International Journal of Applied Social Science
Volume 9 (11 & 12), November & December (2022) : 544-549
Received : 30.10.2022; Revised : 14.11.2022; Accepted : 29.11.2022

RESEARCH ARTICLE
ISSN : 2394-1405
DOI: 10.36537/IJASS/9.11&12/544-549

Exploring Nonlinear Time Series Cycle Analysis in the Context of Big Data with Practical Applications

PARVEEN KAUR
Assistant Professor
Department of Computer Science, Guru Nanak College, Ferozepur Cantt (Punjab) India

ABSTRACT

Over the last several decades, nonlinear time series analysis methods have evolved beyond their time series roots to include a variety of approaches that try to close a gap in the modeling and forecasting of certain kinds of data sets, such as chaotic determinate systems. These systems may be found in a wide range of both the natural and human domains. This paper describes some of the predictive methodologies as well as the historical context in which these approaches evolved and their underlying principles. The purpose of this research is to provide some light on their potential and processes. This work presents a different method for analyzing time series data, where the size of the time series is used to study the data generation process. Here, we talk about the model's order-dependent convergence. To demonstrate the convergence of the model, an empirical study is performed on the time series of the sectoral indices of the National Stock Exchange (India). This convergence also supports the idea that we get a constant model when we fit a series beyond a certain size. This demonstrates that the genuine model correctly predicted the future value and that data is inherently convergent. According to the research, the realization of large data time series may take the series' size into account, therefore the DGPs that are discovered may be the most appropriate intervals for the relevant study period.

Key Words : Time series, Big Data, Velocity Big Data, ARIMA model, GARCH model

INTRODUCTION

Data of unprecedented complexity is being produced by AI, ushering in a new era of big data. Data that is very large, very fast, or very diverse presents significant hurdles for researchers from different disciplines who must work together. Machine learning, data mining, measurements, regular language processing, and text processing are just not many of the high level multidisciplinary data investigation apparatuses that assistance to catch the underlying designs within otherwise turbulent data points. As a result, we are able to take advantage of the unstructured datasets to learn quickly and make more informed judgments in less time. There is a lot of room for innovation in data analysis tools within these procedures. Complex networks and complexity theory are often mentioned in the same breath as these techniques when discussing their application to dynamical systems and statistical physics.

Attention this study, we zero attention on a sector of nonlinear time series analysis that has seen a lot of recent activity: the incorporation of complex network theory methods. The observations in a time series are made at regular intervals, giving rise to a time-indexed sequence of data points. This means that there is an inherent chronological discreteness to time series data. Time series examples span a wide range of fields, many of which may have practical applications in people's daily lives. This includes (i) data gathered from meteorological stations or satellites, like surface-air temperatures, ocean level strain, and wind speeds; (ii) data gathered from the financial business sectors, for example, the everyday closing costs of financial exchange indices like the Dow Jones Industrial Normal, individual resources, or trade

How to cite this Article: Kaur, Parveen (2022). Exploring Nonlinear Time Series Cycle Analysis in the Context of Big Data with Practical Applications. *Internat. J. Appl. Soc. Sci.,* **9** (11 &12) : 544-549.

rates; and (iii) data gathered from bio-clinical monitoring frameworks, like electroencephalogram (EEG) gadgets or high-goal brain imaging to survey human physiological and clinical circumstances. Instead of looking at individual numerical values at various temporal occurrences, time series analysis looks at the full collection of data as a whole.

Not the same as data analysis where there is no regular ordering of the perceptions, (for example, cross-sectional investigations of explaining individuals' wages by reference to their particular training levels or spatial data analysis where house costs are represented by area notwithstanding the intrinsic qualities of the houses), time series analysis exploits the normal fleeting ordering of the perceptions. Data mining calculations have been recommended in the field of software engineering examination to uncover stowed away examples in such more expansive gigantic data sets from numerous sources, and these devices have additionally found many purposes in the context of time series mining. Indexing, clustering, arrangement, division, theme ID, and forecasting are the main goals of time series mining in this context. The approach of big data and distributed computing has started a whirlwind of action in the formation of data mining devices, which in turn reflects the growing quantity and complexity of accessible datasets over the past decade. To gain from and produce expectations on the big data sets, headways in administered and solo learning calculation configuration are one model. Inaddition; there is a growing movement towards integrating complicated network methods with data mining tools, which offer a wealth of cutting-edge research ideas for unearthing previously unseen patterns in massive datasets. By selecting feature vectors of smaller dimension, the grouping issue of data mining empowers a rich portrayal of a few muddled frameworks, for example, a meaningful reproduction of practical organizations from exceptionally enormous data sets. Therefore, using highlight choice strategies extends our insight into the elements of organization models. Mining innovations have been effectively used to complex organization analysis using both synthetic and exploratory data, most notably in the field of illness categorization.

This examination centers around nonlinear time series analysis using complex organization draws near, a subject that has gotten surprisingly little consideration notwithstanding the conspicuous practical significance of time series mining calculations. While regular data mining strategies are grounded on static models, the time series network approaches displayed here are cutting-edge instances of nonlinear time series analysis. According to the point of view of perplexing organization research, the issues examined in this paper are instances of how network theory has been effectively applied to the investigation of synthetic and exploratory series from an assortment of utilization domains. Any place it is applicable, we will discuss potential speculations of the different time series network approaches in the context of data mining strategies.

Literature Review:

Zou *et al.* (2019) the use of complex network approaches for the characterization of time-series-based dynamical systems has been the subject of increasing amounts of published work over the past decade. Albeit both nonlinear time series analysis and complex organization theory are deep rooted areas of mind boggling frameworks sciences with binds to nonlinear elements and measurable physical science, nonlinear time series analysis has emerged as an active field in its own right due to the fruitful combination of the two approaches, which has allowed researchers to successfully treat a wide range of problems and answer fundamental questions about the structural organization of nonlinear dynamics. In this study, we examine the history of time series networks and present a comprehensive overview of the most up-to-date methods, interpretations, and practical issues. We begin by providing a short outline of the present status of nonlinear time series analysis and the theory of perplexing organizations, and afterward we focus in on three main organization draws near: stage space based repeat organizations, perceivability diagrams, and Markov chain based progress organizations. Specific aspects, potentials, and limits of these three ideas, as well as various versions thereof, will be examined at length. In addition, we highlight the novel, basic insights that complex network techniques provide to nonlinear time series analysis. We likewise give a concise synopsis of ongoing applications of these techniques across a wide assortment of fields, including climatology, liquid elements, neurophysiology, engineering, and financial matters, highlighting the extraordinary possibilities of time series networks in solving pressing, modern scientific problems. The purpose of this paper is to teach its audience how to implement complicated network methods in real-world time series analysis.

Runge *et al.* (2019) Disciplines that concentrate on complex dynamical frameworks, similar to the Earth framework or the human body, has the test of identifying causal linkages and estimating their solidarity using observational time series data. Data-driven causal inference in such frameworks is troublesome because of the great dimensionality, nonlinearity, and little example measures often connected with data sets of this sort. To gauge causal organizations from enormous scope time series datasets, we give an extraordinary methodology that might combine linear or nonlinear restrictive independence tests with a causal revelation strategy. We test the method on true time series from the environment framework and the human heart, as well as for enormous scope synthetic datasets intended to reproduce the typical attributes of genuine data. Tests show that our methodology beats cutting edge approaches in identification power, providing novel chances to uncover and evaluate causal organizations from time series in a wide assortment of fields of review.

Donner *et al.* (2010) to more readily comprehend the primary qualities of time series from complex frameworks, a clever technique is introduced in this work. The repeat lattice of a time series is viewed as the nearness framework of a connected complex organization that associates distinct minutes in time assuming the states viable are spatial neighbors. The original strategy might be viewed as a unifying structure for translating time series into complex organizations, with other current methodologies as unique instances, and it offers critical calculated benefits when contrasted with similar organization-based procedures. Here, we show that the nontrivial measurable highlights of the stage space thickness of the underlying dynamical framework are intimately associated with a wide assortment of topological parts of repeat organizations. Thus, this new point of view on the repeat framework produces novel quantitative qualities (like normal way length, clustering coefficient, or centrality proportions of the repeat organization) connected with the dynamical intricacy of a time series, most of which are not yet given by other strategies for nonlinear time series analysis.

Donges *et al.* (2015) to utilize and combine cutting-edge data analysis and modeling systems from complex organization theory and nonlinear time series analysis, we give the pyunicorn (Pythonic bound together complex organization and repeat analysis tool stash) open source software bundle. Python's pyunicorn library is profoundly parallelizable and highlights total item direction. It prepares for the improvement of environment networks in climatology or practical brain networks in neuroscience, the two of which address the construction of factual interrelationships in enormous data sets of time series and which can then be investigated using cutting-edge methods from the field of perplexing organization theory, including measures and models for spatial organizations, organizations of interacting networks, hub weighted measurements, and organization proxies. What's more, pyunicorn's repeat evaluation analysis, repeat organizations, perceivability charts, and development of proxy time series shed light on the nonlinear elements of mind-boggling frameworks as kept in uni-and multivariate time series from a clever point. Several examples, mostly from the climate science industry, are used to illustrate the library's versatility.

Laarne *et al.* (2022). Exploring and comprehending multivariate datasets can be difficult due to the non-linear nature of the relationships between atmospheric variables. Here, we detail a mutual information method for identifying the strongest connections in this environment. After some light data quality testing, this approach is able to discover linear and non-linear relationships with high confidence. There is no need for rigidly specified models to account for contextual variation and seasonality. We provide two examples of this strategy in action. The first is an example of depersonalization of a straightforward time series, with outcomes that are consistent with the traditional approach. The second one looks for patterns among several variables, some of which have lognormal distributions (trace gas concentrations) and some of which have circular distributions (wind direction), in a bigger dataset. Our 'ennemi' Python module is used in the examples.

METHODOLOGY

In the context of defining Big Data, velocity is crucial since it expands the size of time series. Concerning the practical issue of high dimensionality, Fan et al. have made the following observations: (1) as the dimensionality grows, the corresponding heavy computing cost also increases; (2) it is not necessary to have the analysis of whole data; and (3) it is not equally contributing. Researchers have been concerned with enhancing analytical methodologies and decreasing processing time since the beginning of conversation about big data analytics. Consider the following questions in the context of analytics:

– Is it true that larger time series tend to yield more reliable results when applied to a given model?

– Does increasing the number of observations in the sample series improve the model's performance?

– Can we gather better data on the system or thing we're studying, if at all possible?

Since market indices play such a crucial part in determining a country's economic standing, every nation has its own system of official declaration for these numbers. Indexes are announced by several exchanges throughout the world, including the BSE and NSE in India, the NASDAQ OMX and NYSE EURONEXT in the United States, the TMX GROUP in Canada, the Shanghai and Shenzhen stock exchanges in China, the London Stock Exchange Group in the United Kingdom, etc. . The National Stock Exchange (NSE) is the primary venue for trading on India's derivatives market, and it has declared sixteen indices, including broad market, sectoral, thematic, and strategy indices, that correspond to various market segments and corporate profiles. Auto, Bank, Energy, FS, FMGC, IT, Metal, PSU, etc., are all examples of sectoral indexes.

DATA ANALYSIS

Eight indexes were utilized, with data ranging from January 1994 to January 2017 from the National Stock Exchange (NSE) website, to conduct an empirical investigation of a suggested technique for Big Data time series modelling. Since they are reported every day, their rate of growth may be measured relative to the amount added each day. We may assume that in this process, addition occurs in the parent series less frequently than once every fraction of a second. When velocities are high, traditional statistical methods are inadequate for analysing the resulting time series. We are realising the series separately taking into account a large data time series using conventional computing equipment, while the series grows by one new observation each day, representing the pace of large data. Although we are interested in training the approach for analysis, practically speaking advanced data come in the situation in the large numbers or MBs. We are modelling the series after adding the data at a monthly interval;therefore there is room for debate about the relative subjectivity of the word "Big" in this context.

$y_t = (\log p_t - \log p_{t-1}) * 100$

Time series modelling software and high-performance computers have made it possible to apply these models to an increasing variety of forecasting problems. Stationarity was checked using the Augmented Dickey Fuller (ADF) test, and the resulting test measurements values are recorded in Table 1 with enlightening insights. The Box Jenkins method is used to create the optimal ARIMA model.

Table 1 : Sectoral Index of the New York Stock Exchange Log Return

Series	Mean	SD	Skewness	ADF-test
MARUTI	1.1346	2.347	-1.3454	-16.4454
POWER	1.1437	2.369	-1.3767	-16.4117
FMCG	1.1756	2.437	-1.7367	-15.3356
IT	1.1437	2.596	-51.345	-14.4506
ENERGY	1.1437	2.470	-1.4558	-13.4158
COAL	1.1276	2.325	-1.4748	-16.1506
METAL	1.1436	2.704	1.2447	-13.7688
PSU	1.1142	2.470	-1.2158	-12.6119

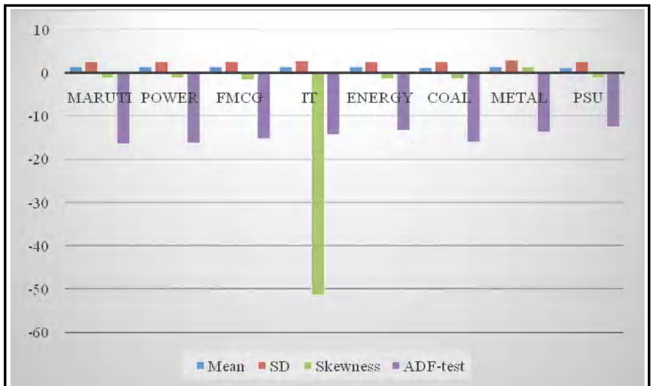

Fig. 1 : Graphical representation of Sectoral Index of the New York Stock Exchange Log Return

Table 1 shows that at the 5% degree of importance, the test measurements esteem is critical. When contrasted with other industries, the IT area's return series has a negative mean and a critical standard deviation. The worth of Skewness of the IT area isn't near nothing (- 42.236), subsequently all series are around typically circulated with the exception of the IT area bring series back. The stationarity of the return series will be tested by rejecting the null hypothesis of a unit root or non-stationary behavior. We use the ADF test to confirm that all sectorial return series are indeed stationary. In the wake of adding the one-month data and recorded request of the best fitted model, we are fitting the series in chronological order, which is getting huger in real time. Model convergence is achieved when there is no change in the series' order. This is a critical stage all the while, since it is as of now that we decide to cease fitting the series once we have confirmed that the model order has not changed. Because of its flexibility to account for trends, seasonality, cycles, error, and non-stationarity, we employed the ARIMA model in our empirical research. The R-software's forecast package allows us to obtain the optimal ARIMA model. After accumulating the original series' monthly observations, a cumulative series will emerge. Cumulative series fit best by ARIMA model, and ARCH/GARCH effect tested in residual series.

Conclusion:

As system-level phenomena, convergence is something we would look for in a time-series model. If new data cannot change the overall direction of the series, or if they add nothing to the model's current parameter values, then we may say that the series is stable. Altering the model's standard error or information criteria can provide different results, but that might not matter much in terms of the overall goal of the modeling exercise. Large data sets and extended time periods are no problem for Big Data analysis tools since they don't need a scalability check. This work, rather than focusing on the ideal number of observations or sample size, examines the role that size plays in time series analysis. The methodology used to collect the NSE data is defended. If the model converges, we may be assured that our predictions will fall inside the predicted interval; otherwise, we'll have to keep looking for a convergent one. Extra costs associated with data recording, analytics, and hardware/software tools may be incurred if series sizes are not properly managed. Additional research into the generalized model considering the other models and other related factors is possible.

REFERENCES

Ana Beatriz Galvao (2000). Univariate Non-Linear Time Series Models and the Business Cycle Stylised Facts.

Andreas Seyfang, Silvia Miksch, Werner Horn, Michael S. Urschitz, Christian Popow and Christian F. (2001). Poets Using Time-Oriented Data Abstraction Methods to Optimize Oxygen Supply for Neonates, Lecture Notes in Computer Science.

Ben Goertzel (1999). Delay Vectors as Perceptual Chunks: Understanding Nonlinear Time Series Analysis as Self Organising Cognition Dynamical Psychology. *An Internat., Interdisciplinary J. Complex Mental Processes*.

Donges, J. F., Heitzig, J., Beronov, B., Wiedermann, M., Runge, J., Feng, Q. Y., ... and Kurths, J. (2015). Unified functional network and nonlinear time series analysis for complex systems science: The pyunicorn package. *Chaos: An Interdisciplinary J. Nonlinear Science*, **25**(11).

Donner, R.V., Zou, Y., Donges, J. F., Marwan, N., &Kurths, J. (2010). Recurrence networks—A novel paradigm for nonlinear time series analysis. *New Journal of Physics*, **12**(3), 033025.

Hegger, R., Kantz, H. and Schreiber, T. (1998). Practical Implementation of nonlinear time series methods: The TISEAN package, CHAOS 9, 413, Winter School on Nonlinear Time Series Analysis.

Hoekstra, B.P.T., Diks, C.G.H., Allessie, M.A. and DeGoede, J. (2001). Non Linear Time Series Analysis: Methods and Applications to Atrial Fibrillation, Analysis and Processing of Cardiac Electrograms in Atrial Fibrillation, Annalidell' Istituto Superioredi Sanita', Roma, Ref Ann. 1st Super. *Sanita',* **37** (3).

Holger Kantz (2001). Time Series Analysis in Reconstructed State Spaces, Stochastics and Dynamics, **1** (1).

Kantz, H. and Scheiber, T. (1997). Nonlinear Time Series Analysis, Cambridge University Press.

Kulkarni, D.R., Pandya, A.S. and Parikh, J.C. (1997). Dynamic Predictions from Time Series Data — An Artificial Neural Network Approach, International Journal of Modern Physics C, in press.

Kulkarni, D.R., Pandya, A.S. and Parikh, J.C. (1998). Modeling and Predicting Sunspot Activity — State Space Reconstruction + Artificial Neural Network Methods, Geo Phys Lett., **25**, 457.

Laarne, P., Amnell, E., Zaidan, M. A., Mikkonen, S. and Nieminen, T. (2022). Exploring Non-Linear Dependencies in Atmospheric Data with Mutual Information. *Atmosphere*, **13**(7) : 1046.

Magni, P., Bellazi, R. and Locatelli, F. (1998). Using uncertainty management techniques in medical therapy planning: a decision theoretic approach. In Applications of Uncertainty Formalisms, Springer-Verlag, Berlin.

Pakalapati, Rajani (2000). Network Traffic Analysis: Identifying Invariant Features in Ethernet Traffic, M.Sc. Thesis, Dept of Elec. Eng. Univ of Massachusetts Lowell.

Price, C.P., Prichard, D. and Hogenson, E.A., Do the Sunspot Numbers Form a "Chaotic" Set? Available from citeseer.ist.psu.edu/price92do.html.

Runge, J., Nowack, P., Kretschmer, M., Flaxman, S. and Sejdinovic, D. (2019). Detecting and quantifying causal associations in large nonlinear time series datasets.

Science Advances, **5**(11), eaau4996.

Scott Armstrong, J. (2001). Standards and Practices for Forecasting taken from Principles of Forecasting: A handbook for Researchers and Practitioners, Kluwer Academic Publishers.

Sello Stefano (1999). Time Series Forecasting: A Multivariate Stochastic Approach, Topic Note Nr. NSG/260199, Los Alamos National Laboratories Preprint Archive, Physics/ 9901050.

Sello Stefano, Time Series Forecasting: A nonlinear Dynamics Approach, Los Alamos National Laboratories, preprint archive Physics, 9906035, or Thermo Fluid Dynamics Research Centre Enel Research USG/180699.

Zou, Y., Donner, R. V., Marwan, N., Donges, J. F. and Kurths, J. (2019). Complex network approaches to nonlinear time series analysis. *Physics Reports*, **787**, 1-97.

****