Market Microstructure and Algorithmic Trading

University of Oxford

Fayçal Drissi
faycal.drissi@eng.ox.ac.uk
http://www.faycaldrissi.com

lecture notes

March 2, 2026

Optimal order routing of aggressive orders

The decision problem

- A “**marketable**” order is a buy (resp. sell) order at a price higher than the best ask (resp. lower than the best bid).
- Often, operators have to be split a large order over N available venues.

Optimal order routing of aggressive orders

The decision problem

- A “**marketable**” order is a buy (resp. sell) order at a price higher than the best ask (resp. lower than the best bid).
- Often, operators have to be split a large order over N available venues.

The model of the environment

- Let ($Q^\star$ at max price $P^\star$) be a marketable **buy** order.
  Let $Q_n(p)$ be the **visible quantity** that is **available** at **price** p in **trading venue** n.

Optimal order routing of aggressive orders

The decision problem

- A “**marketable**” order is a buy (resp. sell) order at a price higher than the best ask (resp. lower than the best bid).
- Often, operators have to be split a large order over N available venues.

The model of the environment

- Let ($Q^\star$ at max price $P^\star$) be a marketable **buy** order.
  Let $Q_n(p)$ be the **visible quantity** that is **available** at **price** p in **trading venue** n.

The optimisation problem : Choose $(p_1, \cdots, p_n)$ to minimize $\sum_{n=1}^{N} p_n \cdot Q_n(p_n)$ such that $Q^\star = \sum_{n=1}^{N} Q_n(p_n)$ with the constraint that $P^\star \geq p_n$ for all $n \in \{1, \cdots, N\}$.

The solution:

- Lagrangian: $Q_n(p_n) + p_n\, Q_n^{'}(p_n) = \lambda\, Q_n^{'}(p_n)$ for $n \in \{1, \cdots, N\}$.
- Assume the linear form $Q_n(p) = q_n + c_n \cdot p$.

$$\Longrightarrow \ (q_n + c_n \cdot p_n^\star) + p_n^\star\, c_n = \lambda\, c_n$$
$$\Longrightarrow \ p_n^\star = \frac{\lambda}{2} - \frac{q_n}{2\, c_n}\,.$$

- Inject λ in the constraint $Q^\star = \sum_{n=1}^{N} Q_n(p_n)$:

$$\Longrightarrow \ Q^\star = \sum_{n=1}^{N} \{q_n + c_n \cdot p_n^\star\} = \sum_{n=1}^{N} q_n/2 + c\,\lambda/2, \quad c = \sum_n c_n$$

- Finally

$$\boxed{p_n^\star = \frac{Q^\star}{c} - \frac{q_n}{2\, c_n}\left(1 + \frac{c_n}{q_n} \cdot \frac{\overline{q}}{\overline{c}}\right)} \quad \overline{c} = \frac{1}{N}\sum_n c, \ \ \overline{q} = \frac{1}{N}\sum_n q$$

Some issues with our (simple) model:

- There is **latency** in high frequency markets.
- There are **cancellations** in limit order books.
- The shape of the book **hides** the truth sometimes: iceberg orders + hidden orders.
- etc..

The decision problem

- An operator splits a large LO over N venues: there is structurally **uncertainty** on limit orders splitting; waiting on a bad queue generates **opportunity costs**. Formulate an optimal trading problem ?

The decision problem

- An operator splits a large LO over N venues: there is structurally **uncertainty** on limit orders splitting; waiting on a bad queue generates **opportunity costs**. Formulate an optimal trading problem ?

The model:

- The best queue in each venue has size Q_n.
- Each queue is **consumed** according to a Poisson P_t^n with intensity λ_n.

The objective: find LO quantities $(q_1, \cdots, q_N)$ to minimize, **on average**, the time $t^\star$ to execute the quantity $Q^\star = \sum_n q_n$.

Solution:

- After our LOs, venue n has new queue quantity $Q_n + q_n$.

- Queue is consumed in t_n:

$$\int_0^{t^n} dP_t^n = q_n + Q_n \implies \mathbb{E}\left[P_{t^n}^n\right] = t_n \lambda_n = q_n + Q_n.$$

- We minimize the maximum of all t^n, so $t^\star = t^n$ for all $n \in \{1, \cdots, N\}$. So:

$$t^\star = t_n = Q^\star / \sum_n \lambda_n + \sum_n Q_n / \sum_n \lambda_n \implies \boxed{q_n^\star = \rho_n \frac{Q^\star}{N} + \left(\rho_n \overline{Q} - Q\right)}$$

$$\text{where } \rho_n = \lambda_n / \overline{\lambda}, \quad \overline{\lambda} = \frac{1}{N} \sum_n \lambda_n, \quad \overline{Q} = \frac{1}{N} \sum_n Q_n\,.$$

Optimal execution: Large operators need optimal execution models. The first approach is in (Bertsimas and Lo 1998).

Two seminal papers: (Almgren and Chriss 1999) and (Almgren and Chriss 2001). Considered to be the pioneers in optimal execution.

- A model taking into account both the expected cost of execution and the risk that the price moves.
- Large orders are split into smaller ones, that are executed progressively over a given time window.
- A trader **executing fast** pays high execution costs: bid-ask spread and limited available liquidity at each price in the order book.
- **Slow execution** exposes to possible adverse price fluctuations.
- There is a **trade-off** between execution costs and price risk.

- An agent with a single stock.
  At time $t = 0$, the position (in number of shares) is q_0.

- The **objective**: **unwind the position by time** T. Find the optimal **schedule**.

- The **model**: We assume a regular temporal grid

$$t_0 = 0 < \cdots < t_n = n\,\Delta t < \cdots < t_N = N\,\Delta t = T.$$

  At the start of each $[t_n, t_{n+1}]$, the agent sends an MO of size $\nu_{n+1}\,\Delta t$.

  If $\nu_{n+1} \leq 0$ then the agent sells shares, if $\nu_{n+1} \geq 0$, then the agent buys shares.

- The dynamics for the **inventory** (number of shares in the agent's portfolio)

$$\boxed{q_{n+1} = q_n + \nu_{n+1}\,\Delta t}$$

- The **mid-price** follows a Brownian motion

$$\boxed{S_{n+1} = S_n + \underbrace{\sigma\sqrt{\Delta t}\,\epsilon_{n+1}}_{\text{market risk}} + \underbrace{k\,\nu_{n+1}\,\Delta t}_{\text{linear perm. impact}}}$$

  - ϵ_n are i.i.d $\mathcal{N}(0,1)$ variables
  - $\sigma > 0$ is the arithmetic volatility
  - $k > 0$ scales the magnitude of the **linear permanent impact**. (more on this later ...)

Permanent price impact: large MOs leave a long-term effect on the midprice.

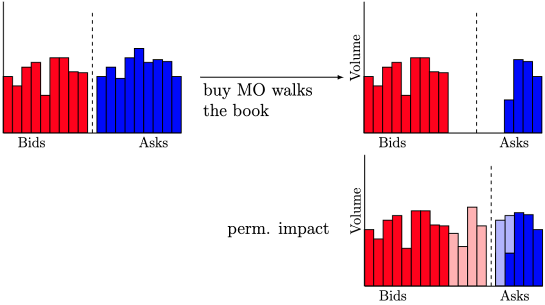
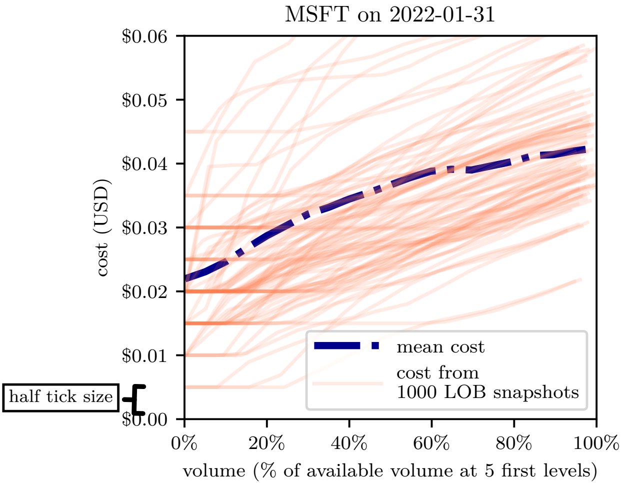


Figure 7: In the first two panels, an MO walks the book so the next midprice exhibits the temporary price impact. Immediately after the MO, market participants replenish the LOB. The difference between the midprice in the last panel and that of the first panel is the permanent impact.

- **Execution costs** (temporary impact):
  - We introduce the deterministic **market volume** V_{n+1}, which is the volume traded by other agents throughout $[t_n, t_{n+1}]$.
  - The price $\tilde{S}_{n+1}$ obtained for each share in $[t_n, t_{n+1}]$ depends on the quantity $\nu_{n+1}\,\Delta t$ and on the market volume V_{n+1}. We assume the linear form

$$\boxed{\tilde{S}_{n+1} = S_n + \eta\,\nu_{n+1}/V_{n+1}}.$$

    $\eta > 0$ so the agent buys (sells) at prices higher (lower) than the mid-price S_n.
  - ν_n/V_n is the **participation rate** and it is very important to estimate η. e.g., the costs are 1 ticks (or bid-ask spread) spread per 5% of participation rate.



Execution costs defined as a function of participation rate for multiple snapshots of the LOB of MSFT quoted on Nasdaq. The total trade volume is approximated by the total available liquidity.

Market volume

The market volume V_{n+1} corresponds to the total volume of the market over a time slice $[t_n, t_{n+1}]$. In practice, it is difficult to consider this deterministic ...

However, market activity depends on the time of day. On average, it is deterministic and has a characteristic U-shape.

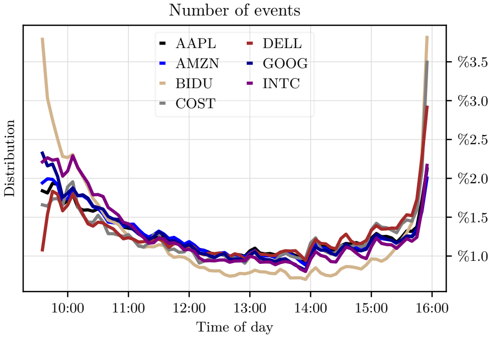


Figure: Distribution of the activity throughout the trading day, measured in portion of LOB events, averaged through trading days between October and December 2022. Source: (Cartea et al. 2023).

- The amount paid (received) for $\nu_{n+1}\,\Delta t$ shares bought (sold) between t_n and t_{n+1} is

$$\nu_{n+1}\,\tilde{S}_{n+1}\,\Delta t = \nu_{n+1}\,(S_n + \eta\,\nu_{n+1}/V_{n+1})\,\Delta t.$$

So the dynamics of the cash account X are

$$\boxed{X_{n+1} = X_n - \nu_{n+1}\,S_n\,\Delta t - \eta\,\frac{\nu_{n+1}^2}{V_{n+1}}\,\Delta t}$$

- The execution costs paid are relative to S_n over $[t_n, t_{n+1}]$, so no price risk within each slice ...

Find a **liquidation strategy** $v = (\nu_1, \ldots, \nu_n)$ maximizing the mean-variance objective function

$$\boxed{\mathbb{E}[X_N] - \frac{\gamma}{2}\mathbb{V}[X_N]}.$$

We focus on deterministic admissible strategies (more on this later ...)

$$(\nu_n)_n \in \mathcal{A}^d = \left\{ (\nu_1, \ldots, \nu_N) \in \mathbb{R}^n, \ \sum_{n=0}^{N-1} \nu_{n+1}\, \Delta t = -q_0 \right\}$$

To solve the problem, we compute the value of the terminal wealth X_N :

$$
\begin{aligned}
X_N &= X_0 - \sum_{n=0}^{N-1} (q_{n+1} - q_n)\; S_n - \eta \sum_{n=0}^{N-1} \frac{\nu_{n+1}^2}{V_{n+1}} \Delta t \\
&= X_0 - \sum_{n=0}^{N-1} q_{n+1} \left(S_{n+1} - \sigma\sqrt{\Delta t}\, \epsilon_{n+1} - k\, \nu_{n+1} \Delta t \right) + \sum_{n=0}^{N-1} q_n\, S_n - \eta \sum_{n=0}^{N-1} \frac{\nu_{n+1}^2}{V_{n+1}} \Delta t \\
&= X_0 + q_0\, S_0 + \sigma \sqrt{\Delta t} \sum_{n=0}^{N-1} q_{n+1}\, \epsilon_{n+1} + \underbrace{k \sum_{n=0}^{N-1} q_{n+1}\, \nu_{n+1}\, \Delta t}_{\star} - \eta \sum_{n=0}^{N-1} \frac{\nu_{n+1}^2}{V_{n+1}} \Delta t.
\end{aligned}
$$

Observe that:

$$
\begin{aligned}
\star &= k \sum_{n=0}^{N-1} \left(\frac{q_{n+1} + q_n}{2} + \frac{q_{n+1} - q_n}{2} \right) (q_{n+1} - q_n) \\
&= \frac{k}{2} \sum_{n=0}^{N-1} \left(q_{n+1}^2 - q_n^2 \right) + \frac{k}{2} \sum_{n=0}^{N-1} (q_{n+1} - q_n)^2 \\
&= -\frac{k}{2} q_0^2 + \frac{k}{2} \sum_{n=0}^{N-1} \nu_{n+1}^2\, \Delta t^2\,.
\end{aligned}
$$

The terminal wealth is

$$X_N = X_0 + q_0\,S_0 - \frac{k}{2}q_0^2 + \sigma\sqrt{\Delta t}\sum_{n=0}^{N-1} q_{n+1}\,\epsilon_{n+1} + \frac{k}{2}\sum_{n=0}^{N-1}\nu_{n+1}^2\,\Delta t^2 - \eta\sum_{n=0}^{N-1}\frac{\nu_{n+1}^2}{V_{n+1}}\Delta t$$

$$= X_0 + q_0\,S_0 - \frac{k}{2}q_0^2 + \sigma\sqrt{\Delta t}\sum_{n=0}^{N-1} q_{n+1}\,\epsilon_{n+1} - \sum_{n=0}^{N-1}\nu_{n+1}^2\,\left(\frac{\eta - \frac{k}{2}\,V_{n+1}\,\Delta t}{V_{n+1}}\right)\Delta t\,.$$

To obtain analytic formulae, **assume** a flat volume curve $V_n = V, \forall n$.

Next, either we neglect the term in Δt^2 (define $\tilde{\eta} = \eta$) or we assume $\eta >> \frac{k}{2}\,V\,\Delta t$ (define $\tilde{\eta} = \eta - \frac{k}{2}\,V\,\Delta t$)

The controls are deterministic, so X_N is normally distributed with mean

$$\boxed{\mathbb{E}\left[X_N\right] = X_0 + q_0\,S_0 - \frac{k}{2}q_0^2 - \tilde{\eta}\sum_{n=0}^{N-1}\frac{\nu_{n+1}^2}{V}\Delta t}$$

and variance

$$\boxed{\mathbb{V}\left[X_N\right] = \sigma^2\,\Delta t\sum_{n=0}^{N-1} q_{n+1}^2}$$

The problem **reduces** to minimising the following functional over $\mathcal{A}^d$

$$\tilde{\eta} \sum_{n=0}^{N-1} \frac{\nu_{n+1}^2}{V} \Delta t + \frac{\gamma}{2} \sigma^2 \Delta t \sum_{n=0}^{N-1} q_{n+1}^2 \,,$$

which is equivalent to minimising J over $\mathcal{C}_d = \{q = (q_0, \dots, q_N)\,, q_0 = q_0, q_N = 0\}$

$$\boxed{J : q \in \mathbb{R}^{N+1} \mapsto \tilde{\eta} \sum_{n=0}^{N-1} \frac{(q_{n+1} - q_n)^2}{V\,\Delta t} + \frac{\gamma}{2} \sigma^2 \Delta t \sum_{n=0}^{N-1} q_{n+1}^2}$$

The Legendre-Fenchel transform of $g : x \mapsto \tilde{\eta} \frac{x^2}{V\,\Delta t}$ is easily found (FOC) to be

$$g^\star : p \mapsto \sup_x p\,x - \tilde{\eta} \frac{x^2}{V\,\Delta t} = \frac{V\,\Delta t}{4\,\tilde{\eta}} p^2.$$

The **optimal trading curve** $q^\star$ is characterized by the Hamiltonian system

$$\begin{cases} p_{n+1} &= p_n + \gamma\,\sigma^2\,\Delta t\, q_{n+1}^\star\,, \quad 0 \le n < N-1 \\ q_{n+1}^\star &= q_n^\star + \frac{V}{2\,\tilde{\eta}}\,\Delta t\, p_n\,, \qquad 0 \le n < N \end{cases}$$

The **optimal inventory** $q^\star$ to hold is the solution of the second-order recursive equation

$$q^\star_{n+2} - \left(2 + \frac{\gamma\,\sigma^2\,V}{2\,\tilde{\eta}}\Delta t^2\right) q^\star_{n+1} + q^\star_n = 0\,,$$

with boundary conditions

$$q^\star_0 = q_0 \quad \text{and} \quad q^\star_N = 0.$$

Solving the equation gives

$$\boxed{q^\star_n = q_0\,\frac{\sinh\left(\alpha\,(T - t_n)\right)}{\sinh(\alpha T)}}$$

where α solves

$$2\,\cosh\left(\alpha\,\Delta t\right) = \frac{\gamma\,\sigma^2\,V}{2\,\tilde{\eta}}\Delta t^2\,.$$

- Consider an asset with volatility $1\$ \cdot \text{day}^{-1/2}$ (approx. 32% annualized vol) with $S_0 = 100$.
- Assume the market trades $V = 4,000,000$ shares per day, and assume $\eta = 0.1\$ \cdot \text{share}^{-1}$.
- The initial inventory to liquidate is $q_0 = 200,000$ corresponding to 5% participation rate.

Effect of model parameters: η, σ, γ, and V ?

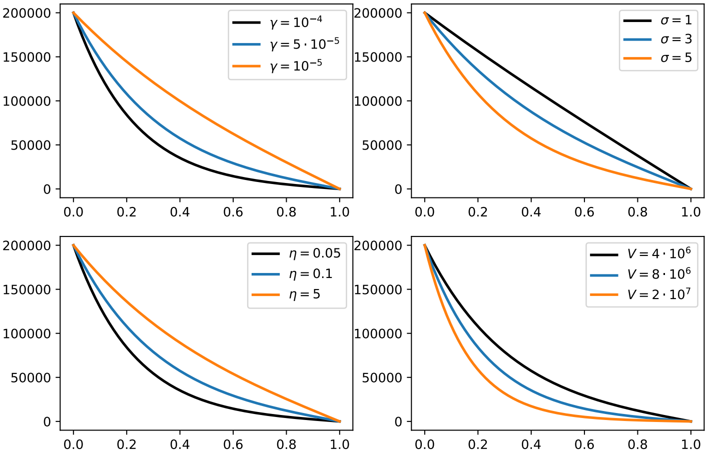


Figure: Optimal trading curves.

Execution costs

Setting temporary market impact parameter η in practice ((Almgren and Chriss 2001)):

Traders suppose that the additional cost incurred per share when trading a given volume is **proportional** to the **participation rate**.

For each percent of participation rate, a cost corresponding to some % of the **the bid-ask spread** is incurred.

Objective functions

- There is a-priori no reason to believe agents have a mean-variance risk profile.
- In practice, the choice of **utility function** is difficult. In the simple case of mean-variance, choosing the **risk aversion parameter** γ is complex.
- If the agent is a Brokerage firm or a cash trader executing for clients, which value of γ to use ?
  - The client must, somehow, indicate the value of γ.
  - One interpretation: the risk aversion parameter encodes the **urgency** of trading.
  - Or the broker needs to tailor γ to the size of the trade / participation rate ($\gamma = \overline{\gamma}/|q_0\, S_0|$)

Extensions

The framework is flexible:

- The criterion can be challenged: PoV, VWAP, TWAP, TC, etc.
- The variance term has a strong influence.
  More interesting risk measures: CVaR, etc ..
- In practice: traders use **participation constraints** to their trading flow.

Some problems / extensions:

- We don't deal with orderbook dynamics $\implies \eta$ and k are stochastic.
- The market impact model is far from being realistic. The market price is partially resilient (it does not vanish instantaneously).

Almgren, Robert and Neil Chriss (1999). “Value under liquidation”. In: **Risk** 12.12, pp. 61–63.

— (2001). “Optimal execution of portfolio transactions”. In: **Journal of Risk** 3, pp. 5–40.

Bertsimas, Dimitris and Andrew W Lo (1998). “Optimal control of execution costs”. In: **Journal of financial markets** 1.1, pp. 1–50.

Cartea, Álvaro, Fayçal Drissi, and Pierre Osselin (2023). “Bandits for algorithmic trading with signals”. In: **Available at SSRN 4484004**.