Big Data in securities markets

By Gabriel Callsen

The purpose of this article is to explore the concept of Big Data, its potential use in the securities markets and the implications for international debt capital markets. The article is divided into four sections and reflects ICMA's monitoring of relevant developments. The first section provides background and definitions of Big Data and associated terms such as "AI". The second section summarises regulators' views on Big Data, which has been a focus area over the last two years. The third section describes Big Data in securities markets and the use of data in fixed income markets. Finally, the fourth section provides a snapshot of potential applications of (big) data and machine learning in fixed income markets. The paper is primarily based on research, but also conversations with selected ICMA members and market participants representing the sell side, buy side, a supranational issuer, as well as data and technology providers.

(i) Defining Big Data, artificial intelligence, machine learning and algorithms

While there is no uniform definition of the term Big Data, it can be described as "data that contains greater variety arriving in increasing volumes and with ever-higher velocity".[1] This is also referred to as the three "Vs" ie variety, volume, and velocity.[2] In other words, Big Data is comprised of large and/or complex datasets, in particular from new data sources, and often includes unstructured (such as text, image, or language) or semi-structured data (a combination of data in table-format and unstructured formats).[3] Additional attributes associated with the concept of Big Data are data quality (or veracity) and data value.[4] A key feature is that Big Data cannot be processed by traditional software, but requires more advanced data processing capabilities, including applications of artificial intelligence (AI), or machine learning (ML).[5] This is also known as "Big Data analytics".

AI is broadly understood to refer to computer systems which "perform human-like tasks, such as learning, reasoning, and problem solving".[6] Machine learning (ML) can be considered a subset of AI which enables computers to "learn directly from examples, data, and experience" and "carry out complex processes by learning from data, rather than following pre-programmed rules".[7] An algorithm in mathematics and computer science can be described as "an unambiguous specification of how to solve a class of problems. Algorithms can perform calculation, data processing, automated reasoning, and other tasks".[8] However, it is worth noting that there does not appear to be a consensus on a single, consistent definition of AI and ML and to what extent ML forms part of AI.[9]

(ii) Regulators' views on Big Data

Big Data is a theme that has clearly captured financial regulators' attention. Over the last two years, a number of reports have been published and consultations launched on Big Data, exploring the potential benefits and risks arising from the use of Big Data, in particular from a financial

1. https://www.oracle.com/uk/big-data/guide/what-is-big-data.html
2. https://www.gartner.com/it-glossary/big-data/
3. FSB (2017), Artificial intelligence and machine learning in financial services. Market developments and financial stability implications, p.4.
4. BaFin (2018), Big data meets artificial intelligence - Challenges and implications for the supervision and regulation of financial services, p.17.
5. Further background on the shift between data processing systems can notably be found in Infosys (2012), Use of Big Data Technologies in Capital Markets, Viewpoint, p.2.
6. https://news.sap.com/2018/03/what-is-artificial-intelligence/
7. The Royal Society (2017): Machine learning: the power and promise of computers that learn by example, p. 16.
8. https://en.wikipedia.org/wiki/Algorithm
9. WEF (2018), The New Physics of Financial Services. Understanding how artificial intelligence is transforming the financial ecosystem, p.10.

stability and retail/consumer perspective. The overview below aims to provide a concise, albeit non-exhaustive, insight into the thinking of key regulators in relation to Big Data.

The Financial Stability Board (FSB) released a report on *Artificial Intelligence and Machine Learning in Financial Services* on 1 November 2017. As the title suggests, the focus of the paper is on AI and ML, and notably on market developments and financial stability implications. This is relevant for Big Data insofar as it is considered a key supply factor for the adoption of such technologies. Applications of AI and ML are considered to be in early stages and fast-evolving, but while potential benefits include greater operational efficiency, for instance for regulatory reporting purposes, the "use of big data from new sources" may have unexpected consequences and "lead to greater dependencies on previously unrelated macroeconomic variables and financial market prices".[10]

The Joint Committee of the European Supervisory Authorities (ESAs) published its *Final Report on Big Data* on 15 March 2018. The report provides a summary of the responses received from its consultation following a *Discussion Paper on the Use of Big Data by Financial Institutions* released on 19 December 2016. The reports describe the phenomenon of Big Data, the applicable regulatory framework, as well as potential benefits and risks for consumers and financial institutions. A key observation is the "continued increase in the use of Big Data across the banking, insurance and securities sectors, ie the collection, processing and use of high volumes of different types of data from various sources".[11]

BaFin, in the 195-page report *Big Data Meets Artificial Intelligence - Challenges and Implications for the Supervision and Regulation of Financial Services*, released on 16 July 2018, undertakes an in-depth analysis of Big Data in conjunction with artificial intelligence and machine learning (referred to as BDAI), and its use by banks, insurance companies and in capital markets. The report explores the use of Big Data and AI, its potential benefits (eg efficiency gains, personalisation of product offers, development of new products) and risks (eg discrimination, market concentration, fragmentation). A common key concern of the ESAs and BaFin relates to retail clients, in particular to profiling and segmentation practices, which could result in "differentiated", or in other words, unfair, pricing practices.

With regard to securities markets, it is acknowledged that the usage of large and diverse data sets, for instance, for high-frequency trading strategies, is not a new phenomenon. A key observation in the BaFin report is that in light of the high quality of existing models, the "increased use of BDAI may thus only offer small increases in model quality".[12] However, with regard to the impact on market structure, it is worth noting that "BigTechs" are already profiting from BDAI usage in capital markets because they are important infrastructure providers supplying cloud computing".[13] and "could quickly become systemically important".[14] Even though regulatory guidance has been provided on outsourcing to cloud service providers[15], BigTechs have generally not been subject to supervision. Looking ahead, the use of Big Data and AI is expected to lead to further automation and use of algorithms. Consequently, the report concludes that the use of Big Data and AI in capital markets is likely to be "more of the same, only faster and better".[16]

The FCA, in a speech given on 13 February 2019 by Julia Hoggett, Director of Market Oversight at the FCA, highlighted the risks of market misconduct arising from the use of (big) data and artificial intelligence. According to a commentary by Clifford Chance, in light of ever-growing volumes of data in electronic format, "it is becoming increasingly difficult to distinguish between data which is publicly available and data which is non-public and therefore potentially inside information".[17]

(iii) Big Data in capital markets

While the concept of Big Data is normally associated with retail clients, where increasing volumes of data are

10. FSB (2017), Artificial intelligence and machine learning in financial services. Market developments and financial stability implications, p.31.

11. Joint Committee of the ESAs (2016), Discussion Paper on the use of Big Data by financial institutions, p.5.

12. BaFin (2018), Big data meets artificial intelligence - Challenges and implications for the supervision and regulation of financial services, p.135.

13. Ibid. p. 141. BigTech may also be well positioned to leverage their customer networks, data collected outside the financial sector combined with data analytics capabilities to venture into the distribution of financial retail products.

14. Ibid. p.8.

15. EBA (2018), Recommendations on outsourcing to cloud service providers.

16. BaFin (2018), Big data meets artificial intelligence - Challenges and implications for the supervision and regulation of financial services, p.137.

17. Clifford Chance (2019), Big data and artificial intelligence – evolving market misconduct risks, 14 March 2019, p.3.

Big Data is a theme that has clearly captured financial regulators' attention.

generated, for example, from the use of mobile devices, social media, or connected devices (internet of things), Big Data in capital markets is not a new phenomenon. The ingestion of large amounts of data, for instance, algorithmic or high-frequency trading in equities markets, or data-driven investment strategies by certain market participants such as hedge funds are not new *per se*. Indeed, market participants tend to make use mostly of structured data sets (provided by third-party data providers, for example), and to a lesser extent unstructured data (for sentiment analysis or market surveillance).[18]

Use of (Big) data in fixed income markets

In fixed income markets, electronification has created increasingly large volumes of data, in particular over the last two years. This has entailed ever-growing capacity requirements to process and store data. Accessibility has improved significantly through the use of cloud networks, which has enabled firms that do not have the required capacity to access and make use of data.

Access to data, but also the ability to manage ever-increasing volumes of data, is therefore of critical importance for a range of purposes. First, besides liquidity, data is a key component for algorithmic trading, as is the case in equities markets, and has resulted in an increase in automated market-making. Streaming prices, and not only seeing actionable prices on screen but also receiving data in electronic format, result from improved use of data and technological progress. Transparency reporting requirements under MiFID II/R have generated vast amounts of data, which can be considered "Big Data". However, low data quality, data dispersion across different sources, lack of standardisation, and divergent national deferral regimes for the publication of trading data, have prevented market participants from making meaningful use of MiFID II/R data. Second, data is equally important for human use or manual trading. For example, mid prices, historical data and CRM data are paramount in order to make informed trading decisions. Third, data is key from a business management perspective, for example, to estimate market share, measure performance, or analyse market trends. And fourth, data is used for regulatory reporting, for example, under MiFID II/R and the forthcoming SFTR reporting regime, and for compliance purposes, for instance, to detect conduct issues or suspicious activities. However, a preliminary step is to aggregate and standardise the data for machine use, and make the data accessible in a user friendly and meaningful way if used manually. This process remains a resource intensive precondition to exploit the full potential of data.

That said, in comparison to large technology companies, but also other sectors such as aviation or other retail sectors, capital markets appear to be far behind when it comes to processing and using data. This is surprising considering market participants, notably from a sell-side perspective, generate large volumes of data. Given the data is recorded and linked to unique identifiers such as ISINs, it should be easy to use. However, to what extent BigTech firms will play a role in international debt capital markets going forward remains rather questionable. The competitive edge is perceived not to be generic cutting-edge technology, but the combination of specialist financial market expertise with cutting-edge technology. And this is an area traditionally dominated by hedge funds rather than BigTechs.

The cost of data

A key factor for the use of data is cost. Market participants are paying to aggregate data (eg using specific software to combine data from different sources in different formats), process and report data for regulatory reporting purposes (eg publication of post-trade data under MiFID II/R through APAs), and incur additional costs for using their own data provided by trading venues or data providers. Reducing the cost of data would create incentives for the producers of data to make greater use of it. However, behaviour will only change if incentives change. While the benefits of data need to outweigh the costs, it requires prior investment and leads to a typical "chicken and egg" dilemma.

18. AITE (2014), Big Data in Capital Markets: At the Start of the Journey, p.11-13.

(iv) Selected examples of "Big Data" and machine learning applications in fixed income markets

While it is not possible to make a qualitative judgement of the following examples, the aim is to provide a snapshot of recent (big) data-related developments and highlight potential use cases in fixed income markets.

ESM: Predicting investor behaviour in European bond markets through machine learning[19]

The quant team of ESM is developing, in cooperation with the Zurich University of Applied Sciences, a machine-learning based application to predict investor demand for syndicated bond issuances. A key reason for using machine learning algorithms is the ability to analyse complex and high-dimensional data sets with widely unknown structures, to capture complex dependencies and relations of variables and identify any kind of patterns in the data.

The analysis comprises diverse datasets, including transaction-related data such as orderbooks, internal and external primary and secondary market data, including secondary market transactions reported by primary dealers. In addition, it comprises internal and external investor-specific data as well as macroeconomic data. The applied ML methodology is promising. First results show a prediction power of well above 50% of investor demand by investor type (such as banks, brokers, fund managers, pension funds or insurance). While results for individual investors were overall less accurate due to smaller data sets, qualitative information and behaviour patterns of specific investors could be detected. These results can help better to understand and address investor needs and consider this in the transaction planning and execution.

This machine-learning application to predict investor demand is considered work in progress. Further improvements of data quality, inclusion of further data sources, and a refinement of the used ML algorithms are expected to improve forecasts substantially. However, a key limiting factor is the availability of data despite access to diverse data sources, including primary dealer reporting. The inclusion of further primary dealer data can play an important role. ML technology can help solve potential confidentiality issues.

Other use cases

Apart from this use case, there is a growing number of different machine learning applications that have been developed in fixed income markets. These include applications to predict pricing of new issues and existing securities, match issuers with potential investors, aggregate siloed data and provide access in real-time across asset classes, or predict financial markets movements based on alternative data to enhance investment decisions.

Conclusion

Big Data, or data characterised by greater variety, increasing volumes and ever-higher velocity (known as three "Vs") is often associated with applications of advanced analytics based on artificial intelligence or machine learning. Regulators have focused extensively on this topic, highlighting potential benefits, but also risks in relation to retail clients in particular. In capital markets, the use of large volumes of data and advanced analytics is not new *per se*. In fixed income markets, electronification has created increasingly large volumes of data. While data is used for a range of key functions, cost is a limiting factor in fixed income markets. Challenges relate in particular to data normalisation and quality. Predictive analytics based on machine-learning algorithms seem promising, but such applications are still in early stages. That said, Big Data analytics and data-driven trading strategies will certainly become more and more widespread in fixed income markets and ICMA will continue to monitor these developments closely.

Contact: Gabriel Callsen
gabriel.callsen@icmagroup.org

19. With kind permission from the European Stability Mechanism (ESM) and special thanks to Martin Hillebrand.