Bivariate modified hotelling's T^2 charts using bootstrap data

**Firas Haddad[1], Mutasem K. Alsmadi[2], Usama Badawi[3], Tamer Farag[4], Raed Alkhasawneh[5],
Ibrahim Almarashdeh[6], Walaa Hassan[7]**
[1,5]Department of General courses, College of Applied Studies and Community Service,
Imam Abdulrahman Bin Faisal University, Saudi Arabia
[2,3,4,6,7]Department of Management Information Systems, College of Applied Studies and Community Service,
Imam Abdulrahman Bin Faisal University, Saudi Arabia

Article Info	**ABSTRACT**

The conventional Hotelling's T^2 charts are evidently inefficient as it has resulted in disorganized data with outliers, and therefore, this study proposed the application of a novel alternative robust Hotelling's T^2 charts approach. For the robust scale estimator S_n, this approach encompasses the use of the Hodges-Lehmann vector and the covariance matrix in place of the arithmetic mean vector and the covariance matrix, respectively. The proposed chart was examined performance wise. For the purpose, simulated bivariate bootstrap datasets were used in two conditions, namely independent variables and dependent variables. Then, assessment was made to the modified chart in terms of its robustness. For the purpose, the likelihood of outliers' detection and false alarms were computed. From the outcomes from the computations made, the proposed charts demonstrated superiority over the conventional ones for all the cases tested.

Corresponding Author:

Firas Haddad,
College of Applied Studies and Community Service,
Imam Abdulrahman Bin Faisal University,
Al-Dammam, Saudi Arabia.
Email: fshaddad@iau.edu.sa

1. INTRODUCTION

In manufacturing, Statistical control charts have been known as the most tool for monitoring the process of production. In monitoring the characteristics of product quality, in the beginning, the employment of control charts was facilitated. Considering the existence of various characteristics of quality in product quality determination, this approach is insufficient in terms of practicality. Nonetheless, multivariate control charts (MVCC) with the capacity in identifying the changes in covariance matrix Σ and the mean vector μ in order to achieve optimal performance of the product [1, 2].

The Hotelling's T^2 chart is among the most common MVCC methods. [3, 4]. With the capacity in detecting multiple outliers, mean shifts and deviations in the dispersal of control distribution [5]. The statistic T_i^2 employs the estimators $\bar{\mathbf{x}}$ and $\mathbf{S}$ which are directly impacted by the presence of outliers in the case of false alarms, resulting in failure in imposing control in the processes of production. The increase in the complexity of the manufacturing of the products in addition to their characteristics that generally contain outliers, contribute to the failure of the chart in performing its designated task.

In order to overcome the impact of outliers on the formed control chart, the application of robust estimators would be an appropriate solution. These robust estimators should be employed in place of the mean vector $\bar{\mathbf{x}}$ and the variance covariance matrix $\mathbf{S}$ as in the conventional Hotelling's T^2 chart.

Accordingly, a chart is considered robust if it could produce strong reaction to the changes to the production process, and this reaction is compelled by the controlled false alarms and the probabilities values of detection outliers are large enough and tends to 100%.

The bootstrap method encompasses a nonparametric technique due to its independence from the presumptions of data parametric distribution. Nonetheless, in the monitoring of a single process, this technique can generate single variable control charts. In this regard, Phaladiganon et al. in [6] mentioned the possibility in integrating multivariate control charts with the charts of the bootstrap technique that have proven their effectiveness.

Considering that it is possible to gather sample that is small in size, violation to the normality assumption distribution is possible. Besides that, in general, estimation to the in-control state of the control charts has to be carried out, but this will have adverse impact on the performance of the control chart. As indicated in Mostajeran et al. in [7], non-parametric bootstrap control charts are appropriate for an unidentified distribution or when making estimation on the process parameters from Phase I dataset or when it is impractical to gather sample of large size.

For observation purpose, control charts generally require normal distribution. Non-parametric control charts including charts of sign control are appropriate for non-normal distributions case. For this situation, the parameters of control chart could be computed with the use of the algorithm of non-parametric bootstrap. In the situation where assumptions of distribution are not required, original observations could be employed.

Jones and William in [8] are among those who have applied bootstrap in the formation of the control charts. In their study, bootstrap was described as a statistical technique which employs power of computing in place of the conventional parametric assumption. The proposed control chart was presented alongside the extensive results of computer simulation, and each control chart was assessed performance wise according to the average length of run.

Niaki and Abbasi in [9], a novel bootstrap-based methodology for deriving the limits of control on the attributes was proposed and formulated. The use of the methodology allows the simultaneous creation of confidence intervals on the attributes. The performance of the proposed method was then examined, in accordance with the in-control and out-of-control average run length criteria. The authors also made a simulation based comparison with a comparable work performed by Bonferroni and Sidak, and the results of the proposed method appeared to be better. Lastly for attributes, the authors made comparison between the bootstrap method and the T^2 control chart.

The application of a bootstrap-based multivariate T^2 control chart was demonstrated in Phaladiganon et al. in [6]. This chart can competently monitor a process in data distribution that is non-normal or unknown. With the application of a simulation study, the authors evaluated the performance of the control chart proposed in their study. The kernel density estimation (KDE)-based T^2 control chart and the conventional Hotelling's T^2 control chart were compared in terms of performance, and from the results of the simulation study, the proposed method demonstrated better performance as opposed to the conventional T^2 control chart. As opposed to the KDE-based T^2 control chart, the proposed method shows comparable performance.

Gandy and Kvaløy in [10] proposed a method grounded upon the bootstrapping concept, where the data were bootstrapped and then employed in the estimation of the in-control state. The use of this method appears to be appropriate for diverse types of control charts. It is also applicable for charts that are based upon regression models. For non-parametric bootstrap, this method is deemed as robust. The author employed large sample properties of the adjustment. The advantages of using the proposed approach were demonstrated using a simulation study.

Edopka and Ogbeide in [11], the authors employed a non-parametric approach in the assessment of the cumulative sum (Cusum) and the exponentially Weighted Moving Average (EWMA) control limits for certain dataset. In the determination of the control limits, the authors employed the underlying dataset conditional distribution. In evaluating the control limits and also in identifying the in-control and out of control of the distribution, the authors applied the method of bootstrap. Here, there was no rigid assumption, for instance, the normality condition for the statistical process control to be dispersed.

In Mostajeran et al. [1], the authors demonstrated the application of a new bootstrap algorithm in the construction of Hotelling's T^2 control chart. In assessing the performance of the proposed method, the authors employed a simulation study. Then, the authors made a comparison between the results of the proposed method and those obtained from the conventional Hotelling's T^2 control chart and also the results of bootstrap reported by Phaladiganon with the application of in-control and out-of-control average run lengths respectively represented by ARL0 and ARL1,.

In Mostajeran et al. [7], the authors presented the use of non-parametric bootstrap multivariate control charts |S|, W, and G, and this method is grounded upon the use of bootstrapped data in the estimation of the in-control state. In this study, the authors succeeded in obtaining satisfactory performance of bootstrap

control charts. Comparison was also made between the performance of the proposed charts and that of the Shewhart-type control charts.

Based on Hotelling's T^2 statistic, Mostajeran et al. in [12] demonstrated the application of a bootstrap multivariate control chart and compared it with a Hotelling's T^2 parametric multivariate control chart, a multivariate sign control chart, and a multivariate Wilcoxon control chart. A simulation study was employed for the purpose.

This study attempts to improve the performance of Hotelling's T^2 chart and therefore, a new method is proposed. In particular, modification on the sensitivities towards outliers is to be carried out. Further, in the construction of the new methodology, this study applies the robust estimator of location as follows: the Hodges-Lehmann estimator and the covariance matrix of the robust scale estimator S_n. Meanwhile, in resampling the data from the normally distributed data, the method of bootstrap is employed.

Accordingly, the concept of Hodges-Lehmann estimator and the properties of the scale estimator of S_n will be highlighted in the next section (Section 2). This is followed by the description of the construction of the Hotelling's T^2 charts. Then, the ensuing section (Section 4) will describe the findings of simulation in a summary form. The final section concludes the paper.

2.   ROBUST LOCATION AND SCALE ESTIMATORS

This paper demonstrates the application of a novel robust location estimator and three robust scale estimators. Aside from allowing easy implementation in the calculation and construction of the Hotelling's T^2 chart, these methods appear to be appropriate technically when dealing with multivariate data. The following section highlights the properties of each estimator.

2.1. Robust location estimator: hodges-lehmann estimator

The location estimation for a sample containing n observations was first introduced in Hodges and Lehmann (1963). This estimator takes a median of the averages of the $\frac{1}{2}n(n-1)$ potential observation pairs. As provided by Brown and Kildea (1978), the estimator is defined as follows:

"A simple Hodges-Lehmann estimator for that $X_j = \theta + Y_j$ for j = 1, 2,…, n where Y_j are i.i.d random vector's symmetric about zero, with density function G and continuous bounded density g. The H-L estimator of θ is the median of $\left\{ \dfrac{X_i + X_j}{2}, 1 < i, j < n \right\}$ and an asymptotically equivalent estimator $\hat{\theta}_n$ is the median of $\left\{ \dfrac{X_i + X_j}{2}, 1 \leq i, j \leq n \right\}$."

The significance of properties of this location estimator has 29% breakdown, symmetric about the parameter θ, about 0.955 asymptotic relative efficiency and it requires O (n^2) operation at minimum.

2.2. Robust scale estimator: S_n

In Rousseeuw and Croux [13], the estimator S_n for the sample $x_1,..., x_n$ was defined as follows:

$$S_n = c \times med_i\{med_j \mid x_i - x_j \mid\} \text{ for } i, j = 1,2,…,n; i < j \tag{1}$$

Where: $c = 1.1926$ denotes a correction factor in making S_n unbiased for predetermined samples. For S_n, its primary properties are as follows: carries 50% maximum breakdown, 58% efficient at normal distribution, limited function of influence, and encompasses an affine equivariance estimator. Accordingly, the work by Rousseeuw and Croux [13] presented more specifics regarding S_n.

3.   CONSTRUCTION OF THE ROBUST HOTELLING'S T^2 CHARTS

The proposed robust Hotelling's T^2 charts have simple construction. The location measure of Hodges-Lehmann estimator is employed in place of the parametric mean vector, while the variance covariance matrices, that is, S_n is employed in place of the variance covariance matrix S.

Given $x_{i1},...,x_{ip}$ for $i = 1,2,..,n$ entails an $n \times p$ matrix where n denotes the size of the bootstrap individual observations, while p =2 denotes the amount of quality characteristics. Accordingly, the calculation of the Hotelling's T^2 charts follows the steps below:

1. For each column within the $n \times p$ matrix, compute the Hodges-Lehmann estimator.
2. Then, the robust variance covariance matrices are computed, follows the steps as described below:
   Considering the symmetric characteristic of variance covariance matrix of S_n, the main diagonal encompasses the matrix of sample variance covariance signified by $S_{jj} = S_j^2$ where j=1 and 2, as demonstrated in [14-18]. For other elements of this matrix, they encompass the covariance between each pair of two variables including X_j, X_g whose computation follows the steps below:
   a. Calculate $S_n(X_j), S_n(X_g), j = 1,2; g = 1,2\ j \neq g$
   b. For ranks $corr(\mathrm{X}_j, \mathrm{X}_g)$ between X_j and X_g, calculate the spearman correlation [19]).
   c. For the scale estimator S_n, repeat steps (i-ii).

The computation of sample covariance between the variables X_j and X_g for 2×2 variance covariance matrix of S_n is based on the formulas shown below:

$$S_n(X_j, X_g) = S_n(X_j)\ S_n(X_g)\ corr(\mathrm{X}_j, \mathrm{X}_g) \tag{2}$$

3. Finally, the new charts of the proposed Hotelling T^2 are formed using the equation below:

$$T^2{}_{HL\_Sn}(X_i) = (X_i - HL)^T S_n^{-1}(X_i - HL) \tag{3}$$

Evaluated was made to the proposed robust Hotelling's T^2 chart with the application of simulated datasets in 5000 replications. As for the simulation, it follows the settings as follows:
a. The general likelihood of false alarm is established at $\alpha = 0.05$,
b. The number of variables encompasses p = 2, and,
c. The sizes of sample n = 20, 30, 40, 50 and 100.

Meanwhile, the chart is formed and assessed in two phases as follows:
a. Phase I produces 5000 datasets from $N_p(\mathbf{0}, \mathbf{I}_p)$ in two circumstances, that is, Case A includes independent variables, while Case B contains dependent variables. Further, the estimators of Hodges and Lehmann (*HL*) and the robust scale covariance matrix for S_n for the conventional and robust charts are calculated.
b. Phase II includes the creation of fresh observation for each dataset in order to allow the performance of assessment.

The performances of the new robust chart is evaluated concerning its false alarms and its likelihood of detecting outliers, and such performances are equivalent to the fraction of the amount of values of robust statistics for new observations which are greater than the upper control limit (UCL) to the amount of replications (5000). For all processes of computation, they are executed with MATLAB version 2015.

4. RESULTS

The outcomes generated by the conventional Hotelling's T^2 and the modified robust Hotelling's T^2 charts, correspondingly labeled as $T^2_{\bar{x}-S}$ and T^2_{HL-Sn} are shown in Tables 1-2. For Case A which contains independent variables are, the results demonstrate the superiority of performance of the modified Hotelling's T^2 chart over the traditional Hotelling's T^2 chart particularly with respect to false alarms. Additionally, according to the outliers' detection, comparable result is identifiable. In this regard, the modified Hotelling's T^2 chart fully supersedes the conventional Hotelling's T^2.

Table 1. False alarm rates and outliers' detecting probability for the conventional and the modified hotelling's T^2 control charts, where: p=2 and α=0.05, ε denotes the percentage of outliers and several values of non-centrality parameters 3 and 5 in case A (case with independent variables)

	n	ε	μ	$T^2_{\bar{x}-S}$	T^2_{HL-sn}
Case(A)	20	0	(0,0)	(3.4)	(3.04)
		0.1	(3,3)	(2.54) 49.7	(6.18) 70
			(5,5)	(3.48) 94	(5.02) 98.5
		0.2	(3,3)	(2.74) 4.22	(4.14) 40.74
			(5,5)	(1.42) 29.66	(3.06) 84.48
	30	0	(0,0)	(2.16)	(5.78)
		0.1	(3,3)	(1.66) 42.54	(2.64) 87.8
			(5,5)	(2.04) 92.02	(2.6) 99.96
		0.2	(3,3)	(0.5) 2.08	(0.94) 65.5
			(5,5)	(0.26) 15.24	(0.78) 96
	40	0	(0,0)	(2.84)	(1.86)
		0.1	(3,3)	(1.66) 40.62	(0.46) 57.8
			(5,5)	(1.02) 97.3	(0.46) 98.84
		0.2	(3,3)	(0.52) 0.82	(0.18) 28.9
			(5,5)	(0.12) 16.1	(0.80) 84.16
	50	0	(0,0)	(5.26)	(7.84)
		0.1	(3,3)	(1.94) 60.06	(2.26) 86.26
			(5,5)	1.02 99.6	2.18 99.98
		0.2	(3,3)	(0.54) 4.86	(1.04) 67.24
			(5,5)	(0.12) 54.7	0.9 98.1
	100	0	(0,0)	(12.06)	(16.44)
		0.1	(3,3)	(1.04) 76.4	(8.7) 96.48
			(5,5)	(0.14) 100	(8.58) 100
		0.2	(3,3)	(2.88) 4.3	(4.74) 80.8
			(5,5)	(0.08) 87.72	(4.56) 99.86

As also can be observed, the modified Hotelling's T^2 chart shows improved rates of false alarms with the increase of the sizes of samples (*n*). Furthermore, for the alternative chart, the probabilities rates for outliers' detection increase with the increase of sample size. However, for the control chart of the modified Hotelling's T^2, the changes of the rates values of the outliers' detection probability appears to be smaller as opposed to the changes demonstrated by the charts of the conventional Hotelling's T^2 especially when the outliers' percentage increases from 0.1 to 0.2 notwithstanding the shifted mean and the size of sample. As can be construed from the generated result, the modified chart is more robust in reacting to changes in the process of production.

For Case B that contains dependent variables as shown in Table 2, the rates of false alarms and those of outliers' detection of the robust charts appear to be superior compared to the exact rates in the conventional chart when there are outliers notwithstanding the n, ε and μ. Notably, the false alarms rates decreases with the increase of the sample size (*n*). Also, the probability detection rates appear to be increasing nearly 100%.

Table 2. Rates of false alarm and probability of detecting outliers for the traditional and the modified hotelling's T^2 control charts when p=2 and α=0.05, ε is the percentage of outliers and several values of non-centrality parameters 3 and 5 in case of dependent variables case (B)

	n	ε	μ	$T^2_{\bar{x}-s}$	T^2_{HL-sn}
Case(B)	20	0	(0,0)	(0.08)	(22.2)
		0.1	(5,5)	(0.5) 88	(8.26) 98.78
		0.2	(5,5)	(0.42) 22.7	(4.94) 88.1
	30	0	(0,0)	(0.06)	(1.34)
		0.1	(5,5)	(0.24) 76.7	(0.6) 98.36
		0.2	(5,5)	(0.14) 8.54	(0.26) 85.96
	40	0	(0,0)	(0)	(0.5)
		0.1	(5,5)	(0.02) 90.5	(0) 87.36
		0.2	(5,5)	(0) 8.94	(0.0) 39.26
	50	0	(0,0)	(0)	(0.36)
		0.1	(5,5)	(0.04) 95.78	(0.06) 96.18
		0.2	(5,5)	(0) 28.02	(0.18) 80.16
	100	0	(0,0)	(0)	(5.12)
		0.1	(5,5)	(0.02) 99.88	(1.76) 99.98
		0.2	(5,5)	(0) 71.58	(0.58) 98.42

5. EMPIRICAL CASE

We used the example from Vargas, Queensberry data sets in order to compare and evaluate results of the performance of both the conventional and modified control charts. Their data comprises of two characteristics, random variables, namely X_1 and X_2 on 30 different products taken from the production process. In Vargas, Queensberry data set's two variables were used. The observations of both random variables are shown in Table 3 (Appendix). The table also shows the values of the new Hotelling's T^2 statistics along with the conventional T^2 statistics.

We calculated the *UCL* using the simulation for the robust and the conventional T^2 charts to be 8.03 and 6.4619 respectively. We set the value of all *UCL* for the robust charts and the conventional for α=0.05. This case has false alarm probability with 30 observations. The final results show that in the case of conventional chart, the production process is not in control at two observations, the second and twentieth observations, whereas the process is out of control only on second observation in case of robust charts.

6. CONCLUSION AND DISCUSSION

The modified robust alternatives Hotelling's T^2 chart demonstrates superiority in performance as opposed to the conventional Hotelling's T^2 chart particularly concerning false alarms. Also, as opposed to the conventional Hotelling's T^2 chart, the modified robust alternatives Hotelling's T^2 chart appears better at outliers detection.

REFERENCES

[1] A. Mostajeran, N. Iranpanah, and R. Noorossana, "A New Bootstrap Based Algorithm for Hotelling's T2 Multivariate Control Chart," *Journal of Sciences, Islamic Republic of Iran,* vol. 27, pp. 269-278, 2016.

[2] I. Almarashdeh, K. E. Eldaw, M. AlSmadi, U. Badawi, F. Haddad, O. A. Abdelkader, G. Jaradat, A. Alkhaldi, and Y. Qawqzeh, "Search Convenience and Access Convenience: The Difference Between Website Shopping and Mobile Shopping," in *International Conference on Soft Computing and Pattern Recognition*, 2018, pp. 33-42.

[3] F. B. Alt, Ed., *Multivariate quality control,* Encyclopedia of Statistical Sciences New York, John Wiley & Sons, 1985.

[4] D. C. Montgomery, Ed., *Introduction to Statistical Quality Control,* John Wiley and Sons Inc, 2005.

[5] J. D. Williams, W. H. Woodall, J. B. Birch, and J. H. Sullivan, "Distribution of Hotelling's T2 Statistics Based on the Successive Differences Estimator," *American Statistical Association,* vol. 38, 2006.

[6] P. Phaladiganon, S. B. Kim, V. C. Chen, J.-G. Baek, and S.-K. Park, "Bootstrap-based T 2 multivariate control charts," *Communications in Statistics—Simulation and Computation,* vol. 40, pp. 645-662, 2011.

[7]   A. Mostajeran, N. Iranpanah, and R. Noorossana, "An evaluation of the multivariate dispersion charts with estimated parameters under non-normality," *Applied Stochastic Models in Business and Industry,* vol. 33, pp. 694-716, 2017.

[8]   L. A. Jones and W. H. Woodall, "The performance of bootstrap control charts," *Journal of Quality Technology,* vol. 30, pp. 362-375, 1998.

[9]   S. T. A. Niaki and B. Abbasi, "Bootstrap method approach in designing multi-attribute control charts," *The International Journal of Advanced Manufacturing Technology,* vol. 35, pp. 434-442, 2007.

[10] A. Gandy and J. T. Kvaløy, "Guaranteed conditional performance of control charts via bootstrap methods," *Scandinavian Journal of Statistics,* vol. 40, pp. 647-668, 2013.

[11] I. Edopka and E. Ogbeide, "Bootstrap approach control limit for statistical quality control," *International Journal of Engineering Science Invention,* vol. 2, pp. 2319-6734, 2013.

[12] A. Mostajeran, N. Iranpanah, and R. Noorossana, "An explanatory study on the non-parametric multivariate T2 control chart," *Journal of modern applied statistical methods,* vol. 17, pp. 12, 2018.

[13] P. J. Rousseeuw and C. Croux, "Alternatives to the Median Absolute Deviation," *American Statistical Association,* vol. 88, 1993.

[14] P. J. Rousseeuw, *Robust estimation and identifying outliers*, ed: McGraw-Hill, New York, pp. 16.1-16.24, 1990.

[15] F. Haddad, S. Syed Yahaya, and J. L. Alfaro, "Robust Hotelling's Charts based on High Breakdown Points Estimators," in *Icoqsia*, 2010, 2-4 November, Penang Malaysia, 2010.

[16] F. Haddad, S. Syed Yahaya, and J. L. Alfaro, "Alternative Hotelling's T2 Charts using Winsorized Modified One-Step M-estimator," *Quality and Reliability Engineering International,* vol. 29, pp. 583-593, 2013.

[17] F. Haddad and M. K. Alsmadi, "Improvement of The Hotelling's T2 Charts Using Robust Location Winsorized One Step M-Estimator (WMOM)," *Journal of Mathematics (ISSN 1016-2526),* vol. 50, pp. 97-112, 2018.

[18] F. Haddad, J. Alfaro, and M. K. Alsmadi, "Hotelling's T² Charts Using Winsorized Modified One Step M-Estimator For Individual Non Normal Data," *Journal of Theoretical & Applied Information Technology,* vol. 72, pp. 215-226, 2015.

[19] M. B. Abdullah, "On robust correlation coefficient," *statistician,* vol. 39, pp. 455-460, 1990.

Appendix

Table 3. The two variables X_1 and X_2 of vargas data set with the values of T^2 statistics using the conventional and the winsorized MOM estimator

Product No	X_1	X_2	T_0^2	$T^2_{\text{HL-Sn}}$
1	0.567	60.558	0.807	0.763
2	0.538	56.303	**12.975**	**22.034**
3	0.53	59.524	0.1373	0.331
4	0.562	61.102	1.8375	2.208
5	0.483	59.834	1.5697	1.742
6	0.525	60.228	0.33	0.351
7	0.556	60.756	0.977	1.083
8	0.586	59.823	0.904	0.955
9	0.547	60.1530	0.1269	0.072
10	0.531	60.64	0.801	0.9687
11	0.581	59.785	0.7192	0.785
12	0.585	59.675	0.910	1.0676
13	0.54	60.489	0.483	0.52
14	0.458	61.067	5.2413	6.585
15	0.554	59.788	0.073	0.11
16	0.469	58.64	3.5357	4.6365
17	0.471	59.574	2.2696	2.539
18	0.457	59.718	3.2442	3.5497
19	0.565	60.901	1.398	1.5629
20	0.664	60.18	**6.8326**	6.906
21	0.6	60.493	1.8978	1.767
22	0.586	58.37	3.3564	5.667
23	0.567	60.216	0.427546	0.332
24	0.496	60.214	1.1838	1.343
25	0.485	59.5	1.4968	1.7537
26	0.573	60.052	0.48432	0.41376
27	0.52	59.501	0.28989	0.50
28	0.556	58.476	2.0635	3.8619
29	0.539	58.666	1.38596	2.693
30	0.554	60.239	0.24043	0.1697