FinTech Accelerator Proof of Concept

BMLL Technologies (Bayesian Machine Learning Limited)
Access to massive limit order book data sets through cloud computing services

Background

The Bank of England has established a FinTech Accelerator to work with innovative firms and new technologies. This note describes a project spanning two areas of interest to the Bank: Big Data tools and financial market data.

The Bank is looking at an ever more diverse set of indicators for policy and research. This data must be stored, processed and analysed using a new generation of Big Data tools. Such tools can be expensive to procure and maintain; remote, shared resources – in other words, cloud computing services – are one way to manage projects related to vast data sets.

This proof of concept study focused on limit order book data. Trading exchanges use limit order books to record buyers' and sellers' interest in trading a particular financial instrument. Limit order books can be combined with a matching engine to automatically match trades. Data from limit order books therefore have many potential uses for the Bank, from financial markets research to financial stability studies as well as analysing the impact of policy.

During the study, we evaluated a platform that allows cloud-based analysis of limit order book data from a number of trading exchanges. We wanted to explore hardware and data storage options, in addition to how the technology could be combined with machine learning capabilities.

The Proof of Concept

We ran a Proof of Concept (PoC) with [BMLL](#)- a London-based start-up offering a product that combines:

- order book data from global trading exchanges;
- massive scale hardware based on cloud computing services; and
- a learning and inference software library.

BMLL provided a machine learning API (application programming interface) designed to recognise patterns in limit order books.

We evaluated an alpha version of BMLL's product, accessed via a web browser.

BMLL negotiates with trading exchanges in order to make available previously private historic data. These order books are reformatted and saved on Amazon Web Services storage, and the exchange receives payment for hourly access. Licence terms are agreed between BMLL and the relevant exchanges.

BMLL clients sign a single contract and are billed jointly for data, hardware and software usage. Users access the data sets via a web browser. At present, the platform is only accessible to the Python programming language. The Matlab and R languages will be added in due course. Several example code notebooks are provided and an API manual is available.

Reflections and next steps

The PoC was useful in learning more about this type of technology and how we might use it for our own data in the future. Given the product was in an alpha testing phase, we identified certain areas for further development:

- At the time of testing, only four exchanges had data available via the platform. Although we are aware that BMLL, post the platform launch in April, intend to expand the data sets available;

- The API guide would benefit from further improvement, as there were some bugs in the platform during the period of our PoC. The provider has made us aware that these are being addressed with a regular release cycle of bug fixes;

- The programming editor currently used (Jupyter) might not be intuitive to use for those familiar with other statistical programming environments. For example, code and the output are stacked vertically. We note that there are other browser-based editors that have a more familiar look and feel.

Development points such as these are not uncommon, by their very nature, in alpha products, and are easily addressed. In principle we think the concept is worth exploring, with the following benefits clearly evident:

- A contract with BMLL could, as we understand it, remove the need for licence agreements with individual exchanges. For an organisation wishing to look at data from a number of markets and regions, this could save time and effort spent on negotiating such agreements;

- Using cloud computing services to store and analyse the data might be a viable alternative for organisations that have to procure and maintain their own storage and processing capacity. With data sets often reaching petabytes in size, it may mean the difference between a project being possible or not;

- For some projects, BMLL's analytics library may prove valuable. For example, tools for Bayesian inference and microstructure analysis are under development;

- Finally, because BMLL adopts a common interface, a study could in theory be applied to many data sets with no code changes. Dealing with the exchanges independently would most likely mean data sets with different formats, and thus require additional resources.

In conclusion, we believe the concept of access to data through cloud computing services is certainly one that should be explored further. The nature of the data discussed here indicates that it would be challenging to access the full content by other means.