

Research Article

Robust Control Charts for Monitoring Process Mean of Phase-I Multivariate Individual Observations

Asokan Mulayath Variyath and Jayasankar Vattathoor

Department of Mathematics and Statistics, Memorial University of Newfoundland, St. John's, NL, Canada A1C 5S7

Correspondence should be addressed to Asokan Mulayath Variyath; variath@mun.ca

Received 22 November 2012; Revised 22 March 2013; Accepted 2 April 2013

Academic Editor: Adiel Teixeira de Almeida

Copyright © 2013 A. M. Variyath and J. Vattathoor. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Hotelling's T^2 control charts are widely used in industries to monitor multivariate processes. The classical estimators, sample mean, and the sample covariance used in T^2 control charts are highly sensitive to the outliers in the data. In Phase-I monitoring, control limits are arrived at using historical data after identifying and removing the multivariate outliers. We propose Hotelling's T^2 control charts with high-breakdown robust estimators based on the reweighted minimum covariance determinant (RMCD) and the reweighted minimum volume ellipsoid (RMVE) to monitor multivariate observations in Phase-I data. We assessed the performance of these robust control charts based on a large number of Monte Carlo simulations by considering different data scenarios and found that the proposed control charts have better performance compared to existing methods.

1. Introduction

Control charts are widely used in industries to monitor/control processes. Generally, the construction of a control chart is carried out in two phases. The Phase-I data is analyzed to determine whether the data indicates a stable (or in-control) process and to estimate the process parameters and thereby the construction of control limits. The Phase-II data analysis consists of monitoring future observations based on control limits derived from the Phase-I estimates to determine whether the process continues to be in control or not. But trends, step changes, outliers, and other unusual data points in the Phase-I data can have an adverse effect on the estimation of parameters and the resulting control limits. That is, any deviation from the main assumption (in our case, identically and independently distributed from normal distribution) may lead to an out-of-control situation. Therefore, it becomes very important to identify and eliminate these data points prior to calculating the control limits. In this paper, all these unusual data points are referred to as "outliers."

Multivariate quality characteristics are often correlated, and to monitor the multivariate process mean Hotelling's T^2 control chart [1, 2] is widely used. To implement Hotelling's T^2

control chart for individual observations in Phase-I, for each observation \mathbf{x}_j we calculate

$$T^2(\mathbf{x}_j) = (\mathbf{x}_j - \bar{\mathbf{x}})' \mathbf{S}_1^{-1} (\mathbf{x}_j - \bar{\mathbf{x}}), \quad (1)$$

where $\mathbf{x}_j = (x_{j1}, x_{j2}, \dots, x_{jp})'$ is the j th p -variate observation, ($j = 1, 2, \dots, m$) and the sample mean $\bar{\mathbf{x}}$, sample covariance matrix \mathbf{S}_1 are based on m Phase-I observations. In Phase-I monitoring, the $T^2(\mathbf{x}_j)$ values are compared with the T^2 control limit derived by assuming that the \mathbf{x}_j 's are multivariate normal so that the T^2 control limits are based on the beta distribution with the parameters $p/2$ and $(m - p - 1)/2$. However, the classical estimators, sample mean, and sample covariance are highly sensitive to the outliers, and hence robust estimation methods are preferred as they have the advantage of not being unduly influenced by the outliers. The use of robust estimation methods is well suited to detect multivariate outliers because of their high breakdown points which ensure that the control limits are reasonably accurate. Sullivan and Woodall [3] proposed a T^2 chart with an estimate of the covariance matrix based on the successive differences of observations and showed that it is effective in

detecting process shift. However, these charts are not effective in detecting multiple multivariate outliers because of their low breakdown point.

Vargas [4] introduced two robust T^2 control charts based on robust estimators of location and scatter, namely, the minimum covariance determinant (MCD) and minimum volume ellipsoid (MVE) for identifying the outliers in Phase-I multivariate individual observations. Jensen et al. [5] showed that T^2_{MCD} and T^2_{MVE} control charts have better performance when outliers are present in the Phase-I data. Chenouri et al. [6] used reweighted MCD estimators for monitoring the Phase-II data, without constructing Phase-I control charts. However, in many situations Phase-I control charts are necessary to assess the performance of the process and also to identify the outliers. We propose T^2 control charts based on the reweighted minimum covariance determinant (RMCD)/reweighted minimum volume ellipsoid (RMVE) ($T^2_{\text{RMCD}}/T^2_{\text{RMVE}}$) for monitoring Phase-I multivariate individual observations. RMCD/RMVE estimators are statistically more efficient than MCD/MVE estimators and have a manageable asymptotic distribution. We empirically arrive at Phase-I control limits for the $T^2_{\text{RMCD}}/T^2_{\text{RMVE}}$ control chart for some specific sample sizes and fitted a nonlinear model to determine control limits for any sample size for dimensions 2 to 10. Our simulation studies show that $T^2_{\text{RMCD}}/T^2_{\text{RMVE}}$ control charts are performing well compared to $T^2_{\text{MCD}}/T^2_{\text{MVE}}$ control charts for monitoring the Phase-I data.

The organization of the remaining part of the paper is as follows. In Section 2, we discuss the properties of a good robust estimator and we briefly explain the MCD/MVE estimators and their reweighted versions. The proposed $T^2_{\text{RMCD}}/T^2_{\text{RMVE}}$ control charts are given in Section 3 along with the control limits arrived at based on Monte Carlo simulations. We assess the performance of the proposed control charts in Section 4, and the implementation of the proposed methods is illustrated in a case example in Section 5. Our conclusions are given in Section 6.

2. Robust Estimators

The affine equivariance property of the estimator is important because it makes the analysis independent of the measurement scale of the variables as well as the transformations or rotations of the data. The breakdown point concept introduced by Donoho and Huber [7] is often used to assess the robustness. The breakdown point is the smallest proportion of the observations which can render an estimator meaningless. A higher breakdown point implies a more robust estimator, and the highest attainable breakdown point is $1/2$ in the case of median in the univariate case. For more details on affine equivariance and breakdown points one may refer to Chenouri et al. [6] or Jensen et al. [5].

An estimator is said to be relatively efficient compared to any other estimator if the mean square error for the estimator is the least for at least some values of the parameter compared to others. A robust estimator is considered to be good if it carries the property of affine equivariance along with a higher breakdown point and greater efficiency. In addition to the

above three properties of a good robust estimator, it should be possible to calculate the estimator in a reasonable amount of time to make it computationally efficient.

It is difficult to get an affine equivariant and robust estimator as affine equivariance and high breakdown will not come simultaneously. Lopuhaä and Rousseeuw [8] and Donoho and Gasko [9] showed that the finite sample breakdown point of $(m - p + 1)/(2m - p + 1)$ is difficult for an affine equivariant estimator. The largest attainable finite sample breakdown point of any affine equivariant estimator of the location and scatter matrix with a sample size m and dimension p is $(m - p + 1)/2m$ [10]. Therefore relaxing the affine equivariance condition of the estimators to invariance under the orthogonal transformation makes it easy to find an estimator with the highest breakdown point.

The classical estimators, sample mean vector, and covariance matrix of location and scatter parameters are affine equivariant but their sample breakdown point is as low as $1/m$. The MCD and MVE estimators have the highest possible finite sample breakdown point $(m - p + 1)/2m$. However, both of these estimators have very low asymptotic efficiency under normality. But the reweighted versions of MCD and MVE estimators have better efficiency without compromising on the breakdown point and rate of convergence compared to MCD and MVE. In the next two subsections, we discuss in detail about the MCD and MVE estimators and their reweighted versions.

2.1. MCD and RMCD Estimators. The MCD estimators of location and scatter parameters of the distribution are determined by a two-step procedure. In step 1, all possible subsets of observations of size $h = (m * \gamma)$, where $0.5 \leq \gamma \leq 1$ are obtained. In step 2, the subset whose covariance matrix has the smallest possible determinant is selected. The MCD location estimator \bar{x}_{MCD} is defined as the average of this selected subset of h points, and the MCD scatter estimator is given by $S_{\text{MCD}} = a_{\gamma,p} * b_{\gamma,p}^m * C_{\text{MCD}}$, where C_{MCD} is the covariance matrix of the selected subset, the constant $a_{\gamma,p}$ is the multiplication factor for consistency [11], and $b_{\gamma,p}^m$ is the finite sample correction factor [12]. Here $(1 - \gamma)$ represents the breakdown point of the MCD estimators. The MCD estimator has its highest possible finite sample breakdown point when $h = (m + p + 1)/2$ and has an $m^{-1/2}$ rate of convergence but has a very low asymptotic efficiency under normality. Computing the exact MCD estimators (\bar{x}_{MCD} , S_{MCD}) is computationally expensive or even impossible for large sample sizes in high dimensions [13], and hence various algorithms have been suggested for approximating the MCD. Hawkins and Olive [14] and Rousseeuw and van Driessen [15] independently proposed a fast algorithm for approximating MCD. The FAST-MCD algorithm of Rousseeuw and van Driessen finds the exact MCD for small datasets and gives a good approximation for larger datasets, which is available in the standard statistical software SPLUS, R, SAS, and Matlab.

MCD estimators are highly robust, carry equivariance properties, and can be calculated in a reasonable time using the FAST-MCD algorithm; however, they are statistically not efficient. The reweighted procedure will help to carry both

TABLE 1: Estimates of the model parameters $a_{1(p,\alpha)}$, $a_{2(p,\alpha)}$, $a_{3(p,\alpha)}$ for $T_{\text{RMCD}}^2/T_{\text{RMVE}}^2$ control charts.

p	$\alpha = 0.05$			$\alpha = 0.01$			$\alpha = 0.001$		
	\hat{a}_1	\hat{a}_2	\hat{a}_3	\hat{a}_1	\hat{a}_2	\hat{a}_3	\hat{a}_1	\hat{a}_2	\hat{a}_3
T_{RMCD}^2									
2	17.223	41102	2.647	21.134	38170	2.329	27.051	192909	2.508
3	20.134	35844	2.209	24.287	128924	2.344	31.350	1144947	2.718
4	23.152	269357	2.548	28.181	1272773	2.773	35.575	5989325	2.973
5	24.685	467949	2.524	28.437	1417059	2.632	31.013	2666196	2.593
6	26.962	1762051	2.746	29.654	3061216	2.711	31.662	5414248	2.669
7	24.892	1099128	2.493	22.882	1585224	2.416	19.058	3465278	2.444
8	27.236	2908821	2.667	27.245	4922576	2.644	28.326	12134778	2.710
9	23.974	2447649	2.534	21.420	4726835	2.554	18.772	14096595	2.676
10	31.894	12572909	2.914	37.085	34375654	3.033	56.573	172176786	3.301
T_{RMVE}^2									
2	17.442	29553	2.494	21.365	31571	2.244	27.594	148747	2.434
3	20.286	22497	2.066	24.387	59096	2.13	31.326	338665	2.402
4	23.095	108855	2.286	27.549	291064	2.372	35.109	1255429	2.576
5	24.796	238966	2.334	28.302	508097	2.367	32.008	1063783	2.377
6	27.585	1041090	2.606	31.126	1882888	2.601	37.136	4714353	2.671
7	28.151	1541634	2.598	30.936	3183762	2.635	39.357	12199414	2.827
8	34.917	14798692	3.127	45.767	75616029	3.419	70.875	840512379	3.904
9	39.191	59094377	3.415	50.271	275604839	3.679	72.768	1960966919	4.039
10	50.733	950607720	4.099	68.154	4696452032	4.379	110.587	56398461817	4.881

robustness and efficiency. That is, first a highly robust but perhaps an inefficient estimator is computed, which is used as a starting point to find a local solution for detecting outliers and computing the sample mean and covariance of the cleaned data set as in Rousseeuw and van Zomeren [16]. This consists of discarding those observations whose Mahalanobis distances exceed a certain fixed threshold value. MCD is the current best choice for the initial estimator of a two-step procedure as it contains the robustness, equivariance, and computational efficiency properties along with its $m^{-1/2}$ rate of convergence. Hence RMCD estimators are the weighted mean vector

$$\bar{\mathbf{x}}_{\text{RMCD}} = \frac{\left(\sum_{j=1}^m w_j \mathbf{x}_j\right)}{\left(\sum_{j=1}^m w_j\right)}, \quad (2)$$

and the weighted covariance matrix

$$\mathbf{S}_{\text{RMCD}} = c_{\alpha,p} * d_{\gamma,\alpha}^{m,p} * \frac{\sum_{j=1}^m w_j (\mathbf{x}_j - \bar{\mathbf{x}}_{\text{RMCD}})(\mathbf{x}_j - \bar{\mathbf{x}}_{\text{RMCD}})'}{\sum_{j=1}^m w_j}, \quad (3)$$

where $c_{\alpha,p}$ is the multiplication factors for consistency [11], $d_{\gamma,\alpha}^{m,p}$ is the finite sample correction factor [12], and the weights w_j are defined as

$$w_j = \begin{cases} 1 & \text{if } \text{RD}(\mathbf{x}_j) \leq \sqrt{q_\alpha}, \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

where the robust distance $\text{RD}(\mathbf{x}_j) = \sqrt{(\mathbf{x}_j - \bar{\mathbf{x}}_{\text{MCD}})' \mathbf{S}_{\text{MCD}}^{-1} (\mathbf{x}_j - \bar{\mathbf{x}}_{\text{MCD}})}$ and q_α is $(1 - \alpha)100\%$ quantile of the chi-square distribution with p degrees of freedom.

This reweighting technique improves the efficiency of the initial MCD estimator while retaining (most of) its robustness. Hence the RMCD estimator inherits the affine equivariance, robustness, and asymptotic normality properties of the MCD estimators with an improved efficiency.

2.2. MVE and RMVE Estimators. Determining the MVE estimators of location and scatter parameters of the distribution is almost in line with that of the MCD estimator. As in the case of MCD, all the possible subsets of data points with size $h = (m * \gamma)$ (where $0.5 \leq \gamma \leq 1$) is obtained first. Then the ellipsoid of minimum volume that covers the subsets are obtained to determine the MVE estimators. The MVE location estimator is the geometrical center of the ellipsoid, and the MVE scatter estimator is the matrix defining the ellipsoid itself, multiplied by an appropriate constant to ensure consistency [13, 16]. Thus MVE estimator does not correspond to the sample mean vector and the sample covariance matrix as in the case of the MCD estimator. Here $(1 - \gamma)$ represents the breakdown point of the MVE estimators, as in the case of MCD, and it has the highest possible finite sample breakdown point when $h = (m + p + 1)/2m$ [8, 17]. The MVE estimator has an $m^{-1/3}$ rate of convergence and a nonnormal asymptotic distribution [17].

As in the case for MCD estimators, MVE estimators are also not efficient. Hence, a reweighted version similar to that

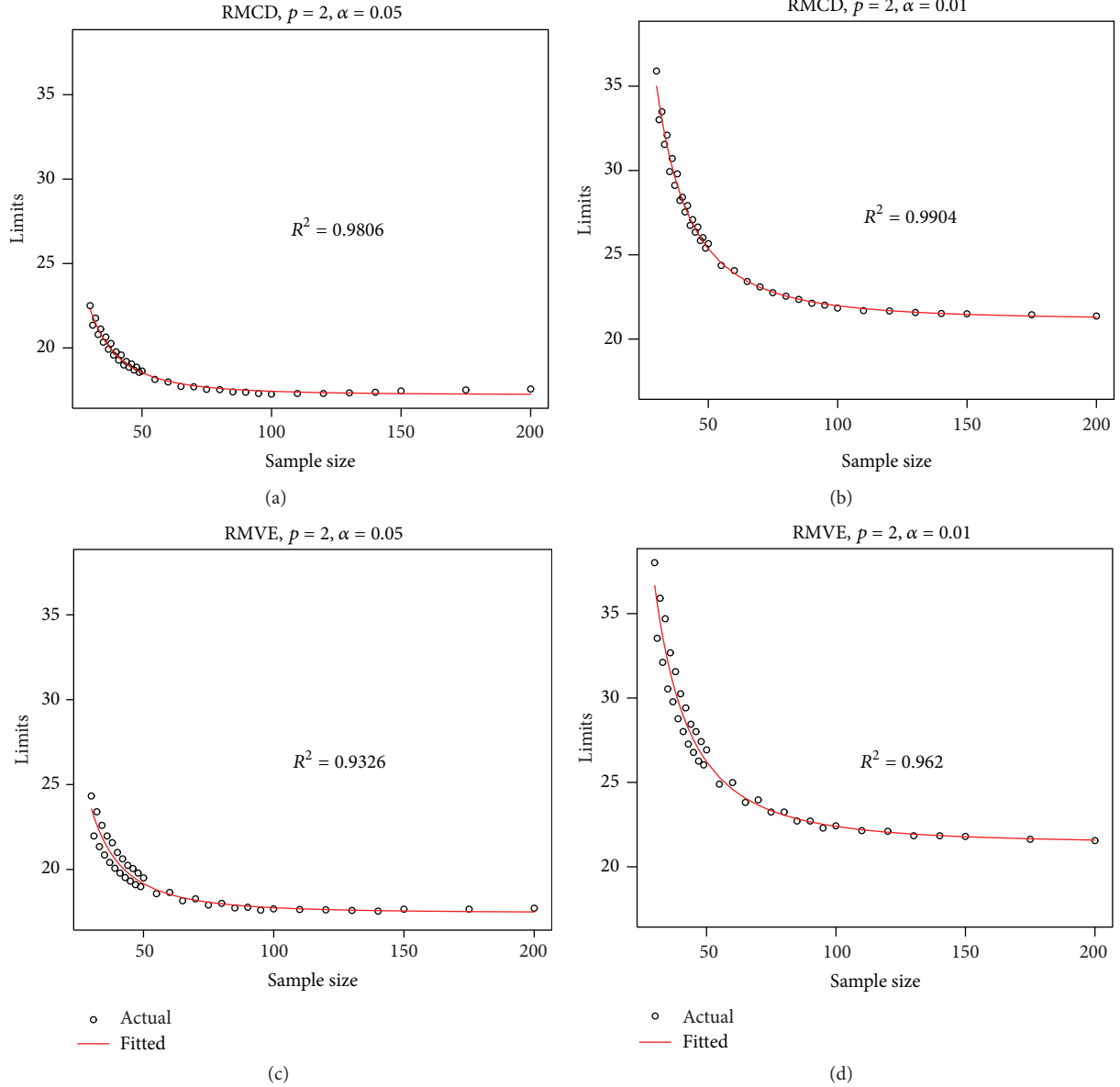


FIGURE 1: Scatter plot of $T_{\text{RMCD}}^2/T_{\text{RMVE}}^2$ control limits and the fitted curve for $p = 2$.

for MCD has been proposed by Rousseeuw and van Zomeren [16]. Note that it has been shown more recently that the RMVE estimators do not improve on the convergence rate (and thus the 0% asymptotic efficiency) of the initial MVE estimator [8, 12]. Therefore, as an alternative, a one-step M-estimator can be calculated with the MVE estimators as the initial solution [13, 18] which results in an estimator with the standard $m^{-1/2}$ convergence rate to a normal asymptotic distribution. For more details on MCD/MVE estimators one may refer to Chenouri et al. [6] or Jensen et al. [5]. The algorithm to determine the MVE/RMVE estimators is available in the statistical software SPLUS, R, SAS, and Matlab.

3. Robust Control Charts

We propose to use T^2 charts with robust estimators of location and dispersion parameters based on RMCD/RMVE

for monitoring the process mean of Phase-I multivariate individual observations. RMCD/RMVE estimators inherit the nice properties of initial MCD estimators such as affine equivariance, robustness, and asymptotic normality while achieving a higher efficiency. We now define a robust T^2 control chart with RMCD and RMVE estimators for i th multivariate observation as

$$\begin{aligned} T_{\text{RMCD}}^2(\mathbf{x}_i) &= (\mathbf{x}_i - \bar{\mathbf{x}}_{\text{RMCD}})' \mathbf{S}_{\text{RMCD}}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}_{\text{RMCD}}), \\ T_{\text{RMVE}}^2(\mathbf{x}_i) &= (\mathbf{x}_i - \bar{\mathbf{x}}_{\text{RMVE}})' \mathbf{S}_{\text{RMVE}}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}_{\text{RMVE}}), \end{aligned} \quad (5)$$

where $\bar{\mathbf{x}}_{\text{RMCD}}$, $\bar{\mathbf{x}}_{\text{RMVE}}$ are the mean vectors and \mathbf{S}_{RMCD} , \mathbf{S}_{RMVE} are the dispersion matrices under the RMCD/RMVE methods based on m multivariate observations.

The exact distribution of $T_{\text{RMCD}}^2/T_{\text{RMVE}}^2$ estimators not available, hence the control limits for Phase-I data are

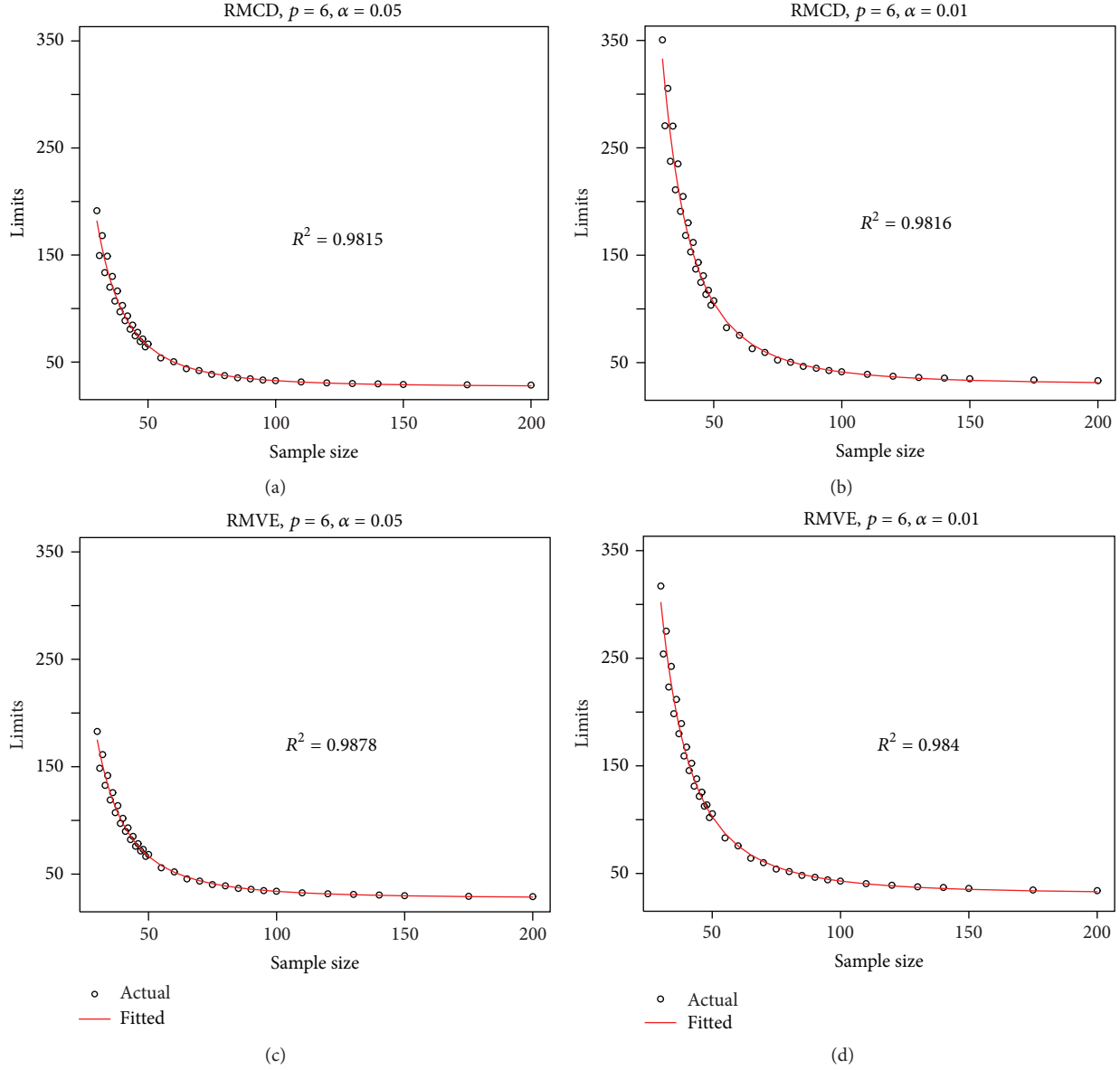


FIGURE 2: Scatter plot of $T_{\text{RMCD}}^2/T_{\text{RMVE}}^2$ control limits and the fitted curve for $p = 6$.

obtained empirically. In the next subsection we apply Monte Carlo simulation to estimate quantiles of the distribution of T_{RMCD}^2 and T_{RMVE}^2 for several combinations of sample sizes and dimensions. For each dimension, we further introduce a method to fit a smooth nonlinear model to arrive, the control limits for any given sample size.

3.1. Computation of Control Limits. We performed a large number of Monte Carlo simulations to obtain the control limits. We generated $n = 200,000$ samples of size m from a standard multivariate normal distribution $\text{MVN}(0, I_p)$ with dimension p . Due to the invariance of the T_{RMCD}^2 and T_{RMVE}^2 statistics, these limits will be applicable for any values of μ and Σ . Using the reweighted MCD/MVE estimators $\bar{\mathbf{x}}_{\text{RMCD}}$, \mathbf{S}_{RMCD} , $\bar{\mathbf{x}}_{\text{RMVE}}$, and \mathbf{S}_{RMVE} with a breakdown value

of $\gamma = 0.50$, $T_{\text{RMCD}}^2/T_{\text{RMVE}}^2$ statistics for each observation in the data set were calculated using (5), and the maximum value attained for each data set of size m was recorded. The empirical distribution of maximum of T_{RMCD}^2 and T_{RMVE}^2 was inverted to determine the $(1 - \alpha)100\%$ quantiles. We used the R-function “CovMcd()” in the “rrcov” package written by Torodov [19] to ascertain the RMCD/RMVE estimators.

We have constructed the empirical distribution of $T_{\text{RMCD}}^2/T_{\text{RMVE}}^2$ as above for $m = [30(1)50, 55(5)100, 110(10)200]$, $p = (2, 3, \dots, 10)$ when $\gamma = 0.50$ and arrived at the control limits for $\alpha = (0.05, 0.01, \text{ and } 0.001)$. The scatter plots of the quantiles and sample sizes for different dimensions suggest a family of nonlinear models of the form

$$f_{p,\alpha,\gamma,m} = a_{1,(p,\alpha,\gamma)} + \frac{a_{2,(p,\alpha,\gamma)}}{m^{a_{3,(p,\alpha,\gamma)}}}, \quad (6)$$

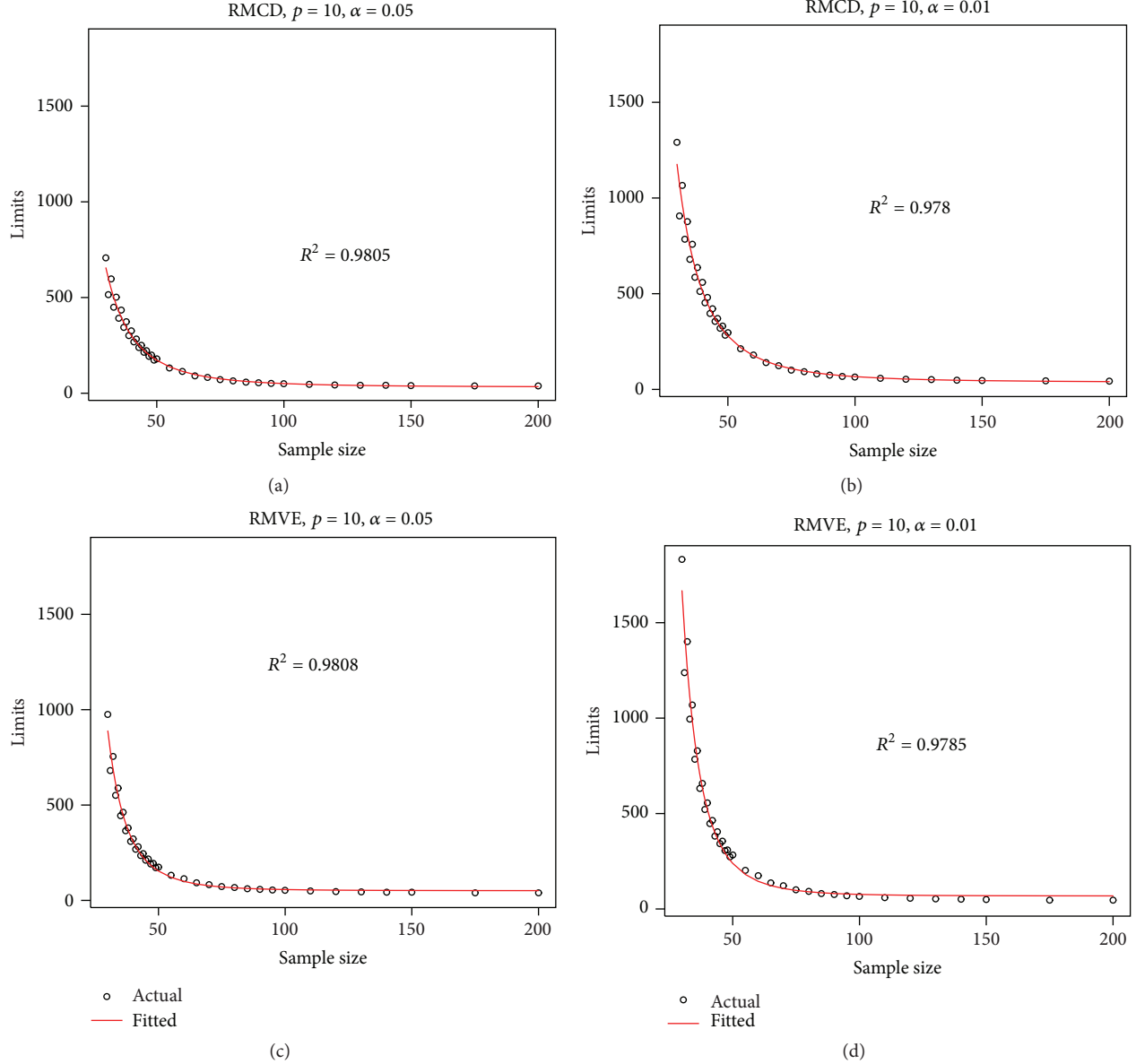


FIGURE 3: Scatter plot of $T_{\text{RMCD}}^2/T_{\text{RMVE}}^2$ control limits and the fitted curve for $p = 10$.

where $a_{1(p,\alpha,\gamma)}$, $a_{2(p,\alpha,\gamma)}$, and $a_{3(p,\alpha,\gamma)}$ are the model parameters. For clarity, the scatter plot of the actual and the fitted values of the quantiles of T_{RMCD}^2 and T_{RMVE}^2 for $p = 2, 6$, and 10 are given in Figures 1, 2, and 3; other plots are omitted to save space.

From Figures 1, 2, and 3, we can see that the nonlinear fit is very well supported by the high R^2 values, which help us to determine the T_{RMCD}^2 and T_{RMVE}^2 control limits for any given sample size. The least square estimates of the parameters $a_{1(p,\alpha)}$, $a_{2(p,\alpha)}$, and $a_{3(p,\alpha)}$ when $\gamma = 0.50$ for dimensions $p = (2, 3, \dots, 10)$ and $\alpha = (0.05, 0.01 \text{ and } 0.001)$ for $T_{\text{RMCD}}^2/T_{\text{RMCD}}^2$ control charts are given in Table 1. Using these estimates, the

control limits for T_{RMCD}^2 and T_{RMVE}^2 can be found using (6) for any sample size.

For the implementation of a robust control chart, first collect a sample of m multivariate individual observations with dimension p . Compute robust estimates of mean and covariance matrix using R or any other software with $\gamma = 0.50$, and determine $T_{\text{RMCD}}^2/T_{\text{RMVE}}^2$. Outliers can be determined by comparing the $T_{\text{RMCD}}^2/T_{\text{RMVE}}^2$ values with control limits obtained using (6) for specific values of α, m, p , and the constants given in Table 1. The outlier free data can be used to construct the standard T^2 control chart for monitoring the Phase-II observations.

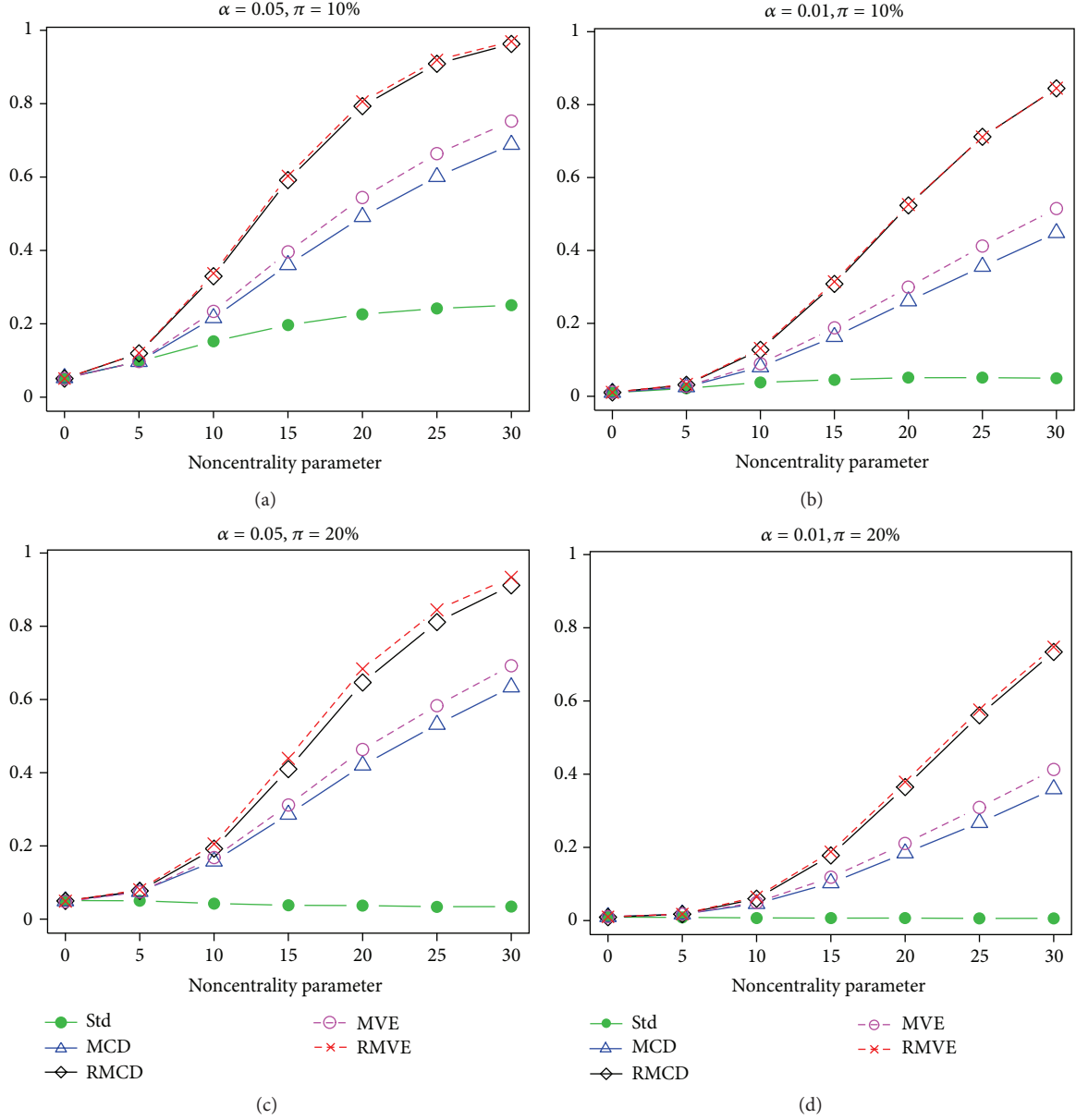


FIGURE 4: Probability of signal for T^2 control chart with different estimation methods for $p = 2$, $m = 50$.

4. Performance Analysis

We assess the performance of the proposed charts when outliers are present due to the shift in the process mean. In their study, Jensen et al. [5] concluded that the T^2_{MCD}/T^2_{MVE} control charts had better performance in terms of probability of signal. Hence, we compare the performance of our proposed method with T^2_{MCD}/T^2_{MVE} charts as well as the standard T^2 charts based on classical estimators. Our study compares more combinations of dimension p , sample size m , and π . For a particular combination of p , m , and π , a number of datasets are generated. Out of the m observations generated, $m * \pi$ of them are random data points generated from the out-of-control distribution, and the remaining $m * (1 - \pi)$ observations are generated from the in-control distribution so that the sample of m data points may contain some

outliers. We set $\pi = 0.10$ and 0.20 to ensure that the sample contains few outliers. Without loss of generality, we consider the in-control distribution as $N(0, I_p)$. The out-of-control distribution is a multivariate normal with a small shift in the mean vector with same covariance matrix. The amount of mean shift is defined through a noncentrality parameter (δ), which is given by

$$\delta = (\mu_1 - \mu)' \Sigma^{-1} (\mu_1 - \mu), \quad (7)$$

where $(\mu_1 - \mu)$ is the shift in the mean vector. The larger the value of δ is, the more extreme the outliers are. The proportion of datasets that had at least one T^2_{RMCD} or T^2_{RMVE} statistic greater than the control limit was calculated, and this proportion becomes the estimated probability of signal. We compared the performance of these charts with standard T^2

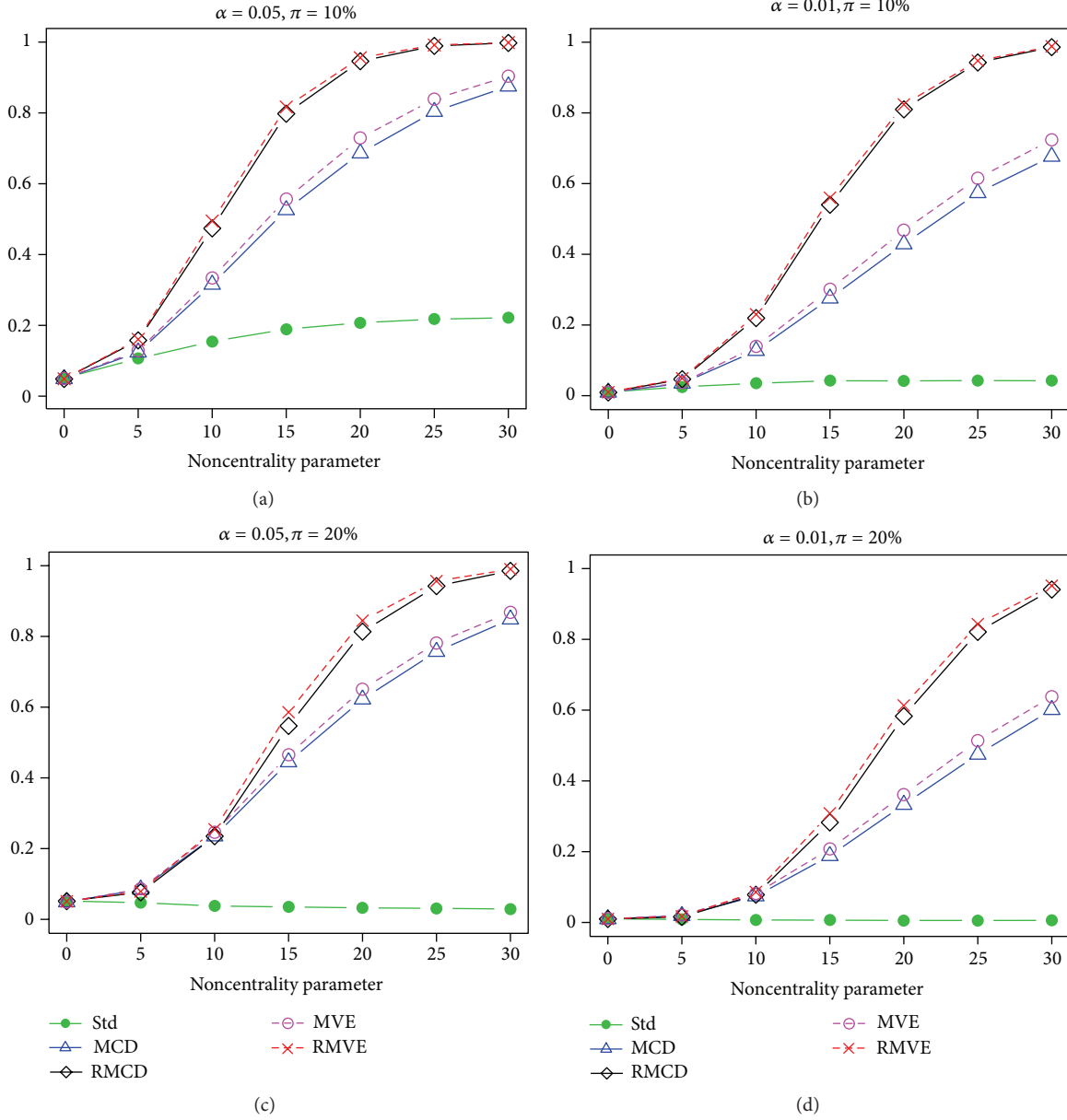


FIGURE 5: Probability of signal for T^2 control chart with different estimation methods for $p = 2$, $m = 100$.

charts, T^2_{MCD} , and T^2_{MVE} charts. The standard T^2 chart was included in our performance study as a reference because of its common usage.

The probability of a signal for different values of $\delta = (0, 5, 10, 15, 20, 25, 30)$ and for some of the values of $m = (30, 50, 100, 150)$, $p = (2, 6, 10)$ and $\pi = (10\%, 20\%)$ was considered in our study. Fifty thousand datasets of size m were generated for each combination of p , π , and δ , and the probability of signal was estimated for $\alpha = 0.05, 0.01$, and 0.001 . We considered various combinations of μ_1, μ_2 , and ρ which determine δ as per (7) and found that the probability of signal is the same irrespective of the combination of μ_1, μ_2 and ρ . Hence we have considered $\mu_1 = \mu_2$ and $\rho = 0$ for various

values of δ . We have presented only a selected set of plots to save space. The plots of probability of signal for $\alpha = 0.05$ and 0.01 , $p = 2$ and 6 , and $m = 50$ and 100 are given in Figures 4, 5, 6, and 7 for easier understanding. For dimension $p = 10$, we used $m = 100$ and 150 , and the plots of probability of signal are given in Figures 8 and 9.

From Figures 4–9, we can see that when the value of the noncentrality parameter is zero or close to zero, the probability of signal is close to α which is expected for an in-control process. As the value of the noncentrality parameter increases the probability of signals also increases. Using this criterion, we select the best method for identifying the outliers. If the probability of signal does not increase for

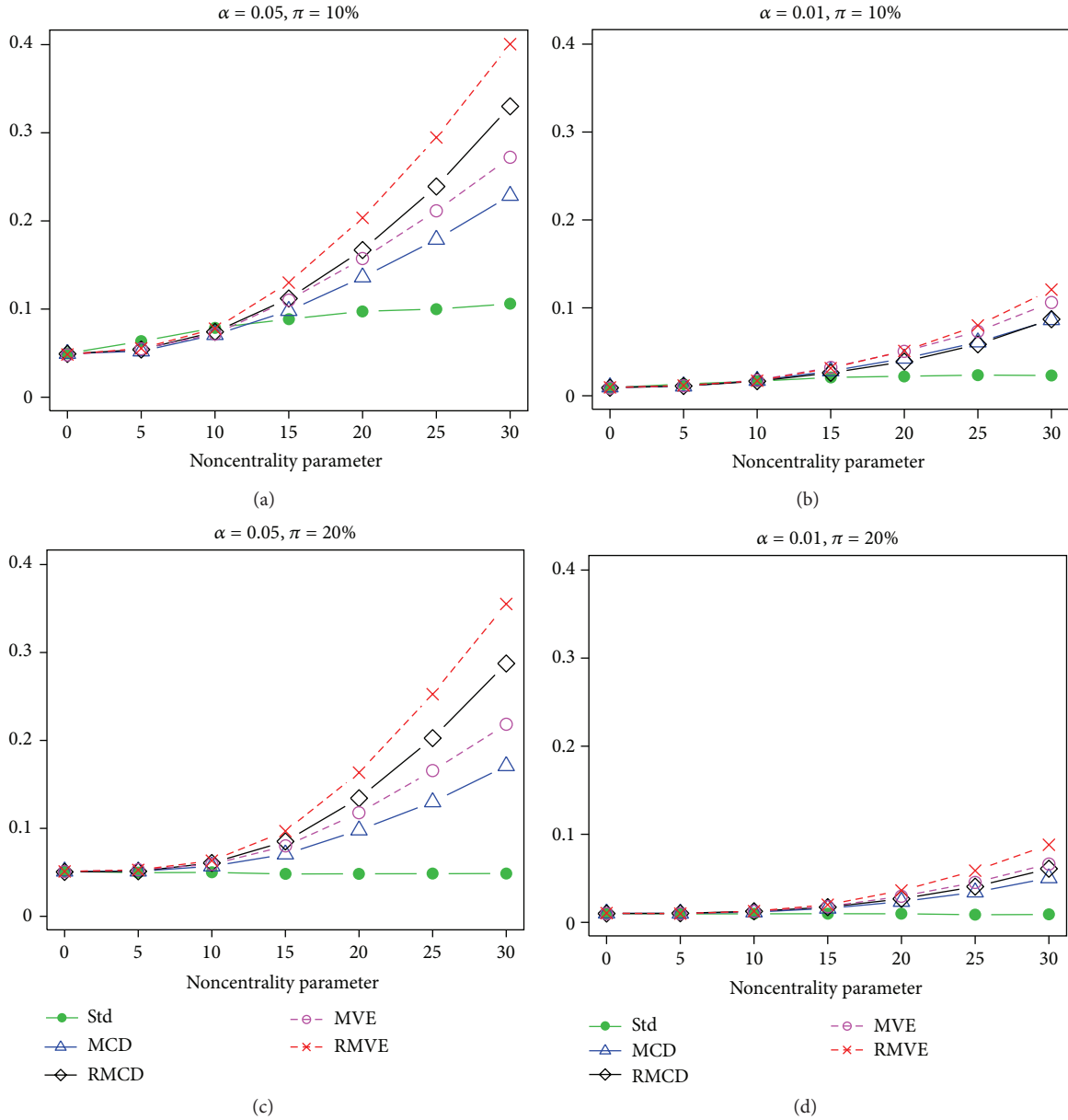


FIGURE 6: Probability of signal for T^2 control chart with different estimation methods for $p = 6$, $m = 50$.

increase in noncentrality parameter, then it is clear that the estimator has broken down and is not capable of detecting the outliers.

A careful examination of these plots of probability of signals corresponding to various values of p , m , and π indicates that for small values of p and m , T^2_{RMVE} performs well. As m and p increase, T^2_{RMCD} chart is superior. For example, from Figures 4 and 5 we see that T^2_{RMVE} has slight advantage over T^2_{RMCD} . But compared to T^2_{MCD}/T^2_{MVE} charts, T^2_{RMCD}/T^2_{RMVE} charts are performing well which is evident from all the plots presented here. When p is large (see Figures 8 and 9), the T^2_{RMCD} has clear advantage compared to T^2_{RMVE} . From these figures, we see that standard T^2 control chart

possesses little ability to detect the outliers and the T^2_{MVE} , and T^2_{MVE} stands below the T^2_{RMCD}/T^2_{RMVE} charts throughout all the values of δ .

As p increases for a fixed value of m , the breakdown points of RMCD and RMVE get smaller as the breakdown value is given by $(m - p + 1)/2m$. This suggests that the larger p is, the larger m will need to be in order to maintain the breakdown point, which is very well demonstrated in Figures 8 and 9. In general, there was always one estimator, RMCD or RMVE, that was found to be superior across all the values of the noncentrality parameter as long as the proportion of outliers was not so big as to cause the estimators to break down. This greatly simplifies the conclusions that

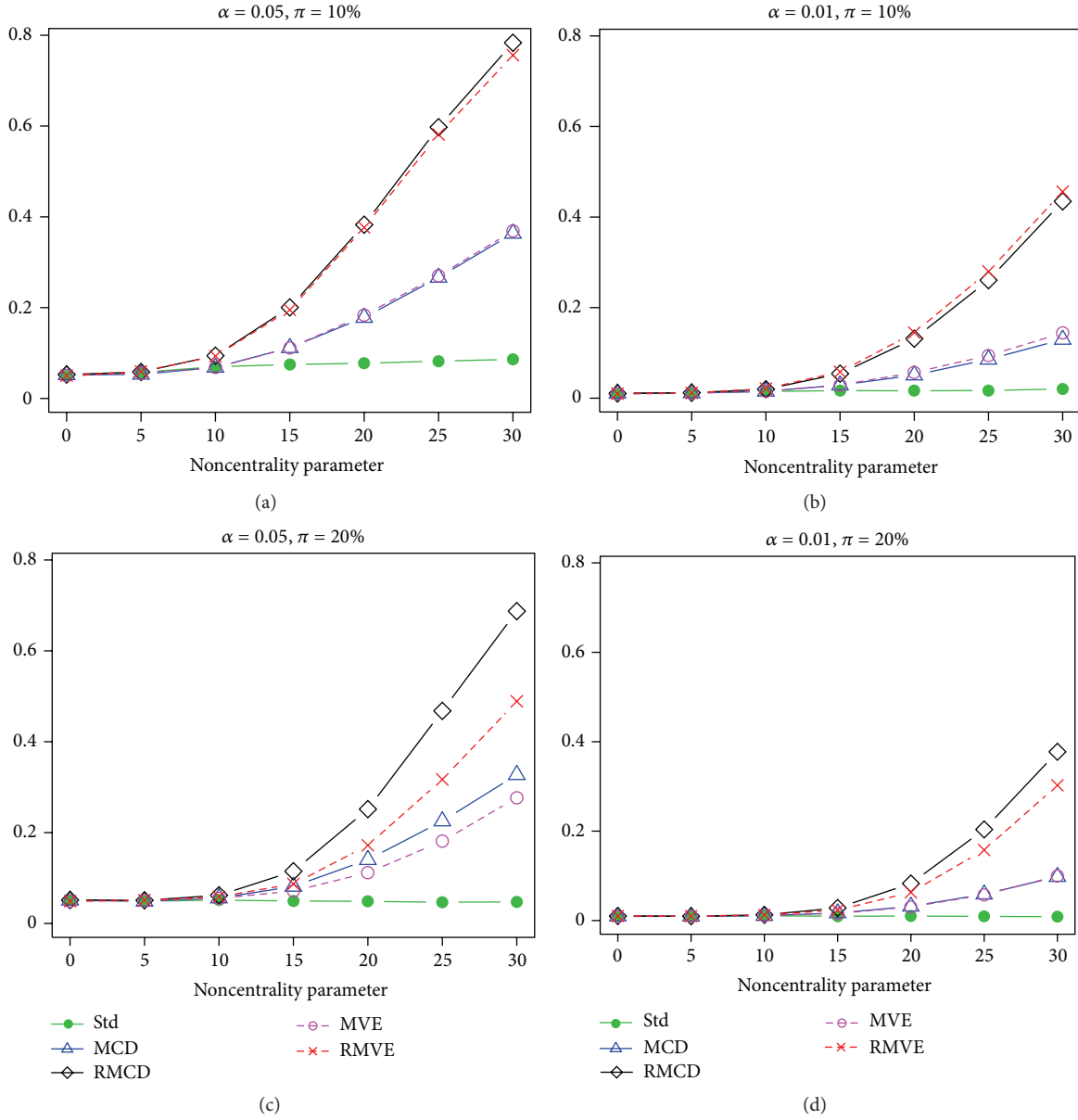


FIGURE 7: Probability of signal for for T^2 control chart with different estimation methods for $p = 6$, $m = 100$.

can be made about when the RMCD or RMVE estimators are preferred to the MCD and MVE estimators.

Nevertheless, T^2_{RMCD} and T^2_{RMVE} charts are preferred for the various combinations of m , p , and π , and some broad recommendations can be made on the selection among these two charts. When $m < 100$, the T^2_{RMVE} will be the best for small dimension. When $m \geq 100$, the T^2_{RMCD} is preferred. As p increases, then the percentage of outliers that can be detected by the T^2_{RMVE} chart decreases. It is true for both the charts that when p is higher, the number of outliers that can be detected decreases for smaller sample sizes. Thus

for Phase-I applications where the number of outliers is unknown, T^2_{RMVE} should be used only for smaller sample sizes, and it is also computationally feasible. T^2_{RMCD} should be used for larger sample sizes or when it is believed that there is a large number of outliers. When the dimension is large, larger sample sizes are needed to ensure that the estimator does not break down and lose its ability to detect outliers. Hence for larger dimension cases, T^2_{RMCD} is preferred with large sample sizes. For very small samples ($m < 30$), one may opt for higher values of γ , for which control limits need to be developed.

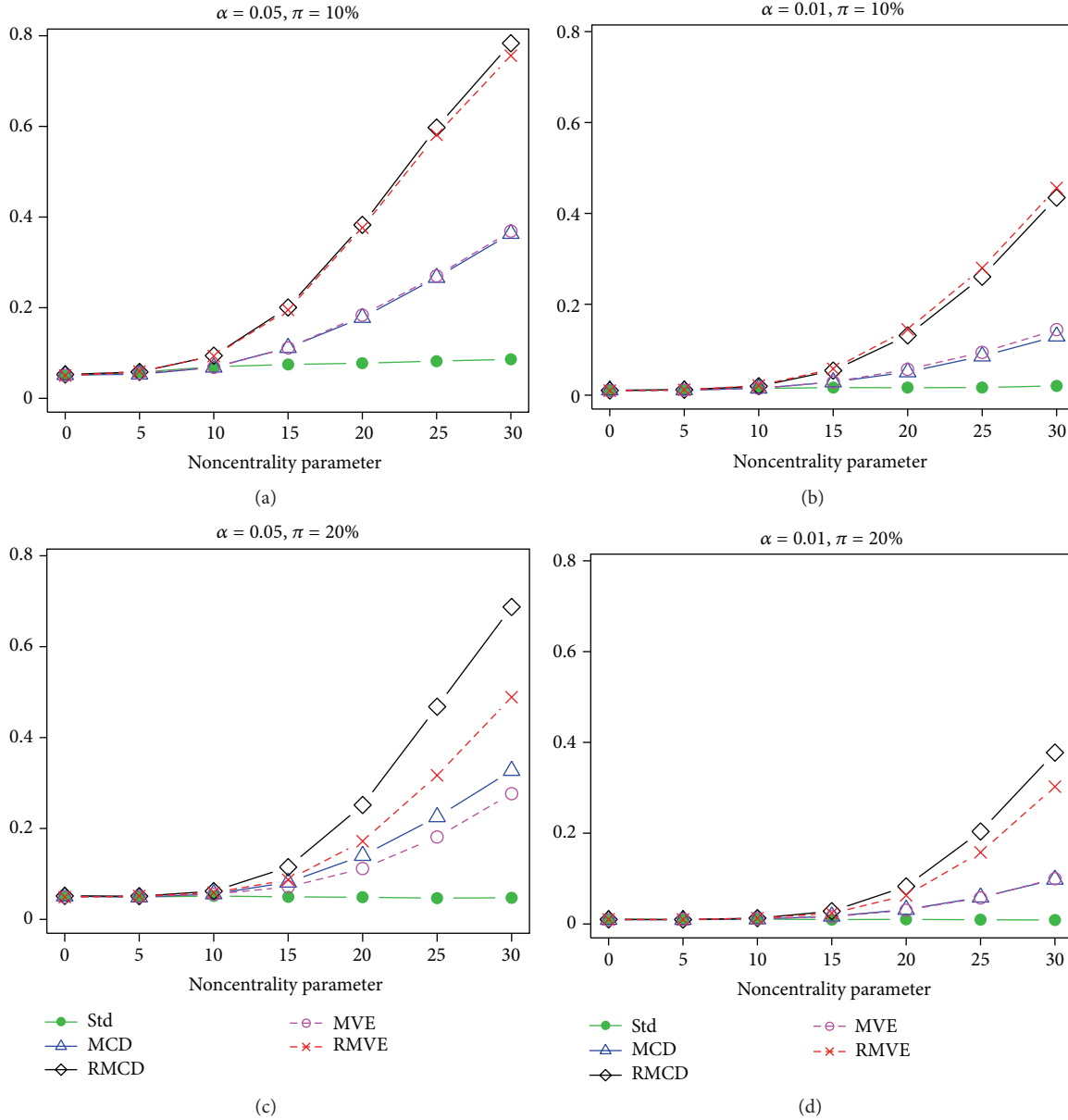


FIGURE 8: Probability of signal for T^2 control chart with different estimation methods for $p = 10$, $m = 100$.

5. Case Example

To illustrate the applicability of the proposed control chart method, we discuss a real case example taken from an electronic industry. The data gives 105 measurements of 3 axial components of acceleration measured by accelerometer on a e-compass unit fixed on the objects. The mean vector and covariance matrix under the classical, RMCD, and RMVE methods of the sample data considered are given by

$$\bar{X} = \begin{pmatrix} 6.3143 \\ 5.7339 \\ 5.7527 \end{pmatrix},$$

$$S = \begin{pmatrix} 3.4022 & -1.1524 & -1.0746 \\ -1.1524 & 1.9249 & 1.1209 \\ -1.0746 & 1.1209 & 2.2004 \end{pmatrix},$$

$$\bar{X}_{\text{RMCD}} = \begin{pmatrix} 5.7125 \\ 6.2643 \\ 6.0837 \end{pmatrix},$$

$$S_{\text{RMCD}} = \begin{pmatrix} 2.8549 & 0.1901 & -0.2926 \\ 0.1901 & 1.1175 & 0.4433 \\ -0.2926 & 0.4433 & 2.3115 \end{pmatrix},$$

$$\bar{X}_{\text{RMVE}} = \begin{pmatrix} 5.7790 \\ 6.1379 \\ 5.9894 \end{pmatrix},$$

$$S_{\text{RMVE}} = \begin{pmatrix} 2.7837 & -0.2642 & -0.5745 \\ -0.2642 & 1.2778 & 0.6080 \\ -0.5745 & 0.6080 & 2.2909 \end{pmatrix}.$$

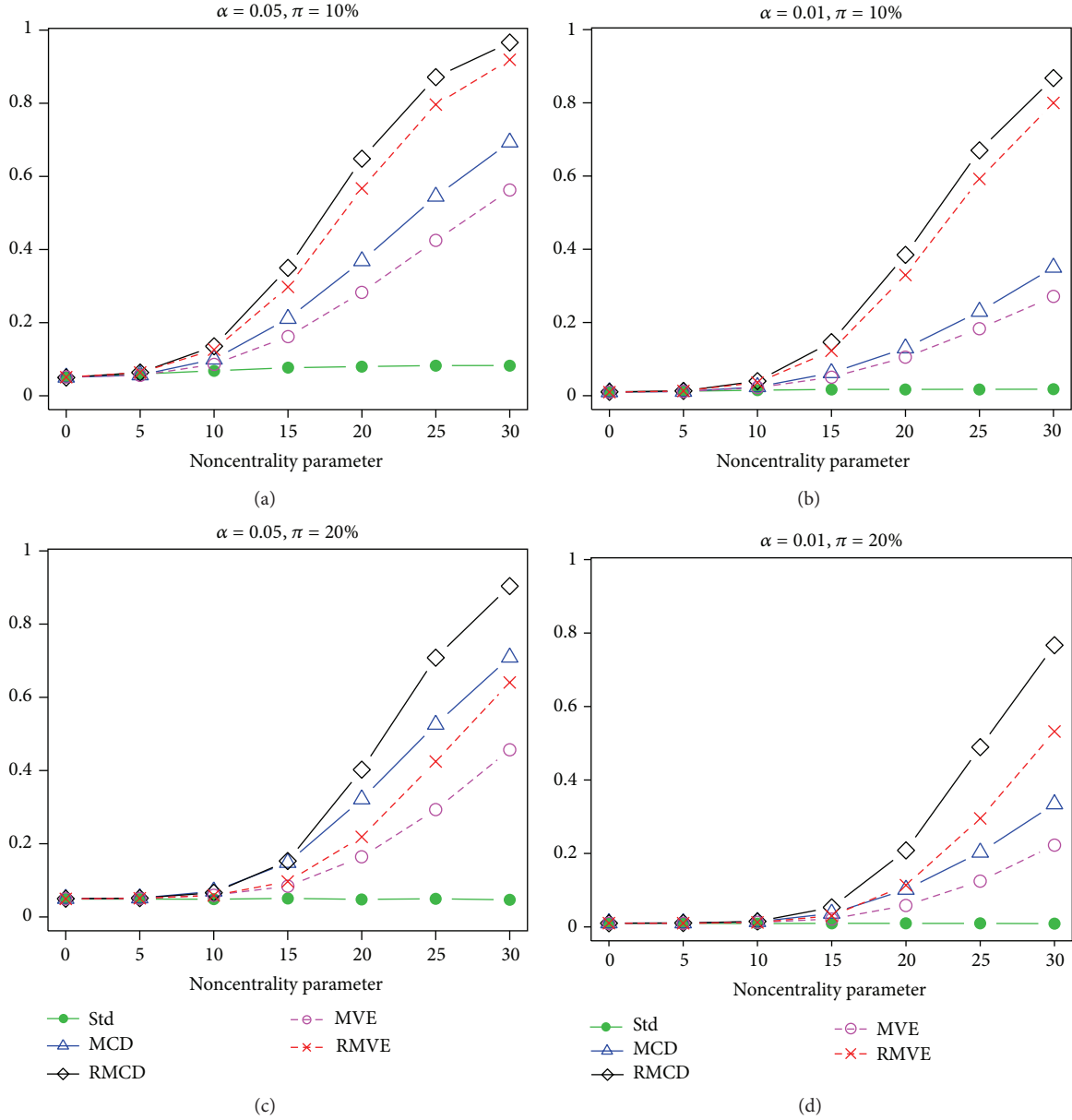


FIGURE 9: Probability of signal for T^2 control chart with different estimation methods for $p = 10$, $m = 150$.

A simple comparison of these estimators indicates that there are outliers in the Phase-I data. The plots of T^2 , T^2_{RMCD} , and T^2_{RMVE} values along with the respective control limits at 99% confidence level for the sample data are given in Figure 10.

The control limits for T^2 are arrived at based on beta distribution, and T^2_{RMCD}/T^2_{RMVE} are calculated using (6) for $p = 3$ and $m = 105$. From Figure 10, it is very clear that both T^2_{RMCD} and T^2_{RMVE} control chart alarms signal for 3 outliers whereas the standard T^2 control chart alarm signals for none even though all the charts are having the same pattern. This indicates the effectiveness of the proposed robust control charts in identifying the outliers.

6. Conclusions

Use of robust control chart in Phase-I monitoring is very important to assess the performance of the process as well as detecting outliers. We propose T^2_{RMCD}/T^2_{RMVE} control charts for Phase-I monitoring of multivariate individual observations. The control limits for these charts are arrived empirically and a non-linear regression model is used for arriving control limits for any sample size. The performance of the proposed charts were compared under various data scenarios using large number of Monte Carlo simulations. Our simulation studies indicate that T^2_{RMVE} control charts are performing well for smaller sample sizes and smaller

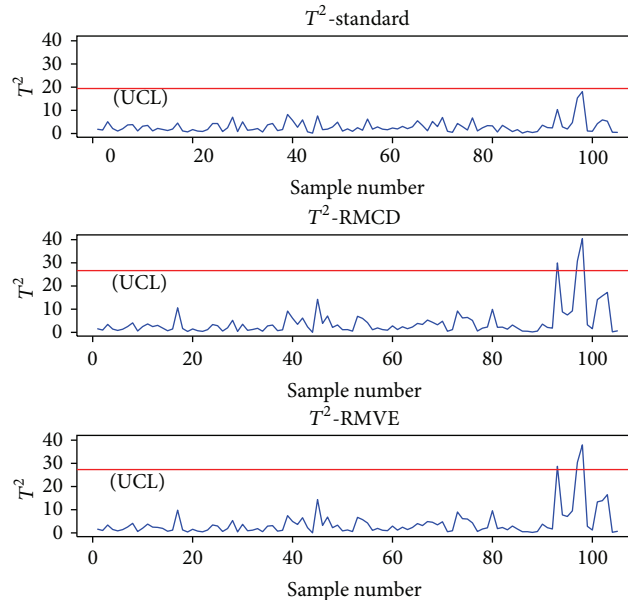


FIGURE 10: T^2 , T_{RMCD}^2 , and T_{RMVE}^2 control charts for the sample data.

dimension where as T_{RMCD}^2 control charts are performing well for larger sample sizes and larger dimensions. We illustrated our proposed robust control chart methodology using a case study from the electronic industry.

Acknowledgments

The authors would like to thank the editor and two anonymous referees for their valuable comments and suggestions that substantially improved the overall quality of an earlier version of this paper. The research is supported by a grant from the Natural Science and Engineering Council of Canada.

References

- [1] H. Hotelling, "Multivariate quality control," in *Techniques of Statistical Analysis*, C. Eisenhart, H. Hastay, and W. A. Wallis, Eds., pp. 111–184, McGrawHill, New York, NY, USA, 1947.
- [2] N. D. Tracy, J. C. Young, and R. L. Mason, "Multivariate control charts for individual observations," *Journal of Quality Technology*, vol. 24, pp. 88–95, 1992.
- [3] J. H. Sullivan and W. H. Woodall, "A comparison of multivariate control charts for individual observations," *Journal of Quality Technology*, vol. 28, no. 4, pp. 398–408, 1996.
- [4] J. A. N. Vargas, "Robust estimation in multivariate control charts for individual observations," *Journal of Quality Technology*, vol. 35, no. 4, pp. 367–376, 2003.
- [5] W. A. Jensen, J. B. Birch, and W. H. Woodall, "High breakdown estimation methods for phase I multivariate control charts," *Quality and Reliability Engineering International*, vol. 23, no. 5, pp. 615–629, 2007.
- [6] S. Chenouri, S. H. Steiner, and A. M. Variyath, "A multivariate robust control chart for individual observations," *Journal of Quality Technology*, vol. 41, no. 3, pp. 259–271, 2009.
- [7] D. L. Donoho and P. J. Huber, "The notion of breakdown point," in *A Festschrift for Erich Lehmann*, P. Bickel, K. Doksum, and J. Hodges, Eds., pp. 157–184, Wadsworth, Belmont, Calif, USA, 1983.
- [8] H. P. Lopuhaä and P. J. Rousseeuw, "Breakdown points of affine equivariant estimators of multivariate location and covariance matrices," *The Annals of Statistics*, vol. 19, pp. 229–248, 1991.
- [9] D. L. Donoho and M. Gasko, "Breakdown properties of location estimates based on halfspace depth and projected outlyingness," *The Annals of Statistics*, vol. 20, pp. 1803–1827, 1992.
- [10] P. L. Davies, "Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices," *The Annals of Statistics*, vol. 15, pp. 1269–1292, 1987.
- [11] C. Croux and G. Haesbroeck, "Influence function and efficiency of the minimum covariance determinant scatter matrix estimator," *Journal of Multivariate Analysis*, vol. 71, no. 2, pp. 161–190, 1999.
- [12] G. Pison, S. van Aelst, and G. Willems, "Small sample corrections for LTS and MCD," *Metrika*, vol. 55, no. 1-2, pp. 111–123, 2002.
- [13] D. L. Woodru and D. M. Rocke, "Computable robust estimation of multivariate location and shape in high dimension using compound estimators," *Journal of American Statistical Association*, vol. 89, pp. 888–896, 1994.
- [14] D. M. Hawkins and D. J. Olive, "Improved feasible solution algorithms for high breakdown estimation," *Computational Statistics and Data Analysis*, vol. 30, no. 1, pp. 1–11, 1999.
- [15] P. J. Rousseeuw and K. van Driessen, "A fast algorithm for the minimum covariance determinant estimator," *Technometrics*, vol. 41, no. 3, pp. 212–223, 1999.
- [16] P. J. Rousseeuw and B. C. van Zomeren, "Unmasking multivariate outliers and leverage points," *Journal of American Statistical Association*, vol. 85, pp. 633–639, 1990.
- [17] P. L. Davies, "The asymptotics of Rousseeuw's minimum volume ellipsoid estimator," *The Annals of Statistics*, vol. 20, no. 4, pp. 1828–1843, 1992.

- [18] C. Croux and G. Haesbroeck, "An easy way to increase the finite-sample efficiency of the resampled minimum volume ellipsoid estimator," *Computational Statistics and Data Analysis*, vol. 25, no. 2, pp. 125–141, 1997.
- [19] V. Torodov, "rrcov: Scalable Robust Estimators with High Breakdown point," R package version 0.5-03, 2009, <http://cran.r-project.org/web/packages/rrcov/>.

