The Effect of Neglecting Autocorrelation on the Performance of T^2 Control Charts in Monitoring of Logistic Profiles

Amirhossein Amiri[1], Mehdi Koosha[2]
[1, 2] Shahed university, Department of industrial engineering
[1]Amiri@shahed.ac.ir [2] Koosha@shahed.ac.ir

Abstract- **There are many researches in the area of profile monitoring and most of them have two basic assumptions including the normality of response variable as well as independency of observations within each profile. In recent years there are some researches investigating the violation of the assumptions. This paper shows the effect of ignoring autocorrelation on the performance of T^2 control chart when the response variable follows a binomial distribution. Besides, a simple method is proposed to account for this effect in the T^2 control chart. At the end, the performance of the proposed method is evaluated using simulation studies.**

Key words: **Profile monitoring, Logistic regression, Autocorrelation, T^2 control chart.**

I. INTRODUCTION

In some statistical process control applications, quality of a process or product is characterized by a relationship between a response variable and one or more explanatory variables known as profile. There are many researches in this area but in most of them, there is a basic assumption that the observations are independent from each other. For example see [1], [2], [3], and [4]. A number of researches have investigated the violation of this assumption. Reference [5] used Linear Mixed Models (LMM) to eliminate the effect of autocorrelation in a simple linear profile. Reference [6] proposed a method based on transforming of the observations. A method based on Non-Linear Mixed Models (NLMM) in order to remove the effect of autocorrelation in a nonlinear profile is proposed by [7]. In all of the aforementioned papers, the distribution of the observations in a profile is assumed to be normal.

In the area of profile monitoring, sometimes, the response variable follows a binomial distribution. In this regard, [8] have presented some T^2 based control charts to monitor logistic profiles in which the observations are assumed to have a Bernoulli distribution. Reference [8] has assumed that the observations in different levels of explanatory variables are independent from each other. However, in some cases the independency of observations in different levels of explanatory variable is violated. This paper specifically concentrates on Phase I monitoring of autocorrelated logistic regression profiles and analyses the effect of neglecting the autocorrelation on the performance of T^2 control chart. Moreover, a simple method to consider this effect is proposed. The rest of this paper is structured as follows: Section II introduces the method for estimating parameters of a logistic profile which applied by [8]. T^2 Hoteling control chart will be defined in Section III. Section IV analyses the impact of presence of autocorrelation on the performance of T^2 Hoteling by simulation and section V proposes a method to account for this effect. Our concluding remarks are given in the final section.

II. PARAMETER ESTIMATION OF A LOGISTIC PROFILE

When the observations are assumed to be normal we can use Least Square Error (LSE) method to estimate the parameters of a profile. But in the case that the response variable follows a Bernoulli distribution, this method leads to unbiased estimators which do not have minimum variance. So the maximum likelihood estimation (MLE) method can be used in this case. To obtain the parameters estimators in logistic regression profiles by using MLE method, an iterative algorithm should be used [9].

Suppose that there are m independent profiles (i=1,2,…,m) and the observations in each level of the explanatory variable (j=1,2,…,p) are repeated for n times. For estimating of parameters in a logistic profile, [8] have applied an approach that is explained below:

Start the algorithm with an initial vector of $\hat{\beta}$ which could be obtained by any methods such as LSE.

Step 1: Set l=0. Determine the value of X (Vector of explanatory variables) and observe the response variables in each level.

Step 2: Compute a vector of $\pi_i^{(l)}$ that contains elements $\pi_{ij}^{(l)}$ which shows the probability of success in an experiment in j^{th} level of X for i^{th} profile in l^{th} iteration.

Step 3: Set l=l+1. Calculate $\eta_i^{(l)} = X_i^T \hat{\beta}^{(l)}$ for any iteration of the algorithm.

Step 4: Update the values of $\pi_i^{(l)}$ with Eq. (1)

$$\pi_i^{(l)} = \frac{e^{\eta_i^{(l)}}}{1+e^{\eta_i^{(i)}}} \tag{1}$$

Logit link function is used in Eq. (1) to relate the values of $\eta_i^{(l)}$ and the expected value of each observation ($\pi_i^{(l)}$)

$$\ln\left(\frac{\pi_i^{(l)}}{1-\pi_i^{(l)}}\right) = \eta_i^{(l)} \Rightarrow \pi_i^{(l)} = \frac{e^{\eta_i^{(l)}}}{1+e^{\eta_i^{(i)}}} \tag{2}$$

$\pi_i^{(l)}$ should be between 0 and 1. Hence, this link function could be applied to map any given X to a value between 0 and 1.

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

μ is the vector of expected values for response variables.

$$W_i^{(l)} = diag[n\pi_{ij}^{(l)}(1-\pi_{ij}^{(l)})] \tag{3}$$

$$\mu_i^{(l)} = [n\pi_{ij}^{(l)}] \tag{4}$$

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

$$q_i^{(l)} = \eta_i^{(l)} + (W_i^{(l)})^{-1}(Y_i - \mu^{(l)}) \tag{5}$$

Y_i is the vector of observations in i^{th} profile.

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

$$\hat{\beta}^{(l+1)} = (X^T W_i^{(l)} X)^{-1} X^T W_i^{(l)} q_i^{(l)} \tag{6}$$

If $\hat{\beta}^{(l+1)} - \hat{\beta}^{(l)} < \varepsilon$, then $\hat{\beta}^{(l+1)}$ can be used as the final estimation of β.

Else set $l=l+1$ and go to step 3

III. T² HOTELING CONTROL CHART

In the parametric approach of profile monitoring one can monitor the regression parameters. Since the regression parameters estimators in logistic regression are dependent, multivariate control charts should be used for this purpose. There are some types of multivariate control charts such as Multivariate Cumulative Sum (MCUSUM), Multivariate Exponentially Weighted Moving Average (MEWMA) and T^2 Hoteling control charts in the literature of SPC [10]. Different T^2-based control charts are proposed by [8] (T_i^2, T_r^2, T_h^2, T^2_{MCD}, T^2_{MVE}) to monitor logistic regression profiles in Phase I. This reference shows that the performance of T_i^2 control chart is better than the other methods in detecting either small or large step shifts in the parameters. So this control chart is used in our paper for comparisons and validating the results. The variables used in the T^2 control chart should follow a multivariate normal distribution [10]. It should be noted that the estimators of the regression parameters follow a p-variate normal distribution with mean β and variance-covariance matrix $(X^T W X)^{-1}$ [8]. The T_i^2 statistic is given in Eq. (7).

$$T_i^2 = (\hat{\beta}_i - \bar{\beta})^T S_i^{-1} (\hat{\beta}_i - \bar{\beta}), \tag{7}$$

where $\hat{\beta}_i$ is the estimated vector of regression parameters for i^{th} profile. $\bar{\beta}$ and S_i are the mean vector and covariance matrix of the parameters estimators and computed by Eq.(8) and Eq.(9), respectively.

$$\bar{\beta} = \frac{\sum_{i=1}^{m} \beta_i}{m} \tag{8}$$

$$S_i = \frac{1}{m}\sum_{i=1}^{m} \mathrm{var}(\hat{\beta}_i) = \frac{1}{m}\sum_{i=1}^{m} (X^T W_i X)^{-1} \tag{9}$$

The UCL for this control chart is obtained by simulation to achieve a specific probability of Type I error.

IV. THE EFFECT OF AUTOCORRELATION

In the methods proposed by [8] including T_i^2 control chart, the independency of the observations with each other in different levels of explanatory variables in a profile is an assumption. However, sometimes this assumption is violated. In this section we evaluate the effect of autocorrelation on the performance of T_i^2 when it is neglected.

For simulating a profile with autocorrelated binomial response variables, we should first generate random variables based on multivariate binomial distribution. Let m (the number of profiles generated in each iteration) and n (the number of observations in each level) be 30. There are many standard structures for correlation matrix. Here we use Exchangeable structure in which the correlation matrix is as follows:

$$\rho = \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \rho & \dots\dots & 1 & \rho \\ \rho & \dots\dots & \rho & 1 \end{bmatrix} \tag{10}$$

For generating multivariate binomial variables with a predetermined correlation matrix we have used NORTA algorithm [11].

We consider 4 values for ρ (the correlation between each two observations in a specific level of a profile). Then we generate autocorrelated binomial distribution by using NORTA algorithm. We assume that there are m=30 profiles and the replication in each level of a profile is equal to n=30. In addition, 9 levels and the same x-values proposed by reference [8] are used as log 0.1, log 0.2,…, and log 0.9. The shift is also a step shift in the second half of the samples.

After generating the dataset, the T_i^2 control chart proposed by [8] is used for Phase I monitoring of autocorrelated logistic regression profile. The results are computed in the terms of probability of signal and summarized in Table I in comparison with the results in [8] when there is no autocorrelation within each profile. The simulation results in Table I showed that the presence of autocorrelation increases the probability of type I error. This affects the performance of the T^2 control chart when there is shift in the regression parameters. Due to increasing in the probability of Type I error, the probability of Type II error decreases. As a result, the power in detecting shifts increases. However, the increase in the power does not imply that the performance of the control chart has improved because this improvement is due to increasing in probability of Type I error.

Note that shifts in Table I are in the scale of σ. For example (1, 2) means that $\beta_{0,out} = \beta_{0,in} + 1\sigma_0$ and $\beta_{1,out} = \beta_{1,in} + 2\sigma_1$ in which σ_0 and σ_1 are computed by Eq. (11)

$$S_0 = \begin{pmatrix} \sigma_0^2 & \rho\sigma_0\sigma_1 \\ \rho\sigma_0\sigma_1 & \sigma_1^2 \end{pmatrix} = (X^T W X)^{-1} \qquad (11)$$

TABLE I
The effect of autocorrelation on the performance of T_i^2 control chart in various shifts in profile parameters (β_0 and β_1)

shift \ ρ	0	0.1	0.15	0.2	0.25
0,0	0.0534	0.2532	0.3134	0.4480	0.5924
1,1	0.1064	0.2964	0.3534	0.4814	0.6178
1,0	0.1802	0.5620	0.6218	0.7176	0.8062
1,2	0.2588	0.5214	0.5762	0.6692	0.7644
2,3	0.4320	0.6188	0.6614	0.7316	0.8138
2,0	0.6882	0.9380	0.9550	0.9758	0.9874
1,3	0.7196	0.9060	0.9220	0.9532	0.9662
2.5,0	0.9236	0.9912	0.9910	0.9950	0.9984
3,0.3	0.9652	0.9978	0.9968	0.9986	0.9992
3,0	0.9926	0.9992	0.9998	0.9998	1

The results of Table I are also shown in Figure 1. It is clear that as the correlation coefficient increases, the probability of out-of-control signal is also increases due to increasing the probability of type I error.

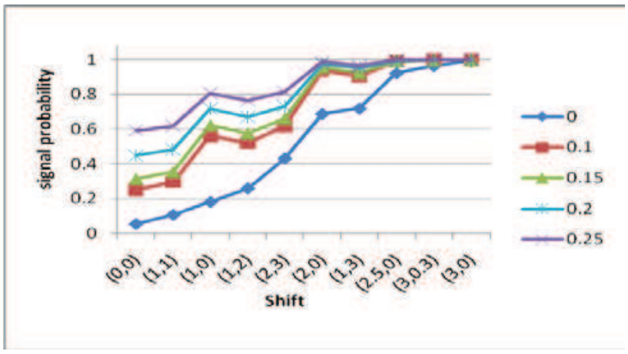

Fig 1 - The effect of autocorrelation on the performance of T_i^2 control chart

V. PROPOSED METHOD

One of the simplest methods for considering the effect of autocorrelation is increasing the value of UCL by simulation in order to make the probability of type I error equal to a predetermined level. Here, for any values of correlation coefficient (ρ) UCL has been determined by simulation to obtain 0.05 for probability of Type I error. The corresponding UCL for each correlation coefficient and probability of out-of-control signals under various step shifts are shown in Table II and Fig. 2.

Table II
Comparison between T^2_i Control charts in the presence of various correlation coefficient and the modified *UCL*

ρ	0	0.1	0.15	0.2	0.25
Shift \ UCL	15.0796	21	25	29.2	34.2
0,0	0.0534	0.0512	0.0526	0.053	0.0506
1,1	0.1064	0.065	0.064	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
2.5,0	0.9236	0.7956	0.7122	0.6024	0.5320
3,0.3	0.9652	0.8626	0.7664	0.6796	0.5868
3,0	0.9926	0.93	0.8636	0.7782	0.6730

As shown in Fig 2, by performing this modification to values for UCL, the probability of Type I error will be equal to 0.05 for all of the values of correlation coefficient. Using this method, the real values for probability of out-of-control signal is also given. In addition, the results in Table II shows that the increasing the correlation coefficient deteriorates the performance of T_i^2 control chart.

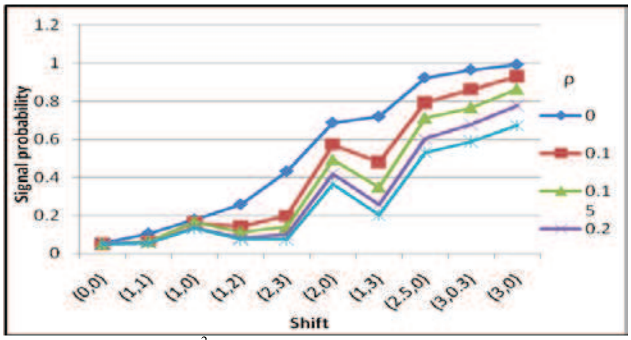

Fig 2 - The power of Ti^2 control chart with modified UCL under different correlation coefficients

VI. CONCLUSIONS

In this paper, we investigated the effect of neglecting autocorrelation with exchangeable structure on the performance of the T_i^2 control chart proposed by [8]. The results showed that as the correlation coefficient increases the probability of Type I error increases. Moreover, the power of the control chart increases due to increasing in probability of Type I error. Finally, a simple method based on the modification of the UCL proposed to account for this effect. This method was performed under different correlation coefficients. The results showed that the problem with probability of Type I error can be solved. However, power of the control chart deteriorates especially when the correlation coefficient is large. One can use statistical methods such as generalized linear mixed model (GLMM) as a future research to improve the power and accounts for the autocorrelation in logistic regression profile as a future research. Different structures of autocorrelation such as AR(1) can also be investigated by researchers.

REFRENCES

[1] Kang, L., Albin, S. L., On-Line monitoring when the process yields a linear profile, Journal of Quality Technology, 2000, 32, 418-426.

[2] Kim, K., Mahmoud, M.A., Woodall, W.H., On the monitoring of linear profiles, Journal of Quality Technology, 2003, 35, 317-328.

[3] Mahmoud, M.A., Parker, P.A., Woodall, W.H., Hawkins, D.M., A change point method for linear profile data, Quality and Reliability Engineering International, 2007, 23, 247-268.

[4] Kazemzadeh, R. B., Noorossana, R., Amiri, A., Phase I monitoring of polynomial profiles, Communications in Statistics, Theory and Methods, 2008, 37, 1671-1686.

[5] Jensen W.A., Birch J. B., WOODALL W.H., Monitoring correlation within linear profiles using mixed models, ASQ, 2008, 40,168-183.

[6] Soleimani, P., Noorossana R., Amiri A., Simple linear profiles monitoring in the presence of within profile autocorrelation, computers and industrial engineering, 2009, 57, 1015-1021.

[7] Jensen, W.A., Birch, J.B., Profile monitoring via nonlinear mixed models, Journal of Quality technology, 2009, 41, 18-34.

[8] Yeh, A. B., Huwang, L., Li, Y. M., Profile monitoring for a binary response, IIE Transactions, 2009, 41, 931-939.

[9] Kleinbaum, D., Klein, M., Logistic regression: a self-learning Text, Springer, 2002.

[10] Montgomery, D.C., Introduction to statistical quality control, Wiley, 2005.

[11] Chen, H., Initialization for NORTA: Generation of Random Vectors with Specified Marginals and Correlations, INFORMS Journal on Computing, 2001, 13, 312-331.