

## A COMPARISON BETWEEN MODIFIED EWMA CONTROL CHARTS USING DIFFERENT ROBUST ESTIMATORS

A.S. RAZALEE, N. MOHAMED ALI<sup>1</sup>, K.S. KIEM, AND N. ALI

**ABSTRACT.** The assumption most often made to construct a control chart is that the observations are from a normal distribution. However, there are many instances in which this underlying assumption is violated. Hence, in this study median absolute deviation (MAD),  $S_n$ , interquartile range (IQR) and Biweight A estimators are proposed in constructing EWMA control chart. The average run length ( $ARL$ ) is used to evaluate the performance of the control charts by determining the number of samples needed before an out of control point is detected when the system is said to be faulty. As a result, the EWMA-Biweight control chart is the most effective as it can detect the out of control at the quicker time frame.

### 1. INTRODUCTION

Control charts are used in statistical process control (SPC) as a method to visualize and control the process performance by reducing product variability and improve production efficiency. The commonly used memory-type control chart are the cumulative sum (CUSUM) control chart that was introduced by Page [1] and the exponentially weighted moving average (EWMA) control chart developed by Roberts [2]. Normality assumption is needed in order to produce an efficient control chart. If the normality assumption cannot be reached or lack of information in the data, therefore, the control charts will be less practical to be used and the

---

<sup>1</sup>corresponding author

2020 *Mathematics Subject Classification.* 62J07, 90-10.

*Key words and phrases.* EWMA control chart, normality, robust estimator, average run length, skewness.

efficiency of the control chart will reduce [3]. Hence, robust control charts are preferred and of more practical use when designing a control chart.

The identified robust scale estimators that have been used in this study are MAD,  $S_n$ , IQR and Biweight A estimators. MAD is a robust scale estimator that uses median as the estimator and is resilient with outliers compared to mean. Rousseeuw and Croux [4] proposed a few other estimators as alternative to MAD, for instance  $S_n$  and  $Q_n$  estimator. However, Rousseeuw and Croux [4] also state that they prefer  $S_n$  than  $Q_n$  although  $Q_n$  is more effective because  $S_n$  is very robust because of its low gross sensitivity. IQR is another robust measure of scale which used to replace the standard deviation in computation of control limits. Biweight A estimator was shown in a major study by Lax [5] to perform well compared to other robust univariate scale estimator.

The main goal of this paper is to modified the classical EWMA control chart using different robust estimators and evaluate each control chart performance using their *ARL* values.

## 2. METHODOLOGY

In this study, data with a rational subgroup of sample size,  $n > 1$  is used. The mean is defined as  $\bar{x}_i = \frac{\sum_{i=1}^n x_i}{n}$  The plotting statistic  $z_i$  is defined as  $z_i = \lambda \bar{x}_i + (1 - \lambda)z_{i-1}$  for  $i = 1, 2, \dots$ , and  $z_0 = u_0$ . While  $\sigma$  is replaced by  $\bar{\sigma}_x = \frac{\sigma}{\sqrt{n}}$  The center limit (CL), upper control limit (UCL) and lower control limit(LCL) for the EWMA control chart are:

$$(2.1) \quad UCL = \mu_0 + L\sigma \sqrt{\frac{\lambda}{(2-\lambda)}[1 - (1-\lambda)^{2i}]},$$

$$(2.2) \quad LCL = \mu_0 - L\sigma \sqrt{\frac{\lambda}{(2-\lambda)}[1 - (1-\lambda)^{2i}]},$$

where  $0 \leq \lambda \leq 1$  is the smoothing parameter that serves as weight given to the data. In this study, the value of  $\lambda$  used is 0.1, recommended by Montgomery [6].

MAD for a sample is defined as  $P|X - X_{0.5}| \leq 0.5$ . MAD is a very robust estimator than the sample standard deviation [7]. The formula is shown as

$$MAD_j = 1.4826[\text{Median}|x_{ij} - \text{Median}(x_{ij})|],$$

where  $i = 1, 2, 3, \dots$ , and  $j = 1, 2, 3, \dots$ . The  $x_{ij}$  is the  $i^{th}$  observation from subgroup  $j$  and the 1.4826 is the consistency factor. The unbiased estimator of  $\sigma$  for MAD is  $\hat{\sigma} = b_n \overline{MAD}$ . Where  $\overline{MAD} = \frac{1}{m} \sum_{j=1}^m MAD_j$  and  $b_n$  is the correction factor of the sample size  $n$ . The value of  $b_n$  is the smallest sample correction depends on  $n$  proposed by Rousseeuw [4] as shown in Table 1. For  $n > 9$ ,  $b_n$  can be calculated using  $b_n = \frac{n}{n-0.8}$ .

TABLE 1. Derived constant for  $MAD(b_n)$ ,  $IQR(d_n)$ ,  $S_n(c_n)$ 

n	2	3	4	5	6	7	8	9
$b_n$	1.196	1.495	1.363	1.206	1.200	1.140	1.129	1.107
$d_n$	1.1284	1.6926	0.5940	0.9900	1.2835	1.5147	0.9456	1.1439
$c_n$	0.743	1.851	0.954	1.351	0.993	1.198	1.005	1.131

The IQR was proposed by Roberts [2] as an estimator. The IQR is a measure of variability, based on dividing data set into quartiles. These quartiles divide a range-ordered data set into four equal parts. The values that divide each part are called the first, second and third quartiles; and they are denoted by  $Q_1$ ,  $Q_2$ , and  $Q_3$  respectively. The population IQR for a continuous distribution is defined to be:

$$IQR = \frac{Q_3 - Q_1}{1.34898},$$

where  $Q_3$  and  $Q_1$  are found by solving the following integral  $0.75 = \int_{-\infty}^{Q_3} f(x)dx$  and  $0.25 = \int_{-\infty}^{Q_1} f(x)dx$ . The unbiased robust estimator for  $\sigma$  is  $\hat{\sigma} = \frac{\overline{IQR}}{d_n}$  where  $\overline{IQR} = \frac{1}{m} \sum_{j=1}^m IQR_j$ ,  $d_n$  is a correction factor of the sample size  $n$ . The values of  $n$  and its respective  $d_n$  values are given as in Table 1.

$S_n$  estimator is an alternative estimator of MAD introduced by Rousseeuw [4].  $S_n$  is appropriate to be used in both cases of non-normality distributed data and normality distributed data and is given as:

$$S_n = 1.1926 \text{ median}_i \{ \text{median}_i |x_i - x_j| \}.$$

The median of  $\{|x_i - x_j|, j = 1, 2, \dots, n\}$  is calculated for each  $i$ .  $S_n$  is the median of these values and the constant 1.1926 is determined, in order to make  $S_n$  as a consistent estimator and approximately unbiased for finite samples. The unbiased robust estimator for  $\sigma$  is  $\hat{\sigma} = c_n \overline{S_n}$  where  $\overline{S_n} = \frac{1}{m} \sum_{j=1}^m S_{nj}$ ,  $c_n$  is a correction factor of the sample size  $n$ . The values of  $n$  and its respective  $c_n$  are displayed in Table 1.

For  $n > 9$ ,  $c_n$  can be calculated using  $c_n = \frac{n}{n-0.8}$  when  $n$  is odd and  $c_n = 1$  when  $n$  is even. The value of  $\hat{\sigma}$  estimated using this estimator will then be used in the control limits formulae in order to construct the resulting EWMA.

As given by Lax [5], the Biweight A estimator of the standard deviation for a sample of size  $n$  is

$$\text{Biweight} = \frac{n}{(n-1)^{\frac{1}{2}}} \frac{(\sum_{|U_i| < 1} (x_i - T)^2 (1 - U_i^2)^4)^{\frac{1}{2}}}{|\sum_{|U_i| < 1} (1 - U_i^2)(1 - 5U_i^2)|},$$

where  $T$  is the sample median and  $U_i = (x_i - T)/(cMAD)$ . The unbiased robust estimator for  $\sigma$  is  $\hat{\sigma} = \overline{\text{Biweight}}$ .

Modified EWMA chart for robust scale estimators are EWMA-MAD, EWMA-IQR, EWMA- $S_n$  and EWMA-Biweight A. For each modified EWMA charts, the value of  $\sigma$  in (2.1) and (2.2) will be replaced by  $\hat{\sigma}$ .

### 3. CONTROL CHART PERFORMANCE

Two measures that are commonly used to compare the performance of control charts are the  $ARL$  for in control process ( $ARL_0$ ) and for out of control process ( $ARL_1$ ). In order to reduce the number of false out of control, signal  $ARL_0$  must be large. A small ( $ARL_1$ ) is required in order to reduce the time that the process is out of control. In this study,  $ARL$  is used it as a measuring tool for comparing the performance of control charts. Monte Carlo simulation is conducted in R programming. The simulations are conducted by manipulating three conditions which are  $\lambda$ , width of control limit ( $L$ ) and type of population distribution. Width of control limit between  $2.1 \leq L \leq 3.5$  is used to determine the best combination ( $L, \lambda$ ) to obtain  $ARL_0 \approx 500$  for data that follow normal distribution. Whereby width of control limit  $4.5 < L < 5.0$  is used for skewed data.

The data used in this study is obtain from Wadsworth et al. [8]. Firstly, the data undergoes a normality test by two different methods that are Shapiro Wilk test and QQ plots. Then, the data is used to construct a classical control chart and four robust EWMA control charts. The UCL and LCL of the five control charts are computed. The statistics for each sample is computed and plotted in the EWMA control charts. If the plotting statistics exceed the UCL and LCL, the process is deemed to be out of control and vice versa. Lastly, the control chart which has the

smallest  $ARL_1$  is selected as the best chart. Many studies such as Tapang et al. [9] and Nazir et al. [10] used  $ARL$  to compare the performance of control charts.

#### 4. RESULTS AND DISCUSSION

The simulated data sets are being tested for all five types of robust EWMA control charts and the performance of each control chart are measured using their  $ARL$  values.

##### **Simulation Results for Normal Distribution Data**

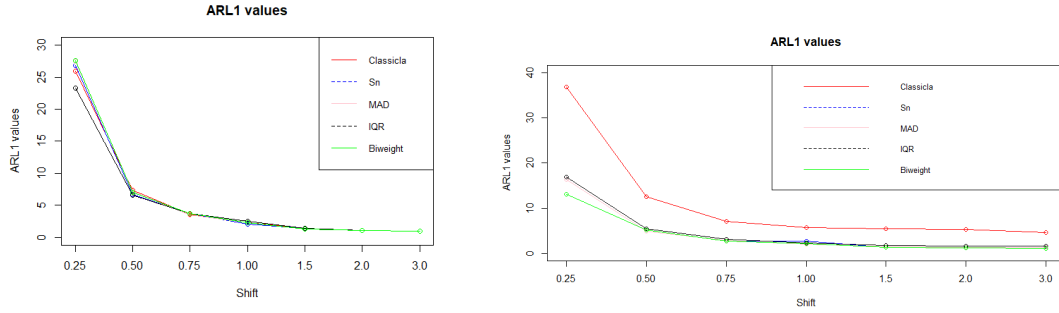
Figure 1(A) displays the five EWMA control charts under Normal distribution with prefix  $ARL_0 \approx 500$  and  $n = 5$  by varying values of width of control limit between  $L = 2.2$  and  $L = 3.5$  with fixed  $\lambda = 0.1$ . From the result, the  $ARL$  value under Normal (0, 1) for both of control charts in steady mode demonstrated expected results where the  $ARL$  results decrease as the shift in mean increases. Additionally, the output shows that EWMA-IQR generally produce smaller value of  $ARL_1$  which indicate that EWMA-IQR control chart is the best control chart in detecting the out of control points if the case of normality assumption is fulfilled.

##### **Simulation Results for Mild Skewness Data**

Figure 1(B) displays the five EWMA control charts under mild skewed Normal distribution with prefix  $ARL_0 \approx 500$  and  $n = 5$  by varying values of width of control limit between  $L = 3.0$  and  $L = 4.3$  with fixed  $\lambda = 0.1$ .  $ARL$  values shows a decreasing trend when the shift in mean increases. From the outputs of the  $ARL_1$ , robust estimator Biweight A is significantly the most effective in promptly detecting the out of control (OOC) point in a quicker time frame when there is a shift in mean.

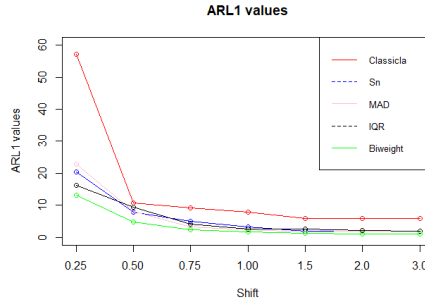
##### **Simulation Results for Heavy Skewness Data**

The  $ARL_1$  output of heavy skewness data is plotted as shown in Figure 1(C). The  $ARL$  values shows a decreasing trend when the shift in mean increases. From the outputs of the  $ARL_1$ , Biweight A control chart is significantly the most effective in promptly detecting the OOC point in a quicker time frame when there is a shift in mean. Contrastingly, classical control charts will take a longest period of time to detect the OOC, hence is it deemed to be ineffective in the aim of detecting the shift especially in the situation where the data do not follow normal distribution.



(A)  $ARL_1$  classical EWMA ( $L = 3.5$ ), EWMA- $S_n$  ( $L = 2.17$ ), EWMA-MAD ( $L = 2.9$ ), EWMA-IQR ( $L = 2.9$ ), and EWMA-Biweight ( $L = 3.5$ )

(B)  $ARL_1$  classical EWMA ( $L = 3.0$ ), EWMA- $S_n$  ( $L = 3.8$ ), EWMA-MAD ( $L = 4.2$ ), EWMA-IQR ( $L = 4.0$ ), and EWMA-Biweight ( $L = 4.3$ )



(C)  $ARL_1$  classical EWMA ( $L = 4.5$ ), EWMA- $S_n$  ( $L = 5$ ), EWMA-MAD ( $L = 5$ ), EWMA-IQR ( $L = 5$ ), and EWMA-Biweight ( $L = 5$ )

FIGURE 1.  $ARL$  values

### Application to real life data

The data has been tested for normality using QQ plot and also Shapiro-Wilk test, and it can be concluded that the data does not come from a normal population. Figure 2(A) and 2(D) show the performance of classical EWMA and EWMA-IQR control chart respectively. No point goes beyond the limit of UCL and LCL. Therefore, the process is deemed to be in control. The resulting chart for EWMA-MAD, EWMA- $S_n$  and EWMA-Biweight are shown in Figure 2(B), 2(C) and 2(E). The first point of the data goes beyond the LCL. It means the system will detect the OOC on the first sample.

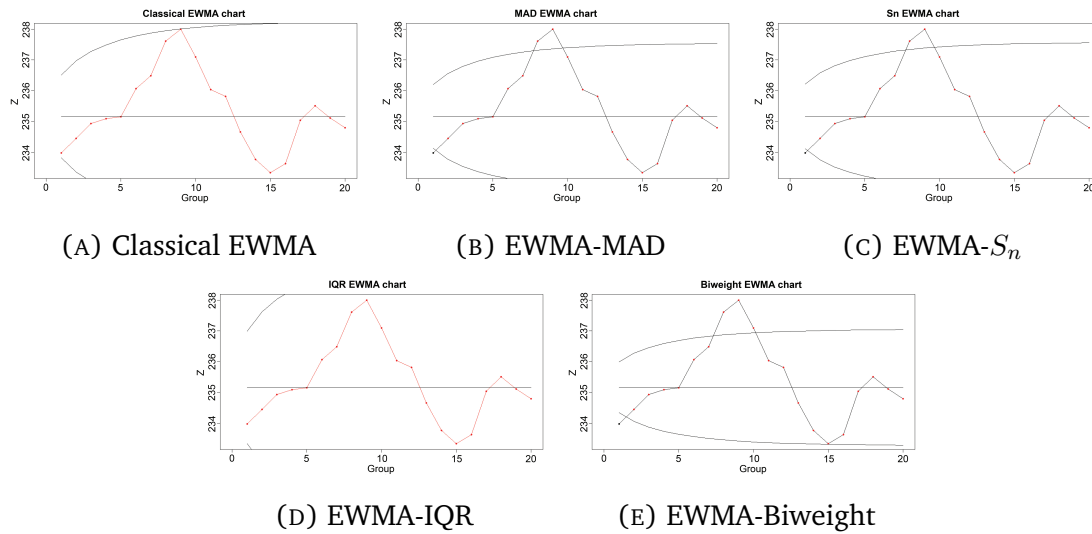


FIGURE 2. Control Charts Performance

## 5. CONCLUSION

The output of simulation under standard Normal distribution shows that EWMA-IQR generally can detect the out of control points in a quicker time compared to other control charts. In the case of violation of normality assumption, the EWMA-Biweight A can detect the out of control points in a quicker time comparing to other control charts. The second method is done by applying real data set in constructing the control charts. It is found that classical EWMA and EWMA-IQR failed to detect the out of control. It is justifiable to conclude that the most effective robust control chart is the EWMA-Biweight control chart.

## REFERENCES

- [1] E.S. PAGE: *Continuous inspection scheme*, Biometrika, **14**(1/2) (1954), 100–115.
- [2] S.W. ROBERTS: *Control chart tests based on geometric moving averages*, Technometrics, **1**(3) (1959), 239–250.
- [3] N. DAS: *A comparison study of three non-parametric control charts to detect shift in location parameters*, The International Journal of Advanced Manufacturing Technology, **41**(7) (2009), 799–807.
- [4] P.J. ROUSSEUW, C. CROUX: *Alternatives to the median absolute deviation*, Journal of the American Statistical Association, **88**(424) (1993), 1273–1283.

- [5] D.A. LAX: *Robust Estimators of Scale: Finite-sample performance in long-tailed symmetric distributions*, Journal of the American Statistical Association, **80**(391) (1985), 736–741.
- [6] D.C. MONTGOMERY: *Introduction to Statistical Quality Control*, John Wiley and Sons, 2009.
- [7] M.O. ABU-SHAWIESH, M.B. ABDULLAH: *A new robust bivariate control chart for location*, Communications in Statistics-Simulation and Computation, **30**(3) (2001), 513–529.
- [8] H.M. WADSWORTH, K.S. STEPHENS, A.B. GODFREY: *Modern Methods for Quality Control and Improvement*, John Wiley and Sons, 1986.
- [9] W. TAPANG, P. ADISAK, S. SUKUMAN: *Three nonparametric control chart based on ranked set sampling*, Chiang Mai Journal of Science, **43**(4) (2016), 914–929.
- [10] H.Z. NAZIR, M. RIAZ, R.J. DOES, N. ABBAS: *Robust CUSUM control charting*, Quality engineering, **25**(3) (2013), 211–224.

DEPARTMENT OF MATHEMATICS, FACULTY OF SCIENCE,  
UNIVERSITI PUTRA MALAYSIA, 43400 SERDANG, SELANGOR, MALAYSIA.  
*Email address:* ainaasalfariena@gmail.com

DEPARTMENT OF MATHEMATICS, FACULTY OF SCIENCE,  
UNIVERSITI PUTRA MALAYSIA, 43400 SERDANG, SELANGOR, MALAYSIA.  
*Email address:* nazihanma@upm.edu.my

DEPARTMENT OF MATHEMATICS, FACULTY OF SCIENCE,  
UNIVERSITI PUTRA MALAYSIA, 43400 SERDANG, SELANGOR, MALAYSIA.  
*Email address:* siew\_kim@yahoo.com

DEPARTMENT OF MATHEMATICS, FACULTY OF SCIENCE,  
UNIVERSITI PUTRA MALAYSIA, 43400 SERDANG, SELANGOR, MALAYSIA.  
*Email address:* norhaslinda@upm.edu.my