Research Article

Quality and Reliability Engineering International

(www.interscience.wiley.com) DOI: 10.1002/qre.1066 Published online 1 September 2009 in Wiley InterScience

Statistical Monitoring of Multivariate Multiple Linear Regression Profiles in Phase I with Calibration Application

R. Noorossana[a], M. Eyvazian[a], Amirhossein Amiri[b]*† and Mahmoud A. Mahmoud[c]

In some statistical process control applications, there are some correlated quality characteristics which can be modeled as linear functions of some explanatory variables. We refer to this structure as multivariate multiple linear regression profiles. When the correlation structure between quality characteristics is ignored and profiles are monitored separately then misleading results could be expected. Hence, developing methods to account for this multivariate structure is required. In this paper, we specifically focus on phase I monitoring of multivariate multiple linear regression profiles and develop four methods for this purpose. The performance of the developed methods is compared through simulation studies in terms of probability of a signal. In addition, a diagnostic scheme to find the out-of-control samples is developed. Finally, the application of the proposed methods is illustrated using a calibration application at the National Aeronautics and Space Administration (NASA) Langley Research Center. Copyright © 2009 John Wiley & Sons, Ltd.

Keywords: calibration; multivariate multiple regression; phase I; principal components analysis; statistical process control

1. Introduction

Sometimes the quality of a process or product is characterized by a relationship or function between two or more variables, which is referred to as a '*profile*' in the literature. This relationship can be modeled by a simple linear regression, multiple linear, or polynomial regression or even more complicated models such as nonlinear regressions. Several applications of profiles monitoring have been introduced by authors including Stover and Brill[1], Kang and Albin[2], Mahmoud and Woodall[3], Woodall *et al.*[4], Wang and Tsung[5], Montgomery[6], Woodall[7], and Zou *et al.*[8]. Simple linear profiles monitoring has been studied well in both Phases I and II. In Phase I, one investigates the process stability. On the other hand, detecting the out-of-control state as quickly as possible is the main purpose of Phase II. Authors including Kang and Albin[2], Kim *et al.*[9], Noorossana *et al.*[10], Gupta *et al.*[11], Zou *et al.*[12], Zhang *et al.*[13], Saghaei *et al.*[14], and Mahmoud *et al.*[15] have studied Phase II monitoring of simple linear profiles. Phase I monitoring of simple linear profiles has also been investigated by researchers such as Mestek *et al.*[16], Stover and Brill[1], Kang and Albin[2], Mahmoud and Woodall[3], and Mahmoud *et al.*[17]. Autocorrelated simple linear profiles were also studied by Noorossana *et al.*[18] and Soleimani *et al.*[19]. The effect of non-normality on the performance of control charts for Phase II monitoring of simple linear profiles was investigated by Noorossana *et al.*[20, 21].

Some authors investigated more complicated models including multiple linear regression profiles and polynomial profiles. Mahmoud[22] investigated Phase I monitoring of multiple linear regression profiles. Jensen *et al.*[23] proposed a linear mixed model (LMM) to model the autocorrelation structure in general linear profiles in Phase I. Zou *et al.*[8] proposed a multivariate exponentially weighted moving average (MEWMA) control chart for monitoring general linear profiles in Phase II. Kazemzadeh *et al.*[24] extended three phase I methods for monitoring polynomial profiles and compared the performance of the extended methods under both sustained and outlier shift. Kazemzadeh *et al.*[25] used a transformation based on an orthogonal polynomial regression to make the regression parameters independent and then used an EWMA-based monitoring procedure in Phase II. Kazemzadeh *et al.*[26] also studied Phase II monitoring of autocorrelated polynomial profiles where there is an AR(1) structure between successive profiles.

[a]*Department of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran*
[b]*Faculty of Engineering, Department of Industrial Engineering, Shahed University, Tehran, Iran*
[c]*Faculty of Economics and Political Science, Department of Statistics, Cairo University, Egypt*
**Correspondence to: Amirhossein Amiri, Faculty of Engineering, Department of Industrial Engineering, Shahed University, Tehran, Iran.*
†*E-mail: amirhossein.amiri@gmail.com*

Contract/grant sponsor: Iran National Science Foundation for Dr Noorossana's research

Nonlinear profiles monitoring has also been discussed by several authors including Jin and Shi[27], Walker and Wright[28] Ding *et al.*[29], Williams *et al.*[30], Moguerza *et al.*[31], Vaghefi *et al.*[32] and Jensen and Birch[33].

In some applications, the quality of a process or a product can be characterized by a multivariate regression. In this case, there are some dependent quality characteristics as response variables, which are modeled as functions of one or more explanatory variables. Noorossana *et al.*[34, 35] investigated the case where there are relationships between some dependent response variables and one explanatory variable in both Phases I and II. We refer to this case as multivariate simple linear profiles monitoring. They proposed four approaches including principal components analysis, Wilks' lambda, which is an extension of the F method by Mahmoud and Woodall[3] to the multivariate case, and two approaches based on T-squared statistics for monitoring multivariate simple linear profiles in Phase I. Noorossana *et al.*[35] extended the traditional control charts by Kang and Albin[2] and the EWMA charts procedure by Kim *et al.*[9] to the multivariate case and developed some methods to monitor multivariate simple linear profiles in Phase II.

This paper is motivated by a calibration case discussed by Parker *et al.*[36] and Mahmoud *et al.*[17] in which the relationships between six response variables and six explanatory variables are investigated. The replicated calibrations of a force balance used in wind tunnel experiments are considered at the NASA Langley Research Center. In this case, the relationships between response variables and explanatory variables can be modeled by multiple linear regression models. Owing to the dependency of response variables, we deal with a multivariate regression structure. We refer to the case as a multivariate multiple linear regression profiles. In this paper, we specifically focus on Phase I and develop four methods to monitor multivariate multiple linear regression profiles. The extended methods are compared under both sustained and unsustained shifts. The structure of the paper is outlined as follows. In Section 2, the model and the assumptions are discussed. Four Phase I methods are developed in Section 3. In Section 4, a diagnostic aid is developed to find the out-of-control samples. In Section 5, the performance of the proposed methods is evaluated through simulation studies under both sustained step and unsustained isolated shifts. The proposed methods are illustrated by a calibration application at NASA in Section 6. Our concluding remarks are given in the last section.

2. Model and assumptions

We assume for the kth sample collected over time that we have observations $(x_{1ik}, x_{2ik}, \ldots, x_{qik}, y_{1ik}, y_{2ik}, \ldots, y_{pik})$ $i=1,2,\ldots,n_k$, $k=1,2,\ldots,m$, where n_k is the size of sample k, m is the number of samples and p and q are the number of response and explanatory variables, respectively. The model that relates the response variables with explanatory variables is a multivariate multiple linear regression and can be given as follows:

$$\mathbf{Y}_k = \mathbf{X}_k \mathbf{B}_k + \mathbf{E}_k \tag{1}$$

where $\mathbf{Y}_k$ is an $n_k \times p$ matrix of response variables for the kth sample, $\mathbf{X}_k$ is an $n_k \times (q+1)$ matrix of explanatory variables, $\mathbf{B}_k$ is a $(q+1) \times p$ matrix of the regression parameters, and $\mathbf{E}_k$ is an $n_k \times p$ matrix of the error terms. It is assumed that the error terms have a p-variate multivariate normal distribution with p-variate mean vector of $\underline{\mathbf{0}}$ and $p \times p$ variance–covariance matrix of $\boldsymbol{\Sigma}$. The ordinary least-square (OLS) estimator of the matrix $\mathbf{B}_k$ is as follows:

$$\hat{\mathbf{B}}_k = (\mathbf{X}_k^{\mathrm{T}} \mathbf{X}_k)^{-1} \mathbf{X}_k^{\mathrm{T}} \mathbf{Y}_k \tag{2}$$

3. Phase I proposed methods

Four methods for phase I monitoring of multivariate multiple linear regression profiles are proposed in this section. The first one is a T^2-based method proposed by Williams *et al.*[30]. The second method is an extension of the LRT method by Mahmoud *et al.*[17]. Wilks' lambda approach, which is an extension of the F method by Mahmoud and Woodall[3], is proposed as the third method. Finally, a principal components approach is discussed as the fourth method. It should be noted that the first, third, and fourth methods are extensions of methods proposed by Noorossana *et al.*[34] for monitoring multivariate simple linear profiles.

3.1. T-squared approach based on successive difference estimator

We first rewrite the matrix $\hat{\mathbf{B}}_k$ as a $1 \times (q+1)p$ multivariate normal random vector and denote it by $\hat{\boldsymbol{\beta}}_k$ as follows:

$$\hat{\boldsymbol{\beta}}_k = (\hat{\beta}_{01k}, \hat{\beta}_{02k}, \ldots, \hat{\beta}_{0pk}, \hat{\beta}_{11k}, \hat{\beta}_{12k}, \ldots, \hat{\beta}_{1pk}, \ldots, \hat{\beta}_{q1k}, \hat{\beta}_{q2k}, \ldots, \hat{\beta}_{qpk})^{\mathrm{T}} \tag{3}$$

Then, we construct the T^2 statistic for the kth sample as below:

$$T_k^2 = (\hat{\boldsymbol{\beta}}_k - \bar{\hat{\boldsymbol{\beta}}})^{\mathrm{T}} \mathbf{S}_{\hat{\beta}}^{-1} (\hat{\boldsymbol{\beta}}_k - \bar{\hat{\boldsymbol{\beta}}}) \tag{4}$$

where the average vector $\bar{\hat{\boldsymbol{\beta}}}$ is computed by using $\bar{\hat{\boldsymbol{\beta}}} = (1/m)\sum_{k=1}^{m} \hat{\boldsymbol{\beta}}_k$ and $\mathbf{S}_{\hat{\beta}}$ is the estimator of the covariance matrix of $\hat{\boldsymbol{\beta}}_k$, which is determined based on successive differences of the regression parameters vector estimators.

Similar to Williams *et al.*[30], we define $\hat{\mathbf{v}}_k = \hat{\boldsymbol{\beta}}_{k+1} - \hat{\boldsymbol{\beta}}_k$, $k = 1, 2, \ldots, m-1$ and stack the transpose of these $m-1$ difference vectors into the matrix $\hat{\mathbf{V}}$ as $[\hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2, \ldots, \hat{\mathbf{v}}_{m-1}]^{\mathrm{T}}$, then $\mathbf{S}_{\hat{\beta}}$ is given by

$$\mathbf{S}_{\hat{\beta}} = \frac{\hat{\mathbf{V}}^{\mathrm{T}} \times \hat{\mathbf{V}}}{2 \times (m-1)} \tag{5}$$

The upper control limit for this control chart is chosen to give a specified probability of Type I error. The exact distribution for the T^2 statistic in Equation (4) is not reported in the literature and just some approximate distributions have been proposed for it. For further information, refer to Williams *et al.*[37].

3.2. Likelihood ratio approach

This method is the extension of the method proposed by Mahmoud *et al.*[17] in multivariate regression case. With a step shift in one or more regression parameters in a multivariate regression profile data set after sample m_1, we assume that

$$\begin{aligned} \mathbf{Y}_k &= \mathbf{X}_k \mathbf{B}_1 + \mathbf{E}_{k1} \quad k = 1, 2, \ldots, m_1 \\ \mathbf{Y}_k &= \mathbf{X}_k \mathbf{B}_2 + \mathbf{E}_{k2} \quad k = m_1 + 1, \ldots, m \end{aligned} \tag{6}$$

where $\mathbf{E}_{k1} \sim MN(\underline{\mathbf{0}}, \boldsymbol{\Sigma}_1)$ and $\mathbf{E}_{k2} \sim MN(\underline{\mathbf{0}}, \boldsymbol{\Sigma}_2)$.

Considering the following null and alternative hypotheses:

$$\begin{aligned} &H_0 : \mathbf{B}_1 = \mathbf{B}_2 = \mathbf{B} \quad \text{and} \quad \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma} \\ &H_1 : H_0 \text{ is not true.} \end{aligned} \tag{7}$$

It can be shown (see appendix) that the likelihood ratio statistic is given by the following equation:

$$lrt_{m1} = N \log |\hat{\boldsymbol{\Sigma}}| - N_1 \log |\hat{\boldsymbol{\Sigma}}_1| - N_2 \log |\hat{\boldsymbol{\Sigma}}_2| \tag{8}$$

where N is the total number of observations, N_1 is the total number of observations prior to the change point and N_2 is the total number of observations following the change point. Here, $|\hat{\boldsymbol{\Sigma}}|$, $|\hat{\boldsymbol{\Sigma}}_1|$, and $|\hat{\boldsymbol{\Sigma}}_2|$ are the determinants of the error terms variance–covariance matrix estimator for the multivariate regression model fitted for all of the samples pooled into one sample of size N, the error terms variance–covariance matrix estimator for the multivariate regression model fitted for all of the samples prior to m_1 pooled into one sample of size N_1, and the error terms variance–covariance matrix estimator for the multivariate regression model fitted for all of the samples following m_1 pooled into one sample of size N_2, respectively. The formulas used for computing these matrices are

$$\hat{\boldsymbol{\Sigma}} = \frac{(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})^{\mathrm{T}} (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})}{N} \tag{9}$$

$$\hat{\boldsymbol{\Sigma}}_1 = \frac{(\mathbf{Y}_1 - \mathbf{X}_1\hat{\mathbf{B}}_1)^{\mathrm{T}} (\mathbf{Y}_1 - \mathbf{X}_1\hat{\mathbf{B}}_1)}{N_1} \tag{10}$$

$$\hat{\boldsymbol{\Sigma}}_2 = \frac{(\mathbf{Y}_2 - \mathbf{X}_2\hat{\mathbf{B}}_2)^{\mathrm{T}} (\mathbf{Y}_2 - \mathbf{X}_2\hat{\mathbf{B}}_2)}{N_2} \tag{11}$$

The multivariate regression parameter estimator matrices $\hat{\mathbf{B}}$, $\hat{\mathbf{B}}_1$, and $\hat{\mathbf{B}}_2$ are calculated based on all the samples pooled into one sample, samples prior to the change point pooled into one sample, and samples following the change point pooled into one sample, respectively.

The corrected likelihood ratio statistic based on Mahmoud *et al.*[17] is given by

$$lrtc_{m1} = \frac{lrt_{m1}}{E(lrt_{m1})} \tag{12}$$

where $E(lrt_{m1})$ is the expected value of lrt_{m1} statistic under H_0 and can be determined by simulation. An upper control limit is chosen by simulation to give a specified probability of a Type I error.

3.3. Wilks' lambda approach

This approach is an extension of the F-approach by Mahmoud and Woodall[3] to the multivariate regression structure. This approach was first proposed by Noorossana *et al.*[34] for monitoring multivariate simple linear profiles. Here, we extend the approach to the multivariate multiple linear regression profiles. We assume that we have m samples each with size of n_k observations. The first step is to pool all the m samples into one sample of size $N = \sum_{k=1}^{m} n_k$. Then, $m-1$ indicator variables are defined as

$$\begin{cases} Z_{ki} = 1 & \text{if observation } i \text{ is from sample } k \\ Z_{ki} = 0 & \text{otherwise } i = 1, 2, \ldots, N, \; k = 1, 2, \ldots, m-1 \end{cases} \tag{13}$$

The mth sample is called the reference sample. Finally, the following regression model is fitted to the pooled data:

$$y_i=\hat{\beta}_0+\hat{\beta}_1x_{1i}+\cdots+\hat{\beta}_qx_{qi}+\hat{\beta}_{01}Z_{1i}+\hat{\beta}_{11}Z_{1i}x_{1i}+\cdots+\hat{\beta}_{q1}Z_{1i}x_{qi}+\cdots+\hat{\beta}_{0m'}Z_{m'i}+\hat{\beta}_{1m'}Z_{m'i}x_{1i}+\cdots+\hat{\beta}_{qm'}Z_{m'i}x_{qi}+\varepsilon_i \quad i=1,2,\ldots,N \tag{14}$$

where $m'=m-1$, $\mathbf{y}_i$ is a $1\times p$ vector of response variables for the ith observation in the pooled sample, $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_q$, $\boldsymbol{\beta}_{01},\ldots,\boldsymbol{\beta}_{qm'}$ are $1\times p$ vector of regression parameters, and ε_i is a $1\times p$ vector of random errors.

The following hypotheses are tested to check for the equality of the m multivariate regression equations.

$$\begin{aligned}&H_0:\boldsymbol{\beta}_{01}=\boldsymbol{\beta}_{11}=\cdots=\boldsymbol{\beta}_{q1}=\cdots=\boldsymbol{\beta}_{0m'}=\boldsymbol{\beta}_{1m'}=\cdots=\boldsymbol{\beta}_{qm'}=\mathbf{0}\\&H_1:H_0 \text{ is not true}\end{aligned} \tag{15}$$

Under the null hypothesis, the reduced model is given by:

$$\mathbf{y}_i=\hat{\boldsymbol{\beta}}_0+\hat{\boldsymbol{\beta}}_1x_{1i}+\cdots+\hat{\boldsymbol{\beta}}_qx_{qi}+\varepsilon_i \quad i=1,2,\ldots,N \tag{16}$$

To test the hypotheses in Equation (15), the following statistic is used (Rencher[38], p. 347):

$$\Lambda=\frac{|\mathbf{Y}^{\mathrm{T}}\mathbf{Y}-\hat{\mathbf{B}}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{Y}|}{|\mathbf{Y}^{\mathrm{T}}\mathbf{Y}-\hat{\mathbf{B}}_r^{\mathrm{T}}\mathbf{X}_r^{\mathrm{T}}\mathbf{Y}|} \tag{17}$$

where $\mathbf{Y}$ is an $N\times p$ matrix of response variables, $\mathbf{X}$ is an $N\times[m(q+1)]$ matrix of observations, $\hat{\mathbf{B}}$ is an $[m(q+1)]\times p$ matrix of the regression parameters for full model in Equation (14). $\mathbf{X}_r$ is an $N\times(q+1)$ matrix of observations and $\hat{\mathbf{B}}_r$ is a $(q+1)\times p$ matrix of the regression parameters for the reduced model in Equation (16).

Under the null hypothesis the statistic in Equation (17) follows Wilks' lambda distribution with p, $(q+1)(m-1)$, and $N-m(q+1)$ degrees of freedom. If $\Lambda<\Lambda_{\alpha,p,(q+1)(m-1),N-m(q+1)}$, then H_0 is rejected (see Rencher[38], p. 347 for more information).

In conjunction with the Wilks' lambda statistic, Noorossana *et al.*[34] proposed the statistic of Lee *et al.*[39], $-2\log(M)$, to test for the equality of the covariance matrices. The exact upper percentage points for statistic$-2\log(M)$ are also given in Lee *et al.*[39]. The M statistic is given by the following equation:

$$M=\frac{|\mathbf{S}_1|^{(n_1-q-1)/2}|\mathbf{S}_2|^{(n_2-q-1)/2}\ldots|\mathbf{S}_m|^{(n_m-q-1)/2}}{|\mathbf{S}_{pl}|^{(N-m(q+1))/2}} \tag{18}$$

where $\mathbf{S}_k$ $(k=1,2,\ldots,m)$ is an unbiased estimator of variance–covariance matrix of sample k, $\boldsymbol{\Sigma}_k$, and $\mathbf{S}_{pl}$ is the pooled sample covariance matrix. These matrices are computed by

$$\mathbf{S}_k=\frac{\mathbf{Y}_k^{\mathrm{T}}\mathbf{Y}_k-\hat{\mathbf{B}}_k^{\mathrm{T}}\mathbf{X}_k^{\mathrm{T}}\mathbf{Y}_k}{n_k-(q+1)} \tag{19}$$

$$\mathbf{S}_{pl}=\frac{\sum_{k=1}^{m}[n_k-(q+1)]\mathbf{S}_k}{\sum_{k=1}^{m}[n_k-(q+1)]} \tag{20}$$

where $\mathbf{Y}_k$ is an $n_k\times p$ matrix of response variables for the kth sample, $\mathbf{X}_k$ is an $n_k\times(q+1)$ matrix of explanatory variables, and $\hat{\mathbf{B}}_k$ is the least-squares estimator of $\mathbf{B}$ for the kth sample. The M statistic varies between 0 and 1, with values close to 1 indicating the equality of the covariance matrices.

3.4. *Principal components analysis approach*

The fourth method is the method proposed by Noorossana *et al.*[34] for monitoring multivariate simple linear profiles. In this method, they used principal components analysis to make the correlated response variables independent and to reduce the number of response variables. They first pooled all the m samples into one sample of size N and used the principal components of the response variables, which explain the most variations in the response variables. These new components are linear combinations of the p response variables and follow normal distributions. Then, the scores of these principal components were computed for each observation. After that, they modeled the relationships between these new variables and the explanatory variables. Since the new variables are independent, they can be monitored separately. Finally, they used the likelihood ratio method by Mahmoud *et al.*[17] to monitor the relationships between the principal components and the explanatory variable in Phase I. The upper control limits for each control chart are determined by simulation to obtain a specified overall probability of Type I error. We use this method in our paper considering the fact that we deal with a multiple linear regression instead of a simple linear regression. It should be noted that this method can only be used if the x variables and their values corresponding to each profile are the same at all m samples.

4. Diagnostic aid

As discussed in the Introduction section, the purpose of Phase I studies is to check the stability of the process as well as to determine the out-of-control samples. Four approaches to test for the process stability are introduced in the previous section. In this section, we propose a diagnostic procedure to find the out-of-control samples. In the proposed scheme, Wilk's lambda approach is used. If the equality of covariance matrices and equality of regression parameters are rejected, each sample is compared with the mth sample to find the out-of-control sample. In other words, the equality of the regression parameters based on jth and mth samples as well as the equality of their covariance matrices ($\boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}_m$) are tested. It should be noted that if one of the above hypotheses is rejected, then its corresponding diagnostic test is performed. If the hypothesis of equality of covariance matrices is rejected, we use the statistic $-2\log(M)$, where M is computed as follows:

$$M = \frac{|\mathbf{S}_j|^{(n_j-(q+1))/2}|\mathbf{S}_m|^{(n_m-(q+1))/2}}{|\mathbf{S}_{pl}|^{(n_j+n_m-2(q+1))/2}} \tag{21}$$

where $\mathbf{S}_k$ and $\mathbf{S}_{pl}$ are computed using the following equations, respectively:

$$\mathbf{S}_k = \frac{\mathbf{Y}_k^{\mathrm{T}}\mathbf{Y}_k - \hat{\mathbf{B}}_k^{\mathrm{T}}\mathbf{X}_k^T\mathbf{Y}_k}{n_k-(q+1)} \tag{22}$$

$$\mathbf{S}_{pl} = \frac{[n_j-(q+1)]\mathbf{S}_j + [n_m-(q+1)]\mathbf{S}_m}{[n_j-(q+1)]+[n_m-(q+1)]} \tag{23}$$

where $\mathbf{Y}_k$ is an $n_k \times p$ matrix of response variables for the kth sample, $\mathbf{X}_k$ is an $n_k \times (q+1)$ matrix of explanatory variables, and $\hat{\mathbf{B}}_k$ is the least-squares estimator of $\mathbf{B}$ for the kth sample.

A method similar to the approach by Mahmoud[22] is proposed if the equality of regression parameters is rejected. In this method, Mahmoud[22] combined the jth and mth samples and proposed an F approach to find the source of variation and to determine the out-of-control samples. In this approach, the F-statistic is decomposed into $m-1$ F-statistics, each representing whether or not the jth sample is out-of-control. Similar to this approach, we first combine the jth and mth samples and then decompose the Wilk's lambda statistic to $m-1$ Wilk's statistics with p, $q+1$ and $n_j+n_m-2(q+1)$ degrees of freedom, each obtained by comparing the jth sample with the last sample.

5. Performance comparisons

In this section, we compare the performance of the proposed methods for Phase I monitoring of multivariate multiple linear regression profiles in terms of the overall probability of a signal. Each signal probability is estimated using 10 000 simulated sets of multivariate multiple profile data. These methods are Method A: *multivariate T^2 control chart*, Method B: likelihood ratio approach, Method C: Wilks' lambda approach and Method D: principal components approach. The underlying multivariate profile model that we use in this paper is

$$Y_{i1k} = 0+1x_{i1}+1x_{i2}+\varepsilon_{i1k}, \quad Y_{i2k} = 0+1x_{i1}+1x_{i2}+\varepsilon_{i2k}, \quad i=1,2,\ldots,n, \; k=1,2,\ldots,m$$

where $n=10$, $m=20$, $(\varepsilon_{i1k}, \varepsilon_{i2k})$ is a bivariate normal random variable with mean vector zero and variance–covariance matrix of

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

The ordered pairs of (0, 0.2), (0.2, 0.8), (0.4, 0.7), (0.6, 0.6), (0.8, 1), (1, 1.5), (1.2, 1.2), (1.4, 1.4), (1.6, 1.6), and (1.8, 1.2) are considered for the independent variables x_1 and x_2 as a set of 10 observations.

The decision rule of methods was set to produce an overall probability of Type I error equal to 0.04. For Method A, the upper control limit is set to 25.3 to produce an overall probability of Type I error of 0.04. For Method B, first, $E(lrt_{m1})$ are computed by 10 000 simulation runs and then a threshold of 2.75 is set to achieve a probability of a Type I error of approximately 0.04. For Method C, the global Wilks' lambda test was performed at a significance level of $\alpha_2 = 1-\sqrt{1-\alpha} = 0.02020$ and the control limits of the chart for monitoring the variance–covariance matrix were set based on a probability of Type I error equal to $\alpha_3 = 1-\sqrt[m]{1-\alpha_2} = 0.00102$. For Method D, the simulation studies showed that at least 85% of the variance is explained by the first principal component. Hence, the first principal component was considered as the response variable and the relationship between the response variable and the explanatory variables, x_1 and x_2, was modeled by a multiple linear profile. Finally, the LRT method proposed by Mahmoud *et al.*[17] was used to monitor the new profile. The simulation studies showed that a threshold of 3.73 produce an overall probability of Type I error of $\alpha = 0.04$. The types of shifts investigated in our study are the sustained step shifts taking place after sample $k (k<m)$ for $k = 10$, 15, and 18 and unsustained shifts in sample 5 as well as in samples 5 and 13.

Figure 1 shows the simulated overall probabilities of an out-of-control signal for sustained shifts in the Y-intercept of the first profile from β_{01} to $\beta_{01}+\lambda_0\sigma_1$. Method B has a uniformly better performance than the competing Methods under this type of

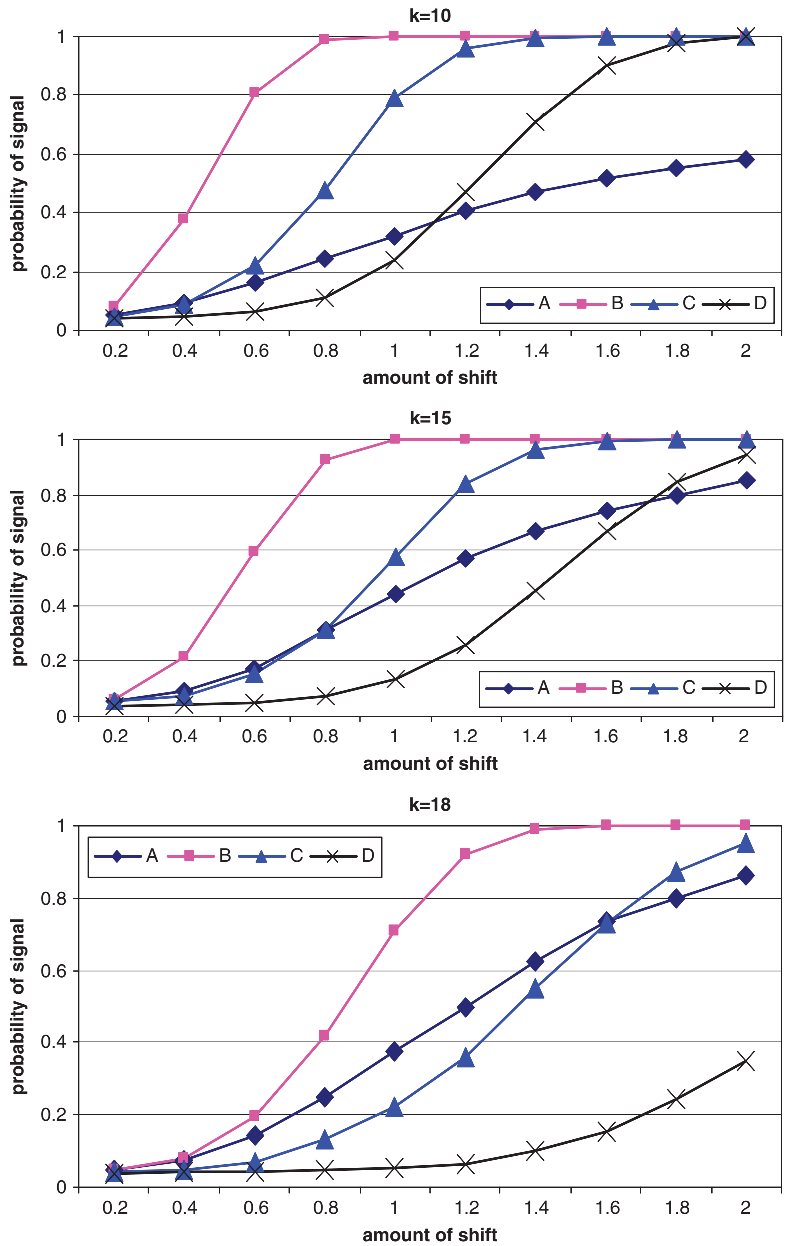

Figure 1. Probability of an out-of-control signal under the sustained shifts from β_{01} to $\beta_{01}+\lambda_0\sigma_1$. This figure is available in colour online at www.interscience.wiley.com/journal/qre

shift for all values of k and λ_0 considered. Method C is better than Method D for all values of k and λ_0. It also performs better than Method A except in small and medium shifts when the location of the shift is after sample 18. Method A performs uniformly better than Method D when the location of the shift is after sample 18. However, for the other shift locations considered in the simulation study, $k=10$ and $k=15$, the performance of Method A is worse than that of Method D as the magnitude of the shifts increases.

The simulated overall probabilities of an out-of-control signal for sustained shifts in the linear coefficient of the first profile from β_{11} to $\beta_{11}+\lambda_1\sigma_1$ are illustrated in Figure 2. Method B is uniformly better than the other methods under the shifts considered. Method C works better than Method D for all values of k and λ_1, and works better than Method A except in small and medium shifts for k equal to 18. Method A has better performance than Method D for $k=18$ and small values of shifts for $k=10$ and $k=15$.

Figure 3 shows the simulated overall probabilities of an out-of-control signal for shifts in the standard deviation of the first profile from σ_1 to $\gamma\sigma_1$. Method B performs better than all the other methods and Method C performs better than Methods A and D. No Methods perform well when the shifts occur after sample 18.

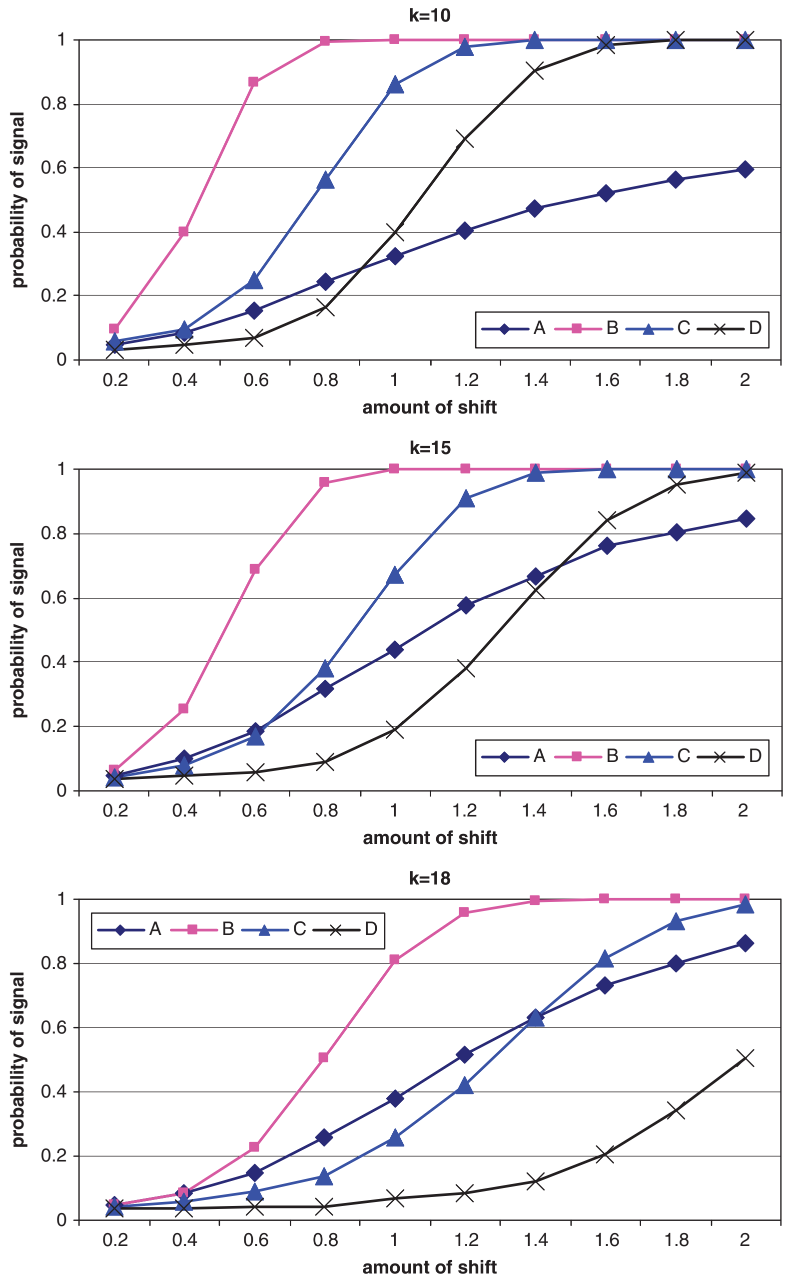

Figure 2. Probability of an out-of-control signal under the sustained shifts from β_{11} to $\beta_{11}+\lambda_1\sigma_1$. This figure is available in colour online at www.interscience.wiley.com/journal/qre

Our simulation results (not reported here) show that when the correlation coefficient of the response variables, ρ_{ij}, increases, the performance of Methods A, B, and C gets better while the performance of Method D deteriorates.

When unsustained shifts occur in the regression parameters, the performance of Methods A and C does not change. However, as shown in Figures 4 and 5, Method D has a uniformly poor performance and Method B does not perform as well as the case of sustained shifts.

Figure 4 shows the performance of the four methods when unsustained shifts in the intercept of the first profile in sample 5 were considered. Method C performs better than the other methods and Method A outperforms Methods B and D. Our simulation studies showed that when the shift takes place in the middle of the data set, the performance of Methods A and C does not change. However, the performance of the Method B gets worse.

Figure 5 shows the performance of the competing methods when unsustained shifts in the intercept of the first profile in samples 5 and 13 are considered. Methods A and C perform better than Method B, and Method D has a uniformly poor performance for all the shifts considered.

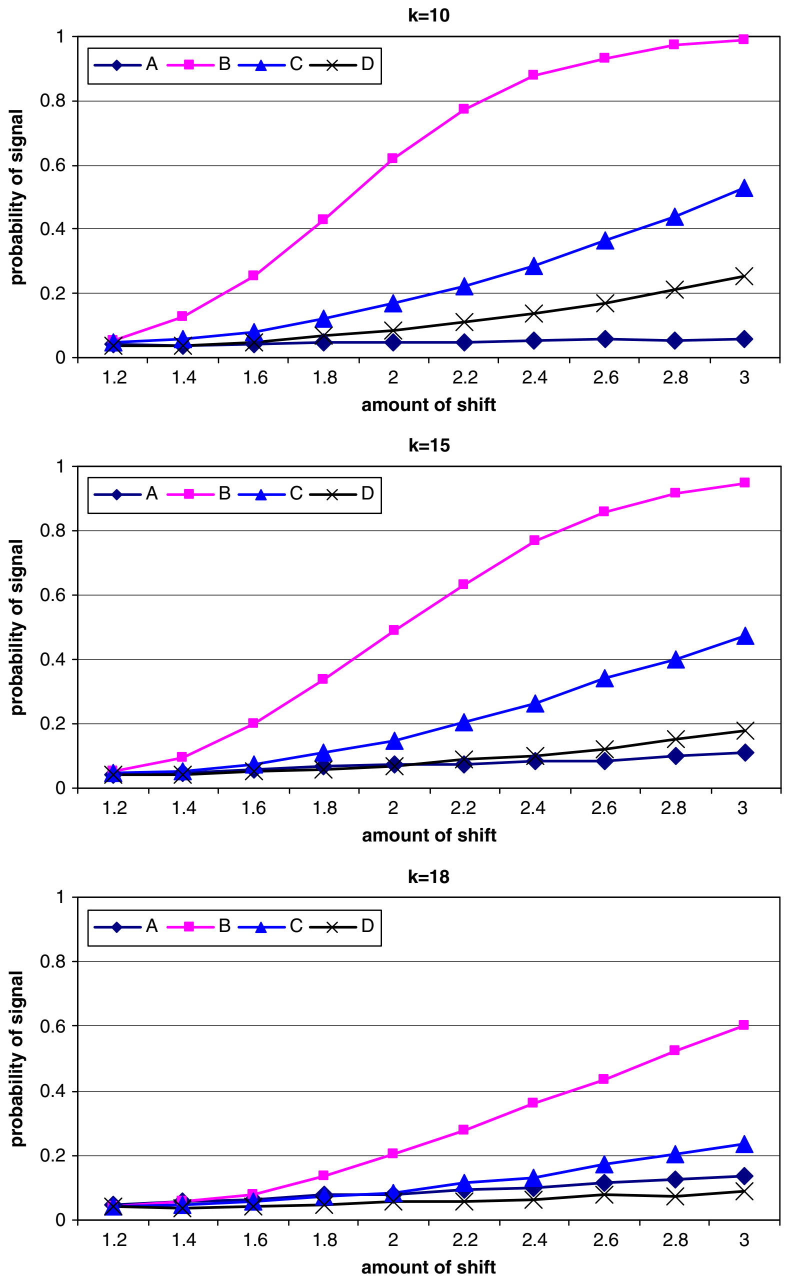

Figure 3. Probability of out-of-control signal under sustained shifts from σ_1 to $\gamma\sigma_1$. This figure is available in colour online at www.interscience.wiley.com/journal/qre

Generally, our simulation studies showed that Method B, which performs well in the presence of sustained shifts, performs worse than Method C in the presence of unsustained shifts. Similar results (not reported here) were obtained for shifts in the linear coefficient of the profiles.

Similar results (not reported here) for both sustained and unsustained shifts were obtained under different values of m and n and the independent variables, x_1 and x_2.

6. Example

In this section, we apply the proposed Phase I multivariate multiple linear profile methods to a calibration application at the NASA Langley Research Center. As discussed in the Introduction section, there are six response variables that are modeled as functions of six explanatory variables. In this case, the replicated calibrations of a force balance used in wind tunnel experiments are investigated. A force balance provides six measurements of forces and torques including normal, axial, and side forces and pitching, rolling, and yawing moments. These forces and moments are illustrated in Figure 6.

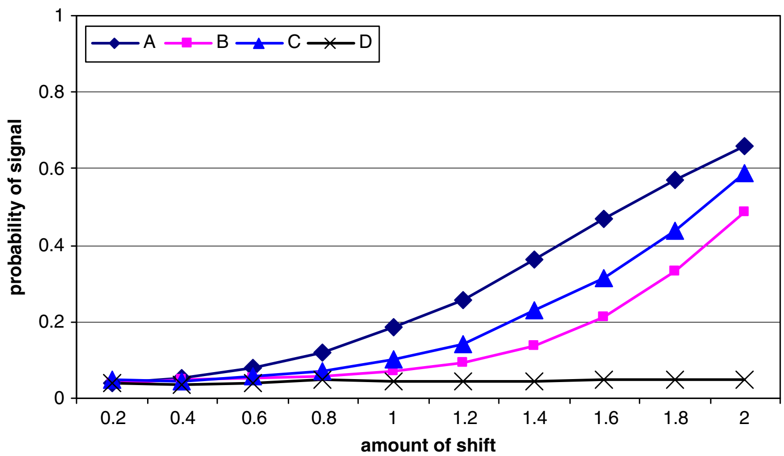

Figure 4. Probability of an out-of-control signal under one unsustained shift from β_{01} to $\beta_{01}+\lambda_1\sigma_1$ in sample 5. This figure is available in colour online at www.interscience.wiley.com/journal/qre

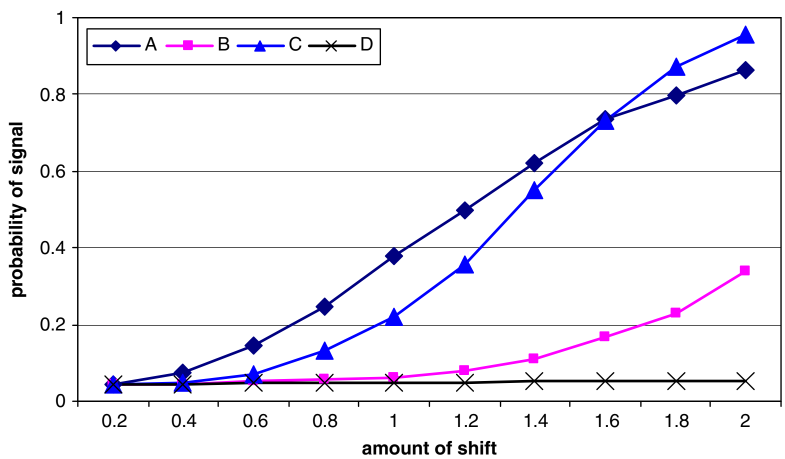

Figure 5. Probability of an out-of-control signal under two unsustained shifts from β_{01} to $\beta_{01}+\lambda_1\sigma_1$ in samples 5 and 13. This figure is available in colour online at www.interscience.wiley.com/journal/qre

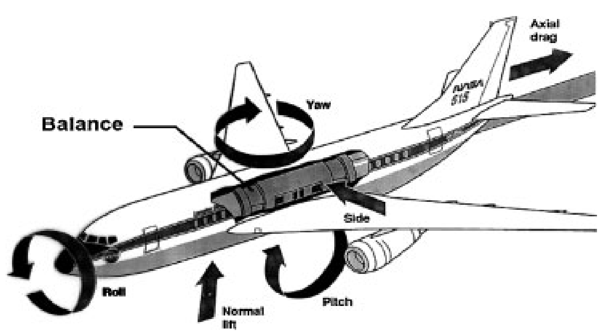

Figure 6. Forces and moments (extracted from Parker *et al.*[36] and reprinted with permission of the American Institute of Aeronautics and Astronautics, Inc.)

Six electrical responses are produced that correspond to the six aforementioned forces and moments. To model the relationship between the applied forces and moments (explanatory variables) and the electrical response variables, the calibration experiment is performed. In this case, response variables are correlated and they are explained by six explanatory variables. Hence, the case can be modeled by a multivariate structure.

We first modeled the relationship between response and explanatory variables as multivariate multiple linear regression by using real data set. The results showed serious violation of some assumptions such as normality and linearity. Since our proposed approaches presented in Section 3 are sensitive to the departure from normality and linearity, the values of the response variables

Table I. In-control regression parameters

	Intercept	Normal	Axial	Pitch	Roll	Yaw	Side
Y_1	−0.05	0.10	−0.01	−0.03	0.26	0	0.03
Y_2	0.48	0.24	21.01	−0.09	0.03	−0.12	0.01
Y_3	0.37	0.09	0.01	6.81	0.04	0.02	−0.03
Y_4	0.04	0	0	0	10.53	−0.47	0.21
Y_5	0.09	−0.02	0	0.01	0.02	7	−0.37
Y_6	0.01	0.04	0	−0.01	0.18	−0.34	11.46

Table II. The values of *lrtc* statistics for the 11 samples

m_1	1	2	3	4	5	6	7	8	9	10
$lrtc_{m1}$	36.081	47.133	59.702	70.999	81.646	92.211	103.5	114.21	45.749	34.764

Table III. The values of the Wilk's statistics and corresponding *p*-values

Sample	Wilk's statistic	*p*-value
1	0.683	0.267
2	0.752	0.854
3	0.691	0.331
4	0.731	0.742
5	0.745	0.821
6	0.233	**0.000**
7	0.637	0.152
8	0.674	0.349
9	0.000	**0.000**
10	0.734	0.760

were simulated. We generated the multivariate multiple linear regression data set by using the real *x*-values used in NASA. The parameters of the first multivariate multiple linear regression profiles were used as the in-control regression parameters. These parameter values are presented in Table I.

The errors were generated from a multivariate normal distribution with mean vector of zero and covariance matrix with unit variances and covariance of 0.2 for each pair of variables. The data set consists of 11 samples each with sample sizes equal to 64, 73, or 74.

To check the performance of the proposed approaches, two out-of-control situations were considered. In the first out-of-control situation, we changed the slope of the independent variable 'Normal' for the first profile from 10 to 0.26 in sample 9. In the second out-of-control situation, the intercepts of the six profiles in sample 6 are changed to 1, 1, 2, 2, 1, and 2, respectively. The data set is available upon request from the authors.

Owing to their superior performance as illustrated in Section 3, we applied only the Wilk's lambda (Method C) and the likelihood ratio (Method B) approaches to this data set. The values of the *lrtc* statistics are computed by using Equation (12) and the results are summarized in Table II as follows:

The threshold limit for the *lrtc* statistics is set equal to 1.52 to achieve a probability of Type I error of 0.04. As the maximum of the $lrtc_{m1}$ statistics is greater than the threshold limit, then we conclude that the process is out-of-control.

The value of the Wilk's lambda statistic for checking the equality of the regression parameters is equal to 0.0000047 and the corresponding *p*-value is equal to zero. These results show that the process is out-of-control. The value of the statistic $(-2\log(M))$ is equal to 204.07 with a *p*-value of 0.848, which indicates that the hypothesis of equal variance–covariance matrices is not rejected.

As discussed in Section 5, since the hypothesis of equal covariance matrices is not rejected, we just decompose the Wilk's statistic to determine the out-of-control sample. The values of Wilk's statistics and the corresponding *p*-values are given in Table III. The *p*-values show that the 6th and 9th samples are out-of-control, which conforms to the simulated data set.

7. Conclusions

In this paper, we extended four methods including likelihood ratio, Wilk's lambda, T^2, and principal components to monitor multivariate multiple linear regression profiles in Phase I. The results showed that the likelihood ratio approach and Wilk's lambda approach are the best competing approaches in detecting both sustained and outlier shifts. Likelihood ratio approach is superior to Wilk's lambda statistics in detecting step shifts in the process parameters. However, the Wilk's lambda approach is better than the likelihood ratio approach in detecting outliers. We also illustrated the use of the proposed approach under a calibration case at the NASA Langley Research Center. In addition, a similar approach proposed by Mahmoud[22] was used to determine the out-of-control samples. The illustrative example showed that the proposed approach performs well in detecting the locations of the out-of-control samples.

Acknowledgements

The authors acknowledge the precious comments of the editor and the anonymous reviewer that led to the improvement of this paper. The authors are also grateful to Peter A. Parker, Senior Research Scientist at the NASA Langley Research Center for providing the initial data set used in the example. Dr Noorossana's research is partially supported by a grant from Iran National Science Foundation.

References

1. Stover FS, Brill RV. Statistical quality control applied to ion chromatography calibrations. *Journal of Chromatography A* 1998; **804**(1–2):37–43.
2. Kang L, Albin SL. On-line monitoring when the process yields a linear profile. *Journal of Quality Technology* 2000; **32**(4):418–426.
3. Mahmoud MA, Woodall WH. Phase I analysis of linear profiles with calibration applications. *Technometrics* 2004; **46**(4):380–391.
4. Woodall WH, Spitzner DJ, Montgomery DC, Gupta S. Using control charts to monitor process and product quality profiles. *Journal of Quality Technology* 2004; **36**(3):309–320.
5. Wang K, Tsung F. Using profile monitoring techniques for a data-rich environment with huge sample size. *Quality and Reliability Engineering International* 2005; **21**(7):677–688.
6. Montgomery DC. *Introduction to Statistical Quality Control*. Wiley: New York, 2005.
7. Woodall WH. Current research on profile monitoring. *Revista Producão* 2007; **17**(3):420–425.
8. Zou C, Tsung F, Wang Z. Monitoring general linear profiles using multivariate exponentially weighted moving average schemes. *Technometrics* 2007; **49**(4):395–408.
9. Kim K, Mahmoud MA, Woodall WH. On the monitoring of linear profiles. *Journal of Quality Technology* 2003; **35**(3):317–328.
10. Noorossana R, Amiri A, Vaghefi SA, Roghanian E. Monitoring quality characteristics using linear profile. *Proceedings of the 3rd International Industrial Engineering Conference,* Tehran, Iran, 2004.
11. Gupta S, Montgomery DC, Woodall WH. Performance evaluation of two methods for online monitoring of linear calibration profiles. *International Journal of Production Research* 2006; **44**(10):1927–1942.
12. Zou C, Zhang Y, Wang Z. Control chart based on change-point model for monitoring linear profiles. *IIE Transactions* 2006; **38**(12):1093–1103.
13. Zhang J, Li Z, Wang Z. Control chart based on likelihood ratio for monitoring linear profiles. *Computational Statistics and Data Analysis* 2009; **53**:1440–1448.
14. Saghaei A, Mehrjoo M, Amiri A. A CUSUM-based method for monitoring simple linear profiles. *The International Journal of Advanced Manufacturing Technology* 2009; DOI: 10.1007/s00170-009-2063-2.
15. Mahmoud MA, Morgan JP, Woodall WH. The monitoring of simple linear regression profiles with two observations per sample. *Journal of Applied Statistics* 2009; to appear.
16. Mestek O, Pavlik J, Suchánek M. Multivariate control charts: Control charts for calibration curves. Fresenius'. *Journal of Analytical Chemistry* 1994; **350**(6):344–351.
17. Mahmoud MA, Parker PA, Woodall WH, Hawkins DM. A change point method for linear profile data. *Quality and Reliability Engineering International* 2007; **23**(2):247–268.
18. Noorossana R, Amiri A, Soleimani P. On the monitoring of autocorrelated linear profiles. *Communications in Statistics—Theory and Methods* 2008; **37**(3):425–442.
19. Soleimani P, Noorossana R, Amiri A. Simple linear profiles monitoring in the presence of within profile autocorrelation. *Computers and Industrial Engineering* 2009; DOI: 10.1007/j.cie.2009.04.005.
20. Noorossana R, Vaghefi SA, Amiri A. The effect of non-normality on monitoring linear profiles. *Proceedings of the 2nd International Industrial Engineering Conference*, Riyadh, Saudi Arabia, 2004.
21. Noorossana R, Vaghefi A, Dorri M. The effect of non-normality on performance of linear profile monitoring. *Proceedings of the 2008 IEEE Industrial Engineering and Engineering Management*, Singapore, 8–11 December 2008.
22. Mahmoud MA. Phase I analysis of multiple linear regression profiles. *Communications in Statistics, Simulation and Computation* 2008; **37**(10):2106–2130.
23. Jensen WA, Birch JB, Woodall WH. Monitoring correlation within linear profiles using mixed models. *Journal of Quality Technology* 2008; **40**(2):167–183.
24. Kazemzadeh RB, Noorossana R, Amiri A. Phase I monitoring of polynomial profiles. *Communications in Statistics—Theory and Methods* 2008; **37**(10):1671–1686.
25. Kazemzadeh RB, Noorossana R, Amiri A. Monitoring polynomial profiles in quality control applications. *The International Journal of Advanced Manufacturing Technology* 2009; **42**(7):703–712.
26. Kazemzadeh RB, Noorossana R, Amiri A. Phase II monitoring of autocorrelated polynomial profiles in AR(1) processes. *International Journal of Science and Technology, Scientia Iranica* 2009; to appear.
27. Jin J, Shi J. Feature-preserving data compression of stamping tonnage information using wavelets. *Technometrics* 1999; **41**(4):327–339.
28. Walker E, Wright S. Comparing curves using additive models. *Journal of Quality Technology* 2002; **34**(1):118–129.
29. Ding Y, Zeng L, Zhou S. Phase I analysis for monitoring nonlinear profiles in manufacturing processes. *Journal of Quality Technology* 2006; **38**(3):199–216.
30. Williams JD, Woodall WH, Birch JB. Statistical monitoring of nonlinear product and process quality profiles. *Quality and Reliability Engineering International* 2007; **23**(8):925–941.
31. Moguerza JM, Muñoz A, Psarakis S. *Monitoring Nonlinear Profiles Using Support Vector Machines* (*Lecture Notes in Computer Science*, vol. 4789). Springer: Berlin, 2007; 574–583.
32. Vaghefi A, Tajbakhsh SD, Noorossana R. Phase II monitoring of nonlinear profiles. *Communications in Statistics—Theory and Methods* 2009; **38**(11):1834–1851.
33. Jensen WA, Birch JB. Profile monitoring via nonlinear mixed model. *Journal of Quality Technology* 2009; **41**(1):18–34.

34. Noorossana R, Eyvazian M, Vaghefi A, Amiri A. Phase I monitoring of multivariate simple linear profiles. *Journal of Quality Technology* 2009; under revision by the authors to resubmit.
35. Noorossana R, Eyvazian M, Vaghefi A. Phase II monitoring of multivariate simple linear profiles. *Computers and Industrial Engineering* 2009; under revision by the authors to resubmit.
36. Parker PA, Morton M, Draper NR, Line WP. A single-vector force calibration method featuring the modern design of experiments. *Proceedings of the American Institute of Aeronautics and Astronautics 39th Aerospace Sciences Meeting and Exhibit 2001*, Reno, Nevada, 2001.
37. Williams JD, Woodall WH, Birch JB, Sullivan JH. Distribution of Hotelling's T^2 statistic based on the successive differences estimator. *Journal of Quality Technology* 2006; **38**(3):217–229.
38. Rencher AC. *Methods of Multivariate Analysis*. Wiley: New York, 2002.
39. Lee JC, Chang TC, Krishnaiah PR. Approximations to the distributions of the likelihood ratio statistics for testing certain structures on the covariance matrices of real multivariate normal populations. *Multivariate Analysis* 1977; **4**:105–118.

Appendix A

Assuming H_0 in Equation (7) to be true, the logarithm of the likelihood function based on the model in Equation (6) is as follows:

$$l_0 = -\frac{N}{2}[p\log(2\pi)+\log|\Sigma|] - \frac{1}{2}\sum_{i=1}^{N}(\mathbf{y_i}-\mathbf{x_i}\mathbf{B})\Sigma^{-1}(\mathbf{y}_i-\mathbf{x}_i B)^{\mathrm{T}} \tag{A1}$$

We obtain the following equation by replacing the parameters by their maximum likelihood estimators:

$$l_0 = -\frac{N}{2}[p\log(2\pi)+\log|\hat{\Sigma}|] - \frac{1}{2}\sum_{i=1}^{N}(\mathbf{y}_i-\mathbf{x}_i\hat{\mathbf{B}})\hat{\Sigma}^{-1}(\mathbf{y}_i-\mathbf{x}_i\hat{\mathbf{B}})^{\mathrm{T}} \tag{A2}$$

leading to

$$l_0 = -\frac{N}{2}[p\log(2\pi)+\log|\hat{\Sigma}|] - \frac{1}{2}\sum_{i=1}^{N}(\mathbf{y}_i-\mathbf{x}_i\hat{\mathbf{B}})\left[\frac{(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}})^{\mathbf{T}}(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}})}{N}\right]^{-1}(\mathbf{y}_i-\mathbf{x}_i\hat{\mathbf{B}})^{\mathrm{T}} \tag{A3}$$

where $(\mathbf{y}_i-\mathbf{x}_i\hat{\mathbf{B}})$ is the ith row of the matrix $(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}})$.

In each $N\times p$ matrix such as $\mathbf{Z}$, it can be shown that the following equation is true:

$$\sum_{i=1}^{N} z_i[\mathbf{Z}^{\mathrm{T}}\mathbf{Z}]^{-1}z_i^{\mathrm{T}} = p \tag{A4}$$

where z_i is the ith row of the matrix $\mathbf{Z}$.

Replacing Equation (A4) into Equation (A3), we obtain the following equation:

$$l_0 = -\frac{N}{2}[p\log(2\pi)+\log|\hat{\Sigma}|] - \frac{Np}{2} \tag{A5}$$

Similarly, for all the samples prior to and following the change point m_1, the maximum of the log-likelihood functions are as follows, respectively:

$$l_1 = -\frac{N_1}{2}[p\log(2\pi)+\log|\hat{\Sigma}_1|] - \frac{N_1 p}{2} \tag{A6}$$

$$l_2 = -\frac{N_2}{2}[p\log(2\pi)+\log|\hat{\Sigma}_2|] - \frac{N_2 p}{2} \tag{A7}$$

Under H_1, the maximum of the log-likelihood function based on the model in Equation (6) is as follows:

$$l_a = l_1 + l_2 \tag{A8}$$

The likelihood ratio test statistic is given as follows:

$$lrt_{m1} = -2(l_0 - l_a) \tag{A9}$$

replacing Equations (A5) and (A8) in Equation (A9), we obtain

$$lrt_{m1} = N\log|\hat{\Sigma}| - N_1\log|\hat{\Sigma}_1| - N_2\log|\hat{\Sigma}_2| \tag{A10}$$

Authors' biographies

Rassoul Noorossana is Professor of Statistics at Iran University of Science and Technology. He received his BS in Engineering from the Louisiana State University in 1983 and his MS and PhD in Engineering Management and Statistics from the University of Louisiana in 1986 and 1990, respectively. His primary research interests include statistical process control, process optimization, and statistical analysis. He is the editor of the *Journal of Industrial Engineering International* and serves on the editorial review board of many journals. He is a senior member of the American Society for Quality, Iranian Society for Quality, Iranian Statistical Association, and Industrial Engineering Society.

Majid Eyvazian is a PhD candidate in Industrial Engineering at Iran University of Science and Technology. His primary research interests include applied probability and engineering statistics and its applications in statistical quality control. He has been teaching Probability Theory and Engineering Statistics in the industrial engineering departments at Iran University of Science and Technology and Khajeh Nasir Toosi University of Technology.

Amirhossein Amiri is a lecturer at Shahed University. He holds a BS, MS, and PhD in Industrial Engineering from Khajeh Nasir Toosi University of Technology, Iran University of Science and Technology and Tarbiat Modares University, respectively. He is a member of Iranian Statistical Association. His research interests are statistical quality control, profile monitoring, and Six Sigma.

Mahmoud A. Mahmoud is Assistant Professor of Statistics at Cairo University. He holds a BS (1992) and an MS (1997) in Statistics from Cairo University, and a PhD (2004) in Statistics from Virginia Tech. His primary area of interest is statistical quality control and improvement.