Latent liquidity in limit order driven markets

Julius Bonart

Imperial College London, Department of Mathematics, CFM-Imperial Institute of Quantitative Finance

10 septembre 2015
Limit order books

More than half of world’s exchanges now operate LOBs as the market clearing mechanism:

- Other market clearing mechanisms are imaginable (for instance: Centralized market maker).
Limit order books

More than half of world's exchanges now operate LOBs as the market clearing mechanism:

- Other market clearing mechanisms are imaginable (for instance: Centralized market maker).
- Are markets neutral as to the way market clearing is implemented?
Limit order books

More than half of world’s exchanges now operate LOBs as the market clearing mechanism:

- Other market clearing mechanisms are imaginable (for instance: Centralized market maker).
- Are markets neutral as to the way market clearing is implemented?
- I will argue here that they are not. LOBs create specific incentives for market participants, in particular with respect to the liquidity provision.
What is liquidity?

Recall: Liquidity in LOBs
Liquidity taking = submission of market orders
Liquidity provision = submission of limit orders

Three criteria for a liquid market:

Tightness – Small bid-ask spread, i.e. market orders incur small instantaneous costs (instant liquidity taking)

Depth – Available volume at the best quotes are large.

Latency – A sequence of market orders can be executed within a short time horizon, i.e. executed volume is quickly refilled by liquidity providers.
What is liquidity II

Are “liquid” securities liquid?

Daily traded volume of Apple Inc. ≈ 6.000.000.000$.

Market capitalization of Apple Inc. ≈ 600.000.000.000$ ≈ 100 × V_D. Imagine a large investment wishes to acquire 1% of Apple; it takes certainly several days to execute this large order!

Therefore “liquid” stocks are in fact not that liquid at all. High-volume traders need to split a large order into small child orders and execute them incrementally.

This is a very widespread behaviour (optimal execution problem), so why do markets not provide sufficient liquidity for these traders?
Empirical evidence for liquidity rationing: Average and typical shapes of the LOB

Conclusion: Liquidity in LOBs bends down and LOBs are sparse (for small-tick stocks). Aggregate demand and supply curves should increase!  

Empirical evidence for liquidity rationing II: Selective liquidity taking

The probability that the volume of a market order exceeds the volume at the best quote is very small. A large fraction of market orders exactly match the available volume at the best quote thereby executing the whole limit order queue.

<table>
<thead>
<tr>
<th>Stock</th>
<th>$P_&gt;$</th>
<th>$P_=$</th>
<th>tick</th>
</tr>
</thead>
<tbody>
<tr>
<td>APP</td>
<td>7.5%</td>
<td>48.7%</td>
<td>small</td>
</tr>
<tr>
<td>AMZ</td>
<td>6.6%</td>
<td>53.0%</td>
<td>small</td>
</tr>
<tr>
<td>GOO</td>
<td>14.4%</td>
<td>25.4%</td>
<td>small</td>
</tr>
<tr>
<td>MST</td>
<td>0.9%</td>
<td>16.3%</td>
<td>large</td>
</tr>
<tr>
<td>CIS</td>
<td>0.19%</td>
<td>7.8%</td>
<td>large</td>
</tr>
<tr>
<td>ORA</td>
<td>0.12%</td>
<td>28.3%</td>
<td>large</td>
</tr>
</tbody>
</table>

**Conclusion**: Liquidity takers condition the size of their order on the available volume. The large values of $P_=$ suggest that traders frequently wish to execute more volume.
Empirical evidence for liquidity rationing III: Virtual/true market depth

Method of Rosenow & Weber:

Virtual market depth = (inverse) impact if only the initial volume in the LOB would be executed.

True market depth = (inverse) measured impact
Empirical evidence for liquidity rationing IV: Order splitting

High-volume traders need to split large orders into small child orders to execute them incrementally (over hours/days). This creates a correlated market order flow.

$$\epsilon_t = +1 \text{ if MO at } t \text{ was a buy; } \epsilon_t = -1 \text{ if MO was a sell.}$$
Empirical evidence for liquidity rationing IV: Order splitting

High-volume traders need to split large orders into small child orders to execute them incrementally (over hours/days). This creates a correlated market order flow.

\[ \epsilon_t = +1 \text{ if MO at } t \text{ was a buy; } \epsilon_t = -1 \text{ if MO was a sell.} \]

The order flow is highly predictable:

\[ \mathbb{P}[\epsilon_{t'} = +1|\epsilon_t = +1] > \frac{1}{2} \forall t' > t. \]
Order splitting and the bid-ask spread

Observed average Price impact (API) of a single market order:

\[ \mathbb{P}[\epsilon_{t'} = +1|\epsilon_t = +1] > \frac{1}{2} \forall t' > t. \]

Therefore, the measured API of a market order includes the *future* imbalance of the order flow:

API of market orders increases until reaching a plateau value \(\in [s/2, s]\). Price impact of one-shot market orders that penetrate the book is the same.
Reasons for liquidity rationing in LOBs

- A human marketplace is often conceptualized with the Walrasian paradigm: Repeated auctions lead to market clearing at the intersection between the supply and demand functions.
- LOBs implement an asymmetry between liquidity providers and takers: A transaction can only take place after the prior submission of a limit order, which is a commitment to trade without the guarantee of execution.
- Liquidity providers therefore face information leakage costs and adverse selection risks.
- Conclusion: The precise market clearing mechanism has a great effect on the behaviour of market participants. The LOB creates an incentive not to submit limit orders!
Liquidity rationing in LOBs

Market participants conceal their trading intentions. Therefore, there is a large pool of “latent liquidity” behind the visible submitted liquidity. This latent liquidity is gradually revealed and financial markets never truly clear.
The micro-macro model of financial markets

- We need to take into account the trading intentions of market participants which are not necessarily visible in the LOB.
The micro-macro model of financial markets

- We need to take into account the trading intentions of market participants which are not necessarily visible in the LOB.
- News feeds contain a few messages per day per stock. Also, hedging of portfolios occurs (max) twice a day. These are the fundamental time scales that change trading intentions of investors.
The micro-macro model of financial markets

- We need to take into account the trading intentions of market participants which are not necessarily visible in the LOB.
- News feeds contain a few messages per day per stock. Also, hedging of portfolios occurs (max) twice a day. These are the fundamental time scales that change trading intentions of investors.
- Separation of time scales (high-frequency versus fundamental news) and separation of liquidity (submitted limit orders versus latent liquidity)
Definition of latent liquidity

- If $p(t)$ decreases sufficiently, investor $k$ starts buying (the opposite is true when the price increases): Thresholds $x^{(k)}_{S/B}$.

- The latent order book collects the expected intended trading volumes $\mathbb{E}[Q^{(k)}_{S/B}|\mathcal{F}^k_t]$ at time $t$ of all investors, once a trading intention is triggered.

- Consider now functions $\mathcal{L}_S$ and $\mathcal{L}_B$, the *latent buy and sell order books*, defined on a discrete grid $\tau\mathbb{Z}$ (with $\tau$ the tick size) and image in the positive real values, defined as

$$
\mathcal{L}_B, \mathcal{L}_S : \tau\mathbb{Z} \times \mathbb{R}^+ \mapsto \mathbb{R}^+
$$

$$
\mathcal{L}_B(x, t) = \sum_{k : x^{(k)}_B(t) = x} \mathbb{E}[Q^{(k)}_B(t)|\mathcal{F}^k_t],
$$

$$
\mathcal{L}_S(x, t) = \sum_{k : x^{(k)}_S(t) = x} \mathbb{E}[Q^{(k)}_S(t)|\mathcal{F}^k_t].
$$
“Dormant” versus “outstanding” latent liquidity

When $p(t)$ moves latent liquidity becomes “outstanding” and transforms into a trading intention.
Exogeneous sell order flow: Dynamics of outstanding latent liquidity

Outstanding latent buy liquidity:

\[ n^B(p_t) = \int_{p_t}^{\infty} d x \rho_b(x) \]

Clearing of outstanding latent liquidity:

- Executed limit orders are refilled.
- A fraction \( \lambda_L \) of outstanding latent liquidity is submitted via market orders (per unit time).
- All orders have fixed size \( \omega \).

Linear and constant latent liquidity (no exogeneous news):

\[ \rho_b(x) \approx L(p(0) - x) \]

Conservation equation for outstanding buy liquidity:

\[ \dot{n}^B(t) = L \dot{p}(t)p(t) - \lambda_L n^B(t) - \lambda_S(t). \]
The micro-macro model of financial markets. Price impact of an exogeneous sell order flow

Linear price impact $\rho$ of the order flow which comprises of an \textbf{exogeneous} and an \textbf{endogeneous} component :

$$\dot{p}(t) = -\rho \lambda_S + \rho \lambda_L n^B(t).$$
Impact formula

We find an inhomogeneous Riccati equation for the average impacted price:

\[ \dot{p}(t) + \omega \lambda_L p(t) - \frac{\rho L \lambda_L}{2} p(t)^2 = 2\rho \lambda_L Q_S(t) - \rho \lambda_S(t), \]

with \( Q_S(t) = -\omega \int_0^t dt' \lambda_S(t') \) the aggregate traded volume. When markets clear efficiently, i.e. \( \lambda_L \to \infty \), we obtain

\[ p(t) \approx \frac{\omega}{\rho L} - \sqrt{\frac{\omega^2}{\rho L^2} - \frac{4}{L} Q_S(t)}, \quad Q_S(t) < 0. \]

For large \( Q_S \) we find square-root impact.
Empirical square-root law

Impact of large sequential orders on the Bitcoin/USD exchange market:

Impact in %
Volume in Bitcoins
Trading rate: 10 BTC/s

Impact in %
Volume in Bitcoins
Trading rate: 3 BTC/s

Non-Markovian optimal execution

Linearized price impact equation for $\lambda_L$ large:

$$p(t) \simeq \frac{\omega}{L} - \sqrt{q_t} + \frac{1}{2\rho L \lambda_L} \frac{\dot{q}_t}{q_t} - \frac{1}{4\rho \omega \lambda_L} \frac{\dot{q}_t}{\sqrt{q_t}},$$

$$q_t = \frac{\omega^2}{\rho^2 L^2} - \frac{4Q_S(t)}{L}.$$

Optimal execution problem:

$$\arg\max_{\lambda_S} \left[ -\omega \int_0^T dt' \lambda_S(t)p(t) \right], \text{ under the constraint}$$

$$Q = \omega \int_0^T dt' \lambda_S(t).$$
Non-Markovian optimal execution

Optimal execution problem:

$$\arg \max_{\lambda_S} \left[ -\omega \int_{0}^{T} dt' \lambda_S(t)p(t) \right], \text{ under the constraint}$$

$$Q = \omega \int_{0}^{T} dt' \lambda_S(t).$$

We find the Euler-Lagrange equations:

$$\frac{1}{L} \ddot{q}_t q_t - \frac{1}{2L} q_t^2 - \frac{1}{2\omega} \ddot{q}_t q_t^{3/2} + \frac{1}{8\omega} \dot{q}_t^2 \sqrt{q_t} = 0,$$

with the boundary condition

$$q_0 = \frac{\omega^2}{\rho^2 L^2}, \quad q_T = \frac{4Q_T}{L} + \frac{\omega^2}{\rho^2 L^2}.$$
Conclusion

- LOBs implement an asymmetry between liquidity providers and takers. This leads to liquidity rationing.

- We have developed a macroscopic approach to market clearing which takes into account the strategic behaviour of investors who are interested in trading (order splitting).

- An external order flow generates an endogeneous order flow in the opposite direction.

- The resulting price impact is non-linear and non-Markovian.