Market impact models and optimal execution algorithms

Fabrizio Lillo

https://fabriziolillo.wordpress.com

Scuola Normale Superiore, Pisa (Italy)

Imperial College London, June 7-16, 2016

Outline

- June 7: Microstructure of double auction markets
  1. Market impact(s): origin and phenomenology
  2. Impact of single trades
  3. Order flow: phenomenology and models
- June 9: Market impact models.
  1. Transient impact models
  2. History dependent impact models
  3. Order book models
- June 16: Market impact of large trades and optimal execution.
  1. Phenomenology of large trade executions
  2. Models of optimal execution
- References

Market impact models

A fixed permanent impact model

- 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

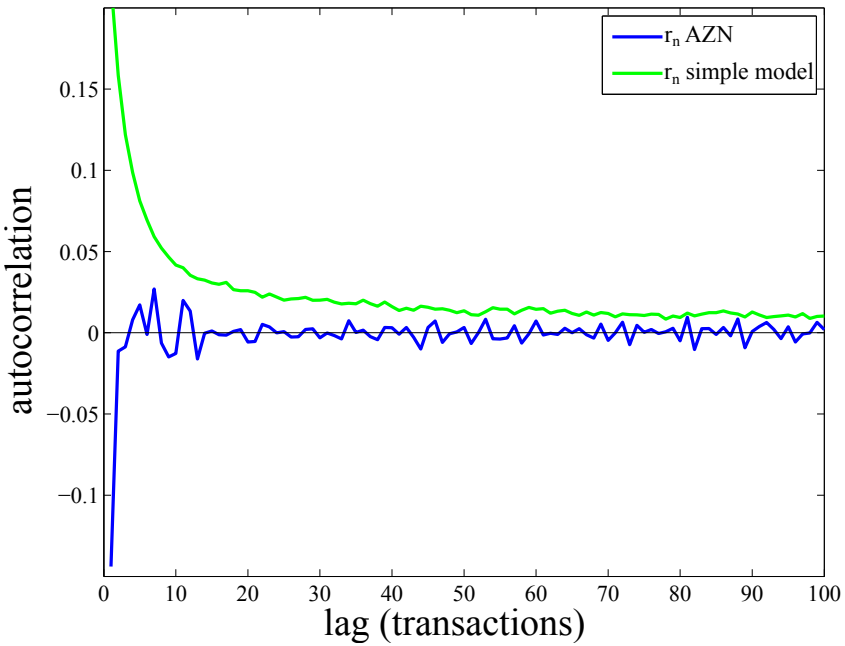
$$\mathcal{R}(\ell) \equiv E[\epsilon_n \cdot (m_{n+\ell} - m_n)] = E[f]; \; \mathcal{V}(\ell) \equiv E[(m_{n+\ell} - m_n)^2] = \left(E[f^2] + \Sigma^2 \right) \ell, \tag{34}$$

  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 .

Visualizing the paradox



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?

Madhavan, Richardson and Roomans (MRR) model (1997)

- Assumptions: (i) all trades have the same volume $v_n = v$ and (ii) the ϵ_n's are generated by a Markov process with correlation ρ, thus $E[\epsilon_n|\epsilon_{n-1}] = \rho\epsilon_{n-1}$
- In this model, correlations decay exponentially fast, i.e. $C(\ell) = E[\epsilon_i \epsilon_{i+\ell}] = \rho^{\ell}$ which does not conform to reality.
- The MRR model postulates that the mid-point m_n evolves only because of unpredictable external shocks (or news) and because of the surprise component in the order flow. This postulate of course automatically removes any predictability in the price returns and ensures efficiency.

$$m_{n+1} - m_n = \theta[\epsilon_n - \rho\epsilon_{n-1}] + \eta_n, \tag{36}$$

where η is the shock component, and the constant θ measures the size of trade impact.

Madhavan, Richardson and Roomans (MRR) model (1997)

One may write:

$$m_{n+\ell} - m_n = \sum_{j=n}^{n+\ell-1} \eta_j + \theta \sum_{j=n}^{n+\ell-1} [\epsilon_j - \rho\epsilon_{j-1}], \tag{37}$$

The full impact function is found to be constant, equal to:

$$\mathcal{R}(\ell) = \theta(1-\rho^2), \quad \forall \ell \tag{38}$$

We can also define the 'bare' impact of a single trade $G(\ell)$, which measures the influence of a single trade at time $n-\ell$ on the the mid-point at time n. In terms of $G(\ell)$, the mid-point is therefore written as:

$$m_n = \sum_{j=-\infty}^{n-1} \eta_j + \sum_{j=-\infty}^{n-1} G(n-j-1)\,\epsilon_j, \tag{39}$$

is here found to given by $G(\ell = 0) = \theta$ and $G(\ell \geq 1) = \theta(1-\rho)$: a part $\theta\rho$ of the impact instantaneously decays to zero after the first trade, whereas the rest of the impact is permanent.
Finally, the volatility, within this simplified version of the MRR model, reads:

$$\mathcal{V}(\ell) = \theta^2(1-\rho^2)\ell. \tag{40}$$

A transient impact model

- The long term memory of trades is *a priori* paradoxical and hints towards a non trivial property of financial markets, which can be called *long-term resilience*.
- Take again Eq. (39) with the assumption that single trade impact is lag independent: $G(\ell) = G$ and that volume fluctuations can still be neglected. The mid-price variance is easily computed to be:

$$\mathcal{V}(\ell) \equiv \langle (m_{n+\ell} - m_n)^2 \rangle = [\Sigma^2 + G^2]\ell + 2G \sum_{j=1}^{\ell} (\ell - j) C(j). \tag{41}$$

- When $\gamma < 1$, the second term of the rhs can be approximated, when $\ell \gg 1$, by $2c_0 G \ell^{2-\gamma}/(1-\gamma)(2-\gamma)$, which grows faster than the first term. In other words, the price would *super-diffuse*, or trend, at long times, with a volatility diverging with the lag ℓ. This of course does not occur: The market reacts to trade correlations so as to prevent the occurrence of such trends.

The MRR model with a bid-ask spread

- The original MRR model assumes that it is the 'true' fundamental price p_n, rather than the midpoint m_n, which is impacted by the surprise in order flow, and hence

$$p_{n+1} - p_n = \eta_n + \theta[\epsilon_n - \rho\epsilon_{n-1}]. \tag{42}$$

- Market makers cannot guess the surprise of the next trade, and post a bid price b_n and an ask price a_n given by:

$$a_n = p_n + \theta[1 - \rho\epsilon_{n-1}] + \phi; \qquad b_n = p_n + \theta[-1 - \rho\epsilon_{n-1}] - \phi, \tag{43}$$

  where ϕ is the extra compensation claimed the market maker, covering processing costs and the shock component risk.

- The midpoint $m \equiv (a + b)/2$ immediately before the nth trade is:

$$m_n = p_n - \theta\rho\epsilon_{n-1}, \tag{44}$$

  whereas the spread is given by $S = a - b = 2(\theta + \phi)$

The MRR model with a bid-ask spread

- Neglecting ϕ for arbitrary correlations between signs:

$$m_{n+\ell} - m_n = \sum_{j=n}^{n+\ell-1} \eta_n + \theta \sum_{j=n}^{n+\ell-1} \{\epsilon_j - E_j[\epsilon_{j+1}]\}\,, \tag{45}$$

- The impact function, in the general case, reads

$$\mathcal{R}(\ell) = \theta\,[1 - C(\ell)]\,. \tag{46}$$

- The long term profit of market makers is zero, because $\mathcal{R}(\infty) = \theta = S/2$. Spread and impact are two sides of the same coin.
- However the long time impact is enhanced compared to the short term impact by a factor

$$\lambda \equiv \frac{\mathcal{R}(\infty)}{\mathcal{R}(1)} = \frac{1}{1 - C(1)} > 1. \tag{47}$$

The MRR model with a bid-ask spread

- When $\phi \neq 0$

$$S = 2(\theta + \phi) = 2(\mathcal{R}(\infty) + \phi) = 2\lambda\mathcal{R}(1) + 2\phi \tag{48}$$

where $\lambda = (1-\rho)^{-1}$.

- The mid-point volatility on scale ℓ

$$\sigma_\ell^2 = \frac{1}{\ell}\langle (m_{\ell+i} - m_i)^2 \rangle. \tag{49}$$

is the sum of a trade induced volatility $\theta^2(1-\rho)^2$ and a 'news' induced volatility Σ^2:

$$\sigma_1^2 = \langle (m_{n+1} - m_n)^2 \rangle = \Sigma^2 + \theta^2(1-\rho)^2 \tag{50}$$

and

$$\sigma_\infty^2 = \Sigma^2 + \theta^2(1-\rho)^2\left(1 + 2\frac{\rho}{1-\rho}\right) = \Sigma^2 + \theta^2(1-\rho^2) \geq \sigma_1^2. \tag{51}$$

- The MRR model leads to two simple relations between spread, impact and volatility per trade

$$S = 2\lambda\mathcal{R}(1) + 2\phi; \qquad \sigma_1^2 = \mathcal{R}(1)^2 + \Sigma^2. \tag{52}$$

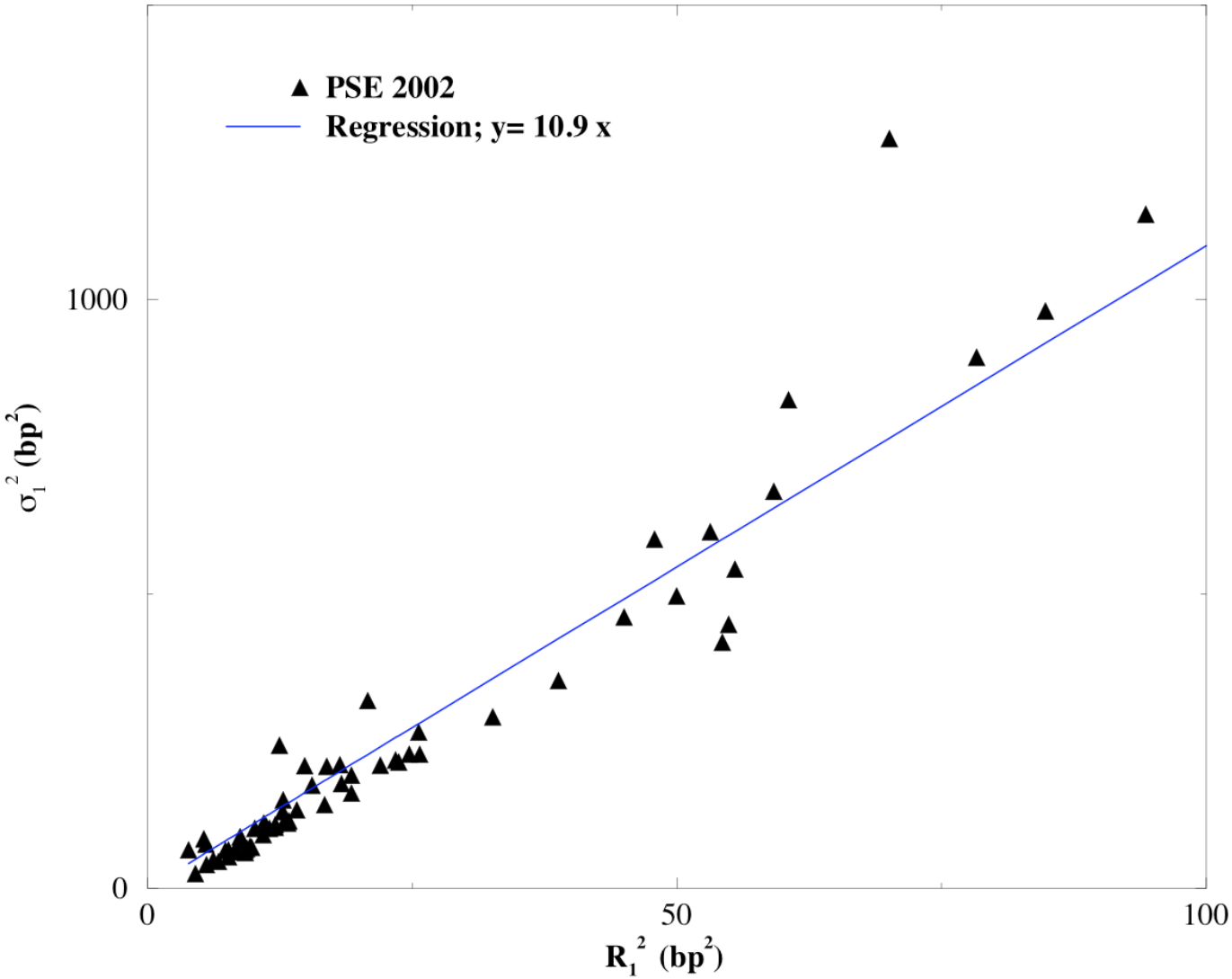


Figure: From Wyarth et al 2008. Plot of σ_1^2 vs. $\overline{\mathcal{R}}_1^2$, showing that the linear relation holds quite precisely with $\Sigma^2 = 0$ and $a \approx 10.9$. (The intercept of the best affine regression is even found to be slightly negative). Data here corresponds to the 68 stocks of the PSE in 2002. The correlation is very high: $R^2 = 0.96$.

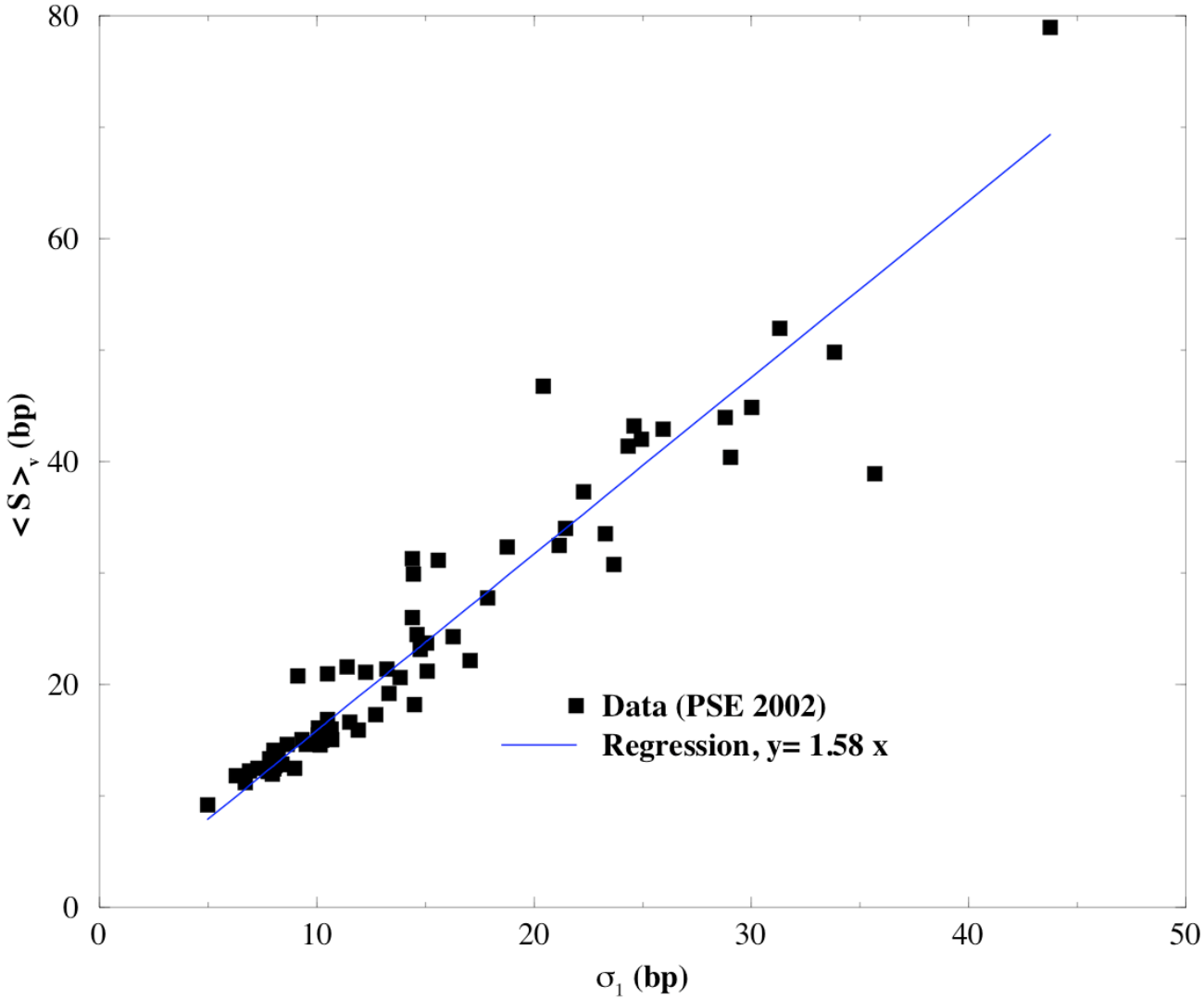


Figure: From Wyarth et al 2008. Relation between spread and volatility per trade for 68 stocks from the Paris Stock Exchange in 2002, averaged over the entire year. The value of the linear regression slope is $c \approx 1.58$, with $R^2 = 0.96$

Comments

- The linear relation between spread and volatility per trade is not expected to hold for the volatility *per unit time* σ, since it involves an extra stock-dependent and time-dependent quantity, namely the the trading frequency f, through:

$$\sigma = \sigma_1 \sqrt{f}. \tag{53}$$

- Note that there are two complementary economic interpretations of the relation $\sigma_1 \sim S$ in small tick markets:
  - (i) Since the typical available liquidity in the order book is quite small, market orders tend to grab a significant fraction of the volume at the best price; furthermore, the size of the 'gap' above the ask or below the bid is observed to be on the same order of magnitude as the bid-ask spread itself which therefore sets a natural scale for price variations. Hence both the impact and the volatility per trade are expected to be of the order of S, as observed.
  - (ii) The relation can also be read backward as $S \sim \sigma_1$: when the volatility per trade is large, the risk of placing limit orders is large and therefore the spread widens until limit orders become favorable.

A transient impact model (TIM)

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)

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

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

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

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

Hence past order flow affects future *returns*.

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

A transient impact model (TIM)

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

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

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

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

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:

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

where $I > 0$ is a certain numerical integral.

A transient impact model (TIM)

- 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 ($\beta < 1$).

Calibration of TIM

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

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

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:

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

allowing to recover the response function from its differential form.

Empirical propagator

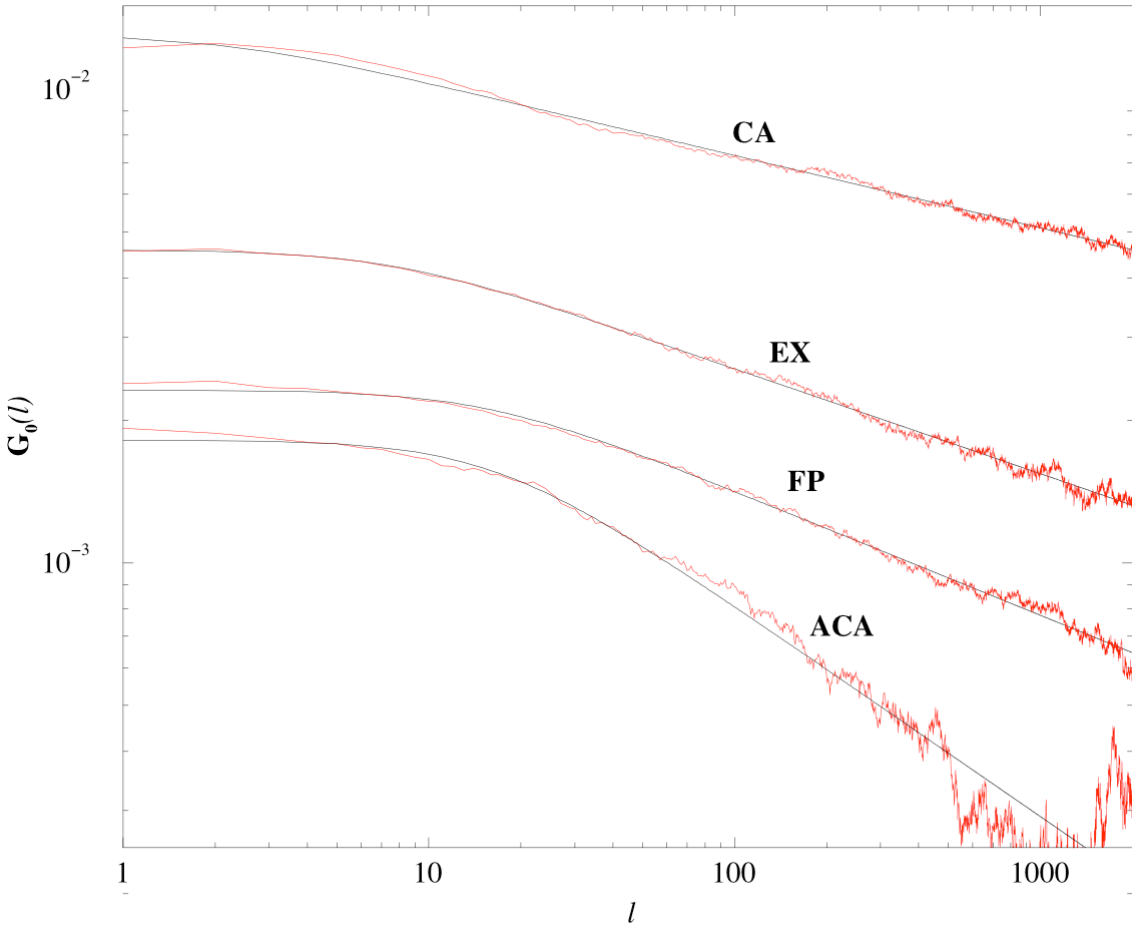


Figure: Comparison between the empirically determined $G(\ell)$, extracted from $\mathcal{R}$ and $\mathcal{C}$ using Eq.(59), and the power-law fit $G^f(\ell) = \Gamma_0/(\ell_0^2 + \ell^2)^{\beta/2}$, for a selection of four stocks: ACA, CA, EX, FP.

Response function of TIM

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

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

  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.

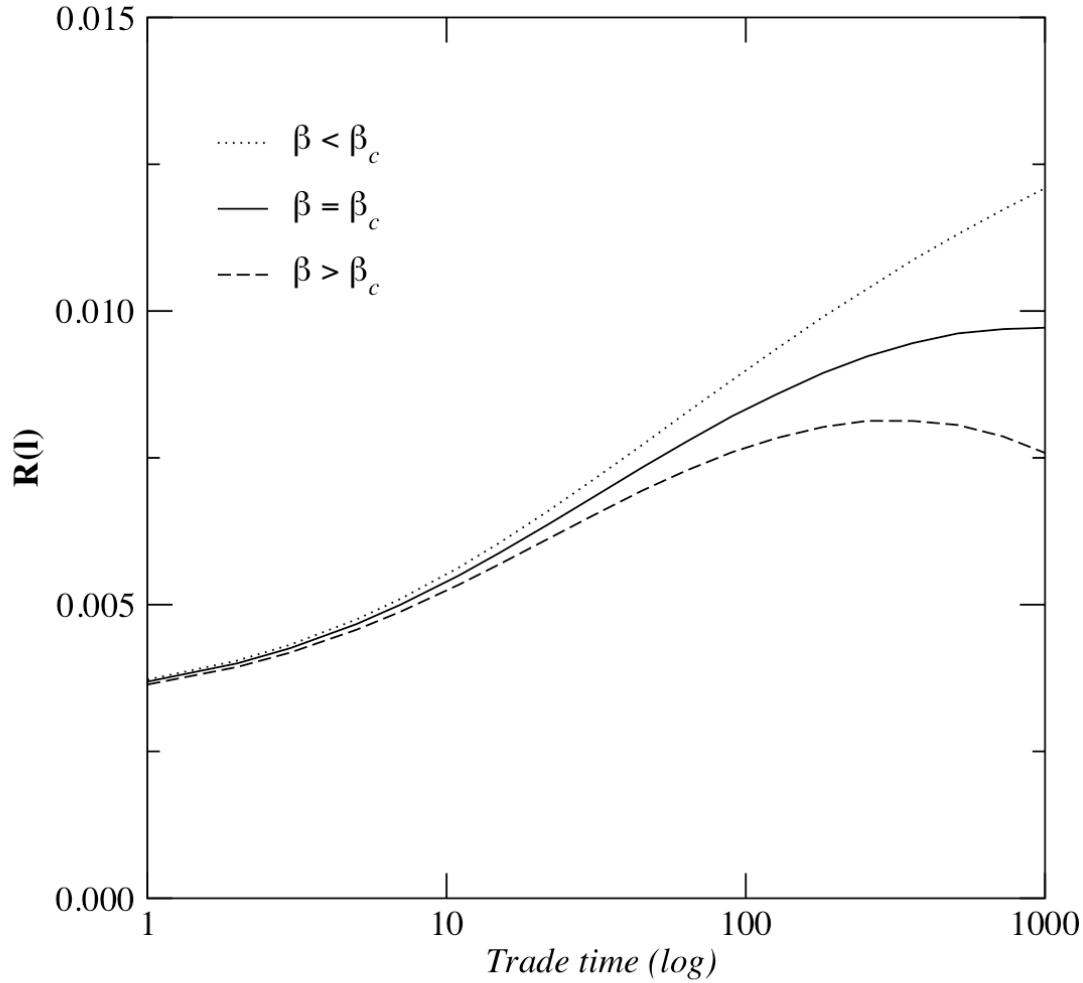


Figure: Theoretical impact function $\mathcal{R}(\ell)$, from Eq. (59), and for values of β close to β_c. When $\beta = \beta_c$, $\mathcal{R}(\ell)$ tends to a constant value as ℓ becomes large. When $\beta < \beta_c$ (slow decay of G), $\mathcal{R}(\ell \to \infty)$ diverges to $+\infty$, whereas for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

Empirical response function

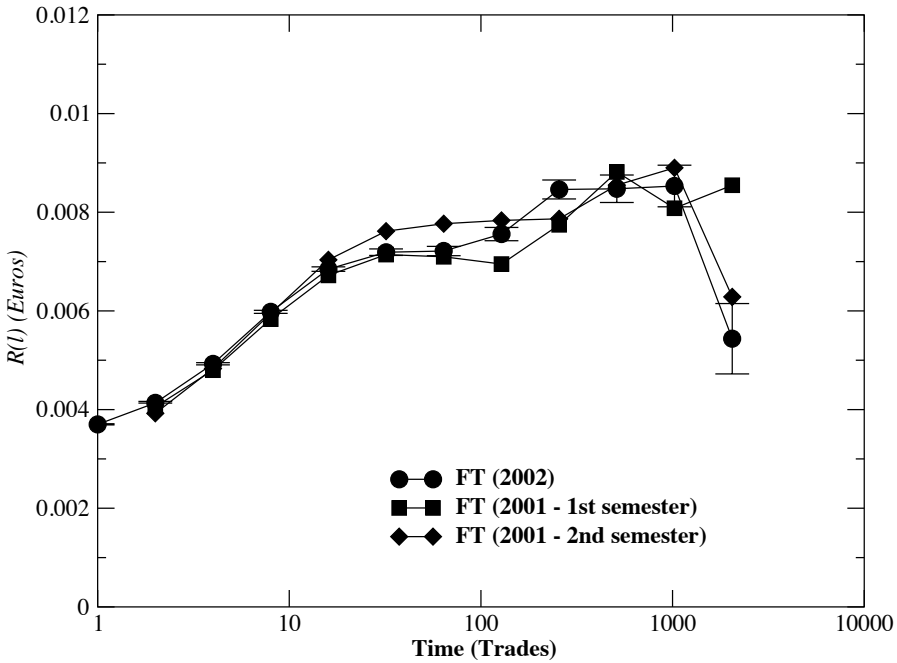


Figure: Average empirical response function $\mathcal{R}(\ell)$ for FT, during three different periods (first and second semester of 2001 and 2002). We have given error bars for the 2002 data. For the 2001 data, the $y-$axis has been rescaled such that $\mathcal{R}(1)$ coincides with the 2002 result.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

Decoupling the contribution of different traders to response function

We use brokerage data from LSE. MO=market order not changing the price MO'= market orders changing the price
The average behavior of the price ℓ time steps after an event of a particular type π_1 is

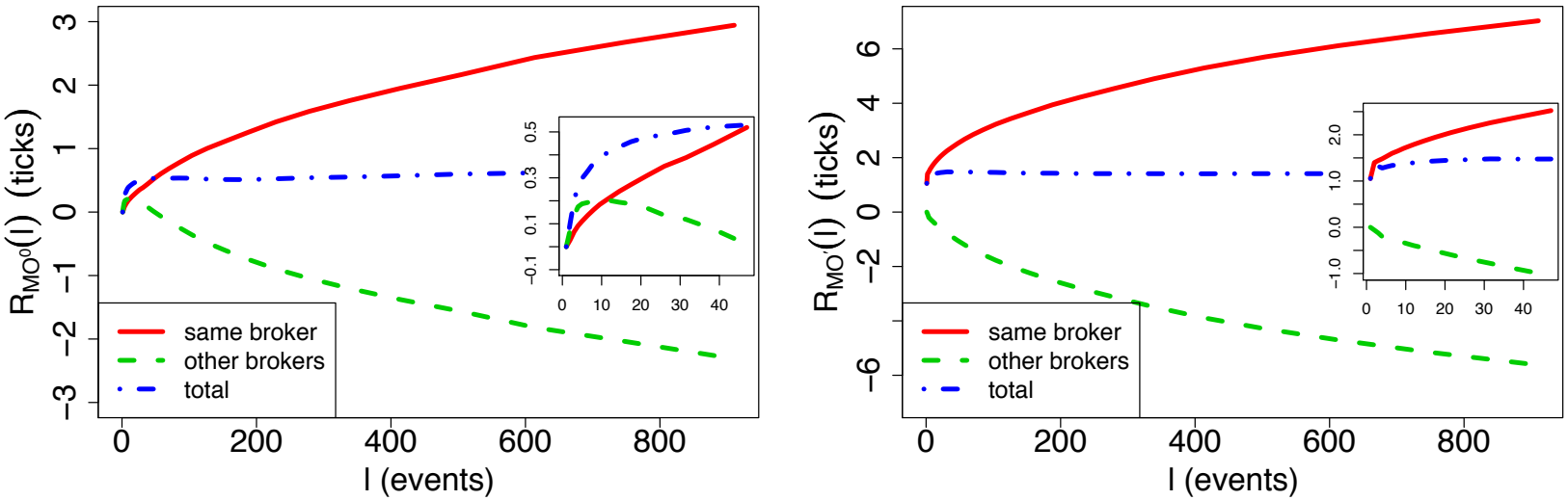
$$\mathcal{R}_{\pi_1}(\ell) = \frac{\langle (m_{n+\ell} - m_n) I(\pi_n = \pi_1) \epsilon_n \rangle}{P(\pi_1)}. \tag{62}$$

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.

$$\mathcal{R}^{\text{same}}_{\pi_1}(\ell) = \frac{\left\langle \sum_{n'=n}^{n+\ell-1} (p_{n'+1} - p_{n'}) I(b_{n'} = b_n) I(\pi_n = \pi_1) \epsilon_n \right\rangle}{P(\pi_1)}. \tag{63}$$

$$\mathcal{R}^{\text{diff}}_{\pi_1}(\ell) = \frac{\left\langle \sum_{n'=n}^{n+\ell-1} (p_{n'+1} - p_{n'}) I(b_{n'} \neq b_n) I(\pi_n = \pi_1) \epsilon_t \right\rangle}{P(\pi_1)}. \tag{64}$$

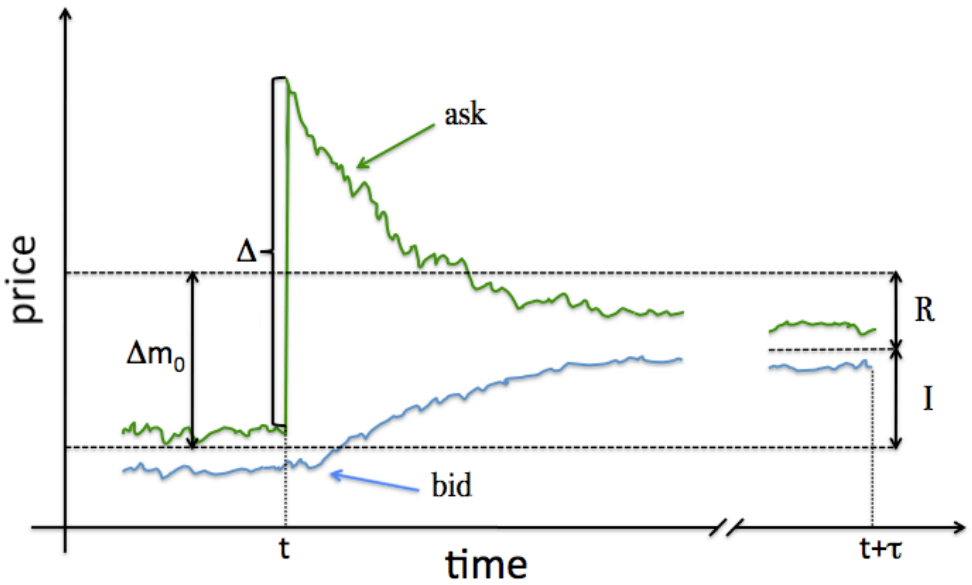
Impact (or response) is the result of a delicate balance

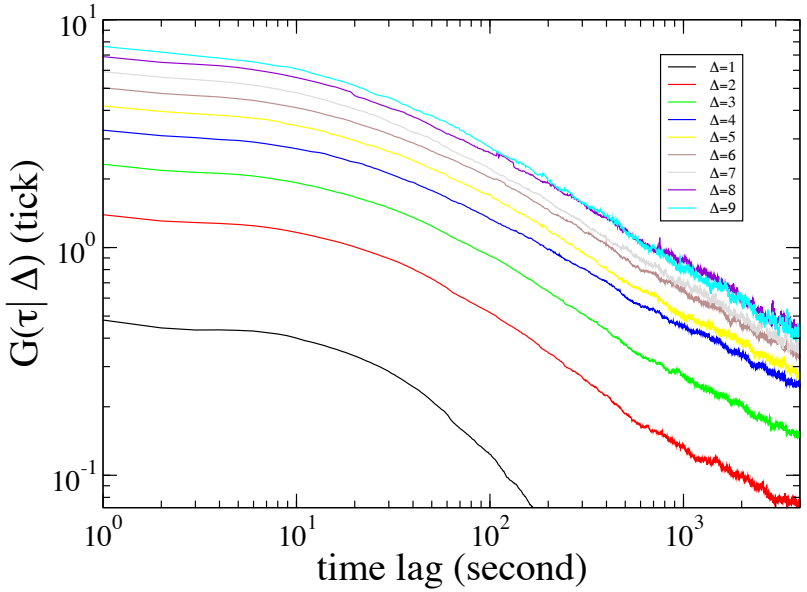


From Toth et al 2012. The two contributions very nearly offset each other, leading to a total impact that is nearly constant in time and much smaller than both these contributions.

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)

Spread decay





From Ponzi et al. 2009.

- The quantity

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

measures the average spread dynamics $s(t+\tau)$ conditional to a shock $\Delta s = \Delta$ at time t.

- Empirical data are consistent with $G(\tau|\Delta) \sim \tau^{-0.37}$ when $\Delta > 1$, not too different from the propagator exponent.

Transient impact model: summary

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 $G(k)$

History dependent impact model (HDIM)

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:

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

  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,

History dependent, permanent impact

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

$$\varphi_n^+ r_n^+ - \varphi_n^- r_n^- = 0. \tag{68}$$

  i.e.

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

- 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.

Equivalence between the two models

- 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.

Equivalence between the two models

- 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

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

- The equivalence is obtained with the relation:

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

Building a predictor: The DAR(p) model

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 $\Omega_{n-1} = \{X_{n-1}, \ldots, X_{n-p}\}$:

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

Asymmetric liquidity (Lillo and Farmer 2004)

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.

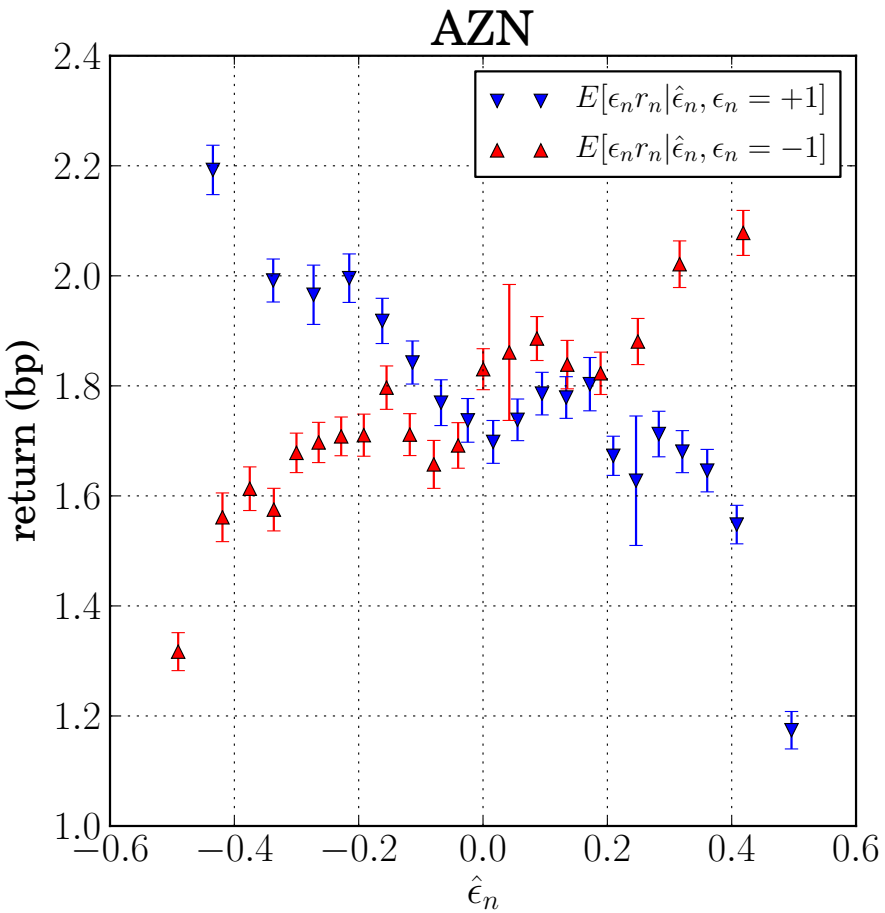


Figure: Expected return behavior as a function of an autoregressive sign predictor $\hat{\epsilon}_n \equiv \mathbb{E}[\epsilon_n|\epsilon_{n-1}, \epsilon_{n-2}, ...]$ for Astrazeneca (from Taranto et al. JSTAT 2014).

Direct tests of the model

- Negative lag response function

$$\mathcal{R}(-\ell) = -\sum_{0<i\leq\ell} \mathcal{S}(-i) = -\mathbb{E}[(m_t - m_{t-\ell}) \cdot \epsilon_t]. \tag{74}$$

which for the TIM is

$$\mathcal{R}^{\mathrm{TIM1}}(-\ell) = -\sum_{0<i\leq\ell} \sum_{n\geq 0} \mathcal{G}(n) C(n+i) < 0. \tag{75}$$

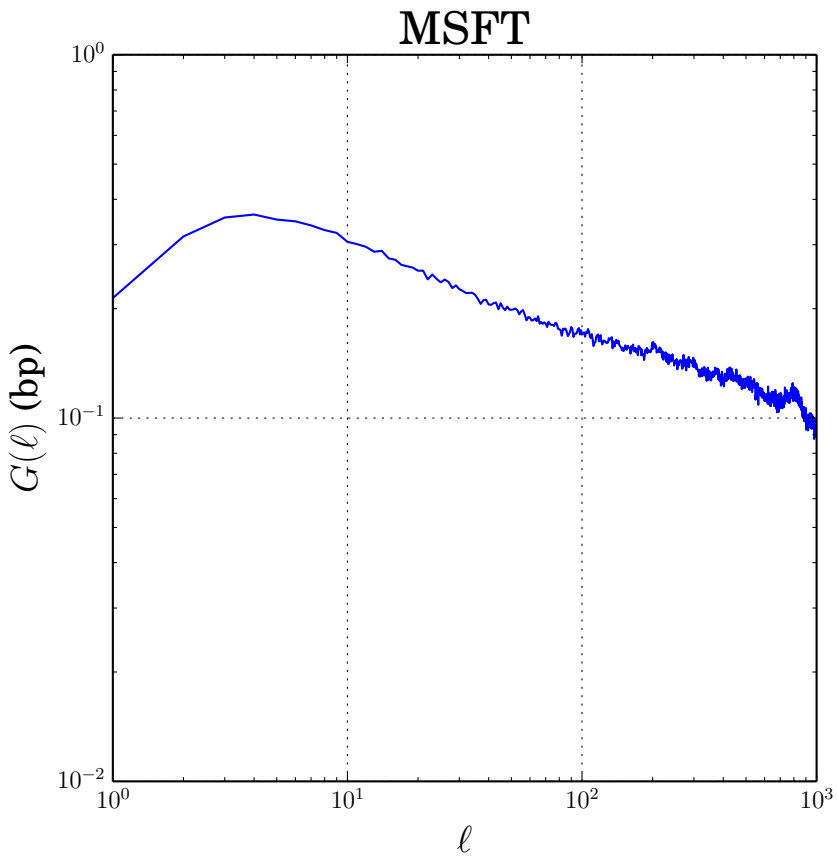
- Signature plot

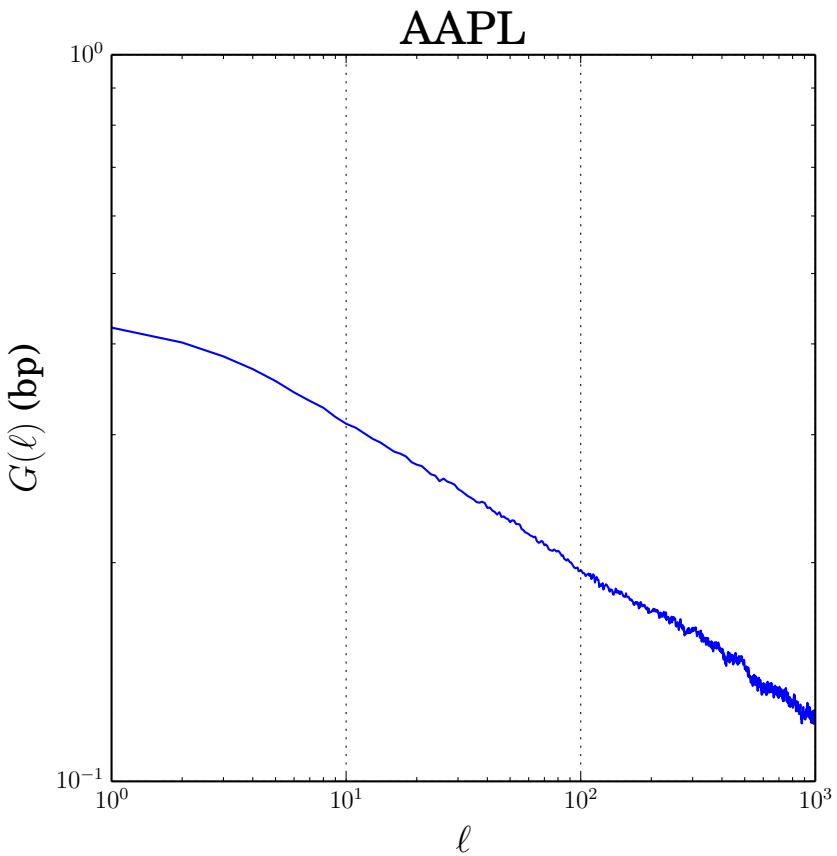
$$D(\ell) = \frac{1}{\ell}\mathbb{E}[(m_{t+\ell} - m_t)^2].$$

$$D^{\mathrm{TIM1}}(\ell) = \frac{1}{\ell}\sum_{0\leq n<\ell} G^2(\ell-n) + \frac{1}{\ell}\sum_{n>0}[G(\ell+n) - G(n)]^2 + 2\Psi(\ell) + \frac{D_{\mathrm{HF}}}{\ell} + D_{\mathrm{LF}},$$

where $\Psi(\ell)$ is the correlation-induced contribution to the price diffusion and we have added a high frequency (HF, e.g. microstructure noise) and low frequency (LF, e.g. news) component to the noise.

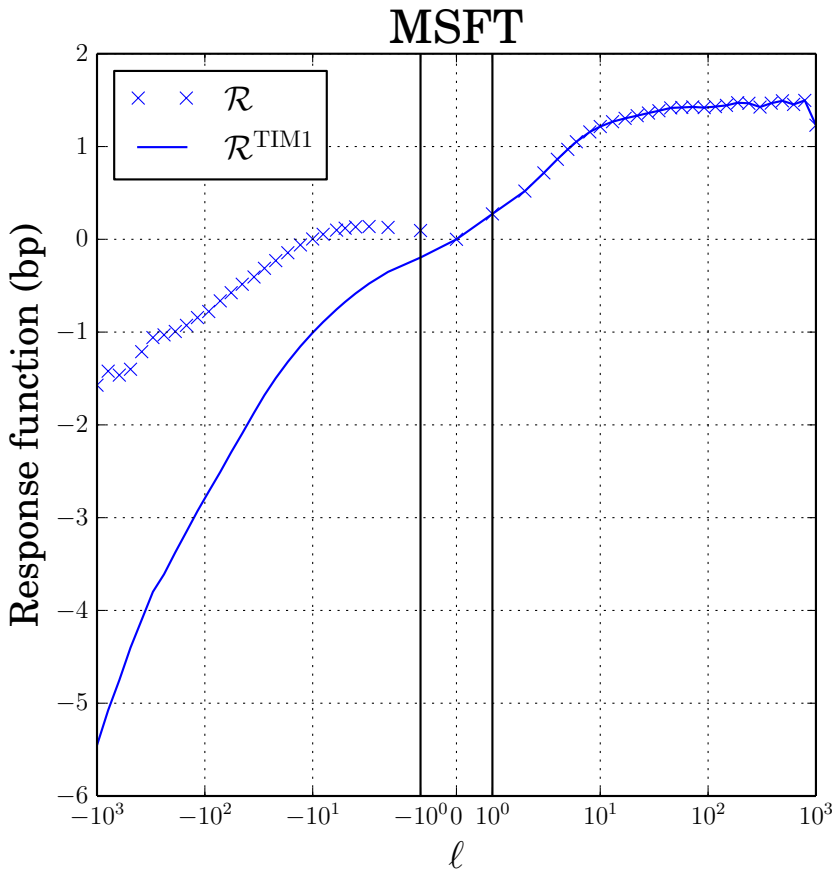
Fitted propagators

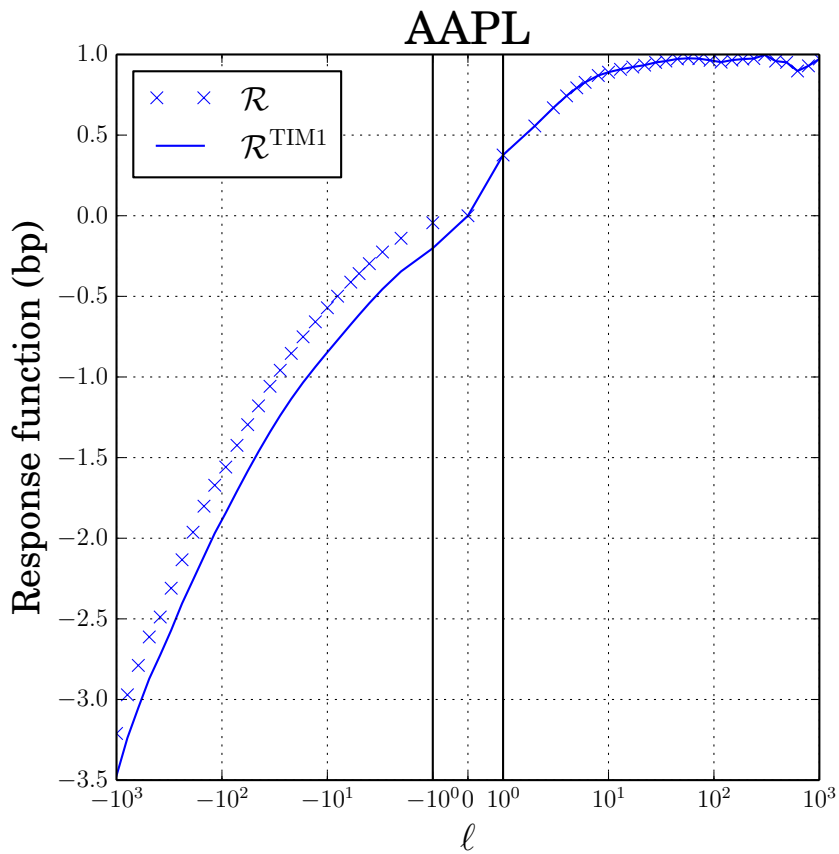




- Nasdaq 2013
- Slow decay ($\sim 1,000$ trades), effect of long memory
- For large tick stocks (MSFT) non-monotonic $\rightarrow$ inefficiency $\rightarrow$ dependence of order flow on price movement

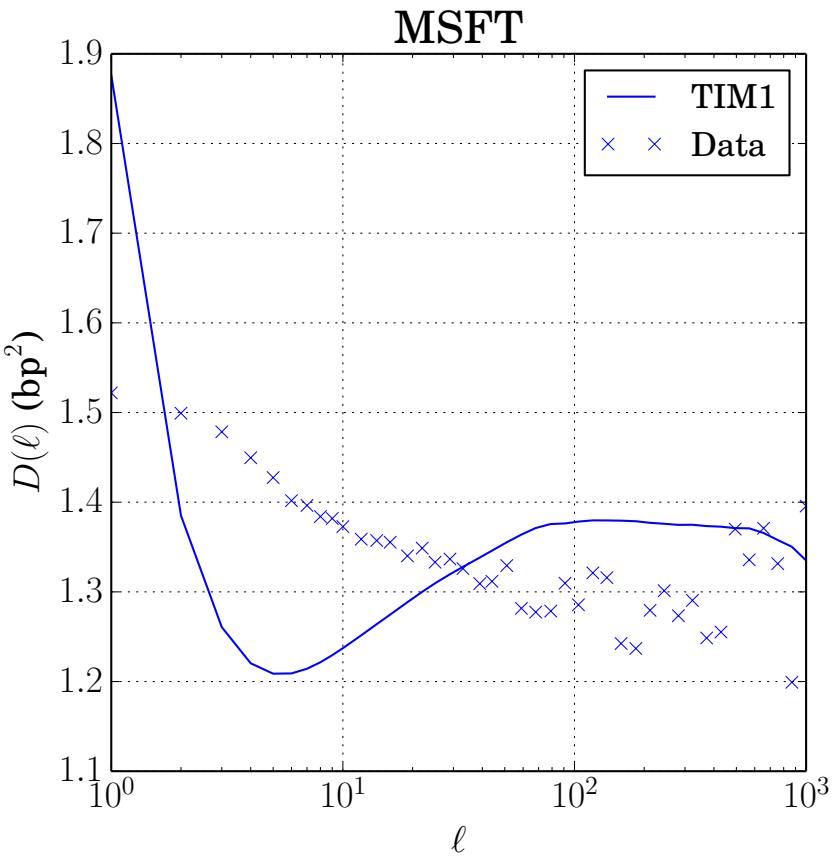
Negative lag response function

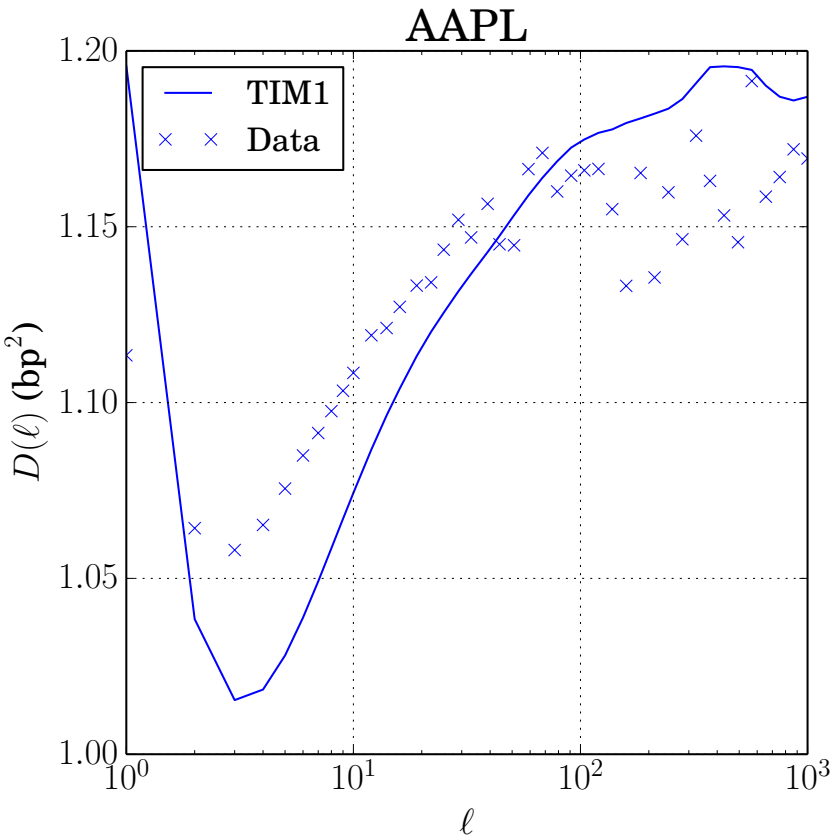




- Difference due to a rigid order flow not depending on price changes. This means that in the data there exists an additional anti-correlation between past returns and the subsequent order flow.
- Effect much stronger for large tick stocks

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.

Generalized linear impact models

Generalized linear impact models

The linear impact model has several limitations

- All market orders have the same impact, since G only depends on $t - t'$ and not on t and t' separately, which is certainly very crude. For example, some market orders are large enough to induce an immediate price change, and are expected to impact the price more than smaller market orders. One furthermore expects that depending on the specific instant of time and the previous history, the impact of market orders is different.
- Limit orders and cancellations should also impact prices, but their effect is only taken into account through the time evolution of $G(\ell)$ itself that phenomenologically describes how the flow of limit orders opposes that of market orders and reverts the impact of past trades.
- The model assumes a linear addition of the impact of past trades and neglect any non-linear effects which are known to exist. For example, the total impact of a metaorder of size Q is now well known to grow as $\sim \sqrt{Q}$, a surprising effect that can be traced to non-linearities induced by the deformation of the underlying supply and demand curve.

Two propagator models

- Idea: different types of event can have different impact on price (see Eisler et al 2012).
- We limit here to two types of events π_t defined as:

$$\pi_t = \begin{cases} \text{NC} & \text{if } r_t = m_{t+1} - m_t = 0 \\ \text{C} & \text{if } r_t = m_{t+1} - m_t \neq 0. \end{cases}$$

- Note: we consider only trades (differently from Eisler et al 2012), hence our returns include the reaction to the trade.
- Taranto et al 2016a and 2016b

Two propagator transient impact model (TIM2)

- The model is

$$r_t = \sum_{\pi} G_{\pi}(1) I(\pi_t = \pi)\epsilon_t + \sum_{t'<t} \sum_{\pi'} \mathcal{G}_{\pi'}(t-t') I(\pi_{t'} = \pi')\epsilon_{t'} + \eta_t,$$

  where $\pi = \{\mathrm{NC}, \mathrm{C}\}$ and $\mathcal{G}_{\pi'}(\ell) \equiv G_{\pi'}(\ell+1) - G_{\pi'}(\ell)$.

- It can be calibrated from

$$S_{\pi_1}(\ell) = \sum_{\pi_2} \mathbb{P}(\pi_2) \sum_{n \geq 0} \mathcal{G}_{\pi_2}(n) C_{\pi_1,\pi_2}(\ell - n). \tag{76}$$

  where

$$\mathcal{S}_{\pi}(\ell) = \mathbb{E}[r_{t+\ell} \cdot \epsilon_t | \pi_t = \pi] = \frac{\mathbb{E}[r_{t+\ell} \cdot \epsilon_t I(\pi_t = \pi)]}{\mathbb{P}(\pi)},$$

  and

$$C_{\pi_1,\pi_2}(\ell) = \frac{\mathbb{E}[\epsilon_t I(\pi_t = \pi_1) \cdot \epsilon_{t+\ell} I(\pi_{t+\ell} = \pi_2)]}{\mathbb{P}(\pi_1)\mathbb{P}(\pi_2)} \tag{77}$$

Two propagator history dependent impact model (HDIM2)

- The model is

$$r_t = \sum_{\pi} G_{\pi}(1) I(\pi_t = \pi)\epsilon_t + \sum_{t'<t} \sum_{\pi',\pi} \kappa_{\pi',\pi}(t-t') I(\pi_t = \pi) I(\pi_{t'} = \pi')\epsilon_{t'} + \eta_t,$$

  i.e. the expected sign for an event of type π is a linear regression of past signed events, with an "influence kernel" κ that depends on both the past event type π' and the current event π.

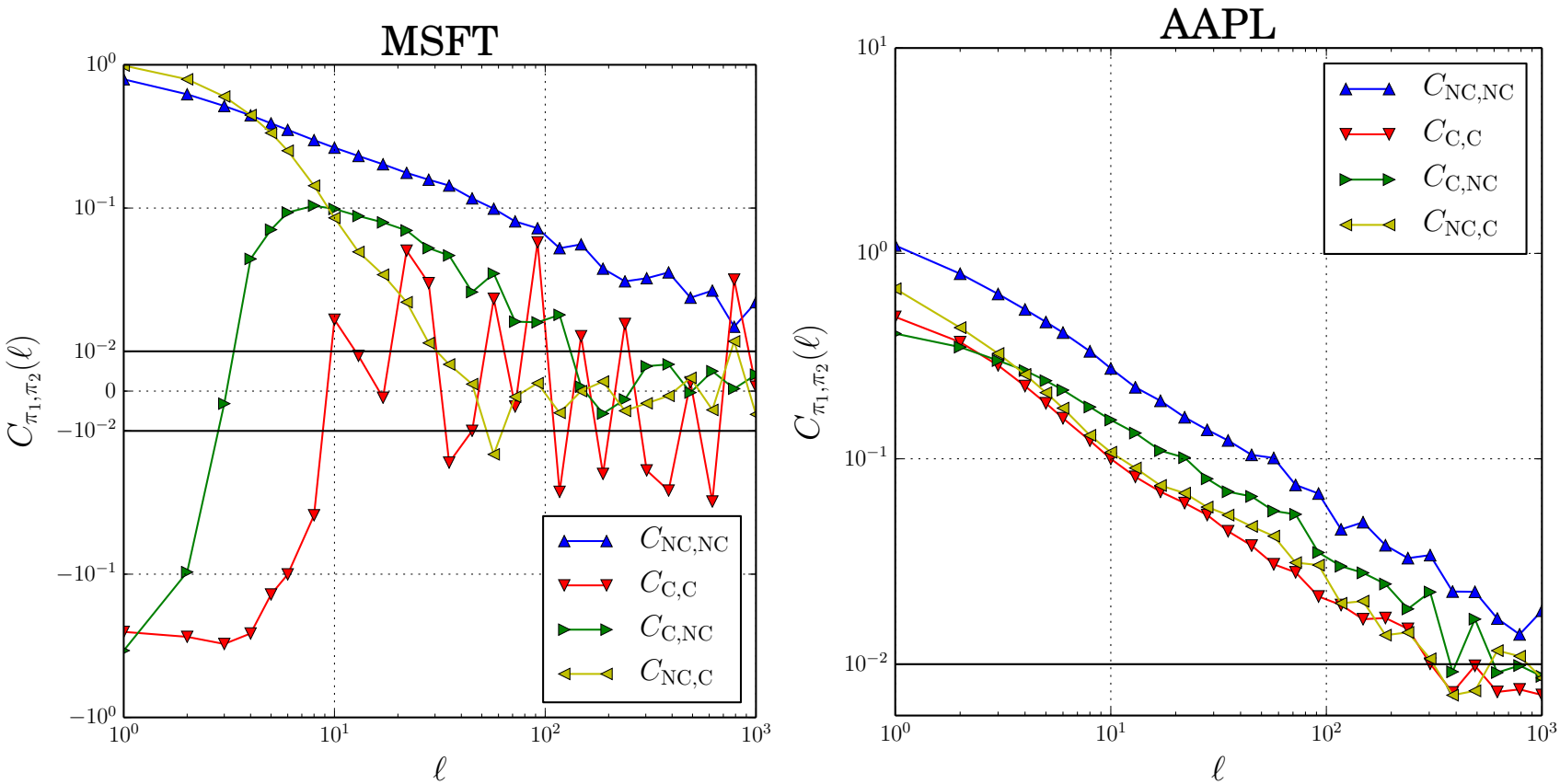
- The TIM2 is a special case of HDIM2 when

$$\kappa_{\pi',\pi}(\ell) = \mathcal{G}_{\pi'}(\ell), \qquad \forall \pi, \tag{78}$$

  i.e. only the type of the past event π' matters.

- Calibration is more subtle and requires the approximation of three-point and four-point correlations in terms of two-point correlations

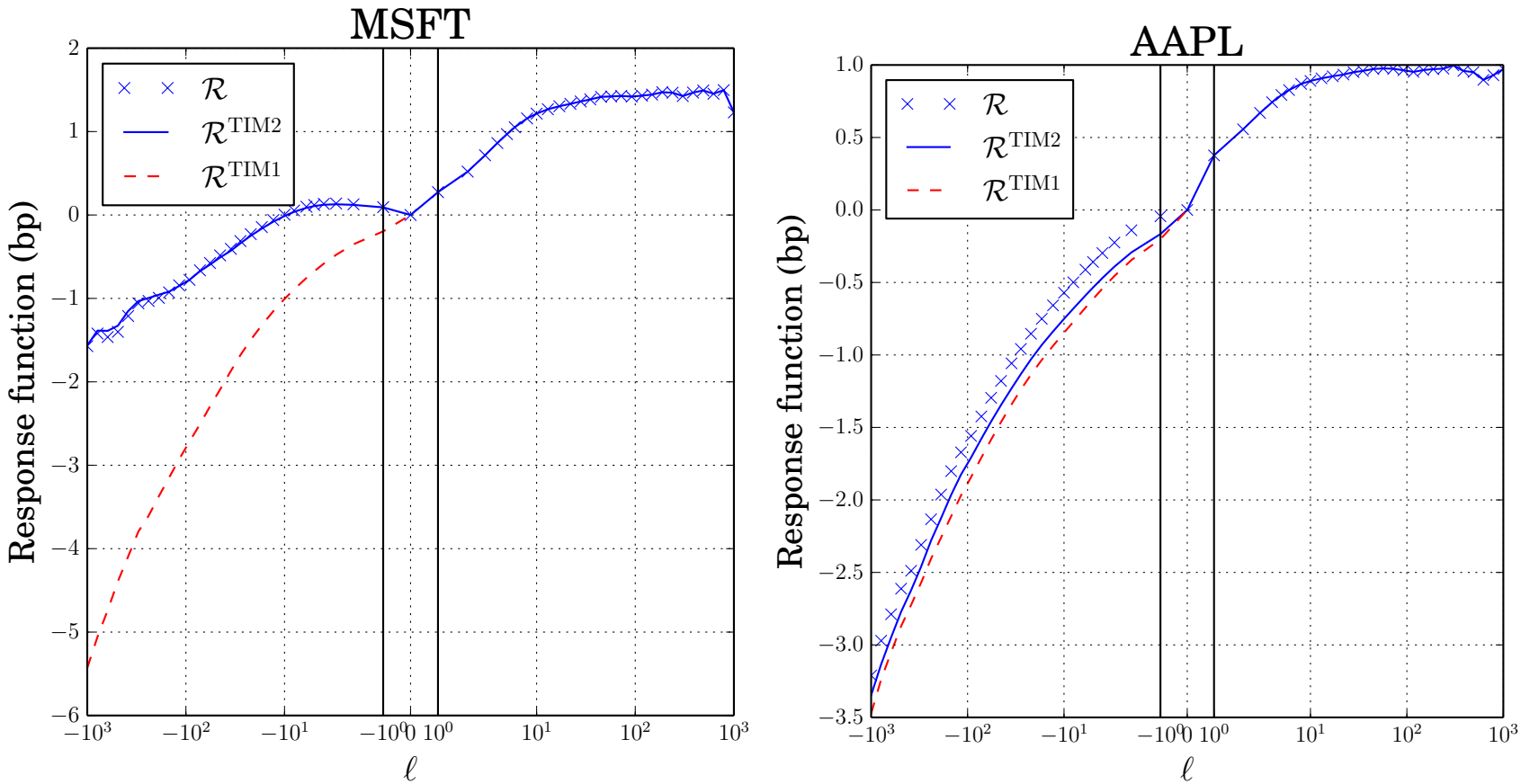
Conditional correlations



Note that the first subscript corresponds to the event that happened first chronologically

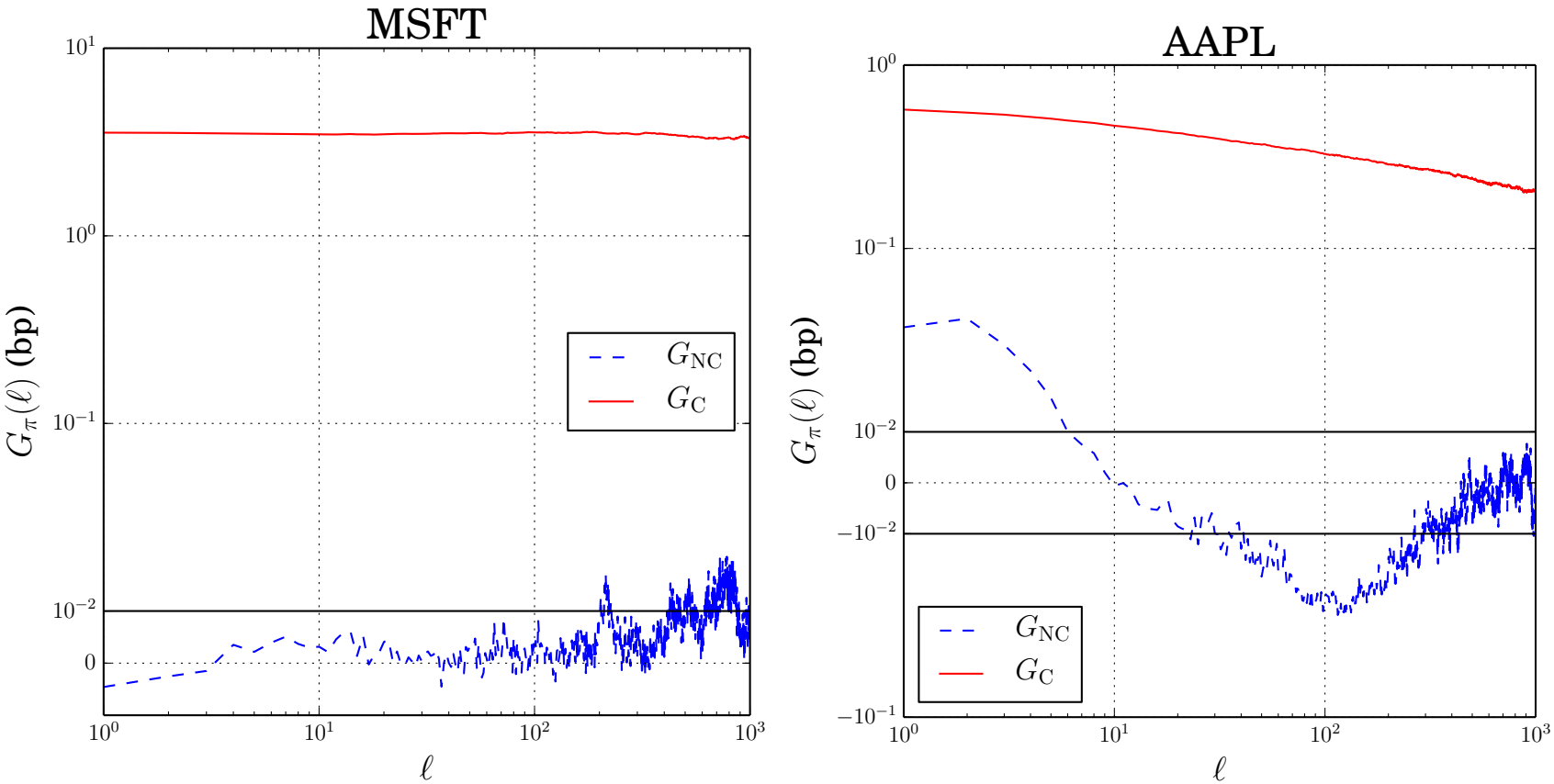
- Small tick (AAPL): C and NC events are not radically different and correlations are all similar.
- Large tick (MSFT): $C_{C,C}$ and $C_{C,NC}$ both start negative, i.e. after a price changing event, it is highly likely that the subsequent order flow (either C or NC) will be in the other direction (Note however that $\mathbb{P}(\pi = \mathrm{C}) = 0.08$).

Fitted propagators



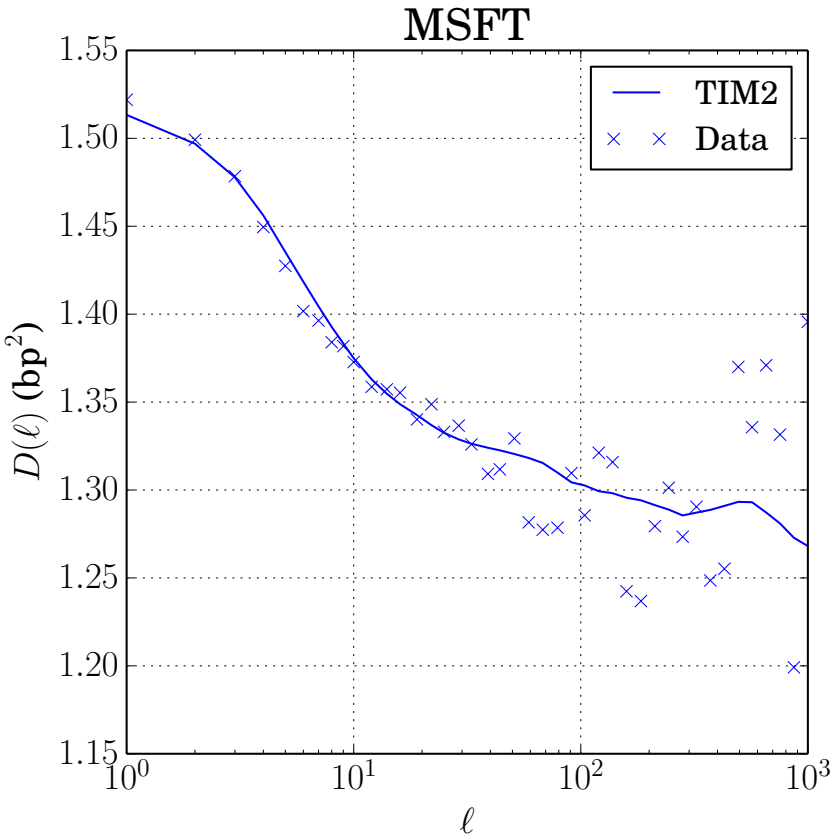
- In the case of large tick stocks the empirical curves are perfectly reproduced, whereas for small tick stocks some little deviation still persists.
- HDIM2 performs slightly better than TIM2 in capturing the excess anti-correlation measured from the small tick data between past returns and future order signs (not shown).

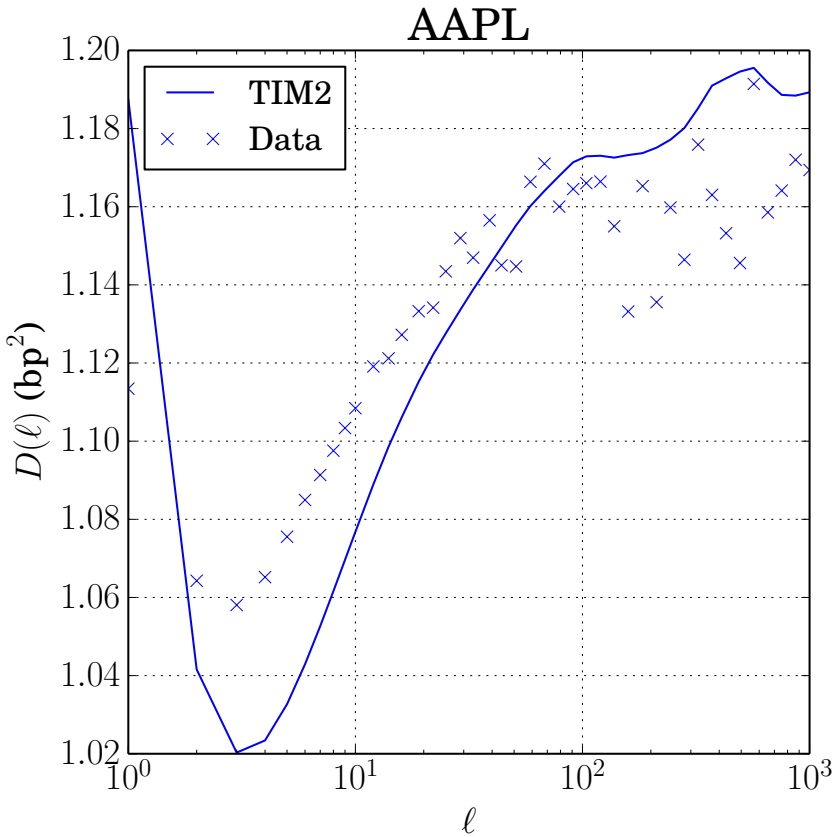
Negative lag response function



- G_C is equal to the spread, whereas G_{NC} is equal to zero.
- The dynamics of the price is completely determined by the sequence of random variables $\{(\epsilon_t, \pi_t)\}_{t\in\mathbb{N}}$, and the temporal structure of their correlations.

Signature plot of TIM2 (and HDIM2)





- Clear improvement with respect to TIM1

Recap

- We cannot model the order flow dynamics as exogenous (as in the TIM model).
- A two propagator model is able to reproduce satisfactorily well the negative lag response function and the signature plot.
- Propagators are constant, hence the sequence of random variables $\{(\epsilon_t, \pi_t)\}_{t\in\mathbb{N}}$ reproduces the price dynamics.
- We need a stochastic model describing the joint dynamics of order flow and prices.
- One possibility VAR (Hasbrouck 1991). However
  - VAR is not suited for discrete variables
  - Linear relation between the variables, vs linear relation between past variables and future probabilities.

Linear models

- m state Markov chain X_t; $\mathbf{Q}$ is the $m \times m$ time-invariant transition matrix, and let $\chi_t = (x_t(1), ..., x_t(m))$ be a row vector such that $x_t(i) = 1$ if $X_t = i$ and zero otherwise.
- The probability vector $\hat{\chi}_t = (\mathbb{P}(X_t = 1), ..., \mathbb{P}(X_t = m))$ is determined by the *linear* system of equations

$$\hat{\chi}_t = \chi_{t-1}\mathbf{Q}\,.$$

- Idea: Markov process with $m = 4$ states, $(\epsilon_t, \pi_t) \in \{(-1, \mathrm{C}), (-1, \mathrm{NC}), (+1, \mathrm{NC}), (+1, \mathrm{C})\}$, corresponding to buys $(\epsilon_t = +1)$ and sells $(\epsilon_t = -1)$ and price changing $(\pi_t = \mathrm{C})$ and not changing $(\pi_t = \mathrm{NC})$ trades.
- Problem: the long memory of correlations requires a large number of parameters

(Generalized) Mixture Transition Distribution Model (MTDg)

- MTDg(p) model: Raftery 1985, Berchtold 1995 $\{X_t\}_{t\in\mathbb{N}}$ be a sequence of random variables taking values in the finite set $\mathcal{X} = \{1, \ldots, m\}$
- $\forall t > p$ and $\forall (i, i_1, \ldots, i_p) \in \mathcal{X}^{p+1}$,

$$\mathbb{P}(X_t = i | X_{t-1} = i_1, \ldots, X_{t-p} = i_p) = \sum_{g=1}^{p} \lambda_g q^g_{i_g, i}\,, \tag{79}$$

where the vector $\lambda = (\lambda_1, \ldots, \lambda_p)$ is subject to the constraints:

$$\lambda_g \geq 0, \qquad \forall g \in \{1, \ldots, p\}\,, \tag{80}$$

$$\sum_{g=1}^{p} \lambda_g = 1\,. \tag{81}$$

- The matrices $\left\{\mathbf{Q}^g = \left[q^g_{i,j}\right]\,;\, i,j \in \mathcal{X}; 1 \leq g \leq p\right\}$ are positive $m \times m$ stochastic matrices, i.e. they satisfy

$$q^g_{i,j} \geq 0 \quad \text{and} \quad \sum_{j=1}^{m} q^g_{i,j} = 1 \qquad \forall g \in \{1, \ldots, p\}, \forall i, j \in \mathcal{X}\,. \tag{82}$$

(Generalized) Mixture Transition Distribution Model (MTDg)

- The number of parameters is $O(m^2 p)$ rather than $O(m^p)$ as in a Markov chain of order p.
- This model can be interpreted as a probabilistic mixture of Markov processes.
- However the model has still a probabilistic interpretation when $(\lambda_g)_{g=1,\ldots,p}$ is not probability vector and $\mathbf{Q}^g$ are not stochastic matrices, provided that

$$0 \leq \sum_{g=1}^{p} \lambda_g q^g_{i_g,i} \leq 1, \qquad \forall (i, i_1, \ldots i_p) \in \mathcal{X}^{p+1}, \tag{83}$$

- We shall assume the matrices $\mathbf{Q}^g$ share the same stationary state, i.e. the same left eigenvector $\hat{\eta}$ corresponding to the eigenvalue 1

Theorem

Suppose that a sequence of random variables $\{X_t\}_{t\in\mathbb{N}}$ taking values in the finite set $\mathcal{X} = \{1, \ldots, m\}$ is defined by

$$\mathbb{P}(X_t = i | X_{t-1} = i_1, \ldots, X_{t-p} = i_p) = \sum_{g=1}^{p} \lambda_g q^g_{i_g, i} \,,$$

where $\mathbf{Q}^g = \left[q^g_{i,j}\right]_{i,j\in\mathcal{X}}$ are matrices with normalized rows, $\sum_j q^g_{i,j} = 1$, $\sum_{g=1}^p \lambda_g = 1$, and assume that $\hat{\eta}\mathbf{Q}^g = \hat{\eta}, \forall g$. If the vector $\hat{\eta}$ is such that $\hat{\eta}_i > 0, i \in \mathcal{X}$ and $\sum_i \hat{\eta}_i = 1$, and

$$0 < \sum_{g=1}^{p} \lambda_g q^g_{i_g, i} < 1, \qquad \forall (i, i_1, \ldots, i_p) \in \mathcal{X}^{p+1} \,, \tag{84}$$

then

$$\lim_{\ell\to\infty} \mathbb{P}(X_{t+\ell} = i | X_{t-1} = i_1, \ldots, X_{t-p} = i_p) = \hat{\eta}_i \,.$$

Maximum Likelihood.

$$\left(\hat{\lambda}_g, \hat{\mathbf{Q}}^g\right)_{1\leq g\leq p} = \underset{(\lambda_g,\mathbf{Q}^g)_{1\leq g\leq p}}{\operatorname{argmax}} \sum_{t=p+1}^{n} \log\left\{\sum_{g=1}^{p} \lambda_g q^g_{x_{t-g},x_t}\right\},$$

$$\text{s.t} \quad \sum_{g=1}^{p} \lambda_g = 1\,,$$

$$\lambda_g \geq 0, \qquad \forall g \in \{1,\ldots,p\}$$

$$q^g_{i,j} \geq 0 \quad \text{and} \quad \sum_{j=1}^{m} q^g_{i,j} = 1 \qquad \forall g \in \{1,\ldots,p\}, \forall i,j \in \mathcal{X}\,. \tag{85}$$

Large number of constraints, hard to solve unless suitably (and strongly) parametrized.

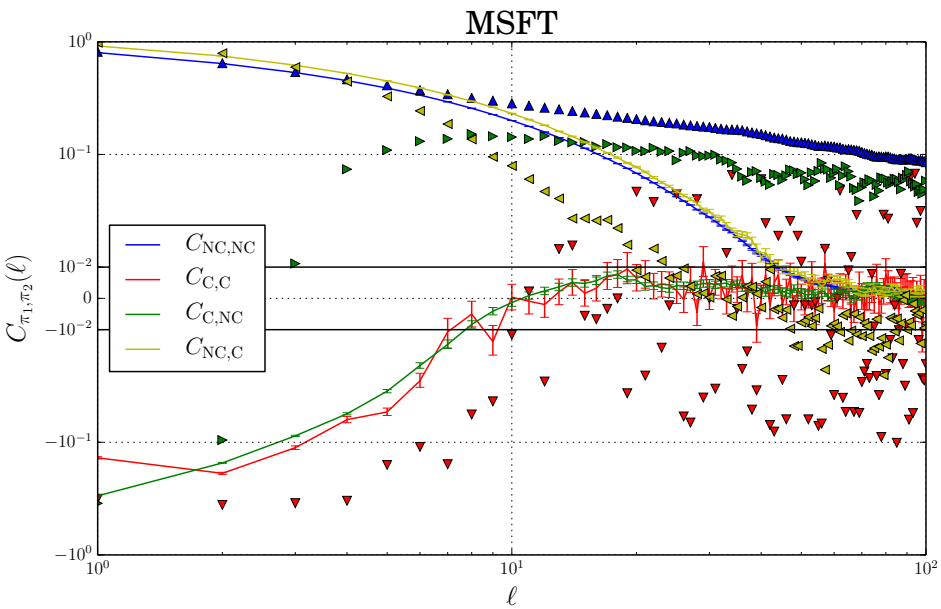
Estimation: MLE

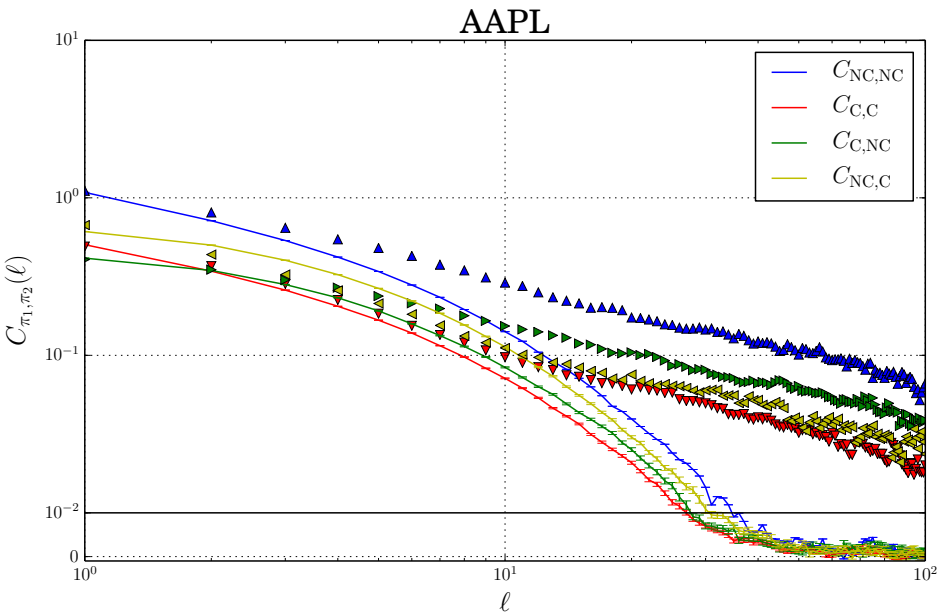
- We choose a 11 parameter model
- Impose buy-sell symmetry
- $\lambda_g = N_\beta g^{-\beta}$, where $N_\beta^{-1} = \sum_{i=1}^{p} g^{-\beta}$
- Writing $\mathbf{Q}^g = \mathbf{Q} + \tilde{\mathbf{Q}}^g$, we make the following strongly parametrized ansatz:

$$\mathbf{Q} = \begin{pmatrix} B_1 & A_1 & A_1 & B_1 \\ B_2 & A_2 & A_2 & B_2 \\ B_2 & A_2 & A_2 & B_2 \\ B_1 & A_1 & A_1 & B_1 \end{pmatrix}, \tilde{\mathbf{Q}}^g = \begin{pmatrix} -\mu_1 e^{-\alpha_{11} g} & -\nu_1 e^{-\alpha_{12} g} & \nu_1 e^{-\alpha_{12} g} & \mu_1 e^{-\alpha_{11} g} \\ \mu_2 e^{-\alpha_{21} g} & \nu_2 e^{-\alpha_{22} g} & -\nu_2 e^{-\alpha_{22} g} & -\mu_2 e^{-\alpha_{21} g} \\ -\mu_2 e^{-\alpha_{21} g} & -\nu_2 e^{-\alpha_{22} g} & \nu_2 e^{-\alpha_{22} g} & \mu_2 e^{-\alpha_{21} g} \\ \mu_1 e^{-\alpha_{11} g} & \nu_1 e^{-\alpha_{12} g} & -\nu_1 e^{-\alpha_{12} g} & -\mu_1 e^{-\alpha_{11} g} \end{pmatrix}, \tag{86}$$

  where $\alpha_{ij} \geq 0$.
- $\boldsymbol{\theta} = \{\beta, B_i, \mu_i, \nu_i, \alpha_{ij}\}$

Estimation: MLE





Generalized Method of Moments.

Proposition.

Suppose that a sequence of random variables $\{X_t\}_{t\in\mathbb{N}}$ taking values in the finite set $\mathcal{X} = \{1, \ldots, m\}$ is defined by Eq. 79 and is stationary. Let $\mathbf{B}(k)$ be a $m \times m$ matrix with elements

$$b_{i,j}^k = \mathbb{P}(X_t = i, X_{t+k} = j), \qquad i, j \in \mathcal{X}; k \in \mathbb{Z}$$

and $\mathbf{B}(0) = \mathrm{diag}(\hat{\eta}_1, \ldots, \hat{\eta}_m)$. Then

$$\mathbf{B}(k) = \sum_{g=1}^{p} \lambda_g \mathbf{B}(k-g)\mathbf{Q}^g \,. \tag{87}$$

$m^2 p$ different equations, which can be reduced to $p(m^2 - 2m + 1)$

Estimation: GMM

- We introduce MTD(p) models where

$$\mathbf{Q}^g = 1^T \hat{\eta} + \tilde{\mathbf{Q}}^g \tag{88}$$

  and $\hat{\eta}\tilde{\mathbf{Q}}^g = 0$. Similar to DAR(p) model.
- The GMM equations become

$$\mathbf{B}(k) - \hat{\eta}^T \hat{\eta} = \sum_{g=1}^{p} \mathbf{B}(k-g)\mathbf{A}^g. \tag{89}$$

  where $\mathbf{A}^g \equiv \lambda_g \tilde{\mathbf{Q}}^g$.
- The knowledge of $\mathbf{A}^g$ gives not uniquely λ_g and $\tilde{\mathbf{Q}}^g$. The generating model is uniquely determined.

Estimation: GMM

- We solve the optimization problem

$$\hat{q} = \underset{\mathbf{q}\in\mathbb{R}^{p(m^2-2m+1)}}{\operatorname{argmin}} \|\mathbf{d} - \mathbf{K}\cdot\mathbf{q}\|^2$$

$$\text{s.t.} \quad \hat{\eta}_i + \sum_{g=1}^{p} \max_{i_g} \left(a^g_{i_g,i}\right) < 1, \qquad \forall i \in \mathcal{X}$$

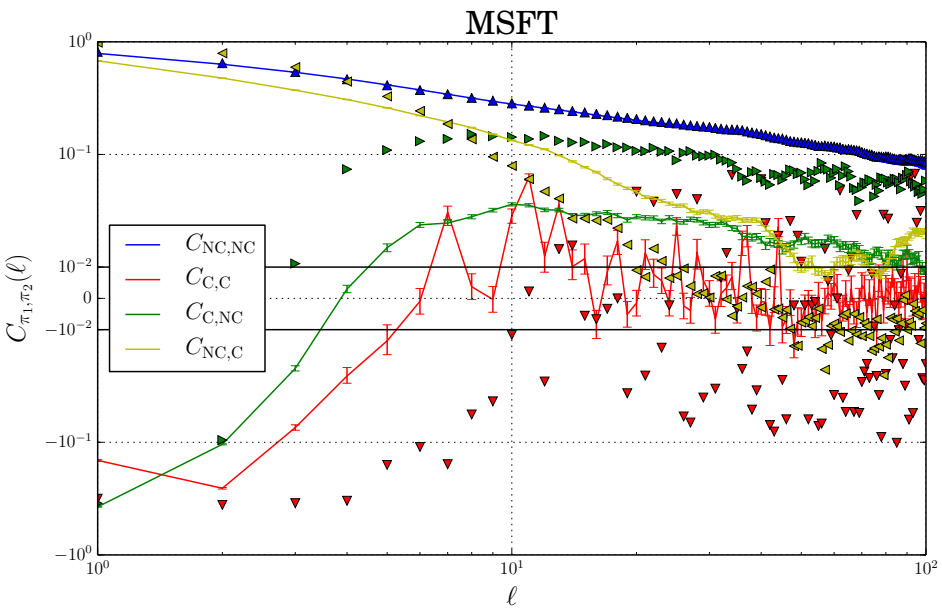
$$\hat{\eta}_i + \sum_{g=1}^{p} \min_{i_g} \left(a^g_{i_g,i}\right) > 0, \qquad \forall i \in \mathcal{X} \tag{90}$$

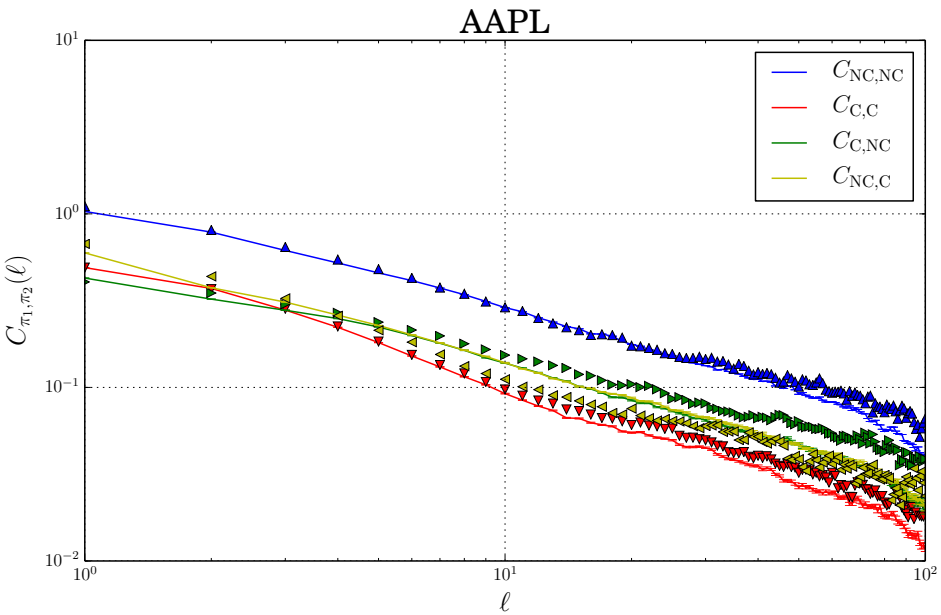
  where $\mathbf{d}$ and $\mathbf{K}$ depend on $\mathbf{B}(k)$ and $\hat{\eta}$ and $\mathbf{q}$ depend on $\mathbf{A}^g$.
- The inequalities guarantee the uniqueness of the stationary solution

Proposition

If $\mathbf{K}$ is not singular, the optimization program of Eq. (90) is strictly convex in $\mathbb{R}^{p(m^2-2m+1)}$.

Estimation: GMM





Out of sample analysis

		Model A	Model B	Model C
MSFT	EPE	1.928	1.199	1.181
	SE	0.003	0.004	0.004
BAC	EPE	1.744	0.799	0.785
	SE	0.003	0.004	0.004
GE	EPE	1.922	1.169	1.153
	SE	0.004	0.005	0.005
CSCO	EPE	1.919	1.112	1.098
	SE	0.004	0.005	0.005
AAPL	EPE	2.643	2.211	2.192
	SE	0.001	0.002	0.002
AMZN	EPE	2.579	2.196	2.183
	SE	0.002	0.004	0.004

EPE values

$$\mathrm{EPE}(\boldsymbol{\theta}) = \mathbb{E}[L(X_t, \hat{X}_t^{\boldsymbol{\theta}})]\,,$$

and standard errors (SE) for MSFT, BAC, GE, CSCO, AAPL and AMZN data. *Model A*: Unconditional probabilities as predictor. *Model B*: Strongly constrained MTDg(100) estimated via MLE. Total number of parameters: 11. *Model C*: Weakly constrained MTDg(100) model estimated via GMM with matrices. Total number of parameters: 500.

Market impact of metaorders under TIM-HDIM models

Market impact laws

- 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

$$\Delta \ln p \equiv \mathcal{I}(Q) = \pm Y \sigma_D \left(\frac{Q}{V_D} \right)^{\delta} \tag{91}$$

  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)

Impact of metaorders from proprietary data

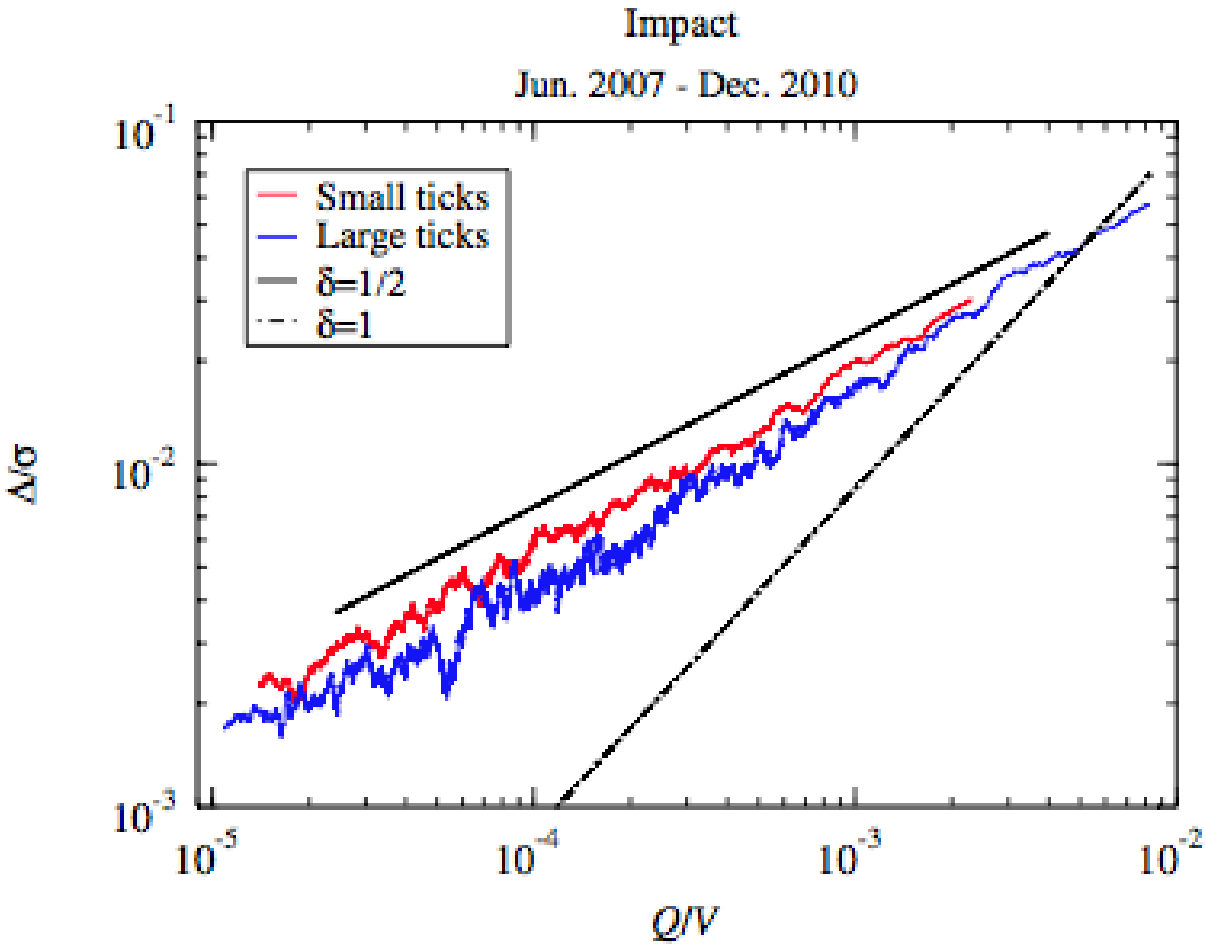


Figure: From Toth et al. 2011. The impact of metaorders for Capital Fund Management proprietary trades on futures markets, Impact is measured here as the average execution shortfall of a metaorder of size Q. The data base contains nearly $500,000$ trades. We show $\mathcal{I}(Q)/\sigma_D$ vs Q/V_D on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

Temporary and permanent impact

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 $[0, 1]$.

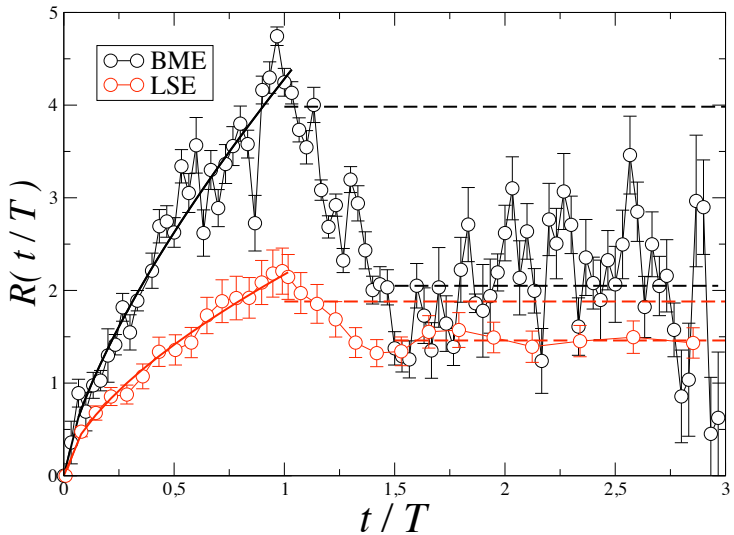


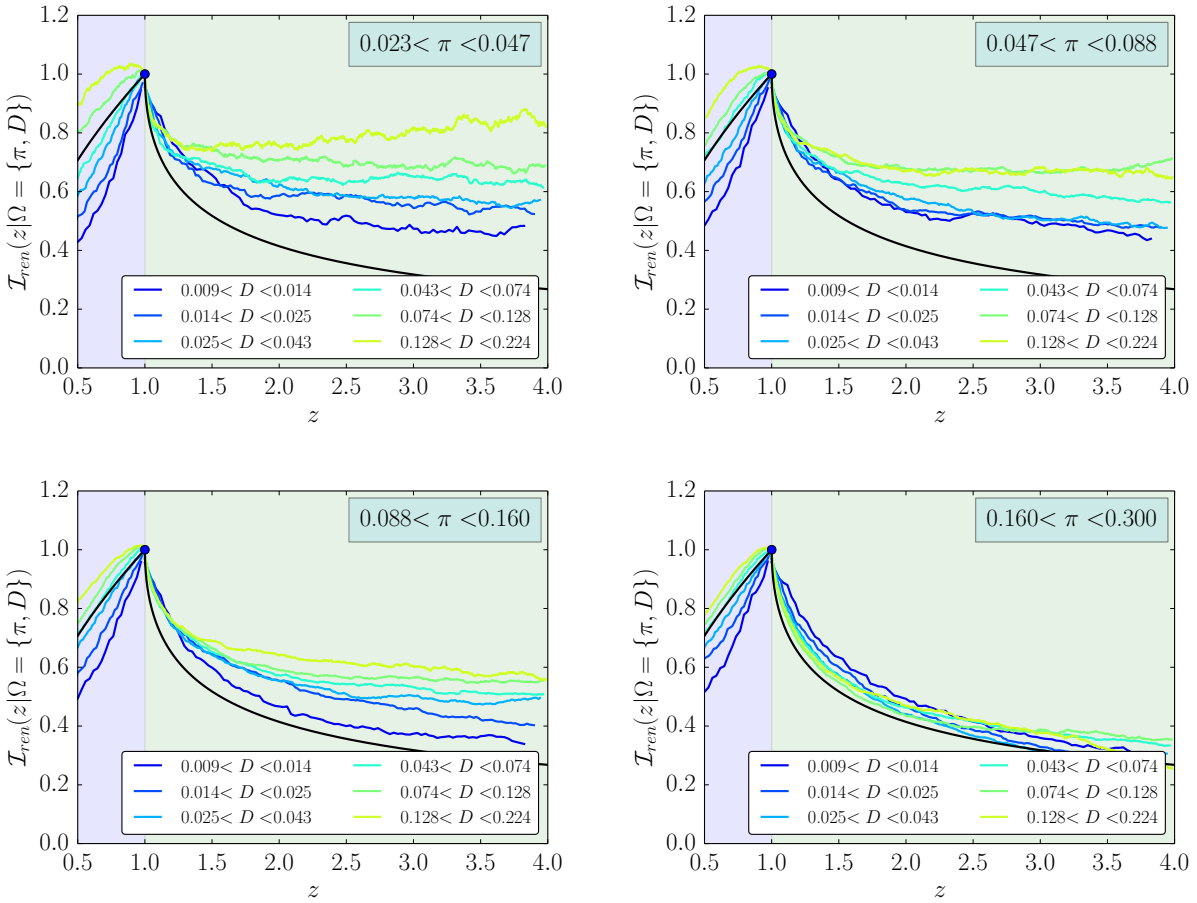
Figure: Market impact versus time. The symbols are the average value of the market impact of the metaorder as a function of the normalized time to completion t/T. The rescaled time $t/T = 0$ corresponds to the starting point of the metaorder, while $t/T = 1$ corresponds to the end of the metaorder.

We find approximately the square root law

$$E[r|N] = A\epsilon s N^{\beta} \qquad \beta \simeq 1/2 \tag{92}$$

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

Temporary and permanent impact



From Zarinelli et al 2015. Decay of temporary market impact after the execution of a metaorder. Within each panel the solid lines correspond to the average market impact trajectory for metaorders with different durations D; the four panels correspond to different participation rates π. The black line corresponds to the prediction of the transient impact model with $\delta = 0.5$.

Impact under propagator/asymmetric liquidity model

- Only one metaorder order active at a given time, made of N trades (constant volume), and executed with a flow of uncorrelated orders with a constant participation rate π
- The total time needed to execute the hidden order is then $T = N/\pi$.
- If m_0 the price at the beginning of the metaorder, the total impact is

$$E[m_N] - m_0 = \epsilon\theta \sum_{i=1}^{N} \left(1 - \sum_{k=1}^{i/\pi} a_k \right) . \tag{93}$$

Impact under propagator/asymmetric liquidity model

- We assume that the participants observing public information model the time series with a FARIMA process. For large k the best linear predictor coefficients of a FARIMA process satisfy $a_k \approx k^{-\beta-1}$ where $\beta = (1-\gamma)/2$. For large k we can pass into the continuum limit and from Eq. 93 the impact is

$$E[m_N] - m_0 = \epsilon\theta \left[1 + \sum_{i=1}^{N-1} \left(1 - \left(1 - (n/\pi)^{-\beta} \right) \right) \right]. \tag{94}$$

  Converting the sum to an integral gives

$$E[m_N] - m_0 \approx \epsilon\theta \left(1 + \frac{2^{\beta-1}\pi^{\beta}}{1-\beta} [(2N-1)^{1-\beta} - 1] \right) \sim \pi^{\beta} N^{1-\beta}. \tag{95}$$

- For a fixed participation rate, the market impact asymptotically increases with the length of the hidden order as $N^{1-\beta} \sim N^{0.75}$.
- The size of the impact varies as π^{β}. This means that the slower an order is executed, the less impact it has, and in the limit as the order is executed infinitely slowly the impact goes to zero.
- Note however that if the execution time $T = N/\pi$ is *fixed*, the impact become linear with N but decays as $T^{-\beta}$

Impact decay under propagator/asymmetric liquidity model

- No noise traders ($\pi = 1$), FARIMA model on the past K trades

$$\hat{\epsilon}_n = \sum_{i=1}^{K} a_i^{(K)} \epsilon_{n-i} \tag{96}$$

where

$$a_i^{(K)} = -\binom{K}{i} \frac{\Gamma(i - H + 1/2)\Gamma(K - H - i + 3/2)}{\Gamma(1/2 - H)\Gamma(K - H + 3/2)} \tag{97}$$

and $H = 1/2 - \beta$ is the Hurst exponent of the FARIMA process.

- Permanent impact is

$$E[m_\infty] - m_0 = \epsilon\theta N(1 - \sum_{j=1}^{K} a_j^{(K)}) = \tag{98}$$

$$\epsilon\theta N \frac{4^{H-1}\sqrt{\pi}\Gamma[H]\sec[(K - H)\pi]}{\Gamma(3/2 + K - H]\Gamma[2H - 1 - K]}$$

which can be approximated as

$$E[m_\infty] - m_0 \sim \epsilon\theta \frac{N}{K^\beta}. \tag{99}$$

Impact decay under propagator/asymmetric liquidity model

- If K is infinite, then $E[m_\infty] - m_0 = 0$, i.e. the impact is completely temporary (as in a pure propagator model).
- For a FARIMA forecast model with finite K (or equivalently if the sign autocorrelation function decays fast beyond time scale K), the permanent impact is non zero and is linear in N. Even if for large K the permanent impact is small, the convergence to zero with the memory K is very slow.
- Immediately after the end of the metaorder the initial drop for $t \ll N$ is in fact very sharp for $\beta < 1$: $m_{N+t} - m_N \propto -t^{1-\beta}$, such that the slope of the decay is infinite when $t \to 0$ (in the continuous limit)

Which microstructural mechanisms are responsible for asymmetric liquidity?

Data

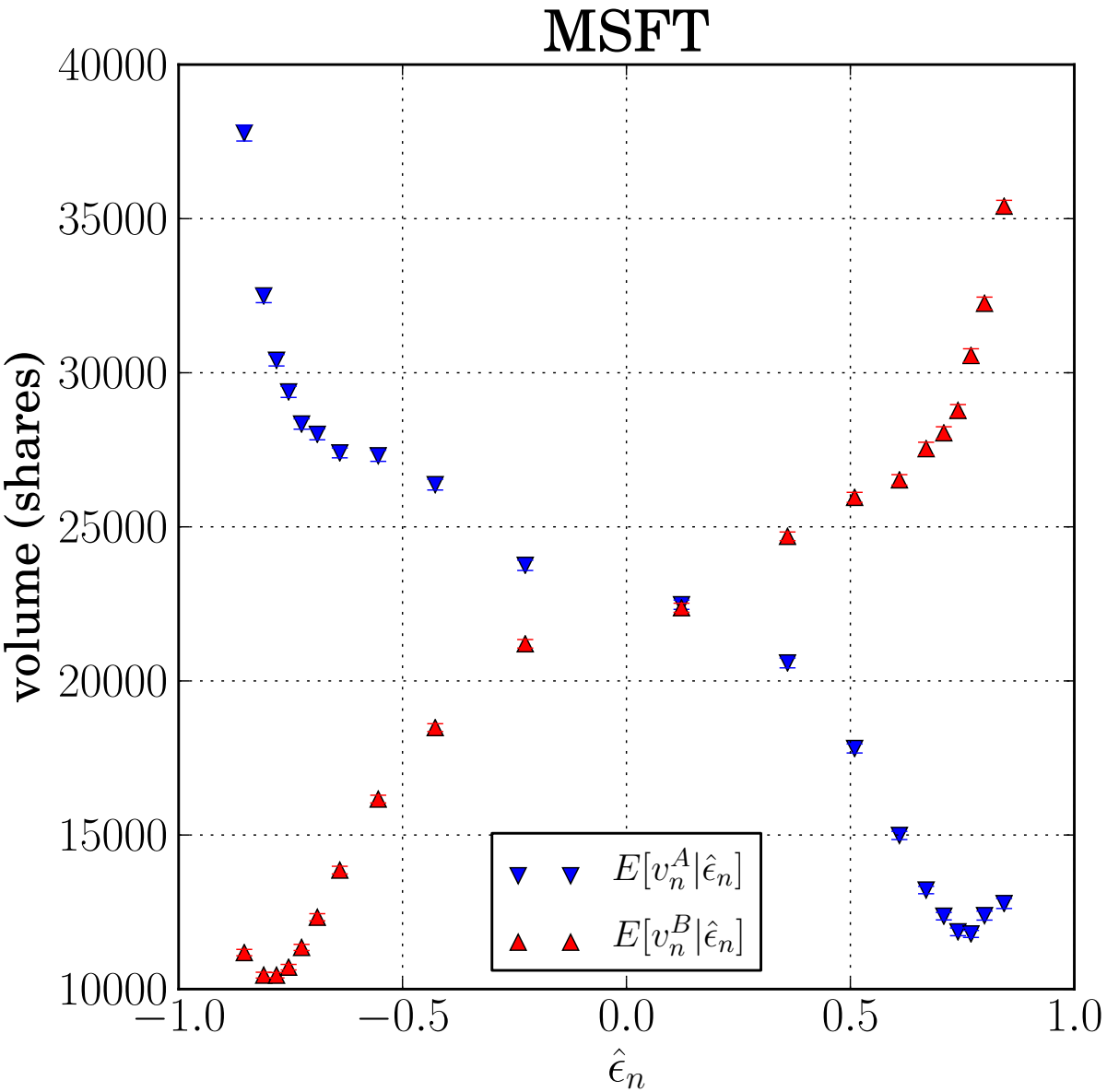
Two datasets

- LSE dataset includes Astrazeneca (AZN) and Vodafone (VOD) stocks in 2004 (254 trading days)
- NASDAQ dataset includes Apple (AAPL) and Microsoft (MSFT) in July-August 2009 (42 trading days)

Symbol	Number of trades	Average intertrade time	Average stock price	Average tick size-price ratio
AAPL	857925	1.1 s	157.17 USD	0.6 bp
MSFT	575040	1.7 s	23.74 USD	4.2 bp
AZN	405481	23.1 s	24.38 GBP	4.1 bp
VOD	411736	22.9 s	1.34 GBP	18.7 bp

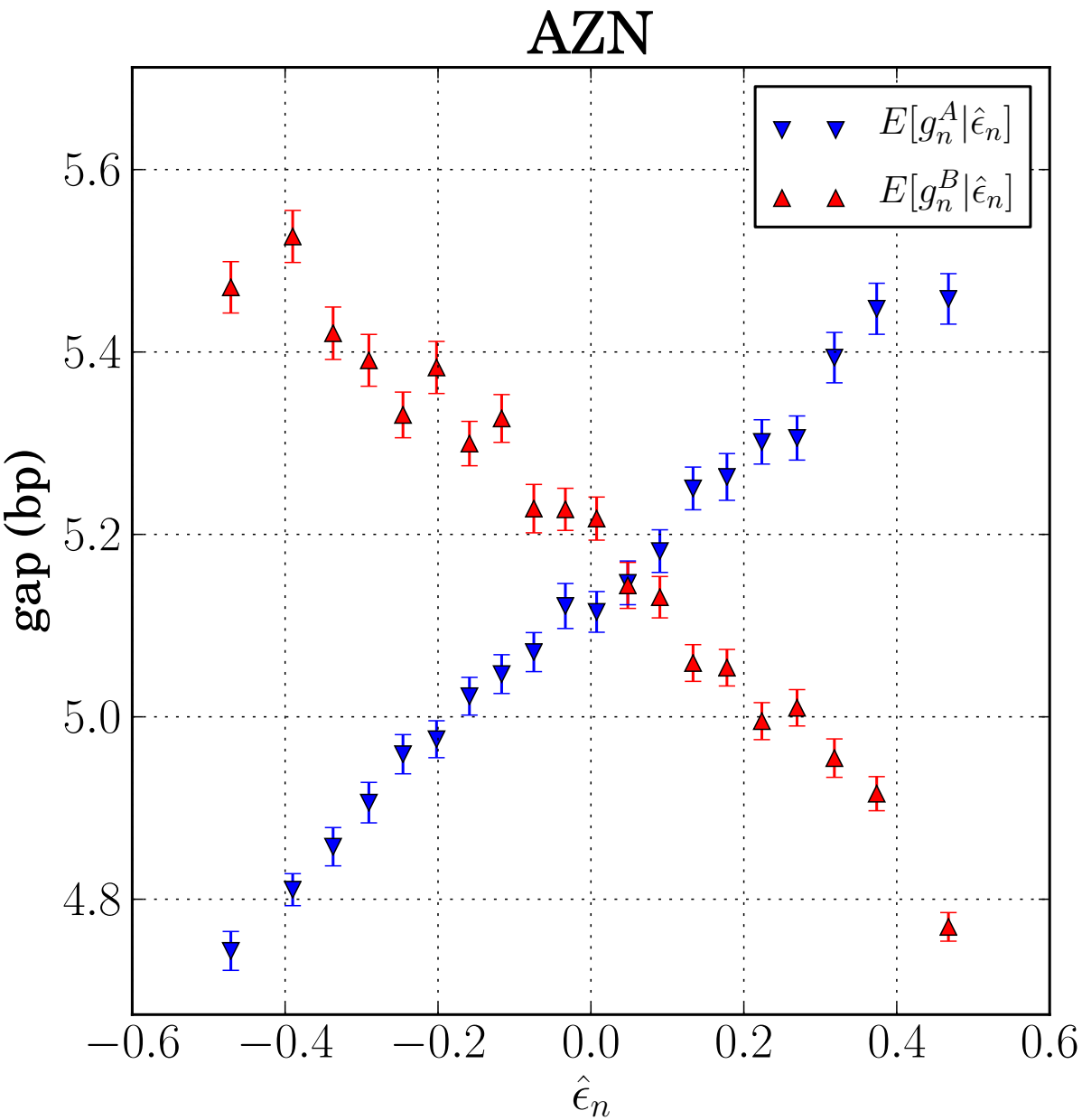
Table: The average stock price is expressed in U.S. Dollars for AAPL and MFST, whereas it is expressed in Great Britain Pounds for AZN and VOD. The average intertrade time and tick size-price ratio are given in seconds and in basis points, respectively.

Volume at best quotes



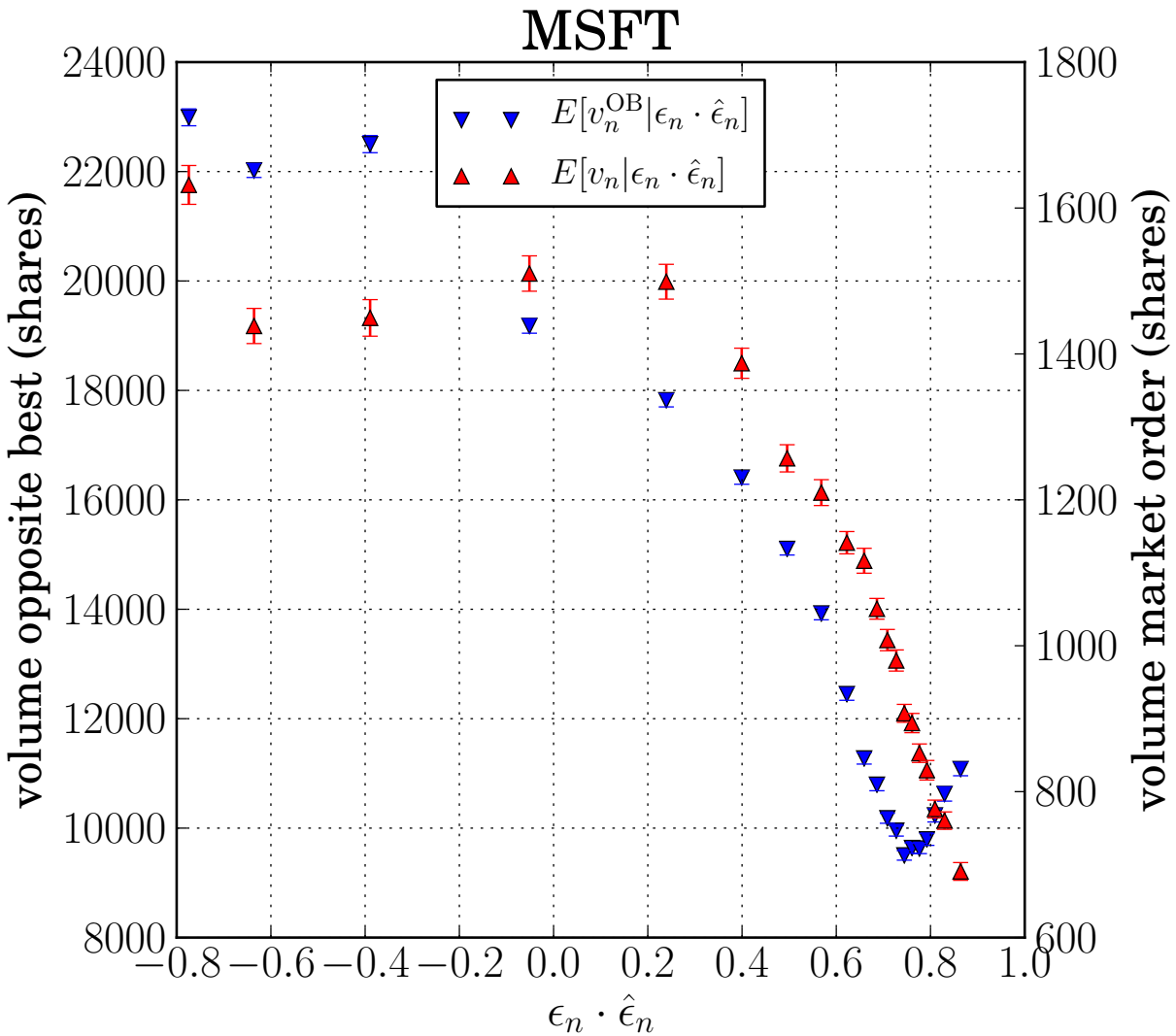
- Conditional best ask volumes $\mathbb{E}[v_n^A|\hat{\epsilon}_n]$ and conditional best bid volumes $\mathbb{E}[v_n^B|\hat{\epsilon}_n]$ as a function of the sign predictor.
- If a buy is more likely, there is more volume at the best bid than at the best ask
- Opposite to asymmetric liquidity
- When predictability is very high, volume at the opposite side increases slightly

Depth: Sparsity of of the book



- Conditional ask gap $\mathbb{E}[g_n^A|\hat{\epsilon}_n]$ and conditional bid gap $\mathbb{E}[g_n^B|\hat{\epsilon}_n]$ as a function of the sign predictor.
- For small tick stocks: if a buy is more likely, the ask side is more sparse
- Opposite to asymmetric liquidity
- For large tick stocks, the book is full on both sides.

Best opposite volume and market order volume

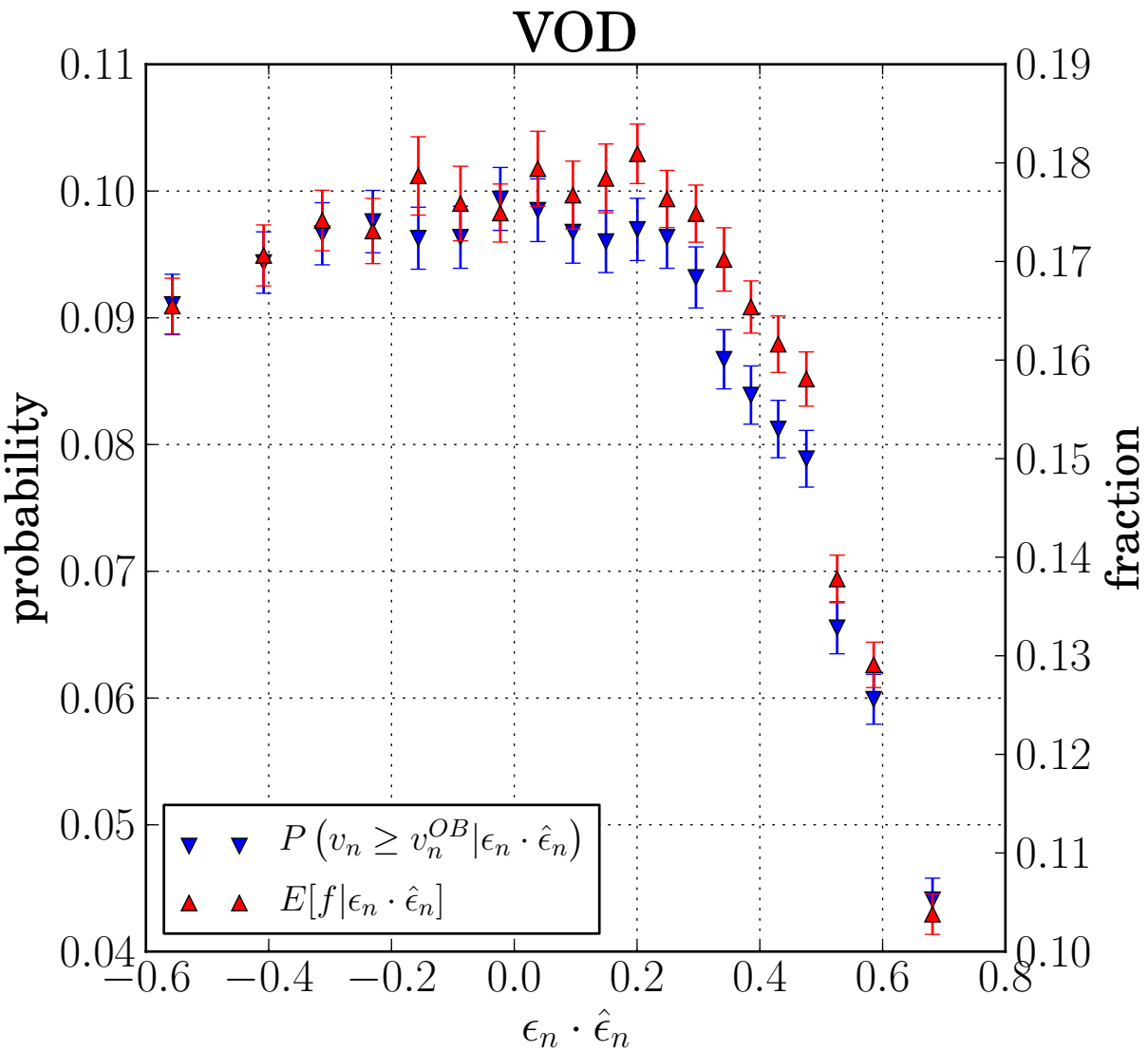


(Neg)Surprise of the trade

$$\epsilon_n \cdot \hat{\epsilon}_n \in [-1, 1]$$

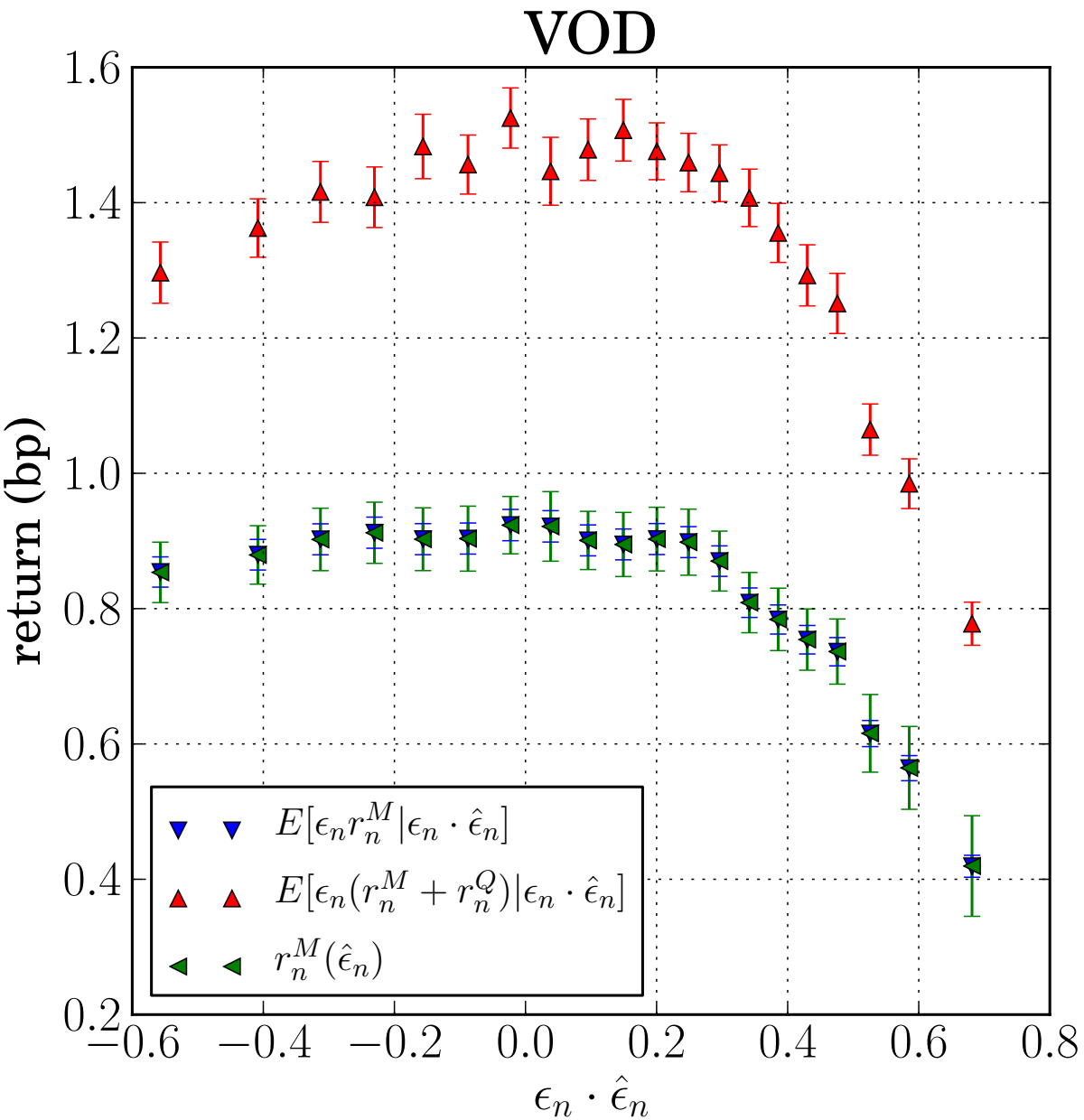
- Conditional best opposite volumes $\mathbb{E}[v_n^{OB}|\epsilon_n \cdot \hat{\epsilon}_n]$ and conditional market order volumes $\mathbb{E}[v_n|\epsilon_n \cdot \hat{\epsilon}_n]$
- If the most likely sign occurs, the volume at the opposite best is small AND the volume of the market order is smaller
- In agreement with asymmetric liquidity

Penetration probability



- Conditional penetration probabilities of the market orders and conditional average ratio between market order volumes and best opposite volumes.
- The most likely market order has a smallest probability of penetrating the best and moving the price
- In agreement with asymmetric liquidity

Mechanical and total impact



- Conditional mechanical impact $\mathbb{E}[\epsilon_n r_n^M | \epsilon_n \cdot \hat{\epsilon}_n]$, the approximate expression $r_n^M(\hat{\epsilon}_n)$, and conditional returns $\mathbb{E}[\epsilon_n (r_n^M + r_n^Q) | \epsilon_n \cdot \hat{\epsilon}_n]$.
- Quote revision in the same direction of mechanical impact
- In agreement with asymmetric liquidity
- Small quote revision for very likely events

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-Hoppé (editors), *Handbook of Financial Markets: Dynamics and Evolution*, 57-160 (2009).
- F. Lillo J. D. Farmer and R. N. Mantegna, Master curve for the price-impact function, *Nature* **421**,129-130 (2003).
- F.Lillo and J.D. Farmer, The long memory of efficient market. *Studies in Nonlinear Dynamics and Econometrics* **8**, 1 (2004).
- F. Lillo, S. Mike e J. Doyne Farmer, Theory for long memory in supply and demand. *Physical Review E* **71**, 066122 (2005).
- 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).
- E. Busseti and F. Lillo, Calibration of optimal execution of financial transactions in the presence of transient market impact, *J. Stat. Mech.* P09010, 2012.
- J. D. Farmer, A. Gerig, F. Lillo, H. Waelbroeck, How efficiency shapes market impact *Quantitative Finance* **13**, 1743-1758 (2013).
- D. E. Taranto, G. Bormetti and F. Lillo, The adaptive nature of liquidity taking in limit order books. *J. Stat. Mech.*, doi:10.1088/1742-5468/2014/06/P06002 (2014)
- B. Toth, I. Palit, F. Lillo, J. D. Farmer, Why is equity order flow so persistent?, Journal of Economic Dynamics & Control 51 (2015) 218-239

Bibliography (II)

- E. Zarinelli, M. Treccani, J.D. Farmer, F. Lillo. Beyond the square root: Evidence for logarithmic dependence of market impact on size and participation rate, *Market Microstructure and Liquidity* **1**, 1550004 (2015)
- G. Curato, J. Gatheral, F. Lillo, Optimal execution with nonlinear transient market impact, Quantitative Finance, (in press 2016)
- G. Curato, J. Gatheral, F. Lillo, Discrete homotopy analysis for optimal trading execution with nonlinear transient market impact, *Communications in Nonlinear Science and Numerical Simulation*, **39**, 332 342 (2016)
- D. E. Taranto, G. Bormetti, J.-P. Bouchaud, F. Lillo, B. Toth, Linear models for the impact of order flow on prices I. Propagators: Transient vs. History Dependent Impact, http://arxiv.org/abs/1602.02735 (2016)
- D. E. Taranto, G. Bormetti, J.-P. Bouchaud, F. Lillo, B. Toth, Linear models for the impact of order flow on prices II. The Mixture Transition Distribution model, http://arxiv.org/abs/1604.07556 (2016)