

FAST AND ROBUST BIVARIATE CONTROL CHARTS FOR INDIVIDUAL OBSERVATIONS

Sajesh T A

Department of Statistics, St. Thomas College (Autonomous), Thrissur, Kerala, India
*sajesh.t.abraham@gmail.com

Abstract

There are various circumstances where it is important to simultaneously monitor or control two or more related quality characteristics. Independently tracking these quality characteristics might be quite deceptive. Hotelling's T² chart, in which the T² statistics are generated using the classical estimates of location and scatter, is the most well-known multivariate process monitoring and control approach. It is well known that the existence of outliers in a dataset has a significant impact on classical estimators. Any statistic that is computed using the classical estimates will be distorted by even a single outlier. The non-robustness issue is investigated in this study, which also suggests four robust bivariate control charts based on the robust Gnanadesikan-Kettenring estimator. This study employs four highly robust scale estimators, with the best breakdown point, namely the Qn estimator, Sn estimator, MAD estimator, and τ estimator, in order to robustify the Gnanadesikan-Kettenring estimator. Through the use of a Monte Carlo simulation and a real-life data, the performance of the suggested control charts is assessed. The four techniques all outperform the traditional method and provide greater computing efficiency.

Keywords: Gnanadesikan- Kettenring estimator, Qn estimator, Sn estimator, MAD, τ estimator.

1. Introduction

Bivariate control charts are specifically designed for situations where two variables are observed simultaneously. It enables the detection of patterns or trends that signal a shift or alteration in the process. There are two separate phases, namely Phase I and Phase II when constructing the control chart [1]. Historical data is used in Phase I to determine control limits, estimate the unknown parameters of the in-control process, and evaluate the process' stability. Phase II involves applying the estimated parameters and control limits discovered in Phase I to the data gathered during the actual production process in order to analyse it and find any deviations or out-of-control signals. Phase II's goal is to keep track of and maintain the process' stability in accordance with the defined control limits.

The most frequently used multivariate control chart for monitoring the variability of a multivariate industrial process is the Hotelling's T² control chart. Let \mathbf{x}_i , $i = 1, 2, \dots, n$ be a two-dimensional vector of measurements made on a process at the time period i , then for the sample $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, the Hotelling's T² statistic is defined as

$$T^2(\mathbf{x}_i) = (\mathbf{x}_i - \boldsymbol{\theta})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\theta}). \quad (1)$$

It is assumed that, when the process is in statistical control, \mathbf{x}_i 's are independent bivariate normal random vectors with mean vector $\boldsymbol{\theta}$ and covariance matrix. If both $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ are known, T_i^2 follows a Chi-square distribution with 2 degrees of freedom. When the population parameters $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ are unknown, the T² statistic is constructed using the classical estimators of mean vector ($\bar{\mathbf{x}}$) and covariance matrix (S) as given below,

$$T^2(\mathbf{x}_i) = (\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{S}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}), \tag{2}$$

where $\bar{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix}$ and $\mathbf{S} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$, $S_{12} = Cov(X_1, X_2)$. When constructing control charts using phase I data, the classical estimators, which are unfortunately highly susceptible to the influence of outliers, can produce inaccurate findings (known as the masking problem). Researchers have suggested several methods in the literature to lessen the negative effects of outliers in response to this problem. The control charts can be made more robust and reliable in the presence of outliers by using robust estimators. Through the use of robust estimators, which are more resistant to the presence of outliers, the conventional estimators are swapped out in these alternative methods.

This paper proposes a robust bivariate control chart that can effectively handle spurious outliers. The proposed control chart makes use of the covariance estimator introduced by Gnanadesikan and Kettenring [2], which is based on the identity

$$cov(X, Y) = \frac{1}{4} (\sigma(X + Y)^2 - \sigma(X - Y)^2), \tag{3}$$

where σ is the standard deviation and X, Y is a pair of random variables. By replacing σ by a robust scale estimator, one can easily robustify the Gnanadesikan and Kettenring (GK) estimator. The robust GK estimator is defined as

$$cov_R(X, Y) = \frac{1}{4} (s_R(X + Y)^2 - s_R(X - Y)^2), \tag{4}$$

where s_R is a robust scale estimator. This study has considered four robust scale estimators with an optimal breakdown point to robustify the GK estimators. These robust GK estimators were then used for the construction of bivariate control charts. The Lower control limits (LCL) of these control charts are set to zero and the Upper Control limits (UCL) were estimated by fitting the quantiles. The performance of these control charts is examined and compared with that of classical bivariate control chart through Monte Carlo simulation.

2. Robust scale estimators

For a variety of applications, from genuine scale problems to outlier identification, and as auxiliary factors for more complicated analysis, robust estimates of scale are crucial. The broad population of users of statistical methods seems to have a somewhat lower level of acceptance for robust estimation of scale. Previously, the interquartile range, which has a breakdown point of 25%, was the only robust scale estimator to be found in the majority of statistical software packages.

i) Median Absolute Deviation (MAD)

A preliminary or auxiliary estimate of scale is frequently required in robust estimation. A very robust scale estimator is the median absolute deviation about the median (MAD), given by

$$MAD = b \text{ med}_i \{ |x_i - \text{med}_j(x_j)| \}, \tag{4}$$

where 'med' denotes median. The MAD has the best possible breakdown point (50%, twice as much as the interquartile range), and its influence function is bounded, with the sharpest possible bound among all scale estimators. The MAD was first promoted by Hampel [3], who attributed it to Gauss. The constant b in equation (4) is needed to make the estimator consistent for the parameter of interest. In the case of the usual parameter σ at Gaussian distributions, we need to set $b = 1.4826$. In spite of many advantages, the MAD also has some drawbacks. Its efficiency at Gaussian distributions is very low;

whereas the location median's asymptotic efficiency is still 64%, the MAD is only 37% efficient. Also, it takes a symmetric view on dispersion.

ii) τ Estimator

Yohai and Zamar [4] introduced a new class of robust scale estimates called τ estimates which possesses optimal breakdown value. Let ρ be a real function satisfying the following properties:

1. $\rho(0) = 0$.
2. $\rho(-u) = \rho(u)$.
3. $0 \leq u \leq v$ implies that $\rho(u) \leq \rho(v)$.
4. ρ is continuous.
5. Let $a = \text{Sup } \rho(u); 0 < a < \infty$
6. If $\rho(u) < a$ and $0 \leq u < v$, then $\rho(u) < \rho(v)$.
- 7.

Let ρ_1 and ρ_2 be two functions satisfying above assumptions. Then for a given sample $\mathbf{u} = (u_1, \dots, u_n)$ the τ estimate for scale is defined as

$$\tau^n(\mathbf{u}) = s^2(\mathbf{u}) \frac{1}{n} \sum_{i=1}^n \rho_2\left(\frac{u_i}{s(\mathbf{u})}\right), \tag{5}$$

where s be a M estimate of scale based on ρ_1 . This estimator possesses approximately 80% efficiency when $c = 3$. Moreover, τ estimate is asymptotically normal and has bounded influence function. Maronna and Zamar [5] used this estimate for introducing a multivariate outlier detection technique in which they have considered $s = \text{MAD}$ and $\rho_2(x) = \min(x^2, c^2)$. This study also used the considerations of Maronna and Zamar for s and ρ_2 .

iii) S_n and Q_n Estimator

To address the lower efficiency drawback of MAD, Rousseeuw and Croux [6] introduced two robust scale estimators with optimal breakdown value of 50%, namely, S_n estimator and Q_n estimator. S_n estimator is defined as

$$S_n = c \text{med}_i \{ \text{med}_j |x_i - x_j| \}. \tag{6}$$

The factor c is for consistency, and its default value is 1.1926. Moreover, the asymptotic efficiency S_n is 58.23% which is much higher than MAD. A drawback of MAD, and S_n , is that their influence functions have discontinuities. The Q_n estimator is solution for this drawback. It is defined as

$$Q_n = d \{ |x_i - x_j|; i < j \}_{(k)}, \tag{7}$$

where d is a constant factor and $k = \binom{h}{2} \approx \frac{\binom{n}{2}}{4}$, where $h = \lfloor \frac{n}{2} \rfloor + 1$ is roughly half of the observations. The estimator Q_n , shares the attractive properties of S_n simple and explicit formula, a definition that is equally suitable for asymmetric distributions, and a 50% breakdown point. In addition, we will see that its influence function is smooth, and that its efficiency at Gaussian distributions is very high (about 82%). Rousseeuw and Croux [6] showed that, although Q_n is more efficient, S_n is more preferable in most of the applications because of its low gross-error sensitivity.

3. Proposed Bivariate Robust Control Charts

Let $\{x_1, \dots, x_n\}$ be a set of Phase I data follows bivariate normal distribution with mean vector θ and covariance matrix Σ . Let $\mathbf{y} \notin \{x_1, \dots, x_n\}$ be a Phase II observation, then it is known that

$$T^2(\mathbf{y}) \sim \left[\frac{2(n^2-1)}{n(n-2)} \right] F_{(2,n-2)} \tag{8}$$

where $T^2(\mathbf{y})$ is as defined in Equation (1) and $F_{(v_1,v_2)}$ is F distribution with (v_1, v_2) degrees of freedom [7]. Since this statistic is not robust to the presence of outliers, in the proposed robust control chart we use robust Hotelling's T^2 statistic, denoted by T_R^2 . It is obtained by replacing the classical estimates used for the computation of Hotelling's T^2 statistic by their robust counterpart. Suppose \mathbf{x}_{med} and \mathbf{S}_{GK} represent the component wise median vector and robust GK covariance matrix, respectively. We define a robust Hotelling's T^2 for \mathbf{y} based on these estimates by

$$T_R^2(\mathbf{y}) = (\mathbf{y} - \mathbf{x}_{med})^T \mathbf{S}_{GK}^{-1} (\mathbf{y} - \mathbf{x}_{med}), \tag{9}$$

where $\mathbf{S}_{GK} = \begin{bmatrix} s_R(X_1)^2 & cov_R(X_1, X_2) \\ cov_R(X_1, X_2) & s_R(X_2)^2 \end{bmatrix}$. By applying Slutsky theorem [8], the asymptotic distribution of T_R^2 can be obtained as Chi square distribution with 2 degrees of freedom. As $n \rightarrow \infty$

$$(\mathbf{y} - \mathbf{x}_{med})^T \mathbf{S}_{GK}^{-1} (\mathbf{y} - \mathbf{x}_{med}) \xrightarrow{D} (\mathbf{y} - \boldsymbol{\theta})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\theta}) \sim \chi_{(2)}^2. \tag{10}$$

This asymptotic distribution, though, only holds true for large sample sizes. We employ Monte Carlo simulations to estimate quantiles for different sample sizes in order to determine the control limits for the suggested control charts. The sample size and quantiles of T_R^2 were then fitted with a smooth curve. For modest Phase I sample sizes, these fits can be utilised to determine the proper control limits of the proposed control charts.

3.1. Estimation of Control Limits for the proposed control charts

Upper control limits of the proposed control charts are obtained by modelling the quantiles of T_R^2 , for a given Phase I sample size n , computed from $N=10000$ trials. Phase I samples are generated from a standard bivariate normal distribution $N_2(\mathbf{0}, \mathbf{I})$ and 99%, 95%, 99.73% and 99.9% quantiles of T_R^2 are computed using S_n, Q_n, MAD and τ estimates. In each trial, for each data set, we also generate a new random observation \mathbf{y}_i from $N_2(\mathbf{0}, \mathbf{I})$ (treated as a Phase II observation) and calculate the corresponding $T_R^2(\mathbf{y}_i)$ value using robust GK covariance estimates. By inverting the empirical distribution function of $T_R^2(\mathbf{y}_i)$, computed for $n = 10, 15, 20, \dots, 500$, we obtain Monte Carlo estimates of the 99%, 95%, 99.73% and 99.9% quantiles.

Scatter plots of the empirical quantiles of $T_R^2(\mathbf{y}_i)$ versus the sample size n suggest that we could model the quantiles using a family of regression curves of the form $f(n) = a + \frac{b}{n^c}$. Scatter plots of the empirical 99% quantiles of $T_R^2(\mathbf{y}_i)$ computed using the four robust scale estimates are shown in figure 1. Since the T_R^2 statistic asymptotically follow $\chi_{(2)}^2$ distribution, following two parameter family of curves is used for both robust GK control charts:

$$f_{1-\alpha}(n) = \chi_{(2,1-\alpha)}^2 + \frac{b_{1-\alpha}}{n^{c_{1-\alpha}}}, \tag{11}$$

where $\chi_{(2,1-\alpha)}^2$ is the $1 - \alpha$ quantile of the χ^2 distribution with 2 degrees of freedom and $b_{1-\alpha}$ and $c_{1-\alpha}$ are constants with overall false alarm probability α . Fitting this curve to the data will help us to estimate the desired upper control limits of the proposed control charts for any Phase I sample of size n . Note that, as n increases, $f_{1-\alpha}(n)$ approaches $\chi_{(2,1-\alpha)}^2$. Table 1 gives the least-square estimates of the parameters $b_{1-\alpha}$ and $c_{1-\alpha}$. Using Table 1 and Equation (11), we can compute the 99%, 95%, 99.73% and 99.9% quantiles of $T_R^2(\mathbf{y}_i)$ for Phase I sample size n . The regression curves given by Equation (9) fit well to all the cases in Table 1, yielding R^2 values of at least 90.7%.

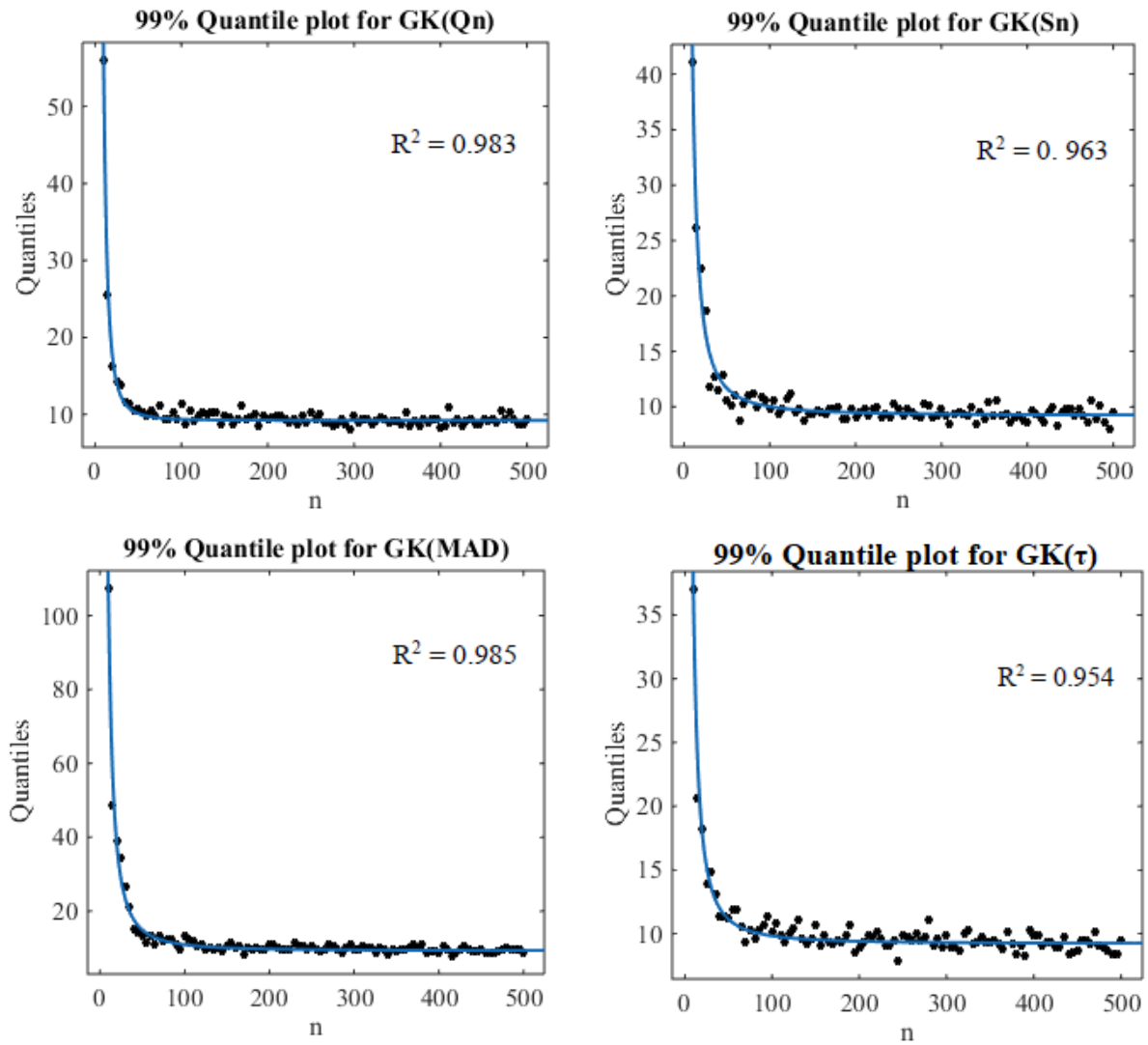


Figure 1: Simulated Quantiles of T_{GK}^2 and the Fitted Curves for $\alpha = 0.01$

Table 1: The Least-Squares Estimates of the Regression Parameters $b_{1-\alpha}$ and $c_{1-\alpha}$ for Confidence Levels $1 - \alpha = 0.99, 0.95, 0.9973$ and 0.999 .

α	Parameters	GK(Q_n)	GK(S_n)	GK(MAD)	GK(τ)
0.01	$b_{1-\alpha}$	13120.00	1281.00	5496.00	1145.00
	$c_{1-\alpha}$	2.451	1.598	1.758	1.632
	R^2	0.983	0.963	0.985	0.954
0.05	$b_{1-\alpha}$	228.80	249.50	462.60	324.20
	$c_{1-\alpha}$	1.437	1.399	1.373	1.447
	R^2	0.916	0.930	0.976	0.942
0.0027	$b_{1-\alpha}$	120600.00	11440000.00	721800.00	2610.00
	$c_{1-\alpha}$	2.945	4.66	3.22	1.656
	R^2	0.987	0.990	0.990	0.907
0.001	$b_{1-\alpha}$	556600.00	26090.00	41240.00	3973.00
	$c_{1-\alpha}$	3.526	1.947	1.769	1.826
	R^2	0.968	0.920	0.950	0.930
0.1	$b_{1-\alpha}$	31530.00	248.30	219.4	118.60
	$c_{1-\alpha}$	1.65	1.574	1.31	1.281
	R^2	0.951	0.919	0.967	0.947

3.2. Computational algorithm for obtaining the proposed control chart

A step-by-step approach for constructing proposed robust control chart is given as follows:

Phase I

1. Select the confidence level $1 - \alpha$ and sample size n .
2. Collect the Phase I data $\{x_1, x_2, \dots, x_n\}$ at predetermined periodic intervals and compute robust estimates of location and scale parameters using this data.
3. Use Table 1 to select the least square estimates of the parameters $b_{1-\alpha}$ and $c_{1-\alpha}$ for the desired α , then use equation (11) to determine the upper control limit.

Phase II

4. Compute T_R^2 for each of the new observation as per Equation (11) and plot it on a control chart with the control limits determined in Phase I.
5. Analyse the chart and look for any out-of-control points or non-random patterns.

4. Performance of the Proposed Control Charts

We carried out numerous simulations to assess and compare the efficiency of the proposed control charts under normal and contaminated situations. The control chart's efficacy is assessed by analysing its ability to detect changes and the rate of false detections in the process behaviour using different estimators in Phase I. We measure the performance of the control chart by *success rate* (SR)- which is the proportion of statistic values that exceed the control limits across 1000 replications, which provides an estimate of the likelihood of detecting changes and *false alarm rate* (FAR) – which is the false detections in the process behaviour based on the Phase II data. Phase I data are generated using the following contaminated model:

$$(1 - \varepsilon) N_2(\theta_0, \Sigma_0) + \varepsilon N_2(\theta_1, \Sigma_1),$$

where ε is the proportion of outliers, θ_0 and Σ_0 are the in-control parameters and θ_1 and Σ_1 are the out-of-control parameters of location and scatter. Without loss of generality, θ_0 is set to be a zero vector. We have generated the phase I data sets for $n = 25, 50, 100$ and 1000 , $\varepsilon = 0, 0.1$ and 0.2 and $\alpha = 0.001, 0.0027, 0.01, 0.05$ and 0.1 . The probability of detecting a change depends on the values of θ_1, Σ_0 and Σ_1 and hence we consider three different contaminated models in which the values of θ_1, Σ_0 and Σ_1 vary.

Case A: Independent Variables

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

$$(1 - \varepsilon) N_2(0, I_2) + \varepsilon N_2(\theta_1, I_2),$$

where I_2 is the identity matrix of size 2. In this case, we compare the behaviour of different robust alternatives when there are different-sized changes in the average of all the variables if the variables are independent.

Case B: Correlated Variables

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

$$(1 - \varepsilon) N_2(\mathbf{0}, \Sigma_0) + \varepsilon N_2(\theta_1, \Sigma_0),$$

where $\Sigma_0 = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$. We have used this value of Σ_0 to analyse whether the correlation level affects the detection probability of each alternative.

Table 2: SR (FAR) obtained for Case A with different values of θ_1 and ε

n	θ_1	ε	GK(Q_n)	GK(S_n)	GK(MAD)	GK(τ)	Classical
50	(0,0)	0	100 (1.9)	100 (1.0)	99.8 (0.8)	100 (1.3)	100 (1.3)
		10	100 (1.4)	100 (1.0)	99.9 (0.4)	100 (1.1)	100 (1.1)
		20	100 (1.3)	100 (0.9)	99.8 (0.6)	100 (0.6)	100 (0.6)
	(5,5)	0	100 (2.0)	100 (1.6)	99.7 (1.3)	100 (1.2)	100 (1.5)
		10	98.8 (0.3)	98.8 (0.3)	97.7 (0.4)	98.6 (0.1)	22.1 (0.2)
		20	86.2 (0.0)	80 (0.0)	76.8 (0.0)	74.3 (0.1)	2 (0.2)
	(10,10)	0	100 (1.1)	99.9 (0.3)	99.7 (0.5)	100 (1.7)	100 (0.8)
		10	99.4 (0.3)	99.1 (0.1)	97.9 (0.3)	99.1 (0.4)	0.6 (0.4)
		20	88.3 (0.0)	80.8 (0.0)	79.8 (0.1)	75.8 (0.2)	0.4 (0.3)
100	(0,0)	0	100 (1.2)	100 (1.0)	100 (1.0)	100 (1.2)	100 (1.2)
		10	100 (1.5)	100 (1.4)	100 (1.3)	100 (1.3)	100 (1.3)
		20	100 (1.4)	100 (1.3)	100 (1.2)	100 (0.7)	100 (0.7)
	(5,5)	0	100 (0.6)	100 (0.7)	100 (0.9)	100 (1.5)	100 (0.6)
		10	99.7 (0.2)	99.6 (0.1)	99.8 (0.3)	99.8 (0.4)	30.2 (0.4)
		20	92.5 (0.0)	90.4 (0.0)	93.5 (0.0)	83.9 (0.0)	2.5 (0.1)
	(10,10)	0	100 (1.5)	100 (1.2)	99.9 (0.7)	100 (0.7)	100 (1.1)
		10	99.8 (0.2)	99.8 (0.3)	99.8 (0.5)	99.6 (0.4)	0.1 (0.2)
		20	91.1 (0.1)	89.4 (0.1)	92.5 (0.1)	85.8 (0.2)	0.4 (0.4)
500	(0,0)	0	100 (1.9)	100 (1.9)	100 (1.8)	100 (1.8)	100 (1.8)
		10	100 (1.5)	100 (1.4)	100 (1.4)	100 (1.4)	100 (1.4)
		20	100 (0.8)	100 (0.9)	100 (0.9)	100 (0.8)	100 (0.8)
	(5,5)	0	100 (1.4)	100 (1.2)	100 (1.4)	100 (1.2)	100 (1.2)
		10	99.8 (0.0)	99.8 (0.0)	99.9 (0.0)	99.8 (0.1)	34.5 (0.0)
		20	94.8 (0.0)	95.5 (0.0)	98.1 (0.0)	89.3 (0.0)	2.4 (0.3)
	(10,10)	0	100 (0.6)	100 (0.7)	100 (0.8)	100 (1.8)	100 (0.6)
		10	99.8 (0.1)	99.9 (0.1)	99.9 (0.2)	99.9 (0.2)	0.7 (0.1)
		20	94.6 (0.0)	95.1 (0.0)	97.8 (0.1)	88.1 (0.0)	0.4 (0.2)
1000	(0,0)	0	100 (1.0)	100 (0.9)	100 (0.8)	100 (0.9)	100 (0.9)
		10	100 (0.8)	100 (0.8)	100 (0.9)	100 (0.8)	100 (0.8)
		20	100 (0.6)	100 (0.5)	100 (0.5)	100 (0.7)	100 (0.7)
	(5,5)	0	100 (0.8)	100 (0.8)	100 (1.0)	100 (0.7)	100 (0.6)
		10	99.9 (0.2)	99.9 (0.2)	100 (0.3)	100 (0.0)	34.9 (0.3)
		20	94.9 (0.0)	95.8 (0.1)	98.3 (0.1)	89.5 (0.0)	2 (0.4)
	(10,10)	0	100 (1.8)	100 (1.9)	100 (1.7)	100 (0.8)	100 (1.8)
		10	99.8 (0.1)	99.8 (0.1)	100 (0.2)	99.8 (0.1)	0.8 (0.2)
		20	93.8 (0.0)	94.7 (0.0)	98.3 (0.1)	89 (0.0)	0.3 (0.2)

Case C: Correlated Variables and Regression Outliers

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

$$(1 - \varepsilon) N_2(\mathbf{0}, \Sigma_0) + \varepsilon N_2(\theta_1, \Sigma_1),$$

where $\Sigma_0 = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$ and $\Sigma_1 = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}$. In this case, we analyse and compare the proposed robust methods in terms of the so-called good leverage and regression outliers.

In all the three cases, we consider θ_1 as a vector of size 2 where the elements are all 0 (when there is no change), 5 or 10 (which is a good leverage point). This process is repeated 1000 times, and in each trial, a random observation, z_{1k} from $N_2(\mathbf{0}, \Sigma_u)$ and another observation, z_{2k} from $N_2(\theta_c, \Sigma_c)$ are generated, for $k = 1, 2, \dots, 1000$. Here, Σ_u is the scale estimator used for generating uncontaminated observations in Phase I data, and θ_c and Σ_c are the location and scale estimates used for generating contaminated observations in Phase I model. The success rates are computed as the percentages of z_{2k} 's that are successfully detected, and the false alarm rates are computed as the percentages of z_{1k} 's that are falsely detected. The results obtained for $\alpha = 0.01$ is presented here.

Table 3: SR (FAR) obtained for Case B with different values of θ_1 and ϵ

n	θ_1	ϵ	GK(Q_n)	GK(S_n)	GK(MAD)	GK(τ)	Classical
50	(0,0)	0	95.6 (4.1)	90.6 (4.6)	84.6 (2.7)	97 (2.8)	97.4 (0.8)
		10	95.6 (4.3)	91.4 (3.2)	84.6 (2.7)	96.3 (2.7)	96.3 (0.9)
		20	95.5 (3.4)	90.9 (4.8)	82.4 (1.7)	95.7 (2.6)	96.4 (1.4)
	(5,5)	0	94.7 (4.5)	91.2 (3.8)	83.6 (3.0)	96.9 (1.4)	97 (0.9)
		10	84.7 (3.0)	77.1 (2.9)	67.2 (1.7)	79.7 (3.1)	17.9 (0.4)
		20	57 (2.3)	48.2 (1.9)	40.5 (1.7)	40.2 (1.1)	2.3 (0.2)
	(10,10)	0	95.3 (3.4)	91.2 (3.5)	82.4 (2.8)	97.1 (2.5)	98.3 (1.2)
		10	84.4 (3.9)	76.2 (1.9)	67.4 (2.6)	80.3 (2.6)	0.7 (0.2)
		20	57.1 (2.5)	46.3 (1.3)	43.8 (0.8)	41.1 (2.1)	0.8 (0.4)
100	(0,0)	0	98.6 (3.3)	96.3 (4.7)	94.2 (3.9)	98.6 (2.4)	98.4 (1.3)
		10	97.9 (2.0)	95 (3.7)	92.2 (3.4)	97.6 (2.2)	97.4 (1.1)
		20	98.5 (1.8)	95.8 (3.4)	92.4 (3.5)	97.9 (0.9)	97.8 (0.4)
	(5,5)	0	98.5 (2.7)	96 (4.8)	93.2 (3.6)	98.1 (0.9)	98.1 (1.0)
		10	90.4 (1.5)	86.9 (2.6)	86.7 (3.1)	85.8 (2.0)	23.5 (0.9)
		20	63.3 (0.9)	56.8 (2.0)	61.4 (2.7)	49.6 (1.9)	2.8 (0.2)
	(10,10)	0	98.8 (2.1)	96 (3.7)	93 (3.6)	98.5 (1.7)	98.7 (1.1)
		10	91.2 (0.7)	87.4 (2.2)	87.6 (2.6)	88.2 (2.1)	0.8 (0.3)
		20	62.4 (0.7)	57.5 (2.5)	60.1 (2.0)	51.2 (1.6)	0.7 (0.6)
500	(0,0)	0	99.2 (1.0)	99 (2.2)	98.8 (2.8)	99.3 (1.2)	99.2 (1.0)
		10	99.1 (1.1)	98.9 (2.7)	98.1 (3.5)	99 (1.1)	99 (1.1)
		20	98.6 (1.3)	98.6 (2.6)	97.8 (3.5)	98.7 (1.5)	98.6 (1.5)
	(5,5)	0	98.6 (1.1)	98.7 (2.9)	98.3 (3.9)	99 (1.2)	98.4 (1.1)
		10	90.8 (0.7)	91.9 (2.6)	94 (3.2)	90 (0.7)	25.6 (0.6)
		20	62.4 (0.0)	63.8 (0.8)	74.8 (1.6)	53.5 (0.8)	3.1 (0.3)
	(10,10)	0	98.3 (1.7)	98.2 (2.2)	97.6 (3.8)	98.5 (1.4)	98.4 (1.7)
		10	90.9 (0.1)	92.4 (1.5)	93.1 (2.5)	91.1 (0.4)	0.7 (0.2)
		20	64.6 (0.0)	66.6 (0.9)	77.1 (1.4)	55 (0.3)	0.6 (0.4)
1000	(0,0)	0	99 (0.2)	98.9 (0.5)	98.7 (1.3)	99 (0.3)	98.9 (0.2)
		10	98.1 (1.3)	98.1 (1.6)	97.8 (1.7)	98 (1.1)	97.9 (1.1)
		20	99.1 (0.6)	98.9 (1.6)	98.8 (2.1)	99.1 (0.8)	98.9 (0.7)
	(5,5)	0	98.8 (1.3)	98.7 (1.5)	98.7 (2.6)	98.7 (1.2)	98.8 (1.1)
		10	91.3 (0.0)	92.4 (0.5)	93.9 (2.0)	91.8 (0.5)	25.8 (0.2)
		20	64.7 (0.0)	66.7 (0.1)	76.4 (1.1)	53.4 (0.3)	3.2 (0.6)
	(10,10)	0	98.3 (1.6)	98.1 (2.1)	98 (2.7)	98.7 (1.2)	98.3 (1.6)
		10	90.1 (0.1)	91.8 (0.4)	93.8 (1.4)	90.8 (0.4)	0.5 (0.2)
		20	64.8 (0.0)	67.2 (0.0)	77.3 (1.1)	53.1 (0.3)	0.7 (0.1)

Table 2 presents the results obtained from simulations conducted for Case A samples. The findings clearly indicate that when the phase I data contains outliers, the control charts based on robust methods exhibit a high success rate with minimal FAR compared to the classical control chart. Furthermore, the success rates of the proposed control charts improve with increasing sample size. Conversely, when the phase I data is uncontaminated, the robust control charts demonstrate similar performance to their classical counterpart.

Table 4: SR (FAR) obtained for Case C with different values of θ_1 and ϵ

n	θ_1	ϵ	GK(Q_n)	GK(S_n)	GK(MAD)	GK(τ)	Classical
50	(0,0)	0	97.9 (4.7)	96.2 (4.9)	91.6 (2.3)	99.6 (2.4)	100 (1.2)
		10	98.8 (3.8)	97.1 (3.1)	95.3 (2.4)	99.3 (2.7)	100 (1.3)
		20	99 (4.1)	98.8 (4.1)	96.5 (3.9)	99.1 (4.3)	100 (1.8)
	(5,5)	0	98.3 (3.7)	96.8 (3.3)	90 (2.7)	99.2 (1.9)	100 (0.6)
		10	95.1 (2.1)	85.6 (2.8)	75.9 (2.1)	91.5 (3.7)	8.4 (0.2)
		20	78.6 (0.2)	45.9 (1.3)	38 (1.1)	36.7 (2.0)	0.9 (0.3)
	(10,10)	0	98.8 (5.2)	94.9 (4.0)	90.8 (2.9)	99.1 (2.6)	100 (1.0)
		10	96.3 (3.2)	87.8 (2.5)	77.4 (1.7)	92.7 (2.9)	0.8 (0.6)
		20	76.6 (1.7)	47.6 (1.4)	37.1 (1.0)	38 (1.5)	0.6 (0.4)
100	(0,0)	0	100 (2.5)	98.2 (3.3)	96 (3.7)	99.9 (1.6)	100 (1.0)
		10	100 (2.1)	99.2 (2.6)	97.7 (3.5)	99.9 (1.7)	100 (0.6)
		20	99.9 (2.8)	99.7 (3.1)	99.5 (5.2)	99.8 (3.3)	100 (1.9)
	(5,5)	0	99.8 (2.6)	98.4 (2.9)	95.7 (2.5)	100 (1.6)	100 (0.5)
		10	99.7 (0.6)	94.6 (3.5)	93.3 (2.0)	98.4 (2.7)	10.6 (0.1)
		20	91.2 (0.1)	62.4 (2.3)	68.4 (1.6)	47.4 (1.3)	1.2 (0.4)
	(10,10)	0	99.9 (3.5)	98 (4.5)	96.4 (4.3)	100 (1.9)	100 (1.5)
		10	99.9 (0.7)	96.1 (3.4)	94 (3.2)	98 (2.7)	0.6 (0.1)
		20	92.9 (0.2)	65.5 (1.8)	68.9 (1.5)	50.6 (1.5)	0.2 (0.1)
500	(0,0)	0	100 (1.0)	99.9 (1.9)	99.4 (3.4)	100 (1.0)	100 (0.8)
		10	100 (1.5)	100 (1.8)	99.8 (3.5)	100 (1.6)	100 (1.3)
		20	100 (2.1)	100 (2.4)	100 (3.8)	100 (2.1)	100 (1.7)
	(5,5)	0	100 (1.3)	100 (2.5)	99.9 (3.3)	100 (1.7)	100 (1.4)
		10	100 (0.4)	99.6 (1.5)	98.7 (2.3)	100 (0.5)	11 (0.6)
		20	98.9 (0.0)	85.3 (1.0)	95.7 (1.0)	56.4 (1.4)	1.8 (0.0)
	(10,10)	0	100 (1.3)	99.9 (2.5)	99.6 (3.3)	100 (1.5)	100 (1.4)
		10	100 (0.2)	99.8 (0.8)	98.9 (2.6)	100 (0.4)	0.9 (0.4)
		20	99.2 (0.0)	87.9 (1.2)	95.4 (1.9)	58.8 (1.0)	0.4 (0.2)
1000	(0,0)	0	100 (1.1)	100 (1.5)	99.9 (1.9)	100 (1.3)	100 (1.3)
		10	100 (1.7)	100 (2.0)	100 (3.3)	100 (2.0)	100 (1.7)
		20	100 (1.9)	100 (2.5)	100 (3.5)	100 (1.7)	100 (1.7)
	(5,5)	0	100 (0.9)	100 (1.2)	99.9 (1.8)	100 (0.8)	100 (0.8)
		10	100 (0.3)	99.9 (0.8)	100 (1.2)	100 (0.5)	10 (0.3)
		20	99.7 (0.0)	89.9 (0.3)	98.9 (0.9)	57.1 (0.4)	2 (0.3)
	(10,10)	0	100 (1.0)	100 (1.0)	100 (2.2)	100 (1.3)	100 (1.0)
		10	100 (0.1)	100 (0.8)	99.8 (1.6)	100 (0.0)	0.6 (0.2)
		20	99.9 (0.0)	88.2 (0.1)	98.5 (1.1)	59.4 (0.0)	0.2 (0.1)

Results obtained for Case B samples are presented in table 3. It is clear that the proposed control charts perform well with respect to SR and FAR in almost all cases. Even though, their performance is not very satisfactory in cases of small samples with high contamination, their SR and FAR are getting better with increasing sample size.

In all the three cases the proposed robust GK based control charts are performing better than the classical control chart in contaminated situations and they show similar performance as that of classical chart in uncontaminated cases. It is interesting to see that, among the robust control charts, the one based on $GK(Q_n)$ shows better success rate than other robust control charts when $n < 500$ while the control chart based on $GK(MAD)$ outperform others for large samples when the data from Case A and Case B. But, when the data come from Case C, in all the cases $GK(Q_n)$ based control chart shows superior performance.

4.1. Time Complexity

The proposed methods' significantly faster computation times are a key benefit. To compare the computation times of these approaches for various sample sizes, a simulation study was done. The simulation study consists of 10,000 trials, and the average running time is presented in table 5. Among the robust methods, the control chart based on $GK(S_n)$ is multiple times faster than the other three methods when $n \leq 500$ while $GK(MAD)$ -based control chart performs faster than the other compared methods when $n > 500$.

Table 5: Running time (in seconds)

n	$GK(Q_n)$	$GK(S_n)$	$GK(MAD)$	$GK(\tau)$
50	0.000128	0.000080	0.000207	0.000361
100	0.000222	0.000093	0.000231	0.000361
500	0.001065	0.000290	0.000304	0.000516
1000	0.002231	0.000526	0.000397	0.000992

4.2. Real life data

A data set given by Quesenberry [9] has been used to evaluate the performance of the proposed methods in real life data. The original data consists of 11 quality variables measured on 30 products from a production process. For our comparison purposes, we consider the third and fourth variables as our bivariate data. Bivariate control charts using the proposed robust methods and the classical methods are developed for this data and presented in figure 3. From charts it is clear that the data is outlier-free, and none of the methods, including classical method, commit any false detection. In order to evaluate performance in a contaminated situation, we artificially created two outlying observations. Observations 7 and 16 are changed from (21.5, 5.08) and (21.5, 15.32), to (22.75, 5.08) and (22.75, 15.32) respectively, by adding a very small shift of 1.25 in the first variable. Control charts of the contaminated data set are given in figure 3, and it is clear that except for the classical control chart, all the proposed control charts detected these outliers.

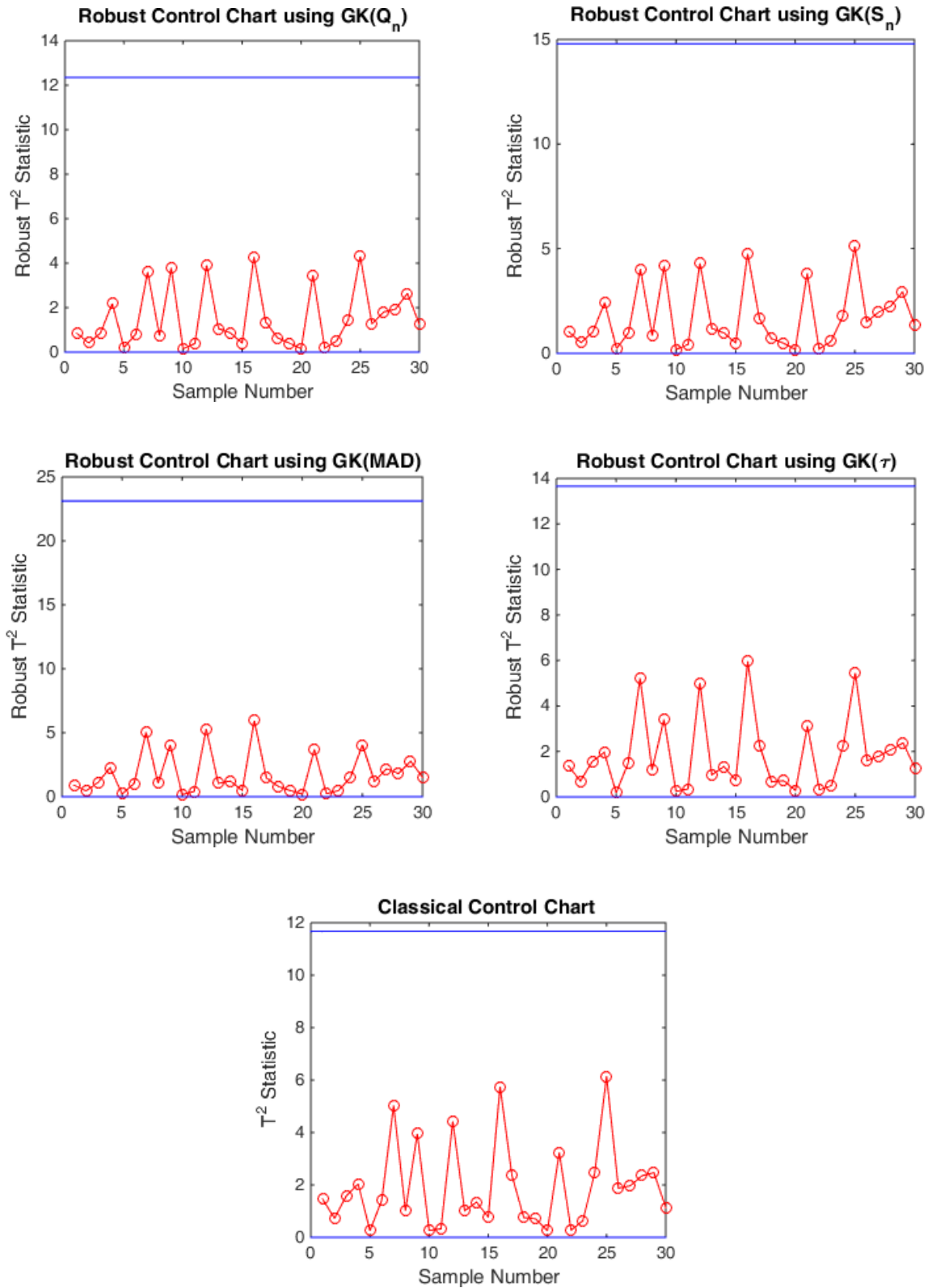


Figure 2: Control charts for uncontaminated Quesenberry data

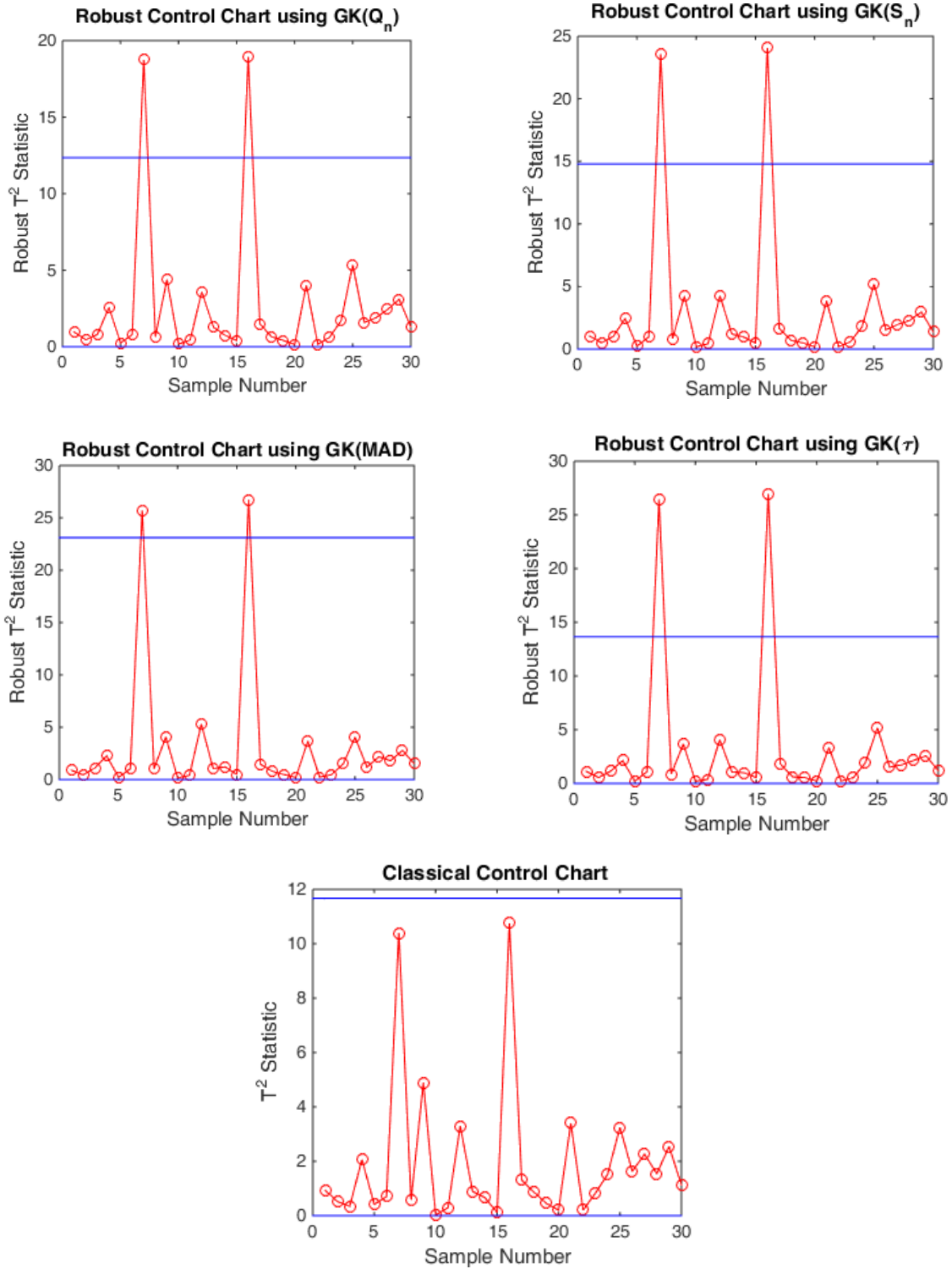


Figure 3: Control charts for contaminated Quesenberry data

5. Conclusions

When quality characteristics are interdependent, monitoring them simultaneously is crucial. Using univariate techniques to monitor or analyse these data is frequently ineffective. The Hotelling's T² control chart, which is created using the classical estimates of location and scatter, is the most well-known multivariate process monitoring and control method. Unfortunately, the existence of outliers

has a significant impact on the classical estimates, which results in incorrect T^2 statistics computation. In light of this, using Hotelling's T^2 control chart based on classical estimates will be quite deceptive. We address this problem for the bivariate case in this paper and suggest four robust control charts to handle outliers. The proposed control charts make use of the covariance estimator introduced by Gnanadesikan and Kettenring [2]. Four highly robust estimators - the Q_n estimator, S_n estimator, MAD estimator, and τ estimator - were used to robustify the GK estimator. These estimators have a bounded influence function and an ideal break down value.

Four different robust control charts for bivariate quality characteristics were introduced using GK(Q_n), GK(S_n), GK(MAD) and GK(τ) estimators. The upper control limits of these robust control charts are obtained by simulating the empirical quantiles of robust T^2 statistics, while the lower control limits are set to zero. Performance of the proposed methods is evaluated through Monte Carlo simulation. Three cases of contaminated Phase I datasets were considered for various amounts of contaminations and in all the cases the proposed methods outperform the classical chart. The proposed methods were also applied to real-life datasets, and even though the data contained outliers, they were still able to identify the out-of-control observation. Another advantage of the proposed methods is their fast computation which will reduce computational complexity.

References

- [1]. Alt, F.B. Multivariate quality control. In S. Kotz and N. Johnson (Eds.), *The Encyclopaedia of Statistical Sciences*, Volume 6 (110-122), John Wiley & Sons, 1985.
- [2]. Gnanadesikan, R. and Kettenring, J.R. (1972), Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics*, 28:81-124.
- [3]. Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69:383-393.
- [4]. Yohai, V. J. and Zamar, R. H. (1986). High breakdown point estimates of regression by means of the minimization of an efficient scale. *Journal of the American Statistical Association*, 86:403-413.
- [5]. Maronna, R.A. and Zamar, R.H. (2002). Robust estimates of location and dispersion for high dimensional datasets, *Technometrics*, 44(4): 307-317.
- [6]. Rousseeuw, P. J. and Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, 88(424):1273-1283.
- [7]. Wilks, S. S. *Mathematical Statistics*. John Wiley & Sons, 1962
- [8]. Serfling, R.J. *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, 1980.
- [9]. Quesenberry, C.P. (2001). The multivariate short-run snapshot Q chart. *Quality Engineering*, 13(4):679-683.