Understanding the role of Exogenous Information in High Frequency market Regimes

Charles-Albert Lehalle

Version 0.1, printed the July 7, 2023

Abstract

This project targets to get an understanding of orderbook dynamics that goes further the current state of the art.

To start with, student will learn how to fit a Queue Reactive Model on tick by tick data (step 1). Then they will study the main weakness of the model: its focus on endogenous dynamics and not on exogenous ones. That for, timestamped News will be used and the sensitivity of the price formation to the occurrence of News will be understood (step 2).

The group will then try to reflect this understanding of the influence of News in the modelling, and potentially use Machine Learning that for (step 3).

At the end of the project, students will have a very deep understanding of high frequency dynamics on limit orderbooks. They will have gained an understanding of the different roles of endogenous and exogenous effects in HF price dynamics; it will give the knowledge needed to later try causal inference for orderbook dynamics that is suggested by Webster (2023).

Modelling HF dynamics at the LOB scale. There is now a well established literature around Limit Orderbook (LOB) dynamics. Following the Queue Reactive Model (QRM) of Huang et al. (2015), one of the valid way to model the intensities of occurrences of events (insert, cancel, trade) on a LOB is to make each of them a function of the size of the queues. One simple way is to do it using a grid:

- Each queue size Q_i (where i stands for the $|i|^{th}$ limit, on the bid side when $i > 0$ and on the ask side otherwise) is discretized in *Average Even Size*;
$$D_i := \frac{Q_i}{\text{AES}}.$$
- The intensities λ_i^e of event $e \in \{\text{insert, cancel, trade}\}$ occurring on queue i is modelled by the average intensity given the discretized size of the same queue and the one "in front of it":[1]
$$\hat{\lambda}_i^e(D_i = d_i, D_{i\odot} = d_{i\odot}) := \mathbb{E}(\lambda_i^e | D_i = d_i, D_{i\odot} = d_{i\odot}),$$
when discretized quantity d_j is observed on queue j.

[1] the operator $\odot$ is such that $i\odot = i + 1$ when $i < -1$, $i\odot = i - 1$ when $i > +1$, and $1\odot = -1$ and $-1\odot = 1$.

This approach is know to capture very efficiently *liquidity dynamics*, i.e. when there is no sudden price jump or microtrend, but is too mean-reverting when the high frequency dynamics are driven by (probably exogenous) price changes. Academic papers seems to evaluate the fraction of sudden price changes at 10 to 20% of the time.

Going further by understanding the conditions of price changes. It is not clearly established today that sudden price changes are triggered by exogenous events. Even if the natural meaning of "exogenous event" is an information that comes from "the outside", it is not always clear at this time scale; for instance, are events occurring on another LOB (of a correlated security) exogenous or not? Nevertheless all authors agree that *News are exogenous*: they inject information in the price formation process independently from the current shape of the orderbook. The more unexpected this information, the larger the move (the discovery of the covid vaccine moved Pfizer price, see Mason and Elkassabgi (2022)). Nevertheless it maybe that some sudden changes are endogenous, like Joulin et al. (2008) or Marcaccioli et al. (2022) is describing.

Expected work

1. Estimate and study a Queue Reactive Model (QRM) on real data. UC Berkeley Haas has access to the `databento` tick by tick dataset. It is hence possible to reproduce the numerical steps of Huang et al. (2015), focusing on the first limits of the LOB:

1. study the average event size and find a realistic discretization for the first queues of the orderbook,
2. understand how to estimate the intensities of basic events occurring on each queue,
3. record the sizes of the 2 first queues and study the intensity of the 3 basic events on each queue given its size and the one of the queue in front of it (producing charts similar of (Huang et al., 2015, Figure 6 and 8)),
4. fit a QRM and simulate it to produce charts as (Huang et al., 2015, Figure 5).
5. simulate it to get an estimate of the volatility and measure how much it underestimates the volatility, like in (Huang et al., 2015, Figure 10). Following the QRM paper, use a exogenous parameter θ to recover the desired volatility by "resetting" the orderbook with a probability θ after each depletion of a first limit.

This first step is a starting point. It is not easy and will enable the group to understand an advanced way to model orderbook dynamics; it will also be interesting to compare the results on US stocks and the one on French stocks that have been used in the paper. In any case, the study should be primarily restricted to "large tick stocks" in the sense of Huang et al. (2016).

It will be valuable to focus on stocks on which News are available (see next step).

2. Measuring the effect of News. UC Berkeley Haas has access to Bloomberg labelled News with accurate timestamps. Thanks to labelled News, it is possible to split the trading days in intervals "around News related to the stock"[2] vs. other intervals "away from News related to the stock".

The goal of this step is to compare the HF price dynamics around News with the dynamics away from news:

1. Measure the quantiles of the price changes around News and look how many price changes of the same amplitude are observed away from News (that is a simple version of what has been done in Joulin et al. (2008));

2. Estimate volatility around News and away from News and compare; if there is a gap, is it of the sample amplitude of the underestimated volatility by the QRM?

We could look for other metrics at the level of the orderbook, like the basic 3 intensities on the first bid and ask queues. Comparing them around News and away from News will be a good preparation for the next step.

3. Going further than existing models. The goal would be to find a way to improve the QRM incorporating exogenous effects a realistic way. Different paths can be tried

1. To understand what the QRM is missing, compute the likelihood of the QRM in the intervals away from News vs. the intervals around News. We can expect the likelihood to be larger away from News if the missing volatility comes from exogenous events.
   It will be interesting to see is the likelihood decreases slowly before or during News or if the change is sudden.

2. A more direct way (especially if the change is sudden) is to fit two QRM: one away from News and one around News. It will be equivalent to compare the conditional intensities of events away from News or around News.

3. Depending on the length of the memory that will have been identified, a non linear multilayers perceptron or an LSTM can be feed with the inputs of the QRM, adding a term related to the recovery of the volatility in the loss function, to see how this regularization improves the fit in the direction of volatility without being too detrimental to the liquidity dynamics.

Contact. Charles-Albert Lehalle – `charles.lehalle@adia.ae`
__________ +971 24 155 259.

References

W. Huang, C.-A. Lehalle, and M. Rosenbaum. Simulating and analyzing order book data: The queue-reactive model. *Journal of the American Statistical Association*, 110(509):107–122, 2015.

[2]"around" can mean 5 minutes before and 20 minutes after, to be investigated.

W. Huang, C.-A. Lehalle, and M. Rosenbaum. How to predict the consequences of a tick value change? evidence from the tokyo stock exchange pilot program. *Market Microstructure and Liquidity*, 2(03n04):1750001, 2016.

A. Joulin, A. Lefevre, D. Grunberg, and J.-P. Bouchaud. Stock price jumps: news and volume play a minor role. *arXiv preprint arXiv:0803.1769*, 2008.

R. Marcaccioli, J.-P. Bouchaud, and M. Benzaquen. Exogenous and endogenous price jumps belong to different dynamical classes. *Journal of Statistical Mechanics: Theory and Experiment*, 2022(2):023403, 2022.

A. N. Mason and A. Elkassabgi. Evidence of abnormal trading on covid-19 pfizer vaccine development information. *Journal of Risk and Financial Management*, 15(7):299, 2022.

K. T. Webster. *Handbook of Price Impact Modeling*. CRC Press, 2023.