International Journal of Social Science and Economic Research

ISSN: 2455-8834

Volume: 04, Issue: 07 "July 2019"

BIG DATA IN ECONOMIC ANALYSIS: ADVANTAGES AND CHALLENGES

BALAR Khalid[1] and CHAABITA Rachid[2]

[1,2] Department of Economic Science and Management, Hassan II University/Faculty of Economic and Social Legal Sciences Ain Chock, Casablanca, Morocco

ABSTRACT

In this paper, we focus on the impact of big data in the economic field. The potential for big data and Predictive Analytics to improve outcomes is tremendous. We discuss some of the latest (and most interesting) methods currently available for analyzing and utilizing big data when the objective is improved prediction. Our discussion includes a summary of various so-called dimension reduction and machine learning methods as well as a summary of recent tools that are useful for ranking prediction models associated with the implementation of these methods. We also provide a brief empirical illustration of big data in action, in which we show that the granularity and multidimensionality of big data offers advantages to economists in identifying economic trends when they occur (nowcasting), testing the behavioral theories of previously untested agents, and creating a set of tools for manipulating and analyzing this data. However, even though these new databases and statistical techniques open up many opportunities, they also pose many challenges for economists: access to these data, the ability to replicate them, and the development of technical skills to manipulate them. Also, closer collaboration between big data companies and researchers working on Big Data would be highly beneficial for the advancement of economic discipline.

Keywords: Big Data, Predictive Analytics, Economic Analysis, Prediction models

1. INTRODUCTION

Big data and predictive analytics are the process of collecting, organizing and analyzing large sets of data to discover patterns and other useful information. Big data analytics is typically performed using specialized software tools and applications for predictive analytics, data mining and data optimization Also Big data analytics can help organizations to better understand the information contained within the data and will also help identify the data that is most important to the economic decisions in the future. Analysts working with big data basically want the knowledge that comes from analyzing the data. The term of Big Data often refers simply to the

ISSN: 2455-8834

Volume: 04, Issue: 07 "July 2019"

use of predictive analytics or certain other advanced data analytics methods that extract value from data, and seldom to a particular size of data set. [1]

But beyond the debates on what this "new asset class" (as it is called by the World Economic Forum[2]) will be able to bring to the economy, one can wonder how these data can improve the way in which one analyzes the economic activity, and how the development of new methods of data analysis and predictive modeling developed in statistics and computer science can be useful in economic analysis.

- The place of economics at the intersection of the university and the applied sciences to the firm, as well as its important theoretical and methodological corpus make it an ideal candidate for the use more important and richer data while maintaining the robustness and representativeness that characterize this discipline. [3]

- Big Data applied to the economy corresponds to a radical change in the scale and extent of the resources (and the tools to manipulate them) available for the object of study; this definition differs from the more practical one used in the corporate world where the notions of "volume, variety and speed" of data help to create an advantage over the competition. [4]

Otherwise, you may use this as just an instruction set. It is remarked that you do not have to follow this style file when your works is submitted for the initial review stage.

2. GENERALITY ABOUT BIG DATA ANALYTICS

2.1 Characteristics and Specifications

In general rule, Specialists say that Big Data is characterized by 5 V and can be described by the following specifications: [5], [6]

Volume: The quantity of generated and stored data. The size of the data determines the value and potential insight- and whether it can actually be considered big data or not.

Variety: The type and nature of the data. This helps people who analyze it to effectively use the resulting insight.

Velocity: In this context, the speed at which the data is generated and processed to meet the demands and challenges that lie in the path of growth and development.

ISSN: 2455-8834

Volume: 04, Issue: 07 "July 2019"

Value: The most important element of the big data we call the Sage Blue Book is value. Value that includes a large volume and variety of data that is easy to access and delivers quality analytics that enables informed decisions.

IBM society has added another V called the veracity to ensure the accuracy of data and the quality of captured data can vary greatly, affecting accurate analysis, the Fig1 shows the 5VS of Big Data.

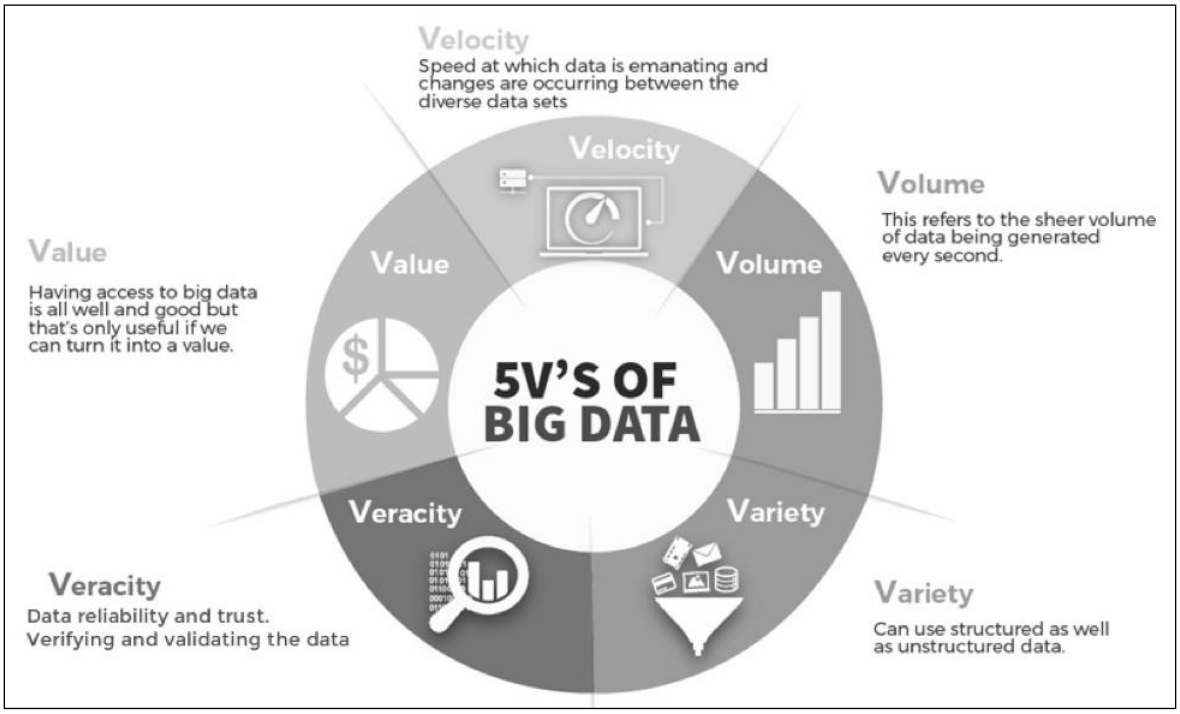

Fig. 1: BIG DATA illustration with 5VS.

2.2 Big Data in Economic Analysis

The application of Big Data in economic analysis could therefore be associated with notions:

A- "Multidimensionality": in terms of number of variables per observation, number of observations or both

B- "Granularity": Big data sets often provide useful micro-level data for analyzing agent behavior.

The advantages for research and economic policy are the following :

ISSN: 2455-8834

Volume: 04, Issue: 07 "July 2019"

1) Improve monitoring and forecasting of economic activity at the government level. Central and local public administrations collect vast amounts of administrative data at the micro level, in areas such as tax collection, social programs, education or demography, among others.

2) A level of periodicity and granularity often higher than traditional survey data. The use of new data to track private sector economic activity, sometimes even in real time (eg the MIT Billion Prices Project [7] which collects prices from several hundred online sales sites to obtain an accurate proxy for inflation, or the SpendingPulse [8] tool from Master Card that tracks household consumption via credit card payments), are powerful tools for monitoring economic activity with a level of frequency and granularity often higher than traditional survey data.

3) Proxies of economic indicators. Indirect measures such as online searches or social media publications can also be used as proxies for economic indicators such as employment or household confidence.[9] "Trends" on Google to "predict the present", suggesting that queries on Google for a specific product accurately reflect the demand for this product). The availability of "real-time" data can provide an advantage in terms of "nowcasting" or identifying economic trends as they unfold.

4) A significant amount of data that would contribute to a significant improvement in measurements. The gradual availability of large-scale administrative and private data could lead to better ways of measuring economic effects through broader and more granular data, particularly with respect to the behavior of individual agents [10]; the large size of the new databases could also solve the statistical problem of the limited number of observations and make the analysis more robust and accurate.

5) A better perception of the effects of different policies and economic shocks. These new data could encourage economists to ask new questions and research themes, in areas as diverse as labor market dynamics [11], the effects of preschool education on future earnings (Chetty et al., See below), stock market dynamics [12] and the functioning of online markets [13]. The ability to combine different databases expands the range of research, as shown for example in the Chetty, Friedman and Rockoff [14] study that combines administrative data on 2.5 million New York schoolchildren with their incomes as that adults 20 years later to show the "added value" of having benefited from a "good" teacher; in this case, the high level of granularity in the data makes it possible to link the individual scores to the school tests and the corresponding tax records for a large sample, which would have been impossible with aggregate data or a smaller sample. Many aspects of individual behaviors, such as social relationships (with data from social

ISSN: 2455-8834

Volume: 04, Issue: 07 "July 2019"

networks) or geolocation, may also become easier to observe and analyze; the example of Scott Keeter [15], from Pew Research Center, who advances the idea of using data collected on social networks as a supplement or substitute for public survey data, proves this idea.

6) Make possible "natural experiments". For example, switching from weekly data to significantly higher frequency data (up to minute per minute), or consumer or individual product data, can detect micro-level details or variations, would be more difficult to isolate and exploit with more aggregated data. The study by Einav, Farronato and Levin [16], which offers an analysis of pricing and sales strategies on the Internet, is a concrete example of the advantage of benefiting from granular data in order to obtain a rich information on the individuals studied and to explore a variety of consequences for a given experience (for example, substitution to other products in the case of a price change). These advantages are of particular interest when applied to the case of companies, and more particularly to online platforms for which it becomes increasingly simple and inexpensive to experiment when they have granular and personalized pricing strategies and automated methods always easier to capture (and apply) the results of these experiments.

7) New opportunities could also come from new statistical techniques and machine learning[17], which can help build more robust predictive models, particularly in the field of empirical microeconomics. Einav's study,

Jenkins and Levin[18] is an example of the use of Big Data techniques in predictive modeling to incorporate heterogeneity into their econometric model; In this study, the use of predictive modeling techniques allows the construction of "credit risk scores" that help researchers model consumer borrowing behaviors and how lenders should rate loans and set limits for credit risk. borrowing for the different types of borrowers divided according to their risk of default. Grasping heterogeneity through Big Data techniques and new statistical methodologies could also benefit many other sectors, because of the possibility of going beyond the "average effects" measure and being able to link measurable heterogeneity to specific treatment effects and specific policies; the example of the Safeway[19] food chain, which offers specific discounts to individual customers based on individual price elasticities, shows the progressive ability of companies to go beyond simple price elasticities and develop algorithms to estimate the elasticity and optimal prices specific to each type of consumer; similarly for governments in the development of their economic policies, with the possibility of developing more user-oriented policies (for example, health policies adjusted to the medical environment and the characteristics of the patient, education policies adjusted by level, teacher or mix of students, etc.).

ISSN: 2455-8834

Volume: 04, Issue: 07 "July 2019"

2.3 Big Data challenges for economists.

However, even though these new databases and statistical techniques open up many opportunities, they also pose many challenges for economists.

access to data: a large part of this new data on which researchers work belongs to companies (which aggregate them from their clientele), and the benefits for these companies of benefiting from researchers' knowledge of these data are not always comparable to the costs of disclosing the data.

The unstructured nature of the data, which represents a challenge in econometric terms - just to separate the dependencies between the series studied; this is the most important technical challenge with this type of data, which requires the development of new regression tools.

The need for economists brought to use this data to develop new skills - and more specifically at the level of advanced software and languages (SQL, R and Xlstat) as well as machine learning algorithms - in order to be able to combine the framework conceptual of economic research with the ability to apply ideas on massive databases; the highly publicized profession of "data scientist", which consists of analyzing data in order to find empirical models, is exactly at the crossroads of computer science and econometric analysis. The extraction and synthesis of the various variables and the search for relations between them will therefore become important parts of the work of economists and require new skills in computer science and databases.

As this article has described, we can think that the emergence of Big Data can change the landscape of research and economic policy quite strongly. However, this evolution can not be a substitute for economic theory; [20], the usual practice of economic forecasting (theory - simulation - calibration - prediction) can not be changed because "theory is needed to understand the mechanisms or at least to suggest what the 'we would hope to find in the first place'. Indeed, even if mass data is very useful for detecting correlations, including subtle correlations that a smaller database analysis might miss, it does not tell us which ones are relevant; similarly, the magnitude of the data can lead to "misleading" correlations between series that have nothing in common. To summarize, Big Data can not replace the theoretical research phase; indeed, no economic problem is solved via a simple "data crunching", and there is always need to understand beforehand the problem we are working on.

The ability of Big Data and associated statistical techniques to reduce large data sets to a unitary statistic is only the appearance of accuracy and therefore does not replace a thorough scientific analysis.

ISSN: 2455-8834

Volume: 04, Issue: 07 "July 2019"

The over-reliance on big data can even lead to perverse effects, as these databases often aggregate data in different ways and for different purposes; this risk is particularly significant for data collected from Internet searches, as shown by the example of Google Flu Trends, whose responsibility for data collection in relative failure was pointed out by the Harvard statistician Kaiser Fung [21]; too much reliance on web data has even resulted in what Marcus and Davis[22] call an "echo chamber effect", with the example of Google Translate results based on page translations Wikipedia ... and vice versa. Another danger comes from the growing difficulty in replicating data and research paper programs as data becomes more and more massive, as Barry Eichengreen [23] shows in Project Syndicate, which calls for greater focus on historical analysis of data. economic phenomena only on the development of more and more advanced statistical methods.

3. CONCLUSION

As described in this article, the benefits of using Big Data for economic analysis are legion. In terms of public policy recommendations and education, the "seam" of Big Data is fully in line with the exponential development of NICTs in everyday life and represents another argument for the development of computer science education, especially in university courses in economics and sociology; the recent integration of a "Big Data" module into the CFA® exam is only an illustration of this phenomenon[24]. The development of public laboratories focused on Big Data could also represent a solution to the lack of representativeness that this discipline suffers from researchers.

Similarly, a closer collaboration between researchers and companies with mass data would be beneficial to all actors and would, on the one hand, allow companies to benefit from external points of view and decision support. and economists to benefit from a "material" that can be used to develop new models and test new theories.

REFERENCES

[1] New Horizons for a Data-Driven Economy - Springer. doi:10.1007/978- 3-319-21569-3

[2] https://www.weforum.org/reports/personal-data-emergence-new-asset-class

[3] Taylor, Shroeder and Meyer http://journals.sagepub.com/doi/pdf/10.1177/2053951714536877

[4] Schroeder, http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199661992.001.0001/acprof- 9780199661992-chapter-11

ISSN: 2455-8834

Volume: 04, Issue: 07 "July 2019"

[5] K. Balar and A. Naji, A Model for Predicting Ischemic Stroke Using Data Mining Algorithms IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 11, November 2015.

[6] Hilbert, Martin. "Big Data for Development: A Review of Promises and Challenges. Development Policy Review.". martinhilbert.net. Retrieved 2015-10-07.

[7] http://www.thebillionpricesproject.com/

[8] https://www.mastercardadvisors.com/en-us/solutions/spendingpulse.html

[9] Choi and Varian's, http://onlinelibrary.wiley.com/doi/10.1111/j.1475-4932.2012.00809.x/full

[10] Brynjolffson's http://digital.mit.edu/bigdata/agenda/slides/Brynjolfsson%20Big%20Data%20MIT%20CDB%202012-12- 12.pdf

[11] Choi and Varian, http://onlinelibrary.wiley.com/doi/10.1111/j.1475-4932.2012.00809.x/full

[12] Moat et al. https://www.nature.com/articles/srep01801

[13] Einav et al. http://siepr.stanford.edu/sites/default/files/publications/10-033_Paper_Einav_10.pdf

[14] Friedman and Rockoff, http://pubs.aeaweb.org/doi/pdfplus/10.1257/aer.104.9.2633

[15] Scott Keeter, http://www.pewresearch.org/2012/05/24/survey-research-its-new-frontiers-and-democracy/

[16] Einav, Farronato and Levin, https://web.stanford.edu/~leinav/pubs/AR2016.pdf

[17] Einav, Les lecteurs intéressés par ces nouvelles techniques pourront consulter le papier de Hal Varian, “Big Data: New Tricks for Econometrics” (http://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.28.2.3),

[18] Jenkins and Levin, http://web.stanford.edu/~leinav/pubs/ECMA2012.pdf

[19] https://retailleader.com/pulling-more-meaning-big-data

ISSN: 2455-8834

Volume: 04, Issue: 07 "July 2019"

[20] Sascha Becker, http://journals.sagepub.com/doi/full/10.1177/2053951714536877

[21] Kaiser Fung, https://hbr.org/2014/03/google-flu-trends-failure-shows-good-data-big-data

[22] Marcus and Davis, https://www.nytimes.com/2014/04/07/opinion/eight-no-nine-problems-with-big-data.html

[23] Barry Eichengreen, https://www.theguardian.com/business/economics-blog/2013/may/17/economic-big-data-rogoff-reinhart

[24] https://www.bloomberg.com/news/articles/2017-05-09/cfa-exam-to-include-big-data-artificial-intelligence-as-topics

