A Multivariate CUSUM Chart Handling Auto- and Cross-correlated Observations

Kyu Young Lee[1], Chuljin Park[2], and Mi Lim Lee[3]

[1,2] *Department of Industrial Engineering, Hanyang University, Seoul 04763, Republic of Korea*

kaylee@hanyang.ac.kr
parkcj@hanyang.ac.kr

[3]*College of Business Administration, Hongik University, Seoul 04066, Republic of Korea*
mllee@hongik.ac.kr

ABSTRACT

Multivariate CUSUM charts have been widely used as statistical-process-control tools to detect out-of-control states of monitoring variables. Most of earlier studies regarding multivariate CUSUM charts assume that observations of monitoring variables are independent or correlated with a limited structure. In this paper, we suggest a multivariate CUSUM chart that can handle the observations auto- and cross-correlated. The chart estimates its control limit analytically and captures small mean shifts faster when compared to existing charts.

1. INTRODUCTION

With recent advances of sensors and computers, it becomes possible to acquire and manage a large amount of data in real time. In order to monitor such real time data, statistical process control (SPC) has been studied popularly and used to raise an alarm when a monitoring process is in an out-of-control state. Since the real time data monitored often have auto-correlation as well as cross-correlation, it is important for SPC methods to be designed with robustness and sensitivity to both correlations.

Multivariate-CUSUM(MCUSUM) control charts have been used as SPC tools to detect mean changes in vectors observed. Many researchers have studied multivariate CUSUM chart (Healy, 1987; Crosier,1988; Pignatiello & Runger, 1990;) and it is well-known that MCUSUM chart is sensitive to detect small mean shifts in observation vectors. However, most of earlier studies for MCUSUM charts assume that the observations are independent and identically distributed (i.i.d.), and warn that the charts may not work if the assumption is violated. Noorossana and Vaghefi (2006) show that auto-correlated observation data are inappropriate to be applied to MCUSUM charts directly.

Some MCUSUM charts have been studied to deal with cross-correlation in observation data. Woodall and Ncube (1985) monitor multiple univariate CUSUM charts that work well if there is little or no cross-correlation among the observations. Rogerson and Yamada (2004) find that monitoring an MCUSUM chart is better than monitoring multiple univariate CUSUM charts in terms of detection performance. Later, Jiang *et al.* (2011) propose an MCUSUM chart that can be used even with cross-correlated observation vectors. Motivated by the MCUSUM chart from Jiang *et al.* (2011), Lee *et al.* (2014) introduce Separated-MCUSUM(S-MCUSUM) chart with the techniques to approximate its control limit analytically. However, the approximation of the control limit is not suitable for auto-correlated observation data because average run length (ARL) of the MCUSUM chart becomes much shorter than the expected ARL with the auto-correlation.

A traditional approach to address auto-correlation in observation data is using model-based control charts (Loredo, 2002; Kalgonda, 2004; Noorossana and Vaghefi, 2006;). Model-based control charts first set a model to represent the auto-correlated observation data, and use the residuals as the input data of control charts for monitoring. A typical selection for the model is a time-series model and, since the residuals are not auto-correlated usually, MCUSUM charts are often used as the control charts for monitoring the residuals. To improve detection performance with small mean shifts, Arkat *et al.* (2007) and Issam and Mohamad (2008) propose model-based MCUSUM charts using an artificial neural network (ANN) and a support vector regression (SVR), respectively. Kim *et al.* (2012) compare various model-based MCUSUM charts under 9 scenarios with different numbers of dimensions and different degree of auto-correlation. In their study, Kim *et al.* (2012) show that the SVR and ANN based MCUSUM charts perform better than other charts considered.

In this paper, we proposed a model based MCUSUM chart that can handle the auto- and cross-correlated observation data. For addressing the autocorrelation, we adapt Vector auto-regression (VAR) to model the data and apply the SMCUSUM chart to monitor the residuals. Section 2 describes our MCUSUM chart proposed. In Section 3, we compared the performances of the proposed chart to those of existing charts. Finally, Section 4 provides conclusions of this paper.

2. ALGORITHM DESCRIPTION

In order to deal with observations auto- and cross- correlated, we use a VAR to obtain residuals of the observations first and then apply the SMCUSUM chart (Lee *et al.*, 2014) to the residuals.

Suppose that we monitor a process of $\boldsymbol{X}_t = (x_{1t}, x_{2t}, \cdots, x_{mt})'$, an $(m\times1)$ observation vector of m variables at each time *t*. Under in-control state, we model $\boldsymbol{X}_t$ as a VAR with p lags as follows.

$$\boldsymbol{X}_t = \boldsymbol{\nu} + \boldsymbol{A}_1\boldsymbol{X}_{t-1} + \boldsymbol{A}_2\boldsymbol{X}_{t-2} + \cdots + \boldsymbol{A}_p\boldsymbol{X}_{t-p} + \boldsymbol{\varepsilon}_t, \quad (1)$$

where $\boldsymbol{A}_i$ for $i = 1, 2, \cdots, p$ are $(m\times m)$ coefficient matrices, $\boldsymbol{\nu} = (\nu_1, \nu_2, \cdots, \nu_m)'$ is a vector of intercept terms, and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \cdots, \varepsilon_{mt})'$ is a white noise vector that follows an i.i.d. multivariate normal distribution with mean $E[\boldsymbol{\varepsilon}_t] = \boldsymbol{0}$ and a known time invariant covariance matrix $\boldsymbol{\Sigma}_r$. Note that $\boldsymbol{\varepsilon}_t$ is not auto-correlated but cross-correlated.

Suppose that $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_m)' = E[\boldsymbol{X}_t]$ when the monitoring process of $\boldsymbol{X}_t$ is in in-control state. With $\boldsymbol{\mu}$, Eq. (1) can be reformulated as

$$\boldsymbol{X}_t = \boldsymbol{\mu} + \boldsymbol{A}_1(\boldsymbol{X}_{t-1} - \boldsymbol{\mu}) + \cdots + \boldsymbol{A}_p(\boldsymbol{X}_{t-p} - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}_t, \quad (2)$$

and the estimated vector $\widehat{\boldsymbol{X}}_t$ is represented by

$$\widehat{\boldsymbol{X}}_t = \boldsymbol{\mu} + \boldsymbol{A}_1(\boldsymbol{X}_{t-1} - \boldsymbol{\mu}) + \cdots + \boldsymbol{A}_p(\boldsymbol{X}_{t-p} - \boldsymbol{\mu}). \quad (3)$$

Then we can obtain the residual vector as follows

$$\boldsymbol{R}_t = (r_{1t}, r_{2t}, \cdots, r_{mt})' = \boldsymbol{X}_t - \widehat{\boldsymbol{X}}_t. \quad (4)$$

When the monitoring process is under out-of-control state, the mean vector $\boldsymbol{\mu}$ is shifted by $\boldsymbol{\delta}_x = (\delta_{x_1}, \delta_{x_2}, \cdots, \delta_{x_m})'$. Based on the mean shift $\boldsymbol{\delta}_x$ in observations and Eqs. (2)-(4), we can calculate the corresponding mean shift in residuals $\boldsymbol{\delta}_R$,

$$\boldsymbol{\delta}_R = (\delta_{r_1}, \delta_{r_2}, \cdots, \delta_{r_m})' = (\boldsymbol{I} - \boldsymbol{A}_1 - \cdots - \boldsymbol{A}_p)\boldsymbol{\delta}_x. \quad (5)$$

$\boldsymbol{R}_t$ is expected to follow i.i.d. multivariate normal distribution with the mean $\boldsymbol{0}$ in the in-control state and $\boldsymbol{\delta}_R$ in the out-of-control state. The variance-covariance matrix of $\boldsymbol{R}_t$ can be estimated by $\boldsymbol{\Sigma}_r$.

With the baseline mean value 0 under in-control state and the marginal variance σ_j^2 of each residual component r_{jt} for $j = 1, 2, \cdots, m$, we generate the standardized residual vector $\boldsymbol{Y}_t = (y_{1t}, y_{2t}, \cdots, y_{mt})'$ where $y_{jt} = \frac{(r_{jt} - 0)}{\sigma_j}$, the standardized residual mean shift vector $\boldsymbol{\delta}_y = \left(\frac{\delta_{r_1}}{\sigma_1}, \cdots, \frac{\delta_{r_m}}{\sigma_m}\right)'$, and the variance-covariance matrix of $\boldsymbol{Y}_t$, $\boldsymbol{\Sigma}_y$. Since $\boldsymbol{Y}_t$ is a standardized residual vector that contains cross-correlation only, we can apply the SMCUSUM chart for $\boldsymbol{Y}_t$.

Motivated by the statistic $S_t^{c,r}$ from Lee *et al.* (2014) and Jiang *et al.* (2011), we set our monitoring MCUSUM statistic:

$$S_t = \max(0,\ S_{t-1} + l_t) \quad (6)$$

where

$$l_t = \boldsymbol{\delta}_y'\boldsymbol{\Sigma}_y^{-1}\boldsymbol{y}_t - \frac{\boldsymbol{\delta}_y'\boldsymbol{\Sigma}_y^{-1}\boldsymbol{\delta}_y}{2} \quad (7)$$

As soon as S_t hits a predetermined control limit, an alarm is raised to announce the detection of out-of-control state. The control limit is calibrated to achieve a target average run length in in-control state (ARL_0).

Lee *et al.* (2014) inspired by Kim *et al.* (2008) provide an equation that enables us to analytically estimate the control limit of S_t for a target ARL_0:

$$\text{ARL} \approx \begin{cases} \dfrac{\Omega^2}{2d^2}\left(\exp\left[-\dfrac{2d(H+1.166\Omega)}{\Omega^2}\right] - 1 + \dfrac{2d(H+1.166\Omega)}{\Omega^2}\right) & \text{if } d \neq 0 \\ \left(\dfrac{H+1.166\Omega}{\Omega}\right)^2 & \text{if } d = 0 \end{cases} \quad (8)$$

where $d \equiv E[l_t]$, $\Omega^2 = Var[\boldsymbol{\delta}_y'\boldsymbol{\Sigma}_y^{-1}\boldsymbol{y}_t]$, and H is the control limit of the MCUSUM statistics S_t.

3. EXPERIMENTAL STUDY

In this section, we compare the performances of our proposed MCUSUM chart to those of existing charts under various experimental configurations.

As a performance measure, we consider an average run length in the out-of-control state (ARL_1). For a common ARL_0, the chart that provides the shorter ARL_1 is the better.

3.1. Experiment setup

For the comparison of the charts, we generate a multivariate autoregressive data set using a simulation technique. Specifically, the multivariate autoregressive data set follows a VAR(1) which is modeled as follows (Arkat *et al.*, 2007; Issam and Mohamed, 2008):

$$\boldsymbol{X}_t = \boldsymbol{\mu} + \boldsymbol{A}_1(\boldsymbol{X}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}_t, \quad (9)$$

with

$$\boldsymbol{\mu} = \begin{pmatrix} 260 \\ 470 \end{pmatrix},\ \boldsymbol{\Sigma}_r = \begin{pmatrix} 99.91 & 63.99 \\ 63.99 & 69.52 \end{pmatrix},$$

and

$$\boldsymbol{A}_1 = \begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{pmatrix} = \begin{pmatrix} 0.0146 & 0.0177 \\ 0.6493 & 0.0958 \end{pmatrix}.$$

For existing charts, the target ARL_0 is set to 205 based on 1,000 replications of each shift which is the same as past studies. Note that we can analytically calculate the control limit of our chart under the target ARL_0 which is 205.

When generating the out-of-control state data, we simply change values in the mean vector of Eq. (9) by $\boldsymbol{\delta}_x$ while using the same values of the variance-covariance matrix and the coefficient matrix.

As competitive existing charts, we select the MCUSUM chart with an ANN model (Arkat *et al.*, 2007) and the MCUSUM chart with a SVR model (Issam and Mohamed, 2008) because they empirically outperform other existing charts under various settings as in Arkat *et al.* (2007), Issam and Mohamed (2007), and Kim *et al.* (2012).

3.2. Simulation results

We denote VAR-SMCUSUM as our chart, ANN-MCUSUM as the MCUSUM chart with an ANN model, and SVR-MCUSUM as the MCUSUM chart with an SVR model in Table 1 and Figure 1.

Table 1 shows ARL_1 values of ANN-CUSUM, SVR-CUSUM, and VAR-SMCUSUM for ten different settings of the mean shifts. For the case with relatively small mean shifts (i.e., from case 1 to case 5), VAR-SMCUSUM achieves ARL_1values which are 45~75% of ARL_1 values of ANN-CUSUM and SVR-CUSUM. For the rest of the cases, VAR-SMCUSUM results in ARL_1 values which are similar to or smaller than the ARL_1 values of ANN-CUSUM and SVR-CUSUM.

Figure 1 shows ARL_1 values of ANN-CUSUM, SVR-CUSUM, and VAR-SMCUSUM according to ten different cases of the mean shift. As the amount of the mean shift increases, ARL_1 values of all three charts decrease, and the differences between ARL_1 values of VAR-SMCUSUM and ANN-CUSUM or SVR-CUSUM tend to decrease.

Table 1. Empirical ARL_1 for three different charts

Case #	Mean Shift $\boldsymbol{\delta}_x$	ANN-MCUSUM (Arkat *et al.*, 2007)		SVR-MCUSUM (Issam and Mohamed, 2008)	VAR-SMCUSUM	
		Shift in N-N residuals $\boldsymbol{\delta}_N$	ARL_1	ARL_1	Shift in VAR residuals $\boldsymbol{\delta}_y$	ARL_1
1	$\begin{bmatrix} 0.50 \\ 0.30 \end{bmatrix}$	$\begin{bmatrix} 0.58 \\ 0.59 \end{bmatrix}$	142.60	145.18	$\begin{bmatrix} 0.009734 \\ -0.001962 \end{bmatrix}$	106.70
2	$\begin{bmatrix} 1.00 \\ 0.70 \end{bmatrix}$	$\begin{bmatrix} 1.10 \\ 1.24 \end{bmatrix}$	101.40	117.27	$\begin{bmatrix} 0.009734 \\ -0.001962 \end{bmatrix}$	68.58
3	$\begin{bmatrix} 1.50 \\ 1,00 \end{bmatrix}$	$\begin{bmatrix} 1.45 \\ 1,68 \end{bmatrix}$	86.10	91.82	$\begin{bmatrix} 0.146106 \\ -0.008365 \end{bmatrix}$	47.22
4	$\begin{bmatrix} 1.60 \\ 0.00 \end{bmatrix}$	$\begin{bmatrix} 2.05 \\ 0.05 \end{bmatrix}$	56.10	42.36	$\begin{bmatrix} 0.157735 \\ -0.124598 \end{bmatrix}$	25.70
5	$\begin{bmatrix} -1.35 \\ 1.00 \end{bmatrix}$	$\begin{bmatrix} -2.13 \\ 0.91 \end{bmatrix}$	28.60	34.73	$\begin{bmatrix} -0.134860 \\ 0.213574 \end{bmatrix}$	18.42
6	$\begin{bmatrix} 2.00 \\ -2.80 \end{bmatrix}$	$\begin{bmatrix} 3.80 \\ -2.75 \end{bmatrix}$	6.90	8.13	$\begin{bmatrix} 0.202127 \\ -0.459393 \end{bmatrix}$	7.87
7	$\begin{bmatrix} -8.00 \\ 0.00 \end{bmatrix}$	$\begin{bmatrix} -8.45 \\ -0.37 \end{bmatrix}$	5.10	3.58	$\begin{bmatrix} -0.788675 \\ 0.622989 \end{bmatrix}$	2.39
8	$\begin{bmatrix} 6.60 \\ -7.50 \end{bmatrix}$	$\begin{bmatrix} 10.27 \\ -7.84 \end{bmatrix}$	1.60	1.99	$\begin{bmatrix} 0.637376 \\ 0.299371 \end{bmatrix}$	1.46
9	$\begin{bmatrix} 9.00 \\ -10.00 \end{bmatrix}$	$\begin{bmatrix} -16.57 \\ 7.99 \end{bmatrix}$	1.20	1.63	$\begin{bmatrix} -0.904967 \\ 1.785313 \end{bmatrix}$	1.14
10	$\begin{bmatrix} 14.00 \\ -14.00 \end{bmatrix}$	$\begin{bmatrix} 22.89 \\ -12.91 \end{bmatrix}$	1.00	1.20	$\begin{bmatrix} 1.404972 \\ -2.608461 \end{bmatrix}$	1.00

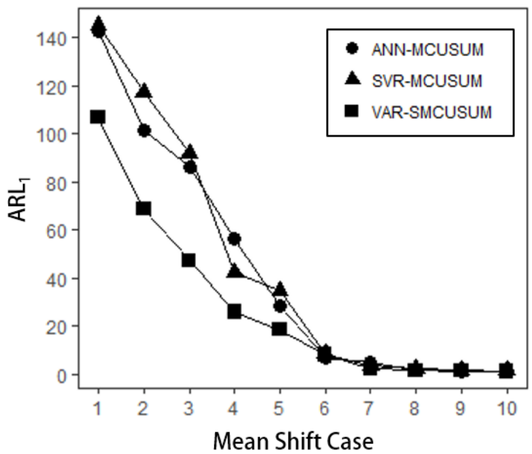

Figure 1. Change of ARL_1 for three different control chart

4. CONCLUSION

This study proposes a model based SMCUSUM chart that can deal with multivariate auto- and cross-correlated data. Our chart is effective in two points of views mainly. Fist, while other existing MCUSUM charts need to calibrate their control limits by trial-and-error to achieve the target ARL_0, our chart offers an analytical way to approximate its control limit that can save a great deal of time and efforts. Second, while other charts based on data-mining models (such as ANN and SVR) take long time to find their parameters by training the models, our chart based on VAR and SMCUSUM can be set up in relatively short time. Because it takes a long time to find parameters and train the model. When compared to other existing MCUSUM charts considered, the proposed chart provides significant improvement in ARL_1 performance under various experimental configurations, especially with the small mean shifts.

REFERENCES

Arkat, J., Niaki, S. T. A., & Abbasi, B. (2007). Artificial neural networks in applying MCUSUM residuals charts for AR (1) processes. Applied Mathematics and Computation, 189(2), 1889-1901.

Crosier, R. B. (1988). Multivariate generalizations of cumulative sum quality-control schemes. Technometrics, 30(3), 291-303.

Healy, J. D. (1987). A note on multivariate CUSUM procedures. Technometrics, 29(4), 409-412.

Issam, B. K., & Mohamed, L. (2008). Support vector regression based residual MCUSUM control chart for autocorrelated process. Applied mathematics and computation, 201(1), 565-574.

Jiang, W., Han, S. W., Tsui, K. L., & Woodall, W. H. (2011). Spatiotemporal surveillance methods in the presence of spatial correlation. Statistics in Medicine, 30(5), 569-583.

Kalgonda, A. A., & Kulkarni, S. R. (2004). Multivariate quality control chart for autocorrelated processes. Journal of Applied Statistics, 31(3), 317-327.

Kim, S. B., Jitpitaklert, W., Park, S. K., & Hwang, S. J. (2012). Data mining model-based control charts for multivariate and autocorrelated processes. Expert Systems with Applications, 39(2), 2073-2081.

Lee, M. L., Goldsman, D., Kim, S. H., & Tsui, K. L. (2014). Spatiotemporal biosurveillance with spatial clusters: control limit approximation and impact of spatial correlation. IIE Transactions, 46(8), 813-827.

Loredo, E. N., Jearkpaporn, D., & Borror, C. M. (2002). Model-based control chart for autoregressive and correlated data. Quality and Reliability Engineering International, 18(6), 489-496.

Noorossana, R., & Vaghefi, S. J. M. (2006). Effect of autocorrelation on performance of the MCUSUM control chart. Quality and Reliability Engineering International, 22(2), 191-197.

Pignatiello, J. J., & Runger, G. C. (1990). Comparisons of multivariate CUSUM charts. Journal of quality technology, 22(3), 173-186.

Rogerson, P. A., & Yamada, I. (2004). Monitoring change in spatial patterns of disease: comparing univariate and multivariate cumulative sum approaches. Statistics in Medicine, 23(14), 2195-2214.

Woodall, W. H., & Ncube, M. M. (1985). Multivariate CUSUM quality-control procedures. Technometrics, 27(3), 285-292