Market Microstructure & Algorithmic Trading

Jan Obłój
jan.obloj@maths.ox.ac.uk

MSc in Mathematical and Computational Finance
Mathematical Institute, University of Oxford

Hilary Term 2022

Outline

Market Microstructure and Algorithmic Trading

Main aim is to learn about

- Market Microstructure: How electronic markets operate, the price formation mechanisms, choices available to market participants and their (potential) information signals
- Algorithmic Trading (AT): The use of computer algorithms that make trading decisions, submit orders, and manage those orders after submission.
  - NYSE 2005 to 2009: consolidated av daily share vol up 181%; av speed of exec for small (marketable) orders down 10.1 to 0.7 sec; consolidated av daily trades up 662%; consolidated av trade size dwon 724 to 268 shares, SEC (2010).
  - US/EU: 2003 AT $\sim$ 15 % of market volume, in 2010 about 60–70%. FX: from 25% in 2006 to 80% in 2016.
- High Frequency (HF) Trading: refers to the subset of AT trading strategies that are characterised by their reliance on speed differences relative to other traders to make profits based on short-term predictions and also by the objective to hold essentially no asset inventories for more than a very short period of time.

Why AT? One example

- Institutional or large players need to trade (buy and sell) large amounts of securities. These quantities are too large for the market to process without prices moving in the 'wrong direction' (slippage).
- Thus, large orders are broken up in small ones and these are traded over time (minutes, hours, days, weeks, or even months) and across different venues.
- Deciding how to break up and execute a large order can mean saving millions of pounds for large players

Outline

Reading

There is a wealth of ongoing research and growing body of publications.
For market impact modelling, a good start are two survey papers:

- Lehalle, "Market Microstructure Knowledge Needed for Controlling an Intra-Day Trading Process"
- Gatheral and Schied, "Dynamical models of market impact and algorithms for order execution"

both in 2013 *Handook on Systemic Risk* (ed. Fouque and Langsam) and on arXiv.

For market microstructure, I suggest two review papers and four books:

- Chakraborti et al, "Econophysics review" (parts I and II), in *Quantitative Finance*, 2011
- "How markets slowly digest changes in supply and demand", Bouchaud et al (2009)
- O'Hara, *Market Microstructure Theory,* 1995
- Hasbrouck, *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading,* 2006
- Lehalle and Laruelle, *Market Microstructure in Practice*, 2014.
- Cartea, Jaimungal and Penalva, *Algorithmic and High-Frequency Trading*, 2015.

Some questions to the audience...

Market Microstructure usually refers to

A The structure of prices across different stocks

B The price formation mechanisms from trading actions

C The micro movements of the stock price within one day

D All of the above

Outline

Models in Quantitative Finance

- "All models are wrong but some are useful" (G. Box '78)
- Models need to be tailored to
  - the available inputs
  - the intended outputs
- Models need to
  - conform to stylised facts
  - produce reasonably useful and robust outputs
  - avoid creating arbitrage opportunities

Brief history of modelling in QF

- Fair price (fundamental economics)
  model: fundamentals
- Option pricing & optimal investment
  model: the underlying price process (exogenous)
  Samuelson '65, B&S and Merton '73
- Further option pricing: Exotics or FI options
  model: a high- or ∞- dimensional system of underlyings
  e.g.: HJM '92 and LMM '97 in Fixed Income; Market models of Schweizer & Wissel '08, Carmona & Nadtochiy '09

Brief history of modelling in QF – cont.

- Optimal execution of planned trades
  model: impact of trades on price dynamics or
  model: supply & demand dynamics
  Bertsimas & Lo '98, Almgren & Chriss '00;
  Obizhaeva & Wang '13, Alfonsi et al. '08
- Price formation via market microstructure
  model: LOB dynamics (zero intelligence)
  model: Agent trades (agent based)
  Cont et al. '10, Smith et al. '03, Farmer et al. '05;
  Kyle '85, O'Hara '95

... and MANY more references...

Frictionless modelling setting

The classical modelling framework in mathematical finance, like the one postulated by Black and Scholes '73, assumes infinite liquidity:

- asset traded at uniquely given and known prices
- buying and selling in arbitrary quantities possible
- trading at no cost possible
- trading has no impact on the price

This is unrealistic and unsatisfactory: in reality we have market frictions.

Market Frictions

Market frictions – cont.

Many frictions either part of the game (opportunity cost) or well-defined (taxes). For many traders other frictions satisfactory summarised in

- proportional transaction costs: pay ϵS_t for trading one unit of S_t.

However this is not acceptable for

- large trades (relative to volume & time horizon)
- frequent trading (relative to liquidity)

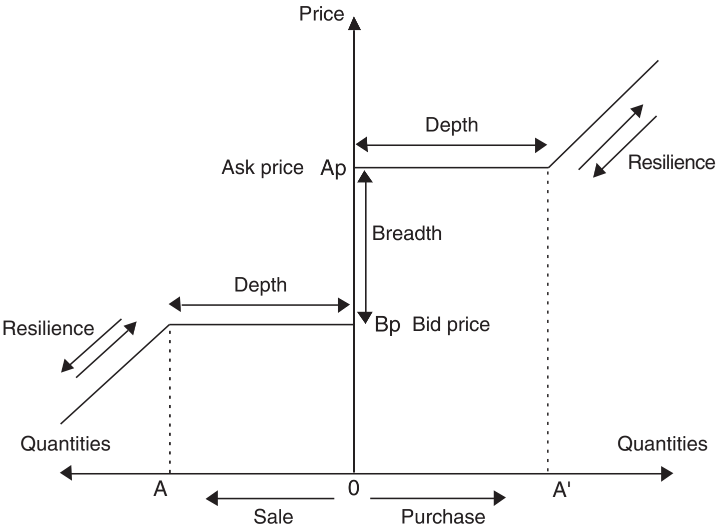
which require understanding of

- liquidity provision and
- price formation.

Aspects of liquidity (Kyle '85)

- **Tightness** (Breadth): measures how wide the bid-ask is, i.e. measures the cost of a position reversal at a short notice for a standard amount
- **Market depth**: corresponds to the volume which may be bought/sold without immediately affecting the price
- **Market resilience**: describes the speed at which prices revert to previous level (equilibrium) after a random shock in the order flow
- **Time delay**: measures the delay between processing and executing an order

Aspects of liquidity (Kyle '85)



Source: Bervas '06

Summary so far

- Models are build taking into account available inputs and desirable outputs
- In QF models postulate exogenous dynamics for different underlyings depending on what is traded and what one wants to price
- Traditional models assume a frictionless setting with ∞ liquidity
- In practice this fails. A lot can be accounted for using proportional transactions costs.
- Large and/or frequent trading requires modelling of liquidity and/or price impact.

Outline

Price impact modelling

We saw that large and/or frequent trades may affect the price. We may need to split and spread large orders in practice. To answer how to do it we need to understand:

- how to model/quantify the impact of trading on the price?
- what are the desirable/undesirable properties of such models?
- how to compute optimal execution trading strategies?

There are two natural approaches to model price impact:

I: postulate fair price dynamics and the price impact of trading

II: be serious about modelling Market Microstructure, i.e. model supply and demand and their interaction.

We focus first on I. Then we use the LOB discussion to tackle II.

Trade execution setup

Goal: buy/sell $\mathfrak{N}$ shares by time T.

Trade execution strategy:

- $Q = (Q_t)_{t \leq T}$, where Q_t is the number of shares held at time t
- The initial position $Q_0 = q = \mathfrak{N}$ is positive for a sell strategy and negative for a buy strategy
- The final condition $Q_T = 0$ indicates the position is liquidated at T
- The path will be monotone for a pure buy or pure sell strategy. In general it is of finite variation.

We think of T as around $5 - 10$, and up to 30, minutes.
For now, we are ignoring problems from higher(+) or lower(-) levels:

\+ How a large desired trade position is split into chunks allocated their time horizons.

\- What orders (market vs limit) are used and to which venues these are routed.

Price impact model

Price impact model quantifies the feedback effect of trading strategy Q on the asset price. A typical setup is:

- Exogenously specified price process $S^0 = (S^0_t : t \leq T)$ for fair (unaffected) price dynamics.
  S is a semimartingale (usually a martingale) on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ and we assume Q is predictible
- Given Q, a model prescribes S^Q the price process realised when implementing trading strategy Q.
- Typically $dQ_t = \nu_t dt$ and we then write $S^Q = S^\nu$.
- Typically, a buy strategy increases the prices and a sell strategy decreases the prices: if $\nu_t \geq 0$ for all $t \leq T$ then $S^\nu_t \geq S^0_t$, $t \leq T$. However this is not necessarily true for a fixed t since S^ν_t may be affected by all of $(Q_u : u \leq t)$.

Revenues and costs

Suppose $dQ_t = \nu_t dt$ and S_t^ν depends continuously on Q. Then at time t, the infinitesimal amount of $-d\nu_t dt$ shares is sold at price S_t^ν. Thus

$$\text{revenues from strategy } Q \text{ are} \quad \mathcal{R}(Q) = -\int_0^T S_t^Q dQ_t$$

(when Q is not absolutely continuous adjustments are necessary)

Objective: Maximise some performance functional of $\mathcal{R}(Q)$.
For example:

- maximise the expected value $\mathbb{E}[\mathcal{R}(Q)]$
- maximise a mean-variance criterion $\mathbb{E}[\mathcal{R}(Q)] - \lambda \operatorname{var}(\mathcal{R}(Q))$
- maximise the expected utility $\mathbb{E}[U(\mathcal{R}(Q))]$
- ...

Revenues and costs – cont.

Alternatively: Minimise functional of implementation shortfall (i.e. cost of liquidation), which is the difference between the book value $Q_0 S_0^0$ and the revenues (or the capture):

$$\text{liquidation cost of } Q \text{ is} \qquad \mathcal{C}(Q) = Q_0 S_0^0 - \mathcal{R}_T(Q).$$

If we write $S_t^Q = S_t^0 + I_t^Q$ then

$$\begin{aligned}\mathcal{R}(Q) &= -\int_0^T S_t^Q dQ_t = -\int_0^T S_t^0 dQ_t - \int_0^T I_t^Q dQ_t \\ &= S_0^0 Q_0 + \underbrace{\int_0^T Q_t dS_t^0}_{=-\mathcal{C}^{vol}(Q)} - \underbrace{\int_0^T I_t^Q \dot{Q}_t dt}_{=\mathcal{C}^{exec}(Q)}\end{aligned}$$

The total liquidation cost $\mathcal{C}(Q)$ has two components:

- $\mathcal{C}^{vol}$ expresses the volatility risk of trading over time instead of instantly
- $\mathcal{C}^{exec}$ expresses the effect of price impact

Almgren–Chriss type price impact

The unaffected price follows a Brownian motion:

$$S_t^0 = S_0^0 + \sigma W_t.$$

Then, the price impact has two components:

- permanent impact: $\int_0^t g(\nu_s)ds$
- temporary impact: $h(\nu_t)$

for nondecreasing functions $g, h : \mathbb{R} \to \mathbb{R}$ and $\dot{Q}_t = \frac{dQ_t}{dt} = \nu_t$ the trading speed. The affected price is given by

$$S_t^Q = S_t^0 + \int_0^t g(\nu_s)ds + h(\nu_t).$$

In the special case of linear impacts: $g(x) = \gamma x$ and $h(x) = \eta x$

$$S_t^Q = S_t^0 + \gamma \int_0^t dQ_s + \eta\nu_t = S_t^0 + \gamma(Q_t - Q_0) + \eta\nu_t.$$

A–C model with linear price impact

In the special case of linear impacts: $g(x) = \gamma x$ and $h(x) = \eta x$

$$S_t^Q = S_t^0 + \gamma \int_0^t dQ_s + \eta \nu_t = S_t^0 + \underbrace{\gamma (Q_t - Q_0) + \eta \nu_t}_{= I_t^Q}.$$

The revenues are then given by

$$\begin{aligned} \mathcal{R}(Q) &= -\int_0^T S_t^Q dQ_t = S_0^0 Q_0 + \int_0^T Q_t dS_t^0 - \int_0^T I_t^Q \nu_t dt \\ &= S_0^0 \mathfrak{N} + \sigma \int_0^T Q_t dW_t - \frac{\gamma}{2} \mathfrak{N}^2 - \eta \int_0^T \nu_t^2 dt, \end{aligned}$$

since $Q_T = 0$.

A–C model with linear price impact (cont.)

Assuming Q is bounded, the expected revenues are

$$\mathbb{E}[\mathcal{R}(Q)] = S_0^0\mathfrak{N} - \frac{\gamma}{2}\mathfrak{N}^2 - \eta\,\mathbb{E}\left[\int_0^T \nu_t^2 dt\right].$$

The last term is an integral w.r.t. $\mathbb{P}(d\omega)\otimes dt$ of the square of $\nu_t(\omega)$. It follows that it is minimised, and hence $\mathbb{E}[\mathcal{R}(Q)]$ is maximised, by the strategy

$$\nu_t^* = -\frac{\mathfrak{N}}{T}$$

which sells (or buys) the shares at constant speed (to see this simply apply Jensen's inequality). In particular the solution is independent of the volatility! (Bertsimas & Lo '98)

The resulting expected liquidation cost of $\mathfrak{N}$ shares is

$$\mathbb{E}[\mathcal{C}(Q)] = \left(\frac{\gamma}{2} + \frac{\eta}{T}\right)\mathfrak{N}^2$$

quadratic in number of shares and independent of volatility σ.

A–C model so far – summary

Proposition

In the Almgren–Chriss price impact model with linear permanent impact, $g(x) = \gamma x$, and $xh(x)$ convex, for any given $\mathfrak{N} \in \mathbb{R}$ the strategy

$$Q_t^* = \frac{\mathfrak{N}(T-t)}{T}, \quad t \le T,$$

maximises the expected revenues $\mathbb{E}[\mathcal{R}(Q)]$ in the class of all adapted and bounded trade execution strategies Q.

The strategy Q^* spreads the execution evenly over the time horizon $t \in [0, T]$. It is often referred to as the time-weighted average price strategy or TWAP. When the time is relative and t corresponds to traded volume the Q^* is called volume-weighted average price strategy or VWAP. Both are used as industry benchmarks.

Almgren et al. '05 argued these assumptions are consistent with empirical observations and suggested $xh(x) \approx |x|^{1.6}$.

A–C model with mean-variance criterion

So far we only looked at expected revenues. Almgren and Chriss '00 propose to consider

$$\max_Q \mathbb{E}[\mathcal{R}(Q)] \quad \text{subject to } \operatorname{var}(\mathcal{R}(Q)) \leq v_*$$

which, introducing a Langrange multiplier, turns into an unconstrained problem

$$\max_Q \left(\mathbb{E}[\mathcal{R}(Q)] - \lambda \operatorname{var}(\mathcal{R}(Q))\right).$$

This is a **hard** problem. However assuming Q is deterministic it turns into

$$\max_Q \left(\mathfrak{N} S_0^0 - \frac{\gamma}{2}\mathfrak{N}^2 - \eta \int_0^T \dot{Q}_t^2 dt - \lambda\sigma^2 \int_0^T Q_t^2 dt\right)$$

which can be solved explicitly as a standard variational calculus problem.

A–C model with mean-variance criterion

Indeed, the problem is equivalent to

$$\min_{Q} \int_0^T \left(\frac{\lambda\sigma^2}{2} Q(t)^2 + \eta Q'(t)^2 \right) dt$$

Setting the first variation to zero:

$$0 = \int_0^T \left(g(t)\lambda\sigma^2 Q(t) + 2g'(t)\eta Q'(t) \right) dt, \quad \forall g \in C^1 : g(0) = g(T) = 0.$$

Integrating by parts:

$$0 = \int_0^T g(t) \left(\lambda\sigma^2 Q(t) - 2\eta Q''(t) \right) dt, \quad \forall g \in C^1 : g(0) = g(T) = 0$$

which gives the Euler-Lagrange equation

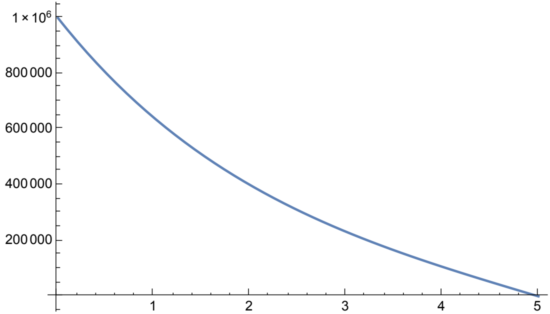
$$Q''(t) = \frac{\lambda\sigma^2}{2\eta} Q(t), \quad \text{s.t. } Q(0) = \mathfrak{N}, Q(T) = 0.$$

Solving the ODE we obtain

A–C model with mean-variance criterion

The solution is given by

$$Q_t^* = \mathfrak{N}\frac{\sinh(\kappa(T-t))}{\sinh \kappa T} \quad \text{for } \kappa = \sqrt{\frac{\lambda\sigma^2}{2\eta}}.$$

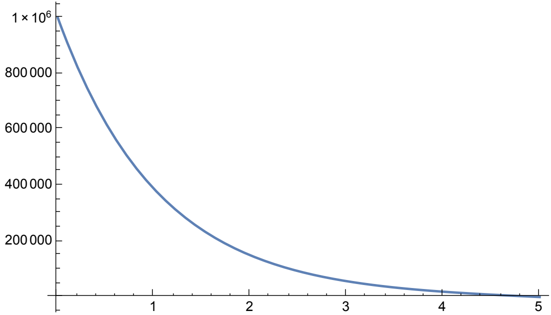


Optimal liquidation strategy of 10^6 shares over 5 days under 30% annual vol and impact 1% of daily volume = bid-ask. Moderate λ.

A–C model with mean-variance criterion

The solution is given by

$$Q_t^* = \mathfrak{N}\frac{\sinh(\kappa(T-t))}{\sinh \kappa T} \quad \text{for } \kappa = \sqrt{\frac{\lambda\sigma^2}{2\eta}}.$$



Optimal liquidation strategy of 10^6 shares over 5 days under 30% annual vol and impact 1% of daily volume = bid-ask. High λ.

A–C model with other criteria

Mean-variance is **not** amenable to dynamic programming and leads to time-inconsistent strategies. In analogy to optimal investment, other criteria are natural:

- Maximise expected utility: $\max_Q \mathbb{E}[U(\mathcal{R}(Q))]$
  The problem can be reformulated as a stochastic control problem with non-standard (finite fuel) constraint: $Q_0 = \mathfrak{N}$ and $Q_T = 0$. Leads to an **HJB** equation. Solution known for $U(x) = -\exp(-\lambda x)$ ... the same as for mean-variance! (Schied, Schöneborn & Tehranchi '10).
- Maximise
  $$\mathbb{E}\left[\mathcal{R}(Q) - \lambda \int_0^T Q_t S_t^Q dt\right]$$
  Gatheral & Schied '11
- And more... see the next section, where we look at these using stochastic control methods!

Criticism of A–C setting

- Price process can go negative; impact additive & in absolute terms. Bertsimas & Lo '98 suggest

$$S_t^Q = S_t^0 \exp\left(\int_0^t g(\nu_s)ds + h(\nu_t)\right), \quad S_t^0 = S_0^0 \exp\left(\sigma W_t - \frac{\sigma^2}{2}t\right)$$

  but computing optimal strategies more involved.
- Price impact simplistic, in reality transient effect, see Moro et al. '09 (cf. resilience)
- Computed optimal strategies are deterministic and do not react to price changes
- No modelling of feedback effects between the seller and the market (e.g. Flash Crash 06/05/10)

$\Longrightarrow$ Need to understand price formation better!

Summary of A–Ch-type market impact modelling

- Revenues from a large sell/buy order may depend crucially on its execution
- The optimal execution strategy in turn may depend crucially on the criterion
- Almgren–Chriss models involve permament and temporary impact of trades on prices
- Under linear impacts and maximising revenues, it is optimal to sell at a constant speed
- Under linear impacts and among deterministic strategies, optimising mean-variance criterion, it is optimal to use a specific convex programme.

Stochastic Control approach to Optimal Execution

Model Setup

- $\nu = (\nu_t)_{\{0\leq t\leq T\}}$ is the trading rate, the speed at which the agent is liquidating or acquiring shares. It is also the variable the agent controls in the optimisation problem,
- $Q^\nu = (Q_t^\nu)_{\{0\leq t\leq T\}}$ is the agent's inventory, which is clearly affected by how fast she trades,
- $S^\nu = (S_t^\nu)_{\{0\leq t\leq T\}}$ is the midprice process, and is also affected in principle by the speed of her trading (permanent impact)
- $\hat{S}^\nu = (\hat{S}_t^\nu)_{\{0\leq t\leq T\}}$ corresponds to the price process at which the agent can sell or purchase the asset, i.e. the execution price, by walking the LOB (temporary impact), and
- $X^\nu = (X_t^\nu)_{\{0\leq t\leq T\}}$ is the agent's cash process resulting from the agent's execution strategy.

Model Setup

$$dQ_t^\nu = \pm\nu_t\,dt\,, \qquad Q_0^\nu = q\,, \tag{1a}$$

while the midprice is assumed to satisfy the following SDE,

$$dS_t^\nu = \pm g(\nu_t)\,dt + \sigma\,dW_t\,, \qquad S_0^\nu = S\,, \tag{1b}$$

where

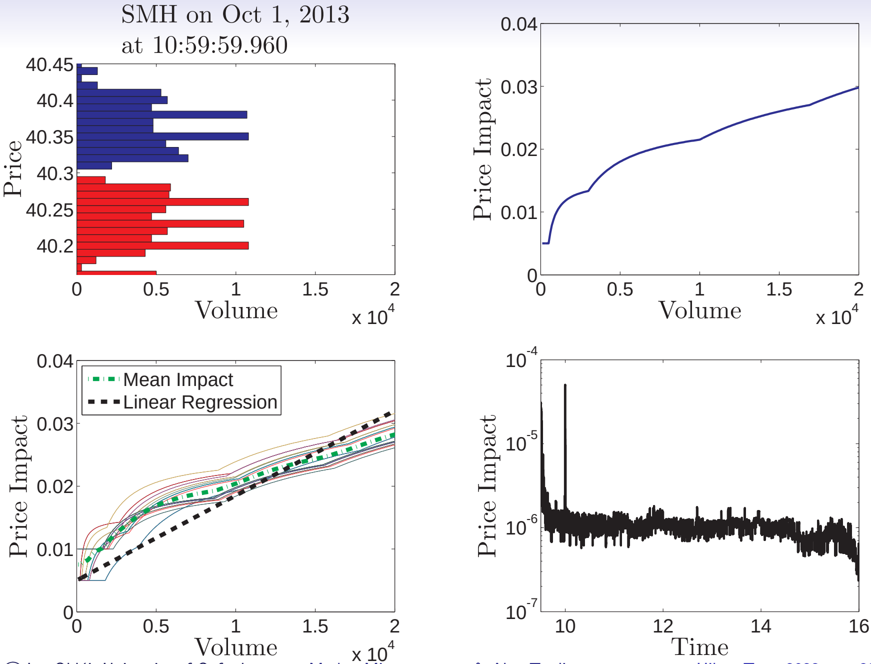
- $W = (W_t)_{\{0\le t\le T\}}$ is a standard Brownian motion,
- $g : \mathbb{R}_+ \to \mathbb{R}_+$ denotes the permanent impact that the agent's trading action incurs on the midprice.

The execution price satisfies the SDE

$$\hat{S}_t^\nu = S_t^\nu \pm \left(\tfrac{1}{2}\Delta + f(\nu_t)\right)\,, \qquad \hat{S}_0^\nu = \hat{S}\,, \tag{1c}$$

where *(note the small change of notation from h to f)*

- $f : \mathbb{R}_+ \to \mathbb{R}_+$ denotes the temporary impact that the agent's trading action has on the price they can execute the trade at.
- $\Delta \ge 0$ is the bid-ask spread, assumed here to be a constant.


SMH on Oct 1, 2013
at 10:59:59.960
Price
Volume
x 10^4
Price Impact
Volume
x 10^4
Mean Impact
Linear Regression
Price Impact
Volume
x 10^4
Price Impact
Time

Optimal Liquidation with Temporary Impact

Liquidation

$$H(t,S,q) = \sup_{\nu\in\mathcal{A}} \mathbb{E}_{t,S,q}\left[\int_t^T (S_u - k\,\nu_u)\,\nu_u\,du\right],$$

where $\mathbb{E}_{t,S,q}[\,\cdot\,]$ denotes expectation conditional on $S_t = S$ and $Q_t = q$, and $\mathcal{A}$ is the set of admissible strategies. Here $dQ_t = -\nu_t dt$.

Assume $f(\nu) = k\,\nu$, and $g(\nu) = 0$, i.e. only temporary price impact (walking the LOB).

Use the DPP which suggests that H satisfies the DPE (HJB PDE):

$$\partial_t H + \tfrac{1}{2}\sigma^2\,\partial_{SS} H + \sup_{\nu}\left\{(S - k\,\nu)\,\nu - \nu\,\partial_q H\right\} = 0\,. \tag{2}$$

The agent must liquidate all the inventory by time T, hence impose

$$H(T,S,q) \xrightarrow{t\to T} -\infty, \quad \text{for} \quad q > 0\,, \qquad \text{and} \qquad H(T,S,0) \xrightarrow{t\to T} 0\,.$$

Liquidation

The FOC applied to DPE (2):

$$\nu^* = \tfrac{1}{2k}\left(S - \partial_q H\right) , \tag{3}$$

which is the optimal trading speed in feedback control form. Now

$$\partial_t H + \tfrac{1}{2}\,\sigma^2\,\partial_{SS} H + \tfrac{1}{4\,k}\left(S - \partial_q H\right)^2 = 0 \tag{4}$$

for the value function. Propose ansatz (trial solution)

$$H(t, q) = q\,S + h(t, q)\,, \tag{5}$$

where $h(t, q)$ is still to be determined, though we know that it must blow up as $t \to T$.

Thus

$$\partial_t h + \tfrac{1}{4\,k}\left(\partial_q h\right)^2 = 0\,.$$

Interestingly, the volatility of the asset's midprice drops out of the problem.

Again, make ansatz $h(t,q) = q^2\, h_2(t)$ allows us to factor out q and obtain

$$\partial_t h_2 + \tfrac{1}{k}\, h_2^2 = 0\,, \tag{6}$$

which we solve by integrating between t and T to obtain

$$h_2(t) = \left(\frac{1}{h_2(T)} - \frac{1}{k}(T-t)\right)^{-1}.$$

The optimal strategy must ensure that the terminal inventory is zero and this is equivalent to requiring $h_2(T) \to -\infty$ as $t \to T$. In this way the value function heavily penalises non-zero final inventory. An alternative way to obtain this condition is to calculate the inventory path along the optimal strategy and impose that the terminal inventory be zero. To see this, use the ansatz (5) to reduce (3) to

$$\nu_t^* \;\; = \;\; -\tfrac{1}{k}\, h_2(t)\, Q_t^{\nu^*}\,, \tag{7}$$

then integrate $dQ_t^{\nu^*} = -\nu_t^* \, dt$ over $[0, t]$ to obtain the inventory profile along the optimal strategy:

$$\int_0^t \frac{dQ_s^{\nu^*}}{Q_s^{\nu^*}} = \int_0^t \frac{h_2(s)}{k} ds \quad \Rightarrow \quad Q_t^{\nu^*} = \frac{(T-t) - k/h_2(T)}{T - k/h_2(T)} \mathfrak{N} \, .$$

To satisfy the terminal inventory condition $Q_T^{\nu^*} = 0$, and also ensure that the correction $h(t, q)$ is negative, we must have

$$h_2(t) \to -\infty \quad \text{as} \quad t \to T \, . \tag{8}$$

Hence

$$h_2(t) = -k \, (T - t)^{-1} \, ,$$

so that

$$Q_t^{\nu^*} = \frac{T - t}{T} \mathfrak{N} \, , \tag{9}$$

and

$$\nu_t^* = \frac{\mathfrak{N}}{T} \, . \tag{10}$$

The result is TWAP.

Optimal Acquisition with Temporary Impact

Acquisition

Objective is to acquire $\mathfrak{N}$ shares by time T. We let $Q_0 = 0$ and *aim* to have $Q_T = \mathfrak{N}$.

Thus, the agent's expected costs from strategy ν_t are

$$\mathbb{E}\Big[\underbrace{\int_t^T \hat{S}_t \, \nu_u \, du}_{\text{Terminal Cash}} + \underbrace{(\mathfrak{N} - Q_T^\nu) \, S_T}_{\text{Terminal execution at mid}} + \underbrace{\alpha \, (\mathfrak{N} - Q_T^\nu)^2}_{\text{Terminal Penalty}} \Big] . \tag{11}$$

To simplify notation, let $Y = (Y_t)_{0 \le t \le T}$ denote the shares remaining to be purchased:

$$Y_t^\nu = \mathfrak{N} - Q_t^\nu \,, \qquad \text{so that} \qquad dY_t^\nu = -\nu_t \, dt \,,$$

and write the value function as

$$H(t, S, y) = \inf_{\nu \in \mathcal{A}} \mathbb{E}_{t,S,y} \left[\int_t^T \hat{S}_u \, \nu_u \, du + Y_T^\nu \, S_T + \alpha \, (Y_T^\nu)^2 \right] .$$

Applying the DPP, H satisfies the DPE

$$0 = \partial_t H + \tfrac{1}{2}\sigma^2 \partial_{SS} H + \inf_{\nu} \left\{ (S + k\nu)\,\nu - \nu \partial_y H \right\} \,, \tag{12}$$

with $H(T, S, y) = y\,S + \alpha\,y^2$. Optimal speed:

$$\nu^* = \tfrac{1}{2k} \left(\partial_y H - S \right) \,, \tag{13}$$

and upon substitution into the DPE above, we obtain

$$\partial_t H + \tfrac{1}{2}\sigma^2 \partial_{SS} H - \tfrac{1}{4\,k} \left(\partial_y H - S \right)^2 = 0 \,.$$

Make ansatz

$$H(t, S, y) = y\,S + h_0(t) + h_1(t)\,y + h_2(t)\,y^2 \,, \tag{14}$$

where $h_2(t)$, $h_1(t)$, $h_0(t)$ and

$$h_2(T) = \alpha \quad \text{and} \quad h_1(T) = h_0(T) = 0 \,.$$

Moreover, upon substituting the ansatz into the above non-linear PDE we find that

$$0 = \left\{\partial_t h_2 - \tfrac{1}{k}\, h_2^2\right\}\, y^2 + \left\{\partial_t h_1 - \tfrac{1}{2k}\, h_2\, h_1\right\}\, y + \left\{\partial_t h_0 - \tfrac{1}{4k}\, h_1^2\right\}\, .$$

- Equation must be valid for each y, so each term in curly braces must individually vanish.
- Due to $h_1(T) = 0$, we obtain $h_1(t) = 0$.
- Similarly, since $h_0(T) = 0$ and $h_1(t) = 0$ we obtain $h_0(t) = 0$.
- Finally, because $h_2(T) = \alpha$ we have

$$h_2(t) = \left(\tfrac{1}{k}\,(T - t) + \tfrac{1}{\alpha}\right)^{-1}\, .$$

Thus, the optimal speed is

$$\nu_t^* = \left((T - t) + \tfrac{k}{\alpha}\right)^{-1}\; Y_t^{\nu^*}\, . \tag{15}$$

Inventory path

We obtain the optimal inventory path by solving

$$dY_t^{\nu^*} = -\left((T-t)+\tfrac{k}{\alpha}\right)^{-1} \, Y_t^{\nu^*}\, dt$$

for $Y_t^{\nu^*}$. Recall that $Y_t^{\nu} = \mathfrak{N} - Q_t^{\nu}$, so

$$Q_t^{\nu^*} = \frac{t}{T+\frac{k}{\alpha}}\,\mathfrak{N}\,. \tag{16}$$

- For any finite $\alpha > 0$ and finite $k > 0$,
  - it is always optimal to leave some shares to be executed at the terminal date, and
  - the fraction of shares left to execute at the end decreases with the relative price impact at the terminal date, k/α.

Substitute for $Q_t^{\nu^*}$ into the expression for ν_t^*, so that

$$\nu_t^* = \frac{\mathfrak{N}}{T+\frac{k}{\alpha}}\,. \tag{17}$$

Optimal Liquidation with Temporary and Permanent Price Impact

Liquidation with Permanent Price Impact

The agent's performance criterion is

$$H^{\nu}(t,x,S,q) = \mathbb{E}_{t,x,S,q}\Big[\underbrace{X_T^{\nu}}_{\text{Terminal Cash}} + \underbrace{Q_T^{\nu}\,(S_T^{\nu} - \alpha\, Q_T^{\nu})}_{\text{Terminal Execution}} - \underbrace{\phi \int_t^T (Q_u^{\nu})^2\, du}_{\text{Inventory Penalty}} \Big] \tag{18}$$

and the value function

$$H(t,x,S,q) = \sup_{\nu \in \mathcal{A}} H^{\nu}(t,x,S,q)\,.$$

The DPP implies that the value function should satisfy the HJB equation

$$\begin{aligned} 0 = &\left(\partial_t + \tfrac{1}{2}\,\sigma^2\,\partial_{SS}\right) H - \phi\, q^2 \\ &+ \sup_{\nu} \left\{ \left(\nu\,(S - f(\nu))\,\partial_x - g(\nu)\,\partial_S - \nu\,\partial_q\right) H \right\}\,, \end{aligned} \tag{19}$$

subject to the terminal condition $H(T,x,S,q) = x + S\,q - \alpha\,q^2$.

Assume that $f(\nu) = k\,\nu$ and $g(\nu) = b\,\nu$ for constants $k \geq 0$ and $b \geq 0$ and obtain

$$\nu^* = \tfrac{1}{2\,k}\frac{(S\,\partial_x - b\,\partial_S - \partial_q)H}{\partial_x H}\,. \tag{20}$$

Hence

$$0 = \left(\partial_t + \tfrac{1}{2}\sigma^2\,\partial_{SS}\right)H - \phi\,q^2 + \tfrac{1}{4\,k}\frac{[(S\,\partial_x - b\,\partial_S - \partial_q)H]^2}{\partial_x H}\,.$$

Make ansatz $H(t,x,S,q) = x + S\,q + h(t,S,q)$ with terminal condition $h(T,S,q) = -\alpha\,q^2$. Hence

$$0 = \left(\partial_t + \tfrac{1}{2}\sigma^2\,\partial_{SS}\right)h - \phi\,q^2 + \tfrac{1}{4k}\left[b\,(q + \partial_S h) + \partial_q h\right]^2\,.$$

Since the above PDE contains no explicit dependence on S and the terminal condition is independent of S, then $\partial_S h(t,S,q) = 0$ and

$$0 = \partial_t h(t,q) - \phi\, q^2 + \tfrac{1}{4k}\left[b\, q + \partial_q h(t,q)\right]^2 \;.$$

Use separation of variables $h(t,q) = h_2(t)\, q^2$ where $h_2(t)$ to write

$$0 = \partial_t h_2 - \phi + \tfrac{1}{k}\left[h_2 + \tfrac{1}{2} b\right]^2 \;, \tag{21}$$

subject to $h_2(T) = -\alpha$. This ODE is of Riccati type and can be integrated exactly. Let $h_2(t) = -\frac{1}{2}b + \chi(t)$, then

$$\frac{\partial_t \chi}{k\phi - \chi^2} = \frac{1}{k}\,,$$

s.t. $\chi(T) = \frac{1}{2}b - \alpha$. Next, integrating over $[t,T]$

$$\log\frac{\sqrt{k\,\phi} + \chi(T)}{\sqrt{k\,\phi} - \chi(T)} - \log\frac{\sqrt{k\,\phi} + \chi(t)}{\sqrt{k\,\phi} - \chi(t)} = 2\gamma\,(T-t)\,,$$

so that

$$\chi(t) = \sqrt{k\,\phi}\,\frac{1 + \zeta\, e^{2\gamma\,(T-t)}}{1 - \zeta\, e^{2\gamma(T-t)}}, \quad \text{where } \gamma = \sqrt{\frac{\phi}{k}} \quad \text{and} \quad \zeta = \frac{\alpha - \frac{1}{2}b + \sqrt{k\,\phi}}{\alpha - \frac{1}{2}b - \sqrt{k\,\phi}}\,. \tag{22}$$

From (20), the optimal speed to trade at is

$$\nu_t^* = -\sqrt{\frac{\phi}{k}} \frac{1+\zeta\, e^{2\gamma\,(T-t)}}{1-\zeta\, e^{2\gamma(T-t)}}\, Q_t^{\nu^*} \ . \tag{23}$$

In addition

$$dQ_t^{\nu^*} = \frac{\chi(t)}{k}\, Q_t^{\nu^*}\, dt\,, \quad \text{so that} \quad Q_t^{\nu^*} = \mathfrak{N} \exp\left\{ \int_0^t \frac{\chi(s)}{k}\, ds \right\} \ .$$

To obtain the inventory process

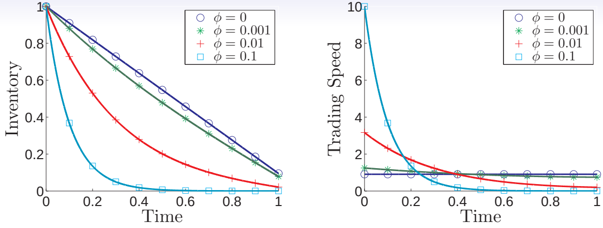
$$\begin{aligned} \int_0^t \frac{\chi(s)}{k}\, ds \quad &= \quad \frac{1}{k} \int_0^t \sqrt{k\phi} \frac{1+\zeta e^{2\gamma(T-s)}}{1-\zeta e^{2\gamma(T-s)}}\, ds \\ &= \quad \log \frac{\zeta\, e^{\gamma(T-t)} - e^{-\gamma(T-t)}}{\zeta\, e^{\gamma T} - e^{-\gamma T}} \ , \end{aligned}$$

hence

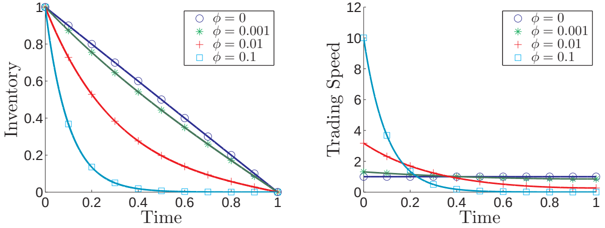
$$Q_t^{\nu^*} = \frac{\zeta\, e^{\gamma(T-t)} - e^{-\gamma(T-t)}}{\zeta\, e^{\gamma T} - e^{-\gamma T}}\, \mathfrak{N} \ . \tag{24}$$

The limit in $\alpha \to +\infty$ is independent of b:

$$Q_t^{\nu^*} \xrightarrow[\alpha\to+\infty]{} \frac{\sinh\left(\gamma(T-t)\right)}{\sinh\left(\gamma T\right)}\, \mathfrak{N} \ .$$



(a) $\alpha = 0.01$



(b) $\alpha = +\infty$

Figure: Other model parameters are $k = 10^{-3}$, $b = 10^{-3}$.

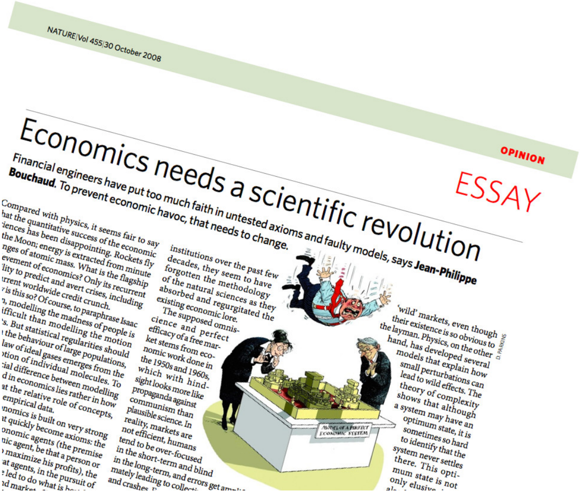
Outline

Market Microstructure: A Bottom-Up Approach

"*Market microstructure is the study of the process and outcomes of exchanging assets under a specific set of rules. While much of economics abstracts from the mechanics of trading, microstructure theory focuses on how specific trading mechanisms affect the price formation process.*" – M. O'Hara (1995)

In contrast to the top-down models where price is specified exogenously, market microstructure models attempt to explain how price formation *emerges* from the actions and interactions of the many different traders in a market.

Market Microstructure: A Bottom-Up Approach



NATURE|Vol 455|30 October 2008

OPINION

ESSAY

Economics needs a scientific revolution

Financial engineers have put too much faith in untested axioms and faulty models, says Jean-Philippe Bouchaud. To prevent economic havoc, that needs to change.

Compared with physics, it seems fair to say hat the quantitative success of the economic iences has been disappointing. Rockets fly the Moon; energy is extracted from minute nges of atomic mass. What is the flagship evement of economics? Only its recurrent lity to predict and avert crises, including rrent worldwide credit crunch.

is this so? Of course, to paraphrase Isaac n, modelling the madness of people is ifficult than modelling the motion s. But statistical regularities should the behaviour of large populations, law of ideal gases emerges from the otion of individual molecules. To ial difference between modelling d in economics lies rather in how at the relative role of concepts, empirical data.

nomics is built on very strong t quickly become axioms: the onomic agents (the premise nic agent, be that a person or maximize his profits), the at agents, in the pursuit of led to do what is

institutions over the past few decades, they seem to have forgotten the methodology of the natural sciences as they absorbed and regurgitated the existing economic lore.

The supposed omniscience and perfect efficacy of a free market stems from economic work done in the 1950s and 1960s, which with hindsight looks more like propaganda against communism than plausible science. In reality, markets are not efficient, humans tend to be over-focused in the short-term and blind in the long-term, and errors get ampli mately leading to collecti and crashes

'wild' markets, even though their existence is so obvious to the layman. Physics, on the other hand, has developed several models that explain how small perturbations can lead to wild effects. The theory of complexity shows that although a system may have an optimum state, it is sometimes so hard to identify that the system never settles there. This optimum state is not only elusive

Quote Driven Markets

Quote Driven Markets

Traditional markets operate thanks to Market Makers.

- Market Makers (Dealers) centralise buy and sell orders providing set bid and ask quotes
- They are patient: they publicise prices and wait for others to trade with them
- There are significant barriers to becoming a market maker: hold stock, have large bankroll, develop infrastructure...
- Market Makers offer a crucial service: provision of liquidity
- They make profit from crossing the bid-ask spread

Order Driven Markets

Electronic platforms aggregate all (or many) buy and sell orders in a

Limit Order Book (LOB)

Traders

- post limit orders (LO) to sell or buy a given quantity at a given price
  $\rightsquigarrow$ stored in the LOB
- post market orders (MO) to sell or buy a given quantity instantly
  $\rightsquigarrow$ matched against orders in the LOB
- can cancel their outstanding active limit orders

Order Driven Markets

New York Stock Exchange
Hong Kong Stock Exchange
Shenzhen Stock Exchange
Australian Securities Exchange
Helsinki Stock Exchange
Euronext

Limit order books are used to match buyers and sellers in more than half of the world's financial markets

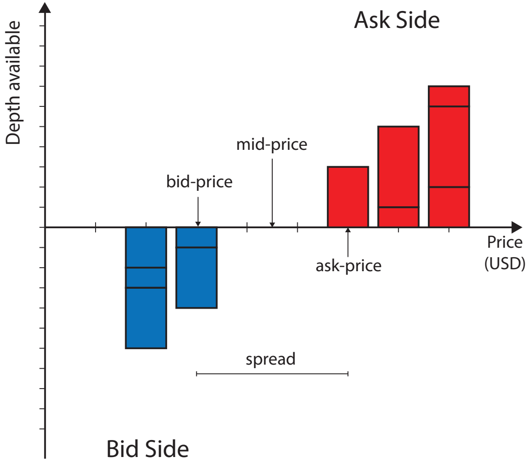
Toronto Stock Exchange
London Stock Exchange
Tokyo Stock Exchange
Vancouver Stock Exchange
Swiss Stock Exchange

LOB vocabulary

- tick size – smallest possible interval between consecutive prices
- minimum order size (lot) – smallest quantity of shares which can be traded
- Ask Side – all sell orders in the LOB.
  Ask Price a_t – the lowest price among active sell orders.
- Bid Side – all buy orders in the LOB.
  Bid Price b_t – the highest price among active buy orders.
- **Mid Price** $m_t = \frac{a_t + b_t}{2}$
  Bid–Ask Spread $= a_t - b_t$
- **Depth** at a given price level p – the aggregate volume of shares to trade in orders at price p.

Price Formation in a Limit Order Book

A Dynamic Example



All LOB graphics from Martin Gould '15

LOB – auctions & queues

LOB is composed of two sets of active orders:

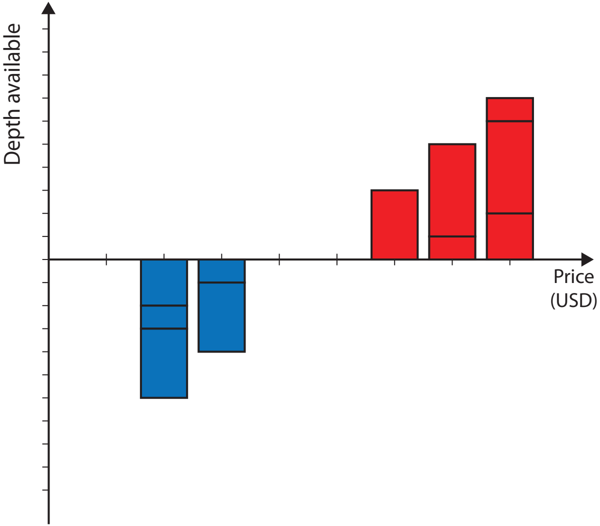
- buy orders (quantity < 0)
- sell orders (quantity > 0)

LOBs are often referred to as continuous double-auction mechanisms.

A LOB can also be regarded as a set of queues, each of which contains active buy or sell orders at a specified price.

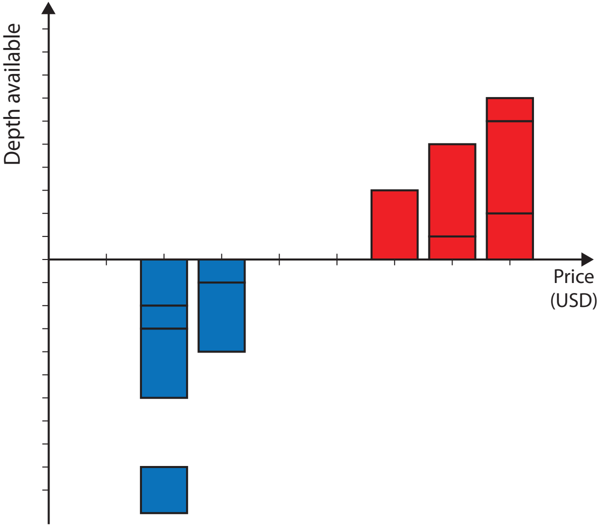
Price Formation in a Limit Order Book

A Dynamic Example



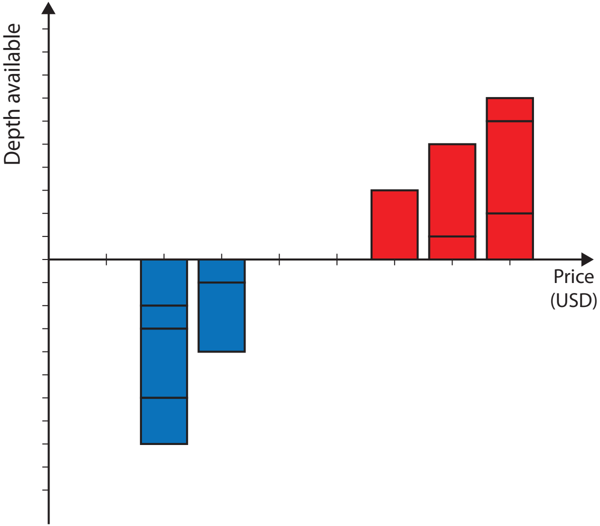
Price Formation in a Limit Order Book

A Dynamic Example



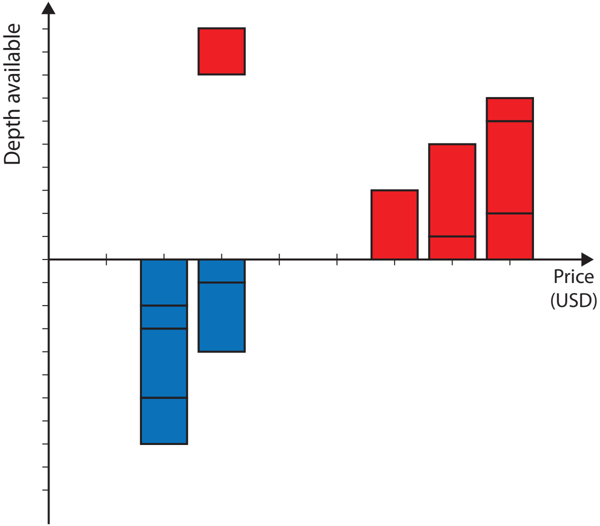
Price Formation in a Limit Order Book

A Dynamic Example



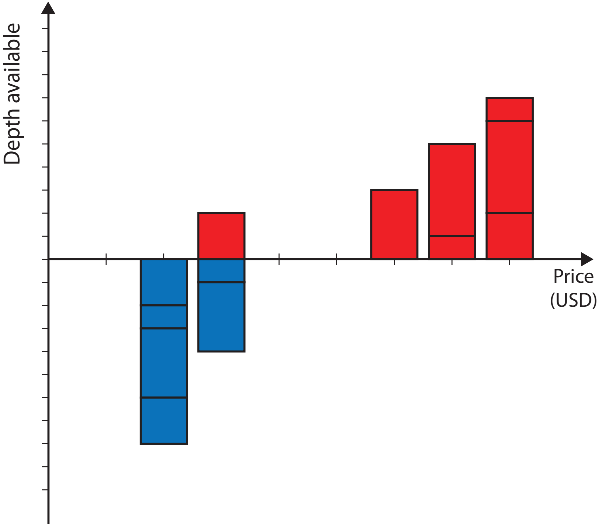
Price Formation in a Limit Order Book

A Dynamic Example



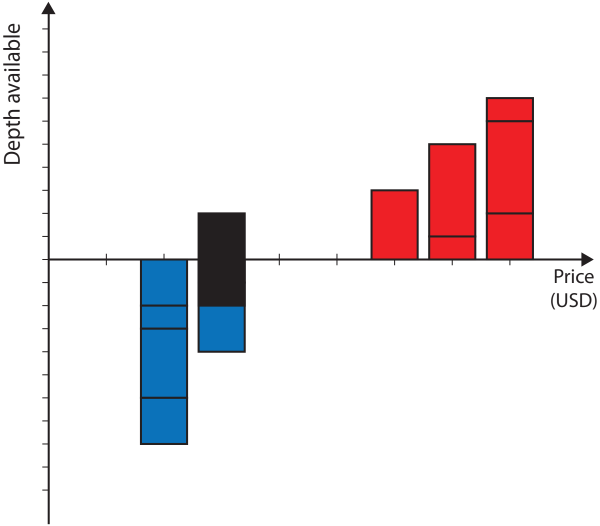
Price Formation in a Limit Order Book

A Dynamic Example



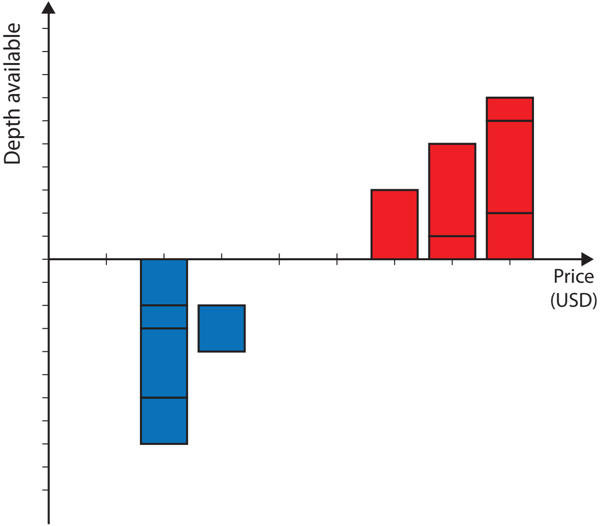
Price Formation in a Limit Order Book

A Dynamic Example



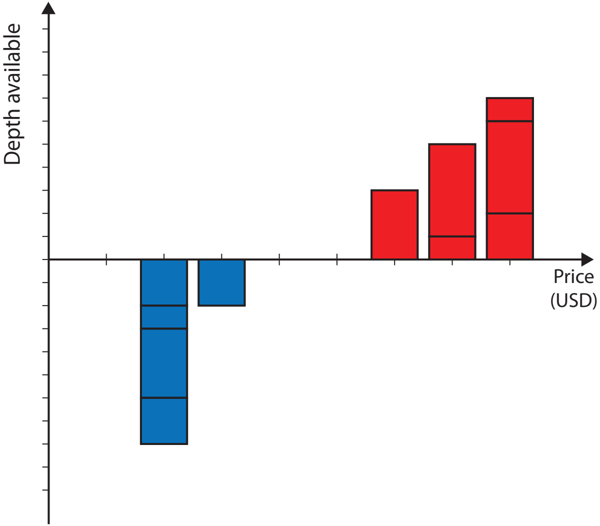
Price Formation in a Limit Order Book

A Dynamic Example



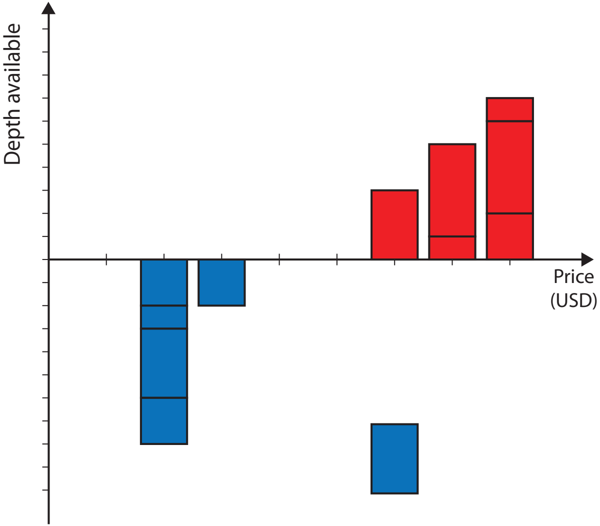
Price Formation in a Limit Order Book

A Dynamic Example



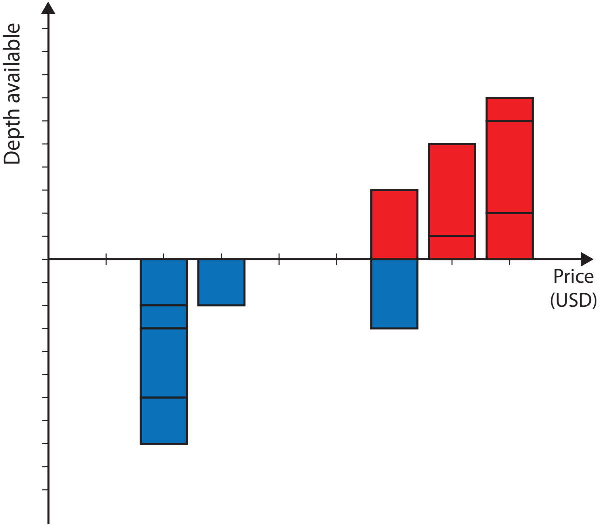
Price Formation in a Limit Order Book

A Dynamic Example



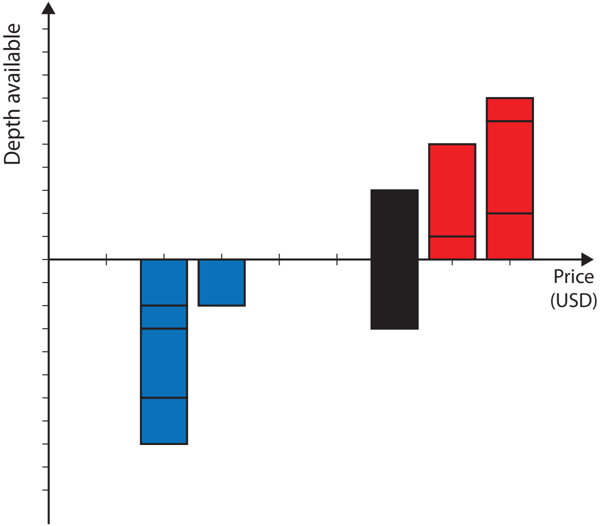
Price Formation in a Limit Order Book

A Dynamic Example



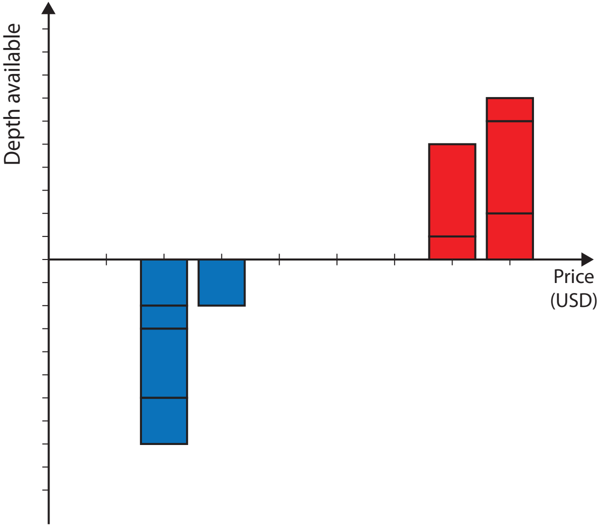
Price Formation in a Limit Order Book

A Dynamic Example

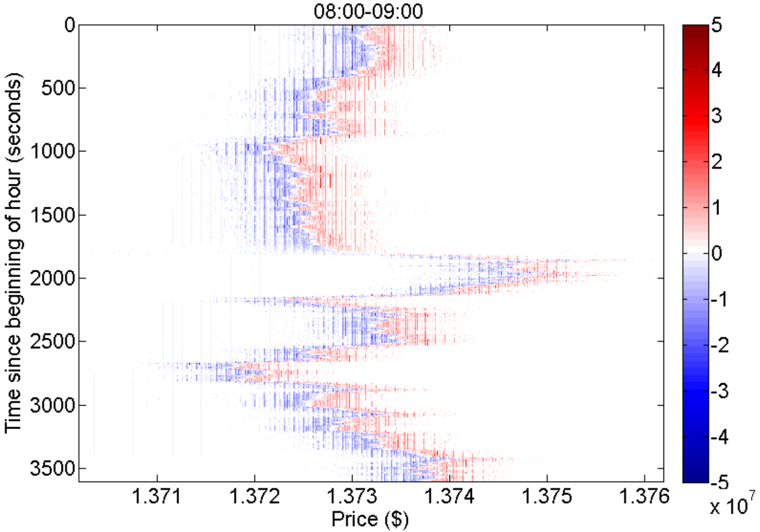


Price Formation in a Limit Order Book

A Dynamic Example



Price Formation in a Limit Order Book



LOB Dynamics Summary

- LOB holds all active buy and sell orders
- Traders may post limit orders in the LOB:
  - order to buy/sell a given quantity at given price
  - known execution cost
  - unknown execution time
- Traders may issue market orders:
  - order to buy/sell a given quantity against liquidity in the LOB
  - unknown execution cost
  - instant execution (if book deep enough)
- Traders may cancel their active limit orders in the LOB
- When current best bid/ask queue depleted by market order or cancellations, the price moves to the next level of the LOB.

Price Changes in a LOB

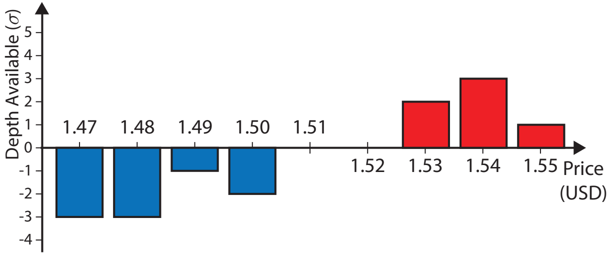
If limit buy/sell order x arrives with price $p_x \leq b_t$ / $p_x \geq a_t$ then x joins the LOB.

If limit buy/sell order x arrives with price $b_t < p_x < a_x$ then x joins the LOB and bid/ask changes.

If limit buy/sell order x arrives with price $p_x \geq a_t$ / $p_x \leq b_t$ then x is a market order that immediately matches to one, or more if needed, active sell (respectively, buy) orders upon arrival.
Whenever such a matching occurs, it does so at the price of the active order, which is not necessarily equal to the price of the incoming order.

An incoming market order x matches to the highest priority active order y of opposite type, and then the following one ... until x is fulfilled or book depleted.

Price Changes in a LOB



Arriving order x	Values after arrival (USD)			
	$b(t_x)$	$a(t_x)$	$m(t_x)$	$a(t_x) - b(t_x)$
Initial Values	1.50	1.53	1.515	0.03
$(\$1.48, -3\sigma, t_x)$	1.50	1.53	1.515	0.03
$(\$1.51, -3\sigma, t_x)$	1.51	1.53	1.52	0.02
$(\$1.55, -5\sigma, t_x)$	1.50	1.55	1.525	0.05
$(\$1.47, 4\sigma, t_x)$	1.48	1.53	1.505	0.05
$(\$1.50, 4\sigma, t_x)$	1.49	1.50	1.495	0.01

Order Imbalance

Depth (Volume) available at the best bid/ask – and its imbalance – is often informative

$$\text{microprice} = \text{volume weighted midprice} = \rho_t a_t + (1 - \rho_t) b_t$$

where ρ_t is the order imbalance

$$\rho_t = \frac{V_t^b}{V_t^a + V_t^b}$$

It is a good predictor of trade direction
(ORIT June 21, 2011)

ρ	# Buy Orders	# Sell Orders
All	**756 (67%)**	396 (33 %)
> 0.5	**568 (79%)**	155 (21%)
> 0.75	**320 (84%)**	60 (16%)
< 0.5	168 (43%)	**225 (57%)**
< 0.25	39 (25%)	**116 (75%)**

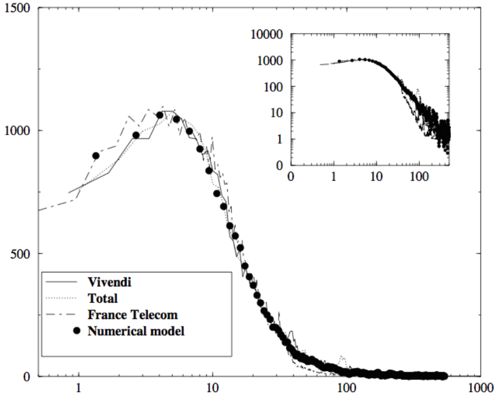
Order Imbalance

A slice of imbalance for MSFT 10:00am to 10:10am on 22 Mar 2011

The Depth Profile

Most traders assess the state of the LOB via the depth profile

Depth displays stylised facts when measured in ticks from the bid/ask



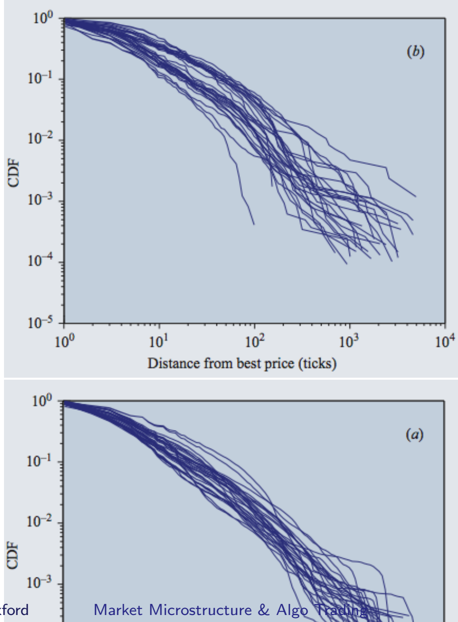
Mean relative depths profiles for Vivendi, Total, and France Telecom stocks (from "Statistical Properties of Stock Order Books...", Bouchaud *et al.* '02).

Relative Price

The distribution of relative prices of incoming orders appears to exhibit power-law behaviour in all markets studied.

This may be because some traders place limit orders deep into LOBs, under the optimistic belief that large price swings could occur.

Relative Price



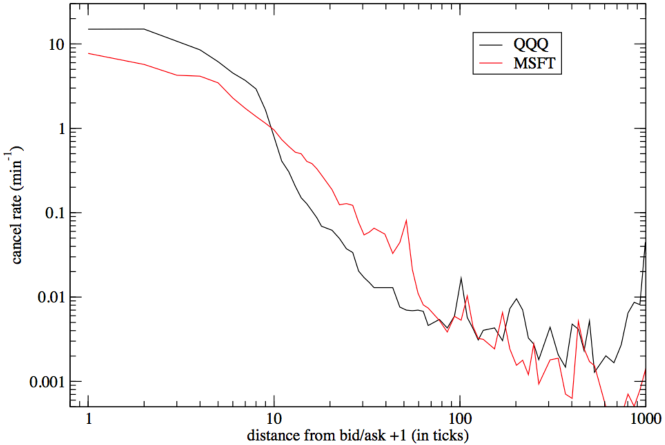
Order Cancellations

Cancellations play a major role in the evolution of the LOB.

Several empirical studies covering a wide range of different markets have concluded that the vast majority of active orders end in cancellation rather than matching. This may be between 70% and 80% and possibly up to 99.9% of orders more recently.

This is often attributed to surge of popularity of electronic trading algorithms, across all markets, which often submit and cancel vast numbers of limit orders over short periods as part of their trading strategies.

Order Cancellations



Cancellation rate for NASDAQ Index and Microsoft Stock (from “More Statistical Properties of Order Books and Price Impact”, Potters and Bouchaud ’03).

LOBs Impact – pros

The shift from traditional markets to electronic LOB driven markets had many consequences. Some positive:

- competition leading to lower fees and smaller tick sized
- more information available
- democratised trading process
- choice of patient (limit) or impatient (market) orders available to everyone
- computerised/algorithmic trading possible
- high frequency trading possible
  - HFT $\approx$ duration of order of seconds, reaction within milliseconds
  - accounts for $60 - 75\%$ of traded volume
- extra provision of liquidity $\rightsquigarrow$ market efficiency

LOBs Impact – cons

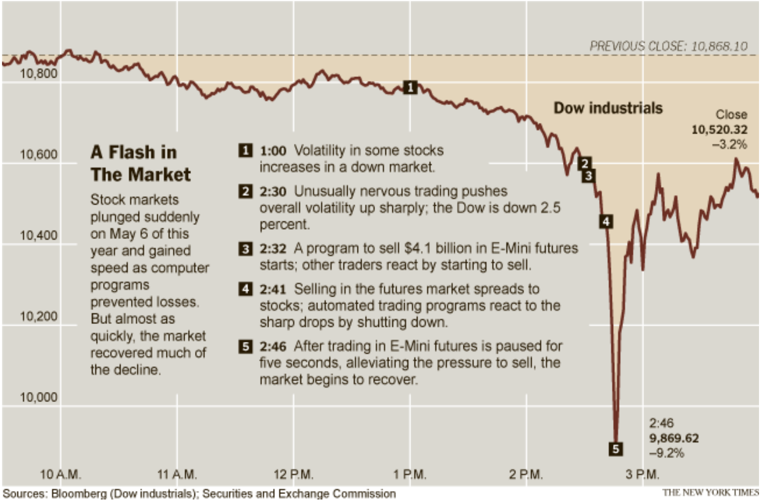
And some negative:

- Technological armsrace
- Little human oversight
- Predatory trading

This led to the infamous Flash Crash of May 6, 2010 when Dow Jones IA (DJIA) dived almost 1000 points (just to recover in minutes). All this because:

- A mutual fund activated a program to sell $75,000$ E-Mini S&P 500 contracts (≈ 4.1 billion USD) using VWAP algorithm at 9%
- HFT began to quickly buy and resell these contracts to each other generating more volume: between 2:45:14 and 2:45:27, HFT traded $27,000$ contracts (about 49% of total volume) while buying only 200 contracts net.
- This led the original program to rapidly sell the whole position

Flash Crash of May 6th, 2010



Why Model/Study LOBs?

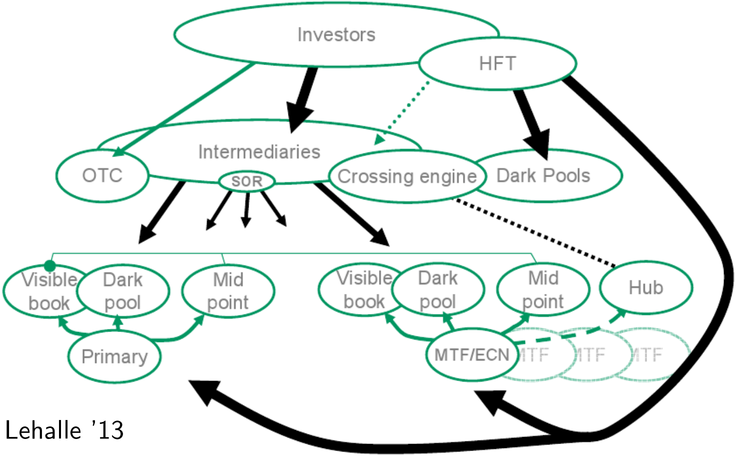
The benefits of having a good model of a LOB include:

- Enhancing the effectiveness of electronic trading algorithms
- Developing optimal execution strategies for traders
- Understanding costs of delayed trading
- Optimising market-making (liquidity providing) algorithms
- Understanding stylised facts about LOBs
- Developing detailed but tractable models of bid-ask spread and transaction costs
- Providing insight into fundamental economic questions
- Testing theories regarding complex systems as a whole

(some) LOB studies

- Empirical studies:
  Bouchaud et al '02, Farmer et al. '04, Hollifield et al. '04...
- Equilibrium models:
  Kyle '85, Parlour '98, Foucault et al. '05, Rosu '09...
- Reduced form models:
  - Stochastic dynamic models (LOB a Markov process in a high-dim state space):
    Smith et al. '03, Bovier et al. '06, Bouchaud et al. '08...
  - Queuing theory models:
    Cont et al. '10...

LOBs: Reality is more complex...



Source: Lehalle '13

MTF=Multilateral Trading Facility; ECN = Electronic Communications Network;

SOR = Smart Order Routing; OTC = Over The Counter

LOBs: Reality is more complex...

- the same shares traded on many venues – smart routing necessary
- consolidated LOB for all venues may (US) or may not (EU) be readily available
- various execution order conventions: price-time priority (FIFO), price-size priority, pro-rata priority $\rightsquigarrow$ strategic order posting
- hidden (iceberg orders) or invisible (dark pool) liquidity $\rightsquigarrow$ fishing, price manipulation, predatory trading

Priority

Price-time priority

- For active buy orders, priority is given to the active orders with the highest price.
- For active sell orders, priority is given to the active orders with the lowest price.
- Ties are broken by selecting the active order with the earliest submission time.

Price-time priority is an effective way to encourage traders to place limit orders. Without a priority mechanism based on time, there is no incentive for traders to show their hand by submitting limit orders earlier than is absolutely necessary.

Priority

Price-size priority

- For active buy orders, priority is given to the active orders with the highest price.
- For active sell orders, priority is given to the active orders with the lowest price.
- Ties are broken by selecting the active order with the largest size.

Price-time priority is an effective way to encourage traders to place large limit orders, thereby providing liquidity to the market.

Priority

Pro-rata priority

- For active buy orders, priority is given to the active orders with the highest price.
- For active sell orders, priority is given to the active orders with the lowest price.
- When a tie occurs at a given price, each relevant active order receives a share of the matching proportional to the fraction of the depth available that it represents at that price.
- Traders in pro-rata priority LOBs are faced with the substantial difficulty of optimally selecting limit order sizes, because posting limit orders with larger sizes than the quantity that is really desired for trade becomes a viable strategy to gain priority.

Priority

Pro-rata priority

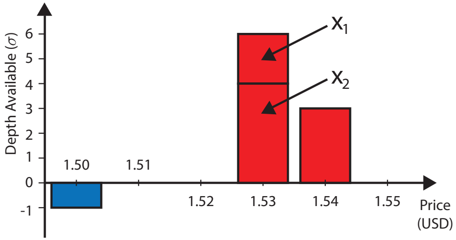


Figure: An LOB with pro-rata priority.

If a buy market order of size -3 lots arrived, then -1 lot would match to active order x_1 and -2 lots would match to active order x_2, because they correspond, respectively, to $1/3$ and $2/3$ of the depth available at a_t.

Priority–triggered behaviour

Different priority mechanisms encourage traders to behave in different ways:

- Price-time priority encourages traders to submit limit orders early
- Price-size and pro-rata priority reward traders for placing large limit orders and thus for providing greater liquidity to the market

Traders' behaviour is closely related to the priority mechanism used, so LOB models need to take priority mechanisms into account when considering order flow. Furthermore, priority plays a pivotal role in models that attempt to track specific orders.

Hidden Liquidity

An iceberg order is a type of limit order that specifies not only a total size and price but also a visible size. Other market participants only see the visible size.

Rules regarding the treatment of the hidden quantity vary greatly from one exchange to another:

- In some cases, once a quantity of at least the visible size matches to an incoming market order, another quantity equal to the visible size becomes visible, with time priority equal to that of a standard limit order placed at this time.
- Some other trading platforms, such as Currenex and Hotspot FX, allow entirely hidden limit orders. These orders are given priority behind both entirely visible active orders at their price and the visible portion of iceberg orders at their price, but they give market participants the ability to submit limit orders without revealing any information whatsoever to the market.

Dark Pools

Recently, there has also been an increase in the popularity of so-called dark pools, particularly in equities trading.

- electronic engine matching buy and sell order without routing to lit exchanges
- no information about market participants' trading intentions is available to other market participants
- some dark pools are essentially LOBs in which all limit orders are entirely hidden
- other dark pools are time-priority queues of buy/sell orders (no prices specified), trading at mid-point of a reference (lit) exchange
- allow to trade large amounts without impacting the price – over 30% of all trades!

LOBs Summary

- Electronic markets operate without designated market maker.
- Instead, the Limit Order Book (LOB) holds all active buy and sell orders
- Traders may post orders in the LOB: limit or market, as well as cancel active limit orders.
  When current best bid/ask queue depleted by market order or cancellations, the price moves to the next level of the LOB.
- LOBs democratised trading, made High Frequency Trading (HFT) possible and increased provision of liquidity
- but also led to technological armsrace, predatory trading, algorithm-triggered crashes
- In reality large part of liquidity hidden in iceberg orders and/or dark pools
- Practice is more complex with many layers and implementation challenges

Outline

Market Making

- $S_t = S_0 + \sigma\, W_t$, $\sigma > 0$ and $W = (W_t)_{0 \le t \le T}$ is a standard Brownian motion,
- $\delta^{\pm}$ depth at which the agent posts LOs. Sell LOs are posted at a price of $S_t + \delta_t^+$ and buy LOs at $S_t - \delta_t^-$.
- $M^{\pm}$ counting processes corresponding to the arrival of other participants' buy $(+)$ and sell $(-)$ market orders (MOs) which arrive at Poisson times with intensities $\lambda^{\pm}$.
- $N^{\delta,\pm}$ denote the counting processes for the agent's filled sell $(+)$ and buy $(-)$ LOs.
- Conditional on an MO arrival, the LO is filled with probability $e^{-\kappa^{\pm}\,\delta_t^{\pm}}$ with $\kappa^{\pm} \ge 0$.
- X^{δ} denotes the MM's cash process

$$dX_t^{\delta} = (S_{t-} + \delta_t^+)\, dN_t^{\delta,+} - (S_{t-} - \delta_t^-)\, dN_t^{\delta,-}\,. \tag{25}$$

- Q^{δ} denotes the agent's inventory process and satisfies the SDE

$$Q_t^{\delta} = N_t^{\delta,-} - N_t^{\delta,+}\,. \tag{26}$$

Market Maker's Control Problem

The MM's performance criterion is

$$H^{\delta}(t,x,S,q) = \mathbb{E}_{t,x,q,S}\left[X_T + Q_T\,(S_T - \alpha\, Q_T) - \phi \int_t^T (Q_u)^2\, du\right],$$

where $\alpha \geq 0$ represents the fees for taking liquidity (i.e. using an MO) as well as the impact of the MO walking the LOB, and $\phi \geq 0$ is the running inventory penalty parameter. The MM's value function is

$$H(t,x,S,q) = \sup_{\delta^{\pm} \in \mathcal{A}} H^{\delta}(t,x,S,q)\,, \tag{27}$$

and the MM caps her inventory so that it is bounded above by $\overline{q} > 0$ and below by $\underline{q} < 0$.

DPE

A DPP holds and the value function satisfies the DPE

$$\begin{aligned} 0 = {} & \partial_t H + \tfrac{1}{2}\sigma^2 \partial_{SS} H - \phi q^2 \\ & + \lambda^+ \sup_{\delta^+} \left\{ e^{-\kappa^+\delta^+} \left(H(t, x + (S + \delta^+), q - 1, S) - H \right) \right\} \mathbb{1}_{q > \underline{q}} \\ & + \lambda^- \sup_{\delta^-} \left\{ e^{-\kappa^-\delta^-} \left(H\left(t, x - (S - \delta^-), q + 1, S\right) - H \right) \right\} \mathbb{1}_{q < \overline{q}} \,, \end{aligned} \tag{28}$$

where $\mathbb{1}$ is the indicator function, with terminal condition

$$H(T, x, S, q) = x + q(S - \alpha q)\,. \tag{29}$$

Recall that inventory is bounded, thus when $q_t = \overline{q}$ ($\underline{q}$) the optimal strategy is to post one-sided LOs which are obtained by solving (28) with the term proportional to λ^- (λ^+) absent as stated by the indicator function $\mathbb{1}$ in the DPE. Alternatively, one can view these boundary cases as imposing $\delta^- = +\infty$ and $\delta^+ = +\infty$ when $q = \overline{q}$ and $\underline{q}$ respectively.

Solving HJB

Make an ansatz for H. In particular, write

$$H(t, x, q, S) = x + q\,S + h(t, q)\,. \tag{30}$$

The first term is the accumulated cash, the second term is the book value of the inventory marked-to-market, and the last term is the added value from following an optimal market making strategy up to the terminal date. Thus,

$$\begin{aligned}\phi\, q^2 = \partial_t h(t,q) &+ \lambda^+ \sup_{\delta^+} \left\{ e^{-\kappa^+\delta^+} \left(\delta^+ + h(t, q-1) - h(t,q) \right) \right\} \mathbb{1}_{q>\underline{q}} \\ &+ \lambda^- \sup_{\delta^-} \left\{ e^{-\kappa^-\delta^-} \left(\delta^- + h(t, q+1) - h(t,q) \right) \right\} \mathbb{1}_{q<\overline{q}}\,,\end{aligned} \tag{31}$$

with terminal condition $h(T, q) = -\alpha q^2$.

Optimal Controls

Then the optimal depths in feedback form are given by

$$\delta^{+,*}(t,q) = \frac{1}{\kappa^+} - h(t,q-1) + h(t,q)\,, \quad q \neq \underline{q}\,, \tag{32a}$$

$$\delta^{-,*}(t,q) = \frac{1}{\kappa^-} - h(t,q+1) + h(t,q)\,, \quad q \neq \overline{q}\,, \tag{32b}$$

and the boundary cases are $\delta^{+,*}(t,q) = +\infty$ and $\delta^{-,*}(t,q) = +\infty$ when $q = \underline{q}$ and $\overline{q}$ respectively.

Substituting the optimal controls into the DPE we obtain

$$\begin{aligned}\phi\, q^2 = \partial_t h(t,q) &+ \tfrac{\lambda^+}{\kappa^+} e^{-1} e^{-\kappa^+(-h(t,q-1)+h(t,q)}\, \mathbb{1}_{q>\underline{q}} \\ &+ \tfrac{\lambda^-}{\kappa^-} e^{-1} e^{-\kappa^-(-h(t,q+1)+h(t,q))}\, \mathbb{1}_{q<\overline{q}}\,.\end{aligned} \tag{33}$$

Symmetric fill probability

It is possible to find an analytical solution to the DPE if the fill probabilities of LOs is the same on both sides of the LOB. In this case if $\kappa = \kappa^+ = \kappa^-$ then write

$$h(t, q) = \frac{1}{\kappa} \log \omega(t, q)\,,$$

and stack $\omega(t, q)$ into a vector

$$\boldsymbol{\omega}(t, q) = \left[\omega(t, \overline{q}),\, \omega(t, \overline{q} - 1), \ldots, \omega(t, \underline{q})\right]'\,.$$

Now, let A denote the $(\overline{q} - \underline{q} + 1)$-square matrix whose rows are labeled from $\overline{q}$ to $\underline{q}$ and whose entries are given by

$$\mathrm{A}_{i,q} = \begin{cases} -\phi\kappa\, q^2\,, & i = q\,, \\ \lambda^+\, e^{-1}\,, & i = q - 1\,, \\ \lambda^-\, e^{-1}\,, & i = q + 1\,, \\ 0\,, & \text{otherwise.} \end{cases} \tag{34}$$

(33) can now be rewritten as the coupled system of equations

$$\partial_t \boldsymbol{\omega}(t) + \mathsf{A}\boldsymbol{\omega}(t) = 0. \tag{35}$$

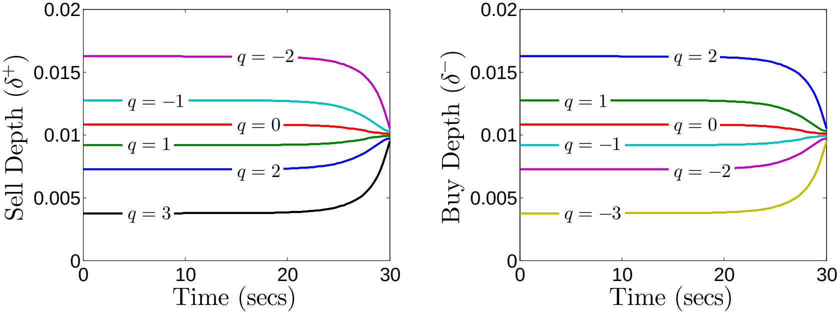
with terminal and boundary conditions $\omega(T, q) = e^{-\alpha\kappa q^2}$.
This is solved giving

$$\boxed{\boldsymbol{\omega}(t) = e^{\mathsf{A}(T-t)}\mathsf{z}\,,} \tag{36}$$

where z is a $(\overline{q} - \underline{q} + 1)$-dim vector where each component is $z_j = e^{-\alpha\kappa\, j^2}$, $j = \overline{q}, \ldots, \underline{q}$.

Optimal Postings

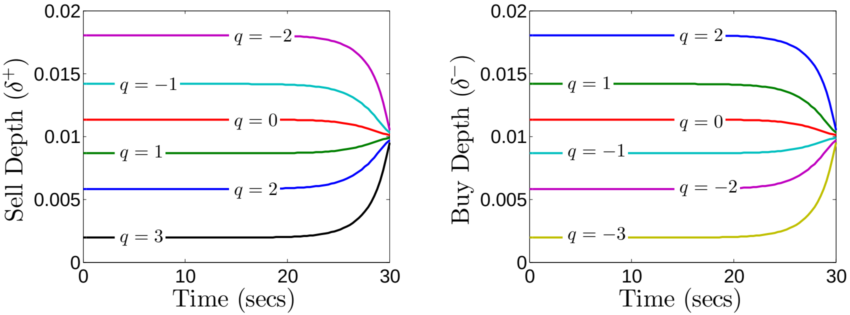
Optimal postings $\phi = 0.001$



(a) $\phi = 0.001$

Figure: The optimal depths as a function of time for various inventory levels and $T = 30$. The remaining model parameters are: $\lambda^{\pm} = 1$, $\kappa^{\pm} = 100$, $\overline{q} = -\underline{q} = 3$, $\phi = 0.001$ and $\phi = 0.02$, $\alpha = 0.0001$, $\sigma = 0.01$, $S_0 = 100$.

Optimal postings $\phi = 0.02$



(a) $\phi = 0.02$

Figure: The optimal depths as a function of time for various inventory levels and $T = 30$. The remaining model parameters are: $\lambda^{\pm} = 1$, $\kappa^{\pm} = 100$, $\overline{q} = -\underline{q} = 3$, $\phi = 0.001$ and $\phi = 0.02$, $\alpha = 0.0001$, $\sigma = 0.01$, $S_0 = 100$.

Mean reversion in inventory

Given the pair of optimal strategies $\delta^+(t,q), \delta^-(t,q)$, the expected drift in inventories q_t is given by

$$\boxed{\begin{aligned}\mu(t,q) &\triangleq \lim_{s \downarrow t} \frac{1}{s-t} \mathbb{E}\left[Q_s - Q_t \,|\, Q_{t^-} = q\right] \\ &= \lambda^- e^{-\kappa^- \delta^{-,*}(t,q)} - \lambda^+ e^{-\kappa^+ \delta^{+,*}(t,q)} .\end{aligned}} \tag{37}$$

Note that the drift $\mu(t,q)$ depends on time. For instance it is clear that for the same level of inventory the speed will be different depending on how near of far is the strategy from the terminal date because at time T the strategy tries to unwind all outstanding inventory.

Inventory

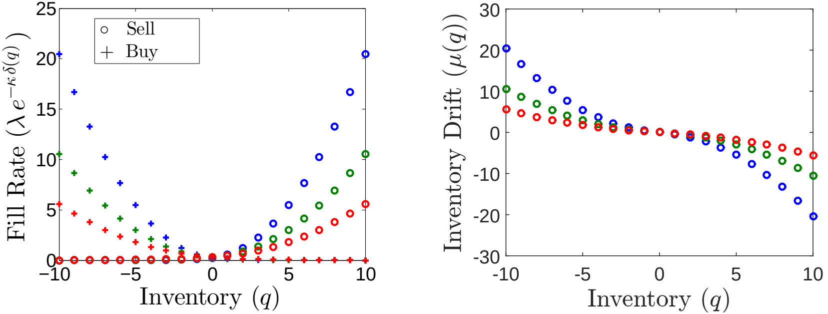


Figure: Long-term inventory level. Model parameters are: $\lambda^{\pm} = 1$, $\kappa^{\pm} = 100$, $\overline{q} = -\underline{q} = 10$, $\alpha = 0.0001$, $\sigma = 0.01$, $S_0 = 100$, and $\phi = \{2 \times 10^{-3}, 10^{-3}, 5 \times 10^{-4}\}$.

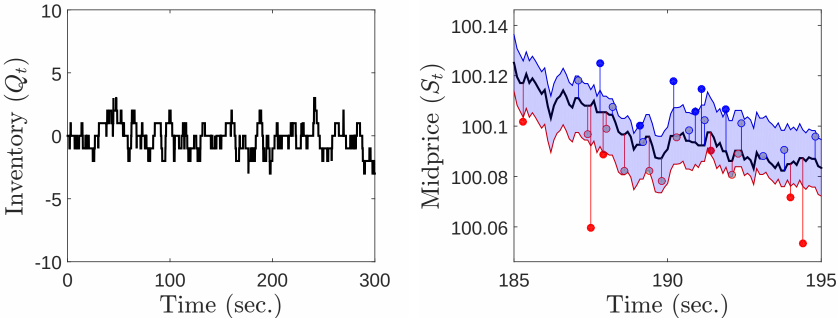
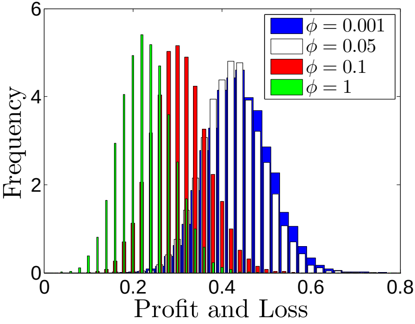


Figure: Inventory and midprice path. Model parameters are: $\lambda^{\pm} = 1$, $\kappa^{\pm} = 100$, $\overline{q} = -\underline{q} = 10$, $\phi = 0.02$, $\alpha = 0.0001$, $\sigma = 0.01$, $S_0 = 100$.

Profit and Loss



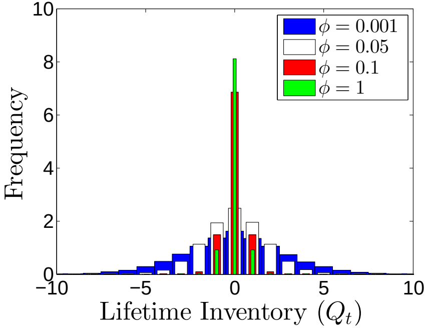


Figure: P&L and Life Inventory of the optimal strategy for 10,000 simulations. The remaining model parameters are: $\lambda^{\pm} = 1$, $\kappa^{\pm} = 100$, $\overline{q} = -\underline{q} = 10$, $\alpha = 0.0001$, $\sigma = 0.01$, and $S_0 = 100$.

Market Making at-the-touch

Market Making at-the-touch

Throughout we assume that the spread is constant and equal to Δ. Next, let $\ell_t^{\pm} \in \{0,1\}$ denote whether the agent is posted on the sell side $(+)$ or buy side $(-)$ of the LOB. In this way, the agent may be posted on both sides of the book, only the sell side, only the buy side, or not posted at all. Her performance criteria is

$$H^{\boldsymbol{\ell}}(t,x,S,q) = \mathbb{E}_{t,x,S,q}\left[X_T^{\boldsymbol{\ell}} + Q_T^{\boldsymbol{\ell}}\ \left(S_T - \left(\tfrac{\Delta}{2} + \varphi\, Q_T^{\boldsymbol{\ell}}\right)\right) - \phi \int_t^T \left(Q_u^{\boldsymbol{\ell}}\right)^2 du\right],$$

where her cash process $X_t^{\boldsymbol{\ell}}$ now satisfies the SDE

$$dX_t^{\boldsymbol{\ell}} = \left(S_t + \tfrac{\Delta}{2}\right)\, dN_t^{+,\boldsymbol{\ell}} - \left(S_t - \tfrac{\Delta}{2}\right)\, dN_t^{-,\boldsymbol{\ell}},$$

where $N_t^{\pm,\boldsymbol{\ell}}$ denote the counting process for filled LOs. We also further assume that, if she is posted in the LOB, when a matching MO arrives her LO is filled with probability one. In this case, $N_t^{\pm,\boldsymbol{\ell}}$ are controlled doubly stochastic Poisson processes with intensity $\ell_t^{\pm}\,\lambda^{\pm}$.

The agent is not posted on the buy (sell) side if her inventory is equal to the upper (lower) inventory constraints $\overline{q}$ ($\underline{q}$) and her value function is denoted by

$$H(t,x,S,q) = \sup_{\boldsymbol{\ell}\in\mathcal{A}} H^{\boldsymbol{\ell}}(t,x,S,q)\,.$$

The Resulting DPE

Applying the DPP, we find the agent's value function H should satisfy the DPE

$$\begin{aligned} 0 = & \left(\partial_t + \tfrac{1}{2}\sigma^2\partial_{SS}\right)H - \phi\, q^2 \\ & + \lambda^+ \max_{\ell^+\in\{0,1\}} \left\{\left(H\left(t, x + \left(S + \tfrac{\Delta}{2}\right)\, \ell^+, S, q - \ell^+\right) - H\right)\right\} \mathbb{1}_{q>\underline{q}} \\ & + \lambda^- \max_{\ell^-\in\{0,1\}} \left\{\left(H\left(t, x - \left(S - \tfrac{\Delta}{2}\right)\, \ell^-, S, q + \ell^-\right) - H\right)\right\} \mathbb{1}_{q<\overline{q}}\,, \end{aligned}$$

subject to the terminal condition

$$H(T, x, S, q) = x + q\, \left(S - \left(\tfrac{\Delta}{2} + \varphi\, q\right)\right)\,.$$

Ansatz:

$$H(t, x, S, q) = x + q\,S + h(t, q)\,,$$

and on substituting this ansatz into the above DPE we find that h satisfies

$$\begin{aligned} 0 &= \partial_t h - \phi\, q^2 \\ &\quad + \lambda^+ \max_{\ell^+ \in \{0,1\}} \left\{ \left(\ell^+ \tfrac{\Delta}{2} + \left[h(t, q - \ell^+) - h(t, q) \right] \right) \right\} \mathbb{1}_{q > \underline{q}} \\ &\quad + \lambda^- \max_{\ell^- \in \{0,1\}} \left\{ \left(\ell^- \tfrac{\Delta}{2} + \left[h(t, q + \ell^-) - h(t, q) \right] \right) \right\} \mathbb{1}_{q < \overline{q}}\,, \end{aligned}$$

subject to the terminal condition

$$h(T, q) = -q\, \left(\tfrac{\Delta}{2} + \varphi\, q \right)\,.$$

When $\ell = 0$ both terms that are being maximised are zero, hence,

$$\boxed{\begin{aligned} \ell^{+,*}(t, q) &= \mathbb{1}_{\left\{ \frac{\Delta}{2} + [h(t,q-1) - h(t,q)] > 0 \right\} \cap \{q > \underline{q}\}}\,, \\ \ell^{-,*}(t, q) &= \mathbb{1}_{\left\{ \frac{\Delta}{2} + [h(t,q+1) - h(t,q)] > 0 \right\} \cap \{q < \overline{q}\}}\,. \end{aligned}} \tag{38}$$

The agent posts an LO on the appropriate side of the LOB by ensuring that she only posts if the arrival of an MO, which hit/lifts her LO, produces a change in her value function larger than $-\frac{\Delta}{2}$.

Market Making with No Terminal Penalty

Solving HJB with $\alpha = \phi = 0$

Assume no penalties for liquidating inventories at time T. Thus the ansatz is

$$H(t, x, q, S) = x + q\,S + g(t)\,. \tag{39}$$

Note that the function $g(t)$ does not depend on q. In the problem above the Ansatz contained $h(t, q)$ because the optimal strategy had to manage inventory risk which is something that is not a problem when $\alpha = 0$ here. Thus,

$$0 = g_t(t) + \lambda^+ \sup_{\delta^+} \left\{ e^{-\kappa^+ \delta^+}\, \delta^+ \right\} + \lambda^- \sup_{\delta^-} \left\{ e^{-\kappa^- \delta^-}\, \delta^- \right\}\,, \tag{40}$$

and the optimal postings are:

$$\delta^{*,+} = \frac{1}{\kappa^+} \tag{41}$$

and

$$\delta^{*,-} = \frac{1}{\kappa^-}\,. \tag{42}$$

Solving HJB with $\alpha = \phi = 0$

Alternatively note that

- For a risk-neutral MM, who does not penalise inventories, seeks to maximise the probability of being filled at every instant in time.
- The MM chooses $\delta^{\pm}$ to maximise the expected depth conditional on a market order hitting or lifting the appropriate side of the book: maximises $\delta^{\pm} e^{-\kappa^{\pm}\delta^{\pm}}$. The FOC

$$e^{-\kappa^{\pm}\delta^{\pm}} - \kappa^{\pm}\delta^{\pm} e^{-\kappa^{\pm}\delta^{\pm}} = 0\,. \tag{43}$$

Thus, we see that the optimal half spreads are as in (41) and (42).

Outline

Transient Price Impact Models

- Obizhaeva–Wang type models
- Non-robustness w.r.t. decay kernel
- Regularity of market models

Types of price impact

We have previously considered A-Ch setting with:

- permanent price impact and
- temporary price impact.

In reality, transactions interact with the LOB. Market orders will eat into the book but new liquidity will then come as markets are resilient.

We now consider optimal execution in a model with

- transient price impact.

Modelling transient price impact

Idea: model transient price impact by:

- stochastic dynamics of LOB
  $\rightsquigarrow$ e.g. constant depth λ, model only bid B_t & ask A_t
- a buy (market) order eats into the ask side of the book
  $\rightsquigarrow$ a buy order of $\Delta Q_t > 0$ moves ask $A_{t+} = A_t + \Delta Q_t/\lambda$
- book then reverts back at some speed
  $\rightsquigarrow$ according to a decay kernel $G(\text{delay})$, e.g. $\mathrm{e}^{-\rho t}$, $(1+t)^{-\alpha}$

Obizhaeva & Wang '13, Alfonsi et al. '08, Gatheral '10, Gatheral et al. '12...

Simple transient price impact (Obizhaeva & Wang '13)

- Assume no bid-ask spread, $S_t^0 = B_t = A_t$ is a martingale
- Constant book depth of $\lambda = 1/G(0)$
- A discrete order $Q_{t+} - Q_t =: \Delta Q_t$ moves price

$$S_{t+}^Q = S_t^Q + \Delta Q_t G(0)$$

and is executed at cost of (= - expected revenue of)

$$\frac{1}{G(0)} \int_{S_t^Q}^{S_{t+}^Q} v dv = \frac{1}{2G(0)} \left((S_{t+}^Q)^2 - (S_t^Q)^2 \right) = \frac{G(0)}{2} (\Delta Q_t)^2 + \Delta Q_t S_t^Q .$$

- The market is resilient and trade impact wanes away. So that

$$S_t^Q = S_t^0 + \sum_{s<t: |\Delta Q_s|>0} G(t-s) \Delta Q_s$$

Simple transient price impact – cont.

- Assume now trading is only possible at some given time points: $0 = t_0 < t_1 < \ldots < t_n = T$, Q_0 given, $Q_T = 0$ and

$$Q_t = Q_0 + \sum_{i:t_i<t} \Delta_i, \quad \text{where } \Delta_i := Q_{t_i+} - Q_{t_i}$$

- The mid-price resulting from strategy Q is

$$S_t^Q = S_t^0 + \sum_{i:t_i<t} G(t-t_i)\Delta_i$$

- The total cost of executing Q is

$$\mathcal{C}(Q) = S_0^0 Q_0 - \mathcal{R}(Q) = S_0^0 Q_0 + \sum_{i=0}^{n}\left(\frac{G(0)}{2}\Delta_i^2 + \Delta_i S_t^Q\right)$$

$$= S_0^0 Q_0 + \sum_{i=0}^{n} S_{t_i}^0 \Delta_i + \sum_{i=0}^{n}\left(\frac{G(0)}{2}\Delta_i^2 + \Delta_i \sum_{j<i} G(t_i - t_j)\Delta_j\right)$$

Simple transient price impact – cont.

$$S_0^0 Q_0 + \sum_{i=0}^{n} S_{t_i}^0 \Delta_i = S_0^0 Q_0 + \int_0^T S_t^0 dQ_t = -\int_0^T Q_{t-} dS_t^0$$

which has zero expectation (assuming Δ_i bounded). Further,

$$\begin{aligned}
&\sum_{i=0}^{n} \left(\frac{G(0)}{2} \Delta_i^2 + \Delta_i \sum_{j<i} G(t_i - t_j) \Delta_j \right) \\
&= \sum_i \frac{G(0)}{2} \Delta_i^2 + \sum_i \sum_{j<i} G(t_i - t_j) \Delta_i \Delta_j \\
&= \frac{1}{2} \sum_i \sum_j G(|t_i - t_j|) \Delta_i \Delta_j
\end{aligned}$$

In consequence, the total expected cost of liquidation following Q is

$$\mathbb{E}[\mathcal{C}(Q)] = \frac{1}{2} \sum_i \sum_j G(|t_i - t_j|)\, \mathbb{E}\left[\Delta_i \Delta_j\right]$$

Simple transient price impact – solution

It is then enough to look for Q among deterministic strategies:

$$\text{minimise} \sum_i \sum_j G(|t_i - t_j|)\Delta_i \Delta_j \quad \text{over } \Delta \in \mathbb{R}^{n+1} : \Delta^{\mathrm{T}} 1 = -Q_0$$

Rk: value invariant under $\Delta \to -\Delta \implies$ Optimal Buy = - Optimal Sell.

If G is strictly positive definite then the optimal solution Δ^* is

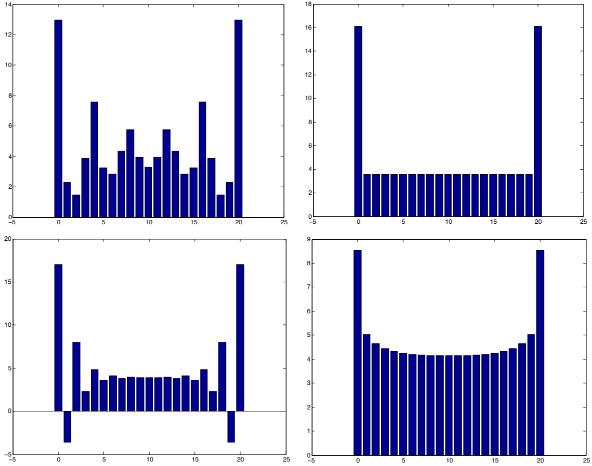
$$\Delta^* = \text{const} \cdot \Gamma^{-1} 1, \quad \text{where } \Gamma_{ij} = G(|t_i - t_j|).$$

Let us take equidistant steps: $t_{i+1} - t_i = \frac{T}{N}$ and look at different examples of G.

Optimal strategy – examples

Optimal Δ_i^* for $t \in [0,1]$, $N = 20$, $Q_0 = -100$ and four decay kernels:

$$G_1(t) = \mathrm{e}^{-5t}, \; G_2(t) = (0.5 - 2.7t)^+, \; G_3(t) = \frac{1}{(1+10t)^2}, \; G_4(t) = \frac{1}{1+(10t)^2}.$$



Which one is which?

$$A = \left| \begin{array}{c|c} 1 & 2 \\ \hline 3 & 4 \end{array} \right|$$

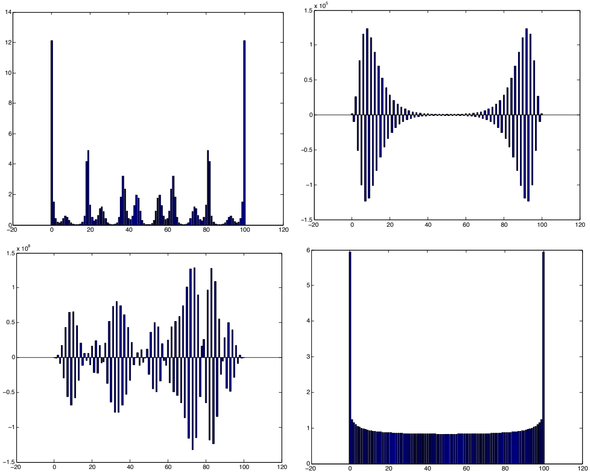
$$B = \left| \begin{array}{c|c} 2 & 1 \\ \hline 4 & 3 \end{array} \right|$$

$$C = \left| \begin{array}{c|c} 2 & 3 \\ \hline 4 & 1 \end{array} \right|$$

Optimal strategy – examples

Optimal Δ_i^* for $t \in [0,1]$, $N = 100$, $Q_0 = -100$ and four decay kernels:

$$G_1(t) = (0.5{-}2.7t)^+,\ G_2(t) = \frac{1}{(1+5t)^2},\ G_3(t) = \frac{1}{1+(10t)^2}\ G_4(t) = \frac{1}{1+(7t)^2}.$$



Which one is which?

$$A = \left|\begin{array}{c|c} 1 & 3 \\ \hline 4 & 2 \end{array}\right|$$

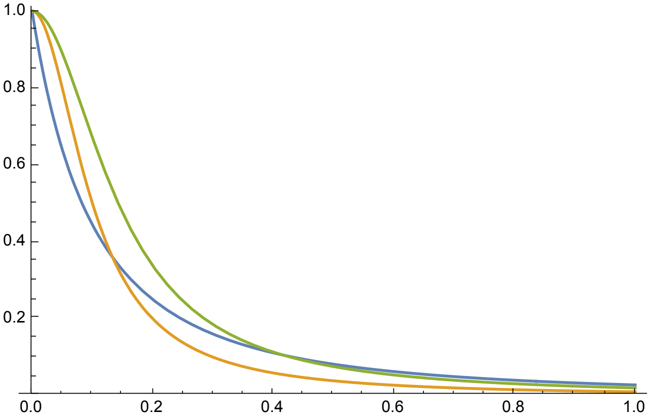
$$B = \left|\begin{array}{c|c} 1 & 4 \\ \hline 3 & 2 \end{array}\right|$$

$$C = \left|\begin{array}{c|c} 2 & 3 \\ \hline 4 & 1 \end{array}\right|$$

Non-robustness w.r.t. decay kernel

The optimal Δ_i^* for $t \in [0,1]$, $N = 100$, $Q_0 = -100$ and three decay kernels:

$$G_2(t) = \frac{1}{(1+5t)^2}, \;\; G_3(t) = \frac{1}{1+(10t)^2} \;\; G_4(t) = \frac{1}{1+(7t)^2}.$$



differ dramatically...

Notion of "price manipulation strategy"

We saw that very similar decay functions may lead to drastically different optimal portfolios, including round-trip-taking trading. Clearly requires further studies.

Definition

A round trip strategy Q, $Q_0 = Q_T = 0$ with strictly negative expected cost $\mathbb{E}[\mathcal{C}(Q)] < 0$ is called a price manipulation strategy.

Note that this is not the usual arbitrage since profit is not a.s. but in expectation. However in some models rescaling and repeating price manipulation leads to (weak) arbitrage.

We first extend our previous analysis to arbitrary strategies Q.

Transient price impact with arbitrary strategies

With discrete Q, the impacted price process was

$$S_t^Q = S_t^0 + \sum_{i:t_i<t} G(t-t_i)\Delta_i = S_t^0 + \int_{s<t} G(t-s)dQ_s$$

and the last term extends to arbitrary Q (predictable, left-continuous, of bounded variation). The revenues of a continuous strategy are given as previously

$$-\int_0^T S_t^Q dQ_t = -\int_0^T S_t^0 dQ_t - \int_0^T \int_{s<t} G(t-s)dQ_s dQ_t.$$

In the case of discrete Q we had

$$\begin{aligned} &-\sum_{i=0}^{n} S_{t_i}^0 \Delta_i - \frac{1}{2}\sum_i \sum_j G(|t_i-t_j|)\Delta_i\Delta_j \\ &= -\int_0^T S_t^0 dQ_t - \frac{1}{2}\int\int G(|t-s|)dQ_s dQ_t \end{aligned}$$

Transient price impact with arbitrary strategies

Combining, the execution cost of Q are

$$\mathcal{C}(Q) = S_0^0 Q_0 - \mathcal{R}(Q) = \int_0^T Q_{t-} dS_t^0 + \frac{1}{2}\int_0^T \int_0^T G(|t-s|) dQ_s dQ_t.$$

composed of volatility risk and price impact cost

$$\mathcal{C}^{exec}(Q) = \frac{1}{2}\int_0^T \int_0^T G(|t-s|) dQ_s dQ_t.$$

Price manipulation $\iff \mathbb{E}[\mathcal{C}^{exec}(Q)] < 0.$

Let's start with understanding when $\mathcal{C}^{exec}(Q) \geq 0$ a.s.

Bochner's theorem and positive costs

Proposition

We have $\mathcal{C}^{exec}(Q) \geq 0$ for all strategies Q iff G is positive definite, i.e. can be represented as the Fourier transform of a positive finite Borel measure μ on $\mathbb{R}$. Further, if G is strictly positive definite (μ is not discrete) then $\mathcal{C}^{exec}(Q) > 0$ for all nonzero Q.

We may also formalise the case of deterministic discrete strategies.

Proposition (Gatheral, Schied and Slynko '12)

Suppose G is positive definite. Then among deterministic strategies trading at given times (t_i), an optimal one Q^* satisfies a generalised Freedholm integral equation

$$\int G(|t_i - s|)dQ^*_s = \lambda, \quad i = 0, 1, \ldots, N$$

for some constant λ.

Rk.: We wrote this equation as $\Gamma\Delta = \text{const} \cdot 1$ before.

Is absence of price manipulation enough?

We have

$$\text{positive definite } G \implies \text{no price manipulation strategy.}$$

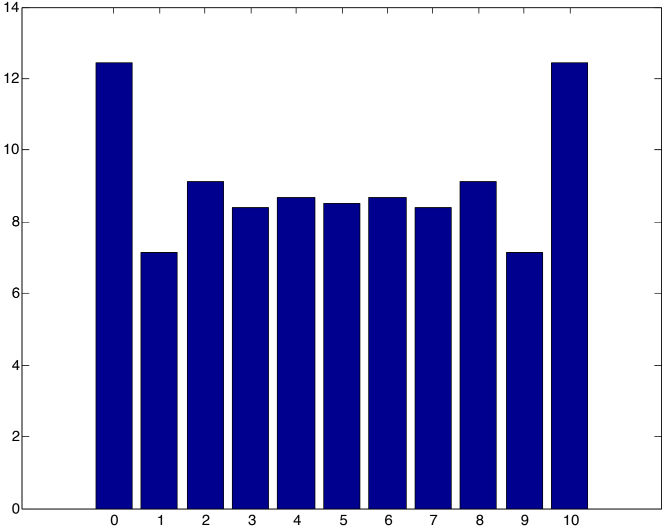
Is this enough? Take

$$G(t) = \mathrm{e}^{-t^2}$$

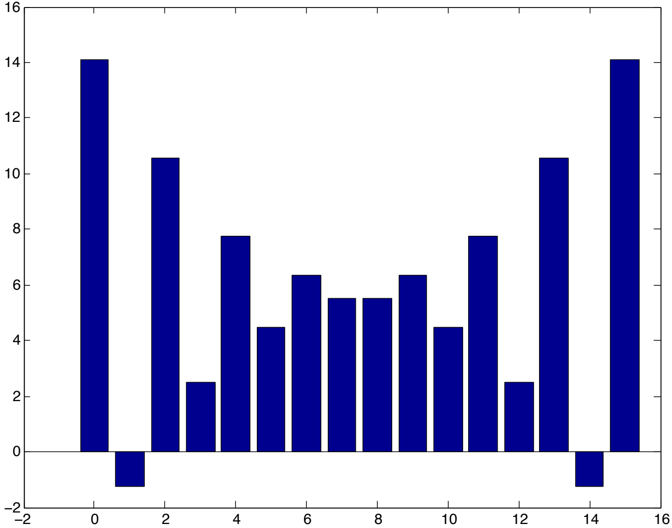
which, up to scaling, is its own Fourier transform and hence positive definite.

Let's look at the optimal strategy for $T = 10$, $Q_0 = -100$ and vary N.

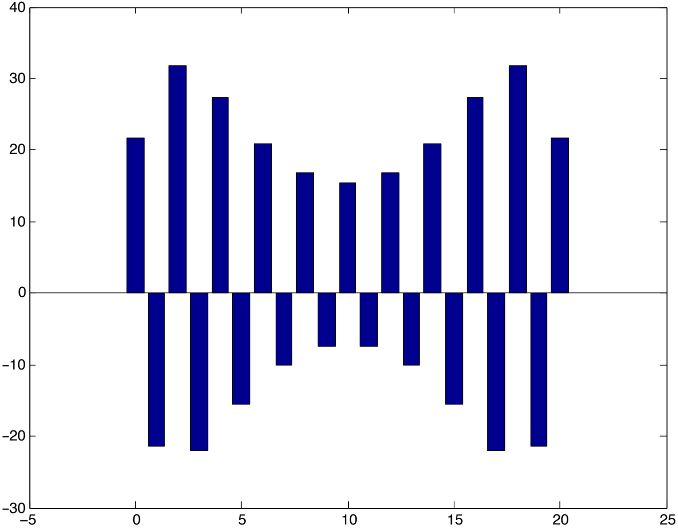
Optimal trading with Gaussian decay $G(t) = \exp(-t^2)$, $T = 10$, $Q_0 = -100$, $N = 10$



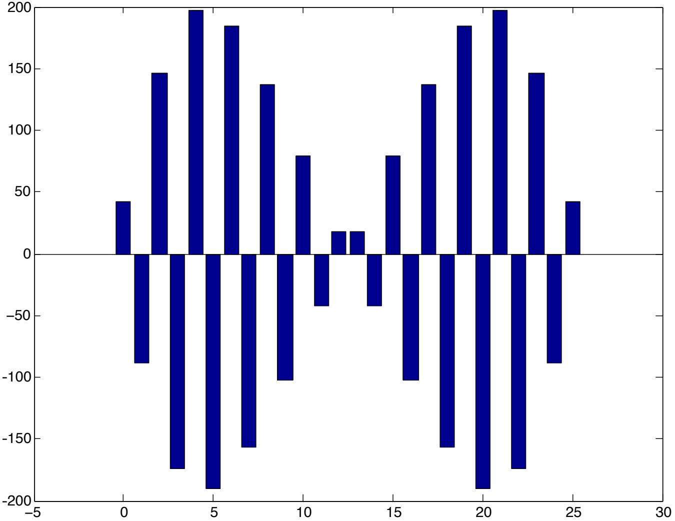
Optimal trading with Gaussian decay $G(t) = \exp(-t^2)$, $T = 10$, $Q_0 = -100$, $N = 15$



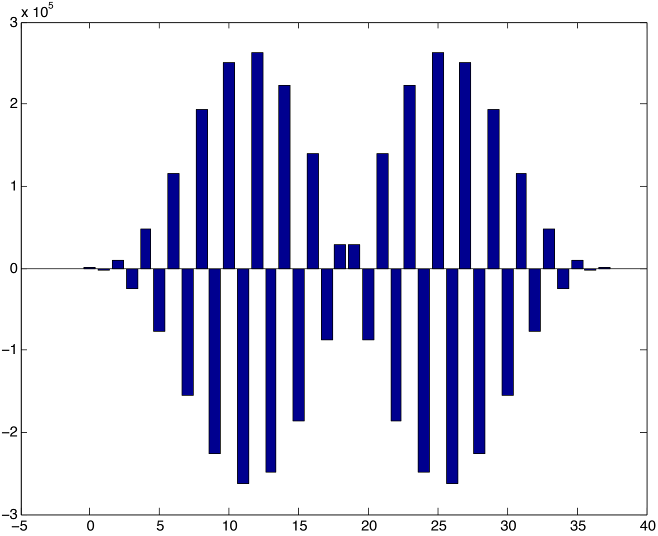
Optimal trading with Gaussian decay $G(t) = \exp(-t^2)$, $T = 10$, $Q_0 = -100$, $N = 20$



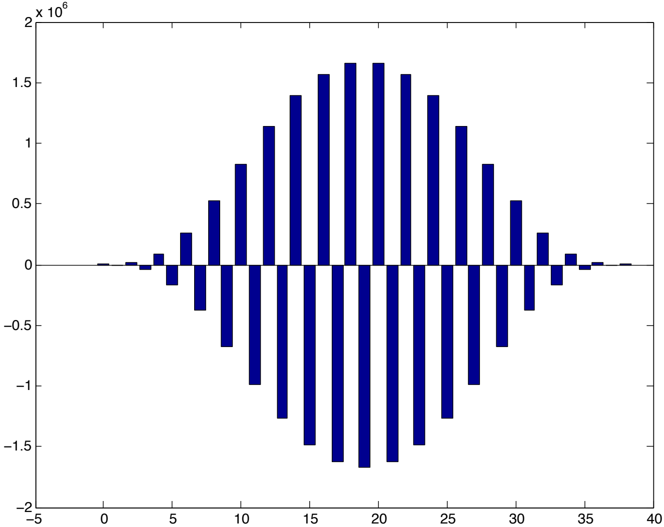
Optimal trading with Gaussian decay $G(t) = \exp(-t^2)$, $T = 10$, $Q_0 = -100$, $N = 25$



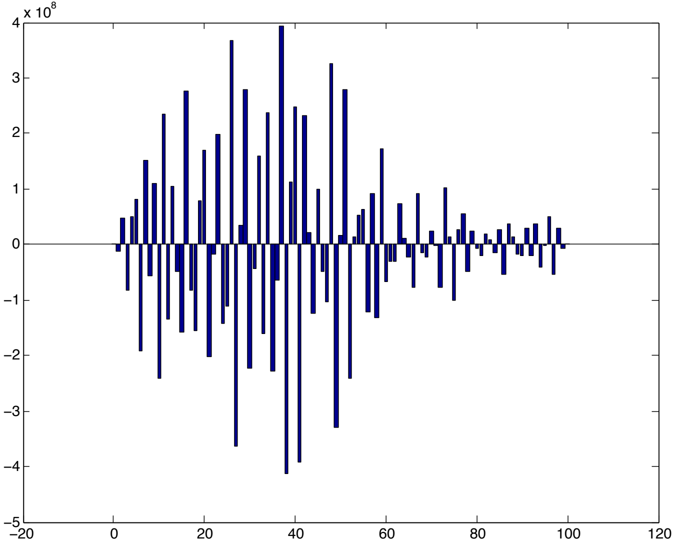
Optimal trading with Gaussian decay $G(t) = \exp(-t^2)$, $T = 10$, $Q_0 = -100$, $N = 37$



Optimal trading with Gaussian decay $G(t) = \exp(-t^2)$, $T = 10$, $Q_0 = -100$, $N = 38$



Optimal trading with Gaussian decay $G(t) = \exp(-t^2)$, $T = 10$, $Q_0 = -100$, $N = 100$



Price manipulation strategies

Definition

A market model admits price manipulation if there exists a round trip strategy Q, $Q_0 = Q_T = 0$ with strictly positive expected revenues $\mathbb{E}[\mathcal{R}(Q)] > 0$.

Definition

We say that a market impact model admits transaction–triggered price manipulation if the expected revenues of a sell (resp. buy) program can be increased by intermediate buy (resp. sell) orders.

Remark: in a sensible model (i.e. if buying increases prices and selling decreases prices) absence of *transaction–triggered price manipulation* implies absence of the usual *price manipulation*.

Regularity of Almgren–Chriss type models

Recall that in A-CH framework, the impacted price is

$$S_t^Q = S_t^0 + \int_0^t g(\dot{Q}_s) ds + h(\dot{Q}_t).$$

Proposition (Huberman & Stanzl '04, Gatheral '10)

If the model above does NOT admit price manipulation for all $T > 0$ then $g(x) = \gamma x$ for some $\gamma \geq 0$.
Further, if g is linear and $x \to xh(x)$ is convex than the model does NOT admit transaction–triggered price manipulation.

Rk: the second part is clear since in this setting the optimal Q^* is linear.

Regularity of Obizhaeva–Wang type models

Proposition (Alfonsi, Schied & Slynko '12)

A transient price impact model with decay kernel G s.t.

$$G(0) - G(s) < G(t) - G(t+s), \quad \textit{for some } s \neq t,$$

admits transaction-triggered price manipulation trading at $\{0, s, t+s\}$*.*

In particular, it is enough that G is NOT convex for small t.

Proposition (Alfonsi et al. '12, Gatheral et al. '12)

A transient price impact model with convex, decreasing, non-negative decay kernel G admits a unique optimal Q^* which is monotone in time. In particular the setup does NOT admit transaction-triggered price manipulation.

Other developments

- Non-linear transient price impact models: the book has varying depth according to a given shape f, see Alfonsi & Schied '10
- A combination of impacts, e.g. Gatheral '10

$$S_t^Q = S_t^0 + \int_0^t h(-\dot{Q}_t) G(t-s) ds$$

- Stochastic models of LOB where the shape f is a stochastic process in space of curves and/or stochastic resilience, see Alfonsi & Infante Acevedo '12, Klöck '12, Fruth, Schöneborn & Urusov '11, Müller & Keller-Ressel '15.
- ...

Summary of transient market impact models

- Transient price impact models take into account the interaction of orders with the LOB and market resilience
- Under constant LOB depth, discrete trading at (t_i) and maximising expected revenues the optimal strategy explicit for many impact decay kernels G
- More generally the problem quickly becomes very hard...
- Even in simple setting, the optimal strategies may often involve round trips. Solution is non-robust with respect to G.
- Possible to study, and provide sufficient conditions for, the absence of price-triggered manipulation strategies.

Outline

Predatory trading and HF hot-potatos

Multi-agent frameworks

- In reality many agents interact in a market.
- Mathematically best modelled as game. When number of players $n \to \infty$, sometimes possible to analyse as a mean field game.
- Interesting as it allows to study
  - Interaction of one large player with n small players (e.g. predatory trading)
  - Global market implications of interactions between small players
  - Properties of markets which facilitate different phenomena

Predatory Trading

Large Trader facing a forced liquidation

$+$

other (HF) traders aware of this fact

$\Downarrow$

Predatory Trading

Examples of "targets":

- Index-replicating funds at rebalancing dates
- Institutional investors subject to regulatory constraints (e.g. when an instrument is downgraded)
- Traders using portfolio insurance or stop-loss strategies
- Hedge funds close to a margin call
- Recalled short-seller

Predatory Trading

"... if lenders know that a hedge fund needs to sell something quickly, they will sell the same asset – driving the price down even faster. Goldman, Sachs & Co. and other counterparties to LTCM did exactly that in 1998."

Business Week, 26 Feb 2001

"When you smell blood in the water, you become a shark ... when you know that one of your number is in trouble ... you try to figure out what he owns and you start shorting those stocks ... "

Cramer, 2002

Predatory Trading – mechanisms

When a need of a large trader (prey) to liquidate is recognised, the strategic traders (predators) might

- first trader **in the same direction**
  - withdraw liquidity instead of providing it
  - market impact is greater leading to price overshooting
  - may further enforce distressed trader's need to liquidate
- then **reverse direction** to profit from the overshoot
- closing the roundtrip at a profit.

However when strategic traders have a longer horizon than the liquidation, their behaviour may depend on market characteristics:

- could act as predators as above $\rightsquigarrow$ large trader tries to keep intentions hidden (stealth trading)
- could act as liquidity providers $\rightsquigarrow$ large trader announces intentions (sunshine trading)

see Brunnermeier & Pedersen '05, Carlin, Lobo & Viswanathan '05, Schied & Schöneborn '08.

One-period game model with A–Ch price impact

- $n+1$ players with portfolios $Q_0(t), \ldots, Q_n(t)$, $t \in [0, T]$, assumed cont. diff. in time
  - one prey (seller): $Q_0(0) = q_0 > 0$, $Q_0(T) = 0$
  - n predators: $Q_i(0) = Q_i(T) = 0$, $i = 1, \ldots, n$
- and the above is common knowledge
- players are risk-neutral and maximise their expected profit

$$\mathcal{R}^i(Q) = -\mathbb{E}\left[\int_0^T S_t dQ_i(t)\right]$$

- one risk-free and one risky asset, continuous trading, Almgren–Chriss linear price impact model

$$S(t) = S(0) + \sigma W_t + \gamma \sum_{i=1}^{n} (Q_i(t) - Q_i(0)) + \eta \sum_{i=1}^{n} \dot{Q}_i(t)$$

- Solved by searching for Nash equilibrium.

One-period game model with A–Ch price impact

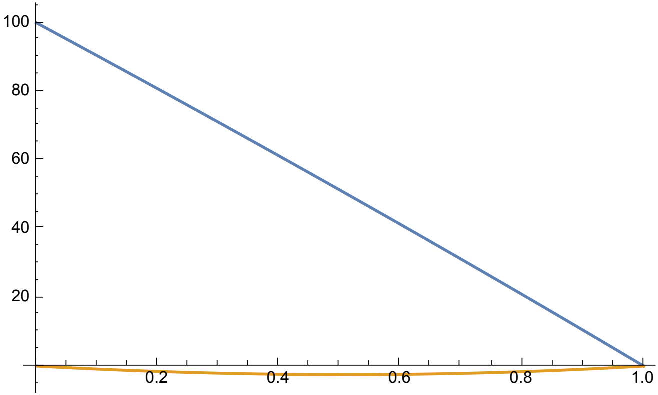
Assuming all Q_i are deterministic this can be solved explicitly giving

$$\dot{Q}_i^*(t) = \alpha \mathrm{e}^{-\frac{n}{n+2}\frac{\gamma}{\eta}t} + \beta_i \mathrm{e}^{\frac{\gamma}{\eta}t},$$

where

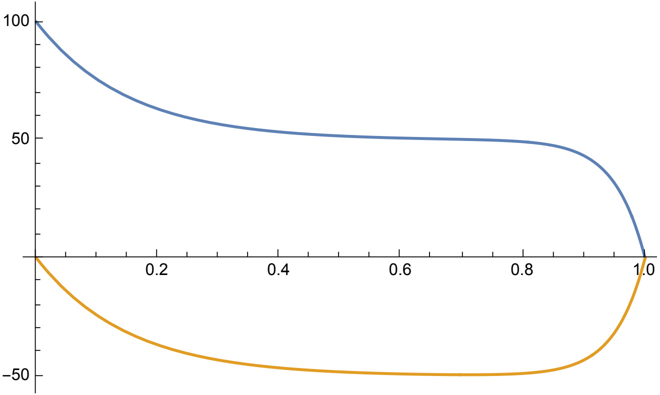
$$\begin{aligned}
\alpha &= \frac{-n}{n+2}\frac{\gamma}{\eta}\left(1 - \mathrm{e}^{-\frac{n}{n+2}\frac{\gamma}{\eta}T}\right)^{-1}\frac{x_0}{n+1}, \\
\beta_i &= \frac{\gamma}{\eta}\left(\mathrm{e}^{\frac{\gamma}{\eta}T} - 1\right)^{-1}\left(Q_i(T) - Q_i(0) + \frac{x_0}{n+1}\right)
\end{aligned}$$

Optimal strategies with $n = 1$, $T = 1$, $\frac{\gamma}{\eta} = 0.3$



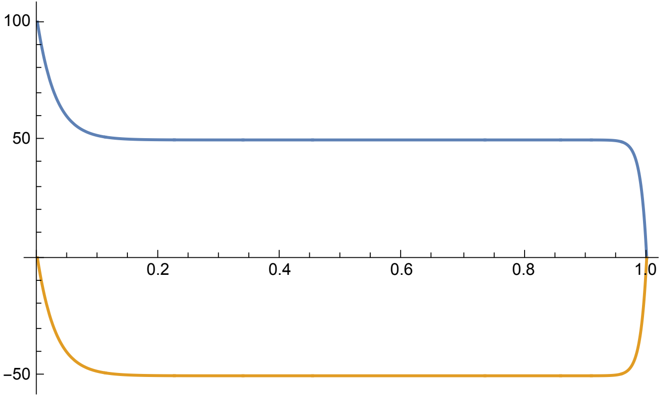
Distressed trader (blue) and one predator in a elastic market (i.e. temporary impact $>$ permanent impact)

Optimal strategies with $n = 1$, $T = 1$, $\frac{\gamma}{\eta} = 20$



Distressed trader (blue) and one predator in an plastic market (i.e. permanent impact $>$ temporary impact)

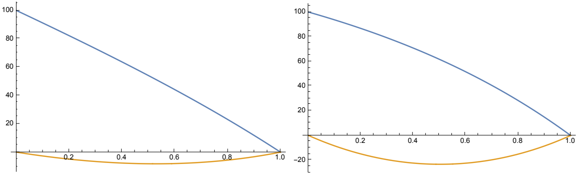
Optimal strategies with $n = 1$, $T = 1$, $\frac{\gamma}{\eta} = 100$



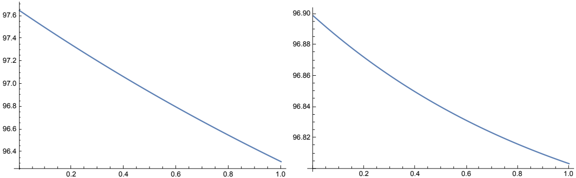
Distressed trader (blue) and one predator in a highly plastic market.

Effect of predators, $T = 1$, $x_0 = 100$, $S_0 = 100$, $\gamma = \eta = 2\%$

Comparison of $n = 1$ and $n = 40$ predators. Aggregated Holdings:



Expected market price:

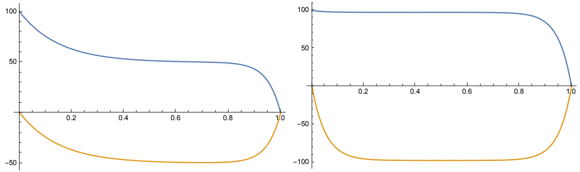


Expected execution cost $\mathbb{E}[\mathcal{C}(Q)]$: 3.1% and 3.2% (*compare with* 3% *when* $n = 0$)

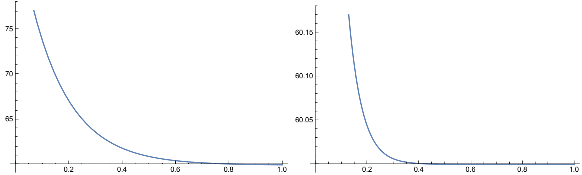
Expected revenue per predator: 7.27 and 0.4.

Effect of predators, $T = 1$, $x_0 = 100$, $S_0 = 100$, $\gamma = 20 * \eta = 2\%$

Comparison of $n = 1$ and $n = 40$ predators. Aggregated Holdings:



Expected market price:



Expected execution cost $\mathbb{E}[\mathcal{C}(Q)]$: 33.3% and 40% (*was* 22% *when* $n = 0$)
Expected revenue per predator: 665 and 2.3.

Price and execution costs scale linearly with costs when keeping $\frac{\gamma}{\eta}$ fixed.

HF hot-potato game

Schied & Zhang '13 considered the following setup:

- two HF players Q and Y trading in an Obizhaeva & Wang market with $G(t) = \mathrm{e}^{-\rho t}$
- trading at an equidistant discrete time grid
- with opposite initial positions $Q_0 = -Y_0$.

Using a Nash equilibrium analysis, they show that

- the optimal behaviour, if trading is frequent enough, involves a highly oscillatory trading
- hot-potato effect with volume passed between traders
- the effect can be eliminated if transaction costs present and high enough compared to LOB depth

Multi-agent setup summary

- Detailed analysis of market behaviour may require models with interacting agents
- Mathematically, often done using game theory and searching for Nash equilibria
- Predatory trading can be described as a game between one large seller (prey) and n strategic traders (predators)
- Both from the theory and practice, we see that predators often first trade in the same direction as the large trader leading to price overshoot of which they then take advantage.
- The optimal behaviour highly dependent on the market characteristic (e.g. which type of price impact dominates)
- More involved situations (e.g. strategic traders having longer trading horizon) may lead to qualitatively different solutions
- Many other situations in which game analysis is interesting, e.g. high trade volume (hot-potato) effect of trading between two agents.