Journal of Mathematics and Statistics Studies
ISSN: 2709-4200
DOI: 10.32996/jmss
Journal Homepage: www.al-kindipublisher.com/index.php/jmss

JMSS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Time Series Analysis and Some Applications in Medical Research

Saadettin Aydin
Health Sciences University, Gülhane Faculty of Medicine, Turkey
Corresponding Author: Saadettin Aydin, **E-mail**: saadettin.aydin@sbu.edu.tr

| ABSTRACT

Time series have applications in many fields. Time Series Analysis is used for many applications, such as economic forecasting, sales forecasting, budgetary analysis, stock market analysis, yield projections, process and quality control, etc. In this study, we examined the applications of time series in the field of health. We gave some examples of applications in the field of health. We mentioned that many more studies could be done in the field of health with time series analysis. In this paper, we reviewed recent works and present a summary and a brief theoretical introduction to time series ARIMA modeling. We have seen that this model is the most widely used method in research in the field of health. As a result, time series analysis, a statistical technique that may be appropriate when data are measured repeatedly and at nearly equal intervals of time, has potential research applications in the study of chronic diseases such as diabetes, hypertension, and herpes simplex. We believe that this study will guide researchers working in this field in their research.

| KEYWORDS

Time series, time series analysis, ARIMA method, machine learning.

| ARTICLE INFORMATION

ACCEPTED: 20 October 2022 **PUBLISHED:** 30 October 2022 **DOI:** 10.32996/jmss.2022.3.2.3

1. Introduction

For years, time series has been used mostly in the field of economics. Time series techniques have been very successful in making many forecasts in the field of economics. The reason that pushed us to this study is, are time series used in studies in the field of medicine? If so, what has been done so far? To find examples of these studies and to introduce the time series method to those who will do research in the field of medicine. In this study, first of all, what the time series is and what it is used for is explained. As long as we accumulate data, "time" remains a very important factor. Essentially, time is an important element of data in time series analysis. A time series is a collection of time-ordered observations of a particular process for which the time interval between each observation is constant, e.g., weeks, months, and years; in some cases, small deviations in time intervals are acceptable. Time series data helps to investigate how an asset, security, or economic variable has changed over time.

The data of a time series are under the influence of various factors, and they record some fluctuations depending on the effects of the factors in different directions and intensities. In time series models, it is assumed that these fluctuations are due to the simultaneous and combined effects of four different types of movements. These movements, called components (elements) of the time series:

• Trend: Long cycle trend; It is the tendency to increase, decrease or remain unchanged in the value of the variable in the long run due to structural reasons affecting the variable. The trend is the factor that reveals the character of the time series.

• Cyclical fluctuations: These are the constantly recurring fluctuations in the value of the variable, observed at equal periodic intervals, occurring with the economic conjuncture. The value reaches a maximum with an improvement, then regresses and reaches a minimum, then increases again and reaches a maximum. These fluctuations are repeated periodically.

• Seasonal fluctuations: These are fluctuations that occur with natural factors such as climate or socio-economic effects. They are repeated in the same or similar way each time within a certain period of time.

• Incidental – random movements: These are irregular fluctuations that do not occur continuously. They can be caused by natural events such as earthquakes or social events such as wars. They are difficult to predict.

According to the way the observation values are obtained, time series, continuous and discrete series; Stationary and non-stationary series are examined according to whether the observation values show large deviations from the mean value of the series and, finally, seasonal or non-seasonal according to the cyclical movements they have shown. Stationary time series are a trend, conjuncture, and seasonally adjusted series. For this reason, it is possible to observe the clear movements of the series in this type of series.

Time series analysis has been a known method for many years. Box and Jenkins provided a method for constructing time series models in practice [Box, 1976; Jenkins, 1968]. Their methods are often referred to as the Box-Jenkins approach and autoregressive integrated moving average models (ARIMA). The method is based on the use of the autocorrelation structure of time series. It can be used if there is autocorrelation. This method was initially applied in fields such as industry and economics and later in medical research and fields such as [Helfenstein et al. 1991]. The Box-Jenkins method, as a univariate model, is one of the methods for predicting the future. It is an important assumption of this method that the series in which this method is applied, which is very successful in short-term estimation, is a discrete and stationary series consisting of observation values obtained at equal time intervals. The concept of stationarity in such a series is one of the important assumptions of the Box-Jenkins method. Models included in the Box-Jenkins method have been developed based on the assumption that time-dependent random events and time series related to these events are stochastic processes. In addition, internal dependence is taken into account in the most effective way. Because of these features, Box-Jenkins models are also called linear stationary stochastic models. Box-Jenkins models can be examined in three groups. These; are linear stationary stochastic models, non-stationary linear stochastic models, and seasonal models.

The seasonal time series analysis method can be used in various fields of medicine. With such time series, the periodic trend of the occurrence of a particular disease can be detected [Fleming, 2000; Saynajakangas, 2001, Lani, 2001]. The seasonal periodicity of childhood lymphoid leukemia, among other diseases, was also analyzed using statistical methods [Cohen, 1987 Harris, 1987]. Konca, et al. (2015) studies are also studies on the application of time series to the field of health.

2. Time Series Analysis

In its simplest form, a time series analysis is a process of analyzing an observation of data points, i.e., time series data collected over a time period. In time series analysis, data analysts record data observations at fixed intervals for a series of time periods rather than randomly recording data observations. The rate of observation (time interval) can be from milliseconds to several years.

To study "how variables change over time," time series data describes the phenomenon studied at specific time points to analyze fluctuations in variables over time. Parameters of interest may vary between fields. For example, daily recorded values with scientific tools, daily hits on some websites, weekly stock values in the stock market, the number of rainy days per year, and so on. In addition, time series analysis deals with multiple data points to ensure consistency and reliability. The enormous amount of data reflects a good sample size, ensuring that any trends or patterns discovered are not outliers. Also, time series data can be used to predict future outcomes based on previous data.

One of the most important and widely used time series models is the ARIMA model. The ARIMA model is popular because of its statistical properties (for example, moving averages) for finding its parameters. ARIMA models generally use the past values of a univariate time series to predict the future values of the time series. A special case of the ARIMA model is a persistence model in which the future value in a time series is equal to the previous value (hence, a persistence model has only one autoregressive term and no moving average term). Given its simplicity, the persistence model can serve as a good basis for comparing other models. However, a challenge in persistence and ARIMA models involves nonlinear time series data in various practical scenarios (for example, health datasets) involving nonlinear time series data [Newbold, 1983; Zhang, 2003].

3. Time Series Forecast

Time series analysis helps companies detect the causes of fluctuations in trends or key patterns over time using a variety of data visualization techniques, and organizations can examine seasonal trends and conduct further research to understand the reasons for these trends. Organizations use time series forecasting to predict upcoming events as they periodically analyze time series data. In simple terms, time series forecasting is a technique of predicting future events by analyzing past data behavior or trends while taking into account assumptions that future trends will bear similarities with past trends.

The purpose of predicting time series data is to predict how the data observation will continue or change in the future. Forecasting methods using time series data are used in fundamental analysis and technical analysis. In simple words, fundamental analysis is the evaluation process of analyzing a business at its most fundamental or fundamental financial level. For example, the basic ratios of the business are examined to determine its financial health. In addition, this analysis can give an idea of the estimated value the company stock will trade based on the approximate assessment of similar companies. Income, wealth management, business return, interest rates, etc., takes into account various factors such as. For example, at an organizational level, the fundamental analysis includes researching financial data, management, business information, and competition. At the industrial level, an examination is made of the forces of supply and demand for the products offered. Fundamental analysis in accounting and finance is a method of determining the intrinsic value of a security by examining many microeconomic and macroeconomic factors. It aims to measure the true value of a security that can be correlated to the current market price to assist in making investment decisions.

Technical analysis is a trading discipline that analyzes statistical patterns derived from trading behavior, such as price change and volume, to identify trades and find trading opportunities. It helps to develop a perspective on a particular stock by defining entry, exit, and other trading parameters. Technical analysis is only concerned with the actions of market participants. It helps to visualize stock charts with a focus on the patterns that appear above them, and these patterns help identify trading opportunities. Technical analysis does not focus on the reason behind a stock's movement. A technical analyst is more concerned with knowing "how" the price is reacting to a market participant's movement than with "why" the movement occurs.

In general, time series are characteristic of complex datasets recorded over a fixed period of time. Time series analysis, modeling, and forecasting play a vital role in real-life practical examples and in many research fields, such as business, economics, medicine, astronomy, engineering, politics, and more.

4. Time Series in Medicine

Medicine has evolved as a data-driven field and continues to contribute to human knowledge with tremendous advances in time series analysis. Time series datasets, such as electronic health records (ESCs) and other records, represent valuable sources of information spanning a patient's entire lifespan. Whether intentional or not, it captures genetic and lifestyle risks, signals the onset of diseases, indicates the emergence of new morbidities and comorbidities, indicates the time and stage of diagnosis and documents the evolution of treatment plans and their effectiveness.

Using patient data generated at each of these points in the care pathway, we can develop machine-learning models that give us a much deeper and more interconnected understanding of individual health and disease trajectories, including the number of conditions needed to provide an accurate representation. The state of a disease, how to extract a patient's current state, what triggers transitions from one state to another, and much more.

Machine learning is a branch of artificial intelligence and computer science that focuses on the use of data and algorithms to mimic the way humans learn, gradually increasing its accuracy. Machine learning is an important component of the growing data science field. Through the use of statistical methods, algorithms are trained to make classifications or predictions and to uncover key information in data mining projects. In the branch of artificial intelligence called machine learning, computer software learns from experience. The results teach medical researchers and clinicians new ways to study diseases, manufacture drugs, and treat patients.

Some laboratories use time-series datasets to make new discoveries and develop an understanding of the progression and clinical trajectories in a wide variety of diseases, including cancer, cystic fibrosis, Alzheimer's, cardiovascular disease, and COVID-19.

Armed with a purely quantitative and scientific understanding of the progression of multiple diseases over time, the full capabilities of machine learning can be unlocked to create comprehensive long-term patient management programs that evolve with each individual's changing context and history and not just a single disease. This type of truly personalized end-to-end medical care is called bespoke medicine. While current approaches to precision or personalized medicine tend to fit the patient into a pattern (e.g., based on genetic information), bespoke medicine attempts to recognize and adapt to changes in patterns resulting from age, lifestyle changes, and the onset of new conditions.

4.1. *Predicting Disease Trajectories*

Chronic diseases such as cardiovascular disease, cancer, and diabetes progress slowly throughout a patient's life. This progression can be broken down into stages, as manifested by clinical observations. The goal here is to build disease progression models from electronic health records and other informative datasets, learn model parameters at training time, and then publish personalized dynamic estimates. In addition to providing accurate estimates for the patient at hand, it is to make new discoveries about the mechanisms of disease progression at the population level, subgroup level, and individualized level.

4.2. Case Analysis

Today, technology companies and institutions work on big data. Extracting useful information from a large chunk of data is a strenuous task. Considering the gains obtained as a result of mining, it has become an extremely sensitive issue for companies to be able to protect and process not only the data they own but also the data received from outside.

The process of obtaining valuable data that will meet the need of software systems that are defined as large-scale and have millions of data in institutions is called Data Mining. In this way, it becomes possible to reveal the relationships between the data and to make accurate predictions when necessary. In Data Mining, billions of data can be worked on. We can say that the main purpose of mining is to reveal valuable data for the systems called decision support mechanisms in institutions after certain methods and processing processes.

In the medical field, it is important to examine the transformation of behavior over time as compared to making inferences based on absolute values in time series. For example, diagnosing respiratory heart rate variability based on sensor readings is a characteristic example of linking case-based monitoring with time series. However, time series in the context of the field of epidemiology has emerged very recently and gradually as approaches to time series analysis demand record-keeping systems where records must be linked together over time and collected precisely at regular intervals.

Since there are sufficient scientific tools to collect good and long-term data, healthcare applications using time series analysis yield useful results for the industry as well as for the health diagnoses of individuals.

4.3. Time to Event and Survival Analysis

Survival analysis (often referred to as time-to-event analysis) refers to the study of time until one or more events occur. This is vital for many predictive tasks in numerous application areas, including economics, finance and engineering, and of course, healthcare.

In the medical setting, survival analysis is often applied to the discovery of risk factors that affect survival, the comparison between the risks of different subjects at a given time of interest, and decisions about obtaining cost-effective information (e.g., cancer screening). Specifically, the goal is to dynamically estimate the probability of occurrence of various types of future events occurring at a particular time in the future, taking into account competing risks. To do this, it is essential to incorporate longitudinal measures of biomarkers and risk factors into a model. This enables better risk assessments of clinical events rather than discarding valuable information recorded over time.

4.4. *Medical Instruments and Some Applications*

Time series analysis entered medicine with the advent of medical devices. Invented in 1901, electrocardiograms (ECGs) diagnose heart conditions by recording electrical impulses passing through the heart. The electroencephalogram (EEG) was invented in 1924: It measures electrical activity/impulses in the brain. These inventions created more opportunities for medical practitioners to use time series for medical diagnosis.

With the advent of wearable sensors and smart electronic health devices, people can now take regular measurements automatically with minimal input, resulting in a consistently good collection of medical data for both sick and healthy individuals (for more detail, see [Sreenivas et al. 2001].

Electrocardiogram (ECG) signals are the best way to monitor the functionality and health of the cardiovascular system and identify related disorders. Abnormal heartbeats are reflected in the EKG pattern, and such abnormal signals are called arrhythmias. Automatic classification and identification of the ECG arrhythmia signal, which provides faster and more accurate results, is becoming an increasing need. Various machine learning capabilities have been applied to improve the accuracy of the results and increase the speed and robustness of the models. There is a lot of focus on the architectures and datasets used, but preprocessing of the data is equally important.

Kanani (2020) proposes a modified deep learning architecture that contributes to training stability and a preprocessing technique that significantly improves the accuracy of deep learning models used for ECG classification. With this preprocessing technique and deep learning model, the system reached over 99% accuracy levels without overfitting the model.

The few clinical experience studies available have concluded that smoking may be associated with ischemic heart disease and acute coronary events that can be reflected in the electrocardiogram (ECG). However, there was no formal scientific evidence of a significant association between smoking and electrocardiogram results. In [Tseng, 2020], the relationship between electrocardiogram and smoking was investigated using unsupervised neural network techniques, and the relationship was proven.

Electrocardiograph (ECG) signal analysis has been widely used for decades to examine a patient's heart and detect problems such as arrhythmias. Manual analysis of ECG in real time is cumbersome and, therefore, impractical for physicians. Deep learning helps make this task much easier due to faster learning of signal properties and event prediction. Deep Learning classifiers can help doctors distinguish between normal and abnormal ECG signals based on basic and advanced characteristics of ECG signals. Sreenivas (2019) focused on constructing a Convolutional Neural Network (CNN) to classify arrhythmia in dual-channel ECG signals based on images generated from time series according to image coding techniques. ECG time series signals were converted to images using Gramian Angular Fields (GAF) and Markov Transition Fields (MTF), these images were fed as input to a deep learning classifier that further classifies the signals into various types, and the model showed 97% for GAF images and 97% for MTF images. It achieved 85 accuracies.

Time series estimation has been a prosperous field of science due to its popularity in real-world applications, but it poses a challenge in method development. In medical applications, time series prediction models have been successfully applied to predict disease progression, predict mortality, and assess time-dependent risk. However, the wide availability of many different techniques, where each species excels in certain situations, makes the process of choosing a suitable model more difficult. Bui (2017) summarized and reviewed different prediction models that have been enormously cultured for medical purposes using time-series-based prediction methods. Relevant research papers available for each type of model are listed, their underlying theories are briefly explained, and their advantages and disadvantages in different clinical situations are discussed. In addition, a robust and purposeful classification of approximately 60 different forecasting models has been made. Therefore, this study provides a comprehensive reference for scientists and researchers to identify predictive models appropriate for their situation.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
ORCID IDhttps://orcid.org/0000-0002-9559-0730

References

[1] Allard, R., (1998), Use of Time Series Analysis in Infectious Disease Surveillance. Bull. World Health Organ, 76, 327–333.
[2] Bui, C., Pham, N., Vo, A., Tran, A., Nguyen, A. & LE, T. (2017), Time Series Forecasting for Healthcare Diagnosis and Prognostics with the Focus on Cardiovascular Diseases, 6th International Conference on the Development of Biomedical Engineering in Vietnam (BME6), 809–818.
[3] Box, G.E.R. & Jenkins, G.M., (1976). Time Series Analysis, Forecasting, and Control, Holden-Day, San Francisco 1976.
[4] Cohen, P. (1987), The Influence on Survival of Onset of Childhood Acute Leukaemia (ALL).Chronobiol Int, 4, Num. 2 291–297
[5] Fleming, D. M. – Cross, K .W. – Sunderland, R. – Ross, A.M., (2000), Comparison of the Seasonal Pattern of Asthma Identified in General Practitioner Episodes, Hospital Admissions, and Deaths. Thorax, 8, 662–665.
[6] Hellfenstein, U., (1990), Detecting Hidden Relationships between Time Series of Mortality Rates. Methods Inf. Med., 29, 57–60.
[7] Helfenstein, U., Ackermann-Liebrich, U. , Braun-Fahrlander, C. & Uhrs Wanner, H., , (1991), The Environmental Accident at 'Schweizerhalle' and Respiratory Diseases in Children: A Time Series Analysis. Statistics in Medicine, 10 1481–1492.
[8] Harris, R.E., HarrelL, F. E., Patil, K. D. & Al-Rashid, R., (1987), The Seasonal Risk of Paediatric/Childhood Acute Lymphocyte Leukaemia in the United States. J Chronic Dis, 40, Num. 10 915–923.https://www.vanderschaar-lab.com/time-series-in-healthcare
[9] Jenkins, D.M. and Watts, D.G.. (1968). Spectral Analysis, and its Applications, Holden-Day, San Francisco 1968.
[10] Karakaş, E. (2019). Çocuk Yoğun Bakım Ünitesine Olan Talebin Zaman Serisi Yöntemleri ile Tahmin Edilmesi, European Journal of Science and Technology , 17 454-462.
[11] Konca, Ç., Tekin, M. ve Karakoç, F. & Turgut, M. (2015),, Çocuk Yoğun Bakım Ünitesinde Yatan 770 Hastanın Değerlendirilmesi: Tek Merkez Deneyimi, Türkiye Çocuk Hastalıkları Dergisi, 2 90-95.
[12] Kanani, P. & Padole, M**.** **(**2020**).** ECG Heartbeat Arrhythmia Classification Using Time-Series Augmented Signals and Deep Learning Approach, Procedia Computer Science 171**,** 524-531**.**
[13] Lani, L., Rios, M. and Sanchez, J. (2001). Meningococcal Disease in Spain: Seasonal Nature and Resent Changes. Gac Sanit, 15, Num. 4 336–340.
[14] Newbold, P. (1983), ARIMA model building and the time series analysis approach to forecasting, *J. Forecast,* 2, 23–35. DOI: 10.1002/for.3980020104
[15] Özüdoğru, A. G., Görener, A., (2015). Sağlık Sektöründe Talep Tahmini Üzerine Bir Uygulama, İstanbul Ticaret Üniversitesi Sosyal Bilimleri Dergisi, 14 (27) 37-53.
[16] Rios, M., Garcia, J. M., Cubedo, M. &Perez, D. (1996), Time Series in the Epidemiology of Typhoid Fever in Spain. Med. Clin., 106, Num. 18, 686–689.
[17] Saynajakangas, P., Keistinen, T.and Tuponen, T., (2001), Seasonal Fluctuations in Hospitalisation for Pneumonia in Finland. Int J Circumpolar Health, 60, Num. 1 34–40.
[18] Saynajakangas, P., Keistinen, T.and Tuponen, T., (2001), Seasonal Fluctuations in Hospitalisation for Pneumonia in Finland. Int J Circumpolar Health, 60, Num. 1 34–40
[19] Sreenivas, K. V., Ganesan, M., Lavanya, R. (2021), Classification of Arrhythmia in Time Series ECG Signals Using Image Encoding And Convolutional Neural Networks, **DOI:** 10.1109/ICBSII51839.2021.9445177.

[20] Sreenivas, K . V., Ganesa N, M. & Lavanya, R (2019), Classification of Arrhythmia in Time Series ECG Signals Using Image Encoding And Convolutional Neural Networks, 18th IEEE International Conference On Machine Learning And Applications (ICMLA)
[21] Tseng, K, Lİ, J, Tang Y, Yang C W & Lin F (2020), BMC Medical Informatics and Decision Making, Article number 127 https://doi.org/10.1186/s12911-020-1107-2.
[22] Zhang, G. P (2003). Time series forecasting using a hybrid ARIMA and neural network model, *Neurocomputing,* 50, 159–175. DOI: 10.1016/S0925-2312(01)00702-0.