IIE Transactions (2012) **44**, 1–15

ISSN: 0740-817X print / 1545-8830 online
DOI: 10.1080/0740817X.2011.649386

A distribution-free robust method for monitoring linear profiles using rank-based regression

XUEMIN ZI[1], CHANGLIANG ZOU[2,*], and FUGEE TSUNG[3]

[1]*School of Science, Tianjin University of Technology and Education, Tianjin, China*
[2]*LPMC and Department of Statistics, School of Mathematical Sciences, Nankai University, Tianjin, China*
E-mail: chlzou@yahoo.com.cn
[3]*Department of Industrial Engineering and Logistics Management, Hong Kong University of Science and Technology, Hong Kong*

Received September 2010 and accepted November 2011

Profile monitoring is a technique for checking the stability of the functional relationship between a response variable and one or more explanatory variables over time. Linear profile monitoring is particularly useful in practice due to its simplicity and flexibility. The existing monitoring methods suffer from a drawback in that they all assume the error distribution to be normal. When the underlying distribution is misspecified, the efficiency of the commonly used Least Squares Estimation (LSE) is likely to be low and as a consequence the detection ability of procedures based on LSE is reduced. To overcome this drawback, this article develops a non-parametric methodology for monitoring the linear profile, including the regression coefficients and profile variations. The Multivariate Sign exponentially weighted moving average (MSEWMA) control scheme is applied to the estimated profile parameters obtained using a rank-based regression approach. Benefiting from certain favorable properties of MSEWMA and the efficiency of rank-based regression estimators, the proposed chart is robust from the point of view of the in-control and out-of-control average run length, particularly when the process distribution is heavily tailed. An example with real data from a manufacturing facility shows that it performs well in application.

Keywords: Heavily-tailed distributions, leptokurtic distributions, MEWMA, MSEWMA, non-parametric procedure, profile monitoring, sign-based EWMA, Wilcoxon rank estimator

1. Introduction

Due to recent progress in sensing and information technology, automated data acquisition has been widely adopted in various industries. Consequently, large amounts of quality-related data have become available. Statistical Process Control (SPC) based on such data is an important component of process monitoring and control. In many applications, the quality of a process is characterized by the relationship between a response variable and one or more explanatory variables. A collection of data points of these variables can be observed at each sampling stage, which can be represented by a curve (i.e., a profile). In some calibration applications, the profile can be described adequately by a linear regression model. In other applications, however, more flexible models are necessary if the profiles are to be described properly. An extensive discussion of the related research problems has been given by Woodall *et al.* (2004).

Studies focusing on simple linear profiles have flourished. See, for instance, Kang and Albin (2000), Kim *et al.* (2003), Mahmoud and Woodall (2004), Zou *et al.* (2006), and Mahmoud *et al.* (2007), among several others. Multiple and polynomial regression profile models are considered by Zou, Tsung, and Wang (2007); Kazemzadeh *et al.* (2008); Mahmoud (2008); and Jensen *et al.* (2008). Non-linear profile models are investigated by Williams *et al.* (2007), Jensen and Birch (2009), Zhang and Albin (2009), etc. Yeh *et al.* (2009) studied Phase I profile monitoring for binary responses that could be represented by the logistic regression model. Recently, profile monitoring for the general profile model has also attracted considerable attention. See Zou *et al.* (2008) and Qiu *et al.* (2010) for Phase II methods based on non-parametric regression and Lada *et al.* (2002), 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).

This article focuses on monitoring linear profiles parametrically over time. In particular, we consider the linear profile model given by Zou, Tsung, and Wang (2007), and concentrate on Phase II. Most of the aforementioned work on linear profile monitoring relies on an unambiguous

*Corresponding author

specification of the error distribution, e.g., a normal distribution, and the corresponding control schemes are constructed based on integrating conventional SPC procedures with the Least Squares Estimate (LSEs) or, equivalently, Maximum Likelihood Estimate (MLEs) of the parameters of each profile. While the normal assumption and LSE are useful and popular in applications, questions have arisen about the appropriateness of this assumption and about the potential impact of misspecifying the distribution in statistical analyses. In the literature of non-parametric univariate or multivariate SPC (see Chakraborti *et al.* (2001), Qiu and Hawkins (2001), Qiu (2008), Zhou *et al.* (2009), Zou and Tsung (2010) and the references therein), it is well recognized that the underlying process distribution in many applications is not normal, so that the statistical properties of commonly used charts, which perform best under the normal distribution, could potentially be (highly) affected. This is also true in the present profile monitoring problem.

The effects of misspecifying the error distribution on profile monitoring are two-fold. On one hand, when the underlying probability distribution is misspecified, LSE is likely to be inefficient (see Hettmansperger and McKean (2010)). Consequently, the detection ability of the mentioned procedures based on LSE would be reduced in certain situations (see the simulation results in Tables 3 and 4 in Section 3). In such situations, robust linear regression will be particularly useful and desirable. On the other hand, the control charts for monitoring the estimated profile parameters, designed under the normality assumption, usually take quadratic forms (the Multivariate exponentially weighted moving average (MEWMA) chart; Zou, Tsung, and Wang (2007)) or adopt the max operator (multi-charts; Kim *et al.* (2003)) of the related parameter estimates. In heavy-tailed situations, these control charts would result in some biases from the nominal value of the In-Control (IC) Average Run Length (ARL); i.e., the IC ARL tends to be smaller than intended if the smoothing parameter is not chosen appropriately (see Table 1 in Section 3).

It should be emphasized that the issue of not being able to attain the nominal IC ARL is actually inherited from the conventional multivariate control charts when the multinormality assumption is violated because parametric profile monitoring can be regarded as a special case of multivariate SPC to some extent (Woodall, 2007; Zou, Tsung, and Wang, 2007). Therefore, it is natural to consider certain non-parametric multivariate schemes instead of the MEWMA chart. However, Stoumbos and Sullivan (2002) argued that multivariate non-parametric control charts "are less powerful, more computationally intensive, and generally do not apply to skewed distributions." As a solution, Zou and Tsung (2011) recently proposed a new multivariate control chart, called the multivariate sign EWMA (or simply MSEWMA), for monitoring location parameters. The MSEWMA adapts a powerful multivariate sign test proposed by Randles (2000) to online sequential monitoring and can be seen as a non-parametric counterpart of the classic MEWMA. It is efficient in detecting shifts in location parameters, particularly small or moderate shifts when the process distribution is heavy-tailed or skewed. More important, it is appealing in that for a broad class of distributions, its IC run length distribution can reach (or is always very close to) the nominal one when the same control limit designed for a multinormal distribution is used.

Here we develop a non-parametric SPC methodology for monitoring linear profiles, including the regression coefficients and profile variations. We apply MSEWMA to the estimated profile parameters that are obtained by using a rank-based regression approach. Due to certain favorable properties of MSEWMA and the efficiency of rank-based regression estimators, the proposed chart is robust from both the IC and out-of-control (OC) ARLs' point of view, particularly when the process distribution is very heavy-tailed. The remainder of this article is organized as follows: our proposed methodology is described in detail in Section 2 and its numerical performance is investigated in Section 3. The method is demonstrated using a real example in Section 4 and several remarks conclude the article in Section 5. Some technical details are provided in the Appendix. A supplemental file, which contains several other results and dataset, can be found on the webpage of the corresponding author.

2. Methodology

We describe the proposed control chart in four parts. In Section 2.1, the linear model formulation and the associated assumptions are introduced. Then, a brief introduction of the rank-based regression is presented in Section 2.2. A new EWMA control chart combined with rank-based estimation is derived in Section 2.3. Practical guidelines for design and computation are given in Section 2.4.

2.1. *Model and assumption*

Consider the linear profile model given by Zou, Tsung, and Wang (2007). Assume that for the jth ($j \geq 1$) random sample collected over time we have the observations $(\mathbf{X}_j, \mathbf{y}_j)$, where $\mathbf{y}_j = (y_{j1}, \ldots, y_{jn_j})$ is an n_j-variate response vector and $\mathbf{X}_j$ is an $n_j \times p$ regressor matrix. It is assumed that the profile is a linear combination of the covariates without time delay, i.e., the process observations are collected over time using the following profile model:

$$y_{ji} = \alpha_j + \mathbf{x}_{ji}^{\mathrm{T}}\boldsymbol{\beta}_j + \varepsilon_{ji}, \quad i = 1, \cdots, n_j; \quad j = -m_0 + 1, \ldots, 0, 1, \ldots, \tau, \tau + 1, \ldots, \qquad (1)$$

where τ is the unknown change-point, $\{(\mathbf{X}_j, \mathbf{y}_j)\}_{j=-m_0+1}^{0}$ denotes the historical dataset, $\mathbf{x}_{ji}^{\mathrm{T}}$ denotes the ith row of $\mathbf{X}_j$, α_j is the intercept parameter, $\boldsymbol{\beta}_j$ is a p-dimensional coefficient vector, and the ε_{ji} are independent and identically distributed (i.i.d.) from a distribution F (its density is

denoted as f) and satisfy $E[\varepsilon_{ji}] = 0$, and $\sigma_j^2 = \mathrm{Var}[\varepsilon_{ji}] < \infty$. Let us suppose that:

$$\alpha_j = \alpha_{(0)}, \boldsymbol{\beta}_j = \boldsymbol{\beta}_{(0)}, \sigma_j^2 = \sigma_{(0)}^2 \quad \text{for} \quad j \le \tau,$$
$$\alpha_j = \alpha_{(1)}, \boldsymbol{\beta}_j = \boldsymbol{\beta}_{(1)}, \sigma_j^2 = \sigma_{(1)}^2 \quad \text{for} \quad j > \tau,$$

and $\alpha_{(0)} \neq \alpha_{(1)}$ and/or $\boldsymbol{\beta}_{(0)} \neq \boldsymbol{\beta}_{(1)}$ and/or $\sigma_{(0)}^2 \neq \sigma_{(1)}^2$, where the parameters with subscripts (0) and (1) denote the IC and OC profile parameters, respectively.

In what follows, for ease of exposition, the n_js are taken to be equal (denoted as n), and the explanatory variable matrix, $\mathbf{X}_j$, is assumed to be fixed for different values of j (denoted as $\mathbf{X}$). This is usually the case in practical calibration applications in industrial manufacturing and is also consistent with the literature, such as Kim *et al.* (2003) and Zou, Tsung, and Wang (2007, 2008). For profile monitoring with more complex covariate designs, see Qiu *et al.* (2010). Without loss of generality, let us suppose that all of the columns of $\mathbf{X}$ are orthogonal to $\mathbf{1}$ where $\mathbf{1}$ is an n-variate vector of all ones. We can also obtain this form through some appropriate transformations. We mainly consider the Phase II case in which the necessary IC parameters are assumed to be known, as is a common convention in the literature. It is essentially equivalent to saying that m_0 is sufficiently large. Once the IC models are established as the baseline in Phase II, we would want to detect any change in the regression coefficients and the profile variance as quickly as possible. In a Phase I study, although the methods based on the normal assumption may not be the most efficient, they can still be applied and usually work just as well. Interested readers can refer to Jensen *et al.* (2008), Kazemzadeh *et al.* (2008), and Mahmoud (2008).

2.2. *A brief review of rank-based regression*

For notational convenience, we choose to suppress the index j in the notation of Equation (1) and consider only one fixed profile sample to illustrate the method. It is well understood that the LSE

$$\widehat{\boldsymbol{\beta}}_{\mathrm{L}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left(y_i - \alpha - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)^2, \tag{2}$$

is sensitive to outliers and is less efficient if the error distribution has heavier tails than the normal distribution. In particular, the efficiency of the LSE is zero if the error distribution is Cauchy (Hettmansperger and McKean, 2010). On the other hand, formulation of an efficient MLE requires estimating the closed form of f, which creates extra technical difficulties and inconvenience in practice.

A natural alternative seems to be the Least Absolute Deviation Estimator (LADE; Bloomfield and Steiger, 1983):

$$\widehat{\boldsymbol{\beta}}_{A} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left|y_i - \alpha - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right|,$$

which may be more robust because it is resistant to heavy-tailed errors and/or extreme observations (outliers). Nevertheless, the asymptotic efficiency of LADE compared with the LSE is proportional to the density at the median (Pollard, 1991); that is,

$$\mathrm{ARE}(\widehat{\boldsymbol{\beta}}_{\mathrm{A}}, \widehat{\boldsymbol{\beta}}_{\mathrm{L}}) = 4\sigma^2 f^2(\theta),$$

where $\mathrm{ARE}(E_1, E_2)$ denotes the Asymptotic Relative Efficiency (ARE) of estimator E_1 with respect to E_2 and θ is the median of f. For the normal distribution, this quantity is only 0.637. And worse still, the efficiency is arbitrarily small if the density value at the median is close to zero. Please refer to Section 1.5 of Hettmansperger and McKean (2010) for a thorough discussion and some finite-sample analysis.

To overcome the inefficiency issue of LADE, a commonly used method is the so-called Wilcoxon-type rank-based estimate, $\widehat{\boldsymbol{\beta}}_{\mathrm{R}}$, which minimizes:

$$W_n(\boldsymbol{\beta}) := \sum_{i=1}^{n} \left\{ \frac{R\left(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)}{n+1} - \frac{1}{2} \right\} \left(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right), \tag{3}$$

where $R(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})$ denotes the rank of $y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}$ among $\{y_1 - \mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}, \ldots, y_n - \mathbf{x}_n^{\mathrm{T}}\boldsymbol{\beta}\}$. This estimator, which was proposed by Jaeckel (1972), is asymptotically equivalent to the rank estimator of Jureăková (1971). Clearly, the above minimization is analogous to the least squares procedure except that the Euclidean norm is substituted by a Wilcoxon-type rank norm. Moreover, the objective function (3) is a nonnegative convex function and provides a robust measure of the dispersion of the residuals (see Theorems 2.5.1 on page 99 of Hettmansperger and McKean (2010)). McKean and Schrader (1980) further revealed an intuitive geometric interpretation of Equation (3).

Under some mild assumptions, the Wilcoxon rank estimator has an approximate normal distribution (see Theorem 3.5.4 and Corollary 3.5.1 of Hettmansperger and McKean (2010)):

$$\widehat{\boldsymbol{\beta}}_{\mathrm{R}} \to N_p(\boldsymbol{\beta}, \eta^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}),$$

where $\eta^2 = 1/[12(\int f^2(t)\mathrm{d}t)^2]$. Consequently, its ARE with respect to LSE is $\mathrm{ARE}(\widehat{\boldsymbol{\beta}}_{\mathrm{R}}, \widehat{\boldsymbol{\beta}}_{\mathrm{L}}) = \sigma^2/\eta^2$. Note that this ARE is closely related to that of the signed-rank Wilcoxon test with respect to the t-test in the context of classic nonparametric statistics. Hence, $\widehat{\boldsymbol{\beta}}_{\mathrm{R}}$ is quite robust and highly efficient in both testing and estimation when the error distribution is heavy-tailed. For instance, it has a lower bound of 0.864 among all symmetric distributions (see Hettmansperger and McKean (2010, Section 1.7.2)). For the normal error distribution, this quantity is 0.955. For a comprehensive presentation of the rank-based analysis of linear models, see Hettmansperger and McKean (2010). In our profile monitoring problem, we are inclined to consider the Wilcoxon Rank Estimator (WRE) based on the above reasoning.

We emphasize here that the WRE belongs to the family of general rank-norm estimators (see Hettmansperger and McKean (2010, Section 3)), which is defined as

$$\arg\min \sum_{i=1}^{n} a\big(R(y_i - \alpha - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})\big)(y_i - \alpha - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}),$$

where $a(1) \le a(2) \le \cdots \le a(n)$ is a set of scores generated as $a(i) = \psi(i/(n+1))$, satisfying $\sum_{i=1}^{n} a(i) = 0$, for some non-decreasing score function $\psi(u)$ defined on the interval $(0, 1)$ and standardized such that $\int \psi(u)\mathrm{d}u = 0$ and $\int \psi^2(u)\mathrm{d}u = 1$. The Wilcoxon norm is generated by the linear score function $\psi(u) = \sqrt{12}(u - 1/2)$. It can also be seen that using $\psi(u) = \mathrm{sign}(u - 1/2)$ leads to the LADE. With this general rank-norm loss function, by choosing some appropriate $\psi(\cdot)$, we may define various estimators with better efficiencies for some specific error distributions, such as heavy-tailed distributions. These estimators are certainly suitable candidates other than the WRE for tackling the present profile monitoring problem. However, as demonstrated in the literature, the WRE is one of the most popular and efficient rank-norm estimators and has been shown to perform remarkably well in a wide variety of settings; e.g., see Leng (2010) for variable selection and Wang *et al*. (2009) for non-parametric regression. Thus, we adopt the WRE here, but further research on using some other rank-norm estimators would be useful.

It is worth pointing out that the intercept term α is not involved in Equation (3). This is because α does not affect the value of the Wilcoxon-type rank-based loss function when we use $y_i - \alpha - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}$ instead due to

$$\sum_{i=1}^{n} \left\{ \frac{R\big(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\big)}{n+1} - \frac{1}{2} \right\} = 0.$$

Hence, the Wilcoxon-type rank-based method cannot be used to estimate the intercept. In contrast, the LSE and LADE methods are able to estimate intercept and regression coefficients simultaneously. Thus, the WRE should be used together with some other estimate of α, which will be discussed later.

2.3. *Control charts for monitoring the linear profile*

To monitor the linear profile (1), the p regression coefficients, the intercept, and the standard deviations σ must be controlled simultaneously. For the jth profile collected over time, denote $\widehat{\boldsymbol{\beta}}_{j\mathrm{L}}$ and $\widehat{\boldsymbol{\beta}}_{j\mathrm{R}}$ as the corresponding LSE and WRE for the regression coefficient using Equations (2) and (3), respectively. Then, we could use

$$\widehat{\alpha}_j = \frac{1}{n}\sum_{i=1}^{n} y_{ji}, \quad \widehat{\sigma}^2_{j\mathrm{R}} = \frac{1}{n-p-1}\sum_{i=1}^{n}\big(y_{ji} - \widehat{\alpha}_j - \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{j\mathrm{R}}\big)^2,$$

as the estimates of the intercept α and variance σ^2, respectively. When using $\widehat{\boldsymbol{\beta}}_{j\mathrm{L}}$ in the above equation instead of $\widehat{\boldsymbol{\beta}}_{j\mathrm{R}}$, we obtain the corresponding $\widehat{\sigma}^2_{j\mathrm{L}}$. By Theorem 3.9.5 of Hettmansperger and McKean (2010), we know that both $\widehat{\sigma}^2_{j\mathrm{R}}$ and $\widehat{\sigma}^2_{j\mathrm{L}}$ are consistent estimators of σ^2. Note that the median, which may be more robust in certain cases, could be used to estimate the intercept. Nevertheless, its asymptotic efficiency compared with the mean is the same as $\mathrm{ARE}(\widehat{\boldsymbol{\beta}}_{\mathrm{A}}, \widehat{\boldsymbol{\beta}}_{\mathrm{L}})$. Thus, using the median instead of the mean would make charts largely inefficient in detecting the change of intercept except for very heavy-tailed distributions and we choose to use the sample mean here.

Zou, Tsung, and Wang (2007) defined a working vector,

$$\mathbf{z}_{j\mathrm{L}} = \big[\widehat{\alpha}_j, \widehat{\boldsymbol{\beta}}_{j\mathrm{L}}, \Phi^{-1}\big\{\Psi\big((n-p-1)\widehat{\sigma}^2_{j\mathrm{L}}/\sigma^2_{(0)}; n-p-1\big)\big\}\big]^{\mathrm{T}},$$

where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal cumulative distribution function and $\Psi(\cdot;\nu)$ is the chi-square distribution function with ν degrees of freedom (χ^2_ν). When the process is IC and the ε_{ji} values are normally distributed, the vector is multivariate normally distributed with mean $(\alpha_{(0)}, \boldsymbol{\beta}_{(0)}, 0)^{\mathrm{T}}$ and covariance matrix $\boldsymbol{\Omega} = \mathrm{diag}\{n^{-1}\sigma^2_{(0)}, \sigma^2_{(0)}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}, 1\}$. Zou, Tsung, and Wang (2007) applied the MEWMA chart to $\mathbf{z}_{j\mathrm{L}}$, which we denote as the MEWMA-L chart hereafter. The advantage of transforming $\widehat{\sigma}^2_{j\mathrm{L}}$ into a normal variable is that the control chart can be made sensitive to decreases in the standard deviation as well. However, when the ε_{ji} values are heavy-tailed, this transformation would yield a variable whose distribution will deviate greatly from normality since $\widehat{\sigma}^2_{j\mathrm{L}}$ is no longer chi-square distributed. The above covariance matrix $\boldsymbol{\Omega}$ is also inappropriate in such situations.

Even though $\mathbf{z}_{j\mathrm{L}}$ is not normally distributed, the MEWMA is still applicable after the mean and covariance of $\mathbf{z}_{j\mathrm{L}}$, denoted as $\boldsymbol{\mu}_{(0)}$ and $\boldsymbol{\Sigma}_{(0)}$, are estimated from the historical dataset. As demonstrated by Stoumbos and Sullivan (2002), the MEWMA chart would still be more appealing than some multivariate non-parametric schemes because MEWMA charts are quite robust when the weighting parameter λ is sufficiently small. 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. In fact, the IC run length distribution for a continuous non-normal process is quite close to the distribution for a multivariate normal process with the same control limit. However, how small λ should be depends on the deviation of the actual measurement distribution from the multinormal distribution, which may not be easy to measure in practice.

Alternatively, we consider applying the MSEWMA chart proposed by Zou and Tsung (2011) to the estimated parameter vector:

$$\mathbf{z}_{j\mathrm{R}} = [\widehat{\alpha}_j, \widehat{\boldsymbol{\beta}}_{j\mathrm{R}}, \widehat{\sigma}^2_{j\mathrm{R}}]^{\mathrm{T}}.$$

As illustrated in Section 2.2, $\widehat{\boldsymbol{\beta}}_{j\mathrm{R}}$ is more robust than $\widehat{\boldsymbol{\beta}}_{j\mathrm{L}}$ for some error distributions, especially for heavy-tailed

distributions. In turn, we may expect the use of $\mathbf{z}_{j\mathrm{R}}$ to yield a more efficient test than the use of $\mathbf{z}_{j\mathrm{L}}$ if the error distribution is very heavy-tailed. Note that $\mathbf{z}_{j\mathrm{R}}$ is not exactly multinormally distributed in a small-sample situation, even when the error distribution is normal. Hence, the MSEWMA chart is employed to deal with its non-normality.

Next, we present a succinct description of how to construct the MSEWMA chart based on $\mathbf{z}_{j\mathrm{R}}$, but readers are advised to refer to Zou and Tsung (2011) for more theoretical and empirical justifications about the efficiency of this chart in multivariate settings. The proposed control scheme contains two steps. The first step is to obtain an IC multivariate median, $\boldsymbol{\theta}_{(0)}$, and a corresponding IC transformation matrix, $\boldsymbol{\Gamma}_{(0)}$, of the vector $\mathbf{z}_{j\mathrm{R}}$. This step is similar to constructing the MEWMA-L chart in which $\boldsymbol{\mu}_{(0)}$ and $\boldsymbol{\Sigma}_{(0)}$ for $\mathbf{z}_{j\mathrm{L}}$ are usually estimated from the historical data before monitoring. We consider using the affine equivariant multivariate median and the associated transformation matrix proposed by Hettmansperger and Randles (2002), given as the solutions of the following equations:

$$E\left(\frac{\boldsymbol{\Gamma}(\mathbf{z}-\boldsymbol{\theta})}{||\boldsymbol{\Gamma}(\mathbf{z}-\boldsymbol{\theta})||}\right)=0,\quad E\left(\frac{\boldsymbol{\Gamma}(\mathbf{z}-\boldsymbol{\theta})(\mathbf{z}-\boldsymbol{\theta})^{\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}}}{||\boldsymbol{\Gamma}(\mathbf{z}-\boldsymbol{\theta})||^2}\right)=\frac{1}{p+2}\mathbf{I}_{p+2}, \tag{4}$$

where $\mathbf{z}$ denotes the random vector, which has the same distribution as $\mathbf{z}_{j\mathrm{R}}$, and $\boldsymbol{\Gamma}$ is a $(p+2)\times(p+2)$ upper triangular positive-definite matrix with a one in the upper left-hand element. In practice, $\boldsymbol{\theta}_{(0)}$ and $\boldsymbol{\Gamma}_{(0)}$ are estimated from the IC historical profile samples, defined as the solutions of the following sample equations:

$$\frac{1}{m_0}\sum_{j=-m_0+1}^{0}\left(\frac{\boldsymbol{\Gamma}(\mathbf{z}_{j\mathrm{R}}-\boldsymbol{\theta})}{||\boldsymbol{\Gamma}(\mathbf{z}_{j\mathrm{R}}-\boldsymbol{\theta})||}\right)=\mathbf{0}, \tag{5}$$

$$\frac{1}{m_0}\sum_{j=-m_0+1}^{0}\left(\frac{\boldsymbol{\Gamma}(\mathbf{z}_{j\mathrm{R}}-\boldsymbol{\theta})(\mathbf{z}_{j\mathrm{R}}-\boldsymbol{\theta})^{\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}}}{||\boldsymbol{\Gamma}(\mathbf{z}_{j\mathrm{R}}-\boldsymbol{\theta})||^2}\right)=\frac{1}{p+2}\mathbf{I}_{p+2}. \tag{6}$$

Hettmansperger and Randles (2002) presented an iterative procedure to solve Equations (5) and (6) simultaneously that is efficient in obtaining $(\boldsymbol{\theta}_{(0)}, \boldsymbol{\Gamma}_{(0)})$ from a given sample. A detailed step-by-step description of the algorithm can also be found in the Appendix of Zou and Tsung (2011).

After $(\boldsymbol{\theta}_{(0)}, \boldsymbol{\Gamma}_{(0)})$ is specified or estimated, we standardize and transform our online estimated parameter vector $\mathbf{z}_j$ from the jth collected profile to obtain the unit vector $\mathbf{v}_j$ through:

$$\mathbf{v}_j=\frac{\boldsymbol{\Gamma}_{(0)}(\mathbf{z}_{j\mathrm{R}}-\boldsymbol{\theta}_{(0)})}{||\boldsymbol{\Gamma}_{(0)}(\mathbf{z}_{j\mathrm{R}}-\boldsymbol{\theta}_{(0)})||}. \tag{7}$$

The unit vectors of the transformed data have a variance–covariance structure like the one for a random variable that is uniform on the unit p-sphere. Define an EWMA sequence in a similar way to Lowry *et al.* (1992) and Zou, Tsung, and Wang (2007),

$$\mathbf{w}_j=(1-\lambda)\mathbf{w}_{j-1}+\lambda\mathbf{v}_j, \tag{8}$$

where $\boldsymbol{w}_0=\mathbf{0}$. Finally, the proposed control chart triggers a signal if

$$Q_j=\frac{2-\lambda}{\lambda}p\mathbf{w}_j^{\mathrm{T}}\mathbf{w}_j>L, \tag{9}$$

where $L>0$ is a control limit chosen to achieve a specific IC run length distribution. We also make use of the fact that $\mathrm{Cov}(\mathbf{w}_j)\approx\lambda\mathrm{Cov}(\mathbf{v}_j)/(2-\lambda)=p^{-1}\lambda\mathbf{I}_p/(2-\lambda)$ by Equation (4).

As has been shown by Zou and Tsung (2011), the MSEWMA chart is distribution free in the sense that its IC run length distribution is the same for all distributions with elliptical directions in which random variables are generated via $\mathbf{z}_i=r_i\mathbf{D}\mathbf{u}_i$, where the $\mathbf{u}_i$ are i.i.d. uniform on the unit p sphere, $\mathbf{D}$ is a non-singular matrix, and the r_i are positive scalars. The elliptical directions family contains all of the elliptically symmetric distributions, such as multinormal and multivariate t distributions and certain skewed distributions. Moreover, its IC run-length performance is quite robust under various process distributions, including very skewed distributions. We can expect these properties to remain (approximately) valid in the present profile monitoring example. We shall call the MSEWMA chart that uses $\mathbf{z}_{j\mathrm{R}}$ the MSEWMA-R chart. Although its exact charting performance is difficult to derive because the small-sample distributional properties of $\mathbf{z}_{j\mathrm{R}}$ are not available to us, the IC and OC performance of MSEWMA-R will be assessed in Section 3 to verify its effectiveness.

2.4. *Design and implementation of the proposed schemes*

2.4.1. *On computation*

For online detection, after obtaining the vectors of estimated parameters from each profile, the computation burden of the MSEWMA-R chart is similar to that of the MEWMA-L chart since both only require computing a working EWMA sequence and a quadratic form. In comparison with LSE, the MSEWMA-R chart is more complicated because the minimization of Equation (3) has no closed-form solution. The WRE can be obtained with the function "wwest" in R software developed by Terpstra and McKean (2005). Alternatively, $\widehat{\boldsymbol{\beta}}_{j\mathrm{R}}$ can be calculated by applying the iterative reweighted least squares algorithm of Sievers and Abebe (2004), with the help of major software packages. The detailed algorithm is provided in the Appendix. Note that the computation task here is not too difficult by virtue of the massive computing and data storage capabilities of modern computers. For instance, for $p=8$ and $n=20$, usually less than 1/1000 of a second is required to complete the iterative procedure to obtain an estimate using a Pentium-M 2.4 MHz CPU. The MSEWMA-R chart is generally applicable to online monitoring.

In addition, by using the efficient algorithms provided by Hettmansperger and Randles (2002), the convergence of $(\boldsymbol{\theta}_{(0)}, \boldsymbol{\Gamma}_{(0)})$ from the historical profile data with any practical p and m_0 is guaranteed and is usually quite fast. Since this is a one-time computation before Phase II online process monitoring, it is easy to accomplish in practice. The Fortran codes for implementing the proposed schemes, including solving Equations (5) and (6) and obtaining $\widehat{\boldsymbol{\beta}}_{j\mathrm{R}}$, are available from the authors upon request.

2.4.2. *On the control limits and robustness*

As shown in Propositions 2 and 3 of Zou and Tsung (2011), the MSEWMA chart is distribution free in the sense that its IC run-length distribution is the same for all distributions with elliptical directions. The Q_j process is a Markov chain if the underlying distribution has elliptical directions. Zou and Tsung (2011) claimed that the MSEWMA shares a similar key property with its parametric counterpart, the MEWMA, which is that the IC ARL of MSEWMA for distributions with elliptical directions can be calculated via the Markov chain model, which greatly facilitates the search for the control limits of the MSEWMA-R chart. The detailed algorithm can be obtained from Zou and Tsung (2011).

Table 1 tabulates the control limits of the MSEWMA-R chart for various commonly used combinations of λ, p, and IC ARL obtained using a Markov chain with $m = 200$ transition states. The simulation results shown in the next section demonstrate that the IC run-length performance of the MSEWMA-R chart is robust under various process distributions including some skewed error distributions. Therefore, the control limits tabulated in Table 1 could be used in practice when there is little knowledge about error distributions. Note that the control limits with p correspond to the control limits presented in Table 1 of Zou and Tsung (2011) with $p + 2$ because in MSEWMA-R there are actually $p + 2$ parameters to be monitored.

2.4.3. *On choosing the smoothing weight*

λ: The MSEWMA-R chart is robust under IC for $\lambda \in (0, 0.2]$. In general, a smaller λ leads to a quicker detection of small shifts (see, for example, Lucas and Saccucci (1990)). Based on our simulation results, we suggest choosing $\lambda \in [0.05, 0.2]$ in practice.

3. Comparison of numerical performance

In order to see when the proposed MSEWMA-R chart should be considered for use, we present some simulation results in this section regarding its run-length performance, as compared with that of the MEWMA-L chart. We do not consider the chart proposed by Kim *et al.* (2003), which is a combination of three charts, because Zou, Tsung, and Wang (2007) have shown that it has similar performance to MEWMA-L for the simple linear model and is not directly applicable to more complex models, such as multiple regression models and polynomial models. As is the convention in the Phase II literature, we assume that m_0 is sufficiently large so that all parameters are estimated with negligible error. The control limits of the MEWMA-L chart are determined using the Markov chain method given by Zou, Tsung, and Wang (2007) to attain the nominal IC ARL under the standard normal error distribution, while the control limits given in Table 1 are used for MSEWMA-R. Since the zero-state and Steady-State ARL (SSARL) results are similar, only the results for 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 (see Hawkins and Olwell (1998)). Because a similar conclusion holds for other cases, here we only present the results for when IC ARL $= 200$ and $\tau = 50$ for illustration. All of the ARL results in this section are obtained from 10 000 replications.

Table 1. The control limits of the MSEWMA-R chart with IC ARL = 200, 370, and 500 for the p-variate linear profile model

IC ARL	λ	$P = 1$	$P = 2$	$P = 3$	$P = 4$	$P = 5$	$P = 6$	$P = 7$	$P = 8$
200	0.4	7.920	9.668	11.321	12.911	14.448	15.946	17.402	18.841
	0.2	9.830	11.674	13.414	15.084	16.708	18.284	19.812	21.329
	0.1	10.052	11.896	13.636	15.310	16.911	18.489	20.027	21.532
	0.05	9.177	10.963	12.646	14.264	15.819	17.340	18.837	20.288
	0.025	7.691	9.345	10.906	12.408	13.864	15.290	16.694	18.066
370	0.4	8.294	10.125	11.847	13.505	15.083	16.633	18.140	19.628
	0.2	10.687	12.626	14.448	16.192	17.876	19.514	21.093	22.649
	0.1	11.303	13.249	15.077	16.828	18.511	20.150	21.750	23.310
	0.05	10.700	12.607	14.404	16.110	17.774	19.371	20.939	22.472
	0.025	9.392	11.205	12.918	14.551	16.124	17.668	19.176	20.644
500	0.4	8.459	10.329	12.083	13.772	15.388	16.951	18.489	19.983
	0.2	11.074	13.058	14.924	16.705	18.409	20.068	21.688	23.284
	0.1	11.887	13.877	15.750	17.525	19.247	20.929	22.549	24.147
	0.05	11.417	13.375	15.216	16.971	18.663	20.293	21.903	23.462
	0.025	10.198	12.081	13.852	15.536	17.165	18.755	20.293	21.812

To illustrate the effectiveness of our proposed monitoring scheme, the simulations cover two models: a simple linear model and a multiple regression model.

Scenario 1: The straight line regression model:

$$y_{ji} = \alpha + \beta x_i + \varepsilon_{ji}, \quad i = 1, \ldots, 7,$$

where $(\alpha, \beta) = (3, 2)$ and the design points x_i are fixed as

$$\{-0.429, -0.286, -0.143, 0, 0.143, 0.286, 0.429\},$$

in each profile. In fact, these points are generated from $x_i = (i - 0.5)/n$, $i = 1, \ldots, 7$, by centering so that their mean is zero. This model has been used for performance evaluation by Kang and Albin (2000), Kim et al. (2003), Zou, Tsung, and Wang (2007), etc. In this scenario, $p = 1$.

Scenario 2: A multiple regression model (Zou, Zhou, Wang, and Tsung, 2007; Mahmoud 2008):

$$y_{ji} = \alpha + \beta_{11} x_{i1} + \beta_{21} x_{i2} + \beta_{12} x_{i1}^2 + \beta_{22} x_{i2}^2 + \varepsilon_{ji}, \quad i = 1, \ldots, 15, \tag{10}$$

where $(\alpha, \beta_{11}, \beta_{21}, \beta_{12}, \beta_{22}) = (1, 2, 4, 3, 6)$. In this scenario, the design points $\mathbf{x}_i$ are

$$\begin{aligned}(x_{11}, \ldots, x_{15,1}) = {} & (0.374, -0.394, 0.461, -0.045, 0.115, \\ & -0.291, -0.268, 0.437, -0.367, -0.243, \\ & 0.337, 0.145, -0.458, 0.472, -0.276), \\ (x_{12}, \ldots, x_{15,2}) = {} & (-0.190, 0.000, 0.311, 0.268, -0.323, \\ & 0.304, -0.300, 0.098, 0.240, -0.016, \\ & -0.349, 0.023, 0.104, -0.276, 0.109),\end{aligned}$$

which are independently generated from the uniform distribution $U[0, 1]$ and centered so that their mean is zero. This scenario, containing both multiple and polynomial regression terms, is also quite common in practical applications. It is a higher dimensional case than Scenario 1, with $p = 4$. Following the robustness analyses in the literature (e.g., Zou and Tsung (2010) and the references therein), we consider the following three error distributions: (i) standard normal $N(0, 1)$; (ii) t distribution with ζ degrees of freedom, denoted as t_ζ; and (iii) chi-square distribution with ζ degrees of freedom, denoted as Chi_ζ. The generated variables in both cases (ii) and (iii) are standardized so that they have zero mean and unit variance.

We first compare the IC performance between MSEWMA-R and MEWMA-L. The simulation results for the two charts with different values of λ under Scenarios 1 and 2 are presented in Table 2. In addition to the ARLs, the corresponding Standard Deviation of the Run Lengths

Table 2. IC ARL and SDRL values of the MSEWMA-R and MEWMA-L charts with some non-normal distributions. Numbers in parentheses are SDRL values

			MSEWMA-R			*MEWMA-L*			
	f	ζ	$\lambda = 0.2$	$\lambda = 0.1$	$\lambda = 0.05$	$\lambda = 0.2$	$\lambda = 0.1$	$\lambda = 0.05$	$\lambda = 0.025$
	t_ζ	3	194 (188)	202 (196)	201 (190)	86.4 (84.7)	139 (136)	188 (177)	208 (188)
		4	198 (194)	197 (189)	196 (184)	102 (100)	151 (145)	181 (170)	197 (174)
		5	195 (190)	196 (188)	197 (186)	120 (117)	162 (156)	187 (176)	193 (171)
		7	199 (192)	200 (193)	200 (190)	147 (142)	175 (168)	192 (178)	195 (172)
		10	198 (190)	197 (188)	198 (185)	166 (161)	187 (181)	196 (184)	197 (172)
Scenario 1	Chi_ζ	1	187 (186)	194 (187)	197 (179)	92.4 (89.6)	145 (139)	189 (176)	197 (177)
		2	191 (188)	198 (192)	199 (180)	115 (113)	164 (160)	192 (180)	202 (174)
		3	192 (187)	198 (193)	202 (184)	129 (126)	173 (166)	201 (188)	203 (188)
		5	193 (192)	198 (194)	198 (185)	149 (148)	185 (177)	201 (185)	204 (182)
		8	196 (190)	200 (198)	201 (193)	164 (162)	193 (186)	202 (189)	201 (180)
	Control limit		9.830	10.052	9.177	11.865	10.786	9.376	7.707
	t_ζ	3	194 (188)	197 (190)	202 (190)	68.5 (66.5)	108 (104)	160 (150)	195 (173)
		4	194 (189)	196 (189)	198 (183)	88.1 (84.7)	134 (127)	168 (155)	192 (165)
		5	199 (190)	200 (193)	200 (185)	108 (105)	150 (142)	179 (163)	193 (166)
		7	200 (190)	201 (189)	199 (184)	136 (130)	168 (158)	188 (169)	195 (165)
		10	197 (191)	199 (186)	201 (184)	157 (152)	182 (171)	189 (170)	197 (169)
Scenario 2	Chi_ζ	1	183 (180)	191 (184)	194 (180)	73.6 (70.9)	124 (116)	166 (156)	191 (164)
		2	190 (184)	198 (186)	201 (185)	96.3 (93.8)	149 (141)	184 (166)	193 (165)
		3	191 (188)	197 (191)	201 (187)	116 (112)	164 (152)	190 (176)	199 (170)
		5	194 (188)	197 (186)	198 (182)	137 (133)	172 (166)	189 (172)	196 (168)
		8	196 (189)	200 (190)	199 (184)	146 (142)	175 (167)	191 (173)	196 (168)
	Control limit		15.084	15.310	14.264	17.501	16.264	14.582	12.487

(SDRL) values are also included in this table to give a broader picture of the run-length distribution. The two non-normal error distributions t_ζ and Chi_ζ with various degrees of freedom ζ are considered. The normal distribution is not considered, though, because both the MSEWMA-R and MEWMA-L charts can accurately attain the nominal IC ARL in the two considered scenarios.

From Table 2, we can see that the MSEWMA-R is satisfactorily robust to the heavy-tailed and skewed distributions. When the error distribution is very skewed, as is Chi_1, using $\lambda = 0.2$ results in a slightly smaller IC ARL than the nominal one, but this bias is smaller than that of MEWMA-L. When $\lambda \leq 0.1$, the MSEWMA-R's IC ARL is always close to the nominal one even for the extremely non-normal distribution. Note that in this table, the values with $\lambda = 0.025$ are not presented since they are almost 200 in all cases. The MEWMA-L usually has a certain bias in the IC ARL if λ is not small. For Scenario 2, when λ is 0.025 the MEWMA-L chart is able to maintain approximately a desired IC ARL (based on the criterion given by Jones *et al.* (2001) that the bias should be less than 5% percent of the nominal IC ARL).

Next, we compare the OC performance. Table 3 shows the ARLs for monitoring a shift in the intercept, slope, and variance (including increases and decreases) under Scenario 1. We use a generic notation δ to represent the shift size and consider the OC model with $\alpha_{(0)} + \delta\sigma_\alpha$, $\beta_{(0)} + \delta\sigma_\beta$, or $\delta\sigma$, where σ_α and σ_β are the standard deviations of $\widehat{\alpha}_j$ and $\widehat{\beta}_{j\text{L}}$, respectively. Note that for a relatively fair comparison, the MSEWMA-R with $\lambda = 0.1$ and 0.05 and the MEWMA-L chart with $\lambda = 0.05$ and 0.025 are considered because when $\lambda = 0.2$ or 0.1, the IC ARL of MEWMA-L deviates considerably from 200 as shown in the preceding example. From this table, when the error distribution is normal, we observe that the MEWMA-L chart has better efficiency as we would expect, since the parametric hypothesis is the correct one in this case. The MSEWMA-R chart also offers satisfactory performance for small and moderate shifts and the difference between MSEWMA-R and MEWMA-L is not significant. It should be pointed out that the superiority of MEWMA-L becomes more significant when δ is large. The analogous phenomenon for univariate non-parametric charts has been mentioned in the literature; e.g., by Zou and Tsung (2010). The MSEWMA-R, which is essentially based on signs rather than distances, shares a similar drawback with those rank-based charts for univariate processes. That is, even though the shift is large, the ranks or signs of the observations may not increase. When the error distribution (f) is not normal, the MSEWMA-R chart is more efficient in detecting the small and moderate shifts in slope than is the MEWMA-L chart with the same value of λ. This is mainly due to the benefit of using rank-based estimation. For detecting the shifts in the intercept, the MSEWMA-R chart does not seem to have

Table 3. A comparison of the ARLs of MSEWMA-R and MEWMA-L under Scenario 1 when the nominal IC ARL = 200

		N(0, 1)				*t_3*				*Chi_1*			
		MSEWMA-R		*MEWMA-L*		*MSEWMA-R*		*MEWMA-L*		*MSEWMA-R*		*MEWMA-L*	
	δ	$\lambda = 0.1$	$\lambda = 0.05$	$\lambda = 0.05$	$\lambda = 0.025$	$\lambda = 0.1$	$\lambda = 0.05$	$\lambda = 0.05$	$\lambda = 0.025$	$\lambda = 0.1$	$\lambda = 0.05$	$\lambda = 0.05$	$\lambda = 0.025$
α	0.25	86.4	73.2	67.7	61.6	67.4	57.0	73.6	65.8	29.5	27.2	31.0	30.5
	0.50	34.2	30.7	26.2	27.4	24.1	23.1	28.2	28.4	10.9	11.8	12.0	13.5
	1.00	12.2	13.1	10.6	12.3	9.63	10.7	10.8	12.6	6.43	7.60	5.00	6.07
	1.50	7.84	9.03	6.37	7.70	6.72	7.95	6.46	7.80	5.54	6.70	3.00	3.78
	2.00	6.19	7.39	4.44	5.53	5.72	6.91	4.51	5.59	5.17	6.30	2.10	2.71
	3.00	5.05	6.22	2.69	3.44	5.04	6.16	2.73	3.49	4.83	5.92	1.32	1.69
	5.00	4.58	5.68	1.46	1.90	4.67	5.77	1.47	1.91	4.57	5.64	1.00	1.05
β	0.25	84.2	71.2	67.1	61.4	57.7	48.2	74.7	66.7	27.5	25.8	70.0	64.2
	0.50	32.6	29.4	26.6	27.2	20.6	20.1	28.4	28.9	11.6	12.6	27.1	28.0
	1.00	12.1	12.9	10.5	12.2	8.55	9.71	10.9	12.6	6.51	7.67	10.6	12.3
	1.50	7.81	8.89	6.35	7.64	6.22	7.38	6.47	7.85	5.39	6.51	6.35	7.70
	2.00	6.14	7.27	4.45	5.50	5.37	6.50	4.52	5.61	4.94	6.06	4.43	5.52
	3.00	5.05	6.11	2.71	3.44	4.79	5.88	2.74	3.51	4.61	5.69	2.69	3.46
	5.00	4.55	5.58	1.47	1.91	4.52	5.58	1.47	1.91	4.41	5.48	1.47	1.90
σ	0.40	4.81	5.94	5.56	6.70	5.92	7.14	9.09	10.7	4.70	5.80	5.66	6.82
	0.60	6.52	7.75	9.38	11.0	10.2	11.4	16.7	18.3	6.09	7.24	9.55	11.2
	0.80	18.4	18.3	23.9	25.0	33.9	30.9	53.4	45.7	14.7	15.3	24.6	25.7
	1.20	30.9	27.9	18.6	21.1	48.3	41.6	34.9	36.9	24.1	23.1	20.2	22.1
	1.50	10.7	11.9	6.24	7.61	14.9	15.6	11.6	13.6	9.30	10.6	6.60	8.18
	1.80	7.58	8.79	3.68	4.57	9.51	10.7	6.85	8.29	6.94	8.20	3.91	5.04
	IC ARL	201	200	199	200	202	201	188	208	194	197	189	197

Table 4. A comparison of the ARLs of MSEWMA-R and MEWMA-L under Scenario 2 when the nominal IC ARL = 200

		N(0, 1)				*t_3*				*Chi_1*			
		MSEWMA-R		*MEWMA-L*		*MSEWMA-R*		*MEWMA-L*		*MSEWMA-R*		*MEWMA-L*	
	δ	$\lambda = 0.1$	$\lambda = 0.05$	$\lambda = 0.05$	$\lambda = 0.025$	$\lambda = 0.1$	$\lambda = 0.05$	$\lambda = 0.05$	$\lambda = 0.025$	$\lambda = 0.1$	$\lambda = 0.05$	$\lambda = 0.05$	$\lambda = 0.025$
α	0.25	100	80.6	79.1	70.9	85.2	69.7	79.9	74.5	48.5	40.5	39.4	39.0
	0.50	39.6	34.1	32.2	32.6	30.6	28.0	33.2	33.3	15.6	16.1	15.5	17.4
	1.00	13.6	14.3	12.5	14.5	11.0	12.2	12.9	14.6	6.72	7.95	6.37	7.85
	1.50	8.19	9.43	7.43	9.13	7.00	8.36	7.63	9.16	5.00	6.17	3.84	4.94
	2.00	6.11	7.37	5.20	6.57	5.56	6.76	5.32	6.61	4.41	5.49	2.70	3.54
	3.00	4.59	5.72	3.15	4.11	4.41	5.51	3.21	4.15	3.92	4.93	1.65	2.20
	5.00	3.82	4.83	1.69	2.27	3.80	4.80	1.71	2.27	3.60	4.57	1.02	1.23
β_{11}	0.05	102	86.9	78.4	70.1	69.6	57.1	75.4	69.19	42.0	36.2	76.0	70.5
	0.10	41.4	35.4	30.8	31.8	24.2	22.9	31.0	31.32	14.7	15.4	31.3	32.1
	0.15	20.8	20.4	17.62	19.6	13.0	13.9	17.6	19.59	8.90	10.1	17.9	20.1
	0.30	8.21	9.48	7.21	8.83	6.06	7.25	7.15	8.83	5.00	6.11	7.30	8.94
	0.50	5.15	6.30	3.85	4.91	4.38	5.43	3.80	4.91	4.02	5.03	3.89	4.99
	0.75	4.17	5.22	2.34	3.07	3.83	4.83	2.31	3.06	3.68	4.65	2.35	3.10
	1.00	3.83	4.82	1.65	2.18	3.64	4.59	1.62	2.17	3.56	4.51	1.66	2.22
β_{12}	0.05	106	88.5	80.3	71.9	71.5	58.0	76.2	70.2	43.6	37.0	77.8	71.9
	0.10	43.2	36.2	31.6	32.5	24.8	23.4	31.9	32.1	15.0	15.5	32.2	32.8
	0.15	21.6	21.0	18.0	20.1	13.4	14.1	18.2	20.0	9.11	10.3	18.4	20.5
	0.30	8.40	9.72	7.40	9.04	6.14	7.33	7.34	8.98	5.05	6.14	7.48	9.13
	0.50	5.25	6.42	3.95	5.05	4.40	5.46	3.89	5.00	4.04	5.04	3.97	5.09
	0.75	4.22	5.26	2.39	3.15	3.84	4.80	2.35	3.13	3.68	4.64	2.40	3.17
	1.00	3.84	4.86	1.68	2.23	3.62	4.59	1.65	2.21	3.56	4.49	1.69	2.25
σ	0.60	5.03	6.21	7.83	9.51	8.79	10.1	15.4	17.5	5.36	6.55	9.01	10.9
	0.80	12.6	13.6	18.7	20.8	29.8	27.2	48.3	42.9	14.4	15.2	22.3	24.2
	1.20	19.5	19.6	13.2	15.8	41.8	35.7	26.1	29.5	26.0	24.4	15.9	18.4
	1.50	7.29	8.54	4.36	5.71	12.4	13.3	8.86	10.9	9.09	10.3	5.31	6.81
	1.80	5.33	6.52	2.55	3.46	7.94	9.17	5.21	6.57	6.42	7.68	3.15	4.20
	IC ARL	200	199	199	201	197	202	160	195	191	194	166	191

any advantage over MEWMA-L. It is also worth noting that the MSEWMA-R chart outperforms the MEWMA-L in detecting a decrease in variance, whereas the MSEWMA-R is not as good at detecting an increase in variance as the MEWMA-L.

The above findings on the performance of MSEWMA-R and MEWMA-L are also valid for Scenario 2, which can be seen from the ARL results shown in Table 4. For simplicity, in Table 4, we only report the shift in α, β_{11}, β_{12}, and σ. Again, the OC model is $\alpha_{(0)} + \delta\sigma_\alpha$, $\beta_{11(0)} + \delta\sigma_{\beta_{11}}$, $\beta_{12(0)} + \delta\sigma_{\beta_{12}}$, or $\delta\sigma$, where $\sigma_{\beta_{1i}}$ are the standard deviations of the LSEs of β_{1i} for $i = 1, 2$. Note that in Table 4 the scales of δ for α and β_{1i}, $i = 1, 2$ are different. This setting is to make the non-centrality parameter:

$$\frac{1}{\sigma}\sqrt{(\boldsymbol{\beta}_{(1)} - \boldsymbol{\beta}_{(0)})^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}(\boldsymbol{\beta}_{(1)} - \boldsymbol{\beta}_{(0)})}.$$

for different shifts are comparable as the charting performance of the considered charts is essentially related to this parameter (Zou, Tsung, and Wang, 2007). Similar to Scenario 1, the MSEWMA-R performs better than MEWMA-L in detecting the small or moderate shifts in the regression coefficients β_{11} and β_{12}, which demonstrates the advantage of using $\widehat{\boldsymbol{\beta}}_{\mathrm{R}}$ when f is heavy-tailed. Of course, this advantage would be compromised if larger values of ζ are used (the error distribution is not as heavy-tailed as the considered t_3 or Chi_1). With respect to the intercept, the difference in the performance between MSEWMA-R and MEWMA-L comes mainly from the difference between MSEWMA and MEWMA because both $\mathbf{z}_{j\mathrm{R}}$ and $\mathbf{z}_{j\mathrm{L}}$ use the LSE $\widehat{\alpha}_j$. Also, the MSEWMA-R is outperformed by MEWMA-L when the shift size is very large in all of the cases due to the use of sign-based charts. Simultaneous shifts in multiple parameters are also of interest in practice. The OC ARL values for Scenario 2 with simultaneous shifts in (α, β_{11}), (α, σ), and (β_{11}, σ) are available in the supplemental file. In this situation, the two charts have comparable performance and neither one is better than the other. In words, the MSEWMA-R should be a reasonable alternative to existing charts for non-normal profile processes by virtue of its robustness.

To show the advantage of using rank-based estimators, we also include some comparisons between this chart

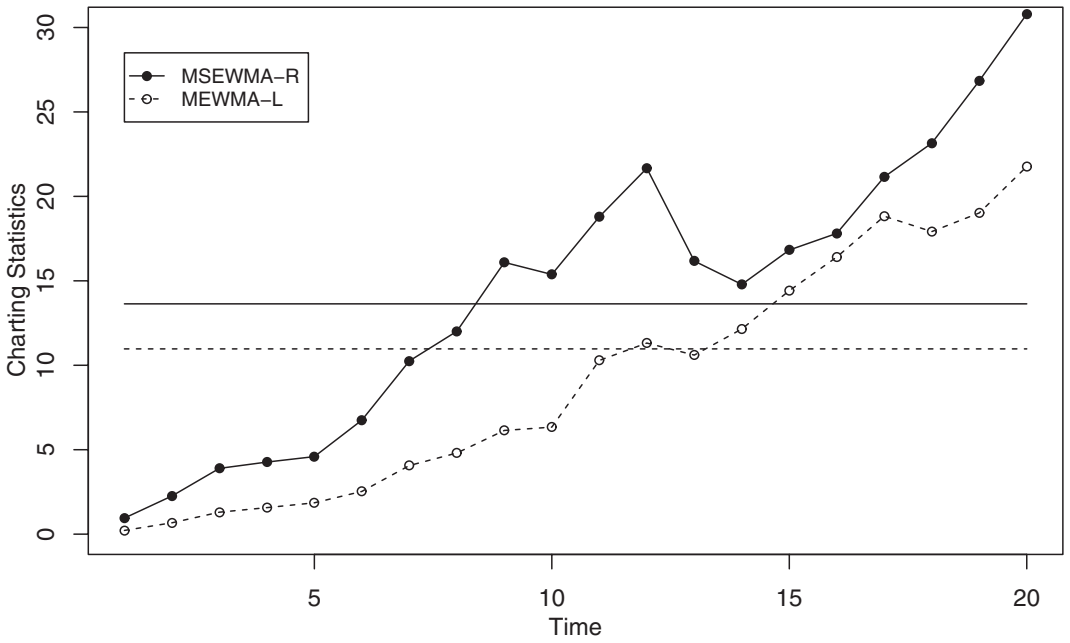

Fig. 1. A comparison of the OC ARLs of the MSEWMA-R and MSEWMA-L charts with $\lambda = 0.1$ and 0.05 for monitoring a shift in β_{11} ($\beta_{11(0)} + \delta$) under Scenario 2 (a multiple regression model with $p = 4$): (a) t_3; (b) Chi_1.

and a modified version of it using the LSE $\widehat{\boldsymbol{\beta}}_{j\text{L}}$. To be specific, different from $\mathbf{z}_{j\text{R}}$ we define $\mathbf{z}'_{j\text{L}} = [\widehat{\alpha}_j, \widehat{\boldsymbol{\beta}}_{j\text{L}}, \widehat{\sigma}^2_{j\text{L}}]^{\text{T}}$ and the corresponding MSEWMA chart (we shall denote it as MSEWMA-L) based on $\mathbf{z}'_{j\text{L}}$ can be constructed by mimicking Equations (5) to (10). Although both MSEWMA-R and MSEWMA-L inherit certain features of MSEWMA, their OC performance would be quite different for different error models due to the use of different parameter estimates. Figure 1 summarizes the ARL curves (in log scale) of the MSEWMA-R and MSEWMA-L charts for monitoring a shift in β_{11} under Scenario 2. In this figure, we consider $\lambda = 0.1$ and 0.05, and the results for error distributions t_3 and Chi_1 are shown in the left and right panels, respectively. It can be readily seen that the MSEWMA-L chart is outperformed by the MSEWMA-R chart with the same value of λ in detecting various magnitudes of shifts, especially for the Chi_1.

4. A real data application

In this section, we apply the proposed methodology to a real dataset from an Aluminum Electrolytic Capacitor (AEC) manufacturing process (provided by ENW Electronics Ltd.). The aim of the process is to transform the raw materials (anode aluminum foil, cathode aluminum foil, guiding pin, electrolyte sheet, plastic cover, aluminum shell, and plastic tube) into AECs with certain specifications. The whole manufacturing process, which is a typical multistage process (see Shi (2007)), includes a sequence of operations, such as clenching, rolling, soaking, assembly, cleaning, aging, and classifying. The quality of unfinished AEC products, which are called *capacitor elements*, in terms of appearance and functional performance is inspected by sampling after each stage. At each stage, certain important characteristics of an AEC, such as the capacitance and loss tangent (or equivalently dissipation factor), are automatically calibrated by an electronic device at some given measuring voltage, frequency, and temperature. The relationship between the characteristics of an AEC from one stage to another stage can often be described by linear models as demonstrated in the literature (see Shi (2007)). The engineers are usually concerned about significant changes in such relationships (or profiles) that may indicate that some assignable causes in the process have occurred.

To illustrate the above, we consider dissipation factor values in the aging stage as the response variable (y) and the values of capacitance and dissipation factor observations (denoted as x_1 and x_2) from the soaking stage as the explanatory variables. The dataset comprises 243 profile samples of size $n = 10$. Among them, 16 profiles are classified as inferior profiles based on physical knowledge and experience of the engineers. We use the other 227 profiles as the historical sample to calibrate the necessary parameters for monitoring. Such a reference sample may not be ideal for fully determining the IC distribution, but it suffices to illustrate the use of the method in a real-world setting. All of the data, including the 227 IC profiles used for calibration and 16 inferior profiles, are available in the supplemental file. In addition, in this example, the sample size is fixed for each profile sample but the $\mathbf{x}_i$ are not exactly the same. However, the MEWMA-L and MSEWMA-R charts are still applicable for this case. To calibrate the model, we pool all 227 profiles into a single sample; i.e., 2270

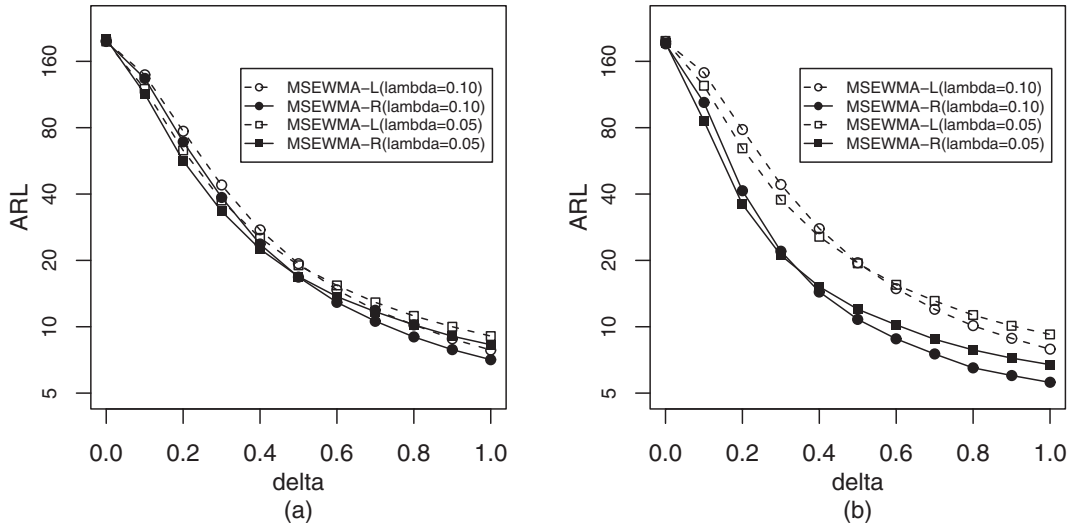

Fig. 2. Checking the normality assumption of the AEC dataset: (a) histogram of the LSE and residuals and (b) Q–Q plot of the LSE residuals.

observations. Using standard linear regression graphic analysis (Cook, 1998) and LSE, we obtain the following estimated IC model:

$$y_i = \alpha + \beta_{11}x_{1i} + \beta_2 x_{2i} + \beta_{12}x_{1i}^2 + \varepsilon_i, \quad i = 1, \cdots, 10, \tag{11}$$

where $\alpha = 26.25$, $\beta_{11} = 7.124$, $\beta_2 = 3.833$, $\beta_{12} = 0.215$, $\sigma^2 = 0.482$ and the covariates have been transformed so that all of the columns of $\mathbf{X}$ are orthogonal to $\mathbf{1}$. The normality assumption on the error distribution may be poor, as suggested by the histogram and the Quantile–Quantile (Q–Q) plot of the standardized LSE residuals in Figures 2(a) and 2(b). In fact, a one-sample Kolmogorov-Smirnov test on the residuals is highly significant (p-value is smaller than 0.001). All of these indications suggest that the normality assumption of f may not be valid. Thus, we may expect the rank-based regression to be more robust for this particular dataset. The estimated mean vector and covariance matrix ($\boldsymbol{\mu}_{(0)}$, $\boldsymbol{\Sigma}_{(0)}$) for MEWMA-L and ($\theta_{(0)}$, $\boldsymbol{\Gamma}_{(0)}$) for MSEWMA-R are presented in Table 5. Note that the first four elements of $\boldsymbol{\mu}_{(0)}$ given in Table 5 are slightly different from those pooled estimates given below Equation (11) since $\boldsymbol{\mu}_{(0)}$ is the average of all 227 individual LSE vectors. In addition, $\sigma^2_{(0)}$ in the definition of $\mathbf{z}_{jL}$ is replaced by its pooled estimate 0.482.

After computing all of the necessary estimates from the IC data, we are ready to construct the proposed chart for Phase II analysis. We set the IC ARL to 200 and λ to 0.1. The control limit is 13.636 ($p = 3$) as given in Table 1. Figure 3 shows the resulting MSEWMA-R chart (solid line connecting the dots) along with its control limit (the solid horizontal line). The corresponding MEWMA-L chart with $\lambda = 0.025$ (dashed line connecting circles) is also presented in the figure, along with its control limit of 10.970 (the dotted horizontal line). Note that $\lambda = 0.025$ is used in MEWMA-L to make it robust to this heavy-tailed data. For illustration, we have generated four IC profiles by adding random errors to the estimated model (11). These random errors were drawn from the LSE residuals based on the 227 IC profiles with replacement. In this example, we assume

Table 5. The estimated ($\boldsymbol{\mu}_{(0)}$, $\boldsymbol{\Sigma}_{(0)}$) for MEWMA-L and ($\theta_{(0)}$, $\boldsymbol{\Gamma}_{(0)}$) for MSEWMA-R for the AEC data

MEWMA-L					*MSEWMA-R*				
$\boldsymbol{\mu}_{(0)}$					$\boldsymbol{\theta}_{(0)}$				
26.25	*6.921*	*3.799*	*0.276*	*−0.181*	*26.26*	*6.876*	*3.859*	*0.248*	*0.410*
$\boldsymbol{\Sigma}_{(0)}$					$\boldsymbol{\Gamma}_{(0)}$				
0.045	0.090	−0.000	−0.010	−0.017	1.000	0.000	0.000	0.000	0.000
0.090	178.3	19.23	−28.79	1.278	0.004	0.254	0.000	0.000	0.000
−0.000	19.23	2.835	−3.389	−0.150	0.004	0.703	0.350	0.000	0.000
0.010	−28.79	−3.389	4.782	−0.047	0.022	2.001	0.241	0.134	0.000
−0.017	1.278	−0.150	−0.047	1.980	0.120	−0.082	0.056	−0.064	0.633

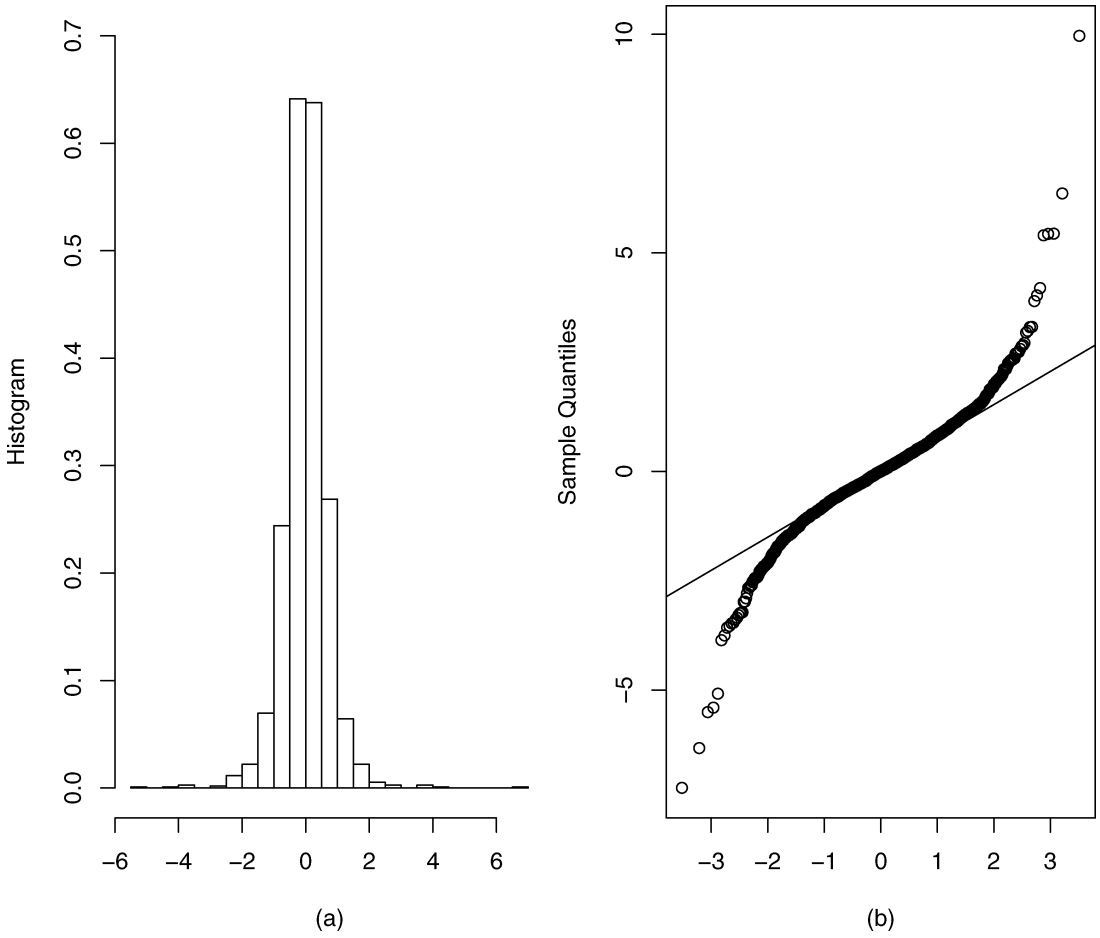

Fig. 3. The MSEWMA-R and MEWMA-L control charts for monitoring the AEC profile, with the solid and dashed horizontal lines indicating their control limits, respectively.

that we first monitor these four simulated IC samples (the IC model (11) plus errors from $N(0, \sigma^2_{(0)})$) and then obtain the 16 OC profiles. From the plot, it can be seen that the MSEWMA-R chart passes its control limit at around the ninth profile and remains above the control limit for the remaining profiles. This excursion suggests that a marked step-change has occurred. In comparison, the MEWMA-L chart does not give any signal until the 12th profile. All of the intermediate results for obtaining this figure, including the $\mathbf{z}_{j\mathrm{R}}$, $\mathbf{z}_{j\mathrm{L}}$ and the charting statistics, are also provided in the supplemental file.

5. Concluding remarks

Linear profile monitoring is particularly useful in practice due to its simplicity and flexibility. The existing monitoring methods suffer from a drawback in that they all assume a normal error distribution. When the underlying distribution is misspecified, the performance of control charts in monitoring the estimated profile parameters, designed under the normality assumption, may deteriorate to a certain degree. Hence, there is a need for SPC schemes to monitor profiles robustly. To achieve this goal, we have developed a non-parametric methodology for monitoring linear profiles, including the regression coefficients and profile variation. We have proposed the use of a WRE instead of LSE. As the resulting estimated profile parameter vectors are non-normal, we apply the MSEWMA control scheme to them. Benefiting from certain favorable properties of MSEWMA and WRE's efficiency, the proposed chart is robust in terms of both the IC and OC ARLs, particularly when the process distribution is heavy-tailed. A real data example from manufacturing has shown that our methodology performs quite well in application.

As we have shown in Section 3, the main advantage of MSEWMA-R in terms of OC ARL lies in its detection ability for small or moderate shifts in $\boldsymbol{\beta}$ when the error distribution is very heavy-tailed. The MSEWMA-R is outperformed by MEWMA-L in detecting very large shifts especially for the change in the intercept. The MSEWMA-R is more sensitive to a decrease in error variance but the MEWMA-L performs better for an increase in variance. In addition, the proposed method requires more

computational effort than a traditional parametric method such as MEWMA-L. In practice, engineers need to choose an appropriate chart according to their major concern. First of all, one should perform some residual analysis to determine whether the underlying error distribution is heavy-tailed. Some technical, empirical, and engineering knowledge about the profile modeling should also be taken into consideration. If the error distribution is close to normal or light-tailed, the MEWMA-L is preferred. Otherwise, one may consider its alternative, the MSEWMA-R, especially when the shifts in the regression coefficient $\boldsymbol{\beta}$ are more important to detect. However, when the process tends to incur a quite large change in some parameters, the MEWMA-L is more favorable because the MSEWMA-R chart is not as efficient as the MEWMA-L chart for very large shifts, as shown in Section 3. This drawback, which is common to almost all rank-based non-parametric charts, is mainly inherited from the way that MSEWMA is constructed where only the direction of multivariate observations from the origin is used. Of course, the benefit of using MSEWMA is its robustness in terms of IC ARL. In particular, when the normality assumption is doubtful and the control of nominal false alarm rates (IC ARL) is important, the MSEWMA-R may be preferred.

The proposed MSEWMA-R chart incorporates three aspects that differ from the existing MEWMA-L method. Most of the improvement derives from the use of the robust rank-based estimator of the regression coefficient rather than the LSEs, which are known to break down with leptokurtic error distributions. Another difference, the use of the sign-based chart instead of conventional MEWMA, is to make the control chart more robust in terms of IC ARL. This aspect seems to be less important since the MEWMA chart would attain the nominal IC ARL provided that the smoothing parameter is chosen small enough (Stoumbos and Sullivan, 2002). A third difference is that MSEWMA-R does not applies the transformation of Zou, Tsung, and Wang (2007) to the estimated variance. Various variants of MSEWMA-R could be considered. For instance, we may expect that the MEWMA chart with the WRE $\widehat{\boldsymbol{\beta}}_{j\mathrm{R}}$ would alleviate the issue of insensitivity to very large shifts at the expense of choosing the smoothing parameter λ more carefully to achieve IC ARLs. A thorough study on those three aspects and variants of MSEWMA-R would be an opportunity for future research.

As we can expect, the performance of both MEWMA-L and MSEWMA-R is affected by the amount of data in the reference dataset. Through simulation, we find that the performances of MSEWMA-R and MEWMA-L are similarly affected when m_0 is small and very large Phase I samples must be collected for both charts to perform as well as those with known parameters. Thus, determination of the Phase I sample size required to remove the effects of estimated parameters is critical and warrants future research. In quality control, besides detecting a process change quickly, it is also critical to diagnose the change and to identify which parameter or parameters in a profile have shifted after an OC signal occurs. Zou, Zhou, Wang, and Tsung (2007) and Mahmoud (2008) have suggested some parametric testing methods that are based on the normality as well. A robust diagnostic aid to locate the change-point in the process and to isolate the type of parameter change in a profile with non-normal error distributions may be desired and deserves further study.

Acknowledgements

The authors would like to thank the Department Editor and three anonymous referees for their many helpful comments that have resulted in significant improvements in the article. This research was supported by the RGC Competitive Earmarked Research Grants 620010, RPC10EG15, and NNSF of China Grants 11001138, 11071128, 10901092, 11131002, and 11101306. Zou also acknowledges support under Nankai Young Grant 65010731.

References

Bloomfield, P. and Steiger, W.L. (1983) *Least Absolute Deviation: Theory, Applications and Algorithms*, Birkhauser, Boston, MA.

Chakraborti, S., Van der Laan, P. and Bakir, S.T. (2001) Nonparametric control charts: an overview and some results. *Journal of Quality Technology*, **33**, 304–315.

Chicken, E., Pignatiello, J.J. and 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*, Wiley, New York, NY.

Ding, Y., Zeng, L. and Zhou, S. (2006) Phase I analysis for monitoring nonlinear profiles in manufacturing processes. *Journal of Quality Technology*, **38**, 199–216.

Hawkins, D.M. and Olwell, D.H. (1998) *Cumulative Sum Charts and Charting for Quality Improvement*, Springer-Verlag, New York, NY.

Hettmansperger, T.P. and McKean, J.W. (2010) *Robust Nonparametric Statistical Methods*, second edition, CRC Press, Boca Raton, FL.

Hettmansperger, T.P. and Randles, R.H. (2002) A practical affine equivariant multivariate median. *Biometrika*, **89**, 851–860.

Jaeckel, L.A. (1972) Estimating regression coefficients by minimizing the dispersion of residuals. *The Annals of Mathematical Statistics*, **43**, 1449–1458.

Jensen, W.A. and Birch, J.B. (2009) Profile monitoring via nonlinear mixed models. *Journal of Quality Technology*, **41**, 18–34.

Jensen, W.A., Birch, J.B. and Woodall, W.H. (2008) Monitoring correlation within linear profiles using mixed models. *Journal of Quality Technology*, **40**, 167–183.

Jones, L.A., Champ, C.W. and Rigdon, S.E. (2001) The performance of exponentially weighted moving average charts with estimated parameters. *Technometrics*, **43**, 156–167.

Jureăková, J. (1971) Nonparametric estimate of regression coefficients. *The Annals of Mathematical Statistics*, **42**, 1328–1338.

Kang, L. and 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. and Amiri, A. (2008) Phase I monitoring of polynomial profiles. *Communications in Statistics: Theory and Methods*, **37**, 1671–1686.

Kim, K., Mahmoud, M.A. and Woodall, W.H. (2003) On the monitoring of linear profiles. *Journal of Quality Technology*, **35**, 317–328.
Lada, E.K., Lu, J.-C. and Wilson, J.R. (2002) A wavelet-based procedure for process fault detection. *IEEE Transactions on Semiconductor Manufacturing*, **15**, 79–90.
Leng, C. (2010) Variable selection and coefficient estimation via regularized rank regression. *Statistica Sinica*, **20**, 167–181.
Lowry, C.A., Woodall, W.H., Champ, C.W. and Rigdon, S.E. (1992) Multivariate exponentially weighted moving average control chart. *Technometrics*, **34**, 46–53.
Lucas, J.M. and 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., Parker, P.A., Woodall, W.H. and Hawkins, D.M. (2007) A change point method for linear profile data. *Quality and Reliability Engineering International*, **23**, 247–268.
Mahmoud, M.A. and Woodall, W.H. (2004) Phase I analysis of linear profiles with calibration applications. *Technometrics*, **46**, 380–391.
McKean, J.W. and Schrader, R.M. (1980) The geometry of robust procedures in linear models. *Journal of the Royal Statistical Society, Series B*, **42**, 366–371.
Pollard, D. (1991) Asymptotics for least absolute deviation regression estimators. *Econometric Theory*, **7**, 186–199.
Qiu, P. (2008) Distribution-free multivariate process control based on log-linear modeling. *IIE Transactions*, **40**, 664–677.
Qiu, P. and Hawkins, D.M. (2001) A rank-based multivariate CUSUM procedure. *Technometrics*, **43**, 120–132.
Qiu, P., Zou, C. and Wang, Z. (2010) Nonparametric profile monitoring by mixed effects modeling (with discussions). *Technometrics*, **52**, 265–277.
Randles, R.H. (2000) A simpler, affine invariant, multivariate, distribution-free sign test. *Journal of the American Statistical Association*, **95**, 1263–1268.
Shi, J. (2007) *Stream of Variation Modeling and Analysis for Multistage Manufacturing Processes*, CRC Press, Boca Raton, FL.
Sievers, G.L. and Abebe, A. (2004) Rank estimation of regression coefficients using iterated reweighted least squares. *Journal of Statistical Computation and Simulation*, **74**, 821–831.
Stoumbos, Z.G. and Sullivan, J.H. (2002) Robustness to non-normality of the multivariate EWMA control chart. *Journal of Quality Technology*, **34**, 260–276.
Terpstra, J. and McKean, J. (2005) Rank-based analysis of linear models using R. *Journal of Statistical Software*, **14**, 1–26.
Wang, L., Kai, B. and Li, R. (2009) Local rank inference for varying coefficient models. *Journal of the American Statistical Association*, **104**, 1631–1645.
Williams, J.D., Woodall, W.H. and 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. and Gupta, S. (2004) Using control charts to monitor process and product quality profiles. *Journal of Quality Technology*, **36**, 309–320.
Yeh, A.B., Huwang, L. and Li, Y.M. (2009) Profile monitoring for binary response. *IIE Transactions*, **41**, 931–941.
Zhang, H. and Albin, S. (2009) Detecting outliers in complex profiles using a χ^2 control chart method. *IIE Transactions*, **41**, 335–345.
Zhou, C., Zou, C., Zhang, Y. and Wang, Z. (2009) Nonparametric control chart based on change-point model. *Statistical Papers*, **50**, 13–28.
Zou, C. and Tsung, F. (2010) Likelihood ratio based distribution-free EWMA schemes. *Journal of Quality Technology*, **42**, 174–196.
Zou, C. and Tsung, F. (2011) A multivariate sign EWMA control chart. *Technometrics*, **53**, 84–97.
Zou, C., Tsung, F. and Wang, Z. (2007) Monitoring general linear profiles using multivariate EWMA schemes. *Technometrics*, **49**, 395–408.
Zou, C., Tsung, F. and Wang, Z. (2008) Monitoring profiles based on nonparametric regression methods. *Technometrics*, **50**, 512–526.
Zou, C., Zhang, Y. and Wang, Z. (2006) Control chart based on change-point model for monitoring linear profiles. *IIE Transactions*, **38**, 1093–1103.
Zou, C., Zhou, C., Wang, Z. and Tsung, F. (2007) A self-starting control chart for linear profiles. *Journal of Quality Technology*, **39**, 364–375.

Appendix
Algorithm for Computing $\widehat{\boldsymbol{\beta}}_{\mathrm{R}}$

For notation convenience, we denote:

$$T_i(\boldsymbol{\beta}) = \frac{R(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})}{n+1} - \frac{1}{2}$$

and let $m(\boldsymbol{\beta})$ be the median of the residuals $\{r_i(b) = y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}, i = 1, \ldots, n\}$. We define:

$$w_i(\boldsymbol{\beta}) = \begin{cases} \dfrac{T_i(\boldsymbol{\beta})}{r_i(\boldsymbol{\beta}) - m(\boldsymbol{\beta})} & \text{if } r_i(\boldsymbol{\beta}) \neq m(\boldsymbol{\beta}), \\ 0 & \text{otherwise .} \end{cases}$$

Since $\sum_{i=1}^{n} T_i(\boldsymbol{\beta}) = 0$ the dispersion function given in Equation (3) may be written as

$$\begin{aligned} W_n(\boldsymbol{\beta}) :&= \sum_{i=1}^{n} T_i(\boldsymbol{\beta})[r_i(\boldsymbol{\beta}) - m_v(\boldsymbol{\beta})] \\ &= \sum_{i=1}^{n} w_i(\boldsymbol{\beta})[r_i(\boldsymbol{\beta}) - m_v(\boldsymbol{\beta})]^2. \end{aligned}$$

The iterated reweighted least squares algorithm given by Sievers and Abebe (2004) for computing $\widehat{\boldsymbol{\beta}}_{\mathrm{R}}$ is described as follows:

Step 1. Start by finding the initial valve of $\widehat{\boldsymbol{\beta}}_{\mathrm{R}}^{(1)}$. We recommend using the LSE $\widehat{\boldsymbol{\beta}}_{\mathrm{L}}$.

Step 2. Given the kth step estimate, $\boldsymbol{\beta}_{\mathrm{R}}^{(k)}$, the $(k+1)$th step estimate of $\boldsymbol{\beta}$ minimizes the kth step dispersion given by

$$W_n^*\big(\boldsymbol{\beta}|\boldsymbol{\beta}_{\mathrm{R}}^{(k)}\big) := \sum_{i=1}^{n} w_i\big(\boldsymbol{\beta}_{\mathrm{R}}^{(k)}\big)\big[r_i(\boldsymbol{\beta}_{\mathrm{R}}) - m_v\big(\boldsymbol{\beta}_{\mathrm{R}}^{(k)}\big)\big]^2.$$

Thus, the $(k+1)$th step estimate, $\boldsymbol{\beta}_{\mathrm{R}}^{(k+1)}$, satisfies $\nabla W_n^*(\boldsymbol{\beta}_{\mathrm{R}}^{(k+1)}|\boldsymbol{\beta}_{\mathrm{R}}^{(k)}) = 0$. This may be written as

$$\sum_{i=1}^{n} w_i\big(\boldsymbol{\beta}_{\mathrm{R}}^{(k)}\big)\big[y_i - m_v\big(\boldsymbol{\beta}_{\mathrm{R}}^{(k)}\big) - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_{\mathrm{R}}^{(k+1)}\big]\mathbf{x}_i = 0. \tag{A1}$$

Assuming that the $p \times p$ matrix $\sum_{i=1}^{n} w_i(\boldsymbol{\beta}_{\mathrm{R}}^{(k)})\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}$ is non-singular, Equation (A1) leads to

$$\boldsymbol{\beta}_{\mathrm{R}}^{(k+1)} = \boldsymbol{\beta}_{\mathrm{R}}^{(k)} + \left(\sum_{i=1}^{n} w_i(\boldsymbol{\beta}_{\mathrm{R}}^{(k)})\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}\right)^{-1} \times \sum_{i=1}^{n} w_i\big(\boldsymbol{\beta}_{\mathrm{R}}^{(k)}\big)\big[r_i\big(\boldsymbol{\beta}_{\mathrm{R}}^{(k)}\big) - m_v\big(\boldsymbol{\beta}_{\mathrm{R}}^{(k)}\big)\big]\mathbf{x}_i.$$

Step 3. Repeat Step 2 until the following condition is satisfied:

$$\big|W_n\big(\boldsymbol{\beta}_{\mathrm{R}}^{(k+1)}\big) - W_n\big(\boldsymbol{\beta}_{\mathrm{R}}^{(k)}\big)\big| \le \epsilon,$$

where ϵ is a pre-specified small positive number (e.g., $\epsilon = 10^{-3}$). Then the algorithm stops at the $(k+1)$th iteration and returns the value of $\boldsymbol{\beta}_{\mathrm{R}}^{(k+1)}$ as the final estimate for $\widehat{\boldsymbol{\beta}}_{\mathrm{R}}$.

Biographies

Xuemin Zi is Associate Professor of the School of Science at Tianjin University of Technology and Education. Her research interests include statistical process control and design of experiments.

Changliang Zou obtained his B.S., M.S., and Ph.D. in Statistics from the School of Mathematical Sciences, Nankai University, in 2003, 2006, and 2008, respectively. He then worked as a Postdoctoral Researcher at the Quality Laboratory, Department of IELM, Hong Kong University of Science and Technology. He joined the School of Mathematical Sciences, Nankai University, as an Assistant Professor in June 2009. His primary research interests include statistical process control, non-parametric regression, and dimension reduction. His research has been published in various refereed journals in Statistics and Industrial Engineering, including *Journal of the American Statistical Association*, *Annals of Statistics*, *Technometrics*, *Journal of Quality Technology*, *IIE Transactions*, *Statistica Sinica*, and *Annals of Operation Research*.

Fugee Tsung is Professor and Head of the Department of Industrial Engineering and Logistics Management, Director of the Quality Laboratory, at the Hong Kong University of Science & Technology. He is a Fellow of the Institute of Industrial Engineers, Fellow of the American Society for Quality, Academician of the International Academy for Quality, and Fellow of the Hong Kong Institution of Engineers. He received both his M.Sc. and Ph.D. from the University of Michigan, Ann Arbor, and his B.Sc. from National Taiwan University. He is currently Department Editor of the *IIE Transactions*, Associate Editor of *Technometrics* and *Naval Research Logistics*, and is on the Editorial Board of the *Journal of Quality Technology*. He has authored over 80 refereed journal publications and was also the winner of the Best Paper Award for *IIE Transactions* in 2003 and 2009. His research interests include quality engineering and management to manufacturing and service industries, statistical process control, monitoring, and diagnosis.