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 of large trades and optimal execution

Market impact of metaorders: phenomenology

The square-root law of price impact

The average relative price change between the first and the last trade of a metaorder of size Q is well described by the so called "square-root" law:

$$\Delta(Q) = Y\sigma\sqrt{\frac{Q}{V}}$$

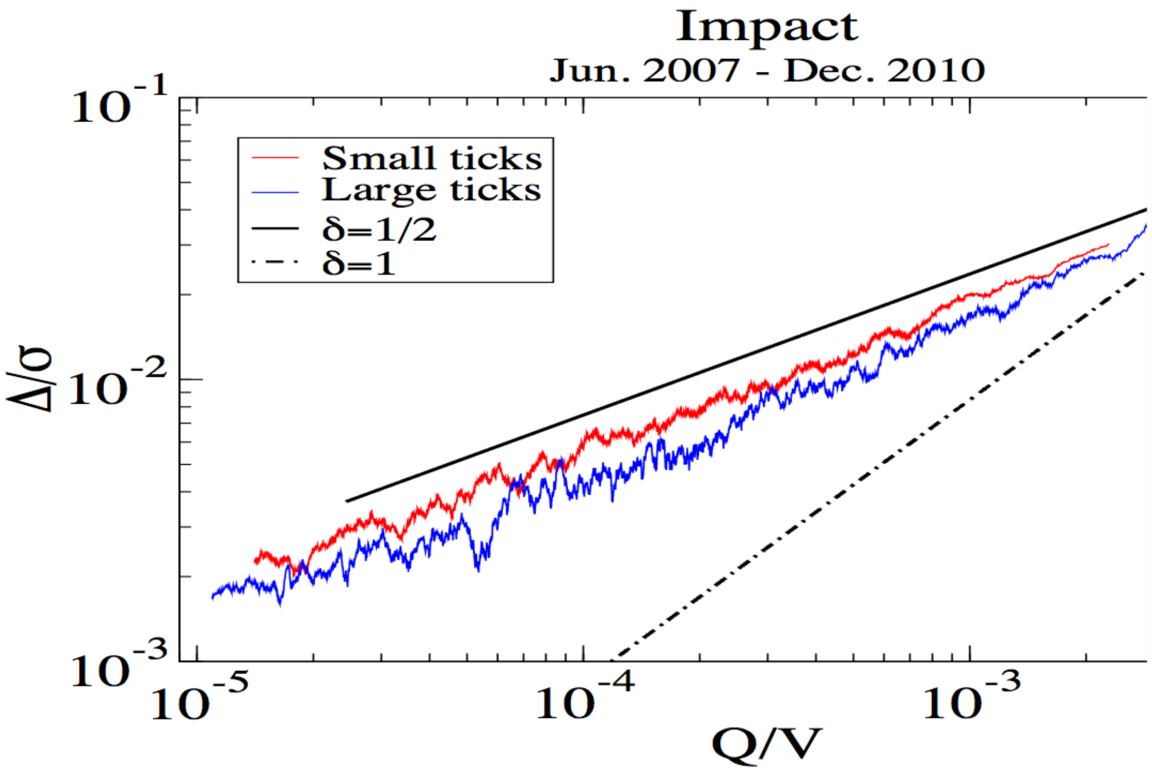


Figure reproduced by:

Bouchaud J.P. et al.

Anomalous Price Impact and the Critical Nature of Liquidity in Financial Markets

Physical Review X 2011

The price impact trajectory during the execution

The price impact trajectory is a concave function of time, i.e. for a given execution size, earlier transactions of the metaorder change the price more than later transactions:

$$R(t/T < 1) \sim C \left(\frac{t}{T}\right)^{\alpha}$$

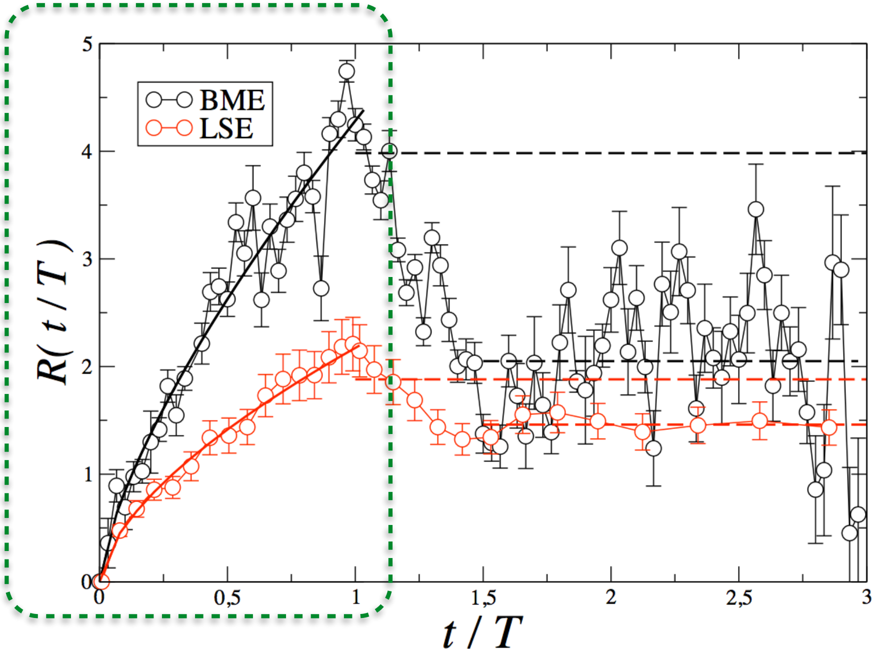


Figure reproduced by:

Moro. at al.

Market Impact and the Trading Profile of Hidden Orders in Stock Markets

Physical Review E 2009

The price impact trajectory after the execution

Several studies indicate that once the metaorder is executed the price impact relaxes from its peak value and converges to a plateau. The reversion indicates that not all the impact is permanent. Even stronger, a recent study suggests that, up to a proper deconvolution of the price impact with respect to the impact of subsequent metaorders and of the the price momentum, the impact relaxes to zero.

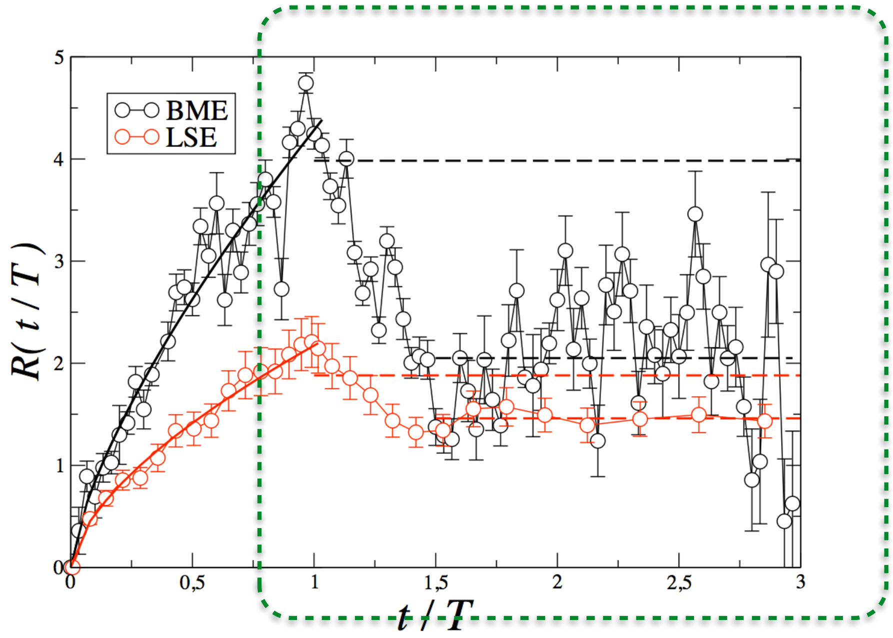


Figure reproduced by:

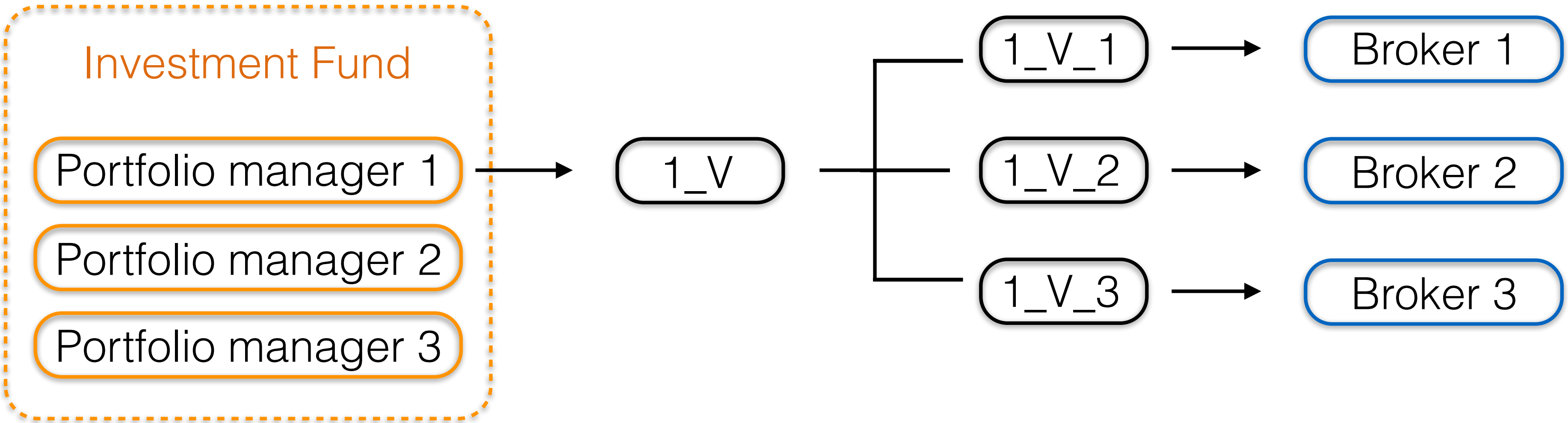
Moro at al.

Market Impact and the Trading Profile of Hidden Orders in Stock Markets

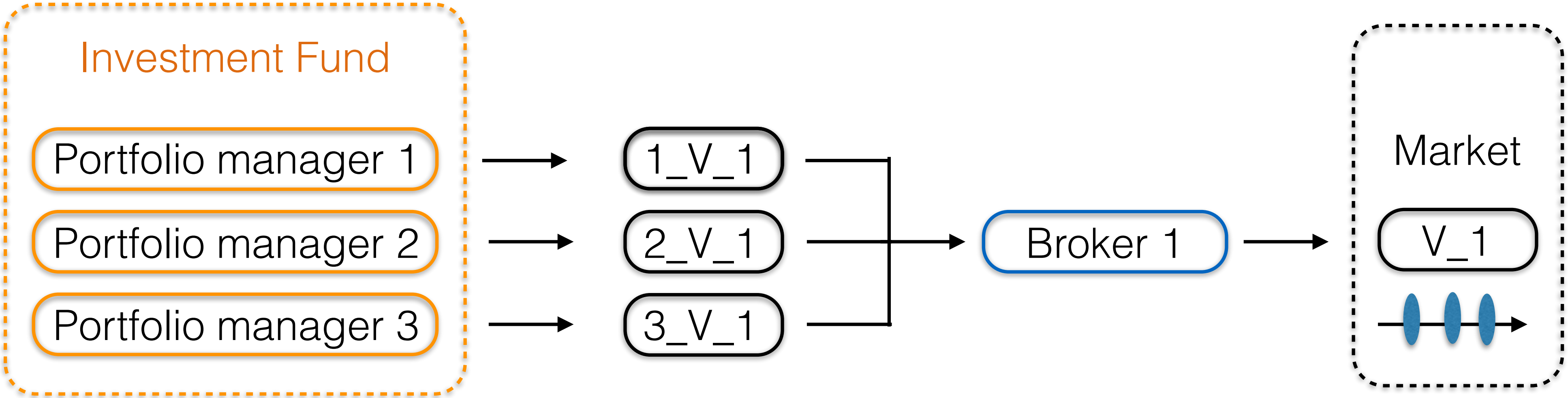
Physical Review E 2009

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:

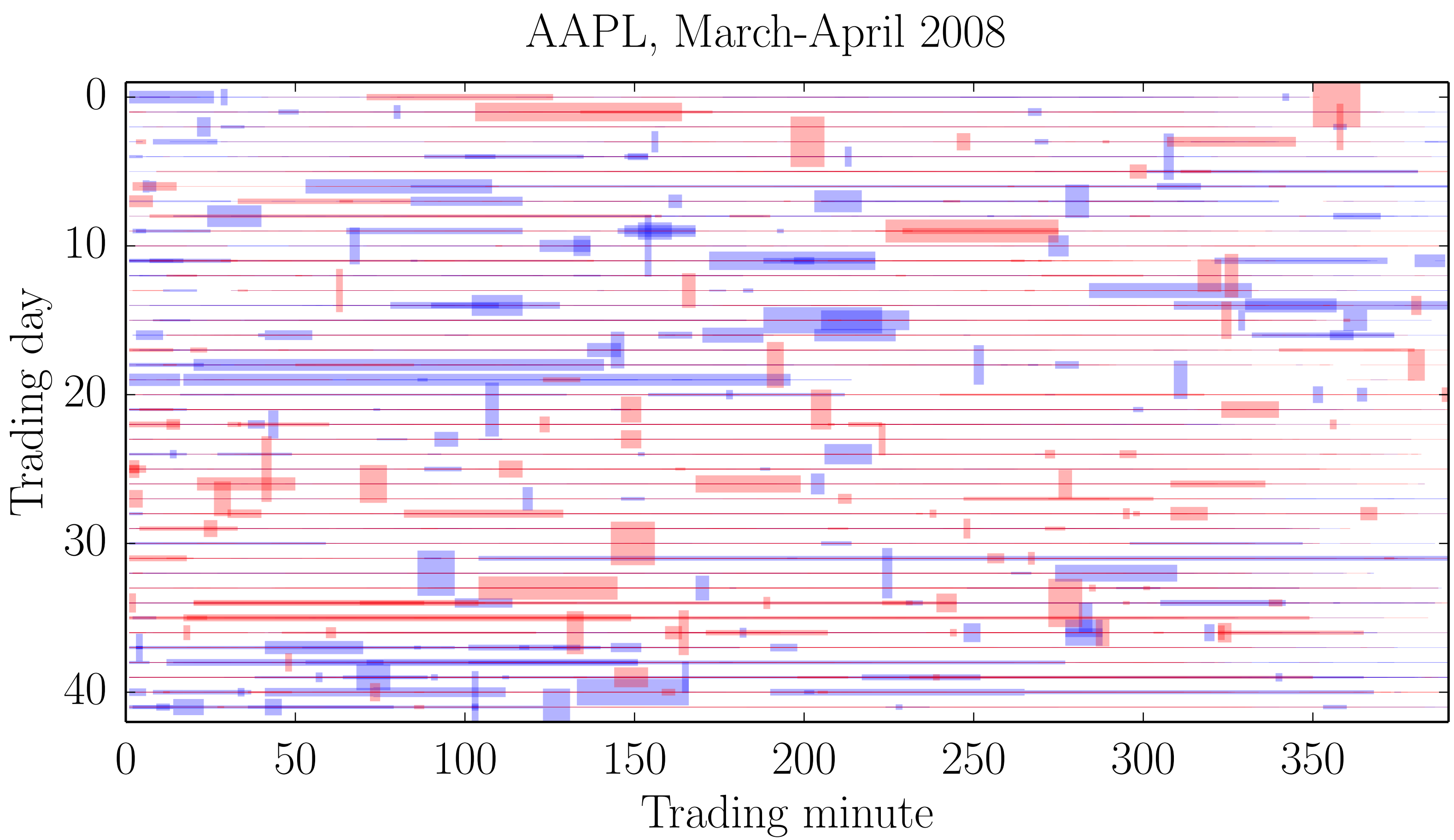
Filter 0	Selecting metaorders traded between January 2007 and December 2009	~ 28,500,000
Filter 1	Selecting metaorders traded on Russell3000 index	~ 23,000,000
Filter 2	Selecting metaorders traded during regular trading section: 09:30 - 16:00	~ 11,000,000
Filter 3	Selecting metaorders with duration longer 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	$\boxed{\rho(v, v_s, v_e)}$

Not available information

$$\pi = \eta F$$

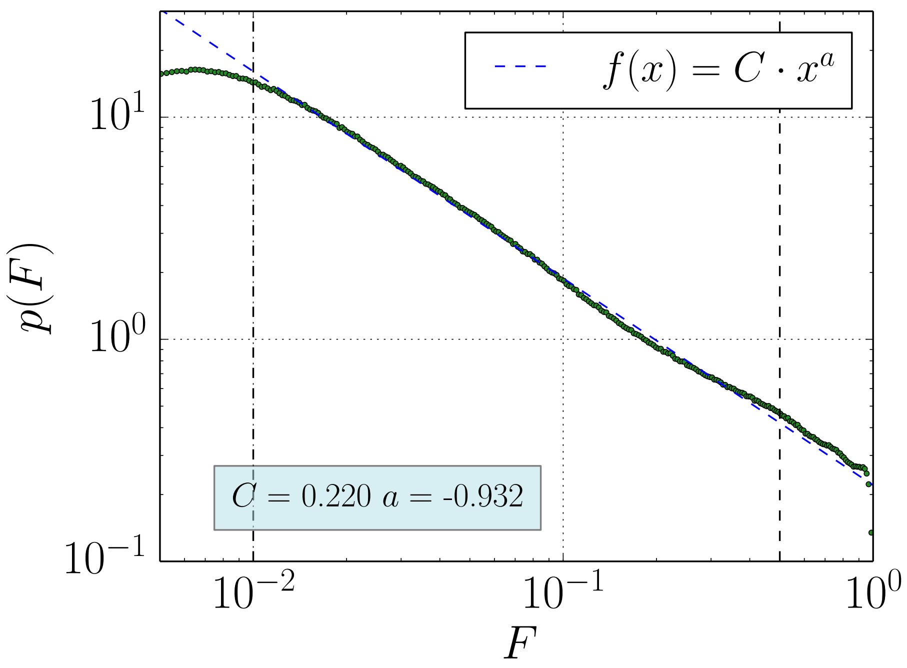
A snapshot of the market



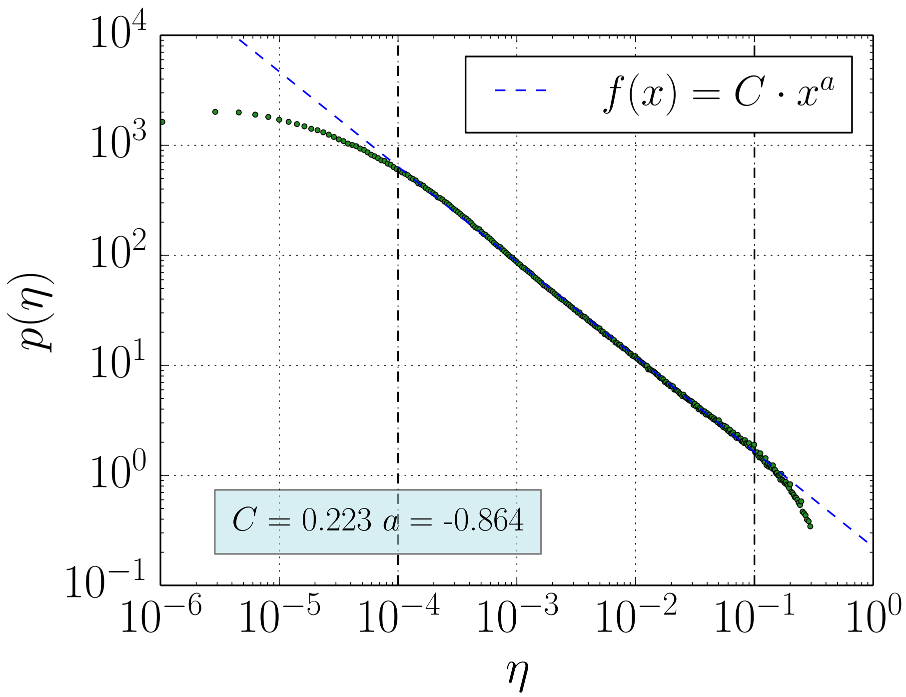
Time series of metaorders active on the market for AAPL in the period March-April 2008. Buy (Sell) metaorders are depicted in blue (red). The thickness of the line is proportional to the metaorder participation rate. More metaorders in the same instant of time give rise to darker colours. Each horizontal line is a trading day. We observe very few blanks, meaning that there is almost always an active metaorder from our database, which is of course only a subset of the number of orders that are active in the market.

Distribution of the describing variables

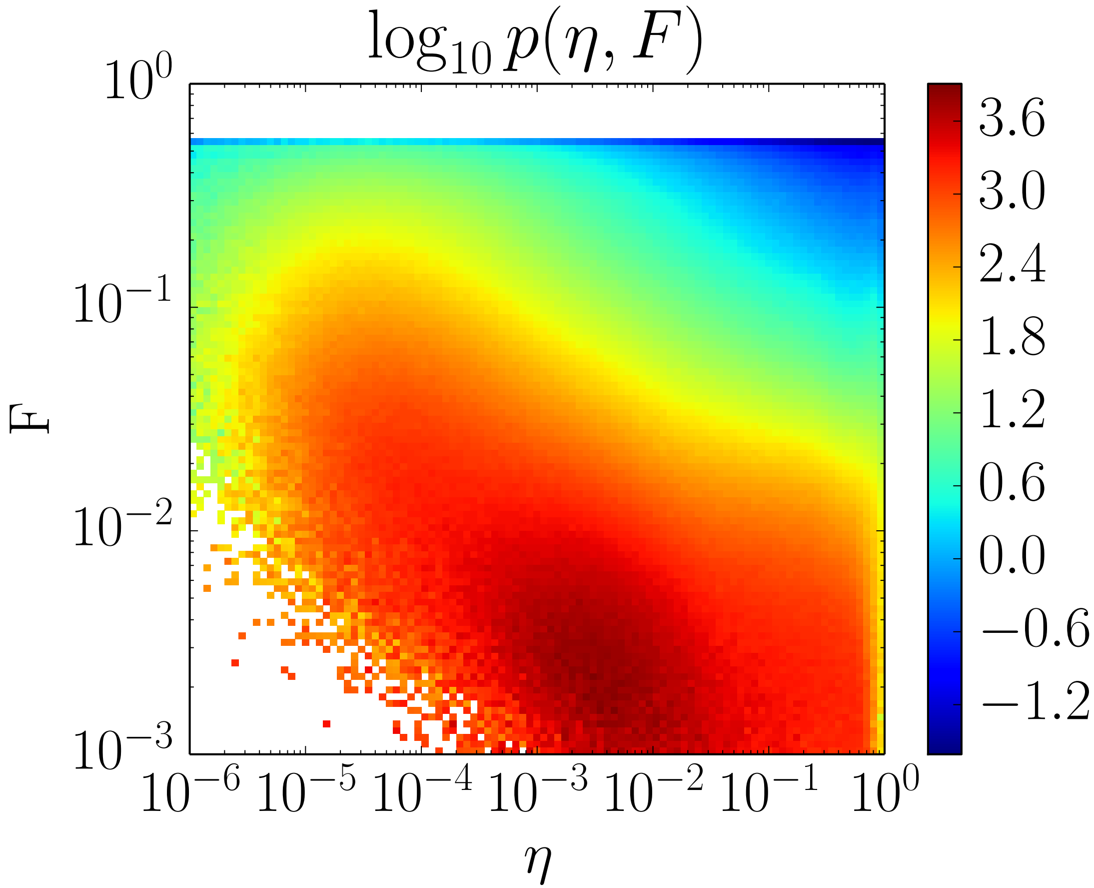
Duration $F := V_P / V_D$



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

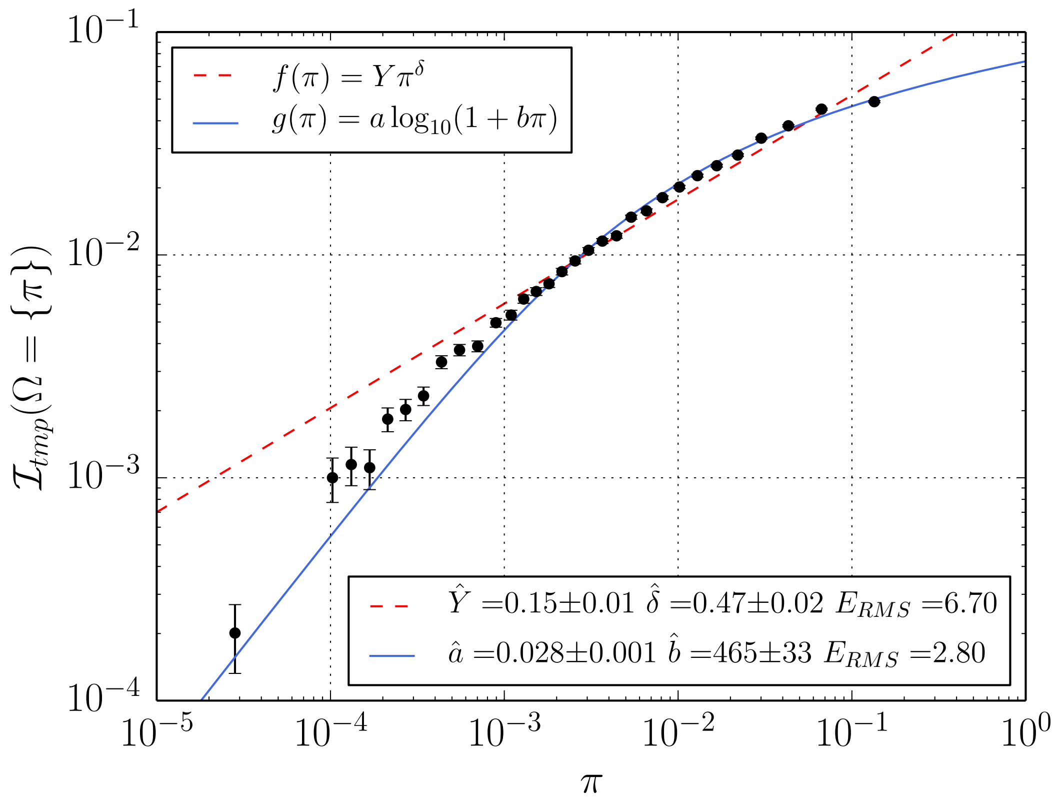


Logarithm of the estimated joint probability density function in double logarithmic scale of the duration F and the participation rate η

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 $\pi := Q/V_D$

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

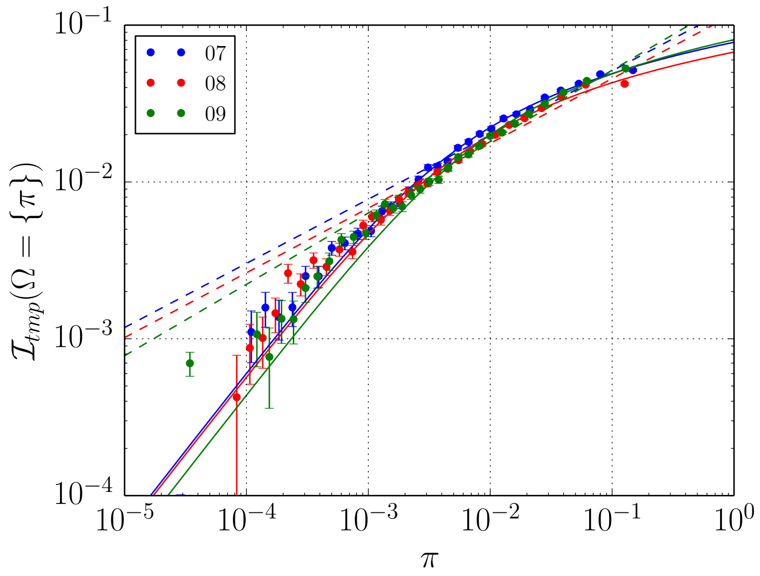


A **square-root** model well describe price impact only in the central region (red curve).

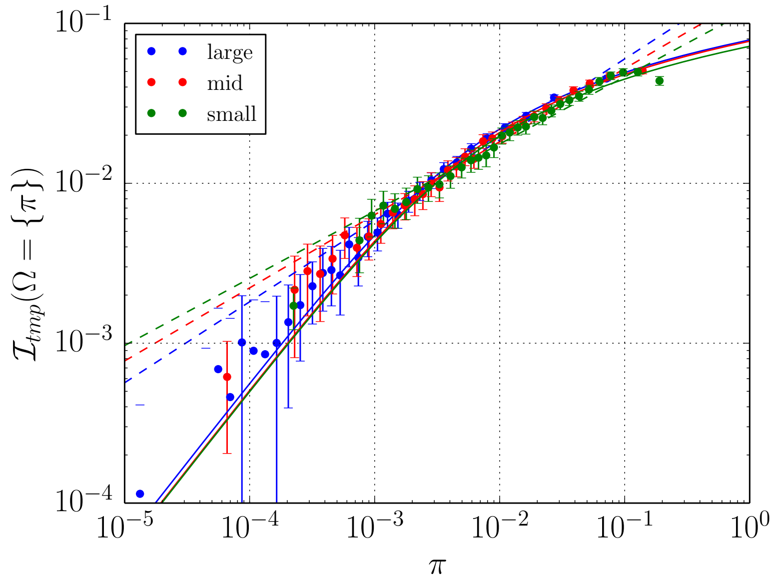
A **logarithmic** (more concave) model allows to capture the **whole shape** of the curve (blue curve).

Further conditioning…

Trading year



Stock market capitalisation



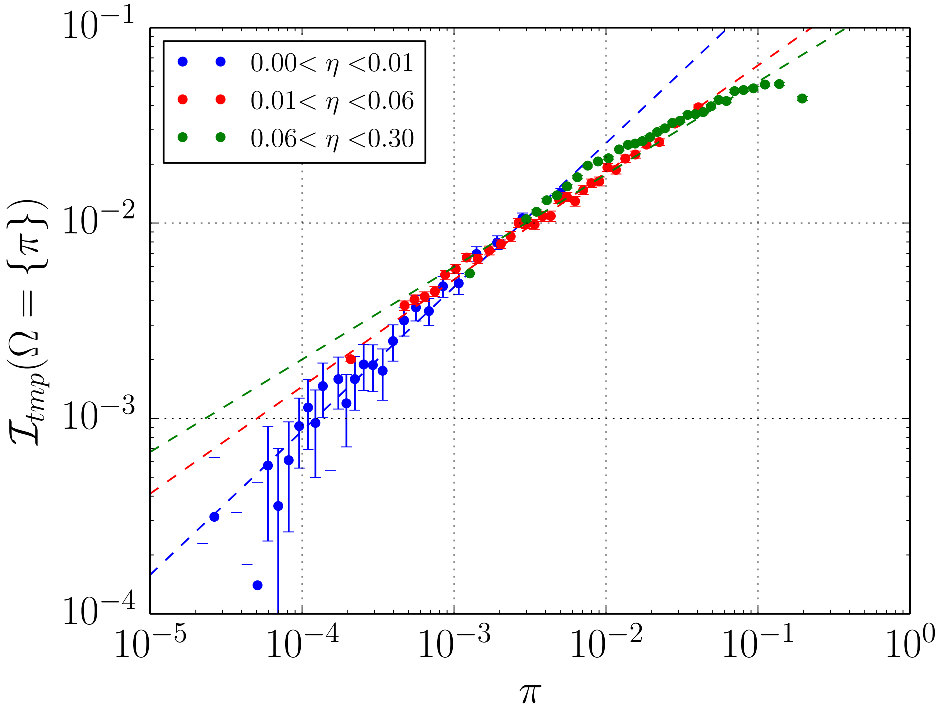
	Year	2007	2008	2009
Power law	$\hat{Y}$	0.13±0.01	0.12±0.01	0.15±0.01
	$\hat{\delta}$	0.41±0.02	0.41±0.02	0.46±0.01
	E_{RMS}	5.47	4.96	3.18
Logarithm	$\hat{a}$	0.029±0.001	0.025±0.001	0.032±0.001
	$\hat{b}$	491±29	547±56	316±25
	E_{RMS}	**1.36**	**2.11**	**2.23**

	Mkt. Cap	Large	Medium	Small
Power Law	$\hat{Y}$	0.19±0.01	0.15±0.01	0.12±0.01
	$\hat{\delta}$	0.51±0.02	0.46±0.02	0.42±0.02
	E_{RMS}	1.38	1.32	**0.69**
Logarithm	$\hat{a}$	0.030±0.001	0.030±0.001	0.027±0.001
	$\hat{b}$	441±21	400±35	428±62
	E_{RMS}	**0.40**	**0.69**	1.03

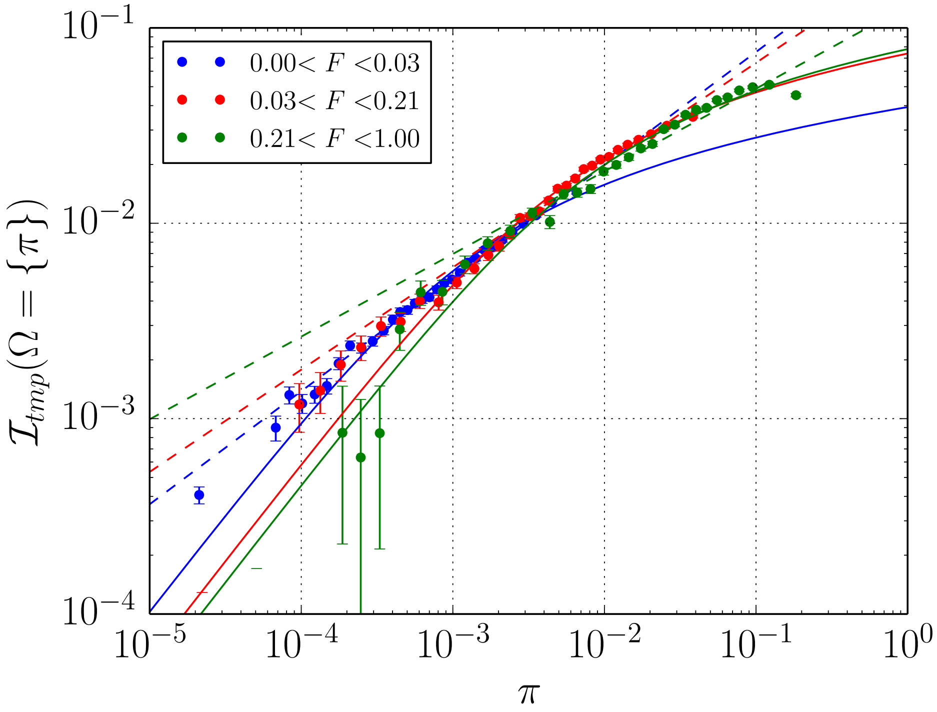
The price impact curve is **quite stable** and, with the exception of the small capitalisation conditioning, the logarithmic function **always better** explains the data

Further conditioning…

Participation rate



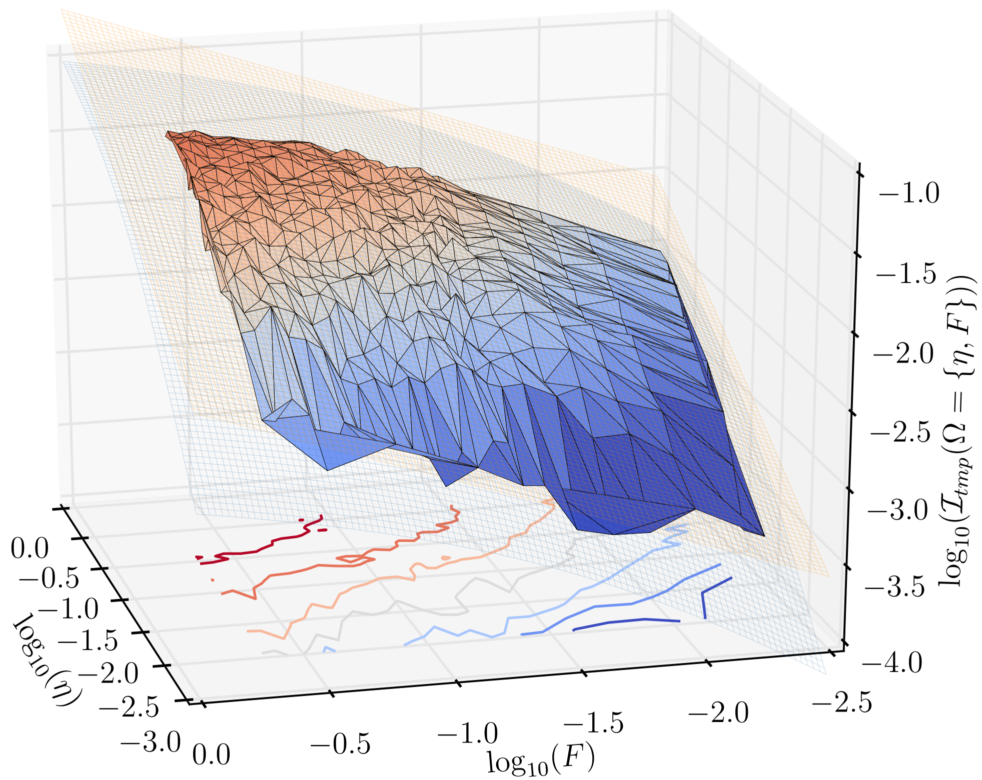
Metaorder duration



The price impact surface

Price impact surface: the average relative price change between the **end** and the **beginning** of the execution, conditioning on the participation rate $\eta := Q/V_P$ and the duration $F := V_P/V_D$

$$\mathcal{I}(\eta, F) := \mathbb{E}\left[\epsilon\left(s(v_e) - s(v_s)\right)\middle|\, \eta, F\right]$$



A **double logarithmic** function (blue surface) better describes the data compared with a double power-law functions (yellow surface)

It is possible to recover the price impact curve by averaging on regions such that $\pi = \eta F$ is constant (diagonal lines in the double logarithmic scale)

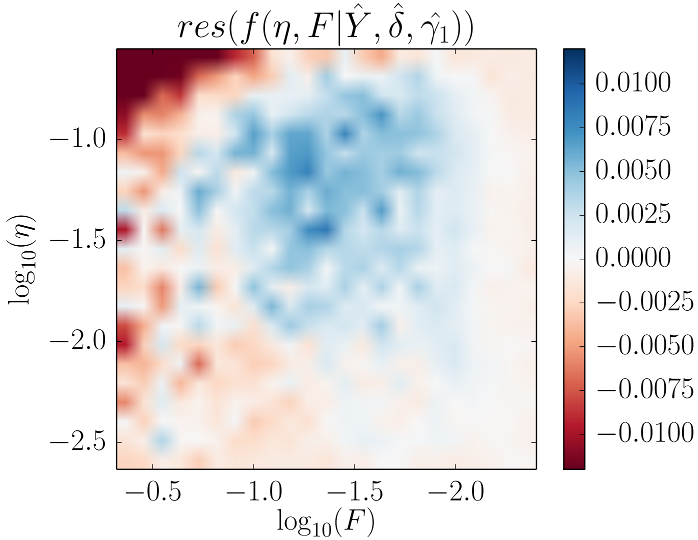
Analysis of the residuals

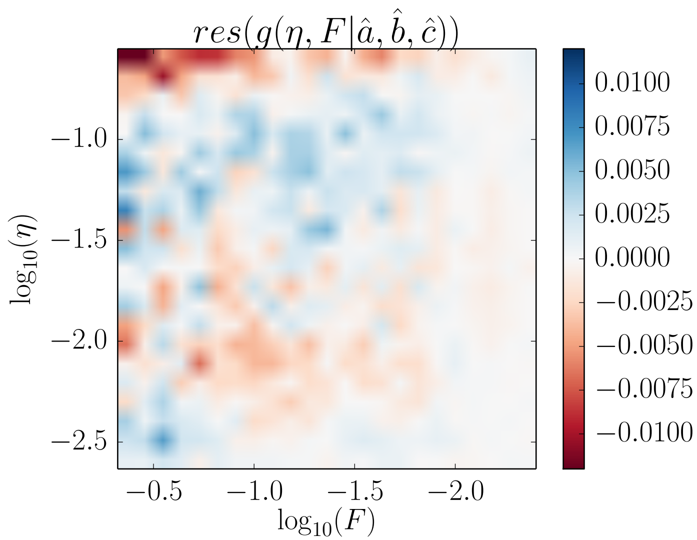
Price impact models:

$$f(\eta, F|Y, \delta, \gamma_1) = Y \cdot \eta^{\delta} \cdot F^{\gamma_1}$$

$$g(\eta, F|a, b, c) = a \cdot \log_{10}(1 + b\eta) \cdot \log_{10}(1 + cF)$$

Residuals of the fitted models:





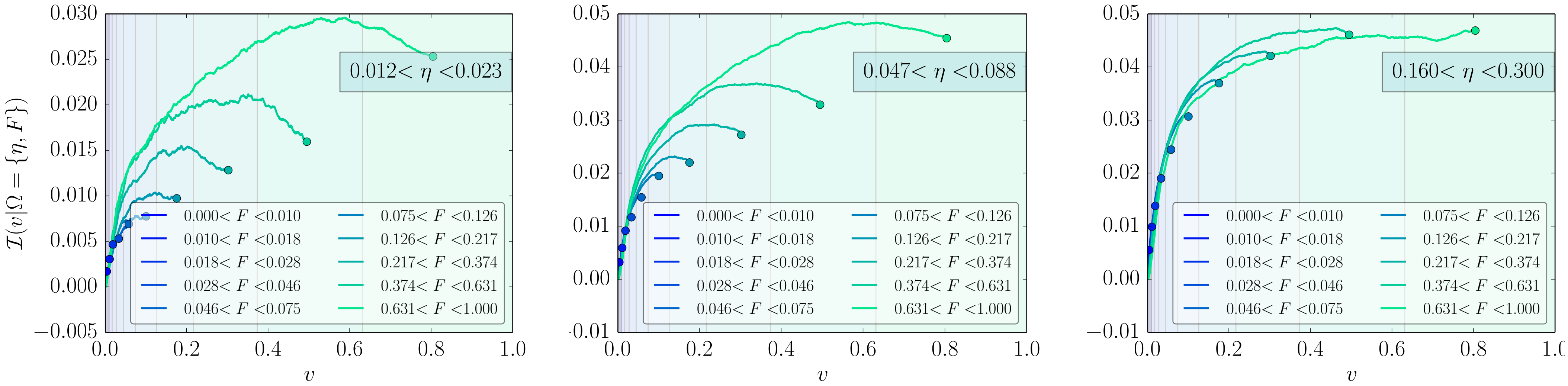
Double power-law: a clear non-random **pattern** emerges: positive residuals in the centre, negative residuals in the periphery

Double logarithm: residuals are **evenly distributed**

The price impact trajectory - during the execution

We fix the participation rate and the duration of metaorders, and we follow the price impact trajectory during the execution:

$$\mathcal{I}(v|\eta, F) := \mathbb{E}\left[\epsilon\left(\widetilde{S}(v) - \widetilde{S}(v_s)\right)\Big|\,\eta, F\right]$$



The price impact trajectories (lines) **deviate** from the price impact curve/ surface (circles)

The price impact trajectories **revert** during the execution of the metaorder

Price dynamics during execution: Almgren-Chriss

Continuous-time stock price model for a trader who impacts the price of the asset in linear permanent way.

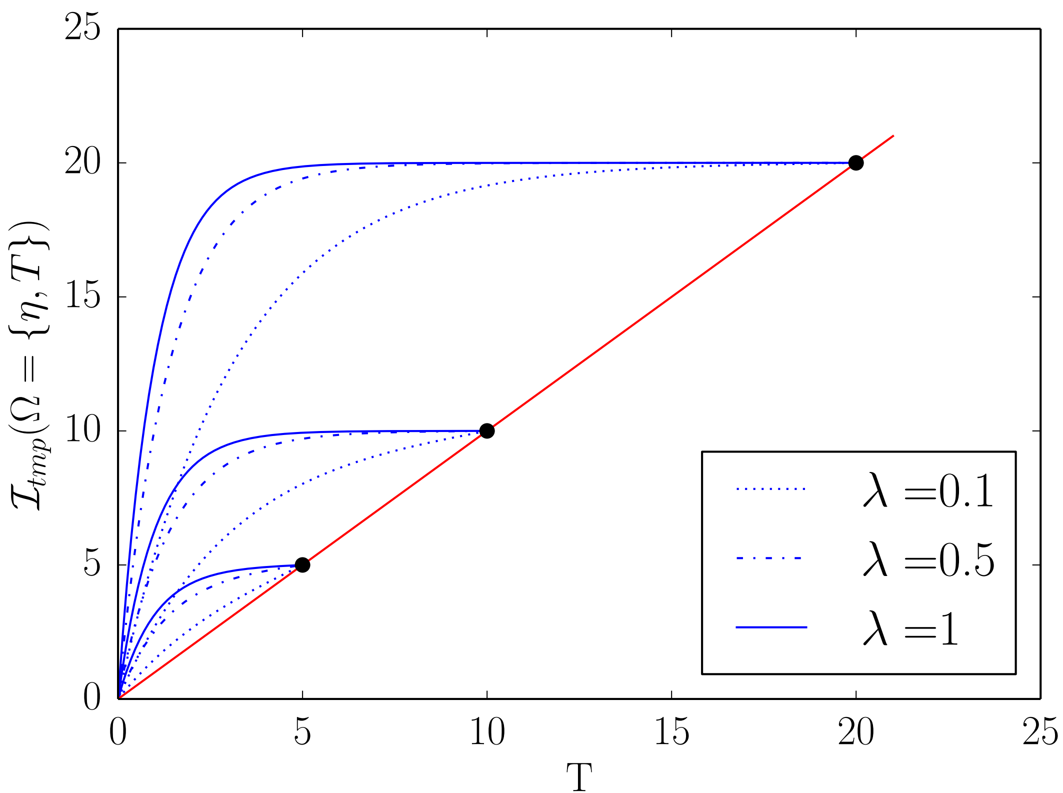
Trading occurs at a rate of q(t) shares per unit time

$$S(t) = S(0) + a\int_0^t q(s)ds + \sigma\int_0^t \mathrm{d}W_s$$

λ is the risk aversion parameter

$$q(t) = Q\frac{\sinh k(T-t)}{\sinh kT}$$

$$k = \sqrt{\lambda\sigma^2/a}$$



Price dynamics during execution: TIM (propagator)

Continuous-time stock price model for a trader who can move the price of the asset. As long as the trader is not active, the asset price is determined by the other market participants and follows a brownian motion.

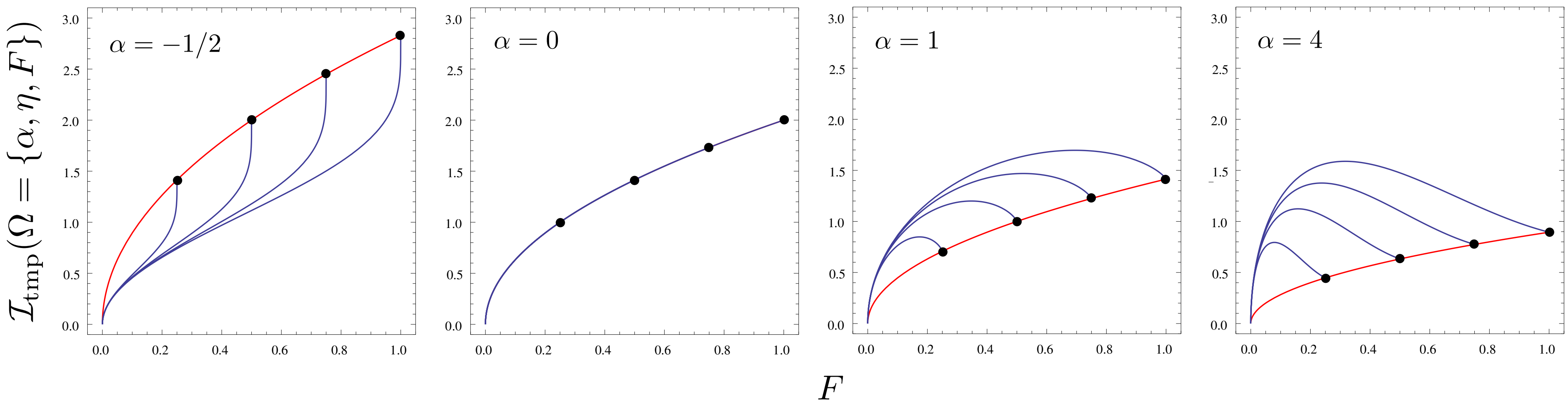
$$S(t) = S(0) + \int_0^t f(q(s))G(t-s)\mathrm{d}s + \int_0^t \sigma(s)\mathrm{d}W_s$$

$$q(s) = \frac{Q}{V(t_c)} \frac{(\alpha+1)}{D^{\alpha+1}}(D-s)^{\alpha}$$

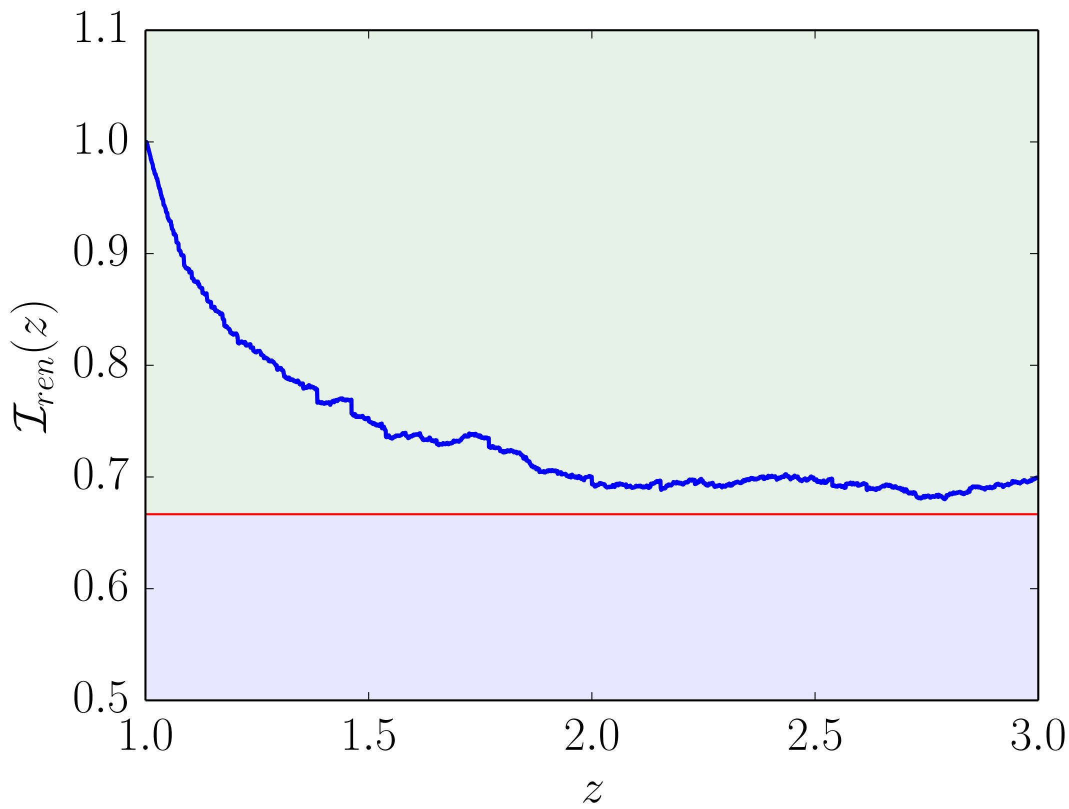
$f(q) = sign(q)|q|^{\delta}$ Single-trade impact function

$G(s) \sim s^{-\gamma}$ Memory kernel

α Measure of front loading: α = 0 is a VWAP; >0 (<0) trade more at the beginning (end) of the execution



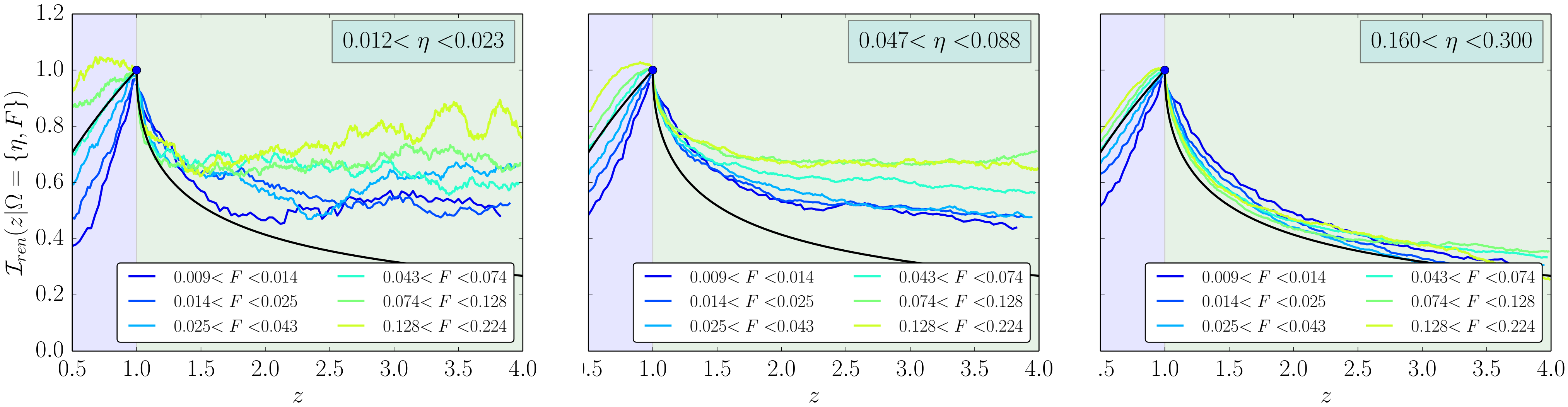
Market impact decay: unconditional results



Decay of temporary market impact after the execution of the meteorder. We follow the normalised market impact path as a function of the time rescaled by the metaorder duration. The red horizontal line corresponds to 2/3, as predicted by the model of Farmer et al. (2013)

The price impact trajectory - after the execution

Decay of temporary price impact after the execution of the meteorder. We follow the normalised price impact as a function of the rescaled variable z = v/F.



For **small participation rates**, the price impact trajectories of longer metaorders relax to levels which are higher than those of shorter metaorders.

For **large participation rates**, the price impact trajectories superimpose quite well one with each other. They relax very slowly and we do not observe any flattening of the curve. Quite interestingly, in this regime the price impact trajectories are well described by the prediction of the propagator model (black curve).

The role of metaorder sign autocorrelation

The picture emerging from the previous analysis can be partly clarified by taking into account the **autocorrelation of the sign of metaorders**. If different metaorders executed consecutively or in the same time period typically have the same sign, we expect that the effect of this correlation is to keep the price impact of a single metaorder relaxation **artificially high**. Moreover the effect of autocorrelation is

- **stronger** for longer metaorders, since the probability of overlapping with other metaorders is larger, and for lower participation rates, since their effect on price can be considerably perturbed by metaorders with larger participation rates.
- **milder** for shorter metaorders, because of the lower probability of overlap, and larger participation rates, because the effect of metaorders with lower participation rates on price becomes negligible.

duration (mins)	number	a.n.o.m.	fraction concordant	fraction discordant
0 - 10	368484	1.7	0.548	0.452
10 - 25	117756	3.0	0.553	0.447
25 - 50	71031	4.6	0.553	0.447
50 - 100	52931	6.6	0.547	0.453
100 - 200	43884	8.0	0.546	0.454
200 - 390	49411	9.3	0.543	0.457
0 - 390	703497	3.54	0.548	0.452

Average fraction of overlapping metaorders with the same sign

Introduction to optimal execution

Introduction to optimal execution: a multiscale problem

- An investor wants to trade (buy or sell) a given number of shares and wants to minimize cost by trading incrementally.
- A three scale decomposition
  - First, the portfolio manager decides how to split the order across the different days.
  - Then, for each day the trader divides the day in "macroscopic" intervals, for example 5 or 15 minutes, and decides how much to trade in each of these intervals.
  - Lastly, one has to decide how to trade in each interval, specifying the type of orders used (e.g. limit versus market orders) and the strategy to follow (for example, when to cross the spread if the price moves in an adverse direction).
- We focus here on the second level of optimization.
  - Discrete time (with data calibration)
  - Continuous time

Discrete time: Objective function

- An investor has X shares to trade in N time periods. Let v_k $(k = 1, ..., N)$ be the (signed) number of shares to be traded in interval k. Let P_k be the price at which the investor trades at interval k and P_0 the price before the start of the execution

- A very used objective function is the *implementation shortfall* defined as

$$C(\mathbf{v}) \equiv \sum_{k=1}^{N} v_k \tilde{p}_k - X p_0 \tag{1}$$

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

- The implementation shortfall is in general a stochastic variable, therefore one often wants to minimize $E[C(\mathbf{v})]$. This assumes a risk neutral profile.

Almgren and Chriss assumed that the price of the stock at step k is equal to the previous price plus a linear market impact term and a random shock

$$p_k = p_{k-1} + \theta v_k + \eta_k \qquad \eta \sim \mathrm{IID}(0, \sigma). \tag{2}$$

They consider the effective price $\tilde{p}_k$ payed as different from the average price p_k in the interval and they model it as

$$\tilde{p}_k = p_k + \rho v_k + \mathrm{sign}(v_k) \cdot S/2, \tag{3}$$

where $\mathrm{sign}(v_k) \cdot S/2$ is the contribution from the bid ask spread S and ρv_k represents a linear *temporary impact*.

The temporary impact accounts for the resilience of the limit orders in the book, which relaxes back to the steady state after a trade-induced price movement.

Almgren and Chriss (2)

The equation for the execution costs becomes

$$C(\mathbf{v}) = \sum_{k=1}^{N} v_k \tilde{p}_k - X p_0 = \sum_{k=1}^{N} (\eta_k + \theta v_k) \sum_{j=1}^{k} v_j + \sum_{k=1}^{N} \left(\text{sign}(v_k) S/2 + \rho v_k\right) v_k \tag{4}$$

and the expected value of the costs to be minimized is

$$E\left[C(\mathbf{v})\right] = \frac{\theta}{2} X^2 + (\rho + \theta) \sum_{k=1}^{N} v_k^2 + \theta \sum_{i \neq j} v_i v_j + S/2 \sum_{k=1}^{N} |v_k|, \tag{5}$$

under the constraint $\sum_{k=1}^{N} v_k = X$

If one assumes that all v_k have the sign of X, the last term becomes constant and equal to $XS/2$.

Almgren and Chriss (3)

Given the symmetry, the solution that minimizes the expected impact costs is

$$\mathbf{v}^* \equiv \arg\min_{\mathbf{v}} E\left[C(\mathbf{v})\right] = \left(\frac{X}{N}, \frac{X}{N}, \ldots, \frac{X}{N}\right)^T. \tag{6}$$

showing that the solution simply consists in trading at a constant rate over the periods.

If the price has a drift

$$p_k = p_{k-1} + \mu + \theta v_k + \eta_k \qquad \eta \sim \mathrm{IID}(0, \sigma). \tag{7}$$

Minimization of implementation shortfall gives

$$v_k^* \propto X\left(\frac{1}{N} + \frac{(N+1) - 2k}{2(2\rho + \theta)}\mu\right) \tag{8}$$

With positive drift, we will accelerate a buy order. The amount of the acceleration depends positively on μ, of course, but it also depends inversely on ρ and θ.

Almgren and Chriss (4)

The second innovation in Almgren and Chriss is that they consider risk aversion for optimal execution: they minimize the sum of expected cost and costs' risk. By mimicking the theory on portfolio optimization (Markowitz mean-variance) , Almgren and Chriss consider as optimal trading schedule the solution of

$$\arg\min_{\mathbf{v}} \left(E[C(\mathbf{v})] + \lambda Var[C(\mathbf{v})] \right). \tag{9}$$

where λ is the coefficient of risk aversion. The higher is the λ, the more important is risk with respect to cost. A risk neutral investor corresponds to $\lambda = 0$. The set of solutions to this problem for different values of λ is called *optimal frontier*.

The variance of the execution cost is

$$Var\left[C(\mathbf{v})\right] = E\left[(C(\mathbf{v}) - E[C(\mathbf{v})])^2\right] = E\left[\left(\sum_{k=0}^{N-1} \eta_k \sum_{j=k+1}^{N-1} v_j\right)^2\right], \tag{10}$$

We assumed the η_k are independent, so we have $E[\eta_i \eta_j] = 0$ for $i \neq j$ and thus:

$$Var\left[C(\mathbf{v})\right] = \sigma^2 \sum_{k=0}^{N-1} (\sum_{j=k+1}^{N-1} v_j)^2. \tag{11}$$

Almgren and Chriss (6)

By using the impact model of Eqs. 2 and 3 one obtains

$$v_k = A\cosh(\beta(N-k)), \tag{12}$$

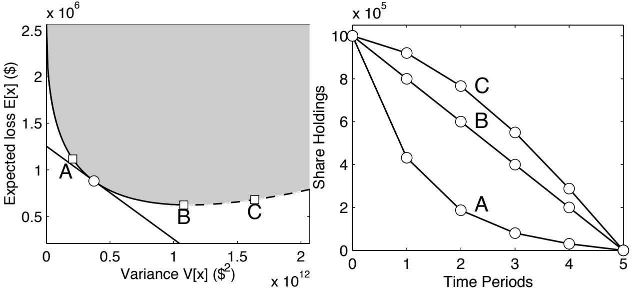
where A is a normalization constant and β solves the equation[1]

$$2[\cosh(\beta) - 1] = \frac{\lambda\sigma^2}{\rho+\theta}. \tag{13}$$

By inverting this equation expressing β as a function of σ and by taking the continuous time limit (i.e. $\sigma \to 0$) we have $\beta \simeq \sqrt{\frac{\lambda\sigma^2}{\rho+\theta}}$.

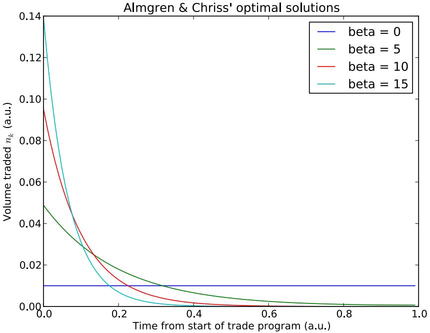
[1] The equation found in the original Almgren-Chriss paper is slightly different because they define the temporary impact and the variance as proportional to the interval length $\tau = T/N$.

Almgren and Chriss (8)



Right. The efficient frontier. Shaded region is attainable by some strategy. Solid curve is the efficient frontier. Dashed line is a strategy with higher variance but same expected cost. Point B is the naive strategy of slice and dice (Bertismas and Lo)
Left. Optimal trajectories $X - \sum_{i=1}^{k} v_i^*$ of the amount of shares still hold at time k.
In both panels we consider $\lambda > 0$ (Point A, risk averse), $\lambda = 0$ (Point B, risk neutral), and $\lambda < 0$ (Point C, risk lover)
From Almgren and Chriss 2001

Almgren and Chriss (8)



The solution of Eq. 12 shows that the more risk averse is the investor (i.e. the higher is β), the more front loaded is the trading schedule. This means that more volume is traded at the beginning of the execution in order to minimize the uncertainty on execution price of the last part of the trading schedule.

Optimal execution with transient market impact

Market impact with transient impact

- Almgren and Chriss assume a market impact which is *linear, fixed, and permanent*.
- We have seen in the previous lectures that, due to the correlation of the order flow, market impact is transient, i.e. *the past order flow affects future price impacts*.
- One way of capturing this effect is through the transient impact model (TIM)
- TIM assumes that

$$p_n = p_{-\infty} + \sum_{k=1}^{\infty} f(v_{n-k})G(k) + \sum_{k} \eta_k \tag{14}$$

where v_n is the *signed order flow*. Thus

$$p_{n+1} - p_n = G(1)f(v_n) + \sum_{k=1}^{\infty}[G(k+1) - G(k)]f(v_{n-k}) + \eta_n \tag{15}$$

- The decay of G is such that prices are diffusive (or approximately efficient) given the correlated order flow
- Efficiency leads to

$$p_{n+1} - p_n = K(v_n - \hat{v}_n) + \eta_n \tag{16}$$

where $\hat{v}_n$ is the best predictor of v_n given $\mathcal{F}_n$.

The propagator model in real time

- We consider 5 minute intervals, $t_0 = 8{:}00$, $t_1 = 8{:}05$, $t_2 = 8{:}10$, ...
- p_n is the log mid price right before time t_n. We define the series of aggregated volumes v_n in terms of the volumes v_i^{tt} of the single transactions, i.e.

$$v_n = \sum_{[t_n, t_{n+1}]} v_i^{tt}. \tag{17}$$

- We consider the *normalized* volume imbalance v_n^{nor} as

$$v_n^{nor} = \frac{\sum_{[t_n, t_{n+1}]} v_i^{tt}}{\sum_{[t_n, t_{n+1}]} |v_i^{tt}|} \tag{18}$$

- The impact function $f(v^{nor})$ of the normalized volume imbalance is

$$f(v^{nor}) = E[r_n | v_n^{nor}]. \tag{19}$$

- The propagator model in real time becomes

$$r_j \equiv p_{k+1} - p_k = \sum_{k=0}^{j-1} \mathcal{G}(k) f(v_{j-k}^{nor}) + \eta_j. \tag{20}$$

where we defined $\mathcal{G}(k) \equiv G(k+1) - G(k)$, and $G(0) = 0$.

Market impact in real time

We compute the impact dependence on normalized volume

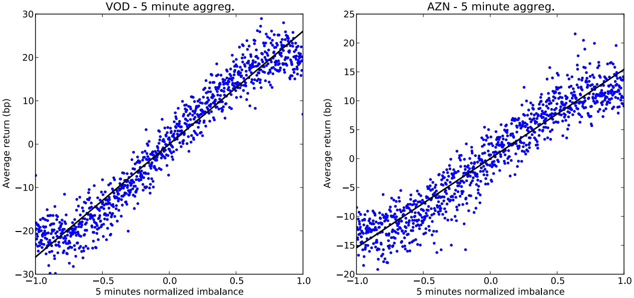
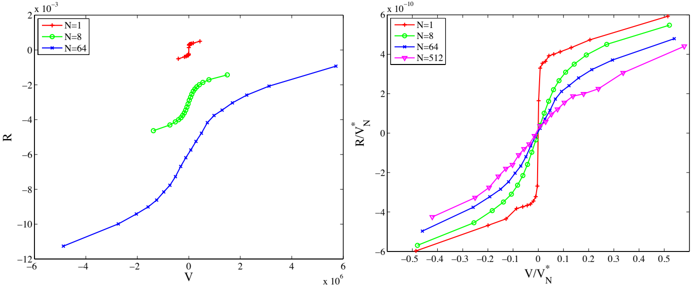


Figure: Impact (top) and propagator (bottom) from 5 min imbalance data

At this level of aggregation, impact is roughly linear

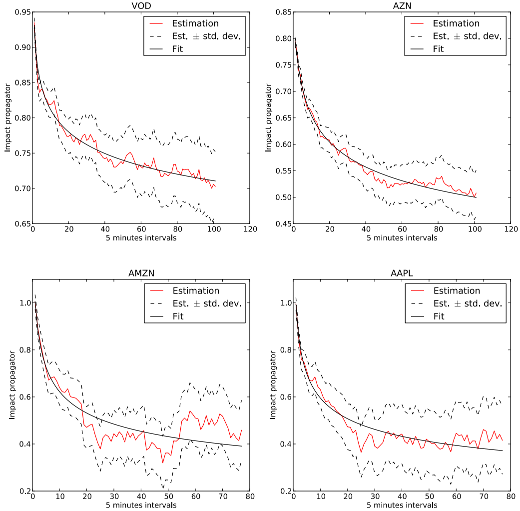
Market impact at different aggregations (# of trades)



Market impact is a strongly concave function of volume at short scales, but becomes progressively more linear on longer scales (Bouchaud, Farmer, Lillo, 2009)

Propagator in aggregated real time

We computed the propagator $G(k)$ over 5 min intervals by linear regression



The TIM fits data quite well also on aggregated (time or trades) data.

- The effective log-midprice $\tilde{p}_k$ is the logarithm of the average mid-price at which we trade the shares v_k between time t_k and time t_{k+1}. We assume that

$$\tilde{p}_k = \frac{p_k + p_{k+1}}{2}. \tag{21}$$

- The equation that describes the dynamics of effective price is therefore

$$\tilde{p}_n = p_0 + \sum_{k=0}^{n} [\eta_k + f(v_k)\tilde{G}(n-k)] \tag{22}$$

where we defined the effective propagator $\tilde{G}_0$ as

$$\tilde{G}(0) = \frac{G(1)}{2}, \quad \tilde{G}(1) = \frac{G(1)+G(2)}{2}, \quad \tilde{G}(2) = \frac{G(2)+G(3)}{2}, \quad \ldots \tag{23}$$

- We define the logarithmic transaction costs $c(\mathbf{v})$ as

$$c(\mathbf{v}) \equiv \sum_{k=0}^{N-1} v_k(\tilde{p}_k - p_0) \approx \sum_{k=0}^{N-1} v_k \log\left(\frac{\tilde{P}_k}{P_0}\right) \simeq \sum_{k=0}^{N-1} v_k \left(\frac{\tilde{P}_k - P_0}{P_0}\right) = \frac{C(\mathbf{v})}{P_0}. \tag{24}$$

Optimal execution with TIM

- The expected implementation shortfall is

$$E[C(\mathbf{v})] = \sum_{n=1}^{N} v_n [\sum_{k=1}^{n} f(v_k)\tilde{G}(n-k)] \tag{25}$$

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

$$E[C(\mathbf{v})] = 2\sum_{k,j} \theta_k \tilde{G}(|k-j|) v_k v_j = \mathbf{v}^T \mathcal{I} \mathbf{v}. \tag{26}$$

where $\mathcal{I}$ is a Toeplitz matrix (diagonal constant)
- We thus have a quadratic optimization problem (as in portfolio optimization)

$$\mathbf{v}^* = \arg\min_{\mathbf{v}} \mathbf{v}^T \mathcal{I} \mathbf{v} \qquad s.t. \quad \sum_k v_k \equiv \mathbf{1}^T \mathbf{v} = X \tag{27}$$

that can be solved with a Lagrange multiplier with solution

$$\mathbf{v}^\star = \frac{X}{\mathbf{1}^T \mathcal{I}^{-1} \mathbf{1}} \mathcal{I}^{-1} \mathbf{1}. \tag{28}$$

Solution

With $G(k) = a(c+k)^{-\beta} \sim k^{-\beta}$

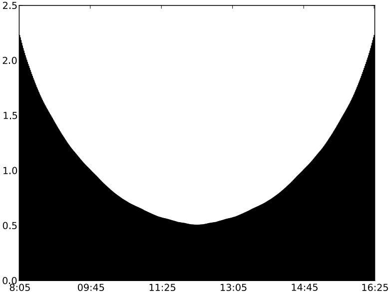


Figure: An example of theoretical optimal solution of the optimal execution problem. As a function of real time, the plot shows the amount of shares to be traded (in arbitrary units).

The solution is symmetric around $N/2$

Note that the U shape does not depend on the intraday profile of volume.

Adding a risk term

The variance of the execution cost under the propagator model is

$$Var[c(\mathbf{v})] = E\left[(c(\mathbf{v}) - E[c(\mathbf{v})])^2\right] = E[(\sum_{k=1}^{N} v_k \sum_{j=0}^{k-1} \eta_j)^2] =$$

$$= E[(\sum_{k=1}^{N} \eta_k \sum_{j=k}^{N} v_j)^2] = \sigma^2 \sum_{k=1}^{N} (\sum_{j=k}^{N} v_j)^2 = \sum_{k,j} \mathcal{V}_{k,j} v_k v_j. \quad (29)$$

where σ^2 is the variance of the residuals. The variance of the cost is a bilinear form. We define $\mathcal{F} \equiv \mathcal{I} + \lambda \mathcal{V}$. By using again Lagrange multipliers, we have therefore the optimal trading schedule

$$\mathbf{v}^\star = z\, \mathcal{F}^{-1}\mathbf{1} = \frac{X}{\mathbf{1}^T \mathcal{F}^{-1}\mathbf{1}} \mathcal{F}^{-1}\mathbf{1}. \quad (30)$$

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

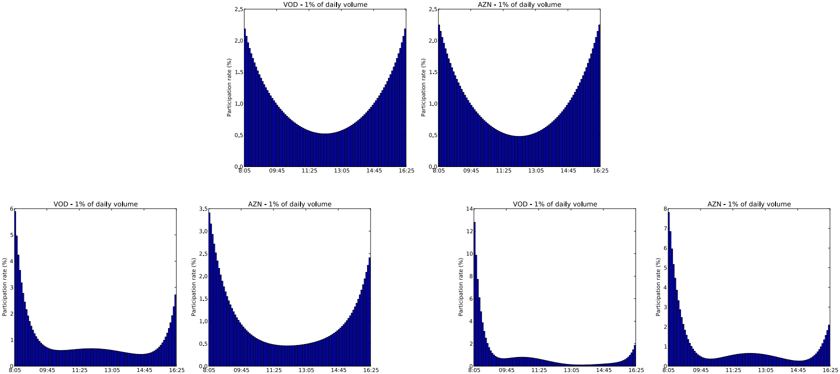


Figure: $\lambda = 0$ (top), $\lambda = 0.2$ (bottom left), $\lambda = 0.9$ (bottom right)

Optimal execution: calibration on real data (no risk aversion)

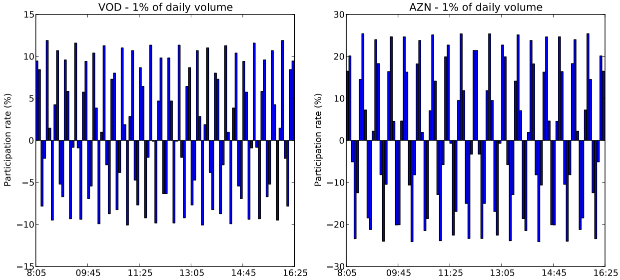


Figure: Optimal solution for two stocks

- The optimal execution for a buy trade includes buys and sells !!
- The cost is positive (no price manipulation), but transaction triggered price manipulation.
- We'll see later conditions leading to this result.

Including spread costs (no risk aversion)

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

$$\tilde{p}_n = p_0 + \sum_{k=0}^{n} [\eta_k + f(v_k)\tilde{G}(n-k)] + \text{sign}(v_k)\delta_k, \tag{31}$$

because we pay half the bid-ask spread on execution. We have defined

$$\delta_k \equiv \frac{s_k/2}{P} = \frac{A_k - B_k}{A_k + B_k}. \tag{32}$$

where A_k and B_k are the ask and the bid.
The objective function of the optimization with spread costs is

$$F[\mathbf{v}] = E[C(\mathbf{v})] + BA(\mathbf{v}) = \mathbf{v}^T \mathcal{I} \mathbf{v} + \mathbf{D}^T |\mathbf{v}|. \tag{33}$$

where $\mathbf{D} = (\{\delta_k\})^T$ is a vector describing the spread cost during execution. We assume for simplicity that $\mathbf{D} = \delta \mathbf{1}$
The absolute value prevents an analytical solution and we use numerical optimization. The shape of the solution does not qualitatively change.

Optimal execution: calibration on real data with spread (no risk aversion)

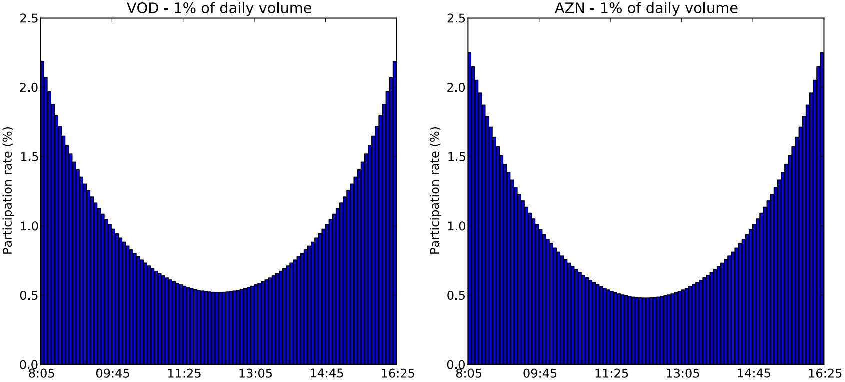


Figure: Spread costs regularize the solution (no sells for a buy program)

Comments

- The alternating (buy-sell) solution and the regularization achieved by the bid ask term is similar to what happens in portfolio optimization.
- It is known that adding to Markowitz objective function a penalty proportional to the sum of the absolute values of the portfolio weights stabilizes the solution and corresponds to an exclusion of short positions (Brodie et al 2009). $\Rightarrow L1$ (or LASSO) regularization
- By choosing a δ parameter much smaller than the fractional spread, one still recovers the U-shaped solution.

Efficient frontier

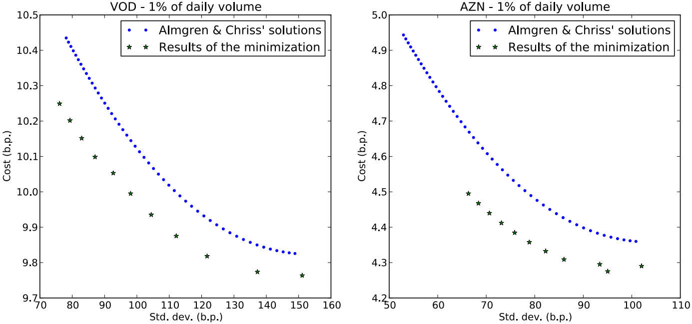


Figure: $2-3\%$ gain with respect to Almgren-Chriss on all the frontier

Optimal execution in real time

- Up to now we have considered the problem in discrete time
- Either transaction time or aggregated real time
- This is mathematically simpler and more applicable to real data, neglecting microstructure
- Some computations and theorems are more easily obtainable by considering continuous time

Problem setting

- $t \in [0, T]$: time interval of execution
- X_t: asset position at time t. **Find the optimal position**
- $X_0 > 0$ ($X_0 < 0$) for a sell (buy)
- $X_{T^+} = 0$
- $S^0 = (S_t^0)_{t \geq 0}$ exogenously given asset price dynamics, here assumed to be a martingale on a probability space
- $S^X = (S_t^X)_{t \geq 0}$ asset price dynamics when the strategy $X = (X_t)_{t \geq 0}$ is used.

Key quantities

- Revenues

$$\mathcal{R}_T(X) = -\int_0^T S_t^X \, dX_t \tag{34}$$

- Liquidation costs

$$\mathcal{C}_T(X) = X_0 S_0^0 - \mathcal{R}_T(X) \tag{35}$$

- Both are stochastic quantities
- dX_t: number of shares traded in $[t, t + dt]$

If the cost depends on the stock price only through the term

$$\int_0^T S_t dX_t \tag{36}$$

and S_t is a martingale, then, integrating by parts

$$E\left[\int_0^T S_t dX_t\right] = E\left[S_T X_T - S_0 X_0 - \int_0^T X_t dS_t\right] = -S_0 X_0 \tag{37}$$

"There is no longer a source of randomness. We may restrict the search for an optimal strategy to non random functions of time".
This means that statically optimal strategies are also dynamically optimal.

Price manipulation

Definition

A round trip is an order execution strategy X with $X_0 = X_T = 0$. A price manipulation strategy is a round trip with

$$E[\mathcal{R}_T(X)] > 0 \tag{38}$$

- It is not possible to open and close a position with a positive expected profit
- Huberman and Stanzl (2004)
- Average profits versus almost sure profits (e.g. in derivative pricing)

Definition

(Alfonsi et al 2012). A market impact model admits transaction-triggered price manipulation if the expected revenues of a sell (buy) program can be increased by intermediate buy (sell) trades. In other words, $\exists X_0, T > 0, \tilde{X}$, such that

$$E[\mathcal{R}_T(\tilde{X})] > \sup\{E[\mathcal{R}_T(X)] | X \textit{ is monotone}\} \tag{39}$$

Definition

A market impact model has negative expected liquidation costs if

$$E[\mathcal{C}_T(X)] < 0 \tag{40}$$

i.e. $E[\mathcal{R}_T(X)] > X_0 S_0$.

Relation between manipulations

Proposition

(Klöck et al 2011)

- *Any market impact model that does not admit negative expected liquidation costs does also not admit price manipulation.*
- *Suppose that asset prices are decreased by sell orders and increased by buy orders. Then the absence of transaction-triggered price manipulation implies that the model does not admit negative expected liquidation costs. In particular, the absence of transaction-triggered price manipulation implies the absence of price manipulation in the usual sense.*

The Almgren-Chriss model

Given two non-decreasing functions h and g with $h(0) = g(0) = 0$, an absolutely continuous strategy $(X_t)_{t \geq 0}$ leads to a price trajectory

$$S_t^X = S_t^0 + \int_0^t g(\dot{X}_s) ds + h(\dot{X}_s) \tag{41}$$

where it is typically assumed that

$$S_t^0 = S_0 + \sigma W_t \tag{42}$$

and W_t is a Wiener process.

- $h(X_t)$ corresponds to temporary price impact
- the term $\int_0^t g(X_s)ds$ describes permanent price impact

The Almgren-Chriss model

The revenues are

$$\mathcal{R}_T(X) = -\int_0^T S_t^X dX_t = \tag{43}$$

$$-\int_0^T S_t^0 dX_t - \int_0^T \dot{X}_t \int_0^t g(\dot{X}_s) ds\ dt - \int_0^T \dot{X}_t h(\dot{X}_t) dt \tag{44}$$

$$= X_0 S_0 + \int_0^T X_t dS_t^0 - \int_0^T \dot{X}_t \int_0^t g(\dot{X}_s) ds\ dt - \int_0^T k(\dot{X}_t) dt \tag{45}$$

where $k(x) \equiv x h(x)$

Proposition

(Huberman and Stanzl 2004, Gatheral 2010). If the Almgren-Chriss model does not admit price manipulation, then $g(x) = \gamma x$ *with* $\gamma \geq 0$

This means that non-linear permanent market impact is inconsistent with the principle of no price manipulation.

Almgren-Chriss with linear permanent impact

If $g(x) = \gamma x$ the revenues simplify to

$$\mathcal{R}_T(X) = X_0 S_0 + \int_0^T X_t dS_t^0 - \frac{\gamma}{2} X_0^2 - \int_0^T k(\dot{X}_t) dt \qquad (46)$$

Proposition

If $g(x) = \gamma x$ and f is convex, then the strategy $v_t = X_0/T$ maximizes the expected revenues.

- The second term in Eq. (46) vanishes in expectation because S^0 is a martingale
- Since f is convex, by Jensen inequality

$$E\left[\int_0^T k(\dot{X}_t) dt\right] \geq \int_0^T k\left(E\left[\dot{X}_t\right]\right) dt = Tk\left(\frac{X_0}{T}\right) \qquad (47)$$

- Almgren et al (2005) estimate $k(x) \propto |x|^{1+\beta}$ with $\beta \simeq 0.6$.
- The constant velocity strategy is called VWAP (or TWAP), i.e. Volume (Time) Weighted Average Price.

Euler-Lagrange solution of the Almgren-Chriss model and linear permanent impact

Let us assume $h(x) = \eta x$ and $g(x) = \gamma x$. The trading velocity is $v_t = -\dot{X}_t$. Minimizing the cost means minimizing

$$E\left[\int_0^T (S_t^X) v_t dt\right] = const + \eta \int_0^T v_t^2 dt \tag{48}$$

In order to find the extremum, let us apply the Eulero-Lagrange equations,

$$\frac{\partial L}{\partial X} - \frac{d}{dt}\frac{\partial L}{\partial v} = 0 \tag{49}$$

with the boundary conditions $X_{t=0} = X_0$ and $X_T = 0$ obtaining

$$v_t = \frac{X_0}{T} \qquad X_t = X_0\left(1 - \frac{t}{T}\right) \tag{50}$$

By adding the variance of the cost as a penalizing risk term

$$Var\left[\int_0^T X_t dS_t^X\right] = \sigma^2 \int_0^T X_t^2 dt \tag{51}$$

we obtain the functional

$$\int_0^T (\eta \dot{X}_t^2 + \lambda\sigma^2 X_t^2) dt \tag{52}$$

The Eulero-Lagrage equation gives

$$\ddot{X}_t - \kappa X_t = 0 \tag{53}$$

with solution

$$X_t = X_0 \frac{\sinh(\kappa(T-t))}{\sinh \kappa T} \qquad \kappa = \sqrt{\frac{\lambda\sigma^2}{\eta}} \tag{54}$$

Transient impact models

- Up to now we have considered fixed and permanent impact models
- Empirical evidences in many markets (equity, futures, etc) shows that the impact is transient, i.e. it decays with time. We now explore optimal execution in such models
- We neglect the temporary impact h and the risk term, present in Almgren and Chriss
- Under these conditions, a possible continuous time generalization of the TIM model is

$$S_t^X = S_0 + \int_0^T f(\dot{X}_s) G(t-s) ds + \int_0^t \sigma dW_t \tag{55}$$

- The expected cost is

$$E[\mathcal{C}_T(X)] = \int_0^T \dot{X}_t dt \int_0^t f(\dot{X}_s) G(t-s) ds \tag{56}$$

Obizhaeva-Wang model

- Impact is linear and decays exponentially in time
- The decay is intepreted as the relaxation of the limit order book when shocked by a trade
- Remembering that $v_t = -\dot{X}_t$, the price during the execution is modeled as

$$S_t^X = S_0 + \eta \int_0^T v_s e^{-\rho(t-s)} ds + \int_0^t \sigma dW_t \tag{57}$$

- The expected cost is

$$E[\mathcal{C}_T(X)] = \eta \int_0^T v_t dt \int_0^t v_s \exp[-\rho(t-s)] ds \tag{58}$$

Obizhaeva-Wang model:solution

The Eulero-Lagrange equation gives

$$\int_0^T v_s e^{-\rho|t-s|} ds = A \tag{59}$$

where A is an integration constant.

This is a Fredholm integral equation of the first kind, whose exact solution is

$$v_t = (X_0 - \rho T)\delta(t) + \rho + (X_0 - \rho T)\delta(t - T) \tag{60}$$

where $\delta(x)$ is the Dirac delta.

Is it possible to generalize to the model

$$S_t^X = S_0 + \eta \int_0^T f(v_s) e^{-\rho(t-s)} ds + \int_0^t \sigma dW_t \tag{61}$$

i.e. a non-linear and exponentially decaying impact?

Gatheral (2010) showed that this is not possible, in fact

Proposition

If temporary market impact decays exponentially, price manipulation is possible unless $f(v) \propto v$

Linear case

Let us assume first that the instantaneous impact is linear, $f(v) = \gamma v$, then

$$S_t^X = S_t^0 + \int_{s<t} G(t-s)dX_s \tag{62}$$

$$C(X) \equiv E[\mathcal{C}_T(X)] = \frac{1}{2}\int_0^T \int_0^T G(|t-s|)dX_s dX_t \tag{63}$$

Proposition

(Bochner Theorem) $C(X) \geq 0$ if and only of $G(|x|)$ can be represented as the Fourier transform of a positive finite Borel measure μ on $\mathbb{R}$, i.e.

$$G(|x|) = \int e^{ixz}\mu(dz) \tag{64}$$

Theorem

Suppose G is positive definite. Then X^* minimizes $C(X)$ if and only if $\exists\lambda$ such that X_t^* solves $\forall t$

$$\int_0^T G(|t-s|)dX_s^* = \lambda \tag{65}$$

and thus $C(X^*) = \frac{1}{2}\lambda X_0$.

(Important) Example: let $G(t-s) = (t-s)^{-\gamma}$ then the integral equation is the Abel equation with solution

$$v_s = \frac{B}{[s(T-s)]^{(1-\gamma)/2}} \tag{66}$$

and

$$X_0 = \int_0^T v_s ds = B\sqrt{\pi}\left(\frac{T}{2}\right)^\gamma \frac{\Gamma\left(\frac{1+\gamma}{2}\right)}{\Gamma\left(1+\frac{\gamma}{2}\right)} \tag{67}$$

Linear case: transaction-triggered price manipulation

Theorem

Suppose G is convex, satisfies $\int_0^t G(t)dt < \infty$, and there is an admissible strategy. Then there exists a unique admissible optimal strategy X_t^* which is monotone, i.e. there is no transaction-triggered price manipulation.

Examples:

- $G(t) = \frac{1}{(1+t)^2}$ is convex, no transaction triggered price manipulation
- $G(t) = \frac{1}{(1+t^2)}$ is concave around zero, the optimal soultion displays transaction triggered price manipulation

Non-linear transient impact

Let us consider now the most general model

$$S_t^X = S_0 + \int_0^T f(\dot{X}_s) G(t-s) ds + \int_0^t \sigma dW_t \tag{68}$$

with

$$\int_0^T v_t dt = X_0 \tag{69}$$

- The optimal solution is not known
- If we consider a VWAP strategy, $V_t = X_0/T$, the expected cost is

$$C^{VWAP} = \frac{1}{T} f\left(\frac{X_0}{T}\right) \int_0^T dt \int_0^t G(t-s) ds \tag{70}$$

Theorem

(Gatheral 2010): If $G(t)$ is finite and continuous at $t=0$ and f is nonlinear, then there is price manipulation

A special, yet important, class

We consider the case where

$$f(v) = c\left(\frac{|v|}{V}\right)^{\delta} sign(v) \qquad\qquad G(t-s) = (t-s)^{-\gamma} \tag{71}$$

Theorem

(Gatheral 2010). If $G(t) = t^{-\gamma}$ with $\gamma \in (0,1)$ and $f(x) \propto |x|^{\delta} sign(x)$ with $\delta > 0$, then price manipulation exists when one of the two following conditions is verified:

$$\gamma + \delta \leq 1 \qquad\qquad \gamma \leq \gamma^* = 2 - \frac{\log 3}{\log 2} \simeq 0.415 \tag{72}$$

A special, yet important, class

- The expected cost of a VWAP is

$$C^{VWAP} = \frac{c}{(1-\gamma)(2-\gamma)} \frac{X_0^{\delta+1}}{V^{\delta}} T^{1-\gamma-\delta} \tag{73}$$

and the impact is

$$E[S_T - S_0] = \frac{c}{1-\gamma} \left(\frac{X_0}{V}\right)^{\delta} T^{1-\gamma-\delta} \tag{74}$$

- Interestingly, if $\delta + \gamma = 1$ the expected impact and cost do not depend on the execution time.
- If $\delta = \gamma = 0.5$, the impact is

$$E[S_T - S_0] = 2c\sqrt{\frac{X_0}{V}} = 2c\sqrt{T_d}\sqrt{\frac{X_0}{ADV}} \tag{75}$$

which is the celebrated square root impact formula

Transforming the cost minimization into an integral equation

- Dang (2014) shows that, given $f \in C^1(\mathbb{R})$ and $G \in L^1[0, T]$, for the class of functions x on $[0, T]$ satisfying
  - x is absolutely continuous on $(0, T)$,
  - $f \circ v \in L^1[0, T]$,

  the following necessary condition for the stationarity of the cost functional holds:

  $$\int_0^t f(v(s))\, G(t-s)\, ds + f'(v(t)) \int_t^T v(s)\, G(s-t)\, ds = \lambda, \tag{76}$$

  where λ is a constant set by the constraint equation.
- In the concave $(\delta < 1)$ case there is no guarantee that the minimum is global
- This is a weakly singular Urysohn integral equations of the first kind (very hard to solve!!)

Integral equation

Equation 76 is a weakly singular Urysohn equations of the first kind

$$\int_0^T G(|t-s|)\, F(v(s), t)\, ds = \lambda \tag{77}$$

where

$$F(v(s), t) = \begin{cases} f(v(s)), & s \leq t \\ v(s) f'(v(t)), & s > t, \end{cases} \tag{78}$$

- Two nonlinearities: one related to f and one related to F.
- The term with the first derivative entangles the susceptibility of response at time t with the future trading rates, i.e. a coupling between present and future values of v. This implies that Eq. 78 cannot be classified as a weakly singular nonlinear Fredholm equation, because the function F depends both on t and on s.
- For concave f, the integral equation becomes meaningless when $v = 0$

The linear case

In the linear case, $f(v) = v$, the integral equation becomes a Friedholm integral equation of the Abel type, which can be solved analytically as we have seen above:

$$v(t) = \frac{c}{[t\,(T-t)]^{\frac{1-\gamma}{2}}}, \tag{79}$$

where c is uniquely determined by the normalization constraint

$$c = X / \left(\sqrt{\pi} \left(\frac{T}{2} \right)^{\gamma} \frac{\Gamma\left(\left(1+\gamma\right)/2\right)}{\Gamma\left(1+\gamma/2\right)} \right), \tag{80}$$

where $\Gamma(\cdot)$ is the Euler's function.

This solution has a U shape and is symmetric under time reversal, i.e. $v(t) = v(T-t)$, $t \in [0, T/2]$.
In the following we will refer to this solution as the GSS solution.

Perturbative approach

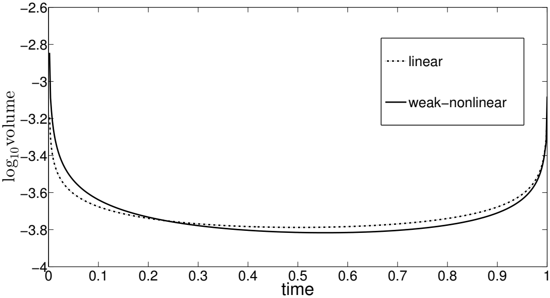


Figure: Solution of the weak nonlinear Fredholm integral equation for $\gamma = 0.5, \epsilon = 0.02$ and $X = 0.1$. The full line represents the solution $v(s) = u(s) + \epsilon\, w(s)$. The solution is not symmetric for time reversal. The dotted line represents the GSS solution, i.e. the solution valid for the linear impact case.

We perform an expansion $f(v) = v^{1-\epsilon}$, with $0 < \epsilon \ll 1$ and we solve exactly the perturbed equation

No symmetry for time reversal: front loading for concave impact ($\delta < 1$), back loading for convex impact ($\delta > 1$).

Homotopy Approach

- We solve the integral equation by means of the Discrete Homotopy Analysis Method (DHAM), i.e. as a *continuous* transformation of a given solution (never crossing $v = 0$)
- Given the following general equation

$$\mathcal{N}\left[v\left(t\right)\right] = 0, \tag{81}$$

we construct the so-called zero-order deformation equation

$$\left(1 - p\right)\mathcal{L}\left[\phi\left(t;p\right) - v^{0}\left(t\right)\right] = p\,\hbar\,H\left(t\right)\mathcal{N}\left[\phi\left(t;p\right)\right], \tag{82}$$

where $p \in [0,1]$ and $v^0(t)$ is an initial guess

- $$\phi\left(t;0\right) = v^{0}\left(t\right), \qquad \phi\left(t;1\right) = v\left(t\right). \tag{83}$$
- Expanding $\phi\left(t;p\right)$ in Maclaurin series with respect to p, we have

$$\phi\left(t;p\right) = v^{0}\left(t\right) + \sum_{m=1}^{\infty} v^{m}\left(t\right)p^{m}, \tag{84}$$

where

$$v^{m}\left(t\right) = \frac{1}{m!}\frac{\partial\phi^{m}\left(t;p\right)}{\partial p^{m}}\bigg|_{p=0}. \tag{85}$$

- The operator is

$$\mathcal{N}\left[v(t)\right] = -\lambda + \int_0^T G\left(|t-s|\right)\, F\left(v(s), t\right)\, ds\,. \tag{86}$$

- Choosing $\mathcal{L}$ and $H(t)$ as the identity operators, the zero-order deformation equation is

$$(1-p)\left[\phi\left(t;p\right) - v^0\left(t\right)\right] = \hbar\, p \mathcal{N}\left[\phi\left(t;p\right)\right]. \tag{87}$$

- Differentiating m times we get

$$v^m\left(t\right) = v^{m-1}\left(t\right) + \hbar\, R^m\left(\mathbf{v}^{m-1}\right), \tag{88}$$

where for $m > 1$

$$\begin{aligned} R^m\left(\mathbf{v}^{m-1}\right) &= \frac{1}{(m-1)!}\frac{\partial^{m-1}\mathcal{N}\left[\phi\left(t;p\right)\right]}{\partial p^{m-1}}\Big|_{p=0} \\ &= \int_0^T G\left(|t-s|\right)\left\{\frac{1}{(m-1)!}\frac{\partial^{m-1}F\left(\phi\left(s;p\right),t\right)}{\partial p^{m-1}}\Big|_{p=0}\right\} ds\,. \end{aligned} \tag{89}$$

Discrete Homotopy Approach: Results

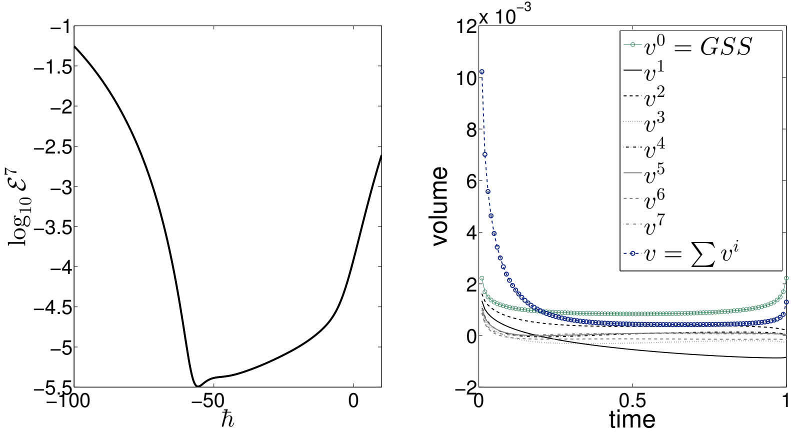


Figure: The logarithm of the squared residual $\mathcal{E}^7(\hbar)$ is illustrated on the left panel, the minimum is attained for $\hbar = -55.7$ where we have $\mathcal{E}^7 = 3.2 \times 10^{-6}$. The GSS guess and the DHAM solution are reported on the right panel respectively by a full green line with circles and a dashed blue line with circles, are reported also the results of the seven deformation equations.

Discrete Homotopy Approach: Costs

	VWAP	GSS	DHAM	VWAP	GSS	DHAM
δ	$\gamma = 0.45$	$\gamma = 0.45$	$\gamma = 0.45$	$\gamma = 0.5$	$\gamma = 0.5$	$\gamma = 0.5$
1.0	0.0117	**0.0116**	**0.0116**	0.0133	0.0132	**0.0131**
0.95	0.0132	**0.0130**	**0.0130**	0.0150	**0.0148**	**0.0148**
0.90	0.0148	0.0146	**0.0143**	0.0168	0.0166	**0.0164**
0.85	0.0166	0.0164	**0.0162**	0.0188	0.0186	**0.0185**
0.80	0.0186	0.0184	**0.0179**	0.0211	0.0209	**0.0204**
0.75	0.0209	0.0206	**0.0198**	0.0237	0.0234	**0.0227**
0.70	0.0234	0.0231	**0.0218**	0.0266	0.0263	**0.0249**
0.65	0.0263	0.0260	**0.0235**	0.0298	0.0295	**0.0274**
0.60	0.0295	0.0291	**0.0251**	0.0335	0.0331	**0.0297**
0.55	0.0331	0.0327	**0.0275**	0.0376	0.0372	**0.0323**
0.50				0.0422	0.0417	**0.0347**

Table: Costs for three different strategies, VWAP, GSS, and DHAM, in the no-dynamic-arbitrage region for $\gamma = 0.45$, 0.5. The numbers in boldface are those achieving the smallest cost. The difference between costs increases with the degree of non-linearity, i.e. $\delta < 1$. In this case we use a GSS initial guess to obtain the DHAM solution.

The DHAM solution has a cost up to 20% smaller than the GSS, while the latter has a cost which is only 1% smaller than the VWAP

Fully numerical solutions

- DHAM solution is a continuous deformation of a VWAP or a GSS solution
- Therefore it is smooth and, more important, has always the same sign
- What happens if we minimize **numerically** the cost on a discrete grid of N intervals in $[0, T]$ (piecewise constant solution)?

$$\arg\min \sum_{i=1}^{N} \sum_{j=1}^{N} v_i^n \, f\left(v_j^n\right) A_{ij} \qquad s.t. \quad \sum_{i=1}^{N} v_i = \frac{NX}{T} \tag{90}$$

where the A_{ij} are elements of a Toeplitz matrix that describes the decay kernel $G\left(t-s\right)$

$$\begin{aligned}
A_{ij} &= 0;\; j > i, \\
A_{ii} &= \frac{1}{\left(1-\gamma\right)\left(2-\gamma\right)}\left(\frac{T}{N}\right)^{2-\gamma}; \\
A_{ij} &= \frac{1}{\left(1-\gamma\right)\left(2-\gamma\right)}\left(\frac{T}{N}\right)^{2-\gamma}\left\{\left(i-j+1\right)^{2-\gamma}-2\left(i-j\right)^{2-\gamma}+\left(i-j-1\right)^{2-\gamma}\right\};\; j \leq
\end{aligned}$$

Fully numerical solution: A $N = 2$ period motivating example

Let us consider the case of a buy program over $N = 2$ periods

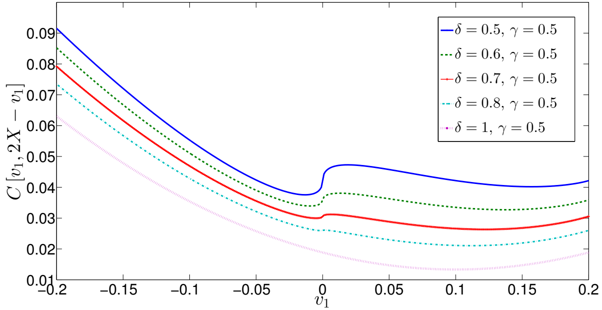


Figure: Cost function $C[v_1, 2X - v_1]$ for $X = 0.1, \gamma = 0.5$. For $\delta = 1$ the minimum is at $v_1 = X$. In the nonlinear case there are two local minima.

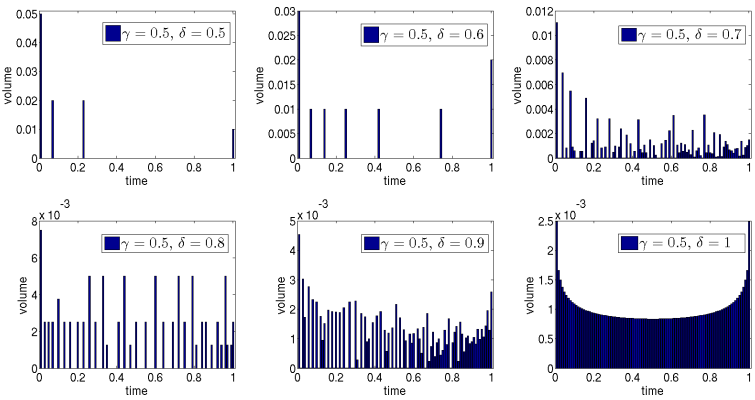
For a buy program with strong nonlinearity ($\delta \gtrsim 0.5$)

- With the constraint $v_i \geq 0$ the optimal solution is to trade only in the second period
- Without constraints, it is optimal to sell in the first period and buy more in the second one

SQP numerical optimization

- We perform an in-depth numerical optimization of the cost function when N is large (~ 100)
- We use Sequential Quadratic Programming (SQP) with a large number of starting points on the simplex $\sum_{i=1}^{N} v_i = NXT^{-1}$
- We find a very large number of distinct extremal points and we select the one with the smallest cost
- We use second order condition to verify that a very large fraction of extremal points are minima
- By computing the eigenvalue spectra of the Hessian of the cost function a the minima, we exclude that the landscape of cost is sloppy (i.e. does not depends strongly on a few number of directions in the state-space).
- The landscape is in fact **rugged** (i.e. composed by many local minima with similar cost)
- Many suboptimal minima correspond to similar trading patterns (see below)

"Optimal" strategies with the constraint $v \geq 0$



The optimal strategy is to trade in bursts separated by no trading periods

The unconstrained "optimal" solution for a buy program

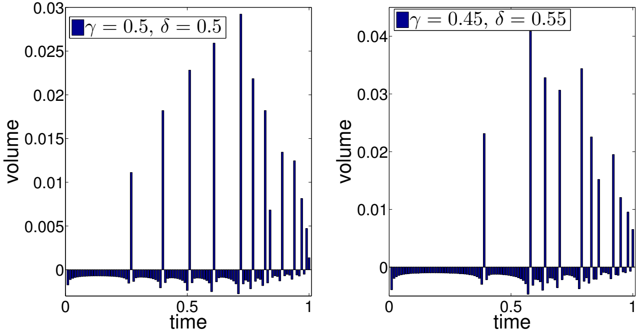


Figure: Optimal solution given by the SQP-algorithm for a buy-program where $X = 0.1$, i.e. 10% of a unitary market volume. We report the volume to be traded in each interval of time.

Under strong non-linearity, the optimal buy program is composed by few intense buying periods interspersed by long weak selling periods

For strong nonlinearity $\rightarrow$ Negative costs!! $\rightarrow$ Possibility of price manipulation

Price manipulation and its regularization

- The no-arbitrage condition $\delta + \gamma > 1$ is not sufficient to guarantee the absence of price manipulation.
- Any non-linear impact leads to price manipulation strategies for sufficiently large discretization.
- It is possible to regularize the solutions with two approaches:
  - Adding a spread cost

$$C = \sum_{i=1}^{N} \sum_{j=1}^{N} v_i \, f\left(v_j\right) A_{ij} + \delta_S \frac{T}{N} \sum_{i=1}^{N} |v_i|, \tag{91}$$

    where δ_S is half the bid-ask spread. This is equivalent to a L_1 or LASSO regularization widely used in computer science.
  - Modifying the impact function $f(v)$ to

$$f_G\left(v\right) = c \; sign\left(v\right) \left\{ \left(\frac{|v|}{|v| + V} \right)^{\delta} + d \, \frac{|v| \left(|v| + V\right)}{V^2} \right\}, \tag{92}$$

    where $V = X_M / T$ is the market volume per unit time. This is concave for small v and convex for large v (illiquidity wall)
- Both regularization succeed (in some parameter regime) to avoid negative costs

Spread regularization

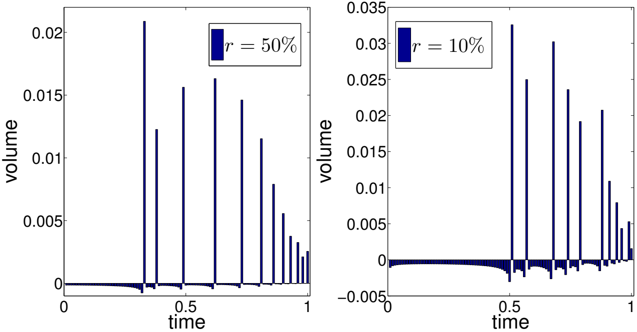


Figure: Optimal solution given by the SQP algorithm for a buy-program where $X = 0.1$, i.e. 10% of a unitary market volume, in presence of a spread cost for $\gamma = 0.45$ and $\delta = 0.55$. The liquidation cost is $C_{SQP} = 0.026$ for large spread cost (left) and $C_{SQP} = 5.9 * 10^{-3}$ for small spread cost (right).

The larger the spread, the stronger the regularization, the smaller the contribution from sell trades

Concave-convex impact

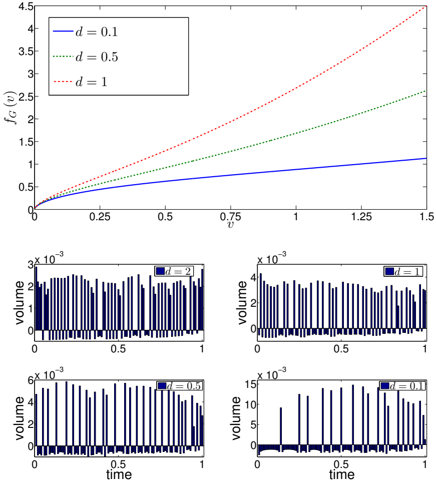


Figure: Top. Concave-convex impact function for values of parameters: $c = 1$, $\delta = 0.55$, $X_M = 1$, $T = 1$. Bottom. Optimal solution given by the SQP-algorithm for a buy-program

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)
- Jim Gatheral, Alexander Schied, Dynamical models of market impact and algorithms for order execution, in *Handbook of Systemic Risk* (J.-P. Fouque and J. A. Langsam editors), Cambridge University Press 2013.