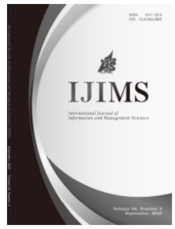
http://ijims.ms.tku.edu.tw/main.php

International Journal of
Information
and
Management
Sciences

International Journal of
Information and Management Sciences

36 (2025), 19-34. DOI:10.6186/IJIMS.202503_36(1).0002



Applying a Simple Test Procedure to Recognize the Quality Variables Responsible for Variance Shifts in a Multivariate Normal Process

Chia-Ding Hou, Rung-Hung Su and Chien-Cheng Sun*
Department of Statistics and Information Science, Fu Jen Catholic University, Taiwan

Keywords

Multivariate control chart, statistical process control, variance shifts.

Abstract

Recognizing the contributors of a multivariate process disturbance is a critical research issue and has recently drawn a great deal of attention. To recognize the quality variables of process faults in multivariate processes, most current studies use machine learning techniques or decomposition approaches. Additionally, most research has attempted to identify the sources of process mean shifts. As opposed to most of the current research, this study proposes to use a simple outlier testing procedure to determine the quality fault variables responsible for the process variance shifts. In this paper, we first introduce a statistical test for identifying outlying variances in a multivariate normal distribution. Then an iterative test method is employed to determine the contributors to process variance shifts. As demonstrated by simulation results and a practical illustrative example, the proposed method is effective and easy for recognizing the quality variables responsible for variance shifts in a multivariate normal process.

1. Introduction

To improve the underlying process, process personnel has always desired to find faults in real time. For several decades, statistical process control (SPC) charts have been successfully applied to monitor process faults. With technological advancements allowing for the use of more and more sophisticated sensors, the process of monitoring multiple quality characteristics has become increasingly popular. The multivariate SPC chart is commonly used for monitoring process faults. Nevertheless, the out-of-control signal simply reveals that process defects have occurred at hand. It is difficult to determine the source(s) of a triggered multivariate SPC signal because of the features of multiple quality variables. Consequently, identifying the contributors of process faults has become a critical issue in multivariate SPC, resulting in the rapid growth of related research.

In the literature, machine learning methods have been suggested as a way to determine the source of process shifts. For example, Cheng and Cheng (2008) used neural networks and support

*corresponding author. Email: 141637@mail.fju.edu.tw

vector machines to identify the source of variance shifts in the multivariate process. Brian Hwarng and Wang (2010) proposed a neural network-based identifier for multivariate autocorrelated processes. Shao et al. (2010) considered that the flexible discriminant analysis using multivariate adaptive regression splines (MARS) and BRUTO features can effectively identify the sources of variance shifts. Shao and Hou (2013) modeled a hybrid artificial neural network for fault identification of a stochastic multivariate process. Shao et al. (2016) employed two computational intelligence approaches, artificial neural networks (ANN) and MARS, to classify the sources of variance shifts in a multivariate normal process. Shao and Lin (2019) presented a time delay neural network (TDNN) classifier to diagnose the quality variables that cause out-of-control signals for a multivariate normal process. Zheng and Yu (2019) proposed a hybrid system that composes support vector machines (SVM) and convolutional neural networks (CNN) techniques. Zhang et al. (2021) proposed a hybrid deep learning model that integrates a one-dimensional convolutional neural network (1-DCNN) and stacked denoising auto-encoders (SDAE) to extract high-level features from complex process signals. Güler et al. (2024) developed a hybrid independent components analysis-support vector machines method to pinpoint the sources of mean shifts in both multivariate normal and non-normal processes. Also, decomposition statistics were applied to identify the contributors to process faults. For example, Runger et al. (1996) proposed the use of different metric distances to decide which variables have shifted. Mason et al. (1997) developed a cause-selecting procedure using the decomposition of the T^2 statistic. Vives-Mestres et al. (2016) considered two distinct methodologies for signals interpretation of T^2 control chart for large and small dimensional compositional data. Kim et al. (2016) proposed an adaptive step-down procedure using conditional T^2 statistic for fault variable identification. Piña Monarrez (2018) applied Hotelling's T^2 decomposition method to the R-chart. Özdemir Güler and Bakır (2022) used independent component analysis to detect and identify the mean shift. Haq and Khoo (2022) proposed an adaptive multivariate EWMA charts based on variable sample size and variable sampling interval techniques to identify the sources of a mean shift in the multivariate normal process. Ahsan et al. (2024) developed a T^2 based PCA mix control chart to determine the source of process shifts.

Although the machine learning approach may provide possible solutions, it may require a large amount of data and huge computing resources to perform better. In addition, there is no standard theory to guide process personnel in choosing the right model, as it has several parameters that need to be tuned through trial and error. Therefore, different personnel may get different results, and the method may be difficult to be adopted by less skilled personnel. While the statistical decomposition methods may offer some solutions, the mathematical difficulty in applying them may limit their application. Therefore, in contrast to the current research studies that use machine learning or decomposition approaches to recognize the sources of process faults, this study seeks to develop a simple test method that can effectively find the quality variables responsible for process variance shifts.

The remainder of the paper is organized as follows. The next section presents the proposed testing procedure for recognizing the quality variables responsible for variance shifts in a multivariate normal process. Section 3 provides the results of simulations, demonstrating the effectiveness of the introduced procedure. In section 4, a real-life example is provided to illustrate the proposed procedure. The study is concluded in the final section.

2. The proposed method

Let

$$\boldsymbol{X}_i = [X_{i1}, X_{i2}, ..., X_{ip}]', \quad i = 1, 2, ..., n \tag{2.1}$$

represent the p characteristics of the i^{th} observation with a multivariate normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = [\mu_1, \mu_2, ..., \mu_p]'$, $\boldsymbol{\Sigma} = Cov(\boldsymbol{X}_i) = [\sigma_{st}]_{p\times p}$, and $\sigma_{st} = \text{cov}(X_{is}, X_{it})$. Suppose that $\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_n$ is a random sample taken from the multivariate normal distribution mentioned above. The sample means vector and covariance matrix are therefore

$$\bar{\boldsymbol{X}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i = (\bar{X}_1, \bar{X}_2, ..., \bar{X}_p)' \tag{2.2}$$

and

$$\boldsymbol{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})' = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1p} \\ S_{21} & S_{22} & \cdots & S_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ S_{p1} & S_{p2} & \cdots & S_{pp} \end{bmatrix}, \tag{2.3}$$

respectively. There have been some multivariate control charts proposed to monitor a variance shift in a multivariate process. An example is Alt (1985) who proposed to use the sample generalized variance $|\boldsymbol{S}|$, and the following control limits:

$$\begin{aligned} UCL &= \left|\Sigma_0\right|(b_1 + 3\sqrt{b_2}) \\ LCL &= \max(0, \left|\Sigma_0\right|(b_1 - 3\sqrt{b_2})), \end{aligned} \tag{2.4}$$

where UCL and LCL represent the upper and lower control limit, $\left|\Sigma_0\right|$ is the determinant of the in-control covariance matrix, and

$$\begin{aligned} b_1 &= \frac{1}{(n-1)^p}\prod_{i=1}^{p}(n-i), \\ b_2 &= \frac{1}{(n-1)^{2p}}\prod_{i=1}^{p}(n-i)\left(\prod_{i=1}^{p}(n-i+2) - \prod_{i=1}^{p}(n-i)\right). \end{aligned} \tag{2.5}$$

Generally, an unbiased estimator can be used to estimate the parameter $\left|\Sigma_0\right|$ when it is unknown. As an out-of-control signal is triggered in such a multivariate control chart, it can be challenging to determine the cause-assignable variables. In the next subsection, we develop a test for detecting outlying variance of the multivariate normal distribution. Then we propose an iterative test approach for identifying the quality variables responsible for process variance shifts.

2.1 A test for the identification of outlying variance

Note that $\boldsymbol{X}_i$ has the distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, therefore X_{ij} has a normal distribution $N(\mu_j, \sigma_{jj})$. For testing the null hypothesis $H_0^{(j)}: \sigma_{jj} = \sigma_{jj}^{(0)}$, we can use the test statistic

$$T_j = \frac{(n-1)S_{jj}}{\sigma_{jj}^{(0)}}, \tag{2.6}$$

which has a chi-squared distribution with $(n-1)$ degrees of freedom. There might be many circumstances in which the hypothesis $H_0: \sigma_{11} = \sigma_{11}^{(0)}, \sigma_{22} = \sigma_{22}^{(0)}, \ldots, \sigma_{pp} = \sigma_{pp}^{(0)}$ needs to be tested, which denotes a prespecified variance vector. For identifying positive outlying variances in the multivariate normal distribution, it may be appropriate to apply the largest ordered statistic

$$\max_j T_j. \tag{2.7}$$

Hence, at significance level α, the test for such an alternative would be to reject the null hypothesis as

$$\max_j T_j > c, \tag{2.8}$$

where c satisfies

$$P\left[\max_j T_j > c \middle| H_0\right] = \alpha. \tag{2.9}$$

Because $T_1, T_2, \ldots, T_p$ are correlated, the distribution of $\max_j T_j$ has no analytical form. As a result, determining the critical value is not an easy task. In order to approximate critical value c, we can use Boole's inequality and obtain

$$\begin{aligned} \alpha &= P\left[\max_j T_j > c \middle| H_0\right] \\ &= P\left[T_1 > c \quad or \quad T_2 > c \ldots \quad or \quad T_p > c \middle| H_0\right] \\ &\le \textstyle\sum_{j=1}^{p} P\left[T_j > c \middle| H_0\right] \\ &= \textstyle\sum_{j=1}^{p} (1 - P\left[T_j \le c \middle| H_0\right]) \\ &= p(1 - F_{\chi^2(n-1)}(c)) \end{aligned} \tag{2.10}$$

where $F_{\chi^2(n-1)}(\cdot)$ is the cumulative distribution function of chi-squared distribution with $n-1$ degrees of freedom. As a result, we have

$$F_{\chi^2(n-1)}(c) \le 1 - \tfrac{\alpha}{p}. \tag{2.11}$$

Accordingly,

$$c \leq \chi^2_{1-\frac{\alpha}{p}}(n-1), \tag{2.12}$$

where $\chi^2_{1-\frac{\alpha}{p}}(n-1)$ is the $100\times(1-\alpha/p)^{\text{th}}$ percentile of chi-squared distribution with $n-1$ degrees of freedom. Consequently, $\chi^2_{1-\frac{\alpha}{p}}(n-1)$ is an upper bound for c. Since the upper bound is easy to compute and provides a conservative result, it could be used as an approximation to the critical value in the above test.

Simulated experiments are performed to assess whether the introduced approximation is effective. To evaluate the accuracy of the approximation, under the null hypothesis, the Monte Carlo estimate for $P\left[\max_j T_j > \chi^2_{1-\frac{\alpha}{p}}(n-1)\right]$ is computed and compared with the nominal level. Let $\boldsymbol{D} = \text{diag}(\sqrt{\sigma^{(0)}_{11}}, \sqrt{\sigma^{(0)}_{22}}, \ldots, \sqrt{\sigma^{(0)}_{pp}})$ be a diagonal matrix, and $\boldsymbol{\mu}_0$ be the in-control mean vector. Since the random vector $\boldsymbol{X}_i$ can be transformed using $\boldsymbol{D}(\boldsymbol{X}_i - \boldsymbol{\mu}_0)$ so that, when the process is in control, the transformed data have a mean vector $\boldsymbol{0}$ and covariance matrix $[\sigma'_{st}]$, where $\sigma'_{ss} = 1$, $s = 1, 2, \ldots, p$; $\sigma'_{st} = \rho_{st}$, $\forall\ s \neq t$, and $\rho_{st} = \text{corr}(X_{is}, X_{it})$ is the correlation of X_{is} and X_{it}. As a consequence, without loss of generality, we assume that $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ are sampled from a normal distribution $N_p(\boldsymbol{0}, \boldsymbol{\Sigma}^*)$, where $\boldsymbol{\Sigma}^* = [\sigma^*_{st}]$, $\sigma^*_{ss} = 1$, $s = 1, 2, \ldots, p$; $\sigma^*_{st} = \rho$, $\forall\ s \neq t$. In this study, p has been evaluated at seven different values: 2, 3, 4, 5, 6, 7, 8. In addition, we conducted simulations with sample sizes of 5, 30, 50, 100, 200, and 300. As the covariance matrix is negative definite for $\rho \leq -0.2$ as $p \geq 6$, the density function does not exist, we can not generate random vectors from such a multivariate normal distribution. We therefore consider the case ρ: -0.1, 0.0, 0.3, 0.5, 0.7, 0.9. From each given multivariate normal population, we simulate 10,000 multivariate normal samples to determine the value of $\max_j T_j$. By calculating the percent of 10,000 $\max_j T_j$'s that are greater than $\chi^2_{1-\frac{\alpha}{p}}(n-1)$, we can estimate $P\left[\max_j T_j > \chi^2_{1-\frac{\alpha}{p}}(n-1)\right]$. Thus, a lower absolute deviation indicates better approximation performance.

For the various values of *p* and *n*, Fig. 1 plots the deviations against ρ values separately. According to Fig. 1, at a significance level of 0.05, about 70% and 50% of the absolute deviations are lower than 0.0064 and 0.0036. At a significance level of 0.01, approximately 90% and 70% of absolute deviations are lower than 0.0031 and 0.0016. Furthermore, 90% of the absolute deviations are less than 0.0005, and 70% are lower than 0.0003, if the significance level is 0.001. As a result, it may be adequate for most applications to use this approximation.

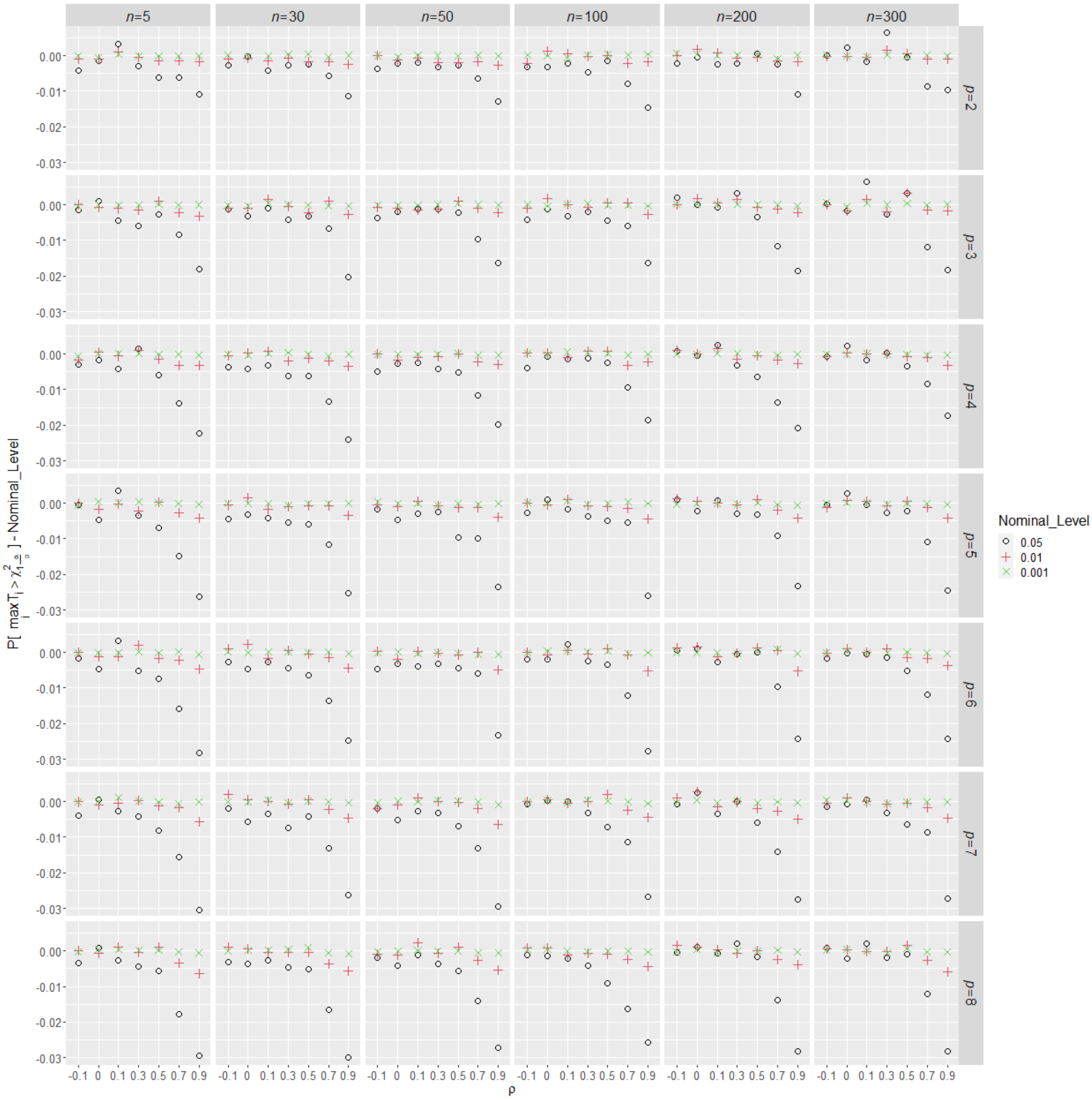


Fig. 1 *Plots of the Deviations Versus* ρ *Values for Different Values of n and p*

2.2 A method for recognizing the quality variables responsible for variance shifts

Using the above method and the following test, it will be able to identify the main contributors to the out-of-control signals:

Step I. Set t to 1;

Step II. Let $\alpha^* = \frac{\alpha}{t+1}$. (Type I error is maintained around the nominal level using the Bonferroni method and the error spending approach. See, Hou et al. (2001));

Step III. Utilize $\max_j T_j$ to test $H_0 : \sigma_{11} = \sigma_{11}^{(0)}, \sigma_{22} = \sigma_{22}^{(0)}, \ldots, \sigma_{pp} = \sigma_{pp}^{(0)}$ at the α^* significance level;

Step IV. If the hypothesis in **Step III** is rejected, remove the variable with the highest test value. To make it easier, consider excluding the pth variable. Return to **Step II** for further testing the rest variables by setting $t = t+1$ and $p = p-1$;

Step V. If the hypothesis in **Step III** is not rejected, terminate and declare that none of the rest variables are responsible for the variance shifts.

Iteratively continue to follow these steps until we are unable to reject the hypothesis for only some variables. Thus, these variables are therefore not the source of process variance shifts, while other variables are regarded as contributors.

3. Simulation Studies

To evaluate the usefulness of the method proposed above, we perform a series of simulation experiments. Without loss of generality, we assume that initially, a multivariate process is in control, and the observations are drawn from a multivariate normal distribution $N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, where

$$\boldsymbol{\mu}_0 = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{3.1}$$

and

$$\boldsymbol{\Sigma}_0 = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & \rho \\ \rho & \cdots & \rho & 1 \end{bmatrix}. \tag{3.2}$$

In addition, we assume that after some time, the covariance matrix changes from $\boldsymbol{\Sigma}_0$ to $\boldsymbol{\Sigma}_1$. Moreover, three out-of-control simulation scenarios are evaluated: (scenario 1) a variance shift occurs at the first quality characteristic; (scenario 2) variance shifts occur at the first two quality characteristics; and (scenario 3) variance shifts occur at the first three quality characteristics. In scenario 1, four possible values of p are considered in this study, namely, 2, 3, 5, and 7. The cases where p=5 and p=7 are examined only in scenarios 2 and 3 as illustrations. Following Cheng and Cheng (2008), the covariance matrixes for these three scenarios are:

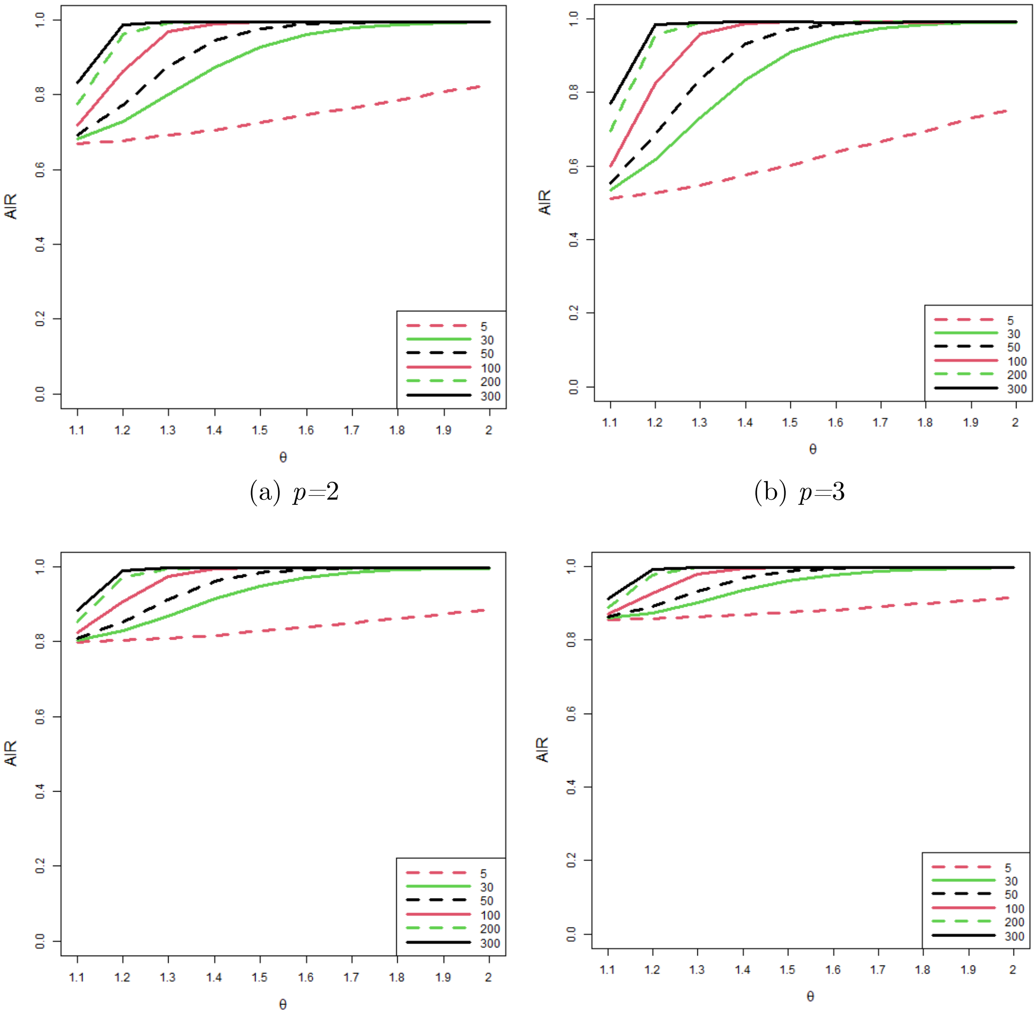
$$\begin{bmatrix} \theta^2 & \rho\theta & \cdots & \cdots & \rho\theta \\ \rho\theta & 1 & \rho & \cdots & \rho \\ \vdots & \rho & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 1 & \rho \\ \rho\theta & \rho & \cdots & \rho & 1 \end{bmatrix}, \text{(scenario 1)} \tag{3.3}$$

$$\begin{bmatrix} \theta^2 & \rho\theta^2 & \rho\theta & \rho\theta & \rho\theta \\ \rho\theta^2 & \theta^2 & \rho\theta & \rho\theta & \rho\theta \\ \rho\theta & \rho\theta & 1 & \rho & \rho \\ \rho\theta & \rho\theta & \rho & 1 & \rho \\ \rho\theta & \rho\theta & \rho & \rho & 1 \end{bmatrix}, \text{(scenario 2)} \tag{3.4}$$

and

$$\begin{bmatrix} \theta^2 & \rho\theta^2 & \rho\theta^2 & \rho\theta & \rho\theta \\ \rho\theta^2 & \theta^2 & \rho\theta^2 & \rho\theta & \rho\theta \\ \rho\theta^2 & \rho\theta^2 & \theta^2 & \rho\theta & \rho\theta \\ \rho\theta & \rho\theta & \rho\theta & 1 & \rho \\ \rho\theta & \rho\theta & \rho\theta & \rho & 1 \end{bmatrix}, \text{(scenario 3)} \tag{3.5}$$

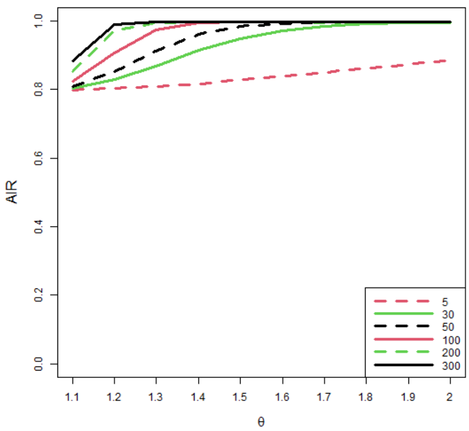
where θ is the inflated ratio. Ten values of θ are taken into account, these are 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, and 2.0. Seven values of ρ are considered, and they are -0.1. 0, 0.1, 0.3, 0.5, 0.7, and 0.9. A sample size of 5, 30, 50, 100, 200, and 300 will be taken. We used a significance level of 0.05. In this simulation experiment, we use the approximation presented above to determine the critical value of the proposed method. In addition, we use the accurate identification rate (AIR) to gauge the effectiveness of the introduced method. The AIR is defined as the percentage of variables classified correctly as noncontributors or contributors. This study conducted simulations under various parameter settings and found that under the same p, θ, n, and scenario, the AIRs corresponding to different correlation coefficients are almost the same. It is found that the correlation coefficient has little effect on the AIR of the proposed procedure. In all cases, the maximum difference in AIR among different correlation coefficients is only approximately 0.0085. Therefore, this study only gives an illustrative example when the correlation coefficient is 0.5. To explore how the proposed method performs when only one quality characteristic is out of control, we explore scenario 1. Fig. 2 presents the results. In order to examine the difference in the AIR performance of the proposed method in three different scenarios, we explore the cases where p=5 and p=7. Fig. 3 and 4 show the results.



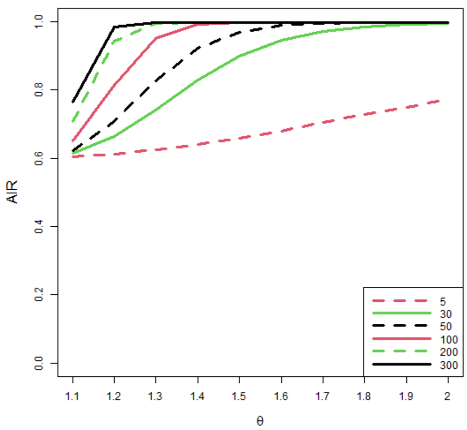
(a) p=2 (b) p=3

(c) p=5 (d) p=7

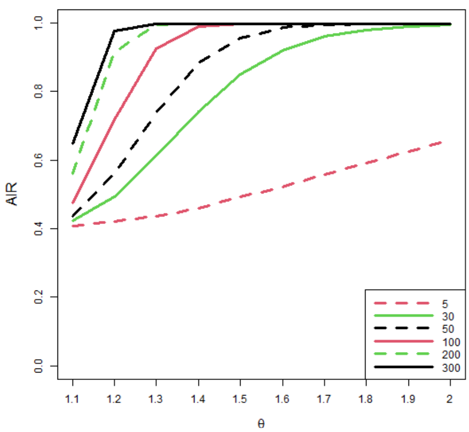
Fig. 2 *The Effect of* θ *on the AIR for Each Sample Size in the Situation of* $\rho = 0.5$ *with Various Values of p in Scenario 1*



(a) scenario 1

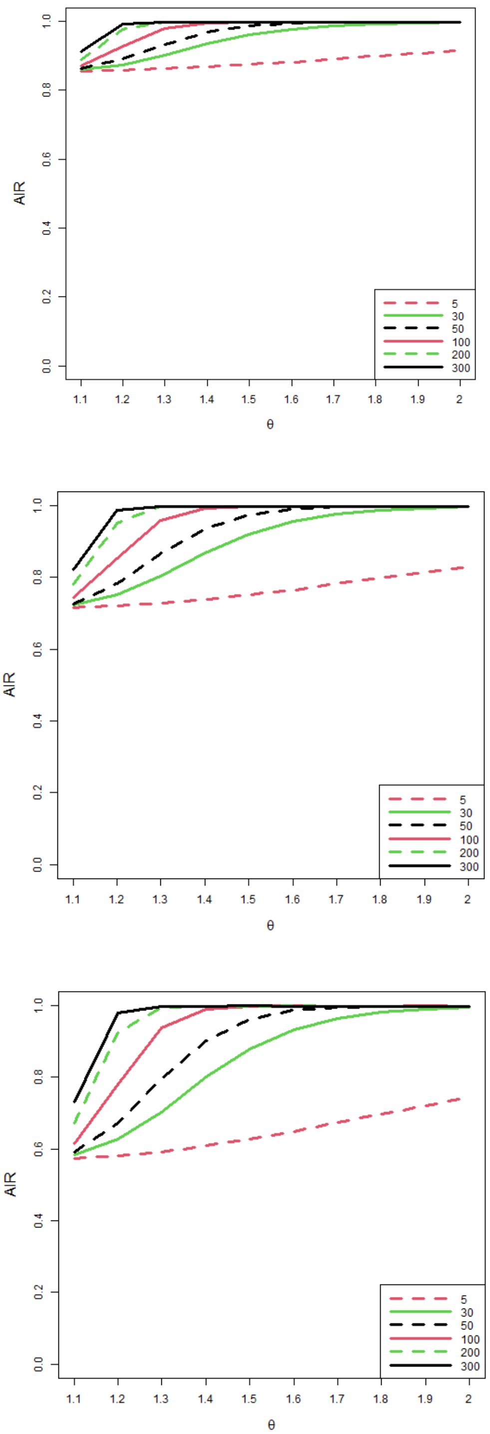


(b) scenario 2



(c) scenario 3

Fig. 3 *The Effect of* θ *on the AIR for Each Sample Size in the Situation of* $\rho = 0.5$ *and p=5 in Different Scenarios*



(a) scenario 1

(b) scenario 2

(c) scenario 3

Fig. 4 *The Effect of* θ *on the AIR for Each Sample Size in the Situation of* $\rho = 0.5$ *and p=7 in Different Scenarios*

Fig. 2 to 4 show the effects of inflated ratio on AIR for various sample sizes. With an increasing inflated ratio or sample size, it is visible that the AIR increases to 1. It is reasonable to expect this. Furthermore, from Fig.3 and 4, it can be seen that the larger the number of out-of-control variables, the larger the sample size required to achieve the same AIR. Hence, using adequate sample sizes, it is evident that the introduced approach can appropriately determine the contributors to process variance shifts. We also obtained similar results from our extensive simulation studies.

4. An illustrative Example

The proposed method is demonstrated by a practical example discussed in Joshi et al. (1997) and Huwang et al. (2007) regarding wafer production in the semiconductor industry. Three critical dimension measurements of the die were taken at three different positions on each drawn wafer from the 74 lots of wafers. For the convenience of explanation, let $\boldsymbol{M}_1$, $\boldsymbol{M}_2$, and $\boldsymbol{M}_3$ represent the three corresponding measurements. According to Alt (1985), the in-control mean vector is estimated as

$$\boldsymbol{\mu}_0 = \begin{bmatrix} 3.135 \\ 3.108 \\ 3.118 \end{bmatrix}$$

In addition, the in-control and out-of-control covariance matrixes used in Alt (1985) are

$$\boldsymbol{\Sigma}_0 = \begin{bmatrix} 0.0093 & 0.0036 & 0.0052 \\ 0.0036 & 0.0085 & 0.0034 \\ 0.0052 & 0.0034 & 0.0088 \end{bmatrix}$$

and

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 0.0186 & 0.0036 & 0.0052 \\ 0.0036 & 0.0170 & 0.0034 \\ 0.0052 & 0.0034 & 0.0088 \end{bmatrix},$$

respectively. Obviously, the measurements $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ are the contributors to variance shifts. Suppose the control chart triggers an out-of-control signal. As for now, the method proposed is applicable to determine the quality variables responsible for variance shifts.

As a convenience, assuming $n = 10$ and based on the distribution $N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_1)$ with $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_1$ mentioned above, an illustrative data set is simulated. Ten observations simulated are (3.2526, 3.1912, 3.2530), (3.1748, 3.1878, 3.2579), (3.0463, 3.1178, 2.9760), (2.8801, 2.7871, 2.9114), (2.9323, 2.9726, 3.1057), (3.0450, 3.4545, 2.9752), (3.3363, 3.0914, 3.0594), (3.2756, 3.1184, 3.1308), (3.2014, 3.0022, 3.1059), (3.0274, 3.0657, 3.2116). Using a significance level of 0.05 and applying the method introduced above, the contributors of variance shifts can be determined. This analysis is summarized in Table 1. As can be seen in Table 1, the test statistic $\max_j T_j$ is greater than 22.18 at the first iteration, therefore we reject the null hypothesis. As $\boldsymbol{M}_2$

has the highest test value, we assert that $\boldsymbol{M}_2$ contributes to the variance shifts. Removing the second variable, set $t = t+1 = 2$, and $p = p-1 = 2$. Similarly, at the second iteration, $\max_j T_j$ also remains above 22.18, and $\boldsymbol{M}_1$ corresponds to the highest test value. Thus we reject the null hypothesis and declare that $\boldsymbol{M}_1$ is the contributor. With the first variable excluded, and set $t = t+1 = 3$, $p = p-1 = 1$. At the third iteration, the test statistic drops to 14.77, which is lower than 21.03, so we do not reject the null hypothesis and stop the test procedure. We declare that the remaining measurement $\boldsymbol{M}_3$ is not the contributor to the variance shift. The results presented in this table demonstrate how the proposed method can easily and effectively detect the contributors of variance shifts. Although this study uses the semiconductor industry as an example to illustrate how to apply the proposed method, this method can also be applied to other industries or other fields, such as automobile production line data discussed by Porzio and Ragozini (2003), and healthcare data examined by Maboudou-Tchao and Diawara (2013).

Table 1 *Illustration of the Proposed Test Procedure* (α=0.05)

Iteration t	p	$\boldsymbol{T} = (T_1,\ldots,T_p)$	Test statistic $\max_j T_j$	Critical value $\chi^2_{1-\frac{\alpha^*}{p}}(n-1)$	Conclusion
	3	(22.64, 31.45, 14.77)	31.45	22.18	$\boldsymbol{M}_2$ is the contributor.
2	2	(22.64, 14.77)	22.64	22.18	$\boldsymbol{M}_1$ is the contributor.
3	1	(14.77)	14.77	21.03	$\boldsymbol{M}_3$ is not the contributor.

5. Conclusion

It is vital for the process industry to quickly and accurately pinpoint the contributors of an out-of-control process. While most established methods use machine learning techniques or decomposition approaches to recognize the sources of process shifts, the present work proposes to use a simple outlier testing procedure to determine the fault variables responsible for the process variance shifts. A method for detecting outlying variance in multivariate normal distributions is proposed. Moreover, an iterative test approach for recognizing the sources of process variance shifts is developed. Results of simulations and a practical illustrative example show that the approach proposed is easy to use and can yield satisfying results when identifying the quality variables responsible for variance shifts in multivariate normal processes.

Multiple testing involves conducting more than one hypothesis test simultaneously. The probability of incurring one or more type I errors (false positives) during multiple statistical tests is known as the family-wise error rate. This study uses the Bonferroni method and error spending approach to keep the overall family-wise error rate near our desired significance level, reducing the risk of false positives in multiple tests. Nonetheless, the Bonferroni correction is not the only technique for handling multiple tests. Other methods, such as the Sidak method (Sidák, 1967), the Holm-Bonferroni method (Holm, 1979), and the Benjamini-Hochberg method (Benjamini and Hochberg, 1995), may also be considered. Further research is needed to determine which method is superior.

In this study, Boole's inequality was used to devise an approximation for calculating the critical value of the proposed test. It is found that the approximation performed well based on our numerical results. For most applications, the approximation may suffice, but a sharper inequality could improve it. Further study is needed on this possibility. Besides, as there are other kinds of process shifts or multivariate processes, there is a need to investigate whether the same approach is applicable to them as well.

Acknowledgements

The authors would like to thank the editor and anonymous reviewers for their valuable and constructive comments, which have led to a significant improvement in the manuscript.

References

Ahsan, M., Mashuri, M., Prastyo, D. D., & Lee, M. H. (2024). Performance of T^2 based PCA mix control chart with KDE control limit for monitoring variable and attribute characteristics. *Scientifc Reports, 14*(1), 7372. https://doi.org/10.1038/s41598-024-58052-4

Alt, F. B. (1985). Multivariate Quality Control. The Encyclopedia of Statistical Sciences, Kotz S., Johnson, NL, Read CR.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple hypothesis testing. *Journal of the Royal Statistical Society: Series B, 57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Brian Hwarng, H., & Wang, Y. (2010). Shift detection and source identification in multivariate autocorrelated processes. *International Journal of Production Research, 48*(3), 835–859. https://doi.org/10.1080/00207540802431326

Cheng, C. S., & Cheng, H. P. (2008). Identifying the source of variance shifts in the multivariate process using neural networks and support vector machines. *Expert Systems with Applications, 35*(1-2), 198–206. https://doi.org/10.1016/j.eswa.2007.06.002

Güler, Z. O., Bakır, M. A., & Kardiyen, F. (2024). A novel hybrid ICA-SVM method for detection and identification of shift in multivariate processes. *Hacettepe Journal of Mathematics & Statistics, 53*(2), 556–576. https://doi.org/10.15672/hujms.1269072

Haq, A., & Khoo, M. B. C. (2022). An adaptive multivariate EWMA mean chart with variable sample sizes and/or variable sampling intervals. *Quality and Reliability Engineering International, 38*(7), 3322–3341. https://doi.org/10.1002/qre.3120

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65–70.

Hou, C. D., Chiang, J., & Tai, J. J. (2001). Identifying chromosomal fragile sites from a hierarchical-clustering point of view. *Biometrics, 57*(2), 435–440. https://doi.org/10.1111/j.0006-341X.2001.00435.x

Huwang, L., Yeh, A. B., & Wu, C. W. (2007). Monitoring multivariate process variability for individual observations. *Journal of Quality Technology, 39*(3), 258–278. https://doi.org/ 10.1080/ 00224065.2007.11917692

Joshi, M., & Sprague, K. (1997). Obtaining and using statistical process control limits in the semiconductor industry (Vol. 25). ASA.

Kim, J., Jeong, M. K., Elsayed, E. A., Al-Khalifa, K. N., & Hamouda, A. M. S. (2016). An adaptive step-down procedure for fault variable identification. *International Journal of Production Research, 54*(11), 3187–3200. https://doi.org/10.1080/00207543.2015.1076948

Maboudou-Tchao, E. M., & Diawara, N. (2013). A lasso chart for monitoring the covariance matrix. *Quality Technology & Quantitative Management, 10*(1), 95–114. https://doi.org/ 10.1080/16843703.2013.11673310

Mason, R. L., Tracy, N. D., & Young, J. C. (1997). A practical approach for interpreting multivariate T^2 control chart signals. *Journal of Quality Technology, 29*(4) 396–406. https://doi.org/ 10.1080/00224065.1997.11979791

Özdemir Güler, Z., & Bakır, M. A. (2022). Detection and identification of mean shift using independent component analysis in multivariate processes. *Journal of Statistical Computation and Simulation, 92*(9), 1920–1940. https://doi.org/10.1080/00949655.2021.2015352

Piña Monarrez, M. R. (2018). Generalization of the Hotelling's T^2 decomposition method to the R-chart. *International Journal of Industrial Engineering, 25*, 200–214. http://cathi.uacj.mx/20.500.11961/4216

Porzio, G. C., & Ragozini, G. (2003). Visually mining off-line data for quality improvement. *Quality and Reliability Engineering International, 19*(4), 273–283. https://doi.org/10.1002/qre.588

Runger, G. C., Alt, F. B., & Montgomery, D. C. (1996). Contributors to a multivariate statistical process control chart signal. *Communications in Statistics-Theory and Methods, 25*(10), 2203–2213. https://doi.org/10.1080/03610929608831832

Shao, Y. E., & Hou, C. D. (2013). Hybrid artificial neural networks modeling for faults identification of a stochastic multivariate process. In *Abstract and Applied Analysis, 2013*, Article ID 386757. https://doi.org/10.1155/2013/386757

Shao, Y. E., & Lin, S. C. (2019). Using a time delay neural network approach to diagnose the out-of-control signals for a multivariate normal process with variance shifts. *Mathematics, 7*(10), 959. https://doi.org/10.3390/math7100959

Shao, Y. E., Hou, C. D., Chou, Y. J., & Lin, S. C. (2016). Classifications of process variance faults using artificial neural networks and multivariate adaptive regression splines. *ICIC Express Letters Part B: Applications, 7*(2), 415–419.

Shao, Y. E., Huang, H. Y., & Chen, Y. J. (2010). Determining the sources of variance shifts in a multivariate process using flexible discriminant analysis. *ICIC Express Letters–An International Journal of Research and Surveys, 4*, 1573–1578.

Sidák, Z. K. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association, 62*(318), 626–633. https://doi.org/ 10.1080/01621459.1967.10482935

Vives-Mestres, M., Daunis-i-Estadella, J., & Martin-Fernandez, J. A. (2016). Signal interpretation in Hotelling's T^2 control chart for compositional data. *IIE Transactions, 48*(7), 661–672. https://doi.org/10.1080/0740817X.2015.1125042

Zhang, C., Yu, J., & Wang, S. (2021). Fault detection and recognition of multivariate process based on feature learning of one-dimensional convolutional neural network and stacked denoised autoencoder. *International Journal of Production Research, 59*(8), 2426–2449. https://doi.org/10.108 0/00207543.2020.1733701

Zheng, X., & Yu, J. (2019). Multivariate process monitoring and fault identification using convolutional neural networks. *In Proceeding of the 24th International Conference on Industrial Engineering and Engineering Management*, Singapore, 197–208. https://doi.org/10.1007/978-981-13-3402-3_22

Chia-Ding Hou
Department of Statistics and Information Science, Fu Jen Catholic University, Taiwan
E-mail address: 046136@mail.fju.edu.tw
Major area(s): Statistical process control, meta analysis, statistical genetics

Rung-Hung Su
Department of Statistics and Information Science, Fu Jen Catholic University, Taiwan
E-mail address: 141637@mail.fju.edu.tw
Major area(s): Inventory, multivariate analysis, statistical process control

Chien-Cheng Sun
Department of Statistics and Information Science, Fu Jen Catholic University, Taiwan
E-mail address: jason0958232345@gmail.com
Major area(s): Statistical process control, biostatistics, applied statistics

(Received September 2023; accepted March 2025)