HOTELLING'S T^2 CONTROL CHARTS BASED ON ROBUST ESTIMATORS

CARTAS DE CONTROL T^2 DE HOTELLING BASADAS EN ESTIMADORES ROBUSTOS

SERGIO YÁÑEZ
Escuela de Estadística, Universidad Nacional de Colombia, Medellín, Colombia, syanez@unalmed.edu.co

NELFI GONZÁLEZ
Escuela de Estadística, Universidad Nacional de Colombia, Medellín, Colombia, ngonzale@unalmed.edu.co

JOSÉ ALBERTO VARGAS
Departamento de Estadística, Universidad Nacional de Colombia, Bogotá, Colombia, javargasn@unal.edu.co

Received for review February 24 th, 2009, accepted June 2 th, 2009, final version June, 22 th, 2009

ABSTRACT: Under the presence of multivariate outliers, in a Phase I analysis of historical set of data, the T^2 control chart based on the usual sample mean vector and sample variance - covariance matrix performs poorly. Several alternative estimators have been proposed. Among them, estimators based on the minimum volume ellipsoid (MVE) and the minimum covariance determinant (MCD) are powerful in detecting a reasonable number of outliers. In this paper we propose a T^2 control chart using the biweight S estimators for the location and dispersion parameters when monitoring multivariate individual observations. Simulation studies show that this method outperforms the T^2 control chart based on MVE estimators for a small number of observations.

KEYWORDS: Multivariate Control Charts, MVE Estimators, Outliers, S Estimators.

RESUMEN: En presencia de outliers multivariados, durante la Fase I de análisis de datos históricos, la carta de control T^2 , basada en los estimadores usuales del vector de medias y de la matriz de varianzas – covarianzas, se comporta de manera deficiente. Varias alternativas se han propuesto. Entre otras, estimadores basados en el elipsoide de mínimo volumen (MVE) y en el determinante de minima covarianza (MCD) son potentes para detectar un número razonable de outliers. En este artículo proponemos una carta de control T^2 usando los estimadores S biponderados para los parámetros de localización y dispersion cuando se monitorean observaciones multivariadas individuales. Estudios de simulación muestran que este método supera las cartas T^2 basadas en los estimadores MVE para un número pequeño de observaciones.

PALABRAS CLAVES: Cartas de Control Multivariadas, Estimadores MVE, Outliers, Estimadores S.

1. INTRODUCTION

Hotelling's T^2 control chart is a widely used tool for monitoring simultaneously several related quality characteristics of a process. See for example [1, 2]. Recently, it has been used for monitoring quality profiles [3, 4]. Following the terminology of [5], in the Stage 1 of Phase I, historical data are studied for determining whether the process was in control and to estimate the in-control parameters of the process. Sometimes is not possible to group these data into rational subgroups [6, 7], so charts are based on individual multivariate observations.

In a Phase I analysis of historical data set, the usual estimates of the process parameters are the sample mean vector and the sample variance-covariance matrix. It is well known that these estimators are sensitive to outlying observations. As a result, statistics plotted on T^2 control charts, based on these estimators, perform poorly when there are several outliers [7, 8, 9]. Alternative estimation methods have been proposed in the literature. One approach consists of calculating the T^2 statistic based on successive differences variance-covariance matrix estimator. See for example, [7, 10, 11]. Though this method is effective in detecting sustained shifts in the mean vector, it fails to detect outliers as is shown in [7, 9]. Another approach uses robust estimators of the process parameters. In [9] is proposed the use of high breakdown estimation methods based on the minimum volume ellipsoid (MVE) estimators of [12] and the minimum covariance determinant (MCD) method of [13]. He showed that a T^2 control chart using MVE estimators was the most effective method to detect out of control signals due to several outliers. In a large simulation study, [14] showed that a T^2 control chart based on MVE estimators performs better than a chart based on MCD estimators when the number of observations is small, but the opposite occurs when the number of observations increases. In [15] three multivariate control charts are compared: the S chart, the MVE chart and the Usual chart. They did it just for the case p=2 and m=30. In this paper we generalize these results by extending our simulations to several values of p and m. We also extend the types of contamination of parameters. It is our purpose to give conclusions under a more general framework.

2. ROBUST ESTIMATORS

Let $\mathbf{x}_1, \ldots, \mathbf{x}_m$ be a set of m observations selected from a p-multivariate normal distribution. The MVE estimator of location and variance-covariance matrix is the pair (t, C) that minimizes the determinant of **C**, subject to

$$\#\left\{i: \left(\mathbf{x}_i-\mathbf{t}\right)^t \mathbf{C}^{-1}\left(\mathbf{x}_i-\mathbf{t}\right) \le \chi^2_{(0.5,p)}\right\} \ge \left[\frac{m+p+1}{2}\right] \quad (1)$$

where the symbol # means the number of points which satisfies the condition, and $\chi^2_{(0.5,p)}$ is the 0.5-quantil of the chi-square distribution with p degrees of freedom. The values, **t** and **C** estimate the center of the smallest ellipsoid containing at least half of the observations and the inverse of the shape matrix of the ellipse, respectively. The cov.mve function of S-PLUS calculates these estimators based on a genetic algorithm.

The biweight S estimator of location and shape is defined as the pair (t, C) that minimizes the determinant $|k^2\mathbf{C}|$, subject to

$$m^{-1}\sum_{i=1}^{m}\rho\left(\sqrt{\left(\mathbf{x}_i-\mathbf{t}\right)^t\left(k^2\mathbf{C}\right)^{-1}\left(\mathbf{x}_i-\mathbf{t}\right)}\right)=b_0 \quad (2)$$

where ρ is Tukey's biweight function. An algorithm proposed by [16] to calculate these estimators is outlined in the Appendix.

MVE and S estimators are good candidates to estimate process parameters because they are affine equivariant and have high breakdown points [16, 17]. The breakdown point is the smallest fraction of contamination that causes an estimator to take on values arbitrarily far away [17]. S estimators properties has been widely studied as a high-efficiency robust estimators. See for example, [16, 18, 19, 20, 21].

3. THE T^2 CONTROL CHART

We assume that the data set from Phase I analysis consists of m statistically independent observations x_i, $i = 1, 2, \ldots, m$ such that $x_i \sim N_p(\mu, \Sigma)$, where p represents the number of quality characteristics being monitored. The usual estimators of μ and Σ are

$$\bar{\mathbf{x}} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{x}_i \quad (3)$$

and,

$$\mathbf{S}_1 = \frac{1}{m-1}\sum_{i=1}^{m}\left(\mathbf{x}_i-\bar{\mathbf{x}}\right)\left(\mathbf{x}_i-\bar{\mathbf{x}}\right)^t \quad (4)$$

respectively. The T^2 statistics based on the usual estimators are

$$T^2_{U,i} = \left(\mathbf{x}_i - \bar{\mathbf{x}}\right)^t \mathbf{S}_1^{-1} \left(\mathbf{x}_i - \bar{\mathbf{x}}\right) \;,\; i = 1, 2, \ldots, m \quad (5)$$

Those based on MVE and S estimators are

$$T^2_{MVE,i} = \left(\mathbf{x}_i - \mathbf{x}_{MVE}\right)^t \mathbf{S}_{MVE}^{-1} \left(\mathbf{x}_i - \mathbf{x}_{MVE}\right) \;,\; i = 1, 2, \ldots, m \quad (6)$$

and

$$T^2_{S,i} = \left(\mathbf{x}_i - \mathbf{x}_S\right)^t \mathbf{S}_S^{-1} \left(\mathbf{x}_i - \mathbf{x}_S\right) \;,\; i = 1, 2, \ldots, m \quad (7)$$

respectively.

Upper control limits (UCL) for the three charts were calculated from 5000 simulations with an overall false alarm probability of 0.05. Once control limits were calculated, the $T^2_{\bullet,i}$ (the $\bullet$ should be replaced by U, MVE, or S) were plotted on the chart.

4. SIMULATION SCHEMES

Sets of m=30, 40 and 50 observations were generated from multivariate normal distributions. Without loss of generality, we assume that the in-control distribution is multivariate normal with mean $\mu_0=0$ and covariance matrix $\Sigma_0=I$, the identity matrix. Let $\delta^2 = \left(\mu_1 - \mu_0\right)^t \Sigma_0^{-1} \left(\mu_1 - \mu_0\right)$ denote the non-centrality parameter, which measures the shift from μ_0 to an out-of-control mean vector μ_1. To generate outliers that contaminated the in-control distribution, k < m observations were randomly generated as follows:

1. Shift the mean vector: k observations were generated from $N_p(\mu_1, \Sigma_0)$ distributions, for δ^2=5, 10, 15, 20, 25, p=2, 3, 5, 10, and k=1, 2, …, 7.
2. Change in the variance-covariance matrix or symmetric contamination: k observations were generated from $N_p(\mu_0, \lambda\Sigma_0)$ distributions, for λ=1.5, 2, 2.5, 3.5, 4.5, 8, 10, 12, 16, p=2, 3, 5, 10, and k=1, 2,…, 7
3. Crossed contamination: k observations were generated from $N_p(\mu_1, \lambda\Sigma_0)$ distributions, for δ^2=5, 10, 15, 20, 25, λ=1.5, 4.5, 8.5, 12.5, p=2, 3, 5, 10, and k=1, 2, …, 7.

To evaluate control chart performance, N=1000 replicates were generated for each combination of the above schemes. The control charts, T^2_U, T^2_{MVE} and T^2_S were then compared by estimating the average proportion of outliers detected (APOD). The APOD is a comparison criterion suggested by [22], which is defined as follows

$$APOD = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{k}\sum_{j=1}^{k} I\left(o_j = 1\right)\right] \quad (8)$$

where $I\left(o_j = 1\right)$ is 1 if $o_j = 1$ and 0 otherwise, and $o_j = 1$ if $T^2_{\bullet,j} > UCL$. Thus, for example, if a data set of size m has 4 outliers and APOD=0.5, then it is expected, on the long run, that the control chart will detect an average of two outliers.

Though the signal probability is the usual criterion for comparing multivariate control charts in Phase I analysis [7, 9], we used the APOD because the expected proportion of the exact number of outliers simultaneously detected seems to be more informative than a signal probability, mainly under the presence of multiple outliers. Moreover, limited simulations, not presented here, showed that plots of APOD and signal probabilities exhibit similar patterns when comparing T^2 control charts.

5. RESULTS

Figure 1 shows APODs under shifts in the mean vector for different non-centrality parameters, p=2, m=30, and k=1, 4, 7 respectively. For these values of p and m, upper control limits were 10.5123, 24.9336 and 20.2441 for the Usual, MVE and S methods respectively. The Usual T^2 control chart performed poorly except when there was just one outlier. For $2 \leq k \leq 7$, estimated APODs for T^2_S control charts were consistently superior to those for T^2_{MVE} and Usual charts. For instance, in the presence of k=4 outliers and δ^2=15, the APODs for the T_S^2, T_{MVE}^2 and T_U^2 charts were 0.2515, 0.1835 and 0.0275, respectively.

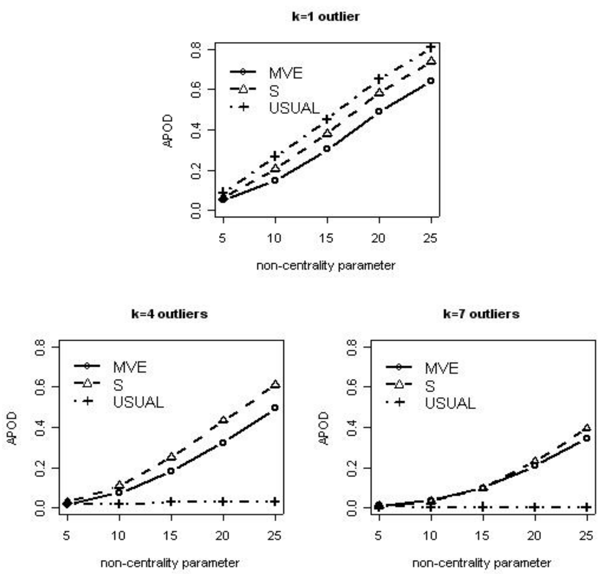

Figure 1. Average proportion of outliers detected under shifts in the mean vector for p=2, m=30, with k=1, 4 and 7 outliers

Table 1 shows estimated APODs for MVE and S methods for p=3, 5, 10, m=30, 40, 50 and k=2, 4, 7. Similar patterns are observed for $p = 3$ and p=2. Otherwise, MVE and S methods had a similar performance.

Table 1. Estimated Average proportions of outliers detected by T_{MVE}^2 and T_S^2 under shifts in the mean vector, for p=3, 5, 10, m=30, 40, 50 with k=2, 4 and 7 outliers and several values of the non-centrality parameter δ^2

δ^2	Esti	p=3			p=5			p=10		
		k =2	4	7	k =2	4	7	k =2	4	7
m=30										
5	MVE	.0250	.0085	.0076	.0130	.0103	.0044	.0060	.0028	.0026
	S	.0325	.0165	.0071	.0245	.0075	.0047	.0095	.0073	.0023
10	MVE	.0740	.0460	.0157	.0520	.0260	.0097	.0110	.0095	.0044
	S	.1185	.0565	.0199	.0665	.0188	.0091	.0190	.0098	.0029
15	MVE	.1725	.1290	.0597	.0965	.0643	.0190	.0230	.0128	.0034
	S	.2520	.1463	.0609	.1420	.0535	.0254	.0395	.0133	.0046
20	MVE	.2990	.2318	.1220	.1660	.1203	.0341	.0385	.0128	.0040
	S	.4170	.3143	.1479	.2465	.1475	.0636	.0715	.0173	.0041
25	MVE	.4675	.3915	.2116	.2750	.2123	.0840	.0645	.0163	.0039
	S	.5965	.4750	.2764	.3735	.2488	.1413	.1045	.0160	.0046
m=40										
5	MVE	.0285	.0193	.0080	.0150	.0095	.0056	.0085	.0048	.0026
	S	.0325	.0243	.0107	.0165	.0090	.0051	.0100	.0053	.0024
10	MVE	.1090	.0745	.0301	.0595	.0370	.0124	.0205	.0100	.0043
	S	.1405	.0850	.0413	.0770	.0295	.0116	.0280	.0058	.0031
15	MVE	.2625	.1855	.1027	.1360	.0963	.0393	.0480	.0218	.0053
	S	.3215	.2208	.1290	.1815	.1118	.0357	.0500	.0123	.0030
20	MVE	.4160	.3248	.2279	.2520	.1965	.1101	.0975	.0378	.0077
	S	.4935	.4205	.2987	.3425	.2323	.1154	.0845	.0138	.0036
25	MVE	.6005	.5215	.3939	.3840	.3153	.2144	.1595	.0635	.0136
	S	.6935	.6035	.4766	.5100	.4198	.2569	.1220	.0143	.0049
m=50										
5	MVE	.0370	.0228	.0110	.0160	.0175	.0043	.0085	.0058	.0034
	S	.0345	.0250	.0124	.0240	.0110	.0046	.0070	.0050	.0026
10	MVE	.1330	.0908	.0513	.0780	.0513	.0216	.0225	.0123	.0059
	S	.1515	.1020	.0564	.0880	.0433	.0151	.0420	.0098	.0034
15	MVE	.3055	.2320	.1559	.1880	.1273	.0690	.0590	.0295	.0080
	S	.3455	.2730	.1873	.2070	.1348	.0543	.0695	.0143	.0050
20	MVE	.4885	.4035	.3250	.3495	.2515	.1753	.1220	.0730	.0279
	S	.5550	.4995	.3839	.3710	.2803	.1697	.1440	.0278	.0051
25	MVE	.6535	.6058	.5249	.4735	.4435	.3420	.2040	.1433	.0501
	S	.7300	.6810	.6089	.5580	.4828	.3399	.2605	.0440	.0044

Figure 2 shows estimated APODs under symmetric contamination for different values of λ, p=2 and m=30, and k=1, 4, 7 respectively. In the presence of a single outlier T_U^2 control charts are most powerful, but their estimated APODs are smallest when there are several outliers. For multiple outliers, estimated APODs for T_S^2 are consistently higher than those for T_{MVE}^2 control charts.

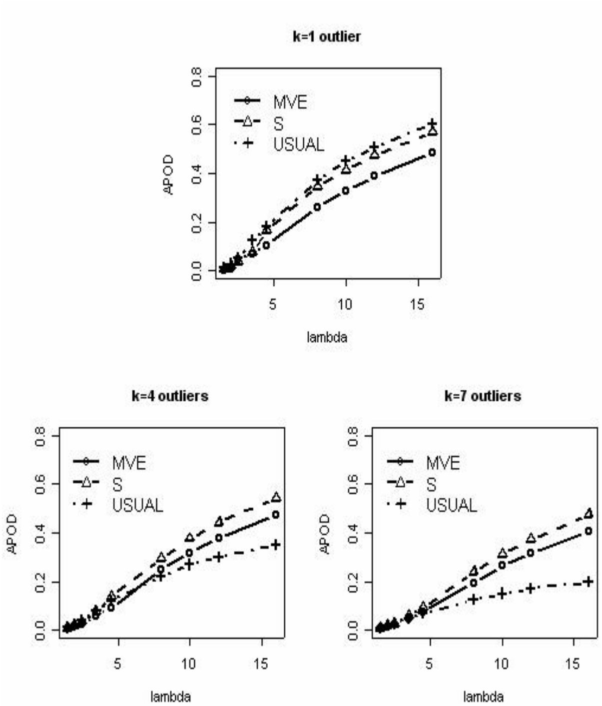

Figure 2. Average proportion of outliers detected under symmetric contamination for p=2, m=30, with k=1, 4 and 7 outliers

Table 2 exhibits estimated APODs under symmetric contamination for T^2_{MVE} and T^2_S control charts and p=3, 5, 10, m=30, 40, 50 and k=2, 4, and 7. Interestingly, estimated APODs increased with p. For instance, for m=50, k=4, and λ=8 APODs for the MVE method were 0.433 for p=3, 0.598 for p=5 and 0.794 for p=10. For the S method the values were 0.454, 0.639 and 0.879, respectively. Under this type of contamination the S performs slightly better than the MVE method. Notice that for m=30, regardless of p, estimated APODs for T^2_S are consistently higher than those for T^2_{MVE} control charts.

Table 2. Estimated average proportions of outliers detected under symmetric contamination, for p=3, 5, 10, m=30, 40, 50 with k=2, 4 and 7 outliers and several values of λ

λ	Esti	p=3			p=5			p=10		
		k=2	4	7	k=2	4	7	k=2	4	7
m=30										
1.5	MVE	.0090	.0053	.0090	.0115	.0083	.0069	.0090	.0090	.0079
	S	.0085	.0085	.0063	.0115	.0108	.0094	.0145	.0120	.0126
2.5	MVE	.0445	.0393	.0284	.0460	.0423	.0339	.0580	.0433	.0373
	S	.0475	.0458	.0319	.0590	.0578	.0461	.0905	.0768	.0511
4.5	MVE	.1355	.1363	.0977	.1920	.1765	.1337	.2460	.2023	.1640
	S	.2040	.1725	.1339	.2475	.2460	.1936	.3725	.3205	.2179
8	MVE	.3525	.3208	.2716	.4855	.4425	.3859	.5510	.4905	.4076
	S	.4305	.3973	.3473	.5400	.5348	.5037	.7510	.6545	.4630
12	MVE	.5220	.4895	.4407	.6650	.6565	.517	.7140	.6703	.5709
	S	.5885	.5805	.5100	.7320	.7433	.7276	.9130	.8383	.6469
m=40										
1.5	MVE	.0040	.0075	.0071	.0075	.0095	.0087	.0125	.0143	.0096
	S	.0080	.0098	.0079	.0095	.0113	.0109	.014	.014	.0113
2.5	MVE	.0440	.0438	.0343	.0500	.055	.0461	.081	.081	.0693
	S	.0685	.0473	.0436	.0695	.0673	.0556	.124	.1025	.0824
4.5	MVE	.1740	.1645	.1409	.2345	.2335	.1957	.4045	.3673	.3304
	S	.2175	.1968	.1741	.3075	.284	.2676	.509	.4448	.3516
8	MVE	.4120	.3998	.355	.5495	.5403	.5217	.7645	.743	.711
	S	.4615	.4430	.4014	.6085	.609	.5853	.824	.8038	.714
12	MVE	.5785	.5723	.5259	.7385	.7603	.7177	.915	.9018	.8577
	S	.6240	.6200	.5996	.7815	.7928	.778	.957	.944	.8944
m=50										
1.5	MVE	.0090	.0080	.0073	.0070	.0098	.0077	.0075	.0115	.0090
	S	.0115	.0083	.0083	.0055	.008	.0086	.0125	.0165	.0126
2.5	MVE	.0500	.0478	.0391	.0655	.0625	.0590	.0880	.0860	.0841
	S	.0615	.0538	.0480	.0810	.0713	.0644	.1465	.1413	.1213
4.5	MVE	.2065	.1828	.1663	.2965	.2785	.2650	.4245	.4148	.4024
	S	.2270	.2075	.1911	.3250	.3078	.2979	.5720	.5413	.4887
8	MVE	.4295	.4333	.4024	.6095	.5978	.5853	.8030	.7938	.8056
	S	.4670	.4535	.4347	.6405	.6385	.6221	.8870	.8788	.8414
12	MVE	.5920	.5925	.5647	.8045	.7838	.7590	.9385	.9320	.9347
	S	.6295	.6155	.6121	.8155	.8143	.7880	.9675	.9700	.9666

Under cross contamination, for different combinations of non-centrality parameters and λs, was observed a similar pattern that in Figures 1 and 2 for δ^2 and λ fixed, respectively.

Simulation results, not shown here, produce similar results for the combinations of p= 3, 5, 10, m= 30, 40, 50 and k= 2, 4, 7.

6. EXAMPLE

Next we illustrate the comparison of the three types of T^2 control charts. We consider the example presented by [23]. The original data set contains 11 quality variables measured on 30 products. Here we consider only the first three variables, whose values are reproduced in the table 3.

Table 3. Variables 1, 2 and 3 of [23] data set

Product No.	X_1	X_2	X_3
1	0.567	60.558	20.7
2	0.538	56.303	20.8
3	0.530	59.524	21.4
4	0.562	61.102	21.2
5	0.483	59.834	21.0
6	0.525	60.228	20.7
7	0.556	60.756	21.5
8	0.586	59.823	20.8
9	0.547	60.153	20.9
10	0.531	60.640	21.2
11	0.581	59.785	21.1
12	0.585	59.675	20.7
13	0.540	60.489	21.2
14	0.458	61.067	21.3
15	0.554	59.788	21.3
16	0.469	58.640	21.5
17	0.471	59.574	20.6
18	0.457	59.718	21.1
19	0.565	60.901	20.8
20	0.664	60.18	20.9
21	0.600	60.493	21.2
22	0.586	58.370	20.9
23	0.567	60.216	20.9
24	0.496	60.214	20.6
25	0.485	59.500	21.7
26	0.573	60.052	20.7
27	0.520	59.501	21.1
28	0.556	58.476	21.4
29	0.539	58.666	21.2
30	0.554	60.239	21.0

Figure 3 shows the obtained charts. The UCLs for the T_S^2, T_{MVE}^2 and T_U^2 charts were 12.275, 30.660, and 25.552, respectively. The Usual and S methods signaled the second observation as out-of-control. In contrast, the MVE method did not.

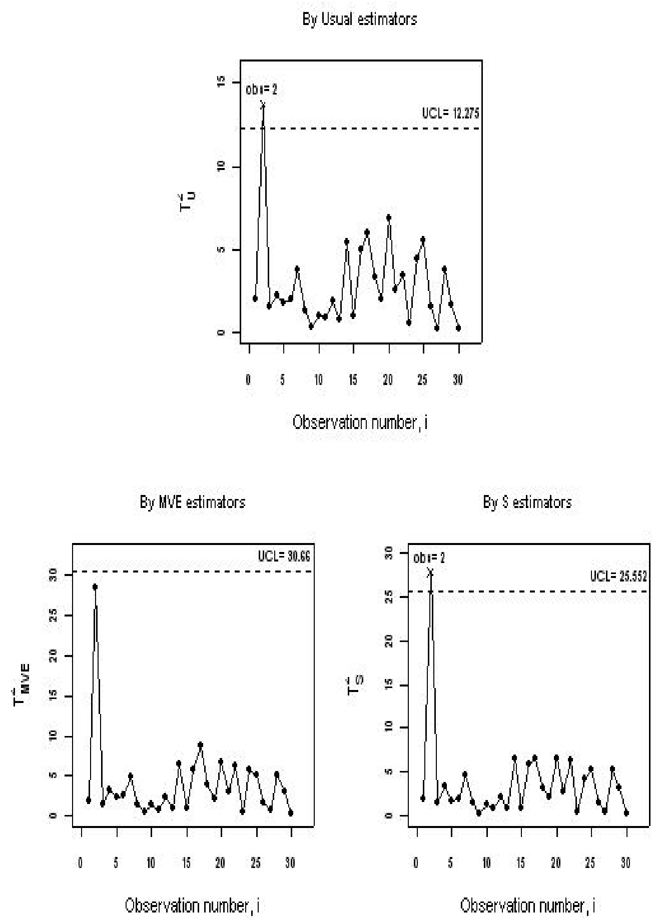

Figure 3. T^2 control charts for the first three variables of [23] data set, using Usual, MVE and S estimators

Next, we replaced the 10^{th} and 25^{th} observations by two outlying observations generated artificially, (0.280, 55.640, 21.2) and (0.485, 55.600, 21.7) respectively. Figure 4 presents the three control charts for the modified data set. The Usual method detected only the 10^{th} observation, whereas the MVE method detected the 10^{th} and 25^{th} observations and the S method detected the 2^{nd}, 10^{th} and 25^{th} observations as out-of-control.

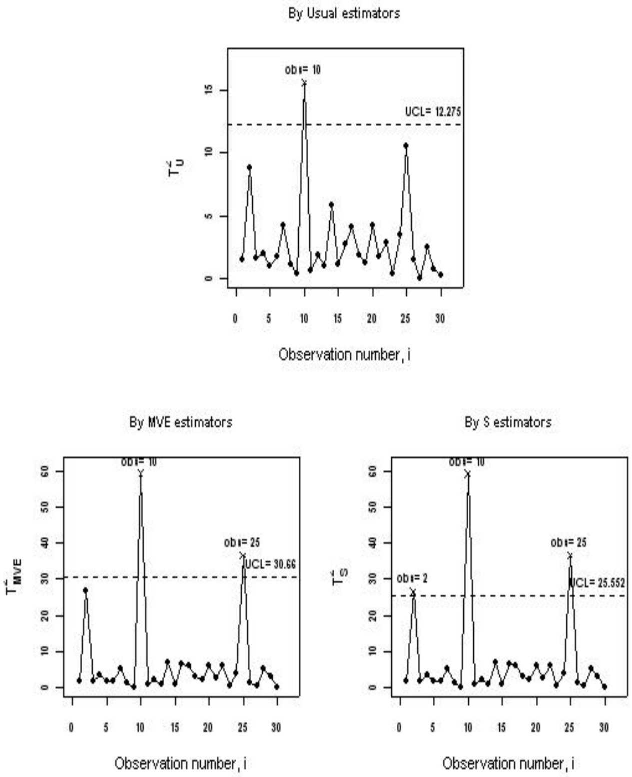

Figure 4. T^2 control charts for the modified data using Usual, MVE and S estimators

7. CONCLUSIONS

During a Phase I analysis of a historical data set, T^2 control charts based on the Usual estimators of the mean vector and covariance matrix of the in-control distribution perform poorly under the occurrence of multiple outliers. We have proposed a T^2 control chart that relies on S estimators. Three control charts were compared via simulation: The usual T^2 control chart, the T^2 control chart with MVE estimators and T^2 control chart with S estimators. Our results show that for a small number of observations, control charts with S estimators perform uniformly better than the other two charts. As the number of observations increases, T^2 control charts based on S and MVE estimators perform similarly. In any case, under the presence of outliers, robust control charts should be used instead of the Usual T^2 control chart.

REFERENCES

[1] LOWRY, C. A, AND MONTGOMERY, D. C. A review of Multivariate Control Charts. IIE Transactions, 27, pp. 800-810, 1995.

[2] MASON, R. L, AND YOUNG, J. C. Multivariate Statistical Process Control with Industrial Applications. SIAM, Philadelphia, P.A, 2002.

[3] KIM, K.; MAHMOUD, M. A, AND WOODALL, W. H. On the Monitoring of Linear Profiles. Journal of Quality Technology, 35, pp. 317-328, 2003.

[4] MAHMOUD, M. A.; AND WOODALL, W. H. Phase I Analysis of Linear Profiles with Calibration Applications. Technometrics, 46 (4), pp. 380-391, 2004.

[5] ALT, F. B, AND SMITH, N. D. Multivariate Process Control. In: Handbook of Statistics, vol. 7, edited by P. R. Krishnaiah and C. R. Rao North-Holland, Amsterdam, pp. 333-351, 1988.

[6] RYAN, T. P. Statistical methods for quality improvement, 2nd ed., John Wiley & Sons, Inc, 2000.

[7] SULLIVAN, J. H.; AND WOODALL, W. H. A Comparison of Multivariate Control Charts for Individual Observations. Journal of Quality Technology, 26, pp. 398-408, 1996

[8] SULLIVAN, J. H.; AND WOODALL, W. H. Adapting Control Charts for the preliminary Analysis of Multivariate Observations. Communications in Statistics – Simulation and Computation, 27, pp. 953-979, 1998.

[9] VARGAS, J. A. Robust Estimation in Multivariate Control Charts for Individual Observations. Journal of Quality Technology, 35, pp. 367-376, 2003.

[10] HOLMES, D. S.; AND MERGEN, A. E. Improving the Performance of the T2 Control Chart. Quality Engineering, 5, pp. 619-625, 1993.

[11] WILLIAMS, J. D.; WOODALL, W. H.; BIRCH, J. B, AND SULLIVAN, J. H. On the distribution of Hotelling's T^2 statistic based on the successive differences covariance matrix estimator. Journal of Quality Technology, 38, pp.217-229, 2006.

[12] ROUSSEEUW, P. J.; AND VAN ZOMEREN, B. C. Unmasking Multivariate Outliers and Leverage Points. Journal of the American Statistical Association, 85, pp. 633-639, 1990.

[13] ROUSSEEUW, P. J.; AND VAN DRIESSEN, K. A fast Algorithm for the Minimum Covariance Determinant Estimator. Technometrics, 41, pp. 212-223, 1999.

[14] JENSEN, W. A.; BIRCH, J. B.; AND WOODALL, W. H. High Breakdown Estimation Methods for Phase I Multivariate Control Charts. Quality and Reliability Engineering International, 23 (5), pp. 615-629, 2007.

[15] YAÑÉZ, S.; VARGAS, J. A. AND GONZÁLEZ, N. Carta T2 con base en estimadores robustos de los parámetros, Revista Colombiana de Estadística, 26(2), pp. 159-179, 2003.

[16] WOODRUFF, D. L.; AND ROCKE, D. M. Computable Robust Estimation on Multivariate Location and Shape in High Dimension Using Compound Estimator. Journal of the American Statistical Association, 89, pp. 888-896, 1994.

[17] ROUSSEEUW, P. J.; AND LEROY, A. M. Robust Regression and Outlier Detection. John Wiley & Sons, Inc, 1987.

[18] DAVIES, P. L. Asymptotic Behaviour of S Estimates of Multivariate Location Parameters and Dispersion Matrices. The Annals of Statistics, 15, pp. 1269-1292, 1987.

[19] LOPUHAÄ, H. P. On the Relation Between S - Estimators and M - Estimators of Multivariate Location and Covariance. The Annals of Statistics, 17, pp. 1662-1683, 1989.

[20] ROCKE, D. M. Robustness Properties of S Estimators of Multivariate Location and Shape in High Dimension. The Annals of Statistics, 24, pp. 1327-1345, 1996.

[21] ROCKE, D. M.; AND WOODRUFF, D. L. Identification of Outliers in Multivariate Data. Journal of the American Statistical Association, 91, pp. 1047-1061, 1996.

[22] KOSINSKI, A. S. A Procedure for the Detection of Multivariate Outliers. Computational Statistics & Data Analysis, 29, pp. 145-161, 1999.

[23] QUESENBERRY, C. P. The Multivariate Short-Run Snapshot Q Chart. Quality Engineering, 13, pp. 679-683, 2001.

[24] MARONNA, R. A.; AND YOHAI, V. J. The Behavior of the Stahel – Donoho Robust Multivariate Estimator. Journal of the American Statistical Association, 90, pp. 330-341, 1995.

APPENDIX

Following [16], this appendix describes an algorithm that computes S estimators. S estimators correspond to the global minimum of the objective function using MCD estimators as initial solutions, $\mathbf{t}^{(0)}$ and $\mathbf{C}^{(0)}$. The intermediate steps of the algorithm are as follows:

(a) Set j=j+1.

(b)Compute

$$d_i^{(j)} = \sqrt{\left(\mathbf{x}_i - \mathbf{t}^{(j-1)}\right)^t \left(\mathbf{C}^{(j-1)}\right)^{-1} \left(\mathbf{x}_i - \mathbf{t}^{(j-1)}\right)}.$$

(c) Find $k^{(j)}$ as a solution of $\frac{1}{m}\sum_{i=1}^{m} \rho\left(d_i^{(j)} / k^{(j)}\right) = b_0$, where

$$\rho(d,c) = \begin{cases} d^2/2 - d^4/\left(2c^2\right) + d^6/\left(6c^4\right) & 0 \le d \le c \\ c^2/6 & d > c \end{cases}$$

is the Tukey's biweight function.

(d) Compute $\tilde{d}_i^{(j)} = d_i^{(j)} / k^{(j)}$.

(e) Set $\mathbf{t}^{(j)} = \dfrac{\sum_{i=1}^{m} w\left(\tilde{d}_i^{(j)}\right)\mathbf{x}_i}{\sum_{i=1}^{m} w\left(\tilde{d}_i^{(j)}\right)}$.

(f) Set $\mathbf{C}^{(j)} = \dfrac{\sum_{i=1}^{m} w\left(\tilde{d}_i^{(j)}\right)\left(\mathbf{x}_i - \mathbf{t}^{(j)}\right)\left(\mathbf{x}_i - \mathbf{t}^{(j)}\right)^t}{\sum_{i=1}^{m} v\left(\tilde{d}_i^{(j)}\right)}$,

where

$$w = \begin{cases} \left(1-\left(d/c\right)^2\right)^2, & 0 \le d \le c \\ 0 & d > c \end{cases}$$

and

$$v = \begin{cases} d^2\left(1-\left(d/c\right)^2\right)^2, & 0 \le d \le c \\ 0 & d > c \end{cases}.$$

The constants b_0 and c are set such that the breakdown point of S estimators is close to 0.5 [17].

At the end of the iterative process (after j steps, according to some convergence criterion) the pair $\mathbf{t}^{(j)}$, $\mathbf{C}^{(j)}$ is obtained. Following [24], the following reweighting is done:

(a) Compute $d_i^2 = \left(\mathbf{x}_i - \mathbf{t}^{(j)}\right)^t \left(\mathbf{C}^{(j)}\right)^{-1} \left(\mathbf{x}_i - \mathbf{t}^{(j)}\right)$.

(b) Compare each d_i^2 with $u = (1+15(m-p))^2 \chi^2_{(0.95,p)} med\{d_i^2\} / \chi^2_{(0.5,p)}$,

where $\chi^2_{(0.95,p)}$ and $\chi^2_{0.5,p}$ are the 0.95 and 0.5 quantiles of a chi-square distribution with p degrees of freedom and $med\left\{d_i^2\right\}$ is the median of the d_i^2 distances.

(c) If $d_i^2 \le u$ then we assign it a weight $\omega_i = 1$. Otherwise, $\omega_i = 0$.

(e) The final estimators are calculated as

$$\mathbf{t} = \left(\sum_{i=1}^{m} \omega_i\right)^{-1} \left(\sum_{i=1}^{m} \omega_i \mathbf{x}_i\right)$$

and

$$\mathbf{C} = \left(\sum_{i=1}^{m} \omega_i - 1\right)^{-1} \sum_{i=1}^{m} \omega_i \left(\mathbf{x}_i - \mathbf{t}\right)\left(\mathbf{x}_i - \mathbf{t}\right)^t.$$