*Journal of the Korean Data & Information Science Society*
*2011*, **22**(4), 803–809

한국데이터정보과학회지

# Control charts for monitoring correlation coefficients in variance-covariance matrix†

Duk-Joon Chang[1] · Sunyeong Heo[2]

[12]Department of Statistics, Changwon National University

Received 27 May 2011, revised 23 June 2011, accepted 29 June 2011

**Abstract**

Properties of multivariate Shewhart and CUSUM charts for monitoring variance-covariance matrix, specially focused on correlation coefficient components, are investigated. The performances of the proposed charts based on control statistic Lawley-Hotelling $V_i$ and likelihood ratio test (LRT) statistic $TV_i$ are evaluated in terms of average run length (ARL). For monitoring correlation coefficient components of dispersion matrix, we found that CUSUM chart based on $TV_i$ gives relatively better performances and is more preferable, and the charts based on $V_i$ perform badly and are not recommended.

*Keywords*: Average run length, false alarm, CUSUM, likelihood ratio test.

## 1. Introduction

In many situations, there exist multiple quality variables to define the quality of output rather than a single quality variable. And shifts in correlation coefficients of quality variables are important when the strength of linear relationship between two or more variables largely affect the quality of product. Especially in chemical industry, changing in correlation coefficients of quality variables in production process is often important.

The quality of a product is often characterized by joint levels of several quality characteristics. Because it is inappropriate to use individual charts to detect any changes of each quality variable or each process parameter in this case, a multivariate quality control procedure for simultaneously monitoring correlated variables is needed. The multivariate procedure to quality control was first introduced by Hotelling (1947) and became popular in recent years. Jackson (1959) and Ghare and Torgersen (1968) presented multivariate Shewhart chart based on Hotelling's $T^2$ statistic. Woodall and Ncube (1985) extended the univariate CUSUM procedure to the multivariate case for monitoring mean vector of quality variables.

† This research is financially supported by Changwon National University in 2009-2010.
[1] Corresponding author: Professor, Department of Statistics, Changwon National University, Changwon 641-773, Korea.
E-mail: djchang@changwon.ac.kr
[2] Associate professor, Department of Statistics, Changwon National University, Changwon 641-773, Korea.

Up to the present, multivariate control charts have been widely used for monitoring process mean vector. But, relatively little attention has been given to the use of multivariate charts for monitoring variance-covariance matrix. In this paper, we propose several different control statistics and evaluate numerical performances of Shewhart and CUSUM charts for monitoring the dispersion matrix $\Sigma$, specially correlation coefficients of correlated quality variables.

## 2. Evaluating control statistics

Suppose that the production process of interest has $p$ quality variables represented by the random vector $\underline{X} = (X_1, X_2, \cdots, X_p)'$, $p = 2, 3, \cdots$, and $\underline{X}$ has a multivariate normal distribution $N_p(\underline{\mu}, \Sigma)$. At each sampling time $i$, we obtain an independent random sample vector $\underline{X}_i$, where $\underline{X}_i = (\underline{X}'_{i1}, \cdots, \underline{X}'_{\in})'$ is a sample of observations and $\underline{X}_{ij} = (X_{ij1}, X_{ij2}, \cdots, X_{ijp})'$. Thus $\underline{X}_i$ is $np \times 1$ vector.

Let $\underline{\theta}_0 = (\underline{\mu}_0, \Sigma_0)$ be the known target values for the process parameters $\underline{\mu}_0$ of $p$ quality variables and $\Sigma_0$ is represented as

$$\underline{\mu}_0 = \begin{pmatrix} \mu_{10} \\ \mu_{20} \\ \vdots \\ \mu_{p0} \end{pmatrix} \text{ and } \Sigma_0 = \begin{pmatrix} \sigma_{10}^2 & \rho_{120}\sigma_{10}\sigma_{20} & \cdots & \rho_{1p0}\sigma_{10}\sigma_{p0} \\ \rho_{120}\sigma_{10}\sigma_{20} & \sigma_{20}^2 & \cdots & \rho_{2p0}\sigma_{20}\sigma_{p0} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1p0}\sigma_{10}\sigma_{p0} & \rho_{2p0}\sigma_{20}\sigma_{p0} & \cdots & \sigma_{p0}^2 \end{pmatrix},$$

where the target covariance component of $X_r$ and $X_s$ is $\sigma_{rs0} = \rho_{rs0}\sigma_{r0}\sigma_{s0}$ for $r, s = 1, 2, \cdots, p$.

In univariate case, the process dispersion $\sigma^2$ can be monitored by $S^2$ chart under normality assumption, where $S^2$ denotes an unbiased sample variance for a random sample of size $n$ from a process. The $S^2$ chart signals for large values of $S_i^2$ or equivalently for large values of $T_i = (n-1)S_i^2/\sigma_0^2$ where $\sigma_0^2$ is target value of the process dispersion $\sigma^2$ and $S_i^2$ is obtained at sampling time $i$.

For multivariate case, one possible multivariate version of $T_i$ is

$$V_i = \sum_{j=1}^{n} (\underline{X}_{ij} - \overline{\underline{X}_i})' \Sigma_0^{-1} (\underline{X}_{ij} - \overline{\underline{X}_i}) = tr(A_i \Sigma_0^{-1}) \tag{2.1}$$

where $A_i = \sum_{j=1}^{n} (\underline{X}_{ij} - \overline{\underline{X}}_i)(\underline{X}_{ij} - \overline{\underline{X}}_i)'$.

The distribution of the Lawley-Hotelling statistic $V_i$ was studied by Lawley (1938) and Hotelling (1951). When the process is in-control, the dispersion matrix $\Sigma$ is $\Sigma_0$ and the control statistic $V_i$ has a chi-squared distribution with $(n-1)p$ degrees of freedom. Hotelling (1947) proposed that the statistic $V_i$ can be used to monitor the process dispersion matrix of $p$ quality variables.

Hui (1980) proposed the sample generalized variances for monitoring the process dispersion matrix using the following statistic $L_i$ as

$$L_i = \frac{|S_i|}{|\Sigma_0|} \tag{2.2}$$

where $p \times p$ sample dispersion matrix $S_i$ is $A_i/(n-1)$. It is known that the statistic $\sqrt{n-1}(L_i - 1)$ is asymptotically normally distributed with mean 0 and variance $2p$ (Anderson, 1958). Alt (1982) also proposed the use of sample generalized variance $|S_i|$ to monitor dispersion matrix $\Sigma$.

Another control statistic for monitoring $\Sigma$ can be obtained from the likelihood ratio test (LRT) statistic for testing $H_0 : \Sigma = \Sigma_0$ vs $H_1 : \Sigma \neq \Sigma_0$ where target mean vector $\underline{\mu}_0$ of the quality variables is known, since the general statistical quality control procedures can be considered as a series of repetitive significant tests.

The region above the UCL corresponds to the LRT rejection region. For the $i$th sample, likelihood ratio $\lambda$ can be expressed as

$$\lambda = n^{-np/2} \cdot |A_i \Sigma_0^{-1}|^{n/2} \cdot \exp\left[-\frac{1}{2} tr(\Sigma_0^{-1} A_i) + \frac{1}{2} np\right].$$

Let $TV_i$ be $-2\ln\lambda$. Then

$$TV_i = tr(A_i \Sigma_0^{-1}) - n\ln|A_i| + n\ln|\Sigma_0| + np\ln n - np. \tag{2.3}$$

Hence, $TV_i$ can be used as the control statistic for monitoring $\Sigma$.

# 3. Multivariate control charts

## 3.1. Multivariate Shewhart charts

Shewhart chart is widely used to display sample data from a process for the purpose of determining whether a production process is in-control, for bringing an out-of-control process into in-control, and for monitoring a process to make sure that it stays in-control. A Shewhart chart has a good ability to detect large changes in monitored parameter quickly. The basic Shewhart chart, although simple to understand and apply, uses only the information in the current sample and is thus relatively inefficient in detect small shifts in control parameter.

Since the control limits for a multivariate Shewhart chart based on the control statistic $V_i$ would be set as $\{0, \chi^2_{1-\alpha}[(n-1)p]\}$, a Shewhart chart based on $V_i$ signals whenever

$$V_i \geq \chi^2_{1-\alpha}[(n-1)p]. \tag{3.1}$$

If the process shifts from $\Sigma_0$ then it is difficult to obtain the exact distribution of $V_i$. Thus in order to obtain the percentage points of $V_i$ when the process is out-of-control state, it is necessary to use simulations.

For a Shewhart chart based on the LRT statistic $TV_i$ would be set by using percentage point of $TV_i$, a Shewhart chart based on $TV_i$ signals whenever

$$TV_i \geq h_{TV(S)} \tag{3.2}$$

where $h_{TV(S)}$ can be obtained to satisfy a specified in-control ARL by simulation.

Since it is difficult to obtain the exact distribution of $TV_i$ when the process is in-control or out-of-control states, $h_{TV(S)}$ and performances of this chart are obtained by simulations.

### 3.2. Multivariate CUSUM charts

The CUSUM chart is a good alternative to the Shewhart chart and is often used instead of standard Shewhart chart when detection of small shifts in a production process is important. A CUSUM chart directly incorporates all of the information in the sequence of sample values by plotting the cumulative sum of the deviation of the sample values from the target value.

A multivariate CUSUM chart based on the statistic $V_i$ in (2.1) is given by

$$Y_{V,i} = \max\{Y_{V,i-1}, 0\} + (V_i - k_V) \tag{3.3}$$

where $Y_{V,0} = \omega_V$ ($\omega_V \geq 0$) and reference value $k_V \geq 0$. This chart for dispersion matrix signals whenever $Y_{V,i} \geq h_V$.

When the process parameters are on-target, decision interval $h_V$ can be evaluated by the Markov chain or integral equation approach to satisfy a specified in-control ARL. And when the process parameters in $\Sigma$ have changed, the performances of this chart can be evaluated by simulation.

And for a CUSUM chart based on the statistic $TV_i$ in (2.3) can also be constructed as

$$Y_{TV,i} = \max\{Y_{TV,i-1}, 0\} + (TV_i - k_{TV}) \tag{3.4}$$

where $Y_{TV,0} = \omega_{TV}$ ($\omega_{TV} \geq 0$) and $k_{TV} \geq 0$. This chart signals whenever $Y_{TV,i} \geq h_{TV}$.

Since it is difficult to obtain the exact performances of multivariate CUSUM scheme based on $TV_i$, the percentage point and properties of this chart can be evaluated by simulation under the assumption that the process parameters of the process are on-target or changed.

## 4. Numerical performances and concluding remarks

The ability of a control chart to detect any shifts in the production process is determined by the length of time required to signal. Thus, a good control chart detects shifts quickly in the process when the process is out-of-control state, and produce few false alarms when the process is in-control state. To compare the performances of the proposed control charts, $P$(signal is given $|\Sigma_0$) was fixed for all proposed charts.

In order to evaluate the performances and compare the proposed multivariate Shewhart and CUSUM charts fairly, it is necessary to calibrate each schemes so that on-target ARL $E(N|\Sigma_0)$ be the same for all proposed schemes. In our computation, each scheme was calibrated so that the on-target ARL was approximately equal to 200 and the sample size for each characteristic was five for $p = 3$ and $p = 4$. For convenience, we let that the sampling interval of unit time $d = 1$ and known target mean vector $\underline{\mu}_0 = \underline{0}$. The performance of the charts for monitoring variance-covariance matrix depends on the components of $\Sigma$. For computational simplicity in our computation, we assume that $\sigma_{r0}^2 = 1$, $\rho_{rs0} = 0.3$ for $r, s = 1, 2, \cdots, p$.

Since it is not possible to investigate all of the different ways in which $\Sigma$ could change, we consider the following typical types of shifts for comparison in the process parameters :

(1) $SD_i$ : $\sigma_{10}$ in $\Sigma_0$ is increased to $[1 + 0.2i]$, $i = 1, 2, \cdots, 6$.
(2) $C_i$ : $\rho_{120}$ and $\rho_{210}$ of $\Sigma_0$ are changed to $[0.3 + 0.05i]$, $i = 1, 2, \cdots, 6$.

After the reference value of the proposed CUSUM chart based on $V_i$, decision interval $h_V$ was calculated by Markov chain method with the number of transient states $r = 100$.

A good choice for reference valus $k$ of CUSUM charts depends on the number of quality variables in the proposed control scheme and the size of shift of interesting.

The design parameters $h_{TV(S)}$, $h_V$, $h_{TV}$ values, and the ARL performances of the proposed Shewhart and CUSUM charts were obtained by simulation with 10,000 iterations. Numerical performances of the proposed charts in this study are given in Figures 4.1 through 4.4.

Figures 4.1-4.2 show that a shift for only variance components in $\Sigma$ has occurred, the chart based on the Lawley-Hotelling $V_i$ control statistic seems more efficient than $TV_i$ control statistic, and CUSUM procedure is more efficient than Shewhart procedure.

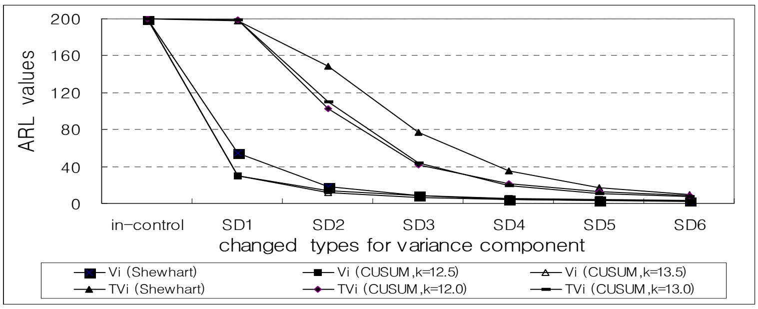


**Figure 4.1** ARL performances when $\sigma_1$ in $\Sigma$ has changed (p =3)

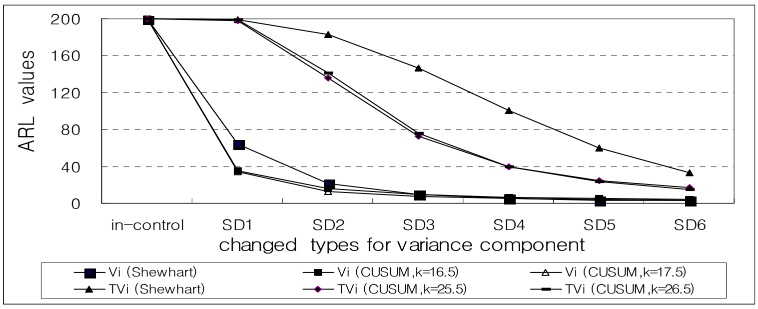


**Figure 4.2** ARL performances when $\sigma_1$ in $\Sigma$ has changed (p =4)

Figures 4.3-4.4 show the performances of the charts based on $V_i$ and $TV_i$ when some shift on correlation coefficient component $\rho_{12}$ in $\Sigma$ has occurred. We can note that ARL values of the charts based on $V_i$ considerably increase compared with in-control state ARL 200 when some shifts on correlation coefficients have occurred.

It is common that if there are small or moderate shifts in the production process then the ARL values of CUSUM are smaller than Shewhart chart's ones. However, from the Figures 4.3-4.4, ARL values of CUSUM charts are greater than that of Shewhart charts when small or moderate shifts on correlation coefficients of quality variables have occurred. Therefore, when any shifts in correlation coefficient component in $\Sigma$ are anticipated, the charts based on $V_i$ are not recommendable, and the CUSUM chart based on $TV_i$ gives relatively better performances and is more preferable.

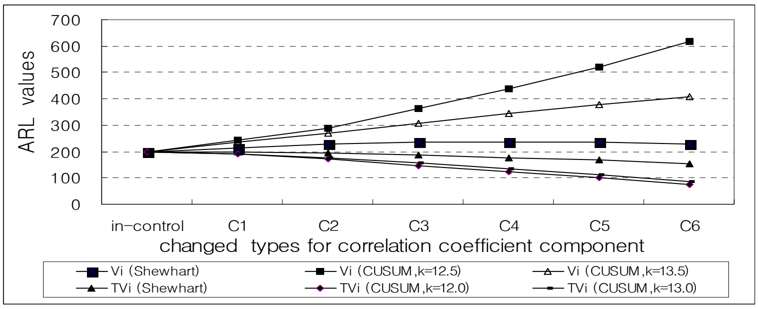


**Figure 4.3** ARL performances when $\rho_{12}$ in $\Sigma$ has changed (p =3)

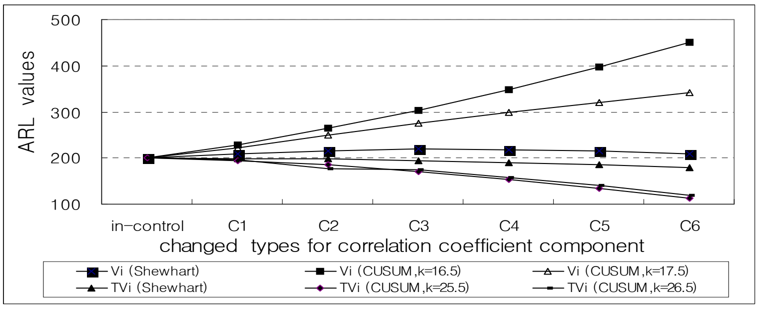


**Figure 4.4** ARL performances when $\rho_{12}$ in $\Sigma$ has changed (p =4)

Based on Figures 4.1-4.4, when some shifts both on variance components and correlation coefficient components are anticipated, we recommend to use the CUSUM charts based on control statistic $TV_i$ since the control charts based on $V_i$ is bad at detecting the changes in correlation coefficient of quality variables, even though in some process changes the ARL values from $V_i$ may be possibly smaller than those from $TV_i$.

However, if the changes of quality characteristics' correlation coefficient do not nearly effect on the product quality, then the CUSUM charts based on $V_i$ may be applicable.

## References

Alt, F. B. (1982). Multivariate quality control. In *The Encyclopedia of Statistical Sciences*, edited by S. Kotz and N. L. Johnson, Wiley, New York.

Anderson, T. W. (1958). *An introduction to multivariate statistical analysis*, Wiley, New York.

Chang, D. J. and Heo, S. (2010). Properties of variable sampling interval control charts. *Journal of the Korean Data & Information Science Society,* **21**, 819-829.

Ghare, P. H. and Torgersen, P. E. (1968). The multicharacteristic control charts. *Journal of Industrial Technology,* **19**, 269-272.

Hui, Y. V. (1980). *Topics in statistical process control*, Ph.D. dissertation, Virginia Politechnic Institute and State University, Blacksberg, Virginia.

Hotelling, H. (1947). Multivariate quality control. In *Techniques of Statistical Analysis*, McGraw-Hill, New York, 111-184.

Hotelling, H. (1951). A generalized T test and measure of multivariate dispersion. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Pprobability*, U. of California Press, Berkeley, 23-42.

Jackson, J. S.(1959). Quality control methods for several related variables. *Technometrics*, **1**, 359-377.

Kim, M. and Lee, S. (2010). The CUSUM test for stochastic volatility models. *Journal of the Korean Data & Information Science Society,* **21**, 1305-1310.

Lawley, D. N. (1938). A generalization of Fisher's Z-test. *Biometrica*, **30**, 180-187.

Lim, C and Cho, G. Y. (2008). A new EWMA control chart for monitoring the covariance matrix of bivariate processes. *Journal of the Korean Data & Information Science Society*, **19**, 677-683.

Na, O., Ko, B. and Lee S. (2010). Cusum of squares test for discretely observed sample from multidimensional diffusion processes. *Journal of the Korean Data & Information Science Society*, **21**, 555-560.

Wooda ll, W. H. and Ncube, M. M. (1985). Multivariate CUSUM quality control procedure. *Technometrics*, **27**, 285-292.