

# Portfolio Management- Part II

# Fintech in Investment Management

Study Session 18

Reading No – 55

Version 2022

# Learning Outcome Statements

## **The candidate Should be able to:**

- a. describe “fintech;”
- b. describe Big Data, artificial intelligence, and machine learning;
- c. describe fintech applications to investment management;
- d. describe financial applications of distributed ledger technology.

# los a: Describe Fintech

- Fintech: Technological innovation in the design and delivery of financial services and products in the financial industry
- Fintech can also refer to companies (often new, startup companies) involved in developing the new technologies and their applications, as well as the business sector that comprises such companies
- Areas of fintech development that are more directly relevant to the investment industry include the following:
  - Analysis of large datasets: In addition to growing amounts of traditional data, such as security prices, corporate financial statements, and economic indicators, massive amounts of alternative data generated from non-traditional data sources, such as social media and sensor networks
  - Analytical tools: For extremely large datasets, techniques involving artificial intelligence (AI)—computer systems capable of performing tasks that previously required human intelligence may be better suited to identify complex, non-linear relationships than traditional quantitative methods and statistical analysis
  - Automated trading. Executing investment decisions through computer algorithms or automated trading applications may provide a number of benefits to investors, including more efficient trading, lower transaction costs, anonymity, and greater access to market liquidity

# los a:Describe Fintech

- 4. Automated advice. Robo-advisers or automated personal wealth management services provide investment services to a larger number of retail investors at lower cost than traditional adviser models can provide
- 5. Financial record keeping. New technology, such as DLT, may provide secure ways to track ownership of financial assets on a peer-to-peer (P2P) basis. By allowing P2P interactions—in which individuals or firms transact directly with each other without mediation by a third party

# los b. describe Big Data, artificial intelligence, and machine learning;

- Big Data: The vast amount of data being generated by industry, governments, individuals, and electronic devices that arises from both traditional and non-traditional data sources
- Alternative data: Non-traditional data types generated by the use of electronic devices, social media, satellite and sensor networks, and company exhaust
- Traditional data also include data that are generated in the financial markets, including trade prices and volumes. Because the world has become increasingly connected, we can now obtain data from a wide range of devices, including smart phones, cameras, microphones, radio-frequency identification (RFID) readers, wireless sensors, and satellites that are now in use all over the world
- The term Big Data typically refers to datasets having the following characteristics:
  - Volume: The amount of data collected in files, records, and tables is very large, representing many millions, or even billions, of data points
  - Velocity: The speed with which the data are communicated is extremely great. Real-time or near-real-time data have become the norm in many areas
  - Variety: The data are collected from many different sources and in a variety of formats, including structured data (e.g., SQL tables or CSV files), semi-structured data (e.g., HTML code), and unstructured data

# los b. describe Big Data, artificial intelligence, and machine learning;

- Big Data, therefore, encompasses data generated by
- financial markets (e.g., equity, fixed income, futures, options, and other derivatives)
- businesses (e.g., corporate financials, commercial transactions, and credit card purchases)
- governments (e.g., trade, economic, employment, and payroll data)
- individuals (e.g., credit card purchases, product reviews, internet search logs, and social media posts)
- sensors (e.g., satellite imagery, shipping cargo information, and traffic patterns), and, in particular
- the Internet of Things, or IoT (e.g., data generated by “smart” buildings, where the building is providing a steady stream of information about climate control, energy consumption, security, and other operational details)

# los b. describe Big Data, artificial intelligence, and machine learning;

- There are three main sources of alternative data:
- data generated by individuals
- data generated by business processes
- data generated by sensors
- Data generated by individuals are often produced in text, video, photo, and audio formats and may also be generated through such means as website clicks or time spent on a webpage. This type of data tends to be unstructured
- Sensor data are collected from such devices as smart phones, cameras, RFID chips, and satellites that are usually connected to computers via wireless networks. Sensor data can be unstructured, and the volume of data is many orders of magnitude greater than that of individual or business process datastreams
- Internet of Things: A network arrangement of structures and devices whereby the objects on the network are able to interact and share information

# Los b:Big Data Challenges

- Big Data poses several challenges when it is used in investment analysis, including the quality, volume, and appropriateness of the data
- The data must be sourced, cleansed, and organized before analysis can occur
- This process can be extremely difficult with alternative data owing to the unstructured characteristics of the data involved, which are more often qualitative (e.g., texts, photos, and videos) than quantitative in nature
- Given the size and complexity of alternative datasets, traditional analytical methods cannot always be used to interpret and evaluate these datasets



# los b: artificial intelligence and machine learning

- Artificial intelligence computer systems are capable of performing tasks that have traditionally required human intelligence
- AI technology has enabled the development of computer systems that exhibit cognitive and decision-making ability comparable or superior to that of human beings.
- Financial institutions have made use of AI particularly, neural networks, programming based on how our brain learns and processes information to detect abnormal charges or claims in credit card fraud detection system
- Machine learning (ML) is a technology that has grown out of the wider AI field. ML algorithms are computer programs that are able to “learn” how to complete tasks, improving their performance over time with experience
- ML involves splitting the dataset into a training dataset and validation dataset (evaluation dataset). The training dataset allows the algorithm to identify relationships between inputs and outputs based on historical patterns in the data
- Analysts must also be cognizant of errors that may arise from overfitting the data, because models that overfit the data may discover “false” relationships or “unsubstantiated” patterns that will lead to prediction errors and incorrect output forecasts. Overfitting occurs when the ML model learns the input and target dataset too precisely

# los b: Types of Machine Learning

- In supervised learning, computers learn to model relationships based on labeled training data. In supervised learning, inputs and outputs are labeled, or identified, for the algorithm. After learning how best to model relationships for the labeled data, the trained algorithms are used to model or predict outcomes for new datasets
- In unsupervised learning, computers are not given labeled data but instead are given only data from which the algorithm seeks to describe the data and their structure.
- In deep learning, (or deep learning nets), computers use neural networks, often with many hidden layers, to perform multistage, non-linear data processing to identify patterns. Deep learning may use supervised or unsupervised machine learning approaches

# los b: extracting information from big data

- Data science: An interdisciplinary field that brings computer science, statistics, and other disciplines together to analyze and produce insights from Big Data
- Data Processing Methods
- Capture—Data capture refers to how the data are collected and transformed into a format that can be used by the analytical process
- Curation—Data curation refers to the process of ensuring data quality and accuracy through a data cleaning exercise. This process consists of reviewing all data to detect and uncover data errors bad or inaccurate data and making adjustments for missing data when appropriate
- Storage—Data storage refers to how the data will be recorded, archived, and accessed and the underlying database design.
- Search—Search refers to how to query data. Big Data has created the need for advanced applications capable of examining and reviewing large quantities of data to locate requested data content
- Transfer—Transfer refers to how the data will move from the underlying data source or storage location to the underlying analytical tool.

# los b: Data Visualization

- Data visualization is an important tool for understanding Big Data. Visualization refers to how the data will be formatted, displayed, and summarized in graphical form
- Traditional structured data can be visualized using tables, charts, and trends, whereas non-traditional unstructured data require new techniques of data visualization.
- These visualization tools include, for example, interactive three-dimensional (3D) graphics
- Another valuable Big Data visualization technique that is applicable to textual data is a “tag cloud,” where words are sized and displayed on the basis of the frequency of the word in the data file
- A “mind map” is another data visualization technique; it is a variation of the tag cloud, but rather than displaying the frequency of words, a mind map shows how different concepts are related to each other

# los b:Programming Languages and Databases

- Python: Python is an open source, free programming language that does not require an in-depth understanding of computer programming
- R: R is an open source, free programming language traditionally used for statistical analysis
- Java: Java is a programming language that can run on different computers, servers, and operating systems. Java is the underlying program language used in many internet applications
- C/C++: C/C++ is a specialized programming language that provides the ability to optimize source code to achieve superior calculation speed and processing performance
- Excel VBA: Excel VBA helps bridge the gap between programming and manual data processing by allowing users to run macros to automate tasks, such as updating data tables and formulas, running data queries and collecting data from different web locations, and performing calculations

# Los c: Text Analytics and Natural Language Processing

- Text analytics: The use of computer programs to analyze and derive meaning from typically large, unstructured text- or voice-based datasets
- More analytical usage includes lexical analysis, or the analysis of word frequency in a document and pattern recognition based on key words and phrases. Text analytics may be used in predictive analysis to help identify indicators of future performance, such as consumer sentiment
- Natural language processing: Computer programs developed to analyze and interpret human language
- Automated tasks using NLP include translation, speech recognition, text mining, sentiment analysis, and topic analysis
- For example, NLP may be used to monitor analyst commentary to aid investment decision making. Financial analysts may generate earnings-per-share (EPS) forecasts reflecting their views on a company's near-term prospects. Focusing on forecasted EPS numbers could mean investors miss subtleties contained in an analyst's written research report

# Los c: Robo-Advisory Services

- In the United States, robo-advisers must be established as registered investment advisers, and they are regulated by the Securities and Exchange Commission
- Robo-advice tends to start with an investor questionnaire, which may include many of the categories and subcategories
- Once assets, liabilities, risk preferences, and target investment returns have been digitally entered by a client, the robo-adviser software produces recommendations, based on algorithmic rules and historical market data, that incorporate the client's stated investment parameters
- Two types of wealth management services dominate the robo-advice sector: fully automated digital wealth managers and adviser-assisted digital wealth managers.
- Fully Automated Digital Wealth Managers: The fully automated model does not rely on assistance from a human financial adviser. These services seek to offer a low-cost solution to investing and recommend an investment portfolio, which is often composed of ETFs.
- Adviser-Assisted Digital Wealth Managers: Adviser-assisted digital wealth managers provide automated investment services along with a virtual financial adviser, who is available to offer basic financial planning advice and periodic reviews by phone

# Los c: Risk Analysis

- The global investment industry has undertaken major steps in stress testing and risk assessment that involve the analysis of vast amounts of quantitative and qualitative risk data
- Required data include information on the liquidity of the firm and its trading partners, balance sheet positions, credit exposures, risk-weighted assets, and risk parameters
- There is increasing interest in monitoring risk in real time. To do so, relevant data must be taken by a firm, mapped to known risks, and identified as it moves within the firm
- ML techniques may be used to help assess data quality. To help ensure accurate and reliable data that may originate from numerous alternative data sources
- Portfolio risk management often makes use of scenario analysis—analyzing the likely performance of the portfolio and liquidation costs under a hypothetical stress scenario or the repeat of a historical stress event



# Los c: Algorithmic Trading

- Algorithmic trading is the computerized buying and selling of financial instruments, in accordance with pre-specified rules and guidelines
- Algorithmic trading is often used to execute large institutional orders, slicing orders into smaller pieces and executing across different exchanges and trading venues
- Algorithmic trading provides investors with many benefits, including speed of execution, anonymity, and lower transaction costs
- High-frequency trading (HFT) is a form of algorithmic trading that makes use of vast quantities of granular financial data (tick data, for example) to automatically place trades when certain conditions are met
- HFT algorithms decide what to buy or sell and where to execute on the basis of real-time prices and market conditions, seeking to earn a profit from intraday market mispricings

# los d: distributed ledger technology

- Distributed ledger technology based on a distributed ledger represents a fintech development that offers potential improvements in the area of financial record keeping
- DLT networks are being considered as an efficient means to create, exchange, and track ownership of financial assets on a peer-to-peer basis
- The processes underlying DLT generally require massive amounts of energy to verify transaction activity
- A distributed ledger is a type of database that may be shared among entities in a network. In a distributed ledger, entries are recorded, stored, and distributed across a network of participants so that each participant has a matching copy of the digital database
- Features of DLT include the use of cryptography, an algorithmic process to encrypt data, making the data unusable if received by unauthorized parties, which enables a high level of network security and database integrity
- DLT has the potential to accommodate “smart contracts,” which are computer programs that self-execute on the basis of pre-specified terms and conditions agreed to by the parties to a contract

# los d: distributed ledger technology

- Blockchain is a type of digital ledger in which information, such as changes in ownership, is recorded sequentially within blocks that are then linked or “chained” together and secured using cryptographic methods.
- Each block contains a grouping of transactions (or entries) and a secure link (known as a hash) to the previous block
- New transactions are inserted into the chain only after validation via a consensus mechanism in which authorized members agree on the transaction and the preceding order, or history, in which previous transactions have occurred
- The consensus mechanism used to verify a transaction includes a cryptographic problem that must be solved by some computers on the network (known as miners) each time a transaction takes place

# los d: Permissioned and Permissionless Networks

- DLT can take the form of permissionless or permissioned networks
- Permissionless networks: are open to any user who wishes to make a transaction, and all users within the network can see all transactions that exist on the blockchain
- The main benefit of a permissionless network is that it does not depend on a centralized authority to confirm or deny the validity of transactions, because this takes place through the consensus mechanism
- Once a transaction has been added to the blockchain, it cannot be changed, barring manipulation; the distributed ledger becomes a permanent and immutable record of all previous transactions
- A well-known example of an open, permissionless network is bitcoin. Using blockchain technology, Bitcoin was created in 2009 to serve as the public ledger for all transactions occurring on its virtual currency. Since the introduction of bitcoin, many more Cryptocurrencies, or digital currencies, which use permissionless DLT networks, have been created
- In permissioned networks, network members may be restricted from participating in certain network activities. Controls, or permissions, may be used to allow varying levels of access to the ledger, from adding transactions (e.g., a participant) to viewing transactions only (e.g., a regulator) to viewing selective details of the transactions but not the full record

# los d: Cryptocurrencies

- A cryptocurrency, also known as a digital currency, operates as electronic currency and allows near-real-time transactions between parties without the need for an intermediary, such as a bank
- As electronic mediums of exchange, cryptocurrencies lack physical form and are issued privately by individuals, companies, and other organizations
- Most issued cryptocurrencies utilize open DLT systems in which a decentralized distributed ledger is used to record and verify all digital currency transactions
- Cryptocurrencies have not traditionally been government backed or regulated
- Cryptocurrencies have proven to be an attractive means for companies looking to raise capital. An initial coin offering (ICO) is an unregulated process whereby companies sell their crypto tokens to investors in exchange for fiat money or for another agreed upon cryptocurrency
- ICOs may have lower associated issuance costs and shorter capital raising time frames. However, most ICOs do not typically have attached voting rights

# los d: Applications of Distributed Ledger Technology to Investment

- Tokenization: the process of representing ownership rights to physical assets on a blockchain or distributed ledger
- DLT has the potential to streamline this process by creating a single, digital record of ownership with which to verify ownership title and authenticity, including all historical activity.
- Real estate transactions that require ownership and identify verification may be one area to benefit from tokenization
  
- Post-Trade Clearing and Settlement
- Post-trade processes to confirm, clear, and settle transactions are often complex and labor intensive, requiring multiple interactions between counterparties and financial intermediaries
- DLT has the ability to streamline existing post-trade processes by providing near-real-time trade verification, reconciliation, and settlement, thereby reducing the complexity, time, and costs associated with processing transactions

# los d: Applications of Distributed Ledger Technology to Investment

- Compliance: Regulators worldwide have imposed more stringent reporting requirements and demand greater transparency and access to data
- To meet these requirements, many firms have added staff to their post-trade and compliance groups. But these functions remain predominantly manual. To comply with regulations, firms need to maintain and process large amounts of risk-related data
- DLT may allow regulators and firms to maintain near-real-time review over transactions and other compliance-related processes
- Improved post-trade reconciliation and automation through DLT could lead to more accurate record keeping and create operational efficiencies for a firm's compliance and regulatory reporting processes, while providing greater transparency and auditability for external authorities and regulators

# CFA Curriculum Questions

1 .A correct description of fintech is that it:

- a) is driven by rapid growth in data and related technological advances
- b) increases the need for intermediaries
- c) is at its most advanced state using systems that follow specified rules and instructions

2.A characteristic of Big Data is that:

- a) one of its traditional sources is business processes.
- b) it involves formats with diverse types of structures.
- c) real-time communication of it is uncommon due to vast content



# CFA Curriculum Questions

3. In the use of machine learning (ML):

- a) some techniques are termed “black box” due to data biases.
- b) human judgment is not needed because algorithms continuously learn from data.
- c) training data can be learned too precisely, resulting in inaccurate predictions when used with different datasets.

4. Text Analytics is appropriate for application to:

- a) economic trend analysis.
- b) large, structured datasets.
- c) public but not private information.

# CFA Curriculum Questions

5. In providing investment services, robo-advisers are most likely to:

- a) rely on their cost effectiveness to pursue active strategies.
- b) offer fairly conservative advice as easily accessible guidance.
- c) be free from regulation when acting as fully-automated wealth managers.

6. Which of the following statements on fintech's use of data as part of risk analysis is correct?

- a) Stress testing requires precise inputs and excludes qualitative data.
- b) Machine learning ensures that traditional and alternative data are fully segregated.
- c) For real-time risk monitoring, data may be aggregated for reporting and used as model inputs.

# Solution

- A is correct. Drivers of fintech include extremely rapid growth in data (including their quantity, types, sources, and quality) and technological advances enabling the capture and extraction of information from it.
- B is correct. Big Data is collected from many different sources and is in a variety of formats, including structured data (e.g., SQL tables or CSV files), semi-structured data (e.g., HTML code), and unstructured data (e.g., video messages).
- C is correct. Overfitting occurs when the ML model learns the input and target dataset too precisely. In this case, the model has been “over trained” on the data and is treating noise in the data as true parameters. An ML model that has been overfitted is not able to accurately predict outcomes using a different dataset and may be too complex.
- A is correct. Through the Text Analytics application of natural language processing (NLP), models using NLP analysis may incorporate non-traditional information to evaluate what people are saying—via their preferences, opinions, likes, or dislikes—in the attempt to identify trends and short-term indicators about a company, a stock, or an economic event that might have a bearing on future performance.

# Solution

- 5. B is correct. Research suggests that robo-advisers tend to offer fairly conservative advice, providing a cost-effective and easily accessible form of financial guidance to underserved populations, such as the mass affluent and mass market segments.
- 6. C is correct. There is increasing interest in monitoring risk in real-time. To do so, relevant data must be taken by a firm, mapped to known risks, and identified while moving within the firm. Data may be aggregated for reporting purposes or used as inputs to risk models

# Summary

- The term “fintech” refers to technological innovation in the design and delivery of financial services and products.
- Areas of fintech development include the analysis of large datasets, analytical techniques, automated trading, automated advice, and financial record keeping.
- Big Data is characterized by the three Vs—volume, velocity, and variety—and includes both traditional and non-traditional (or alternative) datasets.
- Among the main sources of alternative data are data generated by individuals, business processes, and sensors.
- Artificial intelligence computer systems are capable of performing tasks that traditionally required human intelligence at levels comparable (or superior) to those of human beings.
- Machine learning seeks to extract knowledge from large amounts of data by “learning” from known examples and then generating structure or predictions. Simply put, ML algorithms aim to “find the pattern, apply the pattern.” Main types of ML include supervised learning, unsupervised learning, and deep learning.

# Summary

- Natural language processing is an application of text analytics that uses insight into the structure of human language to analyze and interpret text- and voice-based data.
- Robo-advisory services are providing automated advisory services to increasing numbers of retail investors. Services include asset allocation, portfolio optimization, trade execution, rebalancing, and tax strategies.
- Big Data and ML techniques may provide insights into real-time and changing market circumstances to help identify weakening or adverse trends in advance, allowing for improved risk management and investment decision making.
- Algorithmic traders use automated trading programs to determine when, where, and how to trade an order on the basis of pre-specified rules and market conditions. Benefits include speed of executions, lower trading costs, and anonymity.
- Blockchain and distributed ledger technology (DLT) may offer a new way to store, record, and track financial assets on a secure, distributed basis. Applications include cryptocurrencies and tokenization. Additionally, DLT may bring efficiencies to post-trade and compliance processes through automation, smart contracts, and identity verification.