CHILEAN JOURNAL OF STATISTICS

Edited by Víctor Leiva and Carolina Marchant

A free open-access journal indexed by

Volume 12 Number 2
December 2021
ISSN: 0718-7912 (print)
ISSN: 0718-7920 (online)

Published by the
Chilean Statistical Society
SOCHE
SOCIEDAD CHILENA DE ESTADÍSTICA

Aims

The Chilean Journal of Statistics (ChJS) is an official publication of the Chilean Statistical Society (`www.soche.cl`). The ChJS takes the place of *Revista de la Sociedad Chilena de Estadística*, which was published from 1984 to 2000.

The ChJS covers a broad range of topics in statistics, as well as in artificial intelligence, big data, data science, and machine learning, focused mainly on research articles. However, review, survey, and teaching papers, as well as material for statistical discussion, could be also published exceptionally. Each paper published in the ChJS must consider, in addition to its theoretical and/or methodological novelty, simulations for validating its novel theoretical and/or methodological proposal, as well as an illustration/application with real data.

The ChJS editorial board plans to publish one volume per year, with two issues in each volume. On some occasions, certain events or topics may be published in one or more special issues prepared by a guest editor.

Chilean Journal of Statistics

Volume 12, Number 2
December 2021

ISSN: 0718-7912 (print)/ISSN 0718-7920 (online)

http://www.soche.cl/chjs

Contents

Chilean Journal of Statistics
Vol. 12, No. 2, December 2021, 171–187

Statistical Quality Control
Research Paper

Robust Hotelling T^2 control chart using adaptive reweighted minimum covariance determinant estimator

Esra Polat[1,*]

[1]Department of Statistics, Hacettepe University, Ankara, Turkey

(Received: 13 November 2020 · Accepted in final form: 14 May 2021)

Abstract

The classical Hotelling T^2 control chart using classical mean and covariance estimators is not efficient in case of outliers existence in data. To overcome this issue, robust mean and covariance estimators are used in literature. Hence, a robust Hotelling T^2 control chart is proposed based on the adaptive reweighted minimum covariance determinant estimator which is a good option to the classical multivariate T^2 chart in case of outliers presence. The new proposed chart's performance is evaluated by false alarm rates and probability of detection/percentage of outliers detection, later a comparison is made with the performance of the classical Hotelling T^2 chart and the chart obtained using the minimum covariance determinant estimator. Simulation and real data application results are indicated that proposed control chart has better performance in comparison to robust control chart based on the minimum covariance determinant especially in terms of false alarm rates and it performs better than classical chart in terms of probability of detection.

Keywords: Hotelling T^2 · Minimum covariance determinant · Multivariate control chart · Robust estimator · Statistical quality control.

Mathematics Subject Classification: Primary 62P30 · Secondary 62F35.

1. Introduction

In manufacturing process, multiple quality characteristics of a product are generally observed. Hence, multivariate control charts may be a suitable tool to observe the process. The Hotelling T^2 chart is the most commonly known one because it's application is easy, it is flexible, it is sensitive to little process modifications and the software for it's application is available. The Hotelling T^2 uses the classical mean and covariance matrix, is reactive to the outlier. Because in case of more quality characteristics are considered, the risk of multiple outliers existence is getting higher. In case of outliers existence, the classical control chart's performance decreases. Because of the masking effect, the classical method is not effective for multiple outliers case (Alfaro and Ortega, 2009). The masking effect in the monitoring process occurs as a result of the outlier, which cannot be detected by the control chart. To overcome the problem that arises, several robust methods have been proposed for reducing

*Corresponding author. Email: espolat@hacettepe.edu.tr

ISSN: 0718-7912 (print)/ISSN 0718-7920 (online)

http://www.soche.cl/chjs

the effect of multiple outliers by substituting the existing estimators with the more robust ones. Furthermore, the performance of Hotelling T^2 control chart, in detecting the shift of mean vector, is increasing when the robust covariance matrix estimator is implemented (Williams et al., 2006; Ahsan et al., 2019).

Similar to control charts observing variability in a process, its structure arises from Phase I, Phase II (Alt, 1985), as well called retrospective and prospective analysis, in order of. The significant point of Phase I is the analysis of historical data for determining if the process is under control or not by estimation of the in-control parameters and control limits of the process. However, in case of Phase II, the focus is to monitor on-line data for rapidly finding shifts of process from the estimated in-control parameter values in Phase I. Outliers in Phase I may cause the increment of control limits and decrease of power for the detection of process changes in Phase II. Hence, Phase II analysis achievement based on a success in Phase I analysis in the estimation of in-control mean, variance and covariance parameters.

Ordinal Hotelling T^2 chart is a safe method when the underlying process data really has the normal distribution. In contrast to this, in case of outliers presence in data it is not a safe method for detecting out of control points properly. Because classical mean and covariance estimators in the original formulae cannot resist the outliers. Thereby, the classical Hotelling $T^{2'}$s chart ability for monitoring future process data is debatable. One of the way of getting rid of this issue is using control chart which is robust in case of outliers existence.

Up to now, many robustified versions of the Hotelling T^2 control chart have been proposed by utilizing from robust estimators. Abu-Shawiesh and Abdullah (2001) estimated the mean vector using Hodges-Lehmann and the variance-covariance matrix using Shamos-Bickel-Lehman. Vargas (2003) and Jensen et al. (2007) presented robust control charts using minimum covariance determinant (MCD) and minimum volume ellipsoid (MVE) estimators. They detected and omitted the outliers in Phase I data and later compute the traditional estimators using the remained clean observations in case of Phase II data. Although in this method, the breakdown point and calculation of the estimators become more significant, however, statistical efficiency does not become as critical as the extremely robust estimators change place with by classical estimators in Phase II case. When MVE and MCD are used in Phase I, they recognized some problems, such as T^2 obtained by using MVE performed badly under large sample size. However, T^2 obtained by using MCD requires more sample size if a lot of outlying observations is skeptical to guarantee that MCD estimator loses its ability especially in case of monitoring with higher dimensions (p) and it does not breakdown. Alfaro and Ortega (2008) introduced robust Hotelling T^2 control charts by changing the arithmetic mean with trimmed one and sample covariance with sample trimmed covariance. Chenouri et al. (2009) presented a robust chart based on reweighted MCD (RMCD) estimator that it is not overly affected by outlying observations and has better efficiency than MCD. The difference of their method from Vargas (2003) and Jensen et al. (2007) that they use RMCD estimators instead of traditional estimators in establishing Hotelling T^2 chart for Phase II data set. Alfaro and Ortega (2009) compared the performance of Hotelling T^2 control charts using robust MVE, trimmed, MCD and RMCD estimators. The result of this study was that the recommendation of the use of T^2 charts obtained by using RMCD and trimmed estimator in case of not many outlying observations in the production process since these two methods are able to control false alarm rates (FAR).

In the producing of products that concentrates mostly on determining the outlying observations compared to the false alarms, that is, a point outside the control limits for an in-control process (Da Silva et al, 2019), T^2 obtained by using MCD may be taken into consideration as the best option. Because Hotelling T^2 control charts based on MCD has a better performance in terms of probability of detection (POD). In theory, if the POD gets higher, the chart could also control the overall FAR α (Jensen et al., 2007). In spite of this, the results in Alfaro and Ortega (2009) revealed a discordance between the capability of

robust control chart in controlling the overall FAR and POD in case of specific situations. Yañez et al. (2010) constructed the T^2 control chart by using the biweight S estimators for mean and covariance estimators. Their chart outperformed the T^2 chart based on MVE for a small number of observations. Yahaya et al. (2011) presented the minimum variance vector (MVV) estimator in T^2 chart in order to observe the Phase II data. Overall, the robust control chart gave a quick detection in out-of-control status and at the same time, capable for controlling the overall FARs nevertheless as the p is increased. The only disadvantage was a large upper control limits (UCLs) in comparison to the classical T^2 chart. An improved version of the MVV chart was further suggested by Ali et al. (2013) to obtain desired UCLs whilst still it performs well in terms of FAR and POD. This was achieved by making the MVV estimators consistent at normal distribution as well as unbiased for finite samples. Ali et al. (2014) investigated the performance of reweighted version of MVV (RMVV) in constructing the Hotelling T^2 chart. Yahaya et al. (2019) introduced three robust Hotelling T^2 control charts using trimmed estimators. The modified Mahalanobis distance with median used as the location measure and one of the scale estimators MAD_n, S_n (mean absolute average) or T_n as the scale measure. As a consequence of these alternatives, three dissimilar trimmed estimators are introduced. The findings of their study revealed that their three control charts performance is moderate in terms of false alarms and magnificently for POD, outperform the classical control chart in any case of conditions. In case of outliers existence or samples deviation from normality, all of the studies revealed that the robust control charts surpass the classical Hotelling T^2 control chart.

In this study, following the robust Hotelling T^2 control chart literature, a robust Hotelling T^2 control chart is introduced that uses a robust adaptive reweighted minimum covariance determinant (ARWMCD) estimator. The new control chart's performance is evaluated by FARs and POD by doing a simulation and a real data application. Moreover, the performance of the new method is compared with robust control chart using MCD estimator and the classical chart.

The rest of the paper is organized as follows. Section 2 reviews the ARWMCD estimator. In Section 3, we present the new proposed Hotelling T^2 control chart. Section 4 contains a simulation study where the performance of the new robust Hotelling T^2 chart using ARWMCD estimator is compared to classical Hotelling T^2 chart and the robust Hotelling T^2 chart using MCD. In Section 5, we illustrate the performance of the new proposed robust Hotelling T^2 chart based on ARWMCD on the real data that is given in Ali et al. (2013). Finally, Section 6 collects some conclusions about the present study.

2. Adaptive reweighted minimum covariance determinant estimator

In addition to maximum robustness against to outliers, robust multivariate estimators must also propose a sensible efficiency for the normal distribution and a controllable asymptotic distribution. Nevertheless, MCD and MVE estimators do not satisfy that condition. Gervini (2003) expressed that considering the both of being robust and efficient, the best way utilizing a two-stage process. Rousseeuw and Van Zomeren (1990) also expressed that in this process, first of all, a tremendously robust nevertheless maybe not efficient estimator is calculated and it is used for observing outlying observations and calculating the sample location and covariance of the good data. This process comprises of omitting sample points whose Mahalanobis distances go beyond a certainly fixed threshold value. As beginning estimator for that processes, Rousseeuw and Van Driessen (1999) suggested an algorithm for computing MCD estimator, which does not ensure that the precise estimator is obtained, it is quicker and more precise than formerly obtained algorithms also for highly bigger data ($n \gg p$). The advantage of the $1/\sqrt{n}$ convergence rate, in addition to this truth, could indicate that the MCD technique uses the FAST-MCD algorithm is the best preference when compared to MVE for beginning estimator of a two-step process (Gervini, 2003).

MCD is investigating for those h observations for which the determinant of the traditional covariance matrix is minimum. Therefore, the MCD estimators are the location and covariance matrix of that h observations. The computation of MCD estimation is hard. The application of MCD estimator on data sets could merely be in case of the number of observations exceeds the number of variables ($n > p$). Because in case of $p > n$ then also $p > h$, and often the covariance matrix of any h observations is going to be singular, tends to zero determinant. Henceforth, each subset of h observations would tend to the minimum feasible determinant, resulting in a non-unique solution (Filzmoser et al., 2009). FAST-MCD algorithm can handle with larger sizes of sample such as tens of thousands. This algorithm obtains precise solution for small sizes of data and it is quicker and more precise than formerly proposed algorithms, yet for extremely big data sets. Since it is efficient and fast in calculation, Rousseeuw and Van Driessen (1999) proposed of using FAST-MCD algorithm for estimating mean and covariance. Since the raw MCD estimators of mean and covariance are reweighted for improving the finite sample efficiency, named as reweighted MCD (RMCD) estimators (Hubert and Vanden-Branden, 2003). Since it is very popular algorithm for robust literature, a brief information about FAST-MCD is given. Any interested reader could find for detailed information in Rousseeuw and Van Driessen (1999). The algorithms steps for p dimensional vector x_i, for $i = 1, \ldots, n$, as follows.

Step 1: The MCD estimates could withstand $(n - h)$ outlying observations, therefore h (or equally the fraction $\alpha = h/n$) specifies the robustness of the estimator. $(1-\alpha)$ measures the fraction of outliers the algorithm should resist. Any value between 0.5 and 1 may be specified (default = 0.75). In FAST-MCD algorithm by taking $[(n+p+1)/2]$ as the accepted value of h, highly resist against outliers. Nevertheless, any integer h in the interval $[(n+p+1)/2] \leq h < n$ could be used by researcher. In case of a huge fraction of outliers is assumed in data set, thereby, h must be selected as $h = [0.5n]$. Also, if it is correct that the data includes not much than 25% of outliers that is often the condition, a better balance between statistical efficiency and breakdown value is captured by choosing $h = [0.75n]$ (Rousseeuw and Van Driessen, 1999). In this study, we have also used the default value of $h = [0.75n]$.

Step 2: From here on $h < n$ and $p \geq 2$. If n is small (say, $n < 600$) then:

- Repeat (say) 500 times:
  - ✓ Construct an initial h-subset H_1 using method given in Rousseeuw and Van Driessen (1999), that is, starting from a random $(p + 1)$-subset.
  - ✓ Carry out two C-steps described in Rousseeuw and Van Driessen (1999).
- For the 10 results with lowest $\det(\widehat{\boldsymbol{\Sigma}}_3)$:
  - ✓ Conduct C-steps until convergence
- Report the solution $(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}})$ with lowest $\det(\widehat{\boldsymbol{\Sigma}})$.

Step 3: If n is larger (say, $n \geq 600$), then:

- Construct up to five disjoint random subsets of size n_{sub} according to given in Rousseeuw and Van Driessen (1999) (say, subsets of size $n_{\text{sub}} = 300$).
- Inside each subset, repeat 500/5=100 times:
  - ✓ Construct an initial subset H_1 of size $h_{\text{sub}} = [n_{\text{sub}}(h/n)]$.
  - ✓ Carry out two C-steps, using n_{sub} and h_{sub}.
  - ✓ Keep the 10 best results $(\widehat{\boldsymbol{\mu}}_{\text{sub}}, \widehat{\boldsymbol{\Sigma}}_{\text{sub}})$.
- Pool the subsets, yielding the merged set (say, of size $n_{\text{merged}} = 1500$).
- In the merged set, repeat for each of the 50 solutions $(\widehat{\boldsymbol{\mu}}_{\text{sub}}, \widehat{\boldsymbol{\Sigma}}_{\text{sub}})$:
  - ✓ Conduct two C-steps, using n_{merged} and $h_{\text{merged}} = [n_{\text{merged}}(h/n)]$.
  - ✓ Keep the 10 best results $(\widehat{\boldsymbol{\mu}}_{\text{merged}}, \widehat{\boldsymbol{\Sigma}}_{\text{merged}})$.

- In the full data set, repeat for the m_{full} best results:
  - ✓ Take several C-steps, using n and h.
  - ✓ Keep the best final result $(\widehat{\boldsymbol{\mu}}_{\text{full}}, \widehat{\boldsymbol{\Sigma}}_{\text{full}})$.

Here, m_{full} and the number of C-steps (preferably, until convergence) depend on how large the data set is (Rousseeuw and Van Driessen, 1999; Polat and Gunay, 2019). This algorithm is called as FAST-MCD. It is affine equivariant: when the data are translated or subjected to a linear transformation, the resulting $(\widehat{\boldsymbol{\mu}}_{\text{full}}, \widehat{\boldsymbol{\Sigma}}_{\text{full}})$ transforms accordingly. For convenience, the computer program contains two more steps (Rousseeuw and Van Driessen, 1999).

Step 4: In order to obtain consistency when the data come from a multivariate normal distribution, $\widehat{\boldsymbol{\mu}}_{\text{MCD}} = \widehat{\boldsymbol{\mu}}_{\text{full}}$ and $\widehat{\boldsymbol{\Sigma}}_{\text{MCD}} = (\text{med}_{\text{i}}\, d^2_{(\widehat{\boldsymbol{\mu}}_{\text{full}}, \widehat{\boldsymbol{\Sigma}}_{\text{full}})}(i)/\chi^2_{p,0.5})\widehat{\boldsymbol{\Sigma}}_{\text{full}}$ are placed.

Step 5: One-step reweighted estimates could be obtained by reweighting each observation as

$$w_i = \begin{cases} 1, & \text{if } (\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}_{\text{MCD}})^\top \widehat{\boldsymbol{\Sigma}}^{-1}_{\text{MCD}} (\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}_{\text{MCD}}) \leq \chi^2_{p,0.975}, \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, using the weights w_i, the RMCD estimators are calculated as

$$\widehat{\boldsymbol{\mu}}_{\text{RMCD}} = \frac{\sum_{i=1}^n w_i \boldsymbol{x}_i}{\sum_{i=1}^n w_i} \quad \text{and} \quad \widehat{\boldsymbol{\Sigma}}_{\text{RMCD}} = \frac{\sum_{i=1}^n w_i (\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}_{\text{RMCD}})(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}_{\text{RMCD}})^\top}{\sum_{i=1}^n w_i}.$$

If it is desirable that the estimator to be robust and efficient, a two-step process is suggested as a best preference. Gervini (2003) suggested basically enhancement above Rousseeuw and Van Zomeren (1990) that a reweighted one-stage estimator using adaptive threshold values. This adaptive reweighting system can keep the outlier robustness of the starting estimator in bias and breakdown, at the same time, reach 100% efficiency for the normal distribution. For the first time, Gervini and Yohai (2002) suggested this type of adaptive reweighting for the linear regression model. This conception is widened by Gervini (2003) that he suggested an adaptive technique for multivariate mean and covariance estimation.

Since $x_1, \ldots, x_n$ is a sample of under consideration in $\Re^p$ and beginning robust estimators of mean and covariance are $\widehat{\boldsymbol{\mu}}_{0n}, \widehat{\boldsymbol{\Sigma}}_{0n}$ (in our study, they are obtained by MCD estimator using FAST-MCD algorithm) then the Mahalanobis distances are stated as (Gervini, 2003; Polat and Gunay, 2019).

$$d_i := d\Big(\boldsymbol{x}_i, \widehat{\boldsymbol{\mu}}_{0n}, \widehat{\boldsymbol{\Sigma}}_{0n}\Big) = \Big\{(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}_{0n})^\top \widehat{\boldsymbol{\Sigma}}^{-1}_{0n} (\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}_{0n})\Big\}^{1/2}.$$

Under normality assumption, d_i^2 nearly have a χ^2_p distribution and logically, being suspicious about data points with $d_i^2 \geq \chi^2_{p,0.975}$ as an outlier. Rousseeuw and Van Zomeren (1990) suggested to omit those outlying data points and calculated the sample mean and covariance matrix of left of the data set. Hence, by this method, they obtained new estimators $(\widehat{\boldsymbol{\mu}}_{1n}, \widehat{\boldsymbol{\Sigma}}_{1n})$; see Gervini (2003).

Gervini (2003) expressed that MCD estimators can be taken under consideration as the beginning robust estimators of mean and covariance in the adaptive reweighted procedure because the MCD technique computed using FAST-MCD algorithm is developed as a good option instead of MVE. Therefore, similar as in Polat and Gunay (2019), adaptive reweighted technique including the MCD estimators $(\widehat{\boldsymbol{\mu}}_{\text{MCD}}, \widehat{\boldsymbol{\Sigma}}_{\text{MCD}})$ is used as beginning robust estimators of mean and covariance $(\widehat{\boldsymbol{\mu}}_{0n} = \widehat{\boldsymbol{\mu}}_{\text{MCD}}, \widehat{\boldsymbol{\Sigma}}_{0n} = \widehat{\boldsymbol{\Sigma}}_{\text{MCD}})$. This technique had been named as ARWMCD and robust estimators, denoted as $\widehat{\boldsymbol{\mu}}_{\text{ARWMCD}}, \widehat{\boldsymbol{\Sigma}}_{\text{ARWMCD}}$.

This reweighting stage rises up the beginning estimators efficiency and also keeps its robustness mostly. The threshold $\chi^2_{p,0.975}$ is a subjective value. Although they show a normal distribution, in case of big data sets noticeable number of data points must to be omitted out of analysis. For this issue, it is found that the best option constructing an adaptive threshold values, which gets higher related to n in case of the data are uncontaminated, however, stays bounded in case of outliers presence in the sample. The procedure of this method is as in follows. Note that the expression stated as

$$G_n(u) = \frac{1}{n}\sum_{i=1}^{n} I\Big(d^2\Big(\boldsymbol{x}_i, \widehat{\boldsymbol{\mu}}_{\mathrm{MCD}}, \widehat{\boldsymbol{\Sigma}}_{\mathrm{MCD}}\Big) \leq u\Big),$$

where $G_p(u)$ is the χ^2_p distribution function, shows the experimental distribution of the squared Mahalanobis distances (Gervini, 2003; Polat and Gunay, 2019).

The approximation of G_n to G_p is assumed in case of the sample has normal distribution. Hence, comparing the tails of G_n with the tails of G_p is a technique of detection for outliers. In case of $\eta = \chi^2_{p,1-\alpha}$ for a fixed small α, for example $\alpha = 0.025$, we have (Gervini, 2003; Polat and Gunay, 2019)

$$\alpha_n = \sup_{u \geq \eta}\{G_p(u) - G_n(u)\}^+, \tag{1}$$

where $\{\cdot\}^+$ denotes the positive part. Note that α_n could be considered as an outlier measurement in the sample. It only allows positive differences in Equation (1) because a negative difference does not show existence of outliers. If $d^2_{(i)}$ shows the ith order statistic of the squared Mahalanobis distances and $i_0 = \max\{i\colon d^2_{(i)} < \eta\}$, then Equation (1) comes down to as (Gervini, 2003; Polat and Gunay, 2019)

$$\alpha_n = \max_{i > i_0}\left\{G_p\Big(d^2_{(i)}\Big) - \frac{i-1}{n}\right\}^+.$$

Those data points giving the largest $\lfloor \alpha_n n \rfloor$ distances are taken under consideration as outlying points and omitted in the reweighting stage, with $\lfloor a \rfloor$ showing the largest integer that is $\leq a$. The cut-off value is given by

$$c_n = G_n^{-1}(1 - \alpha_n),$$

where $G_n^{-1}(u) = \min\{s\colon G_n(s) \geq u\}$, $c_n = d^2_{(i_n)}$, with $i_n = n - \lfloor \alpha_n n \rfloor$ and that $i_n > i_0$ as a outcome of the description of α_n. Therefore, $c_n > \eta$. To describe the reweighted estimator, weights are stated as (Gervini and Yohai, 2002; Polat and Gunay, 2019)

$$w_{i,n} = w\left(\frac{d^2\Big(\boldsymbol{x}_i, \widehat{\boldsymbol{\mu}}_{\mathrm{MCD}}, \widehat{\boldsymbol{\Sigma}}_{\mathrm{MCD}}\Big)}{c_n}\right). \tag{2}$$

The weight function defined as $w\colon [0,\infty) \to [0,1]$ is non-increasing, with $w(u) = 0$ when $u \in [1,\infty)$ and $w(u) > 0$ when $u \in [0,1)$, $w(0) = 1$. The simplest choice among those functions satisfying it is the hard-rejection function $w(u) = I(u < 1)$, which is the one most commonly used in practice.

Once the weights in Equation (2) are calculated, the one-stage reweighted $(\widehat{\boldsymbol{\mu}}_{\text{ARWMCD}}, \widehat{\boldsymbol{\Sigma}}_{\text{ARWMCD}})$ are given as

$$\widehat{\boldsymbol{\mu}}_{\text{ARWMD}} = \frac{\sum_{i=1}^{n} w_{i,n}\boldsymbol{x}_i}{\sum_{i=1}^{n} w_{i,n}} \tag{3}$$

and

$$\widehat{\boldsymbol{\Sigma}}_{\text{ARWMCD}} = \frac{\sum_{i=1}^{n} w_{in}(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}_{\text{ARWMCD}})(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}_{\text{ARWMCD}})^{\top}}{\sum_{i=1}^{n} w_{i,n}}. \tag{4}$$

3. The new proposed Hotelling T^2 control chart

The p dimensional random sample of n observations of prior data in case of Phase I is shown by $\boldsymbol{x}_i = \{x_1, \ldots, x_n\}$, where $\boldsymbol{x}_i$ are supposed to be independent and have a multivariate normal distribution with covariance matrix $\boldsymbol{\Sigma}$ and mean vector $\boldsymbol{\mu}$. In case of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are not known then the estimation of them utilizing an in-control data set is needed. The procedure of describing the in-control data set from $\boldsymbol{x}_i$ is mentioned as Phase I action. Using preliminary data set, $\overline{x}$ and S are computed. These estimates are used to compute $T^2_{(i)}$ for $i = 1, \ldots, n$ based on

$$T^2_{(i)} = (\boldsymbol{x}_i - \bar{\boldsymbol{x}})\boldsymbol{S}^{-1}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^{\top},$$

To obtain an in-control data set, describe outliers utilizing UCL established on the beta distribution given by

$$\text{UCL}_1 \sim \left[\frac{(n-1)^2}{n}\right] B_{\left(\alpha, \frac{p}{2}, \frac{n-p-1}{2}\right)},$$

where $B(\alpha, p/2, (n-p-1)/2)$ is the $100 \times (1-\alpha)\%$ quantile of the beta distribution with $p/2$ and $(n-p-1)/2$ degrees of freedom, whereas α is the overall FAR.

The sample points where $T^2_{(i)} > \text{UCL}_1$ are omitted that since they are outliers. The clean data set that the outlying observations are omitted (n_c) is then used for computing the new estimations, $\overline{x}_N$ and S_N. These estimations are used for computing $T^2_{(g)}$ statistic for Phase II observation, where $x_g \notin \boldsymbol{x}_i$, such that

$$T^2_{(g)} = (\boldsymbol{x}_g - \overline{\boldsymbol{x}}_N)\boldsymbol{S}_N^{-1}(\boldsymbol{x}_g - \overline{\boldsymbol{x}}_N)^{\top}.$$

By using the desired values of α, p and n_c, compute the LCL and UCL using the F distribution as

$$\text{UCL} \sim \left[\frac{p(n_c+1)(n_c-1)}{n_c(n_c-p)}\right] F_{(\alpha,p,n_c-p)} \quad \text{and} \quad \text{LCL} = 0,$$

where $F_{(\alpha,p,n_c-p)}$ is the $100(1-\alpha)\%$ quantile of the F distribution with p and $n-p$ degrees of freedom and α is the overall FAR. Nevertheless, this classical procedure is merely effective in excluding very unusual outlying observations and observing large shift in the mean vector in small sample sizes, however, it is not successful for detecting more moderate outlying observations specifically when number of variables inflated (Vargas, 2003; Jensen et al., 2007; Chenouri et al., 2009). To overcome this issue of the process, in this study, ARWMCD

estimator is used in Phase I data of $\boldsymbol{x}_i$. As it is known that ARWMCD gives robust estimators of covariance and mean, those are used as in-control estimators in Phase II, where the Phase II observations are $\boldsymbol{x}_g = \{x_{n+1}, x_{n+2}, \dots\}$, $\boldsymbol{x}_g \notin \boldsymbol{x}_i$.

The procedure for new robust chart is as follows. First, from the Phase I data set, $\boldsymbol{x}_i$, the ARWMCD location vector and covariance matrix estimators $\overline{x}_{\text{ARWMCD}}(\widehat{\boldsymbol{\mu}}_{\text{ARWMCD}})$ and $S_{\text{ARWMCD}}(\widehat{\boldsymbol{\Sigma}}_{\text{ARWMCD}})$ are obtained as in Equations (3) and (4). Then, a robust Hotelling T^2 ($T^2_{\text{ARWMCD}(g)}$) for Phase II data, x_g, is defined based on these ARWMCD estimates (obtained from Phase I data) as

$$T^2_{\text{ARWMCD}(g)} = (\boldsymbol{x}_g - \overline{\boldsymbol{x}}_{\text{ARWMCD}})\boldsymbol{S}^{-1}_{\text{ARWMCD}}(\boldsymbol{x}_g - \overline{\boldsymbol{x}}_{\text{ARWMCD}})^{\top}. \tag{5}$$

The UCL, FAR and POD calculations are explained under Section 4 in detail.

4. Simulation study

Robust estimators are used in place of the traditional mean and covariance in T^2 chart, which causes the replacing of distributional properties of the classical T^2 control chart (Williams et al., 2006). As the sampling distribution of the suggested Hotelling T^2 chart T^2_{ARWMCD} is not known, the UCL is estimated with simulation. Moreover, as the distribution of T^2_{ARWMCD} is not known, simulations were done for estimating the quantiles of the T^2_{ARWMCD}, for few combines of dimensions and sample sizes as shown in Table 1. Even, the finite sample distribution of the MCD estimators is still questionable, thus, the distribution of T^2_{MCD} is also unknown (Vargas, 2003; Jensen et al., 2007; Chenouri et al., 2009; Alfaro and Ortega, 2009). Therefore, quantile is also used in estimating the distribution of T^2_{MCD} obtained via Monte Carlo method.

First of all, data sets are originated from the standard multivariate normal distribution $\text{N}_p(\boldsymbol{0}, \boldsymbol{I}_p)$. Then, robust estimators are computed from this distribution. Next, a new additional sample point from the standard multivariate normal distribution is generated and robust Hotelling T^2 statistic for this new sample point is computed. This process is repeated 5000 times and the $95th$ percentile of the 5000 robust Hotelling T^2 statistics considered as the UCL. For assessing the performance of T^2_{ARWMCD} by comparison with classical T^2_0 and robust T^2_{MCD} control charts, several conditions are generated by changing number of dimensions (p), observations (n) and percentage of outliers (ε) and a variety of mean shifts values ($\boldsymbol{\mu}_1$) as shown in Table 1.

Table 1. The Simulation settings

Variables	Values
Number of quality characteristics (p)	2, 5, 10
Proportion of contamination (ε)	0.1, 0.2
Mean shift ($\boldsymbol{\mu}_1$)	0 (no shift), 3, 5
Group size (n)	50, 100, 200

To estimate the 95% quantile of $T^2_{\text{ARWMCD}(g)}$ firstly, for a Phase I case with a sample size n and dimension $p, K = 5000$ samples of size n from a standard multivariate normal distribution $N_p(0, I_p)$ are generated. For different sample sizes n, the ARWMCD mean vector and covariance matrix estimates are computed, $\boldsymbol{\mu}_{\text{ARWMCD}}(k)$ and $\boldsymbol{S}_{\text{ARWMCD}}(k)$, for $k = 1, \dots, K$. Additionally, for each data set, a new observation $X_{g,k}$ is randomly generated that it is handled as a Phase II sample point from $\text{N}_p(\boldsymbol{0}, \boldsymbol{I}_p)$ and the corresponding $T^2_{\text{ARWMCD}(g,k)}$ values are computed as given by Equation (5). The simulated values as

$T^2_{\text{ARWMCD}(g,1)}, \ldots, T^2_{\text{ARWMCD}(g,K)}$ are used to obtain the empirical distribution function of $T^2_{\text{ARWMCD}(g)}$. Then, $T^2_{\text{ARWMCD}(g,K)}$ values are sorted in ascending order and the UCL is the 95th quantile of the 5000 statistics. The UCL values for classical and MCD control charts are also estimated by using this technique.

4.1 Performance evaluation

The classical and two Robust Hotelling T^2 charts success is evaluated in terms of the FAR and POD for Phase II data. Hence, for Phase II sample points, 1000 new datasets were simulated from the standard normal distribution $\mathrm{N}_p(\mathbf{0}, \boldsymbol{I}_p)$ of various sample sizes (n) and dimensions (p) as shown in Table 1. For deciding the FAR and POD, a Phase II sample point is randomly generated with in-control and out of control parameters respectively from Phase I and the robust Hotelling T^2 statistics are calculated. FAR is calculated using a new sample point from the in-control distribution, however, the POD is computed with a new sample point generated from the out-of-control distribution. The FAR or POD is predicted as the percentages of statistic values which are over the control limits of 1000 repetitions. In case of Phase I, several conditions of data sets are simulated by changing the number of observations, dimensions and proportions of contamination. By mixing normal distributions similar as in Alfaro and Ortega (2009), a contaminated model stated as

$$(1-\varepsilon)\mathrm{N}_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) + \varepsilon \mathrm{N}_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) \tag{6}$$

is used for investigating the effect of outliers on the charts success. Here, ε is the percentage of outlying observations, $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ are the in-control parameters, however, $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_1$ are the out-of-control parameters. Contamination with shift in the mean, however, not any changes in covariance is assumed, henceforth, the covariance matrix $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_1$ in Equation (6) are p dimensional identity matrices ($\boldsymbol{I}_p$). Four variables are changed to investigate the strengths and the weaknesses of the classical and robust Hotelling T^2 charts namely number of quality characteristics (p), proportion of contamination (ε), mean shift ($\boldsymbol{\mu}_1$) and sample size (n). The proportions of outliers as 0.1 or 0.2 and also the clean data set is taken under consideration. As for the POD a modification which is based on the shift in the mean vector $\boldsymbol{\mu}_1$ is a vector of size with value of 0 (in case of not any difference), 3 or 5 (in case of good leverage points) are considered. The setting values for the variables are listed in Table 1 following Alfaro and Ortega (2008), Vargas (2003) and Mohammadi et al. (2011). Changes on the mean shifts and proportions of outlying observations produce 5 dissimilar kinds of contaminated distributions stated as:

- $\mathrm{N}_p(0, I_p)$ – ideal case (clean data set);
- $(0.9)\mathrm{N}_p(0, I_p) + (0.1)\mathrm{N}_p(3, I_p)$ – slight contamination;
- $(0.8)\mathrm{N}_p(0, I_p) + (0.2)\mathrm{N}_p(3, I_p)$ – medium contamination;
- $(0.9)\mathrm{N}_p(0, I_p) + (0.1)\mathrm{N}_p(5, I_p)$ – medium contamination;
- $(0.8)\mathrm{N}_p(0, I_p) + (0.2)\mathrm{N}_p(5, I_p)$ – excessive contamination.

Later, in Phase II, the data are simulated from multivariate normal distribution $\mathrm{N}_p(\boldsymbol{\mu}_1, \boldsymbol{I}_p)$, where $\boldsymbol{\mu}_1$ shows the shift in the mean vector such as the case in Phase I (that is, 0, 3, and 5). Then, the new control chart T^2_{ARWMCD} is compared with robust Hotelling T^2 chart based on MCD (T^2_{MCD}) and the classical Hotelling T^2 control chart. For the classical chart T^2_0, the method, which is without cleaning the outlying observations as stated in Alfaro and Ortega (2009), is considered. The programs and simulations were done using MATLAB. The FAST-MCD algorithm code named as `mcdcov` could be found in MATLAB LIBRA Toolbox (Verboven and Hubert, 2005). The features of computer used for simulation is Intel(R) Core(TM) i5-8250U CPU @ 1.60 GHz 1.80 Ghz.

4.2 Simulation results

Here, the results of performance of the classical T_0^2 and robust T_{MCD}^2, T_{ARWMCD}^2 charts are presented in terms of FARs and POD at $\alpha = 0.05$ in Tables 2 and 3.

4.2.1 False alarm rates

The success of a chart cannot only be evaluated by its capability in diagnosing outliers, however, also in controlling the FAR, which is the probability of out-of-control signal in case of a process is under control. In case of the process instability, the value gets larger because of increment in variability. Expanded FAR could cause unrequired process regulations and loss of confidence in the control chart as an observing instrument (Chang and Bai, 2004). Therefore, a technique that could control the FAR to the wished level is essential. The Bradley liberal criterion of robustness is used for evaluating the robustness of the control charts. According to this criterion, a control chart is evaluated as robust in case of its empirical FAR is within the robust interval between 0.5α and 1.5α (Bradley, 1978). Henceforth, as the nominal value is accepted as $\alpha = 0.05$, the control chart is taken under consideration as robust if its FAR is within robust interval, 0.025 to 0.075. In Table 2, the FAR values lying within the robustness interval are bolded. A control chart, which is considered as best, the one has the ability of controlling the FARs within robust interval and also gives the closest FAR to nominal value, 0.05 (Jamaluddin et al., 2018). For every condition, the FARs given in Table 2 are presented in an ascending number of dimensions such as $p = 2, 5$ and 10, with $\alpha = 0.05$. The sample sizes are given in the first column of this table, in second column the proportions of outliers and in third column non-centrality values are provided.

Table 2. FAR values (%) of the three control charts in case of $\alpha = 0.05$.

			$p = 2$			$p = 5$			$p = 10$		
n	ε	$\boldsymbol{\mu}_1$	T_0^2	T_{MCD}^2	T_{ARWMCD}^2	T_0^2	T_{MCD}^2	T_{ARWMCD}^2	T_0^2	T_{MCD}^2	T_{ARWMCD}^2
50	0	0	**5**.6	**5**.8	**5**.5	**5**.5	**6**.0	**5**.5	**5**.7	**4**.9	**5**.2
	0.1	3	2.1	2.2	**3**.6	**2**.8	1.3	**3**.0	**4**.1	1.9	2.2
		5	1.6	2.3	3.8	**2**.6	1.4	3.1	**4**.0	1.8	2.0
	0.2	3	2.1	0.8	2.2	**2**.8	0.5	1.8	**4**.3	1.9	2.2
		5	1.9	0.6	**3**.1	**2**.7	0.2	1.6	**4**.2	0.8	1.1
100	0	0	**4**.7	**4**.5	**4**.2	**5**.0	**3**.3	**4**.2	**5**.1	**4**.3	**5**.3
	0.1	3	1.9	2.0	**3**.5	**2**.7	1.3	**3**.3	**3**.3	2.3	**4**.0
		5	1.5	2.0	3.8	**2**.7	1.4	3.2	**3**.3	2.2	**3**.8
	0.2	3	2.0	0.5	2.3	**2**.8	0.2	**3**.1	**3**.1	0.8	**3**.4
		5	1.5	0.4	**3**.3	**2**.7	0.2	**3**.1	**3**.2	0.5	**3**.2
200	0	0	**6**.3	**5**.7	**6**.0	**4**.2	**4**.1	**4**.4	**5**.0	**5**.6	**5**.5
	0.1	3	2.3	**3**.1	**4**.9	**2**.5	2.1	**3**.4	**3**.8	2.3	**4**.1
		5	1.9	**3**.1	**5**.6	2.4	2.0	**3**.4	**3**.5	2.1	**4**.0
	0.2	3	2.1	0.5	**3**.1	**2**.7	0.1	**3**.8	**3**.7	0.2	**3**.4
		5	1.9	0.2	**5**.2	**2**.7	0.1	**3**.9	**3**.6	0.2	**3**.3

Table 2 shows robust T_{MCD}^2 and T_{ARWMCD}^2 charts that perform as good as the traditional chart in controlling FAR under ideal condition ($\varepsilon = 0, \boldsymbol{\mu}_1 = 0$), regardless of the sample sizes n, outliers proportions ε and variable sizes p. However, the rates for all charts decrease when contamination exists with some results below the Bradley limit.

If $p = 2$, it is clear in Table 2 all results on FARs demonstrate that the T_{ARWMCD}^2 control chart has better performance than the T_0^2 and T_{MCD}^2 control charts. The T_{ARWMCD}^2 control chart has the capability in controlling FARs for nearly whole of the circumstances explored which is about 86% (13 out of 15) of the circumstances in comparison to T_0^2 and T_{MCD}^2 control charts, which are only effective for 20% (3 out of 15) and 33% (5 out of 15) of the circumstances, respectively. The T_{MCD}^2 control chart is affected bad with high proportion of outlying observations, $\varepsilon = 20\%$ for both moderate and high process mean shifts, that they are confirmed by the proportions of false alarm far below the significance value, $\alpha = 0.05$. Henceforth, T_{ARWMCD}^2 control chart performs superior to the traditional chart T_0^2 and robust control chart T_{MCD}^2 for bivariate case. In case of the dimensions raised to multivariate data, $p = 5$, the FARs for traditional chart (T_0^2) improved. In contrast, the FARs for T_{MCD}^2 chart

worsen with values as small as 0.001. In case of $p = 5$, the T^2_{ARWMCD} control chart is still maintains its good performance that it is still effective for 86 % (13 out of 15) of conditions as compared to the T^2_{MCD} control chart which is merely effective for 20 % (3 out of 15) of the conditions. The results of the FARs for the multivariate case of $p = 10$ shows that the T^2_{ARWMCD} control chart has still a good performance in controlling FARs as it is effective for 73% (11 out of 15) of conditions. Nevertheless, the performance of T^2_{ARWMCD} control chart diminishes in case of multivariate data in comparison with to bivariate data, where it capable in controlling FARs for only 11 simulated conditions as compared to 13 simulated conditions. An interesting result for T^2_0, it is effective 93% (14 out of 15) of conditions for $p = 5$ and 100% (all of 15) of conditions for $p = 10$. T^2_{MCD} control chart which is still merely effective for 20% (3 out of 15) of the conditions that means T^2_{MCD} chart performs badly in controlling FAR in all cases.

4.2.2 Probability detection of outliers

The performance in terms of POD is recorded in Table 3. The results are also presented as graphs for a better visual and comparison, the values in Table 3 are translated into Figures 1, 2 and 3 based on the values of p. For each case, the performance of the control chart is considered as better in detecting changes in case of the probabilities value is nearer to 1.

Table 3. Percentage of detecting outliers at $\alpha = 0.05$.

			$p = 2$			$p = 5$			$p = 10$		
n	ε	$\boldsymbol{\mu}_1$	T^2_0	T^2_{MCD}	T^2_{ARWMCD}	T^2_0	T^2_{MCD}	T^2_{ARWMCD}	T^2_0	T^2_{MCD}	T^2_{ARWMCD}
50	0	0	5.6	5.8	5.5	5.5	6	5.5	5.7	4.9	5.2
	0.1	3	49.8	88.2	91.7	37.8	98.6	99.8	25	100	100
		5	74.8	100	100	46.4	100	100	26.7	100	100
	0.2	3	17.6	69.3	75.4	12.5	95.1	98.1	11	55.2	58.7
		5	17.2	100	100	11.8	100	100	10.9	84.2	85.0
100	0	0	4.7	4.5	4.2	5	3.3	4.2	5.1	4.3	5.3
	0.1	3	49.6	90.3	92.3	38.8	99.8	100	28.1	100	100
		5	76.9	100	100	46.5	100	100	29.4	100	100
	0.2	3	16.4	72.8	73.0	10.5	98.6	99.9	10.2	84.4	84.9
		5	15.7	100	100	10.1	100	100	10.2	95	94.9
200	0	0	6.3	5.7	6.0	4.20	4.1	4.4	5	5.6	5.5
	0.1	3	57.1	93.5	94.6	45.5	100	100	36.7	100	100
		5	84.6	100	100	54.6	100	100	39.7	100	100
	0.2	3	20.1	80.3	79.7	12.3	99.5	100	11.6	98.6	98.6
		5	21.2	100	100	12.6	100	100	12.2	99.9	99.9

Once the values of p and n increased, that could be obviously seen in Figures 2 and 3, the line representing T^2_{ARWMCD} consistently at the top location in the plots with the probability value of almost 1 and overlapping with T^2_{MCD} line under most of the situations. Overall, the robust T^2_{ARWMCD} and T^2_{MCD} control charts steadily succeeded in high probability in diagnosing outlying observations. It is obvious that the line represents traditional T^2_0 charts is always at the lowest, producing a very large space between the other two lines (T^2_{ARWMCD} and T^2_{MCD}). Across Figures 1-3, it is observed that for all of the conditions, the robust charts outperform the traditional chart by a large difference. The robust T^2_{ARWMCD} chart under most conditions achieved the 100% detection with the lowest rate of 58.7% while the lowest rate for the robust T^2_{MCD} chart is 55.2% and for traditional chart is 10.1%. Across different dimensions (p), there is no clear pattern of changes in performance among the charts. Generally, the robust charts as well as traditional chart show decrease in POD when ε increases, however, especially in case of the shift is $\boldsymbol{\mu}_1 = 5$ and dimensions $p = 2$ or $p = 5$, POD values do not differ than the value of 100% for robust charts. The shift in mean ($\boldsymbol{\mu}_1$) shows positive effect on the POD performances of two robust charts regardless of the proportion of contamination (ε). However, for the traditional chart, positive effect only occurs when $\varepsilon = 0.1$. Moreover, the increase in sample sizes (n) brings some positive effect on the POD values for all the charts in some of the conditions.

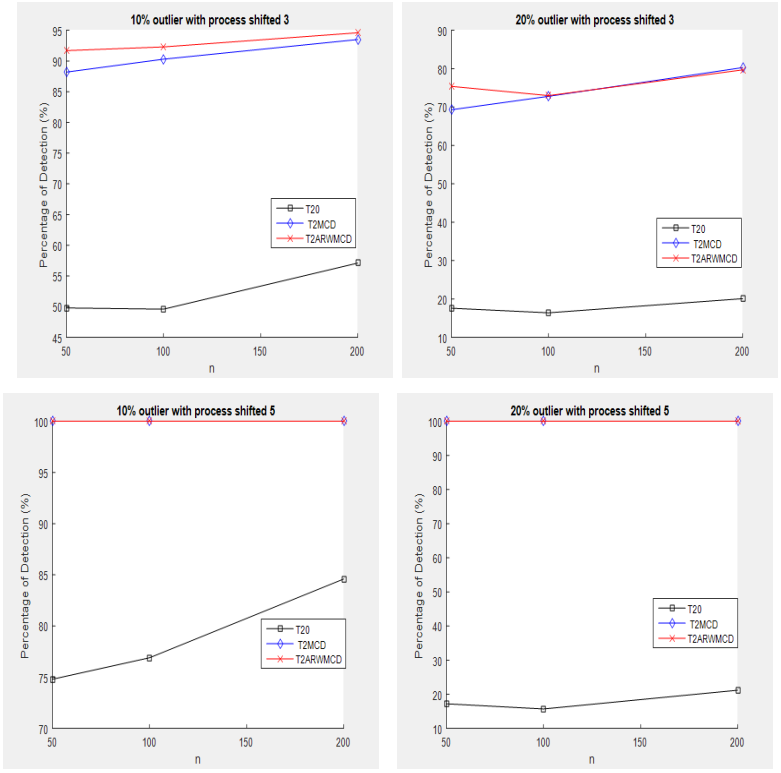

Figure 1. Percentages detection of outliers at $p = 2$.

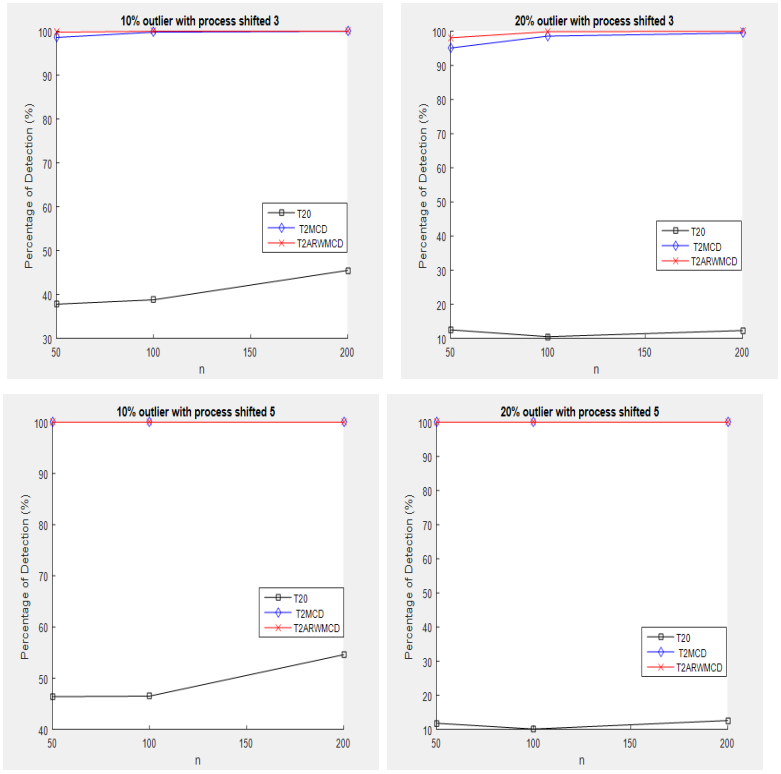

Figure 2. Percentages detection of outliers at $p = 5$.

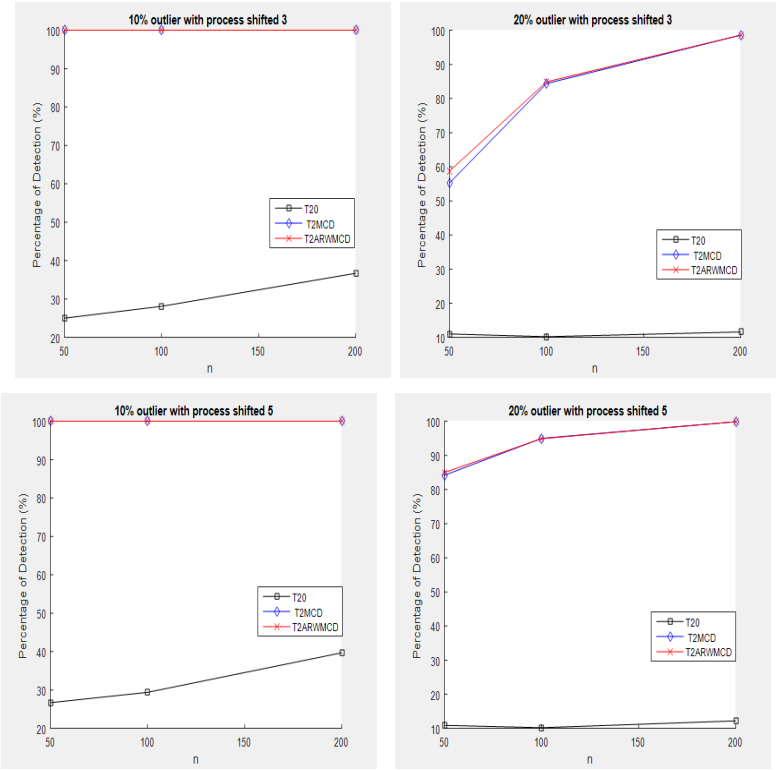

Figure 3. Percentages detection of outliers at $p = 10$.

5. Real data analysis

The proposed robust control chart T^2_{ARWMCD} is applied on real data given by Asian Composites Manufacturing Sdn. Bhd. (ACM) that includes in the production of advanced composite panels for the aircraft industry. ACM produces flat and contoured primary (Aileron Skins, Spoilers and Spars) and secondary (Flat Panels, Leading Edges and MISC: Components) structure composite bond assemblies and subassemblies for aerospace industries (Ali et al., 2013). For demonstrating the Hotelling T^2, the company that the part of the production of advanced for the aircraft industry has supplied the data on spoilers has shown in Table 4. The data set is used before in both Yahaya et al. (2011) and Ali et al. (2013) studies. Spoilers are critical instruments in an airplane which of them function is increasing lifts when the airplane is flying. The products are used in civilian, defense and space applications, which could not compromise any mistakes, albeit a minor one. Therefore, careful monitoring is needed to confirm that none of variation appears in the process. Any small error can risk a human life. A sample of 47 products ($n = 47$) that comprises of a few features like as trim edge (X_1), trim edge spar (X_2), and drill hole (X_3) was provided to Yahaya et al. (2011) by the firm. Note that 21 products were gathered in 2009, however, the rest had been gathered in 2010. Hence, they used the 2009 products data as Phase I historical data and they had taken under consideration the products from 2010 as future data. Hence, following these two studies, this data set is used in this study. The historical and future data are given in Tables 4 and 6, respectively. The products comprise of 3 quality variables (dimensions) as mentioned before known as trim edge, trim edge spar, and drill hole. The location vector ($\overline{x}$) and scatter matrix (S) estimations are given in Table 5. The calculations of the UCLs for $\alpha = 0.05$ based on the estimates are given in the last column of Table 5. The values of the T^2 statistics based on the classic, MCD and ARWMCD estimators are shown in the last three columns of Table 6. The graphical representations of the related control charts are shown in Figure 4.

Table 4. Historical data set (Phase I)

Product	Trim edge (X_1)	Trim edge spar (X_2)	Drill hole (X_3)
1	−0.0011	0.0003	0.0128
2	0.0011	0.0021	0.0246
3	0.0252	0.0308	0.0378
4	−0.0017	0.0109	0.0177
5	−0.0005	−0.0010	0.0106
6	0.0016	−0.0059	0.0128
7	0.0004	0.0001	0.0062
8	0.0078	0.0003	0.0159
9	0.0076	0.0089	0.0097
10	0.0020	0.0005	0.0071
11	0.0108	0.0011	0.0092
12	0.0039	0.0034	0.0425
13	0.0060	−0.0033	0.0160
14	0.0066	0.0100	0.0056
15	0.0045	−0.0067	0.0147
16	0.0110	−0.0207	0.0337
17	0.0047	0.0059	0.0065
18	0.0077	0.0003	0.0191
19	0.0015	0.0123	0.0124
20	0.0011	0.0038	0.0104
21	0.0056	0.0065	0.0063

Table 5. The location vector, covariance matrix and UCL estimations for real data

Types of control chart	Location vector ($\overline{\boldsymbol{x}}$)	Covariance matrix ($\boldsymbol{S}$)	UCL
T_0^2	[0.00504 0.00284 0.01579]	$\begin{bmatrix} 0.000040 & 0.000200 & 0.000030 \\ 0.000020 & 0.000090 & 0.000010 \\ 0.000030 & 0.000010 & 0.000110 \end{bmatrix}$	11.035
T^2_{MCD}	[0.00414 0.00207 0.01096]	$\begin{bmatrix} 0.000022 & 0.000005 & 0.000004 \\ 0.000005 & 0.000053 & -0.000019 \\ 0.000004 & -0.000019 & 0.000030 \end{bmatrix}$	12.5435
T^2_{ARWMCD}	[0.00414 0.00207 0.01096]	$\begin{bmatrix} 0.000012 & 0.000003 & 0.000002 \\ 0.000003 & 0.000028 & -0.000010 \\ 0.000002 & -0.000010 & 0.000016 \end{bmatrix}$	13.0304

Table 6. The Hotelling T^2 values for the future (Phase II) data

Product	Trim edge	Trim edge spar	Drill hole	T_0^2	T^2_{MCD}	T^2_{ARWMCD}
1	0.0041	0.0087	0.0129	0.5582	1.76591	3.32673
2	0.0047	0.0109	0.0124	0.90026	2.46944	4.65208
3	0.0031	0.0057	0.0096	0.49916	0.34367	0.64743
4	0.0035	-0.0020	0.0101	0.54633	0.54563	1.02789
5	0.004	-0.0028	0.0125	0.45922	0.45797	0.86276
6	0.0031	0.0008	0.0061	0.90130	1.25274	2.35998
7	-0.0019	0.0101	0.0112	3.09329	4.44043	8.36515
8	0.0009	0.0039	0.0082	0.80608	0.68370	1.28799
9	-0.0052	0.0090	0.0203	7.36021	14.97663	28.2139
10	-0.0008	0.0110	0.0184	3.61976	9.74168	18.3520
11	-0.0021	0.0139	0.0170	5.38392	11.87166	22.3645
12	-0.0017	0.0092	0.0061	2.73870	2.97882	5.61168
13	-0.0010	0.0133	0.0138	3.80577	7.40398	13.9481
14	-0.003	0.0002	0.0053	2.05480	3.30863	6.23300
15	0.0016	0.0134	0.0151	2.50731	6.80538	12.8204
16	0.0027	0.0086	0.0070	1.19755	1.06789	2.01176
17	0.0004	0.0086	0.0087	1.57979	1.75966	3.31495
18	-0.0036	0.0136	0.0129	5.79103	9.28168	17.4854
19	-0.0028	0.0003	0.0078	1.83044	2.41775	4.55471
20	0.0120	0.0123	0.0768	38.1397	214.923	404.885
21	-0.0015	0.0004	0.0115	1.26507	1.54862	2.9174
22	0.0009	0.0232	0.0202	8.41812	24.6552	46.4468
23	-0.0035	0.0088	0.0107	3.75884	4.87934	9.19198
24	0.0016	0.0061	0.0066	1.06020	0.93200	1.75576
25	-0.0228	-0.0466	0.0231	42.8447	68.63065	129.290
26	0.0037	-0.0038	0.0147	0.4831	0.77959	1.46863

The comparisons of T^2 values in Table 6 with the related control limits in Table 5, it is seen that T^2_{MCD} signals observations {9, 20, 22, 25} as out-of-control, the T^2_{ARWMCD} signals observations {9, 10, 11, 13, 18, 20, 22, 25} as out-of-control, however, T^2_0 only signals 20 and 25 as out-of-control observations and it cannot signal other observations. The result for T^2_0 is not surprising as the analysis on the POD for simulated data revealed that T^2_0 is not as effective as the other two robust charts in diagnosing outliers. For a clearer visualization on the performance of the control charts in diagnosing out or control observations, graphical representation of the three control charts are shown in Figure 4.

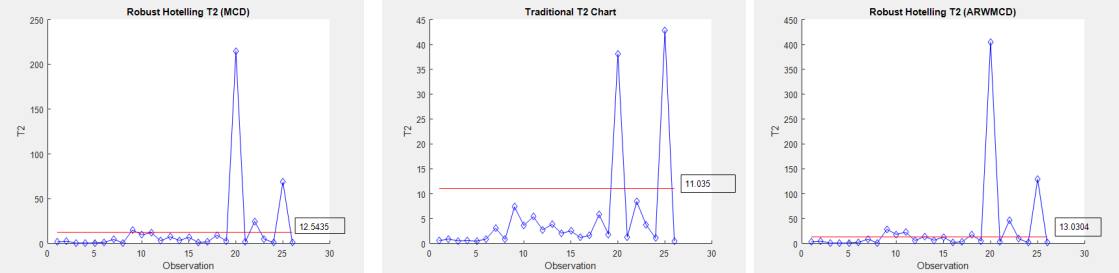

Figure 4. Hotelling T^2 control charts for real data.

6. Conclusion

In this study, an alternative to the classical Hotelling T^2 chart was proposed using a robust mean and covariance estimator called as adaptive reweighted minimum covariance determinant. The performance of proposed robust T^2_{ARWMCD} chart was compared with the robust Hotelling T^2 chart using minimum covariance determinant (T^2_{MCD}) and the classical Hotelling T^2 chart (T^2_0) in terms of false alarm rates and probability of detection.

Simulation results showed that both the robust Hotelling T^2 charts, T^2_{MCD} and T^2_{ARWMCD}, provided the best performances in term of probability of detection when $p = 2$, $p = 5$ or $p = 10$. In terms of false alarms, the best performance was detected by the robust T^2_{ARWMCD} chart when $p = 2$ and T^2_0 chart when $p = 5$ and $p = 10$. Furthermore, the T^2_{ARWMCD} chart was the second one for these dimensions. Alfaro and Ortega (2009) revealed a confusing result between the probability of detection and the overall false alarm rates such that, for both T^2_0 and T^2_{MCD} control charts when the probability of detection values increased, the false alarm rates inflated away from the nominal value. This situation was also observed in this study. Even though the classical T^2_0 control chart performed goodly in terms of false alarm rates, particularly when the number of dimensions is getting larger. However, it fails to achieve good probability of detection. In contrast to the T^2_0 chart, the robust Hotelling T^2_{MCD} control chart performed perfectly in diagnosing outliers, despite that it fails badly in controlling false alarm rates. Nevertheless, the proposed T^2_{ARWMCD} chart performed so well both in terms of diagnosing outliers and in controlling false alarm rates.

Real data analysis results showed that the proposed robust T^2_{ARWMCD} control chart showed best performance in terms of diagnosing outliers and the T^2_{MCD} control chart was the second one. Nonetheless, the classical T^2_0 control chart failed to detect most of the outliers. The real data application results showed consistency with the simulation results.

The overall findings reported that the performance of the robust T^2_{ARWMCD} control chart in controlling false alarm rates was very good. However, the robust T^2_{MCD} control charts performance in terms of controlling false alarm rates was not good. Nevertheless, both of these two robust charts were superior to the classical chart in detecting outliers regardless of the conditions imposed in this study. The traditional chart T^2_0 performed moderately in

lower dimension, but better in higher dimensions in controlling false alarm. In contrast, it reported inability to detect outliers. Overall, the proposed T^2_{ARWMCD} control chart showed the best performance since this control chart produced good values for both false alarm rates and probability of detection.

Author contributions Conceptualization, investigation, methodology, software, writing-original draft preparation, writing review and editing, writing-original: E.P. The author has read and agreed to the published version of the manuscript.

Acknowledgements The author thanks Editors and Referees for this suggestions which allowed us to improve the presentation of this work.

Funding Not applicable

Conflicts of interest The authors declare no conflict of interest.

References

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

Ahsan, M., Mashuri, M., Kuswanto, H., Prastyo, D.D., and Khusna, H., 2019. Multivariate T^2 control chart based on James-Stein and Successive Difference Covariance Matrix Estimators for intrusion detection. Malaysian Journal of Science, 38, 23–35.

Alfaro, J.L. and Ortega, J.F., 2008. A robust alternative to Hotelling's T^2 control chart using trimmed estimators. Quality and Reliability Engineering International, 24, 601–611.

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.

Ali, H., Yahaya, S.S., and Omar, Z., 2013. Robust Hotelling T^2 control chart with consistent minimum vector variance. Mathematical Problems in Engineering, 2013, 401350.

Ali, H., Yahaya, S.S., and Omar, Z., 2014. Robust Hotelling T^2 control chart using reweighted minimum vector variance estimators. AIP Conference Proceedings, 1635, 695–702 https://doi.org/10.1063/1.4903658.

Alt, F.B., 1985. Multivariate quality control. In Kotz, S. and Johnson, N. (eds.), Encyclopedia of Statistical Sciences Vol. 6, pp. 110–122, Wiley, New York.

Bradley, J.V., 1978. Robustness? British Journal of Mathematical and Statistical Psychology, 31, 144–152.

Chang, Y.S. and Bai, D.S., 2004. A multivariate T^2 control chart for skewed populations using weighted standard deviations. Quality and Reliability Engineering International, 20, 31–46.

Chenouri, S., Steiner, S.H., and Mulayath, A., 2009. A multivariate robust control chart for individual observations. Journal of Quality Technology, 41(3), 259-271.

Da Silva, G.V., Taconeli, C.A., Zeviani, W. M., and Do Nascimento, I.A.S., 2019. Performance of Shewhart control charts based on neoteric ranked set sampling to monitor the process mean for normal and non-normal processes. Chilean Journal of Statistics, 10(2), 131-154.

Filzmoser, P., Serneels, S., Maronna, R., and Van Espen, P.J., 2009. Robust multivariate methods In Chemometrics. In Walczak, B., Ferre, R.T. and Brown, S. (eds.), Comprehensive Chemometrics, pp. 681-722. Elsevier, New York, US.

Gervini, D. and Yohai, V.J., 2002. A class of robust and fully efficient regression estimators. Annals of Statistics, 30, 583–616.

Gervini, D., 2003. A robust and efficient adaptive reweighted estimator of multivariate location and scatter. Journal of Multivariate Analysis, 84, 116–144.

Hubert, M. and Vanden-Branden, K., 2003. Robust methods for partial least squares regression. Journal of Chemometrics, 17, 537–549.

Jamaluddin, F., Ali, H.H., and Yahaya, S.S., 2018. The performance of robust multivariate EWMA control charts. The Journal of Social Sciences Research, 6, 52–58.

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

Mohammadi M., Midi H., Arasan J., and Al-Talib B., 2011. High breakdown estimators to robustify phase II control charts. Applied Sciences, 11, 503–511.

Polat, E. and Gunay, S., 2019. A new robust partial least squares regression method based on a robust and an efficient adaptive reweighted estimator of covariance. REVSTAT, 17(4), 449–474.

Rousseeuw, P.J. and Van Zomeren, B.C., 1990. Unmasking multivariate outliers and leverage points. Journal of the American Statistical Association, 85(411), 633–639.

Rousseeuw, P.J. and Van Driessen, K., 1999. A fast algorithm for the minimum covariance determinant estimator. Technometrics, 41, 212–224.

Vargas, J.A., 2003. Robust estimation in multivariate control charts for individual observations. Journal of Quality Technology, 35(4), 367–376.

Verboven, S. and Hubert, M., 2005. LIBRA: a MATLAB library for robust analysis. Chemometrics and Intelligent Laboratory Systems, 75, 127–136.

Williams, J.D., Woodall, W.H., Birch, J.B., Sullivan, J. H., 2006. Distribution of Hotelling's T^2 statistic based on the successive differences estimator. Journal of Quality Technology, 38(3), 217-229.

Yahaya, S.S., Ali, H., and Omar, Z., 2011. An alternative hotelling T^2 control chart based on minimum vector variance (MVV). Modern Applied Science, 5(4), 132–151.

Yahaya, S.S., Haddad, F.S., Mahat, N.I., Rahman, A., and Ali, H., 2019. Robust hotelling T^2 charts with median based trimmed estimators. Journal of Engineering and Applied Sciences, 14(24), 9632–9638.

Yañez, S., Gonzalez, N., and Vargas, J.A., 2010. Hotelling's T^2 control charts based on robust estimators. Dyna, 77(163), 239-247.

Information for authors

The editorial board of the Chilean Journal of Statistics (ChJS) is seeking papers, which will be refereed. We encourage the authors to submit a PDF electronic version of the manuscript in a free format to the Editors-in-Chief of the ChJS (E-mail: `chilean.journal.of.statistics@gmail.com`). Submitted manuscripts must be written in English and contain the name and affiliation of each author followed by a leading abstract and keywords. The authors must include a"cover letter" presenting their manuscript and mentioning: "We confirm that this manuscript has been read and approved by all named authors. In addition, we declare that the manuscript is original and it is not being published or submitted for publication elsewhere".

Preparation of accepted manuscripts

Manuscripts accepted in the ChJS must be prepared in Latex using the ChJS format. The Latex template and ChJS class files for preparation of accepted manuscripts are available at `http://soche.cl/chjs/files/ChJS.zip`. Such as its submitted version, manuscripts accepted in the ChJS must be written in English and contain the name and affiliation of each author, followed by a leading abstract and keywords, but now mathematics subject classification (primary and secondary) are required. AMS classification is available at `http://www.ams.org/mathscinet/msc/`. Sections must be numbered 1, 2, etc., where Section 1 is the introduction part. References must be collected at the end of the manuscript in alphabetical order as in the following examples:

Arellano-Valle, R., 1994. Elliptical Distributions: Properties, Inference and Applications in Regression Models. Unpublished Ph.D. Thesis. Department of Statistics, University of São Paulo, Brazil.

Cook, R.D., 1997. Local influence. In Kotz, S., Read, C.B., and Banks, D.L. (Eds.), Encyclopedia of Statistical Sciences, Vol. 1., Wiley, New York, pp. 380-385.

Rukhin, A.L., 2009. Identities for negative moments of quadratic forms in normal variables. Statistics and Probability Letters, 79, 1004-1007.

Stein, M.L., 1999. Statistical Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York.

Tsay, R.S., Peña, D., and Pankratz, A.E., 2000. Outliers in multivariate time series. Biometrika, 87, 789-804.

References in the text must be given by the author's name and year of publication, e.g., Gelfand and Smith (1990). In the case of more than two authors, the citation must be written as Tsay et al. (2000).

Copyright