Research on Global Olive Oil Price Forecasting

Gangyi Li*

School of Mathematical Sciences, Dalian University of Technology, Dalian, China

*Corresponding author: 1910700330@mail.sit.edu.cn

Abstract. Nowadays, commodity price forecasting has become a popular topic. The study of olive oil prices can inform pricing strategies and price positioning of olive oil. How to analyze the range and magnitude of price changes, as well as how consumers are likely to react to changes in olive oil prices, are important factors that will influence the future price trend of olive oil. Therefore, it is crucial for companies to set prices correctly. Since the ARIMA model has high forecasting accuracy and fewer parameters, the model has been chosen for most similar forecasts and is also chosen as a forecasting tool in this paper. The forecast results show that international olive oil prices will remain wildly high in the future, but the rate of increase will slow down. Since the data of recent years are heavily affected by COVID-19, the outliers of recent years also need to be dealt with if more accurate forecasts are needed, which is an unresolved issue in this paper. This paper also attempts to make some policy recommendations based on previous studies to moderate the rapid increase in olive oil prices.

Keywords: Global olive oil price; forecast; ARIMA model.

1. Introduction

It is well known that olive oil producing agriculture has been globally affected by COVID-19, which has led to a significant drop in production in many sectors worldwide. In addition to inflationary factors, climate change is a factor that cannot be ignored. Data from the Spanish Ministry of Agriculture shows that in Spain, 70.2% of olive plantations are rain-fed.2023 witnessed a continuous decrease in rainfall. Both temperatures and rainfall are anomalous compared to levels over the same period from 1991 to 2020. This is also extremely challenging for olive cultivation and growth. There are many other factors contributing to the global spike in olive oil prices, and olive oil users around the world are suffering the effects of rising prices. This price increase has been going on for several years and will continue.

According to a press release from the International Monetary Fund, olive oil prices have skyrocketed by 356% in the last three years, and international olive oil prices have reached US$5,989.8 per ton in 2022, the highest level since February 1997, and continue to soar. This round of higher olive oil prices globally began in 2020 at the start of the epidemic, when home cooking became mainstream. Due to three consecutive years of COVID-19 and a downturn in the economy, olive oil production continued to decline over the next few years, falling short of what was needed on a global scale. The International Olive Council forecasts that global olive oil consumption will reach 3.05 million tons in the 2022/2023 crop year, while only 2.5 million tons will be produced due to extreme temperatures and lack of rainfall in Spain, the world's largest producer, as well as in Italy and Portugal. It can be inferred that lower production is a major reason why olive oil prices are all the way up [1]. In response to this situation, in China, for example, the Chinese government released an agricultural policy to encourage olive cultivation in 2019. However, will this agricultural policy lead to a rebound in olive production? Will the price of olive oil continue to decrease after the increase in olive production? The discussion of the above questions is very important. This is because accurately grasping the characteristics of the trend of price changes and deeply recognizing the great challenges that price changes bring to the olive industry and economic and social development have an important effect on the long-term development of the olive industry.

During this paper the author uses RStudio to forecast olive oil price trends in order to advise consumers and farmers whether to abandon olive oil in favor of other oils, and tries to analyze other reasons behind the increase in olive oil prices [2].

Abo-Ragab et al. conducted an econometric study on the determinants of Egyptian olive oil production and consumption and the international competitiveness of Egyptian olive oil [3]. George et al. analyzed the competitiveness of the Greek olive oil industry using the Porter's diamond model [4]. Xuejun Wu et al. analyzed the trend of olive oil prices on this basis [5]. In quantitative studies of similar problems, the most widely used are time series models [6, 7]. In this paper, the time series will be analyzed and forecasts will be given using the Arima model with different parameter choices and the seasonal Arima model [8].

However, a multi-decade time series can be affected by many historical events and using the complete time series for forecasting will reduce the accuracy of the results, so it is also intended to use the method of change point analysis to make the results of the study more accurate [9, 10].

In conclusion, the prediction of olive oil price is very important, and the results of olive oil prediction in Egypt and Greece are also very informative predictions. As an extension of the previous studies, this paper is prepared to forecast the future trend of global olive oil price using the global olive oil price data from 1990-2023.

2. Methods

2.1. Data Source and Description

The data used in this paper comes from the FRED database, which is a time series of 405 data points from January 1990 to September 2023 for the Global price of Olive Oil.

2.2. Selection of the Methodology

Figure 1 shows the complete, unprocessed time series plot.

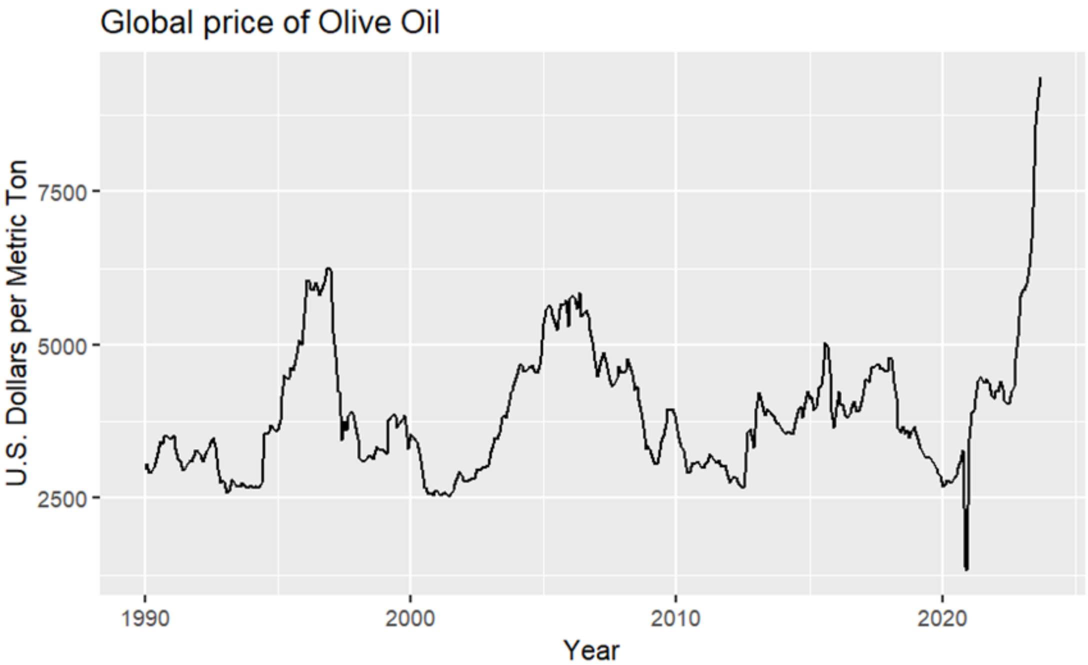

Fig. 1 Global price of Olive Oil

It is not difficult to see from the graph that the global price of olive oil has a cyclicality of about 10 years. Although there has been a clear upward trend in recent years, the time series does not show a clear upward or downward trend when viewing the data as a whole.

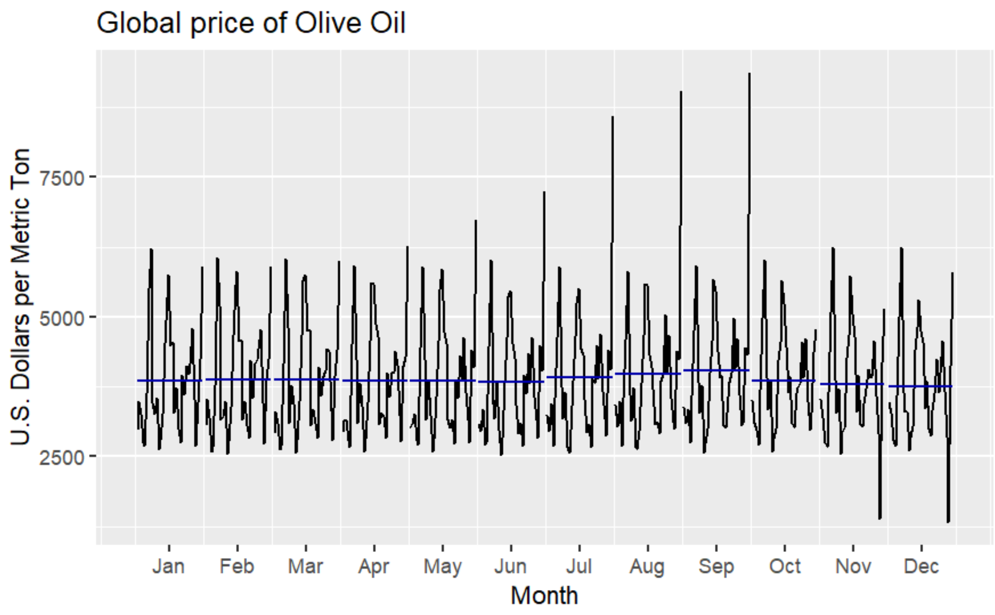

Fig. 2 Subseries plot of international olive oil prices

It can also be seen from Figure 2 that the original data also has a seasonality - the highest value of the year is reached around September each year.

Figure 3 shows the autocorrelation plot for this time series, where the horizontal coordinate is the number of lags and the vertical coordinate is the auto-correlation coefficient at that number of lags, which describes the correlation of the time series with the lagged time series.

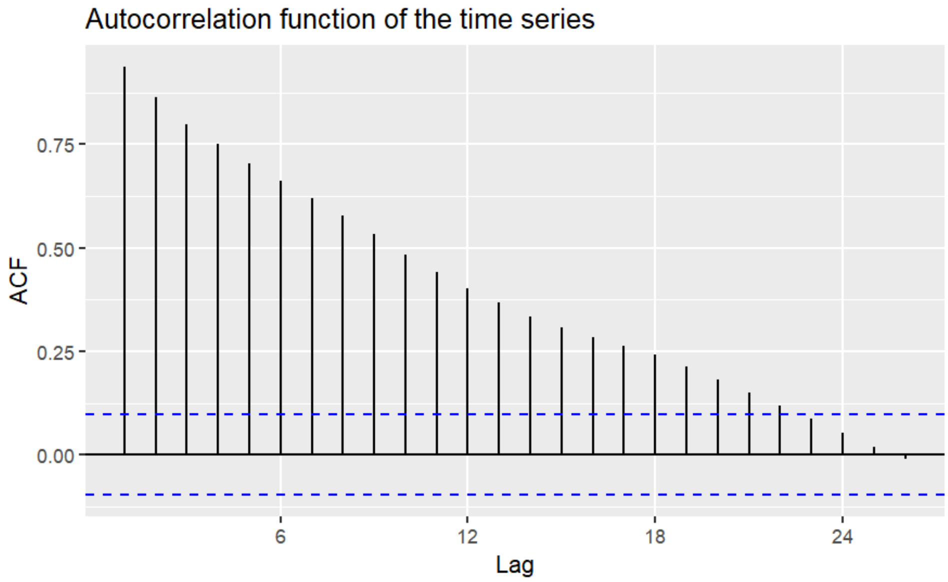

Fig. 3 ACF plot for global olive oil prices

From the ACF plot, it can be seen that the autocorrelation coefficient is decreasing slowly, and the series is clearly not white noise, which allows us to forecast this time series. Therefore, we try to use the ARIMA model to make forecasts.

2.3. Model description

AutoRegressive Integrated Moving Average (ARIMA) is a common and powerful tool used for time series analysis, so we chose ARIMA as the method for model forecasting. The ARIMA model consists of three parts: AR, I and MA, the autoregressive (AR) term means the lag of the difference series, the moving average (MA) term refers to the lag of the error, and the I is the number of differences used to smooth the time series.

The non-seasonal ARIMA model can be written as ARIMA (p,d,q), where p is the number of orders in the autoregressive part, d is the first difference degree, and q is the number of orders in the moving average part. There is also a seasonal ARIMA model: ARIMA(p,d,q)(P,D,Q)[m], whose seasonal part consists of items similar to the non-seasonal part of the model, and m is the number of observations per year.

3. Results and Discussion

3.1. Model Evaluation

We usually use RMSE (Root Mean Square Error) and AIC (An Information Criterion) to evaluate the fit of the model, and the smaller the two values are, the better the model is fitted, so we choose the model parameters p, d, q and P, D, Q by minimizing the RMSE and AIC values. Table 1 demonstrates different selections of the parameters.

Table 1. Selection of model parameters

Model	RMSE	AIC
ARIMA (0,0,3) (1,0,1) [12]	335.9366	5878.757
ARIMA (0,0,2) (1,0,1) [12]	391.2122	6000.694
ARIMA (0,0,3)	350.3789	5907.75
ARIMA (0,0,3) (1,1,1) [12]	332.6663	5740.644

After comparing several ARIMA models with different parameter selections, we found that even the ARIMA (0,0,3) (1,1,1) [12] model with the smallest values of RMSE and AIC fails to pass the residual test, the Ljung-Box test.

The Ljung-Box test for the residuals of ARIMA (0,0,3) (1,1,1) [12] model shows that the p-value is smaller than 2.2e-16, which is considerably lower than the significance level of 0.05, and the original hypothesis is rejected, which indicates that the residuals are not white noise.

Then, we decided to use the method of Change Point Analysis with the intention of identifying the change points in the time series in order to acquire more accurate forecasting.

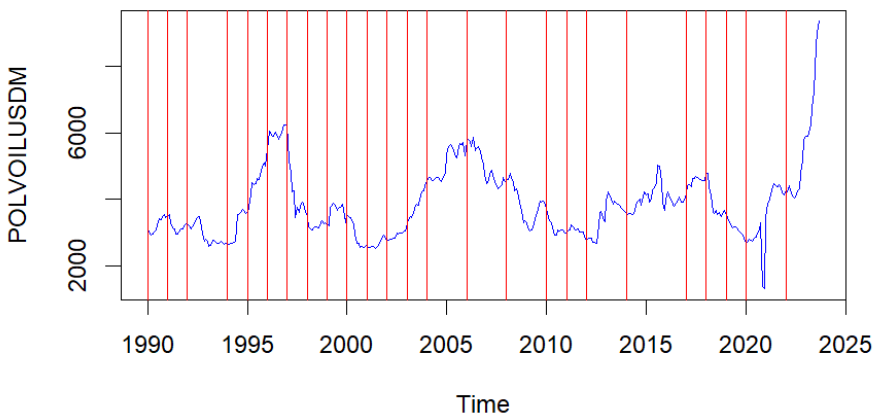

Fig. 4 Global olive oil prices and the change points in the time series

The red line in Figure 4 shows the positions of all change points throughout the time series. Gathering attention on the last period of time, in order to preserve the cyclicality that the original time series possesses, we cut the time series at the data point of January 2017 and focus on the second half of the time series in Figure 5.

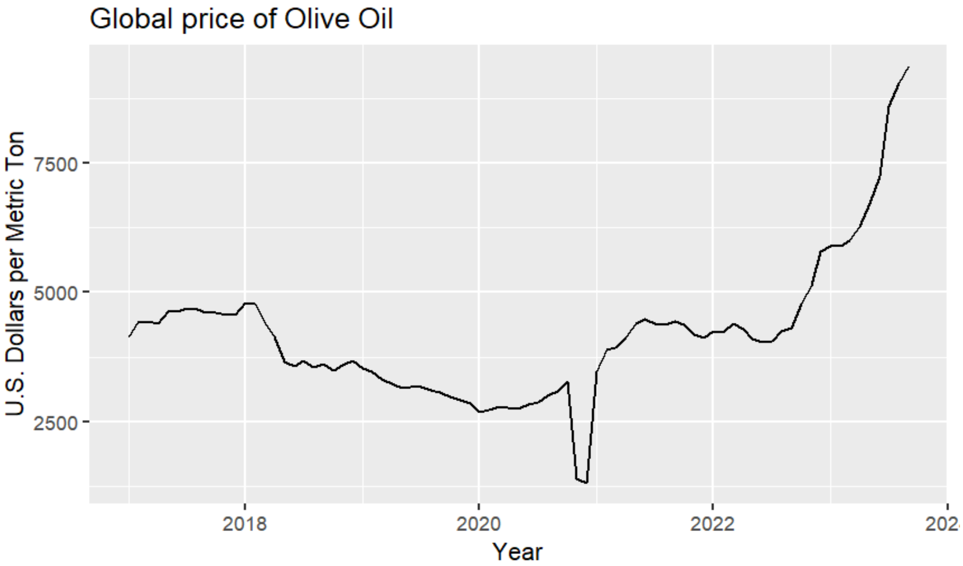

Fig. 5 The time series after cutting

As we have done above, Table 2 gives the RMSE values and AIC values of the ARIMA model with different parameter selections.

Table 2. Selection of model parameters

Model	RMSE	AIC
ARIMA (0,1,3)	358.4697	1178.103
ARIMA (0,2,3)	351.3064	1162.661
ARIMA (1,2,3)	350.8284	1164.517
ARIMA (0,2,2)	376.2744	1171.086

After the comparison in Table 2, ARIMA (1,2,3) has the lowest RMSE value and lower AIC value, so we finally selected ARIMA (1,2,3) as the model for forecasting.

3.2. Prediction Results

Figure 6 presents the results of the predictions using the ARIMA (1,2,3) model, with the blue lines showing the predicted values and the light blue and dark blue regions showing the confidence intervals at 80% and 95% confidence levels, respectively. The contents of Table 3 then show the values predicted by the model.

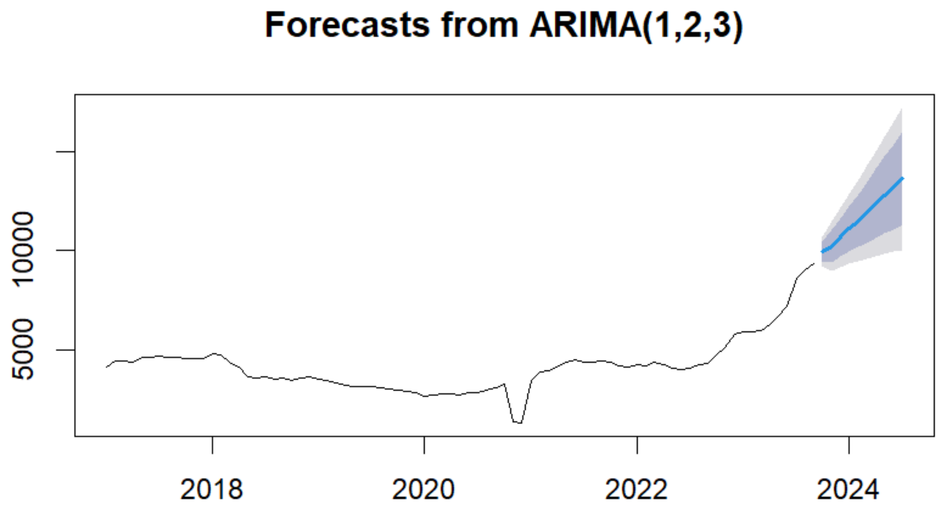

Fig. 6 Prediction results of ARIMA (1,2,3)

Table 3. Prediction results of ARIMA (1,2,3)

Months	Point Forecast	LO 80	HI 80	LO 95	HI 95
2023.10	9968.401	9501.158	10435.64	9253.815	10682.99
2023.11	10237.098	9425.584	11048.61	8995.994	11478.20
2023.12	10682.040	9710.550	11653.53	9196.274	12167.81
2024.1	11104.148	9951.266	12257.03	9340.967	12867.33
2024.2	11529.215	10191.011	12867.42	9482.609	13575.82
2024.3	11953.898	10423.864	13483.93	9613.913	14293.88
2024.4	12378.631	10650.397	14106.86	9735.526	15021.74
2024.5	12803.357	10870.508	14736.21	9847.319	15759.40
2024.6	13228.085	11084.256	15371.91	9949.382	16506.79
2024.7	13652.812	11291.724	16013.90	10041.839	17263.79

3.3. Residual Test

From the ACF plot in Figure 7, it can be seen that the absolute value of all the values is less than the absolute value of critical values, so this residual looks like white noise.

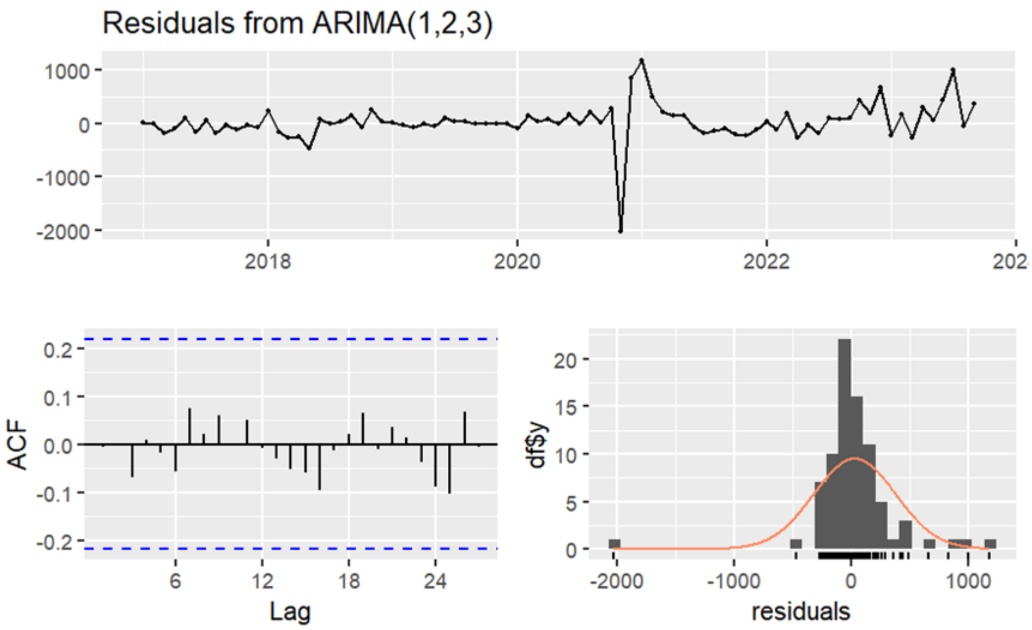

Fig. 7 Residual test of the model

The p-value of the residuals of the ARIMA (1,2,3) model in the Ljung-Box test is 0.9908, which is greater than 0.05, so the original hypothesis is valid and there is no autocorrelation in the residuals, meaning that the residuals are white noise, and our results are appropriate.

However, there are some difficulties in our study. There is a huge trough in the data in 2020, and the huge ups and downs are thought to be the effect brought by COVID-19, and this data may be an outlier that needs to be processed further. That said, our predictions are not yet the most accurate results.

3.4. Critical Thinking

As with most prediction interval calculations, ARIMA-based prediction intervals are often too narrow. This is because only the variation in error is considered. Changes in parameter estimates and models are not included in the calculations. In addition, the calculations assume that the modeled historical patterns will persist through the forecast period. As we know, COVID-19, which lasted for several years, severely affected the trend of the later data, and the impact of the epidemic is difficult to analyze quantitatively, and methods of eliminating outliers are still undetermined.

4. Conclusion

The results of this paper show that although global olive oil prices will continue to increase for some period of time in the future, there is a tendency to slow down. Therefore, the negative impacts of COVID-19 on the olive oil industry are diminishing, and this paper can be of help to olives as well as to those working in the olive oil industry.

Through the analysis of the model and projections, a series of recommendations are also given, including the active use of alternatives such as soybean oil; balancing the overall dietary fatty acid ratio and thus reducing oil intake. In addition, lower prices can be given to other types of cooking oils to induce consumers to reduce their use of olive oil. Governments can also play an important role by encouraging olive oil cultivation and introducing incentives such as tax reductions for farmers. All in all, the problem of soaring olive oil prices needs to be solved, and the key to solving it lies in the ability of governments and people to work together.

References

[1] Guocheng Zhang. Report of market prospective and investment strategy planning on China olive oil industry (2023-2028). Shenzhen, 2022, 32-37.

[2] Unwin A. Graphical data analysis with R. Chapman; Hall/CRC, 2015.

[3] Abo-Ragab S A S, Hany A R S, Essam S S. Econometric analysis of some factors affecting the production and consumption of olive oil in Egypt and the features of its foreign trade. J Appl Sci Res, 2009, 5(10) :1535-1546 .

[4] George G. Analysis of competitiveness of Greek's olive oil sector: using porter's diamond model, Res J Int Stud, 2010, 91(16): 33-46.

[5] Xuejun Wu, Ting Tang. Analysis of World Olive Oil Production and Trade Trends. China Oils and Fats, 2015, 40(12): 1-6.

[6] Hyndman R J, Athanasopoulos G. Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia, 2018.

[7] Wickham H. ggplot2: Elegant graphics for data analysis (2nd ed). Springer, 2016.

[8] Box G E P, Jenkins G M, Reinsel G C, Ljung G M. Time series analysis: Forecasting and control (5th ed). Hoboken, New Jersey: John Wiley & Sons, 2015.

[9] Shilpy S, David A S, Charlie O. Trend analysis and change point techniques: a survey. Energy, Ecology and Environment, 2016, 1: 123-130.

[10] Biometrika. A test for a change in a parameter occurring at an unknown point. E.S. Page, 1955, 42: 523–527.