Revista de Educação e Pesquisa em Contabilidade
Journal of Education and Research in Accounting

Periódico Trimestral, digital e gratuito publicado pela Academia Brasileira de Ciências Contábeis

ISSN 1981-8610

REPeC, Brasília, v. 11, n. 4, art. 2, p. 385-395, Oct./Dec. 2017
Available online at *www.repec.org.br*
DOI: http://dx.doi.org/10.17524/repec.v11i4.1634

Big Data: Epistemological Reflections and Impacts in Finance and Capital Market Studies

Abstract

Objective and method: Access to data series plays a central role in the area of Finance. The increasing availability of large volumes of data, in different formats and at high frequency, combined with the technological advances in data storage and processing tools, have created a new scenario in academic research in general, and in Finance in particular, generating new opportunities and challenges. Among these challenges, methodological issues emerge, which are widely discussed among researchers from different areas, but also epistemological issues that deserve greater space for discussion. Thus, the objective of this theoretical essay is to analyze the conceptual and epistemological aspects of the use of intensive data and its reflec-tions for the area of Finance.
Results and contributions: We consider that the hypothetical-deductive method of empirical research, which is the most recurrent, limits the construction of knowledge in the so-called 'Big data era', as this approach starts from an established theory and restricts research to test-ing the hypothesis(es) proposed. We advocate the use of an abductive approach, as argued in Haig (2005), which converges with the ideas of grounded theory and which seems to be the most appropriate approach to this new context, as it permits greater capacity to collect value information for the data.
Key Words: Big Data, Abductive Method, Epistemology, Finance.

Talieh Shaikhzadeh Vahdat Ferreira
M.Sc. in Business Administration from Universidade Federal da Paraíba (UFPB) and Ph.D. candidate in Business Administration from Universidade Federal da Paraíba (UFPB). Contact: UFPB, Cidade Universitária, s/n, CCSA/PPGA, Castelo Branco, João Pessoa (PB). CEP.: 58051-900.
E-mail: taliehv@gmail.com

Francisco José da Costa
Ph.D. in Business Administration from Fundação Getúlio Vargas (FGV/SP) and Professor, Department of Business Administration, Universidade Federal da Paraíba (UFPB). Contact: Cidade Universitária, s/n, CCSA/PPGA, Castelo Branco, João Pessoa (PB). CEP.: 58051-900.
E-mail: franze.mq@gmail.com

Published in Portuguese and English. Original Version in Portuguese.

Received in 6/24/2017. Ask to Revise on 8/19/2017. Resubmitted on 9/9/2017. Accepted on 9/10/2017 by Dr. Gerlando Augusto Sampaio Franco de Lima (Assistant Editor) and by Dr. Orleans Silva Martins (Editor). Published on 10/30/2017. Organization responsible for the journal: Abracicon

1. Initial Considerations

The advances of the Internet and the expansion of the use of mobile communication technologies have resulted in a substantial increase in the quantity and storage of data. Taking into account the year 2016, according to International Business Machines (IBM), 90% of all data produced in the world had been created in the previous two years. Concurrently, the same company projected that this volume would double every two additional years (IBM, 2016). It was this growing volume of data, combined with the technological advances in processing and storage tools, that have laid the foundations for what is being called 'Big Data' among academics, businessmen and governments (Ekbia et al., 2015).

The advent of Big Data established a new frontier for the development of research (Chen & Zhang, 2014), entailing opportunities and challenges. Opportunities arise as data accessibility increases in a massive way, which drives the continuous improvement of technologies and uses in the most different areas of human life (Demchenko, Grosso, De Laat & Membrey, 2013) . Among these areas that benefit from Big Data, we can mention Finance, an area in which data plays a central role in its academic and professional development (Seth & Chaudhary, 2015).

On the other hand, the 'Big Data era', due to its particularities (which we shall discuss later), brings new dilemmas, primarily associated with technological and statistical questions, and secondly with the epistemological dimension of research and knowledge production (Ekbia et al., 2015). This theoretical essay aims to analyze this last (epistemological) aspect and its reflexes in the area of Finance, discussing possible developments in the capital market.

For this, the paper is divided into five sections, including this introduction. In the second section, we seek to characterize Big Data and point out changes in knowledge construction and in the choice of research method due to epistemological issues that emerge. In the third section, some of the methodological challenges and alternatives to the knowledge production process are explored. In the fourth section, we focus on some issues the use of the Big Data approach brings to Finance. Finally, some considerations are exposed and other challenges are identified that we consider more relevant to the new academic and professional setting.

2. Big Data: characteristics and challenges of knowledge construction

In this item, we present and briefly discuss the conceptual elements and implications of Big Data. Initially, we present its concepts, following the steps defined in the specialized literature, which has not exactly defined 'Big Data', but appointed characteristics associated with the data that are collected and stored. In the second part, and in view of this first construction, we present the discussions about the consequences of a Big Data approach for the academic and professional knowledge construction process.

2.1. Characteristics of Big Data

With the opportunities for research and decision-making that emerge from the enormous volumes of data, the possibilities for using this data have been widely debated both inside and beyond the academy, which contributes to constant advances in its understanding and use (Ekbia et al. al., 2015). No consensus has been established yet on the very definition of Big Data itself though, which has been evolving and maturing over time. Initially, the term was characterized by its association with so-called "3 V's": volume, velocity and variety. The main contributors were technology companies, such as IBM and Oracle. In more recent studies, this form of definition was refined, adding two attributes - value and veracity - to characterize Big Data, constituting the "5 V's" (Demchen-ko et al., 2013; Lau, Zhao, Chen, & Guo, 2016).

The first crucial characteristic of Big Data is thus related to the **volume** of data. Initially, this discussion may be limited to quantifying terabytes, pentabytes, or zettabytes (Kitchin, 2014), but the data production capacity is expected to continue to expand, which may limit the usefulness of this initial definition. Lau et al. (2016) propose a broader discussion on this issue when stating that the volume of Big Data are expected to reach such an extent that current technologies find it difficult to store, recollect, analyze and use.

Speed is related not only to the data production capacity, but also to the processing and analysis capacity of the computer systems involved. In fact, mass data production, such as a national census, for example, is not new. The costs of its processing and the time required for data collection and analytical elaboration have been high though (Miller, 2010). Thus, Big Data is characterized by the continuous production of data, sometimes associated to multiple events, at the detailed level, with great flexibility to analyze its scope, through tools with increasing agility of real-time processing (Kitchin, 2014).

Considering the versatility that Big Data bases may present, the third feature is associated with the **variety** of formats and data sources. With the development of Web 2.0 and Web 3.0, data can be captured not only in its conventional format, such as tables or spreadsheets, but also in semistructured or unstructured formats, or in a mixed way, such as images, sounds, among other possibilities (Demchenko et al., 2013). This heterogeneity of formats demands a new generation of techniques and methods for processing and storage that are constantly being improved.

Value is a key feature when dealing with data in general, since all processing and investment effort is only justifiable when these data add value to the activity or analysis in question (Demchenko et al., 2013). Thus, insofar as the processing tools can transform the collected data in order to produce knowledge, the benefits obtained with this type of technology can reach higher levels of value (Lau et al., 2016).

And, finally, **veracity** is related to the quality and validity of the data. This implies both the consistency that the data must possess, being reliable in terms of its measurement and validity, and the quality of the data itself in terms of integrity, which depends on an entire chain, from its collection to the processing and storage methods (Demchenko et al., 2013). In this context, according to Lau et al. (2016), new technological challenges emerge in the quest to maintain the quality and consistency of the data in enormous bases and that are being updated in real time. In close connection with this challenge of quality and data consistency, the issue of knowledge production based on Big Data emerges, a subject we will discuss in the next section.

2.2. Knowledge construction based on Big Data

The growing application of Big Data in research has altered the conventional forms of knowledge construction (Demchenko et al., 2013, Chen & Zhang, 2014 and Kitchin, 2014). For this essay, we take the view of Chen and Zhang (2014) for reference, who assert that, historically, the construction of scientific knowledge was based on three great paradigms: Empirical Science, Theoretical Science and Computer Science. Hundreds of years ago, science was built on empirical experiments that aimed to test possible intuitions and prove their veracity, constituting the first paradigm. As this evidence grew, theories could be developed, reaching the second paradigm, with the theoretical construction of knowledge.

The complexity of the phenomena under analysis was widening though and researchers needed to use new scientific simulation tools to validate their outcomes. Thus, the concept of the 'Third Paradigm' emerged, which are the computer sciences, which permitted large-scale scientific simulations, creating 'sufficiently robust' results (Hey, Tansley & Tolle, 2009). The case is that the nature of 'Big Data' for the elaboration of simulations and tests demands techniques and technologies completely distinct from the other three paradigms. This gave rise to the idea of the 'Fourth Paradigm', which would be the science of intensive data (Miller, 2010; Hey, Tansley & Tolle, 2009). In this context, more and more researchers and market players need to combine advances in technology with the efficient processing of large volumes of data and the use of more conventional scientific methods to benefit from the collection of desirable data (Demchenko et al., 2013) because, in this scenario, there is plenty of data available. But for these to add value, they need to be processed, transformed into bases to permit the identification of patterns, and finally be interpreted in order to broaden the knowledge about what one is trying to understand (Seth & Chaudhary, 2015). In another perspective, Haig (2005) points out an organization of the evolution of scientific research in two prevailing methods or paradigms, which are the inductive and the hypothetic-deductive. According to Lakatos and Musgrave (1979), the distinction between knowledge and speculation is based on the premise that, while the former was proved by the force of knowledge or the senses, the latter was not. For the inductivists, this evidence starts from an observation that provides the secure basis upon which knowledge can be constructed, so that propositions can be made by induction, from the particular to the whole. On the other hand, the hypothetical-deductive method starts from an established theory, generates hypothesis, and then seeks to test them to broaden their knowledge or identify possible inconsistencies in the theory. In general, the inductive perspective is more associated with qualitative empirical procedures; and the hypothetical-deductive more with quantitative methods, mainly statistical methods that involve inferential estimation techniques and hypothesis testing.

The use of large volumes of data creates a new complexity though, which does not seem to be fully encompassed by these analysis methods. This assertion is based on the possibility of the analysis departing from the data itself, that is, not necessarily starting from a specific hypothesis or theory. Considering the operation of large volumes of data, quantitative techniques will be necessary for procedures that are characteristic of qualitative research. We then approach the inductive view of quantitative procedures typically applied in the hypothetical-deductive view.

Thus, we understand that Big Data entails an effective change in relation to traditional approaches, which requires a different epistemological approach in the development of science in comparison to the current hypothetical-deductive method (Kitchin, 2014). In our view, the best alignment comes from the rescue of a debate that has been ongoing for some decades about grounded theory and the so-called abductive method. In the next item, this discussion will be detailed.

3. Epistemological and methodological changes of the age of Big Data

In this section, we discuss the methodological and epistemological consequences of a real-age research based on abundant data. Initially, we describe the abductive perspective as an appropriate framework of knowledge production in this scenario. Next, we present the methodological and tool challenges most frequently cited in the specialized literature.

3.1. Abductive Method and Big Data

Given the advent of Big Data, it seems to be the appropriate time to rethink the epistemological issues and the research methods in larger units that permit the expansion of knowledge by new perspectives from the data and not by a specific hypothesis. This is justified by the fact that the inductive and hypothetico-deductive method, although presenting different logics, share one fundamental characteristic: they focus on a limited part of the data to perform their analyses (Haig, 2005), so that they can leave important data on the phenomenon or object of research beyond the analysis of results.

It is important to note that the well-structured attempt to construct knowledge from the data is not a new concept in the academy. In 1967, Barney Glaser and Anselm Strauss began discussions on the grounded theory, which is considered to be one of the purest forms of qualitative research (Glaser & Strauss, 1967, Bianchi & Ikeda, 2008). In this method, the researcher does not seek to verify or depart from a specific theory for the elaboration of his research. On the contrary, the researcher is invited to break free from his already established theoretical assumptions and 'moorings' to seek to understand the data, so that questions and knowledge are drawn from them (Glaser & Strauss, 1967).

The main purpose of grounded theory is the construction of theories and, for this, researchers assume two core assumptions: a) researcher, reality and theory are continuous and intrinsic entities and can thus interact continuously; b) the theory evolves throughout the research process, resulting from several interpolations of data and analyses (Bianchi & Ikeda, 2008). Thus, the idea is to ensure that not only the researcher investigates the reality from the data, but that this is a continuous process and that can change at any time, since the data are continuously being re-accessed and analyzed.

Haig (2005) proposed for quantitative research the abductive method (or paradigm, as Haig called it), following the same fundamentals the grounded theory presents as a qualitative method. This method follows the same epistemological logic of grounded theory, that is, it seeks knowledge based on the data, but follows a different methodological execution protocol. The abductive method can be considered a combination of the inductive and hypothetical-deductive methods in the context of quantitative research, as it enables the researcher to discover the empirical facts (by induction) to construct theories that explain these facts (by deduction) (Haig, 2005). Figure 1 illustrates the research dynamics of the abductive method:

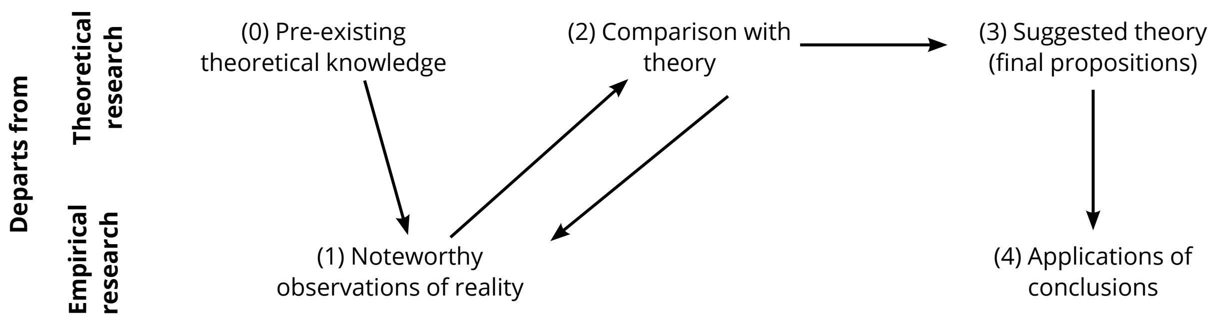

Figure 1. The view of research through the abductive process.

Source: Spens and Kovács (2006).

In this view, phenomena exist to be explained, so that they are not simply understood as objects of prediction or confirmation of theories and hypotheses. In practical terms, this means that, after identifying a phenomenon based on the quantitative and exhaustive exploitation of the data, abductive inferences are used, that is, the best explanation is sought for its occurrence, which is understood as the probable true explanation. Thus, different methods or combinations of them can be used to select the 'best explanation' that, when identified, can serve as the basis for the elaboration of a theory, if it does not fit into any existing theory (Haig, 2005).

Thus, the abductive method can be an interesting alternative as a study method that enables the construction of knowledge from large volumes of data. With the use of the abductive approach, the exhaustive exploration of large databases can foster the generation of knowledge that until then was not available to verification-based methods. As mentioned above, verification methods have the limitation that the research hypothesis itself imposes, so that they are potentially restricted to verifying the data necessary for the analysis of the hypothesis and do not benefit from other information that could be obtained from this same database to enable the understanding of more complex themes in contemporary science (Miller, 2010). Additionally, the abductive method permits that, in the same sample, more than one phenomenon can be identified and that theories can be constructed in a way that better explains reality, using analogies to seek answers to themes that are sometimes subjective or about which no census data are available (Ekbia et al., 2015).

The use of this approach for Big Data analysis also solves one of the pitfalls that may arise when dealing with 'data-intensive science', which means giving exaggerated importance to the data and their causal relationships (e.g. data brokers and data analysis providers). This trap can lead to the false impression that the simple identification of the causal relationship would be sufficient for the production of knowledge, which is, in general, an exaggerated simplification of the phenomenon under analysis (Kitchin, 2014). It is important to keep in mind that the construction of knowledge is guided by concerns, giving shape to the 'research problem', which permit theorizing about why these relationships occur and what their theoretical and practical implications are.

A related issue of this discussion that also raises relevant reflections for academic and professional researchers concerns the methodological decisions of research, which we discuss in the following item.

3.2. Methodological challenges for using Big Data

Research approaches based on large volumes of data entail methodological challenges of two kinds: the first, regarding the technological capacity of data accumulation and processing; and the second on the statistical analysis of these data.

Regarding the first challenge (processing), new processing techniques have been developed, such as data mining, artificial intelligence, statistical learning, machine learning, predictive analytics. These techniques also allow the detection of standards without the need to provide the system with a specific question, which enables the identification of information that might not be easily observable. The capacity of the mentioned techniques comes from computer schedules that permit real-time learning from the data as they are collected. From these 'new' lessons, then, prediction models are constructed that can be constantly reformulated in order to continually improve their results and improve decisions (Seth & Chaudhary, 2015).

Additionally, between 80 and 90% of the data are available in a non-standardized and unstructured form, a fact that creates integration difficulties and quality gaps that increase the challenges for the processing and analysis with the agility needed to add value (Oracle, 2012). There is, therefore, a need for robust database structures that permit the accumulation of data and cross-over possibilities for interdisciplinary research and collaboration (Demchenko et al., 2013).

The Big Data view as accelerated data accumulation alone does not necessarily have value. Analysis tools also need to advance at the same speed. The volume of information and its different formats have affected multiple dimensions of research, since traditional technologies and models fail or present major inconsistencies when accessing such a large volume of data (Einav & Levin, 2014).

It is in this context that the need arises to develop robust techniques that are adapted to this type of situation. In other words, new methods should emerge, including statistical assumptions that consider the use of samples that are very close to the study population (Einav & Levin, 2014), or that are more consistent in informative terms than conventional estimation methods and hypothesis tests. For example, in interval-estimation procedures that involve standard error, this standard error is an inverse function of the sample size, so that, in very large samples, the standard error tends to zero, which makes the interval estimation meaningless. A consequence of this finding is the possibility of an effective return to (or revalidation of) data analyses based on 'magnitude of effect' and less on significance analyses (or p-values, according to Gigerenzer & Marewski, 2015).

Also concerning the analyses, another question emerges from the restrictions that will come in modeling analyses, which, in traditional research, use rigid assumptions, such as normality or homoscedasticity of errors in regression models. With the introduction of large volumes of data, these assumptions are potentially broken, invalidating confidence-building tests of the results, also because the conventional procedures for verifying these assumptions are based on statistical tests, which come with utility problems when the samples are very large, given the tendency to always reject the null hypothesis.

In addition, if not globally, in certain data segments of some variables, it will always be possible to verify spurious correlations, given the large data variance. This results in the need for a distinguished interpretative capacity for users of both the correlation technique and other correlation-based techniques, such as factorial analysis for example.

In short, having the Big Data approach as a driver of technologies and statistical methods, a large number of technological infrastructure components are rethought to overcome related challenges, mainly the increasing volume of data collected from different sources and formats. In addition, multidisciplinary joint effort is needed among areas, such as Administration, Economics, Mathematics, Statistics and Computer Science, in order to redefine the models used for such analyses.

The perspective of academic and professional impacts resulting from this perspective is broad and will find repercussions in several areas of study and research. For this article, and as already indicated in the introduction, we aim to analyze the implications in the specific context of Finance and Capital Markets, as shown in the following item.

4. Big Data and its contribution to Finance and Capital Market studies

Transaction data and activities from the Finance universe, their processing and access play a central role in the area of Finance (Seth & Chaudhary, 2015). Capital markets, as the main source of data for analysis in this area, have significantly changed since the start of the 21st century by reaching increasingly high levels of High-Frequency data (HFD) production, such as the US market, which has about 70% of all HDF trading data (Zervoudakis, Lawrence, Gontikas & Al Merey, 2017). In this context, one can observe a typical configuration of Big Data, given the almost continuous registration of activities, the different forms of manifestation of those activities and the high volume of records.

All areas of finance (investment analysis, econometrics, fraud detection, behavioral finance, etc.) can benefit from this scenario, as the analysis of intensive data enhances the ability to measure both the systematic risks, which cover the market as a whole - and the non-systematic risks - related to each of the companies (Fan, Han & Liu, 2014). In this essay, we chose to restrict the comments to the potential impacts of the Big Data view in the analyses, involving three areas of Finance that have a direct relationship to the capital market: volatility, portfolio development and risk analysis and market transparency. The essay design and our intention to provide a more illustrative than exhaustive reference led us to a brief exposition, limited to pointing out generic implications and potential ramifications.

Volatility is commonly defined as the dispersion of the price of an asset over a certain period of time and historically is an area of significant relevance in Finance surveys, as the measurement and projection of asset volatility are crucial for activities such as asset allocations, derivative pricing, and investment management and analysis options (Seth & Chaudhary, 2015). Indeed, studies that measure market volatility in the context of Big Data can advance in the accuracy of their prediction models by manipulating large volumes of data (Louzis, Xanthopoulos-Sisinis & Refenes, 2013) in order to generate more reliable analyses that provide gains to their investors.

More accurate financial assessments can minimize the likelihood of failures, especially in periods of crisis, when volatility can significantly expand. In addition, volatility presents high explanatory power due to high persistence and conditional dependence, being non-stationary and predictable (Louzis, Xanthopoulos-Sisinis & Refenes, 2013). Given their relevance, many models were developed to estimate volatility, Engel's autoregressive conditional heteroscedasticity model (ARCH) and its variations being one of the most popular in the field (Seth & Chaudhary, 2015). These models were designed to use low frequency data though, with daily returns, losing intraday information that may be valuable in increasing the accuracy of these estimates.

Access to high-frequency data, which is typical of Big Data, has the potential to alter this scenario significantly as statistical models and theories are developed, considering the adequacy necessary for the use of these data (Cartea & Karyampas, 2011). In sum, as far as volatility is concerned, the Big Data approach influences the improvement of measurement, prediction processes and models, based on both the large data volume and the continuity characteristics of the data production.

With regard to the **elaboration of investment portfolios and risk analysis** for the selection of assets, the Big Data perspective has also broadened horizons. Until the early 2010s, asset valuation models have relied heavily on specific data from the company under review or from the economy it is exposed to, basically providing a fundamentalist analysis (Damodaran, 2012). As data availability grew along with the possibility of processing and quantification, business valuation models began to consider a much larger range of information sources than ten years ago, expanding their ability to identify the best opportunities (GSAM, 2016). These innovations include the analysis of texts, images, audios of meetings with shareholders, presentations, and other information in unstructured formats.

The development of machine learning techniques and statistical learning tools (James, Witten, Hastie & Tibshirani, 2013) has also contributed to the creation of more dynamic assessment models, which adapt and learn from the constant processing of large data volumes, speeding up the time it takes for the information to become available for investors or analysts' decision making (GSAM, 2016). In addition, in emerging markets, such as Brazil, where information asymmetry is more pronounced, information deficiencies may lead to increased uncertainty and erratic asset pricing (Martins & Paulo, 2014). Thus, data-intensive models can create an additional advantage for investors by reducing interest rates and directing capital more clearly to the best investments.

In addition to the monetary and financial efficiency benefits these advances can offer to the different market players, there is an area of finance that should be of particular relevance in this context, which is the area of **market transparency**. The use of Big Data tools, such as business intelligence or business analytics, brings a new perspective to the studies and efforts related to increased transparency and, consequently, allows the increase of liquidity, market efficiency (Ye, 2010) and a more stable and reliable environment for its development.

Capital market demands for higher levels of transparency gained momentum especially after the financial scandals of the early 2000s, which involved large corporations such as Enron and WorldCom, prompting different discussions about their content and regulation (Chong & Lopez-de-Silanes, 2007). These events unleashed unprecedented government intervention in the United States capital market, followed by most countries, with a focus on the establishment of the Sarbanes-Oxley Act (Aksu & Kosedag, 2006). More recently, the US housing credit crisis in 2008 revealed that there were still significant limitations and inadequacies of the information disclosure system within the financial system. The lack of standard procedures, quality gaps in information disclosure and a lack of data processing capacity in a timely manner have led regulators not to have the ability to process that information, which has made proactive measures or the more precise identification of what information was missing impossible (Seth & Chaudhary, 2015).

Thus, in this crisis scenario, regulatory demand has played a key role in accelerating the development of Big Data processing solutions in the search for higher degrees of transparency (Oracle, 2012). This movement was not restricted to the capital market of course, as all the companies operating in it have been forced to increase the volume of data disclosed and to redesign their technological infrastructure to support this demand.

These advances make it possible to monitor and identify frauds and other actions that can jeopardize market transparency and reliability more quickly and accurately (Louzis, Xanthopoulos-Sisinis & Refenes, 2013), which would preserve and enhance the market liquidity and efficiency. Thus, the Big Data view offers new possibilities for innovation and growth in the area, which, in view of the increasing accessibility characteristic, allows regulators, companies and researchers to collaborate to obtain solutions together, extracting the maximum of value from the data.

In this scenario, transparency-related discussions have changed their data volume requirement focus, as they are more abundant, to reflect on how to overcome the challenges of their processing (Seth & Chaudhary, 2015). Some examples of such challenges could be: How to process an increasing data volume in a timely manner? How to overcome the absence of standard disclosure protocols? How to define schedules that can fraudulently identify frauds? The search for answers to these and many other questions related to transforming market data into knowledge can generate regulatory tools that are increasingly effective for markets to achieve a new level of transparency and efficiency.

These three aspects we have briefly discussed can benefit not only from the expansion of data the access to Big Data provides, but also from a new epistemological perception of how those data can be analyzed, especially at the academic level. It is here that the use of more abductive methods would permit the expansion of the scope of information the researcher focuses on to know the reality, stimulating the observation of phenomena not restricted to a set of proposed hypotheses for confirmation or rejection. This new look, which at first considers an inductive movement, can create a revolution in an area that has historically departed from purely hypothetical-deductive methods for the construction of knowledge.

5. Final Considerations

The incorporation of the Big Data perspective, in the academy as well as in the professional world, seems to be an inevitable and irreversible move. The possibility of basing the analysis on data with quasi-census magnitudes extends the possibilities of minimizing information asymmetry in all fields of science and business by changing the concerning with data collection to the search for processing and analysis tools. A 'new multidisciplinary language' should emerge to encompass this growing complexity, redesigning and adapting methods and protocols to obtain the maximum value from the data.

This new panorama raises even deeper issues related to the construction of knowledge as it has traditionally taken place. With the establishment of data-driven science (data-drive science, according to Kitchin, 2014), the more traditional epistemology, strongly based on the hypothetical-deductive method, seems to be limited to reach the full value of the data the Big Data perspective provides.

The Big Data discussions on improving the method, the processing and methodological protocols for extracting the largest possible data value are important for the various dimensions of Finance, which is an area where data plays a central role for knowledge production and for managerial and regulatory decisions. We highlight here the studies related to volatility, portfolio elaboration and risk analysis and market transparency, which are areas that have already begun to develop the tools needed to benefit from intensive data, but still present many processing challenges in order to be able to extract the maximum value.

We aimed to analyze how the characteristics of Big Data, particularly volume, speed and variety, reach the research and management practices in these fields. Even for a generic and restricted view of these three themes, we were already able to observe how these areas need improvements as they get hold of more data, with access that is approaching continuity, and with the use or development of better tools to manipulate these growing volumes and greater diversity of data types and formats. The deeper analysis of the implications of Big Data in these fields and in other specialties in the area of Finance which we do not address represent challenges for further research.

This scenario does not yet require a paradigm shift in methodological and epistemological terms. Therefore, we have suggested throughout the text and again emphasize that the incorporation of an abductive approach in the field of Finance research could increase the capacity to extract valuable information and will enable more complete knowledge construction, without limiting the analyses to what is asked as previously defined hypotheses, but accessing all possible information available in the data. The search for a solution to these challenges has required not only a multidisciplinary viewpoint, but also an approach to different actors in society, who are sometimes far from solving challenges, especially governments (regulators), companies and academics. Therefore, we suggest that empirical studies analyze concrete views, challenges and experiences involving these interest groups in the professional field of Finance, in order to generate an increasingly complete and consistent view of best practices and best decisions in the field of Finance in Brazil.

References

Aksu, M., & Kosedag, A. (2006). Transparency and disclosure scores and their determinants in the Istanbul Stock Exchange. *Corporate Governance - an International Review*, *14*(4), pp. 277-296. DOI: https://doi.org/10.1111/j.1467-8683.2006.00507.x/abstract

Bianchi, E. M. P. G., & Ikeda, A. A. (2008). Usos e aplicações da *grounded theory* em administração. *Revista Eletrônica de Gestão Organizacional*, *6*(2), pp.231-248.

Cartea, A., & Karyampas, D. (2011). Volatility and covariation of financial assets: a high-frequency analysis. *Journal of Banking and Finance*, *35*(12), pp.3319-3334. DOI: https://doi.org/10.1016/j.jbankfin.2011.05.012

Chen, C. L. P., & Zhang, C.Y. (2014). Data-intensive applications, challenges, techniques, and technologies: a survey on Big Data. *Information Science*, *27*(5), pp. 314-374. DOI: https://doi.org/10.1016/j.ins.2014.01.015

Chong, A., Lopez-de-Silanes, F. (2007). *Investor protection and corporate governance:* intra-firm evidence across Latin-America. Palo Alto: Stanford University Press.

Damodaran, A. (2012). *Investment valuation*: tools and techniques for determining the value of any assets. 3rd edition. New Jersey: Wiley&Sons.

Demchenko Y., Grosso, P., De Laat C., & Membrey, P. (2013). Addressing Big Data issues in scientific data infrastructure. *Collaboration Technologies and Systems* (CTS), International Conference on 2013.

Einav L. & Levin J. (2014). Economics in the age of big data. *Science*, *346* (6210), pp. 12430891–12430896. DOI: https://doi.org/10.1126/science.1243089

Ekbia, H., Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., Suri, V. R., Tsou, A., Weingart, S., & Sugimoto, C. R. (2015). Big Data, bigger dilemmas: a critical review. *Journal of the Association for Information Science and Technology*, *66*(8), pp. 1523-1545.

Fan, J., Han, F., & Liu, H. (2014). Challenges of Big Data Analysis. *National Science Review*, *1*(2), pp. 293–314. DOI: https://doi.org/10.1093/nsr/nwt032

Gigerenzer, G., & Marewski, J. N. (2015). Surrogate science: the idol of a universal method for scientific inference. *Journal of Management*, *41*(2), pp.421-440. DOI: https://doi.org/10.1177/0149206314547522

Glaser, B, G., & Strauss, A. L. (1967). *The discovery of grounded theory: strategies for qualitative research*. New York: Aldine de Gruyter.

GSAM – Goldman Sachs Asset Management (2016). Perspectives: the role of Big Data in investing. Recuperado em 26 de abril, 2017, de: https://www.gsam.com/.

Haig, B. D. (2005). An abductive theory of scientific method. *Psychological Methods*, *10*(4), pp. 371-388. DOI: https://doi.org/10.1037/1082-989X.10.4.371

Hey, T., Tansley, S., & Tolle K. (2009). *Jim Grey on eScience: A transformed scientific method.* In: Hey T, Tansley S and Tolle K (eds) The fourth paradigm: data-intensive scientific discovery. Redmond: Microsoft Research, pp. xvii–xxxi.

Big Data, New Epistemologies and Paradigm Shift (PDF Download Available). Recuperado em 15 de setembro, 2017, de: https://www.researchgate.net/publication/271525133_Big_Data_New_Epistemologies_and_Paradigm_Shift.

IBM, 2016. Recuperado em 08 de abril, 2107, de: https://www-01.ibm.com/software/data/bigdata/what-is-bigdata.html.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. New York: springer. DOI: https://doi.org/10.1007/978-1-4614-7138-7

Kitchin, R. (2014). Big Data, new epistemologies and paradigms shift. *Big Data & Society*, April-June, pp. 1-12. DOI: https://doi.org/10.1177/2053951714528481

Lakatos, I., & Musgrave, A. (1979). *A crítica e o desenvolvimento do conhecimento*. São Paulo: Editora da Universidade de São Paulo. DOI: http://dx.doi.org/10.11606/issn.2447-9799.cienciaefilosofi.1980.107354

Lau, R. Y.K., Zhao, J.L., Chen, G., & Guo, X. (2016). Big Data commerce. *Information & Management*, *53*, pp. 929-933. DOI: https://doi.org/10.1016/j.im.2016.07.008

Louzis, D.P., Xanthopoulos-Sisinis, S., & Refenes, A.P., (2013). The role of high-frequency intra-daily data, daily range and implied volatility in multi-period value-at-risk forecasting. *Journal of Forecasting*, *32*(6), pp. 561–576.

Martins,O. S. & Paulo E. (2014). Assimetria de informação na negociação de ações, características econômico financeiras e governança corporativa no mercado acionário brasileiro. *Revista Contabilidade & Finanças (Online)*, *25*(64), pp. 33-45. jan./fev./mar./abr. 2014

Miller, H. J. (2010). The data avalanche is here. Shouldn't we be digging? *Journal of Regional Science*, *50*(1), pp. 181-201. DOI: https://doi.org/10.1111/j.1467-9787.2009.00641.x

Oracle (2012). Financial services data management: Big Data technology in financial services. Oracle Financial Services, *An Oracle White paper.*

Seth, T., & Chaudhary, V. (2015). "Big Data in Finance". In: Li, K.; Jiang, H.; Yang, L. T.; Cuzzorea, A. (Eds) "Big Data: algorithms, analytics and applications". Chapman and Hall/ CRC, pp. 329-356.

Spens, K. M., & Kovács, G. (2006). A content analysis of research approaches in logistics research, *International Journal of Physical Distribution & Logistics Management*, *36*(5), pp. 374-390. DOI: https://doi.org/10.1108/09600030610676259

Ye, G. (2010). *High frequency trading models*. NJ: John Wiley & Sons Inc.

Zervoudakis, F., Lawrence, D., Gontikas, G., & Al Merey, M. (2017). *Perspectives on high-frequency trading*. Recuperado em 30 de março, 2017, de: http://www0.cs.ucl.ac.uk/staff/f.zervoudakis/docs/hft.pdf.