Introduction to market microstructure and heterogeneity of investors

Fabrizio Lillo

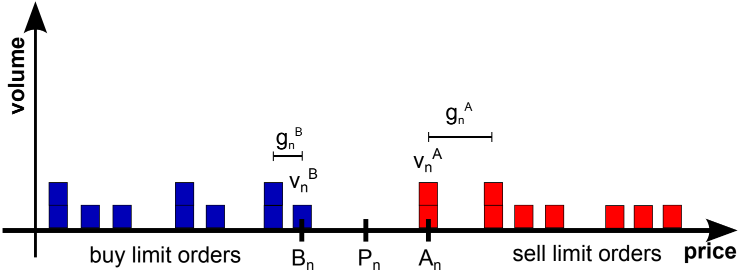
https://fabriziolillo.wordpress.com

Dipartimento di Matematica, Universitá di Bologna
Porta di Piazza San Donato 5, 46126 Bologna, Italy

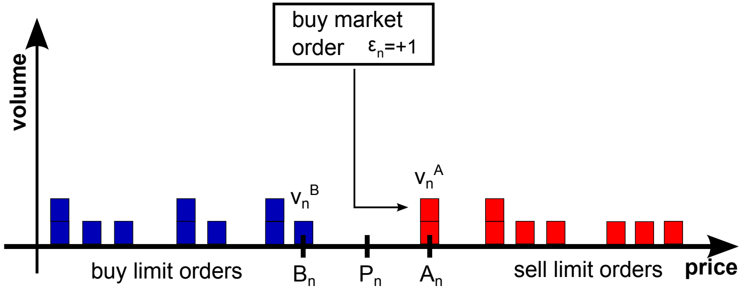
Varenna, July 20-21, 2018

- The mechanism of price formation stems from the complicated interplay between incoming orders (and cancellations) and price change due to these orders
- Market reaction to trades, termed **market impact**, describes how much price change immediately and in the near future in response to order $\rightarrow$ mechanical vs induced response
- The **order flow** is composed by market orders, limit orders and cancellations and depends on the state of the book as well on the past price history

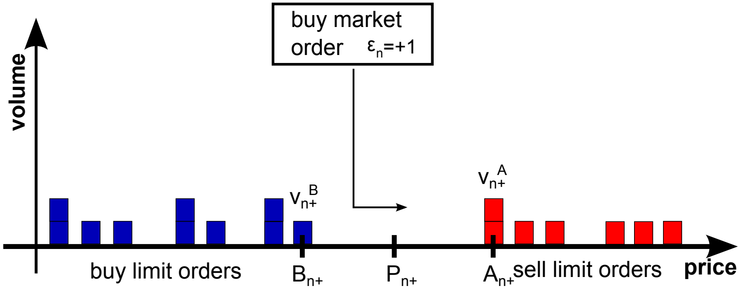
Dynamics of Limit Order Book



Dynamics of Limit Order Book



Dynamics of Limit Order Book



What is market impact?

- Market impact refers to the "correlation" between an incoming order (to buy or to sell) and the subsequent price change.
- Market impact induces extra costs. Indeed, large volumes must typically be fragmented and executed incrementally. The total cost of this large trade is quickly dominated, as sizes become large, by the average price impact
- Monitoring and controlling impact has therefore become one of the most active domains of research in quantitative finance since the mid-nineties.
- Volume dependence of impact (By how much do larger trades impact prices more than smaller trades?), and temporal behavior of impact (is the impact of a trade immediate and permanent, or does the impact decay after one stops trading?).
- Impact is a dynamical quantity since it depends on the available liquidity, but also on the recent history of my trades.

Why is there market impact?

- **Agents successfully forecast short term price movements and trade accordingly.** This does result in measurable correlation between trades and price changes, even if the trades by themselves have absolutely no effect on prices at all. If an agent correctly forecasts price movements and if the price is about to rise, the agent is likely to buy in anticipation of it.'Noise induced' trades, with no information content, have no price impact.
- **The impact of trades reveals some private information.** The arrival of new private information causes trades, which cause other agents to update their valuations, leading to a price change. But if trades are anonymous and there is no easy way to distinguish informed traders from non informed traders, then all trades must impact the price since other agents believe that at least of fraction of these trades contains some private information, but cannot decide which ones.
- **Impact is a purely statistical effect.** Imagine for example a completely random order flow process, that leads to a certain order book dynamics (see, e.g. "zero-intelligence" models). Conditional to an extra buy order, the price will on average move up if everything else is kept constant. Fluctuations in supply and demand may be completely random, unrelated to information, but a well defined notion of price impact still emerges. In this case impact is a completely mechanical – or better, statistical – phenomenon.

There are different types of price impact (often confused even in the specialized literature)

- Impact of an *individual market order* of size v (or more generally of a limit order or even of a cancellation)
- The correlation of the average price change in a given time interval T with the *total* market order imbalance in the same interval (i.e. the sum of the signed volume $\pm v$ of *all individual trades*.)
- Cross impact, i.e. how do trades on asset A impact the price of asset B.
- The impact of a given order of size Q, executed with many trades in a given direction, originating from the same agent.

Terminology

Large orders executed incrementally have different names in the literature

- Large trades
- Large orders
- Hidden orders
- Packages
- Algorithmic executions
- Metaorders (Bouchaud et al.)
- ......

A benchmark model: linear and permanent impact (Kyle)

The average price variation due to a signed volume ϵv is given by:

$$\Delta p = \lambda \epsilon v, \tag{19}$$

where λ is the inverse of liquidity and ϵ is $+1$ (-1) for buyer (seller) initiated trade.
It is direct to show that

- The impact of individual trades is linear in volume and permanent i.e.

$$R_{so}(T) = E\left[(p_T - p_0) \cdot \epsilon_0\right] = \lambda\, E[v], \tag{20}$$

- The impact of aggregated order flow is linear in the volume imbalance

$$p_T = p_0 + \lambda \sum_{n=0}^{N-1} \epsilon_n v_n + \sum_{n=0}^{N-1} \eta_n, \tag{21}$$

- The price impact of a metaorder of total volume Q is linear

$$R_{mo}(T|Q) = E[(p_T - p_0) \cdot \epsilon_{mo} | \sum_{n \in mo} v_n = Q] = \lambda Q, \tag{22}$$

- The time correlation properties of returns are "inherited" by those of order flow $\rightarrow$ market efficiency

Empirical facts: individual impact

Empirical data consistently shows a *sublinear* (concave) volume dependence of impact of individual orders

$$E[\Delta p|v] \equiv R_{so}(T=1|v) \propto v^{\psi}; \qquad \psi \in [0.1, 0.3], \qquad (23)$$

or even a logarithmic dependence $R_{so}(T=1|v) \propto \ln v$.

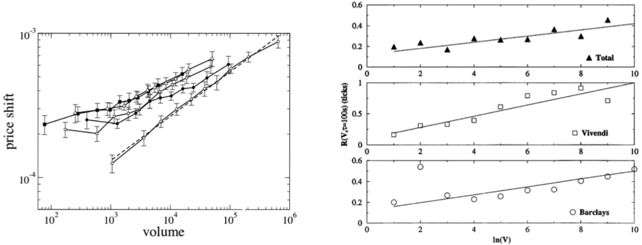


Figure: Impact of individual market orders for London Stock Exchange (left, from Lillo and Farmer 2004) and Paris Bourse (right, from Bouchaud and Potters 2002)

By considering stocks with different market capitalization C on the immediate impact, one can show that impact of individual transactions can be approximately rescaled

$$R_{so}(T=1|v) \approx C^{-0.3} F\left(C^{0.3}\frac{v}{\overline{v}}\right), \qquad (24)$$

where $\overline{v}$ is the average volume per trade for a given stock, and $F(u)$ a master function that behaves as a u^{ψ} for small arguments.

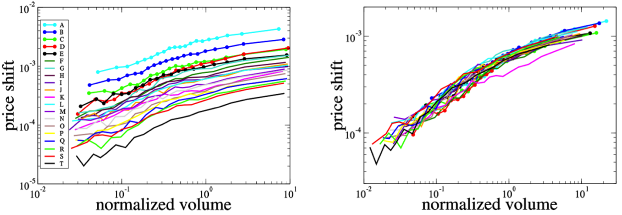
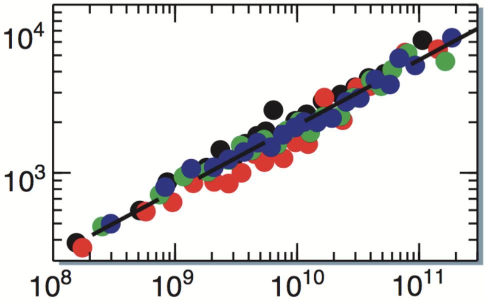


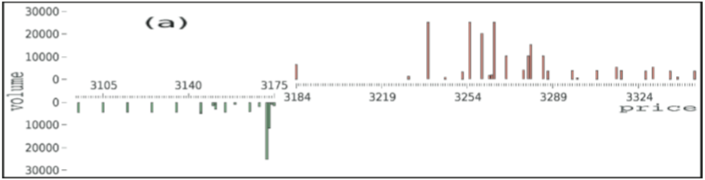
Figure: Impact of individual transactions of groups of stocks with different capitalization (left) and the same curves after rescaling (right) (Lillo et al. 2003).

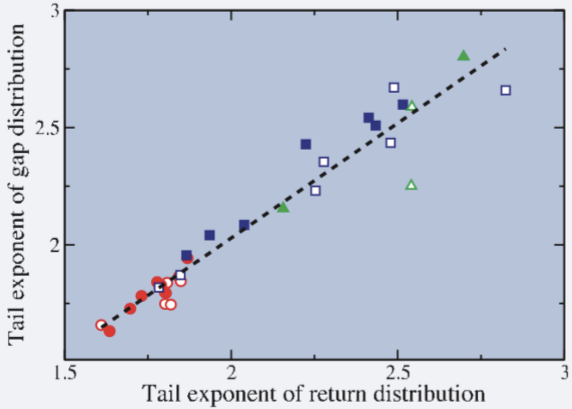


Different colors are different years (Lillo et al. 2003).

Kyle lambda (illiquidity) scales as $\lambda \sim C^{-0.4}$ (note that the dependence of the average volume on market cap has been already considered).

The role of liquidity fluctuations

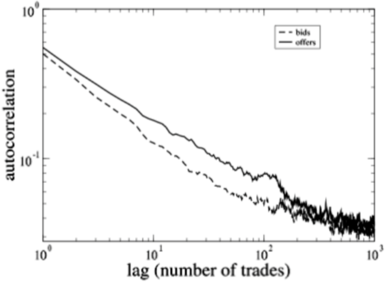




From Farmer et al 2004. Red circles for low liquidity stocks, blue squares for medium liquidity stocks, and green triangles for high liquidity stocks. Empty symbols refer to sell market orders and filled symbols to buy market orders.

Liquidity fluctuations are key determinants of fat tails of return distribution

Persistence of illiquidity (gaps)



From Lillo and Farmer 2005. Autocorrelation function of the first gap size for bids and offers of AstraZeneca in a log-log plot.

- Size of gaps, measuring illiquidity, are very autocorrelated, consistent with long memory processes.
- Similar result for spread.
- The state of the order book is extremely persistent

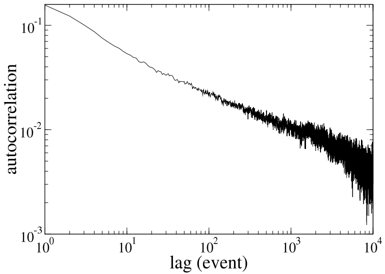
Order flow

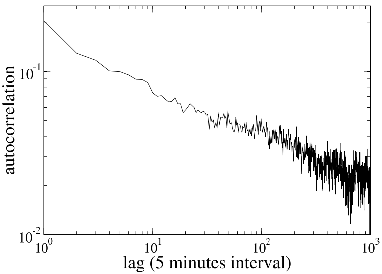
Order flow

- We focus here on orders that trigger transactions, i.e. *market orders*
- A buy market order moves the price up and a sell market order moves the price down (on average)
- The flow of market orders reflects the supply and demand of shares
- A market order is characterized by a volume v and a sign $\epsilon = +1$ for buy orders and $\epsilon = -1$ for sell orders.
- We consider the time series in market order time, i.e. time advances of one unit when a new market order arrives.
- The unconditional sample autocorrelation function of signs is

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

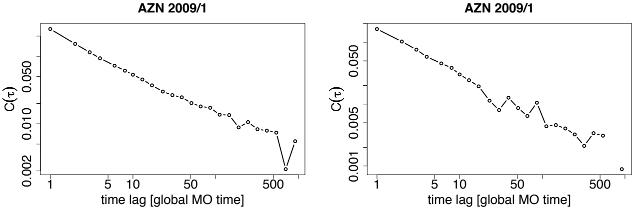
where N is the length of the time series.





From Lillo and Farmer (2004)
Similar plots observed in many different markets, different periods, different asset classes.

It has been shown (Bouchaud *et al.*, 2004, Lillo and Farmer, 2004) that the time series of market order signs is a long memory process.



$C(\tau)$ of market order signs ϵ (left) and signed volumes ϵv (right).
The autocorrelation function decays asymptotically as

$$C(\tau) \sim \tau^{-\gamma} = \tau^{2H-2}$$

where H is the Hurst exponent. For the investigated stocks $H \simeq 0.75$ (i.e. $\gamma \simeq 0.5$).

- Let $\gamma(k)$ be the autocovariance function of a time series X_t. A process is long memory if in the limit $k \to \infty$ it is

$$\gamma(k) \sim k^{-\gamma} L(k) \qquad \gamma \in (0,1) \tag{29}$$

where $L(k)$ is a slowly varying function.

- The Hurst exponent is $H = 1 - \gamma/2$
- Equivalently the spectral density diverges for low frequencies $\omega \to 0$ as

$$g(\omega) \simeq \omega^{1-2H} L(\omega) \tag{30}$$

- The integrated process is superdiffusive $Var(\sum_{s=0}^{t} X_s) \sim t^{2H}$
- Examples: fractional ARIMA (fARIMA)

$$(1-L)^d X_t = \epsilon_t \qquad d = H - 1/2 \tag{31}$$

fractional Brownian motion (in continuous time)

- Frequently observed in finance: volatility, volume, spread,....

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

Two explanations have 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).

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

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

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

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

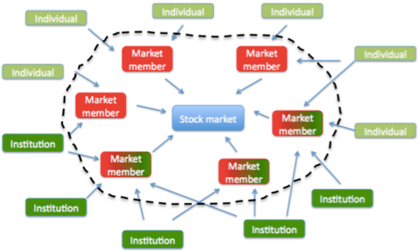
$$C_{herd}(\tau) = \sum_{i \neq j} \left(P^{ij}(\tau) \left[\frac{1}{N^{ij}(\tau)} \sum_t \epsilon_t^i \epsilon_{t+\tau}^j \right] - P^i P^j \left[\frac{1}{N^i} \sum_t \epsilon_t^i \right] \left[\frac{1}{N^j} \sum_t \epsilon_t^j \right] \right)$$

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

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

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	9655	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

- 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)



- 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)

The activity of market members (independently from their trading direction) is characterized by the persistence

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

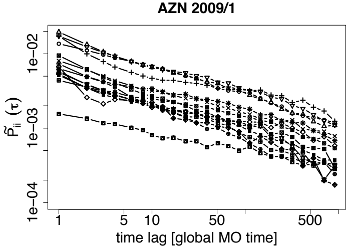


Figure: The diagonal terms of persistence in activity, i.e., $P^{ii}(\tau) - [P^i]^2$ of MO placement for the 15 most active participant codes, the first half of 2009 for AZN.

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.

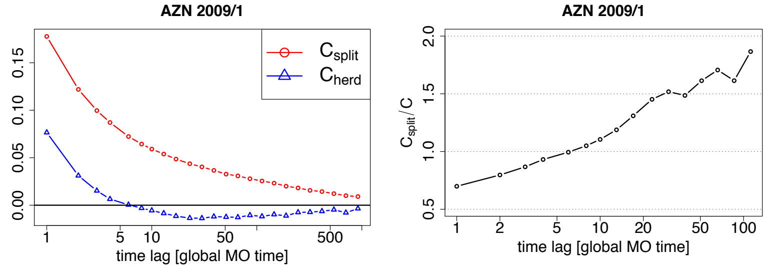


Figure: Left panel. The splitting and the herding term of the correlation of MO signs (the two terms sum up to $C(\tau)$) for the first half year of 2009 for AZN. Right panel. The splitting ratio of MO signs (defined as the ratio of the splitting term in the correlations and the entire correlation) for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

From Toth et al. 2015.

Splitting dominates herding at the broker level (especially for large lags)

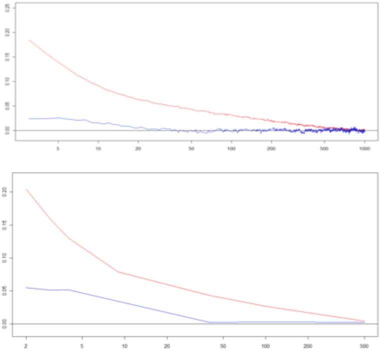


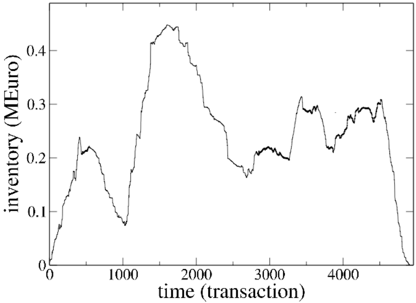
Figure: Splitting and herding component for brokers (top) and agents/accounts (bottom) of a European stock.

Direct evidence for order splitting

- We have seen that correlated order flow is mostly due to order splitting.
- We want to find direct evidence of splitting, characterize the large trades and the splitting characteristics, and to measure the market impact of these large orders.
- The difficulty is, of course, data.
- Some studies use proprietary data of a large financial institution
- We follow a different approach: **statistical identification of large trades from market member data.**

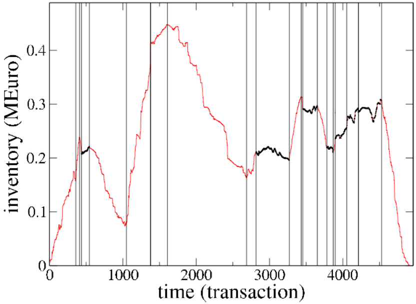
An example: inventory time series

Credit Agricole trading Santander



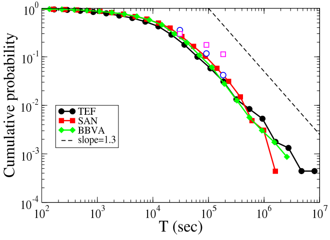
- Clear trends are visible
- The identification of large trades (*metaorders*) must be statistical: a typical regime switching problem

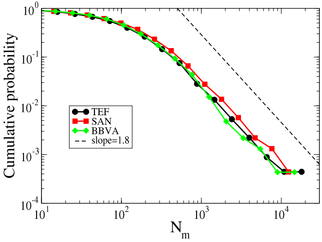
Credit Agricole trading Santander



Different algorithms:

- 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 R. N. Mantegna, New Journal of Physics, **12** 075031 (2010)).





- Metaorder size is asymptotically power law distributed
- Different "size" measures (number of trades, time, total volume) roughly agree on the tail exponent.
- Which model can generate this power law distribution?

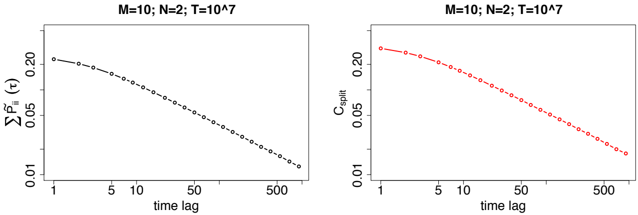
- M funds that want to trade one metaorder each of a size L_i $(i = 1, .., M)$ taken from a distribution p_L, $(L \in \mathbb{N}^+)$.
- The sign of the each metaorder is taken randomly and at each time step one fund is picked randomly with uniform probability.
- The selected fund initiates a trade of the sign of its metaorder, and the size of the metaorder is reduced by one unit.
- When the metaorder is completely traded, a new one is drawn from p_L and assigned a random sign.
- The distribution p_L of metaorder size with the autocorrelation function of trade signs. In particular if the distribution is Pareto

$$p_L = \frac{1}{\zeta(\alpha)} \frac{1}{L^{1+\alpha}}$$

where $\zeta(\alpha)$ is the Riemann zeta function, then the autocorrelation function of trade signs decays asymptotically as

$$\rho_s(\ell) = E[\epsilon_n \epsilon_{n+\ell}] \sim \frac{M^{\alpha-2}}{\ell^{\alpha-1}}$$

- The model connects the exponent of the autocorrelation function of order signs with the tail exponent of metaorder distribution, since $\gamma = 1 - \alpha$.
- There is a growing empirical evidence that the distribution of metaorder size is asymptotically Pareto distributed with a tail exponent close to $\alpha = 1.5$ (Lillo et al. 2005, Gabaix et al. 2006, Vaglica et al. 2008, Bershova et al. 2013).
- Hence the model predicts that $\gamma = \alpha - 1$, i.e. $\gamma \simeq 0.5$, as observed empirically.
- Numerical simulations



Indirect validation of the splitting model

In many markets 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 metaorders

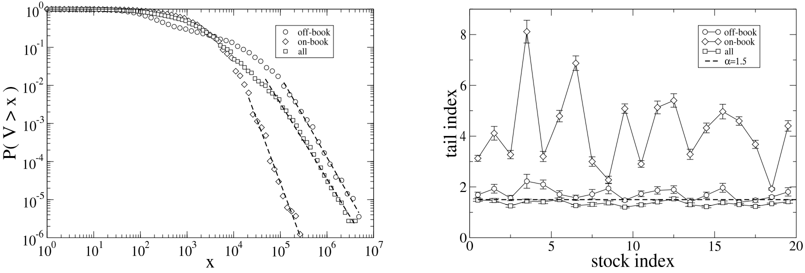
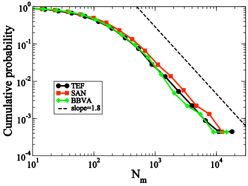


Figure: From Lillo et al 2005. Left. Volume distributions of off-book trades (circles), on-book trades (diamonds), and the aggregate of both (squares). The dashed black lines have the slope found by the Hill estimator and are shown for the largest one percent of the data. Right. Hill estimator of the tail exponent.

The fitted exponent $\alpha \simeq 1.5$ for the metaorder size and the market order sign autocorrelation exponent γ are consistent with the order splitting model ($\gamma = \alpha - 1$).

Direct validation of the splitting model



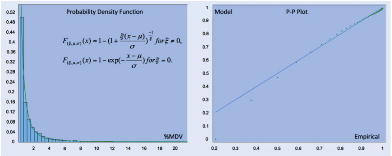


Figure 2. (a) Log–log plot of number of orders vs. order size expressed as %MDV. The table below shows statistics for linear fit model implying that the probability of order size S decays as a power function of S. (b) Distribution of order sizes (as %MDV): Fit to Generalized Pareto with shape parameter $\xi = 0.64$, suggesting $\beta \sim 1/\xi = 1.56$.

- Metaorder size is asymptotically power law distributed (left from Vaglica et al 2008)
- The tail exponent is consistent with the splitting model
- Recently Bershova and Rakhlin (2013) found a tail exponent of 1.56 by investigating metaorders of clients of AllianceBernstein (right)

There is a growing evidence of a power law tailed distribution of time scales of agents

- Distribution of duration of metaorders has a tail exponent of ~ 1.5 (Vaglica et al 2008, Bershova and Rakhlin 2013)
- A multiscale GARCH introduced by Borland and Bouchaud (2005) where agents use stop loss on a given time horizon is consistent with data if the distribution of time scale is power law with tail exponent ~ 1.2
- A simple optimization argument for limit order execution shows that fat tail in limit order prices is consistent with a power law tailed time horizon distribution with exponent ~ 1.5 (Lillo 2007)
- A censored data analysis of the time to fill of limit orders can be used to obtain the distribution of intended lifetime of limit orders. Empirical data are consistent with a power law distribution with tail exponent ~ 1.6 (Eisler et al 2009).

DAR(p) model: a generalization of autoregressive models for discrete valued variates

$$X_n = V_n X_{n-A_n} + (1 - V_n) Z_n,$$

$$Z_n \sim \Xi, \qquad V_n \sim \mathcal{B}(1, \chi), \qquad P(A_n = i) = \phi_i, \quad \sum_{i=1}^{p} \phi_i = 1$$

- Autocorrelation function $\rho_k = Corr(X_n, X_{n+k})$ satisfies:

$$\rho_k = \chi \sum_{i=1}^{p} \phi_i \rho_{k-i}, \qquad k \geqslant 1$$

- Model predictor conditional on $\Omega_{n-1} = \{X_{n-1}, \ldots, X_{n-p}\}$:

$$\hat{X}_{n+s} \equiv \mathbb{E}[X_{n+s}|\Omega_{n-1}] = \chi \sum_{i=1}^{p} \phi_i Y_{n+s-i} + \mathbb{E}[Z](1-\chi), \quad Y_{n+s-i} = \begin{cases} \hat{X}_{n+s-i} & \text{for } i \leqslant s \\ X_{n+s-i} & \text{for } i > s \end{cases}$$

Predictability of order flow has significantly increased from 2004 to 2009

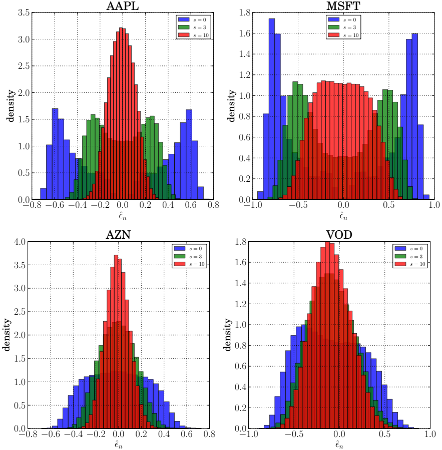


Figure: From Taranto et al. 2014. Distributions of the sign predictor for the stocks AAPL, MSFT, AZN, VOD and $s = 0, 3, 10$ with a DAR(p) model with $p = 500$.

- The exact predictor discriminates orders due to the active metaorder to those due to the noisy background.
- If n trades of the current metaorder has been already traded, the probability that the metaorder continues is (Farmer et al. 2013)

$$\mathcal{P}_n = \frac{\sum_{i=n+1}^{\infty} p_i}{\sum_{i=n}^{\infty} p_i}$$

- For example, if the metaorder size distribution is Pareto

$$\mathcal{P}_n = \frac{\zeta(1+\alpha, 1+n)}{\zeta(1+\alpha, n)} \simeq \left(\frac{n}{n+1}\right)^{\alpha} \sim 1 - \frac{\alpha}{n}$$

- Let us suppose that the active metaorder is a buy and the participation rate is π. The probability that the next order is a buy is

$$p_n^+ = \frac{1-\pi}{2} + \pi\left(\mathcal{P}_n + \frac{1-\mathcal{P}_n}{2}\right) = \frac{1+\pi\mathcal{P}_n}{2}$$

- If s_n indicates the sign of the active metaorder at time n

$$p_n^+ = \frac{1+s_n\pi\mathcal{P}_n}{2}$$

Since $\hat{\epsilon}_n = 2p_n^+ - 1$, it is

$$\hat{\epsilon}_n = s_n\pi\mathcal{P}_n \sim s_n\pi(1-n^{-1})$$

Market impact models

- r_n is the midquote price change between just before the nth trade and just before the $n+1$th trade.
- Immediate impact, $E[r_n|\epsilon_n v_n]$, is non zero and can be written as $E[r|\epsilon v] = \epsilon f(v)$, where f is a function that grows with v
- Impact of a transaction is permanent, like in usual random walks, and the equation for the midquote price m_n at time n is

$$r_n = m_{n+1} - m_n = \epsilon_n f(v_n; \Omega_n) + \eta_n, \tag{32}$$

  where η_n is an additional random term describing price changes not directly attributed to trading itself (e.g. news). We assume that η_n is independent on the order flow and we set $E[\eta] = 0$ and $E[\eta^2] = \Sigma^2$.
- We have included a possible dependence of the impact on the instantaneous state Ω_n of the order book. We expect such a dependence on general grounds: a market order of volume v_n, hitting a large queue of limit orders, will in general impact the price very little. On the other hand, one expects a very strong correlation between the state of the book Ω_n and the size of the incoming market order: large limit order volumes attract larger market orders.

A fixed permanent impact model

- The above equation can be written as:

$$m_n = \sum_{k<n} \epsilon_k f(v_k; \Omega_k) + \sum_{k<n} \eta_k, \tag{33}$$

  which makes explicit the non-decaying nature of the impact in this model: $\epsilon_k \partial m_n / \partial v_k$ (for $k < n$) does not decay as $n - k$ grows.

- The lagged impact function $\mathcal{R}(\ell)$ and the lagged return variance $\mathcal{V}(\ell)$ is

$$\mathcal{R}(\ell) \equiv E[\epsilon_n \cdot (m_{n+\ell} - m_n)] = E[f];\ \mathcal{V}(\ell) \equiv E[(m_{n+\ell} - m_n)^2] = \left(E[f^2] + \Sigma^2\right) \ell, \tag{34}$$

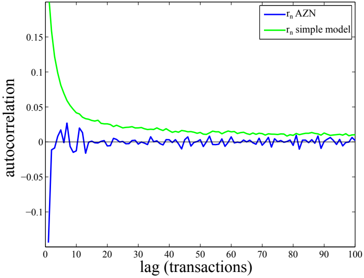
  i.e. constant price impact and pure price diffusion, close to what is indeed observed empirically on small tick, liquid contracts.

- However if we consider the autocovariance of price returns we find that

$$E[r_n r_{n+\tau}] \propto E[\epsilon_n \epsilon_{n+\tau}] \sim \tau^{-\gamma} \tag{35}$$

  which means that price returns are strongly autocorrelated in time. This fact would violate market efficiency because price returns would be easily predictable even with linear methods.

- We therefore come to the conclusion that the empirically observed long memory of order flow is incompatible with the random walk model above if prices are efficient .



A Gerig. *A theory for market impact: How order flow affects stock price*. PhD thesis, University of Illinois, Urbana, Illinois, 2007.

How can the market be statistically efficient (i.e. unpredictable) in the presence of an autocorrelated order flow?

We go back now to the model without spread and consider the consequences of the long memory of order flow.

The Transient Impact Model (or propagator model)

$$m_t = \sum_{t'<t} \left[G(t-t')\epsilon_{t'} + \eta_{t'} \right] + m_{-\infty} \qquad (54)$$

or in differential form, setting $r_t = m_{t+1} - m_t$:

$$r_t = G(1)\epsilon_t + \sum_{t'<t} \mathcal{G}(t-t')\epsilon_{t'} + \eta_t, \qquad \mathcal{G}(\ell) \equiv G(\ell+1) - G(\ell), \qquad (55)$$

where $G(\ell \leq 0) \equiv 0$

Hence past order flow affects future *returns*.

Note that efficiency (i.e. martingale assumption) is not required.

For an arbitrary function $G(\ell)$, the lagged price variance can be computed explicitly and reads:

$$\mathcal{V}(\ell) = \sum_{0 \le j < \ell} G^2(\ell - j) + \sum_{j>0} [G(\ell + j) - G(j)]^2 + 2\Delta(\ell) + \Sigma^2 \ell, \qquad (56)$$

where $\Delta(\ell)$ is the correlation induced contribution:

$$\begin{aligned} \Delta(\ell) &= \sum_{0 \le j < k < \ell} G(\ell - j) G(\ell - k) C(k - j) \\ &+ \sum_{0 < j < k} [G(\ell + j) - G(j)] [G(\ell + k) - G(k)] \, C(k - j) \\ &+ \sum_{0 \le j < \ell} \sum_{k>0} G(\ell - j) [G(\ell + k) - G(k)] \, C(k + j). \end{aligned} \qquad (57)$$

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

$$\Delta(\ell) \approx \Gamma_0^2 c_0 I(\gamma, \beta) \ell^{2-2\beta-\gamma}, \qquad (58)$$

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.
- At the critical value $\beta = \beta_c = (1 - \gamma)/2$, $\Delta(\ell)$ 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'.
- 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 non-summable ($\beta < 1$).

The average impact function $\mathcal{R}(\ell)$ of the model is

$$\mathcal{R}(\ell) = G(\ell) + \sum_{0<j<\ell} G(\ell - j)C(j) + \sum_{j>0} \left[G(\ell + j) - G(j)\right] C(j). \qquad (59)$$

This equation can be used to extract the impact of single trades G from directly measurable quantities, such as $\mathcal{R}(\ell)$ and $C(n)$.

An alternative method of estimation, which is less sensitive to boundary effects, uses the return process of Eq. 55, such that the associated response function $\mathcal{S}(\ell) = \mathbb{E}[r_{t+\ell} \cdot \epsilon_t]$ and $C(\ell)$ are related through:

$$\mathcal{S}(\ell) = \sum_{n \geq 0} \mathcal{G}(n) C(n - \ell),$$

whose solution represents the values of the kernel $\mathcal{G}(\ell)$. The relation between $\mathcal{R}(\ell)$ and $\mathcal{S}(\ell)$ is:

$$\mathcal{R}(\ell) = \sum_{0 \leq i < \ell} \mathcal{S}(i) \qquad (60)$$

allowing to recover the response function from its differential form.

Empirical propagator

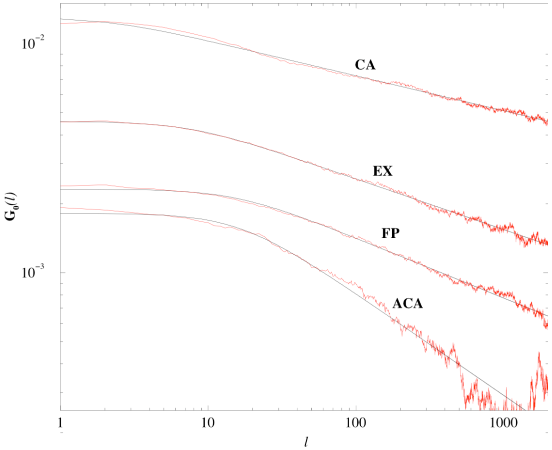


Figure: Comparison between the empirically determined $G(\ell)$, extracted from $\mathcal{R}$ and $\mathcal{C}$ using Eq.(59), and the power-law fit $G^f(\ell) = \Gamma_0/(\ell_0^2 + \ell^2)^{\beta/2}$, for a selection of four stocks: ACA, CA, EX, FP.

- The asymptotic analysis can again be done when $G(\ell)$ decays as $\Gamma_0 \ell^{-\beta}$. When $\beta + \gamma < 1$, one finds:

$$\mathcal{R}(\ell) \approx_{\ell \gg 1} \Gamma_0 c_0 \frac{\Gamma(1-\gamma)}{\Gamma(\beta)\Gamma(2-\beta-\gamma)} \left[\frac{\pi}{\sin \pi\beta} - \frac{\pi}{\sin \pi(1-\beta-\gamma)} \right] \ell^{1-\beta-\gamma}, \tag{61}$$

  Note that numerical prefactor exactly vanishes when $\beta = \beta_c$.
- 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 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 .
- 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.

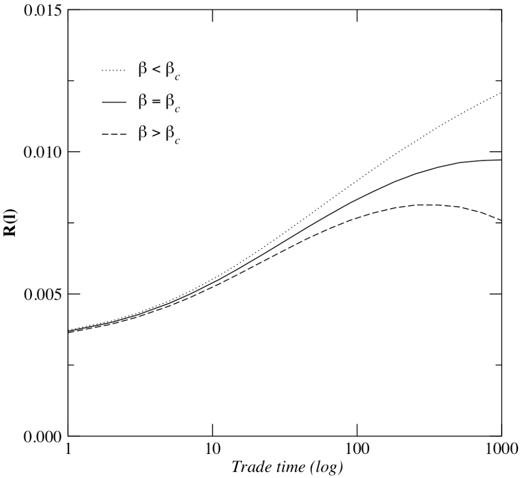


Figure: Theoretical impact function $\mathcal{R}(\ell)$, from Eq. (59), and for values of β close to β_c. When $\beta = \beta_c$, $\mathcal{R}(\ell)$ tends to a constant value as ℓ becomes large. When $\beta < \beta_c$ (slow decay of G), $\mathcal{R}(\ell \to \infty)$ diverges to $+\infty$, whereas for $\beta > \beta_c$, $\mathcal{R}(\ell \to \infty)$ diverges to $-\infty$.

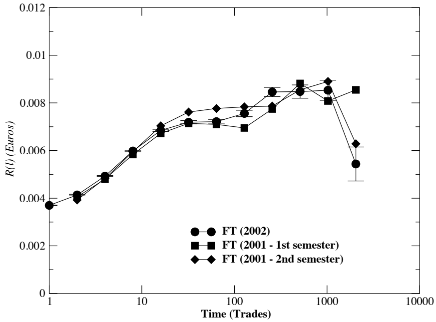


Figure: Average empirical response function $\mathcal{R}(\ell)$ for FT, during three different periods (first and second semester of 2001 and 2002). We have given error bars for the 2002 data. For the 2001 data, the $y-$axis has been rescaled such that $\mathcal{R}(1)$ coincides with the 2002 result. $\mathcal{R}(\ell)$ is seen to increase by a factor ~ 2 between $\ell = 1$ and $\ell = 100$.

TIM assumes that price m_n at transaction time n is

$$m_n = m_{-\infty} + \sum_{k=1}^{\infty} \epsilon_{n-k} f(v_{n-k}) G(k) + \sum_k \eta_k \tag{65}$$

or equivalently

$$m_{n+1} - m_n = G(1)\epsilon_n f(v_n) + \sum_{k=1}^{\infty} [G(k+1) - G(k)] \epsilon_{n-k} f(v_{n-k}) + \eta_n \tag{66}$$

Thus past trades affect future *returns*.

If $C(j) = E[\epsilon_{k+j}\epsilon_k] \simeq j^{-\gamma}$ with $0 < \gamma < 1$, long term diffusivity of prices is recovered only if $G(\ell) \sim \ell^{-\beta}$.

Notice that Eq. 66 suggests to regress price returns on contemporaneous and past order flow to estimate the (increments of the) propagator $G(k)$

An alternative interpretation of the above formalism is to assume that price impact is permanent, but history dependent as to ensure statistical efficiency of prices

- Let us consider a generalized MRR model:

$$r_n = m_{n+1} - m_n = \eta_n + \theta(\epsilon_n - \hat{\epsilon}_n), \qquad \hat{\epsilon}_n = E_n[\epsilon_{n+1}|I] \qquad (67)$$

  where I is the information set available at time n.
- This model implies that $E_{n-1}[r_n|I] = 0$.
- Within the above simplified model, in which we have neglected volume fluctuations, there are only two possible outcomes. Either the sign of the nth transaction matches the sign of the predictor $E_n[\epsilon_{n+1}|I]$, or they are opposite. Let us call r_n^+ and r_n^- the expected ex-post absolute value of the return of the n^{th} transaction given that ϵ_n either matches or does not match the predictor. If we indicate with φ_n^+ and (φ_n^-) the ex ante probability that the sign of the n-th transaction matches (or disagrees) with the predictor ϵ_n,

- We can rewrite $E_{n-1}[r_n|I] = 0$ as:

$$\varphi_n^+ r_n^+ - \varphi_n^- r_n^- = 0. \quad (68)$$

i.e.

$$r_n^+ = \theta(1 - \hat{\epsilon}_n) \quad (69)$$
$$r_n^- = \theta(1 + \hat{\epsilon}_n). \quad (70)$$

- This result shows that the most likely outcome has the smallest impact. We call this mechanism *asymmetric liquidity*: each transaction has a permanent impact, but the impact depends on the past order flow and on its predictability.
- The price dynamics and the impact of orders therefore depend on (i) the order flow process (ii) the information set I available to the liquidity provider, and (iii) the predictor used by the liquidity provider to forecast the order flow.

- Consider the case where the information set available to liquidity providers is restricted to the past order flow. We call this information set *anonymous* because liquidity providers do not know the identity of the liquidity takers and are unable to establish whether or not two different orders come from the same trader.
- We assume also that the predictor used by liquidity takers to forecast future order flow comes from a linear model, namely a K^{th} order autoregressive AR model

$$\hat{\epsilon}_n = \sum_{i=1}^{K} a_i \epsilon_{n-i}, \tag{71}$$

where a_i are real numbers that can be estimated on historical data using standard methods. The MRR model corresponds to an AR(1) order flow, with $a_1 = \rho$ and $a_k = 0$ for $k > 1$, with an exponential decay of the correlation.

- The resulting impact model, Eq. (67) with a general linear forecast of the order flow is in fact *equivalent*, when $K \to \infty$, to the temporary impact model of the previous section. It is easy to show that one can rewrite the generalized MRR model in terms of a propagator as

$$m_n = m_{n-1} + \theta\epsilon_n + \sum_{i=1}^{\infty}[G(i+1) - G(i)]\epsilon_{n-i} + \eta_n, \qquad \theta = G(1). \quad (72)$$

- The equivalence is obtained with the relation:

$$\theta a_i = G(i+1) - G(i) \quad \text{or} \quad G(i) = \theta[1 - \sum_{j=1}^{i-1} a_j]. \quad (73)$$

DAR(p) model: a generalization of autoregressive models for discrete valued variates

$$X_n = V_n X_{n-A_n} + (1 - V_n) Z_n,$$

$$Z_n \sim \Xi, \qquad V_n \sim \mathcal{B}(1, \chi), \qquad P(A_n = i) = \phi_i, \quad \sum_{i=1}^{p} \phi_i = 1$$

- Autocorrelation function $\rho_k = Corr(X_n, X_{n+k})$ satisfies:

$$\rho_k = \chi \sum_{i=1}^{p} \phi_i \rho_{k-i}, \qquad k \geqslant 1$$

- Model predictor conditional on $\Omega_{n-1} = \{X_{n-1}, \ldots, X_{n-p}\}$:

$$\hat{X}_{n+s} \equiv \mathbb{E}[X_{n+s}|\Omega_{n-1}] = \chi \sum_{i=1}^{p} \phi_i Y_{n+s-i} + \mathbb{E}[Z](1-\chi), \quad Y_{n+s-i} = \begin{cases} \hat{X}_{n+s-i} & \text{for } i \leqslant s \\ X_{n+s-i} & \text{for } i > s \end{cases}$$

When it is very likely that the next order is a buy, if a buy occurs the impact is small, while if it is a sell the impact is large.

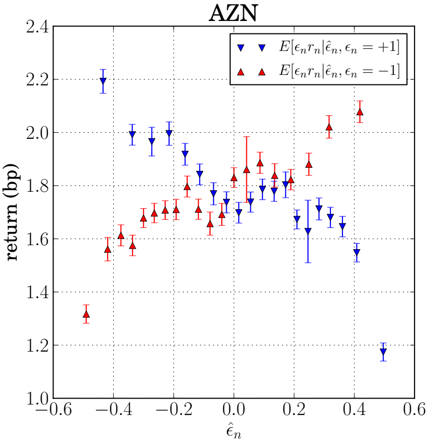
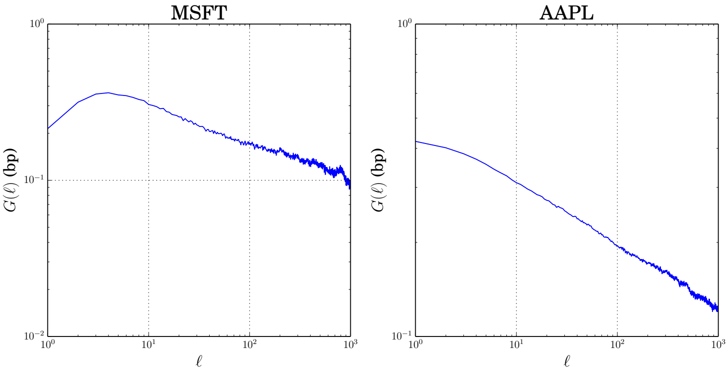
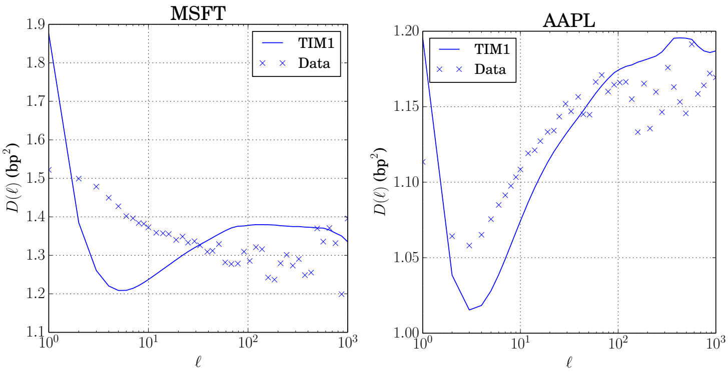


Figure: Expected return behavior as a function of an autoregressive sign predictor $\hat{\epsilon}_n \equiv \mathbb{E}[\epsilon_n|\epsilon_{n-1}, \epsilon_{n-2}, ...]$ for Astrazeneca (from Taranto et al. JSTAT 2014).



- Nasdaq 2013
- Slow decay ($\sim 1,000$ trades), effect of long memory
- For large tick stocks (MSFT) non-monotonic $\rightarrow$ inefficiency $\rightarrow$ dependence of order flow on price movement

Signature plot



- Small tick (AAPL): 'trend-like" behaviour for $\ell \geq 3$ and high frequency activity with the spread leading to a minimum in $D(\ell)$.
- Large tick (MSFT) "mean-reverting" behaviour, with a steadily decreasing signature plot.

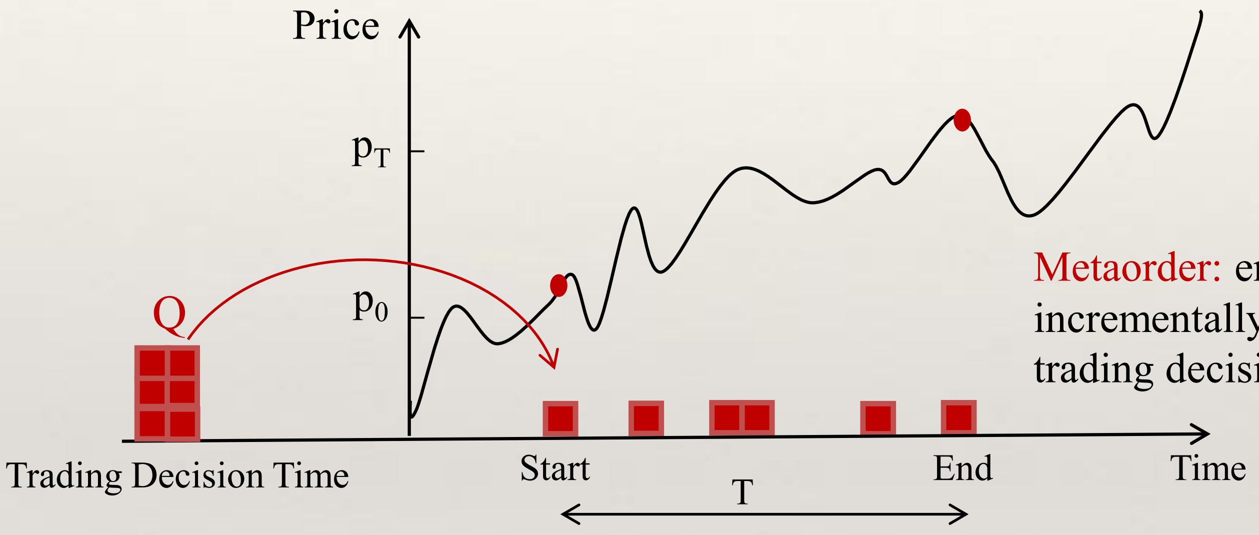
Market impact of metaorders: phenomenology

Market Impact

❖Market Impact (M.I.):

average signed price change conditioned to the sequential execution of a volume Q of contracts (metaorder)

$$I(Q) = \langle \epsilon \cdot \Delta p | Q \rangle = \langle \epsilon \cdot (p_T - p_0) | Q \rangle \qquad \epsilon = \pm 1 \text{ (buy/sell)}$$



Metaorder: ensemble of trades, executed incrementally and that belong to one single trading decision.

❖Why the M.I. is relevant?

- Theory: it is the mechanism through which prices becomes efficients (Supply/Demand).
- Practice (I): for regulators because it influences the market stability.
- Practice (II): it is a cost for traders, which need to accurately control in order to optimize execution.

- Kyle's original model (1985) predicts that price impact should be a linear function of the metaorder size
- Empirical studies have consistently shown that the price impact of a metaorder is a non-linear concave function of its size.
- Market impact $\mathcal{I}$, i.e. the expected average price change between the beginning and the end of a metaorder of size Q is empirically fit by

$$\Delta \ln p \equiv \mathcal{I}(Q) = \pm Y \sigma_D \left(\frac{Q}{V_D} \right)^{\delta} \tag{79}$$

  where σ_D is the daily volatility of the asset, V_D is the daily traded volume, and the sign of the metaorder is positive (negative) for buy (sell) trades. The numerical constant Y is of order unity and the exponent δ is in the range 0.4 to 0.7, but typically very close to 1/2, i.e. to a square root.
- This is the **square-root impact law** (Barra 1997, Almgren et al 2005, Moro et al 2009, Toth et al 2011, Bershova et al 2013)

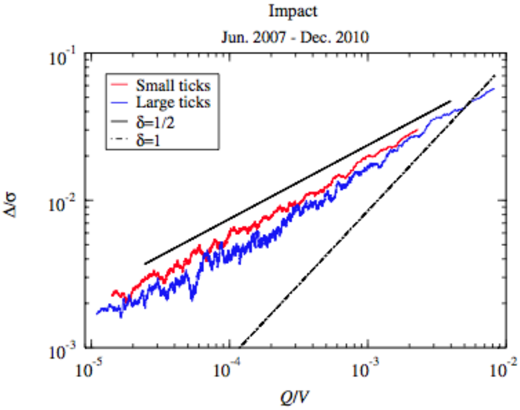


Figure: From Toth et al. 2011. The impact of metaorders for Capital Fund Management proprietary trades on futures markets, Impact is measured here as the average execution shortfall of a metaorder of size Q. The data base contains nearly 500, 000 trades. We show $\mathcal{I}(Q)/\sigma_D$ vs Q/V_D on a log-log scale, where σ and V are the daily volatility and daily volume measured the day the metaorder is executed.

By using brokerage data of LSE and BME, we reconstruct statistically the metaorders and we measure the dynamics of price during the their execution, by rescaling the time in $[0, 1]$.

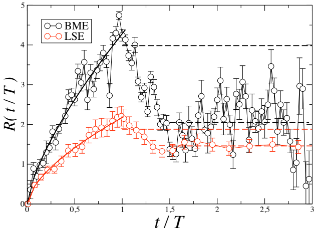


Figure: Market impact versus time. The symbols are the average value of the market impact of the metaorder as a function of the normalized time to completion t/T. The rescaled time $t/T = 0$ corresponds to the starting point of the metaorder, while $t/T = 1$ corresponds to the end of the metaorder.

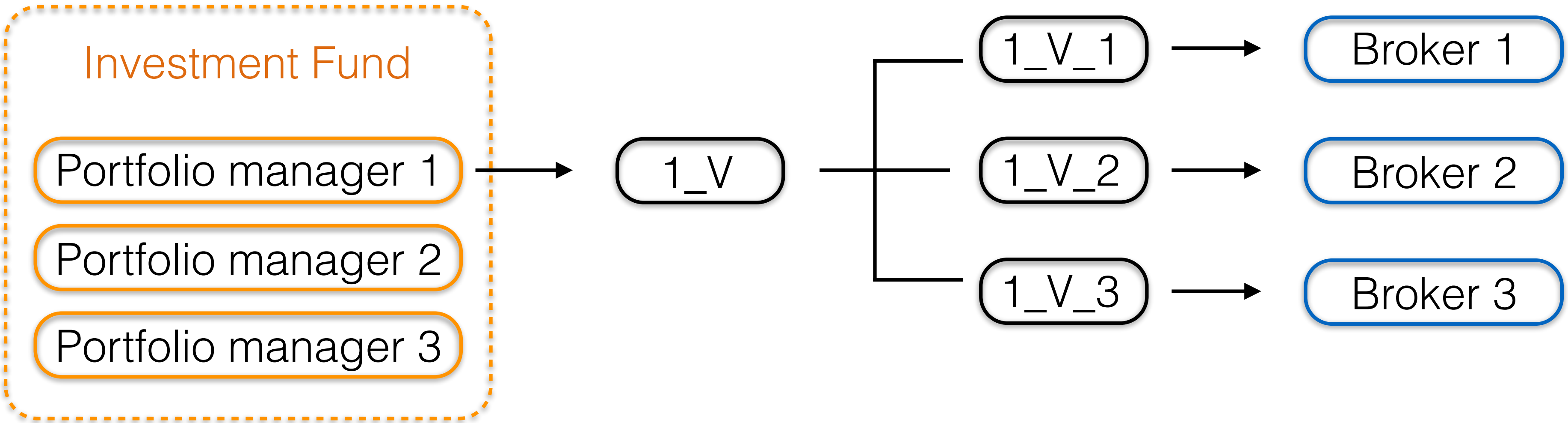
We find approximately the square root law

$$E[r|N] = A\epsilon s N^{\beta} \qquad \beta \simeq 1/2 \tag{80}$$

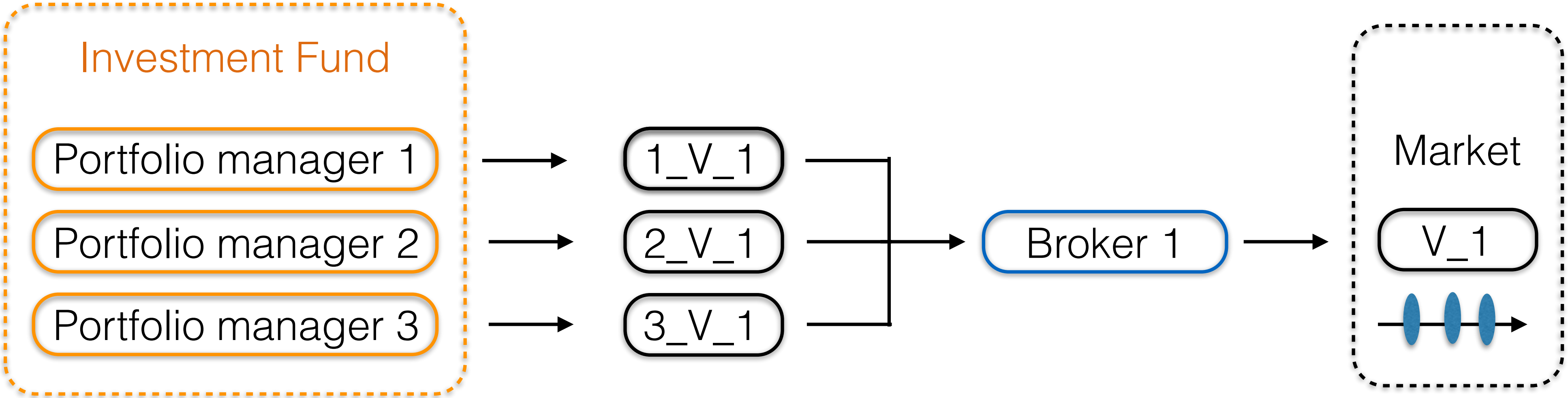
and a decay at approximately 2/3 of the peak impact.

Ancerno Dataset - 1

A portfolio manager liquidates a position and splits its order between brokers.



A broker receives orders from different portfolio managers and bundles them in a unique metaorder.



Ancerno Dataset - 2

- **Metaorder** definition: an execution performed by a single Broker, on a single stock, in a given direction. All metaorders are completed within a trading day.
- The dataset is heterogeneous, containing metaorders traded by many financial institutions for different purposes and it spans several years.
- US Equity in Russel 3000 Index in 2007-2009
- The metaorders account for roughly 5% of ADV for the top 20 stocks
- For each metaorder in the dataset we recover the relative **daily fraction** π, the **participation rate** η, and the **duration** F.
- We work in volume time (intraday patterns)
- We introduce the following filters:

Filter 0	Selecting metaorders traded between January 2007 and December 2009	~ 28,500,000
Filter 1	Selecting metaorders traded on Russell3000 index	~ 23,000,000
Filter 2	Selecting metaorders traded during regular trading section: 09:30 - 16:00	~ 11,000,000
Filter 3	Selecting metaorders with duration longer than 2 minutes	~ 7,500,000
Filter 4	Selecting metaorders whose participation rate is smaller than 0.3	~ 7,000,000

Sign	$\epsilon = \pm 1$
Duration	$F := V_P / V_D$
Participation rate	$\eta := Q / V_P$
Daily rate	$\pi := Q / V_D$
Trading profile	$\rho(v, v_s, v_e)$

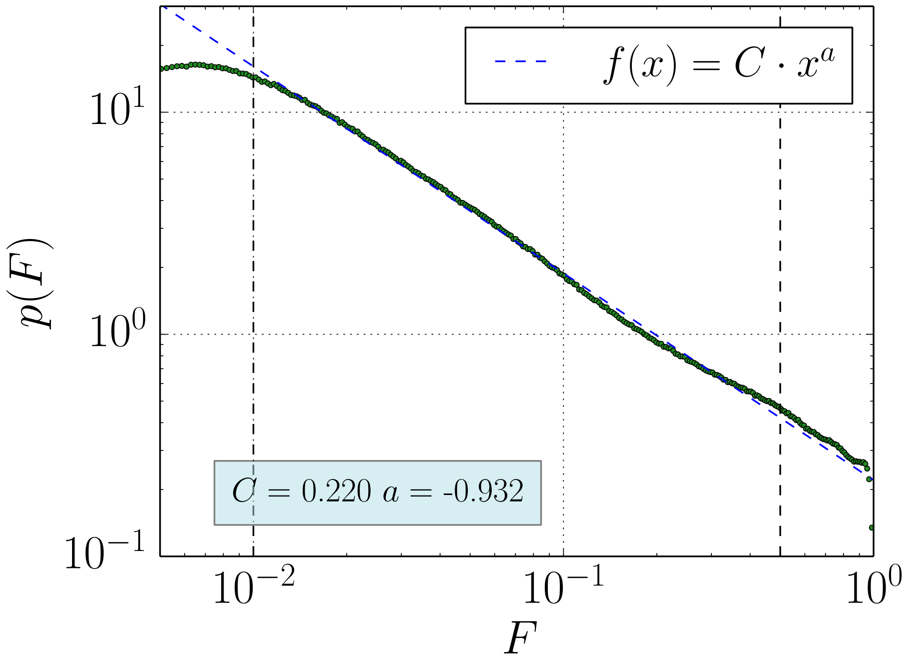
Not available information

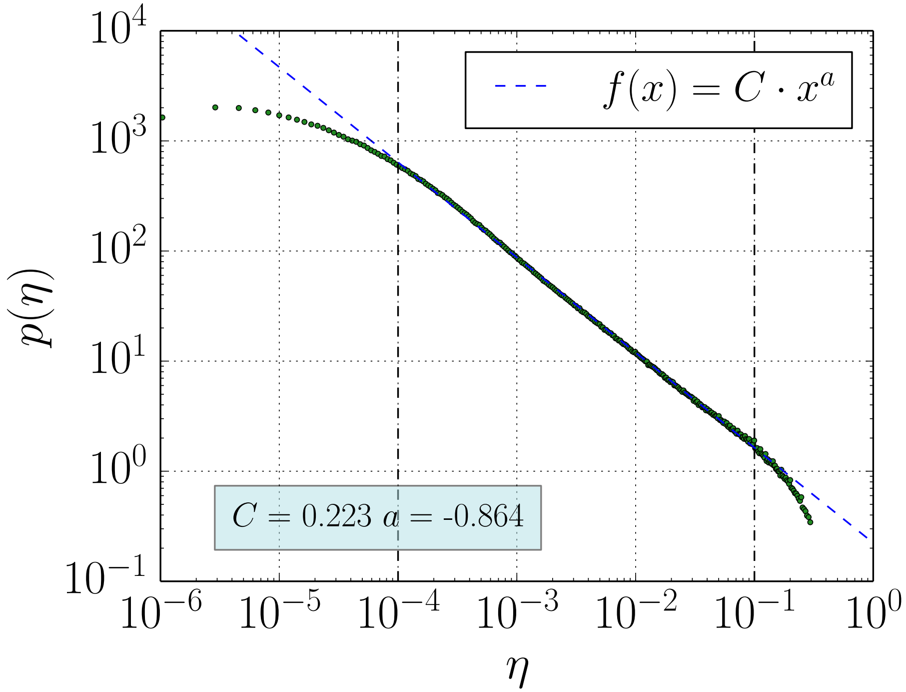
$$\pi = \eta F$$

Distribution of the describing variables

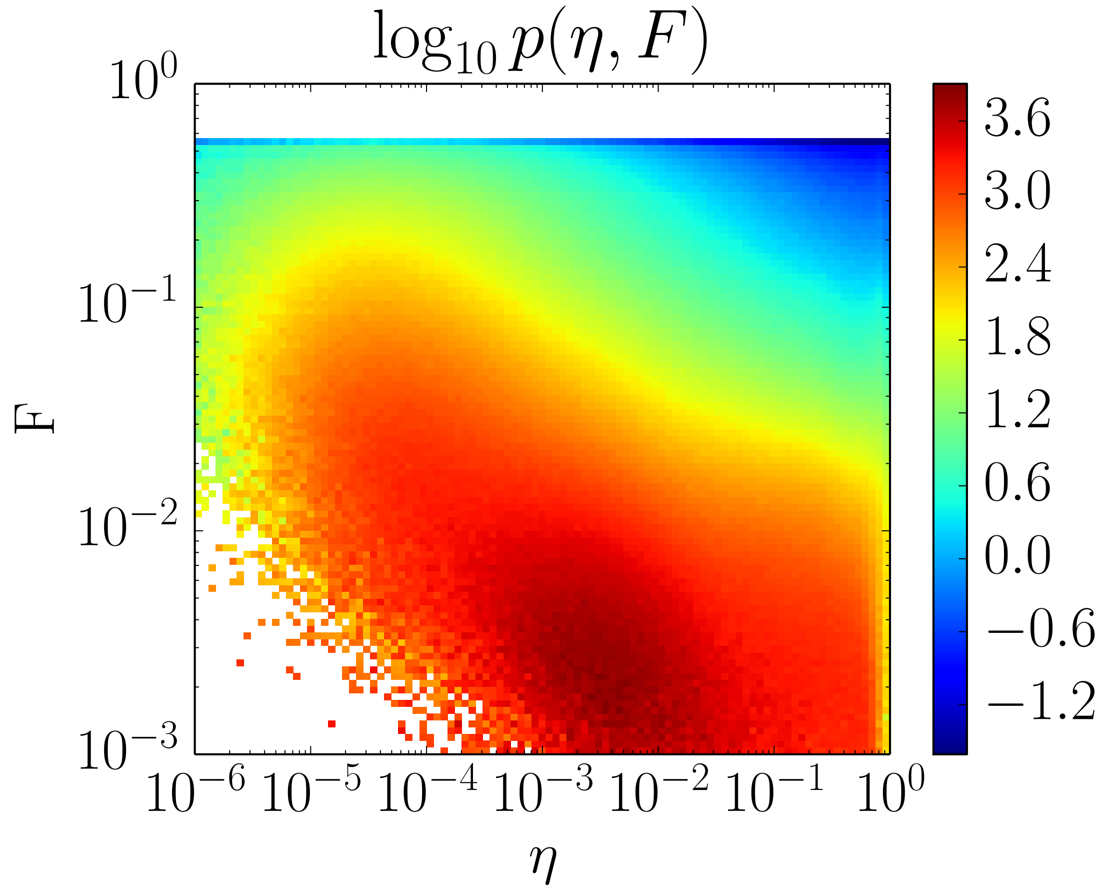
Duration $F := V_P / V_D$

Participation rate $\eta := Q / V_P$





The participation rate η and the duration F are both well approximated by a **truncated power-law distribution** over several orders of magnitude

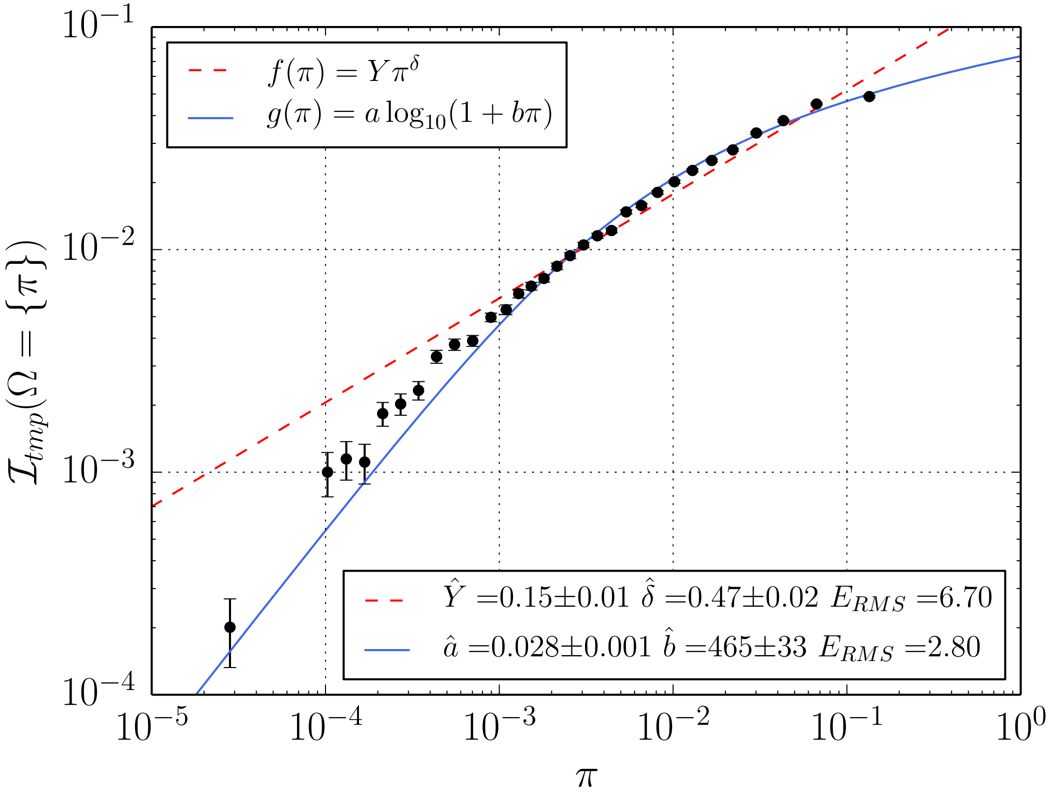


Logarithm of the estimated joint probability density function in double logarithmic scale of the duration F and the participation rate η

The price impact curve: excess concavity

Price impact curve: the average relative price change between the **end** and the **beginning** of the execution, conditioning on the daily rate $\pi := Q/V_D$

$\mathcal{I}(\pi) := \mathbb{E}\left[\epsilon\left(s(v_e) - s(v_s)\right) \middle| \pi\right]$ $\qquad s(v) := \log S(v)/\sigma_D$ Rescaled price



A **square-root** model well describe price impact only in the central region (red curve).

A **logarithmic** (more concave) model allows to capture the **whole shape** of the curve (blue curve).