M2 Research Internship: Interpretable non-linear models of price impact

Laboratory name: CFM Chair of Econophysics & Complex Systems, LadHyX
CNRS identification code: UMR CNRS 7646
Internship location: Ecole Polytechnique, Palaiseau, and Capital Fund Management, Paris.
Thesis possibility after internship: YES
Funding: YES

Supervision: Zoltan Eisler (Element Capital Partners), Iacopo Mastromatteo (CFM) & Michael Benzaquen (X)

Email: michael.benzaquen@polytechnique.edu

It is well known, that the relation between past order flow and future returns is highly non-linear [5]. The objective of the internship is to map out these effects in greater detail, and help identify the economic rationale for them. Depending on the timescale the dominant effect may be selective liquidity taking, adverse selection or the exchange of risk among participants.

The underlying dependencies are difficult to map out, and many simple models are linear. For example the Temporary Impact Model [1] can be written as

$$p_t = \sum_{t'<t} \left[G(t-t') f(v_{t'}) \epsilon_{t'} + \eta_{t'} \right] + p_{-\infty}, \tag{1}$$

where p_t is the market mid-price just before transaction t, ϵ_t is the sign of the transaction (+1 for buyer, −1 for seller initiated). G is the so-called propagator describing the temporal evolution of impact, and f describes the dependence on the size v_t of the transaction.

In a simple setting where f is constant, [3] proposes a fit of the form

$$r_t = p_{t+1} - p_t = G(1)\epsilon_t + \sum_{t'<t} \left[G(t-t'+1) - G(t-t') \right] \epsilon_{t'} + \eta_t, \tag{2}$$

but the full dependence is of course more complicated. As one certainly expects non-linearity and data is abundant, recent machine-learning techniques are likely well suited to do the analysis. One could for example propose the general form

$$r_t = p_{t+1} - p_t = \Phi[\{\epsilon_{t'}\}_{t' \le t}, \{v_{t'}\}_{t' \le t}] + \eta_t, \tag{3}$$

where Φ is a non-linear function, say, a neural network. Such approaches have proven rather successful at predicting returns [2] and they often outperform linear methods. Most of the focus at this high, transaction by transaction frequency however, has been on purely improving prediction performance. Our aim here instead is to obtain interpretable results, understand new effects and better quantify known ones. Such an approach is getting more emphasis recently outside market microstructure [6].

During the internship we shall look specifically at a large tick futures like rates. US Treasuries are traded on the CBOT with price-time priority, but the CME Eurodollar uses pro-rata matching. Volume dependence of price impact should be quite different in the two, and probably exceedingly non-linear for large sizes in the latter case where most resting orders are canceled after partial execution [4]. Good numerical skills are advised.

References

[1] J.-P. Bouchaud, J. Kockelkoren, and M. Potters, *Random walks, liquidity molasses and critical response in financial markets*, Quantitative Finance, 6 (2006), pp. 115–123.

[2] M. Dixon, *Sequence classification of the limit order book using recurrent neural networks*, Journal of Computational Science, 24 (2018), pp. 277–286.

[3] Z. Eisler, J.-P. Bouchaud, and J. Kockelkoren, *Models for the impact of all order book events*, arXiv:1107.3364, (2011).

[4] J. Field and J. Large, *Pro-rata matching in one-tick markets*. 2012.

[5] F. Patzelt and J.-P. Bouchaud, *Universal scaling and nonlinearity of aggregate price impact in financial markets*, Physical Review E, 97 (2018), p. 012304.

[6] C. Rudin, *Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead*, Nature Machine Intelligence, 1 (2019), pp. 206–215.