

Big Data

Subjects: Others

Contributor: Paulo Ferreira

Big data has become a very frequent research topic, due to the increase in data availability. Here we make the linkage between the use of big data and Econophysics, a research field which uses a large amount of data and deals with complex systems.

Keywords: big data ; complexity ; networks ; stock markets

1. Introduction

Big data has become a very popular expression in recent years, related to the advance of technology which allows, on the one hand, the recovery of a great amount of data, and on the other hand, the analysis of that data, benefiting from the increasing computational capacity of devices. Big data has been used in several research areas such as business intelligence ([Chen et al. 2012](#); [Sun et al. 2018](#)), marketing ([Verhoef et al. 2015](#); [Wright et al. 2019](#)), economics ([Glaeser et al. 2018](#); [Sobolevsky et al. 2017](#)), health ([Pramanik et al. 2017](#); [Rose et al. 2019](#)), and psychology ([Matz and Netzer 2017](#); [Adjerid and Kelley 2018](#)), among many other areas and studies which could be mentioned.

Another area where big data is being applied is finance. In this particular case, the existence of large amounts of data allows a very broad type of analysis, from general indices to single specific assets. In particular, the use of big data allows the analysis of complex problems and has attracted the attention of physicists in recent decades. In fact, big data and complexity are intimately related to the emergence of a new research area called Econophysics.

2. The Use of Big Data in Economics and Finance

The use of big data allows the analysis of (very) large datasets, reaching conclusions for some processes which could involve complex analysis. With the growth of web access in recent decades, the amount of data available has increased exponentially. This large amount of available data and its use has influenced society, communication, habits, and even cultural aspects, so understanding, interpreting, and knowing how to use this amount of data has been a challenge for data scientists. In this context, big data was initially described by four different Vs: Velocity, Volume, Veracity, and Variety, as follows.

- Velocity, referring to the speed with which the data is produced, with this parameter being related to processing capacity.
- Volume, related to the amount of data to be processed.
- Veracity, referring to data feasibility.
- Variety, characterized by the heterogeneity, diversity, structures, and scales of the data.

This original V model of the big data paradigm is now added to with other data characteristics. For example, volatility, referring to the validity of the data or value referring to the potential use of the data (see, for example, [Tennant et al. 2017](#)). Other features are found in the literature, both using the initial V letter and others, like the Ps proposed by [Lupton \(2015\)](#), who just asks about the importance of big data in society in general and in the academic world in particular (see, for example, [Kitchin and McArdle 2016](#) for more features). The objective of identifying big data in V was to outline the existence of different interconnected and measurable properties ([Carbone et al. 2016](#)).

The research area of finance is very sensitive to the use of this kind of approach, as there is a huge amount of data available, with some being publicly disclosed and more available in databases via subscription. Platforms such as Google, Twitter, and others give rise to the need for effective verification in terms of whether the large amount of information available on the web can help to analyze and predict financial variables ([Preis et al. 2013](#)). Verifying whether or not this large amount of information influences financial markets requires tools from complex systems ([Arthur et al. 1998](#); [Rosser](#)

1999; Carbone et al. 2016). In this regard, the use of models, methods, and techniques available from the physics of complex systems, such as multifractal analysis, multiscalar analysis, temporal networks, or multilevel networks, have been useful in recent research and can also be useful for future research.

Another application involving big data has been developed by The Observatory of Economic Complexity ([Hausmann et al. 2014](#)), derived from the idea of the Economic Complexity Index (ECI; [Hidalgo and Hausmann 2009](#)). This index provides extensive information about more than 130 countries, making it possible to analyze the role of knowledge in the production of goods for countries' economic growth and development. Since [Solow \(1956\)](#), the theory of endogenous growth has contributed to the understanding of how human capital influences economic growth. However, it is also necessary to consider whether human capital is incorporated into a country's productive system by providing the production of goods with relative complexity and diversity. Thus, [Hidalgo and Hausmann \(2009\)](#) and [Hausmann et al. \(2014\)](#) created the ECI, which considers both the level of knowledge used to export goods and the diversity of production of those goods. A country has high economic complexity if the goods produced incorporate a high level of knowledge and productive diversity. One such example is Japan, which, according to ECI, had the most complex economy in the world between 2011 and 2016 due to the complexity and variety of products exported. In contrast in Botswana, which ranks 129th according to the ECI, 52% of exports comprise only refined copper, that is, a product that requires little knowledge for manufacture.

3. From Big Data to Econophysics

As previously mentioned, finance is one of the areas showing increased work which could be considered as using big data, and has attracted researchers from other research fields, such as physicists, even creating multidisciplinary research fields such as econophysics.

Econophysics is a neologism used in the branch of Complex Systems from Physics seeking to make a complete survey of the statistical properties of financial markets, using the immense volume of available data and the methodologies of statistical physics ([Mantegna and Stanley 1999](#)). The term Econophysics was coined by [Stanley et al. \(1996\)](#) when they analyzed the Dow Jones Index and found that stock returns followed a power law distribution, contributing to the emergence of this new research field. Although use of the neologism is relatively recent, the approximation between physics and finance is not new, beginning probably in the 1960s, when [Mandelbrot \(1963\)](#), analyzing the returns of cotton prices, refutes the condition of normal asset prices ([Jovanovic and Schinckus 2013, 2017](#)). For [Mirowski \(1990, 1991\)](#), neoclassical economics had a strong influence on theoretical physics, contributing to economic theory throughout the 20th century. Mandelbrot's ideas about the non-normality of financial returns remained forgotten, until [Mantegna \(1991\)](#), analyzing the Italian stock market, discovered that returns were compatible with Lévy stable non-Gaussian distributions.

Since its emergence, Econophysics has made a significant contribution to several issues mainly in finance: autocorrelations, long-range dependence, and non-normal asymptotic distributions ([Mantegna and Stanley 1999](#)); assessment of financial risk and asset pricing ([Bouchaud and Potters 2003](#)); the prediction of crises and market crashes ([Sornette 2017](#)); and agent-based modeling ([Farmer and Foley 2009](#); [Lux and Marchesi 1999](#)), systematic risk, and networks ([Battiston et al. 2012a](#); [Tabak et al. 2014](#); [Wang et al. 2017](#); [Wang et al. 2018](#)).

One of the hot topics of Econophysics, where the connection to financial theory is closer, is analysis of the efficient market hypothesis (EMH). Although the EMH was only formalized by [Fama \(1970\)](#), this hypothesis implying that the market reflects all the available information on financial assets' prices, it was formally analyzed by [Bachelier \(1900\)](#) and [Samuelson \(1965\)](#), among others, before the definition of Econophysics. Since [Fama \(1970\)](#), this topic has been studied greatly in the literature (see [Lee 2008](#); [Titan 2015](#) for reviews). On the basis of the EMH, asset prices may be described by a random walk with this possibility being analyzed through the Hurst exponent, which is a very well-known approach in Econophysics. Methodologies like the rescaled range (R/S) analysis of Detrended Fluctuation Analysis (DFA) are widely used to make this estimation and are found in several studies, such as [Costa and Vasconcelos \(2003\)](#), [Di Matteo et al. \(2005\)](#), [Wang et al. \(2011a, 2011b\)](#), [López and Contreras \(2013\)](#), [Kristoufek and Vosvrda \(2013\)](#), [Cao and Zhang \(2015\)](#), [Anagnostidis et al. \(2016\)](#), [Urquhart \(2016\)](#), [Nadarajah and Chu \(2017\)](#), or [Ferreira et al. \(2017\)](#), among many others.

Regardless of the features identified in the previous section, it is important to understand that big data can be used for commercial and financial purposes. In different areas, firms can use their data with the objective of raising their returns (see, for example, [Subrahmanyam 2019](#)). Obviously, thanks to increased computational capacity, the financial sector can also use Econophysics models to identify, test, and evaluate models which could be used, for example, to identify patterns in financial (big) data (see, for example, [Preis et al. 2010, 2012, 2013](#), among many others which use big data trying to anticipate possible warning signs in financial markets).

Regarded with some distrust in mainstream economics at the beginning (see, for example, [Ball 2006](#); [Schinckus 2018](#)), this may now be changing, with some economists adopting methods originating in the field of Econophysics and increasing recognition of its wide applicability in several areas of the economy. Thus, Econophysics is slowly being assured its place. Initially, applications were restricted to financial markets and in rare cases macroeconomics. Currently, Econophysics is used in several areas of the economy, such as energy ([Filip et al. 2016](#)), regional economics ([Gao and Zhou 2018](#)), or environmental economics ([Stolbova et al. 2018](#)), among many others, demonstrating that this discipline is attaining greater importance in solving various economic problems. In complexity, an area that has gained prominence is complex networks.

4. Conclusions

Big data and complexity go hand in hand and Econophysics is a way to use this kind of data, with the particularity of being used mainly in finance. The availability of large datasets, jointly with increased computational processing, made big data and finance very attractive research areas in general, of particular interest in the analysis of crisis events.

Since the global financial crisis, governments around the world have been acting to improve financial stability and reduce the risks of a highly interconnected financial system, using complexity to do so ([Yellen 2013](#)). That said, complexity has been providing methods and models that try to explain the instabilities occurring in different markets, leaving five important lessons for financial markets:

- Extreme events can occur in stock markets.
- The financial markets are interconnected.
- Some sectors or companies are too central to fail.
- Different systems, for example, public health, transport, industry, and finance, are interrelated, increasing the global risk.
- Financial crises can be complex phenomena, when markets are in a transition phase, when any “shock” can cause a crisis and the consequent contagion effect.