

## Robust Hotelling's $T^2$ Charts with Median based Trimmed Estimators

<sup>1</sup>S.S. Syed-Yahaya, <sup>2</sup>F.S. Haddad, <sup>1</sup>N.I. Mahat, <sup>1</sup>A. Abdul Rahman and <sup>1</sup>H. Ali

<sup>1</sup>School of Quantitative Sciences, UUM College of Arts and Sciences, Universiti Utara Malaysia,  
06010 Sintok, Kedah, Malaysia

<sup>2</sup>College of Applied Studies and Community Service, University of Dammam, Dammam, KSA  
sharipah@uum.edu.my

---

**Abstract:** It is well known that the traditional Hotelling's  $T^2$  chart is inefficient when outliers are presented in the data. To alleviate the problem, this study proposed three robust Hotelling's  $T^2$  control charts using trimmed estimators to replace the usual mean vector and the covariance matrix in the traditional  $T^2$  chart. Trimming was done using modified Mahalanobis distance with median as the location measure and one of the robust scale estimators:  $MAD_n$ ,  $S_n$  or  $T_n$  as the scale measure. These modifications consequently produced three different trimmed estimators. Investigation on the performance of the proposed robust charts, measured in terms of false alarm and probability of detection were conducted via simulation study. The results were compared to the traditional chart. It was discovered that the proposed control charts performed moderately in terms of controlling false alarm and excellently in terms of probability of detection, surpassing the performance of traditional control chart regardless of conditions.

**Key words:** Hotelling's  $T^2$  control chart, trimmed estimator, statistics, multivariate, measure, modification

---

### INTRODUCTION

In manufacturing process, multi quality characteristics of a product are often observed. Thus, multivariate control chart can be a suitable tool to monitor the process. However, prudent care shall be exercised in utilizing the tool. This is because along with the increased number of quality characteristics to be monitored, chances of having multiple outliers in the data become greater. In this situation, maximal performance of a control chart may be affected.

The most common approach to simultaneously monitor multivariate measurements is through the Hotelling's  $T^2$  statistic (Alt, 1985; Montgomery, 2005). The Hotelling's  $T^2$  statistic is the multivariate generalization of the student's  $t$ -statistics. To monitor a process over time, the Hotelling  $T^2$  chart is constructed in two phases, namely Phase I and II. In Phase I, the aim is to obtain a stable historical data set. This dataset is then used to estimate the mean vector and variance-covariance matrix. The estimates of these parameters are used together with the control limit to develop the control chart. Meanwhile in Phase II, the constructed control chart is then used to monitor the process.

The traditional Hotelling's  $T^2$  control chart is a reliable tool when the underlying process data actually

follow the normal distribution. However, the control chart is no longer reliable as it will spuriously identify out of control observations when outliers are present in the dataset. This is because the maximum likelihood estimators of the chart, namely the mean and covariance are sensitive to the outliers. Therefore, the capability of the traditional Hotelling's  $T^2$  control chart to monitor future process is arguable. One of the solutions to overcome this problem is to use control chart that is robust to the presence of outliers.

To date, numerous robust control charts have been proposed in the literature. Alloway and Raghavachari employed the so-called trimmed mean and trimmed covariance matrix in place of arithmetic mean and the covariance matrix, respectively. Alternatively, Surtihadi (1994) used median as a robust location estimator when he constructed robust bivariate control charts based on the bivariate sign tests of Blumen and Hodges. In other research, Abu-Shawiesh and Abdullah (2001) estimated the mean vector using Hodges-Lehmann and the variance-covariance matrix using Shamos-Bickel-Lehman. Meanwhile, Alfaro and Ortega (2008) proposed a new alternative robust Hotelling's  $T^2$  control charts by replacing the arithmetic mean with trimmed mean and sample covariance with sample trimmed covariance. Yanez *et al.* (2010) constructed the  $T^2$  control chart based

on the bi weight S estimators for location and dispersion parameters. The proposed chart was shown to outperform the  $T^2$  chart based on Minimum Volume Ellipsoid (MVE) for a small number of observations. Comparison between the MVE and Minimum Covariance Determinant (MCD) approaches in improving the  $T^2$  chart performance under non-normality was engaged by Alfaro and Ortega (2008). For such purpose, the researchers studied the performance of the  $T^2$  charts in Phase II process when Phase I data are distributed as student's t-distribution. The results favoured the MCD approach when the  $T^2$  chart performance was examined under severe non-normality situation but either approach (the MCD or MVE) was deemed suitable under a slight deviation from normality. The outcomes were validated based on the percentage of out-of-control observations detected by these charts. However, the MVE and MCD charts may have poor control of false alarm rates. Alfaro and Ortega (2009) and Yahaya *et al.* (2011) for a review. In view of this conflict, Yahaya *et al.* (2011) introduced the Minimum Variance Vector (MVV) estimator in the  $T^2$  chart for monitoring the Phase II dataset. Overall, the proposed robust chart gave a quick detection in the out-of-control status and at the same time, able to control the overall false alarm rates even as the dimensions increased. The only drawback, however was a large Upper Control Limits (UCLs) as compared to the traditional  $T^2$  chart. An improved version of the MVV chart was further recommended by Ali *et al.* (2013) to obtain desired UCLs whilst still maintaining its good performance in terms of false alarm rate and probability of detection. This was achieved by making the MVV estimators consistent at normal distribution as well as unbiased for finite samples. More recently, Abu-Shawiesh *et al.* (2014) proposed a new bivariate control chart with the same structure as Hotelling  $T^2$  chart but using the sample median, Median Absolute Deviation from the sample median (MAD) and Comedian estimator (COM). This chart outperformed the non-robust  $T^2$  chart for all cases considered in that study.

From the above-mentioned researches, it is apparent that robust Hotelling  $T^2$  chart is far more reliable than the traditional  $T^2$  chart when normality assumption cannot be guaranteed. There is a wide spread choice of robust location and scatter parameters that one may consider in this case. Narrowing down the choices, this study proposed to improve the performance of Hotelling's  $T^2$  control chart by replacing the mean vector and covariance matrix with trimmed mean vector and its corresponding covariance matrix following Alloway and Raghavachari. Three different trimmed means based on three different robust scale estimators namely;  $MAD_n$ ,  $S_n$  and  $T_n$  were proposed.

## MATERIALS AND METHODS

### Multivariate control chart characteristics

**Hotelling's  $T^2$  control chart:** Let  $x_i = \{x_{i1}, x_{i2}, x_{i3}, \dots, x_{ip}\}$  be the p-variate random sample of n observations for  $i = 1, 2, \dots, n$ ,  $\bar{x}$  is the sample mean vector and S is the  $p \times p$  sample covariance matrix. An individual Hotelling's  $T^2$  statistic can generally be expressed as:

$$T^2(x_i) = (x_i - \bar{x})^T S^{-1} (x_i - \bar{x}) \quad (1)$$

The original work on  $T^2(x_i)$  relies on maximum likelihood estimators  $\bar{x}$  and S. Nonetheless, measurements from the  $\bar{x}$  and S are inappropriate when process data are distorted. A more appropriate way is to use robust estimators in the place of sample mean vector and sample covariance matrix.

**Robust scale estimators:** Dispersion of a distribution can be explained via. some scale measures. A fine scale estimator is typically characterized by a high value of breakdown point. Scale estimators such as  $MAD_n$ ,  $S_n$  and  $T_n$  which are employed in this study, possess the highest breakdown point and bounded influence function as proven by Rousseeuw and Croux (1993). Furthermore, they are simple and easy to compute. Let  $x_1, x_2, \dots, x_n$  be a sample set of data and a brief explanation of the estimators is discussed.

**$MAD_n$ :**  $MAD_n$  is median absolute deviation about the median which is given by:

$$MAD_n = 1.4826 \times \text{med} \left\{ |x_i - \text{med} \{x_1, x_2, \dots, x_n\}| \right\} \text{ for } i = 1, 2, \dots, n \quad (2)$$

This scale estimator is very robust with best possible breakdown point and bounded influence function.  $MAD_n$  is the single most useful ancillary estimate of scale due to its high breakdown property.

**$S_n$ :** Despite all the advantages mentioned about  $MAD_n$ , this estimator is confined to some draw backs. It has low efficiency (37%) at Gaussian distributions and it takes a symmetric view on dispersion. Therefore, when dealing with skewed distributions, Rousseeuw and Croux (1993) came up with an estimator which is similar to  $MAD_n$  but not slanted towards symmetry. The estimator which is known as  $S_n$  is defined as:

$$S_n = c \times \text{med}_i \left\{ \text{med}_j |x_i - x_j| \right\} \text{ for } i, j = 1, 2, \dots, n \text{ and } i < j \quad (3)$$

where,  $S_n$  is a location free estimator which looks at typical distance between observations. It has an explicit formula

which means that this formula is always uniquely defined. A modest simulation studied by Rousseeuw and Croux (1993) found that the correction factor  $c = 1.1926$  succeeded in making  $S_n$  unbiased for finite samples.

**T<sub>n</sub>**: Another promising scale estimator is  $T_n$  which is defined as:

$$T_n = 1.38 \times \frac{1}{h} \sum_{k=1}^h \{ \text{med} |x_i - x_j| \}_{(k)} \text{ for } i, j = 1, 2, \dots, n \text{ and } i \neq j$$

Where: (4)

$h = (n/2)+1$

$k =$  The No. of trimmed observations

$T_n$  has all the properties as a robust estimator needs such as a bounded and continuous influence function and a high breakdown point (Rousseeuw and Croux, 1993).

**Trimmed mean:** Trimming is a process that aims to remove the extreme values from each tail of the ordered statistics. Meanwhile, trimmed mean is the mean computed from the trimmed data. There are suggestions on the amount of observations to be trimmed including 20% from each tail of the ordered statistics as recommended by Rosenberger and Gasko (1983) as well as 10-15% as discussed by Othman *et al.* (2004). However, their works deal with univariate data. In the multivariate aspect, the trimming procedure is more complicated. Alloway and Raghavachari examined three methods of trimming. One of the methods suggested the use of Mahalanobis Squared Distance (MSD) where it selects the observations that need to be trimmed. According to Rocke *et al.* (1982) this process resulted in providing the best percentage of the amount of trimming from each side of any ordered statistics in range of 20-25% in symmetric distribution. The MSD formula is given as:

$$MSD(\mathbf{x}) = (\mathbf{x} - \bar{\mathbf{x}})^T S^{-1} (\mathbf{x} - \bar{\mathbf{x}}) \quad (5)$$

where  $\bar{\mathbf{x}}$  and  $S$  depend on the original data. Alloway and Raghavachari selected a set of data pair of individual observations to be trimmed. Similar to the traditional Hotelling's  $T^2$  statistic, this statistic uses the mean and covariance which are known to be sensitive to outliers and it is unlikely to use the MSD to find outliers, since, MSD itself is sensitive to outliers. For such reason, this study modified the MSD procedure to be used as the trimming tool by substituting the  $\bar{\mathbf{x}}$  and  $S$  with robust location vector and scale matrix.

For the location measure, this study proposed using the usual median while for the scale measure,

three different scale estimators from the research by Rousseeuw and Croux (1993) namely  $MAD_n$ ,  $S_n$  and  $T_n$  were suggested. Apart from having highest breakdown point, these estimators performed well in controlling type I error rates when they are integrated in various test statistics (Haddad *et al.*, 2013; Yahaya *et al.*, 2004).

Let,  $x_i = x_{i1}, x_{i2}, \dots, x_{ip}$  be a vector for individual data sets where  $i = 1, 2, \dots, n$  and  $p$  is the number of characteristics. The computation of the proposed MSD for individual observations is as in Algorithm 1.

**Algorithm 1; Computation of the robust MSD:**

- Step 1: Calculate the median for each data set
- Step 2: Estimate the robust covariance matrix according to the following steps
  - Calculate the robust scale estimates  $MAD_n$ ,  $S_n$  and  $T_n$  for each pair of quality characteristics,  $x_j$  and  $x_g$  where  $j, g = 1, 2, \dots, p$  and  $j \neq g$
  - Compute the Spearman Rank correlation between  $x_j$  and  $x_g$  denoted by  $\text{corr}(x_j, x_g)$  (Abdullah, 1990)
  - Compute the covariance between  $x_j$  and  $x_g$  for the three different scale estimators using the followings:

$$S_{MADn}(x_j, x_g) = MAD_n(x_j) \times MAD_n(x_g) \times \text{corr}(x_j, x_g) \quad (6)$$

$$S_{Sn}(x_j, x_g) = S_n(x_j) \times S_n(x_g) \times \text{corr}(x_j, x_g) \quad (7)$$

$$S_{Tn}(x_j, x_g) = T_n(x_j) \times T_n(x_g) \times \text{corr}(x_j, x_g) \quad (8)$$

- Step 3: Compute the robust MSD (termed as RMSD) for each observation as below:

$$RMSD_{MADn}(x_i) = (x_i - \text{med}(x_i))^T S_{MADn}^{-1} (x_i - \text{med}(x_i)) \quad (9)$$

$$RMSD_{Tn}(x_i) = (x_i - \text{med}(x_i))^T S_{Tn}^{-1} (x_i - \text{med}(x_i)) \quad (10)$$

$$RMSD_{Sn}(x_i) = (x_i - \text{med}(x_i))^T S_{Sn}^{-1} (x_i - \text{med}(x_i)) \quad (11)$$

- Step 4: Arrange the values of each RMSD in ascending order and trim the largest 40% of the values following Alloway and Raghavachari
- Step 5: Estimate the location and scale measures for Hotelling  $T^2$  chart. Compute the Hotelling  $T^2$  statistic with trimmed means and winsorized covariance

Determine the trimmed mean for each data set of the individual observations by dividing the total of the remaining observations by  $(m-k)$  where  $m$  denotes the number of individual observations and  $k$  denotes the number of trimmed observations. The use of different scale estimators ( $MAD_n$ ,  $S_n$  or  $T_n$ ) in the computation of covariance matrix of MSD will produce different trimmed means represented by  $\bar{x}_{MADn}$ ,  $\bar{x}_{Sn}$  and  $\bar{x}_{Tn}$ , respectively  
 Create winsorized sample by replacing the trimmed observations with the value of the largest remaining observation  
 Compute the covariance matrix for the Hotelling  $T^2$  statistic by adopting Eq. 12:

$$S_{MADn} = \frac{m-1}{m_1-1} \times S_{wMADn} \quad (12)$$

Where  $S_{wMADn}$  is the winsorized covariance and  $m_1$  is the number of trimmed observations

Calculate the inverse of the covariance matrix

**The robust control charts**

**The Robust Hotelling's T<sup>2</sup> charts:** All the computed robust scale estimators and trimmed mean estimators were used to construct the proposed Hotelling's T<sup>2</sup> charts. This was done by replacing the arithmetic mean vector in the traditional Hotelling's T<sup>2</sup> control chart (Eq. 1) with any of the three trimmed mean vector ( $\bar{x}_{MADn}$ ,  $\bar{x}_{Sn}$  or  $\bar{x}_{Tn}$ ) and its corresponding inverse covariance matrix ( $S_{MADn}^{-1}$ ,  $S_{Sn}^{-1}$  or  $S_{Tn}^{-1}$ ). Such substitutions form the three robust Hotelling's T<sup>2</sup> charts:

$$T_{MADn}^2(x_i) = (x_i - \bar{x}_{MADn})^T S_{MADn}^{-1} (x_i - \bar{x}_{MADn}) \quad (13)$$

$$T_{Sn}^2(x_i) = (x_i - \bar{x}_{Sn})^T S_{Sn}^{-1} (x_i - \bar{x}_{Sn}) \quad (14)$$

$$T_{Tn}^2(x_i) = (x_i - \bar{x}_{Tn})^T S_{Tn}^{-1} (x_i - \bar{x}_{Tn}) \quad (15)$$

**Control limits:** This study focused on independent case of individual observations. To examine the strength and weakness of the proposed robust charts, the mixed normal distributions based on the following contaminated model was used:

$$(1 - \epsilon)N_p(\mu_0, \Sigma_0) + \epsilon N_p(\mu_1, \Sigma_1) \quad (16)$$

Where:

- $\epsilon$  = The percentage of the outliers
- $N_p(\mu_0, \Sigma_0)$  = The in-control distribution with the parameters  $\mu_0$  and  $\Sigma_0$
- $N_p(\mu_1, \Sigma_1)$  = The out-of-control distribution with the parameters  $\mu_1$  and  $\Sigma_1$

For independent case, the contamination model of the mixed normal distribution is:

$$(1 - \epsilon)N_p(0, I_p) + \epsilon N_p(\mu_1, I_p) \quad (17)$$

According to Johnson (2007) the variance covariance matrix  $I_p$  is a homogeneous covariance matrix where the main diagonal is 1 and 0 else where which reflects that there is no correlation among the variables.

Since, the distributions of the proposed Hotelling's T<sup>2</sup> charts are unknown, the estimation of Upper Control Limit (UCL) was done using simulation. First, data sets were generated from the standard normal distribution  $N(0, I_p)$ . Then, robust estimators were computed from this distribution. Next, a new additional observation from the standard normal distribution was generated and robust Hotelling's T<sup>2</sup> statistic for this new observation was computed. This procedure was repeated 5000 times and the 95th percentile of the 5000 robust Hotelling's T<sup>2</sup> statistics is considered as the UCL.

Table 1: Manipulated variables in simulation study

Variables	Values
No. of quality characteristics (p)	2, 5 and 10
Proportion of contamination (ε)	0.1 and 0.2
Mean shift (μ)	0 (no shift), 3 and 5
Group size (m)	50, 100 and 150

**False alarm and probability of detection:** The performance of the control charts was measured in terms of their false alarm rates and probability of detection. To compute these values, 1000 data sets were generated from the standard normal distribution  $N(0, I_p)$ . Then, the data sets were contaminated with outliers and mean shift. Next, robust estimators were obtained from these data sets. The false alarm was computed using a new observation from the in-control distribution whilst the probability of detection was calculated using a new observation generated from the out-of-control distribution.

Four variables were manipulated to investigate the strengths and the weaknesses of the robust Hotelling's T<sup>2</sup> charts namely number of quality characteristics (p), proportion of contamination (ε), mean shift (μ) and group size (m). The setting values for the variables are listed in Table 1 following earlier researches by Alfaro and Ortega (2008), Vargas (2003) and Mohammadi *et al.* (2011). The manipulation of these variables generated five different levels of contaminations which are categorized as:

- $N_p(0, I_p)$ -ideal condition (no contamination)
- $(0.9)N_p(0, I_p) + (0.1)N_p(3, I_p)$ -mild contamination
- $(0.8)N_p(0, I_p) + (0.2)N_p(3, I_p)$ -moderate contamination
- $(0.9)N_p(0, I_p) + (0.1)N_p(5, I_p)$ -moderate contamination
- $(0.8)N_p(0, I_p) + (0.2)N_p(5, I_p)$ -extreme contamination

**RESULTS AND DISCUSSION**

The results of the analysis on the performance of the robust control charts in terms of the false alarm rates and probability of detection at  $\alpha = 0.05$  are summarized in Table 2 and 3, respectively. The robust charts ( $T_{MADn}^2$ ,  $T_{Sn}^2$  and  $T_{Tn}^2$ ) were compared to the traditional control chart ( $T_0^2$ ). The shaded cells in both tables represent robust conditions.

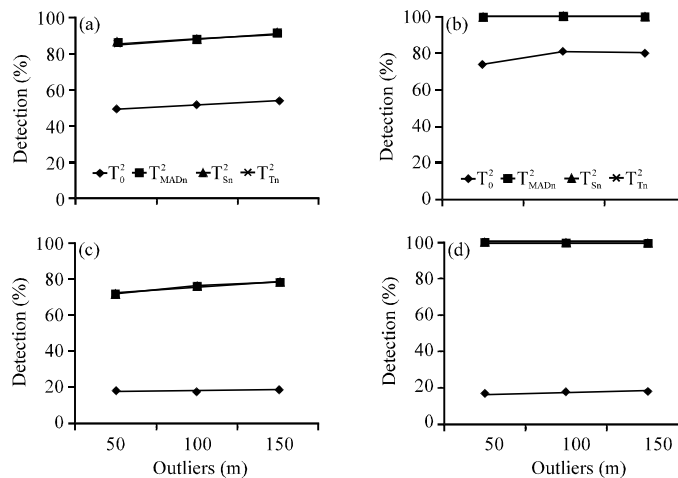
To measure the robustness of the control charts, we adopt Bradley (1978) criterion of robustness interval such that the false alarm rates should be in between 0.025 and 0.075 (2.5 and 7.5%) for a chart to be considered robust at a certain condition. Table 1 shows three robust charts that perform as good as the traditional charts in controlling false alarm under ideal condition ( $\epsilon = 0, \mu = 0$ ), regardless of the group sizes m, ε and p. However, the rates for all charts dwindle when contamination exists. In this situation, some of the results are below the Bradley's robust interval limit. At p = 2, the traditional chart fares

**Table 2: False alarm rates (%) under independent case for  $\alpha = 5\%$**

m	$\varepsilon$	$\mu$	p = 2				p = 5				p = 10			
			$T_0^2$	$T_{MADn}^2$	$T_{Sn}^2$	$T_{Tn}^2$	$T_0^2$	$T_{MADn}^2$	$T_{Sn}^2$	$T_{Tn}^2$	$T_0^2$	$T_{MADn}^2$	$T_{Sn}^2$	$T_{Tn}^2$
50	0.0	0	5.6	4.9	5.4	5.3	5.3	5.5	4.8	4.8	5.7	6.4	6.0	6.1
		3	2.0	3.1	3.4	3.3	2.7	3.6	3.8	3.6	4.1	4.3	4.2	3.9
	0.2	5	1.6	3.0	3.4	3.3	2.6	3.5	3.9	3.7	3.8	4.4	4.1	4.0
		3	2.1	1.9	1.9	1.9	2.6	1.9	1.9	2.0	4.2	2.9	3.0	2.9
		5	1.6	1.8	1.8	1.8	2.5	1.5	2.1	1.7	4.0	3.5	3.4	3.2
100	0.0	0	5.5	4.1	4.1	4.2	5.4	3.7	3.9	3.7	5.5	5.1	4.9	5.1
		3	2.1	2.7	2.7	3.0	2.9	2.5	2.3	2.3	3.3	3.6	3.8	3.9
	0.2	5	1.6	2.6	2.6	3.2	2.8	1.6	2.5	2.2	3.4	3.6	3.4	3.7
		3	2.1	1.1	1.1	1.1	3.0	1.4	1.7	1.7	3.5	1.8	1.8	1.8
		5	1.6	1.1	1.1	1.0	2.9	1.6	1.7	1.6	3.4	2.3	2.4	2.7
150	0.0	0	5.3	4.8	4.8	5.0	4.3	4.4	4.5	4.3	4.6	4.6	7.0	6.2
		3	2.3	3.6	3.6	3.7	2.1	2.8	3.0	2.8	3.7	3.8	3.5	3.5
	0.2	5	1.9	3.4	3.4	3.6	2.1	2.8	2.9	3.0	3.6	3.9	3.5	3.4
		3	2.2	1.6	1.6	1.9	2.1	1.2	1.3	1.5	3.6	2.5	2.6	2.7
		5	1.9	1.8	1.8	1.9	2.1	1.4	1.2	1.3	3.6	2.6	2.4	2.1
%			20.0	60.0	60.0	60.0	73.0	53.0	47.0	47.0	100	87.0	80.0	87.0

**Table 3: Percentage of detecting outliers for independent case at  $\alpha = 5\%$**

m	$\varepsilon$	$\mu$	p = 2				p = 5				p = 10			
			$T_0^2$	$T_{MADn}^2$	$T_{Sn}^2$	$T_{Tn}^2$	$T_0^2$	$T_{MADn}^2$	$T_{Sn}^2$	$T_{Tn}^2$	$T_0^2$	$T_{MADn}^2$	$T_{Sn}^2$	$T_{Tn}^2$
50	0.0	0	5.6	4.9	5.4	5.3	5.3	5.5	4.8	4.8	5.7	6.4	6.0	6.1
		3	49.1	85.7	85.5	85.8	36.4	98.2	98.1	98.3	24.7	100	100	100
	0.2	5	74.7	100	100	100	44.2	100	100	100	26.5	100	100	100
		3	17.6	70.6	71.7	72.4	11.9	89.7	88.8	90.1	10.6	79.8	76.9	79.0
		5	17.1	100	100	100	10.9	100	100	100	10.6	100	100	100
100	0.0	0	5.5	4.1	4.3	4.2	5.4	3.7	3.9	3.7	5.5	5.1	4.9	5.1
		3	52.2	88.5	88.6	88.5	41.7	99.5	99.5	99.5	29.4	100	100	100
	0.2	5	80.6	100	100	100	50.7	100	100	100	31.4	100	100	100
		3	17.5	75.4	75.4	75.3	11.8	93.7	94.3	94.7	10.7	84.1	83.3	84.7
		5	17.5	100	99.9	99.9	12.0	100	100	100	10.8	100	100	100
150	0.0	0	5.3	4.8	4.9	5.0	4.3	4.4	4.5	4.3	4.6	4.6	7.0	6.2
		3	53.7	90.9	91.2	91.3	45.5	100	99.9	99.9	32.0	100	100	100
	0.2	5	80.4	100	100	100	54.2	100	100	100	33.7	100	100	100
		3	18.3	77.6	78	78.1	12.2	96.6	96.7	96.9	11.8	85.9	86.2	86.0
		5	18.3	100	100	100	12.5	100	100	100	12.1	100	100	100



**Fig. 1: a-d) Percentages detection of outliers at p = 2**

the worst with all the contaminated conditions are below the Bradleys's interval limit. The robustness of the charts worsens when  $\varepsilon$  increases but improves when p is large. The increase in m values has no significant impact

on the robustness of the charts. The shift in the mean values on the other hand shows negative effect on traditional chart but no clear effect on the robust charts (Fig. 1).

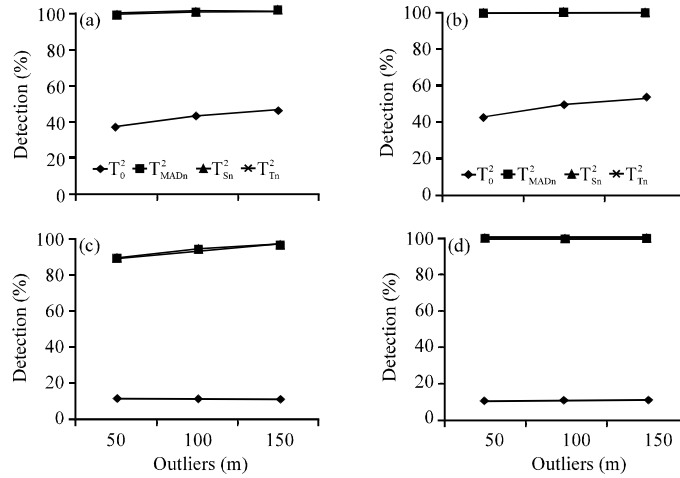


Fig. 2: a-d) Percentage detection of outliers at  $p = 5$

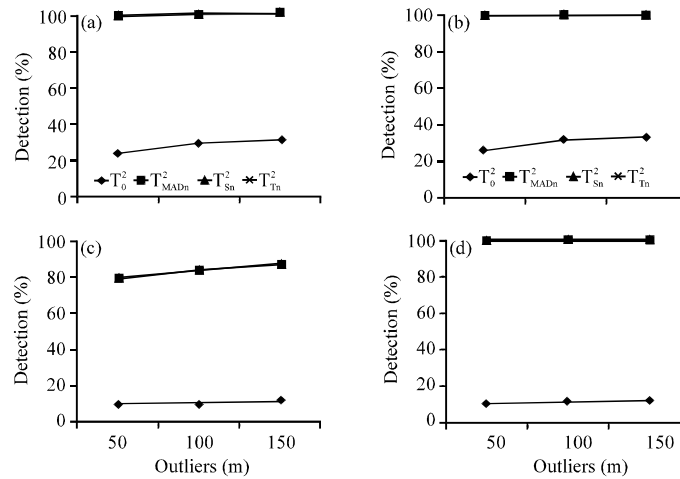


Fig. 3: a-d) Percentage detection of outliers at  $p = 10$

The performance in terms of probability of outlier's detection is recorded in Table 2. For a clearer visual and a better comparison, we translated the values in Table 2 into Fig. 1-3 based on the values of  $p$ . Across Fig. 1-3, we observe that for majority of the conditions, the robust charts outperform the traditional chart by a large difference. The robust charts under most conditions achieved the 100% detection with the lowest rate of 70.6% while the lowest rate for the traditional chart is below 10%. Across different  $p$ , there is no clear pattern of changes in performance among the charts. With regards to  $\epsilon$ , the robust charts as well as traditional chart show decrease in probability of detection when  $\epsilon$  increases. The shift in mean ( $\mu$ ) shows positive effect on the robust charts regardless of the proportion of contamination ( $\epsilon$ ). Meanwhile, for the traditional chart, positive effect only occurs when  $\epsilon = 0.1$ . The performance of the chart

deteriorates when  $\epsilon = 0.2$ . On a side note, the increase in group sizes ( $m$ ) brings some positive effect on the probability of detection for all the charts.

The best performance among the robust Hotelling's  $T^2$  charts in term of probability of detection is  $T_{Sn}^2$  when  $p = 2$  and  $T_{MADn}^2$  when  $p = 5$  and 10. In terms of false alarms, the best performance for the robust charts is  $T_{Sn}^2$  chart when  $p = 2$  and 5 and  $T_{MADn}^2$  when  $p = 10$ . Even though  $T_{Tn}^2$  is not in the league of best performer, its performance in terms of false alarm rates and probability of detection shows not much difference from the other two robust charts.

### CONCLUSION

Three robust Hotelling  $T^2$  control charts based on trimmed means were proposed and their performance in

terms of false alarm rates and probability of detection were investigated. The overall finding shows that the performance of the robust charts in controlling false alarm are moderate. They are however, superior in detecting outliers regardless of the conditions imposed in this study. In contrast, the traditional chart which performs moderately well in controlling false alarm has shown inability to detect outliers. Among the proposed robust control charts, the best performance is given by the  $T^2_{MAD_{DB}}$  since, this control chart produces good values for both false alarm rates and probability of detection.

#### ACKNOWLEDGEMENT

The researchers would like to acknowledge the research that led to this study which was fully funded by the Fundamental Research Grant Scheme (S/O 12801) of Ministry of Higher Education.

#### REFERENCES

- Abdullah, M.B., 1990. On a robust correlation coefficient. *J. R. Stat. Soc. Ser. D. Statistician*, 39: 455-460.
- Abu-Shawiesh, M.O.A. and M.B. Abdullah, 2001. A new robust bivariate control chart for location. *Commun. Stat. Simul. Comput. Series*, 30: 513-549.
- Abu-Shawiesh, M.O.A., B.M.G. Kibria and F. George, 2014. A robust bivariate control chart alternative to the hotelling's  $T^2$  control chart. *Quality Reliab. Eng. Int.*, 30: 25-35.
- Alfaro, J.L. and J.F. Ortega, 2008. A robust alternative to Hotelling's  $T^2$  control chart using trimmed estimators. *Qual. Reliab. Eng. Intl.*, 24: 601-611.
- Alfaro, J.L. and J.F. Ortega, 2009. A comparison of robust alternatives to Hotelling's  $T^2$  control chart. *J. Applied Stat.*, 36: 1385-1396.
- Ali, H., S.S.S. Yahaya and Z. Omar, 2013. Robust Hotelling  $T^2$  control chart with consistent minimum vector variance. *Math. Probl. Eng.*, 2013: 1-7.
- Alt, F.B., 1985. Multivariate Quality Control. In: *The Encyclopedia of Statistical Sciences*, Kotz, S., N.L. Johnson and C.R. Read (Eds.). John Wiley, New York, USA., pp: 110-122.
- Bradley, J.V., 1978. Robustness?. *Br. J. Math. Stat. Psychol.*, 31: 144-152.
- Haddad, F.S., S.S. Syed-Yahaya and J.L. Alfaro, 2013. Alternative Hotelling's  $T^2$  charts using winsorized modified one-step M-estimator. *Qual. Reliab. Eng. Intl.*, 29: 583-593.
- Johnson, N., 2007. A comparative simulation study of robust estimators of standard errors. MSc Thesis, Department of Statistics Statistics, Brigham Young University, Provo, Utah.
- Mohammadi, M., H. Midi, J. Arasan and B. Al-Talib, 2011. High breakdown estimators to robustify phase II multivariate control charts. *J. Applied Sci.*, 11: 503-511.
- Montgomery, D.C., 2005. *Introduction to Statistical Quality Control*. 5th Edn., John Wiley and Sons, New York.
- Othman, A.R., H.J. Keselman, A.R. Padmanabhan, R.R. Wilcox and K. Fradette, 2004. Comparing measures of the typical score across treatment groups. *Br. J. Math. Stat. Psychol.*, 57: 215-234.
- Rocke, D.M., G.W. Downs and A.J. Rocke, 1982. Are robust estimators really necessary?. *Technometrics*, 24: 95-101.
- Rosenberger, J.L. and M. Gasko, 1983. Comparing Location Estimators: Trimmed Means, Medians and Trimean. In: *Understanding Robust and Exploratory Data Analysis*, Hoaglin, D.C., F. Mosteller and J.W. Tukey (Eds.). Wiley Publishing Company, Hoboken, New Jersey, USA., ISBN:9780471097778, pp: 297-336.
- Rousseeuw, P.J. and C. Croux, 1993. Alternatives to the median absolute deviation. *J. Am. Statist. Assoc.*, 80: 1273-1283.
- Surtihadi, J., 1994. Multivariate and robust control charts for location and dispersion. Ph.D Thesis, Rensselaer Polytechnic Institute, Troy, New York, USA.
- Vargas, N.J.A., 2003. Robust estimation in multivariate control charts for individual observations. *J. Qual. Technol.*, 35: 367-376.
- Yahaya, S.S.S., A.R. Othman and H.J. Keselman, 2004. Testing the Equality of Location Parameters for Skewed Distributions Using S1 with High Breakdown Robust Scale Estimators. In: *Theory and Applications of Recent Robust Methods*, Hubert, M., G. Pison, A. Struyf and S. van Aelst (Eds.). Birkhauser, Basel, Switzerland, ISBN: 978-3-0348-7958-3, pp: 319-328.
- Yahaya, S.S.S., H. Ali and Z. Omar, 2011. An alternative hotelling  $T^2$  control chart based on Minimum Vector Variance (MVV). *Mod. Applied Sci.*, 5: 132-151.
- Yanez, S., N. Gonzalez and J.A. Vargas, 2010. Hotelling's  $T^2$  control charts based on robust estimators. *Dyna*, 77: 239-247.