FINANCIAL SCIENCE

Platform technology and infrastructure in algorithmic trading

This is one in a series of white papers intended to help investors educate themselves on funds and investment techniques in order to make the right decisions regarding their assets.

In this paper we examine the current state of the art in terms of technologies being adopted in the finance sector and platforms being used to allow funds and investors to manage their assets. We have grouped these into technologies, which refers to the methodologies being applied, and infrastructure in which we describe both the investor-facing portals and the internal platforms used by funds and their data providers.

1 Financial Investment Technologies

This section reviews the main categories of digital technologies driving FinTech innovation. Our review divides digital technologies into *data technologies, algorithm technologies*, *analytics technologies* and *infrastructure technologies*. However, it is important to recognise that innovation in these areas does not occur in an isolated manner and developments in one area often enable and drive transformation in other spheres.

1.1 Data Technologies

Financial investment is driven by the availability of huge and increasingly comprehensive historic and real-time data sets. This includes both traditional sources such as financial, economic and company filings and alternative data sources gathered from outside of traditional sources, such as social media. The integration of multiple data sources, which are diverse in both content and format, requires careful management; both size and format must be considered alongside data content. With this in mind, developments data can be subdivided into four main areas:

- **Big data** – new insights resulting from the engineering of extremely large data sets analysed computationally reveal patterns, trends, and associations. Innovations in this area especially relate to noisy data which can reveal insights into human behaviours.
- **Novel data sources –** one of the most discussed developments in finance over recent years has been the use of social media data (e.g. mining Twitter or Reddit for investment insights). However, the use of non-standard data is not limited and can include other areas such as image processing techniques mining satellite images for climate data for example.
- **Data standards** – the requirement for data to be easily processed has driven the use of open standards for data representation. These include common data models for industry sectors, for example OpenBanking, FML, markup languages such as eXtensible Markup Language XML, and data interchange formats including Fast Healthcare Interoperability Resources (FHIR). Note that standards such as these, while making data universally readable, do not always guarantee the correctness of the data itself.
- **Machine reading** – development of standards such as those described above can be a slow process with less than universal coverage of the data required. Another area of development

therefore includes machine reading and Natural Language Processing (NLP) in order to ingest data sources which computers have traditionally struggled to use in a meaningful way.

1.2 Algorithmic Technologies– new forms of 'statistics' and artificial intelligence (e.g. neural networks)

The term "algorithmic technologies" in this context describes the use of algorithms to make decisions and automatically perform operations traditionally carried out with human intervention. This includes trade execution, portfolio creation, and asset valuation among others. Although algorithmic technologies are considered a relatively new development, in fact they have been used in some form for several decades. The classification defined in [Koshiyama et al, 2020; Treleaven et al, 2019] distinguishes between relatively well-established methods and more cutting-edge technologies:

- **Computational Statistics** - a large class of modern, computationally intensive statistical methods (e.g., Monte Carlo methods).
- **Complex Systems** - system featuring a large number of interacting components whose aggregate activity is nonlinear (e.g., Agent-Based systems).
- **Artificial intelligence** – AI algorithms mimicking a new form of human learning, reasoning and decision-making.

1.2.1 Automated Algorithmic Trading

An automated trading systems process may be divided [Galas, 2014; Treleaven et al, 2013, Nuti, 2011] into five stages:

- **Data access** – obtaining, storing and cleaning (financial, economic, social, alternative) historic and real-time data that will drive algorithmic trading.
- **Pre-trade analysis** – analysis of properties of assets to identify trading opportunities using market data or financial news etc.
- **Trading signal generation** – identifying a portfolio of assets to be accumulated, based on pre-trade analysis (what & when to trade).
- **Trade execution** – executing orders for a selected asset (how to trade).
- **Post-trade analysis** - analysis of the results of trading activity, such as P&L, the difference between the price when a buy/sell decision was made and the final execution price (slippage and transaction costs analysis) and overall return profiles for live and back-tested trading systems.

1.3 Analytics Technologies

The developments described in the previous sections utilise cutting-edge analytics technologies. The migration of these, often quite theoretical, areas of research into financial applications can take the form of purpose-built algorithmic funds, technology companies moving into the financial area and traditional finance companies developing their own technology.

Some of the main analytics technologies being deployed are:

- **Forecasting** – making predictions of trends based on historic data. Three basic strategies are qualitative techniques, time series analysis/projection often using LSTM neural networks, and causal models.

- **Optimisation** - selection of a best (by some criterion) configuration from a set of alternatives, or maximising (or minimising) a multivariate function subject to a set of constraints. The science of optimisation is concerned both with application (such as best portfolio selection) and operation of the algorithms themselves (for example training a neural network).
- **Natural Language Processing (NLP)** – the analysis and synthesis of natural language and speech [Chowdhary, 2020]. The ubiquitous nature of speech and language means an increasing number of pre-trained NLP models are available for use.
- **Sentiment Analysis** – directly applying statistics, or machine learning methods to the output of an NLP model to extract, identify, or characterise the sentiment content of text or speech [Batrinca et al, 2015; Kolchyna et al, 2015]. This may be done at the macro level (for example identifying market trends from multiple sources) or applied on a more individual level (e.g. to garner information on credit risk).
- **Behavioural/Predictive Analytics** – closely linked to sentiment analysis, this provides insight into the actions of people and determines patterns to predict future outcomes and trends [Kumar & Garg, 2018; Predict, 2020].
- **Computational Psychometrics** – pushing at the boundaries of the developments achieved in the previous points, this interdisciplinary field fuses psychometrics, learning, cognitive sciences, and data-driven AI-based computational models. It is applicable to complex systems which depend on the views and actions of humans, including finance.

1.4 Infrastructure Technologies

To implement the advances described in the previous sections, appropriate infrastructure must be available. Some infrastructure developments have been widely discussed in the financial and mainstream press, for example distributed ledger technologies. However, the importance of other technologies such as architectures supporting big-data analytics are much less widely acknowledged. Important digital infrastructure technologies related to investment include:

- **Big data analytics platforms** – large institutions may choose to create their own environments but cloud computing solutions are also available to allow rapid rollout of a scalable solution to smaller and more agile companies [Galas et al, 2015].
- **Federated Learning** – an infrastructure and machine learning technique that trains an algorithm across multiple decentralised data sources, without direct access to the data [Treleaven et al, 2022]. This can have applications for the preservation of privacy when dealing with financial data, and supports companies that do not wish to share proprietary datasets with competitors. However challenges exist around the standardisation of data from diverse sources.
- **Distributed Ledger Technologies** (DLT) - DLT and Blockchain are distributed databases that secure, validate and process transactional data. Although Blockchain looms largest in the public consciousness, other applications are highly relevant to finance and trading, for example in the area of smart contracts [Treleaven et al, 2017].

2 Trading Platforms

The platform landscape ranges from simple web sites to complete algorithmic trading systems: We distinguish between *investment platforms* - an online service allowing clients to buy, sell and hold assets and funds; and *trading platforms* - software infrastructure provided by a financial intermediary to support investment platforms [Galas et. al, 2012]. Colloquially, investment platforms are analogous to retail sites and trading platforms to logistics.

2.1 Investment Platforms

These are commercial platforms targeted at investors, characterised by ease of use and an assortment of helpful features, such as news feeds and charts, for investor education and research:

- *Marketing web sites* – sites linking clients to fund managers (e.g.); active and passive.
- *Managed funds* - a fund manager or automated trading system that oversees the fund and decides which securities it should hold, in what quantities and when the securities should be bought and sold.
- *Fund supermarkets* - generally more transactional and used for buying funds cheaply online.
- *Assets and Fund wraps* - services where investors access their account details online. Typical asset classes are equities and Exchange Traded Funds.

2.2 Trading Platforms

These proprietary infrastructure platforms are developed by brokerages to suit their own or client requirements:

- *Data providers* - data vendors providing traditional market and alternative data to financial firms, traders, and investors. Data distributed is collected from sources such as stock exchange feeds, brokers and dealer desks or regulatory filings.
- *Systematic trading* – systems defining trade goals, risk controls and rules making investment and trading decisions in a methodical way. This includes both manual, and full or partial algorithmic trading.
- *In-house service* – providing connectivity for a financial institution's staff.
- *Broker services* - connecting buyers and sellers to exchanges to facilitate a transaction (e.g. Interactive Brokers).
- *Market maker* - a brokerage that provides buy/purchase and sell solutions for investors, profiting on the price spread.

3 Financial Science

Professional Investors (e.g. family offices, high-net-worth individuals, public institutions) have traditionally been poorly served with information about the operation and performance of automated investment platforms. In contrast to this, our platform and infrastructure allows investors to work with and review algorithms to achieve the following

- **AI-assisted Investing** – an integration of digital technology for automation and review of iterative investing processes.
- **Big data** – to customise and then harvest comprehensively financial, business, economic, sensors, social media, and alternative data.
- **Algorithms** – to deploy state-of-the-art machine learning and computational statistics algorithms for data analysis and automated trading.
- **Transparent fees structure** – facility for an investor to select the desired fee structure and perform 'what-if' on proposed fee structures based on assets managed.

3.1 Financial Science research platform

The Financial Science research initiative was set up to investigate and optimise an algorithmic investment strategy focused on resilience and long term capital growth. To serve professional investors, the platform is designed to provide a central portal as the key point of contact with comprehensive information on investments including transparent fees calculations and findings of algorithms' performance.

To maximise transparency, tearsheets present data on holdings, performance statistics and back test research. This allows evaluation of performance, including reviews of ongoing enhancements and their results.

The Financial Science core philosophy holds that the interests of investors are best served through transparency in both performance and fee structures (particularly the compounding long term consequences). To support transparency and fairness, an online fee calculator and optimiser, driven by game and contract theory, presents data with long-term fiduciary evaluation as the central objective.

4 References

[1] Koshiyama, Adriano, Nick Firoozye, and Philip Treleaven. "Algorithms in future capital markets." Available at SSRN 3527511 (2020).

[2] Philip Treleaven, Jeremy Barnett, and Adriano Koshiyama. Algorithms:.law and regulation. Computer, 52(2):32–40, 2019.

[3] Galas, Michal. Experimental Computational Simulation Environments for Algorithmic Trading. Diss. UCL (University College London), 2014.

[4] Treleaven, Philip, Michal Galas, and Vidhi Lalchand. "Algorithmic trading review." Communications of the ACM 56.11 (2013): 76-85.

[5] Chowdhary, KR1442. "Natural language processing." Fundamentals of artificial intelligence (2020): 603-649.

[6] Bogdan Batrinca and Philip C Treleaven. Social media analytics: a survey of techniques, tools and platforms. Ai & Society, 30(1):89–116, 2015.

[7] Olga Kolchyna, Tharsis TP Souza, Philip Treleaven, and Tomaso Aste. Twitter.sentiment analysis: Lexicon method, machine learning method and their combination. arXiv preprint arXiv:1507.00955, 2015.

[8] Garg, Harish, and Kamal Kumar. "An advanced study on the similarity measures of intuitionistic fuzzy sets based on the set pair analysis theory and their application in decision making." Soft Computing 22.15 (2018): 4959-4970.

[9] Galas, Michal. "Experimental Computational Simulation Environments for Big Data Analytic in Social Sciences." Handbook of Statistics. Vol. 33. Elsevier, 2015. 259-277.

[10] Treleaven, Philip, Malgorzata Smietanka, and Hirsh Pithadia. "Federated Learning: The Pioneering Distributed Machine Learning and Privacy-Preserving Data Technology." Computer 55.4 (2022): 20-29.

[11] Treleaven, Philip, Richard Gendal Brown, and Danny Yang. "Blockchain technology in finance." Computer 50.9 (2017): 14-17.

[12] Galas, Michal, Brown, Daniel and Treleaven, Philip "A computational social science environment for financial/economic experiments." Proceedings of the Computational Social Science Society of the Americas 1 (2012): 1-13.
[13] Nuti,Giuseppe, Mirghaemi, Giuseppe, Treleaven, Philip and Yingsaeree, Chaiyakorn Algorithmic trading. Computer, 44(11):61–69, 2011.