Original Research

Design of Control Charts for Monitoring Logistic Regression Profiles with Estimated Parameters

Zahra Musavipour, Amirhossein Amiri*, Zahra Jalilibal
Received: 2021-05-01/ Accepted: 2021-11-24/ Published online: 2021-12-03

1- Department of Industrial Engineering, Shahed University, Tehran, Iran
*Corresponding author's email: amirhossein.amiri@gmail.com

Abstract

In many cases, the quality of a process can be described using a regression profile relationship between a response variable and one or more independent variables. So far, much research has been done on the response variables with continuous and normal distribution. While, in real situations, when a product is conforming or nonconforming on the product line, the assumption of normality is violated and a logistic regression model is used to characterize binary response variables. Also, in many cases the parameters used to design control charts for monitoring profiles are unknown and estimated by using IC reference data, which adversely influences the efficiency of control charts in Phase II. In recent years, a few authors have been done on the effect of parameter estimation in monitoring profiles, especially profiles whose response variables do not follow the normal distribution. In this paper, Hotelling's T2 chart and a multivariate exponentially weighted moving average (MEWMA) chart are used to monitor the logistic regression profile in Phase II with estimated parameters. In addition, two criteria including average of average run length (AARL) and standard deviation of average run length (SDARL) are utilized to appraise the effect of parameters estimation in Phase I on the Phase II performance of designed control charts through simulation runs. The results illustrate that the performance of these charts is significantly affected by the estimated parameters in both IC and OC conditions. Also, two methods are utilized to decrease the effect of parameters estimation which include increasing the number of reference profiles in Phase I and modifying the control limits. The results of these methods show that by increasing the number of reference profiles in Phase I, the effect of parameters estimation decreases.

Keywords: Average run length, Control chart, Logistic regression profiles, Parameters estimation, Profile monitoring.

1. INTRODUCTION

The quality of a characteristic in a process can be characterized using a regression profile relationship among a response variable and one or more independent variables. The aim of profile monitoring is to check the stability of a regression profile during the time. The classification of control charts in the literature for monitoring various kinds of profiles is divided into two classes, as Phase I and Phase II methods. Each class has its own purpose where Phase I analysis purpose is to estimate the unknown parameters of a profile by utilizing the historical data from the process and the main purpose of Phase II is detecting the OC parameters of a profile quickly. Refer to the review papers by Woodall (2007) and Maleki et al. (2018) as well as the book by Noorossana et al. (2011) for more information about profile monitoring methods in Phases I and II. Although the values of in-control parameters are presumed to be known in Phase II, this is not true in many cases in real-life manufacturing or non-manufacturing environments. Hence, a historical IC data set is required in Phase I to estimate unknown parameters. For this purpose, usually a large sample size is required to certify that parameters are well estimated. Note that the small sample size leads to inefficient parameters estimation and finally, the efficiency of control charts deteriorates.

Many papers such as Chakraborti and Human (2006), Chakraborti and Human (2008), Maravelakis and Castagliola (2009), Capizzi and Masarotto (2010), Zhang et al. (2011), Castagliola and Wu (2012), Zhang et al. (2013), and Rakitzis and Castagliola (2016) investigated the effect of estimated parameter on the efficiency of control schemes. Refer to the articles by Jensen et al. (2006) and Psarakis et al. (2014) for detail information about parameters estimation's effect on the efficiency of control schemes under different conditions.

Despite the large number of authors on the subject of the effect of parameters estimation in different areas of statistical process monitoring, the number of papers about the effect of parameters estimation on the efficiency of profile monitoring methods is limited. The first related paper was presented by Mahmoud (2012) on the efficiency of three linear profile monitoring methods under estimated parameters of regression in terms of ARL and *SDRL*. Aly et al. (2015) considered *SDARL* criterion to compare the IC performance of three Phase II linear profile monitoring methods. Noorossana et al. (2016) investigated the effect of estimated regression parameters in Phase I on the performance of EWMA-3 chart for monitoring simple linear profiles. A study by Chen et al. (2016) used two types of Phase I estimators to assess the effect of Phase I estimation on Phase II monitoring of processes with profile data. Ahmadi Yazdi et al. (2019a) investigated the effect of parameter estimation on monitoring the multivariate simple linear profiles in Phase II. Ahmadi Yazdi et al. (2019b) studied the effect of parameter estimation on the efficiency of monitoring multivariate profiles in Phase II. Ahmadi Yazdi et al. (2020) investigated the effect of estimated parameters on the monitoring of multivariate linear regression profiles in Phase II based on a novel clustering method.

All the papers reviewed above, concentrate on the effect of Phase I estimated parameters on the performance of Phase II monitoring procedures for profile monitoring with normal response variable. Nevertheless, manufacturing and non-manufacturing situations in real-life may cause a violation in the assumption that the response variable follows normal distribution. When the response variable follows the exponential distributions' family like the binary, Poisson and Gamma distributions, the generalized linear model (GLM) is applied for modeling profiles. Some researchers have been studied on monitoring GLM-based profiles such as Yeh et al. (2009), Shang et al. (2011), Amiri et al. (2011, 2012), Paynabar and Yeh (2012), Saghaei et al. (2012), Koosha and Amiri (2013), Soleymanian et al. (2013), Noorossana et al. (2013), Shadman et al. (2015), Amiri et al. (2015), Panza and Vargas (2016), Qi et al. (2016), Huwang et al. (2016), Amiri et al. (2016), Shadman et al. (2017), Maleki et al. (2017), Amiri et al. (2018), Bandara et al. (2020) Kinat et al. (2020), Moheghi et al. (2021), Piri et al. (2021), Mammadova and Özkale (2021). Maleki et al. (2019) examined the effect of parameters estimation Phase II monitoring of Poisson regression profiles in Phase II.

In this paper, Hotelling's T^2 chart and a MEWMA chart are used to monitor the logistic regression profile in Phase II with estimated parameters. In addition, two criteria including *AARL* and *SDARL* are used to appraise the effect of parameters estimation in Phase I on the Phase II efficiency of designed control schemes through simulation runs. The main contribution of this paper is the monitoring scheme applied on logistic regression profile and the performance measure which is used in the simulation to appraise the efficiency of presented control charts. The results illustrate that the performance of these charts is significantly affected by the estimated parameters in both IC and OC conditions. Also, two methods are used to decrease the effect of parameters estimation which include increasing the number of reference profiles in Phase I and modifying the control limits. The results of these methods show that by increasing the number of reference profiles in Phase I, the effect of parameters estimation decreases. However, since the method of increasing the number of reference profiles in real applications is not economic and has limitations and difficulties, the minimum reference sample size is calculated to achieve the appropriate run length properties for both proposed control charts. Also, by increasing the UCL of the designed control charts under various values of the reference profiles, the desired IC-ARL is obtained. On the other hand, in the case of OC condition, enhancing the number of the reference profiles in Phase I leads to quick discover of shifts in Phase II and improved efficiency of both designed control charts.

The structure of the paper is as: Section 2 discusses a concise description of the logistic regression model along with two control charts, Hotelling's T^2 and MEWMA control chart, for monitoring the Logistic regression profiles in Phase II. Section 3 describes the proposed scheme which evaluates the effect of parameters estimation on Phase II monitoring of the logistic regression profile. In Section 4, simulation studies and remedial approaches are conducted to evaluate the effect of parameters estimation on the efficiency of the proposed control schemes in this paper, namely, the Hotelling's T^2 and MEWMA control charts. Moreover, introduces two remedial approaches which rectify the detrimental effect of parameters estimation on the IC efficiency of the both proposed control charts. Section 5, represents the detection power of the abovementioned control charts with the estimated parameters in Phase I, considering the proposed remedial measures. Eventually, some potentials for further research are recommended in Section 6.

2. Logistic regression profiles monitoring in Phase II

Some researchers like Zand et al. (2013), Amiri et al. (2015) and Izadbakhsh et al. (2018) investigated monitoring the logistic regression profiles in Phase I. However, we focus on monitoring the logistic regression profiles in Phase II. In 2.1, the description of the logistic regression model is given. The Hotelling's T^2 and the MEWMA control charts for Phase II monitoring of logistic regression profiles are clarified in 2.2 and 2.3 subsections, respectively.

2.1. Logistic regression model

The generalized linear model (GLM) is widely used in profile monitoring. But in binary mode (0 and 1), the logistic regression is the most common model used in profile monitoring. In this model, $\mathbf{X}_i$ is the vector of *ps* independent variables related to i^{th} observation $(i=1,2,...,n)$ and Y_i is the corresponding response variable.
The link function that illustrates the relationship among independent variables and the response variable in the logistic regression model, is logit function. Using this function, the logistic regression model is written as:

$$g(\pi_i)=\log(\frac{\pi_i}{1-\pi_i})=\mathbf{X}_i^{\mathbf{T}}\boldsymbol{\beta}=\beta_1X_{i1}+\beta_2X_{i2}+...+\beta_pX_{ip}, \quad (1)$$

where the vector of parameters of model is represented by $\boldsymbol{\beta}=(\beta_1,...,\beta_p)^T$. The intercept of the model is denoted by β_1, which means $X_{i1}=1$. According to this model, the probability of success will be as follows:

$$\pi_i=\frac{\exp(\mathbf{X}_i^{\mathbf{T}}\boldsymbol{\beta})}{1+\exp(\mathbf{X}_i^{\mathbf{T}}\boldsymbol{\beta})}. \quad (2)$$

It is supposed that m_i is the number of observations of the response variable at i^{th} level of independent variables and $M=\sum_{i=1}^{n}m_i$ represents the total number of observations.

Based on this model, $y_i=\sum_{j=1}^{m_j}z_{ij}$ represents the total number of successes in i^{th} level of independent variables and therefore follows a bionomial distribution with (m_i,π_i) parameters, where z_{ij} represents the j^{th} observation at i^{th} level of independent variables. So, $E(y_i)=m_i\pi_i$ and the variance of y_i is as follows:

$$Var(y_i)=m_i\pi_i(1-\pi_i)=m_i\times\frac{\exp(\mathbf{X_i}^{\mathbf{T}}\boldsymbol{\beta})}{1+\exp(\mathbf{X_i}^{\mathbf{T}}\boldsymbol{\beta})}\times\frac{1}{1+\exp(\mathbf{X_i}^{\mathbf{T}}\boldsymbol{\beta})}. \quad (3)$$

The probability function of $y_1,y_2,...,y_n$ is as:

$$L(\pi,\mathbf{y})=\prod_{i=1}^{n}\binom{m_i}{y_i}\pi_i^{y_i}(1-\pi_i)^{m_i-y_i}, \quad (4)$$

$$\log L(\pi,\mathbf{y}) = \sum_{i=1}^{n}\log\binom{m_i}{y_i} + \sum_{i=1}^{n} y_i(\mathbf{X}_i^{\mathbf{T}}\boldsymbol{\beta}) - \sum_{i=1}^{n} m_i \log(1+\exp(\mathbf{X}_i^{\mathbf{T}}\boldsymbol{\beta})), \quad (5)$$

where $\boldsymbol{\mu} = (\mu_1,\mu_2,...,\mu_n)^T = E(\mathbf{y}) = (m_1\pi_1, m_2\pi_2,...,m_n\pi_n)^T$, $\pi = (\pi_1,\pi_2,...,\pi_n)^T$, $\mathbf{y} = (y_1,y_2,...,y_n)^T$ and $\mathbf{X} = (\mathbf{X}_1,\mathbf{X}_2,...,\mathbf{X}_n)^T$ is a $p\times n$ matrix. $\boldsymbol{\beta}$ is obtained by resolving the $\mathbf{X}^{\mathbf{T}}(\mathbf{y}-\boldsymbol{\mu}) = \mathbf{0}_p$ .The Iterative Weighted Least Square (IWLS) estimation approach is used to estimate MLE which is denoted by $\hat{\boldsymbol{\beta}}$. Based on this method, $\hat{\boldsymbol{\beta}}$ follows a p-variate normal distribution, as $N_p(\boldsymbol{\beta}_\mathbf{0},\boldsymbol{\Sigma}_\mathbf{0})$ where $\boldsymbol{\Sigma}_\mathbf{0} = (\mathbf{X}^\mathbf{T}\mathbf{W}\mathbf{X})^{\mathbf{-1}}$ and **W** is defined as:

$$\mathbf{W} = diag\{m_1\pi_1(1-\pi_1), m_2\pi_2(1-\pi_2),.....,m_n\pi_n(1-\pi_n)\}. \quad (6)$$

2.2. Hotelling's T² Control Chart

Some authors used Hotelling's T^2 control chart for monitoring different kinds of GLM profiles such as Amiri et al. (2012), and Maleki et al. (2019). The statistic of Hotelling's T^2 control chart to monitor the logistic regression profile in Phase II is as:

$$T_j^2 = (\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_0)^T \boldsymbol{\Sigma}_\mathbf{0}^{-1}(\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_0), \quad (7)$$

where $\boldsymbol{\beta}_\mathbf{0}$ and $\boldsymbol{\Sigma}_\mathbf{0}$ are the IC mean vector and variance-covariance matrix of the estimated regression parameters and $\hat{\boldsymbol{\beta}}_j = (\hat{\boldsymbol{\beta}}_{j,1},...,\hat{\boldsymbol{\beta}}_{j,p})^\mathbf{T}$ is the vector of the estimated regression parameters for j^{th} profile. The procedure of parameters estimation is done by IWLS approach (Refer to Sharafi et al. (2013) for detailed information). Also, the IC variance-covariance matrix of the regression parameters, which is proposed by Yeh et al. (2009), is:

$$\boldsymbol{\Sigma}_\mathbf{0} = (\mathbf{X}^\mathbf{T}\mathbf{W}\mathbf{X})^{\mathbf{-1}}, \quad (8)$$

where $\mathbf{W} = diag(m_1\pi_1(1-\pi_1),.....,m_n\pi_n(1-\pi_n))$

If $T_j^2 > h_T$, $j = 1,2,...$ an OC alarm is triggered, while h_T is designated in a way that ARL_0 value equal to 200 is attained.

If $\boldsymbol{\beta}_\mathbf{0}$ and $\boldsymbol{\Sigma}_\mathbf{0}$ are unknown, They should be estimated by using historical data in Phase I. Usually, $\boldsymbol{\beta}_\mathbf{0}$ is estimated by $\bar{\hat{\boldsymbol{\beta}}}$ which is an unbiased estimator for $\boldsymbol{\beta}_\mathbf{0}$. Moreover, $\boldsymbol{\Sigma}_\mathbf{0}$ is estimated by $\hat{\boldsymbol{\Sigma}}_\mathbf{0}$ replacing $\hat{\mathbf{W}}$ in Equation (8) where $\hat{\mathbf{W}} = diag(m_1\hat{\pi}_1(1-\hat{\pi}_1),.....,m_n\hat{\pi}_n(1-\hat{\pi}_n))$.

2.3. MEWMA Chart

Zou et al. (2007) presented a MEWMA chart for Phase II monitoring of general linear profiles. Soleymanian et al. (2013) applied Zou et al. (2007)'s method to monitor GLM profiles where the response variable is binary. In this paper, the following MEWMA statistic is used to monitor the logistic regression profile:

$$\mathbf{u}_j = \theta\mathbf{z}_j + (1-\theta)\mathbf{u}_{j-1}, \quad (9)$$

in which

$$\mathbf{z}_j = (\boldsymbol{\Sigma}_0)^{\frac{-1}{2}}(\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_\mathbf{0}). \quad (10)$$

where θ is a smoothing parameter for MEWMA chart and $\theta \in [0, 1]$, $\mathbf{u}_0 = \mathbf{0}$. The control chart alarms a signal of OC for profile $j = 1,2,...$ if $\mathbf{u}_j^\mathbf{T}\mathbf{u}_j > h_E$, where h_E is set in a way that the value of ARL_0 equal to 200 is determined.

If $\boldsymbol{\beta}_\mathbf{0}$ and $\boldsymbol{\Sigma}_\mathbf{0}$ are unknown, they should be estimated by using historical data in Phase I, based on the estimators explained for Hotelling's T^2 control chart.

3. The Effect of Estimated Logistic Regression Parameters

To study the effect of estimated regression parameters on the efficiency of the Hotelling's T^2 and the MEWMA control schemes to monitor logistic regression profiles, the following steps are recommended in this paper:

1. Under the assumption of the known parameters $\boldsymbol{\beta} = (\beta_1,...,\beta_p)^T$ and $\boldsymbol{\Sigma}_\mathbf{0}$, h_r (for the Hotelling's T^2) and h_E (for the MEWMA) values are set such that ARL_0=200 is obtained, based on 10000 replications.
2. Generate m logistic regression profiles based on the known parameters.
3. Estimate $\hat{\boldsymbol{\beta}}_j = (\hat{\boldsymbol{\beta}}_{j,1},...,\hat{\boldsymbol{\beta}}_{j,p})^\mathbf{T}$, $j = 1,2,...,m$, for the samples produced in Step 2.

4. Calculate $\bar{\hat{\boldsymbol{\beta}}} = \frac{1}{m}\sum_{j=1}^{m}(\hat{\boldsymbol{\beta}}_j)$ and also $\hat{\boldsymbol{\Sigma}}_{\mathbf{0}}$ by using $\hat{\mathbf{W}} = diag(m_1\hat{\pi}_1(1-\hat{\pi}_1),.....,m_n\hat{\pi}_n(1-\hat{\pi}_n))$ where $\hat{\pi}_i = \frac{e^{x\bar{\hat{\boldsymbol{\beta}}}}}{1+e^{x\bar{\hat{\boldsymbol{\beta}}}}}$, $i = 1,2,...,n$. Then, set $RL = 1$.
5. Produce a random logistic profile with the known parameters.
6. Calculate both statistics, the Hotelling's T^2 and MEWMA, by using the estimated parameters in step 4, $\bar{\hat{\boldsymbol{\beta}}}$ and $\hat{\boldsymbol{\Sigma}}_{\mathbf{0}}$. Then, compare these values with the control limits h_r and h_E determined in step 1. The statistics for the Hotelling's T^2 and the MEWMA charts are $T_j^2 = (\hat{\boldsymbol{\beta}}_j - \bar{\hat{\boldsymbol{\beta}}})^T \hat{\boldsymbol{\Sigma}}_{\mathbf{0}}^{\mathbf{-1}}(\hat{\boldsymbol{\beta}}_j - \bar{\hat{\boldsymbol{\beta}}})$ and $\mathbf{u}_j = \theta\mathbf{z}_j + (1-\theta)\mathbf{u}_{j\text{-}1}$ where $\mathbf{z}_j = (\hat{\boldsymbol{\Sigma}}_{\mathbf{0}})^{\frac{-1}{2}}(\hat{\boldsymbol{\beta}}_j - \bar{\hat{\boldsymbol{\beta}}})$, respectively.
7. If the amount of the calculated statistics in step 6 are equal to or less than the determined control limits, then set $RL = RL + 1$ and go to step 5; otherwise, go to step 8 as the amount of statistic is bigger than the control limit.
8. Run length values should be recorded.
To obtain the values of ARL_0 for various values of k, steps 5–8 should be repeated 5000 times.
9. Now for computing *AARL* and *SDARL*, steps 2–8 are replicated 1000 times.

4. Simulation studies and remedial approaches

The effect of parameters estimation on the efficiency of the proposed control schemes, Hotelling's T^2 and MEWMA, for logistic regression profile monitoring is appraised by simulation investigations. In this paper, the numerical study presented in Ye et al. (2009) is utilized and the UCL of the charts is adjusted so that ARL_0 is approximately equal to 200 based on the in-control parameters, $\beta_1 = 3$ and $\beta_2 = 2$, in Phase II. Moreover, the smoothing parameter of the MEWMA control chart is selected as $\theta \in (0.05, 0.1, 0.2)$. To appraise the efficiency of both approaches under unknown parameters, different values for k are considered. Therefore, the upper control limit values of the desired control chart are determined for each of the values of k.

Designing both Hotelling's T^2 and MEWMA estimated parameters detrimentally influences the IC performances of these charts. To handle this problem, in the subsections 4.1 and 4.2, two methods are proposed to rectify the adverse effect of parameters estimation of the logistic regression profiles.

4.1. Increasing the Phase I sample size

Based on the results in Table 1, increasing the number of reference data in Phase I analysis, k, decreases the effect of estimated regression parameters on the $AARL_0$ of the proposed control schemes. Thus, it is recommended to increase k to gratify the adverse effect of the estimated parameters.

Table 1: $AARL_0$ and $SDRL_0$ values for various values of k

Chart	Criterion	*K*			
		5	15	30	∞
T²	*AARL*	141.9104	186.1313	198.2378	198.767
	SDARL	2.5335	2.7912	2.9024	2.9598
MEWMA $\boldsymbol{\theta = 0.05}$	*AARL*	41.5515	73.6207	101.3539	200.7379
	SDARL	0.8866	1.3916	1.4894	2.6732
MEWMA $\boldsymbol{\theta = 0.1}$	*AARL*	44.9836	81.4786	111.3561	200.2646
	SDARL	1.0293	1/4862	1.7604	2.6612
MEWMA $\boldsymbol{\theta = 0.2}$	*AARL*	54.0918	96.8898	128.8155	198.946
	SDARL	1.2186	1.6665	2.0261	2.9463

Note that, it is impossible to collect a large reference data as a sample to estimate the parameters of a process due to economic or other limitations. As a result, it is substantial to distinguish the minimum number k of reference data in Phase I to certify a required value of $AARL_0$. Based on the simulation results, the minimum number of reference data in Phase I, k, to achieve $AARL_0$ value of at least $\Delta = 100 \times \left(1 - \frac{ARL_\infty - ARL_m}{ARL_\infty}\right)$ percent for $\Delta \in \{75\%, 80\%, 85\%, 90\%\}$ are calculated in Table 2. According to the results of Table 2, for each value of Δ, the Hotelling's T^2 chart requires a smaller number of Phase I sample data in comparison to the MEWMA control chart. For instance, for the Hotelling's T^2 chart, $k = 11$ to estimate the regression parameters can achieve 90% of the desirable $AARL_0$ value. This percentage can be attained for MEWMA chart by utilizing $k = 164, 139, 104$ for $\theta = 0.05$, $\theta = 0.1$ and $\theta = 0.2$. Moreover, for MEWMA control chart, by increasing θ, k decreases to certify a predefined value of $AARL_0$.

Table 2: (k, ARL_0) values for various values of Δ

Chart	Δ			
	75%	80%	85%	90%
T^2	(6, 149.0752)	(7, 159.0136)	(9, 168.9519)	(11, 178.8903)
MEWMA $\theta = 0.05$	(94, 150.5534)	(111, 160.5903)	(139, 170.6272)	(164, 180.6641)
MEWMA $\theta = 0.1$	(68, 150.1984)	(82, 160.2116)	(111, 170.2249)	(139, 180.2481)
MEWMA $\theta = 0.2$	(44, 149.2102)	(55, 159.1576)	(71, 169.1049)	(104, 179.0523)

4.2. Modifying control limits

According to the results of Tables 1 and 2, by enhancing the number of Phase I sample size (k) the efficiency of the Hotelling's T^2 and the MEWMA control schemes improve for estimated parameters. While, it is impossible to wait for collecting adequate data set to obtain the predetermined value for the IC ARL. Thus, it is substantial to decrease the false alarms' rate without a large sample for Phase I, in such cases. In this paper, by simulation studies, the modified control limits of both the Hotelling's T^2 and MEWMA charts are determined such that the ARL_0 equal to 200 is obtained. Table 3 illustrates for both the Hotelling's T^2 and MEWMA control charts, decreasing the sample size of Phase I, makes the control limits wider. As can be seen, by increasing the value of k in both control charts, the value of UCL decreases and approaches the state where the parameters are known. When the value of k increases from 5 to 30, the percentage of decreasing changes in the UCL of the Hotelling's T^2 control chart is equal to 6.56, whereas this value is much less than the percentage change in the UCL of MEWMA control chart, which are 32.5, 28.57 and 25 percent for smoothing parameters 0.05, 0.1, and 0.2, respectively. According to these results, it is observed that by estimating the parameters and increasing the reference samples, the percentage of changes in the UCL of MEWMA control chart is larger than the Hotelling's T^2 control chart. Also, the percentage of these changes decreases with increasing the value of θ from 0.05 to 0.2.

Table 3: UCL values to attain $AARL_0 = 200$

Chart	K			
	5	15	30	∞
T^2	12.2	11.6	11.4	11.39
MEWMA $\theta = 0.05$	0.4	0.3	0.27	0.2035
MEWMA $\theta = 0.1$	0.7	0.6	0.5	0.482
MEWMA $\theta = 0.2$	1.6	1.4	1.2	1.1282

OUT-OF-CONTROL SHIFT DETECTION FOR PARAMETERS ESTIMATION

The effect of parameters estimation on the efficiency of the presented control schemes is investigated to discover various shifts in the regression parameters utilizing the same data set as in Section 4, in this section. Simulation studies are conducted to appraise the efficiency of the Hotelling's T^2 and MEWMA control charts, in terms of the $AARL_1$ and $SDARL_1$ criteria. The value of $AARL_0$ should be same for competing control charts, in order to have a comparison for different values of k.

$AARL_1$ and $SDARL_1$ criteria in Tables 4 and 5 for different values of k, $k\epsilon\{5،10،15،\infty\}$ are compared to evaluate the OC performances of the Hotelling's T^2 and MEWMA control charts under various changes in the regression parameters.

Table 4: $AARL_1$ and $SDARL_1$ values for various step shifts from β_1 to $\beta_1+\delta_1\sigma_{\beta_1}$

k	Charts	Criterion	δ_1			
			0.5	1	1.5	2
5	T^2	$AARL_1$	70.7986	11.7733	3.005	1.4401
		$SDARL_1$	1.6711	0.2839	0.0483	0.0132
	MEWMA $\theta=0.05$	$AARL_1$	42.3609	8.2279	5.1462	3.845
		$SDARL_1$	3.7829	0.0845	0.0182	0.0103
	MEWMA $\theta=0.1$	$AARL_1$	27.4688	5.9262	3.6395	2.7193
		$SDARL_1$	1.4597	0.0576	0.0147	0.0092
	MEWMA $\theta=0.2$	$AARL_1$	39.4304	5.4226	3.0115	2.2287
		$SDARL_1$	2.0441	0.0961	0.0142	0.0097
15	T^2	$AARL_1$	51.5563	8.5288	2.5145	1.3351
		$SDARL_1$	0.09228	0.1363	0.0286	0.0097
	MEWMA $\theta=0.05$	$AARL_1$	16.7531	6.7483	4.4106	3.3476
		$SDARL_1$	0.2434	0.0265	0.0125	0.0086
	MEWMA $\theta=0.1$	$AARL_1$	14.3402	5.1706	3.3302	2.5137
		$SDARL_1$	0.2358	0.0248	0.0119	0.0081
	MEWMA $\theta=0.2$	$AARL_1$	17.2215	4.5386	2.7571	2.0941
		$SDARL_1$	0.3823	0.0184	0.0113	0.0066
30	T^2	$AARL_1$	46.26	7.8122	2.4408	1.3091
		$SDARL_1$	0.7689	0.1078	0.0279	0.0087
	MEWMA $\theta=0.05$	$AARL_1$	14.5596	6.3329	4.1798	3.1926
		$SDARL_1$	0.1005	0.0236	0.0126	0.0077
	MEWMA $\theta=0.1$	$AARL_1$	11.2223	4.6242	3.0382	2.3143
		$SDARL_1$	0.0973	0.0208	0.0106	0.0072
	MEWMA $\theta=0.2$	$AARL_1$	12.1263	4.0399	2.5402	1.9675
		$SDARL_1$	0.1442	0.022	0.01	0.0064
∞	T^2	$AARL_1$	40.9765	7.24	2.3089	1.2888
		$SDARL_1$	0.5669	0.0985	0.0252	0.0087
	MEWMA $\theta=0.05$	$AARL_1$	11.3692	5.3802	3.6271	2.8099
		$SDARL_1$	0.0575	0.0191	0.01	0.0074
	MEWMA $\theta=0.1$	$AARL_1$	10.1778	4.4619	2.9628	2.2671
		$SDARL_1$	0.0644	0.0184	0.0099	0.0066
	MEWMA $\theta=0.2$	$AARL_1$	10.1789	3.8085	2.4441	1.9115
		$SDARL_1$	0.0867	0.0199	0.0094	0.0066

Table 5: $AARL_1$ and $SDARL_1$ values for different step shifts from β_2 to $\beta_2 + \delta_2\sigma_{\beta_2}$

k	**Charts**	**Criterion**	$\boldsymbol{\delta_2}$			
			0.5	1	1.5	2
5	T^2	$AARL_1$	69.9091	11.6332	3.0039	1.4495
		$SDARL_1$	1.6984	0.2726	0.049	0.0134
	MEWMA $\theta = 0.05$	$AARL_1$	42.4994	8.2353	5.1444	3.8452
		$SDARL_1$	3.6836	0.1332	0.0174	0.0111
	MEWMA $\theta = 0.1$	$AARL_1$	27.3787	5.9283	3.6399	2.7197
		$SDARL_1$	1.4538	0.061	0.015	0.0091
	MEWMA $\theta = 0.2$	$AARL_1$	39.1352	5.4177	3.0131	2.2302
		$SDARL_1$	2.157	0.0896	0.015	0.0078
15	T^2	$AARL_1$	50.6963	8.4277	2.5391	1.3425
		$SDARL_1$	0.8744	0.1369	0.0309	0.0098
	MEWMA $\theta = 0.05$	$AARL_1$	16.7601	6.746	4.4093	3.3486
		$SDARL_1$	0.2384	0.0258	0.0132	0.0086
	MEWMA $\theta = 0.1$	$AARL_1$	14.3273	5.1693	3.3313	2.5154
		$SDARL_1$	0.2281	0.0253	0.0115	0.008
	MEWMA $\theta = 0.2$	$AARL_1$	17.1193	4.5372	2.7588	2.0949
		$SDARL_1$	0.3671	0.0282	0.0111	0.0067
30	T^2	$AARL_1$	44.9908	7.6773	2.3876	1.3127
		$SDARL_1$	0.6984	0.1107	0.0262	0.0094
	MEWMA $\theta = 0.05$	$AARL_1$	14.5489	6.3308	4.1806	3.1929
		$SDARL_1$	0.1038	0.0228	0.0123	0.00081
	MEWMA $\theta = 0.1$	$AARL_1$	11.2105	4.6239	3.0384	2.3164
		$SDARL_1$	0.0968	0.0211	0.0105	0.0072
	MEWMA $\theta = 0.2$	$AARL_1$	12.0928	4.0404	2.5419	1.9669
		$SDARL_1$	0.1431	0.023	0.0101	0.0065
∞	T^2	$AARL_1$	39.9799	7.1071	2.2989	1.2925
		$SDARL_1$	0.5785	0.0955	0.0237	0.0085
	MEWMA $\theta = 0.05$	$AARL_1$	11.3412	5.3752	3.6528	2.808
		$SDARL_1$	0.0564	0.0188	0.0101	0.0077
	MEWMA $\theta = 0.1$	$AARL_1$	10.1493	4.4577	2.9619	2.2693
		$SDARL_1$	0.0656	0.0185	0.01	0.0067
	MEWMA $\theta = 0.2$	$AARL_1$	10.1368	3.8057	2.4448	1.9098
		$SDARL_1$	0.0852	0.0193	0.0094	0.0064

$AARL_1$ and $SDARL_1$ values for various shifts in the intercept and slope parameters in the unit of corresponding standard deviation are presented in Tables 4 and 5. It can be derived that the capability of the proposed control charts in discovering sustained shifts in the intercept and slope parameters based on the estimated parameters for regression, decreases with both $AARL_1$ and $SDARL_1$ criteria. Nevertheless, the detecting power of both the Hotelling's T^2 and MEWMA charts ameliorates, by enhancing the number of Phase I reference data, in terms of $AARL_1$ and $SDARL_1$

criteria. Table 5 represents that for each value of *k*, detecting power of the MEWMA for small changes, $\delta_1 \in \{0.5, 1\}$ outperforms Hotelling's T^2. While the Hotelling's T^2, for large shifts, $\delta_1 \in \{1.5, 2\}$ outperforms the MEWMA control chart.

5. Conclusion and suggestions for future research

In this paper, the effect of parameters estimation in Phase I on the efficiency of two proposed control schemes, Hotelling's T^2 and MEWMA control chart for monitoring logistic regression profiles in Phase II was assessed with *AARL* and *SDARL* criteria through extensive simulation investigations. The results showed that the designed control charts based on the estimated regression parameters, adversely affect the IC performance of both control charts. Furthermore, the minimum number of Phase I reference samples to obtain IC ARL equals to 200 for both control charts was proposed. Moreover, the modified UCL values with $AARL_0$= 200, were calculated for both control charts and the OC efficiency of the Hotelling's T^2 and MEWMA control charts was evaluated in terms of $AARL_1$ and $SDARL_1$ criteria. The simulation results show that by increasing the value of *k* in both control charts, the value of UCL decreases and approaches the state where the parameters are known. Also, by increasing the reference samples with estimated parameters, the percentage of changes in the UCL of MEWMA control chart is larger than the Hotelling's T^2 control chart. Moreover, the efficiency of the proposed control charts in discovering sustained changes in the intercept and slope parameters based on the estimated regression parameters decreases both $AARL_1$ and $SDARL_1$ criteria. Analyzing the parameters estimation's effect in Phase I on the efficiency of control schemes for monitoring the CV in Phase II is considered to be a potential subject in this field.

References

- Aly, A. A., Mahmoud, M. A., & Woodall, W. H. (2015). A comparison of the performance of phase II simple linear profile control charts when parameters are estimated. Communications in Statistics-Simulation and Computation, 44(6): 1432-1440.
- Ahmadi Yazdi, A. & Zeinal Hamadani, A, & Amiri, A. (2019a). Phase II monitoring of multivariate simple linear profiles with estimated parameters. Journal of Industrial Engineering International, 15, 557-570.
- Ahmadi Yazdi, A., Zeinal Hamadani, A., Karimi Gavareshki, M. H., & Amiri, A. (2019b). Phase II monitoring of multivariate profiles with estimated parameters and optimal phase I subgroups. Published online in Communications in Statistics-Simulation and Computation, doi: 10.1080/03610918.2019.1615626.
- Ahmadi Yazdi, A., Zeinal Hamadani, A., & Amiri, A. (2020). Addressing the effect of parameter estimation on phase II monitoring of multivariate multiple linear profiles via a new cluster-based approach. Communication in Statistics-Theory and Methods, 49(17), 4117-4132.
- Amiri, A., Eyvazian, M., Zou, C., and Noorossana, R. (2012). A parameters reduction method for monitoring multiple linear regression profiles. The International Journal of Advanced Manufacturing Technology, 58(5-8): 621-629.
- Amiri, A., Koosha, M. and Azhdari, A. (2011). Profile monitoring for poisson response. Proceedings of IEEM conference, IEEE, Singapore.
- Amiri, A., Koosha, M. and Azhdari, A. (2012). T2 based methods for monitoring gamma profiles. Proceedings of the 2012 International Conference on Industrial Engineering and Operations Management, Istanbul, Turkey.
- Amiri, A., Koosha, M., Azhdari, A., & Wang, G. (2015). Phase I monitoring of generalized linear model-based regression profiles. Journal of Statistical Computation and Simulation, 85(14): 2839-2859.
- Amiri, A., Zou, C., & Doroudyan, M. H. (2014). Monitoring correlated profile and multivariate quality characteristics. Quality and Reliability Engineering International, 30(1): 133-142.
- Amiri, A., Sogandi, F., & Ayoubi, M. (2018). Simultaneous monitoring of correlated multivariate

linear and GLM regression profiles in Phase II. Quality Technology & Quantitative Management, 15(4), 435-458.

- Bandara, K., Abdel-Salam, A. S. G., & Birch, J. B. (2020). Model robust profile monitoring for the generalized linear mixed model for Phase I analysis. Applied Stochastic Models in Business and Industry, 36(6), 1037-1059.
- Chakraborti, S. (2000). Run length, average run length and false alarm rate of Shewhart X-bar chart: exact derivations by conditioning. Communications in Statistics- Simulation and Computation, 29(1): 61-81.
- Chen, G. (1997). The mean and standard deviation of the run length distribution of X2 charts when control limits are estimated. Statistica Sinica, 23(4): 789-798.
- Chen, Y., Birch, J. B., & Woodall, W. H. (2016). Effect of phase I estimation on phase II control chart performance with profile data. Quality and Reliability Engineering International, 32(1): 79-87.
- Cheng, T. C., & Yang, S. F. (2018). Monitoring profile based on a linear regression model with correlated errors. Quality Technology & Quantitative Management, 15(3), 393-412.
- Davidson, J., Summerfelt, S. (2004) Solids flushing, mixing, and water velocity profiles within large (10 and 150 m3) circular 'Cornell-type' dual-drain tanks, Aquacultural Engineering, 32(1), 245-271.
- Ding, Y., Zeng L., and Zhou, S. (2006). Phase I analysis for monitoring nonlinear profiles in manufacturing processes. Journal of Quality Technology, 38: 199–216.
- Eyvazian M., Noorossana R., Saghaei A. and Amiri A. (2011). Phase II monitoring of multivariate multiple linear regression profiles. Quality and Reliability Engineering International, 27(3): 281-296.
- Hakimi, A., Amiri, A., & Kamranrad, R. (2017). Robust approaches for monitoring logistic regression profiles under outliers. International Journal of Quality & Reliability Management, 34(4): 494-507.
- Hu, X., & Castagliola, P. (2017). Guaranteed conditional design of the median chart with estimated parameters. Quality and Reliability Engineering International, 33(8): 1873-1884.
- Izadbakhsh, H., Noorossana, R., Zarinbal, M., Zarinba, l A., Safaian, M. and Chegeni, M. (2011). An EWMA–based method for monitoring polytomous logistic profiles. Proceeding of the 2011 International Conference on Industrial Engineering and Engineering Management (IEEM).
- Izadbakhsh, H., Noorossana, R., & Niaki, S. T. A. (2018). Monitoring multinomial logistic profiles in Phase I using log-linear models. International Journal of Quality & Reliability Management. 35(3), 678-689.
- Jensen, W. A. and Birch, J. B. (2009). Profile monitoring via nonlinear mixed models. Journal of Quality Technology, 41: 18-34.
- Jones, L. A., Champ, C. W., & Rigdon, S. E. (2001). The performance of exponentially weighted moving average charts with estimated parameters. Technometrics, 43(2): 156-167
- Jones, L. A., Champ, C. W., and Rigdon, S. E. (2004). The run length distribution of the CUSUM with estimated parameters. Journal of Quality Technology, 36(1): 95-108.
- Kang, L., & Albin, S. L. (2000). On-line monitoring when the process yields a linear profile. Journal of Quality Technology, 32(4): 418-426.
- Kinat, S., Amin, M., & Mahmood, T. (2020). GLM-based control charts for the inverse Gaussian distributed response variable. Quality and Reliability Engineering International, 36(2), 765-783.
- Kazemzadeh, R. B., Noorossana R. and Amiri A. (2008). Performance evaluation of control charts for detecting outliers in phase I monitoring of polynomial profiles. Proceedings of the 9th Iranian Statistical Conference. Isfahan, Iran, August 20-22.
- Kazemzadeh, R. B., Noorossana, R. and Amiri, A. (2008). Phase I monitoring of polynomial profiles. Communications in Statistics, Theory and Methods, 37: 1671-1686.
- Kazemzadeh, R. B., Noorossana, R., and Amiri, A. (2009). Monitoring polynomial profiles in quality control applications. The International Journal of Advanced Manufacturing Technology, 42(7-8): 703-712.
- Kim, K., Mahmoud, M. A., Woodall, W. H. (2003). On the monitoring of linear profiles. Journal of Quality Technology 35: 317–328.
- Koosha, M., and A. Amiri. (2013). Generalized linear mixed model for monitoring autocorrelated logistic regression profiles. The International Journal of

Advanced Manufacturing Technology, 64 (1–4): 487–95.

- Li, J., Zou, C., Wang, Z., Huwang, L. (2013). A multivariate sign chart for monitoring process shape parameters. Journal of Quality Technology. 45: 149-165.
- Mahmoud, M. A. (2012). The performance of phase II simple linear profile approaches when parameters are estimated. Communications in Statistics-Simulation and Computation, 41(10): 1816-1833.
- Mahmoud, M. A., & Maravelakis, P. E. (2010). The performance of the MEWMA control chart when parameters are estimated. Communications in Statistics-Simulation and Computation, 39(9): 1803-1817.
- Mahmoud, M. A., & Woodall, W. H. (2004). Phase I analysis of linear profiles with calibration applications. Technometrics, 46(4): 380-391.
- Mahmoud, M. A., Morgan, J. P., & Woodall, W. H. (2010). The monitoring of simple linear regression profiles with two observations per sample. Journal of Applied Statistics, 37(8): 1249-1263.
- Mammadova, U., & Özkale, M. R. (2021). Profile monitoring for count data using poisson and conway–maxwell–poisson regression-based control charts under multicollinearity problem. Journal of Computational and Applied Mathematics, 388, 113275.
- Maleki, M. R., A. Amiri, and A. R. Taheriyoun. (2017). Phase II monitoring of binary profiles in the presence of within-profile autocorrelation based on Markov model. Communications in Statistics-Simulation and Computation, 46 (10):7710–32.
- Maleki, M. R., Amiri, A., & Taheriyoun, A. R. (2018). Identifying the time of step change and drift in phase II monitoring of autocorrelated logistic regression profiles. Scientia Iranica, 25(6), 3654-3666.
- Maleki, M. R., Castagliola, P., Amiri, A., & Khoo, M. B. (2019). The effect of parameter estimation on phase II monitoring of poisson regression profiles. Communications in Statistics-Simulation and Computation, 48(7), 1964-1978.
- Maravelakis, P. E., & Castagliola, P. (2009). An EWMA chart for monitoring the process standard deviation when parameters are estimated. Computational Statistics & Data Analysis, 53(7): 2653-2664.
- Moheghi, H. R., Noorossana, R., & Ahmadi, O. (2021). GLM profile monitoring using robust estimators. Quality and Reliability Engineering International, 37(2), 664-680.
- Montgomery, D. C. (2005). Introduction to Statistical Quality Control, 5th edition, John Wiley & Sons.
- Noorossana, R. and Izadbakhsh, H. (2013). Profile monitoring for multinomial responses. International Journal of Industrial Engineering & Production Management, 23(4): 417-429.
- Noorossana, R., Aminmadani, M., & Saghaei, A. (2016). Effect of phase I estimation error on the monitoring of simple linear profiles in phase II. The International Journal of Advanced Manufacturing Technology, 84(5-8): 873-884.
- Noorossana, R., Aminnayeri. M. and Izadbakhsh, H. (2013). Statistical monitoring of polytomous logistic profiles in phase II. Scientia Iranica, 20(3): 958-966.
- Noorossana, R., Eyvazian M., Amiri A. and Mahmoud M.A. (2010). Statistical monitoring of multivariate multiple linear regression profiles in phase I with calibration application. Quality and Reliability Engineering International, 26(3): 291-303.
- Paynabar, K., Jin, J. and Yeh, A. B. (2012). Phase I risk-adjusted control charts for monitoring surgical performance by considering categorical covariates. Journal of Quality Technology, 44(1): 39-53.
- Piri, S., Abdel-Salam, A. S. G., & Boone, E. L. (2021). A wavelet approach for profile monitoring of Poisson distribution with application. Communications in Statistics-Simulation and Computation, 50(2), 525-536.
- Saleh, N. A., Mahmoud, M. A., Jones-Farmer, L. A., Zwetsloot, I. N. E. Z., & Woodall, W. H. (2015). Another look at the EWMA control chart with estimated parameters. Journal of Quality Technology, 47(4): 363-382.
- Shang, Y., Tsung, F. and Zou, C. (2011). Phase-II profile monitoring with binary data and random predictors. Journal of Quality Technology, 43:196-208.
- Shang, Y., Tsung, F., & Zou, C. (2011). Profile monitoring with binary data and random predictors. Journal of Quality Technology, 43(3), 196-208.

- Soleymanian M.E., Khedmati M. and Mahlooji H. (2013). Phase II monitoring of binary response profiles. Scientia Iranica, 20(6): 2238-2246.

- Testik, M. C. (2007). Conditional and marginal performance of the Poisson CUSUM control chart with parameter estimation. International Journal of Production Research, 45(23): 5621-5638.

- Williams, J. D., Woodall W.H. and Birch, J. B. (2007). Statistical Monitoring of Nonlinear Product and Process Quality Profiles. Quality and Reliability Engineering International, 23: 925-941.

- Yeh, A. B., Huwang, L., & Li, Y. M. (2009). Profile monitoring for a binary response. IIE Transactions, 41(11): 931-941.

- Zand, A., Yazdanshenas, N., & Amiri, A. (2013). Change point estimation in phase I monitoring of logistic regression profile. The International Journal of Advanced Manufacturing Technology, 67(9-12), 2301-2311.

- Zhang, H. and Albin, S. (2009). Detecting outliers in complex profiles using $\chi 2$ control chart method. IIE Transactions, 41(4): 335-345.

- Zhang, J., Li, Z., & Wang, Z. (2009). Control chart based on likelihood ratio for monitoring linear profiles. Computational statistics & data analysis, 53(4): 1440-1448.

- Zhang, M., Peng, Y., Schuh, A., Megahed, F. M., & Woodall, W. H. (2013). Geometric charts with estimated control limits. Quality and Reliability Engineering International, 29(2): 209- 223.

- Zou, C., Tsung, F. and Wang, Z. (2008). Monitoring profiles based on nonparametric regression methods, Technometrics, 50(4): 512-526.

- Zou, C., Tsung, F., & Wang, Z. (2007). Monitoring general linear profiles using multivariate exponentially weighted moving average schemes. Technometrics, 49: 395–408.

- Zwetsloot, I. M., & Woodall, W. H. (2017). A head-to-head comparative study of the conditional performance of control charts based on estimated parameters. Quality Engineering, 29(2): 244-253.