ISSN: 2645-5439

www.jonarbset.ir

فصلنامه رویکردهای نوین در تحقیقات علوم پایه، فنی و مهندسی

سال سوم، شماره ۹،تابستان ۱۳۹۹

An Overview of the Applications of Box-Cox Transformation in a Time Series

Reza Cheraghi[۱].Omid Farzi[۲]

[۱]Kermanshah Management and Planning Organization
[۲]Razi University of Kermanshah, Department of Statistics

rezacheraghi.۱۹۹۴@gamil.com

Abstract

In all areas of statistics, one of the most important issues is data collection. Sometimes the data collected does not follow the normal distribution and tests were performed to see if the data followed the normal distribution, The most important of these are the Shapiro-Wilk test and the Kolmogorov-Smirnov test. If the data is abnormal, one of the most suitable methods used for data normalization is the Box-Cox method, which is widely used in the time series. In addition to normalizing the data, this method can help maintain the stability of the variance in the time series. In this article, we will first introduce the Box-Cox method and then introduce similar methods for normalizing data, which are four methods. Finally, we will examine the application of these transformations in a time series in R software.

Keywords: Normal distribution, Box-Cox transformations, Kolmogorov-Smirnov test, Shapiro-Wilk test, Normalization.

1. Introduction

Box-Cox transformations are methods for normalizing abnormal data, which is one of the most powerful methods, first by Box-Cox (1964) [1] was presented. Sakia (1992) [2] also reviewed Box-Cox transformations. There are similar methods for analyzing data normalization, the most important of which are Manly (1971), John and Draper (1980), Bickel and Doksum (1981) and Yeo and Johnson (2000) noted.

In addition to normalizing abnormal data, Box-Cox transformations can help keep variance stable in unstable time series. In Box-Cox transformation, our goal is to make sure that the data is normal, which means that for the amount of y we've seen:

$$y \sim N(XB, \sigma^2 I_n)$$

In fact, we have different transformations and the reason is that not all data is the same, and we have a specific form of transformation for each of them. In other words, we can't clearly find powerful transformations for all data.
Draper and Cox (1969)[3] conducted research on this problem and concluded that even in cases where there are no powerful transformations, A function similar to a normal distribution can be constructed and estimation λ is a satisfactory function of certain constraints, usually normal.
In this article, in the second part, we will examine Box-Cox transformations and inference on parameter λ. In the third section, we will introduce four similar methods for these conversions, and finally, using R software, we will apply the Box-Cox transformations for data in the form of a time series and examine its results.

2. Box-Cox Transformations and Inference on parameter transformation

2.1. Box-Cox Transformation

The main form of Box-Cox transformations is as follows:

$$Y_\lambda = \begin{cases} \frac{y^{\lambda}-1}{\lambda} & , \text{if } \lambda \neq 0 \\ \text{Log}\, y & , \text{if } \lambda = 0 \end{cases} \qquad (١)$$

In some articles, it has been suggested that we use a developed form of this form in which it can contain negative values of λ .

$$Y_\lambda = \begin{cases} \frac{(y+\lambda_2)^{\lambda_1}-1}{\lambda 1} & , \text{if } \lambda \neq 0 \\ \log\,(y+\lambda 2) & , \text{if } \lambda = 0 \end{cases} \qquad (٢)$$

Here $\lambda = (\lambda 1, \lambda 2)$ That we can choose $\lambda 2$ for λ each shape $y + \lambda 2 > 0$. So you can see the value $\lambda 1$ and the value $\lambda 2$ is zero.
In these transformations proposed by Box-Cox, λ is called the transformation parameter and the values that are most commonly used are:

Table 1- Different values for λ

values for λ	Transform
-0.1	$1/Z_t$
-0.5	$1/\sqrt{Z_t}$
0.0	$\ln Z_t$
0.5	$\sqrt{Z_t}$
1.0	Z_t

As can be seen, if λ=0, we must use a logarithmic Transformation, the cause of which is as follows:

$$\lim_{\lambda\to 0} T(z_t) = \lim_{\lambda\to 0} (z_t)^{(\lambda)} = \lim_{\lambda\to 0} \frac{z_t-1}{\lambda} = \ln(z_t)$$

2.2. Inference on parameter Transformation

The main purpose of these analyzes of the **B**ox-Cox is to make an inference on the parameter transformations that Box-Cox proposed in two ways:
The first method and approach is to use the maximum correction method, and the other method is to estimate it normally.
Since the likelihood function is a more convenient method, So it's easier to get a confidence interval Because it has almost the same MLE feature. As stated earlier, we assume that the variable of the transformed response is as follows:

$$Y(\lambda) \sim N(XB, \sigma^2 I_n)$$

Note that x is the raw material of the y sign and data matrix, and the parameter model is as follows $(\lambda, \beta, \sigma^2)$. The density function y (λ) is as follows:

$$F(y(\lambda)) = \frac{\exp\left(-\frac{1}{2\sigma^2}(y(\lambda)-XB))'(y(\lambda)-XB)\right)}{(2\pi\sigma^2)^{\frac{n}{2}}}$$

$J(\lambda, y)$ Jacobin is one of the transformations that goes from y to $y(\lambda)$, so the density y is as follows, which is also the likelihood value for the whole model:

$$L(\lambda,\beta\sigma^2 \mid y,x)=f(y)=\frac{\exp\left(-\frac{1}{2\sigma^2}(y(\lambda)-XB)'(y(\lambda)-XB)\right)}{(2\pi\sigma^2)^{\frac{n}{2}}}\ g(\lambda, y)$$

To obtain the MLE from the previous rectilinear equation, it is observed that for each constant λ, the likelihood equation for observing y (λ) corresponds to the likelihood equation for estimation (β,σ^2). So the value of MLE is (β,σ^2) as follows:

$\hat{\beta}(\lambda)=(xx)^{-}\ x\, y\,(\lambda)$

$\hat{\sigma}2(\lambda)=\left(\frac{y(y(\lambda)(I_n-G)y(\lambda)}{n}\right)$

where in: $G=x(x'x)^{-}x'$

Values $\hat{\beta}(\lambda)$ and $\hat{\sigma}2(\lambda)$ are replaced in the correlation equation. It should be noted that in order for the original form of the Box-Cox transformations to be $J(\lambda,y)=\prod_{i=1}^{n} y_i^{\lambda-1}$ .The logarithmic properties of the likelihood can be obtained for each λ. (Maximum likelihood function on (β,σ^2) for each λ)

$l_p(\lambda)= L(\lambda \mid y, x, \hat{\beta}(\lambda), \hat{\sigma}^2(\lambda))$

$= C - \frac{n}{2} \operatorname{Log}(\hat{\sigma}^2(\lambda) + (\lambda-1)\sum_{i=1}^{n} \operatorname{Log}(y_i)$

g is an average of the vectors of the answer. $(g=\prod_{i=1}^{n} y_i^{1/n})$ also : $Y(\lambda,y)=\frac{y(\lambda)}{g^{\lambda-1}}$

Then we can easily find:

$$l_p(\lambda)=C-\frac{n}{2}\log(S_\lambda^2)$$

Where (S_λ^2) the sum of the remaining squares divided by n to maximize the fit of the linear model:

$$Y(\lambda, g)\sim N(XB,\sigma^2)$$

Using the correction log equation, you only need to estimate the value of λ, which is obtained by minimizing the following function:

$$S_\lambda^2=\frac{y(\lambda g)'(I_n-G)y(\lambda g)}{n}$$

Without adding anything and only using standard correction models, we can easily assume that H0 is: $H0=\lambda=\lambda 0$

Let's do the correlation test, the test statistic is: $W=2[l_p(\hat{\lambda})-l_p(\lambda 0)]$. W is free to distribute Khi-Squares with one degree of freedom. It should be noted that W is a function of both data (via) $\hat{\lambda}$ and λ•: $w\sim X_1^2$

The important point here is that a confidence interval for λ when we have a large sample can be more easily obtained by photographing the likelihood ratio test. It is clear that a value of MLE is for λ, so a confidence interval 100% $(1-\alpha)$ is as follows (for λ):

$$\lambda\{n^*\text{Log}\ (\frac{SSE(\lambda)}{ssE(\lambda)} \leq x^2_{(1)}\ (1\text{-}\alpha)\}$$

where in:

$$SSE(\lambda) = y(\lambda, g)'\ (In - G)y(\lambda, g)$$

With an approximate accuracy, we conclude that:

$$P(w\leq\ x^2_{(1)}\ (1\text{-}\alpha)) = 1 + \alpha + 0\left(n^{-\frac{1}{2}}\right)$$

Therefore, it is not difficult to conclude that using the Rao Atkinson statistical rank test in 1973, it was the first to provide a rank ready for testing H0 = λ = λ0.
Although this argument is not the basis of the theory of correctness, in 1987 Lawrence proposed the result of Atkinson's work using the standard theory of correctness, which we refrain from saying here.

So we conclude that when we use Box-Cox transformations, we can use this test and based on a 0.95% confidence interval to get the right λ values and if the data abnormal, in most cases these conversions will normalize the data.

۳. Similar theories of box-cox transformations

In recent years, there have been complementary theories about Box-Cox transformations, some of which we will mention here:

3.1. Manly

Manly (1971) [4] presented the transformations known as exponential transformations:

$$Y(\lambda)=\begin{cases}\frac{e^{\lambda y}-1}{\lambda} & \text{if } \lambda \neq 0\\ y & if\ \lambda = 0\end{cases}$$

In this model, y values can be negative. These transformations were successful in converting oblique single-mode function to normal function and were not effective only for two-model or U-shaped functions.

3.2. John and Draper

John and Draper (1980) [5] presented the following changes, calling them "measurement transformations."

$$y(\lambda)=\begin{cases}sign(y)\frac{(|y|+1)^{\lambda}-1}{\lambda}\ \text{if } \lambda \neq 0\\ \\ sign(y)Log(|y|+1)if\ \lambda = 0\end{cases}$$

where in:

$$sign(y)=\begin{cases} 1, if \;\; y \geq 0 \\ -1, \;\; if \;\; y < 0 \end{cases}$$

Negative values are also present in these transformations. These transformations work better on functions that are somewhat symmetrical. A powerful conversion that is on a symmetrical function that reduces the amount of skewness.

3.3. Bickel and Doksum

Bickel and Doksum (1981) performed a small change in the symmetry test on functions that are almost symmetrical, using the parameters of the Box-Cox transformation model as follows:

$$Y(\lambda)=\frac{(|y|)^{\lambda} sign(y)^{-1}}{\frac{(|y|+1)^{\lambda}-1}{\lambda}\lambda} \qquad \text{for } \lambda > 0$$

where in:

$$sign(y)=\begin{cases} 1, if \;\; y \geq 0 \\ -1, \;\; if \;\; y < 0 \end{cases}$$

3.4. Yeo and Johnson

Yeo and Johnson (2000) [6] theory is a template for the following transformations:

$$Y(\lambda)=\begin{cases} \frac{(y+1)^{\lambda}-1}{\lambda} \text{ if } \lambda \neq 0 \;\; y \geq 0 \end{cases}$$

$$\text{Log}(y+1) \;\; \text{if } \lambda = 0 \;\; y \geq 0$$

$$\begin{cases} \frac{(1-y)^{2-\lambda}-1}{\lambda-2} \text{ if } \lambda \neq 2 \;\; y < 0 \end{cases}$$

$$-\text{Log}(1-y) \;\; \text{if } \lambda = 2 \;\; y < 0$$

After estimating the above transformations, they came to the conclusion that the minimum value of λ in the Kullback-Liebler interval is between the normal function and these transform functions.

4. Numerical results

4.1. Sunspot.year data

In this section, we will examine the Box-Cox method using the Yearly sunspot.year data from the ts class. Sunspot.year data relates to the Yearly number of sunspots from 1700 to 1988. We first summarize the information from this data:

Table 2 - Summary of sunspot.year data

max	3th-qu	mean	median	1th-qu	min
۱۹۰,۲۰	۶۸,۹۰	۴۸,۶۱	۳۹	۱۵,۶۰	۰

The time series diagram for data on the Yearly number of sunspots from 1700 to 1988 is shown in Figure 1:

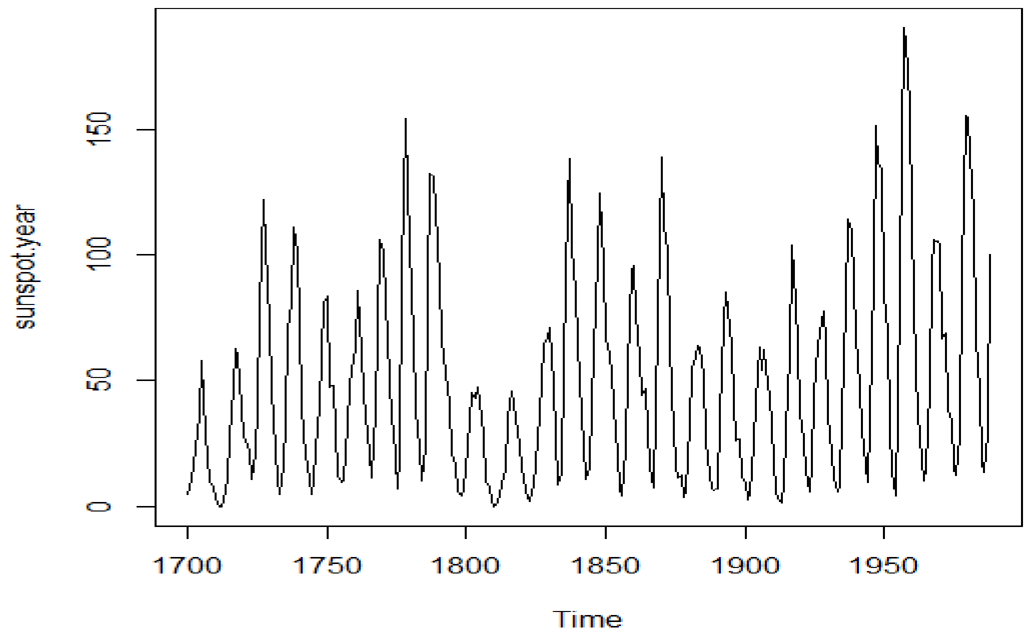

Figure 1. Time series of Yearly numbers of sunspots from 1700 to 1988

At first glance, it seems that the variance of the series (scattering of values) is not constant and changes over time, the series values fluctuate around their average without any specific trend, and an intermittent periodic behavior is seen in the number of sunspots. We use the Box-Cox transformation in the MASS package to stabilize the variable Box-Cox to stabilize the variance.

Figure 2 shows the box-cox transformation on Yearly sunspot data:

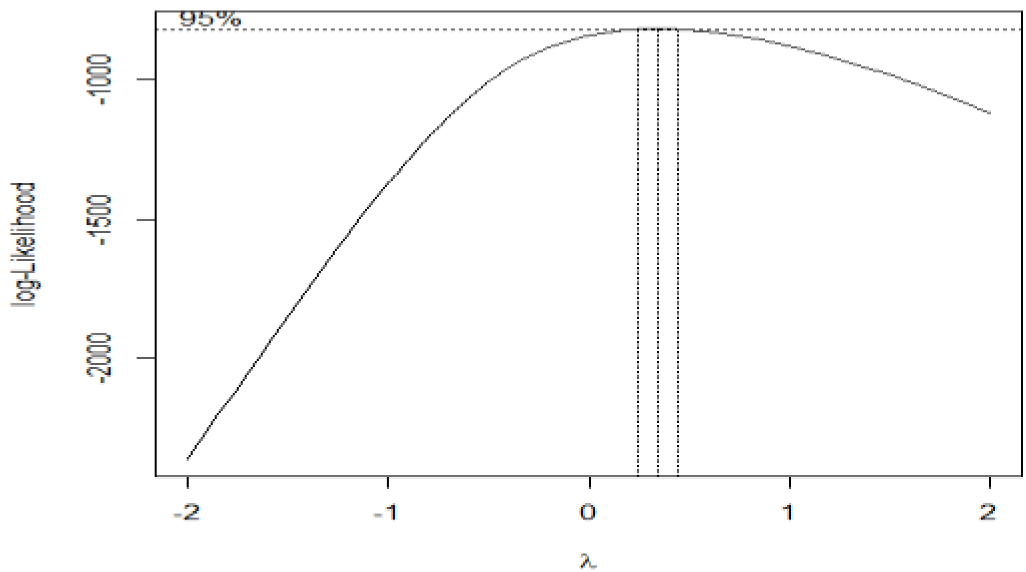

Figure 2. Box-Cox transformation for the Time series of Yearly number of sunspots

Figure 2 shows the logarithmic values of the likelihood function versus the values of the $-2 \leq \lambda \leq 2$ Transformation parameter, and the point at which the function becomes the maximum is the estimate of λ. This point and its confidence interval of 0.95% are marked by dotted lines in the diagram. Based on Figure 2, the estimated value of λ appears to be close to 0.5.
Therefore, using Table 1, it can be concluded that by extracting data from the data, it can help stabilize the series variance. After capturing the transformed series diagram, the following is obtained:

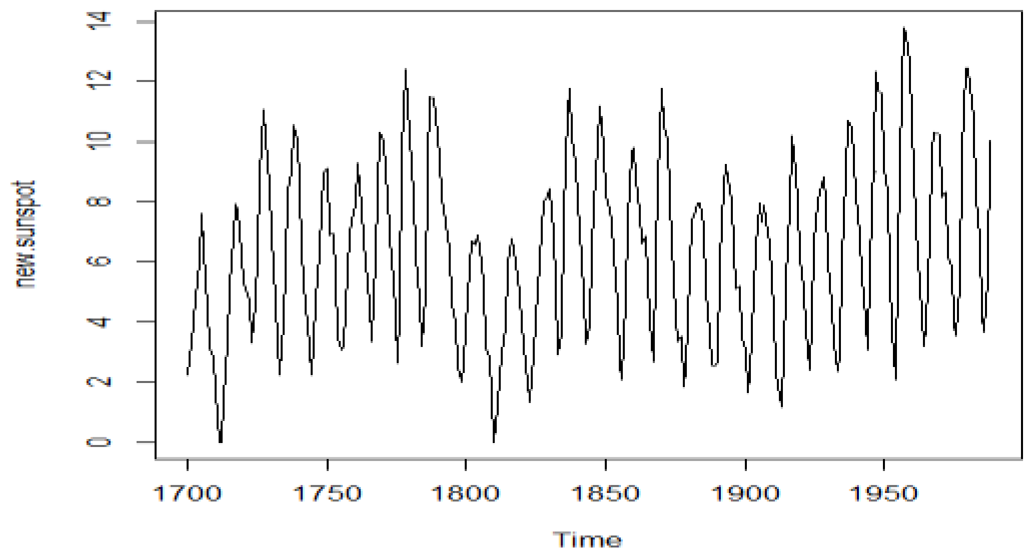

Figure 3. The time series transformed to the Yearly number of sunspots

Figure 3 shows that the modified series changes are much more stable than the original series.

4.2. Johnson.johnson data

In this example, a time series of quarterly earnings per share of Johnson & Johnson from 1960 to 1980 called Johnson.johnson is available in R software.

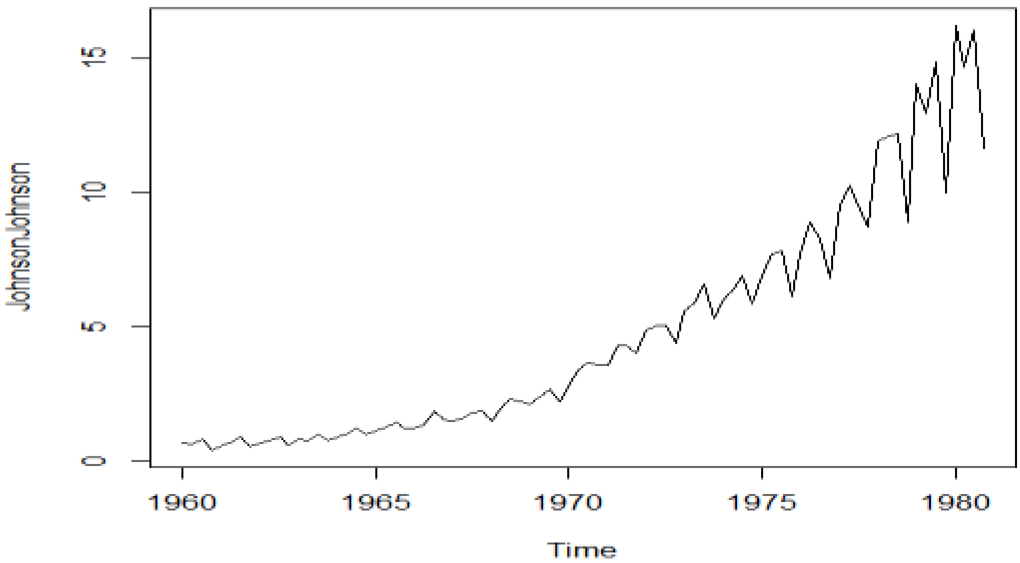

Figure 4. Time series of quarterly earnings per share of Johnson & Johnson from 1960 to 1980

Figure 4 shows the time series of this series showing the increasing trend of this series, which increases with increasing values at the same time as the series changes, which indicates that the variance is not constant.

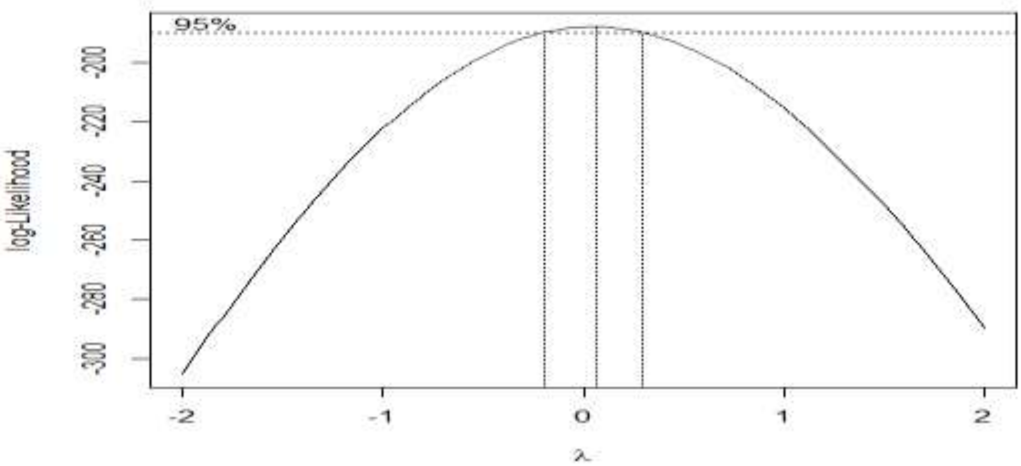

Figure 5. Box-Cox transformations for the time series of Johnson & Johnson's quarterly dividend earnings

According to the box diagram, the coke is at the point $\hat{\lambda}=0$ and this is equivalent to logarithmizing the data. After applying the changes in the data, the time series diagram is drawn as Figure 6, which still has an upward trend.

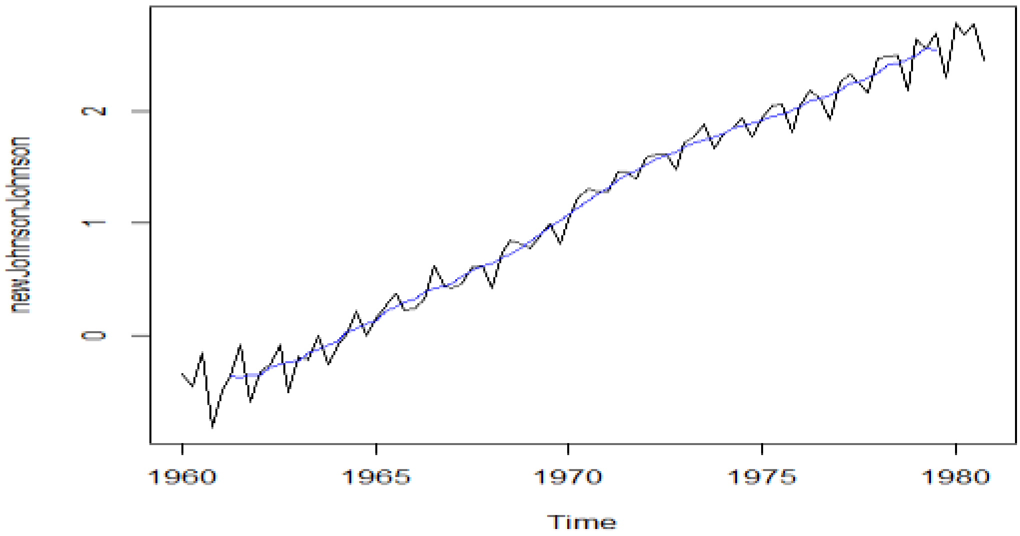

Figure 6 - Johnson and Johnson series logarithm and its moving averages for the fifth time

Due to the increase in this amount of time series and the increase in variance, we conclude that we use various moving averages to check the smoothness of the graph. Figure 6 shows that the moving averages are much smoother than the original series. After distinguishing the first time, the obtained graph will be in the form of Figure 7.

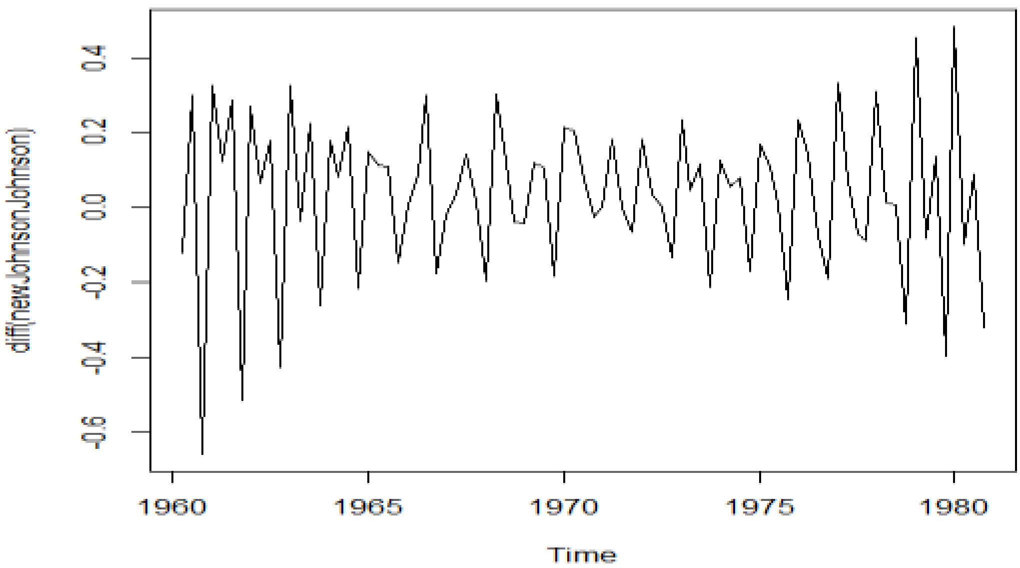

Figure 7- The first time difference between the Johnson and Johnson series logarithms to eliminate the series trend

5. Conclusion

In this paper, it was shown that in addition to normalizing data, Box-Cox transformations are also used to stabilize variance in time series.
Box-Cox transformations are an accurate way to check and stabilize series variations. In addition to this particular type of conversion, there are similar methods, but the most powerful conversion among them is the Box-Cox transformatio.

6. Appreciation

The authors of the article would like to thanks a lot the judges and the editorial board of Magazine.

7. References

1. Box, George E. P. Cox, D. R. *(1964)*. An analysis of transformations. Journal of the Royal Statistical Society, Series B. *26 (2): 211–252*.

2. Sakia, R. M. (1992), "The Box–Cox transformation technique: a review", The Statistician, 41 (2): 169–178

3. Draper, N,R . Cox,D ,R. (1969) On Distributions and Their Transformation to Normality. Journal of the Royal Statistical Society: Series B (Methodological), Volume31, Issue3, Pages 472-476

4. Manly, B. F. (1976). Exponential Data Transformations. **Statistician.** 25(1): 37-42.

5. John, J. A. and Draper, N. R. (1980). An Alternative Family of Transformations. Applied Statistics. 29(2): 190-197.

6. Yeo, Richard A. Johnson.(2000), A New Family of Power Transformations to Improve Normality or Symmetry, Biometrika , Vol. 87, No. 4 (Dec., 2000), pp. 954-959 .