Ann Oper Res (2012) 192:3–19
DOI 10.1007/s10479-010-0797-8

LASSO-based multivariate linear profile monitoring

Changliang Zou · Xianghui Ning · Fugee Tsung

Published online: 19 October 2010

Abstract In many applications of manufacturing and service industries, the quality of a process is characterized by the functional relationship between a response variable and one or more explanatory variables. Profile monitoring is for checking the stability of this relationship over time. In some situations, multiple profiles are required in order to model the quality of a product or process effectively. General multivariate linear profile monitoring is particularly useful in practice due to its simplicity and flexibility. However, in such situations, the existing parametric profile monitoring methods suffer from a drawback in that when the profile parameter dimensionality is large, the detection ability of the procedures commonly used T^2-type charting statistics is likely to decline substantially. Moreover, it is also challenging to isolate the type of profile parameter change in such high-dimensional circumstances. These issues actually inherit from those of the conventional multivariate control charts. To resolve these issues, this paper develops a new methodology for monitoring general multivariate linear profiles, including the regression coefficients and profile variation. After examining the connection between the parametric profile monitoring and multivariate statistical process control, we propose to apply a variable-selection-based multivariate control scheme to the transformations of estimated profile parameters. Our proposed control chart is capable of determining the shift direction automatically based on observed profile data. Thus, it offers a balanced protection against various profile shifts. Moreover, the proposed control chart provides an easy but quite effective diagnostic aid. A real-data example from the logistics service shows that it performs quite well in the application.

Keywords High-dimensional · Least angle regression · MEWMA · Profile monitoring · Statistical process control · Variable selection

C. Zou
LPMC and Department of Statistics, School of Mathematical Sciences, Nankai University, Tianjin, China

X. Ning · F. Tsung (✉)
Department of Industrial Engineering and Logistics Management, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong
e-mail: season@ust.hk

1 Introduction

Because of recent progress in sensing and information technologies, automatic data acquisition is commonly used in manufacturing and service industries. Consequently, a large amount of quality-related data of certain processes has become available. Statistical process control (SPC) of such data-rich processes is an important component for monitoring their performance. In many applications, quality of a process is characterized by the relationship between a response variable and one or more explanatory variables. At each sampling stage, one observes a collection of data points of these variables that can be represented by a curve (or profile). In some calibration applications, the profile can be described adequately by a linear regression model. In other applications, more flexible models are necessary for describing profiles properly. An extensive discussion of research problems on this topic is given by Woodall et al. (2004).

Studies focused on simple linear profiles have been particularly prosperous. See, for instance, Kang and Albin (2000), Kim et al. (2003), Mahmoud and Woodall (2004), Zou et al. (2006), among several others. Multiple and polynomial regression profile models are considered by Zou et al. (2007a), Kazemzadeh et al. (2008), Mahmoud (2008) and Jensen et al. (2008). Nonlinear profile models are investigated by Williams et al. (2007). Recently, profile monitoring for general profile model have also attracted much attention. See Zou et al. (2008, 2009) and Qiu et al. (2010) for the Phase II methods based on nonparametric regression; Ding et al. (2006) and Chicken et al. (2009) for procedures using various dimension-reduction techniques, such as wavelet transformations and independent component analysis. A recent review of the literature has been given by Woodall (2007).

All these recent studies concentrated on the situation with a univariate profile that only contains one response variable. Although such profiles can characterize various applications as described in the literature, multivariate profiles in which multiple response variables are involved simultaneously may be even more representative of most industrial applications in certain real world practices. When the correlation structure between quality characteristics is ignored and profiles are monitored separately, then misleading results may be expected (c.f., Lowry et al. 1992 and Hawkins 1991 for relevant discussions). However, research on the monitoring and diagnosis of multivariate linear profiles is still scanty. A recent research by Noorossana et al. (2010) discusses multivariate linear profile monitoring in Phase I analysis, mainly based on the ordinary least square estimation. In this paper, we focus on a study of the Phase II method for monitoring a general multivariate linear profile.

At first glance, the monitoring problem of multivariate linear profile is trivial if we use a similar method to the general univariate linear profile monitoring one proposed by Zou et al. (2007a), in which a control scheme based on integrating conventional multivariate SPC (MSPC) procedures, such as the multivariate EWMA (MEWMA) chart, with the ordinary LSE or equivalently maximum likelihood estimation (MLE) of the parameters of each profile is constructed. For a multivariate model, it amounts to incorporating certain correlation structures into the charting statistics. However, it should be emphasized that in a multivariate profile problem, we are dealing with a high-dimensional case especially when the number of responses incorporated is large. This issue actually inherits from that of the conventional multivariate control charts when the dimensionality is large because the parametric profile monitoring can be regarded as a special case of MSPC to some extent (Zou et al. 2007a; Woodall 2007). Due to the high dimensionality and the large scale of the data being monitored, existing theories based on conventional estimation and testing methods are usually of limited use or inefficient in practice for real-time, high-dimensional statistical computing and on-line feature selection and dimension reduction.

As a solution, Zou and Qiu (2009) recently propose a new multivariate control chart, the LASSO-based EWMA chart (or LEWMA), for monitoring multiple parameters. It is based on adapting the recent variable selection method LASSO (cf., Tibshirani 1996) to the SPC problem. Their proposed control chart is capable of determining the shift direction automatically based on observed data. Thus, it offers a balanced protection against various shifts. Furthermore, the proposed control chart provides an easy to use, yet quite effective diagnostic aid, mainly due to the sparsity property of a LASSO estimator that some of its components would be exactly zero after an appropriate choice of its tuning parameters.

In light of the LEWMA chart, this paper develops a variable-selection-based SPC methodology for monitoring general multivariate linear profiles. We propose utilizing the LEWMA control scheme for the transformations of estimated parameters, as a single chart to monitor both the coefficients and variances of a multivariate linear profile. Due to certain good properties of LEWMA, the proposed chart is easy to implement, fast to compute, and efficient to detect various profile shifts. As a by-product, this chart also provides a tool for profile diagnosis which is critical for a complicated multivariate linear profile. The remainder of this paper is organized as follows: in Sect. 2, we introduce a loading process example from the logistics industries that motivates this research. Our proposed methodology is described in detail in Sect. 3; Its numerical performance is thoroughly investigated in Sect. 4; The motivating example, which has a profile that fits a multivariate linear regression model well, is used to illustrate the step-by-step implementation of the proposed approach in Sect. 5. The article is concluded with several remarks in Sect. 6.

2 The motivating example: monitoring a loading process

We use an example taken from a logistic company to illustrate the motivation for this research. Everyday, the logistic company has ships from all over the world visiting its ports. The ships either belong to the company itself or its contractors. They release their cargoes at the ports before moving on to the next destinations. One of the job at the ports is to load cargoes into the ports or discharge them. For this logistic company, it is better to complete as many jobs as possible in as little time. Therefore, the quality of the discharging or loading process can be characterized by two variables, the time needed to complete a job and the amount of cargoes in a job. The two variables together represent the efficiency and the capability of the process.

However, having only these two variables is not enough for characterizing the performance. Further study reveals that the time needed to complete a job is influenced by other factors. For instance, the scale of the contractor. If the contractor is of a large scale, the time needed to complete a job is around 27.28 minutes and if the contractor is of a medium scale, the time needed to complete a job is around 22.76 minutes. It can also be influenced by the current workload in the ports. If the ports are too busy, the time needed is longer. The same influences affect the other variable. That is, larger-scale contractors usually have more cargoes while medium-scale contractors have less. Other variables, like the container size, also affects the two process variables. From a technical point of view, it takes more time to transport larger containers than smaller ones. While at the same time, the size of a container reflects partially on the workload.

Therefore, the quality of the discharging or loading process depends not only on the time needed to complete a job and the amount of cargoes in a job, but also on the relationships between the two dependent variables and other independent variables. These relationships are critical to the quality of the process and requires careful monitoring over time. This is a typical multiple or multivariate profile monitoring problem. In the remainder of this paper, we propose an SPC scheme to monitor such a profile and give a step-by-step demonstration of how to implement the proposed scheme in practice in a later section.

3 Methodology

We describe the proposed control chart in four parts. In Sect. 3.1, the considered multivariate linear model formulation and the associated assumptions are introduced. Then, a brief introduction to Zou and Qiu's (2009) multivariate control chart, LEWMA, is presented in Sect. 3.2. A new multivariate profile monitoring and diagnostic scheme using the LEWMA chart is derived in Sect. 3.3. Its practical guidelines regarding design and computational issues are addressed in Sect. 3.4.

3.1 Model and assumptions

Assume that for the jth ($j \geq 1$) random sample collected over time, we have n observations on q responses $\mathbf{y} = (y_1, \ldots, y_q)^T$ and p explanatory variables $\mathbf{x} = (x_1, \ldots, x_p)^T$. The process observations are collected from the following general multivariate linear profile model,

$$\mathbf{Y}_j = \mathbf{X}_j \mathbf{B}_j + \mathbf{E}_j, \quad j = 1, \ldots, \tau, \tau + 1, \ldots, \tag{1}$$

where τ is the unknown change-point, $\mathbf{Y}_j = (\mathbf{y}_{j1}, \ldots, \mathbf{y}_{jq})$ is an $n \times q$ matrix, $\mathbf{X}_j = (\mathbf{x}_{j1}, \ldots, \mathbf{x}_{jn})^T$ is an $n \times p$ matrix, $\mathbf{B}_j = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_q)$ is a $p \times q$ coefficients matrix, $\mathbf{E}_j = (\mathbf{e}_{j1}, \ldots, \mathbf{e}_{jn})^T$ is the regression noise with $\mathbf{e}_{jk} = (\varepsilon_{jk1}, \ldots, \varepsilon_{jkq})^T$ and $\mathbf{e}$'s are independently sampled from a distribution with $(\mathbf{0}, \boldsymbol{\Sigma}_j)$ with the diagonal components of $\boldsymbol{\Sigma}_j$ being $\mathbf{s}_j = (\sigma_{j11}^2, \ldots, \sigma_{jqq}^2)$. It is assumed that after some unknown change-point τ, there is a change in the coefficient matrix and/or variances of the profiles, i.e.

$$\begin{aligned} \mathbf{B}_j &= \mathbf{B}_{(0)}, \qquad \mathbf{s}_j = \mathbf{s}_{(0)} \quad \text{for } j \leq \tau, \\ \mathbf{B}_j &= \mathbf{B}_{(1)}, \qquad \mathbf{s}_j = \mathbf{s}_{(1)} \quad \text{for } j > \tau, \end{aligned}$$

$\mathbf{B}_{(0)} \neq \mathbf{B}_{(1)}$ and/or $\mathbf{s}_{(0)} \neq \mathbf{s}_{(1)}$. We will elaborate on this shift model in this paper because it is of great interest in practice. The extension to the case of monitoring the entire covariance matrices will be discussed in Sect. 6. Here we shall assume that $n > p + 1$ which is not a restrictive one and can easily be satisfied in practical applications.

In what follows, for ease of exposition, the explanatory variable matrix, $\mathbf{X}_j$, is assumed to be fixed for different j (denoted as $\mathbf{X}$). This is usually the case in practical calibration applications in industrial manufacturing and is also consistent with the literatures, such as Kim et al. (2003) and Zou et al. (2007a, 2008). For the profile monitoring with more complex covariate designs, we refer to Qiu and Zou (2010) and Qiu et al. (2010). Without loss of generality, here suppose that the first column of $\mathbf{X}$ is $\mathbf{1}$ where $\mathbf{1}$ is an n-variate vector of all 1s and the other columns are orthogonal to $\mathbf{1}$. Otherwise, we can obtain this form through some appropriate transformations.

We mainly consider the Phase-II case in which the necessary in-control (IC) parameters are assumed to be known as a convention in the literatures. It is essentially equivalent to saying that the number of historical samples is sufficiently large. Once the IC models are established as a baseline, in Phase II, we would want to detect any change in the regression coefficients and profile variance as quickly as possible. For Phase I study, interested readers can refer to Noorossana et al. (2010) for multivariate multiple regression and Kazemzadeh et al. (2008), Mahmoud (2008) and Jensen et al. (2008) for univariate linear regression.

As mentioned above, the parametric profile monitoring can be regarded as a special case of MSPC since we usually apply the multivariate control chart to the estimated parameter (coefficient and variance) vector. Since the conventional MSPC methods may fail under a high-dimensional situation as discussed above, one may first use a dimension-reduction method, such as principal component analysis, before process monitoring. However, these methods usually yield some linear combinations of all original variables, making it much more difficult to exploit the on-line abnormal structures and to interpret a signal, especially when dimension is large or the variables are of different scales. Thus, most recent research on MSPC focus more on directly monitoring all available data/full information instead of skipping or reducing information. We'll propose our method along these lines where we will incorporate all information during on-line monitoring. Of course, it is desirable and necessary to screen irrelevant and redundant predictors from original covariates before monitoring to improve the detection and diagnosis capabilities. Recent development on variable filtering and estimation in multivariate linear regression can be found in Yuan et al. (2007) and the references therein. This is beyond the scope of this paper but should be subjects of future research. In the remainder of this paper, we suppose that all the explanatory variables involved in model (1) are significant in regression, that is, all the irrelevant and redundant covariates have been removed from the model by using some appropriate methods.

3.2 A brief review of MSPC

In this subsection, we review some multivariate schemes and introduce the LEWMA chart suggested by Zou and Qiu (2009). In MSPC, one monitors several quality characteristics of a process. The fundamental tasks of MSPC are to determine whether a multivariate process mean, $\boldsymbol{\mu}$, has changed; to identify when a detected shift in $\boldsymbol{\mu}$ has occurred; and to isolate the shifted components of $\boldsymbol{\mu}$. Methods for accomplishing these tasks are usually derived under the assumption that the observed measurement vectors, $\mathbf{x}_i = (x_{1i}, \ldots, x_{pi})'$, are $N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$ for $i = 1, 2, \ldots, \tau$, and $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ for $i = \tau + 1, \ldots, n$, with $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}$ known. Throughout, we set $\boldsymbol{\mu}_0 = \mathbf{0}$ without loss of generality. Then, a portmanteau test for detecting a mean shift occurring at τ is based on testing $H_0 : \boldsymbol{\mu} = \mathbf{0}$ versus $H_1 : \boldsymbol{\mu} \neq \mathbf{0}$, using the likelihood ratio test statistic, $n\bar{\mathbf{x}}'\Sigma^{-1}\bar{\mathbf{x}}$, where $\bar{\mathbf{x}} = \sum_{i=1}^{n} \mathbf{x}_i/n$. Based on such test statistics, several MSPC control charts have been proposed, in the framework of CUSUM or EWMA, such as the MCUSUM proposed by Croisier (1988) and the MEWMA proposed by Lowry et al. (1992), etc. After a control chart signals a mean shift, a separate diagnostic procedure is often used for identifying the shift time and the component(s) of $\boldsymbol{\mu}$ that shifted. Commonly used diagnostic procedures include the ones based on decomposition of T^2 (e.g., Mason and Young 2002) and various step-down procedures (e.g., Sullivan et al. 2007).

Although conventional MSPC control charts with quadratic charting statistics are powerful in a low-dimensional situation, their shift-detection ability would decline substantially when p increases, due to the well-known "curse of dimensionality". Regarding post-detection diagnosis, conventional approaches (e.g., the decomposition of the T^2 procedure and the step-down test) are theoretically sound, but they are inefficient when p is large. For instance, the decomposition of the T^2 procedure considers $p!$ different decompositions of T^2, which is computationally expensive in such cases. Certain parameters in these approaches (e.g., the threshold values) affect their diagnostic ability significantly, but they are generally difficult to determine (Sullivan et al. 2007).

Zou and Qiu (2009) proposed variable-selection-based MSPC control charts, based on the following assumption: in a high-dimensional process, the probability that all variables shift simultaneously is rather low; Instead, an alarm is more likely to be caused by a hidden source, which affects one or a small set of observable variables. They consider the penalized likelihood function of multivariate observations based on the conventional multinormality assumption. To be specific, Zou and Qiu (2009) integrate a LASSO test statistic into a multivariate EWMA charting scheme for on-line process monitoring, LEWMA, based on the following adaptive LASSO (ALASSO; Zou 2006) penalized likelihood,

$$(\mathbf{U}_j-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{U}_j-\boldsymbol{\mu})+\gamma\sum_{k=1}^{p}\frac{1}{|\mathbf{U}_j^{(k)}|}|\boldsymbol{\mu}^{(k)}|, \tag{2}$$

where

$$\mathbf{U}_j=\lambda\mathbf{x}_j+(1-\lambda)\mathbf{U}_{j-1},\quad \text{for } j=1,2,\ldots, \tag{3}$$

$\mathbf{U}_0=\mathbf{0}$, λ is a weighting parameter in (0, 1], and $\boldsymbol{\mu}^{(k)}$ and $\mathbf{U}_j^{(k)}$ denote the kth components of $\boldsymbol{\mu}$ and $\mathbf{U}_j$, respectively. The proposed charting statistic is

$$Q_j=\max_{k=1,\ldots,q}\frac{W_{j,\tilde{\gamma}_{m_k^{\text{last}}}}-\mathrm{E}(W_{j,\tilde{\gamma}_{m_k^{\text{last}}}})}{\sqrt{\mathrm{Var}(W_{j,\tilde{\gamma}_{m_k^{\text{last}}}})}}>L, \tag{4}$$

where

$$W_{j,\gamma}=\frac{2-\lambda}{\lambda[1-(1-\lambda)^{2j}]}\frac{(\mathbf{U}_j^T\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\mu}}_\gamma)^2}{\widehat{\boldsymbol{\mu}}_\gamma^T\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\mu}}_\gamma},$$

and $\widehat{\boldsymbol{\mu}}_\gamma$ is the minimizer of (2) given that γ, m_k^{last} is the index of the last $\tilde{\gamma}$ in the sequence of the ALASSO transition points $\{\tilde{\gamma}_0,\tilde{\gamma}_1,\ldots,\tilde{\gamma}_K\}$ in which the corresponding active set contains exactly k elements. The random integer K is the total number of transition points and can be larger than p (Efron et al. 2004). This approach is easy to implement using the LARS algorithm (cf., Efron et al. 2004). It is shown that this approach balances protection against various shift levels and shift directions, and hence provides an effective tool for multivariate SPC applications particularly in a high-dimensional setting. We are naturally led to consider the LEWMA in our profile monitoring problem because there are more process parameters involved.

3.3 Control charts for the monitoring and diagnosis of general multivariate linear profiles

Recall model (1) and the associated notation. To monitor a general multivariate linear profile model (1), there are $pq+q$ parameters, pq regression coefficients and q variances $\mathbf{s}$, to be controlled simultaneously. For notation convenience, we define $\mathbf{z}_{(k)}=(\mathrm{Vec}^T(\mathbf{B}_{(k)}),\mathbf{s}_{(k)}^T)^T$ for $k=0,1$ and $\boldsymbol{\delta}=\mathbf{z}_{(1)}-\mathbf{z}_{(0)}=:(\boldsymbol{\delta}_C^T,\boldsymbol{\delta}_V^T)^T$, where $\mathrm{Vec}(\mathbf{S})$ denotes the pq-dimensional vector formed by stacking the columns of a $(p\times q)$-dimensional matrix $\mathbf{S}$, and $\boldsymbol{\delta}_C$ and $\boldsymbol{\delta}_V$ denote the shift vectors for coefficients and variances, respectively.

In the MSPC problem discussed in the previous subsection, an *a priori* assumption is that some components of the mean vector $\boldsymbol{\mu}$ are zero. In high-dimensional cases, it is often reasonable to assume that only a few components are non-zero, which is the so-called *sparsity* characteristic. Analogously, in the profile monitoring problem, we assume in model (1) that only a few profile parameters in the shift vector $\boldsymbol{\delta}$ are expected to be non-zero when a shift occurs. Therefore, the penalized likelihood method, LEWMA, should have a potential in solving the profile problem, which is investigated below.

To motivate our final proposal, we firstly assume that the variances of profiles do not change and the errors are multivariate normally distributed. Similar to Qiu et al. (2010), at any time point t, we consider the following exponential weighted-2log-likelihood function after a constant term is ignored

$$WL(\mathbf{B}; t, \lambda) = \sum_{j=1}^{t} \lambda(1-\lambda)^{t-j} \operatorname{tr}\left((\mathbf{Y}_j - \mathbf{X}\mathbf{B})\boldsymbol{\Sigma}_{(0)}^{-1}(\mathbf{Y}_j - \mathbf{X}\mathbf{B})^T\right),$$

where we use the fact that under IC $\mathrm{Vec}(\mathbf{Y}_j) \sim N_{pq}(\mathrm{Vec}(\mathbf{B}), \boldsymbol{\Sigma}_{(0)} \otimes \mathbf{I}_p)$, $\operatorname{tr}(\cdot)$ stands for the trace operator, $\otimes$ is the Kronecker product, $\mathbf{I}_p$ denotes the $p \times p$ identity matrix and λ is a weighting parameter. From the expression of $WL(\mathbf{B})$ above, we can see that this function makes use of all available observations up to the current (i.e., the tth) profile data, and different profiles are weighted as in a conventional EWMA chart (i.e., more recent profiles get more weight and the weights change exponentially over time). Then, the estimator of $\mathbf{B}$ at time point t, defined as the solution to $\mathbf{B}$ of the minimization problem $\min_{\mathbf{B}} WL(\mathbf{B}; t, \lambda)$, has the expression

$$\tilde{\mathbf{B}}_t = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\tilde{\mathbf{Y}}_t,$$
$$\tilde{\mathbf{Y}}_t = (1-\lambda)\tilde{\mathbf{Y}}_{t-1} + \lambda\mathbf{Y}_t,$$

or equivalently in a standard EWMA recursive form

$$\tilde{\mathbf{B}}_t = (1-\lambda)\tilde{\mathbf{B}}_{t-1} + \lambda\widehat{\mathbf{B}}_t, \tag{5}$$

where $\widehat{\mathbf{B}}_t = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}_t$ and $\widehat{\mathbf{B}}_0 = 0$.

Furthermore, by taking the sparsity of $\boldsymbol{\delta}_C$ into account, we consider the following weighted ALASSO penalized likelihood,

$$AWL(\mathbf{B}; t, \lambda) = \sum_{j=1}^{t} \lambda(1-\lambda)^{t-j} \operatorname{tr}\left((\mathbf{Y}_j - \mathbf{X}\mathbf{B})\boldsymbol{\Sigma}_{(0)}^{-1}(\mathbf{Y}_j - \mathbf{X}\mathbf{B})^T\right) + \gamma \sum_{k=1}^{pq} \frac{|\mathrm{Vec}^{(k)}(\mathbf{B} - \mathbf{B}_0)|}{|\mathbf{z}_{tB}^{(k)}|},$$

where $\mathbf{z}_{tB} = \mathrm{Vec}(\tilde{\mathbf{B}}_t) - \mathrm{Vec}(\mathbf{B}_0)$. By using the properties of trace operator and $\widehat{\mathbf{B}}_t$, it is straightforward to see that

$$\begin{aligned}
\arg\min_{\mathbf{B}} AWL(\mathbf{B}; t, \lambda) &= \arg\min_{\mathbf{B}} \operatorname{tr}\left((\mathbf{X}\tilde{\mathbf{B}}_t - \mathbf{X}\mathbf{B})\boldsymbol{\Sigma}_{(0)}^{-1}(\mathbf{X}\tilde{\mathbf{B}}_t - \mathbf{X}\mathbf{B})^T\right) \\
&\quad + \gamma \sum_{k=1}^{pq} \frac{|\mathrm{Vec}^{(k)}(\mathbf{B} - \mathbf{B}_0)|}{|\mathbf{z}_{tB}^{(k)}|} \\
&= \arg\min_{\mathbf{B}} [\mathrm{Vec}(\tilde{\mathbf{B}}_t - \mathbf{B})]^T \boldsymbol{\Sigma}_{(0)}^{-1} \otimes (\mathbf{X}^T\mathbf{X})[\mathrm{Vec}(\tilde{\mathbf{B}}_t - \mathbf{B})] \\
&\quad + \gamma \sum_{k=1}^{pq} \frac{|\mathrm{Vec}^{(k)}(\mathbf{B} - \mathbf{B}_0)|}{|\mathbf{z}_{tB}^{(k)}|}.
\end{aligned}$$

Then, this minimization can be represented in vector-form as

$$\arg\min_{\boldsymbol{\mu}} (\mathbf{z}_{tB} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}_{(0)}^{-1} \otimes (\mathbf{X}^T\mathbf{X})(\mathbf{z}_{tB} - \boldsymbol{\mu}) + \gamma \sum_{k=1}^{pq} \frac{1}{|\mathbf{z}_{tB}^{(k)}|}|\boldsymbol{\mu}^{(k)}|. \tag{6}$$

Heuristically speaking, by noting that

$$\mathrm{Vec}(\tilde{\mathbf{B}}_t) - \mathrm{Vec}(\mathbf{B}_0) \sim N_{pq}(\boldsymbol{\delta}_C, \boldsymbol{\Sigma}_{(0)} \otimes (\mathbf{X}^T\mathbf{X})^{-1}),$$

we can directly obtain (6) by mimicking the formula (2) with $\mathbf{U}_j$ being replaced by $\mathbf{z}_{tB}$. However, the foregoing arguments provide a formal derivation of our proposed testing statistic.

Following Zou et al. (2007a), we define a working vector

$$\widehat{\mathbf{s}}_t = \sqrt{n-p}\left(\frac{\widehat{\sigma}_{t1}}{\sigma_{(0)11}} - 1, \ldots, \frac{\widehat{\sigma}_{tq}}{\sigma_{(0)qq}} - 1\right)^T,$$

where

$$\widehat{\sigma}_{ti}^2 = (\mathbf{y}_{ti} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{ti})^T(\mathbf{y}_{ti} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{ti})/(n-p),$$

$\widehat{\boldsymbol{\beta}}_{ti}$ is the ith column of $\widehat{\mathbf{B}}_t$ and $\sigma_{(0)11}^2$ is the ith diagonal element of $\boldsymbol{\Sigma}_{(0)}$. Then we could construct an EWMA sequence for $\widehat{\mathbf{s}}_t$ as

$$\mathbf{z}_{tV} = (1-\lambda)\mathbf{z}_{t-1V} + \lambda\widehat{\mathbf{s}}_t,$$

and rewrite $\mathbf{z}_{tB}$ and $\mathbf{z}_{tV}$ into one vector together, say $\mathbf{z}_t = (\mathbf{z}_{tB}^T, \mathbf{z}_{tV}^T)^T$. When the process is in-control, the vector $\mathbf{z}_t$ is asymptotically multivariate normally distributed with mean $\mathbf{0}$ and covariance matrix $\lambda\boldsymbol{\Omega}/(2-\lambda)$, where

$$\boldsymbol{\Omega} = \mathrm{diag}\{\boldsymbol{\Sigma}_{(0)} \otimes (\mathbf{X}^T\mathbf{X})^{-1}, \Gamma_q\}.$$

The asymptotic expression of Γ_q is given in the Appendix.

Thus, we can generalize formula (6) by taking the variances of profiles into account as follows,

$$(\mathbf{z}_t - \boldsymbol{\mu})^T\boldsymbol{\Omega}^{-1}(\mathbf{z}_t - \boldsymbol{\mu}) + \gamma\sum_{k=1}^{pq+q}\frac{1}{|\mathbf{z}_t^{(k)}|}|\boldsymbol{\mu}^{(k)}|, \tag{7}$$

where we still use the notation $\boldsymbol{\mu}$ here which should not cause any confusion. This is exactly a LASSO-type penalized likelihood function. Following Zou and Qiu (2009), for each $\mathbf{z}_t$, we suggest computing q LASSO estimators $\widehat{\boldsymbol{\mu}}_{t,\gamma_k}$, for $k = 1, 2, \ldots, q$, from the penalized likelihood function (7) and obtain the corresponding q regression-adjustment-type testing statistics

$$W_{t,\gamma_k} = \frac{2-\lambda}{\lambda}\frac{(\mathbf{z}_t^T\boldsymbol{\Omega}^{-1}\widehat{\boldsymbol{\mu}}_{t,\gamma_k})^2}{\widehat{\boldsymbol{\mu}}_{t,\gamma_k}^T\boldsymbol{\Omega}^{-1}\widehat{\boldsymbol{\mu}}_{t,\gamma_k}}, \quad k = 1, \ldots, q,$$

and $\widehat{\boldsymbol{\mu}}_{t,\gamma_k}$ is the minimizer of (6) given $\gamma = \gamma_k$, γ_k is the index of the last penalty parameter in the sequence of the ALASSO transition points $\{\gamma_0, \gamma_1, \ldots, \gamma_K\}$ in which the corresponding active set contains exactly k elements. If we have prior information indicating potential shifts in r components, we could use W_{t,γ_r} as the charting statistic which has been shown to have certain optimal properties in detecting such shift patterns (Zou and Qiu 2009). Of course, in practice, r is rarely known in advance. Similar to Zou and Qiu (2009), we suggest combining all these testing statistics into one charting statistic by taking the maximum of their standardizations, that is,

$$Q_t = \max_{k=1,\ldots,q}\frac{W_{t,\gamma_k} - \mathrm{E}(W_{t,\gamma_k})}{\sqrt{\mathrm{Var}(W_{t,\gamma_k})}}, \tag{8}$$

and the proposed chart signals if $Q_t > L$ where $L > 0$ is a control limit chosen to achieve a given IC average run length (ARL). Regarding q, if prior information indicating potential shifts in at most r components, with $1 \le r \le p$, then our numerical studies show that using $q = r + 1$ or $q = r + 2$ provides satisfactory performance in practice. When prior information is unavailable, numerical results (Sect. 4) show that the LEWMA chart with $q = p$ performs reasonably well in all cases considered there.

In the practice of quality control, in addition to being able to detect a process change quickly, it is also critical to identify which parameter(s) have shifted after an out-of-control (OC) signal occurs. A diagnostic aid to isolate the type of parameter change will help an engineer to identify and eliminate the root cause of a problem in a timely manner. After the LEWMA chart gives an OC signal at the tth observation, the LASSO methodology can be used to specify the shifted parameters, by choosing one of the LASSO estimators with a model selection criterion (e.g., C_p, GCV, AIC, or BIC). Namely, we can find $\widehat{\boldsymbol{\mu}}_{\gamma^*}$ so that

$$\gamma^* = \arg\min_{\gamma} \frac{2-\lambda}{\lambda}(\mathbf{z}_t - \widehat{\boldsymbol{\mu}}_\gamma)'\boldsymbol{\Omega}^{-1}(\mathbf{z}_t - \widehat{\boldsymbol{\mu}}_\gamma) + \eta \cdot \widehat{\mathrm{df}}(\widehat{\boldsymbol{\mu}}_\gamma), \tag{9}$$

where $\widehat{\boldsymbol{\mu}}_\gamma$ is the minimizer of (6) given γ and $\widehat{\mathrm{df}}(\widehat{\boldsymbol{\mu}}_\gamma)$ is the number of nonzero coefficients in $\widehat{\boldsymbol{\mu}}_\gamma$. Here η is a parameter indicating which model selection criterion is used, such as $\eta = 2$ for AIC-type and $\eta = \log(n)$ for BIC-type criteria. The choice of η will be carefully examined and certain guidelines will be offered in Sect. 3.4. Since some components of $\widehat{\boldsymbol{\mu}}_{\gamma^*}$ would be exactly zero, we can simply take its nonzero components as the shifted components, without making any extra tests that are necessary in most existing diagnostic methods, such as in the decomposition of T^2 method or in the step-down method. It is convenient to use and is fast in computation as well because the LASSO estimators from $\mathbf{z}_t$ have been computed in the monitoring process. By generalizing the proof of Theorem 3 in Zou et al. (2007), we can obtain results in the following proposition without much difficulty, which greatly facilities the searching procedure.

Proposition 1 *$\widehat{\boldsymbol{\mu}}_{\gamma^*}$ is one of $\widehat{\boldsymbol{\mu}}_{\gamma_1}, \ldots, \widehat{\boldsymbol{\mu}}_{\gamma_K}$, where $\widehat{\boldsymbol{\mu}}_{\gamma_i}$s are the ALASSO solutions to (7) at the transition points.*

With this proposition, we can obtain the diagnostic result easily and quickly after the LEWMA chart triggers a signal. This diagnostic tool will provide a reasonable and efficient alternative to the existing fault isolation method in the literature, such as Sullivan et al. (2007).

3.4 Design and implementation of the proposed schemes

On implementation In calculating the charting statistic Q_t in (8), quantities $\mathrm{E}(W_{j,\gamma_k})$ and $\mathrm{Var}(W_{j,\gamma_k})$ are usually unknown. By the central limit theorem, it can be easily verified that when the process is IC, $\mathbf{z}_t$ is asymptotically distributed as a multivariate normal distribution with mean zero and covariance matrix $\lambda\boldsymbol{\Omega}/(2-\lambda)$, if $n \to \infty$ or $\lambda \to 0$. As a consequence, the following result suggests replacing them by their approximations.

Proposition 2 *When the process is IC, the distribution of W_{t,γ_k} is asymptotically equivalent to that of the corresponding statistic by replacing the $\mathbf{z}_t$ with a multivariate normal observation with mean zero and covariance matrix $\boldsymbol{\Omega}$, as $n \to \infty$ or $\lambda \to 0$.*

By Proposition 2, $\mathrm{E}(W_{t,\gamma_k})$ and $\mathrm{Var}(W_{j,\gamma_k})$ are asymptotically free of λ and t, and thus they can respectively be approximated by the empirical expectation and variance of W_{j,γ_k} computed from simulated multivariate normal measurement vectors with mean zero and covariance matrix $\mathbf{\Omega}$. Since this is a one-time computation before the Phase II online process monitoring, it is convenient to accomplish.

On choosing the smoothing weight, λ As demonstrated by Stoumbos and Sullivan (2002), the MEWMA chart should be more appealing than multivariate nonparametric schemes because MEWMA charts can be quite *robust* in the sense that the IC run length distribution for a continuous non-normal process is quite close to the distribution of a multivariate normal process with the same control limit if the weighting parameter, λ, is small. This is still valid for the LEWMA chart. With a large number of observations and a small smoothing parameter, the central limit theorem would ensure that the accumulation vector has approximately a multinormal distribution, which ensures robustness (c.f., Proposition 2). In fact, some preliminary evidence can be found in Zou and Qiu (2009). Specifically, in the present profile monitoring problem, the sample size n is always quite large in practice, which in turn will further alleviate the problem of degraded statistical performance due to the non-normality (c.f., Montgomery 2005). Our numerical results show that LEWMA's superiority still holds under various non-normal multivariate distribution assumptions.

Of course, similar to MEWMA, in the LEWMA chart, λ should be chosen to balance the robustness to non-normality and the detection ability to various shift magnitudes (c.f., Stoumbos and Sullivan 2002). In general, a smaller λ leads to a quicker detection of smaller shifts (c.f., e.g., Lucas and Saccucci 1990). When λ is too small, the corresponding procedure would not be sensitive to relatively large shifts. Based on our simulation results, we suggest choosing $\lambda \in [0.03, 0.08]$, which is a reasonable choice in practice, and using $\lambda \in [0.005, 0.025]$ when prior information indicates that the underlying distribution is very skewed.

On the control limits and computation For a given λ, $\mathbf{\Omega}$ and a desired IC ARL, computation involved in finding L is not difficult, partly due to the fact that the LARS algorithm used in LASSO computation is efficient. In the searching procedure, some numerical searching algorithms, such as the bisection search, can be applied (cf., e.g., Qiu 2008). For instance, when IC ARL $= 500$ and $pq + q = 15$, it requires about 15 minutes to complete the bisection searching procedure based on 10,000 simulations, using a Pentium-M 2.4 MHz CPU. Fortran code for implementing the proposed procedure is available from the authors upon request.

On post-signal diagnosis In using the diagnostic procedure (9), we need to specify the parameter η. In the literature, it is well demonstrated that AIC (identically C_p) tends to select the model with the optimal predication performance, while BIC tends to identify the true sparse model well if the true model is included in the candidate set (cf., e.g., Wang et al. 2007). As we want to identify the special cause of the OC condition, the sparsity of $\boldsymbol{\mu}^*$ is our primary concern and consequently the BIC criterion may be more relevant. However, as argued by Zou and Qiu (2009), unlike the fixed sample case, in the current SPC problem, we usually do not have a large sample (relative to the shift magnitude) to implement the diagnostic procedure, because the whole point of SPC is to detect the shift (and hence stop the process) as quickly as possible. Based on extensive simulations, we find that the conventional BIC criterion does not perform well in certain cases. Instead, using the recommendation in Zou and Qiu (2009), the risk inflation criterion (RIC) which is proposed by George and Foster (1994) based on the minimax principle, in which $\eta = 2\ln(pq + q)$, the diagnostic procedure (9) would perform reasonably well in most cases. The simulation results can be found in Sect. 4.

To end this subsection, we summarize the detailed steps for implementing the proposed method as follows.

Step 1. Collect the reference data and calibrate IC parameters according to the model (1).
Step 2. Obtain $\mathrm{E}(W_{t,\gamma_k})$ and $\mathrm{Var}(W_{t,\gamma_k})$ by simulation.
Step 3. Choose the desired IC ARL and the smoothing constant, λ. Determine the control limit, L, based on p, q, IC ARL and λ.
Step 4. Start monitoring the process and obtain profile observations sequentially. Whenever obtaining a new profile sample, use the LARS procedure and compute W_{t,γ_k}. Consequently, compute the plot statistic, Q_t, in (8) and compare it with control limit.
Step 5. Then, use (9) to identify the shifted parameters.
Step 6. After correctly identifying the out-of-control parameters and fixing the problem, we will then go back to *Step 3* to re-start the monitoring procedure.

4 Numerical performance comparison

We present some simulation results in this section regarding the numerical performance of the proposed LEWMA chart for Phase II analysis. It is challenging to compare the proposed method with alternative methods, since there is no obvious comparable method in the literature. Here, we consider the control chart based on the generalization of the MEWMA chart proposed by Zou et al. (2007) to multivariate profiles as an alternative method. By this approach, a signal is trigged when $\mathbf{z}_t^T \mathbf{\Omega}^{-1} \mathbf{z}_t$ exceeds the control limit. The combination of three charts proposed by Kim et al. (2003) is not involved because Zou et al. (2007a) has shown that the combination chart has similar performance to MEWMA for the simple linear model but is not directly applicable to more complex models, such as multiple regression models and multivariate profiles. Control limits of the MEWMA chart are determined by Markov chain method given in Zou et al. (2007a) to attain the nominal IC ARL under the standard normal error distribution, while the control limits used for LEWMA are found by the simulation method mentioned in Sect. 3.4. Since the zero-state and steady-state ARL (SSARL) comparison results are similar, only the OC SSARLs are provided. To evaluate the OC SSARL behavior of each chart, any series in which a signal occurs before the $(\tau + 1)$th observation is discarded (c.f., Hawkins and Olwell 1998). Because a similar conclusion holds for other cases, here we only present the results when IC ARL $= 500$, $\lambda = 0.05$ and $\tau = 50$ for illustration. All the ARL results in this section are obtained from 10,000 replications.

The number and variety of covariance matrices and shift directions are too large to allow a comprehensive, all-encompassing comparison. Our goal is to show the effectiveness, robustness and sensitivity of the LEWMA chart, and thus we only choose certain representative models for illustration. For this purpose, we consider the following profile model,

$$y_{ij} = \beta_{1k} + \beta_{2k} x_{i1} + \beta_{3k} x_{i2} + \beta_{4k} x_{i1}^2 + \varepsilon_{ij}, \quad i = 1, \dots, n, \ k = 1, \dots, q \tag{10}$$

where the design points $\mathbf{x}_i$ are generated from uniform distribution by centering so that their means are zero. we fix $n = 25$ and $q = 3$. Without loss of generality, we set

$$(\beta_{1k}, \beta_{2k}, \beta_{3k}, \beta_{4k}) = (0, 1, 2, 3),$$

Table 1 ARL comparisons of MEWMA and LEWMA for model (10) when there is one shifted parameter. The nominal IC ARL = 500 and $\lambda = 0.05$

Shift δ	$\beta_{11}+\delta$		Shift δ	$\beta_{21}+\delta$	
	MEWMA	LEWMA		MEWMA	LEWMA
0.05	162 (1.43)	135 (1.18)	0.05	404 (3.96)	387 (3.79)
0.10	48.4 (0.31)	37.5 (0.22)	0.10	286 (2.74)	253 (2.38)
0.20	16.8 (0.07)	13.5 (0.05)	0.20	114 (0.96)	88.1 (0.70)
0.40	7.23 (0.02)	6.06 (0.02)	0.40	33.9 (0.19)	25.7 (0.13)
0.60	4.74 (0.01)	4.03 (0.01)	0.60	18.4 (0.08)	14.5 (0.06)
0.80	3.59 (0.01)	3.09 (0.01)	0.80	12.5 (0.05)	10.2 (0.05)
1.00	2.93 (0.01)	2.54 (0.01)	1.00	9.65 (0.03)	7.90 (0.04)
1.50	2.08 (0.01)	1.85 (0.01)	1.50	6.16 (0.02)	5.12 (0.03)
2.00	1.71 (0.01)	1.48 (0.005)	2.00	4.60 (0.01)	3.85 (0.01)
Shift δ	$\beta_{41}+\delta$		Shift δ	$\delta\cdot\sigma_{(0)11}$	
	MEWMA	LEWMA		MEWMA	LEWMA
0.05	411 (4.07)	404 (3.94)	1.05	110 (0.85)	92.1 (0.71)
0.10	311 (3.05)	294 (2.84)	1.10	35.1 (0.18)	28.4 (0.15)
0.20	136 (1.20)	113 (0.95)	1.20	13.8 (0.05)	10.6 (0.04)
0.40	40.0 (0.24)	31.4 (0.16)	1.40	6.15 (0.02)	4.61 (0.01)
0.60	21.3 (0.10)	17.1 (0.07)	1.60	3.93 (0.01)	2.94 (0.01)
0.80	14.4 (0.06)	11.7 (0.05)	1.80	2.89 (0.01)	2.18 (0.01)
1.00	10.9 (0.04)	9.01 (0.03)	2.00	2.28 (0.01)	1.75 (0.01)
1.50	6.91 (0.02)	5.79 (0.02)	2.50	1.52 (0.005)	1.17 (0.004)
2.00	5.11 (0.01)	4.32 (0.01)	3.00	1.12 (0.003)	1.02 (0.002)

NOTE: Standard errors are in parentheses

for all k under IC model. This scenario, containing both multiple and polynomial regression terms, is quite common in practical applications. In the interest of brevity, the covariance matrix $\boldsymbol{\Sigma}_{(0)} = (\sigma_{(0)ij})$ is chosen to be $\sigma_{(0)ii} = 1$ and $\sigma_{(0)ij} = 0.5^{|i-j|}$, for $i, j = 1, 2, \ldots, q$.

The OC ARLs of the LEWMA and MEWMA charts for detecting shifts in coefficients β_{11}, β_{21}, β_{41} and variance σ_{11}^2 are presented in Table 1. It can be seen that the proposed LEWMA is almost uniformly superior to its alternative MEWMA in detecting any magnitude of shifts in all the cases considered in this table. Simultaneous shifts in two parameters in model (10) are considered in Table 2. The superiority of LEWMA to its counterpart holds when there are two shifted parameters simultaneously in both cases where the shifts come from one profile (β_{21} and β_{41}) and two profiles (β_{21} and β_{32}). In both tables, λ is fixed at 0.05 for both charts. To diminish the possible effect of λ on the performance of the two charts, we also compute the optimal OC ARLs with respect to λ (i.e., the smallest OC ARLs when λ changes) in the cases considered in Tables 1–2. The results are not reported here but available from the authors. They are slightly better than those in Tables 1–2, as expected; but, the conclusions made from Tables 1–2 are also valid.

We now evaluate the performance of the proposed diagnostic procedure (9). Simulation results in the same chosen representative settings of Tables 1 and 2 are tabulated in Table 3. In this table, the columns labelled "C" present relative frequencies of when the diagnostic procedures identified shifted parameters correctly, the columns labelled "U" present relative

Table 2 ARL comparisons of MEWMA and LEWMA for model (10) when there is two shifted parameter. The nominal IC ARL = 500 and $\lambda = 0.05$

Shifts		$\beta_{21}+\delta_1$ and $\beta_{41}+\delta_2$		$\beta_{21}+\delta_1$ and $\beta_{32}+\delta_2$	
δ_1	δ_2	MEWMA	LEWMA	MEWMA	LEWMA
0.100	0.100	207 (1.96)	195 (1.78)	197 (1.81)	182 (1.67)
0.100	0.200	108 (0.89)	95.3 (0.77)	94.7 (0.75)	80.7 (0.62)
0.100	0.400	37.4 (0.22)	30.3 (0.17)	32.0 (0.17)	26.0 (0.13)
0.100	0.600	20.7 (0.09)	16.9 (0.07)	18.1 (0.07)	14.7 (0.06)
0.200	0.100	97.8 (0.80)	82.4 (0.65)	96.0 (0.78)	81.4 (0.63)
0.200	0.200	67.2 (0.49)	59.6 (0.42)	62.8 (0.44)	55.6 (0.38)
0.200	0.400	32.2 (0.17)	27.6 (0.14)	28.3 (0.14)	24.4 (0.12)
0.200	0.600	19.4 (0.08)	16.4 (0.07)	17.2 (0.07)	14.4 (0.05)
0.400	0.100	33.5 (0.18)	26.1 (0.13)	33.5 (0.18)	26.2 (0.13)
0.400	0.200	29.6 (0.15)	24.6 (0.12)	29.2 (0.15)	24.5 (0.12)
0.400	0.400	21.5 (0.09)	19.4 (0.08)	20.2 (0.08)	18.3 (0.07)
0.400	0.600	15.8 (0.06)	14.2 (0.05)	14.5 (0.05)	13.0 (0.05)
0.600	0.100	18.6 (0.07)	14.9 (0.06)	18.6 (0.07)	14.9 (0.06)
0.600	0.200	17.7 (0.07)	14.5 (0.06)	17.6 (0.07)	14.6 (0.06)
0.600	0.400	15.1 (0.06)	13.3 (0.05)	14.8 (0.05)	13.2 (0.05)
0.600	0.600	12.6 (0.04)	11.4 (0.04)	11.8 (0.04)	10.9 (0.03)

NOTE: Standard errors are in parentheses

Table 3 Diagnostic results of the proposed procedure (9) for model (10) in various shifted cases. The nominal IC ARL = 500 and $\lambda = 0.05$

Shifts	$\beta_{11}+\delta$				$\beta_{31}+\delta$			
	C	U	O	I	C	U	O	I
$\delta = 0.20$	0.538	0.000	0.422	0.041	0.293	0.000	0.218	0.489
$\delta = 0.60$	0.595	0.000	0.391	0.015	0.439	0.000	0.174	0.386
$\delta = 1.00$	0.627	0.000	0.363	0.010	0.452	0.000	0.179	0.369
$\delta = 2.00$	0.684	0.000	0.311	0.005	0.466	0.000	0.179	0.355
Shifts	$\beta_{21}+\delta_1$ and $\beta_{41}+\delta_2$				$\beta_{21}+\delta_1$ and $\beta_{32}+\delta_2$			
	C	U	O	I	C	U	O	I
$\delta_1 = 0.40$; $\delta_2 = 0.40$	0.240	0.243	0.132	0.385	0.265	0.207	0.139	0.388
$\delta_1 = 0.40$; $\delta_2 = 0.60$	0.282	0.232	0.146	0.341	0.278	0.233	0.146	0.344
$\delta_1 = 0.60$; $\delta_2 = 0.40$	0.193	0.317	0.094	0.396	0.219	0.278	0.103	0.399
$\delta_1 = 0.60$; $\delta_2 = 0.60$	0.272	0.218	0.145	0.365	0.288	0.196	0.160	0.356

NOTE: The SE of the rate (π) in each entry, $\sqrt{\widehat{\pi}(1-\widehat{\pi})/10000}$, is typically less than 0.01

frequencies of when some shifted parameters are missed while all identified parameters are indeed OC parameters, the columns labelled "O" denote relative frequencies of when all shifted parameters are identified while some un-shifted parameters are also identified, and the columns labelled "I" denote relative frequencies of when at least one shifted parameter is missed and at least one identified shifted parameter is false. So, for a given diagnostic procedure, it performs better in a given case if its value in column "C" is comparatively larger and its value in column "I" is comparatively smaller. The results show that the proposed LASSO-based approach has reasonable diagnostic ability to identify shifted parameters in all the considered situations. After taking into account its computational advantage, we think that the LASSO-based approach provides a satisfactory diagnosis tool for multivariate profile diagnosis.

5 An illustration of the implementation steps: the loading process profile monitoring case revisited

In this section, we use the data from the logistic company example to illustrate the implementation steps. Two dependent variables are under consideration, y_1, the daily job frequency in the company's ports and y_2, the daily average time length to complete a job. The first dependent variable reflects the workload of the ports while the second reflects the efficiency. Five independent variables are used to explain variations in the dependent variables: x_1 is the daily loading amount of cargoes in the ports, x_2 is the daily discharging amount, x_3 is a dummy variable indicating whether the cargoes belong to the logistic company itself or not, x_4 is a dummy variable indicating whether the cargoes belong to a large-scale contractor or not, and x_5 is a dummy variable indicating whether the cargoes belong to a medium-scale contractor or not. Standard linear regression graphic analysis (Cook 1998) shows that a multivariate multiple regression model is sufficient to capture the relationships between y's and x's.

Here we collect about six months worth of data (the dataset is available from the authors upon request). We choose the first three months profile observations as the historical sample and the others for test. A calibration sample of this size may be smaller than ideal to determine fully the in-control model but it suffices to illustrate the use of the method in a real-world setting. By using the standard outlier detection method in linear regression analysis (Rousseuw and Leroy 1987), we removed a few influential outlying observations from the dataset and then obtained the following estimated IC model based on the LSE,

$$y_{i1} = -81.38 + 0.25x_{i1} + 0.26x_{i2} + 21.47x_{i3} + 98.19x_{i4} + 24.69x_{i5} + \varepsilon_{i1},$$
$$y_{i2} = 5.60 + 0.77x_{i1} - 0.86x_{i2} + 26.82x_{i3} + 22.32x_{i4} + 15.90x_{i5} + \varepsilon_{i2},$$

where $\boldsymbol{\Sigma}_{(0)} = \begin{pmatrix} 2.33 & 1.86 \\ 1.86 & 2.45 \end{pmatrix}$. This covariance matrix shows that the correlation between the two profiles is significant and therefore, using control charts that take this correlation into account should be more appropriate in this example than monitoring each profile individually.

We now apply the LEWMA chart to the remaining profiles of the dataset. In the LEWMA chart, we set $\lambda = 0.05$, and IC ARL $= 500$. By simulation, we can obtain the approximated

$$\begin{aligned} \mathrm{E}(W_{j,\gamma_k}) &= (3.96, 5.86, 7.44, 8.78, 9.92, 10.90, 11.72, \\ &\quad 12.40, 12.94, 13.36, 13.66, 13.87, 13.96, 14.00), \\ \mathrm{Var}^{1/2}(W_{j,\gamma_k}) &= (2.26, 3.00, 3.49, 3.86, 4.17, 4.43, 4.64, \\ &\quad 4.82, 4.97, 5.09, 5.17, 5.23, 5.27, 5.28). \end{aligned}$$

Then, its control limit is computed as $L = 4.281$. Figure 1 shows the resulting LEWMA chart (solid curve connecting the dots), along with its control limit (solid horizontal line).

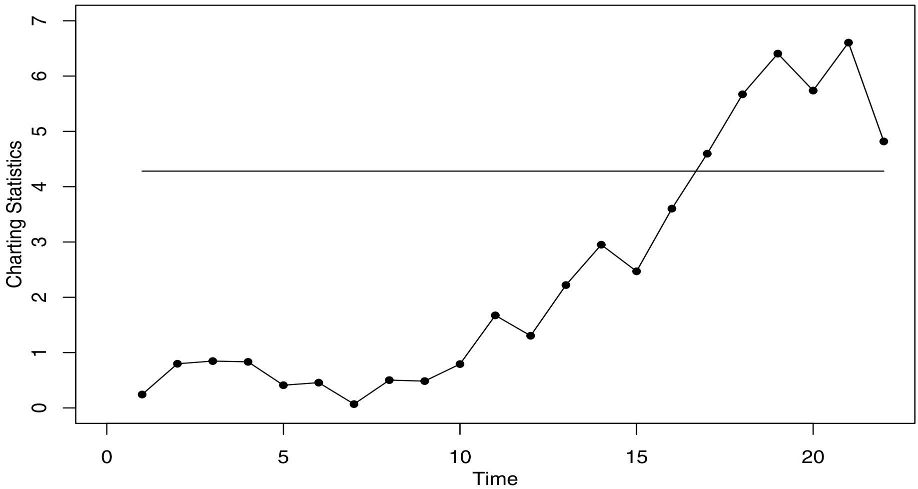

Fig. 1 The LEWMA control chart for monitoring the logistic loading process. The *solid horizontal line* indicates its control limit

The LEWMA chart signals at the 17th profile and remains above the control limits in the remainder of the sequence. Then, we use the proposed diagnostic procedure (9) to identify the shifted parameters and the corresponding values of $\frac{2-\lambda}{\lambda}(\mathbf{z}_{17} - \widehat{\boldsymbol{\mu}}_{\gamma_j})'\boldsymbol{\Omega}^{-1}(\mathbf{z}_{17} - \widehat{\boldsymbol{\mu}}_{\gamma_j}) + 2\ln(14)\cdot j$ are calculated. $\widehat{\boldsymbol{\mu}}_{\gamma^*}$ has twelve zero components but $(\widehat{\boldsymbol{\mu}}_{\gamma^*}^{(8)}, \widehat{\boldsymbol{\mu}}_{\gamma^*}^{(10)}) = (0.44, -2.35)$. These values indicate that the shift may have occurred in the relationships between y_2 and x_2 and that of y_2 and x_4 as well.

6 Concluding remarks

In practical applications, general multivariate linear profile monitoring is particularly useful due to its simplicity and flexibility. However, in such situations, the existing parametric profile monitoring methods suffer from a drawback in that when the profile parameter dimension is large, the detection ability of the procedures commonly used T^2-type charting statistics is likely to decline substantially. Moreover, it is also challenging to identify which parameter(s) have changed after an alarm is signaled in this high-dimensional case. Hence, there is a strong need for designing an efficient SPC scheme for multivariate profile monitoring and diagnosis. This paper develops a new methodology for monitoring general multivariate linear profiles, including the regression coefficients and profile variation. After examining the link between the parametric profile monitoring and multivariate statistical process control, we propose to apply a variable-selection-based multivariate control scheme, which was proposed by Zou and Qiu (2009), to the transformations of estimated profile parameters. Our proposed control chart is capable of determining the shift direction automatically based on observed profile data. Thus, it offers a balanced protection against various profile shifts. As a by-product, the proposed control chart provides an easy to use but quite effective diagnostic aid. A real-data example from manufacturing shows that it performs quite well in applications.

As we can expect, the performance of LEWMA is affected by the amount of data in the reference dataset. Thus, determination of the required Phase I sample sizes to remove the effects of estimated parameters and a general recommendation are needed. An ongoing effort of the authors is to develop a self-starting version of LEWMA that can handle sequential monitoring by simultaneously updating parameter estimates and checking for OC conditions and to compare it with the self-starting scheme designed under the normality assumption proposed by Zou et al. (2007b). Moreover, in the high-dimensional profile monitoring, monitoring all the components of the covariance matrix of the profile may be of practical interest. In fact, this may be handled by constructing a $[q(q+3)/2]$-dimensional working vector $\mathrm{Vec}(\widehat{\boldsymbol{\Sigma}}_j)$ and defining the corresponding $\mathbf{z}_{tV}$.

Acknowledgements The authors thank the editor, associate editor, and two anonymous referees for their many helpful comments that have resulted in significant improvements in the article. This work was completed when Zou was a Postdoctoral Fellow at the Hong Kong University of Science and Technology, whose hospitality is appreciated and acknowledged. This research was supported by the RGC Competitive Earmarked Research Grants 620508 and 620707 and NNSF of China Grants 10771107, 11001138 and 11071128.

Appendix: The asymptotic covariance matrix of $\widehat{s}_t$, $\boldsymbol{\Gamma}_q$, under in-control situation

As we know, the components of $\widehat{\mathbf{s}}_t + \sqrt{n-p}$ are proportional to the square-roots of the diagonal elements of the fitted residual covariance matrix $(\mathbf{Y}_t - \mathbf{X}\widehat{\mathbf{B}}_t)^T(\mathbf{Y}_t - \mathbf{X}\widehat{\mathbf{B}}_t)$ which is distributed as the Wishart distribution $W(\boldsymbol{\Sigma}_{(0)}, n-p)$. Thus, by Anderson (2003, Chap. 7), we have

$$\mathrm{Cov}(\sqrt{n-p}\cdot\widehat{\sigma}_{ti}, \sqrt{n-p}\cdot\widehat{\sigma}_{tj}) \approx \frac{\sigma^4_{(0)ij}}{\sigma_{(0)ii}\sigma_{(0)jj}},$$

where $\sigma^2_{(0)ij}$ is the (i,j)th element of $\boldsymbol{\Sigma}_{(0)}$.

References

Anderson, T. W. (2003). *An introduction of multivariate statistical analysis* (3rd ed.). New York: Wiley.
Chicken, E., Pignatiello, J. J. Jr., & Simpson, J. (2009). Statistical process monitoring of nonlinear profiles using wavelets. *Journal of Quality Technology*, *41*, 198–212.
Cook, R. D. (1998). *Regression graphics: ideas for studying regressions through graphics*. New York: Wiley.
Croisier, R. B. (1988). Multivariate generalizations of cumulative sum quality-control schemes. *Technometrics*, *30*, 243–251.
Ding, Y., Zeng, L., & Zhou, S. (2006). Phase I analysis for monitoring nonlinear profiles in manufacturing processes. *Journal of Quality Technology*, *38*, 199–216.
Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, *32*, 407–489.
George, E. I., & Foster, D. P. (1994). The risk inflation criterion for multiple regression. *The Annals of Statistics*, *22*, 1947–1975.
Hawkins, D. M. (1991). Multivariate quality control based on regression-adjusted variables. *Technometrics*, *33*, 61–75.
Hawkins, D. M., & Olwell, D. H. (1998). *Cumulative sum charts and charting for quality improvement*. New York: Springer.
Jensen, W. A., Birch, J. B., & Woodall, W. H. (2008). Monitoring correlation within linear profiles using mixed models. *Journal of Quality Technology*, *40*, 167–183.
Kang, L., & Albin, S. L. (2000). On-line monitoring when the process yields a linear profile. *Journal of Quality Technology*, *32*, 418–426.
Kazemzadeh, R. B., Noorossana, R., & Amiri, A. (2008). Phase I monitoring of polynomial profiles. *Communications in Statistics: Theory and Methods*, *37*, 1671–1686.
Kim, K., Mahmoud, M. A., & Woodall, W. H. (2003). On the monitoring of linear profiles. *Journal of Quality Technology*, *35*, 317–328.

Lowry, C. A., Woodall, W. H., Champ, C. W., & Rigdon, S. E. (1992). Multivariate exponentially weighted moving average control chart. *Technometrics*, *34*, 46–53.

Lucas, J. M., & Saccucci, M. S. (1990). Exponentially weighted moving average control scheme properties and enhancements. *Technometrics*, *32*, 1–29.

Mahmoud, M. A. (2008). Phase I analysis of multiple linear regression profiles. *Communications in Statistics: Simulation and Computation*, *37*, 2106–2130.

Mahmoud, M. A., & Woodall, W. H. (2004). Phase I analysis of linear profiles with calibration applications. *Technometrics*, *46*, 380–391.

Mason, R. L., & Young, J. C. (2002). *Multivariate statistical process control with industrial application*. Philadelphia: SIAM.

Montgomery, D. C. (2005). *Introduction to statistical quality control* (6th ed.). New York: Wiley.

Noorossana, R., Eyvazian, M., Amiri, A., & Mahmoud, M. A. (2010). Statistical monitoring of multivariate multiple linear regression profiles in Phase I with calibration application. *Quality and Reliability Engineering International*, 291–303.

Qiu, P. (2008). Distribution-free multivariate process control based on log-linear modeling. *IIE Transactions*, *40*, 664–677.

Qiu, P., & Zou, C. (2010). Control chart for monitoring nonparametric profiles with arbitrary design. *Statistica Sinica*, 1655–1682.

Qiu, P., Zou, C., & Wang, Z. (2010). Nonparametric profile monitoring by mixed effects modeling (with discussions). *Technometrics*, *52*, 265–277.

Rousseuw, P. J., & Leroy, A. M. (1987). *Robust regression and outlier detection*. New York: Wiley.

Stoumbos, Z. G., & Sullivan, J. H. (2002). Robustness to non-normality of the multivariate EWMA control chart. *Journal of Quality Technology*, *34*, 260–276.

Sullivan, J. H., Stoumbos, Z. G., Mason, R. L., & Young, J. C. (2007). Step-down analysis for changes in the covariance matrix and other parameters. *Journal of Quality Technology*, *39*, 66–84.

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

Wang, H., Li, R., & Tsai, C. L. (2007). On the consistency of SCAD tuning parameter selector. *Biometrika*, *94*, 553–568.

Williams, J. D., Woodall, W. H., & Birch, J. B. (2007). Statistical monitoring of nonlinear product and process quality profiles. *Quality and Reliability Engineering International*, *23*, 925–941.

Woodall, W. H. (2007). Current research on profile monitoring. *Revista Producão*, *17*, 420–425.

Woodall, W. H., Spitzner, D. J., Montgomery, D. C., & Gupta, S. (2004). Using control charts to monitor process and product quality profiles. *Journal of Quality Technology*, *36*, 309–320.

Yuan, M., Ekici, A., Lu, Z., & Monteiro, R. (2007). Dimension reduction and coefficient estimation in multivariate linear regression. *Journal of the Royal Statistical Society, Series B*, *69*, 329–346.

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

Zou, C., Zhang, Y., & Wang, Z. (2006). Control chart based on change-point model for monitoring linear profiles. *IIE Transactions*, *38*, 1093–1103.

Zou, C., Tsung, F., & Wang, Z. (2007a). Monitoring general linear profiles using multivariate EWMA schemes. *Technometrics*, *49*, 395–408.

Zou, C., Zhou, C., Wang, Z., & Tsung, F. (2007b). A self-starting control chart for linear profiles. *Journal of Quality Technology*, *39*, 364–375.

Zou, C., Tsung, F., & Wang, Z. (2008). Monitoring profiles based on nonparametric regression methods. *Technometrics*, *50*, 512–526.

Zou, C., Qiu, P., & Hawkins, D. (2009). Nonparametric control chart for monitoring profiles using the change point formulation. *Statistica Sinica*, *19*, 1337–1357.

Zou, H. (2006). The adaptive LASSO and its oracle properties. *Journal of the American Statistical Association*, *101*, 1418–1429.

Zou, H., Hastie, T., & Tibshirani, R. (2007). On the 'degrees of freedom' of LASSO. *The Annals of Statistics*, *35*, 2173–2192.