INTRODUCTION TO TIME SERIES ANALYSIS AND ITS APPLICATIONS

Jonath Jose
Christ University

INTRODUCTION TO TIME SERIES ANALYSIS AND ITS APPLICATIONS

JONATH JOSE
CHRIST (Deemed to be University)Bangalore, India

August 1, 2022

Abstract

Time series data typically show patterns including trends, seasonal fluctuations, irregular cycles, and sporadic changes in level or variability in the domains of business, economics, environment, medical, and other sciences. The goals of studying such series are frequently to detect unanticipated interventions, evaluate the impact of known exogenous interventions, and extrapolate the dynamic pattern in the data to forecast future observations. In this article, we give an overview of time series analysis along with its applications.

Keywords: Definition, Objectives, Assumption, Application Stationary.

1 Introduction

Time series analysis is a statistical technique that deals with trend analysis and time series data. Time series analysis made its way into medicine when the first practical electrocardiograms (ECGs), which can diagnose cardiac conditions by recording the electrical signals passing through the heart, were invented in 1901. John Graunt's actuarial tables were one of the first results of time-series-style thinking applied to medical questions. In his book, Grant presented the first life tables, which you may know as actuarial tables. These tables show the probability that a person of a given age will die before their next birthday Data from time series are periodic time periods that have been measured at regular intervals or gathered at certain times. To put it another way, a time series is just a collection of data points arranged chronologically, and time series analysis is the act of interpreting this data. The Gross Domestic Product (GDP), the Consumer Price Index, the SP 500 Index, and unemployment rates are examples of time series data in economics. Time series data in the social sciences could include information on population growth, migration patterns, birth rates, and political variables. Here in the article, we discuss the Definitions, Assumptions, Objectives, Application, Time series models, and Stationarity along with their way to identify.

2 Definitions

Time series analysis is a specific way of analyzing a sequence of data points collected over an interval of time. In time series analysis, analysts record data points at consistent intervals over a set period of time rather than just recording the data points intermittently or randomly.

According to researchers their definition of time series are :

Mooris Hamburg: "A time series is a set of statistical observations arranged in chronological order".

W.Z. Hirsch: "The main objective in analyzing time series is to understand, interpret and evaluate change in economic phenomena in the hope of more correctly anticipating the course of future events".

3 Examples

- **SALES ANALYSIS**
  Time series analysis is frequently used by retailers to examine the long-term trends in their overall sales. Time series analysis is very helpful for examining sales trends on a monthly, seasonal, and annual basis. As a result, retail establishments are better able to forecast their sales for the approaching season as well as the number of workers and merchandise they will require at certain points during the year.

- **STOCK MARKETING**
  In order to better comprehend the trends in different stock prices, time series analysis is also commonly employed by stock traders. Particularly useful are time series charts, which help stock analysts and traders comprehend the trend and direction of a particular stock price.

- **WEATHER FORECASTING**

  Weather forecasters typically utilize time series analysis to forecast the temperatures for various months and seasons of the year. It comprises estimating temperature, observing climate change, identifying seasonal shifts, and forecasting the weather.

- **HEART RATE**

  Time series analysis is also used in the medical field to monitor the heart rate of patients who may be on certain medications to make sure that heart rate doesn't fluctuate too wildly during any given time of the day.

4 Objectives

Primarily there are 4 objectives as Description, Explanation, Prediction, and Control.

- DESCRIPTION

  Plotting the data and obtaining basic descriptive measures of the fundamental characteristics of the series are the initial steps in the analysis. These measurements can be as simple as looking for trends or as complex as analyzing seasonal changes. Although this price trend is inconsistent, the preceding figure shows a constant seasonal pattern of price variation. A graph makes it possible to search for "outlier" observations (not appear to be consistent with the rest of the data). The occurrence of turning points where the ascending trend abruptly turned to a downward trend is made possible by graphing the time series. Different models might need to be attached to the two halves of the series if there is a turning point.

- EXPLANATION

  It was possible to utilize the variation in one-time series to explain the variance in another series since observations were made on two or more variables. This might result in better comprehension. In this situation, a multiple regression model can be useful.

- PREDICTION

  One might want to forecast the time series' future values given an observed time series.

The examination of economic and industrial time series is a crucial responsibility in sales forecasting. Forecasting and prediction are used synonymously.

- CONTROL

  When time series are created to gauge a manufacturing process's quality (with the potential to regulate the process), There are various types of control techniques. The observations are plotted on a control chart during quality control, and the controller responds as a result of looking at the charts. The series is fitted using a stochastic model. The series' anticipated future values are used to change the input process variables and keep the process on track.

5 Assumptions

- Time series is stationary.
- Current data points of the time series depend on the past data points.
- Random error terms are normally distributed.
- Time Series does not contain any missing value

6 Components of Time Series

A time series is essentially composed of the following four components.

- **Trend**

  A trend reveals a regular pattern in the data. Over a certain, extended period of time, it could increase, decrease, go up, or go down. The trend is a consistent, long-term general direction of data movement. The data need not move in the same way for there to be a trend. Over a lengthy period of time, the movement or direction may fluctuate, but the underlying tendency should not change in a trend. The number of schools, agricultural output, population growth, etc is a few instances of trends. It is noteworthy that the trend may change direction over different periods of time, either upward, downward, or steadily. A Trend can be either linear or non-linear.

- **Seasonal Variations**

  Seasonal variations are short-term, typically within less than a year, changes in time series. Over the course of the time series' 12-month span, they typically exhibit the same trend of growth—either upward or downward. Schedules on an hourly, daily, weekly, quarterly, and monthly basis are frequently used to track these variances. Natural or artificial factors or changes can cause seasonal differences. Seasonal changes are greatly influenced by the multiple seasons and artificial alterations. The use of umbrellas increases dramatically during the rainy season, sales of air conditioners increase during the summer, and harvests rely on the season are examples of seasonality. Some norms created by humans plainly display seasonal fluctuations. The seasonal differences are influenced by holidays, traditions, styles, habits, and diverse events like weddings. The seasonal variation period should not be used to judge whether a business is doing better or worse.

- **Cyclical Variations**

  Variations in time series that occur themselves for the span of more than a year are called Cyclical Variations. Such oscillatory movements of time serious often have a duration of more than a year. One complete period of operation is called either a cycle or a 'Business Cycle'. Cyclic variations contain four phases - prosperity, recession, depression, and recovery. It may be regular or non-periodic in nature. Usually, cyclical variations occur due to a combination of two or more economic forces and their interactions.

- **Random Variations**

  There is another kind of movement that can be seen in the case of time series. It is pure Irregular and Random Movement. As the name suggests, no hypothesis or trend can be used to suggest irregular or random movements in a time series. These outcomes are unforeseen, erratic, unpredictable, and uncontrollable in nature. Earthquakes, war, famine, and floods are some examples of random time series components.

7 Time Series Plots

A time series plot is a graph that shows information gathered over time from any process. The chart can be used to determine the historical trends in the data as well as whether the data points

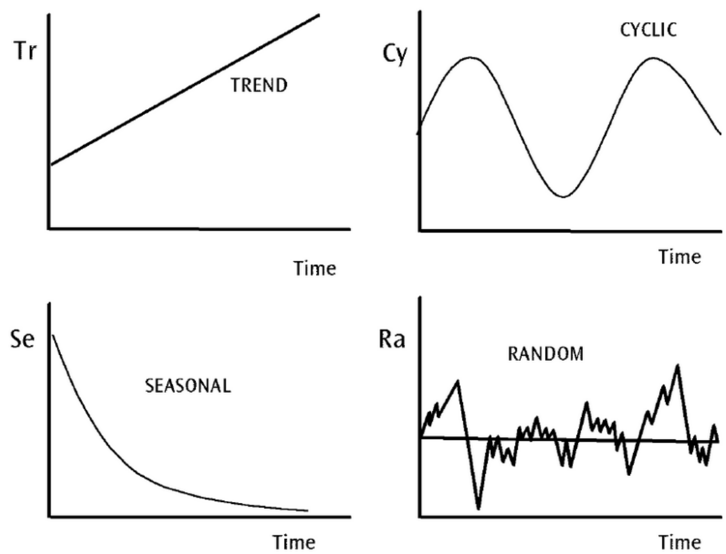


Figure 1: A time series plot with different components

are random or show any patterns. For instance, you could want to check to see if the volume of calls coming into a call center is constant from month to month or if there is any particular pattern in the data, such as a diminishing trend or an increasing trend. A times series plot can also be used to visually detect signs of stability in a process in order to ascertain its stability.

8 Time Series Models

- **Additive Model**

  In the additive model, we represent a particular observation in a time series as the sum of these four components.

  $$O = M + S + C + E$$

  where O represents the original data, M represents the trend. S represents the seasonal variations, C represents the cyclical variations and E represents the irregular variations.In another way, we can write

  $$Z(t) = M(t) + S(t) + C(t) + E(t)$$

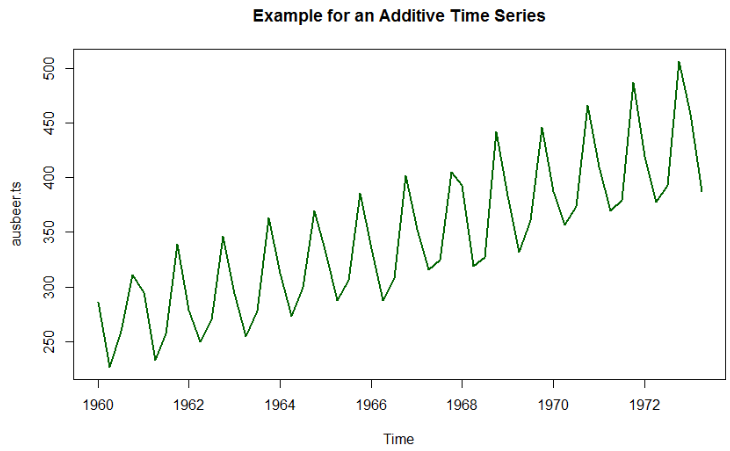


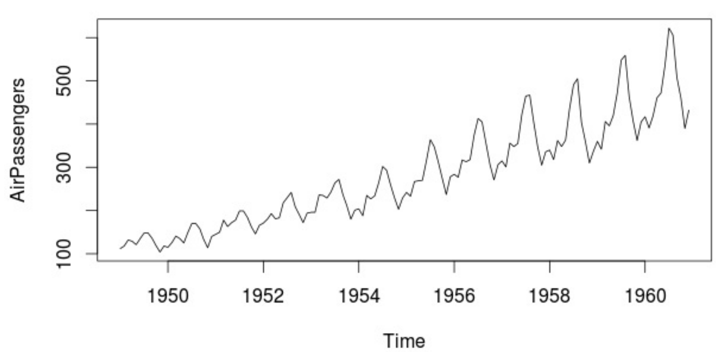
Figure 2: Additive model with M(t) + S(t)+E(t)

- **Multiplicative Model**

  In this model, four components have a multiplicative relationship. So, we represent a particular observation in a time series as the product of these four components:

$$O = MSCE$$

Figure 3: Example for Multiplicative Model



where O, M, S, C, and E represent the terms as in the additive model. In another way, we can write

$$Z(t) = M(t)S(t)C(t)E(t)$$

.This model is the most used model in the decomposition of time series. To remove any doubt between the two models, it should be made clear that in the Multiplicative model S, C, and E are indices expressed as decimal percentages whereas, in the Additive model S, C, and E are quantitative deviations of a trend that can be expressed as seasonal, cyclical, and irregular in nature.

9 Stationarity

Stationarity refers to the time in-variance of some or all, of the statistics of a random process such as mean, autocorrelation, n- order distribution. We can define two types of stationarity strict sense stationary and wide sense stationary. Time series is a realization of..........A time series is said to be stationary if the observations are not dependent on the time that is its statistical properties will not change with time thus they will have constant mean, variance, and covariance. The time series which have trends or seasonality are not stationary. Because trends will have a change in the movement of data concerning time which will cause the change in mean over time. Whereas seasonality occurs when the pattern in time series shows a variation for a regular time interval which will cause the variance to change over time. Our time series data should be stationary in order to forecast and predict. There are different ways to validate the stationarity assumption and convert time series data into stationary.

10 Stationarity Identification

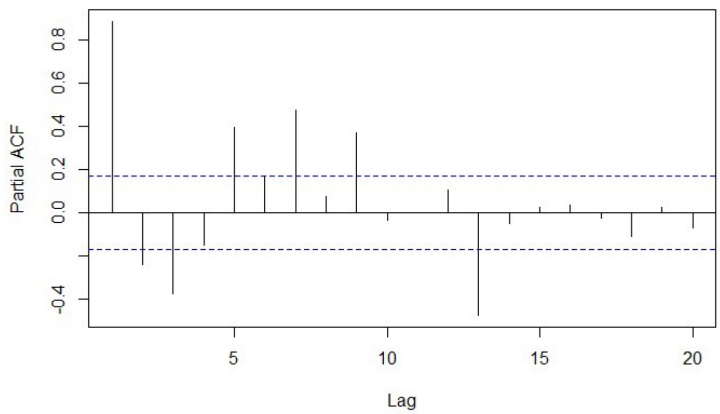
- Checking Mean

  This method first checks the average of the time series for the full observations. Then it takes a sequence of observations from the parent observations for a particular period of time. if the means differ significantly we can conclude that the time series data is not stationary. If the means equate then there might be stationarity present in the time series data.

- Acf Plots

  Autocorrelation represents the degree of similarity between a given time series and a lagged version of itself over successive time intervals. it measures the relationship between a variable's current value and its past values. An autocorrelation of +1 represents a perfect positive correlation, while an autocorrelation of negative 1 represents a perfect negative correlation The ACF plots are known as autocorrelation function plots a display of serial correlation in data that changes over time. The ACF plot can be easily created by using the ACF function. From the plot, if the autocorrelation is more than its threshold limit then the time series data possesses non-stationarity elements.

Figure 4: ACF PLOT



- Adf Test

  The augmented Dickey-Filler test is known as the ADF test. The time series' stationarity is examined using this technique. We need checks like ADF since all forecasting models demand that the time series be stationary. The absence of steady time series is the null hypothesis in ADF. Test Statistics should have a value below the crucial values in order to reject the null hypothesis (1 percent, 5 percent, 10 percent). If this criterion is met, the series can be deemed stationary and the null hypothesis can be rejected. The significance of the test statistics affects the confidence level. For instance, if the value is smaller than the 5 percent critical value, we can declare with 95 percent certainty that the time series is stationary.

11 Transforming into Stationarity

To study time series data our data should be stationary in nature. By nature, every time series data possesses various components such as trend, seasonality, cyclic and random variations. So to achieve stationarity we need to eliminate the trend and seasonal component along with random variation. although the random variation cant is eliminated completely we can minimize it. The methods such as the Method of differencing, Method of least squares help to eliminate the trend component. The method of moving average eliminates the trend component present in the time series.

- Method of Differencing(Detrending)

  Obtaining stationarity with non-stationary data is most frequently done in this manner. By differencing, we formally generate a new data set that contains the variations between the observations made at a given moment in time and at a prior one. Even though sometimes we need to differentiate our data twice to establish stationarity, typically one differencing is sufficient. With larger datasets, the one observation that each difference subtracts from our dataset becomes insignificant.

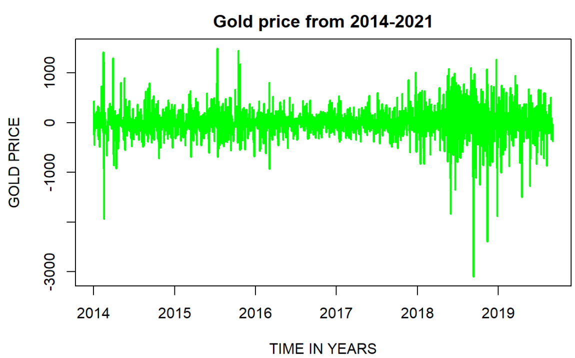


Figure 5: Gold price plot after eliminating trend by Differencing

- Method of Least Squares(Detrending)

  In this method we eliminate the trend component, therefore, making our data to be stationary. Many of the time variables we encounter during time series analysis are reliant on one another, which causes non-stationary. Therefore, we employ the least squares method to eliminate the non-stationary brought on by the trend and other components. The link between a known independent variable and an unknown dependent variable is represented by each point on the fitted curve. In order to fit through the provided points using the least squares method, the time series data is often detrended. If the curve produced by least squares satisfies the ACF plot for stationary, we stop suggesting new models with various degrees; if not, we continue until it does.

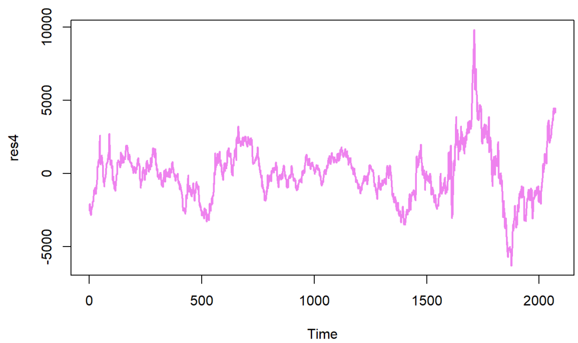


Figure 6: Gold price plot after eliminating trend method of least squares

- Method Of Moving Averages(Deseasonalisation)

  This technique for smoothing a time series is quite effective. It is crucial to eliminate any

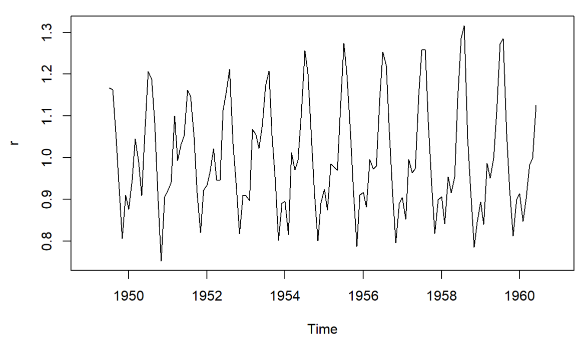


Figure 7: Deseasonalised Air passengers plot

significant variance in the data before forecasting, notably the seasonal effect. The moving averages method evens out short-term variations and clarifies the trend. It is crucial to Before using this procedure, be aware of the data's time span. If there is a duration of time, An n-period moving average is used. The moving average is automatically calculated when n is odd. centered on the data points, but when n is even, we must first center the data. T

12 Conclusion

In the present world, where scientific research is highly valued and digital technology is developing, time is a crucial component that must be taken into account. Time series analysis plays an important part in anticipating all fields, which might include his financial investments, power consumption, spending on e-commerce, or projecting the positive growing stocks in the future. As a result, it is clear that the time series is the dataset with considerable temporal patterning. These patterns might be reproduced repeatedly or not. This article offers solutions to the most fundamental time series questions and aids in learning.

Supplemental Materials

R-package for Time series Analysis: R-package Astsa, Forecast series contain code to perform the diagnostic methods described in the article. The package also contains all datasets used as examples in the article.

Gold Price dataset: Data set used in the illustration of figures in the section is available in the Kaggle repository.

References

Box, G. E., Jenkins, G. M., Reinsel, G. C., amp; Ljung, G. M. (2015).
Time series analysis: forecasting and control. *John Wiley amp; Sons.*

Chatfield, C., amp; Xing, H. (2019)
The analysis of time series: an introduction with R. *CRC Press.* Chatfield, C., amp; Xing, H. (2019). The analysis of time series: an introduction with R.