Identifying the Out of Control Variable in a Multivariate Control Chart

Maravelakis, Petros and Bersimis, Sotiris and Panaretos, John and Psarakis, Stelios

2002

COMMUNICATIONS IN STATISTICS
Theory and Methods
Vol. 31, No. 12, pp. 2391–2408, 2002

QUALITY CONTROL

IDENTIFYING THE OUT OF CONTROL VARIABLE IN A MULTIVARIATE CONTROL CHART

P. E. Maravelakis,[1] S. Bersimis,[2] J. Panaretos,[1,*] and S. Psarakis[1]

[1]Department of Statistics, Athens University of Economics and Business, 76 Patission St., 10434, Athens, Greece
[2]Department of Statistics and Insurance Science, University of Pireaus, 80 Karaoli & Dimitriou St., 18534, Pireaus, Greece

ABSTRACT

The identification of the out of control variable, or variables, after a multivariate control chart signals, is an appealing subject for many researchers in the last years. In this paper we propose a new method for approaching this problem based on principal components analysis. Theoretical control limits are derived and a detailed investigation of the properties and the limitations of the new method is given. A graphical technique which can be applied in some of these limiting situations is also provided.

*Corresponding author. E-mail: jpan@aueb.gr

DOI: 10.1081/STA-120017232
0361-0926 (Print); 1532-415X (Online)

Key Words: Multivariate analysis; Multivariate control charts; Ratio of normals; Principal components; Cusum; Statistical Process Control

1. INTRODUCTION

Multivariate control charts are a powerful tool in Statistical Process Control for identifying an out of control process. Woodall and Montgomery[1] in a review paper emphasized the need for much more research in this area since most of the processes involve a large number of variables that are correlated. As Jackson[2] notes, any multivariate quality control procedure should fulfill four conditions (1) Single answer to the question "Is the process in control?" (2) Specification of an overall Type I error (3) The relationship among the variables must be taken into account and (4) Procedures should be available to answer the question "If the process is out of control, what is the problem?" The last question has proven to be an interesting subject for many researchers in the last years. Woodall and Montgomery[1] state that although there is difficulty in interpreting the signals from multivariate control charts more work is needed on data reduction methods and graphical techniques.

In this paper we propose a new method based on principal components analysis (PCA), for identifying the out of control variable, or variables, when a multivariate control chart for individual observations signals. We investigate this alternative procedure because it leads to a more efficient and easily applied way of identifying the out of control variable. Section 2 describes the existing methods for solving the above stated problem. A presentation of the new method, together with two examples are given in Sec. 3. Some interesting points and discussion on the performance and application of the new method is shown in Sec. 4. Section 5 evaluates the performance of the proposed method in relation to the existing methods that use PCA. A graphical method based on the statistics proposed is presented in Sec. 6. Conclusions are summarized in Sec. 7.

2. LITERATURE REVIEW

If a univariate control chart gives an out-of-control signal, one can easily detect what the problem is and give a solution since a univariate chart is related to a single variable. In a multivariate control chart the problem is more complicated because any chart is related to a number, greater than one, of variables and also correlations exist among them. In this section we present methods for detecting, which of the p variables is out of control.

A first approach to this problem was proposed by Alt[3] who suggested the use of Bonferroni limits. Hayter and Tsui[4] extended the idea of Bonferroni-type control limits by giving a procedure for exact simultaneous control intervals for each of the variable means, using simulation. A similar control chart is the Simulated MiniMax control chart presented by Sepulveda and Nachlas.[5] This chart is based on monitoring the maximum and the minimum standardized sample means of samples taken from a multivariate process. Therefore, an out-of-control signal is directly connected with the corresponding out-of-control variable.

Alt[3] and Jackson[2] discussed the use of an elliptical control region. However, this process has the disadvantage that it can be applied only in the special case of two quality characteristics. An extension of the elliptical control region as a solution to the interpretation problem is given by Chua and Montgomery.[6] They use a multivariate exponentially weighted moving average control chart for detecting the out of control situation and the hyperplane method for identifying the variable, or variables, that caused this problem.

A promising method is the use of T^2 decomposition, which is proposed by Mason et al.[7,8] The main idea of this method is to decompose the T^2 statistic into independent parts, each of which reflects the contribution of an individual variable. The major drawback of this method is that the decomposition of the T^2 statistic into p independent T^2 components is not unique as $p!$ different, non independent partitions are possible. An appropriate computing scheme that can greatly reduce the computational effort is given by Mason et. al.[9] Mason et al.[10] presented an alternative control procedure for monitoring a step process, which is based on a double decomposition of Hotelling's T^2 statistic. Mason and Young[11] showed that by improving the model specification at the time that the historical data set is constructed, it may be possible to increase the sensitivity of the T^2 statistic to signal detection. The methodologies of Murphy,[12] Doganaksoy et al.,[13] Hawkins,[14,15] Wierda,[16] Timm,[17] and Runger et al.,[18] are special cases of Mason et al.[7] partitioning of T^2. Wade and Woodall[19] consider a cause-selecting control chart where incoming variable X_1 is charted regardless of X_2 and outgoing quality X_2 is monitored after using regression adjustment for the incoming quality.

Jackson[2] proposed the use of principal components for monitoring a multivariate process. Since the principal components are uncorrelated, they may provide some insight into the nature of the out of control condition and then lead to the examination of particular original observations. Tracy et al.[20] expanded the previous work and provided an interesting bivariate setting in which the principal components have meaningful interpretations. Kourti and MacGregor[21] expressed the T^2 in terms of normalized principal components scores of the multivariate normal variables. When an out of control signal is received, the normalized scores with high value are detected, and contribution plots are used to find the variables responsible for the signal. A contribution plot indicates how each variable involved in the calculation of that score contributes to it. Computing variable contributions eliminates much of the criticism that principal components lack of physical interpretation. This approach is particularly applicable to large ill conditioned data sets due to the use of principal components.

Fuchs and Benjamini[22] presented the multivariate profile chart, which is a symbolic scatterplot, for simultaneously controlling a process and interpreting an out-of-control signal. Summaries of data for individual variables are displayed by a symbol, and global information about the group is displayed by the location of the symbol on the scatterplot. Finally, Sparks et al.[23] used the Gabriel biplot to detect changes in location, variation, and correlation structure accurately.

3. A NEW METHOD

Let $\mathbf{x}_i = (x_{1i}, x_{2i}, \ldots, x_{pi})^{\mathsf{T}}$ denote the observation (vector) i for the p variables of a process. Assume that $\mathbf{x}_i$ follows a p-dimensional Normal distribution $N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, where $\boldsymbol{\mu}_0$ is the vector $(p \times 1)$ of known means and $\boldsymbol{\Sigma}_0$ is the known $(p \times p)$ variance–covariance matrix. We want to keep this process under control. For this purpose we use a X^2 control chart given by the formula $X_i^2 = (\mathbf{x}_i - \boldsymbol{\mu}_0)^{\mathsf{T}} \boldsymbol{\Sigma}_0^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_0)$. If the value of this statistic plots above $X^2_{p,1-\alpha}$ we get an out of control signal, where $X^2_{p,1-\alpha}$ is the chi-square distribution with p degrees of freedom and a is the Type I error. The next problem is to detect which variable is the one that caused the problem.

A $p \times p$ symmetric, nonsingular matrix, such as the variance–covariance matrix $\boldsymbol{\Sigma}$, may be reduced to a diagonal matrix $\mathbf{L}$ by premultiplying and postmultiplying it by a particular orthonormal matrix $\mathbf{U}$ such that $\mathbf{U}^{\mathsf{T}} \boldsymbol{\Sigma} \mathbf{U} = \mathbf{L}$. The diagonal elements of $\mathbf{L}$, $l_1, l_2, \ldots, l_p$ are the characteristic roots, or eigenvalues of $\boldsymbol{\Sigma}$. The columns of $\mathbf{U}$ are the characteristic vectors, or eigenvectors of $\boldsymbol{\Sigma}$. Based on the previous result the method of PCA was developed (see, e.g., Jackson[2]). The PCA transforms p correlated variables $x_1, x_2, \ldots, x_p$ into p new uncorrelated ones.

The main advantage of this method is the reduction of dimensionality. Since the first two or three PCs usually explain the majority of the variability in a process, they can be used for interpretation purposes instead of the whole set of variables.

The typical form of a PCA model is the following: $Z_k = u_{1k}X_1 + u_{2k}X_2 + u_{3k}X_3 + \cdots + u_{pk}X_p$, where Z_k is the k PC, $(u_{1k}, u_{2k}, u_{3k}, \ldots, u_{pk})^{\mathrm{T}}$ is the corresponding k eigenvector and $X_1, \ldots, X_p$ are the process variables. The score for vector $\mathbf{x}_i$ in PC k is $Y_{ki} = u_{1k}x_{1i} + u_{2k}x_{2i} + \cdots + u_{pk}x_{pi}$. Assuming that the process variables follow a multivariate normal distribution the PCs are also normally distributed.

Our purpose is to use PCA, when we have an out of control signal in the X^2 control chart, to identify the variable or variables that are responsible. For this objective two different methodologies are developed one for the case that the covariance matrix has only positive correlations and the second one for the case that we have both positive and negative ones.

3.1. Covariance Matrix with Positive Correlations

Assume that using one of the existing methods for choosing PCs (see, e.g., Jackson,[2] Runger and Alt[24]) we choose $d \leq p$ significant PCs. The proposed method in this case is based on ratios of the form

$$r_{ki} = \frac{(u_{k1} + u_{k2} + \cdots + u_{kd})x_{ki}}{Y_{1i} + Y_{2i} + \cdots + Y_{di}} \tag{1}$$

where x_{ki} is the ith value of variable X_k, Y_{ji}, $j = 1, \ldots, d$ is the score of the ith vector of observations in the jth PC (Bersimis[25]). In this ratio, the numerator corresponds to the sum of the contributions of variable X_k in the first d PCs in observation (vector) i, whereas the denominator is the sum of scores of observation (vector) i in the first d PCs. Since we have assumed that the variables follow a multivariate normal distribution the ratios are ratios of two correlated normal variables.

The rationale of this method is to compute the impact of each of the p variables on the out of control signal by using its contribution to the total score. It is obvious that the use of only the first d PCs excludes pieces of information. However, a multivariate chart is used when there is at least moderate and usually large correlation between the variables. Under such circumstances the first d PCs account for the largest part of the process variability. The main disadvantage of using PCs in process control, as reported by many authors (see, e.g., Runger and Alt;[24] Kourti and

McGregor[21]), is the lack of physical interpretation. The proposed method eliminates much of that criticism.

Since we have a ratio of two correlated normals its distribution can be computed using the analytical result of Hinkley.[26] Specifically, if X_1, X_2 are normally distributed random variables with means μ_i, variances σ_i^2, and correlation coefficient ρ the distribution function of $R = X_1/X_2$ is given by the formula

$$F(r) = L\left\{\frac{\mu_1 - \mu_2 r}{\sigma_1\sigma_2\alpha(r)}; -\frac{\mu_2}{\sigma_2}; \frac{\sigma_2 r - \rho\sigma_1}{\sigma_1\sigma_2\alpha(r)}\right\} + L\left\{\frac{\mu_2 r - \mu_1}{\sigma_1\sigma_2\alpha(r)}; \frac{\mu_2}{\sigma_2}; \frac{\sigma_2 r - \rho\sigma_1}{\sigma_1\sigma_2\alpha(r)}\right\} \tag{2}$$

where $L(h; k; \gamma) = (1/(2\pi\sqrt{1-\gamma^2}))\int_h^\infty \int_k^\infty \exp\{-(x^2 - 2\gamma xy + y^2)/2(1-\gamma^2)\} \times dx\,dy$ is the standard bivariate normal integral.

However, the proposed method has a correlation problem since the ratios of different variables are interrelated. A simulation study presented in Sec. 4 is implemented to test the effect on the performance of the proposed procedure. In the following we present the method as a stepwise procedure.

- **Step 1.** Calculate the X^2 statistic for the incoming observation. If we get an out of control signal continue with Step 2.
- **Step 2.** Calculate ratios for all the variables using relation (1). Calculate as many ratios for each variable as the number of observations from the beginning of the process. If the proposed process is not used for the first time, calculate as many ratios for each variable, as the number of observations from the last out of control signal till the out of control signal we end up with in Step 1. Alternatively, calculate ratios for only the (last) observation that caused the out of control signal (see Sec. 4).
- **Step 3.** Plot the ratios for each variable in a control chart. Compute the a and $1 - a$ percentage points of distribution (2) with suitable parameters and use them as lower control limit (LCL) and upper control limit (UCL), respectively.
- **Step 4.** Observe which variable, or variables, issue an out of control signal.
- **Step 5.** Fix the problem and continue with Step 1.

In the case where all the variables are positively correlated as Jackson[2] indicates, the first PC is a weighted average of all the variables. Consequently, we can use only this PC for inferential purposes.

3.2. Covariance Matrix with Positive and Negative Correlations

In this case we propose the computation of ratios of the form

$$r_{ki}^{*} = \frac{(u_{k1} + u_{k2} + \cdots + u_{kd})x_{ki}}{\overline{Y}_1 + \overline{Y}_2 + \cdots + \overline{Y}_d} \tag{3}$$

where x_{ki} is the ith value of variable X_k and $\overline{Y}_j$, $j = 1, \ldots, d$ is the score of the jth PC using, in place of each X_k, their in control values. The subscript d stands for the number of significant principal components as in the preceding case.

These ratios are the sum of the contributions of variable X_k in the first d PCs in observation (vector) i, divided by the sum of the in control scores of the first d PCs. Since the denominator of this statistic is constant we actually compute the effect of each of the p variables on the out of control signal. The numerator of the ratios is normally distributed, as already stated, whereas the denominator is just a value. Therefore the ratios Eq. (3) are normally distributed.

Since the variables are correlated the statistic proposed in Eq. (3) for the k different variables may exhibit a correlation problem. As in the previous case a simulation study is presented in Sec. 4 to test for the effect of the correlation on the control limits performance of the proposed procedure. The proposed method in steps is as follows:

- **Step 1.** Calculate the X^2 statistic for the incoming observation. If we get an out of control signal continue with Step 2.
- **Step 2.** Calculate ratios for all the variables using relation (3). Calculate as many ratios for each variable as the number of observations from the beginning of the process. If the proposed process is not used for the first time, calculate as many ratios for each variable, as the number of observations from the last out of control signal till the out of control signal we end up with in Step 1. Alternatively, calculate ratios for only the (last) observation that caused the out of control signal (see Sec. 4).
- **Step 3.** Plot the ratios for each variable in a control chart. Compute the a and $1-a$ percentage points of the normal distribution with suitable parameters and use them as LCL and UCL, respectively.
- **Step 4.** Observe which variable, or variables, issue an out of control signal.
- **Step 5.** Fix the problem and continue with Step 1.

We have to mention that this procedure can not be applied when we have standardized values since the denominator of the ratios in Eq. (3) equals zero.

3.3. Illustrative Examples

Two examples, one for each case are presented in the sequel.

Example 1. Assume that we have a process with known covariance matrix

$$\begin{bmatrix} 100 & & & & & \\ 70 & 100 & & & & \\ 80 & 80 & 100 & & & \\ 75 & 85 & 75 & 100 & & \\ 75 & 80 & 80 & 80 & 100 & \\ 75 & 72 & 75 & 75 & 75 & 100 \end{bmatrix}$$

and in-control vector of means $(100, 100, 100, 100, 100, 100)^{\mathrm{T}}$. We simulated 40 in control observations from a multivariate normal distribution with the preceding parameters. Then, we simulated out of control ones with the same covariance matrix but now with vector of means $(100, 115, 100, 85, 100, 100)^{\mathrm{T}}$, until we get an out of control signal in the X^2 test. The shift is 1.5σ in the means of variables 2 and 4. We get a signal on the first out of control observation and we plot each of the variables in a control chart (Fig. 1) with the control limits from distribution (2) using $\alpha = 0.05$. Note that we used the average root method for simplicity and we ended up with one significant principal component (see, e.g., Jackson[2]).

It is obvious from Fig. 1 that the out of control variables were identified and additionally the direction of the shift was also revealed.

Example 2. Assume that we have a process with known covariance matrix

$$\begin{bmatrix} 100 & & & & & \\ -70 & 100 & & & & \\ 80 & -80 & 100 & & & \\ 75 & -85 & 75 & 100 & & \\ 75 & -80 & 80 & 80 & 100 & \\ 75 & -72 & 75 & 75 & 75 & 100 \end{bmatrix}$$

and in-control vector of means $(100, 100, 100, 100, 100, 100)^{\mathrm{T}}$. We simulated 40 in control observations from a multivariate normal distribution with the

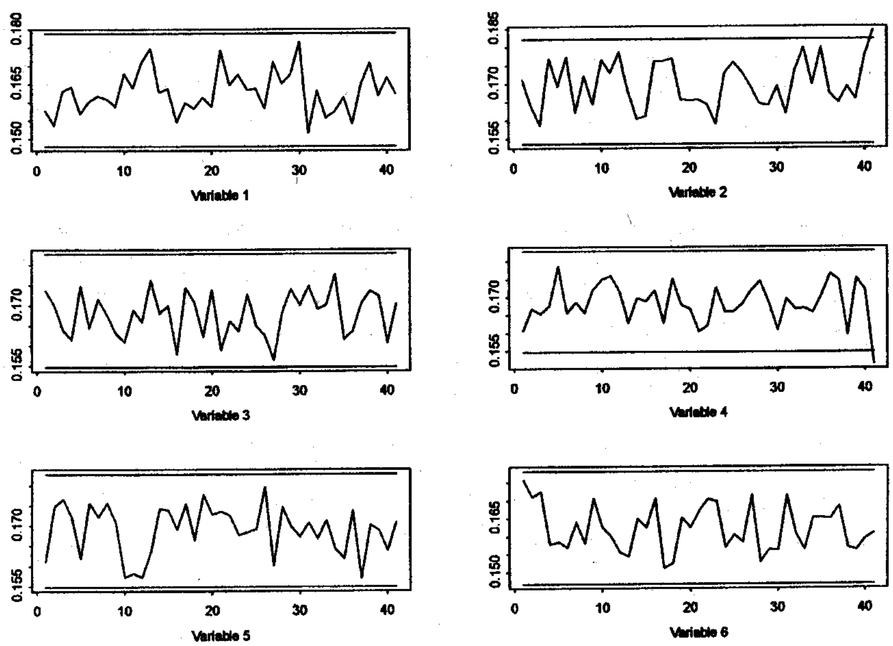

Figure 1. Control charts for positive covariance matrix.

same parameters as in the previous example. Then, we simulated out of control ones with vector of means $(100, 115, 100, 85, 100, 100)^T$ the same covariance matrix till we get an out of control signal in the X^2 test. The shift is again 1.5σ in the means of variables 2 and 4. We plot each of the variables in a control chart (Fig. 2) with the control limits from a normal distribution using $\alpha = 0.05$. As in example 1, we have one significant PC using the average root method again (see, e.g., Jackson[2]). We have to indicate that in Fig. 2 the ratios are standardized and the control limits are properly modified. However, it is not necessary to do this when using this technique.

From Fig. 2, we deduce that the out of control variables were identified but the direction of the shift was not.

4. FURTHER INVESTIGATION

Someone may observe that the ratios in both methods are interrelated. This fact may affect the control limits of the charts. In order to examine this possible correlation we performed a simulation study.

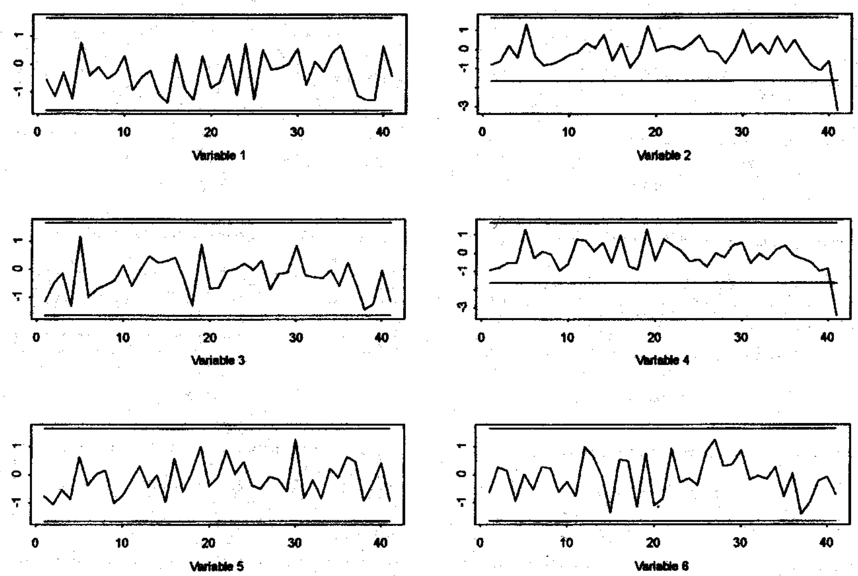

Figure 2. Control charts for positive negative covariance matrix.

Table 1. Type I Errors

Variable	1	2	3	4	5	6
Example 1	5040	5006	5007	4968	5000	5096
Example 2	4941	4983	4898	4882	4906	5005

In particular, we simulated 100000 in control ratios from the known covariance matrices and vector of means of the two examples and we computed the theoretical control limits as proposed in Sec. 3 with $\alpha = 0.05$. Then, we checked if each ratio is in or out of these limits and recorded it. We used this information in order to approximate the Type I error of our limits and compare it with the theoretical one. The results are presented in Table 1.

It should be noted that although the two examples are specific cases, a large number of other cases revealed the same performance. Consequently, we may draw the conclusion that the interrelation does not affect either of the proposed processes. However, we have to point out that after careful examination the procedure proposed for positive correlations can be used only when we have positive values for the in control means. If a process does not have positive in control means, which is a rare event, we may use instead the statistic (3) that is not affected in any case.

Another point which has to be checked is the performance of the processes in identifying the out of control variables and the direction of the shift. For this purpose a simulation study was conducted. Using the covariance matrices and the in control means of the examples we computed in each case the theoretical control limits. Then, we simulated observations from the out of control mean vector used in the examples until we got a signal from the X^2 test using $\alpha = 0.05$. Next, we computed the ratios for each variable and plotted them in a chart with the corresponding control limits. We checked each ratio if it is in or out of the control limits for every variable and recorded which variable, or variables, have given an out of control signal and in which direction. We repeated the whole process 10000 times and the results are presented in Tables 2 and 3. In the first row of the tables (U), we have the number of times the generated ratios crossed the UCL for each of the variables, in the second row (L) we have the corresponding number of times the generated ratios crossed the LCL for each of the variables and in the last line (Total) we have the number of times the generated ratios crossed UCL or LCL for each of the variables. One may observe that in some variables there is an inconsistency, since the sum of rows U and L does not equal the Total. This happens because in one iteration we may generate more than one observations (vectors) until we get an out of control signal in the X^2 test—although this test is sensitive for such shifts—and after computing the ratios for each variable it is possible that for one variable the first ratio crosses UCL and the second ratio crosses LCL. Therefore, we record one value in row U and one in row L but only one in row Total.

From Table 2, we observe that the statistic used is very informative since it is able to detect the out of control variables with very high precision and also to identify the direction of the shift with absolute success. This kind of behavior is similar in other examples also, keeping in mind the limitation about the positive in control means. Note also that the total times the other variables gave a false signal almost coincides with the Type I error of the constructed limits.

In Table 3, we see that the statistic used detected the out of control situation but not with the same degree of success as in the previous case. Moreover, the direction of the shift was not identified. The false alarm rate of the in control variables is not significantly different from the theoretical one.

We already said that the X^2 test is sensitive in the sense that it gives a quick signal when we have out of control observations. In order to examine the ability of the last generated observation (the one that gives

Table 2. Out of Control Performance for Example 1

Variable	1	2	3	4	5	6
U	283	9371	228	0	275	233
L	240	0	299	9429	231	245
Total	523	9371	527	9429	506	478

Table 3. Out of Control Performance for Example 2

Variable	1	2	3	4	5	6
U	364	1	357	2	366	350
L	353	4627	394	4603	357	371
Total	715	4628	748	4605	717	720

Table 4. Out of Control Performance of the Last Observation for Example 1

Variable	1	2	3	4	5	6
U	283	9370	228	0	275	233
L	240	0	299	9429	231	245
Total	523	9370	527	9429	506	478

Table 5. Out of Control Performance of the Last Observation for Example 2

Variable	1	2	3	4	5	6
U	333	1	316	2	324	315
L	312	4279	343	4251	310	333
Total	645	4280	659	4253	634	648

the signal on the X^2 test) to identify the shifted variable we checked their performance on the previous simulation study. The results are displayed in Tables 4 and 5.

From Table 4, we conclude that the performance of the statistic (1) is almost totally explained by the ratios of the last observation meaning that the ratios of the last observation are sufficient to draw a conclusion about the out of control variable. On the other hand, from Table 5 we observe that this does not happen for the statistic (3). One may argue that since the X^2 test is sensitive we produce a small number of observations in each iteration hence the performance in both cases is a result of this fact. When we have small shifts, where we produce more observations, the last observation is not that informative.

The proposed procedures are valid under the assumption of known variance–covariance matrix. However, this is not a case usually met in practice. Tracy et al.[27] examined the performance of multivariate control charts for individual observations when the covariance matrix is known and unknown. They showed that the test statistics used in either case perform the same for a number of observations that depends on the variables involved. This number of observations is small enough, for instance when we have five variables we need 100 observations (vectors) for the two statistics to give a close number. As Woodall and Montgomery[1] note in today's industry we have huge data sets therefore such a number of observations should not be a problem. From this number of observations we estimate the mean vector and the covariance matrix used in our process. Although the control limits computed using the procedures in this paper will not be exact under the estimation process, we expect them to have a satisfactory performance if we use the required number of observations.

5. A COMPARISON

As we already stated in Sec. 2, the competitive methods that use principal components for the specific problem are Jackson's,[2] Tracy et al.'s,[20] and Kourti and MacGregor's.[21] Kourti and MacGregor[21] provide an improved method in relation to Jackson's.[2] Tracy et al.[20] have the disadvantage that their method is applied to a bivariate case. Therefore, it would be interesting to compare the proposed method to the one by Kourti and MacGregor.[21]

In order to do this we perform a simulation study. We apply the method of Kourti and MacGregor[21] to the data of Examples 1 and 2 of Sec. 3.3. As Kourti and MacGregor[21] propose, we use Bonferroni limits on the normalized scores and we calculate the contributions of the variables with the same sign as the score, since contributions of the opposite sign does not add anything to the score (in fact they make it smaller). In the paper of Kourti and MacGregor[21] there is not a specific rule on how to decide whether a contribution is significant or not. Since in the two examples of Sec. 3.3 we have two variables shifted, we choose

Table 6. Performance of Kourti and MacGregor's Method for Example 1

Variable	1	2	3	4	5	6
Largest contribution	2	5204	16	4717	0	0
Second largest contribution	1511	2864	1865	3655	30	10
Total	1513	8068	1881	8372	30	12

Table 7. Performance of Kourti and MacGregor's Method for Example 2

Variable	1	2	3	4	5	6
Largest contribution	23	191	10	2	51	2
Second largest contribution	64	63	47	74	19	12
Total	87	254	57	76	70	14

to record the first and the second larger contributions in each iteration of the simulation study if they exist. The simulation study was conducted 10000 times in order to have a credible estimate of the ability of Kourti and MacGregor's[21] method to identify the out of control variables. In Tables 6 and 7 we have the results of this simulation for Examples 1 and 2, respectively.

From the results in Table 6 for Example 1, we observe that the method of Kourti and MacGregor does not succeed in identifying the out of control variables as many times as the proposed method does (see Tables 2 and 4). Moreover, their mathod has an inherent inability to point if there is an upward or a downward shift. The difference in performance can be seen even better in Table 7 for Example 2. The method of Kourti and MacGregor leads to recordable contributions very rarely, a fact that may lead the practitioner to assume that the signal on the multivariate chart is due to the Type I error. The ability of the new method to operate more effectively is obvious (see Tables 3 and 5).

6. GRAPHICAL TECHNIQUES

The charts proposed in Sec. 3 are Shewhart type. Therefore, they have the ability to identify large shifts quickly but they are not that good for small shifts. An alternative way to plot these statistics is as a Cumulative Sum

(Cusum) chart. The definition of the Cusum chart for detecting upward shifts is

$$S_0^+ = 0$$

$$S_n^+ = \max(0, S_{n-1}^+ + (X_{pn} - k))$$

where X_{pn} is the nth observation of variable X_p and k is called the reference value. The corresponding Cusum chart for detecting downward shifts is

$$S_0^- = 0$$
$$S_n^- = \max(0, S_{n-1}^- + (k - X_{pn}))$$

In the usual concept of Cusum charts we evaluate an optimal value of k depending on the distributional assumption (see Hawkins[28]). This value of k is used along with the value h, which is the control limit, to characterize the Average Run Length performance of a Cusum chart. In our case the application of this theory for Cusum charts is cumbersome due to the underlying distribution. However, we can use the previously defined statistics for upward and downward shifts as a graphical technique solely. The only thing remaining unknown is the value k we have to use. A straightforward selection for k is to use in each case of statistics (1) and (3) their in control counterparts. Specifically, for statistic (1) we use in place of each x_{ki} its in control value both in the numerator and the denominator. A similar action takes place in statistic (3) but only in the numerator this time.

To study the performance of these statistics in practice we applied them to the examples of Sec. 3. We used the same 40 in control observations but now we generated out of control ones with the same covariance matrix as in Sec. 3, and vector of means $(100, 105, 100, 95, 100, 100)^{\mathrm{T}}$, in both examples until we get an out of control signal in the X^2 test. The shift is 0.5σ in the means of variables 2 and 4. In Example 1, we simulated 12 observations till the out of control signal and 14 in Example 2. We computed the Cusum values for the 52 and 54 values for all six variables in Examples 1 and 2 respectively, and we plotted them in the chart given in Fig. 3.

From Fig. 3 we easily deduce that the charts give a clear indication of the out of control variables in both examples. As in the Shewhart type charts of Sec. 3, the statistic (1) used in Example 1 detected also the direction of the shift something that did not happen with statistic (3) in Example 2. We have to mention here that the effectiveness of the Cusum as a graphical device for

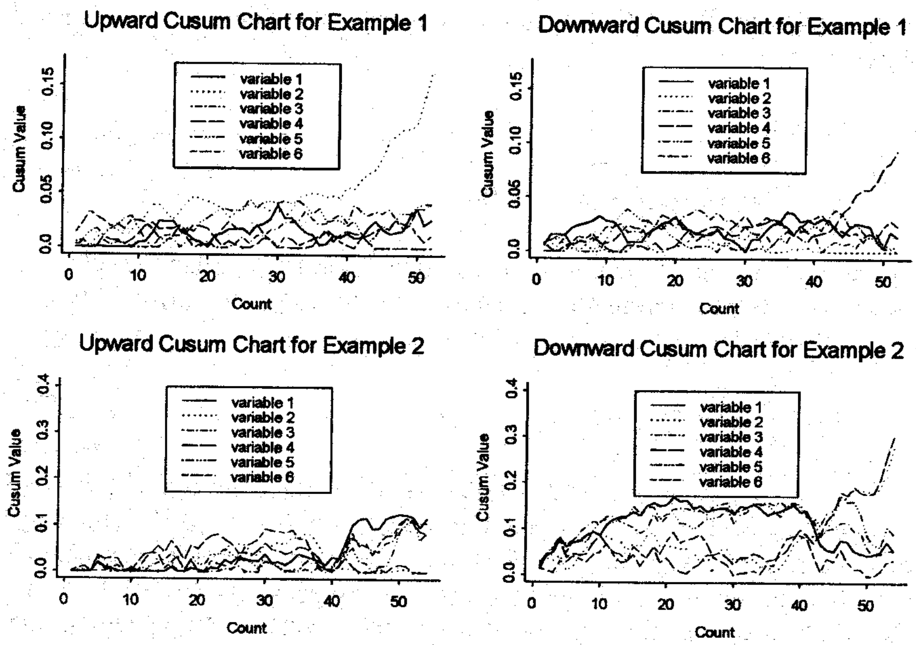

Figure 3. Control charts for Cusum values.

shifts less than 0.5σ is questionable. However, it is an easily interpreted method that can give an indication.

7. CONCLUSIONS

In this paper we presented a method for identifying the out of control variable when a multivariate control chart signals based on the method of principal components. The theoretical control limits were derived as well as the method in steps. We implemented the proposed process in two examples in order to get a clear view of how it works. A detailed investigation of the properties and the limitations of the new method is given. A graphical technique that can be applied in these limiting situations is also provided. Finally, a comparison between the proposed method and the existing ones that use PCA showed that the new one has a better performance.

Summarizing, we note that the charts proposed are an easily applied alternative to most of the existing methods since the computational effort is diminished. Furthermore, we try to give an answer to the problem under a control charting perspective giving operational control limits or design strategies that are not difficult for a practitioner to apply.

REFERENCES

1. Woodall, W.H.; Montgomery, D.C. Research Issues and Ideas in Statistical Process Control. Journal of Quality Technology **1999**, *31*, 376–386.
2. Jackson, J.E. *A User Guide to Principal Components*; John Wiley & Sons: New York, 1991.
3. Alt, F.B. Multivariate Quality Control. *The Encyclopedia of Statistical Sciences*; Kotz, S., Johnson, N.L., Read, C.R., Eds.; John Wiley & Sons: New York, 1985; 110–122.
4. Hayter, A.J.; Tsui, K-L. Identification and Quantification in Multivariate Quality Control Problems. Journal of Quality Technology **1994**, *26*, 197–208.
5. Sepulveda, A.; Nachlas, J.A. A Simulation Approach to Multivariate Quality Control. Computers and Industrial Engineering **1997**, *33*, 113–116.
6. Chua, M-K.; Montgomery, D.C. Investigation and Characterization of a Control Scheme for Multivariate Quality Control. Quality and Reliability Engineering International **1992**, *8*, 37–44.
7. Mason, R.L.; Tracy, N.D.; Young, J.C. Decomposition of T^2 for Multivariate Control Chart Interpretation. Journal of Quality Technology **1995**, *27*, 99–108.
8. Mason, R.L.; Champ, C.W.; Tracy, N.D.; Wierda, R.J.; Young, J.C. Assessment of Multivariate Process Control Techniques. Journal of Quality Technology **1997**, *29*, 140–143.
9. Mason, R.L.; Tracy, N.D.; Young, J.C. A Practical Approach for Interpreting Multivariate T^2 Control Chart Signals. Journal of Quality Technology **1997**, *29*, 396–406.
10. Mason, R.L.; Tracy, N.D.; Young, J.C. Monitoring a Multivariate Step Process. Journal of Quality Technology **1996**, *28*, 39–50.
11. Mason, R.L.; Young, J.C. Improving the Sensitivity of the T^2 Statistic in Multivariate Process Control. Journal of Quality Technology **1999**, *31*, 155–165.
12. Murphy, B.J. Selecting Out-of-Control Variables with T^2 Multivariate Quality Procedures. The Statistician **1987**, *36*, 571–583.
13. Doganaksoy, N.; Faltin, F.W.; Tucker, W.T. Identification of Out-of-Control Multivariate Characteristic in a Multivariable Manufacturing Environment. Communications in Statistics A—(Theory and Methods) **1991**, *20*, 2775–2790.
14. Hawkins, D.M. Multivariate Quality Control Based on Regression-Adjusted Variables. Technometrics **1991**, *33*, 61–75.

15. Hawkins, D.M. Regression Adjustment for Variables in Multivariate Quality Control. Journal of Quality Technology **1993**, *25*, 170–182.
16. Wierda, S.J. Multivariate Statistical Process Control—Recent Results and Directions for Future Research. Statistica Neerlandica **1994**, *48*, 147–168.
17. Timm, N.H. Multivariate Quality Control Using Finite Intersection Tests. Journal of Quality Technology **1996**, *28*, 233–243.
18. Runger, G.C.; Alt, F.B.; Montgomery, D.C. Contributors to a Multivariate SPC Chart Signal. Communications in Statistics A—(Theory and Methods) **1996**, *25*, 2203–2213.
19. Wade, M.R.; Woodall, W.H. A Review and Analysis of Cause-Selecting Control Charts. Journal of Quality Technology **1993**, *25*, 161–170.
20. Tracy, N.D.; Young, J.C.; Mason, R.L. A Bivariate Control Chart for Paired Measurements. Journal of Quality Technology **1995**, *27*, 370–376.
21. Kourti, T.; MacGregor, J.F. Multivariate SPC Methods for Process and Product Monitoring. Journal of Quality Technology **1996**, *28*, 409–428.
22. Fuchs, C.; Benjamini, Y. Multivariate Profile Charts for Statistical Process Control. Technometrics **1994**, *36*, 182–195.
23. Sparks, R.S.; Adolphson, A.; Phatak, A. Multivariate Process Monitoring Using the Dynamic Biplot. International Statistical Review **1997**, *65*, 325–349.
24. Runger, G.C.; Alt, F.B. Choosing Principal Components for Multivariate SPC. Communications in Statistics A—(Theory and Methods) **1996**, *25*, 909–922.
25. Bersimis, S. *Multivariate Statistical Process Control*, M.Sc. Thesis, Department of Statistics, Athens University of Economics and Business: 2001; ISBN 960-7929-45-4.
26. Hinkley, D.V. On the Ratio of Two Correlated Normal Random Variables. Biometrika **1969**, *56*, 635–639.
27. Tracy, N.D.; Young, J.C.; Mason, R.L. Multivariate Control Charts for Individual Observations. Journal of Quality Technology **1992**, *24*, 88–95.
28. Hawkins, D.M. Evaluation of the Average Run Length of Cumulative Sum Charts for an Arbitrary Data Distribution. Communications in Statistics—Simulation and Computation **1992**, *21*, 1001–1020.