IIE Transactions (2013) **45**, 132–146

ISSN: 0740-817X print / 1545-8830 online
DOI: 10.1080/0740817X.2012.663952

Monitoring the covariance matrix via penalized likelihood estimation

BO LI[1], KAIBO WANG[2] and ARTHUR B. YEH[3,*]

[1] *Research Center for Contemporary Management, Key Research Institute of Humanities and Social Sciences at Universities, School of Economics and Management, Tsinghua University, Beijing 100084, People's Republic of China*
[2] *Department of Industrial Engineering, Tsinghua University, Beijing 100084, People's Republic of China*
[3] *Department of Applied Statistics and Operations Research, Bowling Green State University, Bowling Green, OH 43403, USA*
E-mail: byeh@bgsu.edu

Received July 2010 and accepted January 2012

In many industrial multivariate quality control applications, based on the engineering and operational understanding of how the process works, when the process variability is out of control it is typically the case that changes only occur in a small number of elements in the covariance matrix. Under such a premise, we propose a new Phase II Shewhart chart for monitoring changes in the covariance matrix of a multivariate normal process. The new control chart is essentially based on calculating the likelihood ratio of testing the hypothesis that the in-control covariance matrix is equal to a known covariance matrix, where the unknown covariance matrix that appears in the likelihood ratio is replaced by an estimate obtained from a penalized likelihood function. The penalized likelihood function is derived by adding an L_1 penalty function to the usual likelihood. The performance of the proposed chart is evaluated based on simulations and compared with that of several existing Shewhart charts for monitoring the covariance matrix. The simulation results indicate that the proposed chart outperforms existing charts. A real example from the semiconductor industry is presented and analyzed using the proposed chart and other existing charts. Potential future research directions are also discussed.

Keywords: Covariance matrix, L_1 penalty function, likelihood ratio test, penalized likelihood function, Phase II monitoring, sparsity

1. Introduction

1.1. *Literature review and motivation*

The methodological development of statistical control charts continues to flourish in the quality engineering and statistics literature. Over the last two decades, the focus has gradually shifted from univariate to multivariate control charts. Advances in modern data acquisition techniques and computing power mean that multivariate control charts will play an increasingly important role in improving the quality of products and services. As several authors have already pointed out (see, for example, Woodall and Montgomery (1999), Stoumbos *et al.* (2000), and Woodall (2000)), the multivariate control chart will be an important area of research in the 21st century.

Beginning with the seminal work by Hotelling (1947), an extensive literature has been created that considers the use of multivariate control charts to monitor the process mean vector. However, control charts to monitor process variability, as characterized by the covariance matrix of the multivariate random variable representing the correlated quality characteristics being monitored, only began to be developed in the early 1980s. Some early work includes Alt (1984), Alt and Bedewi (1986), Healy (1987), and Alt and Smith (1988). In this period, the representative control chart was based on the Generalized Likelihood Ratio (GLR) statistic for testing $H_0 : \mathbf{\Sigma} = \mathbf{\Sigma}_0$ versus $H_1 : \mathbf{\Sigma} \neq \mathbf{\Sigma}_0$. The GLR charting statistic takes the form

$$W = -(n-1)\left[p + \ln\frac{|\mathbf{S}|}{|\mathbf{\Sigma}_0|} - \mathrm{tr}\left(\mathbf{\Sigma}_0^{-1}\mathbf{S}\right)\right], \qquad (1)$$

where n is the sample size, p is the dimensionality, $\mathbf{S}$ is the sample covariance matrix of a given sample, $|\mathbf{A}|$ is the determinant of a matrix $\mathbf{A}$, and tr($\mathbf{B}$) is the trace of a matrix $\mathbf{B}$.

Between 1990 and 2005, control charts for monitoring the covariance matrix continued to appear in the literature. Yeh *et al.* (2006) gave a detailed account of some of the developments in this period. Since then, several authors have proposed new charting schemes for monitoring the covariance matrix. These include work by Yeh *et al.* (2005), Reynolds and Cho (2006), Williams *et al.* (2006), Huwang *et al.* (2007), Hawkins and Maboudou-Tchao (2008), and Yen and Shiau (2010).

*Corresponding author

All of the existing control charts for monitoring the covariance matrix consider two key issues:

1. Given individual or groups of observations, how can the process covariance matrix be efficiently estimated?
2. Given an estimate of the covariance matrix, how can one devise an effective control charting mechanism?

The consideration of the latter led to control charts based on GLR (Alt, 1984), a determinant (the $|S|$ chart; Alt and Bedewi, 1986), conditional entropy (Guerrero-Cusumano, 1995), and decomposition of the sample covariance matrix (Tang and Barnett, 1996a), among others. As for the consideration of the former first question, for subgroups of data, the sample covariance matrix is typically used to estimate the process covariance matrix, with Cumulative SUM (CUSUM) or Exponentially Weighted Moving Averages (EWMA)-based techniques being typically adopted for individual observations.

However, most of the existing charts do not take into account process knowledge gleaned from engineering and operational understanding of how a process works. It should be noted that from an engineering point of view, not all of the variables in a multivariate quality characteristic are correlated with one another. Instead, subsets of variables exist in which the variables within the same set are correlated, whereas variables between different sets are typically uncorrelated. Such a structure may not be reflected in the sample covariance matrix (or the CUSUM or EWMA version of it) since non-zero entries in places where there should be a zero may result from calculating the sample covariance matrix. Also, when the process covariance matrix changes, it is typically the case that only a smaller set of variables will be affected. For example, in an important process for silicon wafer manufacturing called *lapping*, several correlated variables are considered to be critical factors that determine the uniformity of the thickness of a wafer and therefore need to be closely monitored. Based on engineering knowledge of the process as well as some empirical evidence, these variables can be grouped into two sets of variables and, thus, when the covariances change this only affects some of the variables.

Similar to most of the existing literature on monitoring the covariance matrix, our focus in the current article is on Phase II monitoring. That is, we assume that the in-control covariance matrix is either known or can be estimated using Phase I data. Under such a premise, we can further assume, without loss of generality, that the in-control covariance matrix is equal to a $p \times p$ identity matrix, where p is the number of correlated variables to be monitored. For a process with a general covariance matrix, Hawkins and Maboudou-Tchao (2008) suggested transforming the general covariance matrix into an identity matrix by multiplying the original observations by a lower triangular matrix (inverse-Cholesky root). The same treatment was also used by Huwang *et al.* (2007). Therefore, if only a few of the variances or covariances change when the covariance matrix changes, this means that many of the diagonal elements will continue to have a value of one and off-diagonal elements zero for an out-of-control covariance matrix, especially for moderate to large p. Note that this is also the case for many of the out-of-control scenarios considered for the simulation studies conducted in the existing literature. The fundamental question is this: When only a small number of variances and covariances change in a covariance matrix, are there more efficient ways of estimating the covariance matrix? Consequently, can a more effective control charting mechanism be developed for monitoring the covariance matrix based on more efficient estimates of the covariance matrix? Our main goal in this research is to provide some solutions to these two questions.

1.2. *A brief overview of the proposed methodology*

Statistically speaking, it is less efficient to monitor the covariance matrix of a p-dimensional process that has $p(p+1)/2$ variance and covariance components if we already know that most of the off-diagonal elements are in fact zero. When certain structural information about the covariance matrix is known *a priori*, it is possible to take such information into consideration and obtain a more efficient estimate of the covariance matrix. Dempster (1972) pointed out that the covariance structure of a multivariate normal distribution can be simplified by forcing the off-diagonal elements of the inverse of the covariance matrix to zero. Huang *et al.* (2006) developed a data-driven method to identify parsimony in the covariance matrix. Based on the modified Cholesky decomposition proposed by Pourahmadi (1999, 2000), the authors proposed to add L_1 or L_2 penalties to the normal likelihood in order to shrink the off-diagonal elements to zero. This reparameterization treatment can guarantee that the resulting matrix is positive-definite. However, the decomposition method relies on a hierarchical structure among the variables. Decomposed matrices are obtained by regressing, in the order in which the variables are listed, the variables that appear later in the list on the variables that appear earlier in the list. In situations where the order of variables is not clearly defined, the results of this method become difficult to interpret.

Yuan and Lin (2007) proposed to add a Least Absolute Shrinkage and Selection Operator (LASSO)-based L_1 penalty to the likelihood function when estimating the inverse of the covariance matrix in Gaussian graphical models. This method also leads to a matrix estimate that is positive-definite. Rothman *et al.* (2008) proposed to minimize the same penalized likelihood function in a different context. By using the Cholesky decomposition rather than the modified Cholesky decomposition, this scheme enforces positive definiteness while eliminating the dependency of the resulting estimate on a specific ordering of the variables.

In light of the discussed methods that generate sparsity in the estimate of the inverse of the covariance matrix, this work proposes a likelihood ratio–based method for monitoring the process covariance matrix. Considering the practical issues discussed earlier, we propose to use a penalized likelihood function to force the unchanged covariance components to zero. The control charting mechanism is then based on the likelihood ratio for testing the hypothesis that $H_0 : \mathbf{\Sigma} = \mathbf{I}_p$ versus $H_1 : \mathbf{\Sigma} \neq \mathbf{I}_p$, where the $\mathbf{\Sigma}$ that appears in the likelihood ratio will be replaced by a penalized estimator of $\mathbf{\Sigma}$. It should be pointed out that a LASSO-type penalty function has been applied recently to construct the control charts for monitoring multivariate mean vector in the context of statistical process control as well as linear profile monitoring; see, for example, Wang and Jiang (2009), Zou and Qiu (2009), and Zou *et al.* (2010).

The rest of this article is organized as follows. In Section 2, we discuss in detail the estimation of the inverse of the covariance matrix via a penalized likelihood function. The proposed control chart is also introduced. Section 3 is devoted to the performance evaluation of the proposed chart as well as a comparison with several existing charts. The performance of a control chart is evaluated based on the Average Run Length (ARL), where the run length is defined as the number of samples taken before the first out-of-control signal shows up on a control chart. In Section 4, a real example from the semiconductor industry is discussed and analyzed using the proposed chart and several existing charts. Conclusions are drawn in Section 5, which also contains additional discussions on future research directions.

2. Sparse covariance matrix estimation and monitoring

Let $\mathbf{X} = (X_1, X_2, \ldots, X_p)^{\mathrm{T}}$ represent a p-dimensional quality characteristic that needs to be monitored for quality improvement purposes. We assume that $\mathbf{X}$ follows a p-dimensional normal distribution denoted by $N_p(\mu, \mathbf{\Sigma})$, where μ and Σ are the mean and the covariance matrix of $\mathbf{X}$, respectively. When the process is in control, we assume that $\mu = \mu_0$ and $\mathbf{\Sigma} = \mathbf{\Sigma}_0$, and μ_0 and $\mathbf{\Sigma}_0$ are known or can be estimated from Phase I data. Therefore, without loss of generality, we assume that $\mathbf{X}$ follows $N_p(\mathbf{0}, \mathbf{I}_p)$ when the process is in control, where $\mathbf{I}_p$ is a $p \times p$ identity matrix. When the process is out of control, the covariance matrix changes to a non-identity matrix generally denoted as Σ_{OC}. Our focus in the current article is to propose a new control chart for monitoring changes in the covariance matrix under the premise that the out-of-control $\mathbf{\Sigma}_{\mathrm{OC}}$ remains sparse in the sense that only a few diagonal elements are not equal to one and only a few off-diagonal elements are not equal to zero. Note that the control chart proposed in the current article is entirely data-driven in that no knowledge of the numbers and locations of non-zero off-diagonal and non-unit diagonal elements is required.

We propose a two-step control charting mechanism. In the first step, a penalized likelihood function is used to obtain an estimate of $\mathbf{\Omega} = \mathbf{\Sigma}^{-1}$, the inverse of the covariance matrix, sometimes referred to as the *concentration* or *precision matrix*. The penalty function is chosen such that the resulting estimate of $\mathbf{\Omega}$ is sparse in the sense that off-diagonal elements in the estimate close to zero will be set to zero to ensure sparsity in the estimated $\mathbf{\Omega}$. Once an estimate of $\mathbf{\Omega}$ is obtained, in the second step the charting statistic is then calculated based on the negative log-likelihood ratio of testing $H_0 : \mathbf{\Sigma} = \mathbf{I}_p$ versus $H_1 : \mathbf{\Sigma} \neq \mathbf{I}_p$.

2.1. *The penalized likelihood function*

Let $\mathbf{\Sigma}$ be the true but unknown covariance matrix. Given a sample of size n, $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$, one can write the negative of log-likelihood as, up to a constant:

$$l(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n; \mathbf{\Omega}) = \mathrm{tr}(\mathbf{\Omega} S) - \ln|\mathbf{\Omega}|, \tag{2}$$

where

$$\mathbf{S} = \frac{1}{n} \sum_{j=1}^{n} (\mathbf{X}_j - \overline{\mathbf{X}})(\mathbf{X}_j - \overline{\mathbf{X}})^{\mathrm{T}}$$

is the sample covariance matrix and $\overline{\mathbf{X}} = 1/n \sum_{j=1}^{n} \mathbf{X}_j$. Traditionally, the Maximum Likelihood Estimator (MLE) of $\mathbf{\Omega}$ can be obtained by minimizing Equation (2). That is,

$$\mathbf{\Omega}_{\mathrm{MLE}} = \arg\min_{\mathbf{\Omega}} \{\mathrm{tr}(\mathbf{\Omega}\mathbf{S}) - \ln|\mathbf{\Omega}|\}. \tag{3}$$

It is easy to see that $\mathbf{\Omega}_{\mathrm{MLE}}$ is in fact $n\mathbf{S}^{-1}/(n-1)$.

It is known that a zero off-diagonal element in $\mathbf{\Omega}$ indicates conditional independence (D'Aspremont *et al.*, 2008) between the two variables given all other variables. The conditional independence among variables is often encountered in engineering applications. In the semiconductor manufacturing process we studied (introduced in more detail in later sections), we found that conditional independence exists among many variables, especially if the variables spread across multiple production stages. Moreover, conditional independence (or partial correlation analysis) is helpful in identifying root causes in an engineering system. For example, if X_1 and X_2 are highly correlated but are conditionally independent given X_3, we should not attribute quality deterioration in X_1 to X_2 but should investigate X_3 instead. In such a problem, the conditional independence plays an important role. Therefore, in this article, we focus on the sparsity in $\mathbf{\Omega}$, rather than in $\mathbf{\Sigma}$.

Under the premise discussed in the Introduction, the non-zero off-diagonal elements in $\mathbf{\Omega}$ indicate changes in the covariances of the corresponding variables. However, since $\mathbf{S}^{-1}$ is calculated from sample observations, it is highly likely that the true zero elements in $\mathbf{\Omega}$ are estimated to be non-zero. Therefore, we propose to utilize an estimation method capable of producing an estimate of $\mathbf{\Omega}$ that is sparse,

essentially by forcing elements in the estimated $\mathbf{\Omega}$ with small values to zero.

In the context of variable selection and estimation of a sparse matrix, it can be accomplished by adding a penalty function to the minimization of Equation (2). For the former, the L_1 penalty function has been shown to be capable of removing insignificant variables (Tibshirani, 1996). As for estimating a sparse matrix, incorporating an L_1 penalty in Equation (2) can help shrink estimated elements with small quantities to zero, thereby inducing sparsity in the estimated matrix. Rothman *et al.* (2008) added an L_1 penalty to Equation (2) on all of the off-diagonal elements of $\mathbf{\Omega}$. On the other hand, D'Aspremont *et al.* (2008) and Friedman *et al.* (2008) investigated the use of the L_1 penalty on all elements of $\mathbf{\Omega}$.

As for estimating $\mathbf{\Omega}$ to construct the proposed control chart, we found from conducting our trail simulations that penalizing all elements of $\mathbf{\Omega}$ produced estimates of $\mathbf{\Omega}$ that resulted in a more effective control charting mechanism. Therefore, in this work, we propose to add an L_1 penalty function aimed at penalizing all elements of $\mathbf{\Omega}$. Specifically, given a sample $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$, the penalized negative likelihood function, which is a modified version of what can be found in Rothman *et al.* (2008), can be written as

$$l(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n; \mathbf{\Omega}) = \text{tr}(\mathbf{\Omega} S) - \ln|\mathbf{\Omega}| + \lambda \|\mathbf{\Omega}\|_1, \quad (4)$$

where $||\mathbf{A}||_1 = \sum_{j=1}^{p} \sum_{i=1}^{p} |a_{ij}|$ for $\mathbf{A} = [a_{ij}]_{p\times p}$, and λ is a data-dependent tuning parameter that can be tuned to achieve different levels of sparsity of the $\mathbf{\Omega}$ estimate thus obtained.

Let $\mathbf{\Omega}_\lambda$ be the solution to the objective function in Equation (4) for a given λ; that is,

$$\mathbf{\Omega}_\lambda = \arg\min_{\mathbf{\Omega}>0} \{l(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n; \mathbf{\Omega})\}. \quad (5)$$

Friedman *et al.* (2008) provided algorithms for solving the objective function (4) under the constraint that the solution $\mathbf{\Omega}_\lambda$ is a symmetric positive-definite matrix. It is expected that, by properly tuning λ, one can obtain a more efficient estimate of $\mathbf{\Omega}$ such that if the process is in control, most of the off-diagonal elements are estimated to be zero and the diagonal elements stay unchanged. If the process goes out of control, the off-diagonal non-zero elements and changed diagonal elements in the resulting estimate of $\mathbf{\Omega}$ can truly reflect the changed relationships among the variables.

2.2. *The penalized likelihood ratio control chart*

Having obtained $\mathbf{\Omega}_\lambda$ in Equation (5), we now propose a new control chart for monitoring the covariance matrix. The charting statistic is based on the likelihood ratio for testing the hypothesis $H_0 : \mathbf{\Sigma} = \mathbf{I}_p$ versus $H_1 : \mathbf{\Sigma} \neq \mathbf{I}_p$. The corresponding negative log-likelihood ratio can be written as, up to a constant:

$$\Lambda = \text{tr}(\mathbf{S}) - \text{tr}(\mathbf{\Omega S}) + \ln|\mathbf{\Omega}|. \quad (6)$$

If $\mathbf{\Omega}$ is replaced by the usual MLE, $n\mathbf{S}^{-1}/(n-1)$, the Λ in Equation (6) is equivalent, up to a constant, to the statistic expressed in Equation (1). As it is convenient to detect changes in a covariance matrix by examining the likelihood ratio, the statistic W in Equation (1) and other modified versions of it have been used by many authors to develop control charts for monitoring the covariance matrix (see, for example, Yeh *et al.* (2004) and Hawkins and Maboudou-Tchao (2008)). In this work, we propose to substitute $\mathbf{\Omega}$ in Equation (6) with the estimate $\mathbf{\Omega}_\lambda$ obtained in Equation (5). Specifically, the proposed chart calculates, for each given sample of size n,

$$\Lambda_\lambda = \text{tr}(\mathbf{S}) - \text{tr}(\mathbf{\Omega}_\lambda \mathbf{S}) + \ln|\Omega_\lambda|, \quad (7)$$

where $\mathbf{\Omega}_\lambda$ is obtained from Equation (5) for a given λ. It should be noted that while the negative log-likelihood in Equation (6) is always positive, there is no guarantee that Λ_λ in Equation (7) is always positive since $\mathbf{\Omega}$ in Equation (6) is not replaced by the MLE but by the $\mathbf{\Omega}_\lambda$ obtained in Equation (5). Since the proposed chart is derived from a penalized likelihood ratio, it will be called the Penalized Likelihood Ratio (PLR) chart. The PLR chart signals an out-of-control signal when $\Lambda_\lambda > \text{UCL}_\lambda$, where UCL_λ is chosen, for a given λ, to achieve certain desirable in-control ARL. Note that Λ_λ is intrinsically different from the charting statistics used by Hawkins and Maboudou-Tchao (2008) and other authors in that the $\mathbf{\Omega}_\lambda$ used in calculating Λ_λ is obtained by minimizing a penalized likelihood function. In this study, all of the control limits were obtained using Monte Carlo simulations. The code is available from the second author upon request.

2.3. *The sparsity in* $\mathbf{\Omega}_\lambda$ *and* $\mathbf{S}$

The idea behind the use of $\mathbf{\Omega}_\lambda$ in the likelihood ratio is that since $\mathbf{\Omega}_\lambda$ is obtained based on a penalized likelihood function aimed at forcing small off-diagonal elements to zero, the resulting $\mathbf{\Omega}_\lambda$, unlike the $\mathbf{S}^{-1}$, should produce a sparsity closer to $\mathbf{\Sigma}^{-1}$. To assess the preservation of sparsity in $\mathbf{\Omega}_\lambda$ and in $\mathbf{S}$, we first generated a sample of 50 observations from $N_p(\mathbf{0}, \mathbf{I}_p)$ for $p = 5$. The sample was then used to calculate the sampled covariance matrix $\mathbf{S}$, $\mathbf{\Omega}_\lambda$, and $\mathbf{\Omega}_\lambda^{-1}$, which is an estimate of the covariance matrix, for $\lambda = 0.05$ and 0.2. The resulting $\mathbf{S}$, $\mathbf{\Omega}_\lambda$, and $\mathbf{\Omega}_\lambda^{-1}$ are shown in Fig. 1. It can be seen from Fig. 1 that $\mathbf{S}$ contains no zero elements even though many of the off-diagonal elements have values close to zero. The $\mathbf{\Omega}_\lambda$, on the other hand, does contain zero entries and by choosing $\lambda = 0.2$, $\mathbf{\Omega}_\lambda$ is almost diagonal as expected, except $\text{Cov}(X_3, X_5)$ which was estimated to be 0.01. The sparsity in the covariance matrix—i.e., $\mathbf{I}_5$ in this case—is also preserved in $\mathbf{\Omega}_\lambda^{-1}$, particularly when $\lambda = 0.2$. The results also indicate that increasing λ makes $\mathbf{\Omega}_\lambda$ more sparse. This is intuitive since a larger λ value gives more weight to the penalty function, thus forcing more elements

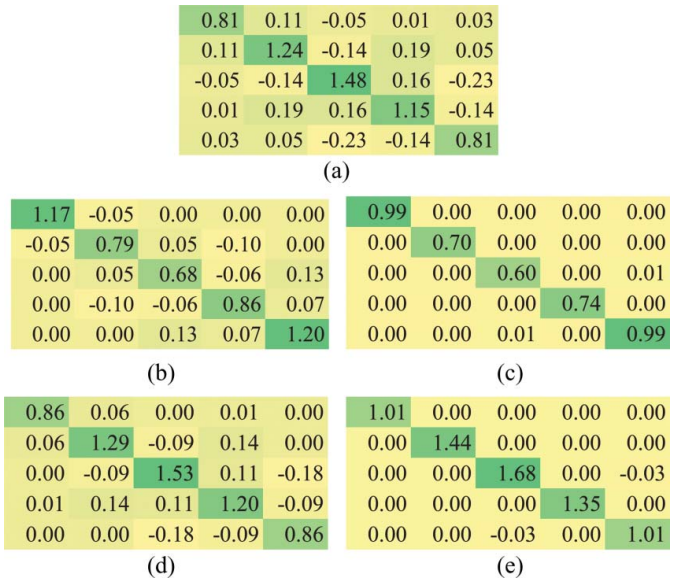

Fig. 1. Comparison of estimated $\mathbf{S}$, $\mathbf{\Omega}_\lambda$, and $\mathbf{\Omega}_\lambda^{-1}$ ($\mathbf{\Sigma} = \mathbf{\Sigma}_0 = \mathbf{I}$): (a) $\mathbf{S}$; (b) $\lambda = 0.05$, $\mathbf{\Omega}_\lambda$; (c) $\lambda = 0.2$, $\mathbf{\Omega}_\lambda$; (d) $\lambda = 0.05$, $\mathbf{\Omega}_\lambda^{-1}$; and (e) $\lambda = 0.2$, $\mathbf{\Omega}_\lambda^{-1}$ (color figure provided online).

with small values to zero. In fact, letting $\lambda = 0.3$ in this example resulted in a perfectly diagonal $\mathbf{\Omega}_\lambda$.

Now suppose that the covariance matrix has changed to

$$\mathbf{\Sigma}_{\mathrm{a}} = \begin{pmatrix} 1 & 0.5 & 0.5 & 0 & 0 \\ 0.5 & 1 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.5 \\ 0 & 0 & 0 & 0.5 & 1 \end{pmatrix}.$$

Again, we generated a sample of 50 observations from $N_5(\mathbf{0}, \mathbf{\Sigma}_{\mathrm{a}})$ and used the sample to calculate $\mathbf{S}$, $\mathbf{\Omega}_\lambda$, and $\mathbf{\Omega}_\lambda^{-1}$. The results are shown in Fig. 2. It can be seen from Fig. 2 that by letting $\lambda = 0.05$, some elements in both $\mathbf{\Omega}_\lambda$ and $\mathbf{\Omega}_\lambda^{-1}$ are approaching zero. By setting $\lambda = 0.2$, $\mathbf{\Omega}_\lambda^{-1}$ successfully generates two blocks in $\mathbf{\Omega}_\lambda$ and $\mathbf{\Omega}_\lambda^{-1}$ and preserves the sparsity of the out-of-control covariance matrix. However, $\mathbf{S}$ still returns non-zero estimates to all entries.

Note that, in general, it is difficult to find a direct connection between the sparsity in $\mathbf{\Omega}_\lambda$ and the sparsity in $\mathbf{\Omega}_\lambda^{-1}$. However, such a connection may be established in certain special cases such as block-wise diagonal matrices. That is, for any given matrix arranged in sub-blocks of matrices, its inverse matrix can be expressed as

$$\begin{bmatrix} \mathbf{A} & 0 \\ 0 & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} & 0 \\ 0 & \mathbf{D}^{-1} \end{bmatrix}.$$

If the resulting $\mathbf{\Omega}_\lambda$ has numerous zero entries, one can rearrange the variables, equivalent to rearranging the columns and rows, so that the matrix thus obtained, say $\mathbf{\Omega}_\lambda^{'}$, is block-wise (one needs to keep track of the original variables). The inverse of $\mathbf{\Omega}_\lambda^{'}$ is also block-wise under the new ordering of variables. This can help in trying to pinpoint which of the

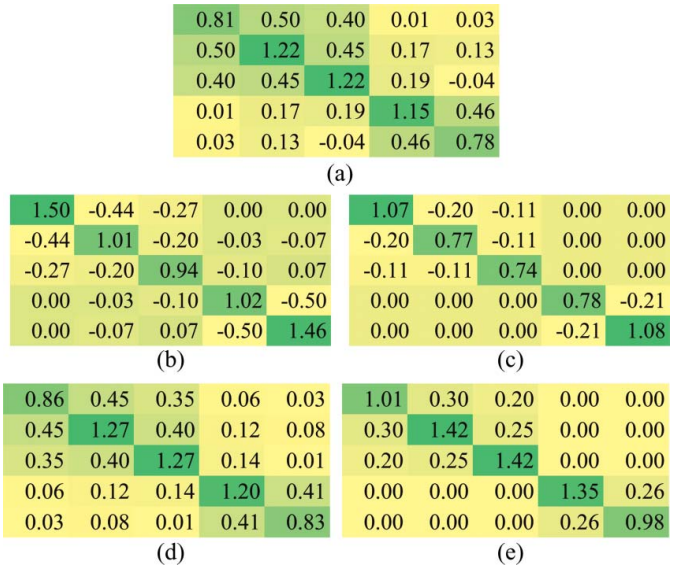

Fig. 2. Comparison of estimated $\mathbf{S}$, $\mathbf{\Omega}_\lambda$, and $\mathbf{\Omega}_\lambda^{-1}$ ($\mathbf{\Sigma} = \mathbf{\Sigma}_a$): (a) $\mathbf{S}$; (b) $\lambda = 0.05$, $\mathbf{\Omega}_\lambda$; (c) $\lambda = 0.2$, $\mathbf{\Omega}_\lambda$; (d) $\lambda = 0.05$, $\mathbf{\Omega}_\lambda^{-1}$; and (e) $\lambda = 0.2$, $\mathbf{\Omega}_\lambda^{-1}$ (color figure provided online).

variances and covariances have changed when the PLR chart signals an out-of-control sample.

3. Performance evaluation and comparison

In this section, we evaluate the performance of the proposed PLR chart in terms of ARL and compare it to that of several existing Shewhart charts. We restrict our attention to existing Shewhart charts since the PLR chart is a Shewhart chart. It should be pointed out that the PLR chart can be readily extended to CUSUM or EWMA procedures by defining some CUSUM or EWMA charting statistics based on Λ_λ. Such extensions, however, will not be investigated in the current article.

The existing Shewhart charts selected for comparison include the traditional likelihood ratio–based chart by Alt (1984; referred to as the LR chart hereafter); the control chart based on the conditional entropy by Guerrero-Cusumano (1995; referred to as the CE chart in the future), which is essentially the sum of the logarithms of the component sample variances; and the control chart by Tang and Barnett (1996a; hereafter referred to as the S_{d} chart), which is based on decomposing $\mathbf{S}$ into a sequence of independent chi-square random variables. The LR chart is chosen since it is capable of detecting general changes in the covariance matrix and is the basis for a number of other existing charts, including the proposed PLR chart. The CE chart is chosen because it is especially effective in detecting changes that only occur in the variance components. We did not include another commonly used chart, the $|S|$ chart, in our study since it was shown in Tang and Barnett (1996b) that the $|S|$ chart is less effective than the S_{d} chart.

3.1. *The simulation settings*

We assume that when the process is in control, $\mathbf{\Sigma} = \mathbf{I}_p$, and when the process is out of control, $\mathbf{\Sigma} = \mathbf{\Sigma}_{\mathrm{OC}} \neq \mathbf{I}_p$ is some specified covariance matrix. We also assume that when the covariance matrix changes from $\mathbf{I}_p$ to $\mathbf{\Sigma}_{\mathrm{OC}}$, such a change is sustained throughout the duration of the monitoring. The sample size n was set at 50 and once a sample was generated, it was used to evaluate all four competing charts. Furthermore, all of the charts were calibrated to have an in-control ARL (ARL_0) value approximately equal to 200. The dimensionality $p = 5$ and 10 were chosen and λ values between 0.05 and 2 were tested. The simulation size was 10 000.

While it is almost impossible to include all possible out-of-control scenarios in the simulation through the selection of $\mathbf{\Sigma}_{\mathrm{OC}}$, we tried to select a number of $\mathbf{\Sigma}_{\mathrm{OC}}$ values for which the PLR chart is expected to be effective, as well as other $\mathbf{\Sigma}_{\mathrm{OC}}$ values for which the existing charts are expected to be effective. We looked at seven specific $\mathbf{\Sigma}_{\mathrm{OC}}$ values. For $p = 10$ and $\delta = 0.01$ to 0.50 (similar for $p = 5$), the seven $\mathbf{\Sigma}_{\mathrm{OC}}$ values, denoted by $\mathbf{\Sigma}_{\mathrm{OC}_j}$, $j = 1, 2, \ldots, 7$, are

$$\mathbf{\Sigma}_{\mathrm{OC}_1} = (1+\delta) \times \mathbf{I}_{10} = \begin{pmatrix} 1+\delta & & & \\ & 1+\delta & & 0 \\ & & \ddots & \\ & 0 & & 1+\delta \end{pmatrix}_{10\times 10},$$

$$\mathbf{\Sigma}_{\mathrm{OC}_2} = \begin{pmatrix} 1+\delta & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}_{10\times 10},$$

$$\mathbf{\Sigma}_{\mathrm{OC}_3} = \begin{pmatrix} 1 & \delta & \delta & \delta & \delta & & & & & \\ \delta & 1 & \delta & \delta & \delta & & & & & \\ \delta & \delta & 1 & \delta & \delta & & 0 & & & \\ \delta & \delta & \delta & 1 & \delta & & & & & \\ \delta & \delta & \delta & \delta & 1 & & & & & \\ & & & & & 1 & & & & \\ & & 0 & & & & \ddots & & & \\ & & & & & & & 1 & & \\ & & & & & & & & \ddots & \\ & & & & & & & & & 1 \end{pmatrix}_{10\times 10},$$

$$\mathbf{\Sigma}_{\mathrm{OC}_4} = \begin{pmatrix} 1+\delta & \delta & \delta & & & & \\ \delta & 1+\delta & \delta & & 0 & & \\ \delta & \delta & 1+\delta & & & & \\ & & & 1 & & & \\ & 0 & & & \ddots & & \\ & & & & & 1 & \\ & & & & & & \ddots \\ & & & & & & & 1 \end{pmatrix}_{10\times 10},$$

$$\mathbf{\Sigma}_{\mathrm{OC}_5} = \begin{pmatrix} 1 & \delta & & & \\ \delta & 1 & & 0 & \\ & & 1 & & \\ & 0 & & \ddots & \\ & & & & 1 \end{pmatrix}_{10\times 10},$$

$$\mathbf{\Sigma}_{\mathrm{OC}_6} = \begin{pmatrix} 1+\delta^2 & \delta & & & \\ \delta & 1+\delta^2 & & 0 & \\ & & 1 & & \\ & 0 & & \ddots & \\ & & & & 1 \end{pmatrix}_{10\times 10},$$

$$\mathbf{\Sigma}_{\mathrm{OC}_7} = \begin{pmatrix} 1+\delta & \delta & & & & & \\ \delta & 1+\delta & & & 0 & & \\ & & 1 & & & & \\ & & & \ddots & & & \\ & 0 & & & 1 & & \\ & & & & & 1+\delta & \delta \\ & & & & & \delta & 1+\delta \end{pmatrix}_{10\times 10}.$$

Note that the existing charts are expected to be effective in detecting changes such as $\mathbf{\Sigma}_{\mathrm{OC}_1}$, and $\mathbf{\Sigma}_{\mathrm{OC}_2}$ since it involves variance changes only. The $\mathbf{\Sigma}_{\mathrm{OC}_2}$, $\mathbf{\Sigma}_{\mathrm{OC}_5}$, and $\mathbf{\Sigma}_{\mathrm{OC}_6}$ were also considered in Hawkins and Maboudou-Tchao (2008). The simulated out-of-control ARL (ARL_1) values for each of the $\mathbf{\Sigma}_{\mathrm{OC}_j}$ values are summarized in Tables 1 to 7. Here we only report the λ value that resulted in the best performing PLR chart for each of the $\mathbf{\Sigma}_{\mathrm{OC}_j}$ values considered. For a detailed account of the simulation results not shown here, please contact the second author.

3.2. *Discussion of simulation results*

As indicated by the simulated ARL_1 values summarized in Table 1, the CE chart has the best performance among the existing charts since it is designed to detect changes in the variance components. However, the PLR chart performs as well as the CE chart when a large tuning parameter was chosen. This forces more off-diagonal elements to zero and thus makes the PLR chart more sensitive to changes in variance components. For $\mathbf{\Sigma}_{\mathrm{OC}_2}$ (Table 2), this is an out-of-control covariance matrix that is difficult to detect (more so for $p = 10$ than for $p = 5$). Among the existing charts, the CE chart and the S_{d} chart are more effective than the very ineffective LR chart, with the CE chart outperforming the S_{d} chart for smaller δ values ($\delta \leq 0.20$) and the S_{d} chart outperforming the CE chart for larger δ values. The PLR chart, on the other hand, outperforms the existing charts, especially for smaller δ values.

For $\mathbf{\Sigma}_{\mathrm{OC}_3}$ (Table 3), the LR chart is expected to be more effective than the CE chart and the S_{d} chart since only

Table 1. The simulated ARL_1 values for $\mathbf{\Sigma}_{OC_1}$

	$p = 5$				$p = 10$			
δ	*LR chart*	*CE chart*	S_d *chart*	*PLR chart* ($\lambda = 1.8$)	*LR chart*	*CE chart*	S_d *chart*	*PLR chart* ($\lambda = 2.0$)
0.05	166.28	44.38	101.86	**43.93**	169.52	26.75	88.10	**26.59**
0.10	99.36	14.14	43.50	**14.06**	106.05	6.66	30.48	**6.62**
0.15	50.00	6.04	19.20	**5.99**	54.66	2.74	11.16	**2.72**
0.20	24.61	3.18	9.24	**3.14**	26.01	1.58	4.75	**1.57**
0.25	12.78	2.05	4.99	**2.02**	12.62	1.20	2.52	**1.19**
0.30	7.07	1.53	3.13	**1.52**	6.49	1.06	1.66	**1.06**
0.35	4.28	1.27	2.16	**1.26**	3.68	1.02	1.27	**1.02**
0.40	2.84	1.14	1.63	**1.13**	2.37	1.00	1.11	**1.00**
0.45	2.05	1.06	1.36	**1.06**	1.70	1.00	1.04	**1.00**
0.50	1.63	1.03	1.20	**1.03**	1.35	1.00	1.01	**1.00**

Note. Numbers in bold represent cases where the proposed PLR-chart is the best performing control chart; i.e., the chart with the smallest ARL.

the covariances change. The results summarized in Table 3 seem to confirm this expectation. The PLR chart slightly outperforms the LR chart, especially for $p = 10$ compared with $p = 5$. The choice of smaller λ value is expected since there are many non-zero entries in $\mathbf{\Sigma}_{OC_3}$. As for $\mathbf{\Sigma}_{OC_4}$ (Table 4), which is similar to $\mathbf{\Sigma}_{OC_3}$ but with added changes in the variance components, the changes in variances are likely to make the CE chart more effective. As can be seen in Table 4, the CE chart outperforms the other two existing charts, except when $\delta \geq 0.20$ and $p = 10$ where the S_d chart outperforms the CE chart. Again, the proposed PLR chart outperforms all the existing charts in detecting $\mathbf{\Sigma}_{OC_4}$, except when $\delta \geq 0.25$ and $p = 10$, where the S_d chart performs slightly better.

For $\mathbf{\Sigma}_{OC_5}$ (Table 5), which is similar to $\mathbf{\Sigma}_{OC_3}$, this is generally difficult for a control chart to detect (more so for $p = 10$ than for $p = 5$ and for $\mathbf{\Sigma}_{OC_5}$ than for $\mathbf{\Sigma}_{OC_3}$). Note that for $p = 5$, the $\mathbf{\Sigma}_{OC_5} = \mathbf{\Sigma}_{OC_3}$. Similar to the results summarized in Table 3, the PLR chart outperforms the best performing existing chart, which is the LR chart. As for $\mathbf{\Sigma}_{OC_6}$ (Table 6), the LR chart is most effective among the existing charts for smaller δ values ($\delta \leq 0.25$), while the S_d chart performs better for larger δ values. The PLR chart outperforms all of the existing charts in detecting $\mathbf{\Sigma}_{OC_6}$, except when $\delta \geq 0.45$ and $p = 10$. Nevertheless, in this case, the performance deteriorates as p increases for all of the charts. In detecting $\mathbf{\Sigma}_{OC_7}$ (Table 7), the PLR chart outperforms, as expected, all of the existing charts. Among the existing charts, the changes in variance components seem to make the CE chart more effective for smaller δ values ($\delta < 0.3$), while the S_d chart is more effective for larger δ values.

In summary, the proposed PLR chart either outperforms or performs as well as the best performing existing chart in all of the out-of-control scenarios examined. The PLR chart is particularly effective when the covariance changes occur in blocks in higher dimensions, such as $\mathbf{\Sigma}_{OC_3}$, $\mathbf{\Sigma}_{OC_4}$, and $\mathbf{\Sigma}_{OC_7}$, as well as when $\mathbf{\Sigma}_{OC}$ has many zero entries, such as $\mathbf{\Sigma}_{OC_2}$ under $p = 10$. Note that we have also conducted simulations for different sample sizes (not reported here). Increasing the sample size generally makes any given control chart more effective. Nevertheless, the comparisons between the proposed PLR chart and the existing charts remain the same.

Table 2. The simulated ARL_1 values for $\mathbf{\Sigma}_{OC_2}$

	$p = 5$				$p = 10$			
δ	*LR chart*	*CE chart*	S_d *chart*	*PLR chart* ($\lambda = 1.0$)	*LR chart*	*CE chart*	S_d *chart*	*PLR chart* ($\lambda = 0.8$)
0.05	193.14	142.79	170.86	**138.80**	195.99	159.29	177.57	**156.11**
0.10	171.63	105.97	134.91	**95.79**	188.25	128.06	148.38	**119.55**
0.15	143.73	81.05	97.99	**68.27**	174.02	104.03	118.15	**90.33**
0.20	113.67	63.13	69.98	**48.10**	156.71	86.78	88.44	**67.47**
0.25	85.70	50.13	50.17	**34.22**	137.46	72.88	66.00	**51.14**
0.30	64.89	40.04	35.64	**24.68**	116.82	62.05	49.39	**38.50**
0.35	47.89	32.79	25.55	**18.24**	98.96	53.43	37.07	**28.86**
0.40	35.75	27.78	18.50	**13.73**	82.07	46.59	27.77	**21.95**
0.45	36.64	23.01	13.68	**10.41**	67.34	40.96	21.29	**16.87**
0.50	20.36	19.53	10.69	**8.21**	55.15	36.14	16.29	**13.09**

Note. Numbers in bold represent cases where the proposed PLR-chart is the best performing control chart; i.e., the chart with the smallest ARL.

Table 3. The simulated ARL_1 values for $\mathbf{\Sigma}_{OC_3}$

	$p = 5$				$p = 10$			
δ	*LR chart*	*CE chart*	S_d *chart*	*PLR chart* ($\lambda = 0.05$)	*LR chart*	*CE chart*	S_d *chart*	*PLR chart* ($\lambda = 0.05$)
0.05	185.61	200.09	193.46	**183.27**	145.90	195.33	177.77	**138.93**
0.10	148.88	196.80	171.95	**146.27**	65.60	183.84	114.28	**57.25**
0.15	103.18	191.87	135.73	**101.81**	24.40	166.88	53.31	**20.16**
0.20	66.65	186.84	96.19	**64.99**	9.50	147.45	22.30	**7.78**
0.25	41.07	179.81	61.63	**39.23**	4.29	126.56	9.77	**3.64**
0.30	24.96	173.07	37.82	**24.17**	2.36	107.58	4.90	**2.05**
0.35	14.93	165.09	22.76	**14.49**	1.55	90.34	2.79	**1.44**
0.40	9.01	155.98	13.45	**8.75**	1.21	75.15	1.82	**1.16**
0.45	5.51	145.92	8.43	**5.40**	1.06	63.04	1.35	**1.05**
0.50	3.49	137.87	5.33	**3.46**	1.01	53.20	1.12	**1.01**

Note. Numbers in bold represent cases where the proposed PLR-chart is the best performing control chart; i.e., the chart with the smallest ARL.

As one referee pointed out, when a control chart signals, an important follow-up question is which of the variables is responsible for the observed out-of-control signal? Unfortunately, unlike the case of monitoring the multivariate process mean, when a covariance matrix monitoring chart, such as the proposed PLR chart, signals, it is very difficult to diagnose the root cause since there are a total of $p(p+1)/2$ parameters. Some recent papers have attempted to provide some diagnostic guidelines and procedures (see, for example, Yeh *et al.* (2006) and Chen and Hong (2010)). Under the premise of the current article, the covariance matrix is assumed to be sparse. Therefore, when the PLR chart signals, an important first step is to estimate, based on the sample that produces the out-of-control signal, the covariance matrix based on the penalized likelihood estimation via $\mathbf{\Omega}_\lambda$. Comparing this estimate with the one based on Phase I data or engineering knowledge could provide the first clues as to which variable may have caused the chart to signal. However, more concrete and systematic diagnostic procedures is an area for future investigations.

The types of changes in the variance components considered in the current article focus on detecting process deterioration, which is a major concern in quality control. However, as suggested by one referee, we also studied the performance of the PLR chart when $\delta < 0$, which indicates a decrease in the variance components with the implication that the process may actually improve. For each of the seven $\mathbf{\Sigma}_{OC_j}$ causes considered, we plotted the ARL_1 value against the values of δ. In terms of the overall trend of these ARL_1 curves, they can be classified into three groups, one group consisting of $\mathbf{\Sigma}_{OC_1}$ and $\mathbf{\Sigma}_{OC_2}$; one group consisting of $\mathbf{\Sigma}_{OC_3}$, $\mathbf{\Sigma}_{OC_5}$, and $\mathbf{\Sigma}_{OC_6}$; and the third group consisting of $\mathbf{\Sigma}_{OC_4}$ and $\mathbf{\Sigma}_{OC_7}$. Here we show in Fig. 3 only $\mathbf{\Sigma}_{OC_2}$, $\mathbf{\Sigma}_{OC_3}$, and $\mathbf{\Sigma}_{OC_4}$ from each of the three groups.

For $\mathbf{\Sigma}_{OC_2}$ (as well as $\mathbf{\Sigma}_{OC_1}$), which only incurs changes in variance, the PLR chart is clearly not effective in detecting the decreases in the variances. For $\mathbf{\Sigma}_{OC_3}$ (as well as $\mathbf{\Sigma}_{OC_5}$ and $\mathbf{\Sigma}_{OC_6}$), which only incurs changes in the covariance components, the ARL_1 curve is highly symmetric, indicating that the PLR chart is equally effective in detecting changes in the covariance in both directions. As for $\mathbf{\Sigma}_{OC_4}$ (as well as $\mathbf{\Sigma}_{OC_7}$), which incurs changes in both variance and covariance components, the ARL_1 curve presents

Table 4. The simulated ARL_1 values for $\mathbf{\Sigma}_{OC_4}$

	$p = 5$				$p = 10$			
δ	*LR chart*	*CE chart*	S_d *chart*	*PLR chart* ($\lambda = 1.0$)	*LR chart*	*CE chart*	S_d *chart*	*PLR chart* ($\lambda = 1.6$)
0.5	193.14	142.79	170.86	**138.80**	174.78	100.70	143.26	**98.51**
0.10	171.63	105.97	134.91	**97.59**	118.02	55.69	78.90	**51.17**
0.15	143.73	81.05	97.99	**68.27**	67.14	33.51	39.11	**28.50**
0.20	113.67	63.13	69.98	**48.10**	37.09	21.47	18.58	**16.73**
0.25	85.70	50.13	50.17	**34.22**	20.37	14.26	9.74	10.20
0.30	64.89	40.04	35.64	**24.68**	12.03	10.04	5.87	6.70
0.35	47.89	32.79	25.55	**18.24**	7.46	7.40	3.81	4.66
0.40	35.75	27.78	18.50	**13.73**	4.98	5.67	2.74	3.49
0.45	26.63	23.01	13.68	**10.41**	3.57	4.45	2.10	2.72
0.50	20.36	19.53	10.59	**8.21**	2.71	3.63	1.71	2.21

Note. Numbers in bold represent cases where the proposed PLR-chart is the best performing control chart; i.e., the chart with the smallest ARL.

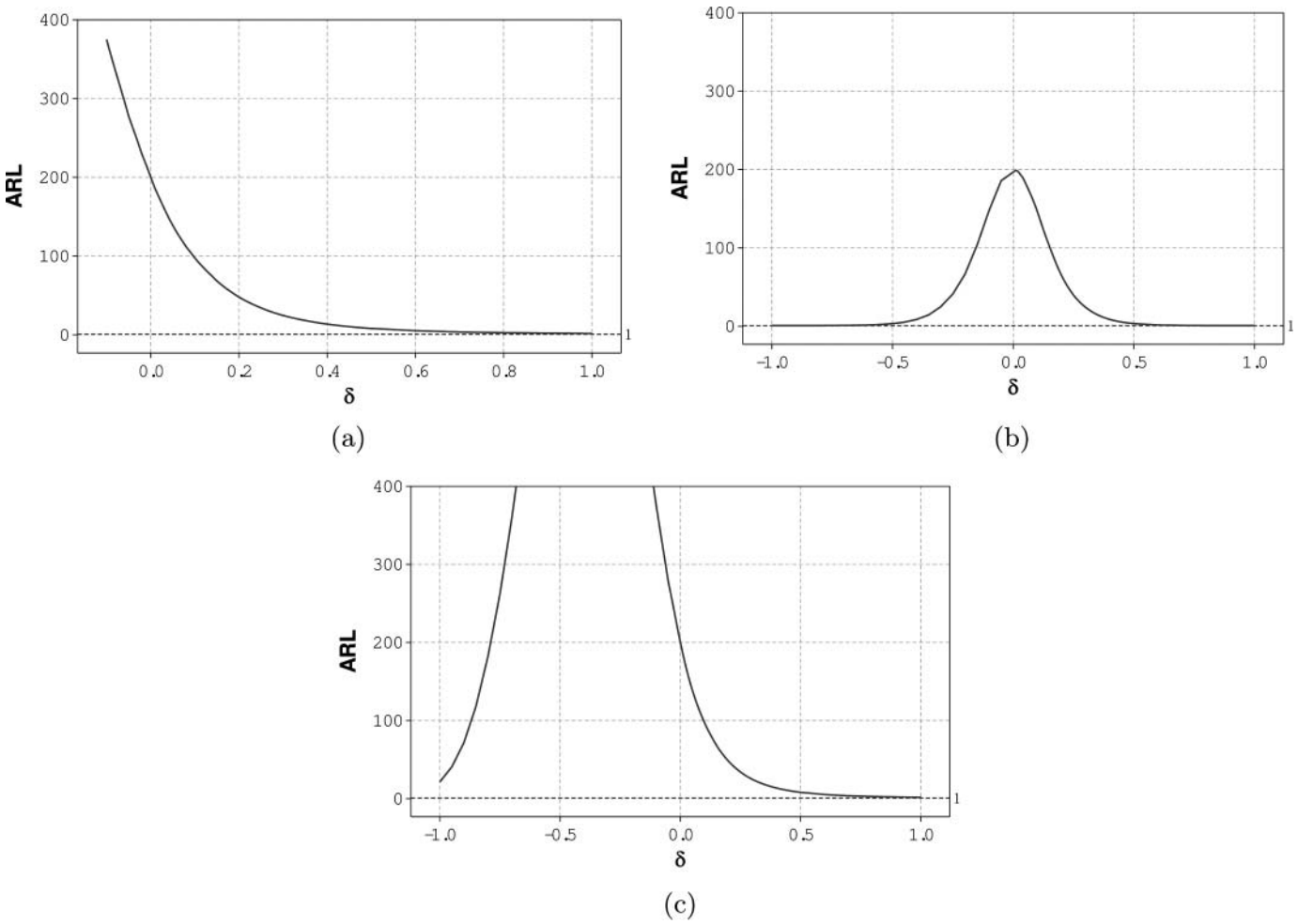

Fig. 3. The ARL_1 curves: (a) $\mathbf{\Sigma} = \mathbf{\Sigma}_{OC_2}$; (b) $\mathbf{\Sigma} = \mathbf{\Sigma}_{OC_3}$; and (c) $\mathbf{\Sigma} = \mathbf{\Sigma}_{OC_4}$.

mixed results. On one hand, a decrease in variance drives up the ARL_1, while on the other hand, a change in the covariance component decreases the ARL_1.

3.3. *Selection of* λ

The choice of λ for the PLR chart, similar to choosing the smoothing constant for an EWMA chart, depends on the change patterns and shift magnitudes. We have simulated the ARL_1 values as a function of δ for different values of $\lambda = 0.05, 0.10, 0.20, 1.0, 1.8$ for $\mathbf{\Sigma}_{OC_1}$, $\mathbf{\Sigma}_{OC_2}$, $\mathbf{\Sigma}_{OC_3}$, and $\mathbf{\Sigma}_{OC_6}$. The results are shown in Fig. 4 (using $\log(ARL_1)$ as the y-axis). Note that for the other OC covariance matrices, the patterns for $\mathbf{\Sigma}_{OC_4}$ and $\mathbf{\Sigma}_{OC_7}$ are similar to that of $\mathbf{\Sigma}_{OC_2}$, while the pattern for $\mathbf{\Sigma}_{OC_5}$ is similar to that of $\mathbf{\Sigma}_{OC_3}$.

The ARL_1 curves for $\mathbf{\Sigma}_{OC_1}$, $\mathbf{\Sigma}_{OC_2}$, and $\mathbf{\Sigma}_{OC_6}$ are basically convex and the ARL_1 generally decreases at a reasonable rate as δ increases. In addition, larger λ values seem to produce better results for $\mathbf{\Sigma}_{OC_1}$, while for $\mathbf{\Sigma}_{OC_2}$ a moderate λ value of 1.0 seems to produce the best results. As for $\mathbf{\Sigma}_{OC_6}$, a smaller value of $\lambda = 0.1$ seems to give better results. Note that in all these OC cases, $\mathbf{\Sigma}_{OC_1}$, $\mathbf{\Sigma}_{OC_2}$ (as well as $\mathbf{\Sigma}_{OC_4}$ and $\mathbf{\Sigma}_{OC_7}$), and $\mathbf{\Sigma}_{OC_6}$, the variances increase when δ increases. These OC scenarios are of greater concern since they generally represent process quality deterioration (as characterized by variance increases).

The ARL_1 curve for $\mathbf{\Sigma}_{OC_3}$ (as well as $\mathbf{\Sigma}_{OC_5}$, which is not shown here) is quite different in that the λ value needs to be more carefully chosen. This is because some of the choices actually produce concave ARL_1 curves that decrease at a much slower rate as δ increases. The graph seems to indicate that smaller λ values are preferable since they result in convex ARL_1 curves. Also note that $\mathbf{\Sigma}_{OC_3}$ and $\mathbf{\Sigma}_{OC_5}$ are the only two out-of-control scenarios in which only the covariances increase as δ increases, but the variances remain unchanged.

The results shown in Fig. 4 suggest that the choice of λ largely depends on the change patterns. It is therefore difficult to recommend one λ value that works well for all out-of-control scenarios. This also reaffirms the importance of failure pattern analysis in practical control chart applications for practitioners.

Table 5. The simulated ARL_1 values for $\mathbf{\Sigma}_{OC_5}$

	$p = 5$				$p = 10$			
δ	*LR chart*	*CE chart*	S_d *chart*	*PLR chart* ($\lambda = 0.5$)	*LR chart*	*CE chart*	S_d *chart*	*PLR chart* ($\lambda = 0.1$)
0.05	185.61	200.09	193.46	**183.27**	194.56	200.88	197.43	**191.97**
0.10	148.88	196.80	171.95	**146.27**	174.89	199.44	187.17	**170.83**
0.15	103.18	191.87	135.73	**101.81**	148.27	196.50	171.33	**141.56**
0.20	66.65	186.84	96.19	**64.99**	116.90	193.94	145.66	**109.03**
0.25	41.07	179.81	61.63	**39.73**	87.27	190.16	113.41	**79.43**
0.30	24.96	173.07	37.82	**24.17**	62.77	186.33	82.14	**55.43**
0.35	14.93	165.09	22.76	**14.49**	43.08	181.14	54.54	**37.38**
0.40	9.01	155.98	13.45	**8.75**	29.09	176.26	34.69	**24.37**
0.45	5.51	145.92	8.43	**5.40**	18.97	170.59	21.75	**15.97**
0.50	3.49	137.87	5.33	**3.46**	12.36	164.73	13.11	**10.37**

Note. Numbers in bold represent cases where the proposed PLR-chart is the best performing control chart; i.e., the chart with the smallest ARL.

4. A real example

In this section, we use a data set taken from the lapping process in semiconductor manufacturing to illustrate the real application of the proposed PLR chart and compare it with the existing charts discussed in previous sections.

It usually takes many steps to produce a wafer for electronic device fabrication. Starting from a single silicon ingot, wafers go through the processes of slicing, lapping, washing, chemical vapor deposition, polishing, etc., before the final products are packaged and shipped to downstream device manufacturers. Among these steps, lapping is a critical step in that it impacts the final wafer quality. Immediately following the slicing that cuts an ingot into wafers, the lapping stage is responsible for removing surface roughness and damage left by the slicing wire, controlling wafer tilt, and reducing thickness to customer specifications (see Chapter 3 of O'Mara *et al.* (2007)).

Values of the following quality variables, Total Thickness Variation (TTV), Total Indicator Reading (TIR), Site TIR (STIR), Bow, and Warp, are the major focus in wafer production. Variable values falling outside the specification limits will affect the yield of downstream device productions. Therefore, these variables are carefully monitored and controlled in production runs. Specifically, TTV measures the difference between the maximum and minimum thickness values of a wafer, TIR measures the minimum distance of two lines that are parallel to a reference plane while encompassing the complete curve of thickness, STIR is the TIR within a small region, Warp measures the difference between the maximum and minimum distances of the median surface from a reference plane, and Bow measures the deviation of the center point of the median surface to a reference plane. Figure 5 gives graphical depictions of the definitions of these variables.

From the physical definitions of these variables, TTV, TIR, and STIR measure different aspects of the thickness uniformity of a wafer, whereas Bow and Warp measure the global curvature of a wafer. Figure 6 shows scatterplots of the five variables. For a normal wafer (see Fig. 6(a)), there

Table 6. The simulated ARL_1 values for $\mathbf{\Sigma}_{OC_6}$

	$p = 5$				$p = 10$			
δ	*LR chart*	*CE chart*	S_d *chart*	*PLR chart* ($\lambda = 0.1$)	*LR chart*	*CE chart*	S_d *chart*	*PLR chart* ($\lambda = 0.1$)
0.05	185.44	192.27	191.71	**181.82**	194.61	196.00	196.04	**190.87**
0.10	148.68	170.53	163.01	**141.05**	174.88	180.96	181.41	**165.89**
0.15	103.41	142.06	119.76	**95.20**	148.27	159.63	156.51	**132.50**
0.20	66.25	112.27	77.59	**57.84**	117.55	134.56	123.08	**97.14**
0.25	40.99	84.60	44.54	**34.37**	88.57	109.85	86.40	**67.89**
0.30	25.36	60.87	25.45	**20.25**	64.15	87.84	54.48	**45.13**
0.35	15.94	43.61	14.38	**12.33**	45.19	67.23	32.49	**29.98**
0.40	10.24	30.22	8.92	**7.91**	31.88	50.93	19.70	**19.62**
0.45	6.89	21.19	5.81	**5.37**	22.29	38.67	12.16	13.21
0.50	4.85	14.68	3.99	**3.84**	15.67	29.20	7.88	9.05

Note. Numbers in bold represent cases where the proposed PLR-chart is the best performing control chart; i.e., the chart with the smallest ARL.

Table 7. The simulated ARL_1 values for $\mathbf{\Sigma}_{OC_7}$

	p = 5				*p = 10*			
δ	*LR chart*	*CE chart*	*S_d chart*	*PLR chart (λ = 1.0)*	*LR chart*	*CE chart*	*S_d chart*	*PLR chart (λ = 1.6)*
0.05	186.35	104.49	152.52	**101.56**	176.15	82.05	135.61	**79.62**
0.10	150.84	60.22	101.35	**54.30**	121.66	39.35	71.14	**35.76**
0.15	106.65	36.40	61.61	**30.04**	71.79	20.89	33.58	**17.89**
0.20	70.72	23.78	37.36	**17.89**	39.55	12.10	16.09	**9.64**
0.25	45.83	16.12	23.13	**11.22**	21.86	7.70	8.40	**5.72**
0.30	29.65	11.54	14.51	**7.41**	12.53	5.28	4.87	**3.81**
0.35	19.68	8.56	9.69	**5.30**	7.59	3.84	3.15	**2.73**
0.40	13.35	6.58	6.77	**3.89**	4.93	2.95	2.23	**2.08**
0.45	9.27	5.19	4.90	**2.48**	3.42	2.27	1.73	**1.69**
0.50	6.64	4.24	3.71	**2.39**	2.53	1.99	1.44	**1.44**

Note. Numbers in bold represent cases where the proposed PLR-chart is the best performing control chart; i.e., the chart with the smallest ARL.

are moderate correlations between TTV, TIR, and STIR and a weak correlation between Bow and Warp. Based on the physical understanding of these variables, it is also reasonable to conclude that there is no direct linkage between these two groups of variables. Figure 6(b) shows one possible scenario of process changes. It is seen that correlations among variables within each group increase significantly, whereas the correlations between the variables of the two

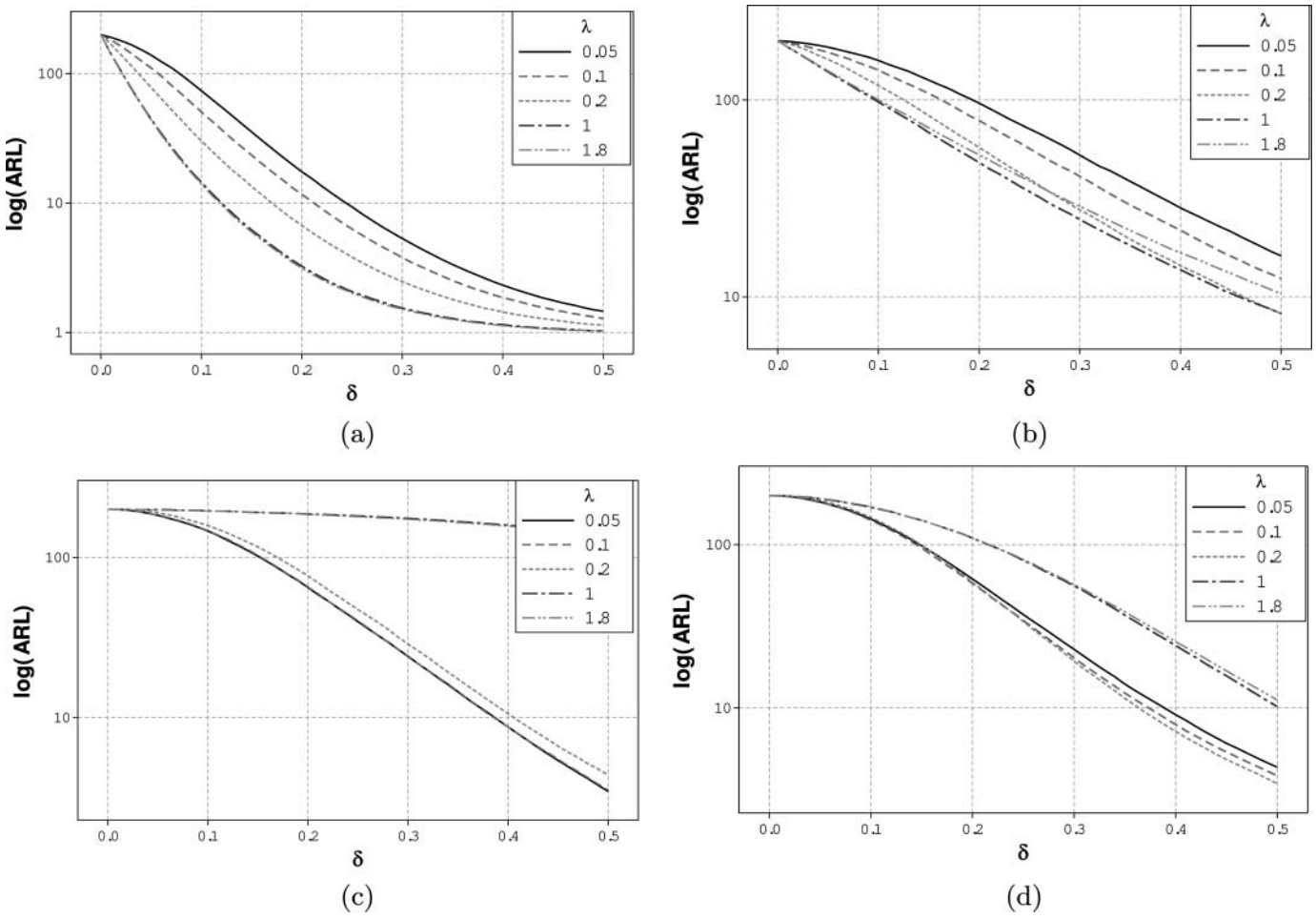

Fig. 4. The ARL_1 curves under different λ values: (a) $\mathbf{\Sigma} = \mathbf{\Sigma}_{OC_1}$; (b) $\mathbf{\Sigma} = \mathbf{\Sigma}_{OC_2}$; (c) $\mathbf{\Sigma} = \mathbf{\Sigma}_{OC_3}$; and (d) $\mathbf{\Sigma} = \mathbf{\Sigma}_{OC_4}$.

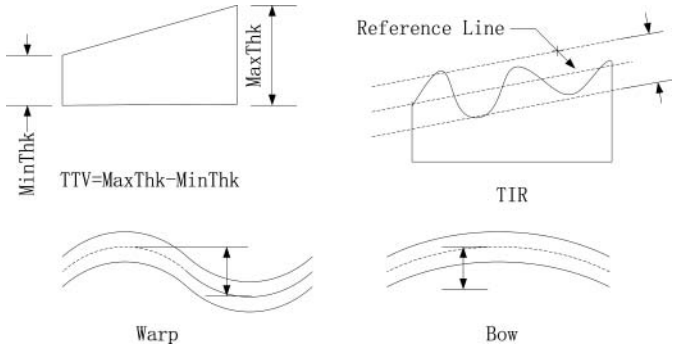

Fig. 5. Definitions of critical wafer quality variables.

the groups remain relatively unchanged. Due to data confidentiality, we have removed the actual scales from all of the figures shown in Fig. 6.

In order to obtain estimates of the in-control and out-of-control covariance matrices, we need to collect samples under both in-control and out-of-control processes. In Statistical Process Control (SPC) applications, Phase I analysis is required to identify samples that are collected when the process is in control (and when the process is out of control). As mentioned earlier, the focus of the current article is on Phase II monitoring. Although the SPC methodologies can help distinguish the in-control and out-of-control processes, the final conclusion about the process status has to be determined in conjunction with engineering knowledge. Therefore, in this example, we collected real data, investigated this data set using statistical methods, and discussed the results with the engineers at the plant. Statistical tools were used to analyze failure patterns in the data set. Furthermore, we identified one data set that the engineers believed was normal and one dataset that the engineers believed was collected when the process was in an abnormal state.

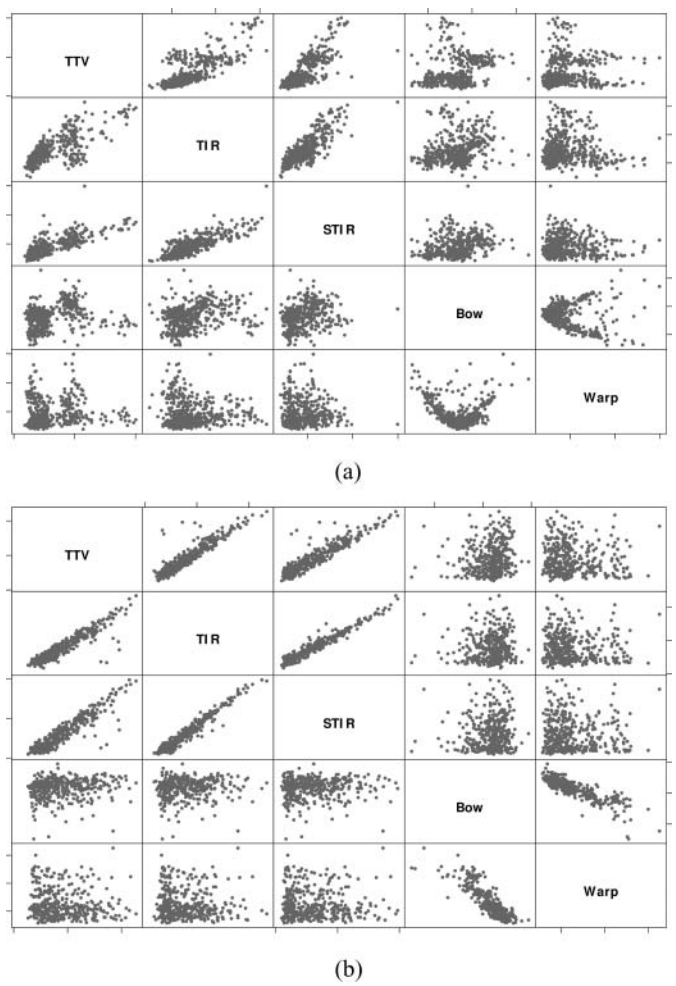

Fig. 6. Scatter plots of wafer quality variables: (a) a normal sample and (b) an out-of-control sample.

The in-control covariance matrix was estimated by the penalized likelihood estimation from the observations shown in Fig. 6(a), after some data manipulation in order to mask the actual scales of the observations in order to maintain data confidentiality. The matrix thus obtained is equal to

$$\boldsymbol{\Sigma}_0 = \begin{pmatrix} 1.30 & 0.46 & 0.51 & 0 & 0 \\ 0.46 & 1.30 & 0.53 & 0 & 0 \\ 0.51 & 0.53 & 1.30 & 0 & 0 \\ 0 & 0 & 0 & 1.30 & 0 \\ 0 & 0 & 0 & 0 & 1.30 \end{pmatrix}.$$

The out-of-control covariance matrix was estimated in a similar way from the observations shown in Fig. 6(b) and is equal to

$$\boldsymbol{\Sigma}_{\mathrm{OC}} = \begin{pmatrix} 1.30 & 0.62 & 0.63 & 0 & 0 \\ 0.62 & 1.30 & 0.68 & 0 & 0 \\ 0.63 & 0.68 & 1.30 & 0 & 0 \\ 0 & 0 & 0 & 1.30 & -0.55 \\ 0 & 0 & 0 & -0.55 & 1.30 \end{pmatrix}.$$

To demonstrate the applications of the proposed PLR chart and other competing charts, we assumed that the process covariance matrix changed from $\boldsymbol{\Sigma}_0$ to $\boldsymbol{\Sigma}_{\mathrm{OC}}$ as previously described. Based on this $\boldsymbol{\Sigma}_{\mathrm{OC}}$, we generated 30 samples each with 50 observations that were used to construct the control charts. For each observation generated, it was first transformed by multiplying the observations by $\boldsymbol{\Sigma}_0^{-1/2}$. The transformed observations were then used to calculate the plotting statistics of the PLR chart ($\lambda = 0.1$) and other existing charts shown in Fig. 7. Note that the in-control covariance matrix of the transformed variable is just the identity matrix, whereas the out-of-control covariance

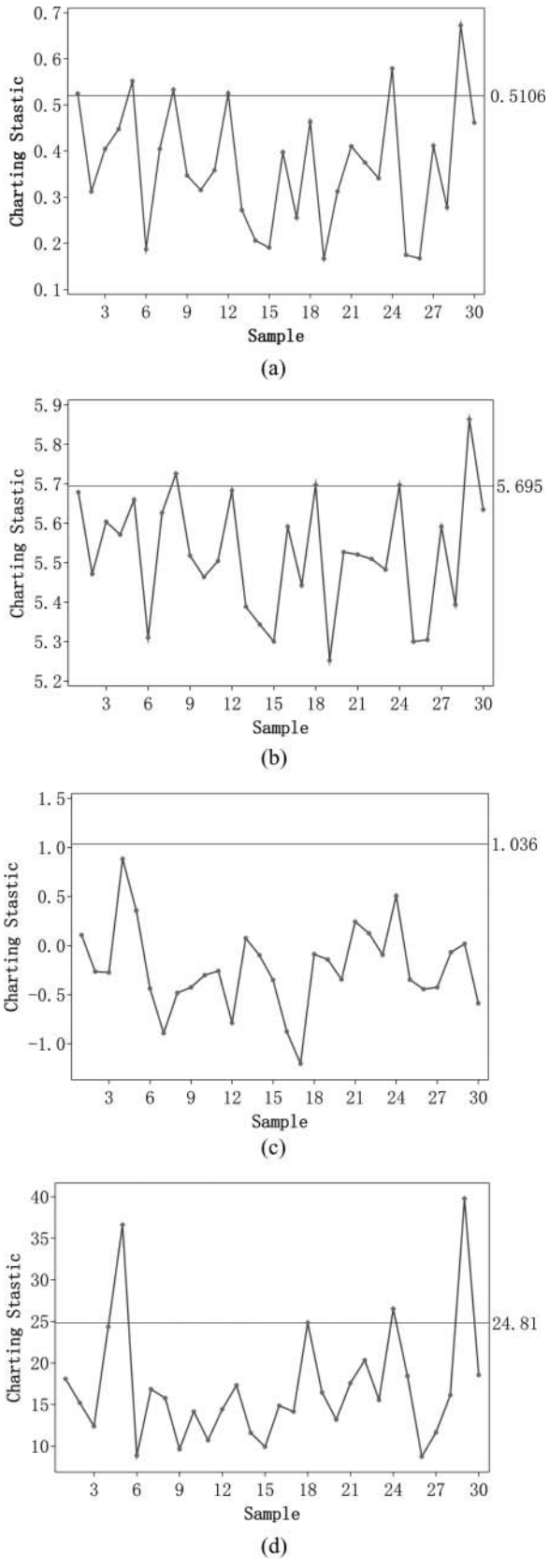

Fig. 7. The control charts of the wafer example: (a) the PLR chart; (b) the LR chart; (c) the CE chart; and (d) the S_d chart.

matrix is equal to

$$\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{\Sigma}_{\mathrm{OC}}\boldsymbol{\Sigma}_0^{-1/2\mathrm{T}} = \begin{pmatrix} 0.83 & -0.03 & 0.02 & 0 & 0 \\ -0.03 & 0.81 & 0 & 0 & 0 \\ 0.02 & 0 & 1.13 & 0 & 0 \\ 0 & 0 & 0 & 1.00 & -0.43 \\ 0 & 0 & 0 & -0.43 & 1.00 \end{pmatrix}.$$

It is seen from Fig. 7 that the proposed PLR chart (Fig. 7(a)) identifies six out-of-control signals and the first out-of-control signal shows up at sample 1. The LR chart (Fig. 7(b)) identifies four out-of-control signals with the first signal showing up at sample 8. The S_{d} chart (Fig. 7(d)) also identifies four out-of-control signals with the first signal showing up at sample 5. The CE chart (Fig. 7(c)), however, does not show any out-of-control signals.

In the performance evaluation, we simply assumed that $\boldsymbol{\Sigma}_0 = \mathbf{I}_p$. In the real example, we first transformed the observations to obtain an identity covariance matrix by multiplying each observation by $\boldsymbol{\Sigma}_0^{-1/2}$. As pointed out by one referee, the ARL_1 performance of the PLR chart will depend on the type of standardization used since the L_1-norm is not affine invariant. Such a dependency of the chart performance on the type of standardization used can also be seen in the existing literature on multivariate control charts. For example, the ARL_1 performance of the regression-adjusted control chart for monitoring the process mean (Hawkins, 1991) also depends on the type of standardization used. The author suggested that the original p variables be ordered in such a way that more important variables appear before less important variables.

In many real applications, there is some existing knowledge of how the variables are ordered. In the example just analyzed, we assumed that there is a partition of the variables of interest such that variables between different sets might be uncorrelated and variables within a set are thought to be correlated, for both the in-control and out-of-control cases. We ordered the variables in such a way that variables in the same set were placed consecutively. Therefore, our assumption led to the same block-wise sparsity pattern for both the in-control and out-of-control covariance matrices. We then applied the L_1 penalty to adaptively find the sparsity pattern in the covariance matrix.

5. Conclusions and discussion

We have proposed and studied a new Phase II control chart, the PLR chart, for monitoring changes in the covariance matrix of a multivariate normal process. The PLR chart was developed under the premise that in practice changes to a covariance matrix usually take place in fewer elements or in smaller blocks of elements. The construction of the PLR chart was based on two steps. First, a penalized likelihood

function (the L_1 penalty function was used in the current article) was used to obtain an estimate of the inverse of the covariance matrix. Such an estimate was then substituted, in step 2, into the calculation of the negative log-likelihood ratio for testing $H_0 : \mathbf{\Sigma} = \mathbf{I}_p$ versus $H_1 : \mathbf{\Sigma} \neq \mathbf{I}_p$.

The performance of the proposed PLR chart was evaluated based on simulations and compared to that of several existing charts. The simulation results demonstrated that the PLR chart outperformed the existing charts for almost all of the out-of-control scenarios considered in the current article. In cases where some of the existing charts were specifically designed to be effective, the PLR chart was as effective as the best performing existing chart. These results seem to indicate that the overall performance of the PLR chart is very attractive, thus making it an appealing new control charting methodology for monitoring multivariate process variability. A real example from the semi-conductor industry concerning the lapping process was presented and analyzed using the PLR chart and several existing charts. The PLR chart seems to return favorable results. In the current article, we suggested using $\mathbf{\Sigma}_0^{-1/2}$ to standardize the original observations and the simulations demonstrated that the PLR chart outperformed a number of representative Shewhart charts for the out-of-control scenarios considered in our study. However, we did not investigate as to whether other types of standardization will have better or worse ARL_1 performance than using $\mathbf{\Sigma}_0^{-1/2}$. Such an investigation will be worthwhile to pursue.

As mentioned earlier, the PLR chart is a Shewhart chart. Nevertheless, from a methodological standpoint, it is rather straightforward to derive CUSUM and EWMA statistics based on Λ_λ. It would be interesting to study the CUSUM and EWMA extensions of the PLR chart and, more important, to examine whether such extensions result in better chart performance than the existing CUSUM and EWMA charts in monitoring the covariance matrix.

The PLR chart is derived from a penalized likelihood function. We believe that the use of a penalty function provides a novel way to integrate engineering or domain knowledge into developing new and possibly improved SPC methodologies. Research efforts directed along this line deserve attention in the future. Here we briefly discuss one such possibility.

Recently, several authors (e.g., Zhang *et al.* (2009)), have studied the problem of simultaneously monitoring the mean vector and covariance matrix on a single chart based on the GLR of testing $H_0 : \mu = \mu_0, \mathbf{\Sigma} = \mathbf{\Sigma}_0$ versus $H_1 : \mu \neq \mu_0$ or $\mathbf{\Sigma} \neq \mathbf{\Sigma}_0$. In practice, changes in both μ and $\mathbf{\Sigma}$ happen in only in a few elements. Wang and Jiang (2009) and Zou and Qiu (2009) have studied penalized likelihood function–based control charts for monitoring the mean vector. Naturally, the question arises as to whether, in the context of simultaneous monitoring, one can obtain more efficient estimates of μ and $\mathbf{\Sigma}$ (or $\mathbf{\Sigma}^{-1}$) by simultaneously penalizing μ and $\mathbf{\Sigma}$ in the likelihood function for estimating $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ (with two separate tuning parameters) and derive new control charts for simultaneously monitoring μ and $\mathbf{\Sigma}$ based on the estimates thus obtained. Note that simultaneously penalizing the mean vector and covariance matrix has been studied in the context of regression analysis; see, for example, Witten and Tibshirani (2009) and Rothman *et al.* (2010). This is a direction worth for this studying and we are currently conducting a follow-up study along this line.

Acknowledgements

We thank the Department Editor and two referees for numerous detailed and insightful suggestions that improved the presentation of the article. Dr. Li's work is supported in part by the National Natural Science Foundation of China (NSFC) grants 10801086, 70831003, and 71072012. Dr. Wang's work is supported by the NSFC under grant 71072012.

References

Alt, F. (1984) Multivariate quality control, in *Encyclopedia of Statistical Sciences*, volume 6, pp. 110–122.

Alt, F.B. and Bedewi, G.E. (1986) SPC for dispersion for multivariate data. *ASQC Quality Congress Transactions*, 248–254.

Alt, F.B. and Smith, N.D. (1988) Multivariate process control, in *Handbook of Statistics*, Krishnaiah, P.R. and Rao, C.R. (eds), Elsevier Science Publishers, New York, NY, pp. 333–351.

Chen, A. and Hong, H.-J. (2010) Fault detection and classification by sample covariance matrix and its application to plasma etcher. Preprint, National Taiwan University, Taiwan.

D'Aspremont, A., Banerjee, O. and El Ghaoui, L. (2008) First-order methods for sparse covariance selection. *Siam Journal on Matrix Analysis and Applications*, **30**, 56–66.

Dempster, A. (1972) Covariance selection. *Biometrics*, **28**, 157–175.

Friedman, J., Hastie, T. and Tibshirani, R. (2008) Sparse inverse covariance estimation with the graphical LASSO. *Biostatistics*, **9**, 432–441.

Guerrero-Cusumano, J.L. (1995) Testing variability in multivariate quality control—a conditional entropy measure approach. *Information Sciences*, **86**, 179–202.

Hawkins, D.M. (1991) Multivariate quality control based on regression adjusted variables. *Technometrics*, **33**, 61–75.

Hawkins, D.M. and Maboudou-Tchao, E.M. (2008) Multivariate exponentially weighted moving covariance matrix. *Technometrics*, **50**, 155–166.

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

Hotelling, H. (1947) Multivariate quality control-illustrated by the air testing of sample bombsights, in *Techniques of Statistical Analysis*, Eisenhart, C., Hastay, M.W. and Wallis, W.A. (eds), McGraw-Hill, New York, NY, pp. 111–184.

Huang, J.Z., Liu, N.P., Pourahmadi, M. and Liu, L.X. (2006) Covariance matrix selection and estimation via penalised normal likelihood. *Biometrika*, **93**, 85–98.

Huwang, L., Yeh, A.B. and Wu, C.-W. (2007) Monitoring multivariate process variability with individual observations. *Journal of Quality Technology*, **39**, 258–278.

O'Mara, W., Herring, R. and Hunt, L. (2007) *Handbook of Semiconductor Silicon Technology*, Crest Publishing House, South Africa.

Pourahmadi, M. (1999) Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation. *Biometrika*, **86**, 677–690.

Pourahmadi, M. (2000) Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. *Biometrika*, **87**, 425–435.

Reynolds, R.M., Jr. and Cho, G.-Y. (2006) Multivariate control charts for monitoring the mean vector and covariance matrix. *Journal of Quality Technology*, **38**, 230–253.

Rothman, A.J., Bickel, P.J., Levina, E. and Zhu, J. (2008) Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics*, **2**, 494–515.

Rothman, A.J., Levina, E. and Zhu, J. (2010) Sparse multivariate regression with covariance estimation. *Journal of Computational and Graphical Statistics*, **19**, 947–962.

Stoumbos, Z.G., Reynolds, M.R., Jr., Ryan, T.P. and Woodall, W.H. (2000) The state of statistical process control as we proceed into the 21st century. *Journal of the American Statistical Association*, **95**, 992–998.

Tang, P.F. and Barnett, N.S. (1996a) Dispersion control for multivariate processes. *Australian Journal of Statistics*, **38**, 235–251.

Tang, P.F. and Barnett, N.S. (1996b) Dispersion control for multivariate processes—some comparisons. *Australian Journal of Statistics*, **38**, 253–273.

Tibshirani, R. (1996) Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society, Series B*, **58**, 267–288.

Wang, K. and Jiang, W. (2009) High-dimensional process monitoring and fault isolation via variable selection. *Journal of Quality Technology*, **41**, 247–258.

Williams, J., Woodall, W., Birch, J. and Sullivan, J. (2006) On the distribution of Hotelling's T^2 statistic based on the successive differences covariance matrix estimator. *Journal of Quality Technology*, **38**, 217–229.

Witten, D.M. and Tibshirani, R. (2009) Covariance-regularized regression and classification for high dimensional problems. *Journal of the Royal Statistical Society, Series B*, **69**, 329–346.

Woodall, W.H. (2000) Controversies and contradictions in statistical process control (with discussions). *Journal of Quality Control*, **32**, 341–378.

Woodall, W.H. and Montgomery, D.C. (1999) Research issues and ideas in statistical process control. *Journal of Quality Technology*, **31**, 376–386.

Yeh, A.B., Huwang, L. and Wu, C.W. (2005) A multivariate EWMA control chart for monitoring process variability with individual observations. *IIE Transactions on Quality and Reliability Engineering*, **37**, 1023–1035.

Yeh, A.B., Huwang, L.C. and Wu, Y.F. (2004) A likelihood-ratio based EWMA control chart for monitoring variability of multivariate normal processes. *IIE Transactions on Quality and Reliability Engineering*, **36**, 865–879.

Yeh, A.B., Lin, D. and McGrath, R. (2006) Multivariate control charts for monitoring covariance matrix: a review. *Quality Technology and Quantitative Management*, **3**, 415–436.

Yen, C. and Shiau, J.J.H. (2010) A multivariate control chart for detecting increases in process dispersion. *Statistica Sinica*, **20**, 1683–1707.

Yuan, M. and Lin, Y. (2007) Model selection and estimation in the Gaussian graphical model. *Biometrika*, **94**, 19–35.

Zhang, J., Li, Z. and Wang, Z. (2009) Multivariate control chart for simultaneously monitoring process mean and variability. *Computational Statistics & Data Analysis*, **54**, 2244–2252.

Zou, C., Ning, X. and Tsung, F. (2010) LASSO-based multivariate linear profile monitoring. *Annals of Operations Research*, **192**, 3–19.

Zou, C. and Qiu, P. (2009) Multivariate statistical process control using LASSO. *Journal of The American Statistical Association*, **104**, 1586–1596.

Biographies

Bo Li received a Ph.D degree in Statistics from the University of California, Berkeley, and a bachelor's degree in Mathematics from Beijing University. He is an Assistant Professor of Statistics at the School of Economics and Management, Tsinghua University. His research interests are statistical methods for high-dimensional data, semi-parametric/nonparametric statistics, applied statistics and econometrics in business, and economics.

Kaibo Wang is an Associate Professor in the Department of Industrial Engineering, Tsinghua University, Beijing, China. He received his B.S. and M.S. degrees in Mechatronics from Xi'an Jiaotong University, Xi'an, China, and his Ph.D. in Industrial Engineering and Engineering Management from the Hong Kong University of Science and Technology, Hong Kong. His research interests include statistical quality control, multivariate statistical process control, and data-driven complex system modeling/monitoring/diagnosis/control.

Arthur B. Yeh is a Professor of Statistics and Chair of the Department of Applied Statistics and Operations Research at Bowling Green State University. Over the years, he has conducted and published research in several areas of industrial statistics, including, among others, optimal experimental designs, computer experiments, univariate and multivariate control charts, multivariate process capability indices, univariate and multivariate run-by-run process control, and statistical profile monitoring. He has also worked as a consultant for various local and international companies in both traditional and modern high-tech manufacturing environments. He currently serves as an Associate Editor for *The American Statistician*. He has also served in the past as the President of the Northwest Ohio Chapter of the American Statistical Association and the Chair of the Toledo Section of the American Society for Quality. He is a senior member of ASQ and an elected member of ISI, in addition to being a member of ASA, IMS, ICSA, and INFORMS.