THE CAPCO INSTITUTE
JOURNAL
OF FINANCIAL TRANSFORMATION

TRANSFORMATION

Financial and data intelligence

CHARLES S. TAPIERO

DESIGN THINKING

#48 NOVEMBER 2018

THE CAPCO INSTITUTE

JOURNAL OF FINANCIAL TRANSFORMATION

RECIPIENT OF THE APEX AWARD FOR PUBLICATION EXCELLENCE

CONTENTS

DESIGN

TRANSFORMATION

SUPERVISION

DEAR READER,

Design thinking, a collaborative, human-focused approach to problem-solving, is no longer just for the creative industries. It has become an important management trend across many industries and has been embraced by many organizations. Its results are hard to ignore. Indeed, design-driven companies regularly outperform the S&P 500 by over 200 percent.[1]

To date, the financial services industry has not led in adopting this approach. However, leaders are recognizing that important challenges, such as engaging with millennial customers, can be best addressed by using design thinking, through the methodology's exploratory approach, human focus, and bias towards action. This edition of the Journal examines the value of design thinking in financial services.

Design thinking introduces a fundamental cultural shift that places people at the heart of problem-solving, which is critical in a technology-driven environment. If the customer's real problems are not fully understood, technological solutions may fail to deliver the desired impact. In this context, design thinking offers a faster and more effective approach to innovation and strategic transformation.

The case studies and success stores in this edition showcase the true value of design thinking in the real world, and how this approach is an essential competitive tool for firms looking to outperform their peers in an increasingly innovation-driven and customer-centric future. At Mastercard, design thinking has become a part of almost all organizational initiatives, from product development, research and employee engagement to solving challenges with customers and partners. Meanwhile, at DBS Bank in Singapore, a data-informed design model has been firmly embedded into the bank's culture, enabling them to successfully move from being ranked last among peers for customer service in 2009, to being named the Best Bank in the World by Global Finance in 2018.

I hope that you enjoy the quality of the expertise and points of view on offer in this edition, and I wish you every success for the remainder of the year.

Lance Levy, **Capco CEO**

[1] http://fortune.com/2017/08/31/the-design-value-index-shows-what-design-thinking-is-worth/

FINANCIAL AND DATA INTELLIGENCE

CHARLES S. TAPIERO | Topfer Chair Distinguished Professor, Department of Finance and Risk Engineering, New York University, Tandon School of Engineering

ABSTRACT

Data and computers on steroids have partnered to transform finance and reengineer its future. Past conventions have defined the role of data to be a complement to financial theories, providing a testing ground and an estimator of future prices, whether of assets, stocks, or derivatives. Theories of finance (such as the Arrow-Debreu framework, Martingale pricing, risk neutral pricing, etc.), while mathematically and theoretically stimulating, also embed a variety of risks and real financial misconceptions. For example, risk is defined by predictable (future states) events while in real finance, uncertainty primes [Knight (1924)]; prices exist only in the present and so on. Further, while conventional finance is an ex-ante approach to the future, data science is an inverse approach that seeks ex-post to estimate causes or models that explain the data so collected and improve their state of knowledge and know-how by learning through a feedback process. This approach is often structured by terms such as "deep learning," "machine learning," and "artificial intelligence." Thus, one approach is defined by hypothetical theories, while the other is an analytic data and inverse approach that implies hypothetical models (not necessarily one) for the purpose of learning and/or deciding. The purpose of this paper is to elaborate on the fundamental elements that are contributing to the transformation of finance and raise its risk consequences.

1. INTRODUCTION

The growth of financial complexity, technology, computing capacities, and services combined with an access to "big" and varied data are currently transforming real finance and challenging its conventional models. Data science has for many generations challenged the practical implementation of theoretical models, statistical learning, numerical techniques, and their approximations. IT software and increasingly powerful computers have contributed to its new computing capacity and their ability to resolve conceived problems hitherto unapproachable. Concurrently, they have upended the search and usefulness of data algorithms based on statistical and computing facilities of various sorts. These are currently providing a range of opportunities to reduce costs and increase profits for financial institutions through the discovered potential of on-line systemic learning, trends discoveries, and their applications to become self-operated, often alluded to as AI (Artificial Intelligence) [Billard and Diday (2003), Callebaut (2012), Chambers (1993), Cleveland (2001), Donoho (2015), Hey et al. (2009), Kirkpatrick and Kurths (2012), National Research Council (2010)].

Developments, improvements, and the control of financial complexity are essential. For example, Ashby's [Ashby (1956)] Principle of Required Variety (The second Law of Cybernetics) already implied long ago that complexity untamed by mathematical intelligence and controls is self-destructive. In this vein, pursuing a financial evolution devoid of "intelligence" will necessarily lead to a financial and technological breakdown.

These approaches have been developed and used ever since the first Industrial Revolution, ushering in industrial automation, cybernetics, robotics, etc. Similarly, mathematical algorithms have developed a multitude of algorithms to search, track, solve, and implement these same problems. Numerical optimization techniques, such as linear programming, stochastic modeling and optimization, and so on, have been devised to revolutionize our capacity to solve practical and complex economic, financial, industrial, service, design, and management problems. Current developments are adapted to a far broader set of applications such as self-driving cars, health care, finance, services, etc.

> "A machine learning algorithm walked into a bar. The bartender asked, "What would you like to drink?" The algorithm replied, "What's everyone else having?""
>
> — (SEEN ON TWITTER)

Models are tested and learning-improved by data, and inversely defined by data upended by learning processes based on both the information that data produces and statistical concepts that estimate and improve the belief in our models. This approach is trumpeted now as a means to confront and manage a complexity hitherto ignored. Terms, such as machine learning, deep learning, robo-advisors, artificial intelligence, and a multitude of algorithms are applied to specific problems deemed far too complex to be defined by just theoretical models. For some, it may seem that models defined ex-ante are "irrelevant" while "ex-post data" (since they are always ascertained "after the fact" or by simulations) is painted as a greater truth seizing the rationality that models portend to present. Namely, measurements that define events or conditions that express only "what is" rather than what we seek in the future to defines: "what to do." Data intelligence in such cases define both "what is" and "what to do" and the potential models (and thus rationalities) that underlie the data we dispose of. The complexity of "what is" and "what is to be" is far too great for one (models and statistics) not to be integrated without the other (the data approach). Models expressing strategic intents are then a "GPS," altered as new data is mined, analyzed, and applied to improve the "GPS" and edit the policies it implies. In this paper, we seek to appreciate ex-ante "modeling" and ex-post data management [Breiman (2001), Diday and Esposito (2003), Goodman and Wong (2009), Guetzkow (1959, 1962), Horton et al. (2015), Krohs and Callebaut (2007), Nyamabuu and Tapiero (2017), Tapiero (2013), Tapiero et al. (1975), Tukey (1962)].

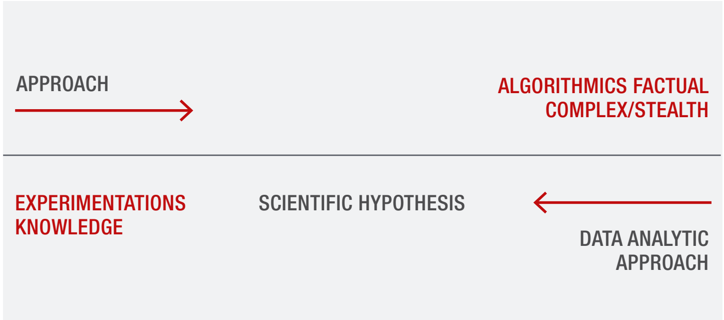
2. FINANCE AND DATA

The globalization of finance and financial technologies, combined with the complexities of financial systems and products led to a finance racing to transform its services, practices, trades, prices, and financial management, to be both data and computer operated and managed. For example, doing away with neighborhood banks and bank tellers, and replacing them with digitalized financial service and so on. It alters the role of traditional banks, financial markets, trade, insurance, etc. As information and technology become more accessible, new competitors are able to perform at relative and competing speeds and costs, increasing the efficiency of loan underwriting and credit scoring for individuals and SMEs. As a result, tasks that were predominantly performed by traditional banks are now computer aided with banks merely providing "financial" infrastructures for services such as trade and robotic advising and investments based on learning and inferential processes that might be designed by clouds of data and a software (algorithmic) intelligence [Albert and Barabási (2002), Overbye (2012), Nyambuu and Tapiero (2017)].

At the same time, banks have increased their push for a "cashless" future, setting the ground for a digitalized, technological, global, monopsonic, and competitive finance (although currently, faced with doubtful crypto-currencies challenges to financial regulation and money). Information Technologies (IT) and algorithms designed to meet the increased demands and the complexity of finance will, necessarily, face a global spiraling complexity challenging both strategic financial systems as well as their regulations. To confront this complexity, greater intelligence is required for both financial models and data. In such an environment, both creative and theoretical financial constructs, data mining and data analytics, statistical treatments that extract information, trends, strategies, decisions, and "models" may provide a finance architecture and the means to remain competitive, profitable, self-managed, and able to adapt to a future technological finance.

Practically, it means that finance is challenged by the intelligence that data provide and require. However, data without models seek a meaning to what they reveal by an inverse rationality. Namely, let data imply a model. Such an approach, necessarily, may provide not one, but many models. Explicitly, given the statistical character of financial data, it implies not one but many "models," maintaining the empirical presumption that all outcomes remain doubtful. For example, financial data prices are, by

Figure 1: Statistical versus data analytic approaches



definition, a present price, reflecting future potential prices and multiple factors such as yields, sentiments, news, macroeconomic trends, politics, the flow of domestic and foreign capital, etc. All of which affect a financial random future and expectations that define prices.

Options are such a case, with known parameters and prices implying a model of future volatility. Similarly, granular (fractional) financial data (reporting prices every day, every hour, second and microseconds) imply a financial randomness and information that data granularity provides [Tapiero and Vallois (2015, 2016, 2017, 2018a, b), Tapiero (2017)]. In such cases, data granularity defines the statistical properties it produces [Tapiero and Vallois (2015, 2016, 2017, 2018a, b), The Economist (2010), Vallois and Tapiero (2007, 2009)].

A complementary approach with one expanding the other [Allen (2001), Breiman (2001), Cleveland (1985, 1993), Donoho (2015), Groenen et al. (2006), Guttman (1944), Hanani and Tapiero (1980), Raveh and Tapiero (1980a, b), Tukey (1962, 1977, 1994)] leads to push-pull challenges that merge traditional ex-ante quants' complete markets models and a real and complex finance. These are, as stated above, defined by multiple factors such as macroeconomics, politics, globalization, complexity, emerging strategic (multi agent) finance, financial gating, and regulations (Nyambuu and Tapiero (2018), Raveh and Tapiero (1980b), Tapiero et al. (1975)].

3. DATA AND STATISTICAL LEARNING

Hardware and software have made it much easier to access mined data and transform it into information. Data access, storage, speed, and the growth of increasingly complex and integrated computer systems have in their wake opened new possibilities to "learn" using inferential software and generate future scenarios. Bayesian models, Copulas, long run (autocorrelation) memory, persistence (short memory), learning wavelets, Bayesian networks, neuro networks etc., provide predictive software, all of which are based on mathematical and statistical models. A comparison of data analytic approaches in Figure 1 emphasizes systemic causal approaches versus algorithmic-data analytic approaches.

These developments are "first generation" financial software combining computational and mathematical techniques with newly found computing and data management capabilities. Current intelligence software are fast mutating into future new generations that are hard to predict, however. For example, based on the presumption that data is never complete, information and the predictive powers it provides are also incomplete. The ancient Greeks, aware of the vagaries of time, claimed that "the likely is unlikely." These beliefs do not negate the fact that predictive models may be used. However, to mitigate their predictive uncertainty, a statistical rationale ought to be applied to qualify the quality of these predictions.

For example, inferences may provide a better appreciation of what data may teach us (revealed by errors and unconfirmed expectations we discover), as well as expand and qualify the breadth of our choices. Statistical learning can make predicting the future more efficient to the extent that the "machine" intelligence never outpaces financial human intelligence – an intelligence needed to be greater than the complexity it purports to confront. Lacking such an intelligence results in chaos [see also Ashby's Cybernetics (1956)].

Statistical models, by contrast to (ex-post) data as the sole mean to assess trends or inferences, are based on what we know in order to mitigate what we don't know applied to some specific and categorized purposes. Confirmed hypotheses are then used to predict future states, calculate and manage risks and their consequences, define "optimum" decisions and policies, and study their robustness. These in turn, produce a feedback mode that revises their hypothetical models.

Breiman (2001) pointed to two cultures. A learning culture that points to a model's (and statistical) uniqueness assessed by a model's statistical fit to improve and update, what we know, and mutate the model (hypotheses or estimated trends). By contrast, the data culture is an inverse approach that provides multiple potential "hypothetical models" that explain the data. Artificial intelligence is then a means to differentiate between these models, either explicitly or implicitly, to define a "common sense" and differentiate it from "nonsense" (based on a "cloud" of experiences and tested inferences).

Further, Breiman states that the roots of statistics, as in science, lie in working and checking theory against data. "I hope that in this (past) century our field will return to its roots (and that in the current century it will integrate data science and statistics)." There are today noticeable moves toward "real world data problems" and their integration with computer science, inverse theories, and to a greater awareness of consequences, social and otherwise.

Data analysis is traditionally associated with a statistical rationality based on learning, adaptive estimates, a long empirical and shared experience, and common theoretical tested principles [for example, see Andersen et al. (2009), Callebaut (2012), Goodman and Wong (2009), Hey et al. (2009), Krohs and Callebaut (2007), McKinsey (2011), Tukey (1962)]. For example, given sampled time series, models of stock prices are used to replicate, track, and infer future prices statistically qualified. Such approaches presume that data is incomplete, and therefore require mathematical and statistical principles to guarantee the quality of estimates and predictions. When data is "big" and presumed complete, it is fed by data analyses and algorithms. For example, Bayesian analysis provides a slow learning process based on pre-posterior estimates. In a changing and unstable environment, financial prices may define time series that are auto-correlated, with mutating trends, i.e., altering their statistical properties as events and time alter the model (providing, therefore, a learning process with very short time spans). Learning from data is, therefore, multi-dimensional, based on trends and the many facets and characteristics data implies. For example, rather than consider a time series trend, its mean and variance, additional factors such as the data samples' granularity, samples range, inverse statistics (of the time series surrogate processes) etc., provide additional dimensions along which data reveals its properties.

Similarly, and practically, data-memory is essentially an abstract filter that produces an image of the past based on data mined, transformed, and statistically treated (thereby, transforming the models that data implies). For example, option prices are implied by a specific "time and future limited horizon." Long run memories and their auto-covariance, co-location, and quantum entanglement altered events (producing a short term memory, with a stochastic mutation due to occurring events) are challenges that require an underlying model as well as a greater assessment of what data does reveal truly.

Breiman predicted these problems by pointing to three elements:

- **The Rashomon effect** (the Japanese film, same data, different perceptions): where a multiplicity of good models may result from the same data.
- **Occam razor:** relates to the conflict between simplicity and accuracy (for example: econometric models, and financial models in general, are not necessarily more accurate the greater their complexity).
- **Bellman:** the curse of dimensionality (and therefore, big data might not be more informative than "small" but "intelligent" data)

For example, Tukey (1962, 1977, 1994) predicted years ago that the future would emphasize the primacy of data [see also Donoho (2015)] and the need to learn from data analytics rather than just "fundamental statistical models." Tukey (1994) states that "For a long time I have thought I was a statistician, interested in inferences from the particular to the general. But as I have watched mathematical statistics evolve, I have had cause to wonder and to doubt. ...I have come to feel that my central interest is in data analysis which I take to include, among other things: procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning and gathering of data to make its analysis easier, more precise or more accurate, and all the machinery and results of (mathematical) statistics which apply to analyzing data."

Statistical and mathematical psychologists have been concerned with similar problems. Guttman (1944), for example, suggested a scaling approach to very large and multidimensional psychological (and educational) tests that led to the development of the "Guttman scale" ("a linearized multidimensional data). Sociology, psychology, and health data are often studied using very large quantities of qualitative data that include a large number of interactions, behavioral patterns, and variables. Each variable is defined in addition by its attributes. Scale can then be used to compare one student knowledge to another, a mental state compared to another by using Guttman's scaling methodology based on quantitative and qualitative data. Further development at the Bureau of Social Research in Jerusalem improved this approach by a dimensional reduction of multivariate datasets [see also Raveh and Tapiero (1980a, b)] and provided apparently more information than standard correlation and studies using factor analysis. Data analysis may then be easily assimilated in the form of "scalogram" providing a visual configuration of qualitative data. Louis Guttman's approach was used successfully in investigating military morale and other problems (the U.S. Army Research Branch Morale Services Division) during WWII. Subsequently, it led to the development of numerous applications by the Bureau of Social Research (Jerusalem, Israel) on voting patterns and a broad range of questionnaires accumulated into large datasets.

Drawing on work of data scientists, a vision of data engineered to be better displayed rather than modeled have contributed also to better "learning from data." In other words, they have produced a friendlier data, visually accessible, providing an easier appreciation of what data means. For example, defining what a data user wants along any number of criteria and articulating a personalized vision of data. "Data science" was defined in terms of six divisions [Donoho (2015)]:

- Data exploration and preparation
- Data representation and transformation
- Computing with data
- Data modeling
- Data visualization and presentation
- Science about data science

These elements are parts and parcel of financial data science, applied to financial time series and other data types commonly used by financial agents. The first and second stages are crucial, especially when considering data preparation and transformation, as the success of some algorithms (such as deep learning and random forest algorithms that require some standardization of the data).

4. RISK MODELS AND DATA FINANCE

Financial theories derived from economic and risk models are hypotheses. Sometimes they are right, sometimes they are incomplete, always in doubt, and never confirmed. Financial risk models are merely partial models of uncertainty that predict future prices. Thus, risk management is based on predicting and accounting for the 'predictable,' rather than managing unpredictable and consequential risks, such as booms and busts, systemic risks, contagious behaviors, and so on, that recur mostly unpredictably. Risks, furthermore are consequences generated by multiple factors, from many sources, some of which are statistically and causally dependent. The following is a summary of some of its elements underlying data finance.

- Increased complexity and uncertainty and a belief in the certainty of data.
- Default models due to incomplete financial models.
- An increasingly strategic finance with dominant agents and a world at risk beset by what we do, by what others do, or we do to each other. In this world, a general equilibrium may no longer be possible, sustainable, or efficient. Risk finance is thus increasingly strategic.
- Financial greed, with TBTF (too big to fail) enterprises, and information and power asymmetries and increasingly aggressive and strategic regulations lead to the tenets of free markets to falter.
- Competing regulations in a global world contribute to increasingly complex financial logistics and to competing financial systems.
- Financial systems increasingly subdued to political and macroeconomic events are also faced with a far more complex risk finance, where risks are derivatives of non-finance risks.

These are partial processes changing tomorrow's world of finance. They may result in extreme behaviors fueled by excessive unfiltered information, far more apt at generating contagious behaviors and, therefore, 'financial runs.' Security risks, networks, and IT are also important candidates that redefine future financial risks. In this environment, conventional financial risk models are no longer relevant.

Though it must be said that financial risk products, such as insurance and lending, and risk markets, such as VIX and Carbone markets, are merely mechanisms for speculating and risk sharing (and an important part of financial activity). For example, credit risks and insurance coverage may be co-dependent, derived from a large number of transparent causes due to networked and IT systems and the availability of "big data" used to better assess borrowers' collaterals, risk history, wealth, etc. For regulators, big data provide a greater transparency of potential non-compliance to an increasingly complex regulatory table.

The promises and the risks of big data in finance are in their infancy. Some may allow the prospective integration of financial models with the many data clouds storing investors and personalized information; a power it can sell and provide to financial institutions; the risks to individual liberties and security clients assume, and so on. Data, thus, fuels a plethora of data analytic techniques to increase profits. An increasing number of software companies and start-ups are proposing 'black boxes' to interpret consumers' sentiments and intents using internet comments on stocks, financial assets, and variables deemed pertinent to the financial environment. Algorithms and learning machines are then created to seek and interpret images to detect a 'flow of sentiments' that are claimed to be related to (and thereby be early predictors of) stock markets performance, and predict consumers' choices and their implied preferences. A rising tide of data driven algorithms is thus emerging and engulfing finance and business to become information and technologically dependent (and therefore, a growing source of risk).

These processes contribute to an extraordinary growth of information asymmetry risks and the misuse of information and insiders' trading risks. For example, say that a company hires a data scientist to determine the public's attitude towards that company and its CEO. It would be like paying a psychiatrist to hear what one wants to hear (since data analysis need not provide one set of conclusive observations). In big data, the chances of finding what one wants to hear and what may be a real fact are equally high. Searching for meaning in large datasets, without theories, may be like seeking the North Pole without a compass. For these and other reasons, big data based on the accumulation of private information is a growing source of risk that contributes to important security problems.

The traditional statistical approach, unlike data intensive treatments, is based on fundamental hypotheses to be refuted or not [Diday and Esposito (2003), Fisher (1936), Albert and Barabási, (2002), Billard and Diday (2003)]. Thus, the 'statistical/scientific' approach reveals 'uncertainty' and its risks from a given and tested knowledge base. Is an evolving process based on a cycle to hypothesize, measure, test, and confirm-or-not? The data driven approach, instead, is a statement of current facts, and a presumed certainty rather than recognizing that all knowledge is partial – embedded in a greater uncertainty that statistics qualifies.

For banks, traders, and suppliers of financial information and advice, data and information are becoming primary assets. The Economist reported that between 1990 and 2005, more than 1 billion people worldwide entered the middle class, and by 2013 the amount of data transferred over the internet will reach 667 exabytes annually. According to Cisco the quantity of data continues to grow faster than the ability of the network to carry. Companies like Amazon's Web Services, AT&T's Synaptic Hosting, AppNexus, GoGrid, Rackspace Cloud Hosting, the HP/Yahoo/Intel Cloud Computing Test bed, the IBM/Google, and Micro Strategy BI Cloud, have provided various types of cloud services to ease these data storage problems (while at the same time setting data at risk). Currently, a variety of corporate clouds and data services have been commercialized, providing an increased access to data to an expanding (and competing) population of data managers and scientists for any purpose, including risk purposes.

The ability of data technologies (with social media having shown a way to handle and analyze vast amounts of unstructured data), to test our abilities to translate complex, diverse, and dynamic data sources into workable financial information remains unproven. At the same time, an expanding digitalized financial system is allowing context-specific analyses.

Information and/or knowledge extracted from digital records render financial banks' jobs easier, by diagnosing and detecting risks accurately, and assessing their clients' propensity to assume risks (and hence improve the overall profitability and "quality" of their services). Similarly, digitalized data may prevent cybercrimes more effectively and thus contribute to the increasingly complex systems of financial networks, e-financial markets, and an increasing financial retailing dependence. Despite the potential for big data and financial data analytics, it may hide risk consequences that have not been revealed. In terms of security, Michael de Crespigny, CEO at ISF, stated a few years ago that: "Only half of organizations surveyed by the ISF are using some form of analytics for fraud prevention, forensics and network traffic analysis, while less than 20% are using it to identify information related to subject matter requests, predict hardware failures, ensure data integrity or check data classification."

Few organizations recognized the benefits of information security, yet many were already using data analytics to support their core business. Currently, security is a prime concern, pointing to a "growing tree" of functions and technologies that renders "security" a dynamic "big business," big data and IT finance challenging. For example, the practical current mismatch of micro and macro financial market arbitrage seeking models – separating micro-economic considerations from macro ones and negating their underlying effects in micro (pricing) financial models. Such a mismatch leads to financial markets becoming 'incoherent.' For example, the mortgage-backed securities (MBS) crisis of 2008 was such a mismatch; combining the conditions of 'a home for everyone' based on low initiation costs and interests, with long run (and unsustainable) individual and systemic risks. Future models in finance may, therefore, be concerned fundamentally with risks they have not yet experienced.

Practically, data science is expanding digitalization, cloud computing, and computing enterprises. IT monopsonic and media sectors such as Google, Microsoft, Amazon, Apple, and a multitude of small firms have discovered that data is an asset that can be mined, sold, and resold. Artificial intelligence, created already in 1958, transformed mathematical models into intelligent software. One example is INRIA's expert optimal control system, self-designing a software used in Tapiero (1988). Software evolutions have also produced obsolescence risks. Cobol and Fortran, although used for some time, were replaced by a new "must" Lisp language and Lisp machines. They too became obsolete with C and C++, the new "kids in town." Today, we have R and Python that will probably mature and be replaced by other languages. Academic and research tendencies to emphasize empirical studies and data at the expense of mathematical models' integrity is also contributing to an additional brand of risk.

Nevertheless, financial IT and data science are providing immense opportunities that can also turn out to be an unwieldy process, victimized by the belief that a larger haystack may help to find a needle in that haystack. Yet, tamed big data can complement the statistical/scientific approach by providing an opportunity to reveal new hypotheses and new opportunities that can set such approaches on a more certain footing. Digitalized financial systems allow automatic context-specific interpretations, aggregation, and analysis of data (e.g., what information is relevant or not to a particular market or stock). For example, information and/or knowledge extracted from digital records can render financial bank jobs easier when diagnosing and detecting risky clients. Of course, it leads at the same time to the removal of human interventions from such processes. These opportunities have significant financial benefits but can also harbor social and financial risks, with society's risks enthralled in "artificially intelligent financial systems."

5. CONCLUSION: DATA SCIENCE AND STATISTICS CHALLENGED

Financial data science at its initial phase is expanding and challenging. The questions below summarize some of the issues it is challenged by:

- Is big data about looking for a needle in a haystack by adding hay?
- Is the future of finance a data science without "models"?
- Is data science merely another IT data-driven tool? Compared to statistics that seeks to justify what we define or conclude based on data? Can data science decide what we are to do? Or merely advise and maintain the freedom of choice?
- Is big data and its businesses the end of privacy?
- Are algorithmic models processing data science transparent models? Are they means or ends? Do they reveal the unexpected or merely the expected?
- Is data science something new? Or the marketing of well-known data analytics tools up-ended with a greater computational efficiency (computers on steroids)?
- Is big data a means to increase or reduce complexity? If so, what are its consequences to regulation, compliance, and safe finance? Is the growth of complexity designed and part of data science?
- Is big data in finance an evolving artificial intelligence for the "war of machines"?
- Is the growth of data and its practical analysis sustainable?

By contrast, there are already payoffs to financial technology and data science including among others:

- Strategic positioning in a global increasingly monopsonic and competing gated world.
- The mass customization of products and services (and their yet undefined consequences).
- Does globalization, complexity, gating, and the transformation of increasing speed of financial and technological finance render finance one of politics and gated national policies or one of markets?
- Immediacy: the need to be here and there and trade everywhere and at all times as well evading regulatory regimes. Do these empower financial corporate firms at the expense on financial investors or vice versa.
- The need to communicate and to sustain a state of instant and mobile communication.
- The need to keep pace and paces away from a future falling upon us faster and faster than the present can handle. Would the relative adaptation of individual investors overwhelm the architecture of corporate financial and banking systems?

These are engines motivating technological and financial growth, and in their wake, the growth of IT networks and services launched daily by the Internet and IT enterprises. In this process, a financial IT infrastructure is:

- Growing ever more complex.
- More diffused, technologically and otherwise.
- Harder to define.
- Network based.
- More difficult to assume and control.

The achievements of industries in integrating IT and data are, by comparison to financial and other services, immense. For example, in industry Internet has contributed to reductions in cycle times, costs, and labor, and helped introduce multiple products, and improve their qualities. It has also maximized and facilitated contacts across national boundaries leading to the globalization of industries and the expansion of supply networks. For these reasons, they incentivized services and financial technologies. Financial enterprises are proceeding in similar directions by adapting data science to theirs needs and expanding into a digitalized finance with networks and services spanning the globe. They do so to reduce financial logistic costs, expand their outreach, and increase their profits and the dominance of finance. Network effects have amplified innovation and creativity, and, in some cases, have empowered clients (although they have also augmented their hold of depositors). By the same token, it has expanded financial services to a far broader set of financial investors and empowered clients to manage their wealth globally. It has integrated financial and data systems and accessed online transactions, support multiple payment methods, and altered the business of banking.

REFERENCES

Albert, R., and Barabási A-L., 2002, "Statistical mechanics of complex networks," Reviews of Modern Physics 74:1, 47-97

Allen, J.F., 2001, "Hypothesis, induction and background knowledge. Data do not speak for themselves," BioEssays 23, 861–862. 33

Andersen T. G., R. A. Davis, J. P. Kreiss, T. Mikosch (eds.), 2009, Handbook of financial time series, Springer Verlag

Ashby W. R., 1956. Introduction to Cybernetics, Chapman & Hall

Billard, L., and E. Diday, 2003, "From the statistics of data to the statistics of knowledge: symbolic data analysis," Journal of the American Statistical Association 98:462, 470-487

Breiman, L., 2001, "Statistical modeling: the two cultures," Statistical Science 16:3, 199–231

Callebaut, W., 2012, "Scientific perspectivism: a philosopher of science's response to the challenge of big data biology," Studies in History and Philosophy of Biological and Biomedical Sciences 43:1, 69–80

Chambers, J. M., 1993, "Greater or lesser statistics: a choice for future research," Statistics and Computing, 3:4, 182-184

Cleveland, W. S., 2001, Data Science: an action plan for expanding the technical areas of the field of statistics," International statistical review, 69:1, 21-26

Cleveland, W. S., 1993, Visualizing data, Hobart Press

Cleveland, W. S. 1985, The elements of graphing data, Wadsworth Advanced Books and Software Monterey

Diday E., and F. Esposito, 2003, "An introduction to Symbolic Data Analysis and the Sodas Software," International Journal on Intelligent Data Analysis 7:6, 583–601

Donoho, D., 2015, 50 years of Data Science, September 18, version 1.00 (on the internet)

Fisher R. A., 1936, "The use of multiple measurements in taxonomic problems," Annals of Eugenics 7:2, 179-188

Goodman A. A., and C. Wong, 2009, "Bringing the night sky closer: discoveries in the data deluge," in Hey, T., The fourth paradigm: data-intensive scientific discovery, Microsoft Research

Groenen, P. J. F., S. Winsberg, O. Rodriguez, and E. Diday, 2006, "I-Scal Multidimensinal scaling of interval dissimilarities," Computational Statistics and Data Analysis 51, 360-378

Guetzkow, H., 1959, "A use of simulation in the study of inter-nation relations," Behavioral Science 4:3, 183-191

Guetzkow, H., 1962, Simulation in social science: readings, Prentice-Hall

Guttman L., 1944, "A basis for scaling qualitative data," American Sociological Review 9:2, 139-150

Hanani, U., and C. S. Tapiero, 1980, "The combined location – districting problem: a dynamic clusters analysis approach," Analyse de Donnees, INRIA, 249-266

Hey, T., S. Tansley, and K. Tolle, (eds.), 2009, The Fourth Paradigm: data-intensive scientific discovery, Microsoft Research

Horton , N. J, B. S. Baumer, and H. Wickham, 2015, "Setting the stage for data science: integration of data management skills in introductory and second courses in statistics," arXiv preprint arXiv:1502.00318

Kirkpatrick, S., and J. Kurths, 2012, "Challenges in network science: applications to infrastructures, climate, social systems and economics," European Physics Journal Special Topics 214:1, 273-293

Krohs, U., and W. Callebaut, 2007, "Data without models merging with models without data," in Boogerd, F. C., F. J. Bruggeman, J. H. S. Hofmeyr, and H. V. Westerhoff (eds.), Systems biology: philosophical foundations, Elsevier

Knight, F. H., 1921, Risk, uncertainty and profit, Cornell University Library

Langlois, R. N., and M. M. Cosgel, 1993, "Frank Knight on risk, uncertainty, and the firm: a new interpretation," Economic Inquiry 31:3, 456–465

McKinsey, 2011, "Big data: The next frontier for innovation, competition, and productivity," https://mck.co/2dDfLuS

National Research Council, 2010, "Technical capabilities necessary for regulation of systemic financial risk: summary of a workshop," The National Academies Press, Washington, D.C.

Nyambuu, U. and C. S. Tapiero, 2018, Globalization, gating and risk finance, Wiley

Pearl, J., 1985, "Bayesian networks: a model of self-activated memory for evidential reasoning," (UCLA Technical Report CSD-850017). Proceedings of the 7th Conference of the Cognitive Science Society, University of California, Irvine

Raveh, A., and C. S. Tapiero, 1980a, "Periodicity, constancy, heterogeneity and the categories of qualitative time series," Ecology 61:3, 715-719

Raveh, A., and C. S. Tapiero, 1980b, "Finding common seasonal patterns among time series: an MDS approach," Journal of Econometrics 12, 353-363

Tapiero, C. S., 2013, Engineering risk and finance, Springer

Tapiero, C., S., 1988, Applied stochastic models and control in management, North-Holland

Tapiero, C. S., M. F. Capobianco, and A. Y. Lewin, 1975, "Structural inference in organizations," Journal of Mathematical Sociology 4, 121-130

Tapiero, C. S., and P. Vallois, 2000, "The inter-event range process and testing for chaos in time series," Neural Network World 10:1–2, 89–99

Tapiero, C. S., and P. Vallois, 2015, "Financial modelling and memory: Mathematical system. Future perspectives in risk models and finance," Springer, Vol. 211, 149-246 (International Series in Operations Research and Management Science; vol. 211)

Tapiero, C. S., 2017, "Data science and intelligence," Risk and Decision Analysis 6:4, 291-298

Tapiero, C. S., and P. Vallois, 2016, "Fractional randomness," Physica A: Statistical Mechanics and its Applications 462, 1161-1177

Tapiero, C. S., and P. Vallois, 2017, Implied fractional rates and default risk distributions, Probability, Uncertainty and Quantitative Risk, Springer.

Tapiero, C. S., and P. Vallois, 2018a, "Fractional randomness and the Brownian bridge," Physica A 503, 835–843

Tapiero, C. S., and P. Vallois, 2018b, "Fractional randomness and alpha-stable distributions," Physica A 511, 54-60

Overbye, D., 2012, "The trouble with data that outpaces a theory," March 26, New York Times, https://nyti.ms/2wTt0AT

The Economist, 2010, "Data, data everywhere," February 25, https://econ.st/2x1TNdw

Tukey, J. W., 1962, "The future of data analysis," The Annals of Mathematical Statistics 33:1, 1-67

Tukey, J. W., 1977, Exploratory data analysis, Pearson

Tukey, J. W., 1994, The collected works of John W. Tukey, Vol. 1, Taylor & Francis

Vallois, P., and C. S. Tapiero, 2007, "Memory-based persistence in a counting random walk process," Physica A 386:1, 303-317

Vallois, P., and C. S. Tapiero, 2009, "A claims persistence process and insurance," Insurance Economics and Mathematics 44:3, 367-373

CAPCO

THE FUTURE. NOW.

ABOUT CAPCO

Capco is a global technology and management consultancy dedicated to the financial services industry. Our professionals combine innovative thinking with unrivalled industry knowledge to offer our clients consulting expertise, complex technology and package integration, transformation delivery, and managed services, to move their organizations forward. Through our collaborative and efficient approach, we help our clients successfully innovate, increase revenue, manage risk and regulatory change, reduce costs, and enhance controls. We specialize primarily in banking, capital markets, wealth and investment management, and finance, risk & compliance. We also have an energy consulting practice. We serve our clients from offices in leading financial centers across the Americas, Europe, and Asia Pacific.

To learn more, visit our web site at **www.capco.com, or follow us on Twitter, Facebook, YouTube and LinkedIn.**

WORLDWIDE OFFICES

APAC
Bangalore
Bangkok
Hong Kong
Kuala Lumpur
Pune
Singapore

EUROPE
Bratislava
Brussels
Dusseldorf
Edinburgh
Frankfurt
Geneva
London
Paris
Vienna
Warsaw
Zurich

NORTH AMERICA
Charlotte
Chicago
Dallas
Houston
New York
Orlando
Toronto
Tysons Corner
Washington, DC

SOUTH AMERICA
São Paulo

WWW.CAPCO.COM