

# **DEPARTMENT OF COMPUTER AND MATHEMATICAL SCIENCES**

Mean Control Charts for Multivariate  
Normal Processes

Pak Fai Tang

(54 EQRM 15)

May, 1995

(AMS : 62N10)

## **TECHNICAL REPORT**

VICTORIA UNIVERSITY OF TECHNOLOGY  
(P O BOX 14428) MELBOURNE MAIL CENTRE  
MELBOURNE, VICTORIA, 3000  
AUSTRALIA

TELEPHONE (03) 9688 4249 / 4492  
FACSIMILE (03) 9688 4050

# Mean Control Charts for Multivariate Normal Processes

PAK FAI TANG

Victoria University of Technology, Melbourne, Australia.

## ABSTRACT

Some methods for monitoring and controlling the mean level of a multivariate normal process are presented in this paper. Those that do not depend on prior estimates of the process parameters are particularly attractive to short-run or low volume manufacturing environments. As the given techniques involve sequences of independent or approximately independent standard normal variables, the resulting control charts can all be plotted using the same scale irrespective of product types, thus simplifying charting administration. A simulation study indicates that these control procedures are particularly useful for 'picking up' a sustained shift in the process mean vector when subgroup data are used, even if prior information about the process parameters are not available. Some illustrative examples are presented.

## INTRODUCTION

In many industrial situations, it is not uncommon to monitor on-going performance of a production process with respect to more than one process or product characteristic simultaneously. If the product characteristics are correlated and no consideration is made of their joint distribution, the use of separate control charts for each of them can be misleading. Specifically, the related variables, when studied separately, may appear to be in statistical control while in fact the manufacturing system is out-of-control or vice versa when considering the quality characteristics in their multivariate context. Montgomery et al.(1972) and Alt et al.(1990) illustrated this for the case of  $p = 2$  variables. Under these circumstances, it seems reasonable to consider the use of multivariate quality control procedures which take into account the covariance structure of the quality characteristics.

Most of the statistical process control (SPC) techniques proposed to date for controlling the mean of a multivariate normal process are based on Hotelling's (1947)  $\chi^2$  or  $T^2$ -type statistic. Other multivariate control procedures, including the use of principal components and multivariate cusum (MCUSUM) techniques, were reviewed by Jackson (1991). A multivariate version of the exponentially weighted moving average chart, referred to as the MEWMA chart, has also been presented by Lowry et al. (1992). Apart from these, some techniques that are designed to provide protection against changes in the process covariance matrix have been presented, for example, in Montgomery et al.(1972) and Alt et al.(1986). Recently, various attempts have also been made to develop control techniques which can both detect any process irregularities and identify the set of out-of-control variables, taking the correlational structure of the quality characteristics into consideration (see for eg.,

Doganaksoy et al.(1991), Hawkins (1993) and Hayter et al. (1994)). However, all of this work essentially assumes that the process mean (or target)  $\mu$  and the process covariance matrix  $\Sigma$  are known or that they can be reliably estimated based on sufficient data prior to full scale production. Thus, these techniques do not readily lend themselves to applications in the short-run or low-volume manufacturing environment where data for estimating the process parameters are usually scarce or unavailable.

The main purpose of this paper is to present some 'self-starting' and unified procedures for controlling the mean vector of a multivariate normal process where prior estimates of the process parameters are not available such as frequently occurs in short-run, low-volume and multi-product manufacturing environments. The proposed techniques also facilitate the control of long-run processes at an earlier stage when the process parameters have to be estimated from the data stream of the current production. For completeness, control procedures are also given for the cases where  $\mu$ ,  $\Sigma$  or both are assumed known in advance of the production runs.

In addition to this introduction, the paper is organized into five sections. In the next section, control statistics for use with individual and subgroup data are given according to various assumptions about the knowledge of the process parameters. Some numerical examples illustrating the use of these techniques are provided in section 3. In section 4, simulation results are presented and statistical performance of the proposed techniques discussed. A brief account of the computational requirements for implementation of the proposed techniques is given in section 5. The final section concludes with a brief summary and further remarks. The total discourse is in the context of discrete items manufacture.

## MULTIVARIATE MEAN CONTROL CHARTS

Like the 'Q' charting approach previously proposed by Quesenberry (1991) for controlling univariate normal processes, the techniques presented in this paper involve the use of the *probability integral transformation* of some (scaled) *quadratic forms* in order to produce sequences of independent or approximately independent standard normal variables. These procedures essentially enable charting to commence with the first units or samples of production whether or not prior knowledge of the process parameters is available. For the case where no relevant data is available prior to a production run, the process parameters  $\mu$  and  $\Sigma$  are 'estimated' and 'updated' sequentially from the current data stream. These dynamic estimates, together with the next observation or subgroup are in turn used to determine whether the process remains stable.

The appropriate control statistics based on both individual observations and subgroup data are presented in the following subsections. The control statistics are given for the cases when both, either or neither of the process parameters  $\mu$  and  $\Sigma$  are assumed known in advance of the production run.

### **(a) Control Charts Based on Individual Measurements**

Let  $X_1, X_2, \dots$  be the vectors of measurements on  $p$  quality characteristics for products produced in time sequence. Assume that these observation vectors are independently and identically distributed having been collected from a  $p$ -variate normal  $N_p(\mu, \Sigma)$  process.

Further, let  $\bar{\mathbf{X}}_k$ ,  $\mathbf{S}_k$  and  $\mathbf{S}_{\mu,k}$  denote respectively the mean vector, the usual and the 'mean-dependent' covariance matrix of the first  $k$  observations as defined below :

$$\begin{aligned}\bar{\mathbf{X}}_k &= \frac{1}{k} \sum_{i=1}^k \mathbf{X}_i \\ \mathbf{S}_k &= \frac{1}{k-1} \sum_{i=1}^k (\mathbf{X}_i - \bar{\mathbf{X}}_k)(\mathbf{X}_i - \bar{\mathbf{X}}_k)' \\ \mathbf{S}_{\mu,k} &= \frac{1}{k} \sum_{i=1}^k (\mathbf{X}_i - \boldsymbol{\mu})(\mathbf{X}_i - \boldsymbol{\mu})'\end{aligned}$$

These values can be updated sequentially using the following recursive formulae :

$$\begin{aligned}\bar{\mathbf{X}}_k &= \frac{1}{k} [(k-1)\bar{\mathbf{X}}_{k-1} + \mathbf{X}_k] & k = 2, 3, \dots \\ \mathbf{S}_k &= \frac{(k-2)}{(k-1)} \mathbf{S}_{k-1} + \frac{1}{k} (\mathbf{X}_k - \bar{\mathbf{X}}_{k-1})(\mathbf{X}_k - \bar{\mathbf{X}}_{k-1})' & k = 3, 4, \dots \\ \mathbf{S}_{\mu,k} &= \frac{k-1}{k} \mathbf{S}_{\mu,k-1} + \frac{1}{k} (\mathbf{X}_k - \boldsymbol{\mu})(\mathbf{X}_k - \boldsymbol{\mu})' & k = 2, 3, \dots\end{aligned}$$

The following additional notation will be used.

$\Phi(\bullet)$  : distribution function of a standard normal variable

$\Phi^{-1}(\bullet)$  : inverse of the standard normal distribution function

$\chi^2_\nu(\bullet)$  : distribution function of a *chi-square* variable with  $\nu$  degrees of freedom

$F_{\nu_1, \nu_2}(\bullet)$  : distribution function of an *F* variable with  $\nu_1$  numerator degrees of freedom and  $\nu_2$  denominator degrees of freedom

The appropriate control statistics are now presented as follows :

**Case (I)** : Both  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  known

$$\begin{aligned}Z_k &= \Phi^{-1}[\chi^2_p(T_k)] \quad k = 1, 2, \dots \\ \text{where } T_k &= (\mathbf{X}_k - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X}_k - \boldsymbol{\mu})\end{aligned} \quad (1)$$

**Case (II)** :  $\boldsymbol{\mu}$  unknown,  $\boldsymbol{\Sigma}$  known

$$\begin{aligned}Z_k &= \Phi^{-1}[\chi^2_p(T_k)] \quad k = 2, 3, \dots \\ \text{where } T_k &= \frac{(k-1)}{k} (\mathbf{X}_k - \bar{\mathbf{X}}_{k-1})' \boldsymbol{\Sigma}^{-1} (\mathbf{X}_k - \bar{\mathbf{X}}_{k-1})\end{aligned} \quad (2)$$

**Case (III)** :  $\mu$  known,  $\Sigma$  unknown

For this case, two alternative control statistics are considered which, as demonstrated later, have slightly different performance. They involve the use of  $S_{\mu,k}$  and  $S_k$  respectively.

(a) Uses  $\mu$  in  $\Sigma$  estimate

$$Z_k = \Phi^{-1} \left[ F_{p,k-p}(T_k) \right] \quad k = p+1, \dots$$

$$\text{where } T_k = \frac{(k-p)}{p(k-1)} (\mathbf{X}_k - \mu)' S_{\mu,k-1}^{-1} (\mathbf{X}_k - \mu) \quad (3a)$$

(b) Uses Sample Variance-Covariance Matrix

$$Z_k = \Phi^{-1} \left[ F_{p,k-1-p}(T_k) \right] \quad k = p+2, \dots$$

$$\text{where } T_k = \frac{(k-1-p)}{p(k-2)} (\mathbf{X}_k - \mu)' S_{k-1}^{-1} (\mathbf{X}_k - \mu) \quad (3b)$$

**Case (IV)** : Both  $\mu$  and  $\Sigma$  unknown

$$Z_k = \Phi^{-1} \left[ F_{p,k-1-p}(T_k) \right] \quad k = p+2, \dots$$

$$\text{where } T_k = \left( \frac{(k-1)(k-1-p)}{kp(k-2)} \right) (\mathbf{X}_k - \bar{\mathbf{X}}_{k-1})' S_{k-1}^{-1} (\mathbf{X}_k - \bar{\mathbf{X}}_{k-1}) \quad (4)$$

Note that for the cases with some unknown parameters, monitoring the process can essentially commence with the first units of production without having to wait until enough process performance data has built up. This is crucial for the control of short-run processes and processes during 'warm up'.

As shown above, for case (I), a value of  $Z_k$  corresponds to  $\mathbf{X}_k$  for all values of  $k = 1, 2, \dots$ . However, no value of  $Z_k$  corresponding to the first observation  $\mathbf{X}_1$  is calculated for case (II). This is due to the fact that the unknown process mean vector  $\mu$  has to be estimated from  $\mathbf{X}_1$  before its constancy can be subsequently monitored, based on further observations. For case (III), if the control statistics given by (3a) are used, the monitoring procedure begins after  $p+1$  observations. For the remaining cases, control is initiated at the  $(p+2)$ th observation. In general, if  $\Sigma$  has to be estimated from the current data stream, the starting periods of the control procedure are such that they yield positive definite estimates of  $\Sigma$ . For example, no value is plotted using the control statistics of (3b) or (4) prior to the  $(p+2)$ th observation because the sample covariance matrix  $S_{k-1}$ , used in the formula, is not of full rank and hence is not invertible for  $k$  less than  $p+2$ .

Except for (3b),  $\{Z_k\}$  for each case is a sequence of independently and identically distributed (*i.i.d*) normal variables with mean 0 and standard deviation 1 under the stable in-control normality assumption. These distributional properties are shown in Appendix A for case (IV) only, which is the most pertinent result for short run SPC. The proofs for other cases can be established in the similar manner. The  $Z_k$  statistics of (3b) are also standard normal variables and are *approximately* independent as  $k$  becomes large. However, as the simulation study indicates, ignoring the issue of dependence among the  $Z_k$ 's has no significant effect on the false signal rate. In particular, the probability of getting a false signal from *any* one of the

first 50  $p$ -variate measurements was simulated for  $p = 2, 3$  and  $5$  based on 10,000 runs when either 3-sigma limits or only the upper 99.73th probability limit is used. The results appear to agree well with the nominal value of  $1 - 0.9973^{50} = 0.1264$ . The reason the above criterion is chosen as a measure of in-control performance instead of the usual average run length is that the control statistics given by (3b) are primarily concerned with short production run situations in which the total number of observations rarely exceeds 50.

As the plot statistics for all cases are sequences of independent (or approximately independent) standard normal variables, the resulting control charts can be constructed using the same scale and with the same Shewhart control limits  $\pm 3$ . In addition, supplementary run rules can be employed to reveal any assignable causes hidden in the point patterns of the charts. Although the argument statistics  $T_k$ 's can be plotted instead of the transformed variables  $Z_k$ 's, this practice is not recommended if additional computational effort can be justified. This is because, for those cases other than case (I), the use of  $T_k$ 's results in varying control limits which can lead to misinterpretation. The use of these procedures is illustrated after the appropriate control statistics based on subgroup data have been presented.

## **(b) Control Charts Based on Subgroup Data**

In practice, the use of subgroup data is often preferable to individual measurements even in situations where only a limited amount of data are available. This is justified on the basis that the resulting control charts are more sensitive to substantial shifts in the process average and that the subgroup mean (vector) is less affected by departure from the underlying normality assumption, by virtue of the central limit theorem.

A little adaptation of the above formulae yields appropriate control statistics for monitoring the stability of the process mean vector based on subgroup data. Discussion will be restricted to the case of *constant* subgroup size, as this is a common phenomenon. Let  $\mathbf{X}_{ij}$ ,  $n$  and  $k$  denote respectively the  $j$ th observation vector of the  $i$ th subgroup, the common subgroup size and the subgroup number, and define the following quantities :

$$\begin{aligned}\bar{\mathbf{X}}_{(i)} &= \frac{1}{n} \sum_{j=1}^n \mathbf{X}_{ij} \\ \bar{\bar{\mathbf{X}}}_{(k)} &= \frac{1}{k} \sum_{i=1}^k \bar{\mathbf{X}}_{(i)} = \frac{1}{k} \left[ (k-1) \bar{\bar{\mathbf{X}}}_{(k-1)} + \bar{\mathbf{X}}_{(k)} \right] \\ \mathbf{S}_{(i)} &= \frac{1}{n-1} \sum_{j=1}^n (\mathbf{X}_{ij} - \bar{\mathbf{X}}_{(i)}) (\mathbf{X}_{ij} - \bar{\mathbf{X}}_{(i)})' \\ \mathbf{S}_{\mu}^{(k)} &= \frac{1}{kn} \sum_{i=1}^k \sum_{j=1}^n (\mathbf{X}_{ij} - \mu) (\mathbf{X}_{ij} - \mu)' \\ &= \frac{1}{k} \left[ (k-1) \mathbf{S}_{\mu}^{(k-1)} + \frac{1}{n} \sum_{j=1}^n (\mathbf{X}_{kj} - \mu) (\mathbf{X}_{kj} - \mu)' \right] \\ \mathbf{S}_{\text{pooled}}^{(k)} &= \frac{1}{k} \sum_{i=1}^k \mathbf{S}_{(i)} = \frac{1}{k} \left[ (k-1) \mathbf{S}_{\text{pooled}}^{(k-1)} + \mathbf{S}_{(k)} \right]\end{aligned}$$



where  $S_{\mu}^{(0)} = S_{pooled}^{(0)} = 0$ . The same technique is now applied to transform the sample mean vectors  $\bar{X}_{(i)}$ 's to standard normal variables for each of the cases considered above for the individual measurements.

**Case (I)** : Both  $\mu$  and  $\Sigma$  known

$$Z_k = \Phi^{-1}[\chi_p^2(T_k)] \quad k = 1, 2, \dots$$

$$\text{where } T_k = n(\bar{X}_{(k)} - \mu)' \Sigma^{-1} (\bar{X}_{(k)} - \mu) \quad (5)$$

**Case (II)** :  $\mu$  unknown,  $\Sigma$  known

$$Z_k = \Phi^{-1}[\chi_p^2(T_k)] \quad k = 2, 3, \dots$$

$$\text{where } T_k = \left(\frac{n(k-1)}{k}\right) (\bar{X}_{(k)} - \bar{\bar{X}}_{(k-1)})' \Sigma^{-1} (\bar{X}_{(k)} - \bar{\bar{X}}_{(k-1)}) \quad (6)$$

**Case (III)** :  $\mu$  known,  $\Sigma$  unknown

Two alternative techniques which employ different process covariance matrix estimates will be considered for this case. The first one incorporates  $\mu$  into its 'running' estimate of  $\Sigma$  whereas the other one is based on the pooled estimate  $S_{pooled}^{(k)}$ . It will be seen in section 4 that these control procedures have comparable performance.

(a) Uses  $\mu$  in estimation of  $\Sigma$

$$Z_k = \Phi^{-1}[F_{p, n(k-1)-p+1}(T_k)] \quad k = 2, 3, \dots, \quad n \geq p$$

$$\text{where } T_k = \left(\frac{n(k-1)-p+1}{p(k-1)}\right) (\bar{X}_{(k)} - \mu)' (S_{\mu}^{(k-1)})^{-1} (\bar{X}_{(k)} - \mu) \quad (7a)$$

(b) Uses Pooled Sample Covariance Matrix

$$Z_k = \Phi^{-1}[F_{p, k(n-1)-p+1}(T_k)] \quad k = 1, 2, \dots, \quad n > p$$

$$\text{where } T_k = \left[\frac{n[k(n-1)-p+1]}{pk(n-1)}\right] (\bar{X}_{(k)} - \mu)' (S_{pooled}^{(k)})^{-1} (\bar{X}_{(k)} - \mu) \quad (7b)$$

**Case (IV)** : Both  $\mu$  and  $\Sigma$  unknown

$$Z_k = \Phi^{-1}[F_{p, k(n-1)-p+1}(T_k)] \quad k = 2, 3, \dots, \quad n > p$$

$$\text{where } T_k = \left(\frac{n(k-1)(k(n-1)-p+1)}{k^2 p(n-1)}\right) (\bar{X}_{(k)} - \bar{\bar{X}}_{(k-1)})' (S_{pooled}^{(k)})^{-1} (\bar{X}_{(k)} - \bar{\bar{X}}_{(k-1)}) \quad (8)$$

Under the assumption that the  $\mathbf{X}_{ij}$ 's are independent observation vectors obtained from a common process with a  $N_p(\mu, \Sigma)$  distribution, the control statistics given by (5), (6) and (7a) are sequences of independently distributed standard normal variables (see Appendix A). For those given by (7b) and (8), successive plotted values also arise from a standard normal distribution but they are correlated due to the use of the pooled sample covariance matrix. Sequences of independently distributed variables can be obtained for these latter cases by replacing the pooled sample covariance matrix  $\mathbf{S}_{pooled}^{(k)}$  by the sample covariance matrix of the current subgroup,  $\mathbf{S}_{(k)}$ . This is not considered, however, because it is found that the performance of the resulting charts are poor even when some additional run rules are used. Furthermore, ignoring the issue of correlation does not appear to have remarkable effect on the false signal rates for the control techniques based on statistics (7b) and (8).

A remark should be made about the computation of the argument statistics for (1) - (8) above which involve evaluation of the inverse of a matrix. In fact, each of these arguments can be expressed as a quotient of two determinants, thus eliminating the need for inverting either the known process covariance matrix or some estimates of it (see for eg., Morrison (1976), p.134 ). For instance, the argument statistic of (4) has the following alternative expression

$$T_k = \frac{\left| \mathbf{S}_{k-1} + \frac{(k-1)(k-1-p)}{kp(k-2)} (\mathbf{X}_k - \bar{\mathbf{X}}_{k-1})(\mathbf{X}_k - \bar{\mathbf{X}}_{k-1})' \right|}{|\mathbf{S}_{k-1}|} - 1$$

Evaluating an expression of this form is much more convenient than that of its original form especially when the number of quality characteristics,  $p$  is large.

For some general guidelines on using the above control charting approach and a discussion of its advantages and disadvantages, readers are referred to the article by Quesenberry (1991) from which the ideas of this present work originate.

## EXAMPLES

In this section, use of the proposed techniques are illustrated by some numerical examples based on simulated data as well as using data from a previously published article. The examples presented here are not intended to cover all possible situations, however, they provide some insight into the behaviour of the proposed techniques under various circumstances.

### Example 1

The first illustration uses the formulae for *individual* measurements. 30 observations have been generated from a bivariate normal distribution with the following parameters

$$\mu = \begin{pmatrix} 10 \\ 15 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & 1.275 \\ 1.275 & 2.25 \end{pmatrix}.$$

These data are shown in TABLE 1, along with the computed values of the control statistics (1) to (4). The corresponding control charts have also been constructed as shown in Figures 1A to



1E. It should be noted that the use of  $3\text{-}\sigma$  control limits in this and subsequent examples is merely for the purpose of illustration. In practice, it may be preferable to use narrower control limits or only the upper control limit in line with the traditional Hotelling  $T^2$  charting approach.

**TABLE 1. Simulated Data and Values of The Control Statistics based on Individual Measurements for Example 1**

Obs. No.	Variable 1	Variable 2	Control Statistic				
			(1)	(2)	(3a)	(3b)	(4)
1	10.39	15.70	-1.27	NA	NA	NA	NA
2	9.02	14.19	-0.08	-0.28	NA	NA	NA
3	9.28	13.71	-0.50	-0.62	0.07	NA	NA
4	8.67	14.04	0.61	-0.19	0.52	0.06	-0.32
5	9.22	14.99	0.40	-0.55	0.46	0.24	-0.21
6	8.82	11.54	1.98	1.99	2.09	1.64	1.56
7	9.07	14.14	-0.24	-1.39	-0.37	0.17	-1.52
8	11.70	17.31	0.73	1.50	0.47	1.29	1.83
9	10.92	16.35	-0.35	0.22	-0.60	-0.57	-0.07
10	9.64	14.84	-1.15	-1.80	-1.21	-1.24	-1.91
11	8.31	12.85	0.70	0.18	0.45	0.44	-0.01
12	9.56	14.23	-1.16	-1.55	-1.29	-1.30	-1.56
13	10.39	14.74	-0.22	0.15	-0.14	-0.16	0.19
14	10.11	15.84	-0.43	-0.30	-0.36	-0.42	-0.33
15	9.53	13.79	-0.42	-0.57	-0.39	-0.44	-0.55
16	11.15	16.77	0.05	0.46	0.01	0.06	0.45
17	8.03	11.89	1.24	0.88	1.05	1.03	0.72
18	9.68	14.04	-0.67	-0.86	-0.62	-0.66	-0.72
19	9.57	14.11	-0.95	-1.48	-1.00	-1.00	-1.39
20	9.05	13.58	-0.29	-0.98	-0.40	-0.38	-1.01
21	12.40	18.15	1.59	1.98	1.37	1.49	1.80
22	10.17	15.17	-2.14	-1.37	-2.12	-2.13	-1.49
23	8.41	13.00	0.58	0.22	0.38	0.37	0.07
24	10.31	16.11	-0.33	-0.07	-0.03	-0.06	0.08
25	9.66	13.57	0.19	0.05	0.55	0.50	0.44
26	9.50	13.78	-0.44	-0.80	-0.35	-0.36	-0.62
27	10.91	17.65	1.22	1.44	1.40	1.38	1.52
28	10.59	16.43	-0.20	0.03	-0.23	-0.25	-0.11
29	9.71	14.05	-0.62	-0.73	-0.54	-0.56	-0.55
30	9.48	13.85	-0.60	-0.87	-0.66	-0.68	-0.86

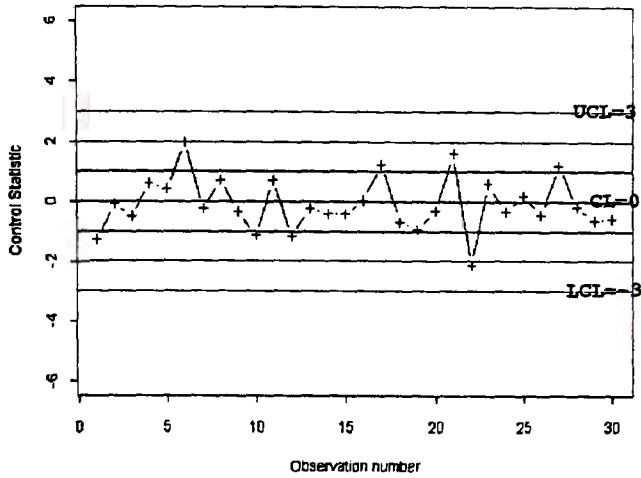


FIGURE 1A. Multivariate Control Chart for Example 1, known Parameters.

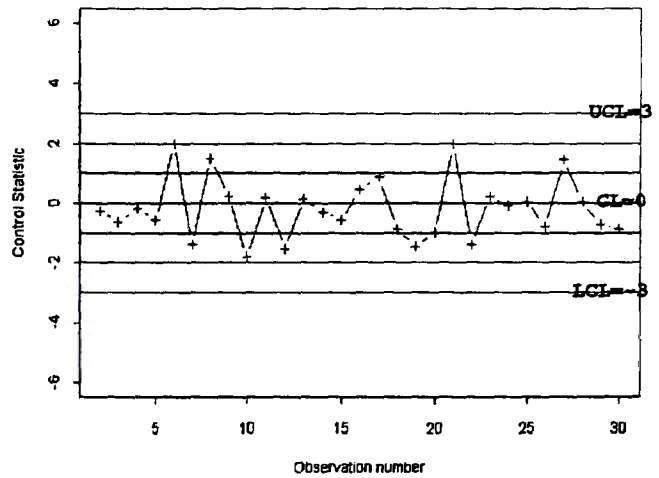


FIGURE 1B. Multivariate Control Chart for Example 1,  $\mu$  Unknown,  $\Sigma$  Known.

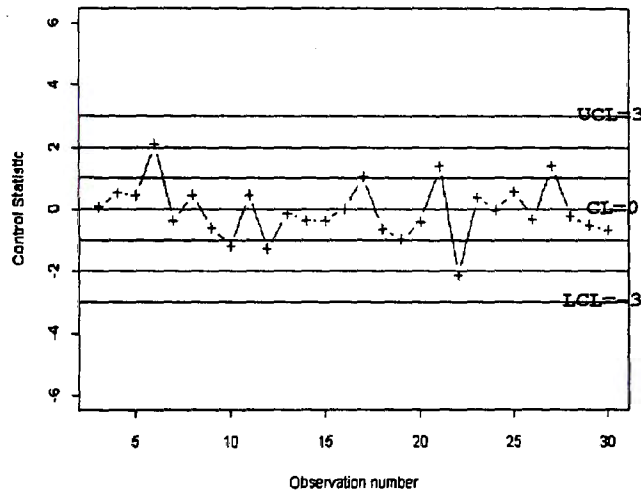


FIGURE 1C. Multivariate Control Chart for Example 1,  $\mu$  Known,  $\Sigma$  Unknown, Statistic (3a).

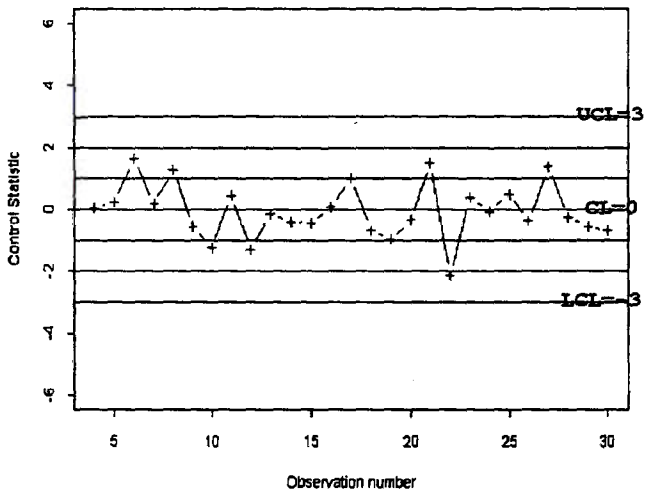


FIGURE 1D. Multivariate Control Chart for Example 1,  $\mu$  Known,  $\Sigma$  Unknown, Statistic (3b).

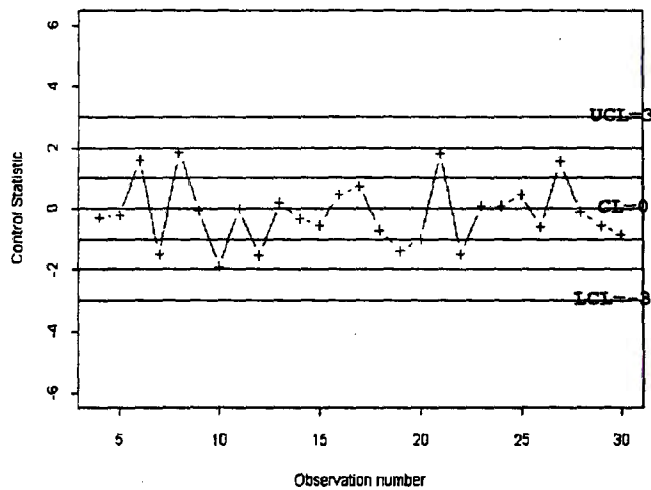


FIGURE 1E. Multivariate Control Chart for Example 1, Unknown Parameters.

Note that for the cases with some unknown parameters, the corresponding statistics were computed using the values of  $\mu$  and  $\Sigma$  as given above. Note also that since  $p = 2$ , calculations of the plotted values of the control statistics (2), (3a), (3b) and (4) have been started with the 2nd, 3rd, 4th and 4th observations respectively. The latter statistic is particularly useful as it enables exercise of control over new, start-up or short-run processes without requiring accumulation of a considerable amount of process data for estimation of the unknown parameters.

As shown in the figures, none of the plotted points exceed the control limits for each of the control charts. This is as expected because the data for this example can be regarded as having been collected from an *in-control* process. It is also interesting to note that after the first few observations, the movement of the charted points are very similar for all cases. This is especially true for the two charts based on formulae (3a) and (3b). This phenomenon is typical for in-control multivariate normal processes.

## Example 2

Next, to demonstrate the behaviour of the control charts based on *subgroup* data for a *stable* process, 120 observations have been generated from a trivariate normal distribution with

$$\mu = \begin{pmatrix} 5 \\ 12 \\ 15 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} 0.0625 & 0.10625 & 0.05 \\ 0.10625 & 0.25 & 0.136 \\ 0.05 & 0.136 & 0.16 \end{pmatrix}$$

These observations are grouped into samples of size  $n = 4$  and the corresponding values of the control statistics (5), (6), (7a), (7b) and (8) were calculated and plotted in Figures 2A to 2E. Note that except for the case with known parameters and when the statistic (7b) is used, charting begins with the 2nd subgroup.

Again, as shown in the figures, the point patterns of the resulting control charts are almost identical after the first few points. This similarity in the point patterns will generally be the case for other in-control processes and the appearance of the control charts will be more similar as the subgroup size increases. Following Quesenberry's (1991) suggestion, since  $\mu$  and  $\Sigma$  are not likely to be known precisely, the safer approach to charting is to use (4) and (8) which do not assume known values for the process parameters.

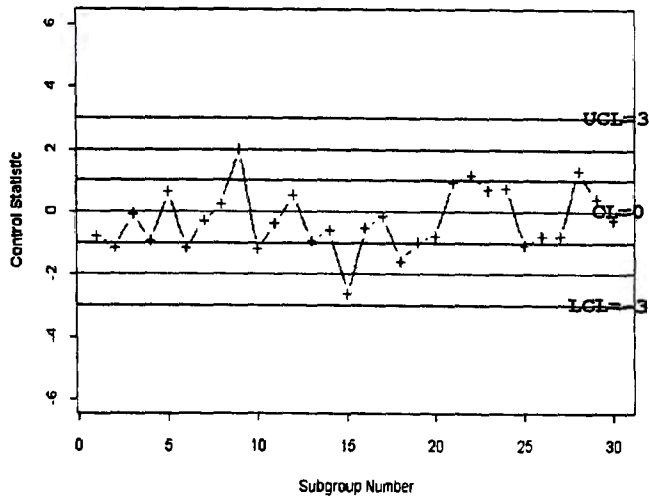


FIGURE 2A. Multivariate Control Chart for Example 2, Known Parameters.

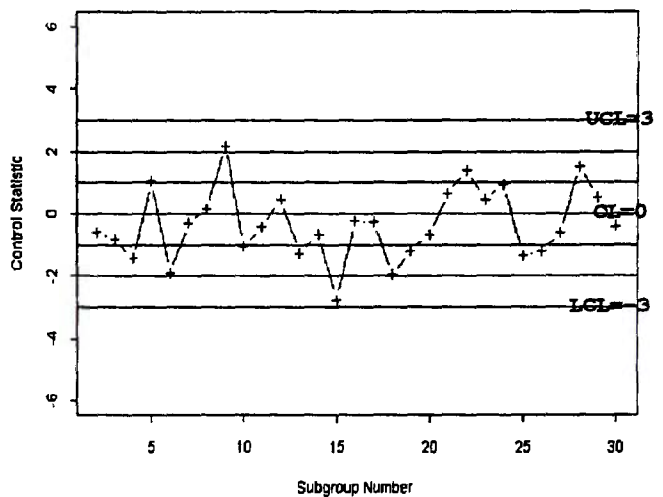


FIGURE 2B. Multivariate Control Chart for Example 2,  $\mu$  Unknown,  $\Sigma$  Known.

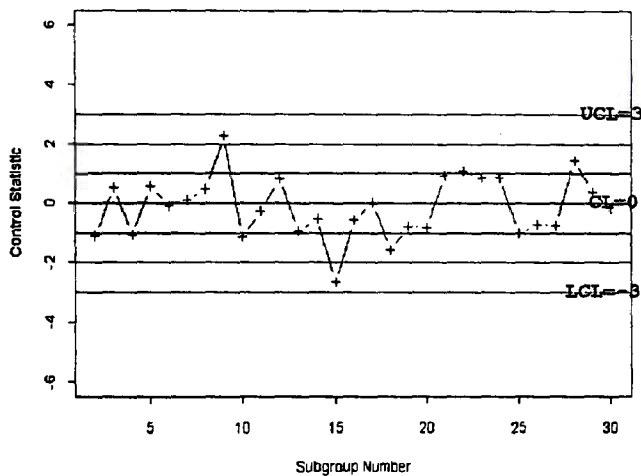


FIGURE 2C. Multivariate Control Chart for Example 2,  $\mu$  Known,  $\Sigma$  Unknown, Statistic (7a).

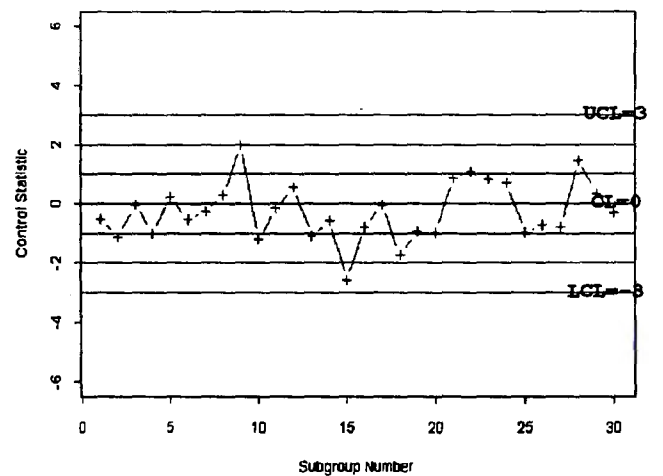


FIGURE 2D. Multivariate Control Chart for Example 2,  $\mu$  known,  $\Sigma$  Unknown, Statistic (7b).

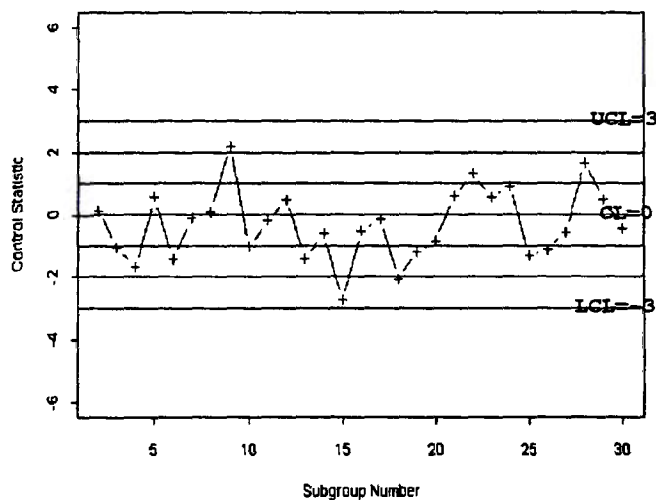


FIGURE 2E. Multivariate Control Chart for Example 2, Unknown Parameters.

### Example 3

In this example, two charts are shown in order to compare the control techniques based on *subgroup* data for the cases when the parameters  $\mu$  and  $\Sigma$  are assumed known and when they are both unknown. Figures 3A and 3B show these control charts for 30 samples of 3 observations each from a bivariate process where the mean of the 1st variable increases by 1.5 standard deviations after the 8th sample. The first 8 samples were generated from a  $N_2(\mu, \Sigma)$  process distribution with

$$\mu = \begin{pmatrix} 10 \\ 23 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} 0.25 & 0.375 \\ 0.375 & 1 \end{pmatrix}$$

whereas samples 9 to 30 were simulated from  $N_2(\mu_{new}, \Sigma)$  where

$$\mu_{new} = \begin{pmatrix} 10.75 \\ 23 \end{pmatrix}.$$

Note that the correct values of  $\mu$  and  $\Sigma$  to use for computing statistic (5) are those before the shift occurs. The figures show that whilst the shift is large enough to trigger out-of-control signals from both charts as soon as subgroup 9 is observed, that based on unknown parameters gradually settles into a pattern indicative of in-control conditions. This is due to the fact that the corresponding control statistic utilizes the current data stream to estimate the unknown values of the process parameters sequentially, causing the effect of parameter changes to 'dilute' as more out-of-control data are incorporated into the parameter estimates. It should be noted, that if an outlier or out-of-control observation (or subgroup) is present, that observation should be removed from subsequent computations. If this is not done, the parameter estimates will be distorted, causing an out-of-control process to appear in-control or vice versa.



FIGURE 3A. Multivariate Control Chart for Example 3, Known Parameters.

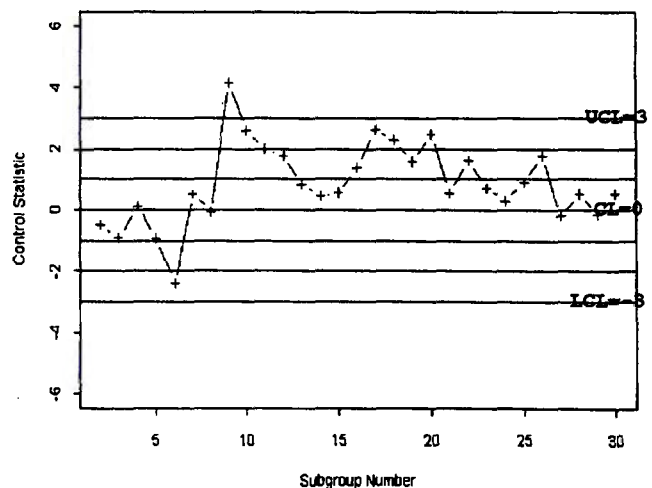


FIGURE 3B. Multivariate Control Chart for Example 3, Unknown Parameters.

## Example 4

As a preliminary investigation of the effect of changes in process standard deviations on the individual values control charts, 30 observations are considered that have been generated from a trivariate normal process where the standard deviations of two of the variables double after the 13th observation. Observations 1 to 13 were generated from  $N_3(\mu, \Sigma)$  with

$$\mu = \begin{pmatrix} 8 \\ 25 \\ 27 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} 0.01 & 0.042 & 0.0825 \\ 0.042 & 1.44 & -0.18 \\ 0.0825 & -0.18 & 2.25 \end{pmatrix}$$

while the remaining subsequent observations were generated from  $N_3(\mu, \Sigma_{new})$  where

$$\Sigma_{new} = \begin{pmatrix} 0.01 & 0.084 & 0.165 \\ 0.084 & 5.76 & -0.72 \\ 0.165 & -0.72 & 9 \end{pmatrix}$$

The *individual* values control charts for the case when both or neither of the process parameters are assumed known were constructed based on these simulated data and these are displayed in Figure 4A and 4B respectively. As shown in the figures, the change in the process standard deviations causes a spike on both the control charts at the 21st observation. However, the signal from the latter is less pronounced than that corresponding to the known parameter case. This example demonstrates that the individual values control chart which does not assume known values of the process parameters can also be used to detect changes in the process covariance matrix  $\Sigma$ .

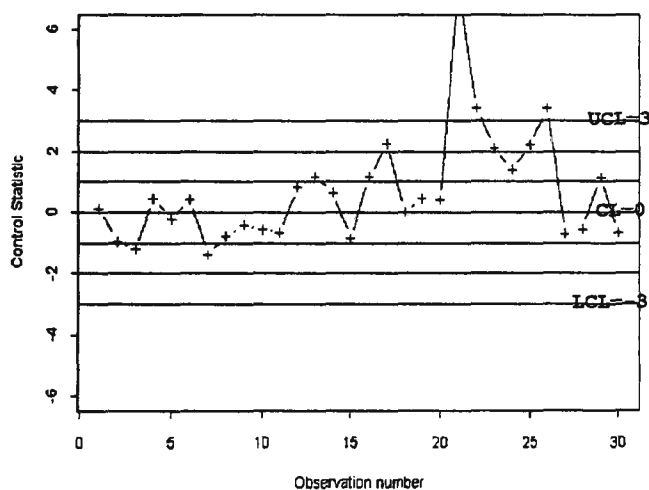


FIGURE 4A. Multivariate Control Chart for Example 4, Known Parameters.

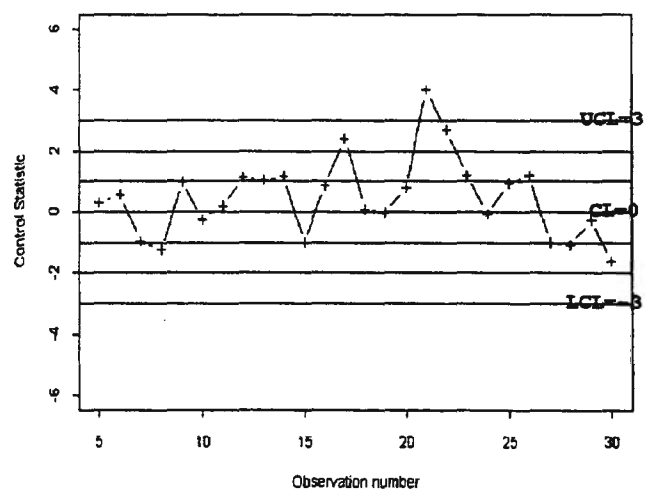


FIGURE 4B. Multivariate Control Chart for Example 4, Unknown Parameters.



## Example 5

To see how the control technique with unknown parameters that is based on subgroup data performs in comparison to an existing method, consider the data given by Alt et al.(1976). These authors presented formulae to compute the control limits for the  $T^2$ -type control chart based on a small number of subgroups, both for retrospective and future testing of the mean level of a multivariate normal process and used the data to illustrate the use of these so-called *small sample probability limits*. The data consists of measurements on  $p=2$  quality characteristics which are grouped into subgroups of size  $n=10$ . Due to limited space, the summary data are not reproduced here. However, for ease of comparison, the values of the  $T^2$ -type statistic, the stage 1 (retrospective) and stage 2 (future) control limits are given in TABLE 2, together with the results obtained using the technique (8) proposed in this paper. Note that the stage 1 and stage 2 control limits are set at  $\alpha=0.001$  and  $\alpha=0.005$  respectively.

**TABLE 2. Values of Alt et al.'s Test Statistic, Small Sample Probability Limits and Control Statistic (8) for Example 5.**

Subgroup No.	Alt et al.'s Stage 1 Control Statistic	Stage 1 UCL	Revised Value of Alt et al.'s Control Statistic	Revised Stage 1 UCL	Control Statistic (8)
1	0.009	1.3268	0.327	0.9546	NA
2	1.147	1.3268	0.264	0.9546	1.0758
3	0.136	1.3268	0.034	0.9546	-1.4909
4	4.901*	1.3268	NA		5.0192*
5	0.632	1.3268	0.057	0.9546	-0.5821†

Subgroup No.	Alt et al.'s Stage 2 ‡ Control Statistic	Stage 2 UCL ‡	Control Statistic (8)
6	0.392	1.5906	0.7766
7	0.197	1.5906	0.5906
8	4.594*	1.5906	4.8528*
9	0.190	1.5906	0.4045†
10	0.226	1.5906	0.3248
11	0.410	1.5906	1.0596
12	0.460	1.5906	0.9490

\* These numbers exceed their respective UCLs indicating the presence of assignable causes.

† These and subsequent values are calculated after removing the out-of control subgroups immediately preceding them.

‡ These are based on subgroups 1, 2, 3 and 5.

As shown in the table, the use of the proposed technique also results in an abnormally large value of the control statistic for subgroup 4 as when Alt et al.'s method is used, indicating that the process was out-of-control when this sample was taken. Similarly, subgroup 8 is also detected as being out-of-control using both control procedures.

# CONTROL PERFORMANCE

It is shown in Appendix B, that the *statistical* performance of the techniques presented above depend on the following parameter(s) (scalar, vector or matrix) for each of the given types of process changes (besides the *change point*  $r$ ) :

(a) A sustained shift in the Mean Vector from  $\mu$  to  $\mu_{new}$  whilst  $\Sigma$  remains unchanged ,

$$\lambda = \sqrt{(\mu_{new} - \mu)' \Sigma^{-1} (\mu_{new} - \mu)}$$

(b) A sustained shift in Covariance Matrix from  $\Sigma$  to  $\Sigma_{new}$  whilst  $\mu$  remains unchanged ,

$$\Omega = \Sigma^{-\frac{1}{2}} \Sigma_{new} \Sigma^{-\frac{1}{2}} \text{ or } \Sigma_{new}^{\frac{1}{2}} \Sigma^{-1} \Sigma_{new}^{\frac{1}{2}}$$

(c) A simultaneous sustained shift in Mean Vector from  $\mu$  to  $\mu_{new}$  and Covariance Matrix from  $\Sigma$  to  $\Sigma_{new}$  ,

$$\eta = \Sigma^{-\frac{1}{2}} (\mu_{new} - \mu)$$

and  $\Omega = \Sigma^{-\frac{1}{2}} \Sigma_{new} \Sigma^{-\frac{1}{2}}$

Note that  $\Sigma^{\frac{1}{2}}$  here denotes the symmetric *square root matrix* of  $\Sigma$  such that  $\Sigma = \Sigma^{\frac{1}{2}} \Sigma^{\frac{1}{2}}$  (see Johnson et al.(1988), p.51) and  $\Sigma^{-\frac{1}{2}} = (\Sigma^{\frac{1}{2}})^{-1}$ . The importance of these results is clear when one realizes that the effort for determining the control performance of the proposed techniques is greatly reduced. For instance, in order to determine the performance under the first type of process changes, it may be assumed, without loss of generality, that  $\mu = (0,0,\dots,0)'$ ,  $\Sigma = \mathbf{I}$  and  $\mu_{new}$  subsequently considered in the form of  $\mu_{new} = (\lambda,0,\dots,0)$  for various values of  $\lambda$ .

In this section, we will consider only the simplest type of process changes, namely, a persistent change in the process mean vector. The performance of the proposed techniques are evaluated on the basis of *probability of detection within  $m=5$  successive observations or subgroups* by means of simulation. This is chosen as the performance criterion instead of the common measure of ARL because, as demonstrated in the examples using simulated data, the first few observations or subgroups after the change appear to predominantly determine whether those techniques with unknown parameters are capable of 'picking up' the shift. In other words, if the mean shift is not detected within the first few observations or subgroups after its occurrence, it is even more unlikely that this will be 'picked up' by subsequent observations or subgroups because of the 'diluting' effect. Another reason is, the run length distributions for the techniques with some unknown parameter are not geometric so that ARL is clearly not a suitable performance criterion (see Quesenberry (1993)). Furthermore, this paper is particularly concerned with short production runs or low volume manufacturing and as such early response of the techniques to any process anomalies or irregularities is a crucial factor.

It should be pointed out that only an upper control limit is used in the simulation. This is chosen because the control techniques are intended primarily for 'picking up' changes in the mean vector and it appears that any change in the mean vector is likely to result in unusually

large values for the control statistics. However, in practice, it might be preferable to use both the lower and upper control limits because the former can provide protection against occasional changes in the variance-covariance matrix and other process disturbances which may cause abnormally small values for the control statistics. The limit is set at the 99.73th percentage point so that the false alarm rate for the proposed techniques equates to that of the traditional Shewhart charts with 3-sigma limits. As a partial check of the simulation, we have included results for those cases with known parameters.

The results for the individual values control techniques obtained through 10,000 simulation runs are tabulated in TABLE 3 for various combinations of  $p$ ,  $\lambda$  and  $r$ . In TABLE 4, the results for those techniques based on subgroup data are given. Note that the subgroup size used in each case is  $n = p + 1$ . This is the minimum common sample size that must be used if  $\Sigma$  is unknown except for technique (7a). Note also that, the exact probabilities for techniques (1) and (5) are obtainable from the *noncentral chi-square* distribution tables or standard statistical software packages. The simulation results for these techniques are found to agree well with the theoretical values. For instance, the theoretical probabilities for technique (1) are 0.0569, 0.3452, 0.8571 and 0.9972 respectively for  $\lambda = 1, 2, 3$  and 4 when  $p = 3$ . These are very close to the corresponding figures in Table 3.

As shown in TABLE 4, the control techniques based on subgroup data for the cases with at least some unknown parameters can be expected to perform as well as the technique with known parameters under this type of process change especially when the *noncentrality parameter*,  $\lambda$  is larger than or equal to 2. For instance, using control statistic (7b) and (8) with a subgroup size of  $n = 6$  when  $p = 5$ , the probabilities of 'picking up' a mean shift of  $\lambda = 2$  which occurs after the 10th subgroup, within 5 consecutive subgroups, are respectively 0.9966 and 0.9402. For smaller values of  $\lambda$ , these control techniques can also be expected to perform reasonably well relative to the technique corresponding to the known parameter case. For example, these probabilities are 0.3812 and 0.2834 respectively when statistics (7b) and (8) are used, as compared to 0.4513 for the known parameter case. As for control based on individual observations, those techniques with some unknown parameters have poor performance relative to that based on known parameters when  $\lambda$  and  $r$  are small and  $p$  is large. However, the performance of these individual values control charts improve with increasing value of  $\lambda$  and  $r$ . As shown in Table 3, the probability of detecting a shift of  $\lambda = 5$  for a bivariate process within 5 successive observations using statistic (4) which does not assume known values for the process parameters is 0.8514 if  $r = 20$ . Apart from these, a number of points can be noted from the tables. It is found that except for control statistic (7b), the proposed techniques decline in performance according to the number of unknown parameters on which they are based. Specifically, those based on known values of the process parameters have the best performance, followed by those with unknown mean vector  $\mu$ , those with unknown covariance matrix  $\Sigma$  and those with both process parameters unknown. Note also that, for the same  $\lambda$  and  $r$ , the performance of the individual values control techniques become worse as  $p$  increases.

**TABLE 3. Probability of Detection Within  $m = 5$  Subsequent Observations.**

$\lambda$	$r$	$p$	Control Statistic				
			(1)	(2)	(3a)	(3b)	(4)
1	10	2	0.0720	0.0492	0.0328	0.0327	0.0245
		3	0.0537	0.0399	0.0260	0.0260	0.0223
		5	0.0417	0.0306	0.0197	0.0179	0.0172
	20	2	0.0736	0.0588	0.0437	0.0432	0.0367
		3	0.0545	0.0432	0.0363	0.0364	0.0304
		5	0.0403	0.0345	0.0267	0.0265	0.0247
2	10	2	0.4430	0.2560	0.0807	0.0792	0.0541
		3	0.3437	0.2033	0.0553	0.0519	0.0364
		5	0.2500	0.1428	0.0342	0.0312	0.0228
	20	2	0.4357	0.3230	0.1660	0.1702	0.1353
		3	0.3456	0.2525	0.1172	0.1164	0.0925
		5	0.2508	0.1830	0.0755	0.0746	0.0614
3	10	2	0.9153	0.7050	0.1887	0.1720	0.1325
		3	0.8550	0.6014	0.1191	0.1080	0.0827
		5	0.7477	0.4752	0.0570	0.0514	0.0408
	20	2	0.9124	0.8074	0.4011	0.3976	0.3314
		3	0.8533	0.7220	0.2863	0.2854	0.2333
		5	0.7530	0.5920	0.1747	0.1662	0.1385
4	10	2	0.9993	0.9687	0.3735	0.3339	0.2728
		3	0.9966	0.9354	0.2483	0.2045	0.1967
		5	0.9880	0.8762	0.1026	0.0882	0.0722
	20	2	0.9994	0.9921	0.7040	0.6954	0.6213
		3	0.9960	0.9796	0.5624	0.5548	0.4837
		5	0.9900	0.9456	0.3576	0.3482	0.2962
5	10	2	1	0.9996	0.6114	0.5532	0.4880
		3	1	0.9982	0.4292	0.3628	0.3076
		5	1	0.9930	0.1560	0.1208	0.0984
	20	2	1	1	0.9096	0.9004	0.8514
		3	1	1	0.8122	0.7986	0.7366
		5	0.9999	0.9988	0.5866	0.5630	0.5074
6	10	2	1	1	0.8180	0.7594	0.6980
		3	1	1	0.6304	0.5430	0.4832
		5	1	0.9998	0.2624	0.1900	0.1568
	20	2	1	1	0.9862	0.9834	0.9702
		3	1	1	0.9500	0.9432	0.9162
		5	1	1	0.8094	0.7870	0.7468

**TABLE 4. Probability of Detection Within  $m = 5$  Subsequent Subgroups.**

$\lambda$	$r$	$p$	$n$	Control Statistic				
				(5)	(6)	(7a)	(7b)	(8)
1	10	2	3	0.2970	0.1794	0.1670	0.2102	0.1276
		3	4	0.3411	0.1981	0.1918	0.2381	0.1356
		5	6	0.4549	0.2692	0.2938	0.3638	0.2063
	20	2	3	0.2925	0.2128	0.2088	0.2192	0.1712
		3	4	0.3410	0.2521	0.2469	0.2787	0.2025
		5	6	0.4513	0.3256	0.3367	0.3812	0.2834
2	10	2	3	0.9864	0.8772	0.7356	0.8680	0.6512
		3	4	0.9966	0.9399	0.8637	0.9582	0.7996
		5	6	1	0.9908	0.9716	0.9966	0.9402
	20	2	3	0.9880	0.9460	0.8992	0.9440	0.8566
		3	4	0.9966	0.9774	0.9582	0.9816	0.9402
		5	6	1	0.9984	0.9952	0.9997	0.9926
3	10	2	3	1	1	0.9924	0.9990	0.9796
		3	4	1	1	0.9991	1	0.9979
		5	6	1	1	1	1	1
	20	2	3	1	1	1	1	0.9994
		3	4	1	1	1	1	0.9979
		5	6	1	1	1	1	1

## COMPUTATIONAL REQUIREMENTS

In this work, evaluation of the standard normal distribution, its inverse, chi-square and  $F$  distribution function are required in order to compute the transformed  $Z_k$  statistics. In addition, computation of the argument statistics  $T_k$ 's involve matrix multiplication and inversion. To implement the proposed control scheme, therefore, requires computing facilities and complex algorithms. Fortunately, these algorithms are widely available and have been built into most of the commercial statistical software packages. The simulated data and the charts in this paper were generated and made by the author using programs written in *S-plus*.

## CONCLUSIONS AND FURTHER REMARKS

Some methods have been presented for controlling the mean vector of a multivariate normal process. The techniques involve a non-linear probability integral transformation of some  $T^2$ -type statistics to sequences of independent or approximately independent standard normal variables. Thus, the resulting control charts can all be plotted in the same scale and with the same Shewhart control limits  $\pm 3$  irrespective of product types. Properly implemented, this standardization of the charting procedures can result in significant savings in time and

costs of charting. Those techniques presented for the case with no prior knowledge of process parameters are particularly attractive for short production runs and low volume manufacturing environments. These control procedures also enable the monitoring of new or start-up processes soon after production commences.

As demonstrated by the examples, the proposed techniques yield similar control chart patterns for a stable in-control process whether or not prior estimates of the process parameters are available. However, for a sustained shift in a parameter, whilst being capable of detecting the change, the control charts based on unknown parameters will eventually settle into an in-control pattern. The sooner the change occurs after the commencement of production, the lower the intensity of the signal and hence it is more likely to go undetected. The last example demonstrates that the control technique based on subgroup data and unknown parameters possesses comparable performance to an existing technique.

The power of the proposed techniques were determined by means of simulation for given sizes of sustained shift in the process mean vector. The results show that the techniques based on subgroup data have desirable performance whether or not the process parameters are assumed known in advance of the production run. As for individual values control techniques, the performance of those which do not assume known values for the process covariance matrix  $\Sigma$  or both the process parameters, improves with increasing value of the non-centrality parameter and point at which the change occurs. This is true irrespective of the number of quality characteristics considered.

This paper has been devoted solely to statistical process control and monitoring of the mean of multivariate processes. In practice, it is also of value to monitor the process spread as measured by the variance-covariance matrix  $\Sigma$  which may be subject to occasional changes. Although some methods have been proposed for this purpose (see for eg., Alt et al.(1990)), the assumption is made that  $\Sigma$  is known or can be estimated based on sufficient relevant data. As such, these methods cannot be readily used in manufacturing situations where the total output of the production processes are low. In view of this, efforts should be made to develop new control procedures that do not depend on prior knowledge of  $\Sigma$ . Another issue of practical importance is the identification of variables which are responsible for out-of-control conditions as signalled by the appropriate multivariate control charts. Although this problem has received considerable attention recently, the techniques proposed for this are again based on the assumption that knowledge of  $\Sigma$  is available prior to production.

## ACKNOWLEDGMENT

The author wish to thank Associate Professor Neil Barnett for his helpful suggestions in preparing this manuscript.



# APPENDIX A

## Distributional Properties of The Control Statistics

As the arguments involved in establishing the distributional properties of all the control statistics are similar, this section will only consider the proof for that given by (4). Suppose  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_r$  are independent  $p$ -variate random observation vectors which have the same covariance structure and are distributed as

$$\mathbf{X}_j \sim N_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}) \quad j = 1, 2, \dots, r$$

where  $\boldsymbol{\Sigma}$  denotes the unknown non-singular variance-covariance matrix. Next, define

$$\mathbf{V}_j = \mathbf{X}_j - \bar{\mathbf{X}} \quad j = 1, 2, \dots, r$$

where

$$\bar{\mathbf{X}} = \frac{1}{r} \sum_{j=1}^r \mathbf{X}_j.$$

We wish to test the hypothesis

$$H_0 : \mu_1 = \mu_2 = \dots = \mu \text{ (say)} \quad \text{vs.} \quad H_A : \text{not all } \mu_j \text{'s are equal}$$

Under  $H_0$  and the assumption of a *constant* process variance-covariance matrix,

$$\mathbf{V} = (\mathbf{V}_1 \quad \mathbf{V}_2 \quad \dots \quad \mathbf{V}_r)' \sim N_{rp}(\boldsymbol{\mu}_V, \boldsymbol{\Sigma}_V)$$

where

$$\boldsymbol{\mu}_V = \mathbf{0}_{rp \times 1}, \quad \boldsymbol{\Sigma}_V = \boldsymbol{\Sigma} \otimes \mathbf{C},$$

$\otimes$  denotes the (*left*) *Kronecker Product* (see Graybill (1983), p.216) and

$$\mathbf{C} = \begin{pmatrix} \frac{r-1}{r} & -\frac{1}{r} & \dots & \dots & \dots & -\frac{1}{r} \\ -\frac{1}{r} & \frac{r-1}{r} & \dots & \dots & \dots & -\frac{1}{r} \\ \vdots & & \ddots & & & \vdots \\ \vdots & & & \ddots & & \vdots \\ -\frac{1}{r} & \dots & \dots & \dots & \dots & \frac{r-1}{r} \end{pmatrix}_{r \times r}$$

Some linear combinations of the  $p \times 1$  component vectors  $\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_r$  that are uncorrelated are now obtained using the standard principal components approach. Consider the following linear combinations :

$$\begin{aligned}
Y_1 &= a_{11}V_1 + a_{12}V_2 + \dots + a_{1r}V_r \\
Y_2 &= a_{21}V_1 + a_{22}V_2 + \dots + a_{2r}V_r \\
&\vdots \\
&\vdots \\
Y_r &= a_{r1}V_1 + a_{r2}V_2 + \dots + a_{rr}V_r
\end{aligned}$$

In matrix notation, these are represented by the following equation :

$$Y = \Gamma V = (\mathbf{I} \otimes \mathbf{A})V \quad (\text{A1})$$

where  $\mathbf{I}$  is a  $p \times p$  identity matrix and

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & \dots & a_{1r} \\ a_{21} & a_{22} & \dots & \dots & \vdots \\ \vdots & & & & \vdots \\ \vdots & & & & \vdots \\ a_{r1} & \dots & \dots & \dots & a_{rr} \end{pmatrix}_{r \times r}$$

Thus,

$$\begin{aligned}
\Sigma_Y &= \Gamma \Sigma_V \Gamma' \\
&= \Sigma \otimes \mathbf{A} \mathbf{C} \mathbf{A}'
\end{aligned} \quad (\text{A2})$$

Recall that the aim here is to produce vectors  $Y_1, Y_2, \dots, Y_r$  which are uncorrelated. This is the case if  $\mathbf{A}$  in (A2) is chosen to diagonalize the symmetric matrix  $\mathbf{C}$ . One choice for  $\mathbf{A}$  is

$$\mathbf{A} = \begin{pmatrix} \frac{1}{\sqrt{r}} & \frac{1}{\sqrt{r}} & \dots & \dots & \dots & \dots & \dots & \frac{1}{\sqrt{r}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & \dots & \dots & \dots & \dots & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & 0 & \dots & \dots & \dots & 0 \\ \frac{1}{\sqrt{4(4-1)}} & \frac{1}{\sqrt{4(4-1)}} & \frac{1}{\sqrt{4(4-1)}} & \frac{-3}{\sqrt{4(4-1)}} & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{1}{\sqrt{r(r-1)}} & \dots & \dots & \dots & \dots & \dots & \dots & \frac{-(r-1)}{\sqrt{r(r-1)}} \end{pmatrix}$$

where the rows of the matrix are the normalized eigenvectors of  $\mathbf{C}$ . Substituting  $\mathbf{A}$  into (A1) results in the following linear combinations :

$$\begin{aligned}
\mathbf{Y}_1 &= \mathbf{0}, \\
\mathbf{Y}_2 &= \frac{1}{\sqrt{2}}(\mathbf{V}_1 - \mathbf{V}_2) = \frac{1}{\sqrt{2}}(\mathbf{X}_1 - \mathbf{X}_2) \\
\mathbf{Y}_3 &= \frac{1}{\sqrt{6}}(\mathbf{V}_1 + \mathbf{V}_2 - 2\mathbf{V}_3) = \frac{1}{\sqrt{6}}(\mathbf{X}_1 + \mathbf{X}_2 - 2\mathbf{X}_3) \\
&\vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\
&\vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\
\mathbf{Y}_r &= \frac{1}{\sqrt{r(r-1)}} \left[ \sum_{j=1}^{r-1} \mathbf{V}_j - (r-1)\mathbf{V}_r \right] = \frac{1}{\sqrt{r(r-1)}} \left[ \sum_{j=1}^{r-1} \mathbf{X}_j - (r-1)\mathbf{X}_r \right]
\end{aligned}$$

Note that the transformation results in a new set of uncorrelated random vectors  $\mathbf{Y}_2, \mathbf{Y}_3, \dots, \mathbf{Y}_r$ , one less than the set of original random vectors. This is due to the fact that the transformation is subject to a constraint, namely, the sum of the component vectors  $\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_r$  is equal to a zero vector leading to  $\text{rank}(\Sigma_{\mathbf{V}}) = rp - p = r(p-1)$ . It is also clear from (A2) that  $\Sigma_{\mathbf{Y}}$  is a 'quasi-diagonal' matrix with diagonal submatrices  $\Sigma$  except for the first one which is a zero matrix.

Since the resulting transformed vectors  $\mathbf{Y}_2, \mathbf{Y}_3, \dots, \mathbf{Y}_r$  are linear combinations of multivariate normal vectors and  $\Sigma_{\mathbf{Y}}$  is a quasi-diagonal matrix as mentioned above, they are mutually independent with common variance-covariance matrix  $\Sigma$ . As the  $k$ th observation vector  $\mathbf{X}_k$ , the  $(k-1)$ th sample mean vector  $\bar{\mathbf{X}}_{k-1}$  and the sample variance-covariance matrix based on the first  $k$  observations  $\mathbf{S}_{k-1}$  are independent,

$$\begin{aligned}
A_k &= \mathbf{Y}_k' (\mathbf{S}_{k-1})^{-1} \mathbf{Y}_k \\
&= \frac{1}{k(k-1)} \left[ \sum_{j=1}^{k-1} \mathbf{X}_j - (k-1)\mathbf{X}_k \right]' (\mathbf{S}_{k-1})^{-1} \left[ \sum_{j=1}^{k-1} \mathbf{X}_j - (k-1)\mathbf{X}_k \right] \\
&= \left( \frac{k-1}{k} \right) (\mathbf{X}_k - \bar{\mathbf{X}}_{k-1})' (\mathbf{S}_{k-1})^{-1} (\mathbf{X}_k - \bar{\mathbf{X}}_{k-1}) \quad k = p+2, \dots
\end{aligned}$$

are easily seen to be distributed as (Anderson (1984), p.163)

$$A_k \sim \frac{(k-2)p}{(k-1-p)} F_{p, k-1-p}$$

To establish mutual independence of successive  $A_k$ 's, it is first proved that they are *pairwise* independent. Clearly,

$$\mathbf{Y}_k \sim N_p(\mathbf{0}, \Sigma), \quad \mathbf{Y}_{k+1} \sim N_p(\mathbf{0}, \Sigma), \quad (k-2)\mathbf{S}_{k-1} \sim \mathbf{W}_{k-2}(\bullet | \Sigma)$$

are mutually independent. The notation  $\mathbf{W}_\nu(\bullet | \Sigma)$  here denotes the *Wishart* distribution with  $\nu$  degrees of freedom and parameter  $\Sigma$ . Let  $\mathbf{D}$  be a non-singular matrix such that  $\mathbf{D}\Sigma\mathbf{D}' = \mathbf{I}$  and define

$$\mathbf{Y}_k^* = \mathbf{D}\mathbf{Y}_k, \quad \mathbf{Y}_{k+1}^* = \mathbf{D}\mathbf{Y}_{k+1}, \quad \mathbf{S}_{k-1}^* = \mathbf{D}\mathbf{S}_{k-1}\mathbf{D}'$$

Thus,

$$\mathbf{Y}_k^* \sim N_p(\mathbf{0}, \mathbf{I}), \quad \mathbf{Y}_{k+1}^* \sim N_p(\mathbf{0}, \mathbf{I}), \quad (k-2)\mathbf{S}_{k-1}^* \sim \mathbf{W}_{k-2}(\bullet|\mathbf{I})$$

and their independence is preserved. Due to the *invariance* property of the transformation,

$$A_k = \mathbf{Y}_k' \mathbf{S}_{k-1}^{-1} \mathbf{Y}_k = \mathbf{Y}_k^* (\mathbf{S}_{k-1}^*)^{-1} \mathbf{Y}_k^*$$

$$A_{k+1} = \mathbf{Y}_{k+1}' \mathbf{S}_k^{-1} \mathbf{Y}_{k+1} = \mathbf{Y}_{k+1}^* (\mathbf{S}_k^*)^{-1} \mathbf{Y}_{k+1}^*$$

Noting that

$$(k-2)\mathbf{S}_{k-1}^* = (k-1)\mathbf{S}_k^* - \mathbf{Y}_k^* \mathbf{Y}_k^{*'}$$

and using identity (2.5.6) for matrix inverses (see Press(1982), Binomial Inverse Theorem, p.23),  $A_k$  may be expressed as follows :

$$A_k = \frac{(k-2)}{(k-1)} \left[ \mathbf{Y}_k^* (\mathbf{S}_k^*)^{-1} \mathbf{Y}_k^* + \frac{(\mathbf{Y}_k^* (\mathbf{S}_k^*)^{-1} \mathbf{Y}_k^*)^2}{(k-1) - \mathbf{Y}_k^* (\mathbf{S}_k^*)^{-1} \mathbf{Y}_k^*} \right]$$

As  $\mathbf{Y}_k^* (\mathbf{S}_k^*)^{-1} \mathbf{Y}_k^*$  is independent of  $\mathbf{S}_k^*$  (see Srivastava and Khatri (1979), Theorem 3.3.6, p.94) and  $\mathbf{Y}_{k+1}^*$ , it is also independent of any function of  $\mathbf{S}_k^*$  and  $\mathbf{Y}_{k+1}^*$ . Thus,  $\mathbf{Y}_k^* (\mathbf{S}_k^*)^{-1} \mathbf{Y}_k^*$  is independent of  $A_{k+1} = \mathbf{Y}_{k+1}^* (\mathbf{S}_k^*)^{-1} \mathbf{Y}_{k+1}^*$  and it follows immediately that  $A_k$  (a function of  $\mathbf{Y}_k^* (\mathbf{S}_k^*)^{-1} \mathbf{Y}_k^*$ ) is independent of  $A_{k+1}$ . Similarly, it can be shown that any pair of  $A_k$ 's are pairwise independent. Note that

$$A_{k+2} = k \mathbf{Y}_{k+2}^* [(k-1)\mathbf{S}_k^* + \mathbf{Y}_{k+1}^* \mathbf{Y}_{k+1}^{*'}]^{-1} \mathbf{Y}_{k+2}^*$$

and since  $\mathbf{Y}_k^* (\mathbf{S}_k^*)^{-1} \mathbf{Y}_k^*$  is independent of  $\mathbf{S}_k^*$ ,  $\mathbf{Y}_{k+1}^*$  and  $\mathbf{Y}_{k+2}^*$ , it is clear that  $A_k$  is also independent of  $A_{k+2}$ . By induction,  $A_{k+m}$  and  $A_{k+n}$  ( $m \neq n$ ) are independent.

Using the result of pairwise independence, it is now possible to proceed to show that they are mutually independent. As  $\mathbf{Y}_k^* (\mathbf{S}_k^*)^{-1} \mathbf{Y}_k^*$ ,  $\mathbf{S}_k^*$ ,  $\mathbf{Y}_{k+1}^*$  and  $\mathbf{Y}_{k+2}^*$  are mutually independent, their joint probability density function is

$$f(\mathbf{Y}_k^* \mathbf{S}_k^{*-1} \mathbf{Y}_k^*, \mathbf{S}_k^*, \mathbf{Y}_{k+1}^*, \mathbf{Y}_{k+2}^*) = g(\mathbf{Y}_k^* \mathbf{S}_k^{*-1} \mathbf{Y}_k^*) h(\mathbf{S}_k^*) q(\mathbf{Y}_{k+1}^*) z(\mathbf{Y}_{k+2}^*)$$

where  $g, h, q$  and  $z$  represent their marginal probability density functions respectively. The moment generating function of

$$A_k = \mathbf{Y}_k^* \mathbf{S}_{k-1}^{*-1} \mathbf{Y}_k^* = f_1(\mathbf{Y}_k^* \mathbf{S}_k^{*-1} \mathbf{Y}_k^*)$$

$$A_{k+1} = \mathbf{Y}_{k+1}^* \mathbf{S}_k^{*-1} \mathbf{Y}_{k+1}^* = f_2(\mathbf{S}_k^*, \mathbf{Y}_{k+1}^*)$$

$$A_{k+2} = \mathbf{Y}_{k+2}^* \mathbf{S}_{k+1}^{*-1} \mathbf{Y}_{k+2}^* = f_3(\mathbf{S}_k^*, \mathbf{Y}_{k+1}^*, \mathbf{Y}_{k+2}^*)$$

where  $f_1, f_2$  and  $f_3$  are some functions of the indicated arguments, is given by

$$M_{A_k, A_{k+1}, A_{k+2}}(t_1, t_2, t_3) = \int_{\substack{\text{all possible} \\ (A_k, A_{k+1}, A_{k+2})}} e^{t_1 A_k + t_2 A_{k+1} + t_3 A_{k+2}} Q(A_k, A_{k+1}, A_{k+2}) dA_k dA_{k+1} dA_{k+2}$$

Integrating over the *original* space gives,

$$M_{A_k, A_{k+1}, A_{k+2}}(t_1, t_2, t_3) = \int_{\substack{\text{all possible} \\ (\mathbf{Y}_k^* \mathbf{S}_k^{*-1} \mathbf{Y}_k^*, \mathbf{S}_k^*, \mathbf{Y}_{k+1}^*, \mathbf{Y}_{k+2}^*)}} e^{t_1 f_1(\mathbf{Y}_k^* \mathbf{S}_k^{*-1} \mathbf{Y}_k^*) + t_2 f_2(\mathbf{S}_k^*, \mathbf{Y}_{k+1}^*) + t_3 f_3(\mathbf{S}_k^*, \mathbf{Y}_{k+1}^*, \mathbf{Y}_{k+2}^*)} f(\mathbf{Y}_k^* \mathbf{S}_k^{*-1} \mathbf{Y}_k^*, \mathbf{S}_k^*, \mathbf{Y}_{k+1}^*, \mathbf{Y}_{k+2}^*) d(\mathbf{Y}_k^* \mathbf{S}_k^{*-1} \mathbf{Y}_k^*) d\mathbf{S}_k^* d\mathbf{Y}_{k+1}^* d\mathbf{Y}_{k+2}^*$$

$$= \int_{\substack{\text{all possible} \\ (\mathbf{Y}_k^* \mathbf{S}_k^{*-1} \mathbf{Y}_k^*, \mathbf{S}_k^*, \mathbf{Y}_{k+1}^*, \mathbf{Y}_{k+2}^*)}} e^{t_1 f_1(\mathbf{Y}_k^* \mathbf{S}_k^{*-1} \mathbf{Y}_k^*) + t_2 f_2(\mathbf{S}_k^*, \mathbf{Y}_{k+1}^*) + t_3 f_3(\mathbf{S}_k^*, \mathbf{Y}_{k+1}^*, \mathbf{Y}_{k+2}^*)} g(\mathbf{Y}_k^* \mathbf{S}_k^{*-1} \mathbf{Y}_k^*) h(\mathbf{S}_k^*) q(\mathbf{Y}_{k+1}^*) z(\mathbf{Y}_{k+2}^*) d(\mathbf{Y}_k^* \mathbf{S}_k^{*-1} \mathbf{Y}_k^*) d\mathbf{S}_k^* d\mathbf{Y}_{k+1}^* d\mathbf{Y}_{k+2}^*$$

As  $\mathbf{Y}_k^* \mathbf{S}_k^{*-1} \mathbf{Y}_k^*$  and  $\mathbf{S}_k^*$  are independent and the region (or space) for two independent variables over which the joint density is not zero must 'factor' (see Lindgren (1973), p.96 and Hogg et al.(1971), p.78), from the above,

$$M_{A_k, A_{k+1}, A_{k+2}}(t_1, t_2, t_3) = \int_{\substack{\text{all possible} \\ \mathbf{Y}_k^* \mathbf{S}_k^{*-1} \mathbf{Y}_k^*}} e^{t_1 f_1(\mathbf{Y}_k^* \mathbf{S}_k^{*-1} \mathbf{Y}_k^*)} g(\mathbf{Y}_k^* \mathbf{S}_k^{*-1} \mathbf{Y}_k^*) d(\mathbf{Y}_k^* \mathbf{S}_k^{*-1} \mathbf{Y}_k^*)$$

$$= \int_{\substack{\text{all possible} \\ (\mathbf{S}_k^*, \mathbf{Y}_{k+1}^*, \mathbf{Y}_{k+2}^*)}} e^{t_2 f_2(\mathbf{S}_k^*, \mathbf{Y}_{k+1}^*) + t_3 f_3(\mathbf{S}_k^*, \mathbf{Y}_{k+1}^*, \mathbf{Y}_{k+2}^*)} h(\mathbf{S}_k^*) q(\mathbf{Y}_{k+1}^*) z(\mathbf{Y}_{k+2}^*) d\mathbf{S}_k^* d\mathbf{Y}_{k+1}^* d\mathbf{Y}_{k+2}^*$$

$$= M_{A_k}(t_1) \cdot M_{A_{k+1}, A_{k+2}}(t_2, t_3)$$

$$= M_{A_k}(t_1) \cdot M_{A_{k+1}}(t_2) \cdot M_{A_{k+2}}(t_3) \quad \because A_{k+1} \text{ and } A_{k+2} \text{ are independent}$$

where  $M_{A_k}$ ,  $M_{A_{k+1}}$  and  $M_{A_{k+2}}$  are the moment generating functions of  $A_k, A_{k+1}$  and  $A_{k+2}$  respectively. Hence,  $A_k, A_{k+1}$  and  $A_{k+2}$  are *mutually independent*. The proof can be extended to any set of  $A_k$ 's in a similar manner.

(Q. E. D.)

## APPENDIX B

### (a) Dependence of Statistical Performance on the Non-Centrality Parameter, $\lambda$

Assume that the process variance-covariance matrix,  $\Sigma$  is *constant* but that the process mean vector may change from  $\mu$  to  $\mu_{new}$  at an arbitrary point in time. It is shown below that for the same *change point*,  $r$ , the joint distribution of the  $T_k$ 's (or equivalently,  $Z_k$ 's) and hence the *statistical performance* of the control techniques presented in section 2 depends on  $\mu$ ,  $\mu_{new}$  and  $\Sigma$  only through the value of the *noncentrality parameter*,

$$\lambda = \sqrt{(\mu_{new} - \mu)' \Sigma^{-1} (\mu_{new} - \mu)}$$

To establish the proof for the above, use is made of the following lemmas and theorems which are adapted from Crosier (1988) and Lowry (1989).

#### Lemma 1

If  $\mathbf{X}_k^* = \mathbf{M}\mathbf{X}_k$ ,  $\mathbf{X}_k^{**} = \mathbf{M}(\mathbf{X}_k - \mu)$ ,  $\mathbf{X}_{kj}^* = \mathbf{M}\mathbf{X}_{kj}$  and  $\mathbf{X}_{kj}^{**} = \mathbf{M}(\mathbf{X}_{kj} - \mu)$ ,  $k=1,2,\dots, j=1,2,\dots,n$  where  $\mathbf{M}$  is a  $p \times p$  matrix of full rank, then the relevant  $T_k$  statistics have the same values whether they are computed from  $\mathbf{X}_k$ ,  $(\mathbf{X}_k - \mu)$ ,  $\mathbf{X}_{kj}$  and  $(\mathbf{X}_{kj} - \mu)$  or the corresponding transformed vectors  $\mathbf{X}_k^*$ ,  $\mathbf{X}_k^{**}$ ,  $\mathbf{X}_{kj}^*$  and  $\mathbf{X}_{kj}^{**}$ , i.e, the  $T_k$  statistics are *invariant* with respect to these transformations. In other words, a full rank linear transformation of the observation vectors or their deviations from target (known mean vector) has no effect on the  $T_k$  statistics.

Proof : Immediate.

#### Lemma 2

If  $\mathbf{X}_k^* = \mathbf{M}\mathbf{X}_k$ ,  $\mathbf{X}_k^{**} = \mathbf{M}(\mathbf{X}_k - \mu)$ ,  $\mathbf{X}_{kj}^* = \mathbf{M}\mathbf{X}_{kj}$  and  $\mathbf{X}_{kj}^{**} = \mathbf{M}(\mathbf{X}_{kj} - \mu)$ ,  $k=1,2,\dots, j=1,2,\dots,n$  where  $\mathbf{M}$  is a  $p \times p$  matrix of full rank, then

$$\mu^* = E(\mathbf{X}_k^*) = E(\mathbf{X}_{kj}^*) = \begin{cases} \mu_{old}^* = \mathbf{M}\mu & , k \leq r \\ \mu_{new}^* = \mathbf{M}\mu_{new} & , k > r \end{cases}$$

$$\mu^{**} = E(\mathbf{X}_k^{**}) = E(\mathbf{X}_{kj}^{**}) = \begin{cases} \mu_{old}^{**} = 0 & , k \leq r \\ \mu_{new}^{**} = \mathbf{M}(\mu_{new} - \mu) & , k > r \end{cases}$$

$$\Sigma^* = \Sigma_{\mathbf{M}\mathbf{X}_k} = \Sigma_{\mathbf{M}\mathbf{X}_{kj}} = \Sigma^{**} = \Sigma_{\mathbf{M}(\mathbf{X}_k - \mu)} = \Sigma_{\mathbf{M}(\mathbf{X}_{kj} - \mu)} = \mathbf{M}\Sigma\mathbf{M}'$$



where  $r$  is the *change point* i.e the observation or subgroup number after which the process mean vector changes from  $\mu$  to  $\mu_{new}$ . Thus,

$$\begin{aligned} (\mu_{new}^* - \mu_{old}^*)' \Sigma^{*-1} (\mu_{new}^* - \mu_{old}^*) &= (\mu_{new} - \mu)' \Sigma^{-1} (\mu_{new} - \mu) \\ (\mu_{new}^{**} - \mu_{old}^{**})' \Sigma^{**^{-1}} (\mu_{new}^{**} - \mu_{old}^{**}) &= (\mu_{new} - \mu)' \Sigma^{-1} (\mu_{new} - \mu) \end{aligned}$$

**Proof** : Immediate.

This result implies that the non-centrality parameter has the same value whether computed from the original dependent variables or from some linearly independent combinations of them (or their deviations from targets or known means).

### **Lemma 3**

If  $(\mu_{1new} - \mu_1)' \Sigma^{-1} (\mu_{1new} - \mu_1) = (\mu_{2new} - \mu_2)' \Sigma^{-1} (\mu_{2new} - \mu_2)$  , there exists a nonsingular matrix  $\mathbf{M}$  such that

$$\begin{aligned} (\mu_{1new} - \mu_1) &= \mathbf{M}(\mu_{2new} - \mu_2) \\ \mathbf{M} \Sigma \mathbf{M}' &= \Sigma \end{aligned}$$

**Proof** :

First, transform each variable of the form  $(\mathbf{X}_{After} - \mathbf{X}_{Before})$  to  $\mathbf{Y}$ , principal components scaled to have unit variances where  $\mathbf{X}_{Before}$  and  $\mathbf{X}_{After}$  respectively denote observation vectors *before* and *after* the change in mean vector. Let  $E(\mathbf{Y}) = \mathbf{V}$ , by lemma 2,

$$\begin{aligned} \mathbf{V}_1' \mathbf{V}_1 &= \mathbf{V}_2' \mathbf{V}_2 \quad \text{where } \mathbf{V}_1 = \mathbf{D}^{-\frac{1}{2}} \mathbf{P}(\mu_{1new} - \mu_1) \\ \mathbf{V}_2 &= \mathbf{D}^{-\frac{1}{2}} \mathbf{P}(\mu_{2new} - \mu_2) \end{aligned}$$

and  $\mathbf{P}$  is an orthogonal matrix that diagonalizes  $\Sigma_{(\mathbf{X}_{After} - \mathbf{X}_{Before})}$  giving

$$\mathbf{P} \Sigma_{(\mathbf{X}_{After} - \mathbf{X}_{Before})} \mathbf{P}' = \mathbf{D}$$

Hence, there exists an orthogonal matrix  $\mathbf{Q}$  such that

$$\mathbf{V}_1 = \mathbf{Q} \mathbf{V}_2$$

Upon substituting  $\mathbf{V}_1 = \mathbf{D}^{-\frac{1}{2}} \mathbf{P}(\mu_{1new} - \mu_1)$  and  $\mathbf{V}_2 = \mathbf{D}^{-\frac{1}{2}} \mathbf{P}(\mu_{2new} - \mu_2)$  into the equation above,  $\mathbf{M}$  is obtainable as follows :

$$\begin{aligned}
\mathbf{D}^{-\frac{1}{2}}\mathbf{P}(\boldsymbol{\mu}_{1_{new}} - \boldsymbol{\mu}_1) &= \mathbf{Q}\mathbf{D}^{-\frac{1}{2}}\mathbf{P}(\boldsymbol{\mu}_{2_{new}} - \boldsymbol{\mu}_2) \\
(\boldsymbol{\mu}_{1_{new}} - \boldsymbol{\mu}_1) &= \mathbf{P}^{-1}\mathbf{D}^{\frac{1}{2}}\mathbf{Q}\mathbf{D}^{-\frac{1}{2}}\mathbf{P}(\boldsymbol{\mu}_{2_{new}} - \boldsymbol{\mu}_2) \\
&= \mathbf{P}'\mathbf{D}^{\frac{1}{2}}\mathbf{Q}\mathbf{D}^{-\frac{1}{2}}\mathbf{P}(\boldsymbol{\mu}_{2_{new}} - \boldsymbol{\mu}_2) \\
\therefore \mathbf{M} &= \mathbf{P}'\mathbf{D}^{\frac{1}{2}}\mathbf{Q}\mathbf{D}^{-\frac{1}{2}}\mathbf{P}
\end{aligned}$$

Now, it is shown that  $\mathbf{M}\boldsymbol{\Sigma}\mathbf{M}' = \boldsymbol{\Sigma}$ . We have

$$\begin{aligned}
\mathbf{M}\boldsymbol{\Sigma}_{(\mathbf{X}_{After}-\mathbf{X}_{Before})}\mathbf{M}' &= \mathbf{P}'\mathbf{D}^{\frac{1}{2}}\mathbf{Q}\mathbf{D}^{-\frac{1}{2}}\mathbf{P}\boldsymbol{\Sigma}_{(\mathbf{X}_{After}-\mathbf{X}_{Before})}\mathbf{P}'\mathbf{D}^{-\frac{1}{2}}\mathbf{Q}'\mathbf{D}^{\frac{1}{2}}\mathbf{P} \\
&= \mathbf{P}'\mathbf{D}^{\frac{1}{2}}\mathbf{Q}\mathbf{D}^{-\frac{1}{2}}\mathbf{D}\mathbf{D}^{-\frac{1}{2}}\mathbf{Q}'\mathbf{D}^{\frac{1}{2}}\mathbf{P} \\
&= \mathbf{P}'\mathbf{D}\mathbf{P} \\
&= \boldsymbol{\Sigma}_{(\mathbf{X}_{After}-\mathbf{X}_{Before})} \\
\Rightarrow \mathbf{M}(2\boldsymbol{\Sigma})\mathbf{M}' &= 2\boldsymbol{\Sigma} \\
\Rightarrow \mathbf{M}\boldsymbol{\Sigma}\mathbf{M}' &= \boldsymbol{\Sigma}
\end{aligned}$$

(Q.E.D.)

This lemma is also applicable to cases with known  $\boldsymbol{\mu}$ , namely, if  $(\boldsymbol{\mu}_{1_{new}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_{1_{new}} - \boldsymbol{\mu}) = (\boldsymbol{\mu}_{2_{new}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_{2_{new}} - \boldsymbol{\mu})$ , there exists a nonsingular matrix  $\mathbf{M}$  such that

$$\begin{aligned}
(\boldsymbol{\mu}_{1_{new}} - \boldsymbol{\mu}) &= \mathbf{M}(\boldsymbol{\mu}_{2_{new}} - \boldsymbol{\mu}) \\
\mathbf{M}\boldsymbol{\Sigma}\mathbf{M}' &= \boldsymbol{\Sigma}
\end{aligned}$$

The proof for this is similar to the above except that  $\boldsymbol{\mu}_1$  and  $\boldsymbol{\mu}_2$  should both be replaced by  $\boldsymbol{\mu}$ .

### Theorem 1

For the cases with unknown  $\boldsymbol{\mu}$ , if  $(\boldsymbol{\mu}_{1_{new}} - \boldsymbol{\mu}_1)' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_{1_{new}} - \boldsymbol{\mu}_1) = (\boldsymbol{\mu}_{2_{new}} - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_{2_{new}} - \boldsymbol{\mu}_2)$ ,

then  $f\left(T_k's \middle| E(\mathbf{X}_k \text{ or } \mathbf{X}_{kj}) = \begin{cases} \boldsymbol{\mu}_1 & , k \leq r \\ \boldsymbol{\mu}_{1_{new}} & , k > r \end{cases} = f\left(T_k's \middle| E(\mathbf{X}_k \text{ or } \mathbf{X}_{kj}) = \begin{cases} \boldsymbol{\mu}_2 & , k \leq r \\ \boldsymbol{\mu}_{2_{new}} & , k > r \end{cases} \right)$  where

$f\left(T_k's \middle| E(\mathbf{X}_k \text{ or } \mathbf{X}_{kj}) = \begin{cases} \boldsymbol{\mu}_1 & , k \leq r \\ \boldsymbol{\mu}_{1_{new}} & , k > r \end{cases} \right)$  denotes the joint pdf of  $T_k's$  given

$$E(\mathbf{X}_k) = E(\mathbf{X}_{kj}) = \begin{cases} \boldsymbol{\mu}_1 & , k \leq r \\ \boldsymbol{\mu}_{1_{new}} & , k > r \end{cases}$$

and  $f\left(T_k's \middle| E(\mathbf{X}_k \text{ or } \mathbf{X}_{kj}) = \begin{cases} \boldsymbol{\mu}_2 & , k \leq r \\ \boldsymbol{\mu}_{2_{new}} & , k > r \end{cases} \right)$  denotes the joint pdf of  $T_k's$  given

$$E(\mathbf{X}_k) = E(\mathbf{X}_{kj}) = \begin{cases} \boldsymbol{\mu}_2 & , k \leq r \\ \boldsymbol{\mu}_{2_{new}} & , k > r \end{cases}$$

where  $r$  is the change point. This theorem implies equivalent performance of the control techniques under the two alternative probability densities.

**Proof :**

Note that the  $T_k$ 's of (2), (4), (6) and (8) are expressible either in terms of random vectors of the form  $(\mathbf{X}_p - \mathbf{X}_q)$  for  $p \neq q$  or  $(\mathbf{X}_{pi} - \mathbf{X}_{qj})$  for  $p \neq q, i, j = 1, 2, \dots, n$  and  $p = q, i \neq j, i, j = 1, 2, \dots, n$ . Let *pdf 1* refer to the multivariate normal density specified by

$$E(\mathbf{X}_k) = E(\mathbf{X}_{kj}) = \begin{cases} \mu_1 & , k \leq r \\ \mu_{1new} & , k > r \end{cases}$$

and *pdf 2* refer to the multivariate normal density specified by

$$E(\mathbf{X}_k) = E(\mathbf{X}_{kj}) = \begin{cases} \mu_2 & , k \leq r \\ \mu_{2new} & , k > r \end{cases}$$

If *pdf 2* is expressed in the transformed variates  $\mathbf{Z}_k = \mathbf{M}\mathbf{X}_k$  (or  $\mathbf{Z}_{kj} = \mathbf{M}\mathbf{X}_{kj}$ ) where  $\mathbf{M} = \mathbf{P}'\mathbf{D}^{\frac{1}{2}}\mathbf{Q}\mathbf{D}^{-\frac{1}{2}}\mathbf{P}$  is as given in lemma 3, then  $Var(\mathbf{Z}_k) = Var(\mathbf{Z}_{kj}) = \mathbf{M}\Sigma\mathbf{M}' = \Sigma$  and

$$\begin{aligned} E(\mathbf{Z}_p - \mathbf{Z}_q) &= E(\mathbf{Z}_{pi} - \mathbf{Z}_{qj}) \\ &= \begin{cases} \mathbf{0} & \text{if } p, q \leq r \text{ or } p, q > r \\ \mu_{1new} - \mu_1 & \text{if } p > r, q \leq r \\ \mu_1 - \mu_{1new} & \text{if } p \leq r, q > r \end{cases} \end{aligned}$$

It can also easily be shown that the covariance of the transformed variates  $(\mathbf{Z}_p - \mathbf{Z}_q)$  and  $(\mathbf{Z}_s - \mathbf{Z}_t)$  under *pdf 2* is the same as the covariance of  $(\mathbf{X}_p - \mathbf{X}_q)$   $[(\mathbf{X}_{pi} - \mathbf{X}_{qj})]$  and  $(\mathbf{X}_s - \mathbf{X}_t)$   $[(\mathbf{X}_{su} - \mathbf{X}_{tv})]$  under *pdf 1*, i.e

$$\begin{aligned} Cov[(\mathbf{Z}_p - \mathbf{Z}_q), (\mathbf{Z}_s - \mathbf{Z}_t)] &= Cov[(\mathbf{X}_p - \mathbf{X}_q), (\mathbf{X}_s - \mathbf{X}_t)] \\ \text{under pdf 2} & \qquad \qquad \qquad \text{under pdf 1} \end{aligned}$$

$$\left\{ \begin{aligned} Cov[(\mathbf{Z}_{pi} - \mathbf{Z}_{qj}), (\mathbf{Z}_{su} - \mathbf{Z}_{tv})] &= Cov[(\mathbf{X}_{pi} - \mathbf{X}_{qj}), (\mathbf{X}_{su} - \mathbf{X}_{tv})] \\ \text{under pdf 2} & \qquad \qquad \qquad \text{under pdf 1} \end{aligned} \right\}$$

for all  $p, q, s$  and  $t$  ( $p, q, s, t, i, j, u$  and  $v$ ). Hence, *pdf 2* expressed in the transformed variates  $\mathbf{Z}_k$  ( $\mathbf{Z}_{kj}$ ) is the same as *pdf 1* expressed in  $\mathbf{X}_k$  ( $\mathbf{X}_{kj}$ ) variates, i.e

$$f\left(T_k' s \middle| E(\mathbf{X}_k \text{ or } \mathbf{X}_{kj})\right) = \begin{cases} \mu_1 & , k \leq r \\ \mu_{1_{new}} & , k > r \end{cases} = f\left(T_k' s \middle| E(\mathbf{Z}_k \text{ or } \mathbf{Z}_{kj})\right) = \begin{cases} \mathbf{M}\mu_2 & , k \leq r \\ \mathbf{M}\mu_{2_{new}} & , k > r \end{cases} \quad (\text{B1})$$

By lemma 1, the values of the  $T_k$ 's are invariant with respect to the transformation from  $\mathbf{X}_k$  ( $\mathbf{X}_{kj}$ ) to  $\mathbf{Z}_k$  ( $\mathbf{Z}_{kj}$ ) so that

$$f\left(T_k' s \middle| E(\mathbf{X}_k \text{ or } \mathbf{X}_{kj})\right) = \begin{cases} \mu_2 & , k \leq r \\ \mu_{2_{new}} & , k > r \end{cases} = f\left(T_k' s \middle| E(\mathbf{Z}_k \text{ or } \mathbf{Z}_{kj})\right) = \begin{cases} \mathbf{M}\mu_2 & , k \leq r \\ \mathbf{M}\mu_{2_{new}} & , k > r \end{cases} \quad (\text{B2})$$

Combining (B1) and (B2) gives

$$f\left(T_k' s \middle| E(\mathbf{X}_k \text{ or } \mathbf{X}_{kj})\right) = \begin{cases} \mu_1 & , k \leq r \\ \mu_{1_{new}} & , k > r \end{cases} = f\left(T_k' s \middle| E(\mathbf{X}_k \text{ or } \mathbf{X}_{kj})\right) = \begin{cases} \mu_2 & , k \leq r \\ \mu_{2_{new}} & , k > r \end{cases}$$

Therefore, the joint distribution of  $T_k$ 's given  $\mu_1$  and  $\mu_{1_{new}}$  (and  $\Sigma$ ) is the same as the joint distribution of  $T_k$ 's given  $\mu_2$  and  $\mu_{2_{new}}$  (and  $\Sigma$ ) if

$$(\mu_{1_{new}} - \mu_1)' \Sigma^{-1} (\mu_{1_{new}} - \mu_1) = (\mu_{2_{new}} - \mu_2)' \Sigma^{-1} (\mu_{2_{new}} - \mu_2). \quad (\text{Q.E.D.})$$

Theorem 1 can be adapted to cases with known  $\mu$ , namely, if  $(\mu_{1_{new}} - \mu)' \Sigma^{-1} (\mu_{1_{new}} - \mu) = (\mu_{2_{new}} - \mu)' \Sigma^{-1} (\mu_{2_{new}} - \mu)$ , then

$$f\left(T_k' s \middle| E(\mathbf{X}_k \text{ or } \mathbf{X}_{kj})\right) = \begin{cases} \mu & , k \leq r \\ \mu_{1_{new}} & , k > r \end{cases} = f\left(T_k' s \middle| E(\mathbf{X}_k \text{ or } \mathbf{X}_{kj})\right) = \begin{cases} \mu & , k \leq r \\ \mu_{2_{new}} & , k > r \end{cases}$$

The proof for this is similar to the above except that  $\mathbf{X}_k$  ( $\mathbf{X}_{kj}$ ) and  $\mu_{i_{new}} - \mu_i$ ,  $i = 1, 2$  should be replaced by  $\mathbf{X}_k - \mu$  ( $\mathbf{X}_{kj} - \mu$ ) and  $\mu_{i_{new}} - \mu$  respectively.

## Theorem 2

For the cases with unknown  $\mu$ , let  $\mathbf{X}_{1k}$ 's ( $\mathbf{X}_{1kj}$ 's) be independent observation vectors from *pdf* 1 and  $\mathbf{X}_{2k}$ 's ( $\mathbf{X}_{2kj}$ 's) be independent observation vectors from *pdf* 2. Let *pdf* 1 be multivariate normal with mean vector  $\mu_1$  and  $\mu_{1_{new}}$  *before* and *after* the change and variance-covariance matrix  $\Sigma_1$ . Let *pdf* 2 be multivariate normal with mean vector  $\mu_2$  and  $\mu_{2_{new}}$  *before* and *after* the change and variance-covariance matrix  $\Sigma_2$ . Denote the *change point* by  $r$ . If  $(\mu_{1_{new}} - \mu_1)' \Sigma_1^{-1} (\mu_{1_{new}} - \mu_1) = (\mu_{2_{new}} - \mu_2)' \Sigma_2^{-1} (\mu_{2_{new}} - \mu_2)$ , then  $f_1(T_k' s) = f_2(T_k' s)$  where  $f_1(T_k' s)$  and  $f_2(T_k' s)$  represent the joint distribution of  $T_k$ 's given *pdf* 1 and *pdf* 2 respectively.

**Proof :**

Let  $\mathbf{X}_{1k}^* = \mathbf{D}_1^{-\frac{1}{2}} \mathbf{P}_1 \mathbf{X}_{1k}$  ( $\mathbf{X}_{1kj}^* = \mathbf{D}_1^{-\frac{1}{2}} \mathbf{P}_1 \mathbf{X}_{1kj}$ ) and  $\mathbf{X}_{2k}^* = \mathbf{D}_2^{-\frac{1}{2}} \mathbf{P}_2 \mathbf{X}_{2k}$  ( $\mathbf{X}_{2kj}^* = \mathbf{D}_2^{-\frac{1}{2}} \mathbf{P}_2 \mathbf{X}_{2kj}$ ) where  $\mathbf{P}_1$  and  $\mathbf{P}_2$  are matrices that diagonalize  $\Sigma_{(X_{1p}-X_{1q})}$  (or  $\Sigma_{(X_{1p'}-X_{1q'})}$ ) and  $\Sigma_{(X_{2p}-X_{2q})}$  (or  $\Sigma_{(X_{2p'}-X_{2q'})}$ ),  $p \neq q$  respectively, i.e

$$\mathbf{P}_1 \Sigma_{(X_{1p}-X_{1q})} \mathbf{P}_1' = \mathbf{D}_1$$

$$\mathbf{P}_2 \Sigma_{(X_{2p}-X_{2q})} \mathbf{P}_2' = \mathbf{D}_2$$

Then,

$$E(\mathbf{X}_{1p}^* - \mathbf{X}_{1q}^*) = \mu_{1p}^* - \mu_{1q}^* = \begin{cases} \mathbf{D}_1^{-\frac{1}{2}} \mathbf{P}_1 (\mu_{1new} - \mu_1) & , p > r, q \leq r \\ \mathbf{D}_1^{-\frac{1}{2}} \mathbf{P}_1 (\mu_1 - \mu_{1new}) & , p \leq r, q > r \\ 0 & \text{elsewhere} \end{cases}$$

$$E(\mathbf{X}_{2p}^* - \mathbf{X}_{2q}^*) = \mu_{2p}^* - \mu_{2q}^* = \begin{cases} \mathbf{D}_2^{-\frac{1}{2}} \mathbf{P}_2 (\mu_{2new} - \mu_2) & , p > r, q \leq r \\ \mathbf{D}_2^{-\frac{1}{2}} \mathbf{P}_2 (\mu_2 - \mu_{2new}) & , p \leq r, q > r \\ 0 & \text{elsewhere} \end{cases}$$

and  $\Sigma_{(X_{1p}^*-X_{1q}^*)} = \Sigma_{(X_{2p}^*-X_{2q}^*)} = \mathbf{I}$ .

By lemma 2,

$$(\mu_{1p}^* - \mu_{1q}^*)' \Sigma_{(X_{1p}^*-X_{1q}^*)}^{-1} (\mu_{1p}^* - \mu_{1q}^*) = (\mu_{1new} - \mu_1)' \Sigma_{(X_{1p}-X_{1q})}^{-1} (\mu_{1new} - \mu_1)$$

$$\Rightarrow (\mu_{1p}^* - \mu_{1q}^*)' \mathbf{I}^{-1} (\mu_{1p}^* - \mu_{1q}^*) = \frac{1}{2} (\mu_{1new} - \mu_1)' \Sigma_1^{-1} (\mu_{1new} - \mu_1)$$

and

$$(\mu_{2p}^* - \mu_{2q}^*)' \Sigma_{(X_{2p}^*-X_{2q}^*)}^{-1} (\mu_{2p}^* - \mu_{2q}^*) = (\mu_{2new} - \mu_2)' \Sigma_{(X_{2p}-X_{2q})}^{-1} (\mu_{2new} - \mu_2)$$

$$\Rightarrow (\mu_{2p}^* - \mu_{2q}^*)' \mathbf{I}^{-1} (\mu_{2p}^* - \mu_{2q}^*) = \frac{1}{2} (\mu_{2new} - \mu_2)' \Sigma_2^{-1} (\mu_{2new} - \mu_2)$$

for  $p > r, q \leq r$  or  $p \leq r, q > r$ . Because the values of  $T_k$ 's are invariant with respect to the transformations, the condition  $(\mu_{1new} - \mu_1)' \Sigma_1^{-1} (\mu_{1new} - \mu_1) = (\mu_{2new} - \mu_2)' \Sigma_2^{-1} (\mu_{2new} - \mu_2)$  is equivalent to  $(\mu_{1p}^* - \mu_{1q}^*)' \mathbf{I}^{-1} (\mu_{1p}^* - \mu_{1q}^*) = (\mu_{2p}^* - \mu_{2q}^*)' \mathbf{I}^{-1} (\mu_{2p}^* - \mu_{2q}^*)$  for  $p > r, q \leq r$  or  $p \leq r, q > r$ . Since  $(\mathbf{X}_{1p}^* - \mathbf{X}_{1q}^*)$  and  $(\mathbf{X}_{2p}^* - \mathbf{X}_{2q}^*)$  are principal components of  $(\mathbf{X}_{1p} - \mathbf{X}_{1q})$  and  $(\mathbf{X}_{2p} - \mathbf{X}_{2q})$  scaled to have the same variance-covariance matrix, the identity matrix  $\mathbf{I}$ , by theorem 1, the joint distribution of  $T_k$ 's given pdf 1 is the same as the joint distribution of  $T_k$ 's given pdf 2.

(Q.E.D.)

This theorem can also be adapted to the cases where  $\mu$  is known, namely, if

$(\mu_{1new} - \mu)' \Sigma_1^{-1} (\mu_{1new} - \mu) = (\mu_{2new} - \mu)' \Sigma_2^{-1} (\mu_{2new} - \mu)$ , then

$$f\left(T_k' s \middle| E(\mathbf{X}_k \text{ or } \mathbf{X}_{kj}) = \begin{cases} \mu & , k \leq r \\ \mu_{1new} & , k > r \end{cases}, \Sigma_1\right) = f\left(T_k' s \middle| E(\mathbf{X}_k \text{ or } \mathbf{X}_{kj}) = \begin{cases} \mu & , k \leq r \\ \mu_{2new} & , k > r \end{cases}, \Sigma_2\right).$$

The proof for this is along the same line of argument as above except that  $\mathbf{X}_{ik}$  ( $\mathbf{X}_{ikj}$ ) and  $\mu_{inew} - \mu_i$  should be replaced by  $\mathbf{X}_{ik} - \mu$  ( $\mathbf{X}_{ikj} - \mu$ ) and  $\mu_{inew} - \mu$  respectively for  $i = 1, 2$ .

### **(b) Dependence of Statistical Performance on $\Omega$ for a Change in Process Covariance Matrix**

Suppose that the process under consideration changes in covariance matrix from  $\Sigma$  to  $\Sigma_{new}$  after the  $r$ th observation, whilst the mean vector  $\mu$  remains constant. It is shown below that the joint distribution of the  $T_k$ 's (or equivalently,  $Z_k$ 's) for each of the control techniques depends on  $\mu$ ,  $\Sigma$  and  $\Sigma_{new}$  through the symmetric matrix

$$\Omega = \Sigma_{new}^{\frac{1}{2}} \Sigma^{-1} \Sigma_{new}^{\frac{1}{2}}$$

As an alternative, it can also be shown that the statistical performance of the control techniques depend on  $\mu$ ,  $\Sigma$  and  $\Sigma_{new}$  through the symmetric matrix  $\Sigma^{-\frac{1}{2}} \Sigma_{new} \Sigma^{-\frac{1}{2}}$ .

#### **Lemma 4**

If  $\Sigma_{1new}^{\frac{1}{2}} \Sigma_1^{-1} \Sigma_{1new}^{\frac{1}{2}} = \Sigma_{2new}^{\frac{1}{2}} \Sigma_2^{-1} \Sigma_{2new}^{\frac{1}{2}}$ , then there exists a full rank matrix  $\mathbf{D}$  such that

$$\begin{aligned} \mathbf{D} \Sigma_2 \mathbf{D}' &= \Sigma_1 \\ \text{and } \mathbf{D} \Sigma_{2new} \mathbf{D}' &= \Sigma_{1new}. \end{aligned}$$

#### **Proof**

From  $\Sigma_{1new}^{\frac{1}{2}} \Sigma_1^{-1} \Sigma_{1new}^{\frac{1}{2}} = \Sigma_{2new}^{\frac{1}{2}} \Sigma_2^{-1} \Sigma_{2new}^{\frac{1}{2}}$ , we have  $\Sigma_1 = \left( \Sigma_{1new}^{\frac{1}{2}} \Sigma_{2new}^{-\frac{1}{2}} \right) \Sigma_2 \left( \Sigma_{1new}^{\frac{1}{2}} \Sigma_{2new}^{-\frac{1}{2}} \right)'$ .

Since  $\mathbf{D} \Sigma_2 \mathbf{D}' = \Sigma_1$ ,  $\mathbf{D} = \Sigma_{1new}^{\frac{1}{2}} \Sigma_{2new}^{-\frac{1}{2}}$ . It can easily be verified that  $\mathbf{D} = \Sigma_{1new}^{\frac{1}{2}} \Sigma_{2new}^{-\frac{1}{2}}$  also satisfies the equation  $\mathbf{D} \Sigma_{2new} \mathbf{D}' = \Sigma_{1new}$ . Furthermore, it can readily be seen that  $\mathbf{D} = \Sigma_{1new}^{\frac{1}{2}} \Sigma_{2new}^{-\frac{1}{2}}$  is of full rank.

(Q.E.D.)

#### **Theorem 3**

Let *pdf1* be the probability density function of the independent multivariate normal observation vectors  $\mathbf{X}_k$ 's (or  $\mathbf{X}_{kj}$ 's) with mean vector  $\mu_1$ , covariance matrix  $\Sigma_1$  and  $\Sigma_{1new}$  *before* and *after* the change. Let *pdf2* be the probability density function of the independent multivariate normal observation vectors  $\mathbf{X}_k$ 's (or  $\mathbf{X}_{kj}$ 's) with mean vector  $\mu_2$ , covariance matrix  $\Sigma_2$  and



$\Sigma_{2new}$  before and after the change. Denote the change point by  $r$ . If  $\Sigma_{1new}^{\frac{1}{2}} \Sigma_1^{-1} \Sigma_{1new}^{\frac{1}{2}} = \Sigma_{2new}^{\frac{1}{2}} \Sigma_2^{-1} \Sigma_{2new}^{\frac{1}{2}}$ , then  $f_1(T_k's) = f_2(T_k's)$  where  $f_1(T_k's)$  and  $f_2(T_k's)$  represent the joint distribution of  $T_k's$  given pdf1 and pdf2 respectively.

### **Proof**

The proof here is similar to that of theorem 1 except that  $\mathbf{M}$  should be replaced by  $\mathbf{D} = \Sigma_{1new}^{\frac{1}{2}} \Sigma_{2new}^{-\frac{1}{2}}$  as given in lemma 4.

### **(c) Dependence of Statistical Performance on $\eta$ and $\Omega$ for a Change in Mean vector and Covariance Matrix**

Suppose that the process under consideration changes in mean vector from  $\mu$  to  $\mu_{new}$  and covariance matrix from  $\Sigma$  to  $\Sigma_{new}$  after the  $r$ th observation,  $\mathbf{X}_r$ . It is shown below that the joint distribution of the  $T_k's$  (or equivalently  $Z_k's$ ) for each of the control techniques depends on  $\mu$ ,  $\mu_{new}$ ,  $\Sigma$  and  $\Sigma_{new}$  through the 'non-centrality vector'

$$\eta = \Sigma^{-\frac{1}{2}}(\mu_{new} - \mu)$$

and the symmetric 'standardized covariance matrix'

$$\Omega = \Sigma^{-\frac{1}{2}} \Sigma_{new} \Sigma^{-\frac{1}{2}} = \left(\Sigma^{\frac{1}{2}}\right)^{-1} \Sigma_{new} \left(\Sigma^{\frac{1}{2}}\right)^{-1}$$

### **Theorem 4**

Let pdf1 be the probability density of the independent multivariate normal observation vectors  $\mathbf{X}_k's$  (or  $\mathbf{X}_{kj}'s$ ) with parameters  $(\mu_1, \Sigma_1)$  and  $(\mu_{1new}, \Sigma_{1new})$  before and after the change. Let pdf2 be the probability density of the independent multivariate normal observation vectors  $\mathbf{X}_k's$  (or  $\mathbf{X}_{kj}'s$ ) with parameters  $(\mu_2, \Sigma_2)$  and  $(\mu_{2new}, \Sigma_{2new})$  before and after the change. Denote the change point by  $r$ . If

$$\Sigma_1^{-\frac{1}{2}}(\mu_{1new} - \mu_1) = \Sigma_2^{-\frac{1}{2}}(\mu_{2new} - \mu_2) \quad \text{and} \quad \Sigma_1^{-\frac{1}{2}} \Sigma_{1new} \Sigma_1^{-\frac{1}{2}} = \Sigma_2^{-\frac{1}{2}} \Sigma_{2new} \Sigma_2^{-\frac{1}{2}}$$

then  $f_1(T_k's) = f_2(T_k's)$  where  $f_1(T_k's)$  and  $f_2(T_k's)$  denote the joint distribution of  $T_k's$  given pdf1 and pdf2 respectively. This theorem implies equivalent performance of the control techniques under the two alternative probability densities.

### **Proof**

The proof here is similar to that of theorem 2.

## REFERENCES

- ALT, F. B. and BEDEWI, G. E. (1986). 'SPC of Dispersion for Multivariate Data'. *ASQC Quality Congress Transaction - Anaheim*. American Society for Quality Control, pp. 248-254.
- ALT, F. B.; GOODE, J. J. and WADSWORTH, H. M. (1976). 'Small Sample Probability Limits For the Mean of a Multivariate Normal Process'. *ASQC Technical Conference Transactions - Toronto*. American Society for Quality Control, pp. 170-176.
- ALT, F. B. and SMITH, N. D. (1990). 'Multivariate Quality Control' in *Handbook of Statistical Methods for Engineers and Scientists*, Ed. H.M. Wadsworth, McGraw-Hill, New York, NY.
- ANDERSON, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed., John Wiley & Sons, New York, NY.
- CROSIER, R. B. (1988). 'Multivariate Generalization of Cumulative Sum Quality-Control Schemes'. *Technometrics* 30, pp. 291-303.
- DOGANAKSOY, N.; FALTIN, F. W. and TUCKER, W. T. (1991). 'Identification of Out Of Control Quality Characteristics in a Multivariate Manufacturing Environment'. *Communications in Statistics - Theory and Methods* 20, pp. 2775-2790.
- GRAYBILL, F. A. (1983). *Matrices with Applications in Statistics*, 2nd ed., Wadsworth Inc., Belmont, CA.
- HAWKINS, D. M. (1993). 'Regression Adjustment for Variables in Multivariate Quality Control'. *Journal of Quality Technology* 25, pp. 170-182.
- HAYTER, A. J. and TSUI, K. (1994). 'Identification and Quantification in Multivariate Quality Control Problems'. *Journal of Quality Technology* 26, pp. 197-208.
- HOGG, R. V. and CRAIG, A. T. (1971). *Introduction to Mathematical Statistics*, 3rd ed., The Macmillan Co., New York, NY.
- HOTELLING, H. (1947). 'Multivariate Quality Control' in *Techniques of Statistical Analysis*, Eds. Eisenhart, Hastay and Wallis. McGraw-Hill, New York, NY.
- JACKSON, J. E. (1991). 'Multivariate Quality Control : 40 Years Later' in *Statistical Process Control in Manufacturing*, Eds. J.B. Keats and D.C. Montgomery, Marcel Dekker NY and ASQC Quality Press Wisconsin.
- JOHNSON, R. A. and WICHERN, D. W. (1988). *Applied Multivariate Statistical Analysis*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ.
- LINDGREN, B. W. (1976). *Statistical Theory*, 3rd ed., Macmillan Publishing Co., New York, NY.

- LOWRY, C. N. (1989). 'A Multivariate Exponentially Weighted Moving Average Control Chart'. *Ph.d thesis*, University of Southwestern Louisiana, U.S.A.
- LOWRY, C. N.; WOODALL W. H.; CHAMP, C. W. and RIGDON, S. E. (1992). 'A Multivariate Exponentially Weighted Moving Average Control Chart'. *Technometrics* 34, pp. 46-53.
- MONTGOMERY, D. C. and WADSWORTH, H. M. (1972). 'Some Techniques For Multivariate Quality Control Applications'. *Annual Technical Conference Transactions*. American Society for Quality Control, pp. 427-435.
- MORRISON, D. F. (1976). *Multivariate Statistical Methods*, 2nd ed., McGraw-Hill, New York, NY.
- PRESS, S. J. (1982). *Applied Multivariate Analysis : Using Bayesian and Frequentist Methods of Inference*, 2nd ed., Robert E. Krieger Publishing Co., Malabar, FL.
- QUESENBERY, C. P. (1991). 'SPC Q Charts for Start-Up Processes and Short or Long Runs'. *Journal of Quality Technology* 23, pp. 213-224.
- QUESENBERY, C. P. (1993). 'The Effect of Sample Size on Estimated Limits for  $\bar{X}$  and X Control Charts'. *Journal of Quality Technology* 25, pp. 237-247.
- SRIVASTAVA, M. S. and KHATRI, C. G. (1979). *An Introduction to Multivariate Statistics*, Elsevier North Holland, New York, NY.

