

RESEARCH ARTICLE

A note on generalization of Z chart

A. A. Kalgonda

 Dept. of Statistics, The New College, Kolhapur-416012, India
 annagmk@rediffmail.com; +919922770979

Abstract

In recent years, due to automation of measurement and data collection system, a process can be sampled at higher rates, which ultimately leads to autocorrelation. Control procedures using Shewhart, MCUSUM and MEWMA charts for first order autoregressive (VAR(1)) processes are available in the literature. Amongst these the Z chart has the practical advantage of diagnostic ability, to indicate which variables are responsible for out of control following a signal. But time series modeling of the process data is not always straight forward and timely revision of order of process is required. Naturally, if the order of the process becomes inadequate to describe the process, the above mentioned control chart procedures for VAR(1) process may fail in its performance. Therefore in this study, a generalized approach of Z chart for k-th order vector autoregressive process is proposed.

Keywords: Automation, data collection system, diagnostic ability, vector autoregressive process.

Introduction

Statistical process control (SPC) methods are extensively used to monitor and improve the quality of manufacturing products. In many instances of production process, the quality of the process depends upon more than one quality characteristic, which jointly determine the quality of the product. In chemical and process industries, it is common to have hundreds of variables to measure quality of product. Hotelling (1947) was the first to apply T^2 statistic to bombsight data during World War II. To name a few for subsequent works in multivariate control procedure I mention: Alt (1985), Crosier (1988), Tracy *et al.* (1992), Woodall and Ncube (1985), Lowery *et al.* (1992).

However, multivariate SPC methods are based on the assumption that the successive observation vectors are independent. Due to automation of data collection system, a process can be sampled at higher rates, which ultimately leads to autocorrelation. This has serious impact on the traditional control chart procedures. Therefore, in recent years, the SPC for autocorrelated processes have received great deal of attention of SPC researchers (Alwan and Roberts, 1988). In this regard, a few multivariate SPC methods have been developed when process observations are autocorrelated. Notably, Theodossion (1993), Bodnar and Schmid (2004) have proposed a MCUSUM chart and Kramer and Schmid (1997) a multivariate EWMA chart. However, there are practical drawbacks of using these methods. A prominent one is when evidence of out of control situation is observed; the above methods fail to provide the variable(s) responsible for the signal.

Further, as suggested by Hayter and Tsui (1994), a good multivariate quality control procedure should not only detect the status of the process but it should also identify the factors responsible for out of control situation. Kalgonda and Kulkarni (2004) have proposed a control chart having diagnostic ability when observation vectors are autocorrelated. The major limitation of their approach is that the process data has to follow VAR(1) model. But time series modeling of the process data is not always straight forward and timely revision of order of process is required. This study provides a generalized approach to monitor process mean of multivariate autocorrelated process data of k-th ordered autoregressive process.

Proposed method

Let $\underline{Y}_t = (Y_{1t}, Y_{2t}, \dots, Y_{pt})'$ be a stationary autocorrelated process denoted by $VAR_p(k)$ model and is given by

$$\underline{Y}_t = \underline{\mu} + \sum_{j=1}^k \Phi_j (\underline{Y}_{t-j} - \underline{\mu}) + \underline{\varepsilon}_t \quad (1)$$

Where $\underline{\mu}$ is the vector of mean values, which is considered to be constant over time, Φ is $p \times p$ matrix is of autocorrelation parameters and $\underline{\varepsilon}_t$ is a vector of independent normal random variables with mean vector of zeros and positive definite covariance matrix Σ of order $p \times p$. I assume that, the parameters of the mean vector are known in an industrial setting over extended period of time. Therefore, consider the case, where the mean vector $\underline{\mu}_0 = (\mu_{10}, \mu_{20}, \dots, \mu_{p0})'$ is known.



Now, let us transform $VAR_p(k)$ process, \underline{Y}_t given in model (1) in the form of $VAR_{kp}(1)$ process (Reinsel, 1993) denoted as \underline{X}_t . For this,

$$\underline{X}_t = \Psi \underline{X}_{t-1} + \underline{e}_t \tag{2}$$

$$\text{With } \underline{X}_t = (\underline{Y}'_t, \underline{Y}'_{t-1}, \dots, \underline{Y}'_{t-k+1})', \tag{3}$$

$$\underline{e}_t = (\underline{\varepsilon}'_t, 0, \dots, 0)' \tag{4}$$

$$\text{And } \Psi = \begin{bmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_k \\ I & 0 & \dots & 0 \\ 0 & I & \dots & 0 \\ \vdots & & & \\ \vdots & & & \\ 0 & 0 & \dots & I & 0 \end{bmatrix} \tag{5}$$

Note that in this $VAR_{kp}(1)$ representation, the stationary condition is equivalent to condition in the corresponding $VAR_p(1)$ representation. Now, using Yule-Walker relationship for covariance matrix of order $kp \times kp$, the cross covariance matrix $\Gamma_{(k)}$ of \underline{X}_t can be obtained when Ψ and Ω are known by solving the following equation

$$\Gamma_{(k)} = \Psi \Gamma_{(k)} \Psi' + \Omega \tag{6}$$

Note that Ω is the $kp \times kp$ covariance matrix defined on the vector $\underline{\varepsilon}_t$.

Let the autoregressive structure $\Gamma_{(k)}$ be defined by

$$\Gamma_{(k)} = \begin{bmatrix} \Gamma_{(0,k)} & \Gamma_{(1,k)} & \dots & \Gamma_{(k-1,k)} \\ \Gamma_{(1,k)} & \Gamma_{(0,k)} & \dots & \Gamma_{(k-2,k)} \\ \vdots & & & \\ \vdots & & & \\ \Gamma_{(k-1,k)} & \Gamma_{(k-2,k)} & \dots & \Gamma_{(0,k)} \end{bmatrix} \tag{7}$$

Now, under the assumption that $\underline{\varepsilon}_t \sim N(0, \Sigma)$,

$$\underline{Y}_t \sim N(\mu, \Gamma_{(0,k)}) \tag{8}$$

Assume that the process goes out of control due to shift in the single variable, rather than in all the variables simultaneously. Generally, in quality control problems this is the case.

Based on this assumption, few multivariate control charts for independent observation vectors are proposed namely: Hawkins (1991) and Timm (1996). Development of control procedure under this assumption includes testing the null hypotheses simultaneously,

$$H_{i0} : \mu_i = \mu_{i0} \quad i = 1, 2, \dots, p \tag{9}$$

Now, using Timm's technique (1996) of single step finite intersection test (FIT) and modifying appropriately for $VAR_p(k)$ process data, one leads to a control statistic

$$Z_{it}(k) = \frac{y_{it} - \mu_{i0}}{\gamma_{i(0,k)}}, \quad i = 1, 2, \dots, p, \tag{10}$$

Where, $\gamma_{i(0,k)}$ is the positive square root of the i -th diagonal element of $\Gamma_{(0,k)}$.

Whenever, $H_{i0} : \mu_i = \mu_{i0} \quad i = 1, 2, \dots, p$ is true, the test statistic $Z_{it}(k)$ follows the standard normal distribution. Now,

$$\Pr_{H0} \left\{ |Z_{it}(k)| \leq C_{\rho(0,k),\alpha}; \quad i = 1, 2, \dots, p \right\} = 1 - \alpha \tag{11}$$

Where, $C_{\rho(0,k),\alpha}$ is a constant to be obtained such that equation (11) is satisfied for a special value α . Essentially, the critical value $C_{\rho(0,k),\alpha}$ indicates dependence structure of $Z_{it}(k)$'s through cross-correlation matrix $\rho(0,k)$ defined on $\Gamma_{(0,k)}$ and α the probability of type I error.

Now, to obtain $C_{\rho(0,k),\alpha}$, I need the joint distribution of $Z_{it}(k)$'s. Since, the underlying data come from autocorrelated process; it is difficult to determine the same analytically. In view of this, one can evaluate $C_{\rho(0,k),\alpha}$ using simulation technique on the lines similar to Kalgonda and Kulkarni (2004) for $VAR_p(1)$ process as follows.

- Generate large number (say 10000) of vectors from \underline{Y}_t , following multivariate normal distribution with zero means and covariance matrix $\rho(0,k)$.
- Compute $Z_t(k) = \text{Max}_{1 \leq i \leq p} |Z_{it}(k)|$ for each of these vectors.
- The $(1 - \alpha)$ th percentile of sample $Z_t(k)$ is the estimate of $C_{\rho(0,k),\alpha}$.



Z-chart

Based on equation (11), the observed process \underline{Y}_t will be in control if $Z_t(k) < C_{\rho(0,k),\alpha}$. On the other hand, if $Z_t(k) > C_{\rho(0,k),\alpha}$, the process will be out of control. A multivariate control chart can be proposed to monitor the mean vector of $VAR_p(k)$ process, in the way similar to Z chart.

1. Set control limits:

$$LCL=0 \text{ and } UCL=C_{\rho(0,k),\alpha}$$

2. Use charting statistic $Z_t(k)$.
3. If value of $Z_t(k)$ falls within control limits, the multivariate $VAR_p(k)$ process is deemed to be in control. Otherwise, the process is considered to be out of control and it is taken as an indication that there is change in some of the μ_i 's.

In particular, if $Z_{it}(k) > C_{\rho(0,k),\alpha}$ the variable y_i is responsible for the out of control situation. In addition, a chart supplying information about the behavior of the individual characteristics can be displayed.

Example

To illustrate the use of the proposed procedure and its diagnostic ability, I consider a bivariate vector $\underline{Y}_t = (Y_{1t}, Y_{2t})'$ following $VAR_2(2)$ model as specified in equation (1). The target mean vector $\underline{\mu}_0$, autocorrelation matrices Φ_1 and Φ_2 with error covariance matrix Σ as:

$$\underline{\mu}_0 = (0,0)' \quad \Phi_1 = \text{diag}(\alpha_1, \beta_1)'$$

$$\Phi_2 = \text{diag}(\alpha_2, \beta_2)'$$

And
$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}.$$

First, I shall represent the $VAR_2(2)$ model into $VAR_4(1)$ as:

$$\underline{X}_t = \Psi \underline{X}_{t-1} + \underline{e}_t$$

Where $\underline{X}_t = (\underline{Y}_t', \underline{Y}_{t-1}')$

$$\Psi = \begin{bmatrix} \Phi_1 & \Phi_2 \\ I & 0 \end{bmatrix} \text{ and } \underline{e}_t = (\underline{\varepsilon}_t', 0)'$$

The stationarity condition for the VAR(1) model is equivalent to the condition that all eigen values of Ψ , that is, all roots of $\det\{\lambda I - \Psi\} = 0$ be less than one in absolute value. This gives

$$\lambda^4 - \lambda^3(\alpha_1 + \beta) - \lambda^2(\alpha_2 + \beta_2 - \alpha_1\beta_1) + \lambda(\alpha_1\beta_2 + \alpha_2\beta_1) + \alpha_2\beta_2 = 0$$

For given values of α 's and β 's by trial and error find the value of λ . If its absolute value is less than one, stationarity will be achieved for the specified values of α 's and β 's. In particular, for analytical purpose after verification of stationarity condition, I take

$$\Phi_1 = \text{diag}(0.5 \quad 0.5)'$$

$$\Phi_2 = \text{diag}(0.4 \quad 0.4)'$$

$$\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

This gives,

$$\Psi = \begin{bmatrix} 3.8961 & 1.94805 & 3.24675 & 1.62338 \\ 1.94805 & 3.8961 & 1.62338 & 3.24675 \\ 3.24675 & 1.62338 & 3.8961 & 1.94805 \\ 1.62338 & 3.24675 & 1.94805 & 3.8961 \end{bmatrix}$$

With usual notations,

$$\Gamma_{(0,k)} = \begin{pmatrix} 3.8961 & 1.94805 \\ 1.94805 & 3.8961 \end{pmatrix} \text{ And } \rho_{(0,k)} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

Three sets of process observations labeled as A, B, and C were simulated. First, in set A, I simulate five observation vectors from the in control situation. Further, in order to illustrate the usefulness of the proposed control chart when there is shift in the mean vector while cross-covariance structure remains the same, I have generated two more sets, labeled as B and C of five samples each as shown in Table 1. The values of the proposed statistic are computed and listed in Table 2. Further, Z chart for each of these samples are plotted with $LCL=0$ and $UCL=C_{\rho(0,k),\alpha}=3.20$.

Table 1. Shift in value of $\underline{\mu}$

Change in element of $\underline{\mu}$	A	B	C
μ_1^0	No change	$\mu_2^0 + 1$	$\mu_1^0 - 1$
μ_2^0	No change	No change	No change

Table 2. Values of $Z_t(k)$ statistics.

Sample set	Observation No.	$Z_t(k)$	$Z_{1,t}(k)$	$Z_{2,t}(k)$
A	1	0.559		
	2	1.069		
	3	1.084		
	4	1.236		
	5	2.878		
B	1	0.101		
	2	1.855		
	3	3.299*	1.131	3.299
	4	4.285	1.341	4.285
	5	3.211		
C	1	1.319		
	2	2.385		
	3	3.371*	-3.371	-0.918
	4	1.409		
	5	1.875		

Results and discussion

- From sample A, one can see that Z values are well below UCL, indicating in control status correctly.
- Sample B and C are out of control due to shift in the mean values.
- For out of control observations in B and C, I consider the values of $Z_{1,t}(k)$, $Z_{2,t}(k)$ and compare with 3.20. The i -th value is considered as out of control if it falls above 3.20. The observation number 3 and 4 of sample B is out of control due to shift in second variable. Further, in sample D observation number 3 is out of control due to shift in first variable.
- These conclusions agree with the fact.

Conclusion

In this study, I extend the control procedure with interpretations of out of control signals where the process data assumes VAR(k) process model. The finite Intersection Test (FIT) and existing control procedure for VAR(1) processes is the primary tool used in this.

References

1. Alt, F.B. 1985. Multivariate control charts. *Encyclopedia Stat. Sci.* 6: 110-122.
2. Alwani, L.C. and Roberts, H.V. 1988. Time-Series modeling for statistical process control. *J. Business Economic Stat.* 6(1): 87-95.
3. Crosier, R.B. 1988. Multivariate generalization of cumulative sum quality control schemes. *Technometrics.* 30: 291-303.
4. Davison, A.C. and Hinkley, D.V. 1997. Bootstrap methods and their applications, Cambridge University Press.
5. Hayter, A.J. and Tsui, K.L. 1994. Identification and quantification in multivariate quality control problems. *J. Quality Technol.* 26: 197-208.
6. Hotelling, H. 1947. Multivariate quality control in techniques of statistical analysis. Edited by Eisenhart, Hastay, and Wallis. McGraw-Hill, New York.
7. Kalgonda, A.A. and Kulkarni, S.R. 2004. Multivariate control charts for autocorrelated processes. *J. Appl. Stat.* 31(3): 317-328.
8. Kramer, H.G. and Schmid, W. 1997. EWMA charts for multivariate time series. *Sequential Analysis.* 16(2): 131-154.