5-2014

A Comparison of Some Robust Bicariate Control Charts for Individual Observations

Moustafa Omar Ahmed Abu Shawiesh
Hashemite University

Florence George
Department of Mathematics and Statistics, Florida International University, fgeorge@fiu.edu

B M Golam Kibria
Department of Mathematics and Statistics, Florida International University, kibriag@fiu.edu

International Journal for Quality Research 8(2) 183–196
ISSN 1800-6450

Moustafa Omar Ahmed Abu-Shawiesh [1]
Florence George
B. M. Golam Kibria

Article info:
Received 01.03.2014
Accepted 29.05.2014

UDC – 65.012.7

A COMPARISON OF SOME ROBUST BIVARIATE CONTROL CHARTS FOR INDIVIDUAL OBSERVATIONS

Abstract: *This paper proposed and considered some bivariate control charts to monitor individual observations from a statistical process control. Usual control charts which use mean and variance-covariance estimators are sensitive to outliers. We consider the following robust alternatives to the classical Hoteling's T2: T2MedMAD, T2MCD, T2MVE A simulation study has been conducted to compare the performance of these control charts. Two real life data are analyzed to illustrate the application of these robust alternatives.*
Keywords: *Bivariate control chart, False Alarm, Hotelling's T2 statistic, Outliers, Robust estimation, Simulation Study, Statistical process control*

1. Introduction

To monitor the quality characteristics in an industrial process, control charts are the most popular tools used in statistical process control (*SPC*). In many of these industrial processes, it is frequently required to monitor several quality characteristics at the same time. For example, the quality of a certain type of tablets may be determined by weight, degree of hardness, thickness, width and length (Liu, 1995). These quality characteristics are clearly correlated and therefore the separate univariate control charts for monitoring individual quality characteristics may not be adequate for detecting outliers and changes in the overall quality of the product. Thus it is desirable to have control charts that can simultaneously monitor multivariate measurements. Because of that, the multivariate control charts are the 2 most common tools used in such cases. These control charts can take into account the simultaneous nature of the control scheme and the correlation structure between the quality characteristics (Alt, 1985).

The multivariate control chart is useful when several quality characteristics of a product are taken to assess quality. The main objective of a multivariate control chart is to detect the presence of special causes of variation and can be used as a tool to detect multivariate outliers, mean shifts, and other distributional deviations from the in-control distribution.

1.1. Effect of outliers in Multivariate Quality Control Charts

In statistical quality control concepts, an outlier is defined as an observation that deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism (Hawkins, 1980).

[1] Corresponding author: Moustafa Omar Ahmed Abu-Shawiesh email: mabushawiesh@hu.edu.jo

Outliers have a big influence on resulting estimates and cause any out-of-control observations to remain undetected.

Outliers can be detected by using univariate or multivariate methods. When, there are more than one outliers the detection situation becomes more difficult due to masking and swamping (Rousseeuw and van Zomeren, 1990). Masking occurs when we fail to detect the outliers while swamping occurs when observations are incorrectly declared as outliers. The identification of outliers in multivariate cases is more difficult than in the univariate case. For instance, the simple graphical methods that can be used to detect outliers in a single dimension are often not available in higher dimensions.

Outliers can heavily influence the estimation of the variance-covariance matrix and subsequently the parameters or statistics that are needed to be derived from it. Hence, a robust estimate of the variance-covariance matrix that will not be affected by outliers is required to obtain valid and reliable results (Hubert and Engelen, 2007). The modern strategy for dealing with masking in the univariate case is to substitute the sample mean and variance with sample median, *MED*, and median absolute deviation from the sample median, *MAD*, respectively (Wilcox and Keselman 2003; Abu-Shawiesh *et al.*, 2009). In multivariate case a popular strategy is to make multivariate approaches more robust by replacing the location and the scale estimators with measures of central tendency and dispersion that are resistant to outliers.

1.2. Constructing the control chart using the Hotelling's T2 Statistic

The Hotelling T2 statistic has widely been used in constructing the multivariate control charts to monitor the individual or subgroups observations. However, it is not robust. In the construction of such control charts, Alt (1985) has defined two phases: Phase I and II. In Phase I, a historical data set of observations is analyzed to determine whether a process is in-control and to estimate the parameters of the in-control process, the control limits and to identify and eliminate multivariate outliers. In Phase II, the estimations and control limits are used to check the data obtained during the industrial process for detecting any departure from the parameter estimates and, as noted by (Woodall *et al.*, 2004), it is important to distinguish between Phase I and Phase II methods and applications.

To construct the control chart using Hotelling's T2 statistic, let us assume that $X_i = (X_{i1}, X_{i2}, \dots, X_{ip})^t$ denote a p x 1 vector that represents the p quality characteristics of the ith observation, and i= 1, 2, … , n, where n is the sample size. We also assume that the $\mathbf{X}$ij's are iid $N_p = (\mu, \sum)$ when the process is in-control. If the process parameter values are unknown, data will be collected when the process in-control. Then, the mean vector μ and the variance-covariance matrix $\sum$ will be replaced respectively by $\bar{\boldsymbol{X}}$ and **S,** where $\bar{\boldsymbol{X}}$ is the sample mean vector and **S** is the sample variance-covariance matrix.

The Hotelling's T2 control chart is then constructed using these estimated parameters. As mentioned before, the control chart is first used to test retrospectively whether the process was in-control (Phase I), then after the initial control chart has been established, the resulted control chart can be used to monitor the process on-line, that is, the values of individual observation are plotted one-at-a time on the chart as each new observation is obtained (Phase II). In this paper, we will consider control charts in Phase I. The statistic plotted on the Hotelling's T^2 control chart for each initial observations is calculated as follows:

$$T_i^2 = (X_i - \bar{X})^t S^{-1} (X_i - \bar{X}), \quad (1)$$
$$i = 1,2,3, \dots, n$$

where $\bar{\boldsymbol{X}}$and **S** are the sample mean vector and sample variance covariance matrix. Then the UCL of this control would be:

$$UCL_{T^2} = \frac{p(n-1)}{(n-p)} F_{p,n-p} \quad (4)$$

where $F_{v_1,v_{2,}a}$ is the $(1-\alpha)^{th}$ percentile point of the F distribution with v1 and v2 degrees of freedom, and α is the desired false alarm probability. The lower control limit (LCL) is usually set to zero.

1.3. Robust Alternatives to Hotelling's T2 Control Chart

Robust estimation has been a useful approach in statistics due to good properties shown under some deviations of distributional assumptions and existence of outliers. Johnson (1987) found that the traditional Hotelling's T^2 statistic cannot resist the departure from the normal distribution. Moreover, Croiser (1988) mentioned that the robustness against the multiple outliers is necessary in the multivariate quality control. Likewise, Brooks (1985) took notice about the outliers; these data errors increase in case of the development of manufacturing system because of the huge collecting of data. It is now evident that the Hotelling's T^2 statistic, which is based on the classical estimators, is easily affected by outliers (Rousseeuw and Leroy, 2003; Sullivan and Woodall, 1996). There have been many robust methods of estimating the variance-covariance matrix of a multivariate data. Such methods include Minimum Volume Ellipsoid (*MVE*), Minimum Covariance Determinant (*MCD*), S-Estimator, M-Estimator and Orthogonalized Gnanadesikan-Kettering (*OGK*) methods. Using these robust methods, various alternatives to Hotelling's T^2 have been proposed in order to avoid the negative effect of outliers on the control chart's behavior. Oyeyemi and Ipinyomi (2010) proposed a robust method for estimating covariance matrix for multivariate data.

Surtihadi (1994) used the median as a robust location estimator. He constructed a robust bivariate control chart based on the bivariate sign tests of Blumen and Hodges. Moreover, he found that this control chart needs fewer assumptions than the traditional control chart. Also, it needs the underlying distribution to be continuous and symmetric; as a result, this control chart has a good protection in the presence of the extreme data error.

Vargas (2003) proposed a control chart based on robust estimators of location and dispersion using the minimum volume ellipsoid (*MVE*) estimators. Simulation studies showed that the robust Hotelling's T^2 statistic that are using the minimum volume ellipsoid (*MVE*) estimators are efficient in detecting the multiple outliers and can deal with the masking effect.

Jensen *et al.* (2007) studied the high breakdown estimation method based on most popular robust estimators the minimum volume ellipsoid (*MVE*) and the minimum covariance determinant (*MCD*). They determined which estimator of them is better to use in the robust control charts in terms of detection of multiple outliers.

Vargas and Lagos (2007) compared four multivariate control charts for process dispersion and among the schemes compared, a new control chart based on robust estimation of the variance-covariance matrix proved to be very effective in detecting changes in the process dispersion matrix.

Alfaro and Ortega (2008) proposed a new alternative robust Hotelling's T^2 controlled charts to the traditional Hotelling's T^2 control charts. They replaced the sample mean vector in the traditional Hotelling's T^2 statistic by the trimmed mean vector, and the variance covariance matrix by the trimmed variance covariance matrix to construct the alternative robust Hotelling's T^2 statistic. They concluded that the new robust Hotelling's T^2 statistic is more effective in detection outliers.

Alfaro and Ortega (2009) has developed four alternatives robust Hotelling's T^2 charts to the traditional Hotelling's T^2 chart, these proposed control charts used minimum volume ellipsoid (*MVE*) estimator, minimum covariance determinant (*MCD*) estimator, reweighted *MCD* estimator and the trimmed mean estimator. They concluded that the robust alternatives Hotelling's T^2 charts behaved better than the traditional Hotelling's T^2 charts in the presences of outliers. Furthermore, they recommended using the Robust Hotelling's T^2 charts that depend on the trimmed mean and the modified of the *MCD* estimators when the amount of outliers is small. They also recommended using the other two robust Hotelling's T^2 statistic of *MVE* and *MCD* when the detection of outliers is more important.

Abu-Shawiesh and Abdullah (2001) developed a new robust Shewart-type control chart for monitoring the location of a bivariate process and examine its behavior based on the Hodges-Lehamnn and Shamos-Bickel-Lehmann estimators. A numerical example is given to illustrate the use of the proposed method. Its performance was investigated using a simulation study.

Abu-Shawiesh *et al.* (2012) proposed a new bivariate control chart for *m* sub-groups based on the robust estimators as an alternative to the Hotelling's T^2 control chart. The location vector and the variance-covariance matrix for the new control chart are obtained using the sample median, the median absolute deviation from the sample median, and the comedian estimator. The performance of the proposed 6 method in detecting outliers is evaluated and compared with the Hotelling's T^2 method by using a Monte-Carlo simulation study.

In some industrial setting we come across with individual observations. This situation occurs frequently in the chemical and process industries. Since these industries frequently have multiple quality characteristics that must be observed, multivariate control chart with individual observation would be of interest in these situations (Montgomery, 2009). Since Alfero and Ortega (2009) suggested MVE and MCD among four methods and Abu-Shewiesh *et al.* (2012) proposed T^2_{MEDMAD} for *m* sub-groups, this paper make an attempt to consider several bivariate control charts, namely, Hotelling's T^2, $\mathrm{T}^2_{\mathrm{MEDMAD}}$, $\mathrm{T}^2_{\mathrm{MVE}}$, and $\mathrm{T}^2_{\mathrm{MCD}}$ to monitor individual observations. The organization of the paper is as follows. The proposed bivariate robust control chart along with the other two methods are discussed in section 2. To compare the performance of the proposed robust control chart, a simulation study has been conducted in section 4. To illustrate the use of the proposed methods, two real life data examples are analyzed in section 4. Section 5 ends up with some conclusions.

2. Bivariate Robust Control Charts

In this section we will review several bivariate control charts, namely, Hotelling's T^2, $\mathrm{T}^2_{\mathrm{MEDMAD}}$, $\mathrm{T}^2_{\mathrm{MVE}}$, and $\mathrm{T}^2_{\mathrm{MCD}}$ to monitor individual observations.

2.1. The proposed Robust Bivariate Robust Control Chart

Following Abu-Shawiesh *et al.* (2012), we present the algorithm for the robust bivariate control chart. We assume that the process characteristics (X_1, X_2) are generally correlated and follow some symmetric and continuous bivariate distribution. The null hypothesis, H0, represents the state of statistical control. In particular, the hypothesis of interest would be:

$$\begin{gathered} H_0: (\mu, v) = (\mu_0, v_0) \\ \text{vs} \\ H_1: (\mu, v) \neq (\mu_0, v_0) \end{gathered} \tag{3}$$

where (μ, v) is the median estimator of the process. We also assume, without loss of generality, that the in-control median (μ_0, v_0) = (0,0). Our proposed control chart constitutes the plotting of the $\mathrm{T}^2_{\mathrm{MEDMAD}}$ statistic computed from successive random samples from the process. The process is in-control if the plots are within the control region defined by the acceptance region of the test. This control region is specified by 7 an upper control limit, UCL, where UCL is defined as the $(1-\alpha)100^{th}$ percentile of the associated statistic under the null hypothesis and α is the probability of a false alarm. For our proposed method, the null hypothesis, H_0, is rejected if the value of the statistic T^2_{MEDMAD} is too large. That is, for the significance level, α, H_0, is rejected if $T^2_{MEDMAD} > T_\alpha$ where T_α is the $(1-\alpha)100^{th}$ percentile of the statistic T^2_{MEDMAD} under the null hypothesis, H_0, and it will represent the UCL value. This value will be determined later by a simulation study for different values of sample size n and significance level α. Suppose that we have p variables x_1, x_2, …, x_n. Each variable consists of *n* observations. In this paper the value of p considered 2, then using individual observation we do the following:

1) Calculate the MED estimators as follows:

$$MED = [MEDj]_{2x1}, \text{ j=1,2}$$

2) Calculate the MAD estimators as follows:

$$MAD_j = 1.4826\, med\{|X_{ij} - med\{x_i\}|\},$$
i=1,2,...,n; j=1,2

3) The 2-by-2 sample variance-covariance matrix for the two variables X_1 and X_2 can be constructed as follows:

$$S_{MAD} = \begin{bmatrix} (MAD(X_1))^2 & COM(X_1, X_2) \\ COM(X_1, X_2) & (MAD(X_2))^2 \end{bmatrix}$$

which is robust and positive definite matrix. The diagonal elements are the square of the MAD of X_j that is, $MAD_j^2 = MAD_{jj}$ and

$$COM(X_1, X_2) = MED[(X_{1i} - MED\{X_{11}, X_{12}, \dots, X_{1n}\})(X_{2i} - MED\{X_{21}, X_{22}, \dots, X_{2n}\})]$$

4) Calculate the inverse of the matrix S_{MAD}, that is S^{-1}_{MAD}.

5) Determine the statistic, $T^2_{MEDMAD_i} = (x_i - MED)^t S^{-1}_{MAD} (x_i - MED), i = 1,2, \dots, n$

6) The control limits will be determined through a Monte-Carlo simulation study.

7) Plot the values of the statistic $T^2_{MEDMAD_i}$ on the control chart.

8) If any value of $T^2_{MEDMAD_i}$ is falling outside the control limits, then the process is considered to be an out of control.

2.2. Robust MVE Control Chart

Following Jensen *et al.* (2007) and Vargas (2003), a robust alternative to Hotelling T^2 statistic is defined as

$$T^2_{MVE_i} = (x_i - \bar{X}_{MVE})^t S^{-1}_{MVE} (X_i - \bar{X}_{MVE})$$

where $\bar{X}_{MVE}$ and S_{MVE} are MVE mean vector and scale estimators respectively**.** The statistical software R is used to calculate MVE estimates based on a genetic algorithm. The computing program is available from the authors upon request. More details on MVE method we refer Vargas (2003), Jensen *et al.* (2007) and Alfaro and Ortega (2009) among others.

Following Jensen *et al*. (2007) and Vargas (2003), a robust alternative to Hotelling T^2 statistic is defined as

$$T^2_{MCD_i} = (x_i - \bar{X}_{MCD})^t S^{-1}_{MCD} (X_i - \bar{X}_{MCD})$$

where $\bar{X}_{MCD}$ and S_{MCD} are MDC location and scale estimators respectively**.** The statistical software R is used to calculate MVE estimates based on a genetic algorithm. The computing program is available from the authors upon request. More details on MCD method we refer Vargas (2003), Jensen *et al*.

(2007) and Alfaro and Ortega (2007) among others.

3. The Simulation Study

Since, the distributions of T^2_{MEDMAD}, T^2_{MVE} and T^2_{MCD} are not known, a theoreticalcomparison among different methods are not possible, a simulation study has beenconducted to compare the performance of the three robust methods. We used the *R* software to conduct this simulation.

3.1. The General Simulation and Results

The simulation series considered the bivariate normal distribution for sample sizes n = 25, 50 and 100. Moreover, we considered the bivariate contaminated normal distribution (10% and 20%). Each simulation run consisted of 5000 replications of size n. The control limits were determined from the 5000 simulations, such that the false alarm probability is 0.05, which is widely used level of significance. We will consider the mean vector $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and the variance-covariance matrix $\sum = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. The simulated upper control limits (UCL) for all methods are given in Table 1. The lower control limits for all methods are set to zero. From Table 1 we observed that as sample size increase the control limits for all robust methods decrease, while the UCL for Hotelling's T^2 remain almost the same. For very large sample size, one may expect a constant UCL for all methods. However, the upper limit of T^2_{MEDMAD} is close to Hotelling T^2 than MCD and MVE. The limits of T^2_{MCD} T^2_{MVE} are very close to each other. It can also be noticed that the control limit of T^2_{MEDMAD} is between control limits of T^2 and T^2_{MCD}/T^2_{MVE}.

Table 1: The Simulated UCL for all T^2 Control Charts

Sample Size (n)	T^2	T^2_{MCD}	T^2_{MVE}	T^2_{MEDMAD}
25	5.989	13.488	14.257	7.724
50	5.987	9.271	9.656	6.705
100	5.96	7.57	7.77	6.28

3.2. Simulation with Outliers

The probability of detecting a change depends on the values of $\mu_1, \sum_0$ and $\sum_1$ but it does not depend on the value of μ_0, therefore, we can, without loss of generality, use $\mu_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. We consider the proportion of outliers (ε) as 0, 0.1 and 0.2. We use 5000 replications for different sample sizes where n = 15, 25, 50 and 100. We have used α=0.05 for the simulation study. We calculated the percentage of detection of all outliers and the percentage of false alarms in parentheses which is estimated as the proportion of statistic values that are above the control limits in the 5000 replications. To perform the simulation study with outliers, we consider the following three cases:

a) **Independent Variables**

In this case, the two variables (*Quality Characteristics*) x_1 and x_2 are assumed to be independent. The contaminated normal model considered is as follows:

$$CN = (1-\varepsilon)N(0, I_2) + \varepsilon N(\mu_1, I_2) \qquad (4)$$

where we consider μ_1 to be a vector of size 2. Its elements are all 0 (there is no change) or 3 or 5 and I_2 is the identity matrix of size 2, that is, there are different sized changes in the average of the two independent variables X_1 and X_2. The results of this simulation study are given in Table 2.

Table 2. Percentage of Detection of all Outliers and the Percentage of False Alarms (within parenthesis) for all methods

Sample Size (n)	ε	μ_1	T^2	T^2_{MCD}	T^2_{MVE}	T^2_{MEDMAD}
25	0	(0,0)	4.92	5.04	4.96	4.88
	0.1	(3,3) (5,5)	44.5(2.7) 57.2(2.2)	73.3(3.4) 98.9(3.5)	74.2(3.4) 99.5(3.4)	73.9(2.7) 98.9(2.5)
	0.2	(3,3) (5,5)	19.5(2.4) 23.2(2.0)	57.8(2.4) 96.9(2.1)	59.6(2.6) 96.9(2.2)	46.8(1.3) 87.3(1.1)
50	0	(0,0)	4.99	5.11	5.14	4.91
	0.1	(3,3) (5,5)	49.6(2.3) 67.7(1.8)	85.6(3.2) 100(3.4)	88.1(3.3) 100(3.5)	83.8(2.3) 100(2.4)
	0.2	(3,3) (5,5)	19.4(2.2) 23.6(1.9)	73.3(2.0) 99.8(2.3)	76.1(2.2) 99.9(2.4)	55.6(1.0) 96.3(0.9)
100	0	(0,0)	5.04	5.15	5.17	5.17
	0.1	(3,3) (5,5)	53.0(2.3) 74.6(1.9)	92.5(3.8) 100(3.9)	93.4(3.8) 100(3.9)	87.7(2.4) 100(2.3)
	0.2	(3,3) (5,5)	20.0(2.1) 22.2(1.9)	83.9(2.6) 100(3.0)	86.7(2.6) 100(2.9)	62.1(0.9) 99.2(0.8)

From Table 2 we observe that if the variables are assumed to be independent and the outliers are present, the proposed robust method, T^2_{MCD} and T^2_{MVE} perform better than the Hotelling's T^2 control chart in the sense of high power. The Hotelling's T^2 has the lower false alarm rate compare to MCD and MVE. However, its power is the worse among four methods. Our proposed robust method T^2_{MEDMAD} has the lowest false alarm rate while keeping high power similar to T^2_{MVE} and T^2_{MCD}.

b) **Correlated Variables**

In this case, two variables, x_1 and x_2 are assumed to be correlated. The contaminated normal model considered is as follows:

$$CN = (1-\varepsilon)N(0,\textstyle\sum_0) + \varepsilon N(\mu_1,\textstyle\sum_0) \quad (5)$$

where we consider μ_1 to be a vector of size 2 where the elements of the vector are all 0 (there is no change) or 5, which shows outliers (*observations out of control*) in the two variables, and $\sum_0$ to be a matrix of size 2 given as $\sum_0 = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}$. This value of $\sum_0$ is used in order to analyze whether the correlation coefficient level affects the detection probability. The results of this simulation study are given in Table 3.

Table 3. Percentage of Detection of all Outliers and the Percentage of False Alarms for all methods Methods

Sample Size (n)	ε	μ_1	T^2	T^2_{MCD}	T^2_{MVE}	T^2_{MEDMAD}
25	0	(0,0)	4.92	5.04	4.96	4.88
	0.1	(3,3) (5,5)	44.5(2.7) 57.2(2.2)	73.3(3.4) 98.9(3.5)	74.2(3.4) 99.5(3.4)	73.9(2.7) 98.9(2.5)
	0.2	(3,3) (5,5)	19.5(2.4) 23.2(2.0)	57.8(2.4) 96.9(2.1)	59.6(2.6) 96.9(2.2)	46.8(1.3) 87.3(1.1)
50	0	0,0)	4.99	5.11	5.14	4.91
	0.1	(3,3)	49.6(2.3)	85.6(3.2)	88.1(3.3)	83.8(2.3)

		(5,5)	67.7(1.8)	100(3.4)	100(3.5)	100(2.4)
	0.2	(3,3) (5,5)	19.4(2.2) 23.6(1.9)	73.3(2.0) 99.8(2.3)	76.1(2.2) 99.9(2.4)	55.6(1.0) 96.3(0.9)
100	0	0,0)	5.04	5.15	5.17	5.17
	0.1	(3,3) (5,5)	53.0(2.3) 74.6(1.9)	92.5(3.8) 100(3.9)	93.4(3.8) 100(3.9)	87.7(2.4) 100(2.3)
	0.2	(3,3) (5,5)	20.0(2.1) 22.2(1.9)	83.9(2.6) 100(3.0)	86.7(2.6) 100(2.9)	62.1(0.9) 99.2(0.8)

From Table 3 we observe that if the variables are assumed to be correlated and the outliers are present, the T^2_{MEDMAD}, T^2_{MCD} and T^2_{MVE} perform better than the Hotelling's T^2 control chart in the sense of high power. The Hotelling's T^2 has the lower false alarm rate compare to MCD and MVE. However, its power is the worse among four methods. For small contamination, T^2_{MEDMAD}, method has the lowest false alarm rate and the highest power. However, in other situations the powers are comparable to MVE and MCD.

c) **Correlated Variables and Regression Outliers**

Here, the two variables x_1 and x_2 are assumed to be correlated and regression outliers are introduced. The contaminated normal model considered is as follows:

$$CN = (1-\varepsilon)N(0,\textstyle\sum_0) + \varepsilon N(\mu_1,\textstyle\sum_1) \qquad (6)$$

where we consider $\sum_0$ to be a matrix of size 2 given as $\sum_0 = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}$ and μ_1 to be a vector of size 2 where the elements of the vector are all 0 (*there is no change*) or 5, or a vector of size 2 with a -1.5 and for the other values, which shows regression outliers, and $\sum_1$ to be a matrix of size 2 given as $\sum_1 = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix}$. This case of comparison is known as regression outliers. Results of this simulation study are given in Table 4.

Table 4. Percentage of Detection of all Outliers and the Percentage of False Alarms for all methods

Sample Size (n)	ε	μ_1	T^2	T^2_{MCD}	T^2_{MVE}	T^2_{MEDMAD}
25	0	(0,0)	4.8	4.8	4.8	4.7
	0.1	(3,3) (5,5)	59.7(2.0) 63.2(2.0)	98.8(3.4) 100(3.4)	98.9(3.4) 100(3.4)	98.6(2.6) 100(2.7)
	0.2	(3,3) (5,5)	29.4(1.9) 28.0(1.8)	98.0(2.0) 99.2(2.1)	98.2(2.0) 99.2(2.2)	96.5(1.1) 99.8(1.1)
50	0	0,0)	5.0	5.2	5.2	5.1
	0.1	(3,3) (5,5)	63.8(1.8) 71.3(1.6)	99.4(3.2) 100(3.4)	99.3(3.4) 100(3.5)	99.2(2.3) 100(2.3)
	0.2	(3,3) (5,5)	30.6(1.7) 29.0(1.7)	99.3(2.2) 100(2.2)	99.3(2.3) 100(2.2)	98.0(0.8) 100(0.8)
100	0	0,0)	5.1	5.5	5.3	5.2
	0.1	(3,3) (5,5)	67.2(1.7) 78.8(1.6)	99.3(4.1) 100(4.0)	99.4(4.0) 100(4.0)	99.1(2.3) 100(2.4)
	0.2	(3,3) (5,5)	31.3(1.7) 28.9(1.7)	99.2(2.9) 100(2.9)	99.2(2.9) 100(2.8)	98.4(0.7) 100(0.7)

From Table 4 we observe that if the variables are assumed to be correlated and the outliers are present, the T^2_{MEDMAD}, T^2_{MCD} and T^2_{MVE} perform better than the Hotelling's T^2 control chart in the sense of high power. The Hotelling's T^2 has the lower false alarm rate compare to MCD and MVE.

However, its power is the worse among four methods. For all possible conditions, our proposed robust method, T2MEDMAD has the lowest false alarm rate and the highest power.

4. Applications

To illustrate the procedure of the robust bivariate control charts, we will consider two real life data examples in this section.

4.1. Example 1

Consider a production process data given by Quesenberry (2001). The original data consists of 11 quality variables (characteristics) and measured on 30 products form a production process. For our comaprison purposes we consider 25 observarions from the first two variables and provided them in Table 5. The first and fourth columns are the production numbers and second, third, fifth and sixth columns are the observed values of production quality variables (X_1, X_2). The sample mean vectors and sample covarinace matrices for all methods are given in Table 6.

Table 5. Variable 1 ans 2 of Quesenberry (2001) data set.

Product Number	X_1	X_2	Product Number	X_1	X_2
1	0.567	60.6	14	0.458	61.1
2	0.538	56.3	15	0.554	59.8
3	0.53	59.5	16	0.469	58.6
4	0.562	61.1	17	0.471	59.6
5	0.483	59.8	18	0.457	59.7
6	0.525	60.2	19	0.565	60.9
7	0.556	60.8	20	0.664	60.2
8	0.586	59.8	21	0.6	60.5
9	0.547	60.2	22	0.586	58.4
10	0.531	60.6	23	0.567	60.2
11	0.581	59.8	24	0.496	60.2
12	0.585	59.7	25	0.485	59.5
13	0.54	60.5			

Table 6. Sample mean vector and covariance matrix

Method	Mean vector	Covarinace matrix
Hotelling T^2	$\begin{pmatrix}0.540\\59.90\end{pmatrix}$	$\begin{pmatrix}0.0015 & 0.0025\\0.0025 & 1.0470\end{pmatrix}$
T^2_{MEDMAD}	$\begin{pmatrix}0.540\\59.90\end{pmatrix}$	$\begin{pmatrix}0.0025 & 0.0032\\0.0032 & 0.4159\end{pmatrix}$
T^2_{MVE}	$\begin{pmatrix}0.540\\59.90\end{pmatrix}$	$\begin{pmatrix}0.0019 & 0.0127\\0.0127 & 0.3343\end{pmatrix}$
T^2_{MCD}	$\begin{pmatrix}0.540\\59.90\end{pmatrix}$	$\begin{pmatrix}0.0021 & -0.0038\\-0.0038 & 0.4006\end{pmatrix}$

Using $\alpha = 0.05$, the upper control limits for the all T^2, T^2_{MEDMAD}, T^2_{MVE} and T^2_{MCD} control charts from Table 1 are found to be 5.989, 7.724, 14.257 and 13.480 respectively. The resulting control charts are given in Figure 1. From Figure 1, we can see that the sample number 2 is out of control by all methods. However, the sample number 22 is out of control by MEDMAD and MVE methods. The sample number 20 is out of control by Hotelling T^2 statistic.

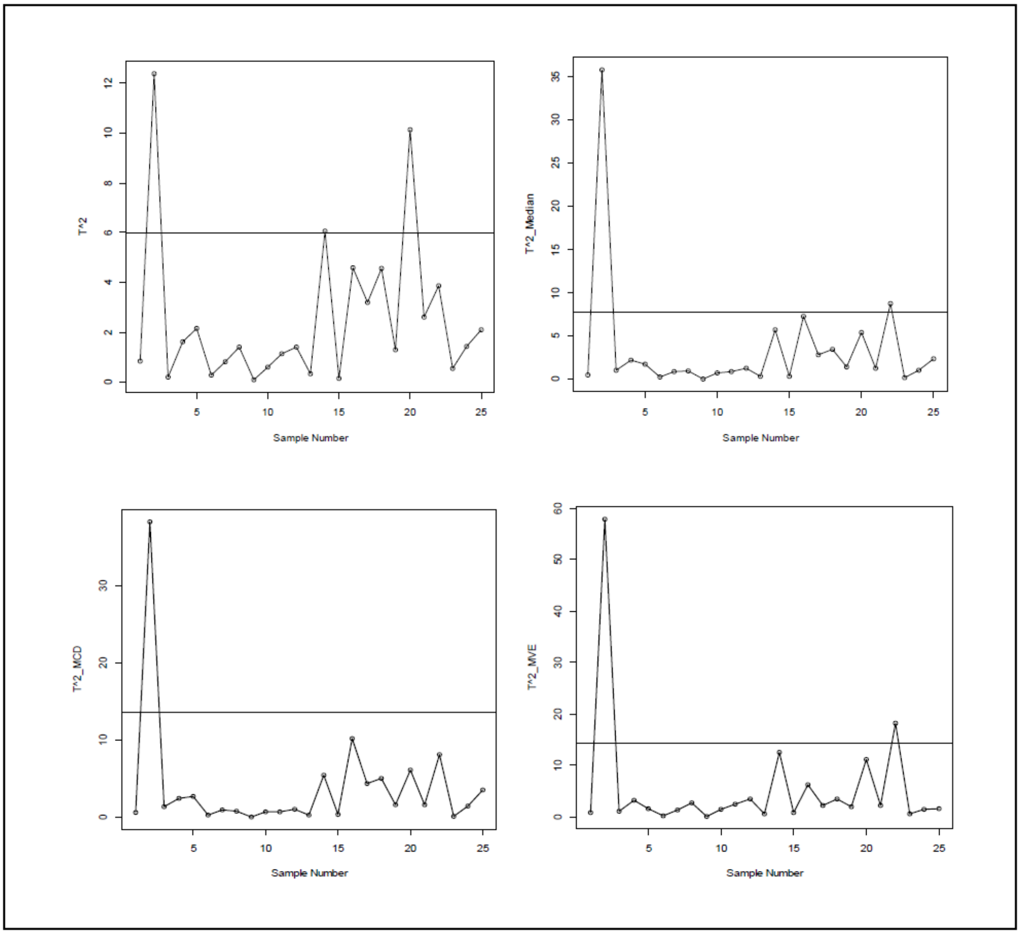

Figure 1. The Control Charts for production process data (Quesenberry, 2011) using the Hotelling's T^2, T^2_{MEDMAD}, T^2_{MVE} and T^2_{MCD} methods

4.2. Example 2

Consider a production process data given by Montgomerry (2009). The original data consists of 4 quality variables (characteristics) and measured on 30 products form a production process. For our comaprison purposes we consider first 25 observarions from the first two variables and provided them in Table 7. The first and fourth columns are the production numbers and second, third, fifth and sixth 15 columns are the observed values of production quality variables (X_1, X_2) The sample mean vectors and sample covarinace matrices for all methods are given in Table 8.

Table 7. Variable 1 ans 2 of Montgomery (2009) data set

Product Number	X_1	X_2	Product Number	X_1	X_2
1	10	20.7	14	10	19.8
2	10.5	19.9	15	8.5	19.2
3	9.7	20	16	9.7	20.1

4	9.8	20.2	17	8.3	18.4
5	11.7	21.5	18	11.9	21.8
6	11	20.9	19	10.3	20.5
7	8.7	18.8	20	8.9	19
8	9.5	19.3	21	9.9	20
9	10.1	19.4	22	8.7	19
10	9.5	19.6	23	11.5	21.8
11	10.5	20.3	24	15.9	24.6
12	9.2	19	25	12.6	23.9
13	11.3	21.6			

Table 8. Sample mean vector and covariance matrix

Method	Mean vector	Covarinace matrix
Hotelling T^2	$\begin{pmatrix}10\\20\end{pmatrix}$	$\begin{pmatrix}1.752 & 1.381\\1.381 & 1.195\end{pmatrix}$
T^2_{MEDMAD}	$\begin{pmatrix}9.97\\20.03\end{pmatrix}$	$\begin{pmatrix}1.407 & 0.800\\0.800 & 1.407\end{pmatrix}$
T^2_{MVE}	$\begin{pmatrix}9.965\\20.035\end{pmatrix}$	$\begin{pmatrix}1.051 & 0.956\\0.956 & 0.984\end{pmatrix}$
T^2_{MCD}	$\begin{pmatrix}9.965\\20.035\end{pmatrix}$	$\begin{pmatrix}1.051 & 0.956\\0.956 & 0.984\end{pmatrix}$

Using $\alpha = 0.05$, the upper control limits for the T^2, T^2_{MEDMAD}, T^2_{MVE} and T^2_{MCD} control charts from Table 1 are found to be 5.989, 7.724, 14.257 and 13.480 respectively. The resulting control charts are given in Figure 2. From Figure 2, we can see that the sample number 24 and 25 are out of control by all methods.

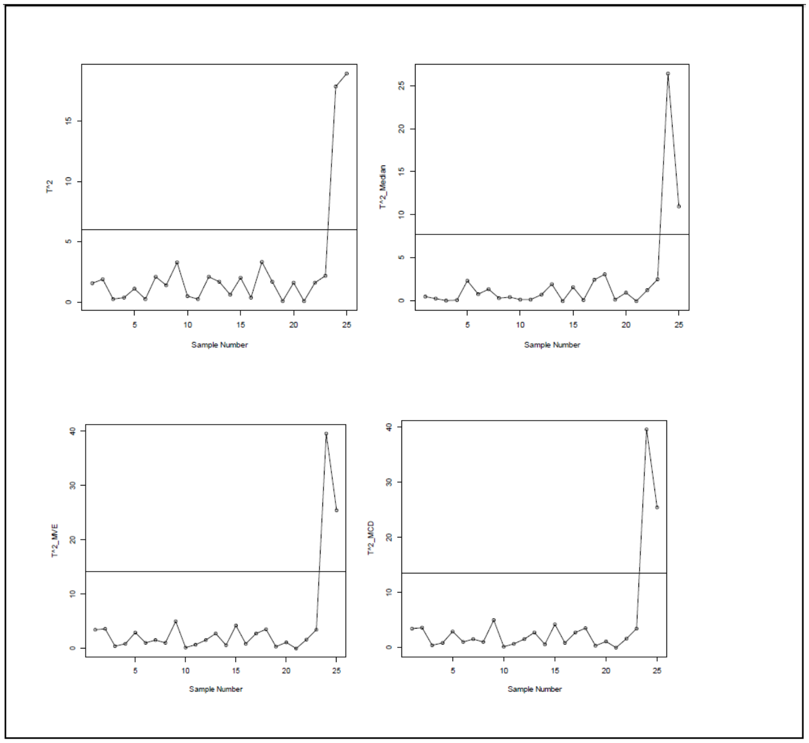

Figure 1. The Control Charts for production process data (Montgomerry, 2009) using the Hotelling's T^2, T^2_{MEDMAD}, T^2_{MVE} and T^2_{MCD} methods

5. Conclusions and Further Research

This paper compared several bivariate control charts which are based on robust estimators as an alternative to the Hotelling's T2 control chart. Since a theoretical comparison is not possible, we have done simulation for three cases: (i) independent variables (ii) correlated variables and (iii) correlated variables and regression outliers. From the simulation study we observed that the robust methods, $\mathrm{T}^2_{\mathrm{MEDMAD}}$, $\mathrm{T}^2_{\mathrm{MVE}}$, $\mathrm{T}^2_{\mathrm{MCD}}$. However, the proposed robust method, $\mathrm{T}^2_{\mathrm{MEDMAD}}$ has the lowest false alarm rate while having the highest power. The research in this paper is limited with two variables and the application of individual observations. The processes with more than two variables and with sub groups might be of interest for the practitioners. Such research possibilities are under our current investigation.

References:

Abu-Shawiesh, M.O., & Abdullah, M.B. (2001). A new robust bivariate control chart for location. *Communications in Statistics-Simulation and Computation*, *30*(3), 513-529.

Abu-Shawiesh, M.O., Al-Athari F.M., & Kittani, H.F. (2009). Confidence interval for the mean of a contaminated normal distribution. *Journal of Applied Sciences*, *9*(15), 2835-2840.

Abu-Shawiesh, M.O., Kibria, B.M.G., & George, F. (2012). A robust bivariate control chart alternative to the Hotelling's T2 control chart. *Quality and Reliablity Engineering International*. In Press. DOI: 10.1002/qre.1474.

Alfaro, J.L., & Ortega, J.F. (2008). A robust alternative to Hotelling's *T*2 control chart using trimmed estimators. *Quality and Reliability Engineering International*, *24*(5), 601-611.

Alfaro, J.L., & Ortega, J.F. (2009). A comparison of robust alternatives to Hotelling's *T*2 control chart. *Journal of Applied Statistics*, *36*(12), 1385-1396.

Alt, F.B. (1985). Multivariate quality control. In S. Kotz & N. Johnson (Eds.), *The Encyclopedia of Statistical Sciences*, *Volume 6* (110-122). New York : John Wiley & Sons..

Brooks, A. (1985). Heat treating shows why SPC is no cure-all for manufacturers. *Production Engineering*, *32*(9) 66-77.

Crosier, R.B. (1988). Multivariate generalizations of cumulative sum quality schemes. *Technometrics*, *30*(3), 291-303.

Hawkins, D.M. (1980). *Identification of Outliers*. New York: Chapman and Hall, Ltd.

Hubert, M., & Engelen, S. (2007). Fast cross-validation of high-breakdown resampling algorithms for PCA. *Computational statistics & data analysis*, *51*(10), 5013-5024.

Jensen, W.A., Birch, J.B., & Woodall, W.H. (2007). High breakdown estimation methods for phase I multivariate control charts. *Quality and Reliability Engineering International*, *23*, 615-629.

Johnson, M.E. (1987). *Multivariate statistical simulation*. New York: John Wiley & Sons.

Liu, R.Y. (1995). Control charts for multivariate processes. *Journal of the American Statistical Association*, *90*(432), 1380-1387.

Montgomery, D.G. (2009). *Introduction to Statistical Quality Control*. Wliey, New York.

Oyeyemi, G.M. & Ipinyomi, R.A. (2010). A robust method of estimating covariance matrix in multivariate data analysis. *African Journal of Mathematics and Computer Science Research*, *3*(1), 001-018.

Quesenberry, C.P. (2001). *The multivariate short-run snapshot Q chart. Quality Engineering*, *13*(4), 679-683.

Rousseeuw, P.J. & van Zomeren, B. (1990). Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association*, *85*, 633-639.

Rousseeuw, P.J., & Leroy, A.M. (2003). *Robust Regression and Outlier Selection*. Wiley, New York.

Sullivan, J.H., & Woodall, W.H. (1996). A comparison of multivariate control charts for individual observations. *Journal of Quality Technology*, *28*(4), 398-408.

Surtihadi, J. (1994). *Multivariate and Robust Control Charts for Location and Dispersion*. Ph. D. Dissertation. Department of Decision Sciences and Engineering Systems, Rensselaer's School of Engineering, Rensselaer Polytechnic Institute, Troy, New York, U.S.A.

Vargas, J.A. & Lagos, C.J. (2007). Comparison of multivariate control charts for process dispersion. *Quality Engineering*, *19*, 191-196.

Vargas, J.A. (2003). Robust Estimation in Multivariate Control Charts for Individual Observations. *Journal of Quality Technology*, *35*, 367-376.

Wilcox, R.R., & Keselman, H.J. (2003). Modern robust data analysis methods: measures of central tendency. *Psychological Methods*, *8*, 254-274.

Woodall, W.H., Spitzner, D.J., Montgomery, D.C., & Gupta, S. (2004). Using Control Charts to Monitor Process and Product Quality Profiles. *Journal of Quality Technology*, *36*(3), 309-320.

Moustafa Omar Ahmed Abu - Shawiesh
Hashemite University
Faculty of Science,
Department of Mathematics
Zarqa 13115
Jordan
mabushawiesh@hu.edu.jo

Florence George
Florida International University
Modesto A. Maidique Campus,
Department of Mathematics and Statistics
Miami, FL 33199
USA
kibriag@fiu.edu

B. M. Golam Kibria
Florida International University
Modesto A. Maidique Campus,
Department of Mathematics and Statistics
Miami, FL 33199
USA