INSIGHT.DATA.CLARITY.

Market Liquidity, Information and High Frequency Trading: Towards New Market Making Practices?

Charles-Albert Lehalle,

joint works with Mathieu Rosenbaum, Pamela Saliba and Othmane Mounjid

Senior Research Advisor (Capital Fund Management, Paris); Visiting Researcher (Imperial College, London)

European Institute of Financial Regulation (EIFR), Paris – April 5, 2018

Outline

Towards New Market Makers?

The title of [Menkveld, 2013] was explicit: *High Frequency Trading and The New-Market Makers*, in modern markets, it seems markets are made by more **technology oriented** and **flow driven** participants:

- Nyse booth, citadelle sec
- MiFID2 (fixed income markets + transparency)
- capital requirements (and liquidity attrition)

It raises questions:

- how do they operate? (Is it really different from before)
- is the liquidity they provide of a **different nature**? (What are the drivers if this liquidity)
- the "liquidity bifurcation theory" emerged, is it true?

Recently academics made a lot of progresses about the understanding of the underlying mechanisms. We will present and discuss them today.

INSIGHT.DATA.CLARITY.

Positioning of This Talk

This presentation relies on 3 papers:

① (Empirical) *The Behaviour of High-Frequency Traders Under Different Market Stress Scenarios*, by N. Megarbane, P. Saliba, C.-A. L and M. Rosenbaum [Megarbane et al., 2017];

② (Theoretical) *Limit Order Strategic Placement with Adverse Selection Risk and the Role of Latency*, by C.-A. L and O. Mounjid [Lehalle and Mounjid, 2016];

- (Theoretical)*Optimal High Frequency Interactions with Orderbooks*, by O. Moundji, C.-A. L and M. Rosenbaum.

The research questions beneath this papers are

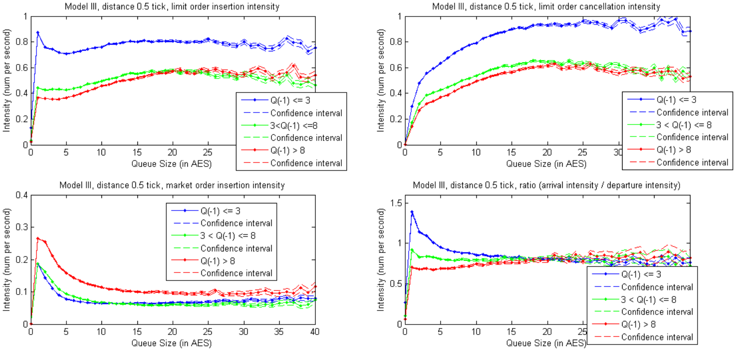
- We know **orderbook dynamics can be modelled with accuracy** (especially the liquidity dynamics [Huang et al., 2015]),
- Can we use this kind of prediction to adjust the now standard optimal trading strategies ? Some attempts have been made on "**trading speed-controlled strategies**" [Cartea and Jaimungal, 2015, Lehalle and Neuman, 2017], but what about "**order-controlled tactics**"?
- Can we have clues about how HFT (or other traders) use such prediction-driven strategies and tactics in practice?

Orderbook Dynamics: The Queue Reactive Model

The **Queue Reactive Model** introduced by Weibing Huang during his PhD thesis [Huang et al., 2015] shows that

- The flows providing liquidity (i.e. limit orders) and consuming liquidity (i.e. cancel and market orders) and a queue of a limit orderbook can be modelled by Poisson processes
- There intensities are functions of the size of the considered queue and its nearest neighbours.

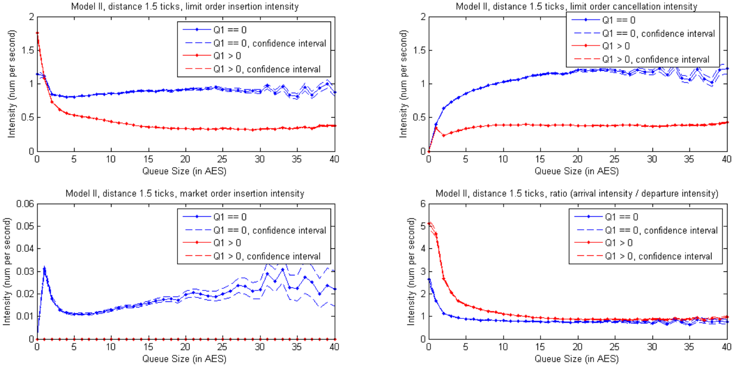
First Limit



The **Queue Reactive Model** introduced by Weibing Huang during his PhD thesis [Huang et al., 2015] shows that

- The flows providing liquidity (i.e. limit orders) and consuming liquidity (i.e. cancel and market orders) and a queue of a limit orderbook can be modelled by Poisson processes
- There intensities are functions of the size of the considered queue and its nearest neighbours.

Second Limit



The Queue Reactive Model Leads To Liquidity Predictability

This means that:

- Given you know the state of the liquidity offer (i.e. size of queues in the book)
- You have a good estimate of the distribution of the sequence of next events.

- Can this be used to **pilot a limit order**?
- In other terms: can market participants looking at orderbook state be **more efficient in providing liquidity**?

$\rightarrow$ paper ②, but this paper is today more on **the influence of exogenous parameters** (market stress).

- When there is no information on the price (just before news), it should be easier to provide liquidity...

$\rightarrow$ paper ①, it addresses **the reaction to endogenous dynamics**.

This is an on-going research program: the two papers are not perfectly aligned yet, any comment / suggestion is welcome.

INSIGHT.DATA.CLARITY.

Outline

Why Focus on HFT Behaviour?

It seems that HFT are **the new (and only) liquidity providers** on Equity markets [Jovanovic and Menkveld, 2010] (in Europe and US for sure, and probably soon in Asia too), hence

- the resiliency of the liquidity they provide is questioned (at least after each –large or small– flash crash)
- the breath of the liquidity they provide is questioned too (cf the “liquidity bifurcation” theory).
- Since regulations are pushing other markets to more electronification, this is important.

They are some studies on US markets [Brogaard et al., 2012, Subrahmanyam and Zheng, 2015], but not that much on European ones.

Two main open questions

- do HFT “ **follow the crowd** ” so much that they provide a liquidity that is not really useful?
- do HFT **provide liquidity** another way when market conditions are stressed (because market participant would need market makers under market stress)?

We would like to address the two questions, the paper presented in this section focusses on the second one. The paper of the other section focuses on the first question.

CFM INSIGHT.DATA.CLARITY.

Empirical Understanding of HFT Usual Behaviour

The data and some descriptive statistics.

The database is provided by the French regulator (AMF), all orders (and transactions) are labelled by the name of the owner, which allows us to identify HFTs. It covers the trades and orders on the most liquid French securities (36 of the CAC 40 stock), from November 2015 to July 2016 (approximatively 40 millions of trades and 1.2 billions of orders to be processed).

❶ HFT are the main liquidity providers in the LOB

Presence in the LOB	Market share in (market depth)
At the best bid and offer	70.8 %
At the two best prices	77.3 %
At the three best prices	79.3 %

❷ And they are very diverse

	A/P ratio below 50%	A/P ratio over 50%
Part in nbe	60%	40%
Part in amount	45%	55%
Avg ratio (std)	25% (18%)	67% (10%)

❸ Everyone trades with everyone

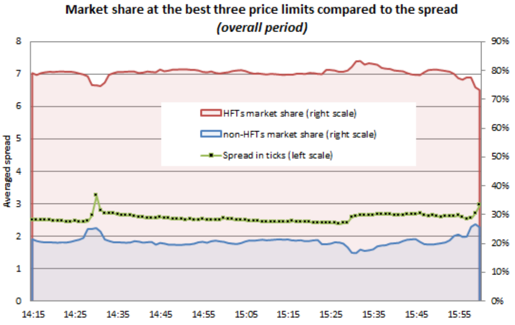
Cons./Prov.	HFTs	non-HFTs	
HFTs	33.6%	31.2%	64.8%
non-HFTs	22.4%	12.8%	35.2%
	56.0%	44.0%	

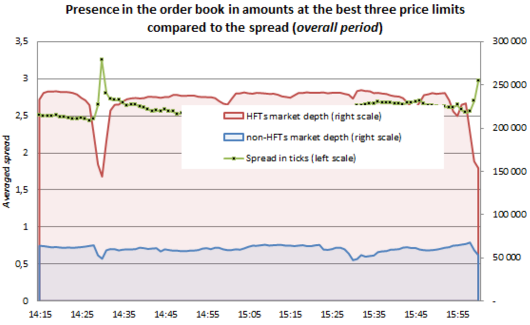
But HFT are not providing that much liquidity to trades

Usual Intraday Behaviour of HFT

- TOP: pct of presence in the first 3 limits and the bid-ask spread,
- BOTTOM: amount in Euro on the first 3 limits and the implicit volatility.
- You can notice the macro news announcements (2:30pm and 4:00pm)

Does This Average Behaviour Changes When There Are News (1/2)



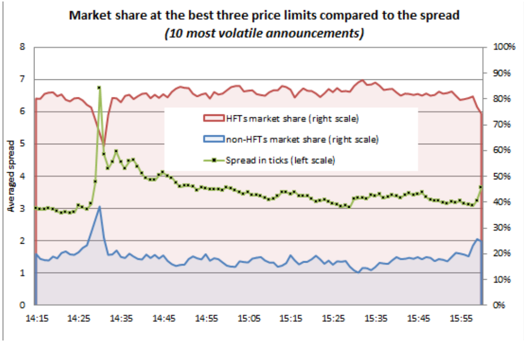


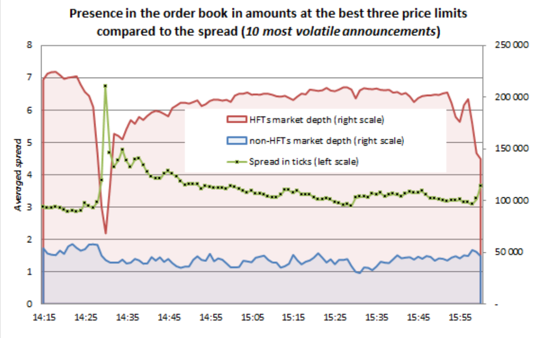
We selected the **10 most impacting News** around 2:30pm.

Left: market share (ie pct), Right: Size of the limit orders (in Euros).

The charts are different: first there is a scaling, second **the liquidity (in Euros) provided by HFT does not come back** after impacting news.

Does This Average Behaviour Changes When There Are News (2/2)





We selected the **10 most impacting News** around 2:30pm.

Left: market share (ie pct), Right: Size of the limit orders (in Euros).

The charts are different: first there is a scaling, second **the liquidity (in Euros) provided by HFT does not come back** after impacting news.

Presence in The Book Around 4:00p.m. Announcements

- We only consider news related to the U.S economy (Bloomberg news): 140 days with announcements, vs. 51 without announcements. Data are restricted between 3:40pm and 4:50pm and we consider 1min bins.
- We create 3 dummy variables: *B* (for *Before*), *D* (for *During*) and *A* (for *After*) 4:00pm.
- The empirical volatility during each 1min bin is renormalized by the avg volatility of the day.
- **Methodology:** Do a model using days without announcements only, work on the residuals of this model and try to explain these residuals on announcement days.

Explaining the pct of HFT liquidity in the book

Variable	Coef.	Std. err.	t	$P > \|t\|$	95% Conf. Int.
Const.	0.7866	0.003	302.564	0	[0.781, 0.792]
Const.	0.011	0.002	6.302	0	[0.008, 0.015]
σ_{norm}	-0.0045	0.002	-2.664	0.008	[-0.008, -0.001]
B	-0.0520	0.004	-14.404	0	[-0.059, -0.045]
D	-0.1507	0.005	-28.941	0	[-0.161, -0.141]
A	-0.0283	0.004	-7.797	0	[-0.035, -0.021]

HFT Agressive/Passive Ratio Around 4:00p.m. Announcements

- We only consider news related to the U.S economy (Bloomberg news): 140 days with announcements, vs. 51 without announcements. Data are restricted between 3:40pm and 4:50pm and we consider 1min bins.
- We create 3 dummy variables: *B* (for *Before*), *D* (for *During*) and *A* (for *After*) 4:00pm.
- The empirical volatility during each 1min bin is renormalized by the avg volatility of the day.
- **Methodology:** Do a model using days without announcements only, work on the residuals of this model and try to explain these residuals on announcement days.

Explaining HFT Agressive/Passive Ratio

Variables	Coef.	Std. err.	t	$P > \|t\|$	95% Conf. Int.
Const.	0.5340	0.002	228.198	0	[0.529, 0.539]
σ_{norm}	0.0111	0.002	5.023	0	[0.007, 0.015]
D	0.0169	0.007	2.494	0.013	[0.004, 0.03]
Const.	0.0113	0.001	9.029	0	[0.009, 0.014]
σ_{norm}	-0.0053	0.001	-4.475	0	[-0.008, -0.003]
B	0.0184	0.003	7.116	0	[0.013, 0.023]
D	0.0268	0.004	7.237	0	[0.02, 0.034]

HFT Market Share on Trades Around 4:00p.m. Announcements

- We only consider news related to the U.S economy (Bloomberg news): 140 days with announcements, vs. 51 without announcements. Data are restricted between 3:40pm and 4:50pm and we consider 1min bins.
- We create 3 dummy variables: *B* (for *Before*), *D* (for *During*) and *A* (for *After*) 4:00pm.
- The empirical volatility during each 1min bin is renormalized by the avg volatility of the day.
- **Methodology:** Do a model using days without announcements only, work on the residuals of this model and try to explain these residuals on announcement days.

Explaining HFT market share on trades

Variables	Coef.	Std. err.	t	$P > \|t\|$	95% Conf. Int.
Const.	0.5557	0.003	208.543	0	[0.551, 0.561]
σ_{norm}	0.0473	0.003	18.740	0	[0.042, 0.052]
Const.	0.0097	0.001	16.799	0	[0.009, 0.011]
B	−0.0346	0.003	−10.228	0	[-0.041, -0.028]
D	−0.0469	0.005	−9.869	0	[-0.057, -0.038]

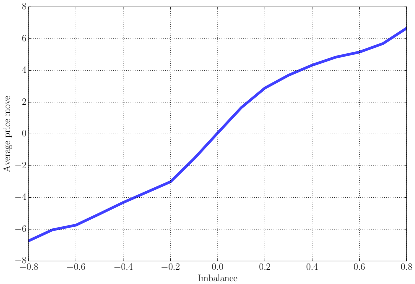
Summary of HFT Behaviour Around News

All these regressions point out in a quantitative way that **the behaviour of HFTs around announcements** cannot be read as a simple reaction to associated variations of volatility.
Around a scheduled announcement, on top of usual reactions to volatility, HFTs:

- provide 15% less liquidity,
- are slightly more aggressive,
- trade less.

On the contrary, **when no announcement is planned**, their attitude towards an increase of volatility goes in the opposite direction (trading more). We thus identify a “change of regime" in the presence of scheduled news.

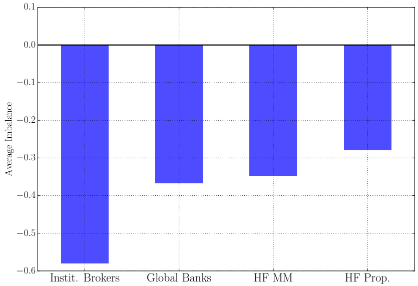
Do Market Participants Look at The Orderbook State?



① The current imbalance predicts future price moves.

We just saw that the "market context" (i.e. expected news) could influence liquidity provision by market participants taking care of orderbooks (i.e. HFT).
① To see if they react to the state of the orderbook (and following the Queue Reactive Model), we can simply try to summarize the state of the book (i.e. queues sizes), by its **Imbalance**: $(Q^{ASK} - Q^{BID})/(Q^{ASK} + Q^{BID})$.

Do Market Participants Look at The Orderbook State?

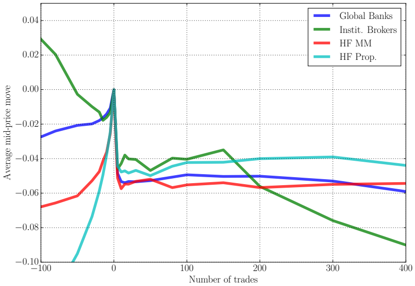


② The state of the imbalance given each type of participants traded with a limit order.

We just saw that the "market context" (i.e. expected news) could influence liquidity provision by market participants taking care of orderbooks (i.e. HFT).
① To see if they react to the state of the orderbook (and following the Queue Reactive Model), we can simply try to summarize the state of the book (i.e. queues sizes), by its **Imbalance**: $(Q^{ASK} - Q^{BID})/(Q^{ASK} + Q^{BID})$.

② We used a dataset of trades on NASDAQ-OMX (Nordic European Equity Markets), on which the identity of the buyer and a seller are know for each transaction, and synchronizing them to CFM's orderbook data. Thanks to this we can compute the average imbalance give each type of participant traded using a limit order.

Do Market Participants Look at The Orderbook State?

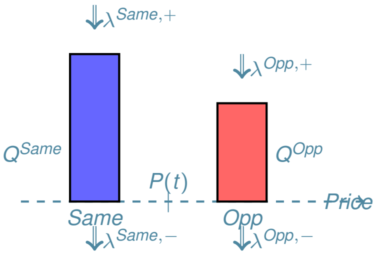


③ Price moves before and after a trade obtained via a limit order for each type of participant.

We just saw that the “market context” (i.e. expected news) could influence liquidity provision by market participants taking care of orderbooks (i.e. HFT).
① To see if they react to the state of the orderbook (and following the Queue Reactive Model), we can simply try to summarize the state of the book (i.e. queues sizes), by its **Imbalance**: $(Q^{ASK} - Q^{BID})/(Q^{ASK} + Q^{BID})$.

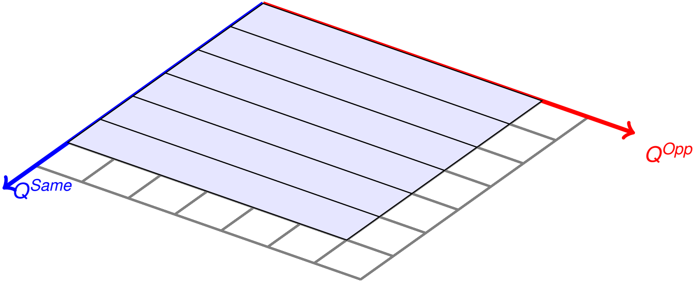
② We used a dataset of trades on NASDAQ-OMX (Nordic European Equity Markets), on which the identity of the buyer and a seller are know for each transaction, and synchronizing them to CFM's orderbook data. Thanks to this we can compute the average imbalance give each type of participant traded using a limit order.

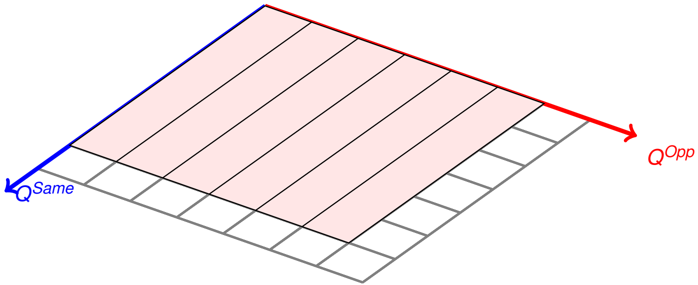
③ **It is efficient**.



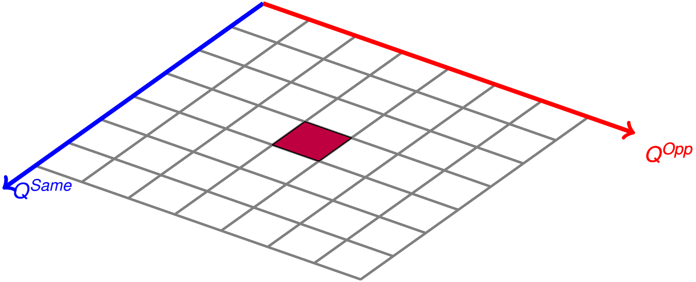
The orderbook state $U_t = \left(Q_t^{Same}, Q_t^{Opp}, P_t\right)$ can be modelled by four counting processes :

- $N_t^{Opp,+}$ (resp. $N_t^{Same,+}$) with an intensity $\lambda^{Opp,+}(Q^{Opp}, Q^{Same})$ (resp. $\lambda^{Same,+}(Q^{Opp}, Q^{Same})$) representing the inserted orders in the opposite limit (resp. same limit).
- $N_t^{Opp,-}$ (resp. $N_t^{Same,-}$) with an intensity $\lambda^{Opp,-}(Q^{Opp}, Q^{Same})$ (resp. $\lambda^{Same,-}(Q^{Opp}, Q^{Same})$) representing the canceled orders in the opposite limit (resp. same limit).


Q^{Same}
Q^{Opp}

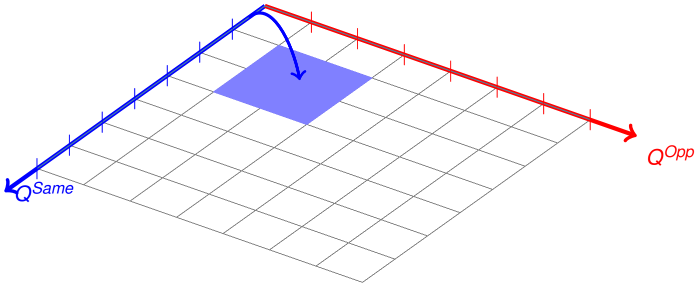

Q^{Same}
Q^{Opp}

Proposed model $\lambda^{Same,\pm}$, $\lambda^{Opp,\pm}$ depend both on the Same and Opp size

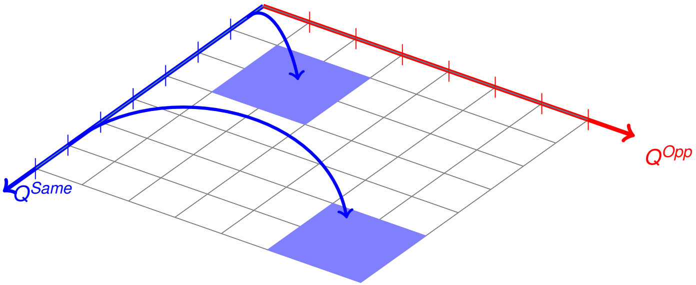


Special case: $\lambda^{Same,\pm}$, $\lambda^{Opp,\pm} = h(\mathrm{Imb}_t)$
are function of the imbalance $\mathrm{Imb}_t = \frac{Q^{Same} - Q^{Opp}}{Q^{Same} + Q^{Opp}}$

Orderbook regeneration depends on the killing state. When one limit is totally consumed, a new price P^{Disc} is discovered, a new limit Q^{Disc} replaces the consumed limit and a new quantity Q^{Ins} is inserted in front of Q^{Disc} by other market participants. P^{Disc}, Q^{Disc} and Q^{Ins} depend on the orderbook state before the price move.

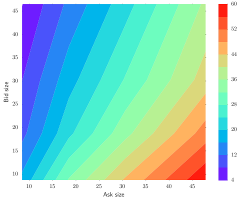


Orderbook regeneration depends on the killing state. When one limit is totally consumed, a new price P^{Disc} is discovered, a new limit Q^{Disc} replaces the consumed limit and a new quantity Q^{Ins} is inserted in front of Q^{Disc} by other market participants. P^{Disc}, Q^{Disc} and Q^{Ins} depend on the orderbook state before the price move.

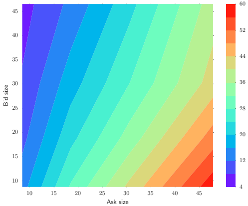


Orderbook Modelling: Comparing Empirics and Models

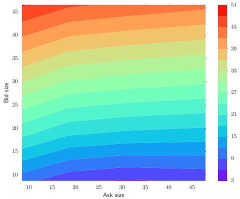
(a) Empirical Q^{Opp} after 20 events



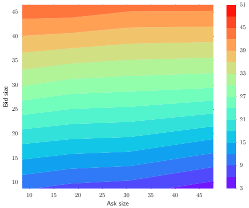
(b) Theoretical Q^{Opp} after 20 events

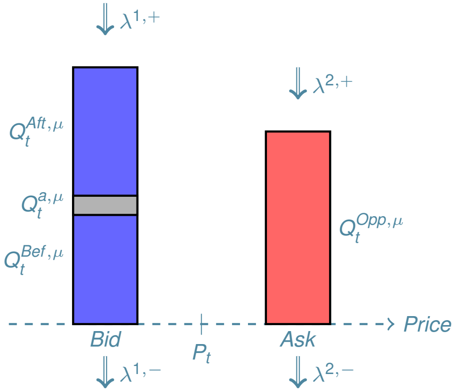


(c) Empirical Q^{Same} after 20 events



(d) Theoretical Q^{Same} after 20 events





Our model will track the position of our limit order (of size Q^a) in the first queue. The flows adding and removing liquidity are similar to the ones of the QR Model (i.e. they are Poisson with intensities conditioned by the sizes of the queues).
The different transitions are:

- if no queue goes to zero, nothing special;
- if a queue goes to zero: a new queue is "discovered" on the same side and another queue is "inserted" on the opposite side. The sizes of these new queues are conditioned by the state of the orderbook.

Using the notation u for a state of the orderbook (including the controlled order), we can show that the process U_t is ergodic under reasonable conditions, and we can show the existence of a "price at infinity":

$$g(u) = \mathbb{E}\left(P_\infty | \mathcal{F}_0, U_0 = u\right).$$

Definition of The Control Problem

The controls μ are taken from:

- Stay in the orderbook
- Cancel (and then reinsert at the top of the queue)
- Convert it in a market order.

You have two versions of the control problem: either the decision can be taken every Δ seconds, either it can be taken at any orderbook move.

Once the order is executed at time T^{μ}_{Exec} at price P, we value the strategy at

$$\sup_{\mu} \mathbb{E}\left[f \circ \mathbb{E}\left(P^{\mu}_{\infty} - P \,\middle|\, \mathcal{F}_{T^{\mu}_{Exec}} \right) - c\, q^{a}\, T^{\mu}_{Exec} \right].$$

Where c is a waiting cost, f can be any (Lipschitz) function, and $\mathbb{E}\left(P^{\mu}_{\infty} \,\middle|\, \mathcal{F}_{t} \right)$ is the price at infinity given the state of the orderbook at t

The controls μ are taken from:

- Stay in the orderbook
- Cancel (and then reinsert at the top of the queue)
- Convert it in a market order.

You have two versions of the control problem: either the decision can be taken every Δ seconds, either it can be taken at any orderbook move.

Once the order is executed at time T^{μ}_{Exec} at price P, we value the strategy at

$$\sup_{\mu} \mathbb{E}\left[f \circ \mathbb{E}\left(P^{\mu}_{\infty} - P \,\middle|\, \mathcal{F}_{T^{\mu}_{Exec}} \right) - c\, q^a\, T^{\mu}_{Exec} \right].$$

Where c is a waiting cost, f can be any (Lipschitz) function, and $\mathbb{E}\left(P^{\mu}_{\infty} \,\middle|\, \mathcal{F}_t\right)$ is the price at infinity given the state of the orderbook at t

Dynamic Programming Equation (for the continuous time version)

Let $u = (q^{bef}, q^a, q^{aft}, q^{opp}, p, p^{exec})$ an initial state . The value function $V(t,u)$ satisfies:

$$\max \begin{pmatrix} g(.) - V(t,.) \\ \mathcal{A}V(t,.) - cq^a\mathbf{1} \\ V^{c-l}(t,.) - V(t,.) \end{pmatrix} = 0, \text{ when } q^a > 0. \quad (1)$$

And $V(t,u) = u$ at execution and $V(T,u) = g(u)$ at T.

We show how to make the numerics to solve (1), and we obtain results like

Estimate of the cost of latency

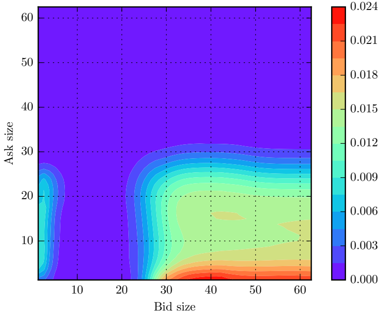
Let $V_T(0,u;\Delta_1)$ the optimal fast agent gain and $V_T(0,u;\Delta_2)$ the optimal slow agent gain.

$$|V_T(0,u;\Delta_1) - V_T(0,u;\Delta_2)| \leq H_1 \left\lceil \frac{T}{\Delta_2} \right\rceil \left\lceil \frac{\Delta_2}{\Delta_1} \right\rceil e^{C_3 T} + H_2 \Delta_2 T,$$

where H_1, H_2 and C_3 are constants involving parameters of the problem.

We fit the model on data and we solve it numerically providing different qualitative results.

With the parameters: $\Delta = 1$ second, $T = 10\Delta$ $Q^{Disc} = 22$, $Q^{Ins} = 3$, $q = 1$, $c = 0$, and the tick is0.01.



Difference between the value of a “join the bid” strategy and the value of the optimal one.

We show how to make the numerics to solve (1), and we obtain results like

Estimate of the cost of latency

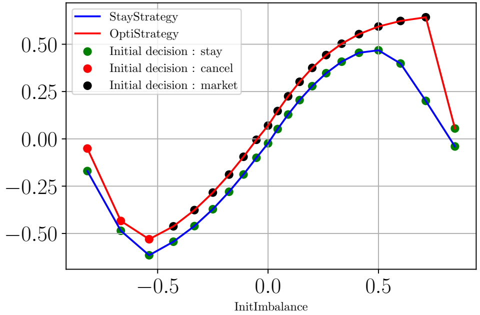
Let $V_T(0,u;\Delta_1)$ the optimal fast agent gain and $V_T(0,u;\Delta_2)$ the optimal slow agent gain.

$$|V_T(0,u;\Delta_1) - V_T(0,u;\Delta_2)| \leq H_1 \left\lceil \frac{T}{\Delta_2} \right\rceil \left\lceil \frac{\Delta_2}{\Delta_1} \right\rceil e^{C_3 T} + H_2 \Delta_2 T,$$

where H_1, H_2 and C_3 are constants involving parameters of the problem.

We fit the model on data and we solve it numerically providing different qualitative results.

With the parameters: $\Delta = 1$ second, $T = 10\Delta$, $\lambda^{Same,+} = \lambda^{Opp,+} = 0.06$, $\lambda^{Same,-} = \lambda^{Opp,-} = 0.12$, $Q^{Disc} = 5$, $Q^{Ins} = 2$, $q = 1$, $c = 0.0085$ and the tick is 0.01. Moreover $Q^{bef}(0) = 1$.



An extreme simulation to compare the "join the bid" strategy and the optimal one.

Conclusion and Open Questions

We have seen different ways market participants aware of orderbook dynamics can interact with the price formation process:

Analyzing **HFT behaviour on real data** (and especially around announced new), we saw that

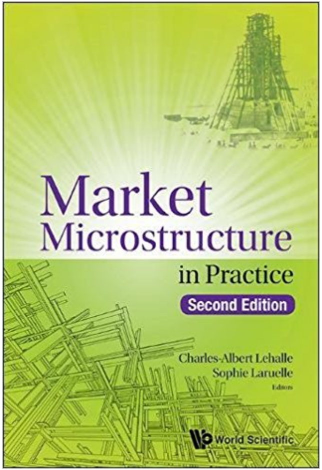
- HFT are the main providers of liquidity available in the orderbook (around 75%)
- but they provide liquidity in 56% of the trades only.
- Moreover, around news, they provide 15% less liquidity, are slightly more aggressive, and trade less.
- when no announcement is planned, their attitude towards an increase of volatility goes in the opposite direction (trading more).

Participants **taking care of orderbook dynamics**

- are better protected **in practice** against adverse selection (being "imbalance-aware")
- **in theory**, using a modified version of the Queue Reactive model fit on real data, it is possible to obtain the observed protection against adverse selection.

Open Questions: What if all participants are following this "optimal strategy" (leading to an MFG [Lachapelle et al., 2016])? In the second part we focus on endogenous dynamics (orderbook state), and in the first part more on exogenous dynamics (external news), how can we link them?

Thank You For Your Attention



To submit papers:
Market Microstructure and Liquidity .

More on market microstructure:
Market Microstructure in Practice by C.-A. L and Sophie Laruelle (World Scientific Publisher, 1s ed. 2013, 2nd ed. 2018).

References I

Brogaard, J., Baron, M., and Kirilenko, A. (2012).
The Trading Profits of High Frequency Traders.
In *Market Microstructure: Confronting Many Viewpoints*.

Cartea, A. and Jaimungal, S. (2015).
Incorporating Order-Flow into Optimal Execution.
Social Science Research Network Working Paper Series.

Huang, W., Lehalle, C.-A., and Rosenbaum, M. (2015).
Simulating and analyzing order book data: The queue-reactive model.
Journal of the American Statistical Association, 10(509).

Jovanovic, B. and Menkveld, A. J. (2010).
Middlemen in Limit-Order Markets.
Social Science Research Network Working Paper Series.

Lachapelle, A., Lasry, J.-M., Lehalle, C.-A., and Lions, P.-L. (2016).
Efficiency of the Price Formation Process in Presence of High Frequency Participants: a Mean Field Game analysis.
Mathematics and Financial Economics, 10(3):223–262.

Lehalle, C.-A. and Mounjid, O. (2016).
Limit Order Strategic Placement with Adverse Selection Risk and the Role of Latency.

Lehalle, C.-A. and Neuman, E. (2017).
Incorporating Signals into Optimal Trading.

Megarbane, N., Saliba, P., Lehalle, C.-A., and Rosenbaum, M. (2017).
The Behaviour of High-Frequency Traders Under Different Market Stress Scenarios.
Technical report, SSRN.

References II

Menkveld, A. J. (2013).
High Frequency Trading and The New-Market Makers.
Journal of Financial Markets, 16(4):712–740.

Subrahmanyam, A. and Zheng, H. (2015).
Limit Order Placement by High-Frequency Traders.
Social Science Research Network Working Paper Series.