dow Jones Industrial Average (DJIA), S&P 500 Index, and the June 2010 E-Mini S&P 500 futures contract on May 6, 2010 between 8:30 and 15:15 CT.

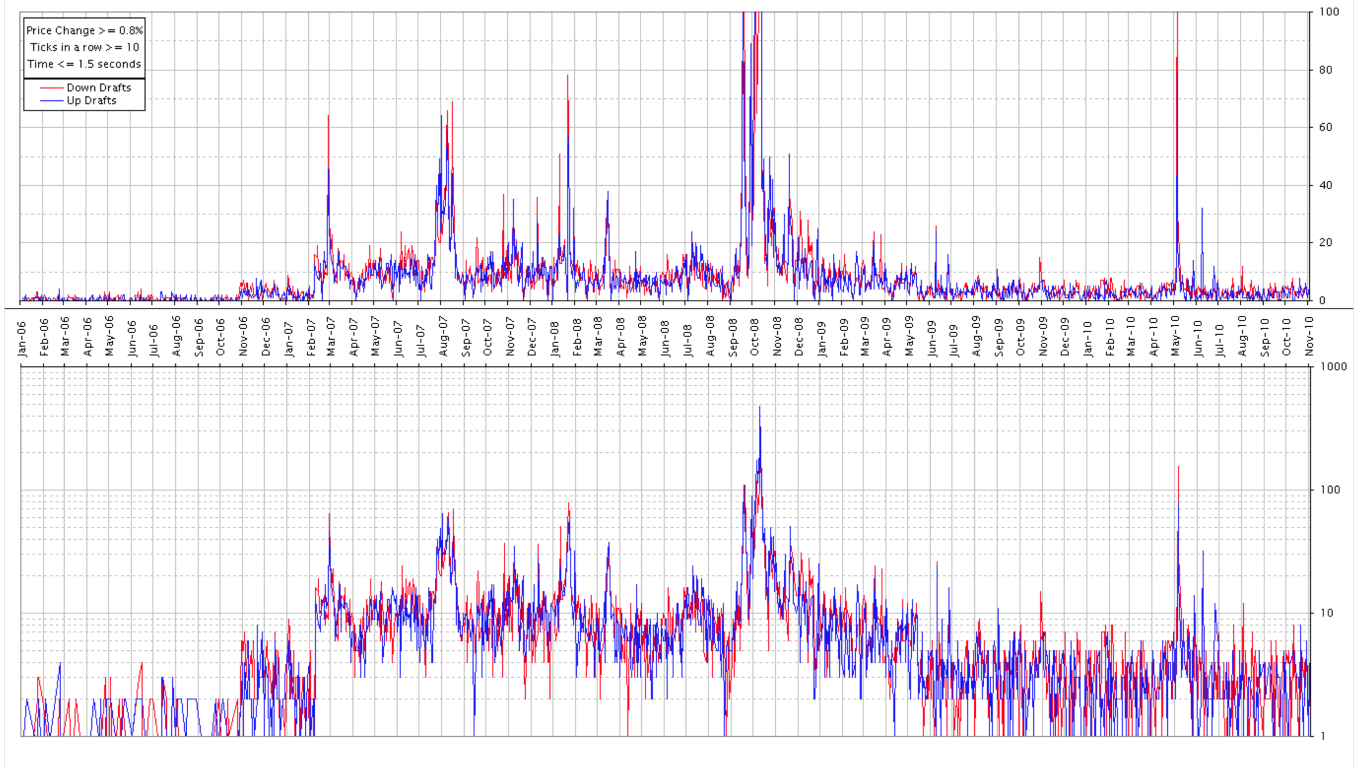
An isolated event? (from Nanex)

- All listed equities for 2006-2011 searching for potential "mini crashes" in individual stocks.
  - **To qualify as a down(up)-draft candidate, the stock had to tick down (up) at least 10 times before ticking up (down)-- all within 1.5 seconds and the price change had to exceed 0.8%.**

	Down Drafts			Up Drafts		
Year	**Count**	**Download All**	**Examples**	**Count**	**Download All**	**Examples**
2011	69+	Download		70+	Download	
2010	1041	Download	View	777	Download	View
2009	1,462	Download	View	1,253	Download	View
2008	4,065	Download	View	4,354	Download	View
2007	2,576	Download	View	2,456	Download	View
2006	254	Download	View	208	Download	View

Systemic instability FROM INSTITUTIONAL DESIGN?

UpDnDrafts.png 1369×766 pixels 5/9/11 9:53 AM



http://www.nanex.net/FlashCrashEquities/UpDnDrafts.png Page 1 of 1

Regulation NMS was implemented in 2007

Consider the NYSE Hybrid Market rollout:

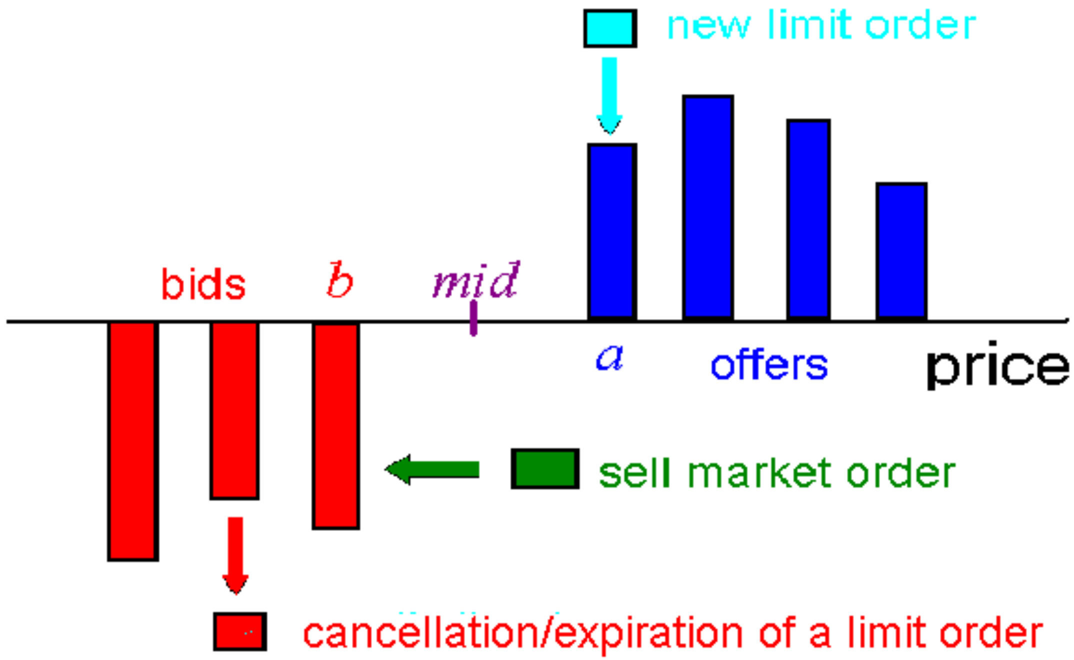
- Hybrid Phase III - COMPLETED rollout January 24, 2007
- Hybrid Phase IV - COMPLETED rollout February 27, 2007

Note that prior to Feb 2007, the NYSE had never been a reporting exchange in any incident.

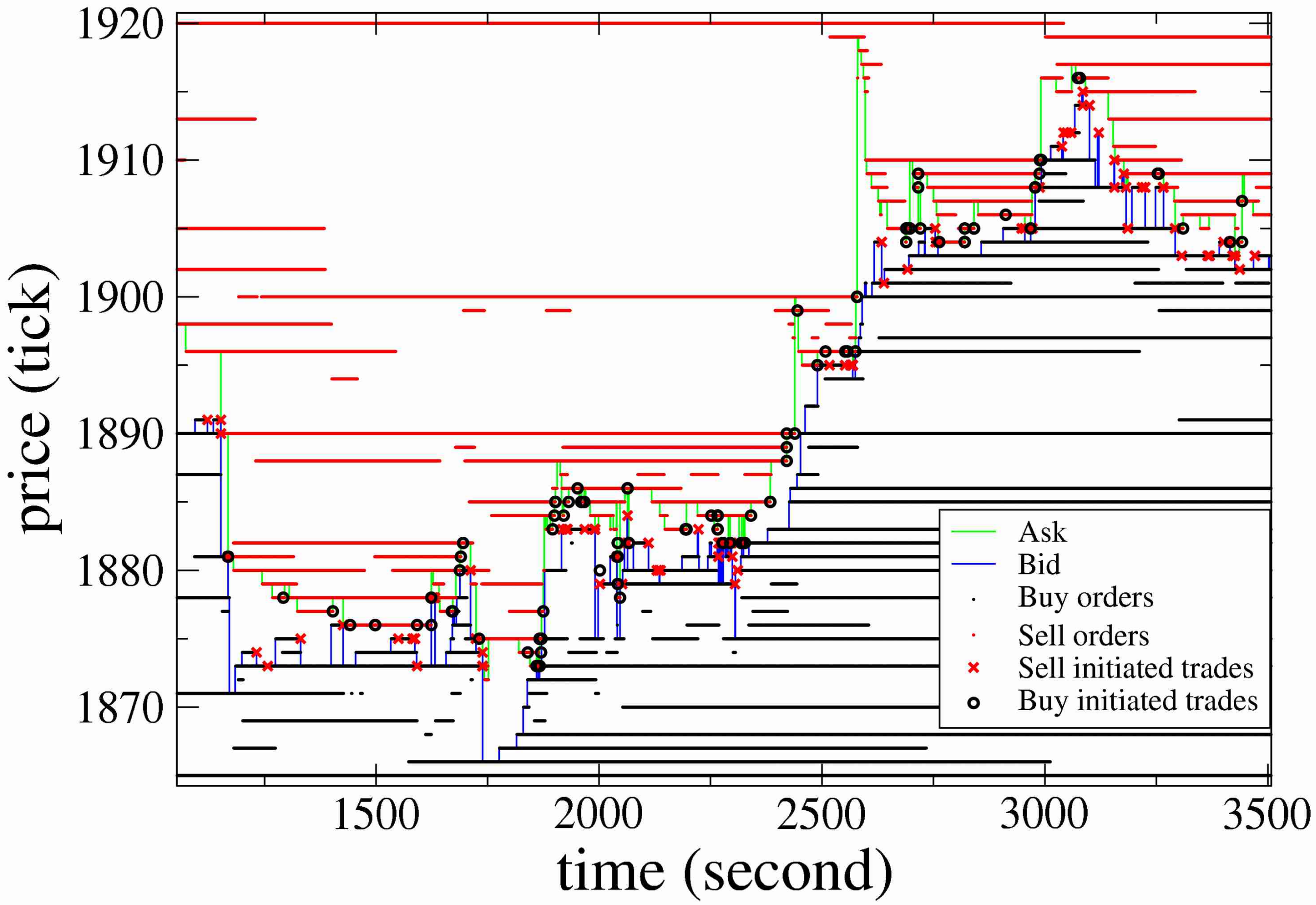
From Nanex

Limit order book

- Many stock exchanges (NYSE, LSE, Paris) works through a double auction mechanism
- Order book data are fundamental to investigate the price formation mechanisms



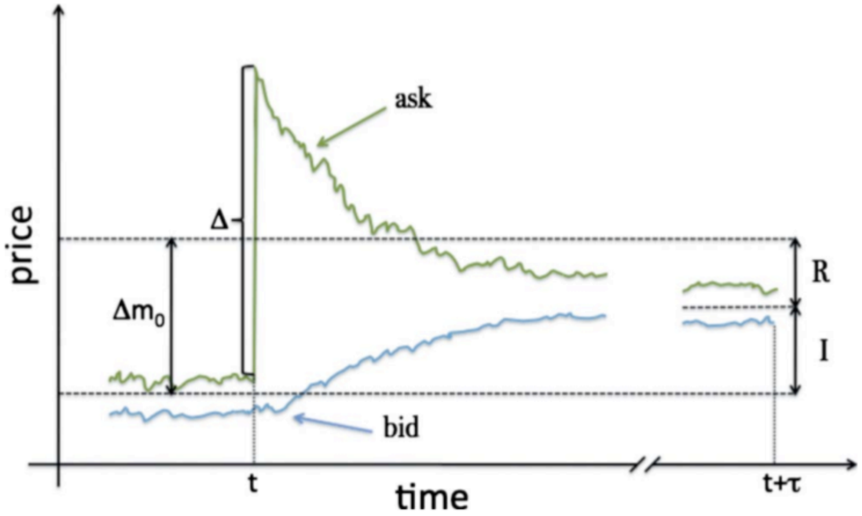
Representation of limit order book dynamics



(Ponzi, Lillo, Mantegna 2007)

MARKET REACTION TO TEMPORARY LIQUIDITY CRISES

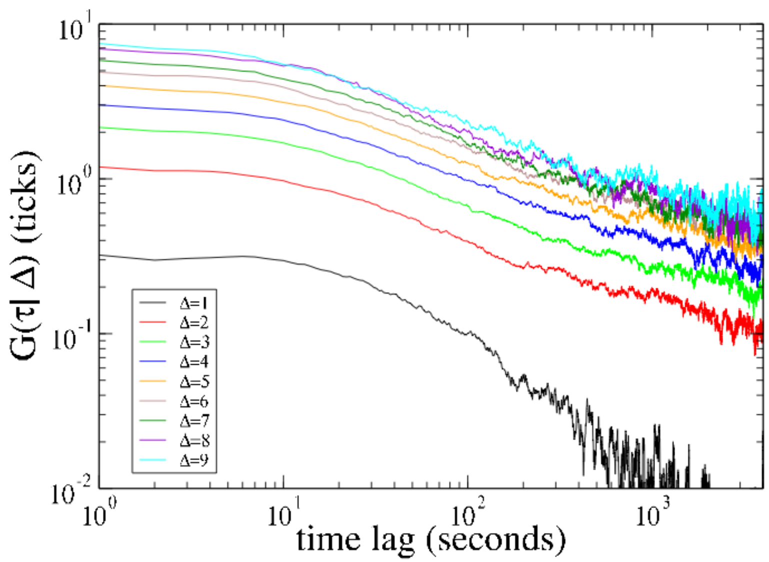
The market reaction to large spread changes (LSE stocks).



Q: How the market relaxes to the "normal" state after a liquidity crisis?

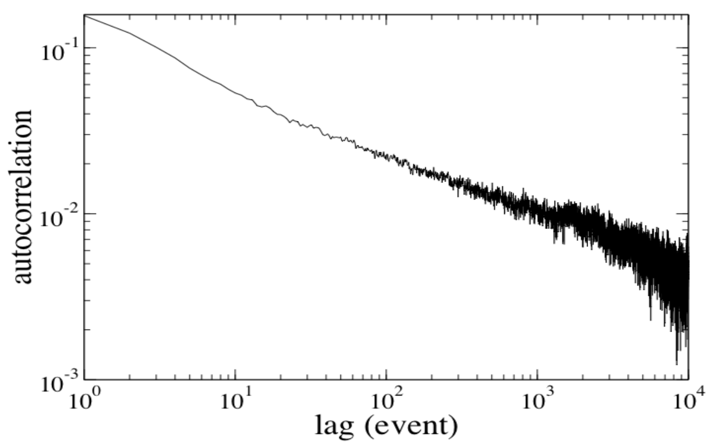
A: The spread (but also the limit order book) decays on average to the "normal" value by following a very slow dynamics

Excess spread

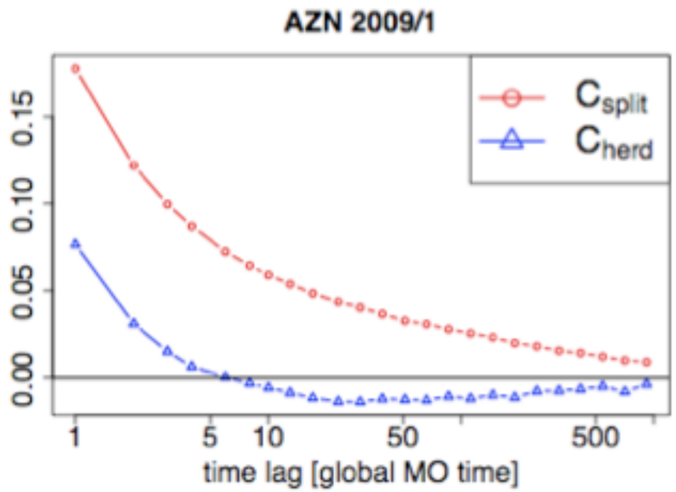


Persistence

- Also order flow (net demand imbalance) is persistent and correlated in time as a result of order splitting and herding



Autocorrelation of signs (buy vs sell) of market orders



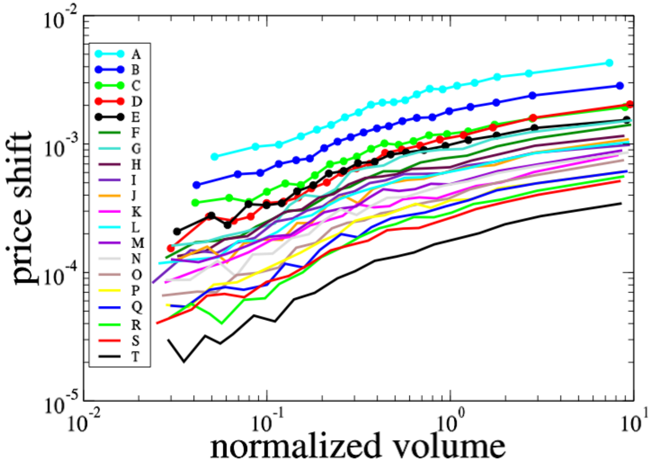
Decomposition of the autocorrelation in a splitting and a herding component (at the broker level)

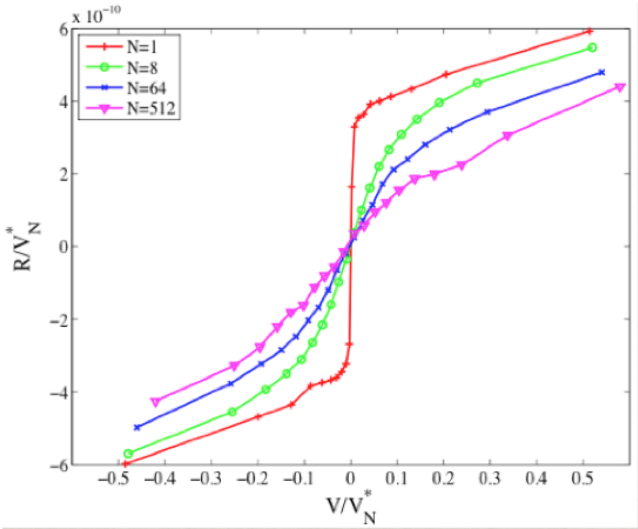
Liquidity RISK

- The practice of marking to market the value of a portfolio might be misleading if either the assets are illiquid or the position must be unloaded quickly
- It has been suggested to use a mark to liquidity approach (Acerbi and Scandolo) to value a portfolio. Up to now an interesting theoretical exercise
- This requires a **market impact model** in
  - Normal situations
  - Distressed market state

MARKET IMPACT

- Market impact is the price reaction to trades
- There are different types of market impact





- Minimizing impact of the execution of a large trade means minimizing cost
- A satisfactory theory of market impact of large trade is still lacking, but it is key for assessing liquidity risk

Why market impact?

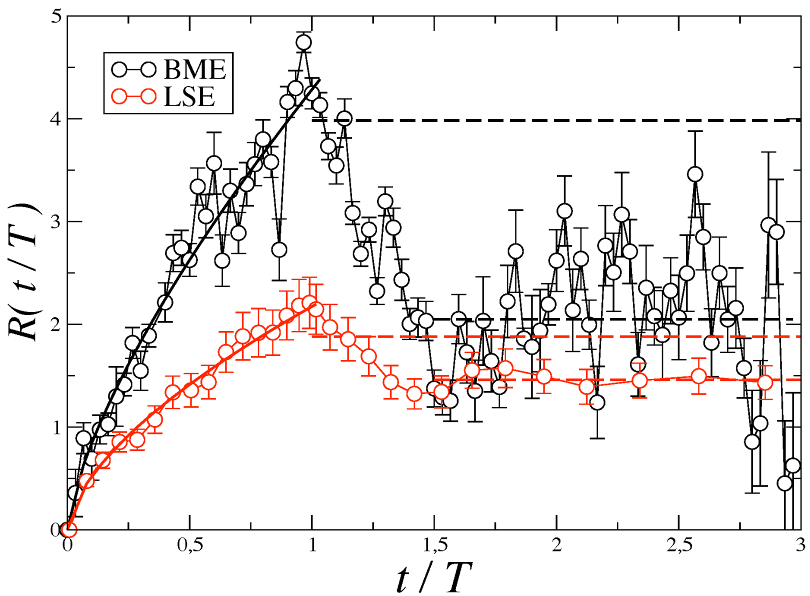
- The impact of trades reveals some private information (Orders do not impact prices. It is more accurate to say that orders forecast prices" (Hasbrouck 2007))
- Impact is a statistical effect due to order flow fluctuations (zero-intelligence models, self-fulfilling prophecy)

Statistically reconstructed price impact temporal profile of large orders by all brokers at LSE and BME (using brokerage data)

Empirical evidence of

- Square root dependence of total impact from order size
- Reversion of price at 2/3 of the peak

We recently developed a model for impact reproducing these facts (Farmer et al 2011) See also Bouchaud et al 2011

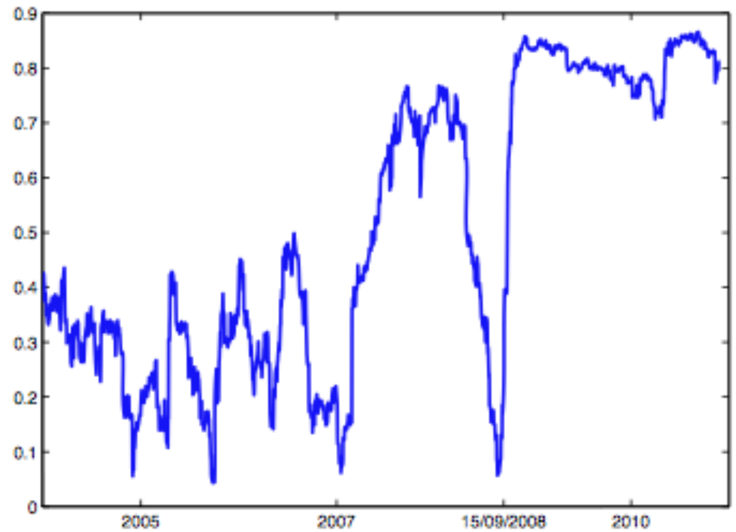


Fire sales or distressed selling

- In an extremely distressed situation a company can be forced to sell large volumes as soon as possible
- Market conditions and liquidity dramatically change and “normal” market impact is not anymore appropriate
- Other subtle effects, such as a dramatic change in correlations during fire sale events (see Cont et al 2011). Uncorrelated assets can become strongly correlated (LTCM, August 2007, etc)
- Understanding market impact in

distressed markets is critical to

assess liquidity risk



EMWA correlation between two ETF of the the S&P 500: SPDR XLE (energy) and SPDR XLK (technology)

A tangled web

- "A transaction in the market affects more than the parties involved in the transaction itself, since the price determined in the transaction is used to price other assets and obligations" (Shin, 2008)
- At a systemic level one must take into account the similarity of portfolios across banks in assessing the effective role of diversification (e.g. hedge funds in August 2007)

Outline

- Why does price move? The Kyle mode
- Market impact: phenomenology and modeling
- The persistence of order flow
  - Market efficiency
  - Optimal execution
  - Origin of correlations
- Toward an ecology of financial markets

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

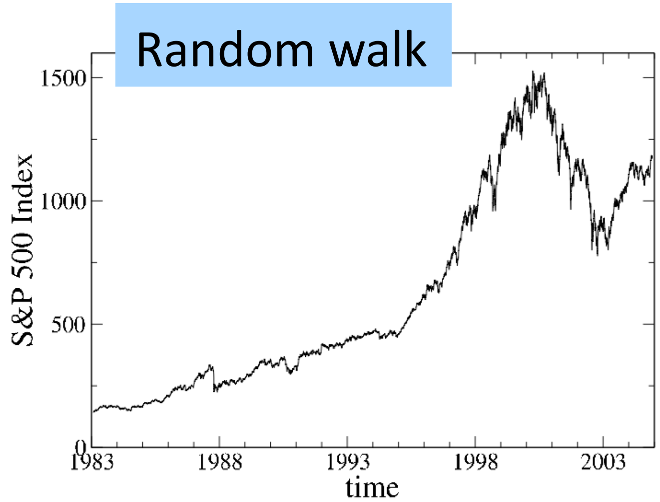
$$E[\pi] = \frac{(v - p_0)^2}{2}\sqrt{\frac{\sigma_u^2}{\Sigma_0}}$$

- 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



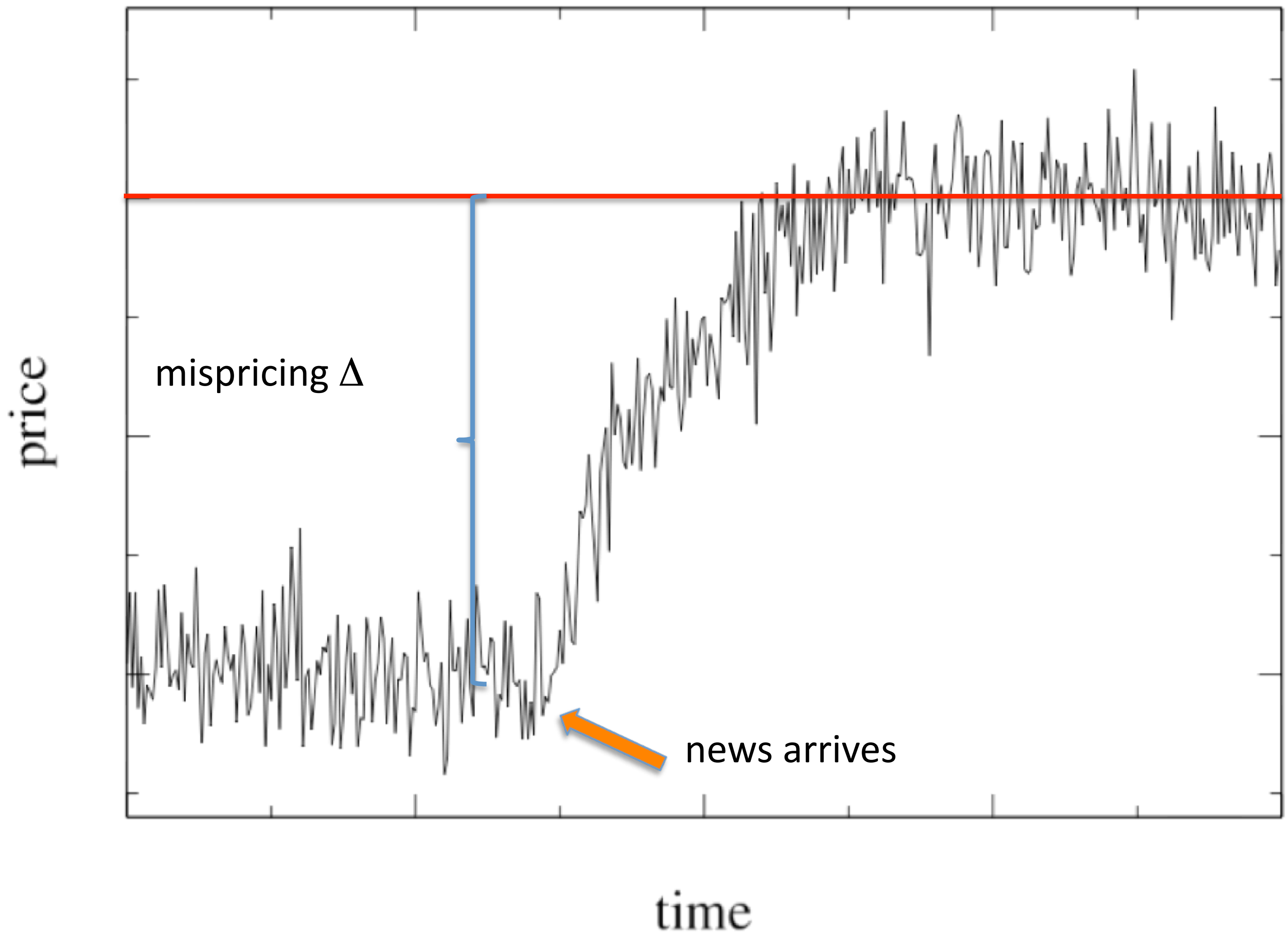
- Price dynamics is often modeled in terms of a random walk
- This process is mechanically determined by the interplay between order flow and price response
- Specifically, from a statistical point of view, price dynamics is determined by three components
  - The market structure
  - The (unconditional) 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

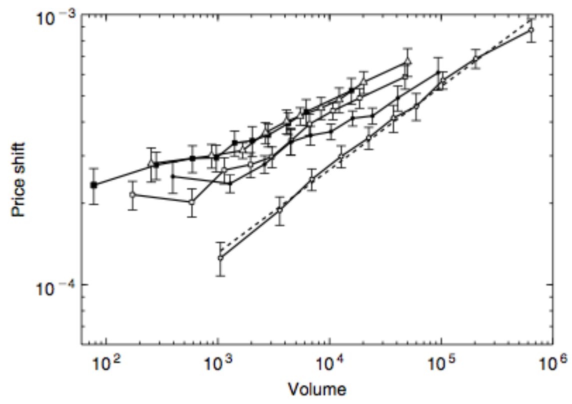


FIGURE 2.4 Market impact function of buy market orders for a set of five highly capitalized stocks traded in the LSE, specifically AZN (*filled squares*), DGE (*empty squares*), LLOY (*triangles*), SHEL (*filled circles*), and VOD (*empty circles*). Trades of different sizes are binned together, and 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 $\psi = 0.3$.

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

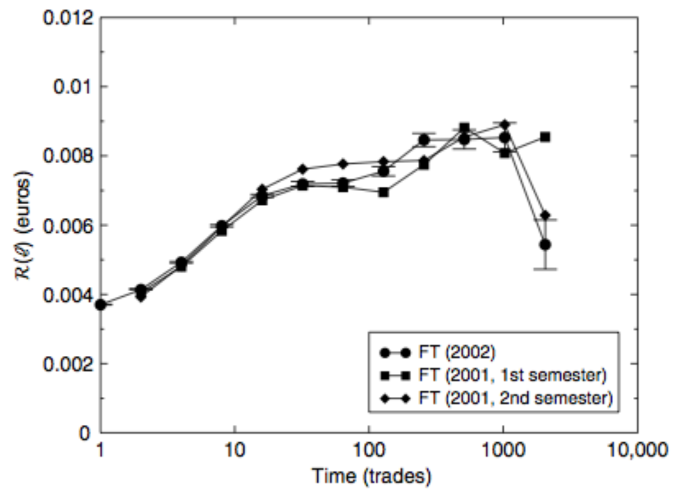


FIGURE 2.8 Average empirical response function $\mathcal{R}_\ell$ for FT during three different periods (1st and 2nd semester of 2001 and 2002); error bars are shown for the 2002 data. For the 2001 data, the y axis has been rescaled such that $\mathcal{R}_1$ coincides with the 2002 result. $\mathcal{R}_\ell$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

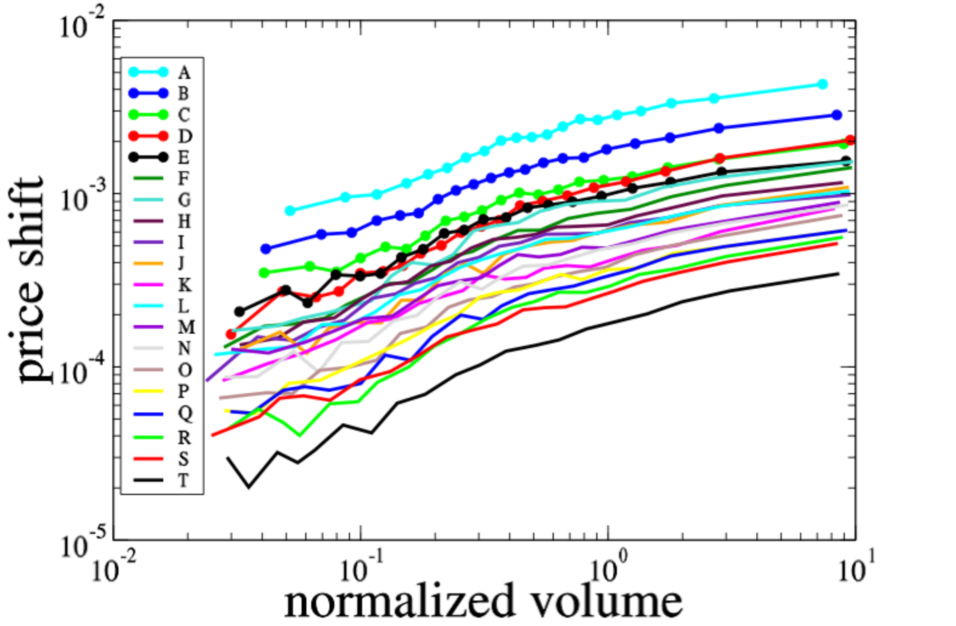
$$R(\ell) = E[(p_{t+\ell} - p_t)\varepsilon_t]$$

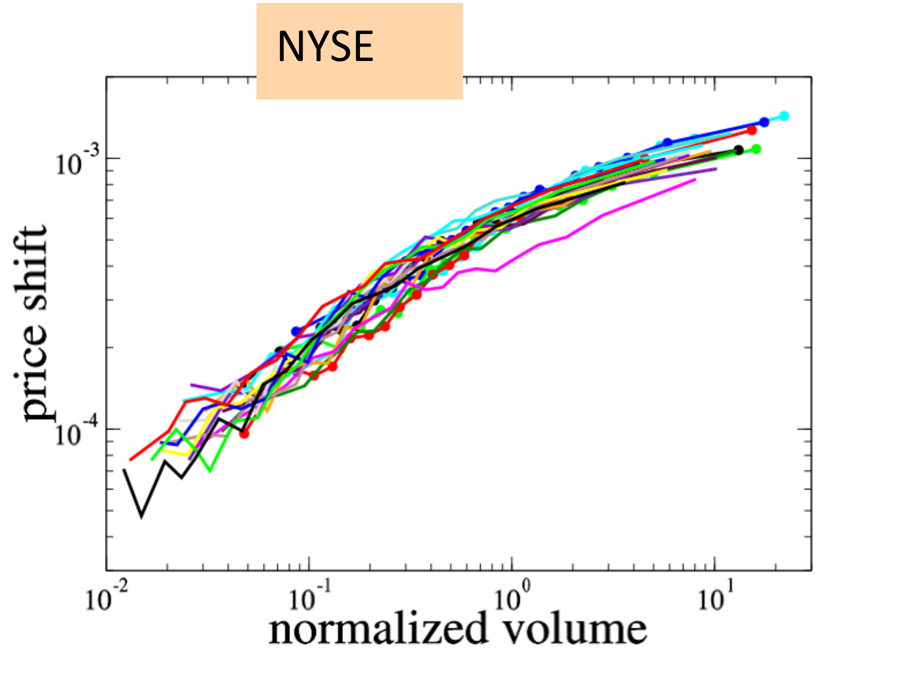
Master curve for individual impact

(Lillo et al. Nature 2003)

GROUP A -> least capitalized group

GROUP T -> most capitalized group

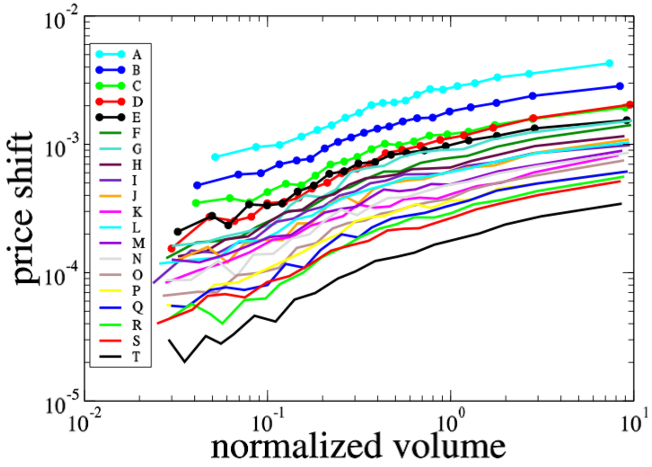




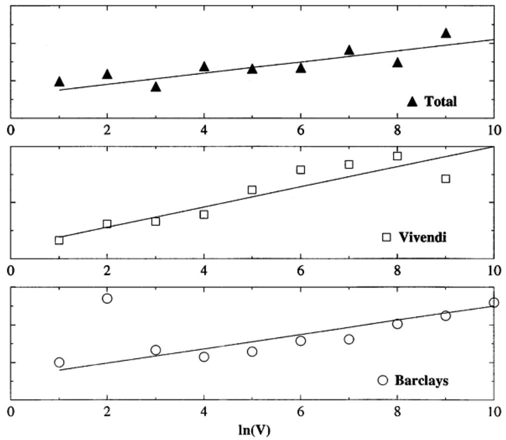
$$r = C^{-\gamma} f(V C^{\delta})$$

Impact of individual transactions

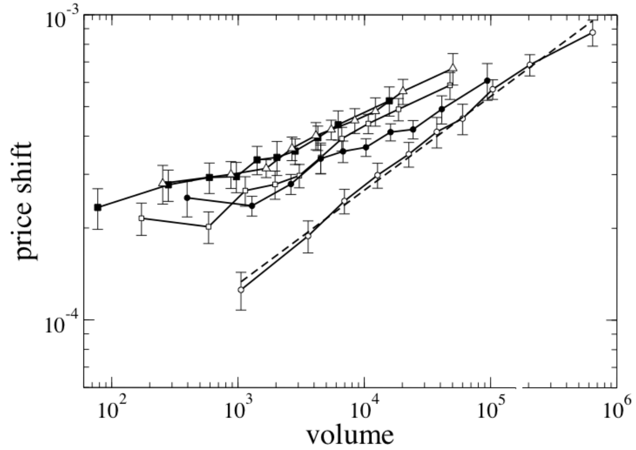
NYSE



Paris Bourse
(Potters et al. 2003)



London Stock Exchange



- Individual transaction market impact is a concave function of the volume

$$f(v) = sign(v)\,|v|^{\beta} \qquad\qquad f(v) = sign(v)\log|v|$$

- Impact of individual transaction is NOT universal
- What is the origin of the functional form of this type of impact?

Price impact from book shape

Let $\Omega(r) = \int_0^r v(x)\, dx$ indicate the cumulative number of shares (depth) up to price return r

A market order of size V will produce a return

$$V = \Omega(r) \quad \longrightarrow \quad r = \Omega^{-1}(V)$$

For example if

$$\Omega(r) = k\, r^{\eta}$$

then the price impact is

$$r \propto V^{1/\eta}$$

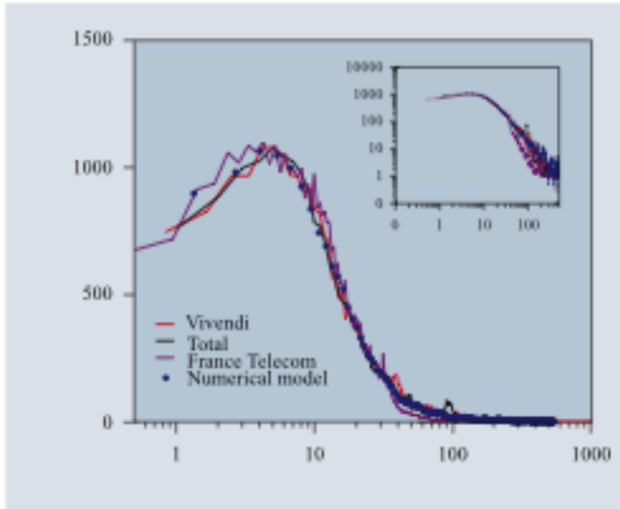


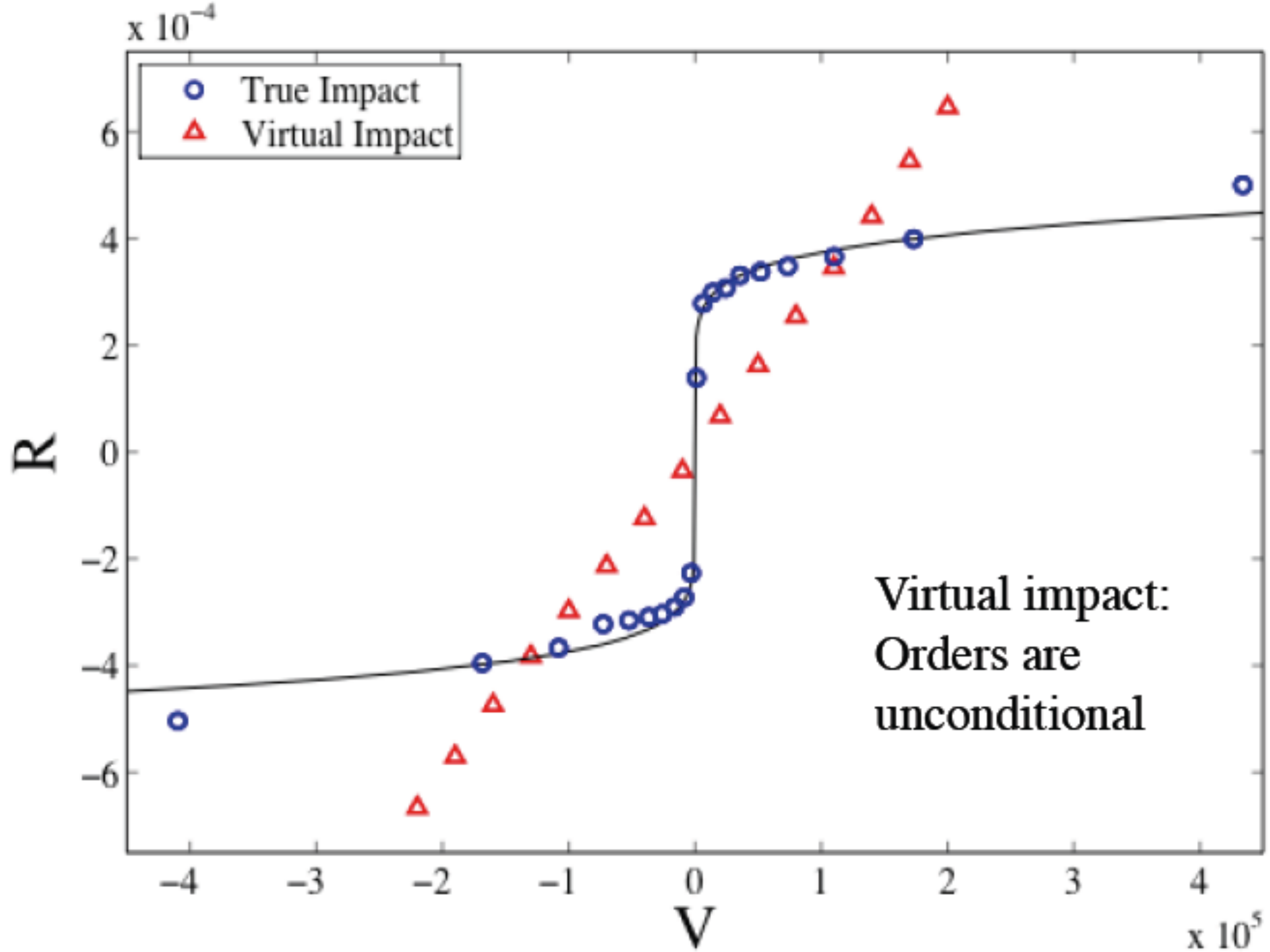
Figure 2. Average volume of the queue in the order book for the three stocks, as a function of the distance Δ from the current bid (or ask) in a log–linear scale. Both axes have been rescaled in order to collapse the curves corresponding to the three stocks. The thick dots correspond to the numerical model explained below, with $\Gamma = 10^{-3}$ and $p_m = 0.25$. Inset: same data in log–log coordinates.

Real and virtual impact

- Is this explanation in terms of the relation between price impact and the limit order book shape correct ?
- The basic assumptions are:
  - the traders placing market orders trade their desired volume irrespectively from what is present on the limit order book
  - the limit order book is filled in a continuous way, i.e. all the price levels are filled with limit orders

- We test the first assumption by measuring the virtual price impact, i.e. the price shift that would have been observed in a given instant of time if a market order of size V arrived in the market
- We test the second assumption by considering the fluctuations of market impact

Market impact f(V) for single trade



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

Fluctuations of the impact

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

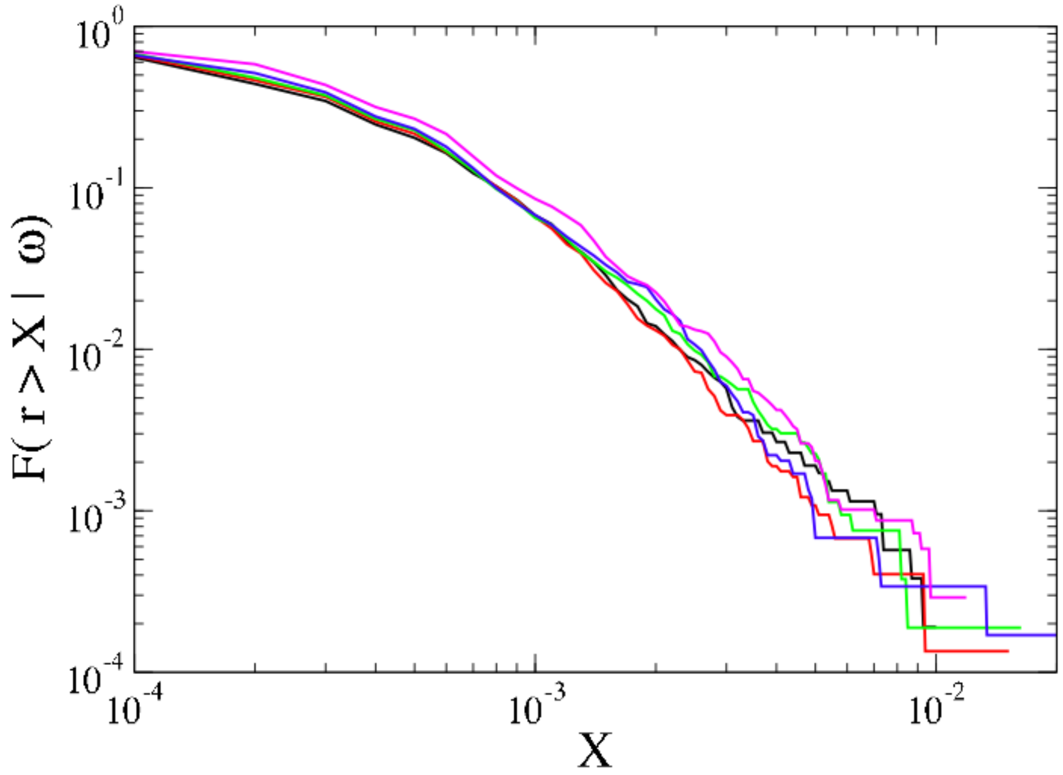
$$p(r|V) = (1 - g(V))\delta(r) + g(V)f(r|V)$$

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

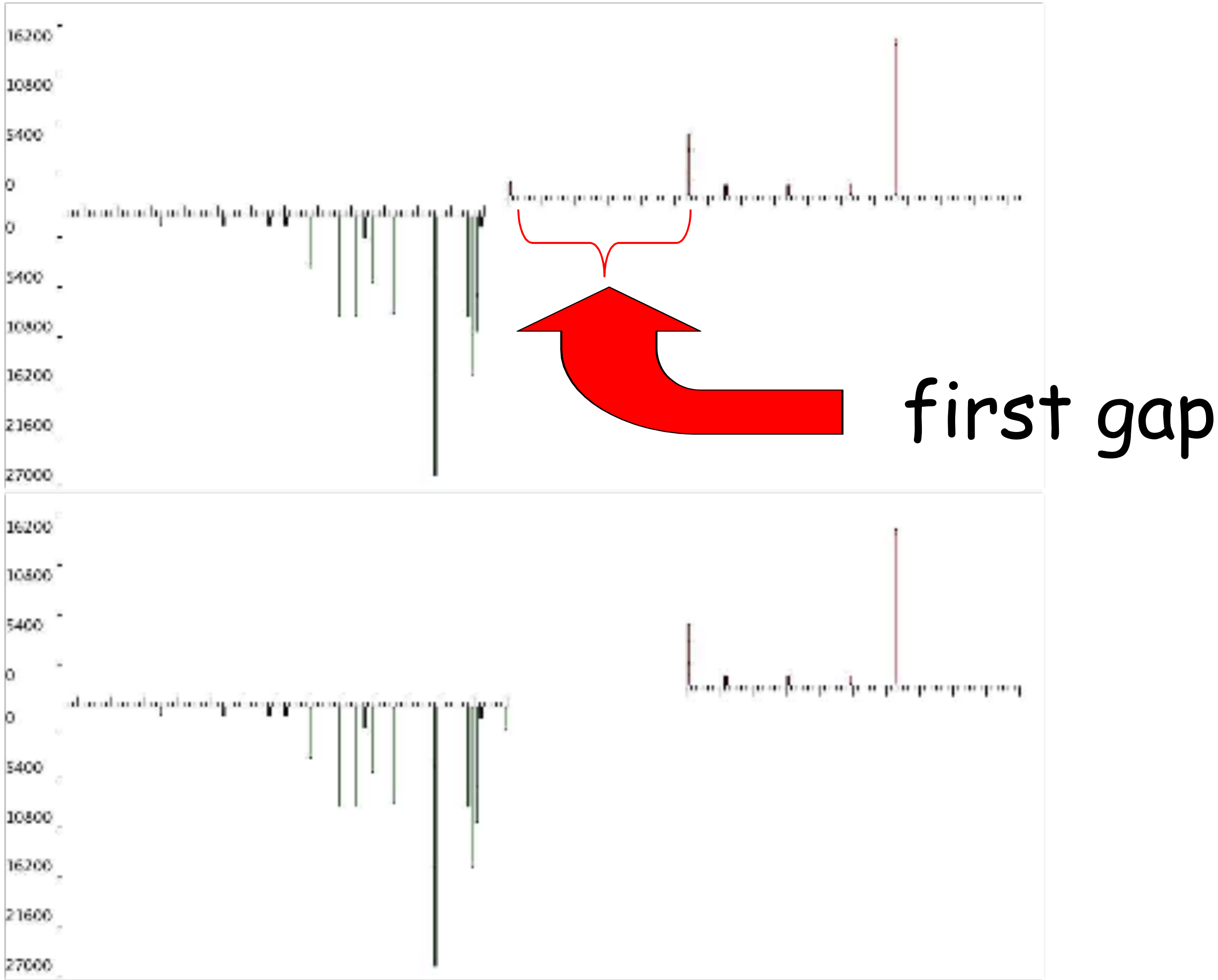
• **The fluctuations in market impact are important**

- 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

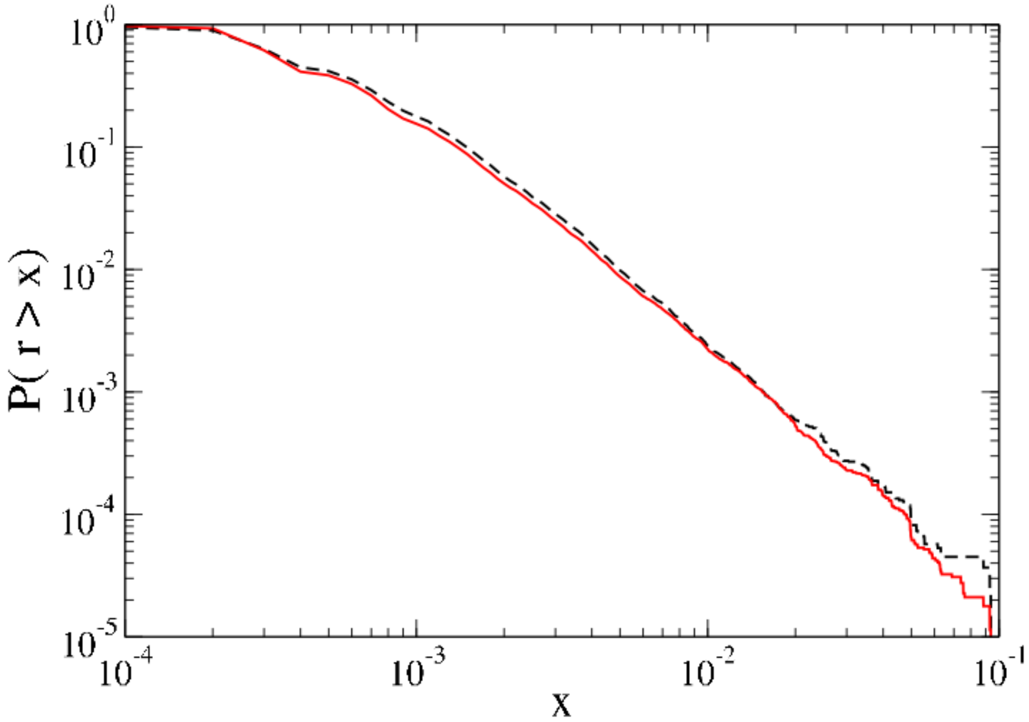
- How can small volume transactions create large price changes ?

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

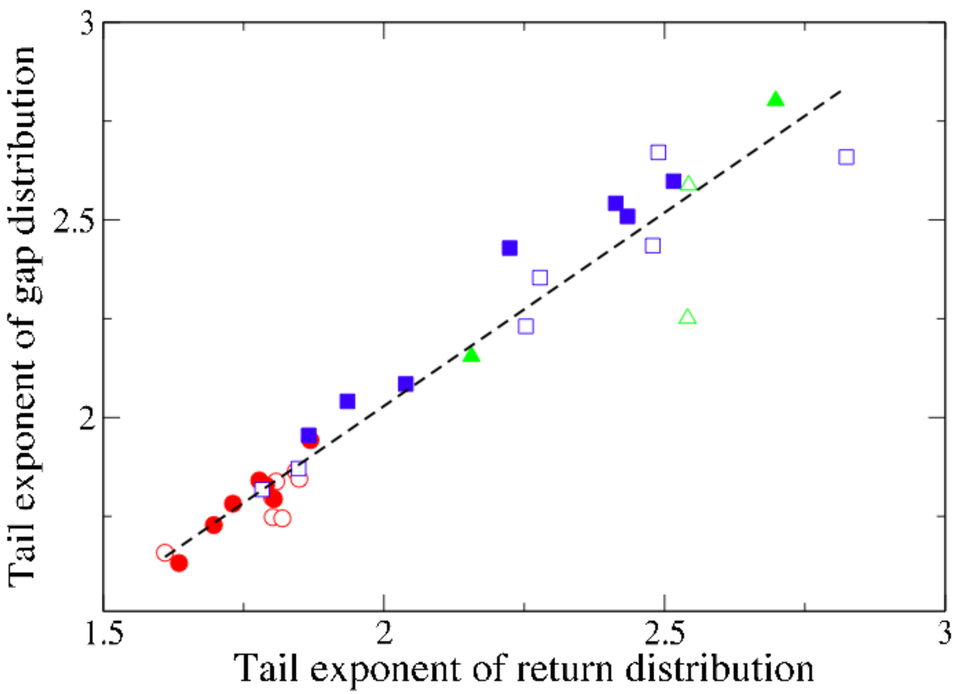
Origin of large price returns



- 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 gap size

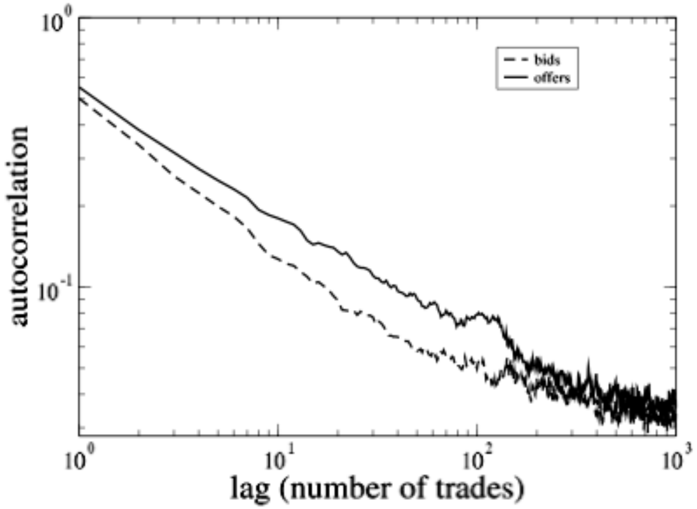


Fig. 4. Autocorrelation function of the first gap size for bids ($g_{-1}(t)$) and offers ($g_1(t)$) in a log-log plot. The data refer to Astrazeneca.

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

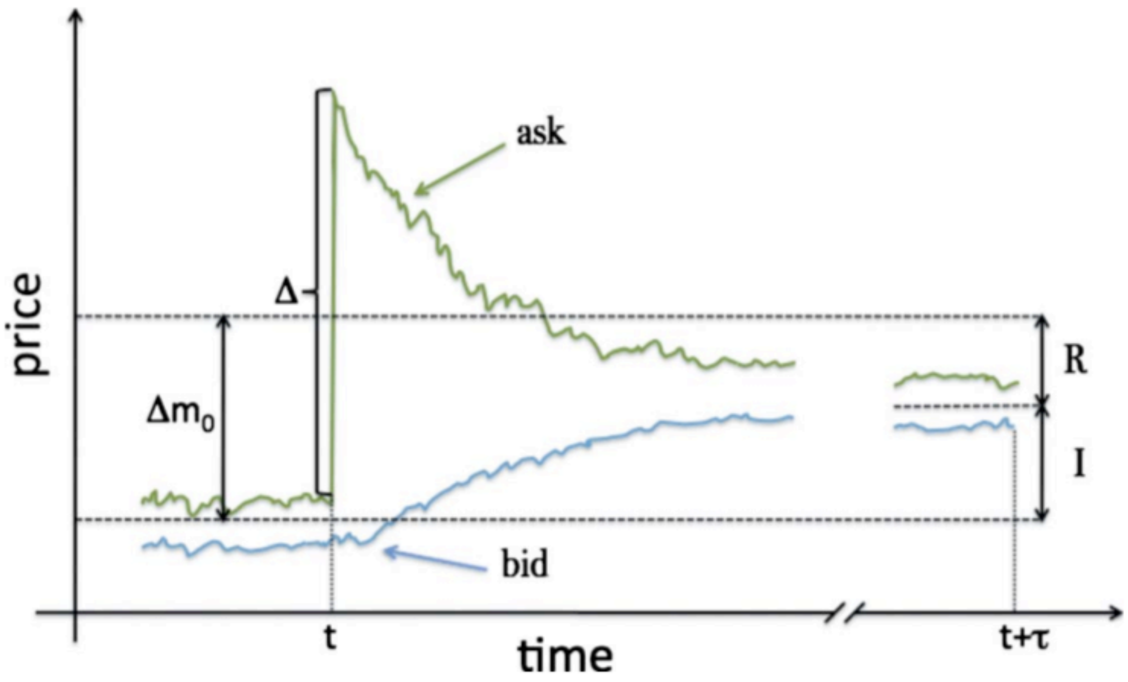
Market reaction to temporary liquidity crises

• 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

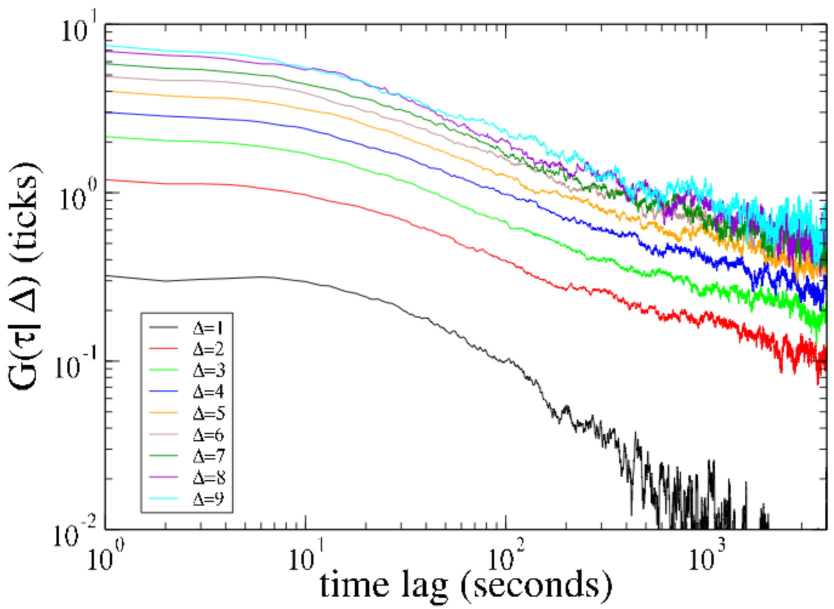
• Liquidity provider can profit of a large spread by placing limit orders and obtaining the best position. However the optimal order placement inside the sprea a nontrivial problem.

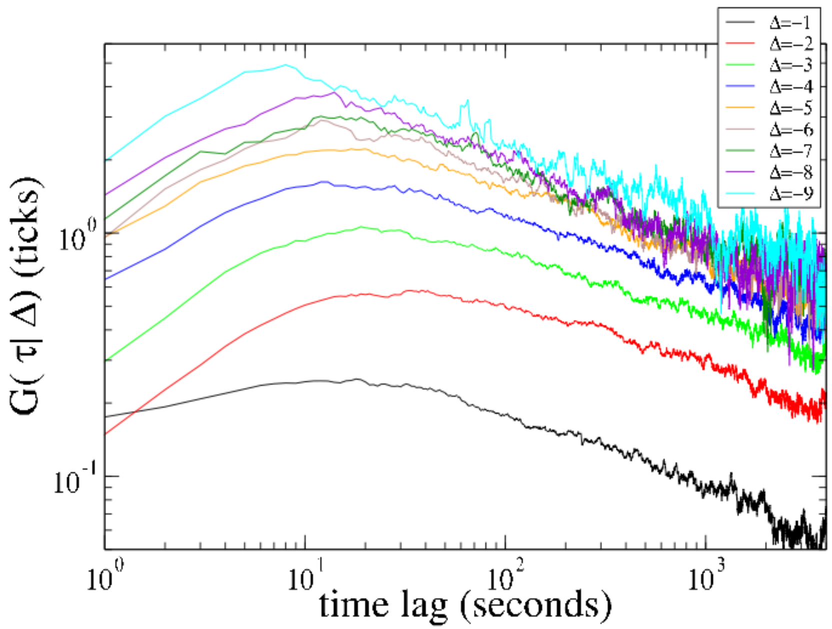


- • Rapidly closing the spread -> priority but risk of informed traders
- • Slowly closing the spread -> "testing" the informed traders but risk of losing priority

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

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

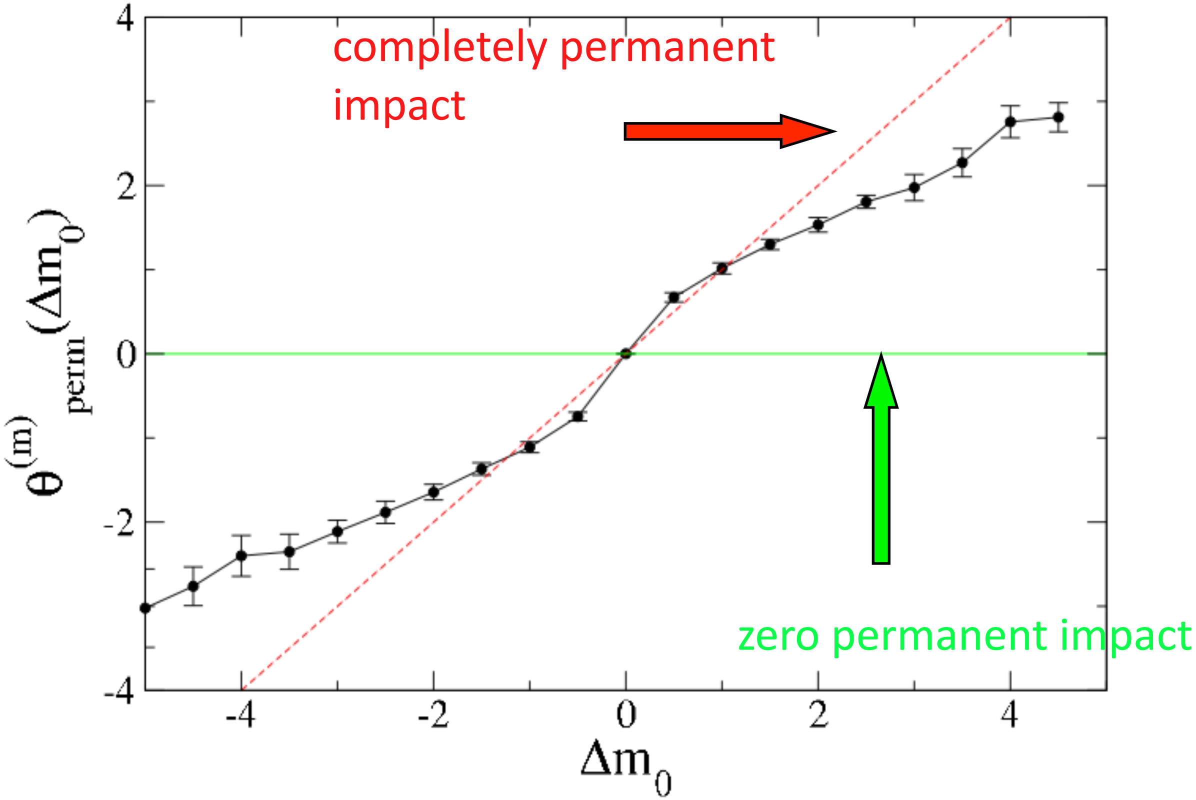




$$G(\tau|\Delta) \sim \tau^{-\delta} \qquad \delta \simeq 0.4 - 0.5$$

Obizhaeva and Wang (2005) postulate an exponential decay

Permanent impact



Permanent impact is roughly proportional to immediate impact

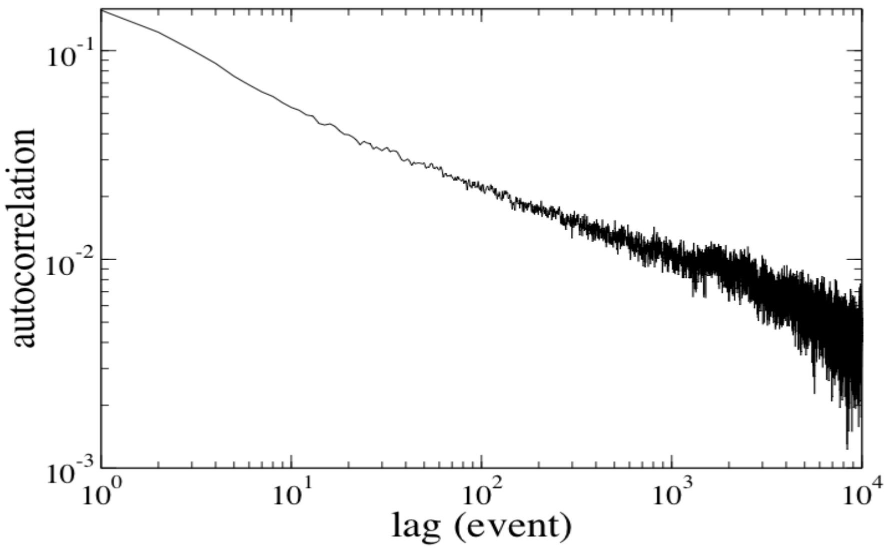
Market order flow

- Let us turn our attention to order flow
- 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,...

- We consider the sign in order to remove any effect of the time persistence of trading volume

Order flow dynamics

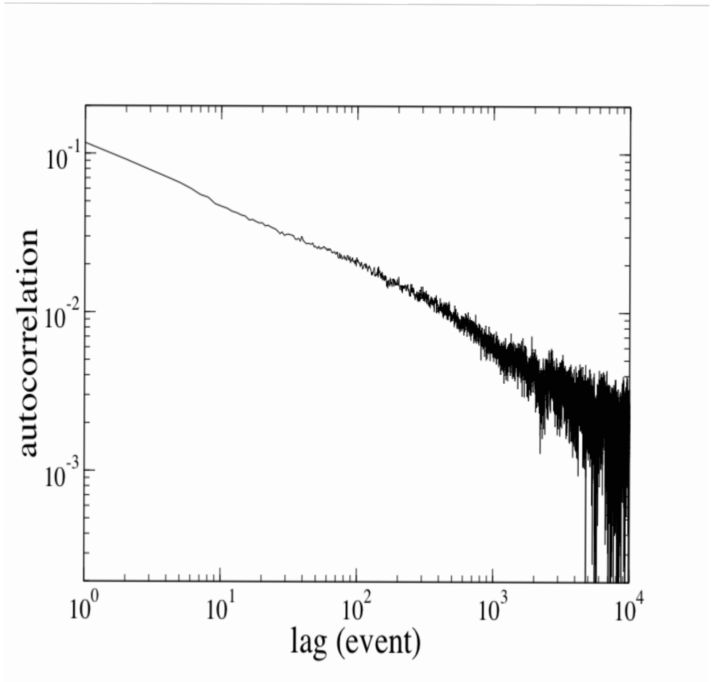


Time series of signs of market orders is a **long memory process** (Lillo and Farmer 2004, Bouchaud et al 2004)

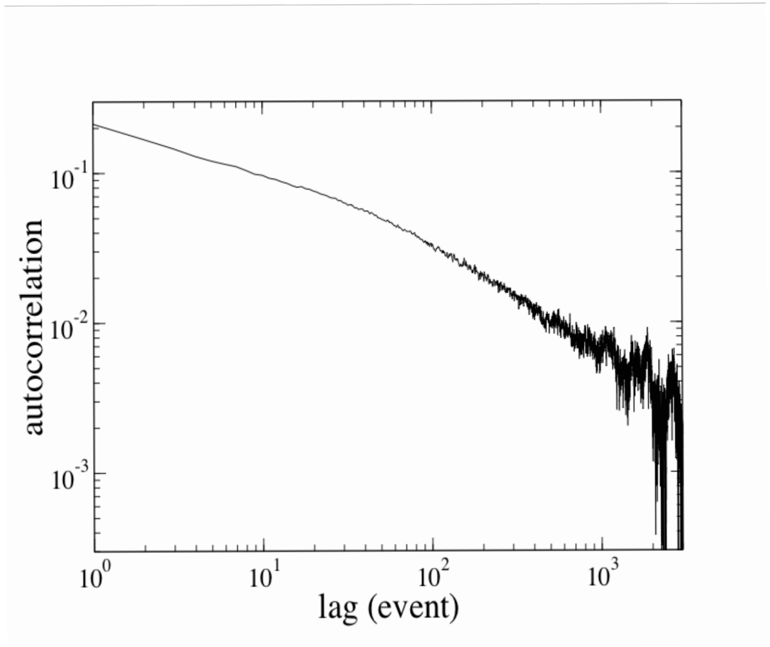
$$C(\tau) \approx \tau^{-\gamma} \approx \tau^{-0.5}$$

- How is this finding consistent with linear unpredictability of returns ?
- Why is the order flow a long-memory process ?

Limit Orders



Cancellations



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

Hurst exponent →

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

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:

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

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.

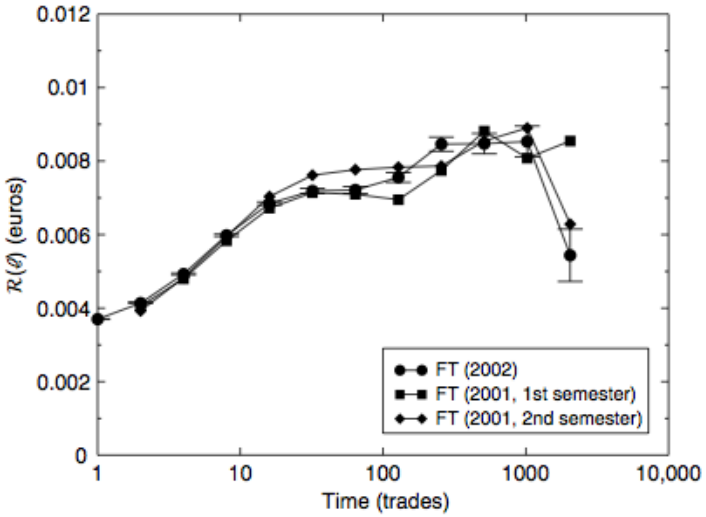


FIGURE 2.8 Average empirical response function $\mathcal{R}_\ell$ for FT during three different periods (1st and 2nd semester of 2001 and 2002); error bars are shown for the 2002 data. For the 2001 data, the y axis has been rescaled such that $\mathcal{R}_1$ coincides with the 2002 result. $\mathcal{R}_\ell$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

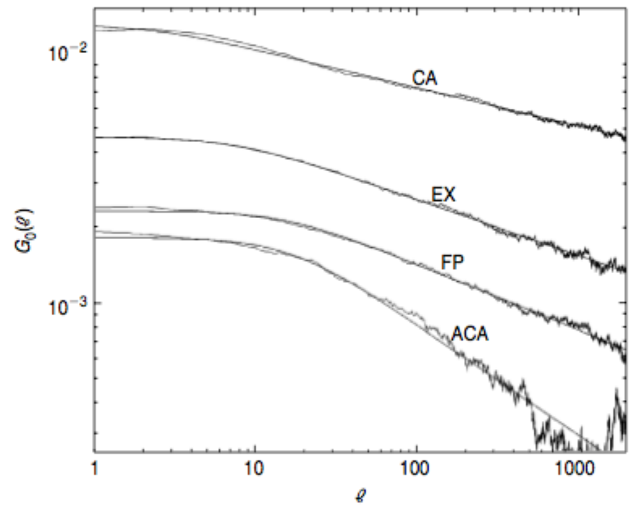


FIGURE 2.9 Comparison between the empirically determined $G_0(\ell)$, extracted from $\mathcal{R}$ and $\mathcal{C}$ using Eq. 2.28, and the power-law fit $G_0^f(\ell) = \Gamma_0/(\ell_0^2 + \ell^2)^{\beta/2}$ for a selection of four stocks: ACA, CA, EX, and FP.

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 q^+_t (q^-_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

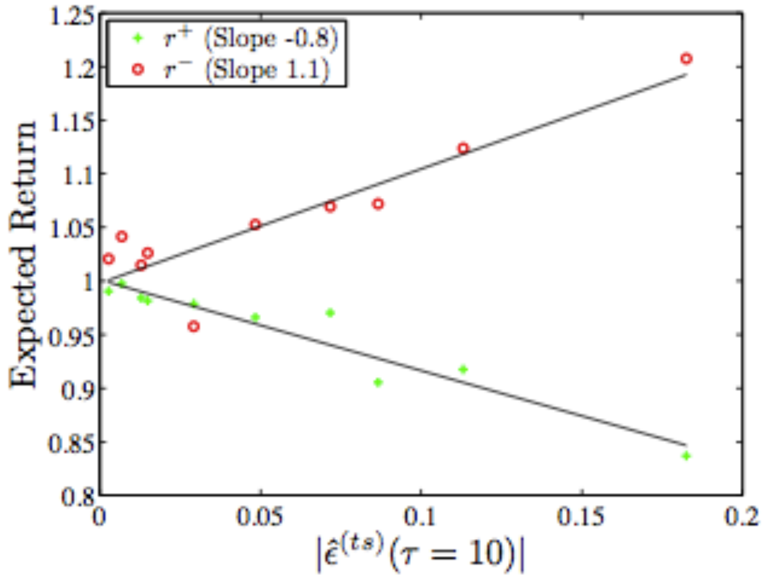
$$q_t^+ r_t^+ - q_t^- r_t^- = 0$$

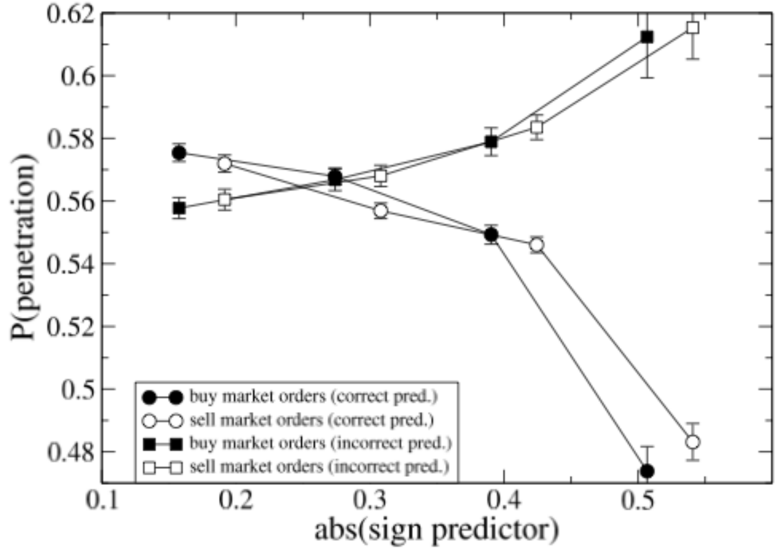
- The market maker has expectations on q^+_t and q^-_t given her information set Ω and adjusts r^+_t and r^-_t in order to make the market efficient

-----> 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. $E\left[\varepsilon_t \varepsilon_{t+\tau}\right] \approx \tau^{-\gamma}$ 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 models and optimal execution

Propagator model

- A persistent order flow is incompatible with market efficiency if the market impact is fixed and permanent
- Two modifications has been proposed:
  - The impact is fixed but temporary (propagator model, Bouchaud et al 2004)
  - The impact is variable (i.e. history dependent) but permanent (Lillo and Farmer 2004)
- The two models are equivalent under some conditions (Bouchaud, Farmer, and Lillo 2009)
- The propagator model assumes that

$$S_n = S_{-\infty} + \sum_{k=1}^{\infty} \epsilon_{n-k} f(v_{n-k}) G_0(k) + \sum_k \eta_k \quad (5)$$

or

$$S_{n+1} - S_n = G_0(1)\epsilon_n f(v_n) + \sum_{k=1}^{\infty} [G_0(k+1) - G_0(k)]\epsilon_{n-k} f(v_{n-k}) + \eta_n \quad (6)$$

Objective function

- An investor has X shares to trade in N time periods. Let v_k $(k = 1, ..., N)$ be the (signed) number of shares to be traded in interval k. Let S_k be the price at which the investor trades at interval k and S_0 the price before the start of the execution
- One very used objective function is the *implementation shortfall* defined as

$$C(\mathbf{v}) \equiv \sum_{k=1}^{N} v_k S_k - X S_0 \quad (7)$$

i.e. the difference between the cost and the cost in an infinitely liquid market.

- The implementation shortfall is in general a stochastic variable, therefore one often wants to minimize $E[C(\mathbf{v})]$. This assumes a risk neutral profile
- Almgren and Chriss (2000) introduced a risk term in optimal execution by setting the problem

$$\mathbf{v}^* = \arg\min_{\mathbf{v}}(E[C(\mathbf{v})] + \lambda Var[C(\mathbf{v})]) \quad (8)$$

where λ is an (investor dependent) risk aversion parameter. By changing λ one obtains an *efficient frontier*

Optimal execution with propagator model

- We assume that propagator model describes market impact. The expected implementation shortfall is

$$E[C(\mathbf{v})] = \sum_{k=0}^{N} v_k [\sum_{j=0}^{k} f(v_j) G_0(k-j)] \tag{9}$$

- We now assume that instantaneous impact is linear, $f(v_k) = \theta_k v_k$. We can rewrite

$$E[C(\mathbf{v})] = 2 \sum_{k,j} \theta_k G_0(|k-j|) v_k v_j = \mathbf{v}^T \mathcal{I} \mathbf{v}. \tag{10}$$

where $\mathcal{I}$ is a Toeplitz matrix (diagonal constant)

- We thus have a quadratic optimization problem (as in portfolio optimization)

$$\mathbf{v}^* = \arg\min_{\mathbf{v}} \mathbf{v}^T \mathcal{I} \mathbf{v} \qquad s.t. \quad \sum_k v_k \equiv \mathbf{1}^T \mathbf{v} = V \tag{11}$$

that can be solved with a Lagrange multiplier with solution

$$\mathbf{v}^\star = \frac{V}{\mathbf{1}^T \mathcal{I}^{-1} \mathbf{1}} \mathcal{I}^{-1} \mathbf{1}. \tag{12}$$

Solution

- The solution is symmetric around $N/2$ (see also Alfonsi, Schied, Slynko 2011)
- For an exponential G_0 one reobtains the solution of Obizhaeva and Wang 2006
- It can be solved analytically also considering a risk term and assuming a Gaussian noise for η in the propagator model

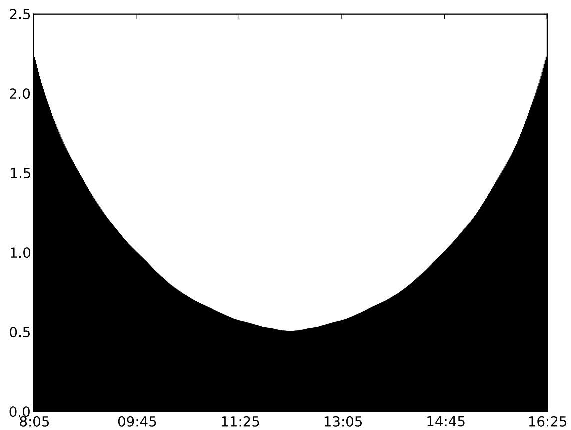


Figure: An example of theoretical optimal solution of the optimal execution problem. As a function of real time, the plot shows the amount of shares to be traded (in arbitrary units).

Calibration on real data

- We calibrate the model on real data from the LSE (two sets, 2000-2002 and 2011 data)
- We consider intervals of 5 minutes and execution for the whole day. We take into account intraday periodicities.

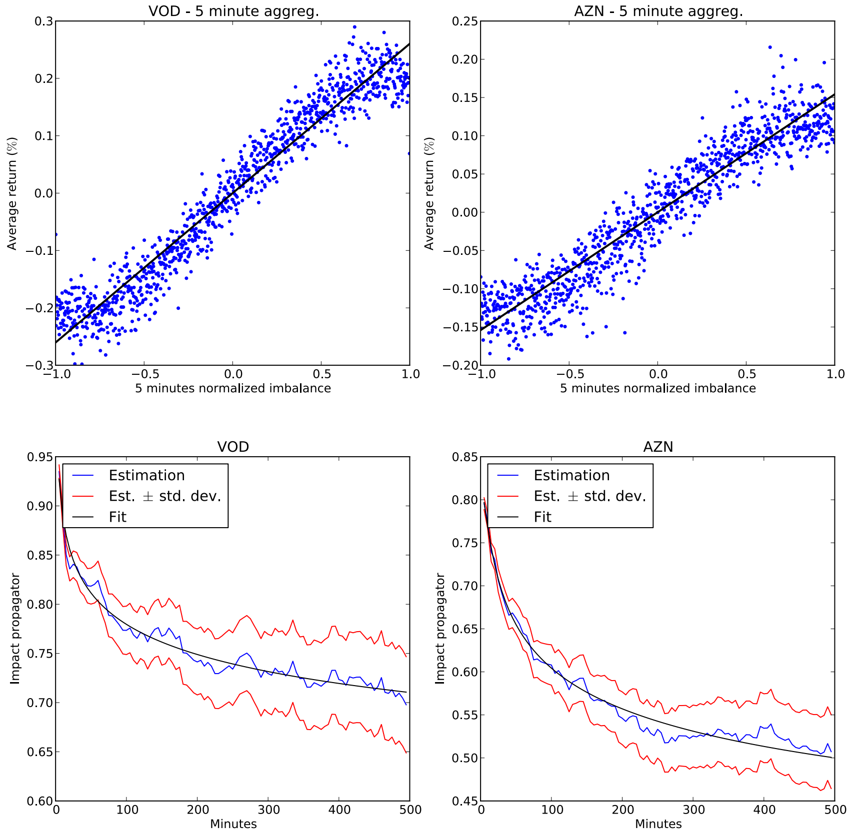


Figure: Impact (top) and propagator (bottom) from 5 min imbalance data

Optimal trading profile

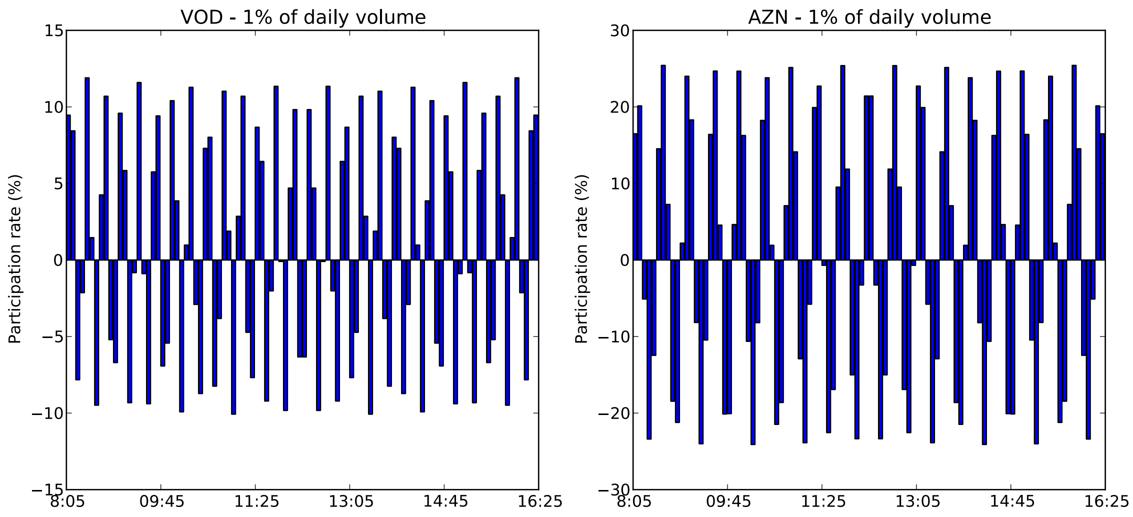


Figure: Optimal solution for two stocks

The optimal execution for a buy trade includes buys and sells !! (see also Alfonsi, Schied, Slynko 2011). The cost is positive (no price manipulation)

Including spread costs

In this derivation we have neglected any cost term related to trading (fees, spread). While fixed and proportional fees do not affect the qualitative properties of the results, spread costs change them significantly.
The optimization with spread costs becomes

$$F[\mathbf{v}] = E[C(\mathbf{v})] + BA(\mathbf{v}) = \mathbf{v}^T \mathcal{I} \mathbf{v} + \mathbf{E}^T |\mathbf{v}|. \tag{13}$$

where $\mathbf{E}$ is a vector describing the spread cost during execution
The absolute value prevents an analytical solution and we use numerical optimization

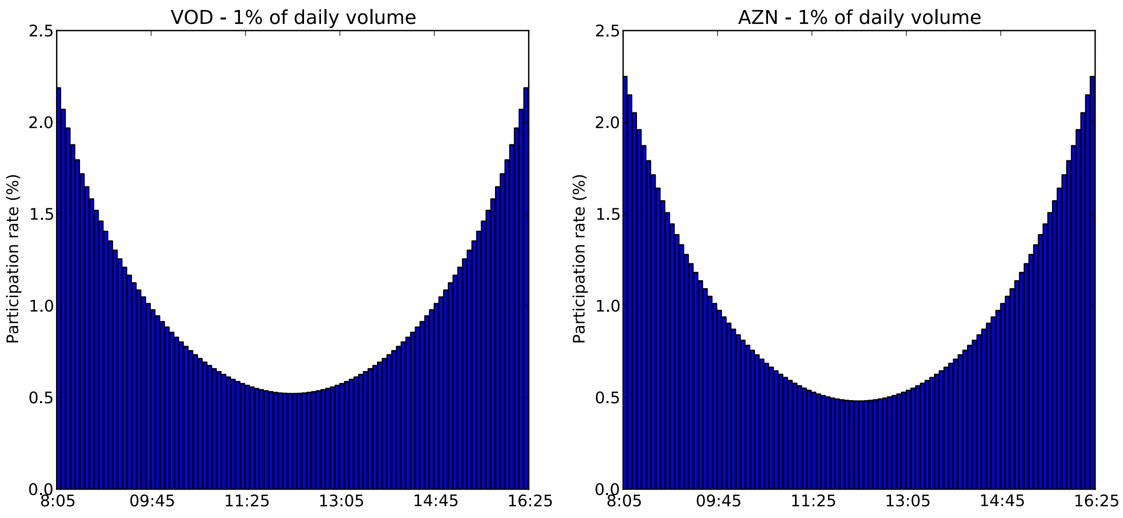


Figure: Spread costs regularize the solution (no sells for a buy program)

Including the risk term

More risk aversion leads to more trading at the beginning of the program

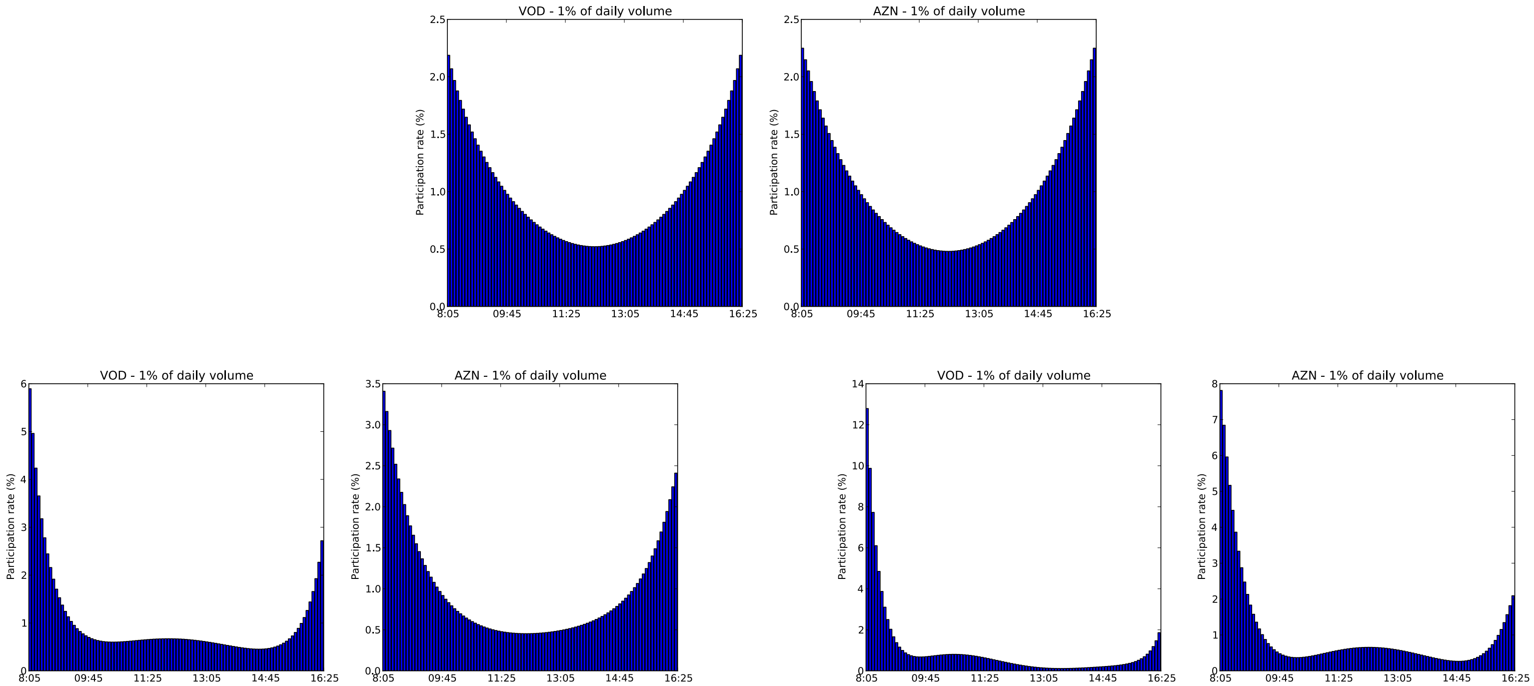


Figure: $\lambda = 0$ (top), $\lambda = 0.2$ (bottom left), $\lambda = 0.9$ (bottom right)

Efficient frontier

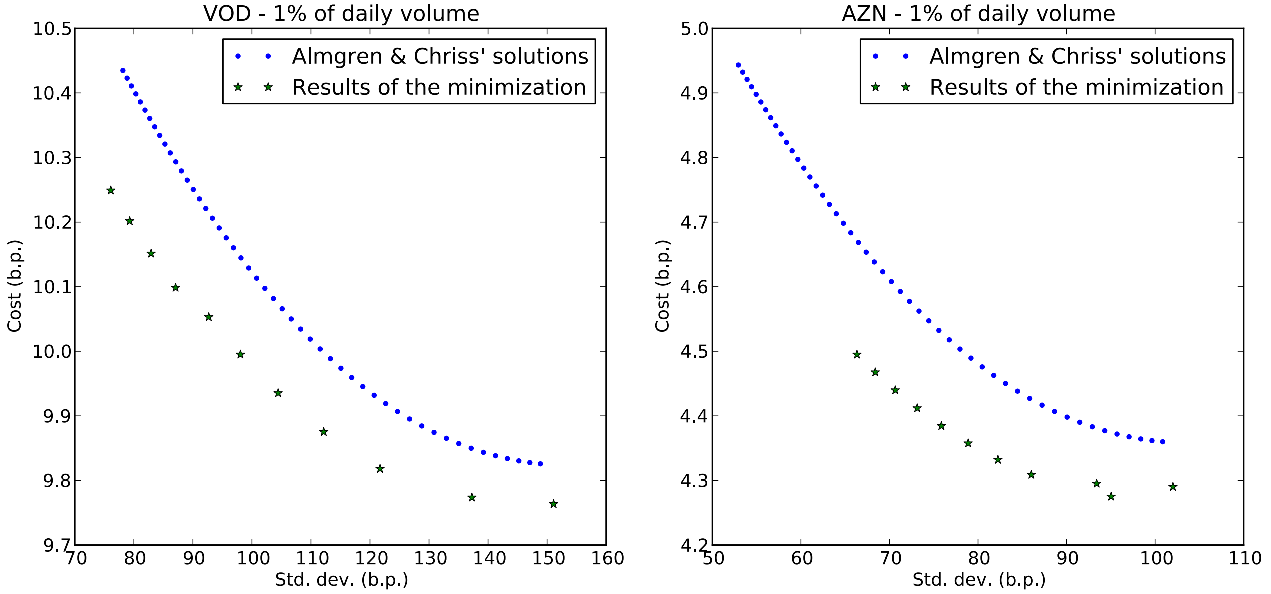


Figure:

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 trader and a contribution from all other trader.

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, 2011)

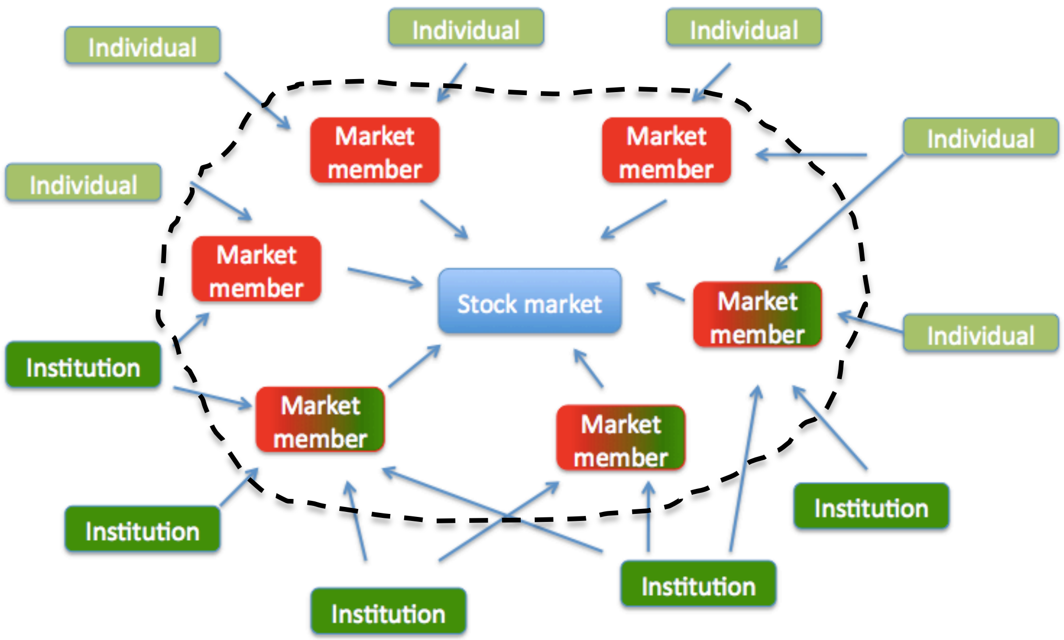
Market member (broker) data

- The investigated markets are:
- Spanish Stock Exchange (BME) 2001-2004
- London Stock Exchange (LSE) 2002-2004

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

❑ Roughly 200 Firms in the BME and LSE (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)

Response function is a delicate balance

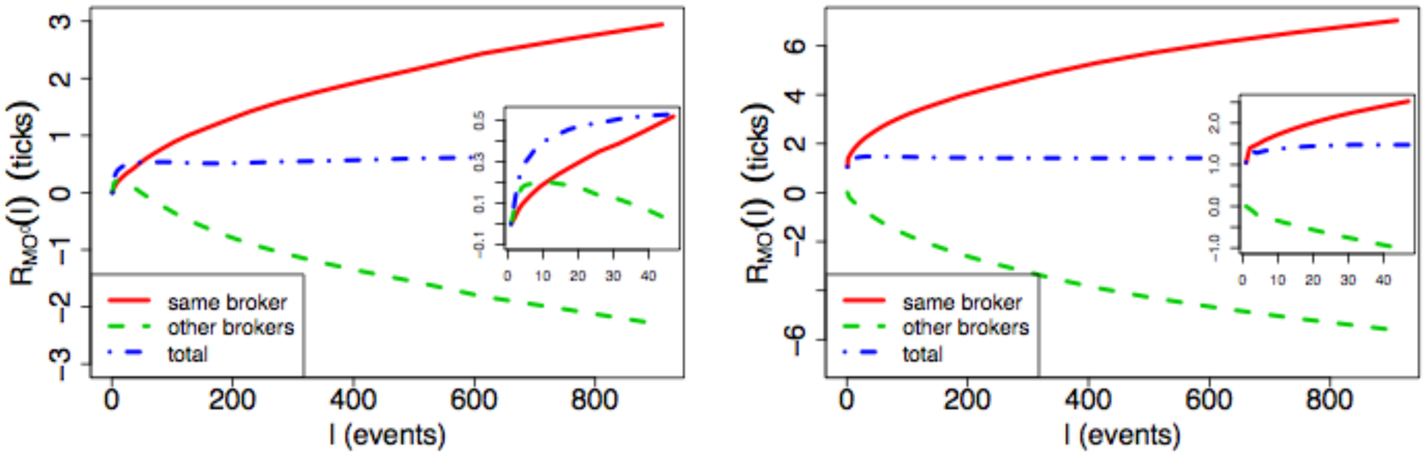


Figure 1: The response function $\mathcal{R}_{\pi_1}(\ell)$ and its contributions coming from the orders of the same broker ($\mathcal{R}^{\rm same}_{\pi_1}(\ell)$) and of different brokers ($\mathcal{R}^{\rm diff}_{\pi_1}(\ell)$). (left) The case of $\pi_1 = \mathrm{MO}^0$. (right) The case of $\pi_1 = \mathrm{MO}'$. 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).

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 $P^{ij}(\tau) = N^{ij}(\tau)/N$

Herding or splitting?

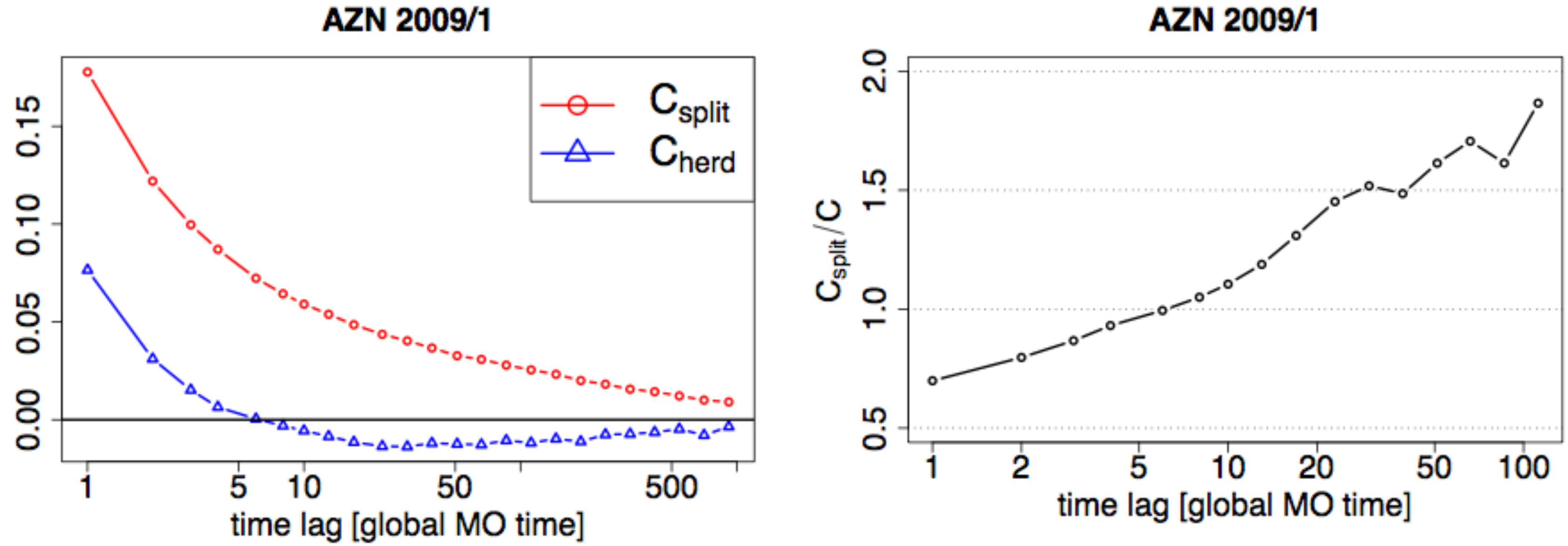


Figure: Left panel. The splitting and the herding term of the correlation of MO signs (the two terms sum up to $C(\tau)$) for the first half year of 2009 for AZN. Right panel. The splitting ratio of MO signs (defined as the ratio of the splitting term in the correlations and the entire correlation) for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

Splitting dominates herding (especially for large lags)

Splitting of metaorders

- 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 metaorders 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 metaorders from the trading profile of the investors and measure their impact

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

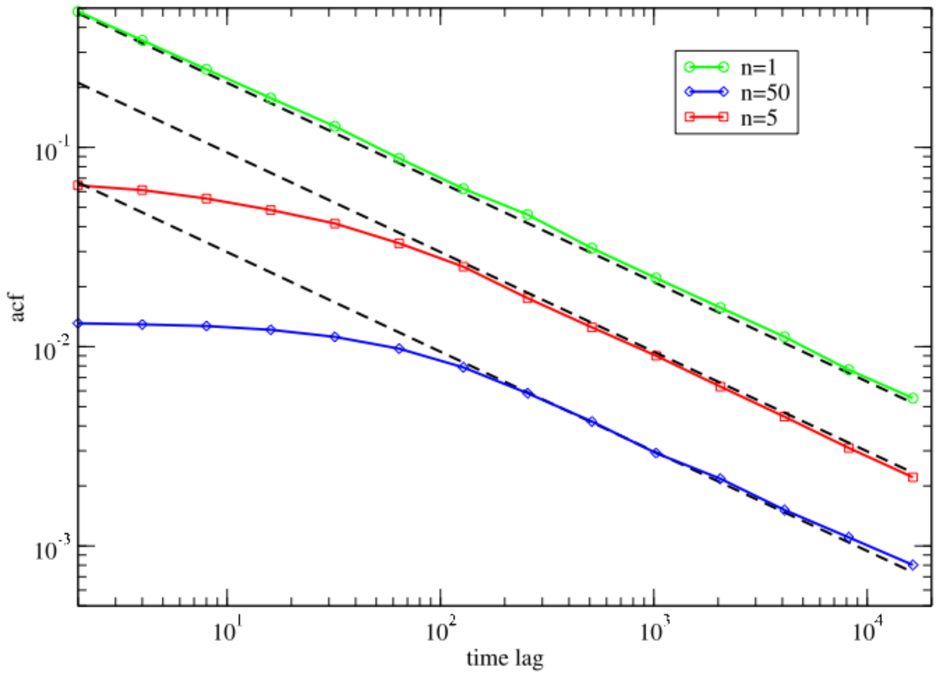
$$p(L) = \frac{\alpha}{L^{\alpha+1}} \qquad L \geq 1 \qquad \alpha > 1$$

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).

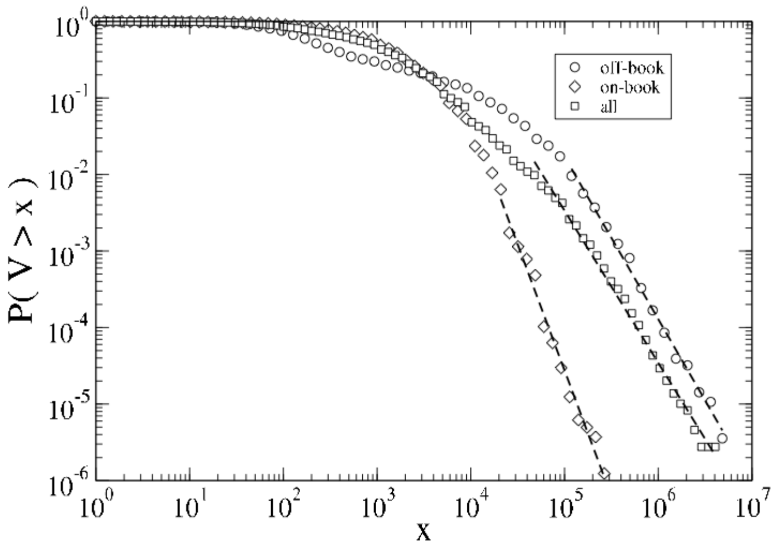
• At LSE there are two alternative methods of trading

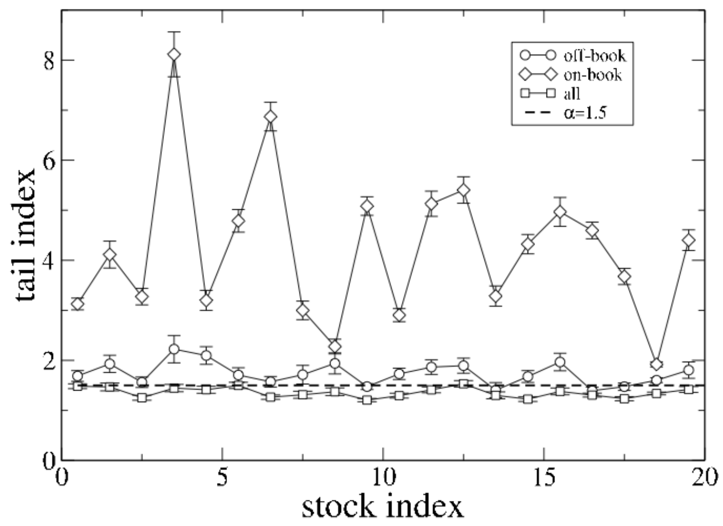
- The on-book (or downstairs) market is public and execution is completely automated (Limit Order Book)
- The off-book (or upstairs) market is based on personal bilateral exchange of information and trading.

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



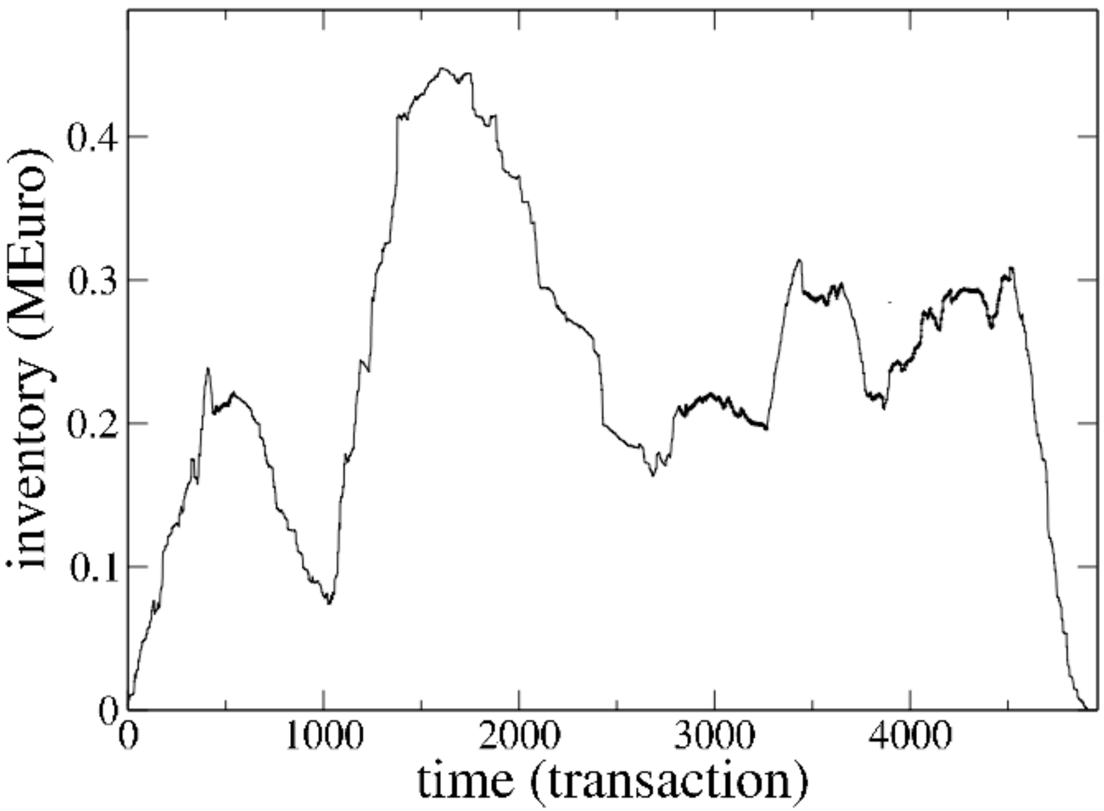


- The exponent α=1.5 for the hidden order size and the market order sign autocorrelation exponent γ are consistent with the order splitting model which predicts the relation $\gamma=\alpha-1$.

Is it possible to identify directly hidden orders?

Brokerage data: A typical inventory profile

Credit Agricole trading Santander at BME

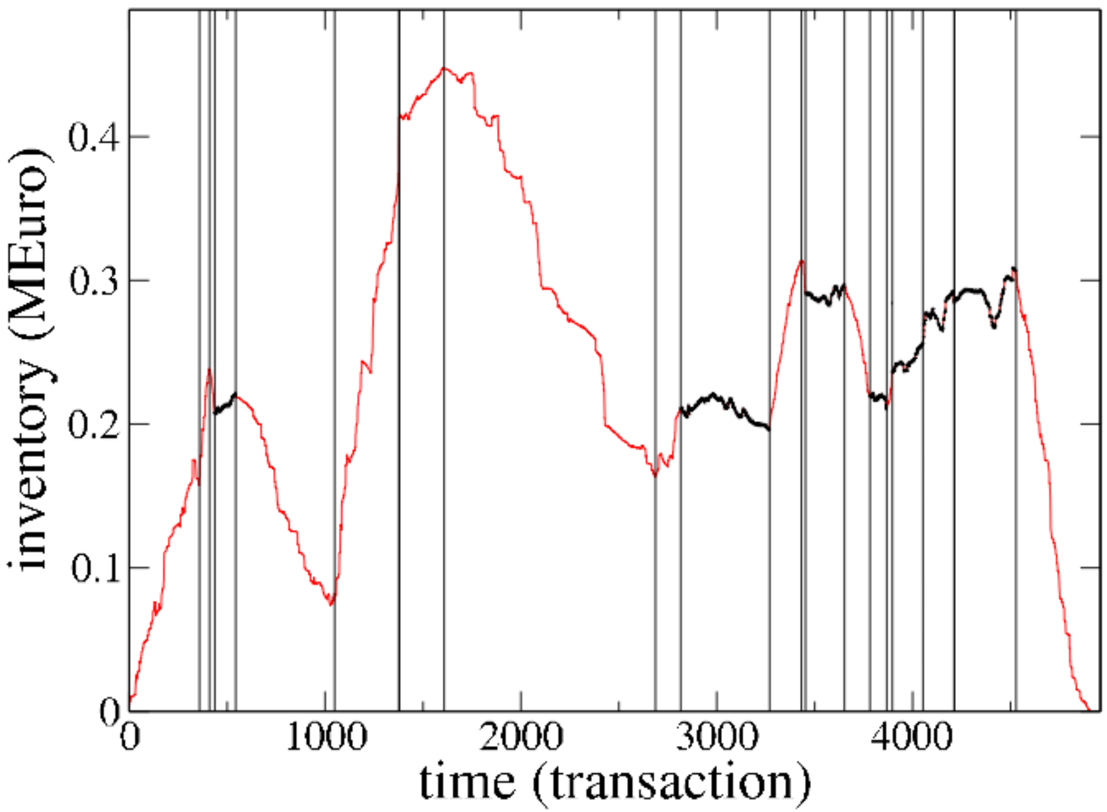


$$I_i^k(t;t_0) = \sum_{\tau=t_0}^{t} \varepsilon_\tau V_\tau$$

A typical regime switching problem

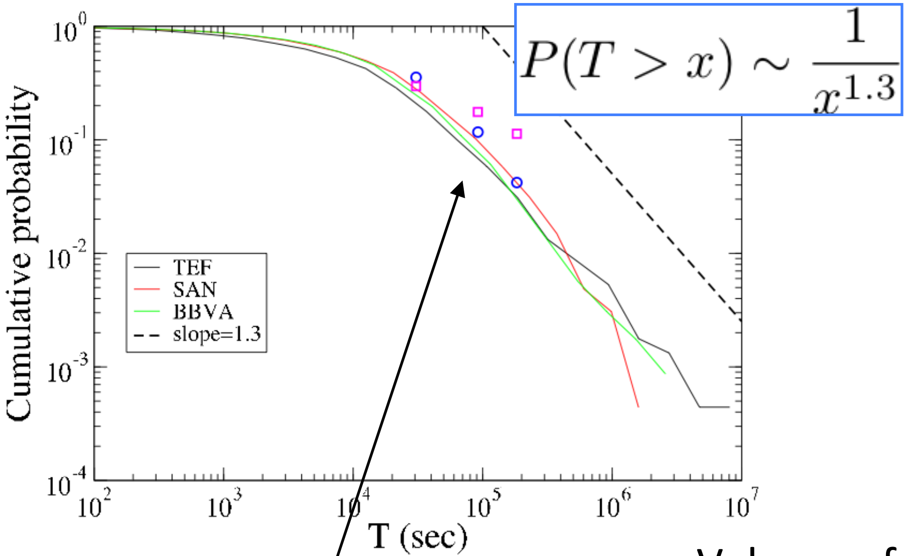
- Modified t-test (G. Vaglica, F. Lillo, E. Moro, and R. N. Mantegna, Physical Review E **77**, 036110 (2008).)
- Hidden Markov Model (G. Vaglica, F. Lillo, and Rosario N. Mantegna, New Journal Of Physics, (in press 2010)).

Detecting hidden orders

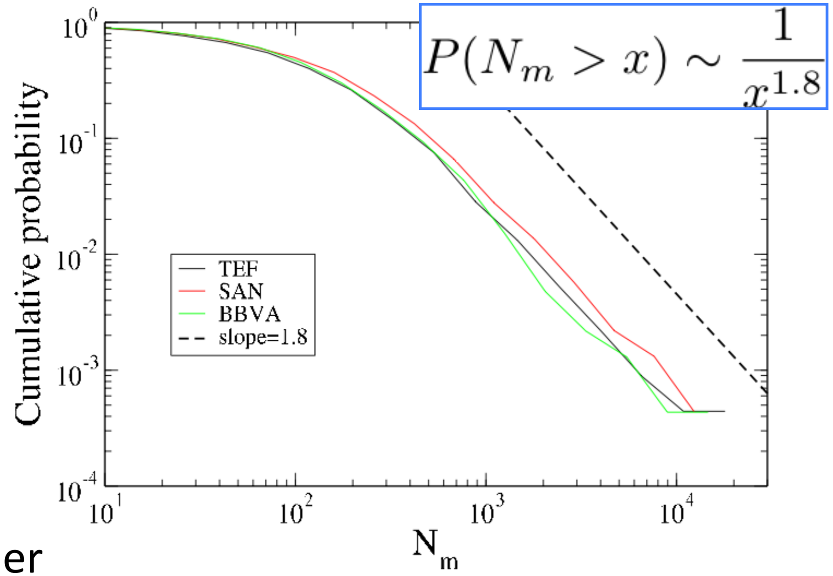


Scaling relations of hidden orders

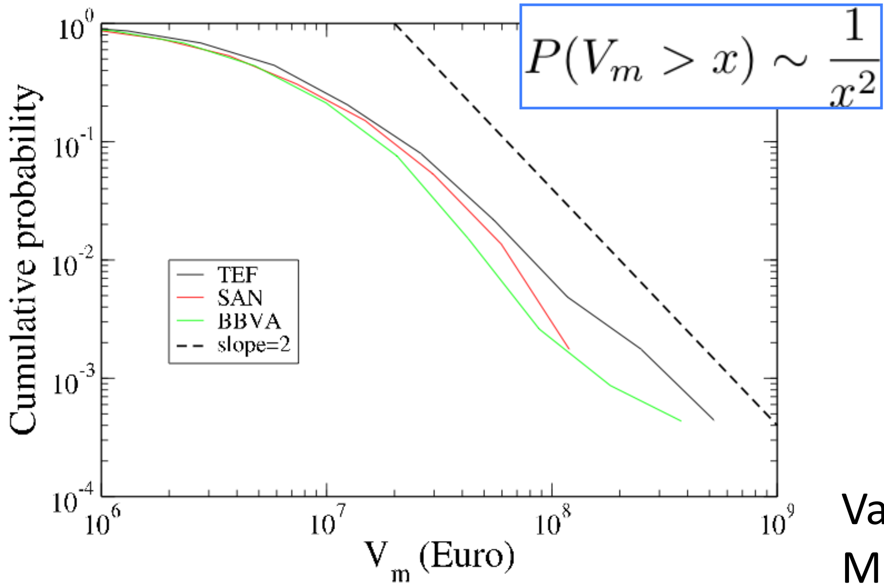
Investment horizon



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)

Vaglica, Lillo, Moro, Mantegna 2008

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

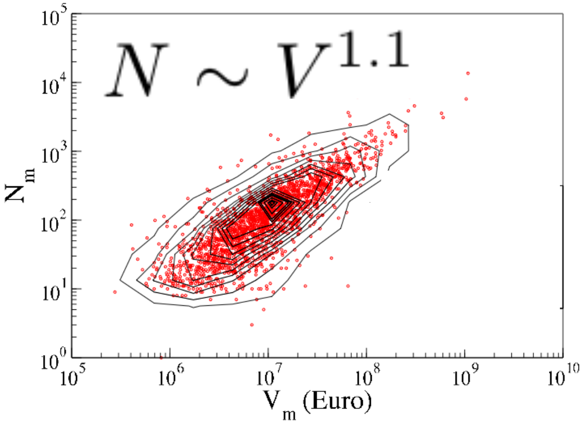
$$P(T > x) \sim \frac{1}{x^{1.3}} \quad P(N_m > x) \sim \frac{1}{x^{1.8}} \quad P(V_m > x) \sim \frac{1}{x^2}$$

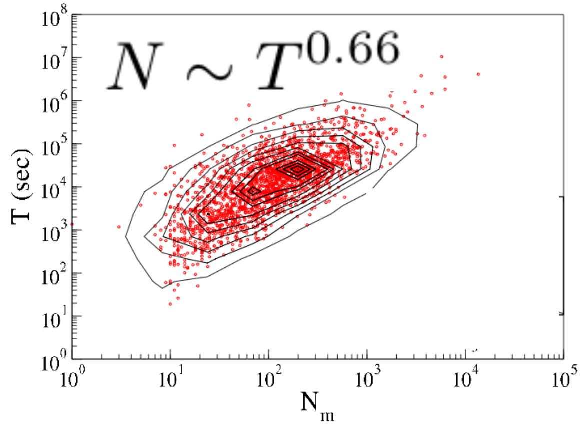
These results are not consistent with the theory of Gabaix et al. Nature 2003)

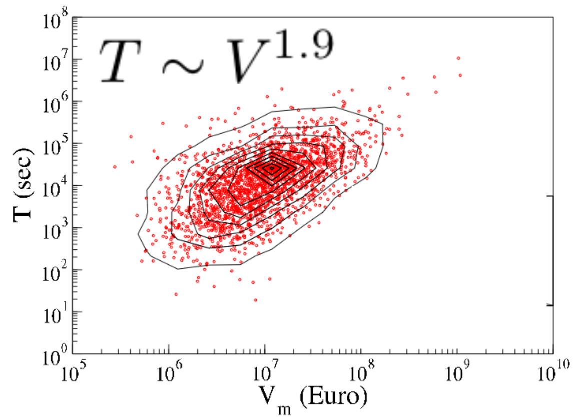
$$P(T > x) \sim \frac{1}{x^3} \qquad P(N_m > x) \sim \frac{1}{x^3} \qquad P(V_m > x) \sim \frac{1}{x^{3/2}}$$

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

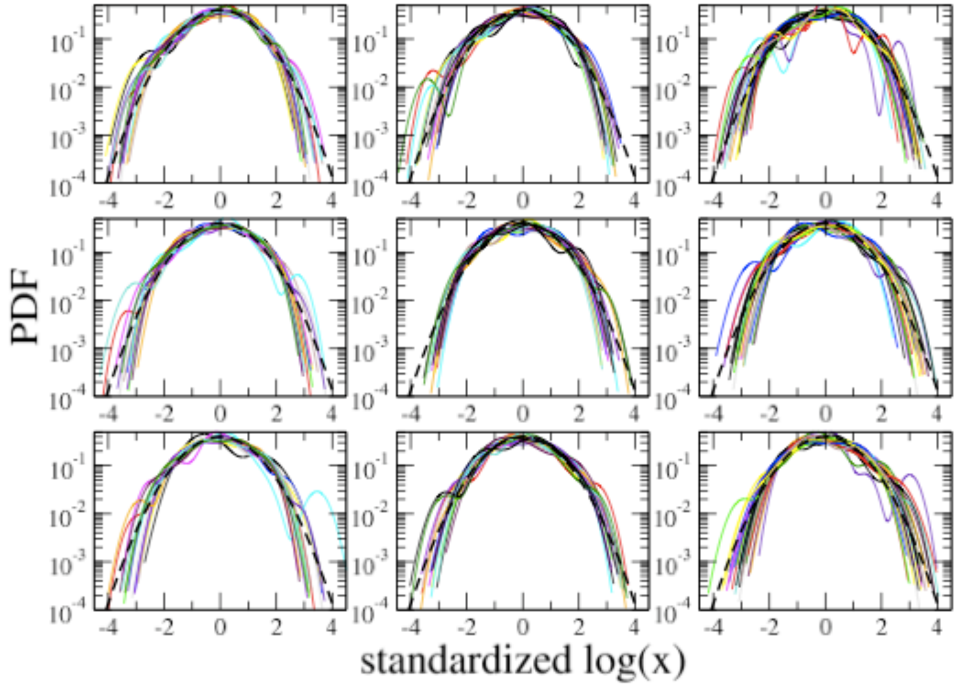


Figure 4.5: Probability density function of the standardized logarithm of the variables T, N_{maj} and V_{maj} of the firms for which the Jarque-Bera test of lognormality cannot be rejected. Specifically, for each stock and each variable we consider the firms for which the lognormal hypothesis cannot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

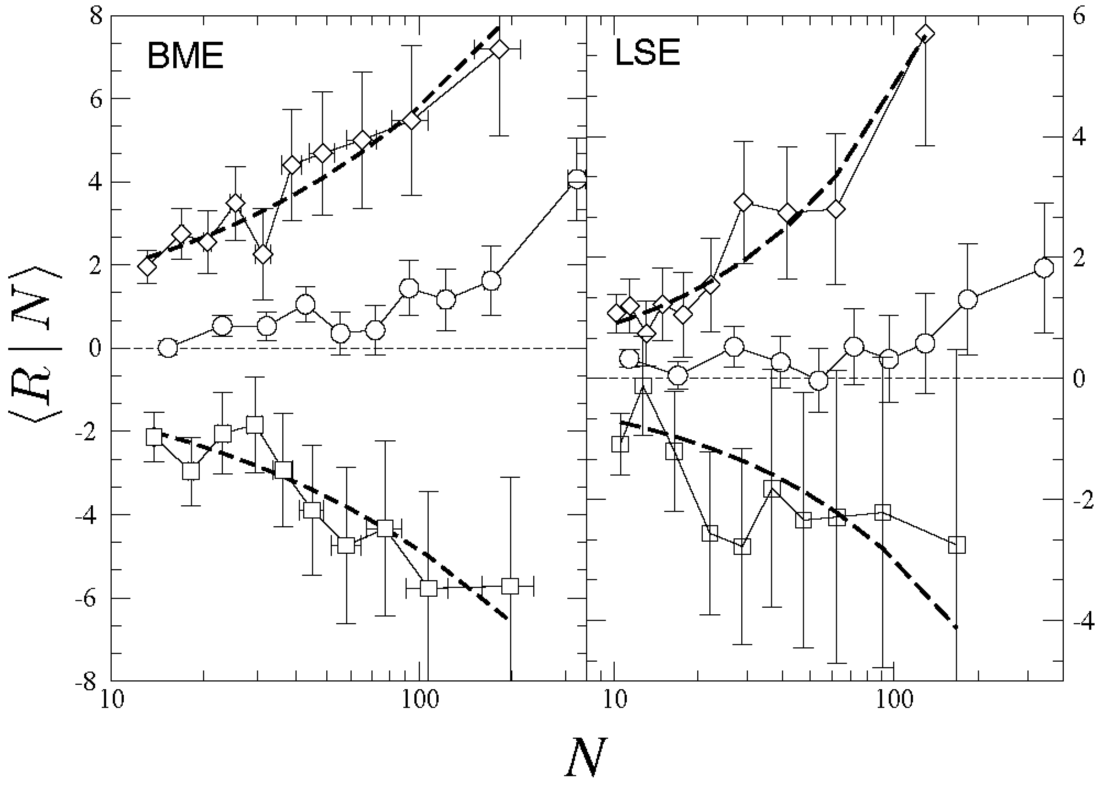


Figure 4: Average rescaled market impact R for hidden orders shorter than 1 day as a function of N for the BME (left) and LSE (right). Circles are the results for all hidden orders, while squares are the results when there is a low fraction of market orders ($f_{mo} < 0.2$) and diamonds are for when 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*

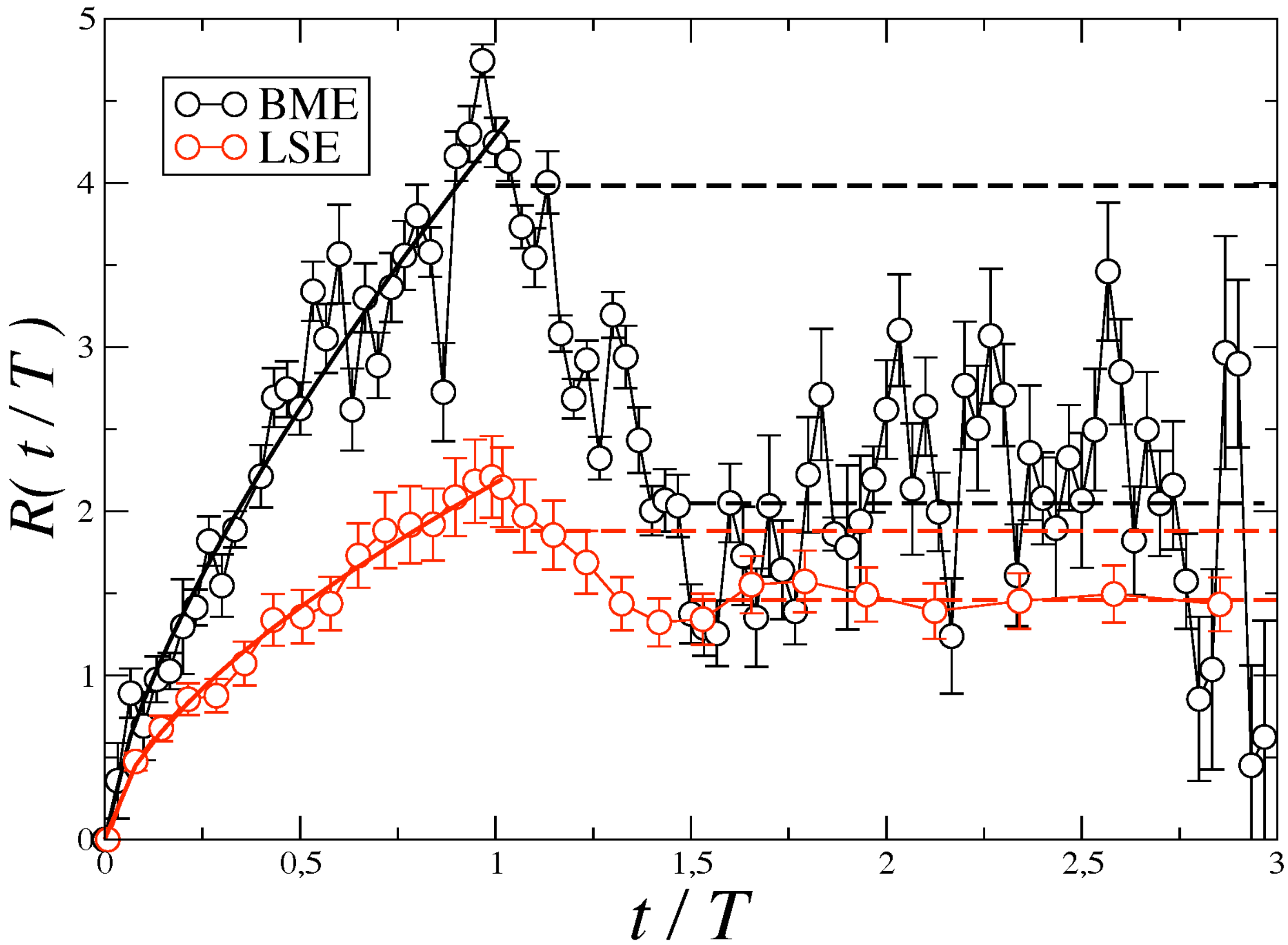
We find that for both groups the relation <R|N> is well described by:

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

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 .

Toward a market ecology

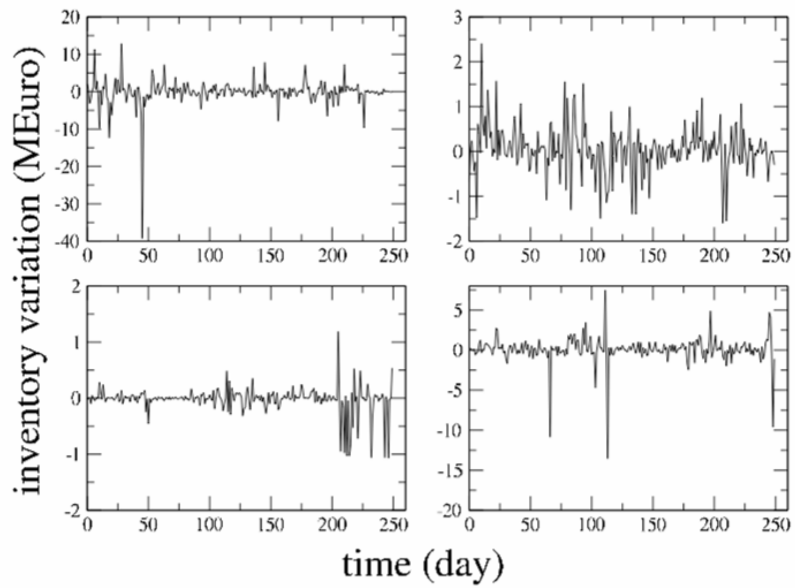
Daily inventory variation time series

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

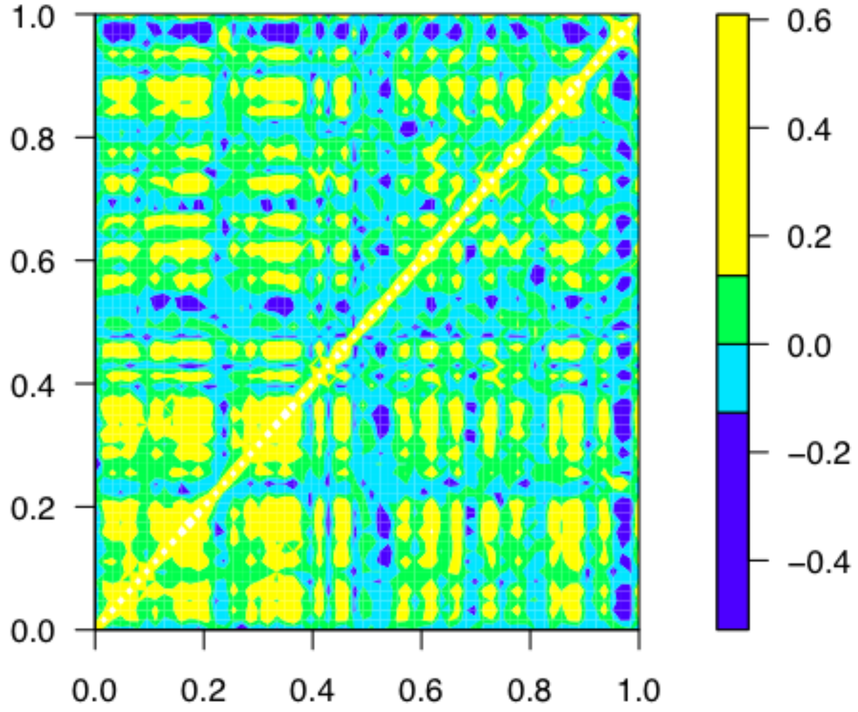
$$v_i(t) \equiv \sum_{s=t}^{t+\tau} \epsilon_i(s) p_i(s) V_i(s)$$

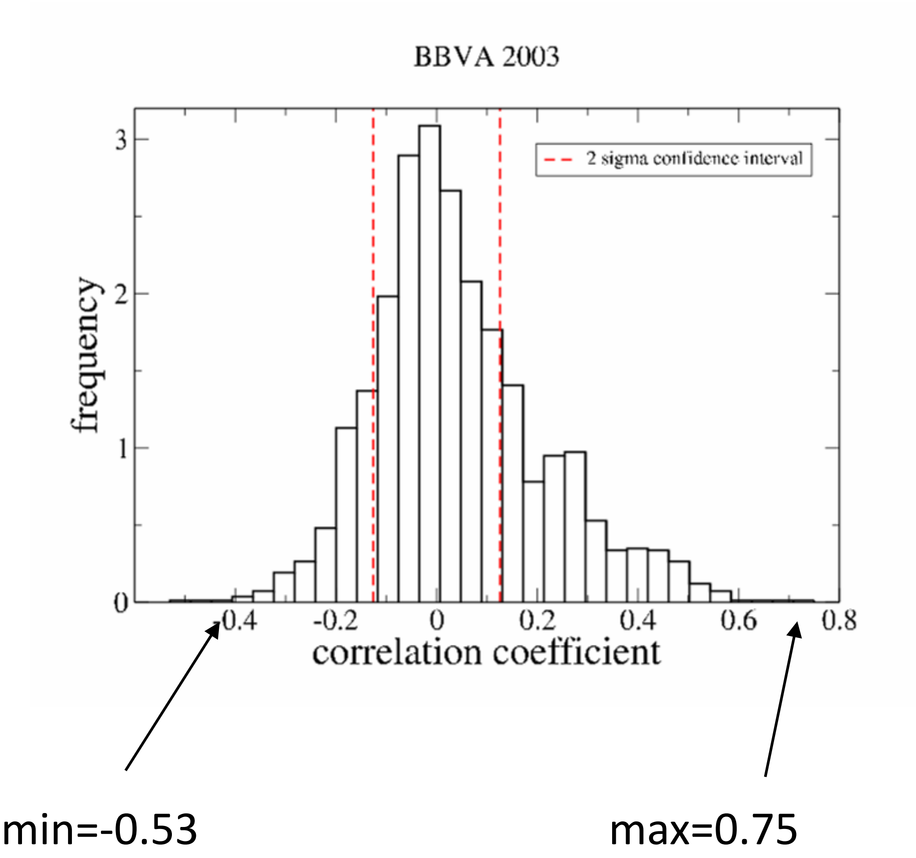
sign　　price　　volume

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



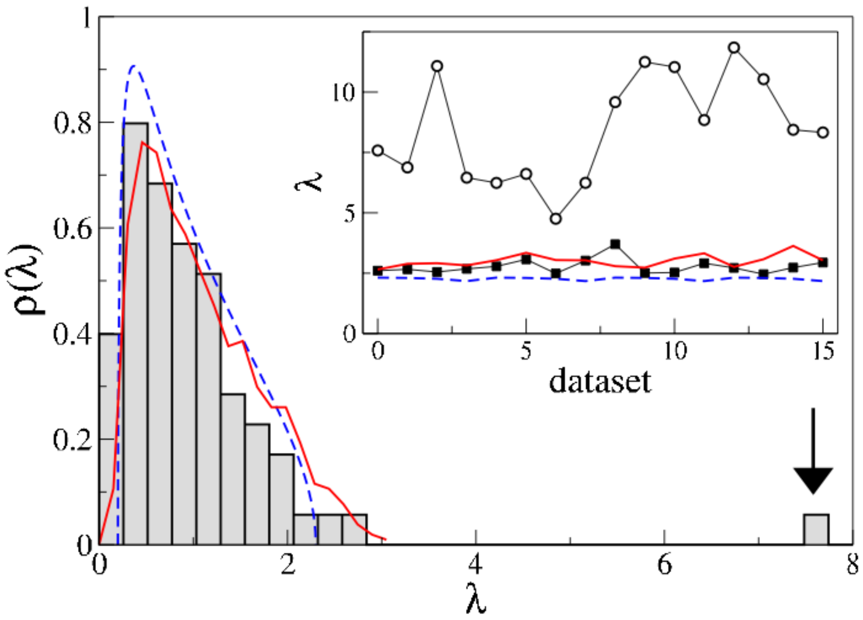
Cross correlation matrix of inventory variation

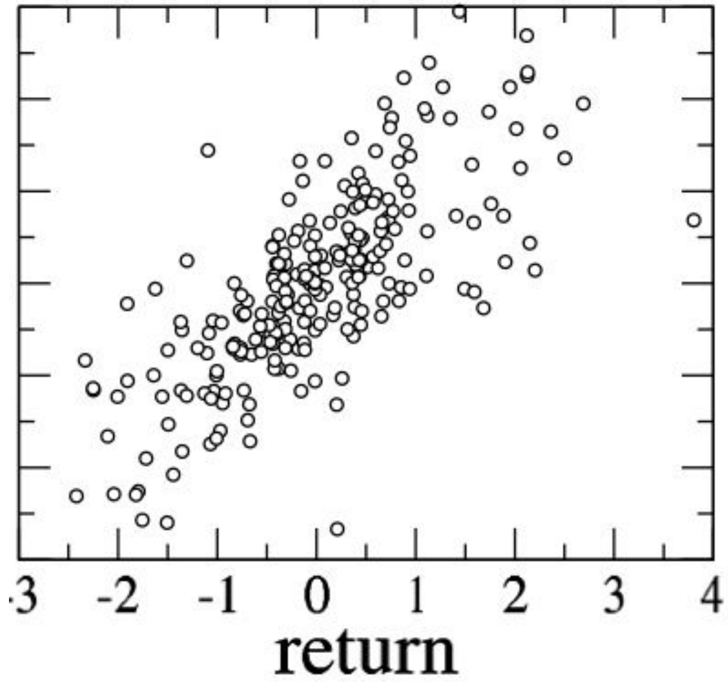




Trading activity is significantly cross correlated among firms

Origin of collective behavior



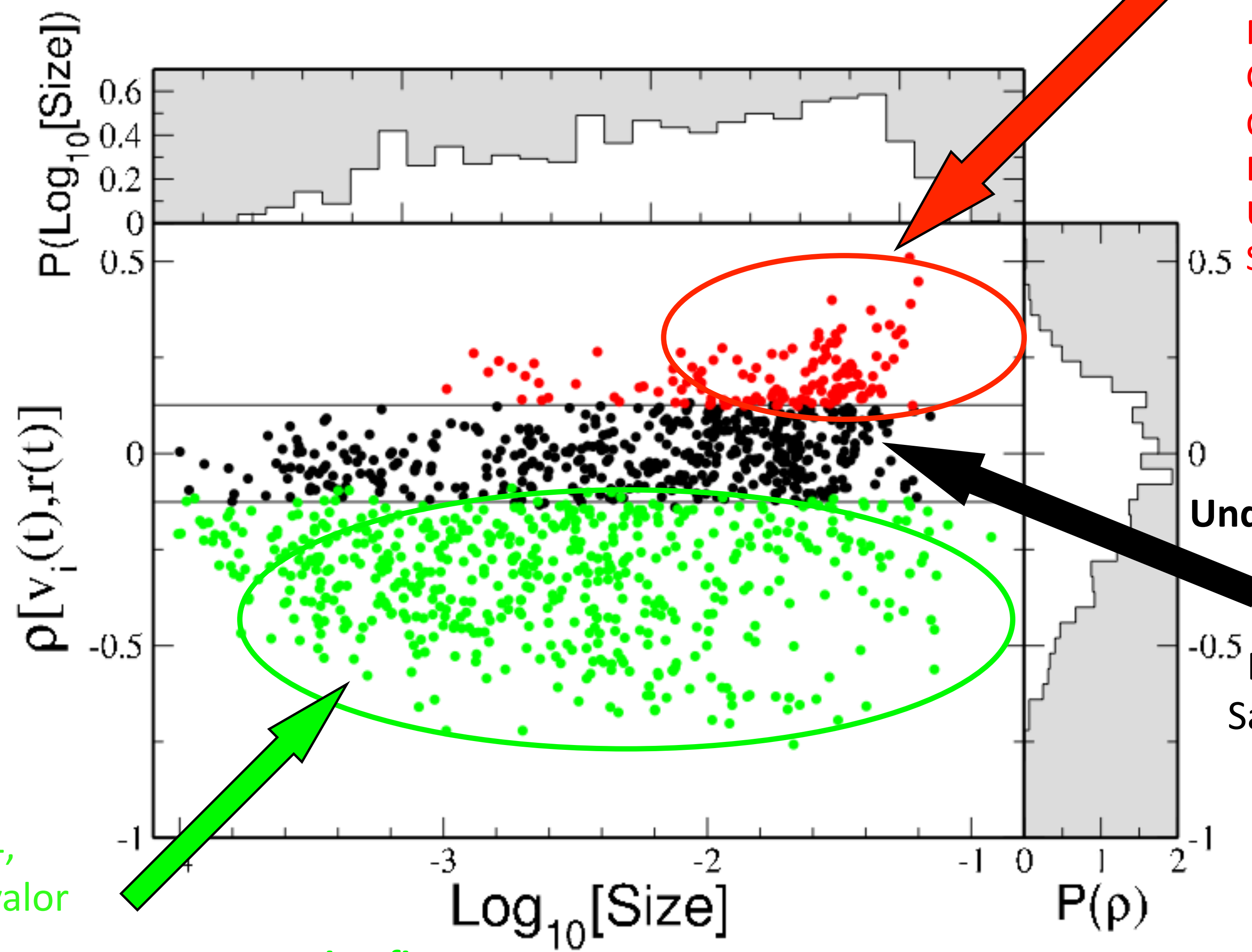


- The first eigenvalue is not compatible with random trading and is therefore carrying information about the collective dynamics of firms.
- The corresponding factor is significantly correlated with price return.
- There are groups of firms having systematically the same position (buy/sell) as the other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

The role of size

Few large trending firms

JP Morgan,
Merryl Lynch,
Credit Suisse,,
Credit Agricole,
BNP Paribas,
UBS Warburg,
Societé General



Undetermined firms

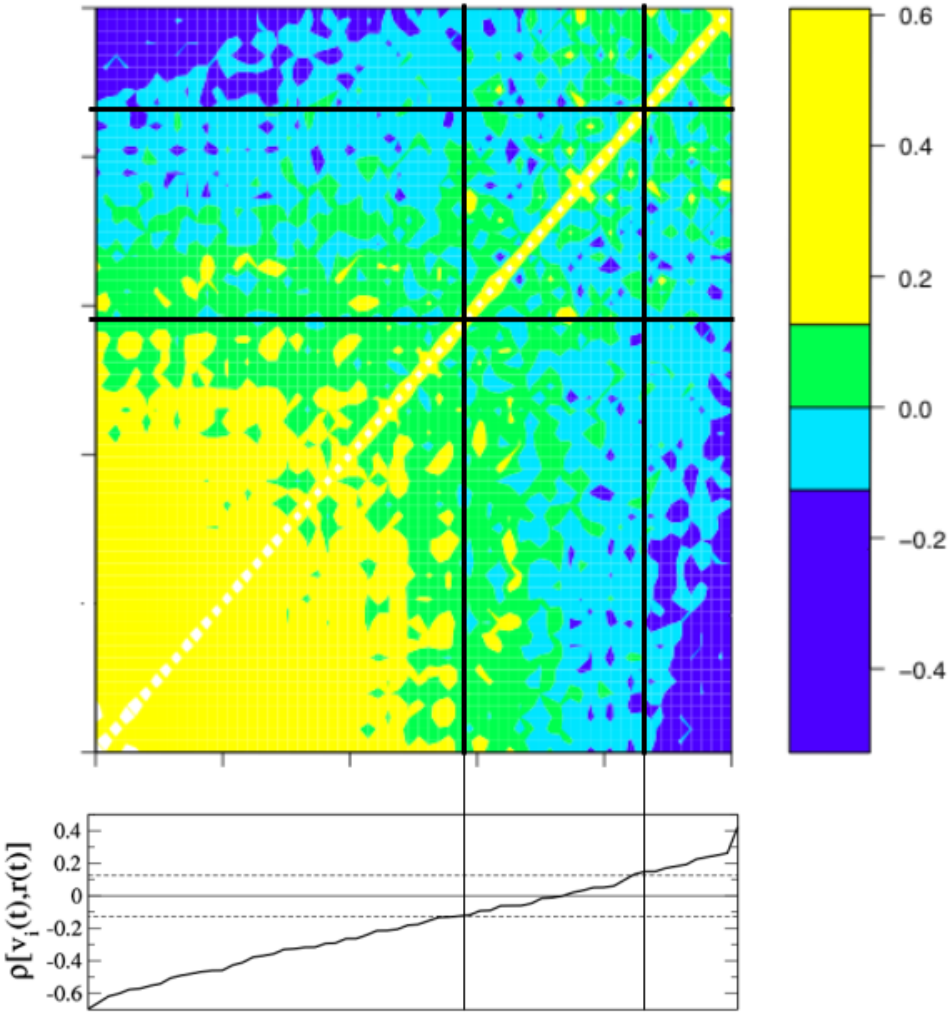
BBVA
Santander

Renta 4,
Mercavalor

Many heterogeneous reversing firms

Size = average daily fraction of volume

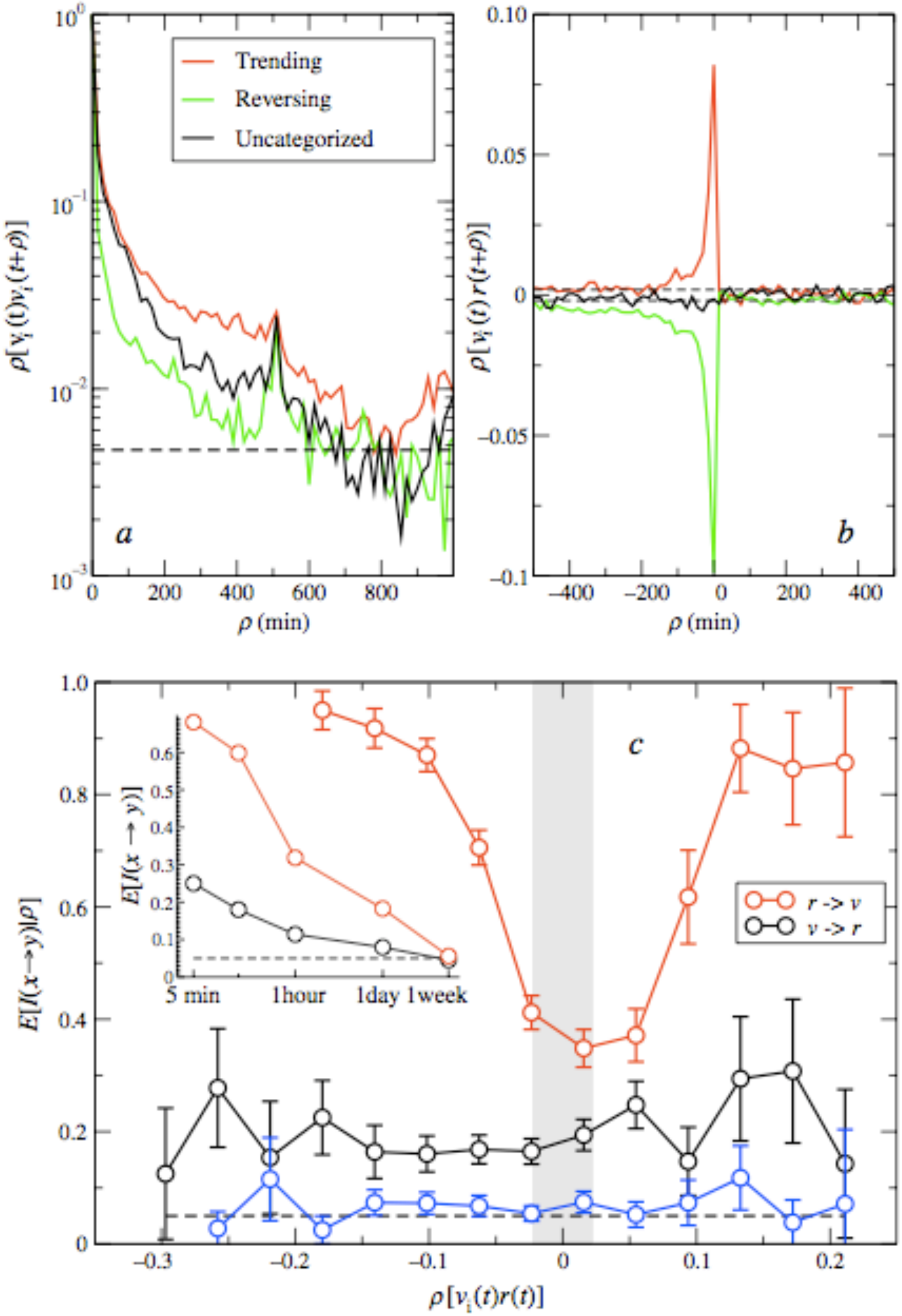
Inventory variation correlation matrix obtained by sorting the firms in the rows and columns according to their correlation of inventory variation with price return



Correlated order flow

Inventory variation is long range correlated

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



Granger causality

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

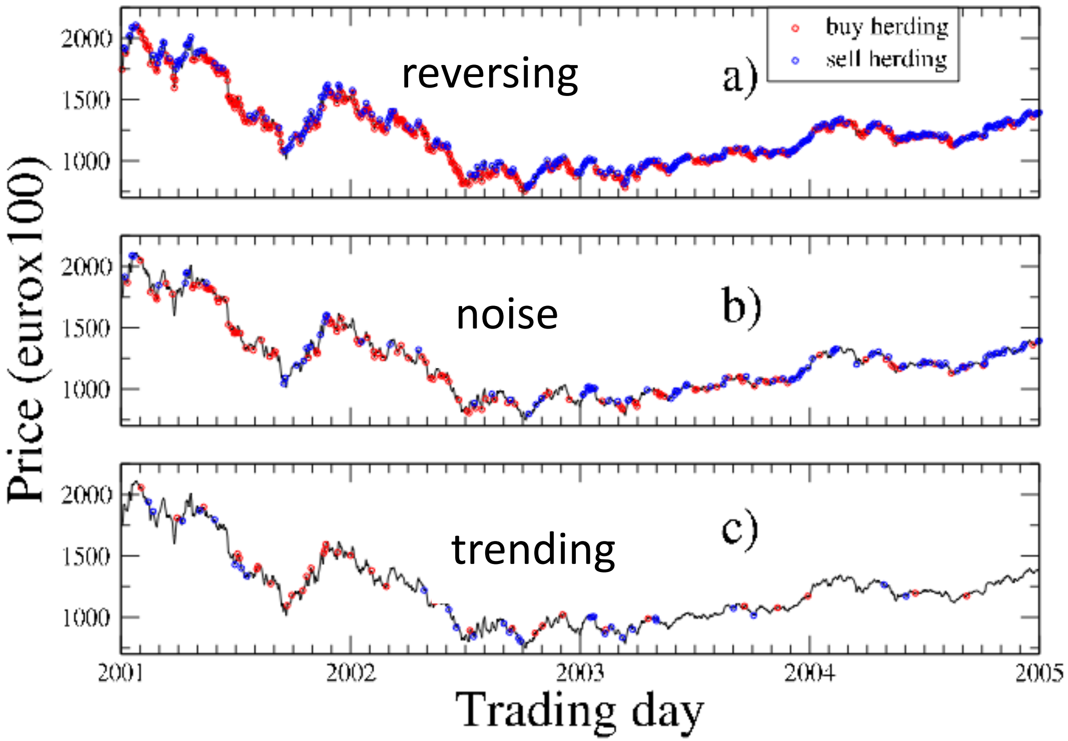
- Returns cause inventory variations
- Inventory variations does not cause returns

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 (almost) open**

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

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).