International Journal of Reliability, Quality and Safety Engineering
Vol. 9, No. 1 (2002) 41–59


# A NEW VARIABLES CONTROL CHART FOR SIMULTANEOUSLY MONITORING MULTIVARIATE PROCESS MEAN AND VARIABILITY

ARTHUR B. YEH and DENNIS K. J. LIN

*Department of Applied Statistics and Operations Research*
*Bowling Green State University, Bowling Green, OH 43403*

*Department of Management Science and Information Systems*
*Penn State University, University Park, PA 16802*

Received 15 December 2000
Revised 8 May 2001

In this paper, we propose a new variables control chart, called the box-chart, to simultaneously monitor, on a single chart, the process mean and process variability for multivariate processes. The box-chart uses a probability integral transformation to obtain two independently and identically distributed uniform distributions. Therefore, a box-shaped (thus the name), two-dimensional control chart can be constructed. We discuss in detail on how to construct the box-chart. The proposed chart is applied to two real-life examples. The performance of the box-chart is also compared to that of the traditional $T^2$- and $|S|$-charts.

*Keywords*: Probability Integral Transformation; Two-Dimensional Control Charts; Uniform Distribution.

## 1. Introduction

Among the techniques that form the core of statistical process control (SPC), control charts are perhaps the most important and widely used tools. First developed by Shewhart,[1] the use of control charts has become a standard practice in industrial applications. Although over the years, other control charts have been developed, the variables control charts, $\bar{X}$- and $R$- (or $S$-) charts, remain the most popular. Recent developments deviate from early ones most notably on the emphasis they place on target value and on simultaneous monitoring of both the process mean and process variability. Examples include the MSE chart (Spiring and Cheng[2]) which uses the mean squares deviation from the target value to plot two non-crossing measurements on a single chart; the $T$ chart (Cheng and Li[3]) which plots the sum of absolute deviations of the extreme data values from the target value on a single

chart; the "process capability plot" (Repco[4] and Van Nuland[5]) which plots

$$W = \left(\frac{\bar{X} - u}{n/\sqrt{n}}\right)^2 + \left(\frac{S - \sigma}{\sigma/\sqrt{2n}}\right)^2$$

on a two-dimensional circle, where $\bar{X}$, $S$, $\mu$ and $\sigma$ are the sample mean, sample standard deviation, process mean and process standard deviation, respectively; and the "semi-circle" chart (Chao and Cheng[6]) which plots

$$T = (\bar{X} - \mu)^2 + \left(\frac{n-1}{n}\right) S^2$$

on a semi-circle.

The main advantages of simultaneously monitoring univariate process mean and variability on a single chart are:

(i) The number of control charts that need to be maintained can be reduced, and the reduction could be significant particularly when there are multiple processes or multiple stream processes under investigation.
(ii) When the process is out of control, the corresponding changes in process mean or process variability or both can be interpreted relatively easily on a single chart.

Nevertheless, the aforementioned control charts are limited to univariate processes and the extension to multivariate processes may not be possible.

In this paper, our motivation is to develop new variables control charts which maintain the ability to simultaneously monitor, on a single chart, the process mean and process variability for multivariate processes. We shall call the proposed chart the box-chart. We first demonstrate our motivation of the box-chart through a simple example in a univariate process. We then discuss how to construct the box-chart for multivariate processes. Two examples are given for illustration. Comparisons between the box-chart and existing $T^2$- and $|S|$-charts are made through simulation. An alternative box-chart which retains the sampling sequence is also discussed. Finally, we discuss some related issues and possible directions for future research in this area.

## 2. The Box-Chart

Suppose that the quality characteristic of interest of a univariate process follows a normal distribution with mean $\mu_0$ and standard deviation $\sigma_0$. Given a random sample of $n$ observations $X_1, X_2, \ldots, X_n$, define

$$U = P\left(Z \leq \frac{\bar{X} - \mu_0}{\sigma_0/\sqrt{n}}\right) \quad \text{and} \tag{1}$$

$$V = P\left(\chi^2_{n-1} \leq \frac{(n-1)S^2}{\sigma_0^2}\right), \tag{2}$$

where, $\bar{X} = \sum_{i=1}^{n} X_i/n$ and $S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)$ are the sample mean and sample variance, respectively. Here, $Z$ and $\chi^2_{n-1}$ denote the standard normal distribution and the chi-square distribution with $n-1$ degrees of freedom, respectively. When the process is in-control, it is clear that $U$ and $V$ are both distributed as $U(0,1)$, a uniform distribution supported on $(0,1)$. Furthermore, $U$ and $V$ are independent since $\bar{X}$ and $S^2$ are independent. Based on these results, a new variables control chart, the box-chart, can be constructed. As in a $3\sigma$ environment, for example, we can take the control limits: $U_{\text{lower}} = V_{\text{lower}} = 0.00135$ and $U_{\text{upper}} = V_{\text{upper}} = 0.99865$. Furthermore, since both $U$ and $V$ have the same scale (between 0 to 1) and are independent, we can combine these two charts together. Take the data in Van Nuland[5] as an example, Fig. 1 shows the combination as a typical box-chart with different regions marked by different letters to indicate whether the process mean is out-of-control ($M$), the process variability is out-of-control ($V$), or both ($B$). The out-of-control samples are marked by their corresponding sample numbers, and some close to out-of-control samples are marked by an additional asterisk. The areas of $M$, $V$ and $B$ are controlled by the probabilities of type-I error for detecting a shift in the process mean and for detecting a change in the process variability. The point $(U, V)$ is plotted on the box-chart. If $(U, V)$ falls in any of the marked regions, it is an indication that the process is out-of-control. For a more detailed discussion of the univariate box-chart, see Yeh, Lin and Venkataramani.[7]

Essentially, the box-chart is constructed using two independent uniform distributions obtained from probability integral transformation. Note that the probability integral transformation is not limited to univariate random variables. Therefore,

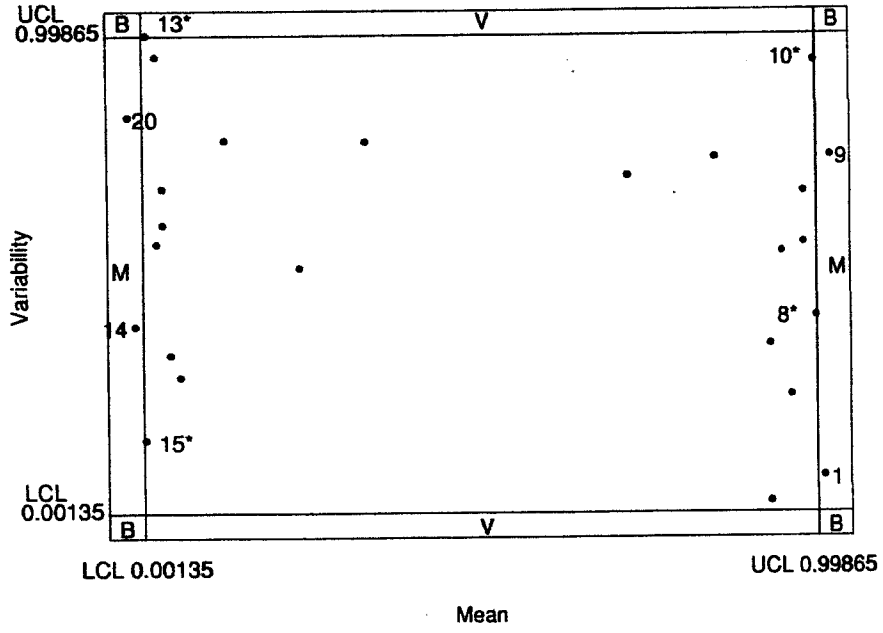


Fig. 1. An example of the box-chart.

one can extend the box-chart to multivariate processes as a way of combining the monitoring of process mean and process variability on a single chart. Specifically, let $\mathbf{X}$ represent the random variable of a multivariate quality characteristic of interest obtained from a process. It is assumed that $\mathbf{X}$ has a $p$-dimensional normal distribution with unknown mean and variance-covariance matrix $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$, respectively. Suppose that in the initial trial samples stage, $m$ samples each of size $n$ are collected. It is assumed that $n > p$, otherwise one may encounter problems when estimating $\boldsymbol{\Sigma}_0$. Let $\mathbf{X}_{ij}$ denote the $i$th observation of the $j$th sample, $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$. Let $\bar{\mathbf{X}}_j = \sum_{i=1}^{n} \mathbf{X}_{ij}/n$ and $\mathbf{S}_j = \sum_{i=1}^{n} (\mathbf{X}_{ij} - \bar{\mathbf{X}}_j)(\mathbf{X}_{ij} - \bar{\mathbf{X}}_j)'/n$ be the sample mean and sample variance-covariance matrix of the $j$th sample, $j = 1, 2, \ldots, m$. Furthermore, define $N = m \times n$, $\bar{\bar{\mathbf{X}}} = \sum_{j=1}^{m} \mathbf{X}_j/m$, $\bar{\mathbf{S}} = \sum_{j=1}^{m} \mathbf{S}_j/m$ and $\bar{\mathbf{S}}(j) = \sum_{i=1, i\neq j}^{m} \mathbf{S}_i/(m-1)$, $j = 1, 2, \ldots, m$.

For $j = 1, 2, \ldots, m$, one computes the following statistics

$$U_j = P\left(F_{p,N-m-p+1} \leq \frac{(N-m-p+1)}{p(m-1)} (\bar{\mathbf{X}}_j - \bar{\bar{\mathbf{X}}})' \bar{\mathbf{S}}^{-1} (\bar{\mathbf{X}}_j - \bar{\bar{\mathbf{X}}})\right), \tag{3}$$

$$V_j = P\left(\prod_{i=1}^{p} F_{n-i,N-m-n+2-i} \leq \left(\prod_{i=1}^{p} \frac{N-m-n+2-i}{n-i}\right) \frac{|n\mathbf{S}_j|}{|(N-n)\mathbf{S}_{(j)}|}\right), \tag{4}$$

where $|A|$ denotes the determinant of the matrix $A$. Here $\prod_{i=1}^{p} F_{n-i,N-m-n+2-i}$ denotes the product of $p$ independent $F$-distributions each with degrees of freedom $n-i$ and $N-m-n+2-i, i = 1, 2, \ldots, p$. Note that here $n$, instead of $n-1$, is used as the divisor in defining the sample variance-covariance matrix. Either choice, the results [Eqs. (3) and (4)] can be derived accordingly (see Anderson[8]). It is shown in Appendix A that $U_j$'s and $V_j$'s are independently and identically distributed (i.i.d.) as $U(0,1)$. Therefore, one can plot $(U_1, V_1), (U_2, V_2), \ldots, (U_m, V_m)$ on the box-chart. Furthermore, we have derived the exact and approximate distributions for computing $V_j$ in Appendix B.

Note that at the trial stage, the objective is to make sure that the process is in-control before the control chart can be used to monitor the future process. If there is any out-of-control point, (i) identify the sample and the source of the problem by looking up the region in which the point falls; (ii) find the assignable causes; and (iii) correct the process if necessary. Note that the box-chart remains unchanged. The values of $\bar{\bar{\mathbf{X}}}, N, \bar{\mathbf{S}}, \bar{\mathbf{S}}_{(j)}$ and $m$ are then recalculated based on the remaining samples. Consequently, the $U_j$'s and $V_j$'s are recomputed for the remaining samples and re-plotted on the box-chart. This is repeated until all the points are in-control. At this point, the estimates obtained from the remaining samples can be used to calculate the $U$'s and $V$'s for future samples. More specifically, denote the updated information by $m_I$ ($m_I \leq m$), $N_I = m_I \times n$, $\bar{\bar{\mathbf{X}}}_I$ and $\bar{\mathbf{S}}_I$. When the monitoring of the future process begins, for any sample $\mathbf{X}_{k1}, \mathbf{X}_{k2}, \ldots, \mathbf{X}_{kn}, k \geq 1$, one plots

$$U_k = P\left(F_{p,N_I-m_I-p+1} \leq \frac{(N_I - m_I - p + 1)}{p(m_I - 1)} (\bar{\mathbf{X}}_k - \bar{\bar{\mathbf{X}}}_I)' \bar{\mathbf{S}}_I^{-1} (\bar{\mathbf{X}}_k - \bar{\bar{\mathbf{X}}}_I)\right), \quad (5)$$

$$V_k = P\left(\prod_{i=1}^{p} F_{n-i,N_I-m_I+1-i} \leq \left(\prod_{i=1}^{p} \frac{N_I - m_I + 1 - i}{n - i}\right) \times \frac{|n\mathbf{S}_k|}{|N_I \bar{\mathbf{S}}_I|}\right) \quad (6)$$

on the box-chart. The process is considered to be out-of-control if $(U_k, V_k)$ falls in any of the marked regions.

## 3. Examples

The first example, taken from Mitra,[9] is related to a component used in the assembly of a transmission mechanism. Two quality characteristics, tensile strength and diameter, are of importance. Twenty samples each of size 4 were obtained from the process. The data set is listed in Table 1. The box-chart, and the corresponding $T^2$- and $|S|$-charts are given in Figs. 2 and 3, respectively. A brief explanation on how the $T^2$- and $|S|$-charts are constructed can be found in Appendix C.

The box-chart indicates that none of the 20 samples is out-of-control, which is consistent with the $T^2$- and $|S|$-charts. Samples 7 and 11 (marked by asterisks) are in-control, although sample 7 is very close to the boundary of region $M$ and sample 11 is close to the upper region $V$. These two samples correspond to the two peaks seen on the $T^2$- and $|S|$-charts, with the peak of sample 7 appearing on $T^2$-chart and the peak of sample 11 appearing on $|S|$-chart, respectively.

Table 1. Data set of the transmission component example.

| Sample number | Tensile strength | | | | Diameter | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | 66 | 70 | 68 | 72 | 16 | 18 | 15 | 20 |
| 2 | 75 | 60 | 70 | 75 | 17 | 22 | 18 | 19 |
| 3 | 65 | 70 | 70 | 65 | 20 | 18 | 15 | 18 |
| 4 | 72 | 70 | 75 | 65 | 19 | 20 | 15 | 17 |
| 5 | 73 | 74 | 72 | 70 | 21 | 21 | 23 | 19 |
| 6 | 72 | 74 | 73 | 74 | 21 | 19 | 20 | 18 |
| 7 | 63 | 62 | 65 | 66 | 22 | 20 | 24 | 22 |
| 8 | 75 | 84 | 75 | 66 | 22 | 20 | 20 | 22 |
| 9 | 65 | 69 | 77 | 71 | 18 | 16 | 18 | 18 |
| 10 | 70 | 68 | 67 | 67 | 18 | 17 | 19 | 18 |
| 11 | 80 | 75 | 70 | 69 | 24 | 18 | 20 | 22 |
| 12 | 68 | 65 | 80 | 50 | 20 | 21 | 20 | 22 |
| 13 | 74 | 80 | 76 | 74 | 19 | 17 | 20 | 21 |
| 14 | 76 | 74 | 75 | 73 | 20 | 17 | 18 | 18 |
| 15 | 71 | 70 | 74 | 73 | 18 | 16 | 17 | 18 |
| 16 | 68 | 67 | 70 | 69 | 18 | 16 | 19 | 20 |
| 17 | 72 | 76 | 75 | 77 | 22 | 19 | 23 | 20 |
| 18 | 76 | 74 | 75 | 77 | 19 | 23 | 20 | 21 |
| 19 | 72 | 74 | 73 | 75 | 20 | 18 | 20 | 19 |
| 20 | 72 | 68 | 74 | 70 | 21 | 19 | 18 | 20 |

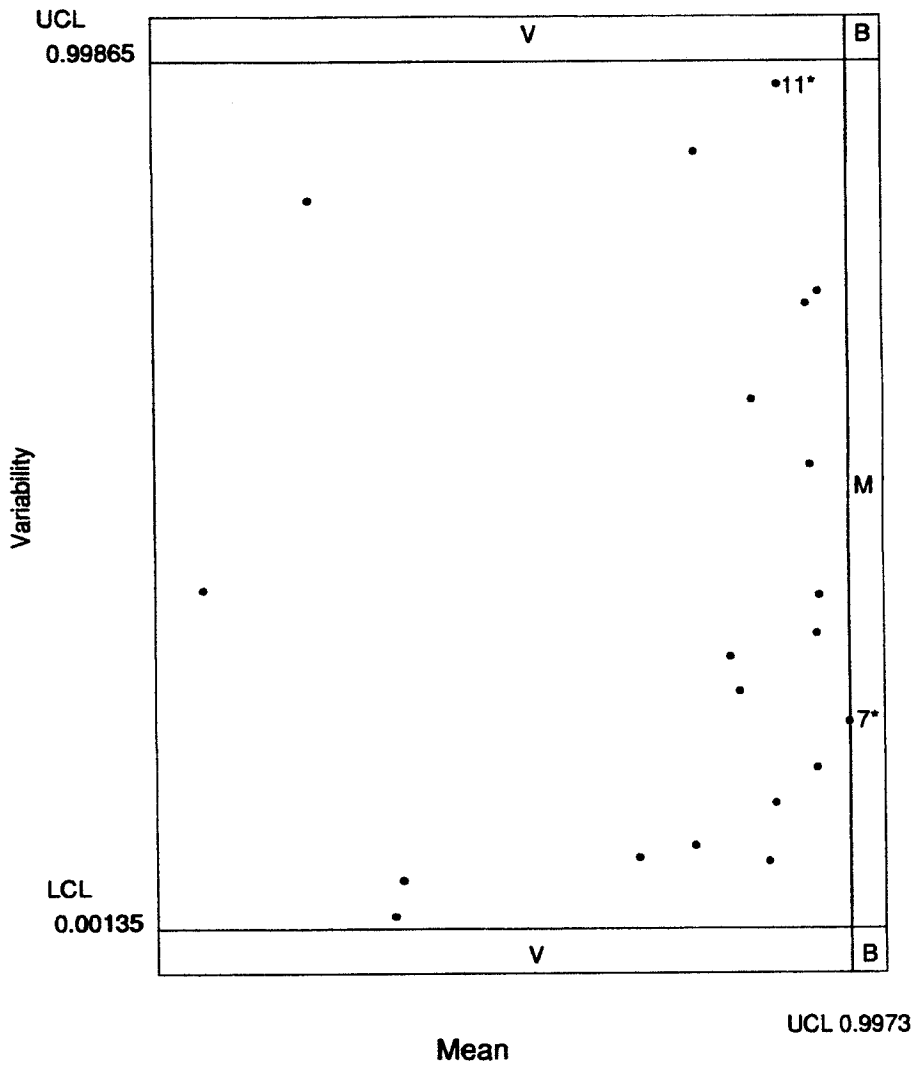


Fig. 2. The box-chart of the transmission component example.

Note that for the process mean, the out-of-control region on the box-chart is one-sided, since the statistic used is based on the Hotelling-$T^2$. As for the process variability, the out-of-control region is two-sided since it is based on the sample generalized variance. Furthermore, caution should be taken when interpreting the $|S|$-chart as well as the region $V$ on the box-chart. If a sample point is plotted out-of-control on the $|S|$-chart or on the box-chart, it is primarily due to the change in the determinant of the sample variance-covariance matrix. However, this does not necessarily imply that there is an increase or decrease in the process variability. It is worth mentioning that Johnson and Wichern[10] gave 3 sample variance-covariance matrices for bivariate data that all have the same determinant and yet have very different correlations.

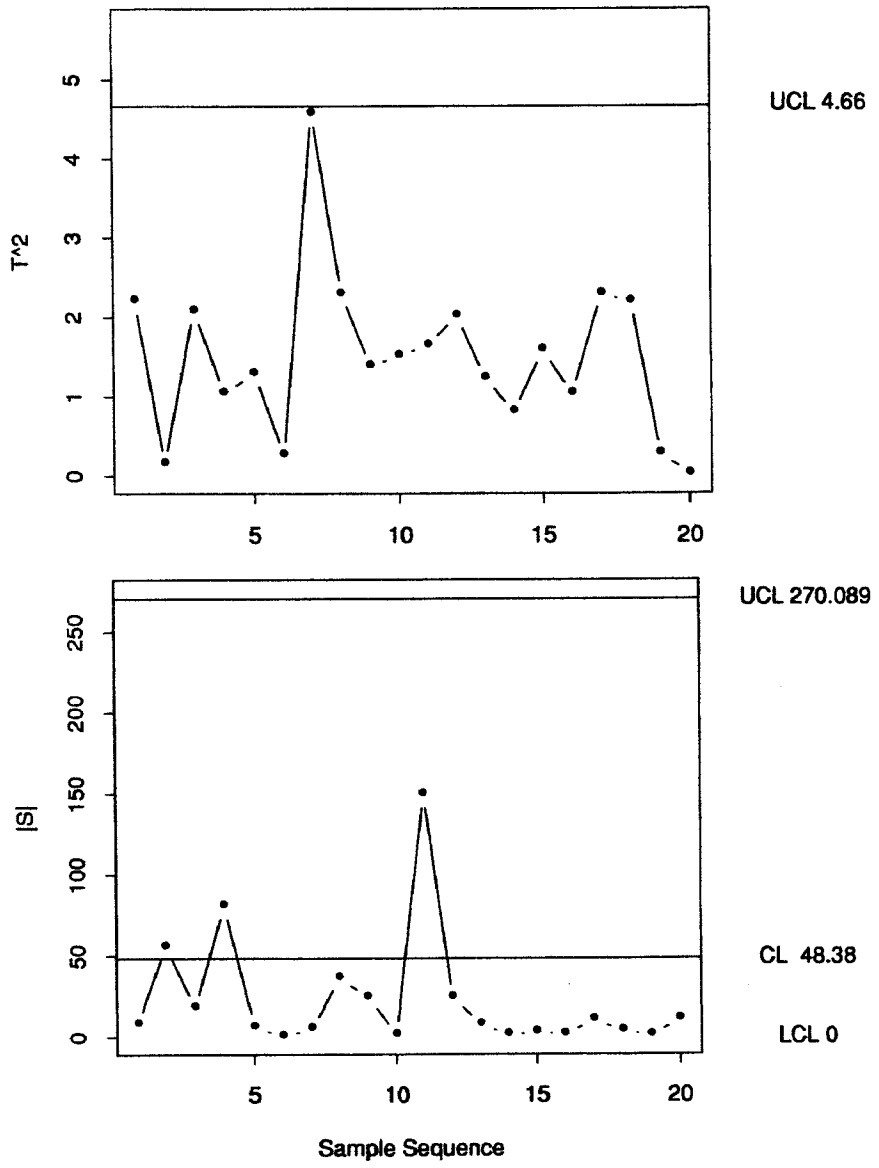


Fig. 3. The $T^2$- and $|S|$-charts of the transmission component example.

The second example, also taken from Mitra,[9] is often encountered in fabric-production process. The variables of interest are the single-strand break factor (a measure of the breaking strength) and the weight of textile fibers. Originally, 20 samples each of size 4 were obtained from the process. The samples are listed in Table 2. The original 20 samples do not indicate any out-of-control signal. Note that in this example

$$\bar{\bar{\mathbf{X}}} = \begin{pmatrix} 82.4625 \\ 20.1750 \end{pmatrix}, \quad \bar{\mathbf{S}} = \begin{pmatrix} 7.5215 & -0.3542 \\ -0.3542 & 3.2917 \end{pmatrix}.$$

Table 2. Data set of the fabric-production example.

| Sample number | Break factor | | | | Weight | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | 80 | 82 | 78 | 85 | 19 | 22 | 20 | 20 |
| 2 | 75 | 78 | 84 | 81 | 24 | 21 | 18 | 21 |
| 3 | 83 | 86 | 84 | 87 | 19 | 24 | 21 | 22 |
| 4 | 79 | 84 | 80 | 83 | 18 | 20 | 17 | 16 |
| 5 | 82 | 81 | 78 | 86 | 23 | 21 | 18 | 22 |
| 6 | 86 | 84 | 85 | 87 | 21 | 20 | 23 | 21 |
| 7 | 84 | 88 | 82 | 85 | 19 | 23 | 19 | 22 |
| 8 | 76 | 84 | 78 | 82 | 22 | 17 | 19 | 18 |
| 9 | 85 | 88 | 85 | 87 | 18 | 16 | 20 | 16 |
| 10 | 80 | 78 | 81 | 83 | 18 | 19 | 20 | 18 |
| 11 | 86 | 84 | 85 | 86 | 23 | 20 | 24 | 22 |
| 12 | 81 | 81 | 83 | 82 | 22 | 21 | 23 | 21 |
| 13 | 81 | 86 | 82 | 79 | 16 | 18 | 20 | 19 |
| 14 | 75 | 78 | 82 | 80 | 22 | 21 | 23 | 22 |
| 15 | 77 | 84 | 78 | 85 | 22 | 19 | 21 | 18 |
| 16 | 86 | 81 | 84 | 84 | 19 | 23 | 18 | 22 |
| 17 | 84 | 85 | 78 | 79 | 17 | 22 | 18 | 19 |
| 18 | 82 | 86 | 79 | 83 | 20 | 19 | 23 | 21 |
| 19 | 79 | 88 | 85 | 83 | 21 | 23 | 20 | 18 |
| 20 | 80 | 84 | 82 | 85 | 18 | 22 | 19 | 20 |
| 21* | 83 | 93 | 83 | 87 | 24 | 21 | 26 | 23 |
| 22* | 79 | 82 | 92 | 84 | 13 | 26 | 21 | 16 |
| 23* | 91 | 76 | 90 | 95 | 23 | 16 | 27 | 20 |

*Samples 21, 22 and 23 are generated samples.

Assuming that the in-control process has a bivariate normal distribution with $\boldsymbol{\mu}_0 = \bar{\bar{\mathbf{X}}}$ and $\boldsymbol{\Sigma}_0 = \bar{\mathbf{S}}$, three samples are generated, where sample 21 (marked by a $\square$) has a shift in $\boldsymbol{\mu}$, sample 22 (marked by a $\bigcirc$) has a change in $\boldsymbol{\Sigma}$, and sample 23 (marked by a $\triangle$) has both. The generated samples, with decimal points rounded off to integers, are listed at the bottom of Table 2. The box-chart, which contains all 23 samples, is shown in Fig. 4 and the corresponding $T^2$- and $|S|$-charts are shown in Fig. 5. All three out-of-control samples are picked up correctly by the box-chart. These results are consistent with those shown on the $T^2$- and $|S|$-charts.

## 4. Simulation Studies

In this section, we compare the performance of the proposed box-chart with the existing $T^2$- and $|S|$-charts for multivariate processes. The comparison is carried out using Monte Carlo simulation. The performance is defined in terms of the average run length (ARL). The simulation run is 50,000, and the subgroup sizes used are $n = 4, 6, 8, 10$. We set the areas of the out-of-control regions of an in-control process to be 0.0027 and 0.00135 for both $M$ and $V$. This leads to in-control ARL's of approximately 185 and 370 respectively for the box-chart. The probabilities of type-I error of both $T^2$-chart and $|S|$-chart are also chosen to be 0.0027 and 0.00135 so that the combined $T^2$- and $|S|$-charts have in-control ARL's approximately the

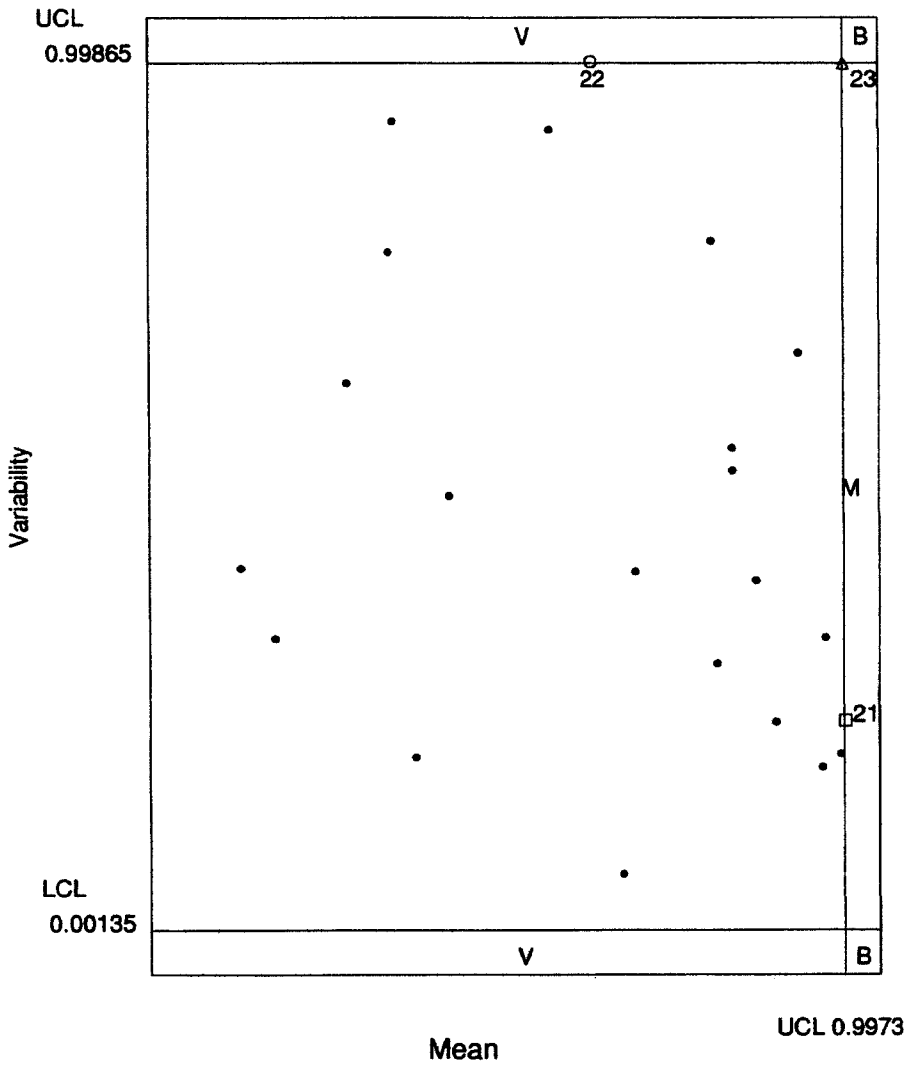


Fig. 4. The box-chart of the fabric-production example.

same as the box-chart. In simulating the out-of-control ARL's, once a sample is generated it is used to evaluate both the box-chart and the $T^2$- and $|S|$-charts.

The in-control process is assumed to have a bivariate normal distribution with $\mu_0 = (0,0)'$ and $\Sigma_0 = I$. We consider three possible changes in the process: a shift in process mean; a change in process variability; and changes both in process mean and process variability. The ARL of the box-chart is compared to that of the $T^2$- and $|S|$-charts.

We first assume that the process variability remains unchanged, i.e., the variance-covariance $\Sigma_0 = I$, but the process mean has been shifted to $\mu = (\mu_1, \mu_2)'$. This is equivalent to shifting the mean of the first variable by $\mu_1$ standard deviations, and shifting the mean of the second variable by $\mu_2$ standard deviations. Summarized in Table 3 are the ARL's of the box-chart and the $T^2$- and $|S|$-charts

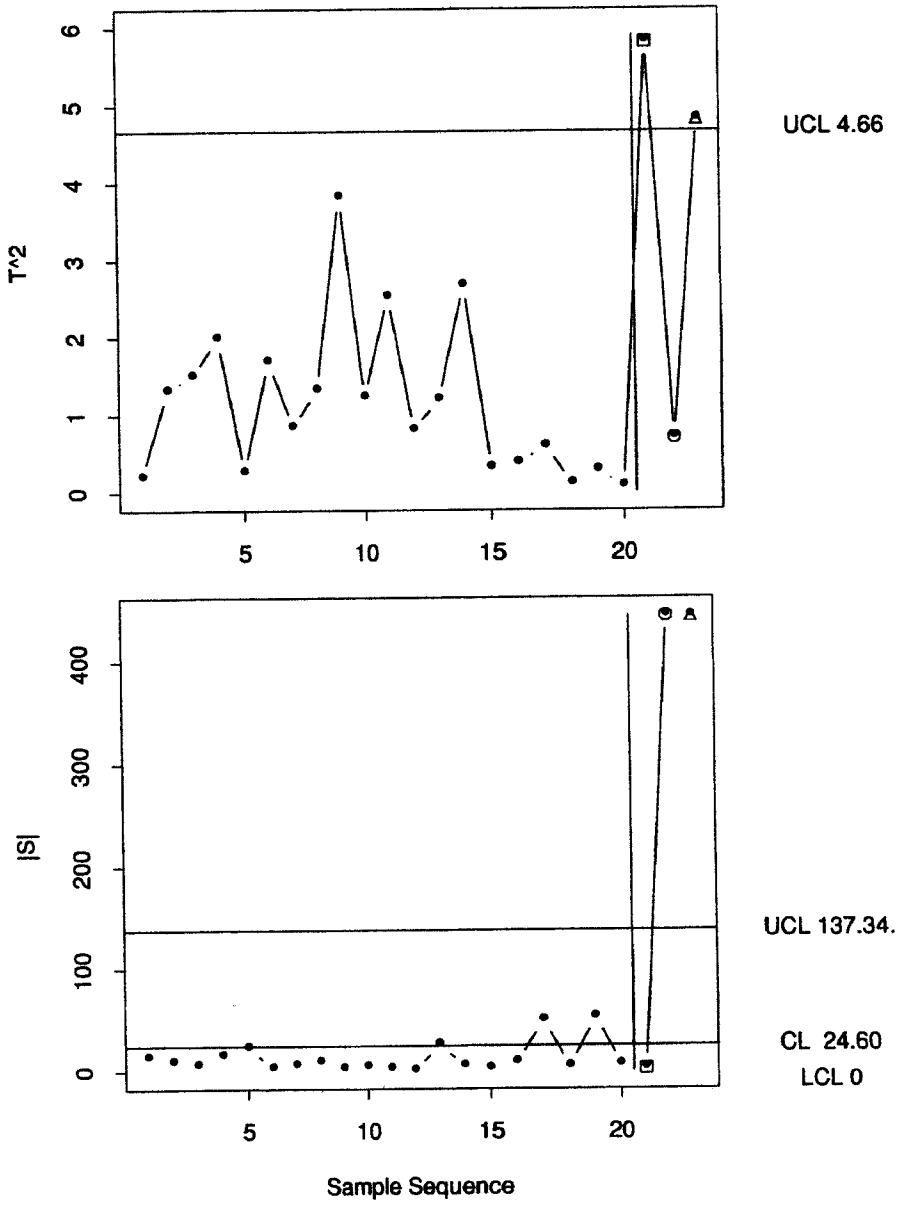


Fig. 5. The $T^2$- and $|S|$-charts of the fabric-production example.

for different choices of $(\mu_1, \mu_2)$. The box-chart is very comparable to $T^2$- and $|S|$-charts except in cases when there is a small shift in first variable and sample size $n$ is small (e.g., $n = 4$, $\mu_1 = 0.5$, 0.75 and $\mu_2 = 0$).

Next, we assume that the process mean has remained at $\boldsymbol{\mu} = (0,0)'$, while the variance-covariance matrix $\boldsymbol{\Sigma}$ is equal to $= \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, where $\sigma_1$, $\sigma_2 > 1$, $-1 \leq \rho \leq 1$. We then simulate the ARL values for a variety of choices of $(\sigma_1, \sigma_2, \rho)$. The simulation results are shown in Table 4. For all cases considered here, the box-chart performs comparably with the $T^2$- and $|S|$-charts. Shown in Table 5 are the

Table 3. Comparisons of ARL for mean shifts.

| | $ARL_0 = 185$ | | $ARL_0 = 370$ | |
|---|---|---|---|---|
| $(\mu_1, \mu_2)$ | Box-chart | $T^2$ and $\|S\|$ | Box-chart | $T^2$ and $\|S\|$ |
| $(n = 4)$ | | | | |
| (0.50, 0.50) | 39.28 | 25.88 | 63.45 | 41.36 |
| (0.75, 0.75) | 10.52 | 7.61 | 14.84 | 10.62 |
| (1.0, 1.0) | 3.89 | 3.06 | 4.90 | 3.85 |
| (0.50, 0) | 79.87 | 54.47 | 145.35 | 100.20 |
| (0.75, 0) | 32.66 | 22.05 | 53.71 | 35.29 |
| (1.0, 0) | 13.10 | 9.28 | 18.78 | 9.28 |
| $(n = 6)$ | | | | |
| (0.50, 0.50) | 21.21 | 14.56 | 32.74 | 21.43 |
| (0.75, 0.75) | 5.06 | 3.90 | 6.84 | 5.16 |
| (1.0, 1.0) | 2.05 | 1.76 | 2.39 | 2.03 |
| (0.50, 0) | 54.70 | 37.20 | 95.79 | 65.88 |
| (0.75, 0) | 17.55 | 11.97 | 26.70 | 18.09 |
| (1.0, 0) | 6.31 | 4.71 | 8.80 | 6.46 |
| $(n = 8)$ | | | | |
| (0.50, 0.50) | 12.56 | 9.05 | 19.44 | 13.48 |
| (0.75, 0.75) | 3.17 | 2.58 | 3.94 | 3.18 |
| (1.0, 1.0) | 1.44 | 1.32 | 1.60 | 1.43 |
| (0.50, 0) | 36.68 | 24.69 | 62.74 | 41.77 |
| (0.75, 0) | 10.63 | 7.68 | 15.03 | 10.67 |
| (1.0, 0) | 3.80 | 3.02 | 5.02 | 3.89 |
| $(n = 10)$ | | | | |
| (0.50, 0.50) | 8.90 | 6.46 | 12.39 | 8.77 |
| (0.75, 0.75) | 2.24 | 1.90 | 2.67 | 2.23 |
| (1.0, 1.0) | 1.20 | 1.14 | 1.28 | 1.20 |
| (0.50, 0) | 28.72 | 19.43 | 46.59 | 30.83 |
| (0.75, 0) | 6.97 | 5.22 | 9.92 | 7.21 |
| (1.0, 0) | 2.64 | 2.21 | 3.29 | 2.69 |

simulated ARL's when shifts in process mean and changes in process variability both occur. Again, the box-chart and the $T^2$- and $|S|$-charts have very similar ARL's.

## 5. An Alternative Box-Chart that Retains Sampling Sequence

Apparently, the box-chart has one drawback in that it loses track of the sampling sequence. To understand the time sequence pattern, however, Balkin and Lin[11] argues that it is better to employ well developed techniques in time series area, such as sacf (sample autocorrelation function) and spacf (sample partial autocorrelation function) plots. The box-chart can easily accommodate the sampling sequence as well. Using the original 20 samples of the fabric-production example, Fig. 6 shows an alternative box-chart which incorporates the sampling sequence. There are two statistics being plotted in Fig. 6, the "u" computed based on (3) is represented by a solid line, while the "v" computed based on (4) is represented by a dashed line.

Table 4. Comparisons of ARL for changes in variability.

| | $ARL_0 = 185$ | | $ARL_0 = 370$ | |
|---|---|---|---|---|
| $(\sigma_1, \sigma_2, \rho)$ | Box-chart | $T^2$ and $\|S\|$ | Box-chart | $T^2$ and $\|S\|$ |
| $(n = 4)$ | | | | |
| $(1.25, 1.25, 0)$ | 25.93 | 22.01 | 39.18 | 32.59 |
| $(1.50, 1.50, 0)$ | 6.92 | 6.20 | 9.19 | 8.24 |
| $(1.25, 1.25, 0.5)$ | 28.39 | 22.81 | 43.98 | 34.85 |
| $(1.50, 1.50, -0.2)$ | 7.09 | 6.34 | 9.51 | 8.36 |
| $(n = 6)$ | | | | |
| $(1.25, 1.25, 0)$ | 18.38 | 16.40 | 27.96 | 25.06 |
| $(1.50, 1.50, 0)$ | 4.34 | 4.06 | 5.48 | 5.14 |
| $(1.25, 1.25, 0.5)$ | 25.51 | 21.51 | 35.79 | 30.32 |
| $(1.50, 1.50, -0.2)$ | 4.56 | 4.29 | 5.75 | 5.37 |
| $(n = 8)$ | | | | |
| $(1.25, 1.25, 0)$ | 13.76 | 12.52 | 20.52 | 18.72 |
| $(1.50, 1.50, 0)$ | 3.09 | 2.98 | 3.75 | 3.61 |
| $(1.25, 1.25, 0.5)$ | 22.20 | 18.70 | 33.42 | 28.23 |
| $(1.50, 1.50, -0.2)$ | 3.25 | 3.11 | 3.98 | 3.81 |
| $(n = 10)$ | | | | |
| $(1.25, 1.25, 0)$ | 10.73 | 9.99 | 15.58 | 14.59 |
| $(1.50, 1.50, 0)$ | 2.42 | 2.36 | 3.84 | 3.76 |
| $(1.25, 1.25, 0.5)$ | 19.67 | 17.09 | 28.69 | 24.59 |
| $(1.50, 1.50, -0.2)$ | 2.54 | 2.47 | 3.02 | 2.93 |

Table 5. Comparisons of ARL for changes both in mean and variability.

| | $ARL_0 = 185$ | | $ARL_0 = 370$ | |
|---|---|---|---|---|
| $(\mu_1, \mu_2)(\sigma_1, \sigma_2, \rho)$ | Box-chart | $T^2$ and $\|S\|$ | Box-chart | $T^2$ and $\|S\|$ |
| $(n = 4)$ | | | | |
| $(1.0, 1.0)(1.50, 1.50, 0)$ | 2.28 | 2.04 | 2.61 | 2.33 |
| $(0.75, 0.75)(1.50, 1.50, 0)$ | 3.26 | 3.88 | 3.75 | 3.40 |
| $(0.50, 0.50)(1.75, 1.75, 0.5)$ | 3.37 | 3.10 | 3.91 | 3.44 |
| $(0.25, 0.25)(1.50, 1.00, -0.2)$ | 19.77 | 15.93 | 28.72 | 23.08 |
| $(n = 6)$ | | | | |
| $(1.0, 1.0)(1.50, 1.50, 0)$ | 1.59 | 1.48 | 1.76 | 1.63 |
| $(0.75, 0.75)(1.50, 1.50, 0)$ | 2.19 | 2.01 | 2.50 | 2.34 |
| $(0.50, 0.50)(1.75, 1.75, 0.5)$ | 3.39 | 2.26 | 2.58 | 2.38 |
| $(0.25, 0.25)(1.50, 1.00, -0.2)$ | 15.41 | 12.76 | 21.98 | 18.15 |
| $(n = 8)$ | | | | |
| $(1.0, 1.0)(1.50, 1.50, 0)$ | 1.30 | 1.25 | 1.41 | 1.33 |
| $(0.75, 0.75)(1.50, 1.50, 0)$ | 1.69 | 1.59 | 1.80 | 1.77 |
| $(0.50, 0.50)(1.75, 1.75, 0.5)$ | 1.87 | 1.80 | 1.92 | 1.84 |
| $(0.25, 0.25)(1.50, 1.00, -0.2)$ | 12.35 | 10.31 | 17.93 | 14.97 |
| $(n = 10)$ | | | | |
| $(1.0, 1.0)(1.50, 1.50, 0)$ | 1.16 | 1.13 | 1.22 | 1.18 |
| $(0.75, 0.75)(1.50, 1.50, 0)$ | 1.42 | 1.36 | 1.57 | 1.49 |
| $(0.50, 0.50)(1.75, 1.75, 0.5)$ | 1.59 | 1.54 | 1.62 | 1.59 |
| $(0.25, 0.25)(1.50, 1.00, -0.2)$ | 10.07 | 8.64 | 14.48 | 12.15 |

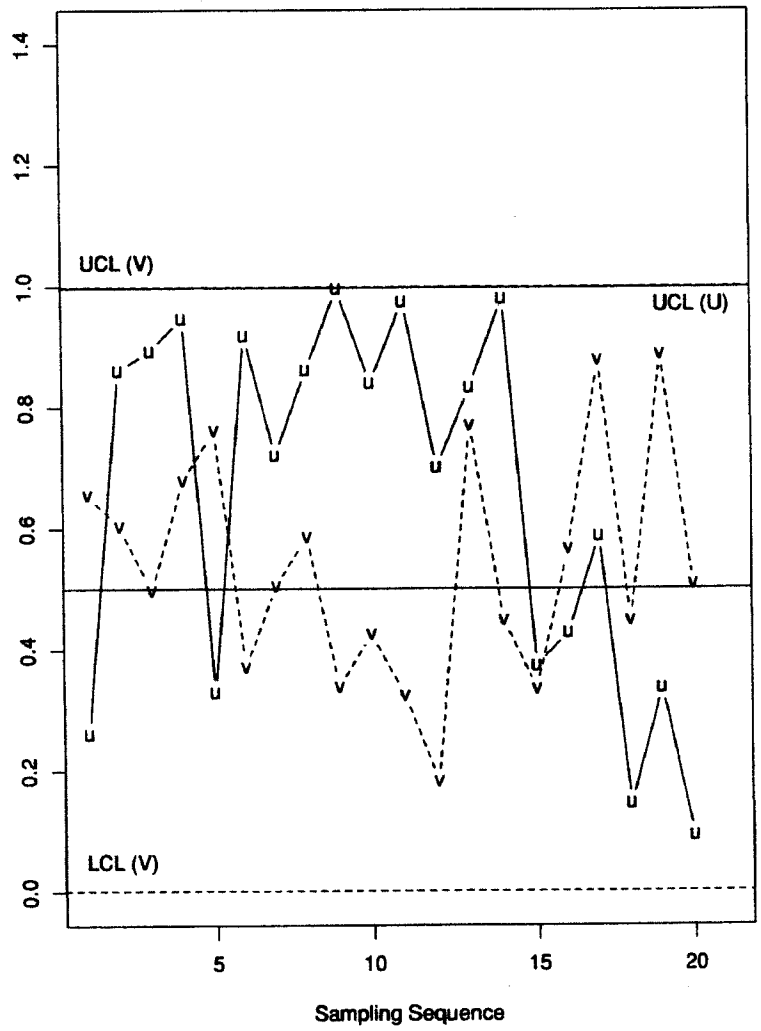


Fig. 6. An alternative box-chart (with time sequence) of the fabric-production example.

Also note that for "u" only upper control limit (denoted by UCL (U) as represented by a solid horizontal line) is needed, while for "v" both control limits (denoted by UCL (V) and LCL (V) as represented by two dashed horizontal lines) are present.

In fact, Fig. 6 brings up a wider point here. After proper probability integral transformation, all performance measures (such as $T^2$ and $|S|$) lie in the range $(0, 1)$. Thus, the (alternative) box-chart can plot more than two performance measures preferably using different colors, if so desire.

## 6. Concluding Remarks

We have proposed and studied the box-chart as a way to combine the $T^2$- and $|S|$-charts into a single chart. The box-chart, based essentially on a probability integral transformation, is developed specifically for multivariate normal processes.

However, the proposed box-chart is not limited to normal processes provided that the distributions of the test statistics used are known. Future research along this line would be worthwhile.

The box-chart can clearly indicate, on a single chart, whether the process mean or the process variability is out-of-control. Therefore, the use of box-chart in practice can reduce the number of control charts that need to be maintained, and the reduction could potentially be significant if there are multiple processes or multiple stream processes to be monitored. Furthermore, since both performance measures $T^2$ and $|S|$ have been transformed to lie in the range $(0, 1)$, additional performance measures, after proper probability integral transformation, could potentially be plotted on the same chart (Fig. 6), thus making it possible for the (alternative) box-chart to incorporate multiple performance measures.

The ARL performance of the box-chart, on the other hand, does not compare favorably with that of the $T^2$- and $|S|$-charts, particularly in cases when the mean shifts are small and the sample size is small. However, in such cases, CUSUM and EWMA control charts may be more appropriate. The other drawback of the box-chart is that it loses track of the sampling sequence. Although the alternative box-chart which retains the sampling sequence, as described in Sec. 5, can be used instead. It should be noted that the out-of-control signals can be interpreted by applying common procedures such as the $T^2$ decomposition techniques (see Mason, Tracy and Young[12] and Mason, Champ, Tracy, Wierda and Young[13]).

The simulation results seem to indicate that the box-chart, like the $T^2$- and $|S|$-charts, is less sensitive in detecting smaller mean shifts or smaller changes in variability or both. One possible way to make the box-chart more sensitive to smaller changes is to develop some type of CUSUM box-chart. Note that the box-chart is constructed essentially by combining two independent $U(0, 1)$ random variables, one derived from the statistic used to monitor the process mean, and the other derived from the statistic used to monitor the process variability. Therefore, a CUSUM-based box-chart can be developed by calculating independently the sums of two sequences of i.i.d. $U(0, 1)$'s, one for detecting smaller mean shifts and the other for detecting smaller changes in process variability. The ideas seem appealing, and we plan to fully explore these ideas in a follow-up study.

## Acknowledgments

We thank two referees for numerous suggestions which improve the presentation of the paper. Dennis Lin is partially supported by the National Science Foundation via Grant DMS-9704711 and the National Science Council of ROC via Contract NSC 87-2119-M-001-007.

## Appendix A

Here we show tht the test statistics described in (3) and (4) are i.i.d. $U(0, 1)$. Write

$$T_1 = (\bar{\mathbf{X}}_j - \bar{\bar{\mathbf{X}}})' \bar{S}^{-1} (\bar{\mathbf{X}}_j - \bar{\bar{\mathbf{X}}}) \quad \text{and} \quad T^2 = \frac{|\mathbf{S}_j|}{|\bar{\mathbf{S}}_{(j)}|}.$$

$T_1$ gives the so-called Hotelling-$T^2$ statistic which can easily be shown to be proportional to an $F$-distribution (Anderson[8]), and therefore $U_j$ is $U(0,1)$. As for $V_j$, note that both $|\mathbf{S}_j|$ and $|\bar{\mathbf{S}}_{(j)}|$ are $|\boldsymbol{\Sigma}|$ times a product of $p$ independent $\chi^2$ with different degrees of freedom. Furthermore, $|\mathbf{S}_j|$ and $|\bar{\mathbf{S}}_{(j)}|$ are independent.

To show that $T_1$ and $T_2$ are independent, it suffices to show that $\bar{\mathbf{S}}$ and $|\mathbf{S}_j|/|\bar{\mathbf{S}}_{(j)}|$ are independent. We next write $N\bar{\mathbf{S}} = n\mathbf{S}_j + n(m-1)\bar{\mathbf{S}}_{(j)}$. By normality,

$$n\mathbf{S}_j \sim W_p(\boldsymbol{\Sigma}, n-1) \quad \text{and} \quad n(m-1)\bar{\mathbf{S}}_{(j)} \sim W_p(\boldsymbol{\Sigma}, (m-1)(n-1))\,,$$

where $W_p(\boldsymbol{\Sigma}, r)$ denotes the Wishart distribution with parameter $\boldsymbol{\Sigma}$, degrees of freedom $r$ and rank $p$. Since $\mathbf{S}_j$ and $\bar{\mathbf{S}}_{(j)}$ are independent, $N\bar{\mathbf{S}} \sim W_p(\boldsymbol{\Sigma}, m(n-1))$. To this end, it suffices to show that, apart from some constants, if $A_1 \sim W_p(\boldsymbol{\Sigma}, m_1)$ and $A_2 \sim W_p(\boldsymbol{\Sigma}, m_2)$ are independent, then their sum $R_1 = A_1 + A_2$ is independent of $R_2 = |A_1|/|A_2|$.

Without loss of generality, we can assume $\boldsymbol{\Sigma} = I$. We shall proceed the proof by induction. When $p = 1$, the result is trivial, since for any two independent $\chi^2$, their sum is independent of their ratio. When $p = 2$, write $A_1 = \sum_1^{m_1} \mathbf{Z}_j\mathbf{Z}_j'$ and $A_2 = \sum_1^{m_1} \mathbf{Y}_j\mathbf{Y}_j'$, where $\mathbf{Z}_j$'s and $\mathbf{Y}_j$'s are independent $N_2(0, I)$ (a bivariate standard normal distribution), and furthermore $\mathbf{Z}_j$'s are independent of $\mathbf{Y}_j$'s. Partition $\mathbf{Z}_j$'s and $\mathbf{Y}_j$'s into two components (both are scalars),

$$\begin{pmatrix} Z_j^{(1)} \\ Z_j^{(2)} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} Y_j^{(1)} \\ Y_j^{(2)} \end{pmatrix}.$$

Note that all these components are i.i.d. $N(0,1)$. Consider the joint distribution of $R_1$ and $R_2$ conditioning on

$$\mathbf{Z}^{(2)} = (Z_1^{(2)}, Z_2^{(2)}, \ldots, Z_{m_1}^{(2)})' \quad \text{and} \quad \mathbf{Y}^{(2)} = (Y_1^{(2)}, Y_2^{(2)}, \ldots, Y_{m_2}^{(2)})'\,.$$

Given $\mathbf{Z}^{(2)}$ and $\mathbf{Y}^{(2)}$, $R_1$ and $R_2$ are independent since $R_1$ is equivalent to the sum of two independent $\chi^2$ and $R_2$ is proportional to their ratio. Furthermore, since neither of the distributions of $R_1$ and $R_2$ depends on $\mathbf{Z}^{(2)}$ and $\mathbf{Y}^{(2)}$, $R_1$ and $R_2$ are independent.

Suppose now the independence result holds for dimension $p-1$, we need to show that it also holds for dimension $p$. Partition $\mathbf{Z}_j$'s and $\mathbf{Y}_j$'s into two components where the first component is a vector of $p-1$ and the second component is a scalar,

$$\begin{pmatrix} \mathbf{Z}_j^{(p-1)} \\ Z_j^{(2)} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \mathbf{Y}_j^{(p-1)} \\ Y_j^{(2)} \end{pmatrix}.$$

Note that the first components of $\mathbf{Z}_j$'s and $\mathbf{Y}_j$'s are i.i.d. $N_{p-1}(0, I)$ and the second components are i.i.d. $N(0,1)$. Given $\mathbf{Z}^{(2)}$ and $\mathbf{Y}^{(2)}$, $R_1$ is equivalent to $W_{p-1}(I, m_1) + W_{p-1}(I, m_2)$ (two independent distributions) and $R_2$ is proportional to $|W_{p-1}(I, m_1)|/|W_{p-1}(I, m_2)|$. Therefore, given $\mathbf{Z}^{(2)}$ and $\mathbf{Y}^{(2)}$, $R_1$ and $R_2$ are independent since the result holds for $p-1$. Finally, $R_1$ and $R_2$ are independent unconditionally since neither distribution depends on the conditions.

## Appendix B

First note that $|n\mathbf{S}_j|$ is distributed as (Anderson[8])

$$|\boldsymbol{\Sigma}_0| \times \chi^2_{n-1} \times \chi^2_{n-2} \times \cdots \times \chi^2_{n-p}$$

and $|(N-n)\bar{\mathbf{S}}_{(j)}|$ is distributed as

$$|\boldsymbol{\Sigma}_0| \times \chi^2_{N-m-n+1} \times \chi^2_{N-m-n} \times \cdots \times \chi^2_{N-m-n-p+2},$$

where the chi-square distributions with various degrees of freedom which appear in the products are all independent. Furthermore, since $\mathbf{S}_j$ and $\bar{\mathbf{S}}_{(j)}$ are independent,

$$\left(\prod_{i=1}^{p} \frac{N-m-n+2-i}{n-i}\right) \times \frac{|n\mathbf{S}_j|}{|(N-n)\bar{\mathbf{S}}_{(j)}|} \sim \prod_{i=1}^{p} F_{n-i,N-m-n+2-i},$$

where $F_{n-i,N-m-n+2-i}$, $i = 1, 2, \ldots, p$, are independent.

For the special case when $p = 2$, the exact distribution of $|n\mathbf{S}_j|/|(N-n)\bar{\mathbf{S}}_{(j)}|$ can be determined. It is based on a simple result (Anderson[8]) which states that if $\chi^2_{n-1}$ and $\chi^2_{n-2}$ are two independent chi-square distributions, then $\chi^2_{n-1} \times \chi^2_{n-2}$ is distributed as $(\chi^2_{2n-4})^2/4$. Therefore, in the case when $p = 2$, the statistic $V_j$ as defined in (4), $j = 1, 2, \ldots, m$, can be computed as

$$V_j = P\left(F_{2n-4,2(N-m-n)} \leq \frac{N-m-n}{n-2} \times \frac{|n\mathbf{S}_j|^{1/2}}{|(N-n)\bar{\mathbf{S}}_{(j)}|^{1/2}}\right).$$

Likewise, when $p = 2$, the statistic $V_k$ as defined in (6) can be computed as

$$V_k = P\left(F_{2n-4,2(N_I-m_I-n)} \leq \frac{N_I-m_I-n}{n-2} \times \frac{|n\mathbf{S}_k|^{1/2}}{|(N_I-n)\bar{\mathbf{S}}_I|^{1/2}}\right).$$

In the case when $p \geq 3$, as suggested in Gnanadesikan and Gupta[14] and in Anderson,[8] one can use the normal distribution to approximate the distributions of

$$\log \frac{|\mathbf{S}_j|}{|\boldsymbol{\Sigma}_0|} \quad \text{and} \quad \log \frac{|\bar{\mathbf{S}}_{(j)}|}{|\boldsymbol{\Sigma}_0|}.$$

More specifically, following the results in Muirhead,[15] it can easily be shown that both

$$\sqrt{\frac{n-1}{2p}} \log \frac{|\mathbf{S}_j|}{|\boldsymbol{\Sigma}_0|} \quad \text{and} \quad \sqrt{\frac{(m-1)(n-1)}{2p}} \log \frac{|\mathbf{S}_{(j)}|}{|\boldsymbol{\Sigma}_0|}$$

are asymptotically distributed as $N(0,1)$, the standard normal distribution. Furthermore, since $\mathbf{S}_j$ and $\bar{\mathbf{S}}_{(j)}$ are independent, it follows that

$$\sqrt{\frac{(m-1)(n-1)}{2pm}} \log \frac{|\mathbf{S}_j|}{|\bar{\mathbf{S}}_{(j)}|} \tag{B.1}$$

is asymptotically distributed as $N(0,1)$. Therefore, in the case when $p \geq 3$, we suggest that $V_j$ as defined in (4) be computed as

$$V_j = P\left(Z \leq \sqrt{\frac{(m-1)(n-1)}{2pm}} \log \frac{|\mathbf{S}_j|}{|\bar{\mathbf{S}}_{(j)}|}\right).$$

Similarly, one can compute $V_k$ as defined in (6) as

$$V_k = P\left(Z \leq \sqrt{\frac{(m_I-1)(n-1)}{2pm_I}} \log \frac{|\mathbf{S}_k|}{|\bar{\mathbf{S}}_I|}\right).$$

It should be noted that in the case when normal approximation is used, $V_j$ (likewise $V_k$) is distributed as $U(0,1)$ asymptotically. For small $n$, the exact distribution of (B.1) tends to have heavier tails than $N(0,1)$ does. This in turn makes the statistic $V_j$ tend to fall more likely in the rejection regions as determined based on $U(0,1)$.

## Appendix C

The construction of the $T^2$-chart is based on Hotelling-$T^2$ (Hotelling[16]) which is defined as, for $j = 1, 2, \ldots, m$,

$$T^2 = (\bar{\mathbf{X}}_j - \bar{\bar{\mathbf{X}}})' \bar{\mathbf{S}}^{-1} (\bar{\mathbf{X}}_j - \bar{\bar{\mathbf{X}}}).$$

Under normality, $T^2$ is distributed as

$$\frac{p(m-1)}{(N-m-p+1)} \times F_{p,N-m-p+1}.$$

Therefore, for a given probability of type-I error $\alpha$, the UCL of the $T^2$-chart can be computed as

$$\text{UCL} = \frac{p(m-1)}{(N-m-p+1)} \times f_{1-\alpha}(p, N-m-p+1),$$

where

$$P(F_{p,N-m-p+1} \leq f_{1-\alpha}(p, N-m-p+1)) = 1-\alpha.$$

As for the $|S|$-chart, the construction is based on the so-called generalized variance, i.e., the determinant of the sample variance-covariance matrix (see, e.g., Alt[17]). More specifically, for $j = 1, 2, \ldots, m$, one plots $|\mathbf{S}_j|$ on the $|S|$-chart with the $3\sigma$ limits defined as

$$\text{LCL} = \frac{|\bar{\mathbf{S}}|}{b_1}(b_1 - 3\sqrt{b_2})$$

$$\text{UCL} = \frac{|\bar{\mathbf{S}}|}{b_1}(b_1 + 3\sqrt{b_2}),$$

where

$$b_1 = \frac{1}{(n-1)^p} \times \prod_{i=1}^{p}(n-i)$$

$$b_2 = \frac{1}{(n-1)^{2p}} \times \prod_{i=1}^{p}(n-i) \times \left[\prod_{j=1}^{p}(n-j+2) - \prod_{j=1}^{p}(n-j)\right].$$

## References


1. W. A. Shewhart, "The applications of statistics as an aid in maintaining quality of a manufactured product," *Journal of the American Statistical Association* (1925), pp. 546–548.
2. F. A. Spiring and S. W. Cheng, "An alternative variables control chart: The univariate and multivariate case," *Statistica Sinica* **8** (1998), pp. 273–287.
3. S. W. Cheng and G.-Y. Li, "A single variables control chart," *Statistica Sinica*, to appear.
4. J. Repco, "Process capability plot," *The Proceedings of the 30th EQQC Conference* (1986), pp. 373–381.
5. Y. Van Nuland, "ISO 9002 and the circle technique," *Quality Engineering* **5** (1992), pp. 269–291.
6. M. T. Chao and S. W. Cheng, "Semicircle control chart for variables data," *Quality Engineering* **8** (1996), pp. 441–446.
7. A. B. Yeh, D. K. J. Lin and C. Venkataramani, "The box-chart: A new variables control chart," Unpublished manuscript (Department of ASOR, Bowling Green State University, 1999).
8. T. W. Anderson, *An Introduction to Multivariate Statistical Analysis* (2nd edition) (John Wiley & Sons, 1984).
9. A. Mitra, *Fundamentals of Quality Control and Improvement* (MacMillan, 1993).
10. R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis* (4th edition) (Prentice-Hall, 1998).
11. S. D. Balkin and D. K. J. Lin, "Performance of sensitizing rules on Shewhart control charts with autocorrelated data," in *International Journal of Reliability, Quality and Safety Engineering* **8**(2) (2001), pp. 159–171.
12. R. L. Mason, N. D. Tracy and J. C. Young, "Decomposition of $T^2$ for multivariate control chart interpretation," *Journal of Quality Technology* **27** (1995), pp. 99–108.
13. R. L. Mason, C. W. Champ, N. D. Tracy, S. J. Wierda and J. C. Young, "Assessment of multivariate process control techniques," *Journal of Quality Technology* **29** (1997), pp. 140–143.
14. M. Gnanadesikan and S. S. Gupta, "A selection procedure for multivariate normal distributions in terms of the generalized variances," *Technometrics* **12** (1970), pp. 103–117.
15. R. J. Muirhead, *Aspects of Multivariate Statistical Theory* (John Wiley & Sons, 1982).
16. H. Hotelling, "Multivariate quality control, illustrated by the air testing of sample bombsights," *Techniques of Statistical Analysis* (1947), pp. 113–184.
17. F. B. Alt, "Multivariate quality control," *Encyclopedia of Statistical Sciences* **6** (1985), pp. 110–122.

## About the Authors

Dr. Arthur B. Yeh is an Associate Professor of Statistics at the Department of Applied Statistics and Operations Research, Bowling Green State University. His areas of interest in research include bootstrap, data mining, multivariate data depth, statistical computing, statistical process control, run-to-run process control.

Dr. Dennis K. J. Lin is a Professor of Management Science and Statistics at the Department of Management Science and Information Systems, Penn State University. His research interests are quality engineering, design of experiments, reliability, statistical process control, quality assurance, response surface methodology and data mining. He has published nearly 100 papers in a wide variety of journals, including *Technometrics*, *Journal of Royal Statistical Society, Ser. C*, *Journal of Quality Technology*, *IEEE Transaction on Reliability*. Dr. Lin is an elected Fellow of the American Statistical Association, an elected member of the International Statistical Institute, a senior member of the American Society for Quality, a lifetime member of the International Chinese Statistical Association, a Fellow of the Royal Statistical Society, and has received the Most Outstanding Presentation Award from SPES of ASA.