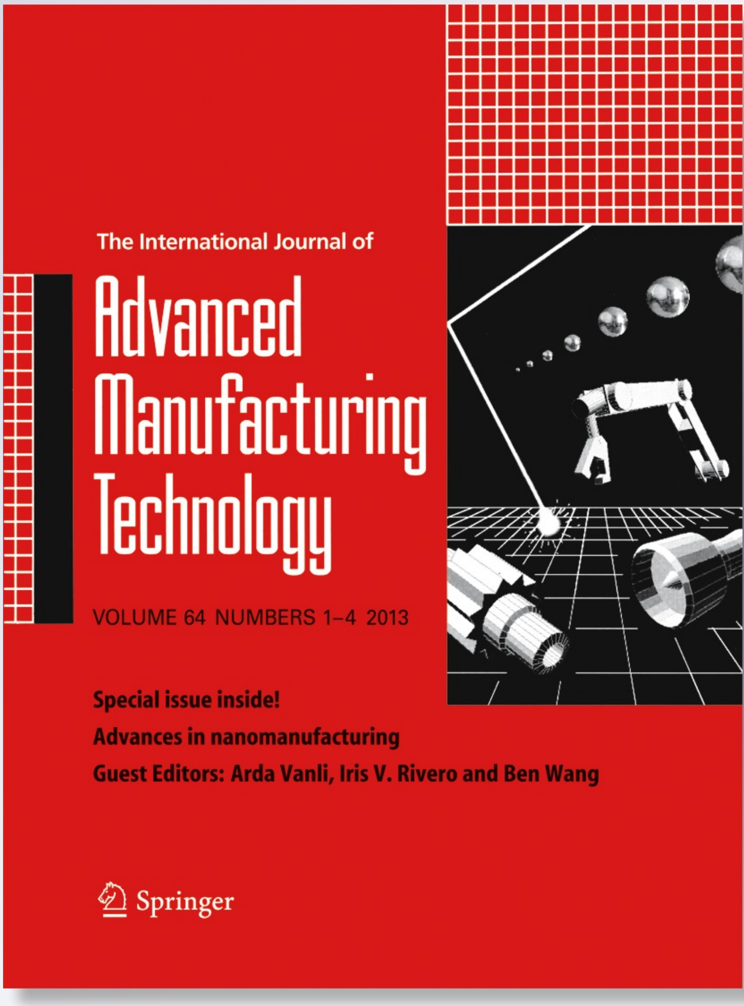
Generalized linear mixed model for monitoring autocorrelated logistic regression profiles

Mehdi Koosha & Amirhossein Amiri

The International Journal of Advanced Manufacturing Technology

ISSN 0268-3768
Volume 64
Combined 1-4

Int J Adv Manuf Technol (2013)
64:487-495
DOI 10.1007/s00170-012-4018-2



Int J Adv Manuf Technol (2013) 64:487–495
DOI 10.1007/s00170-012-4018-2

ORIGINAL ARTICLE

Generalized linear mixed model for monitoring autocorrelated logistic regression profiles

Mehdi Koosha · Amirhossein Amiri

Received: 21 May 2011 / Accepted: 19 February 2012 / Published online: 9 March 2012

Abstract Profile monitoring is used to monitor the regression relationship between a response variable and one or more explanatory variables over time. Many researches have been done in this area, but in most of them, the distribution of the response variable is assumed to be normal. However, this assumption is violated in many real case problems. In these instances, classic methods cannot be used for monitoring the profiles. For example, when the response variable is binary, logistic regression methods should be used rather than ordinary least square or other classic regression methods. There are some methods for monitoring logistic profiles in the literature, but the basic assumption of these methods is the independency of the consecutive observations, while this assumption is violated in some instances for example when the successive samples are taken in short intervals. This paper considers the effect of autocorrelation presence between the observations in different levels of the independent variable in a logistic regression profile on the monitoring procedure (T^2 control chart) and proposes two remedies to account for the autocorrelation within logistic profiles. In one of the remedies, upper control limit of the traditional T^2 control chart is modified. In the second one, we use a generalized linear mixed model (GLMM) to estimate the regression parameters and then use the T^2 control chart for monitoring autocorrelated logistic regression profiles. Simulation studies show the better performance of T^2 control chart when the regression parameters are estimated by the GLMM method under both step shifts and drifts.

M. Koosha · A. Amiri (✉)
Department of Industrial Engineering, Faculty of Engineering,
Shahed University,
Tehran, Iran
e-mail: amirhossein.amiri@gmail.com

A. Amiri
e-mail: Amiri@shahed.acr.ir

M. Koosha
e-mail: koosha.mehdi@gmail.com

Keywords Logistic regression · Profile monitoring · Autocorrelation · Generalized linear mixed models · T^2 control chart · Phase I · Statistical process control

1 Introduction

There are many cases in which monitoring the relationship between response variable and explanatory variables, which is called as profile, is desirable instead of monitoring univariate or multivariate quality characteristics [19]. According to the type of regression relationship between response and predictor variables, profiles are classified to different types such as simple linear profiles, nonlinear profiles, multiple linear profiles, polynomial profiles, logistic profiles, etc. Mestek et al. [18], Stover and Brill [23], Jin and Shi [9], Boeing [3], and Amiri et al. [2] presented some applications of profiles. Different methods by researchers have been proposed for monitoring different types of profiles. Kang and Albin [11], Kim et al. [14], Mahmoud and Woodal [17], and Mahmoud et al. [16] proposed some methods for monitoring simple linear profiles. Other methods for monitoring simple linear profiles are proposed by Saghaie et al. [21], Zhang et al. [28], Chen and Nembhard [5], and Zhu and Lin [29]. Sometimes, this relationship is characterized by more complicated regression functions. For example, Zou et al. [30] and Mahmoud [15] proposed some methods for monitoring multiple linear profiles. Monitoring multivariate profiles was considered by Noorossana et al. [20] and Eyvazian et al. [6]. Authors including Vaghefi et al. [24] and Williams et al. [26] proposed methods to monitor nonlinear profiles. Kazemzadeh et al. [12, 13] proposed some methods to monitor polynomial profiles in Phases I and II, respectively.

In all of the aforementioned researches, consecutive observations either between or within profiles are assumed to be independent. In many real problems, the assumption of independency between the observations is violated. Some methods have been proposed by researchers in order to violate this assumption. Jensen et al. [8] proposed a method based on linear mixed models to monitor autocorrelated linear profiles. A method based on transformation was proposed by Soleimani et al. [22] to eliminate the effect of autocorrelation in monitoring of simple linear profiles. Jensen and Birch [7] have proposed nonlinear mixed models for monitoring nonlinear-autocorrelated profiles.

In all of these researches, it is assumed that the quality characteristic to be monitored is normally distributed. But in many cases, for example when the distribution of response variable is binary, this assumption is violated. Profiles with binary response variables are called logistic profiles. In order to monitor this kind of profiles, Yeh et al. [27] proposed five T^2-based methods. In all of these methods, it is assumed that the successive observations are independent. In this paper, we investigate the effect of autocorrelation on the performance of T^2 procedures proposed by Yeh et al. [27] in monitoring logistic profiles. Then, we propose upper control limit (UCL) modification method as well as GLMM method to account for the effect of autocorrelation on the performance of SPC monitoring procedures in Phase I. In this Phase, the stability of a process is checked and unknown parameters are estimated. The performance of the methods in Phase I is evaluated by power criterion.

The rest of this paper is classified as follows: Section 2 proposes the algorithm of parameter estimation in logistic regression profiles. Section 3 describes T^2 control charts proposed by Yeh et al. [27] for monitoring logistic profiles. Autocorrelation and its prevalent structures are described in Section 4. Section 5 illustrates the effect of neglecting the presence of autocorrelation on the performance of the best T^2 control chart proposed by Yeh et al. [27]. UCL modification method is proposed in Section 6 and applied on all of the methods proposed by Yeh et al. [27] as well as GLMM method. Section 7 compares the performance of the proposed methods. Conclusions and some future researches are given in the final section.

2 Parameter estimation in logistic regression profiles

Least squares error (LSE) method can be applied in order to estimate the parameters of a profile when the observations follow normal distribution. However, in the case that response variables are binary, the LSE method produces the estimators that do not have minimum variance. Hence, the maximum likelihood estimation (MLE) method should be used instead. An iterative algorithm is used to estimate the parameters in logistic regression profiles by using the MLE method [1].

Suppose that there are m independent profiles ($i=1, 2, \ldots, m$) and the observations in each level of the explanatory variable ($j=1,2, \ldots,p$) are repeated for n times. As each observation is binary, the sum of these n observations follows binomial distribution and is considered as the response variable. For parameter estimation in a logistic profile, Yeh et al. [27] applied an algorithm which is explained below.

Start the algorithm with an initial $\widehat{\beta}^{(0)}$ which could be obtained by any method such as LSE.

Step 1 Let $L=0$; Define the value of X (vector of explanatory variables) and observe the response variables in each level. (L is a counter for number of iterations.)

Step 2 A vector of $\pi_i^{(L)}$ is needed that contains elements of $\pi_{ij}^{(L)}$ which shows the probability of success in an experiment in jth level of X in ith profile.

Step 3 Set $L=L+1$ and compute $\eta_i^{(L)} = X_i^T \beta^{(L)}$.

The Logit link function in Eq. (1) is used to make a relationship between the values of $\eta_i^{(L)}$ and expected value of each observation (${\pi_i}^{(L)}$).

$$\ln\left(\frac{\pi_i^{(L)}}{1-\pi_i^{(L)}}\right) = \eta_i^{(L)} \tag{1}$$

Step 4 Update the values of ${\pi_i}^{(L)}$ with Eq. (2) which is derived from Eq. (1)

$${\pi_i}^{(L)} = \frac{e^{{\eta_i}^{(L)}}}{1+e^{{\eta_i}^{(L)}}} \tag{2}$$

Step 5 Compute W and μ by Eqs. (3) and (4), respectively.

$$W_i^{(L)} = \mathrm{diag}\left(n{\pi_{ij}}^{(L)}\left(1-\pi_{ij}^{(L)}\right)\right) \tag{3}$$

and

$$\mu_i^{(L)} = \left[n{\pi_{ij}}^{(L)}\right], \tag{4}$$

where W_i is a diagonal matrix which its main diagonal entries are the variance of various levels of logistic profiles and μ_i is the vector of expected values for response variables.

Step 6 Compute $q_i^{(L)}$ by applying Eq. (5):

$$q_i^{(L)} = \eta_i^{(L)} + \left(W_i^{(L)}\right)^{-1}\left(Y_i - \mu^{(L)}\right), \tag{5}$$

where $\mathbf{Y}_i$ is vector of observations in ith profile.

Step 7 Update the estimation for β by Eq. (6).

$$\beta^{(L+1)} = \left(X^T {W_i}^{(L)} X\right)^{-1} X^T {W_i}^{(L)} {q_i}^{(L)} \tag{6}$$

If $\beta^{(L+1)}-\beta^{(L)}<\varepsilon$, then $\beta^{(L+1)}$ can be used as the final estimation of β. Otherwise go to step 3. ε is a predetermined small value.

We used this parameter estimation procedure for first proposed method of this paper.

3 Existing T^2 methods

Yeh et al. [27] proposedfiveT^2-based methods including T^2 based on sample average and intra-profile pooling (T_{I}^2), T^2 based on sample mean and covariance matrix (T_{H}^2), T^2based on sample average and moving ranges (T_{R}^2), T^2 based on minimum volume ellipsoid (T_{MVE}^2), and T^2 based on minimum covariance determinant (T_{MCD}^2). These methods will be briefly introduced in the following sections.

3.1 T_{I}^2 (T^2 based on sample average and intra-profile pooling)

In this method, the statistic to be monitored on theT^2 control chart is as Eq. (7).

$$T_{\mathrm{I},i}^2 = (\widehat{\beta}_i - \overline{\beta})' S_{\mathrm{I}}^{-1} (\widehat{\beta}_i - \overline{\beta}) \tag{7}$$

In Eq. (7), S_I is calculated with Eq. (8).

$$S_{\mathrm{I}} = \frac{1}{m} \sum_{i=1}^{m} \mathrm{var}(\widehat{\beta}_i) = \frac{1}{m} \sum_{i=1}^{m} (X' W_i X)^{-1} \tag{8}$$

3.2 T_H^2 (T^2 based on sample mean and covariance matrix)

The statistic of this method is shown in Eq. (9):

$$T_{H,i}^2 = (\widehat{\beta}_i - \overline{\beta})' S_H^{-1} (\widehat{\beta}_i - \overline{\beta}), \tag{9}$$

where S_{H} is calculated with Eq. (10)

$$S_{\mathrm{H}} = \frac{1}{m-1} \sum_{i=1}^{m} (\widehat{\beta}_i - \overline{\beta})(\widehat{\beta}_i - \overline{\beta})' \tag{10}$$

3.3 T_{R}^2 (T^2 based on sample average and moving ranges)

The statistic of this method is declared in Eq. (11):

$$T_{\mathrm{R},i}^2 = (\widehat{\beta}_i - \overline{\beta})' S_{\mathrm{R}}^{-1} (\widehat{\beta}_i - \overline{\beta}), \tag{11}$$

where S_{R} should be computed by Eq. (12):

$$S_{\mathrm{R}} = \frac{1}{2(m-1)} \sum_{i=1}^{m-1} (\widehat{\beta}_{i+1} - \widehat{\beta}_i)(\widehat{\beta}_{i+1} - \widehat{\beta}_i)' \tag{12}$$

In the above equations, $\widehat{\beta}_i$ is a vector of estimated coefficients for ith profile and $\overline{\beta}$ is vector of the mean of the estimated logistic regression parameters.

3.4 T_{MVE}^2 (T^2 based on minimum volume ellipsoid)

For this method, the T^2 statistic is given in Eq. (13):

$$T_{\mathrm{E},i}^2 = (\widehat{\beta}_i - \widehat{\beta}_{\mathrm{E}})' S_{\mathrm{E}}^{-1} (\widehat{\beta}_i - \widehat{\beta}_{\mathrm{E}}), \tag{13}$$

where β_{E} and S_{E} are estimated by MVE method. See Vargas [25] for more details about MVE method.

3.5 T_{MCD}^2 (T^2 based on minimum covariance determinant)

The T^2 statistic for this method is as follows:

$$T_{\mathrm{D},i}^2 = (\widehat{\beta}_i - \widehat{\beta}_{\mathrm{D}})' S_{\mathrm{D}}^{-1} (\widehat{\beta}_i - \widehat{\beta}_{\mathrm{D}}) \tag{14}$$

β_{D} and S_{D} are estimated by MCD method. Refer to Vargas [25] for more details.

4 Autocorrelation

In the area of profile monitoring, most researchers assumed that the observations are independent in their approaches. However, this assumption is violated in some cases; for example when the successive observations are taken in short intervals. There are some popular structures for autocorrelation in the literature of statistical process control such as autoregressive in which each observation is dependent to its previous observation with correlation coefficient ρ. Another type of prevalent structures for autocorrelation is symmetric in which each two observations are correlated to each other with the same value of correlation coefficient ρ. So, the correlation matrix for this case is as follows:

$$\begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix} \tag{15}$$

In this paper, we use this kind of autocorrelation structure to show its effect on the performance of the methods proposed by Yeh et al. [27] and to evaluate the performance of the proposed methods.

5 The effect of autocorrelation

The presence of autocorrelation can have a destructive effect on the performance of the T^2 methods when it is neglected. In order to demonstrate this effect, we have generated m=30 profiles in which the observations of each profile in various levels of explanatory variables are correlated. Vector of explanatory variables (design matrix) is similar to the example of Yeh et al. [27]. The entries in the first row of the design matrix are all equal to 1 and the second row entries are equal to log(0.1), log(0.2), …, log(0.9). We investigate the effect of autocorrelation on the performance of T_I^2 as the most powerful method in detecting shifts, and this effect can be expanded to other methods.

In order to generate autocorrelated responses, normal to anything (NORTA) algorithm is used, and a nine-variate binomial response with n=30 and the following vector of π_i is generated for ith profile. NORTA algorithm is explained below.

$$\pi_i = \frac{\exp\left(X_i^T \begin{bmatrix}\beta_0\\ \beta_1\end{bmatrix}\right)}{1+\exp\left(X_i^T \begin{bmatrix}\beta_0\\ \beta_1\end{bmatrix}\right)} \tag{16}$$

5.1 NORTA algorithm

NORTA algorithm is used to generate vectors of multivariate non-normal distributions such as binomial, Poisson, Gamma, and so on [4].

The steps of this algorithm are explained below:

1. Generate a multivariate random vector of normal distribution with mean vector equals to 0 and a random variance–covariance matrix.
2. Compute the cumulative probability of each element of the generated multivariate normal vector.
3. Find the percentile of the desired distribution (for example binomial) in a way that its cumulative probability would be equal to the cumulative probability computed is step 2.
4. Compute the variance–covariance matrix of the vector computed in step 3 and compare it with the predetermined one.
5. If the amount of difference between these two variance–covariance matrices is smaller than a predetermined acceptable error, consider the vector generated in step 3 as a multivariate random vector from the desired distribution. Otherwise, change the elements of variance covariance matrix of the initial multivariate normal vector in such a manner that the resulted variance–covariance matrix of the desired distribution be closer to the predetermined one and repeat the algorithm.

The correlation coefficients (ρ) in this paper are considered equal to 0.1, 0.15, 0.2, and 0.25. Since the assumed correlation structure is symmetric, this coefficient is constant between each two arbitrary observations in each profile. As shown in Fig. 1, as the correlation coefficient increases, the probability of out-of-control signal also increases due to increasing in the probability of type I error for various amounts of step shift.

It should be noted that shifts in Fig. 1 are step shifts in the scale of σ. For example, (1, 2) means that the regression parameters are changed to $\beta_{0,\text{out}} = \beta_{0,\text{in}} + 1\sigma_0$ and $\beta_{1,\text{out}} = \beta_{1,\text{in}} + 2\sigma_1$ in which σ_1 and σ_1 are computed as follows:

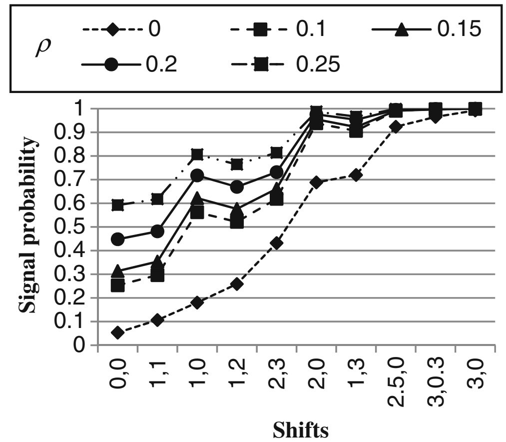


Fig. 1 The effect of autocorrelation on the performance of T_I^2 control chart under step shift

$$S_{i0} = \begin{pmatrix}\sigma_0^2 & \rho\sigma_0\sigma_1\\ \rho\sigma_0\sigma_1 & \sigma_1^2\end{pmatrix} = \left(X_i^T W X_i\right)^{-1} \tag{17}$$

Note that in this simulation study, we have generated the step shift on the second half of the profiles (16th to 30th profile). $\beta_{0,\text{ in}}$ and $\beta_{1,\text{ in}}$ are considered as 3 and 2, respectively.

Figure 2 illustrates the effect of autocorrelation with a drift from second profile. In this type of shift, for example (1, 2) means that the regression parameters are changed as follows:

$$\beta_{0,\text{out},i} = \beta_{0,\text{in}} + 1 \times \frac{i-1}{m-1}\sigma_0$$

$$\beta_{1,out,i} = \beta_{1,in} + 2 \times \frac{i-1}{m-1}\sigma_0$$

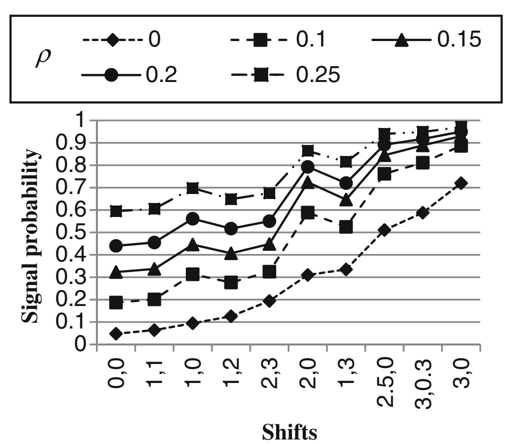


Fig. 2 The effect of autocorrelation on the performance of T_I^2 control chart under drift

Table 1 The performance of the T_{I}^2 control charts with modified UCL under different step shifts

	UCL				
Shift	15.0796 ρ=0	21.051 ρ=0.1	25.2600 ρ=0.15	29.2110 ρ=0.2	34.2090 ρ=0.25
0.0	0.0534	0.0500	0.0526	0.0530	0.0506
1.1	0.1064	0.0650	0.0640	0.0568	0.0568
1.0	0.1802	0.1632	0.1670	0.1366	0.1338
1.2	0.2588	0.1420	0.1142	0.0824	0.0742
2.3	0.4320	0.1976	0.1402	0.1004	0.0752
2.0	0.6882	0.5726	0.4929	0.4194	0.3636
1.3	0.7196	0.4798	0.3490	0.2578	0.2018
25.0	0.9236	0.7956	0.7122	0.6024	0.5320
30.3	0.9652	0.8626	0.7664	0.6796	0.5868
3.0	0.9926	0.9300	0.8636	0.7782	0.6730

Table 3 The performance of the T_{H}^2 control charts with modified UCL under different step shifts

	UCL				
Shift	11.8166 ρ=0	11.8220 ρ=0.1	11.8250 ρ=0.15	11.8290 ρ=0.2	11.8500 ρ=0.25
0.0	0.0507	0.0523	0.0493	0.0536	0.0538
1.1	0.0480	0.0487	0.0513	0.0480	0.0557
1.0	0.0427	0.0500	0.0517	0.0500	0.0577
1.2	0.0355	0.0347	0.0383	0.0303	0.0403
2.3	0.0349	0.0287	0.0350	0.0340	0.0373
2.0	0.0363	0.0443	0.0537	0.0550	0.0543
1.3	0.0360	0.0387	0.0357	0.0330	0.0353
25.0	0.0407	0.0437	0.0530	0.0587	0.0540
30.3	0.0355	0.0573	0.0503	0.0563	0.0637
3.0	0.0356	0.0463	0.0523	0.0607	0.0640

6 Proposed methods

6.1 UCL modification method

As shown in the previous section, the autocorrelation between responses in different levels of a logistic profile leads to increasing the probability of type I error, and hence, the misleading results for probability of type II error. To account for this problem, we propose modifying the upper control limits of the T^2 control charts proposed by Yeh et al. [27]. In this method, we increase the UCL of all T^2 control charts to overcome the increasing in probability of type I error resulted from the autocorrelation between responses. This method adjusts the probability of type I error; therefore, the results of power for all control charts are obtained correctly.

Now, through simulation studies, we modify the UCL of the proposed T^2 methods by Yeh et al. [27] for autocorrelated profiles to obtain a specific probability of type I error equals to 0.05 under different autocorrelation coefficients and different shifts. The corresponding UCL and the power of the control chart for considered autocorrelation coefficients under step shift are summarized in Tables 1, 2, 3, 4, and 5.

Results show that T_{I}^2 has the best performance among the methods for monitoring autocorrelated logistic profiles similar to independent logistic profiles cases in detecting step shifts. In addition to the step shifts, we also consider the performance of the T^2 methods by Yeh et al. [27] under drift in the regression parameters of autocorrelated profiles (Tables 6, 7, 8, 9, and 10).

Table 2 The performance of the T_{R}^2 control charts with modified UCL under different step shifts

	UCL				
Shift	13.4951 ρ=0	13.5550 ρ=0.1	13.5710 ρ=0.15	13.5750 ρ=0.2	13.5800 ρ=0.25
0.0	0.0491	0.0541	0.0520	0.0524	0.0500
1.1	0.1116	0.1068	0.1057	0.1116	0.1156
1.0	0.2522	0.1700	0.1568	0.1540	0.1396
1.2	0.3332	0.3197	0.3410	0.3370	0.3578
2.3	0.4973	0.5030	0.4983	0.5403	0.5580
2.0	0.6687	0.4927	0.4380	0.4013	0.3732
1.3	0.6924	0.6784	0.6577	0.6427	0.6510
25.0	0.8163	0.6643	0.5997	0.5543	0.5120
30.3	0.8446	0.7040	0.6523	0.6130	0.5548
3.0	0.8861	0.7787	0.7283	0.6810	0.6492

Table 4 The performance of the T_{MVE}^2 control charts with modified UCL under different step shifts

	UCL				
Shift	29.1826 ρ=0	29.4436 ρ=0.1	30.0431 ρ=0.15	30.9856 ρ=0.2	31.8671 ρ=0.25
0.0	0.0515	0.0500	0.0489	0.0524	0.0510
1.1	0.0525	0.0502	0.0493	0.0530	0.0516
1.0	0.0459	0.0491	0.0484	0.0516	0.0522
1.2	0.0401	0.0453	0.0456	0.0507	0.0509
2.3	0.0366	0.0409	0.0438	0.0492	0.0501
2.0	0.0438	0.0481	00472	0.0513	0.0520
1.3	0.0442	0.0499	0.0512	0.0541	0.0539
25.0	0.0474	0.0517	0.0522	0.0561	0.0569
30.3	0.0425	0.0502	0.0514	0.0523	0.0535
3.0	0.0471	0.0520	0.0516	0.0519	0.0540

Table 5 The performance of the T^2_{MCD} control charts with modified UCL under different step shifts

	UCL				
Shift	62.0886 ρ=0	62.5123 ρ=0.1	63.1032 ρ=0.15	63.8156 ρ=0.2	64.6856 ρ=0.25
0.0	0.0488	0.0503	0.0537	0.0498	0.0521
1.1	0.0520	0.0514	0.0527	0.0490	0.0512
1.0	0.0465	0.0502	0.0499	0.0509	0.0519
1.2	0.0424	0.0400	0.0486	0.0500	0.0515
2.3	0.0444	0.0460	0.0435	0.0464	0.0498
2.0	0.0566	0.0499	0.0529	0.0563	0.0547
1.3	0.0510	0.0521	0.0543	0.0528	0.0513
25.0	0.0567	0.0553	0.0603	0.0588	0.0593
30.3	0.0600	0.0582	0.0599	0.0583	0.0543
3.0	0.0599	0.0600	0.0602	0.0597	0.0570

The results show that the methods keep their statistical properties and the strength or weakness of each method in detecting step shifts and drifts is preserved in the autocorrelated condition. In addition, the power of the methods in detecting various shifts is decreased by incensing in the correlation coefficient. To solve this deficiency, a method will be proposed afterwards.

6.2 Generalized linear mixed model based method

Generalized linear mixed models are extensions of LLMs to account for the autocorrelation in logistic regression [10]. As mentioned in Section 2, for parameter estimation in logistic regression profiles, the expected value of the response variable is computed by Eq. (2). In the GLMM method, the regression parameters are estimated as follows:

$$\pi_i = \frac{e^{X_i\beta + Z_i u}}{1 + e^{X_i\beta + Z_i u}}, \tag{18}$$

where $\boldsymbol{Z}_i$ is the vector of variables which accounts for the autocorrelation and is considered equivalent to X_i most times. u_i is the vector of random effects regression coefficients which follows a multivariate normal distribution with mean vector equals to 0 and variance–covariance matrix G. An iterative algorithm is used to estimate the vectors of β_i and u_i as follows:

Step 1 Start with an initial value for β_i and u_i which can be computed by any method.

Table 6 The performance of the T^2_I control charts with modified UCL under different drifts

	UCL				
Shift	15.0796 ρ=0	21.051 ρ=0.1	25.2600 ρ=0.15	29.2110 ρ=0.2	34.2090 ρ=0.25
0.0	0.0479	0.0500	0.0526	0.0512	0.0500
1.1	0.0641	0.0518	0.0606	0.0516	0.0502
1.0	0.0945	0.0966	0.0958	0.0910	0.0830
1.2	0.1259	0.0782	0.0692	0.0582	0.0492
2.3	0.1941	0.1040	0.078	0.0703	0.0512
2.0	0.3102	0.2522	0.2228	0.2026	0.1698
1.3	0.3347	0.1834	0.1460	0.1136	0.0914
25.0	0.5109	0.3910	0.3400	0.2806	0.2548
30.3	0.5881	0.4520	0.3888	0.3340	0.2878
3.0	0.7199	0.5338	0.4717	0.4136	0.3332

Table 7 The performance of the T^2_{R} control charts with modified UCL under different drifts

	UCL				
Shift	13.4951 ρ=0	13.5550 ρ=0.1	13.5710 ρ=0.15	13.5750 ρ=0.2	13.5800 ρ=0.25
0.0	0.0491	0.0541	0.0520	0.0524	0.0500
1.1	0.0789	0.0757	0.0777	0.0730	0.0790
1.0	0.1444	0.1127	0.0953	0.0983	0.0920
1.2	0.1967	0.1840	0.1833	0.1817	0.1890
2.3	0.2940	0.3033	0.3293	0.3333	0.3533
2.0	0.4804	0.3133	0.2597	0.2480	0.2360
1.3	0.5146	0.4637	0.4530	0.4437	0.4407
25.0	0.6951	0.4800	0.3733	0.3637	0.3290
30.3	0.7568	0.5427	0.4453	0.4093	0.3370
3.0	0.8451	0.6280	0.5237	0.4947	0.4643

Table 8 The performance of the T^2_{H} control charts with modified UCL under different drifts

	UCL				
Shift	11.8166 ρ=0	11.8220 ρ=0.1	11.8250 ρ=0.15	11.8290 ρ=0.2	11.8500 ρ=0.25
0.0	0.0474	0.0523	0.0493	0.0536	0.0538
1.1	0.0539	0.0563	0.0497	0.0577	0.0513
1.0	0.0493	0.0607	0.0560	0.0520	0.0630
1.2	0.0490	0.0523	0.0527	0.0460	0.0490
2.3	0.0509	0.0437	0.0510	0.0462	0.0410
2.0	0.0428	0.0527	0.0663	0.0630	0.0553
1.3	0.0457	0.0463	0.0440	0.0457	0.0487
25.0	0.0430	0.0460	0.0583	0.0640	0.0623
30.3	0.0421	0.0400	0.0677	0.0613	0.0757
3.0	0.0406	0.0613	0.0637	0.0630	0.0787

Table 9 The performance of the T^2_{MVE} control charts with modified UCL under different drifts

	UCL				
Shift	29.1826 ρ=0	29.4436 ρ=0.1	30.0431 ρ=0.15	30.9856 ρ=0.2	31.8671 ρ=0.25
0.0	0.0507	0.0500	0.0526	0.0512	0.0500
1.1	0.0549	0.0523	0.0586	0.0535	0.0556
1.0	0.0538	0.0519	0.0574	0.0529	0.0549
1.2	0.0524	0.0507	0.0555	0.0514	0.0540
2.3	0.0511	0.0494	0.0539	0.0502	0.0528
2.0	0.0456	0.0462	0.0518	0.0489	0.0514
1.3	0.0433	0.0458	0.0514	0.0464	0.0499
25.0	0.0469	0.0471	0.0521	0.0476	0.0505
30.3	0.0442	0.0436	0.0499	0.0471	0.0508
3.0	0.0424	0.0430	0.0483	0.0462	0.0503

Step 2 Compute W matrix from Eq. (3).

Step 3 Compute $p\times 1$ matrix H_i as follows:

$$H_i = n\,[y_i - \mu_i + \mu_i(1-\mu_i)\eta_i], \tag{19}$$

where y_i is the vector of observations for ith profile. μ_i is the vector of the expected value and $\eta_i = X_i\widehat{\beta}_i$

Step 4 Update the estimation of β_i and u_i as follows until convergence:

$$\begin{pmatrix} X_i' WX_i & X_i' WZ_i \\ Z_i' WX_i & Z_i' WZ_i + G^{-1} \end{pmatrix}\begin{pmatrix} \widehat{\beta}_i \\ \widehat{u}_i \end{pmatrix} = \begin{pmatrix} X_i' H_i \\ Z_i' H_i \end{pmatrix} \tag{20}$$

Table 10 The performance of the T^2_{MCD} control charts with modified UCL under different drifts

	UCL				
Shift	62.0886 ρ=0	62.5123 ρ=0.1	63.1032 ρ=0.15	63.8156 ρ=0.2	64.6856 ρ=0.25
0.0	0.0488	0.0503	0.0537	0.0498	0.0521
1.1	0.0495	0.0508	0.0514	0.0512	0.0519
1.0	0.0546	0.0565	0.0569	0.0555	0.0546
1.2	0.0512	0.0514	0.0531	0.0529	0.0525
2.3	0.0506	0.0503	0.0509	0.0499	0.0512
2.0	0.0450	0.0499	0.0487	0.0503	0.0531
1.3	0.0473	0.0506	0.0498	0.0511	0.0529
25.0	0.0483	0.0520	0.0533	0.0548	0.0553
30.3	0.0460	0.0486	0.0517	0.0527	0.0541
3.0	0.0452	0.0479	0.0510	0.0519	0.0533

Table 11 GLMM method results for step shift T^2_{I}

	UCL			
Shift	5.3800 ρ=0.1	8.0400 ρ=0.15	11.0300 ρ=0.2	14.1650 ρ=0.25
0.0	0.0500	0.0500	0.0500	0.0500
1.1	0.0514	0.0592	0.0520	0.0580
1.0	0.4170	0.3370	0.2574	0.2186
1.2	0.3080	0.2012	0.1476	0.1304
2.3	0.5074	0.2586	0.1440	0.1074
2.0	0.9912	0.9554	0.8684	0.7746
1.3	0.9836	0.8880	0.7208	0.5728
25.0	0.9998	0.9984	0.9798	0.9346
30.3	1.0000	0.9996	0.9910	0.9746
3.0	1.0000	0.9998	0.9990	0.9934

7 Simulation studies

In this section, the performance of the UCL modification method and GLMM method is compared under different correlation coefficients with symmetric correlation structure. As mentioned before, Yeh et al. [27] example data were used for simulations and comparisons of this paper. Matlab 7.8.0 (R2009a) and R 2.12.1 are applied for computational stages. The results of GLMM method for step shift and drift are depicted in Tables 11 and 12, respectively. A comparison between the results of two proposed methods under various step shifts and drifts are shown in Figs. 3 and 4. A comparison between these two methods shows that the GLMM method performs uniformly better than the UCL modification method under both step shifts and drifts. Similar to the results of Section 6 (not reported here), the best T^2 method under autocorrelated situation is T^2_{I}. So the results are compared on this method.

Table 12 GLMM method results for drifts T^2_I

	UCL			
Shift	5.3800 ρ=0.1	8.0400 ρ=0.15	11.0300 ρ=0.2	14.1650 ρ=0.25
0.0	0.0500	0.0500	0.0500	0.0500
1.1	0.0438	0.0510	0.0446	0.0524
1.0	0.1452	0.1366	0.1004	0.1006
1.2	0.0904	0.0800	0.0686	0.0688
2.3	0.1028	0.0812	0.0554	0.0574
2.0	0.5952	0.4734	0.3646	0.2950
1.3	0.4690	0.3262	0.2398	0.1910
25.0	0.8436	0.7190	0.5724	0.4682
30.3	0.9046	0.7904	0.6540	0.5474
3.0	0.9590	0.8812	0.7696	0.6582

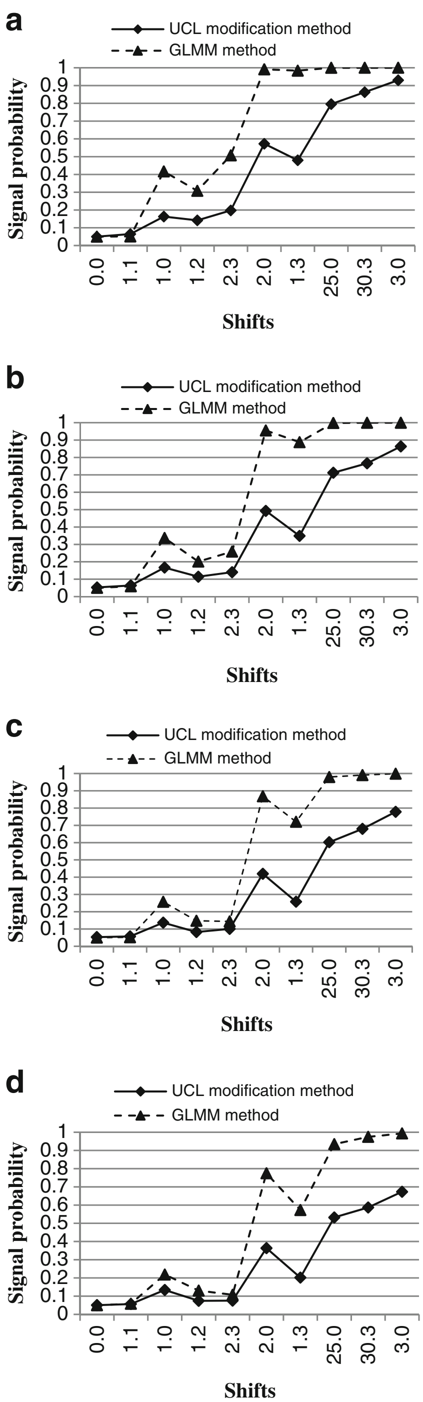


Fig. 3 Comparison between UCL modification method and GLMM method under step shifts in **a** ρ=0.1, **b** ρ=0.15, **c** ρ=0.2, **d** ρ=0.25

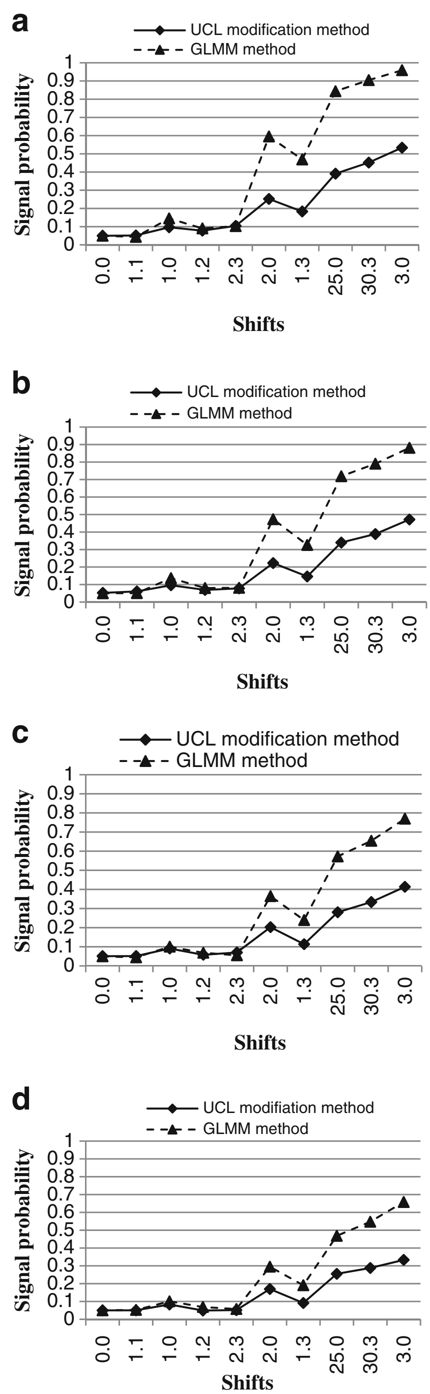


Fig. 4 Comparison between UCL modification method and GLMM method under drift in **a** ρ=0.1, **b** ρ=0.15, **c** ρ=0.2, **d** ρ=0.25

8 Conclusions

In real-world problems, there are many cases in which the response variable of profiles is binary and there is autocorrelation between observations in different level of each profile. In this paper, we showed that neglecting the autocorrelation between observation leads to misleading results on the performance of monitoring procedures. Meanwhile, the performance of five T^2 control charts by Yeh et al. [27] is evaluated under autocorrelated condition for both step shifts and drifts. The results showed that T_I^2 is the best method similar to the results reported by Yeh et al. [27]. In addition, we proposed two methods based on classical logistic regression (the UCL modification method) and GLMM in order to account for the effect of autocorrelation on the performance of T^2 control chart in the monitoring of autocorrelated logistic profiles. Simulation results illustrated that GLMM method is uniformly better than the UCL modification method under both step shifts and drifts. Developing this research for profiles with other distributions such as Poisson and Gamma can be a suggestion for future research.

References

1. Agresti A (2007) An introduction to categorical date analysis. Wiley, New York
2. Amiri A, Jensen WA, Kazemzadeh RB (2010) A case study on monitoring polynomial profiles in the automotive industry. Qual Reliab Eng Int 26:509–520
3. Boeing Commercial Airplane Group (1998) Materiel Division, Procurement Quality Assurance Department, Advanced Quality System Tools (AQS) D1-9000-1, the Boeing Company: Seattle, WA
4. Chen H (2001) Initialization for NORTA: generation of random vectors with specified marginal and correlations. Inf J Comput 13:312–331
5. Chen S, Nembhard HB (2010) A high dimensional control chart for profile monitoring. Qual Reliab Eng Int 27(4):451–464
6. Eyvazian M, Noorossana R, Saghaei A, Amiri A (2011) Phase II monitoring of multivariate multiple linear regression profiles. Qual Reliab Eng Int 27:281–296
7. Jensen WA, Birch JB (2009) Profile monitoring via nonlinear mixed models. J Qual Technol 41:18–34
8. Jensen WA, Birch JB, Woodall WH (2008) Monitoring correlation within linear profiles using mixed models. J Qual Technol 40:168–183
9. Jin J, Shi J (1999) Feature-preserving data compression of stamping tonnage information using wavelets. Technometrics 41:327–339
10. Wolfinger R, O'connell M (1993) Generalized Linear Mixed Models a Pseudo-Likelihood Approach. Journal of statistical Computation and Simulations 48:233-243
11. Kang L, Albin SL (2000) On-line monitoring when the process yields a linear profile. J Qual Technol 32:418–426
12. Kazemzadeh RB, Noorossana R, Amiri A (2008) Phase I monitoring of polynomial profiles. Commun Stat Theory Methods 37:1671–1686
13. Kazemzadeh RB, Noorossana R, Amiri A (2009) Monitoring polynomial profiles in quality control applications. Int J Adv Manuf Technol 42:703–712
14. Kim K, Mahmoud MA, Woodall WH (2003) On the monitoring of linear profiles. J Qual Technol 35:317–328
15. Mahmoud MA (2008) Phase I analysis of multiple linear regression profiles. Commun Stat Simul Comput 37:2106–2130
16. Mahmoud MA, Parker PA, Woodall WH, Hawkins DM (2007) A change point method for linear profile data. Qual Reliab Eng Int 23:247–268
17. Mahmoud MA, Woodall WH (2004) Phase I analysis of linear profiles with calibration applications. Technometrics 46:380–391
18. Mestek O, Pavlik J, Suchánek M (1994) Multivariate control charts: control charts for calibration curves. Fresenius' J Anal Chem 350:344–351
19. Montgomery DC (2005) Introduction to statistical quality control. Wiley, New York
20. Noorossana R, Eyvazian M, Amiri A, Mahmoud MA (2010) Statistical monitoring of multivariate multiple linear regression profiles in phase I with calibration application. Qual Reliab Eng Int 26:291–303
21. Saghaei A, Mehrjoo M, Amiri A (2009) A CUSUM-based method for monitoring simple linear profiles. Int J Adv Manuf Technol 45:1252–1260
22. Soleimani P, Noorossana R, Amiri A (2009) Simple linear profiles monitoring in the presence of within profile autocorrelation. Comput Ind Eng 57:1015–1021
23. Stover FS, Brill RV (1998) Statistical quality control applied to ion chromatography calibrations. J Chromatogr A 804:37–43
24. Vaghefi A, Tajbakhsh SD, Noorossana R (2009) Phase II monitoring of nonlinear profiles. Commun Stat Theory Methods 38:1834–1851
25. Vargas JA (2003) Robust estimation in multivariate control charts for individual observation. J Qual Technol 35:367–376
26. Williams JD, Woodall WH, Birch JB (2007) Statistical monitoring of nonlinear product and process quality profiles. Qual Reliab Eng Int 23:925–941
27. Yeh AB, Huwang L, Li YM (2009) Profile monitoring for a binary response. IIE Trans 41:931–941
28. Zhang J, Li ZH, Wang ZH (2009) Control chart based on likelihood ratio for monitoring linear profiles. Comput Stat Data Anal 53:1440–1448
29. Zhu J, Lin DK (2010) Monitoring the slopes of linear profiles. Qual Eng 33:1–12
30. Zou C, Tsung F, Wang Z (2007) Monitoring general linear profiles using multivariate exponentially weighted moving average schemes. Technometrics 49:395–408