MULTIVARIATE TIME SERIES SIMILARITY-BASED COMPLEX NETWORK IN STOCKS MARKET ANALYSIS: CASE OF NYSE DURING GLOBAL CRISIS 2008

Maman Abdurachman Djauhari
Professor of Statistics
Institute for Mathematical Research (INSPEM)
Universiti Putra Malaysia
43400 UPM Serdang, Selangor Darul Ehsan
Email: maman_abd@upm.edu.my

Abstract

Long before we started with the 21st millennium, Stephen Hawking saw the current millennium as the millennium of complex systems. Until present, he was right due to the fast growing technology in computer. Nowadays, in the era of digital world where big data is our daily menu, we cannot escape from complex systems. As big data is characterized by "4V" (Variety, Velocity, Veracity and Volume), statistics such as practiced in traditional way is not enough and sometime is not apt to be used to understand the most important information contained in big data. What people call now data analytics needs to be used as the only complementary. It is mathematically dominated by multivariate data analysis (MVDA) in the French way. Traditional statistics, which is based on mathematical statistics, is to do confirmatory analysis while data analytics is to do exploratory analysis. The former is to do hypothesis testing (micro analysis) and the latter is for hypothesis generation (macro analysis). Macro analysis is more appropriate to deal with big data. The principal mathematical tool to do macro analysis is MVDA in the French way where big data is considered as a complex system. In this regards, the main problem is to define the similarity among objects of the study such as stocks, economic sectors, currencies, and other commodities in financial industry, which are statistically a multivariate time series. Furthermore, the principal tools to filter the important information contained in a complex system are complex network and social network analysis. To demonstrate the advantages of complex network approach in stocks market analysis, in this paper the behaviour of economic sectors played in NYSE during global crisis in 2008 will be presented and discussed. By nature, all stocks are a multivariate time series. Therefore, in that example, we show that the use of Pearson correlation coefficient is useless to define the similarity among them. We use Escoufier's vector correlation coefficient instead.
