# A suggested weighted running means for smoothing techniques with application

Qasim Nasir Husain[a,*]

[a]Department of Mathematics, College of Education for pure Sciences, Tikrit University, Tikrit, Iraq

(Communicated by Madjid Eshaghi Gordji)

## Abstract

The variety of weighted Average techniques that fall within the structure of exploratory data analysis (EDA) has driven a diversity of competitive results of data analysis. In spite of this variation, the outcomes obtained from the Hanning procedure have some similarities, which makes judging and analysis of data are difficult somehow. Many searches were attitude to get the best performance of technique to give a kind of satisfaction for the researchers who work on the area of EDA. The proposed method is based on calculating the weighted average by applying geometric and harmonic mean to be compared with the existing technique that uses the arithmetic mean. The application of the proposal techniques is done on daily financial data in Malaysia that issues Sukuk of funds in Islamic banking and financial business. The results of applying the proposed techniques on real data sets showed a competitive performance for these standards in the field of comparison of techniques undergoing the smoothing procedures.

*Keywords:* Exploratory Data Analysis, Hanning, Harmonic mean, Geometric mean, Smoothed data.

## 1. Introduction

The research of different fields usually has many advantages for the world in providing information for certain critical areas that need enhancement to keep the accuracy of statistical analyzing and making good decisions for any studied phenomenon. The performance of any data observation, data presentation, and data exploration are usually updated. A Newly proposed technique can be introduced for a better statistical analysis instead of the existent techniques. The mean or the average might have limitations in presenting smoothing data that needs some pattern to be revealed. An

alternative to this, a new technique has been explored. Instead of using a simple mean, the data can be smoothed by other types of statistical measurement tools such as harmonic Hanning and geometric Hanning.

In 2017 Nurul Nisa' Khairol Azmi introduced modifications of running smoothing that are focusing on estimating the middle point of running median for even span by applying geometric, harmonic, quadratic and contraharmonic mean in terms of running smoothing [1]. The different means that give a smaller value than the arithmetic mean , geometric and harmonic, are expected to be more resistant to outliers or extreme data. . Some of the extensions would not work if the data observations have some of zero or negative values. Hence, a constant point should be added to a data to ensure the smoothed data can be given [9]. This paper is fcusing on modifying Hanning technique using harmonic and geometric mean instead of the simple mean.

**Definition 1.1.** *The harmonic mean* H *of the positive real numbers* $y_1, y_2, \ldots, y_n$ *is defined as:*

$$H = \frac{n}{\frac{1}{y_1} + \frac{1}{y_2} + \ldots + \frac{1}{y_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{y_i}} = \left( \frac{\sum_{i=1}^{n} y^{-1}}{n} \right)^{-1} \tag{1.1}$$

Harmonic mean is the altarnating tool of arithmetic mean , which is a traditional way to be used to provide a collection that lying between the maximum and minimum observation data. Harmonic mean is widely used as a tool to calculate central tendency data [5].

**Definition 1.2.** *The geometric mean* G *is defined as the nth root of the product of* n *real numbers* $y_1, y_2, \ldots,$ *which is defined as:*

$$G = \sqrt[n]{y_1 y_2 \ldots y_n} = \left( \prod_{i=1}^{n} y_i \right)^{\frac{1}{n}} \tag{1.2}$$

Geometric mean G is commonly used to get rid of Gaussian outlier data. In general it will help the smoothing process with minimal data losing in compare with arithmetic mean.

An alternative way to existent running median smoothers are running weighted averages which replace each data value with the weighted average of the data values. This is also known as 'Hanning'. Tukey uses a particular form of running weighted average as a gentle smoother of outliers [8].

**Definition 1.3.** *Hanning or running weighted average is defined to be a re-smoothing of the data observations by replacing each data by the mean of its neighborhoods [4, 6].*

The simple Hanning formula can be defined as follows:
Let $n = 1, 2, \ldots$

$$\text{The } 1^{st} \text{ equal Hanning } h_1 \text{ is } \quad h_1 = \frac{1}{3} \left( s_{t-1} + s_t + s_{t+1} \right) \tag{1.3}$$

$$\text{The } 2^{nd} \text{ equal Hanning } h_2 \text{ is } \quad h_2 = \frac{1}{5} (s_{t-2} + s_{t-1} + s_t + s_{t+1} + s_{t+2}) \tag{1.4}$$

Then the $n^{th}$ equal hanning formula can be written as:

$$h_n = \frac{1}{2n+1} (s_{t-n} + s_{t-(n-1)} + \ldots + s_t + \ldots + s_{t+(n-1)} + s_{t+n}) \tag{1.5}$$

One of the three Pythagorean means is the harmonic mean. For all positive data sets that inclode one pair or more of inequal values, the harmonic mean has the lowest value of the three

means. Conversly, the arithmetic mean has the biggest value of the three , while the geometric mean lies always in the middle. If all observations in any dataset are equal, the three means will be equal to one another. For Example, the harmonic, geometric, and arithmetic means for the colletion $\{m, \ m, \ m\}$ are all $m$ for any $m \in \mathcal{R}$ [3].

Because of the effectiveness of the outliers and extreme data on the nature of simple mean, the Hanning procedure must be applied after removing the outliers by one of running smoothing techniques [12].

**Definition 1.4.** *[4] A smoothing index is a tool used to compare the fitted smoothness of curves and its value depends on two scales, namely, the extra balance test ($E_B$), and increase - decrease test ($I_D$) .*

The most reasonable representation of any curves of data observations is that divide the residual points into two equal sets of numbers roughly. The concept of the extra balance is an essential fact in mathematics, so enhancing the extra balance test ($E_B$) and deriving a mathematical formula can help compare smoothing curves and considering the better closest points to the origin [4]. In addition, one of the considerations of choosing among smoothing tools of any smoothing techniques via the degree of the smoothing is $I_D$ test [4, 7]. So the running smoothing index RSI can be defind as:

$$RSI = rank \sum rank(E_B) + rank(I_D) \qquad (1.6)$$

## 2. Data Description

In many projects of exploratory data analysis, the data should be prepared for analysis which is considered as the longest time stage in the procedure. Many obstacles could be appeared in pulling the data together from many sources. in such cases, the data should convert into a suitable shape to be analyzed. Here, accurate dealing with the data should be done [4]. The application of the techniques is done on daily financial data in Malaysia that issues Sukuk of funds in Islamic banking and financial business. The chosen data set to study the level of the performance of data smoothers tools is consists of a long 34-year series of daily US dollar exchange rates is taken from the Reserve Bank of New Zealand from 1980 to 2013. Raw data can be seen in the link `http://www.rbnz.govt.nz/statistics/economic-indicators/b1/discribtion.html,2014`.

In general, big data sets take a long computational time to be analyzed. The analysis can be accelerated by partition and establishing subsets of the data. One method is to take a random subset. Our partition is to divide the data into a group of monthly portions. When the data becomes a table form it should be fully characterized. The data should also be smoothed from extreme and outlier data [6, 7, 11].

To clean data from such values, resorting the smoothed values by taking the maximum value recorded for each month. The details of the data preparation steps needed for analysis should be recorded [2].

## 3. Methodology

According to the weights that can be used in running weighted means, the process can be divided into two directions. First, the weights are unequal such that the weights required by Tukey [4]. Second, the weights are equal such that the weights required by Shitan and Qasim [4, 10]. Via the needs of the proposed measure tools, this study focuses on the equal direction of weights only, so the mathematical formulas of Hanning using the proposed measure tools; harmonic and geometric mean can be arranged as follows:

### 3.1. Proposed Hanning with Harmonic and Geometric Mean

In this section, a modification of Equation (1.5) using the mathematical formula of the harmonic mean and the geometric mean in Equations (1.1) and (1.2) will be introduced respectivly as follows:

$$h_n^* = \frac{2n+1}{\frac{1}{s_{t-n}} + \frac{1}{s_{t-(n-1)}} + \ldots + \frac{1}{s_t} + \ldots + \frac{1}{s_{t+(n-1)}} + \frac{1}{s_{t+n}}} \tag{3.1a}$$

At $n = 1$ the $1^{st}$ equal Hanning

$$h_1^* = \frac{3}{\frac{1}{s_{t-1}} + \frac{1}{s_t} + \frac{1}{s_{t+1}}} \tag{3.1b}$$

At $n = 2$ the $2^{nd}$ equal Hanning

$$h_2^* = \frac{5}{\frac{1}{s_{t-2}} + \frac{1}{s_{t-1}} + \frac{1}{s_t} + \frac{1}{s_{t+1}} + \frac{1}{s_{t+2}}} \tag{3.1c}$$

$$g_n^* = \sqrt[2n+1]{(s_{t-n})(s_{t-n+1}) \ldots s_t \ldots (s_{t+n-1})(s_{t+n})} \tag{3.1d}$$

At $n = 1$ the $1^{st}$ equal Hanning

$$g_1^* = \sqrt[3]{(s_{t-1})(s_t)(s_{t+1})} \tag{3.1e}$$

At $n = 2$ the $2^{nd}$ equal Hanning

$$g_2^* = \sqrt[5]{(s_{t-2})(s_{t-1})(s_t)(s_{t+1})(s_{t+2})} \tag{3.1f}$$

## 4. Applications

Using the statistical software R, a comparison of the results that shown the performance of the proposed Hanning techniques using harmonic and geometric means are introduced. The running smoothing index RSI is applied to some of running smoothing techniques with variant weights of existent Hanning and the proposed Hanning in Equations (3.1b), (3.1c), (3.1d) and (3.1e) with big size of real data observation. The outcomes of the evaluation are shown in Table 1. The chosen Hanning techniques are $H_2$ and $H_4$ taken from [4] with $h_1^*$, $h_2^*$, $g_1^*$ and $g_2^*$ respectively. The application of RSI that gives the results of the procedure can be shown in Table 1 such that $E_B$ coloumn represents the obtained results of the extra balance values of the smoothed data and $\mathbf{I_D}$ coloumn represents the obtained results of the distance between the increasing and decreasing values of the smoothed data [13].

The Figures 1 and 2 shows a comparison among six curves smoothed by EX4253 and RM4253 techniques with different Hanning ways $H_2$, $H_4$, $h_1^*$, $h_2^*$, $g_1^*$ and $g_2^*$ respectively.

Table 1: Showing the outcomes of applying smoothing index RSI on real data set using two of the existent smoothing procedure with two existent and four proposed Hanning

| Technique | Hanning H | $E_B$ | $I_D$ | RSI |
|---|---|---|---|---|
| | $H_2$ | 2.8380 | 3 | 3 |
| | $H_4$ | 2.6766 | 7 | 1.5 |
| EX4253 | $h_1^*$ | 2.8380 | 7 | 6 |
| | $h_2^*$ | 2.6764 | 15 | 4 |
| | $g_1^*$ | 2.8366 | 3 | 1.5 |
| | $g_2^*$ | 2.6768 | 13 | 5 |
| | $H_2$ | 2.60568 | 5 | 4.5 |
| | $H_4$ | 2.49503 | 5 | 3.0 |
| | $h_1^*$ | 2.60497 | 7 | 6.0 |
| RM4253 | $h_2^*$ | 2.49530 | 3 | 1.5 |
| | $g_1^*$ | 2.60424 | 7 | 4.5 |
| | $g_2^*$ | 2.49534 | 1 | 1.5 |
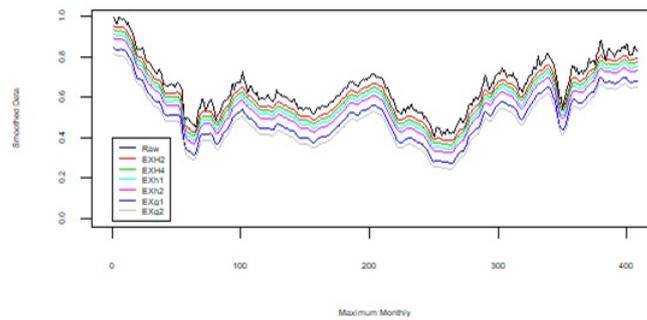


Figure 1: The comparison among different curves showing the support of Hanning Techniques for the visual judgment of smoothed data by EX4253.
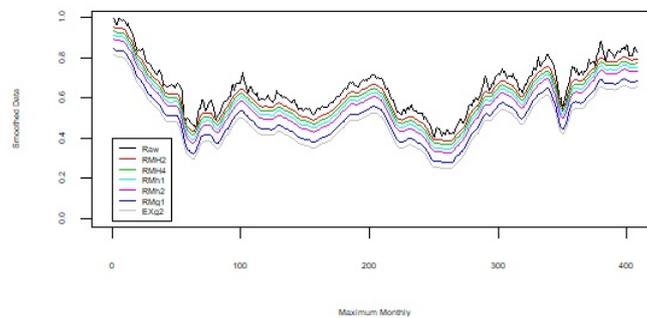


Figure 2: The comparison among different curves showing the support of Hanning Techniques for the visual judgment of smoothed data by RM4253.

## 5. Results and Discussion

In this section, the outcomes of the modification of Hanning technique using harmonic mean and geometric mean will be duscussed. From Table 1 and the Figures 1 and 2 we can conclude the following results

1. For the existent technique of smoothing EX4253 , Hanning by geometric mean with $n = 3$ gives a competative value with compare to the other tested Hanning tools because the two measures give an equal recorded value which is 1.5 that means they share the first ranking.
2. For the improved technique of smoothing RM4253 , Hanning by harmonic mean with $n = 5$ and geometric mean with $n = 5$ perform a competative value with compare to the other tested means because the two measures give an equal recorded value which is 1.5 that means they share the first ranking.

## 6. Conclusions

In this paper, we have presented a new form of Hanning using harmonic and geometric mean. The proposed technique gives the opportunity of comparing the performances among convergence smoothing curves. Reducing the time of the application as seen in Table 1 have respectively shown the outcomes and the comparison of the smoothed curves.

The process of measuring the performance of variant techniques of smoothing has been developed to be more easy and fast in terms of comparing the smoothed data curves. The proposed Hanning of smoothing procedure are proved its efficiency to support the fact of smoothing especially the direct visual judgment on smoothing curves as shown in Figures 1 and 2.

## 7. Recommendation

1. We recommended using another smoothing techniques to guide the smoothing procedure together with another statistical tools like median to compare the performance smoothing more accurate.
2. It is important to put into account enhancing the way of selecting the suitable measure tools that give a better smoothing in terms of time and keeping the data observation so closed to original values
3. very interesting issue to be discussed in future studies is the stopping criteria of smoothing procedure depending on the smoothing index.
4. One of the useful apportuntly of using different statistical measures is to find the best smoothers in terms of reducing the time and to get more acurate smoothed data, we recommend using different techniques of smoothing together with different techniques of Hanning.

## 8. Future Works

In order to reach the best results for the smoothing process while maintaining the accuracy and authenticity of the data, different methods can be used depending on the various statistical measures such as deviations, other averages such as quadratic, contraharmonic with the expansions and modifications that happened to them which leads to modernity and keeps pace with the movement of scientific and statistical development to analyze the data.

## References

[1] N.N.K. Azmi, M.B. Adam, M. Shitan and N.M. Ali, *Modification of smoothing in 4253H [T]*, AIP Conf. Proc. AIP Publishing LLC., 1842(1) 2017.

[2] T. Bernholt, R. Fried, U. Gather and I. Wegener, *Modified repeated median filters*, Stat. Comput. 16(2) (2006) 177–192.

[3] D. Chakrabarty, *Four formulations of average derived from Pythagorean means*, Int. J. Math. Trends Technol. 2021 (2021) 97–118.

[4] Q.N. Husain, M.B. Adam, M. Shitan and A. Fitrianto, *Exploratory extreme data analysis for farmer mac data*, Malay. J. Math. Sci. 11 (2017) 1–16.

[5] W.F. Ferger, *The nature and use of the harmonic mean*, J. Amer. Stat. Assoc. 26(173) (1931) 36–40.

[6] R. Fried, J. Einbeck and U. Gather, *Weighted repeated median smoothing and filtering*, J. Amer. Stat. Assoc. 102(480) (2007) 1300–1308.

[7] K.J. Klassen, *Simultaneous Management of Demand and Supply in Services*, University of Calgary, 1997.

[8] T.C. Mills and T.C. Mills, *Time Series Techniques for Economists*, Cambridge University Press, 1990.

[9] K. Roenfeldt, *Better Than Average: Calculating Geometric Means Using SAS*, Henry. M. Foundation for the Advancement of Military Medicine, 2018.

[10] M. Shitan and T. Vazifedan, *Exploratory Data Analysis for Almost Anyone*, UPM Press, 2011.

[11] H. Tsaknakis and P. Papantoni-Kazakos, *Outlier resistant filtering and smoothing*, In. Comput. 79(2) (1988) 163–192.

[12] P. F. Velleman and D. C. Hoaglin, *Applications, Basics, and Computing of Exploratory Data Analysis*, Duxbury Press, 1981.

[13] L. Zeng, *The optimal degree of smoothing in equipercentile equating with postsmoothing*, Appl. Psychol. Measur. 19(2) (1995) 177–190.