

Modified Multivariate Cumulative Sum Control Chart Based on Robust Estimators

A.S. Razalee, N. M. Ali, N. Ali and N. A. A. Rahmin

Department of Mathematics and Statistics, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor

ABSTRACT

Multivariate cumulative sum (MCUSUM) control charts are one of the popular tools for monitoring multivariate statistical process control aside from the Hotelling T^2 and the multivariate exponentially weighted moving average (MEWMA) control chart. However, these charts are easily affected by outliers or shifts in the dataset. To overcome the problem, this study will integrate several robust approaches to the classical MCUSUM control chart. These approaches used robust location and scale estimator to substitute the usual mean and covariance matrix, respectively into the classical MCUSUM. The two robust location estimators used are the modified one-step M estimator (MOM) and Hodges Lehmann estimator (HL). Then, a scale estimator named Mad_n was introduced and functioned accordingly to the robust location estimators. Altogether, two robust MCUSUM control charts were proposed. The performance of each control chart was monitored based on their probability in detecting mean shifts. Various conditions were created to investigate the performance of proposed and classical control chart, namely the subgroup size (m), number of quality characteristics (p), and the level mean shifts (μ_1). The simulation results show that all the proposed charts are able to outperform the classical chart in term of their probability of detecting mean shift. This shows that the proposed robust MCUSUM charts can be used as an alternative if outliers or shifts happen to present in the dataset.

Keywords: Robust, Control chart, MCUSUM

INTRODUCTION

The statistical process control (SPC) is an important tool that is widely used to detect assignable cause in a process. According to Montgomery (2009) the control chart is very likely the most advanced and popular among researchers and industries apart from other tools like histogram, stem-and-leaf plot, check sheet or scatter diagram. The purpose of control chart is to detect the presence of assignable cause of process shifts quickly. Shewhart control chart is very effective in detecting large shifts but is shows some drawback in detecting small shifts in a process. Hence, the cumulative sum (CUSUM) control chart and exponentially weighted moving average (EWMA) control chart are used by practitioner as alternatives to the Shewhart control chart.

Usually, there are more than one quality characteristics that need to be monitored and most of the quality characteristics in a process are highly correlated. Zhang and Chang (2008) stated that by using the univariate control chart separately, the out-of-control condition cannot be detected because it ignores the correlation between the variables. To overcome this limitation, it is suggested to use the multivariate control chart as this chart can monitor the interactions of several process variable simultaneously (Yang and Trewn, 2004).

There are three multivariate control chart that are well received by researchers. They are the Hotelling T^2 control chart, multivariate exponentially weighted moving average (MEWMA) control chart and multivariate cumulative sum (MCUSUM) control chart. The Hotelling T^2 control chart is the direct counterpart of Shewhart \bar{x} control chart for univariate case (Montgomery, 2009). Therefore, it is less effective to detect small and moderate shift in the mean vector (Yang and Trewn, 2004; Montgomery, 2009). The MEWMA and MCUSUM are the alternatives to the Hotelling T^2 when small or the moderate shift of the process mean is in interest. It is important to mention here that what differ the EWMA and CUSUM control charts is how we obtain the information from the observations in the historical data set. All observations in the

CUSUM procedure are weighted equally while in EWMA we assign less and less weight to the past observations than the current observation.

According to Vargas (2003), the classical multivariate control chart is simply based on the sample mean vector, \bar{x} that represents the centre of quality characteristics and the sample variance-covariance matrix, S that represent the dispersion of data from the \bar{x} . These two parameters are known to be sensitive to outliers and/or shifts and will be greatly influence by their presence. Due to this problem, a modification on the classical control chart based on robust statistics are done by many researchers and they showed better performance as predicted.

Alloway and Raghavachari (1990) proposed a robust Hotelling T^2 control chart based on trimmed mean and trimmed covariance matrix for bivariate case. They tested the proposed robust method on symmetrical distribution and it is proved that the method is robust and resistant to the contamination observations. Later, Abu-Shawiesh and Abdullah (2001) developed a new robust Hotelling T^2 for bivariate data using Hodges-Lehmann estimator as location estimator and Shamos-Bickel-Lehmann estimator as scale estimator. Jamaluddin et al. (2018) developed a new robust MEWMA control chart based on modified one-step M (MOM) estimator and Winsorized modified one-step M(WM) estimator as the location estimator. They also use the Mad_n to replace the classical variance in the covariance matrix, S . Their performance is then compared to the classical MEWMA control chart for bivariate data of normal distribution under different conditions that include proportion of outliers and process mean shifts. The performance of each control chart is monitored through their false alarm rates. Other study about the modified one-step M(MOM) estimator can be found in Wilcox (2003) and Melik et al. (2018). The studies from previous researchers have encourage us to do some modification on the classical MCUSUM control chart using several robust estimators. In this paper, we will use two robust location estimators namely the modified one-step M (MOM) estimator and Hodges-Lehmann (HL) estimator with the median absolute deviation (Mad_n) as the robust scale estimator.

CLASSICAL MULTIVARIATE CUSUM CONTROL CHART

Let $x_j = (x_{1j}, \dots, x_{pj})$, $j = 1, \dots, p$ be a sample from multivariate normal distribution with mean zero and identity covariance matrix I_p , where p is the number of quality characteristics. The multivariate CUSUM statistic is as follows:

$$c_i^2 = D_i^T \sum^{-1} D_i, \text{ where } D_i = \sum_{j=1}^m (x_j - \mu) \quad (1)$$

However, since the value of μ and \sum is unknown, the parameters have to be estimated by using \bar{x} vector and S covariance matrix, respectively as follows:

$$c_i^2 = D_i^T S^{-1} D_i, \text{ where } D_i = \sum_{j=1}^m (x_j - \bar{x}) \quad (2)$$

such that $\bar{x} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix}$, where the arithmetic mean of j -th vector calculated using the following

formula:

$$\bar{x}_j = \frac{1}{m} \sum_{i=1}^m x_{ij}, \text{ for } j = 1, \dots, p. \quad (3)$$

and the covariance matrix, S as follows:

$$S = \begin{bmatrix} s_1^2 & L & s_{1p} \\ M & O & M \\ s_{p1} & L & s_p^2 \end{bmatrix}$$

where the variance and covariance of the x_j as follows respectively:

$$s_j^2 = \frac{1}{m_j} \sum_{i=1}^{m_j} (x_{ij} - \bar{x}_j)^2 \quad (4)$$

$$s = \text{cov}(x_j, x_g) = \frac{1}{1-m} \sum_{i=1}^m (x_{ij} - \bar{x}_j)(x_{ig} - \bar{x}_g) \quad (5)$$

where $j=1, \dots, p$; $g=1, \dots, p$; $j \neq g$.

Robust MCUSUM Control Chart Using Modified One-Step M Estimator (MOM)

The first approach uses the MOM as the location measure to replace the usual mean vector in the classical MCUSUM control chart. The MOM estimator (Wilcox and Keselman, 2003) is defined as follows:

$$MOM_j = \sum_{i=i_1+1}^{m_j-i_2} \frac{x_{ij}}{m_j - i_1 - i_2} \quad (6)$$

where x_{ij} is the i th order statistic in j th characteristics variable.

$$i_1 : \text{Number of } x_{ij} \text{ satisfies the criterion } (x_{ij} - \hat{M}_j) < (K * (Mad_{nj})) \quad (7)$$

$$i_2 : \text{Number of } x_{ij} \text{ satisfies the criterion } (x_{ij} - \hat{M}_j) > (K * (Mad_{nj})) \quad (8)$$

m_j : Denote the group size for j th variable

$$\hat{M}_j = \text{med}\{x_{1j}, \dots, x_{nj}\}, \quad j = 1, \dots, p$$

$$Mad_{nj} = 1.4826 * \text{med}_i \{|x_{ij} - \hat{M}_j|\}. \quad (9)$$

The constant $K=1.4826$ was used so that the efficiency is good under normality especially for small sample size (Wilcox and Keselman, 2003; Syed Yahaya et al., 2006). Wilcox and Keselman (2003) found that the efficiency of MOM estimator is equal to 0.9 for $m=20$ when $K=2.24$. Hence, the value $K=2.24$ will be used throughout this study to MOM approaches. To construct the new robust MCUSUM statistic, the MOM will replace the usual mean vector in the classical MCUSUM statistic, so we get:

$$c_i^2 = D_i^T S^{-1} D_i, \quad \text{where } D_i = \sum_{i=1}^m (x_i - MOM) \quad (10)$$

Robust MCUSUM control chart using Hodges-Lehmann estimator (HL)

The computation of Hodges-Lehmann estimator is showed below (Abu Shawiesh and Abdullah, 2001; Majid, Haron and Midi, 2010).

1. Calculate the Walsh averages w_r by using $w_{rj} = \frac{x_{ij} + x_{kj}}{2}$, where $r = 1, \dots, M$ such as $M = \frac{m(m+1)}{2}$, $i \leq k$ and $i, k = 1, \dots, m$, $j = 1, \dots, p$, m is the group size.

2. Reorder the Walsh averages is ascending order

$$w_{1j} \leq w_{2j} \leq w_{3j} \leq \dots \leq w_{Mj} \quad (11)$$

3. Compute the *HL* estimator

$$HL_j = \text{median}\{w_{1j}, w_{2j}, \dots, w_{Mj}\} \quad (12)$$

Now, we can compute the robust MCUSUM statistic for *HL* estimator:

- i. Let x_{i1}, \dots, x_{ip} , a matrix of $m \times p$ where $i = 1, \dots, m$, with m the number of observations and p in the number of quality characteristics.
- ii. Calculate the mean vector using Hodges-Lehmann estimator for matrix of $m \times p$ as follows:

$$HL_p = \begin{bmatrix} HL_1 \\ \mathbf{M} \\ HL_p \end{bmatrix} \quad (13)$$

- iii. The formula for the new MCUSUM control chart using Hodges-Lehmann as the location estimator is

$$c_i^2 = D_i^T S^{-1} D_i, \text{ where } D_i = \sum_{i=1}^m (x_i - HL) \quad (14)$$

The variance covariance matrix, s will be replaced with robust scale estimator, Mad_n . The computation of the Mad_n will be explain in the next subsection.

Robust MCUSUM control chart using Median Absolute Deviation estimator (Mad_n)

The $p \times p$ covariance matrix of Mad_n is denoted as s_{Mad_n} . First, compute the diagonal element of the $p \times p$ covariance matrix which are the sample variances of each variable which are represented by

$$\text{Mad}_{nj}^2 = \text{Mad}_n(x_{ij})\text{Mad}_n(x_j) \text{ where } j = 1, \dots, p \quad (15)$$

According to Abu-Shaweish and Abdullah (2001), the remaining element of $p \times p$ covariance matrix is calculated as follows:

- i. Compute the Mad_n for vectors x_j and x_g which are denoted by $Mad_n(x_j)$ and $Mad_n(x_g)$ where $j = 1, \dots, p$ and $g = 1, \dots, p$, $j \neq g$.
- ii. Compute the Spearman correlation for ranks between the variables x_j and x_g , which are denoted by $corr(x_j, x_g)$.
- iii. The sample covariance between the variables x_j and x_g is

$$Mad_{jg} = Mad_n(x_j)Mad_n(x_g)corr(x_j, x_g) \quad (16)$$

- iv. Thus, the $p \times p$ covariance matrix is

$$S_{Mad_n} = \begin{bmatrix} Mad_1^2 & L & Mad_{1p} \\ M & O & M \\ Mad_{p1} & L & Mad_p^2 \end{bmatrix} \quad (17)$$

Hence, the formula for the robust MCUSUM control charts will be as follows:

1. $c_{MOMMad}^2(x_i) = D_i^T S_{Mad_n}^{-1} D_i$, where $D_i = \sum_{i=1}^m (x_i - MOM)$ (18)
2. $c_{HLMad}^2(x_i) = D_i^T S_{Mad_n}^{-1} D_i$, where $D_i = \sum_{i=1}^m (x_i - HL)$ (19)

PERFORMANCE OF CONTROL CHARTS

This study deals with the classical MCUSUM control charts, which are sensitive to any shifts in data sets. Hence, the data are generated from standard normal distributions, contaminated with different mean shifts. To monitor the performance and capability of these new robust MCUSUM control charts, the probability of detecting mean shifts is calculated. For the probability of the detecting mean shifts, the higher the probability, the better the charts perform in detecting mean shifts. In order to check the performance of all MCUSUM control charts, we set the values of quality characteristics, p at 2, 3, 5 and 10 with subgroup sizes, $m = 30, 50, 100$ and 400. Five levels of mean shifts (μ_i) were used that are 0.5, 1.0, 1.5, 2.0 and 5.0 with proportion equal to 0.1.

In a multivariate control chart, only the upper control limit (UCL) is required since the lower control limit is always set at 0. For this study, the simulation method is utilised to get the UCL of all MCUSUM as done by Alfaro and Ortega (2009) and Jamaluddin (2018). This is because the underlying distribution is unknown for the robust statistic. The simulated UCL of all control charts is obtained by generating 5000 data set of subgroup size m such that $x_j = (x_{1j}, \dots, x_{ij})$, $j = 1, \dots, p$ from in-control condition, $MVN_p(\mathbf{0}, \mathbf{I}_p)$. Then calculate all the related classical and robust MCUSUM statistics, c^2 for each observation in the generated data set and record the maximum value for the corresponding dataset. In this study, the overall false alarm is set at $\alpha = 0.05$. In order to retain $\alpha = 0.05$, the 95th percentile of the recorded maximum values are declared as the simulated UCLs.

The evaluation of the classical and robust MCUSUM control chart are based on their probability of detecting shifts according to the following steps:

1. Generate new observation vector for each data set from ‘out-of-control’ condition based on mean shifts decided.
2. Calculate the classical and robust MCUSUM statistics, c^2 for each new observation vector.
3. Compare each value of the MCUSUM statistics with their corresponding simulated UCL.
4. Repeat steps 1 to 3 for 5000 replications.
5. The probability of detecting mean shifts is equal to the proportion of data sets that have at least one point greater than the simulated UCL.

Computer program in *R* were developed to calculate the simulated UCL and probability of detecting shift.

RESULTS AND DISCUSSION

The results of the analysis for the probability of detecting mean shifts when $\alpha = 0.05$ are summarised in Table 1 for $p = 2$ and 3 and Table 2 for $p = 5$ and 10. For a clearer and better comparison, we will refer to the graphs in Figure 1 to Figure 4. Figure 1 presents the performance of the investigated charts in terms of their probability in detecting shifts when the quality of characteristics (p) equals to 2.

As shown in Figure 1a and 1b, when small mean shift occurs, the c_{HLMad}^2 chart appears to have the best performance for all subgroup size m , followed by c_{MOMMad}^2 chart and lastly the c^2 chart. However, in Figure 1c and 1d, when the mean shifted to 1.5 and 2.0, the c^2 perform best for smaller subgroup sizes and as the subgroup sizes increase, all control charts give comparable performance. For large mean shift ($\mu_1 = 5$), all control charts are able to detect all outliers cause by mean shifts since the probability of detecting mean shift is equal to 1 as shown in Figure 1e. From all five of mean shift, the performance of each control chart is varied for smaller m , and becoming similar as the m increasing.

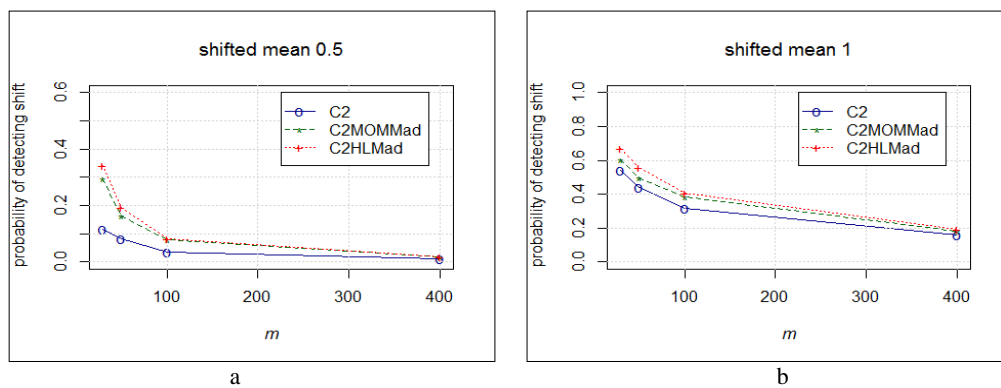
Table 1: The probability of detecting mean shifts for $p = 2$ and 3

m	μ_1	$p = 2$			$p = 3$		
		c^2	c_{MOMMad}^2	c_{HLMad}^2	c^2	c_{MOMMad}^2	c_{HLMad}^2
30	0.5	0.115	0.296	0.343	0.103	0.350	0.428
	1	0.539	0.601	0.672	0.464	0.622	0.720
	1.5	0.937	0.868	0.917	0.894	0.868	0.930
	2	0.998	0.981	0.992	0.996	0.973	0.993
	5	1	1	1	1	1	1
50	0.5	0.083	0.164	0.195	0.067	0.196	0.223
	1	0.439	0.495	0.559	0.370	0.495	0.582
	1.5	0.908	0.849	0.893	0.842	0.848	0.886
	2	0.997	0.986	0.991	0.989	0.981	0.990
	5	1	1	1	1	1	1
100	0.5	0.034	0.079	0.081	0.033	0.079	0.089
	1	0.315	0.386	0.408	0.229	0.360	0.403
	1.5	0.832	0.826	0.852	0.714	0.780	0.822
	2	0.993	0.986	0.991	0.974	0.975	0.982
	5	1	1	1	1	1	1
400	0.5	0.011	0.017	0.020	0.008	0.014	0.013

1	0.162	0.184	0.192	0.087	0.117	0.129
1.5	0.665	0.673	0.676	0.489	0.541	0.580
2	0.967	0.969	0.969	0.918	0.925	0.931
5	1	1	1	1	1	1

Table 2: The probability of detecting mean shifts for $p = 5$ and 10

m	μ_1	$p = 5$			$p = 10$		
		c^2	c^2_{MOMMad}	c^2_{HLMad}	c^2	c^2_{MOMMad}	c^2_{HLMad}
30	0.5	0.099	0.489	0.606	0.072	0.700	0.917
	1	0.359	0.720	0.816	0.237	0.886	0.968
	1.5	0.806	0.892	0.959	0.602	0.955	0.994
	2	0.987	0.983	0.994	0.913	0.992	0.999
	5	1	1	1	1	1	1
50	0.5	0.054	0.241	0.298	0.053	0.401	0.545
	1	0.237	0.513	0.624	0.153	0.633	0.773
	1.5	0.674	0.819	0.900	0.430	0.860	0.941
	2	0.955	0.972	0.989	0.816	0.970	0.994
	5	1	1	1	1	1	1
100	0.5	0.024	0.078	0.094	0.027	0.104	0.147
	1	0.121	0.308	0.349	0.063	0.280	0.365
	1.5	0.516	0.699	0.781	0.252	0.598	0.717
	2	0.906	0.950	0.965	0.606	0.881	0.949
	5	1	1	1	1	1	1
400	0.5	0.006	0.009	0.007	0.005	0.006	0.005
	1	0.045	0.065	0.072	0.011	0.022	0.030
	1.5	0.270	0.358	0.368	0.072	0.146	0.158
	2	0.747	0.798	0.822	0.283	0.472	0.492
	5	1	1	1	1	1	1



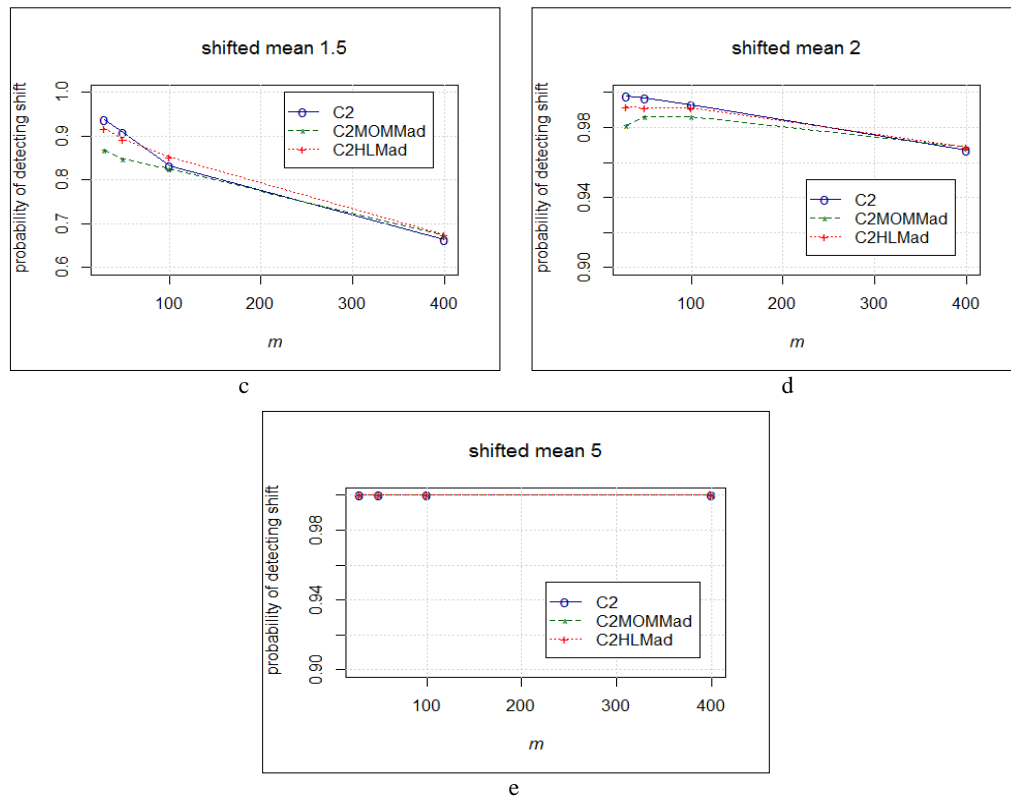
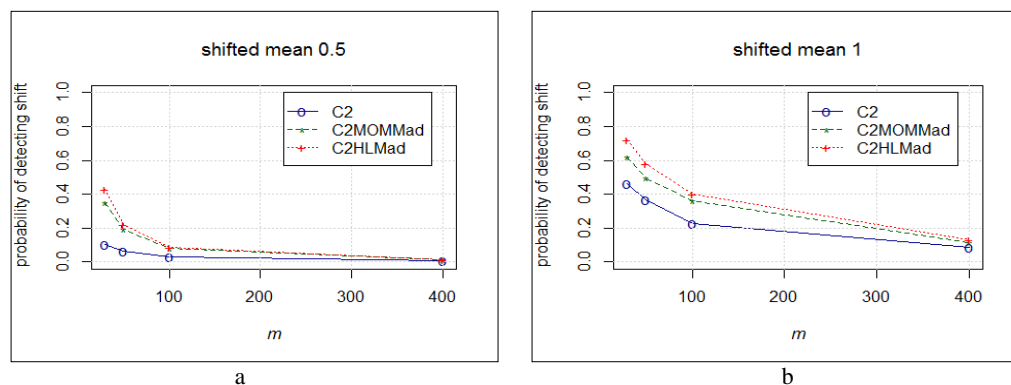


Figure 1: Probability of detecting shift when $p = 2$.

Next, Figure 2 represent the performance of the investigated charts for various mean shift at $p = 3$. Throughout the figures, it could be easily observed that c^2 control chart has the worst performance for most condition while the c^2_{HLMad} mostly showed better performance than another two control charts. At $\mu_1 = 0.5$ and 1.0 as shown in Figure 2a and b, c^2_{HLMad} showed the highest probability followed by c^2_{MOMMad} and c^2 for small subgroup sizes and at $m = 400$, the probability of detecting shifts for all control chart are more or less similar. As mean shift to $\mu_1 = 1.5$ and 2.0 , c^2 gives better performance at smaller m but overtook by other two robust MCUSUM as the m increase. Similar with previous case, for large mean shift ($\mu_1 = 1.5$), all control charts are able to detect all outliers cause by mean shift since the probability of detecting mean shift is equal to 1.0 as shown in Figure 2e.



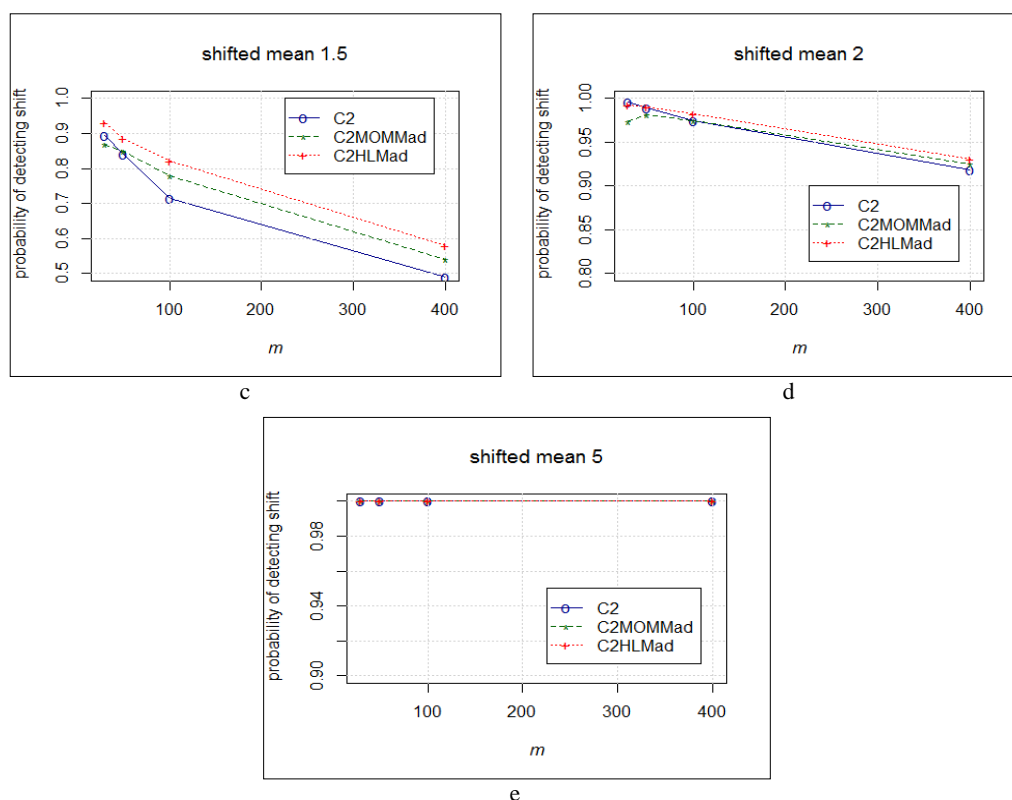
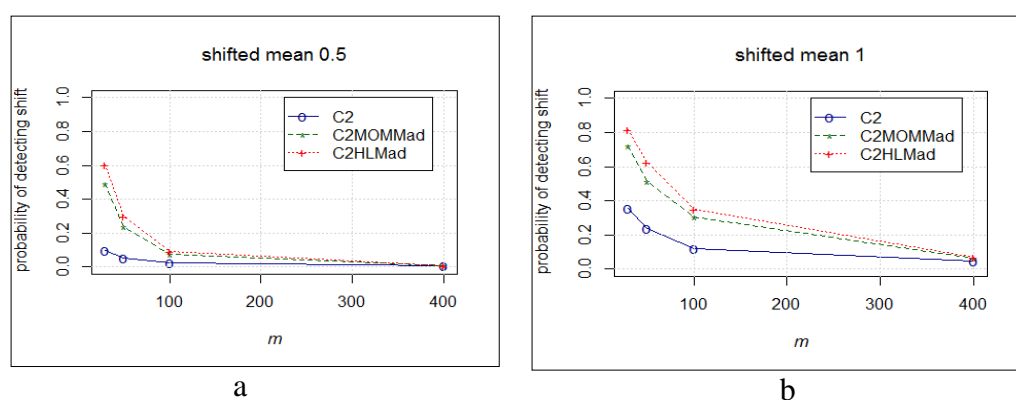


Figure 2: Probability of detecting shift when $p = 3$

Figure 3 represents the performance of the control charts when $p = 5$. The patterns are observed to be almost similar to $p = 2$ and $p = 3$ with c^2_{HLMad} control chart produced the best performance followed by c^2_{MOMMad} and c^2 . For smaller shifted mean as shown in Figure 3a and 3b, the probability of detecting shift for all control charts are always differ at small m but comparable as $m = 400$. Lastly, all control charts are able to detect mean shift when $\mu_1 = 1.5$ as can be seen in Figure 3e.



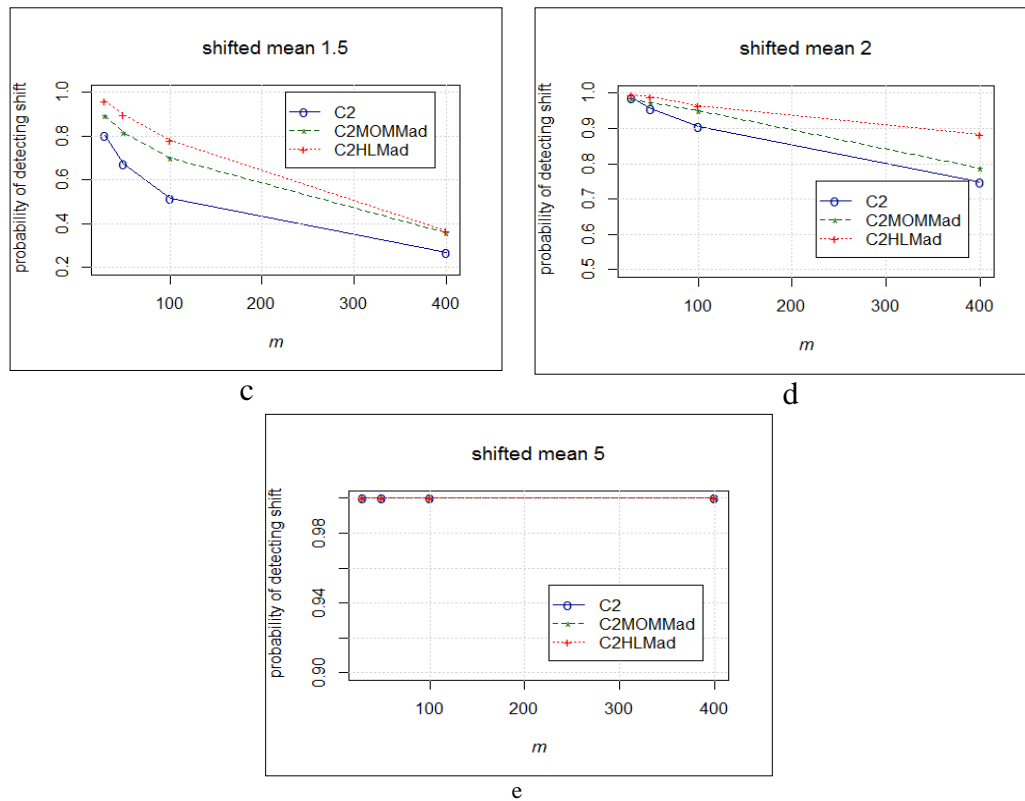
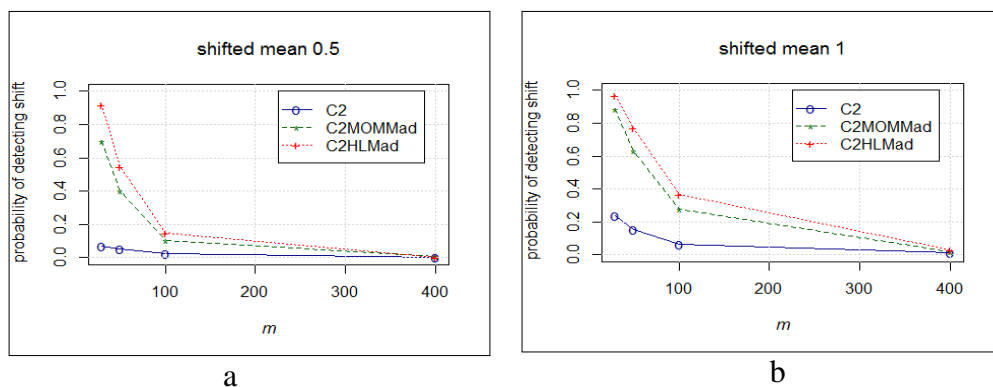


Figure 3: Probability of detecting shift when $p = 5$

Figure 4a, 4b and 4c illustrate the case of mild and moderate contamination when the mean is shifted at $\mu_1 = 0.5, 1.0$ and 1.5 , the two robust control charts perform better than the classical chart. There is a downward trend as the value of the m increase for all control chart. This means that the bigger the subgroup size, the smaller the probability of detecting the mean shift. For condition where shifted mean is at 2.0 based on Figure 4d, when m is small, the performance of all three charts is quite similar. However, as the m increase, the performance gap between the two robust control chart and the classical chart become further. Finally, based on Figure 4e, all control chart is able to detect shifted mean at 5 for all subgroup size.



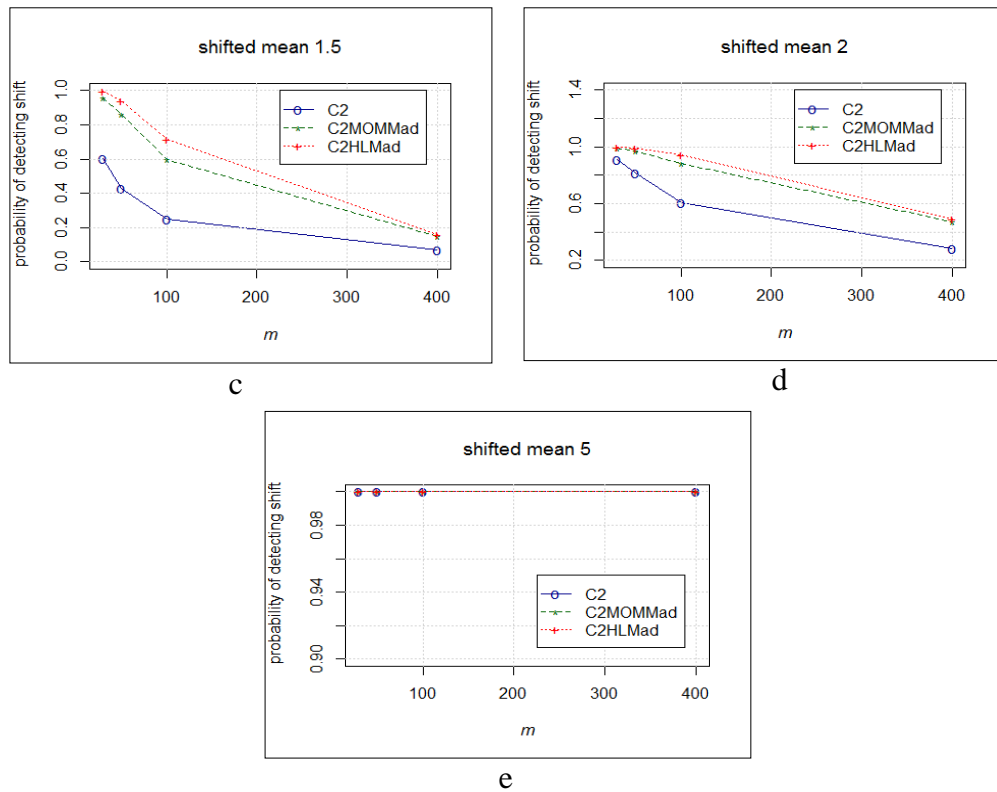


Figure 4: Probability of detecting shift when $p = 10$

CONCLUSIONS

In order to apply and check the performance of control chart, the presence of mean shifts or outliers in the data set should be considered and investigated. The presence of mean shifts or outliers can cause misleading conclusion of the control chart if one uses the classical approach. Therefore, robust statistics are used to overcome this problem. Modified one-step M estimator and Hodges-Lehmann estimator are used by integrating it with the classical MCUSUM control chart to form robust control charts which are the c_{MOMMad}^2 and c_{HLMad}^2 . From the simulation study, the probability of detecting mean shifts of all control chart is computed and compared. Data set from Normal distribution with different mean shifts with proportion 0.1 are used. The subgroup size, $m = 30, 50, 100$ and 400 and number of quality characteristics, p at $2, 3, 5$ and 10 are included in this study. As the conclusion, c_{HLMad}^2 control chart has the best performance among these three control charts when the mean shifts happen to present in the data set. Besides, c_{MOMMad}^2 and c_{HLMad}^2 also perform better than the c^2 control chart for smaller mean shifts.

REFERENCES

- Abu-Shawiesh, M. O. and Abdullah, M. B. (2001). A new robust bivariate control chart for location. *Communications in Statistics-Simulation and Computation*, 30(3): 513-529.
- Alfaro, J. L. and Ortega, J. F. (2009). A comparison of robust alternatives to Hotelling's T^2 control chart. *Journal of Applied Statistics*, 36(12): 1385-1396.
- Alloway, J. J. and Raghavachari, M. (1990). Multivariate control charts based on trimmed means. In *ASQC Quality Congress Transactions—San Francisco* (pp. 449-453).

- Jamaluddin, F., Ali, H. H., Syed Yahaya, S. S. and Zain, Z. (2018). The performance of robust multivariate Ewma control charts. *The Journal of Social Sciences Research*, (6): 52-58.
- Lowry, C. A., Woodall, W. H., Champ, C. W. and Rigdon, S. E. (1992). A multivariate exponentially weighted moving average control chart. *Technometrics*, 34(1): 46-53.
- Majid, A., Haron, K. and Midi, H. (2010). Hodges Lehmann Estimator for Robust Design Using Dual Response Surface Approach. *Australian Journal of Basic and Applied Sciences*, 4(8): 3418-3423.
- Melik, H. N., Ahad, N. A. and Yahaya, S. S. S. (2018). Modified One-Step M-Estimator with Robust Scale Estimator for Multivariate Data. *Journal of Engineering and Applied Science*, 13(24): 10396-10400.
- Montgomery, D. C. (2009). *Introduction to statistical quality control*. John Wiley & Sons.
- Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, 41(1/2): 100-115.
- Roberts, S. W. (1959). Control chart tests based on geometric moving averages. *Technometrics*, 1(3): 239-250.
- Syed Yahaya, S. S., Othman, A. R. and Keselman, H. J. (2006). Comparing the " typical score" across independent groups based on different criteria for trimming. *Metodoloski zvezki*, 3(1): 49-62.
- Vargas, N. J. A. (2003). Robust estimation in multivariate control charts for individual observations. *Journal of Quality Technology*, 35(4): 367-376.
- Wilcox, R. (2003). Multiple comparisons based on a modified one-step M-estimator. *Journal of Applied Statistics*, 30(10): 1231-1241.
- Wilcox, R. R. and Keselman, H. J. (2003). Repeated measures one-way ANOVA based on a modified one-step M-estimator. *British Journal of Mathematical and Statistical Psychology*, 56(1): 15-25.
- Yang, K. and Trewn, J. (2004). *Multivariate statistical methods in quality management*. McGraw Hill Professional.
- Zhang, G. and Chang, S. I. (2008). Multivariate EWMA control charts using individual observations for process mean and variance monitoring and diagnosis. *International Journal of Production Research*, 46(24): 6855-6881.