Directional Control Schemes for Multivariate Categorical Processes

Jian Li[1], Fugee Tsung[1], and Changliang Zou[2*]

[1] *Department of Industrial Engineering and Logistics Management, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong*

[2] *LPMC and Department of Statistics, School of Mathematical Sciences, Nankai University, Tianjin, China*

The statistical process control of multivariate categorical processes is considered. A Phase II log-linear directional control chart is proposed which exploits directional shift information and integrates the monitoring of multivariate categorical processes into the unified framework of multivariate binomial and multivariate multinomial distributions. A diagnostic scheme is suggested for identifying the shift direction. Both the control chart and the diagnostic approach are simple and quick to compute. Numerical simulations and practical guided applications are presented to demonstrate their effectiveness.

Key Words: Contingency Table; EWMA; Generalized Likelihood Ratio Test; Multivariate Multinomial Distribution; Statistical Process Control

Introduction

The overall quality of most modern processes is best described in terms of multiple characteristics. Multivariate statistical process control (SPC) is applied in practical situations where several quality characteristics have to be monitored simultaneously. Much effort has

*Corresponding author. Email: chlzou@yahoo.com.cn

been devoted to the monitoring problem in settings where all the observed variables are numerical and continuous. For instance, tensile strength and diameter are two important quality characteristics of a textile fiber, which must be jointly controlled and have been assumed to follow a bivariate normal distribution. Refer to Lowry and Montgomery (1995) and Bersimis et al. (2007) for thorough reviews of monitoring multivariate continuous processes.

In manufacturing and especially the service industry, however, there are more and more quality characteristics whose values cannot be measured numerically. Obtaining their continuous values is expensive, unnecessary, or even impossible, however, collecting some attribute levels of them may be available and may be a low cost. These classification levels are rough and do not need precise measurements. Examples include items on a production line whose several quality characteristics each are evaluated as conforming or nonconforming to predefined specifications, and multiple indexes in a service flow that may be assessed as excellent, acceptable, or unacceptable. The characteristics all have two or more attribute levels, and the processes are multivariate categorical. Note that this is similar to multiple factors in Design of Experiments (DOE), where each factor also has several specific levels. Here for simplicity we use "factor" to represent a categorical characteristic.

The p-chart and the np-chart for binomial distributed variables, together with the c-chart and the u-chart for Poisson processes are typical SPC tools used with univariate categorical processes. To monitor multiple factors, we may employ multiple univariate categorical control charts as a multi-chart (see Woodall and Ncube (1985)). However, it is usually an unwise choice. The reason is, first, that it is difficult to determine the control limit given a desired average run length (ARL), which is defined as the average number of samples needed for the control chart to signal. The control limits of the separate univariate charts need to be set such that each chart achieves a specific individual in-control (IC) ARL, and the overall IC ARL of the multi-chart therefore attains a pre-specified value. But determining these control limits is nontrivial even for the low-dimensional case of a small number of categorical factors, let alone for a general multivariate categorical distribution. This complexity also increases remarkably when the marginal distributions of categorical factors are not identical, since the individual charts have different run length

distributions. Also, a multi-chart considers individual categorical factors in parallel, and therefore is unable to account for any correlations among them. Naturally, it is desirable to introduce multivariate categorical control charts that can appropriately describe and exploit the relationships among multiple categorical factors.

Woodall (1997) summarized many aspects of the control charts for attribute data, but most of them are univariate categorical ones. In the literature, some efforts have been devoted to the monitoring of multinomial and multiattribute processes. See Topalidou and Psarakis (2009) for a nice overview. Among others, Patel (1973) suggested a χ^2-chart for multivariate binomial or multivariate Poisson populations, which is based on the assumption that the joint distribution of the correlated binomial or Poisson variables can be approximated by a multivariate normal distribution given a sufficiently large sample size. More recent developments include the Shewhart-type *mnp*-chart proposed by Lu et al. (1998) using the weighted sum of the number of nonconforming units with respect to each quality characteristic. There is also the *mp*-chart designed by Chiu and Kuo (2008) for multivariate Poisson count data. However, as with Patel's χ^2-chart, these two charts can only deal with factors with two levels. On the other hand, some methods also focus on monitoring multinomial processes that have only one factor with more than two levels, such as the generalized p-chart developed by Marcucci (1985). This chart extends the traditional p-chart by adopting the Pearson chi-square statistic. The multinomial cumulative sum (CUSUM) chart proposed by Ryan et al. (2011) can also deal with problem, which is based on the likelihood ratio statistic equipped with the CUSUM scheme. Note that the methods of monitoring multinomial processes are actually univariate charts, since only one factor is involved. With several factors where at least one of them has more than two levels, there are no appropriate approaches available. Aside from this deficiency, most of the existing methods focus entirely on the marginal sums with respect to each categorical factor, neglecting the cross-classifications between factors. If some cross-classification probabilities shift to out-of-control (OC) states, these charts may not detect them quickly. To sum up, a general monitoring methodology for multivariate categorical processes is required.

Multi-way analysis of variance (ANOVA) models in DOE may provide some assistance if the observations are assumed to be linearly dependent on the levels of several factors. In

particular, an observation relies on main factor effects and interaction effects. As with the observations in an ANOVA model, the logarithms of the cross-classification probabilities may also depend linearly on the levels of these multiple factors, which forms a log-linear model (see Bishop et al. (2007)). The log-linear model characterizes the association and interaction patterns among the categorical factors appropriately, and therefore can be used to develop multivariate categorical control charts. This is not the first time log-linear models have been applied in SPC. Qiu (2008) dealt with distribution-free monitoring schemes for multivariate continuous processes by dichotomizing continuous multivariate data into categorical data. He then estimated the IC factor distribution using log-linear models in Phase I of SPC, and proposed a Phase II multivariate cumulative sum (CUSUM) chart by employing the Pearson chi-square statistic.

In our study, a systematic directional monitoring mechanism and a diagnostic scheme for multivariate categorical processes were developed based on log-linear models. By analogy with multi-way ANOVA, the cross-classification probabilities are expressed in terms of main factor effects and factor interaction effects. The log-linear model can then be equivalently rewritten as a regression model, leading to one-to-one correspondence between factor effects and coefficient subvectors. The potential shifts to OC states that arise in factor effects therefore appear in their corresponding coefficient subvectors. To monitor a process as efficiently as possible, such practical information formulated as shift directions should be exploited. A Phase II control chart was accordingly developed based on the log-likelihood function of the log-linear model, and an exponentially weighted moving average (EWMA) scheme. The suggested control chart promises to incorporate a unified framework of multiple factors with at least one with more than two levels. Applying some reliable approximations, the resulting chart is easy to construct and convenient to implement. Furthermore, a diagnostic scheme was developed to identify the shift direction once there is an OC signal. Some implementation guidelines are provided and illustrated using a practical example. Monte Carlo simulations were performed to demonstrate the effectiveness of the proposed chart and diagnostic approach.

Multivariate Categorical Processes and Conventional Monitoring Approaches

Multivariate Categorical Processes

We first illustrate the multivariate categorical process with two motivating examples. The aluminium electrolytic capacitor (AEC) manufacturing is a multi-stage process, and the quality of the semi-finished AECs is inspected immediately after each stage. Here for illustration, we concentrate on the quality after the aging stage, which is assessed mainly in terms of leakage current (LC), dissipation factor (DF), and capacity (CAP). Each characteristic is classified automatically as conforming or nonconforming to the specifications by an electronic device at a very high speed. So engineers are reluctant to obtain their specific continuous or numerical values (not impossible but requires more cost). This is a multivariate categorical process with three factors LC, DF and CAP, each with two levels and therefore $2^3 = 8$ level combinations. Without loss of generality, for each factor -1 represents "conforming" and 1 "nonconforming". For example, the combination (-1, 1, -1) means an AEC with conforming LC and CAP and nonconforming DF.

Another example is the quality control of welding rods. One of the key aspects of welding rod inspection is the appearance, which directly reflects the integrated level of welding rod manufacturing and influences the welding performance. The welding rod is composed of a cylindrical metallic core wire and a coating composition (flux) covering the circumference of the metallic core wire. Its appearance has some important indexes, including eccentricity of the core wire, moisture resistance of coating, strength of coating, and rod bend. During testing, each of them is simply evaluated as conforming or nonconforming, and their (latent) continuous values are not considered. With the four factors: eccentricity, moisture resistance, strength, and bend, each two attribute levels, this is also a multivariate categorical process with $2^4 = 16$ cross-classification level combinations. Such processes with multiple categorical factors do not have numerical values, and the existing methods for monitoring multivariate continuous data cannot apply to them.

Now we turn to a general multivariate categorical process. Suppose that there are p

factors $C_1, \ldots, C_p$, and that each classification factor C_i takes a number, say h_i, of possible levels. The overall cross-classifications among all the level combinations of these factors form a p-way $h_1 \times \ldots \times h_p$ contingency table with $h = \prod_{i=1}^{p} h_i$ cells. Each cell corresponds to one level combination of the p factors and stores the count under this level combination.

For a simple $h_1 \times h_2 \times h_3$ three-way table, denote the observed count by n_{ijk} in cell(i, j, k) $(i = 1, \ldots, h_1;\ j = 1, \ldots, h_2;\ k = 1, \ldots, h_3)$, the expected count by m_{ijk}. If observations are made over a period of time without *a priori* knowledge of the total number of observations, it is reasonable to assume that each cell follows an independent Poisson distribution (see Bishop et al. (2007)). During the monitoring process in Phase II of SPC, the total number N of observations is usually fixed. Conditional on this total sample size N, a series of independent Poisson distributions result in a multinomial distribution. So the cell counts in the three-way contingency table therefore follow the multinomial distribution MN$(N; p_{ijk})$ $(i = 1, \ldots, h_1;\ j = 1, \ldots, h_2;\ k = 1, \ldots, h_3)$. Here $p_{ijk} = m_{ijk}/N$ is the probability of an observation falling into cell(i, j, k), and these probabilities must sum up to 1.

To generalize the three-way table to a general p-way $h_1 \times \ldots \times h_p$ contingency table (see Johnson et al. (1997)), let the probability of obtaining the combination of factor levels $a_1, \ldots, a_p$ be $p_{a_1 \ldots a_p}$ $(a_i = 1, \ldots, h_i$ and $i = 1, \ldots, p)$. Furthermore, denote the count of observations among a sample of size N with the combination $a_1 \ldots a_p$ by $n_{a_1 \ldots a_p}$. Clearly, the cell counts $n_{a_1 \ldots a_p}$ jointly follow MN$(N; p_{a_1 \ldots a_p})$. Furthermore, consider for example the group of marginal counts $n_{(i)1}, \ldots, n_{(i)h_i}$ of the factor C_i, which are actually the sums of the cell counts $n_{a_1 \ldots a_p}$ adding along the attribute levels of all the other factors except C_i. It is easy to understand that this group of marginal sums follow the multinomial distribution MN$(N; p_{(i)1}, \ldots, p_{(i)h_i})$. Here $p_{(i)1}, \ldots, p_{(i)h_i}$ are the marginal probabilities of the factor C_i, which can be calculated in a similar way to $n_{(i)1}, \ldots, n_{(i)h_i}$ based on the cell probabilities $p_{a_1 \ldots a_p}$. The joint distribution of the p groups of variables $n_{(i)1}, \ldots, n_{(i)h_i}$ $(i = 1, \ldots, p)$, each being a multinomial distribution, is defined to be a multivariate multinomial distribution (see Johnson et al. (1997)). When each factor has two levels, this reduces naturally to the multivariate binomial distribution. So by applying multivariate binomial or multivariate multinomial distributions and a cross-classified contingency table, multivariate categorical processes can be studied.

Conventional Monitoring Approaches

Generally, statistical process control involves two phases (Montgomery (2009)). In Phase I, a set of process data are collected and examined. Any unusual patterns in the data are identified, and based on this, the data and the process may be adjusted, resulting in a clean dataset collected from stable conditions of the process. This dataset is called the IC dataset and used for estimating the IC parameters representative of the IC operating conditions. Phase I analysis primarily assists in bringing the process into a state of statistical control. Given the IC parameters, in Phase II the process is monitored with control charts to detect and diagnose any deviations from the IC state. Let us review some typical methods for monitoring multivariate binomial and multivariate multinomial processes. Hereafter we use the superscript "(0)" and "(1)" to denote the IC and OC states, respectively.

With a multivariate binomial process, each factor has two levels, shaping a p-way contingency table with 2^p cells for p factors. Denote the two levels of each factor by 1 and 0. Given a sample size N in Phase II, if the process is IC, the Level 1 count of the factor C_i $(i = 1, \ldots, p)$ is binomially distributed as the binomial distribution with the total size N and its IC Level 1 probability. So the IC mean of this count is known, and the IC covariance between any two factors C_i and C_j $(i, j = 1, \ldots, p$ and $i \neq j)$ can also be calculated, hence the IC mean vector and the IC covariance matrix of the Level 1 counts of the p factors. Based on these, Patel (1973) constructed the χ^2 charting statistic of the Hotelling's T^2 form, the expression of which can be found in Appendix A in the supplemental file.

Factors with three or more levels are also common in production and service applications, and they can be treated as multivariate multinomial processes. Take customer attitudes towards a service for instance. Suppose that there are four indexes, each of which may take the values of excellent, acceptable, or unacceptable. This forms a four-way contingency table with 3^4 cells. No appropriate monitoring methods exist that incorporate the cross-classifications among the p factors when at least one of them has more than two attribute levels.

The only feasible approach of monitoring multivariate multinomial processes, albeit a naive one, might be to monitor the p groups of marginal sums of each factor by adopting

p individual charts. If we consider only the group of marginal sums of the factor C_i ($i = 1, \ldots, p$), it is a multinomial process, which could be tackled by applying Marcucci's (1985) generalized p-chart. For the kth sample, the generalized p-chart employs the Pearson chi-square statistic as the charting statistic for the factor C_i, and its expression is also listed in Appendix A in the supplemental file. Finally, there would be p separate charts, which jointly form a multi-chart. Such a multi-chart would signal whenever at least one of the p individual charts signals.

Clearly, the χ^2-chart applies to only multivariate binomial processes, and the multi-chart developed for multivariate multinomial processes is troublesome. Besides, both of them pay sufficient attention to only the one-way marginal sums of each factor, and they almost neglect to a large extent the cross-classification interactions among factors.

New Methodologies for Monitoring and Diagnosis

Log-Linear Models

There is a clear need to model the relationship between each cell count and factor levels associated with it. The cell counts are stored in a multi-way contingency table. As in the multi-way ANOVA where responses to all factor level combinations are also placed in a multi-way table. The ANOVA is based on the assumption that the responses are normally distributed, and it aims to quantify how the responses are influenced by the main factor effects and the factor interaction effects. Take for illustration a three-way ANOVA model, where the three factors take h_1, h_2, and h_3 levels, respectively. Denote the expected response with the first factor at its ith level, the second factor at its jth level, and the third factor at its kth level as y_{ijk} ($i = 1, \ldots, h_1$; $j = 1, \ldots, h_2$; $k = 1, \ldots, h_3$). The three-way ANOVA model is

$$y_{ijk} = u^{(0)} + u_i^{(1)} + u_j^{(2)} + u_k^{(3)} + u_{i,j}^{(1,2)} + u_{i,k}^{(1,3)} + u_{j,k}^{(2,3)} + u_{i,j,k}^{(1,2,3)},$$

where $u^{(0)}$ is the overall mean, $u^{(1)}, u^{(2)}, u^{(3)}$ are the main effects, $u^{(1,2)}, u^{(1,3)}, u^{(2,3)}$ are the two-factor interaction effects, and $u^{(1,2,3)}$ is the three-factor interaction effect. Furthermore,

identifiability requires constraints such as

$$\sum_i u_i^{(1)} = \sum_i u_{i,j}^{(1,2)} = \sum_i u_{i,k}^{(1,3)} = \sum_i u_{i,j,k}^{(1,2,3)} = 0$$

for the first factor along its index i. Similar equations describe the second and third factors along with their indexes j and k, respectively. Therefore, the ANOVA can be represented as a linear regression model. From the generalized linear model (GLM) point of view, a linear regression model where the response is normally distributed has a canonical link function of unity (see McCullagh and Nelder (1989)).

By analogy with the ANOVA, it is possible to build a similar regression model relating the cell counts and their corresponding factor levels in the multi-way contingency table. Note that the response (the cell count) is no longer normally distributed. With no restriction on the total sample size, each cell count is independently Poisson distributed, and a GLM where the response follows a Poisson distribution has a canonical link function of logarithm (see McCullagh and Nelder (1989)). Therefore, the log-linear model can be introduced. For a three-way contingency table of size $h_1 \times h_2 \times h_3$, the log-linear model characterizing the relationship between the expectation m_{ijk} ($i = 1, \ldots, h_1$; $j = 1, \ldots, h_2$; $k = 1, \ldots, h_3$) of the count in cell(i, j, k) and the factor levels indexed with i, j, k, is

$$\ln m_{ijk} = u^{(0)} + u_i^{(1)} + u_j^{(2)} + u_k^{(3)} + u_{i,j}^{(1,2)} + u_{i,k}^{(1,3)} + u_{j,k}^{(2,3)} + u_{i,j,k}^{(1,2,3)},$$

where the u-terms are the main or factor interaction effects defined as in the ANOVA model (see Bishop et al. (2007)). They also satisfy the identifiability constraints. When the total sample size N is fixed, the cell counts jointly follow a multinomial distribution, and it is more convenient to focus on the probability p_{ijk} instead of the expectation $m_{ijk} = Np_{ijk}$. In this case, the log-linear model will be

$$\ln p_{ijk} = u^{(0)} + u_i^{(1)} + u_j^{(2)} + u_k^{(3)} + u_{i,j}^{(1,2)} + u_{i,k}^{(1,3)} + u_{j,k}^{(2,3)} + u_{i,j,k}^{(1,2,3)}, \tag{1}$$

where the probabilities must satisfy $\sum_{i,j,k} p_{ijk} = 1$. Obviously, the interaction terms such as $u^{(1,2)}$ reflect the dependence among the factors, and for a log-linear model without any interaction effects, the factors are independent.

The identifiability constraints applicable to a log-linear model in the form of Equation (1) are somewhat inconvenient to write out, but they can be rewritten equivalently as an-

other form, which is illustrated by a 2×3 contingency table. The identifiability constraints allow setting

$$\begin{array}{lll} u^{(0)} = \beta_0, & u_1^{(1)} = \beta_1, & u_2^{(1)} = -\beta_1, \\ u_1^{(2)} = \beta_2, & u_2^{(2)} = \beta_3, & u_3^{(2)} = -\beta_2 - \beta_3, \\ u_{1,1}^{(1,2)} = \beta_4, & u_{1,2}^{(1,2)} = \beta_5, & u_{1,3}^{(1,2)} = -\beta_4 - \beta_5, \\ u_{2,1}^{(1,2)} = -\beta_4, & u_{2,2}^{(1,2)} = -\beta_5, & u_{2,3}^{(1,2)} = \beta_4 + \beta_5. \end{array}$$

Therefore, the logarithms of the probabilities p_{ij} $(i = 1, 2;\ j = 1, 2, 3)$ can be expressed as a linear combination of the coefficients β_k $(k = 0, 1, \ldots, 5)$. Clearly, β_1 measures the main effect $u^{(1)}$ of the first factor, $[\beta_2, \beta_3]^T$ measures the main effect $u^{(2)}$ of the second factor, and $[\beta_4, \beta_5]^T$ measures the interaction effect $u^{(1,2)}$ of the two factors. That these factor effects can be represented totally by coefficients can be extended to a general scenario.

Actually, the identifiability constraints dictate that the log-linear model for a p-way contingency table where p factors are considered in the form of Equation (1) can be expressed in the following regression form (see Dahinden et al. (2007))

$$\ln \mathbf{p} = \mathbf{1}\beta_0 + \sum_{i=1}^{2^p - 1} \boldsymbol{X}_i \boldsymbol{\beta}_i, \tag{2}$$

where $\mathbf{p}$ is the $h \times 1$ probability vector corresponding to the h cells of the contingency table, $\mathbf{1}$ is a column vector consisting of 1 as all its entries with appropriate dimensions, $\boldsymbol{X}_i$ is an $h \times q_i$ design submatrix corresponding to the ith main or interaction effect and containing 1, 0, or -1 as its elements, and $\boldsymbol{\beta}_i$ is the coefficient subvector of size $q_i \times 1$. Still take the above two-way contingency table of size 2×3 with $p = 2$ factors for illustration. Here $\mathbf{p} = [p_{11}, p_{12}, p_{13}, p_{21}, p_{22}, p_{23}]^T$ and $\mathbf{1} = \mathbf{1}_6$ are both of size 6×1, and $\boldsymbol{\beta}_1 = \beta_1$, $\boldsymbol{\beta}_2 = [\beta_2, \beta_3]^T$, $\boldsymbol{\beta}_3 = [\beta_4, \beta_5]^T$, together with the design submatrixes

$$\boldsymbol{X}_1 = \begin{bmatrix} \mathbf{1}_3 \\ -\mathbf{1}_3 \end{bmatrix},\ \boldsymbol{X}_2 = \begin{bmatrix} \boldsymbol{J}_3 \\ \boldsymbol{J}_3 \end{bmatrix},\ \text{and } \boldsymbol{X}_3 = \begin{bmatrix} \boldsymbol{J}_3 \\ -\boldsymbol{J}_3 \end{bmatrix},$$

where

$$\boldsymbol{J}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \text{ and } \boldsymbol{J}_3 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_2 \\ -\mathbf{1}_2^T \end{bmatrix}.$$

Note that the column sums of $\boldsymbol{J}_2$ and $\boldsymbol{J}_3$ are all zeros, which assures identifiability.

Denote the design matrix by $\widetilde{\boldsymbol{X}} = [\mathbf{1}, \boldsymbol{X}]$ with $\boldsymbol{X} = [\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{2^p-1}]$ and the coefficient vector by $\widetilde{\boldsymbol{\beta}} = [\beta_0, \boldsymbol{\beta}^T]^T$ with $\boldsymbol{\beta} = [\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_{2^p-1}^T]^T$. So Equation (2) can be rewritten as $\ln \mathbf{p} = \widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{\beta}}$. Actually, the log-linear model (2) is at the effect level, and there are in total $2^p - 1$ effects from the main effects up to the p-factor interaction effect. Clearly, we have $\mathbf{p}^T\mathbf{1} = 1$, and β_0 is a scalar representing the intercept. If $\boldsymbol{\beta}$ is known, the first entry β_0 is determined. Therefore, only $h-1$ coefficients are free to vary independently, and β_0 exists merely for accommodating the constraint $\mathbf{p}^T\mathbf{1} = 1$. Hereafter, attention will mainly be paid to the coefficient subvectors $\boldsymbol{\beta}_i$ $(i = 1, \ldots, 2^p - 1)$. In the case of factors all with two levels, the derivation of the design matrix $\widetilde{\boldsymbol{X}}$ is identical to that of the design matrix of 2^k full factorial experiment with 1 and -1 representing the high and low levels, respectively. However, this becomes a little complex if at least one factor has more than two levels. Refer to the additional File 1 in Dahinden et al. (2007) for the general result of deriving $\widetilde{\boldsymbol{X}}$. For convenience of practitioners, we provide the Fortran code for deriving it once the number of factors and their levels are known, which is available from the authors upon request. In addition, an example of four factors is given in Appendix B in the supplemental file.

The design submatrixes, together with their corresponding coefficient subvectors, are arranged from the overall mean, the main effects, up to the effect of the highest order. For three factors C_1, C_2, and C_3 with 2, 3, and 3 levels, respectively, the sequence is 1, C_1, C_2, C_3, C_1C_2, C_1C_3, C_2C_3, and $C_1C_2C_3$. Hence, the coefficient vector

$$\begin{aligned}\widetilde{\boldsymbol{\beta}} \;=\; [\;& \beta_0 \quad \beta_{(1)} \quad \beta_{(2_1)} \quad \beta_{(2_2)} \quad \beta_{(3_1)} \quad \beta_{(3_2)} \\ & \beta_{(1,2_1)} \quad \beta_{(1,2_2)} \quad \beta_{(1,3_1)} \quad \beta_{(1,3_2)} \quad \beta_{(2_1,3_1)} \quad \beta_{(2_1,3_2)} \\ & \beta_{(2_2,3_1)} \quad \beta_{(2_2,3_2)} \quad \beta_{(1,2_1,3_1)} \quad \beta_{(1,2_1,3_2)} \quad \beta_{(1,2_2,3_1)} \quad \beta_{(1,2_2,3_2)} \;]^T.\end{aligned}$$

Following the above arrangement of coefficients, we see that, for example, $\boldsymbol{\beta}_3 = [\beta_{(3_1)}, \beta_{(3_2)}]^T$ measures the main effect of the factor C_3, $\boldsymbol{\beta}_5 = [\beta_{(1,3_1)}, \beta_{(1,3_2)}]^T$ measures the two-factor interaction effect of C_1 and C_3, and $\boldsymbol{\beta}_7 = [\beta_{(1,2_1,3_1)}, \beta_{(1,2_1,3_2)}, \beta_{(1,2_2,3_1)}, \beta_{(1,2_2,3_2)}]^T$ measures the three-factor interaction effect of C_1, C_2, and C_3. Obviously, the ith main or interaction effect, the design submatrix $\boldsymbol{X}_i$, and the coefficient subvector $\boldsymbol{\beta}_i$ $(i = 1, \ldots, 2^p - 1)$ are related. Therefore, the probability vector is essentially determined by the magnitudes of these coefficient subvectors.

The log-linear model (2) is at the effect level, but it can be rewritten equivalently at

the coefficient level as

$$\ln \mathbf{p} = \mathbf{1}\beta_0 + \sum_{i=1}^{h-1} \mathbf{x}_i \beta_i, \tag{3}$$

where $\mathbf{x}_i$ is the ith column vector of the matrix $\boldsymbol{X}$, and β_i as a scalar is its corresponding coefficient. For instance, in the above three-way contingency table of size $2 \times 3 \times 3$, $\beta_1 = \beta_{(1)}$, $\beta_9 = \beta_{(1,3_2)}$, and $\beta_{17} = \beta_{(1,2_2,3_2)}$. Obviously, we see $\boldsymbol{X} = [\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{2^p-1}] = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{h-1}]$ and $\boldsymbol{\beta} = [\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_{2^p-1}^T]^T = [\beta_1, \beta_2, \ldots, \beta_{h-1}]^T$. There is also correspondence between the ith column $\mathbf{x}_i$ of $\boldsymbol{X}$ and the ith coefficient β_i $(i = 1, \ldots, h-1)$. By analogy with a linear regression model, the log-linear model is also essentially a regression model with the probabilities as the responses, the coefficients as the regressors, and the column vectors composing the design matrix.

Proposing a Phase II control chart requires that the IC process parameters be known, which requires estimation of the coefficient vector $\widetilde{\boldsymbol{\beta}}$ in the log-linear model (2) or (3) from an IC dataset. But some simple approximations allow using only the probability vector $\mathbf{p}$ in the IC state, skipping the estimation of the coefficient vector $\widetilde{\boldsymbol{\beta}}$. Once the IC dataset is known, the IC probability vector $\mathbf{p}^{(0)}$ can be obtained immediately by dividing the cell counts by the IC dataset size.

Log-Linear Directional Monitoring

In a log-linear model, the marginal distribution of one factor is mainly determined by its main effect, whereas the dependence among multiple factors is represented by their interaction effect. This provides the explanation of shifts in multivariate categorical processes. According to the one-to-one correspondence between factor effects and coefficient subvectors in a log-linear model, shifts in the marginal distribution of one factor lead to deviations of the coefficient subvector corresponding to its main effect, and shifts in the dependence among multiple factors result in deviations of the coefficient subvector reflecting their interaction effect.

The pre-specified log-linear model can be summarized as

$$\ln \mathbf{p} = \widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{\beta}} \quad \text{and} \quad \mathbf{p}^T\mathbf{1} = 1.$$

Denote this model as $F\big(\widetilde{\boldsymbol{X}};\widetilde{\boldsymbol{\beta}}\big)$. It is reasonable to assume that the jth on-line multivariate sampling observation vector $\mathbf{n}_j$, of size $h\times 1$ and subject to a multinomial distribution with a total size N, is collected over time from the change-point model

$$\mathbf{n}_j \overset{\text{i.i.d.}}{\sim} \begin{cases} F\big(\widetilde{\boldsymbol{X}};\widetilde{\boldsymbol{\beta}}^{(0)}\big), & \text{for} \quad j=1,\ldots,\tau, \\ F\big(\widetilde{\boldsymbol{X}};\widetilde{\boldsymbol{\beta}}^{(1)}\big), & \text{for} \quad j=\tau+1,\ldots, \end{cases} \tag{4}$$

where τ is the unknown change point, and $\widetilde{\boldsymbol{\beta}}^{(0)} \neq \widetilde{\boldsymbol{\beta}}^{(1)}$ are the known IC and unknown OC process coefficient vectors, respectively.

Since β_0 can be determined from $\boldsymbol{\beta}=[\beta_1,\ldots,\beta_{h-1}]^T$, the monitoring problem can be formulated as a hypothesis testing problem:

$$H_0:\ \boldsymbol{\beta}=\boldsymbol{\beta}^{(0)} \ \text{ versus } \ H_1:\ \boldsymbol{\beta}\neq\boldsymbol{\beta}^{(0)}. \tag{5}$$

A natural test for the hypothesis (5) can be constructed by using the idea of a generalized likelihood ratio test (GLRT; Anderson (2003)), which incorporates all possible shifts in $\boldsymbol{\beta}^{(0)}$, and thus is general and robust.

There is correspondence between the ith main or interaction effect and the coefficient subvector $\boldsymbol{\beta}_i$. In practical applications, it is usually reasonable to assume that any changes involve only a few coefficient subvectors, or only a few coefficients in the appropriate model. Suppose, however, that we have some *a priori* knowledge that in the OC state only the coefficient β_i $(1\leq i\leq h-1)$ is incremented by an unknown constant δ_i. The hypothesis then becomes

$$H_0:\ \boldsymbol{\beta}=\boldsymbol{\beta}^{(0)} \ \text{ versus } \ H_1:\ \boldsymbol{\beta}=\boldsymbol{\beta}^{(0)}+\mathbf{d}_i\delta_i,$$

where $\mathbf{d}_i$ is the direction vector of size $(h-1)\times 1$ with 1 at its ith component and 0 elsewhere. Note that here the direction is in terms of the coefficient vector $\boldsymbol{\beta}$, which stands for the index i of the coefficient β_i that shifts. This is different from the multivariate normal distribution, say $\mathrm{N}(\boldsymbol{\mu},\boldsymbol{\Omega})$, where the direction is in terms of the multivariate mean vector $\boldsymbol{\mu}$, and it represents the index i of the shifted mean μ_i.

Next consider the more practical case that in the OC state only one coefficient changes, but its location is unknown. The original alternative in the hypothesis (5) reduces to

$$H_1:\ \boldsymbol{\beta}=\boldsymbol{\beta}^{(0)}+\mathbf{d}_1\delta_1 \text{ or } \ \boldsymbol{\beta}=\boldsymbol{\beta}^{(0)}+\mathbf{d}_2\delta_2\ldots \text{ or } \ \boldsymbol{\beta}=\boldsymbol{\beta}^{(0)}+\mathbf{d}_{h-1}\delta_{h-1}, \tag{6}$$

where δ_i $(i = 1, 2, \ldots, h-1)$ are the unknown shift magnitudes, and the possible shift directions $\mathbf{d}_1$, $\mathbf{d}_2$, $\ldots$, $\mathbf{d}_{h-1}$ are defined in a similar way, applying to β_1, β_2, $\ldots$, β_{h-1}, respectively. The GLRT derived from the hypothesis (6) should be better than that from the hypothesis (5) because it makes full use of more constructive information about potential shift directions. In fact, a GLRT based on the hypothesis (6) that incorporates directional knowledge about potential changes is very much like the GLRTs used in multistage process monitoring and diagnosis which exploit the information of shift directions from the first stage to the last one (see Zou and Tsung (2008) and Zou et al. (2008)).

The hypothesis (6) may be further generalized. Since deviations involving fewer factors may appear more frequently, it is reasonable to believe that in applications most shifts involve lower-order effects rather than higher-order ones. Unlike the hypothesis (6), which considers all one-coefficient-shifts from the main factor effects up to the highest p-factor interaction effect, instead we may focus on effects involving the first few, say q, orders. Denote g as the number of coefficients corresponding to effects of the first q orders. Then the hypothesis (6) can be further extended as

$$H_1 : \ \boldsymbol{\beta} = \boldsymbol{\beta}^{(0)} + \mathbf{d}_1\delta_1 \text{ or } \ \boldsymbol{\beta} = \boldsymbol{\beta}^{(0)} + \mathbf{d}_2\delta_2 \ldots \text{ or } \ \boldsymbol{\beta} = \boldsymbol{\beta}^{(0)} + \mathbf{d}_g\delta_g, \tag{7}$$

With $q = p$, the hypothesis (7) becomes the hypothesis (6). Take the case that all p factors have two levels to illustrate the advantage of considering effects of the first two orders instead of all coefficients. With the first two orders, only $g = \mathrm{C}_p^1 + \mathrm{C}_p^2 = p(p+1)/2$ one-coefficient-shifts are taken into account. However, this number becomes $2^p - 1$ for effects of all p orders, increasing exponentially with p. Since shifts to OC states usually involve low-order effects, it is not difficult to understand that the larger q is, the less powerful the GLRT will be. This follows because the alternative hypothesis will always take into consideration undesired high-order shift directions. If the real shift indeed arises from effects of the first q orders, the GLRT based on the hypothesis (7) should certainly be powerful. Even if a shift occurs in an effect of an order higher than q, however, this change may be reflected to a large extent on the charting statistic indexed by the first q orders. So a GLRT derived from the hypothesis (7) may still be powerful, as we will show later.

Now we give the GLRT statistic for testing the hypothesis (7). For simplicity, we first

define the following quadratic form

$$D(\mathbf{y}, \mathbf{p}, \mathbf{x}, \mathbf{\Sigma}) = \frac{1}{N}\big(\mathbf{y} - N\mathbf{p}\big)^T \mathbf{x}\big(\mathbf{x}^T\mathbf{\Sigma}\mathbf{x}\big)^{-1}\mathbf{x}^T\big(\mathbf{y} - N\mathbf{p}\big)$$

for $\mathbf{y}, \mathbf{p}, \mathbf{x}$ of size $h \times 1$ and $\mathbf{\Sigma}$ of size $h \times h$. With the Phase II sample size N and the observation vector $\mathbf{n}$ of size $h \times 1$ satisfying $\mathbf{n}^T\mathbf{1} = N$, the -2LRT statistic for testing the hypothesis (7) is $Q = \max_{i\in\{1,\ldots,g\}} D(\mathbf{n}, \mathbf{p}^{(0)}, \mathbf{x}_i, \mathbf{\Sigma}^{(0)})$. Its derivation can be found in Appendix C in the supplemental file. Here $\mathbf{p}^{(0)}$ is the IC probability vector, and $\mathbf{\Sigma}^{(0)} = \mathrm{diag}\big(\mathbf{p}^{(0)}\big) - \mathbf{p}^{(0)}(\mathbf{p}^{(0)})^T$ is the IC covariance matrix (see Agresti (2002)), where $\mathrm{diag}(\mathbf{a})$ is the diagonal square matrix with its diagonal elements as the column vector $\mathbf{a}$. Note that g is the number of coefficients corresponding to effects of the first q orders in the log-linear model. The test will reject the null hypothesis if Q is larger than a pre-specified value.

The test statistic Q can be used to construct a Shewhart-type control chart for on-line monitoring. However, this method merely utilizes the information in the current sample, and it will be very inefficient for moderate and small changes. To be more efficient, we combine the hypothesis (7) with an EWMA scheme to properly exploit the information in past and current samples. Denote the observation vector of the jth sample of size N in Phase II by $\mathbf{n}_j$. For any time point k, consider the following recursive exponentially weighted sum of the previous observation vectors $\mathbf{n}_j$ $(j = 1, \ldots, k)$

$$\mathbf{z}_k = (1 - \lambda)\mathbf{z}_{k-1} + \lambda\mathbf{n}_k,$$

where λ is a smoothing parameter. We see that $\mathbf{z}_k$ actually makes full use of all available samples up to the current time point k, and that different samples are weighted as in an EWMA chart (i.e., the more recent samples have larger weights, and the weights decay exponentially over time). Thus the EWMA counterpart of the GLRT statistic Q is

$$R_k = \max_{i\in\{1,\ldots,g\}} D(\mathbf{z}_k, \mathbf{p}^{(0)}, \mathbf{x}_i, \mathbf{\Sigma}^{(0)}). \tag{8}$$

Here R_k can be regarded as the maximum of a series of quadratic D-forms $D(\cdot)$ indexed by the set $\{1, \ldots, g\}$, and it takes into account all the combined directions of potential shifts as well as the exponentially weighted past and current on-line samples. Analogous to Q, a large value of R_k will reject the null hypothesis in (7). Therefore, R_k can be the charting

statistic for a control chart, which will trigger an OC signal if $R_k > L$ for some $k \geq 1$, where $L > 0$ is a control limit chosen to attain a specific IC ARL. Note that by equipping the EWMA scheme, the steady-state covariance $\boldsymbol{\Sigma}^{(0)}$ should be changed into $\frac{\lambda}{2-\lambda}\boldsymbol{\Sigma}^{(0)}$, and correspondingly the charting statistic R_k should be modified as $R'_k = \frac{2-\lambda}{\lambda}R_k$. However, since R_k and R'_k differs only in a constant, and the control limit L will be selected by simulation, this does not have an effect. We still take R_k as the charting statistic.

Since such chart combines the one-coefficient-shifts, it represents the directional changes of multivariate categorical processes, and we call it a log-linear directional (LLD) control chart hereafter. The charting statistic for an LLD chart requires no estimate of the IC log-linear model coefficient vector $\widetilde{\boldsymbol{\beta}}^{(0)}$. Only the IC probability vector $\mathbf{p}^{(0)}$ is needed, which is easy to estimate. In addition, from the derivation of the charting statistic R_k of the LLD chart, we see that it can also be modified as a robust one for detecting other types of shifts, such as general shifts arising in all coefficients in the log-linear model. With only one factor, this robust chart is actually the same as the generalized p-chart proposed by Marcucci (1985) for univariate multinomial processes.

Diagnostic Schemes

Assume that at most one coefficient $\beta_i^{(0)}$ in $\boldsymbol{\beta}^{(0)}$ will change. This can be detected efficiently by the proposed LLD chart. Once the LLD chart triggers an OC alarm, the question naturally arises which coefficient shift is responsible for this signal. This diagnosis is especially important for multivariate categorical processes, identifying the shift direction quickly and accurately.

Diagnosis is required only when there is an OC indication. Here the null hypothesis in (7) has been rejected, and the diagnosis relies on the alternative hypothesis. This alternative hypothesis considers the one-coefficient-shifts of the first q orders ($1 \leq q \leq p$), and the LLD chart monitors the process by checking the charting statistic (8) indexed by g or q, but it is still possible that the real shift occurs in an effect of an order higher than q. In other words, a coefficient $\beta_i^{(0)}$ ($i \notin \{1, \ldots, g\}$) may deviate. Diagnosis attempts to find the actual fault location, unlike monitoring, which only reports whether the process remains IC or not. In

diagnosis, therefore, it is necessary to choose a candidate set of diagnostic shift direction indexes larger than that in monitoring, lest the real shift direction is left out. Thus, we consider one-coefficient-shifts in effects of the first q' orders ($1 \leq q \leq q' \leq p$) for identifying the shift direction. Let g' be the number of coefficients corresponding to effects of the first q' orders. Note that q' should be at least as large as q, and therefore g' is also equal to or larger than g. As with q, the choice of q' will be described later.

Note that in the OC state, the true probability vector and the corresponding covariance matrix are no longer $\mathbf{p}^{(0)}$ and $\mathbf{\Sigma}^{(0)}$. Suppose that it is only the deviation of the coefficient $\beta_i^{(0)}$ in $\boldsymbol{\beta}^{(0)}$ that brings the process into OC. Since the change-point model (4) is assumed to represent the on-line samples, the samples in the OC state are therefore acquired from $F\big(\widetilde{\boldsymbol{X}}; \widetilde{\boldsymbol{\beta}}^{(1,i)}\big)$, where $\widetilde{\boldsymbol{\beta}}^{(1,i)}$ is the OC coefficient vector that deviates from the IC coefficient vector $\widetilde{\boldsymbol{\beta}}^{(0)}$ only in β_i. This OC model is indexed by $i \in \{1, \ldots, g'\}$. Denote also the OC probability vector by $\mathbf{p}^{(1,i)}$ and the corresponding OC covariance matrix by $\mathbf{\Sigma}^{(1,i)} = \mathrm{diag}\big(\mathbf{p}^{(1,i)}\big) - \mathbf{p}^{(1,i)}\big(\mathbf{p}^{(1,i)}\big)^T$.

To recognize the shift direction, one intuition is that some D-forms $D(\cdot)$ indexed by the set $\{1, \ldots, g'\}$, which are similar to those in Equation (8) for monitoring, may help with the identification. Since diagnosis is performed in the OC state, the covariance matrix term contained in these D-forms should be $\mathbf{\Sigma}^{(1,i)}$ instead of $\mathbf{\Sigma}^{(0)}$. If among these forms, the ith one has some kind of property, some one-to-one correspondence may be built between the ith D-form and the current i-indexed OC state model $F\big(\widetilde{\boldsymbol{X}}; \widetilde{\boldsymbol{\beta}}^{(1,i)}\big)$. The diagnosis can then exploit this relationship. Suppose that $\mathbf{n}$ is an observation vector of size N collected from $F\big(\widetilde{\boldsymbol{X}}; \widetilde{\boldsymbol{\beta}}^{(1,i)}\big)$. Define $S_j = D(\mathbf{n}, \mathbf{p}^{(0)}, \mathbf{x}_j, \mathbf{\Sigma}^{(1,i)})$. It is shown in Appendix D in the supplemental file that the $\arg\max_{j \in \{1,\ldots,g'\}} S_j$ will be consistent with the real shift direction index i. So there is indeed a probable correspondence between the largest D-form S_i ($i \in \{1, \ldots, g'\}$) and the real shift direction $\mathbf{d}_i$, and this provides some guidelines for diagnosis.

In applications, the OC probability vector $\mathbf{p}^{(1,i)}$ and the covariance matrix $\mathbf{\Sigma}^{(1,i)}$ are unknown and need to be estimated. Suppose that the LLD chart triggers an alarm at time η. From the change point τ (but not including) to the signal point η, there have already been observation vectors collected from $F\big(\widetilde{\boldsymbol{X}}; \widetilde{\boldsymbol{\beta}}^{(1,i)}\big)$. The best estimation of $\mathbf{p}^{(1,i)}$ seems to

be the average of these observation vectors divided by the sample size N. This, however, relates to estimating the change point τ, which will not be discussed here. Therefore, the exponentially weighted sum of observation vectors $\mathbf{z}_\eta$ at η, which efficiently exploits the information in the OC state, seems to be a reasonable choice for estimating $\mathbf{p}^{(1,i)}$. To this end, the diagnostic procedures can be finalized as

Step 1 Define $\hat{\mathbf{p}} = \mathbf{z}_\eta/N$ and $\widehat{\boldsymbol{\Sigma}} = \mathrm{diag}\big(\hat{\mathbf{p}}\big) - \hat{\mathbf{p}}\hat{\mathbf{p}}^T$ as the estimates of $\mathbf{p}^{(1,i)}$ and $\boldsymbol{\Sigma}^{(1,i)}$, respectively.

Step 2 Replace $\boldsymbol{\Sigma}^{(1,i)}$ with $\widehat{\boldsymbol{\Sigma}}$ and $\mathbf{n}$ with $\mathbf{z}_\eta$ in the D-forms S_j $(j = 1, \ldots, g')$, calculate their values, and denote them as s_j $(j = 1, \ldots, g')$.

Step 3 Find the maximum among these s_j, and its index will be the recognized direction index ζ.

Practical Implementation and Application

Design Parameter Settings

On the parameters q and q': The most possible shift directions are confined in effects of the first q orders $(1 \leq q \leq p)$ when there are p factors. The GLRT (8) will be less powerful if more shift directions are included if indeed most shifts occur in lower-order effects. If q is selected as 1, however, only main effect deviations are considered, ignoring correlations between factors, which is obviously inappropriate. In most real applications, one cares about only means and variances, moments of the first two orders. Furthermore, the monitoring task is to only answer "yes" or "no" about whether the process is IC, rather than specifying shifted effects. Here we should pay attention only to the main effects and two-factor interactions and choose q as 2, focusing on coefficients corresponding to effects of the first two orders. If the shift indeed arises in a main or two-factor interaction effect, the GLRT (8) with $q = 2$ should definitely be powerful. Even if a shift arises in a three-factor or higher-order interaction effect, the effects of the first two orders will be influenced to a fairly large extent, so the OC may still be detected quickly by the LLD chart.

The choice of the diagnostic parameter q' should be considered in combination with the monitoring parameter q. As has been indicated earlier, q' should be at least as large as q with $q' \geq q$, for fear that important shift directions might be missed. Since $q = 2$ has been recommended, should q' be at least 3, confining the candidate subset of diagnostic shift directions to effects in the first three orders. This will still be safe in case a shift appears in a three-factor interaction effect, and it is believed that shifts in four-factor or even higher-order interaction effects are rare. As in monitoring, as q' increases, the diagnostic accuracy will decrease. Therefore, $q' = 3$ is an ideal choice.

In fact, by going one step further, we can choose these index sets for monitoring and diagnosis to contain only the factor effects or coefficients we are interested in. Generally, any subset of the index set $\{1, \ldots, h\}$ can be introduced, with its elements corresponding to the coefficients in which a shift may occur the most likely. The only thing that must be kept in mind is that the index set for diagnosis must be at least as large as the one for monitoring.

On the smoothing parameter λ: Generally, a smaller λ assists in detecting smaller shifts more quickly, while a larger λ leads to quicker detection of larger shifts (see Lucas and Saccucci (1990) and Lowry et al. (1992)). This still applies to LLD charts, as will be confirmed by Figure 2 in the next section. In diagnosis, a smaller λ results in better diagnostic consistency for smaller shifts, while a larger λ helps in identifying the locations of larger shifts. This will also be illustrated in Figure 3 below. Empirical results show that λ between 0.05 and 0.2 is probably a reasonable choice.

On the sample size N: With a small sample size N, to achieve a small OC ARL, a very large magnitude δ is required. This can be seen by comparing Table 1 below and Table A1 in the supplemental file. On the other hand, a large sample size contains more information about observation vectors such as $\mathbf{n}_k$. To detect a small shift as quickly as possible, a very large sample size is needed. Therefore, based on permissible costs or other considerations, we should select a sample size N as large as possible.

In fact, the sample size for multivariate categorical processes cannot be compared with that of multivariate continuous processes, which is relatively small, say, only several or

decades. This follows because the data in multivariate categorical processes are actually the proportions in each cell of a multi-way contingency table, which are acquired by aggregating multivariate categorical observations. Therefore, multivariate categorical processes are different in nature from multivariate continuous ones. Since there is no need of accurate measurements for collecting categorical data, it usually does not cost as much. So a relatively large sample size in multivariate categorical processes is acceptable.

In this paper, for simplicity we assume that the sample size N is fixed. For time-varying sample sizes, we will deal with it in future work. In short, if the distribution of sample sizes can be specified or their trend can be predicted, we can calculate the control limit by simulation before monitoring. Otherwise, it seems that a time-varying control limit that depends on the on-line sample size N_k is required. We can determine the control limit for each sample point k with Monte Carlo simulation by drawing large samples from the multinomial distribution $\mathrm{MN}(N_k; \mathbf{n}_k/N_k)$, making the conditional probability (given there is no alarm up to time point $k-1$) attain $1/\mathrm{ARL}_0$.

On computation: Once the IC probability vector $\mathbf{p}^{(0)}$ has been chosen, the charting statistic R_k of the LLD chart is very simple to compute, entailing only basic arithmetic. Finding the control limit L for a given IC ARL (denoted ARL_0) is trivial. Using a Pentium 3.0GHz CPU the search procedure based on 10,000 replicated simulations takes only a few minutes by bisection search when $\mathrm{ARL}_0 = 370$, $h = 32$, $\lambda = 0.1$, and $N = 1,000$.

Implementation — a Practical Example

Implementation of the proposed methodology is demonstrated in this subsection by revisiting the AEC manufacturing process as the example. This is a multivariate binomial process with three factors LC, DF and CAP, each with two levels. A three-way contingency table can represent the cross-classifications of the three factors in eight cells. The relationship between the cell counts and the factor level combinations can be well characterized by a log-linear model. An LLD chart can be used to monitor the three quality characteristics simultaneously, and the suggested diagnostic method can identify the fault location whenever there is an OC alarm.

Step 1: *Define the multivariate categorical process and derive the design matrix* $\widetilde{\boldsymbol{X}}$. This AEC manufacturing process has $p = 3$ factors each with 2 levels. Denote the three factors LC, DF, and CAP as C_1, C_2, and C_3, respectively. The matrix $\widetilde{\boldsymbol{X}}$ can be formulated in the manner of performing a 2^3 full factorial experiment, and then arranged in the sequence of 1, C_1, C_2, C_3, C_1C_2, ..., $C_1C_2C_3$.

Step 2: *Obtain the Phase I IC probability vector* $\mathbf{p}^{(0)}$. Every day thousands of AECs pass through each workbench in the aging stage, and their LC, DF, and CAP is inspected and classified as conforming or nonconforming. The IC probability vector $\mathbf{p}^{(0)}$ is calculated using an IC dataset of about sixty thousand observations (given in another supplemental file), where the eight level combination counts are summarized as [9, 6, 65, 43, 8, 259, 1830, 61038]. So $\mathbf{p}^{(0)}$ is obtained from the proportions of the eight cell counts in this IC dataset

$$\mathbf{p}^{(0)} = [1.423, 0.9485, 10.28, 6.798, 1.265, 40.94, 289.3, 9649]^T \times 10^{-4}.$$

The IC covariance matrix is then $\boldsymbol{\Sigma}^{(0)} = \mathrm{diag}\big(\mathbf{p}^{(0)}\big) - \mathbf{p}^{(0)}(\mathbf{p}^{(0)})^T$.

Step 3: *Find the control limit* L *for a given* ARL_0. Usually, an ARL_0 of 370 is used in the AEC process. The LLD chart monitors the process by focusing on shifts in the main effects and two-factor interaction effects ($q = 2$ and $g = 6$). An EWMA smoothing parameter $\lambda = 0.1$ and a Phase II sample size $N = 500$ are considered appropriate. The control limit L is set to 0.56 through a bisection search based on 10,000 simulations, achieving $\mathrm{ARL}_0 = 370$.

Step 4: *Phase II monitoring.* The cell counts of the eight level combinations are tabulated for each sample of 500. For the kth sample, the eight cell counts are recorded in the observation vector $\mathbf{n}_k$, and based on the observation vectors up to time k, the exponentially weighted sum $\mathbf{z}_k$ is calculated with the smoothing parameter 0.1. The charting statistic R_k of $\mathbf{z}_k$ is then calculated using Equation (8). As the process proceeds, the charting statistics R_k ($k = 1, 2, \ldots$) are plotted in a control chart and compared with the control limit L. A typical plot is shown in Figure 1. In the example the chart signals OC at the 35th sample and remains above the control limit for the remainder of the sequence.

Step 5: *Identify the shift direction once an OC signal is triggered.* We follow the three

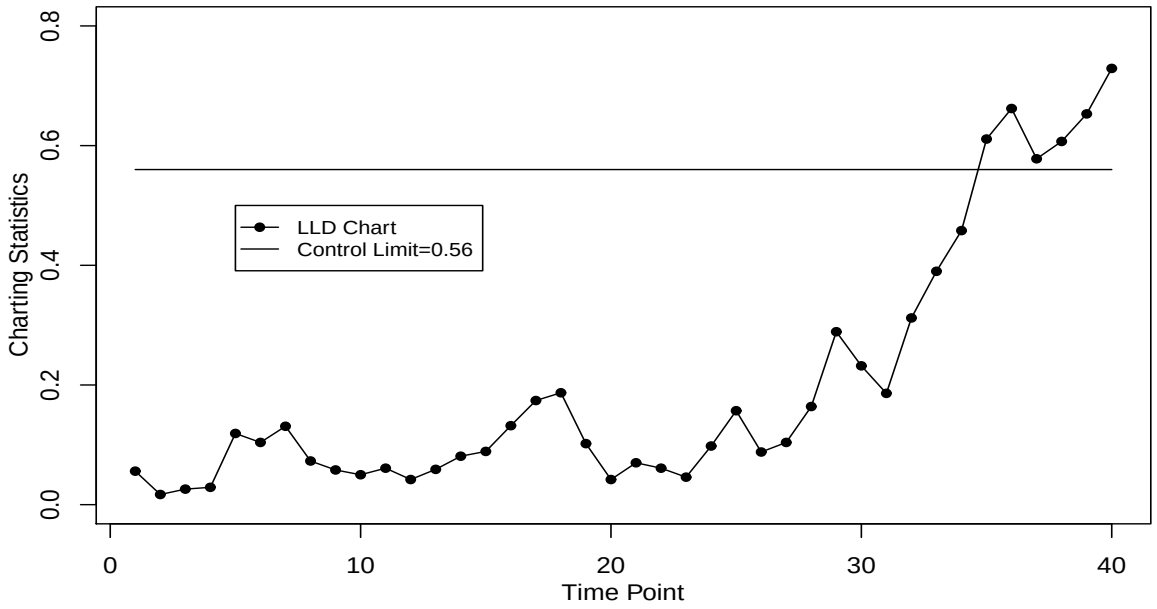

Figure 1: An LLD chart monitoring the AEC process

diagnostic steps stated previously. Based on $\mathbf{z}_{35}$, the OC probability vector is estimated as

$$\hat{\mathbf{p}} = \frac{\mathbf{z}_{35}}{N} = [1.253, 0.2422, 7.838, 1.967, 0.2236, 22.41, 314.9, 9651]^T \times 10^{-4},$$

so the OC covariance matrix may be estimated as $\widehat{\mathbf{\Sigma}} = \text{diag}(\hat{\mathbf{p}}) - \hat{\mathbf{p}}\hat{\mathbf{p}}^T$. As recommended before, the candidate subset of the fault directions is selected as coefficients in the effects of the first three orders with $q' = 3$ and $g' = 7$. The direction estimator chooses the maximum among the values of the seven D-forms, which are 0.29, 0.87, 0.08, 1.11, 0.06, 0.00, and 0.00 corresponding to the seven coefficients $\beta_{(1)}$, $\beta_{(2)}$, $\beta_{(3)}$, $\beta_{(1,2)}$, $\beta_{(1,3)}$, $\beta_{(2,3)}$, and $\beta_{(1,2,3)}$, respectively. Apparently, there has been a shift associated with coefficient $\beta_{(1,2)}$, namely in the interaction effect of the two factors C_1 and C_2, i.e., the interaction of LC and DF.

Performance Comparison and Assessment

Let us now test the performance of the proposed LLD chart and the diagnostic approach through Monte Carlo simulations. Both multivariate binomial processes and multivariate multinomial processes are simulated and the performance of the LLD chart is compared

with competing techniques. Throughout the simulations, ARL_0 is fixed at 370, and all ARL values are averages of 10,000 replicated simulations.

Monitoring a Multivariate Binomial Process

With a multivariate binomial process, the LLD chart can be compared with the χ^2-chart proposed by Patel (1973), which has been introduced as a typical control chart for monitoring multivariate binomial data. Totally ignoring past information, the Shewhart-type χ^2-chart is inadequate for detecting moderate and small shifts. For fair comparison, we extend the χ^2-chart to its EWMA version by replacing the observation vectors in its charting statistic with their exponentially weighted sum vectors. We refer to this chart as the multivariate binomial EWMA (MBE) control chart.

Assume that during a production process five quality characteristics are each assessed as conforming or nonconforming, yielding a multivariate binomial process, which can be arranged into a five-way contingency table with two levels for each factor. The IC log-linear model has the coefficient vector

$$\begin{aligned}\widetilde{\boldsymbol{\beta}}^{(0)} = [\; & \beta_0 \;\; 0.72 \;\; 0.93 \;\; 0.49 \;\; 0.25 \;\; 0.47 \;\; -0.57 \;\; 0.22 \\ & 0.11 \;\; -0.14 \;\; 0.15 \;\; -0.16 \;\; 0.41 \;\; 0.16 \;\; -0.19 \;\; 0.33 \\ & 0.39 \;\; 0.10 \;\; 0.07 \;\; -0.05 \;\; 0.21 \;\; -0.02 \;\; 0.45 \;\; 0.33 \\ & 0.08 \;\; 0.27 \;\; 0.04 \;\; -0.13 \;\; 0.07 \;\; -0.07 \;\; 0.03 \;\; 0.00 \;]^T,\end{aligned}$$

where β_0 is the intercept accommodating the constraint $\mathbf{p}^T\mathbf{1} = 1$. Based on $\widetilde{\boldsymbol{\beta}}^{(0)}$, the IC probability vector $\mathbf{p}^{(0)}$ can be further calculated. It is believed that possible shifts to OC appear in factor effects, and correspondingly in coefficients of the IC log-linear model. According to the design recommendation, construct the LLD chart with $q = 2$ and $g = 15$ in its charting statistic (8), again assuming that most shifts will arise in the main effects or the two-factor interaction effects.

The OC ARLs of the LLD and MBE charts for various shifts of magnitude δ are summarized in Table 1 when the EWMA smoothing parameter $\lambda = 0.1$ and the Phase II sample size $N = 1,000$. Due to space limitations, only some representative results are tabulated (Additional results are available from the authors on request). Table 1 shows

Table 1: OC ARLs for LLD and MBE control charts monitoring one-coefficient-shifts of the first two orders

δ	LLD		MBE		LLD		MBE		LLD		MBE	
	$\beta_{(3)}$				$\beta_{(5)}$				$\beta_{(1,4)}$			
0.01	201	(1.99)	199	(1.90)	241	(2.37)	222	(2.16)	176	(1.67)	249	(2.48)
0.02	71.6	(0.63)	70.5	(0.61)	95.9	(0.86)	86.1	(0.77)	53.0	(0.43)	117	(1.07)
0.05	13.2	(0.07)	13.6	(0.07)	16.3	(0.09)	16.5	(0.09)	10.3	(0.05)	21.6	(0.13)
0.20	2.82	(0.01)	2.90	(0.01)	3.22	(0.01)	3.25	(0.01)	2.39	(0.01)	3.79	(0.01)
−0.01	172	(1.69)	176	(1.69)	200	(1.94)	193	(1.88)	155	(1.50)	230	(2.24)
−0.02	60.5	(0.51)	61.5	(0.52)	79.0	(0.70)	74.3	(0.66)	46.8	(0.38)	101	(0.93)
−0.05	11.6	(0.06)	12.0	(0.06)	14.1	(0.07)	14.2	(0.08)	9.35	(0.04)	18.9	(0.11)
−0.20	2.18	(0.01)	2.23	(0.01)	2.40	(0.01)	2.42	(0.01)	1.96	(0.01)	2.96	(0.01)
	$\beta_{(2,3)}$				$\beta_{(2,5)}$				$\beta_{(3,4)}$			
0.01	193	(1.87)	247	(2.44)	262	(2.55)	318	(3.15)	216	(2.07)	290	(2.85)
0.02	66.0	(0.56)	108	(1.00)	110	(1.00)	201	(1.90)	77.0	(0.67)	158	(1.50)
0.05	12.7	(0.06)	20.2	(0.12)	18.0	(0.10)	47.3	(0.38)	13.8	(0.07)	32.2	(0.23)
0.20	2.76	(0.01)	3.69	(0.01)	3.39	(0.01)	6.14	(0.02)	2.90	(0.01)	4.79	(0.02)
−0.01	167	(1.60)	210	(2.03)	219	(2.15)	279	(2.77)	183	(1.77)	249	(2.44)
−0.02	58.0	(0.50)	93.3	(0.86)	90.1	(0.82)	165	(1.59)	66.0	(0.57)	129	(1.22)
−0.05	11.3	(0.06)	17.2	(0.10)	15.2	(0.08)	36.4	(0.28)	12.0	(0.06)	25.9	(0.18)
−0.20	2.15	(0.01)	2.75	(0.01)	2.50	(0.01)	4.16	(0.01)	2.22	(0.01)	3.48	(0.01)

NOTE: Standard errors are in parentheses. $\lambda = 0.1$. $N = 1,000$.

that the MBE chart outperforms the LLD chart only for small shifts in the main effects such as $\beta_{(3)}$ and $\beta_{(5)}$, which represent the main effects of the factors C_3 and C_5, respectively. This is partly because the MBE chart is based on collecting one-way marginal sums for each factor, whose changes are directly echoed by shifts in the main effects or coefficients of the first order. The LLD chart is developed especially for detecting one-coefficient-shifts, so this superiority of the MBE chart over the LLD chart is not very significant. For the shifts in the main effects, as the magnitudes increase, the situation is reversed, and the LLD chart triggers an OC indication faster than the MBE chart for larger shifts. To say the least, if we select $q = 1$ with its corresponding $g = 5$, the LLD chart is indeed uniformly superior over the MBE chart in detecting one-coefficient-shifts in main effects. These simulation

results are available from the authors.

When it comes to shifts in the two-factor interaction effects, including $\beta_{(1,4)}$ representing the interaction effect of the factors C_1 and C_4, $\beta_{(2,3)}$, $\beta_{(2,5)}$, and $\beta_{(3,4)}$, the LLD chart behaves consistently better than the MBE chart. Additionally, in most situations this advantage is quite substantial. The changes in the high-order interaction effects reflect fewer shifts of the one-way marginal sums, which are usually neglected by the MBE chart. However, the LLD chart is still able to detect possible changes attributable to higher-order interactions effectively.

The OC performance of the LLD and MBE charts for the same coefficients as in Table 1 but with some other λ and N settings is demonstrated in Table A1 in the supplemental file. The results of using shift magnitudes of $\delta = 0.02$ and $\delta = 0.20$ are reported. Table A1 shows patterns similar to those of Table 1 for various changes in the coefficients. With λ fixed, for the same shifts, both charts have greater power when the sample size N increases. For a fixed sample size N and the same coefficients, LLD charts with a smaller λ signal faster for smaller shifts, and those with a larger λ signal faster for larger shifts. This mimics the properties of the conventional EWMA chart (see Lucas and Saccucci (1990) and Lowry et al. (1992)). Figures 2-(a) and -(b) confirm this further. They show the OC ARL curves of the LLD chart for λ values of 0.05, 0.1, 0.2, and 0.5 when there are shifts in $\beta_{(5)}$ and $\beta_{(3,4)}$, respectively.

The LLD chart was constructed by checking the GLRT statistic in Equation (8) indexed by $q = 2$. However, as indicated earlier, shifts may indeed occur in the effects of the third or even higher order. Although such cases should be infrequent, they are considered in Table 2 to test robustness. The OC ARLs of LLD and MBE charts are compared for one-coefficient-shifts in the three-factor interactions $\beta_{(1,2,4)}$, $\beta_{(2,3,4)}$, and $\beta_{(3,4,5)}$ when $\lambda = 0.1$ and $N = 1,000$. Apparently, the advantage of the LLD chart over the MBE chart is maintained, implying that the LLD chart can still deal with higher-order interactions well.

As implied earlier, besides considering cross-classifications among factors, another noteworthy characteristic of the LLD chart is power in discovering one-coefficient-shifts, which is perhaps not robust when dealing with general changes. Therefore, some representative

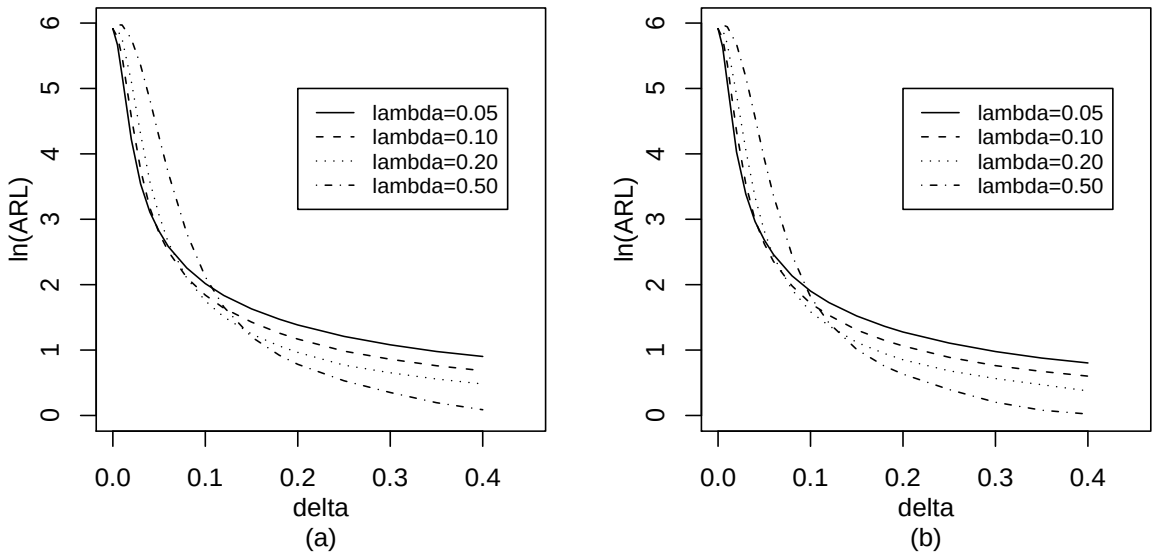

Figure 2: OC ARL curves for the LLD chart with $N = 1,000$ and various values of λ for monitoring multivariate binomial processes when there are shifts in: (a) $\beta_{(5)}$; (b) $\beta_{(3,4)}$

two-coefficient-shift cases are summarized in Table 3 to show that the LLD chart is still powerful in this situation. Table 3 with $\lambda = 0.1$ and $N = 1,000$ contains shifts in two coefficients of the first or second orders, adding magnitudes δ_1 and δ_2. According to Table 3, if the two shifts both appear in main effects of the first order, such as $(\beta_{(1)}, \beta_{(5)})$, the MBE chart performs better than the LLD chart. However, as the two shifts begin to include more effects of the second order. The LLD chart outperforms the MBE chart.

Monitoring a Multivariate Multinomial Process

With a multivariate multinomial process, the LLD can be compared with only the somewhat naive multi-chart described earlier. Similar to the χ^2-chart, the multi-chart should be equipped with the EWMA scheme by replacing the observation vectors in the charting statistic with their exponentially weighted sums. We refer to this new multi-chart as the multivariate multinomial EWMA (MME) chart.

Take a case involving factors with mixed levels for illustration. A service flow has four quality characteristics being monitored, with the first two judged as satisfactory or unsatisfactory and the other two assessed as either excellent, acceptable, or unacceptable.

Table 2: OC ARLs comparing LLD and MBE control charts for cases of one-coefficient-shifts of the third order

δ	LLD		MBE		LLD		MBE		LLD		MBE	
	$\beta_{(1,2,4)}$				$\beta_{(2,3,4)}$				$\beta_{(3,4,5)}$			
0.01	189	(1.84)	262	(2.55)	289	(2.85)	334	(3.32)	275	(2.72)	328	(3.30)
0.02	63.2	(0.54)	125	(1.17)	162	(1.59)	245	(2.44)	144	(1.35)	233	(2.26)
0.05	12.5	(0.06)	24.1	(0.15)	32.9	(0.24)	71.8	(0.61)	29.4	(0.20)	62.0	(0.52)
0.20	2.72	(0.01)	3.99	(0.01)	4.89	(0.02)	8.04	(0.03)	4.77	(0.02)	7.27	(0.03)
−0.01	160	(1.53)	232	(2.26)	251	(2.49)	298	(2.99)	236	(2.30)	290	(2.88)
−0.02	55.5	(0.47)	110	(1.03)	132	(1.25)	200	(1.95)	115	(1.05)	187	(1.81)
−0.05	11.3	(0.05)	20.6	(0.13)	26.5	(0.18)	55.8	(0.47)	24.1	(0.16)	47.8	(0.38)
−0.20	2.22	(0.01)	3.14	(0.01)	3.50	(0.01)	5.34	(0.02)	3.41	(0.01)	4.91	(0.02)

NOTE: Standard errors are in parentheses. $\lambda = 0.1$. $N = 1,000$.

This yield a four-way $2 \times 2 \times 3 \times 3$ contingency table. The IC log-linear model is described by the coefficient vector

$$\widetilde{\boldsymbol{\beta}}^{(0)} = [\begin{array}{rrrrrrrrrr} \beta_0 & 0.73 & 0.72 & 0.70 & 0.12 & 0.71 & 0.10 & 0.17 & 0.12 & \\ -0.15 & 0.19 & -0.14 & 0.23 & 0.07 & 0.16 & -0.14 & 0.23 & -0.30 & \\ -0.17 & 0.14 & 0.10 & 0.06 & 0.09 & -0.12 & 0.19 & -0.15 & 0.11 & \\ 0.22 & 0.24 & 0.24 & -0.08 & -0.16 & 0.07 & -0.11 & 0.05 & 0.03 &]^T, \end{array}$$

where β_0 is the intercept accommodating the constraint $\mathbf{p}^T\mathbf{1} = 1$. Since the service flow process has four factors, the MME chart here is composed of four individual charts. Their IC ARLs are selected by simulation in order to be identical and jointly achieve an overall ARL_0 370.

Before comparison, we should note that in a log-linear model all the elements of a coefficient subvector as a whole represent the corresponding main or interaction effect, and that they should change simultaneously if there is a shift in this effect. In other words, a shift should occur at the effect or coefficient subvector level, instead of the coefficient level. However, here in the simulations about multivariate multinomial processes, for convenience we still consider shifts in only one coefficient. They may be a special type of shifts at the effect level, with the assumption that only one element in the coefficient subvector deviates whereas all the other ones remain unchanged. Even for a shift at the effect level, that

Table 3: OC ARL comparison between the LLD and MBE charts in cases of two-coefficient-shifts of the first two orders, $\lambda = 0.1$ and $N = 1,000$

δ_1	δ_2	LLD		MBE		LLD		MBE	
		$\beta_{(1)} + \delta_1$		$\beta_{(5)} + \delta_2$		$\beta_{(2)} + \delta_1$		$\beta_{(1,3)} + \delta_2$	
0.02	0.02	31.9	(0.23)	31.8	(0.22)	35.1	(0.25)	45.9	(0.36)
0.02	−0.02	43.4	(0.33)	31.3	(0.22)	66.9	(0.57)	81.3	(0.74)
−0.02	0.02	43.1	(0.33)	31.1	(0.22)	70.1	(0.60)	80.2	(0.72)
−0.02	−0.02	29.1	(0.21)	28.4	(0.20)	30.6	(0.22)	39.2	(0.30)
0.20	0.20	2.04	(0.01)	2.01	(0.01)	2.19	(0.01)	2.46	(0.01)
0.20	−0.20	2.10	(0.01)	1.82	(0.01)	2.29	(0.01)	2.67	(0.01)
−0.20	0.20	1.95	(0.01)	1.79	(0.01)	2.74	(0.01)	2.90	(0.01)
−0.20	−0.20	1.71	(0.01)	1.62	(0.01)	1.63	(0.01)	1.76	(0.01)
		$\beta_{(1,4)} + \delta_1$		$\beta_{(4,5)} + \delta_2$		$\beta_{(1,5)} + \delta_1$		$\beta_{(2,3)} + \delta_2$	
0.02	0.02	26.9	(0.18)	45.6	(0.36)	18.7	(0.11)	25.4	(0.17)
0.02	−0.02	47.6	(0.37)	101	(0.93)	76.6	(0.67)	96.8	(0.87)
−0.02	0.02	46.8	(0.37)	100	(0.91)	71.8	(0.64)	88.6	(0.80)
−0.02	−0.02	24.4	(0.16)	40.9	(0.32)	17.0	(0.10)	23.0	(0.15)
0.20	0.20	1.93	(0.01)	2.32	(0.01)	1.69	(0.00)	1.88	(0.01)
0.20	−0.20	2.13	(0.01)	3.23	(0.01)	2.88	(0.01)	3.40	(0.01)
−0.20	0.20	2.07	(0.01)	2.91	(0.01)	2.38	(0.01)	2.73	(0.01)
−0.20	−0.20	1.55	(0.01)	1.91	(0.01)	1.22	(0.00)	1.43	(0.00)

NOTE: Standard errors are in parentheses.

all the elements of its corresponding coefficient subvector shift simultaneously can still be detected powerfully by the LLD chart, which checks individually whether each element in $\boldsymbol{\beta}_i$ shifts or not. The simulation results about this are available from the authors.

Similar to the comparison between the LLD and MBE charts, the OC ARLs when only one coefficient changes by δ are reported in Table 4 with $\lambda = 0.1$ and $N = 1,000$. When the one-coefficient-shifts comes from one of the main effects, for instance, $\beta_{(1)}$, $\beta_{(3_2)}$, or $\beta_{(4_1)}$, the MME chart shows better performance than the LLD chart in terms of giving rise to an OC alert faster. This is not surprising, as each of the four individual charts summarizes efficiently the one-way marginal sums for one of the four factors, and these are decided by

Table 4: OC ARLs of LLD and MME charts in cases of one-coefficient-shifts of the first two orders

δ	LLD		MME		LLD		MME		LLD		MME	
	$\beta_{(1)}$				$\beta_{(3_2)}$				$\beta_{(4_1)}$			
0.02	128	(1.20)	88.2	(0.80)	145	(1.36)	148	(1.39)	110	(0.99)	104	(0.95)
0.05	19.8	(0.11)	15.8	(0.09)	24.5	(0.16)	24.4	(0.16)	17.6	(0.10)	17.2	(0.09)
0.20	3.55	(0.01)	3.17	(0.01)	3.54	(0.01)	3.50	(0.01)	3.27	(0.01)	3.20	(0.01)
−0.02	103	(0.94)	72.7	(0.66)	153	(1.45)	150	(1.43)	92.4	(0.85)	86.5	(0.77)
−0.05	16.5	(0.09)	13.6	(0.07)	25.5	(0.17)	24.6	(0.16)	15.8	(0.09)	15.3	(0.08)
−0.20	2.61	(0.01)	2.35	(0.01)	3.69	(0.01)	3.63	(0.01)	2.64	(0.01)	2.58	(0.01)
	$\beta_{(1,2)}$				$\beta_{(1,3_2)}$				$\beta_{(2,3_1)}$			
0.02	86.3	(0.78)	105	(0.96)	149	(1.41)	224	(2.18)	102	(0.92)	140	(1.25)
0.05	14.3	(0.07)	21.1	(0.13)	24.5	(0.16)	51.7	(0.42)	16.6	(0.09)	26.4	(0.18)
0.20	2.95	(0.01)	3.84	(0.01)	3.51	(0.01)	5.15	(0.02)	3.10	(0.01)	4.11	(0.01)
−0.02	71.7	(0.63)	85.7	(0.77)	148	(1.39)	217	(2.13)	86.7	(0.79)	115	(1.10)
−0.05	12.6	(0.06)	17.1	(0.10)	24.9	(0.16)	53.1	(0.43)	14.9	(0.08)	22.8	(0.15)
−0.20	2.26	(0.01)	2.80	(0.01)	3.62	(0.01)	5.75	(0.02)	2.58	(0.01)	3.42	(0.01)
	$\beta_{(2,4_2)}$				$\beta_{(3_1,4_1)}$				$\beta_{(3_2,4_2)}$			
0.02	165	(1.58)	245	(2.43)	94.6	(0.85)	136	(1.25)	259	(2.54)	338	(3.33)
0.05	27.2	(0.18)	63.3	(0.53)	15.7	(0.08)	25.0	(0.17)	64.7	(0.54)	235	(2.25)
0.20	3.71	(0.01)	5.89	(0.02)	2.98	(0.01)	3.94	(0.01)	5.79	(0.02)	23.7	(0.15)
−0.02	165	(1.61)	235	(2.32)	82.2	(0.75)	114	(1.05)	261	(2.55)	332	(3.31)
−0.05	27.8	(0.19)	62.9	(0.54)	14.5	(0.08)	22.1	(0.14)	68.3	(0.57)	244	(2.45)
−0.20	3.81	(0.01)	6.28	(0.02)	2.59	(0.01)	3.44	(0.01)	5.96	(0.02)	31.1	(0.22)

NOTE: Standard errors are in parentheses. $\lambda = 0.1$. $N = 1,000$.

the coefficients of the first order. When two-factor interaction effects such as $\beta_{(1,2)}$, $\beta_{(1,3_2)}$, $\beta_{(2,3_1)}$, $\beta_{(2,4_2)}$, $\beta_{(3_1,4_1)}$, and $\beta_{(3_2,4_2)}$ are the focus, the superiority of the LLD chart over the MME chart becomes clear. The advantage of the LLD chart lies partly in signalling shifts in higher-order interaction effects and partly in one-coefficient-shifts.

Similar results with other parameter settings for λ and N are presented in Table A2 in the supplemental file. These data exhibit the same trends as in Table 4. The effects of the smoothing parameter λ and the Phase II sample size N on the LLD chart resemble those for the multivariate binomial data in Table A1.

Diagnostic Performance Analysis

The performance of our suggested diagnostic method using the proposed diagnostic procedures is reported in this subsection. In the simulations, the change point τ was set at 50, and 10,000 independent series were generated. Any series for which an OC indication was triggered before time point $\tau + 1$ was discarded. An IC log-linear model was used with the same coefficient vector as in the subsection on monitoring multivariate binomial processes, dealing with five factors, each with two levels.

According to the design recommendation of the diagnostic method, the candidate subset for potential shift directions was selected as coefficients in effects of the first three orders with $q' = 3$ and $g' = 25$, with the assumption that no effect of an order larger than three would deviate. Therefore, in the simulations the real one-coefficient-shifts considered were confined to the first three orders, in particular the main effects $\beta_{(2)}$ and $\beta_{(4)}$, the two-factor interaction effects $\beta_{(1,3)}$, $\beta_{(1,5)}$, $\beta_{(2,3)}$, and $\beta_{(4,5)}$, as well as the three-factor interaction effects $\beta_{(1,4,5)}$ and $\beta_{(2,3,5)}$. The diagnostic performance was investigated for various shift magnitudes δ and different combinations of the EWMA smoothing parameter λ and the sample size N. Table 5 lists the results in terms of the observed matching probability $P(\hat{\zeta} = \zeta)$, where $\hat{\zeta} \in \{1, \ldots, g'\}$ is the estimated shift direction index and $\zeta \in \{1, \ldots, g'\}$ is the real one-coefficient-shift direction index when $\lambda = 0.1$ and $N = 1,000$. For instance, in the case of $\lambda = 0.1$ and $N = 1,000$, when only the coefficient $\beta_{(2)}$ shifted by $\delta = 0.02$, the diagnostic scheme provided a correct prescription in 48% of the simulations, and this percentage rose to 84% if the magnitude δ was 0.05.

The diagnostic performance with other values of λ and N is listed in Table A3 in the supplemental file. At least three conclusions can be drawn with respect to the effects of the shift magnitudes δ, the EWMA smoothing parameter λ, and the sample size N. First, for one coefficient of the first three orders with fixed λ and N, as the magnitude of δ increases, the diagnostic consistency improves progressively with the observed matching probability approaching 1, as might be expected. Second, while keeping λ and δ unchanged, a larger sample size N will improve diagnostic performance, which is also to be expected. Third, a larger λ will be beneficial for identifying larger shifts while a smaller one can assist in

recognizing smaller deviations. Figure 3 further demonstrates this pattern.

Table 5: Observed matching probability in cases of one-coefficient-shifts of the first three orders

Case	δ	$\beta_{(2)}$	$\beta_{(4)}$	$\beta_{(1,3)}$	$\beta_{(1,5)}$	$\beta_{(2,3)}$	$\beta_{(4,5)}$	$\beta_{(1,4,5)}$	$\beta_{(2,3,5)}$
	0.02	0.48	0.46	0.52	0.43	0.46	0.50	0.23	0.18
$\lambda = 0.1$	0.05	0.84	0.71	0.73	0.62	0.73	0.73	0.71	0.64
	0.20	0.95	0.84	0.84	0.77	0.86	0.86	0.87	0.85
	−0.02	0.51	0.46	0.53	0.43	0.46	0.50	0.24	0.20
$N = 1,000$	−0.05	0.82	0.69	0.71	0.61	0.71	0.71	0.71	0.63
	−0.20	0.94	0.85	0.84	0.79	0.86	0.86	0.87	0.85

NOTE: $\lambda = 0.1$. $N = 1,000$.

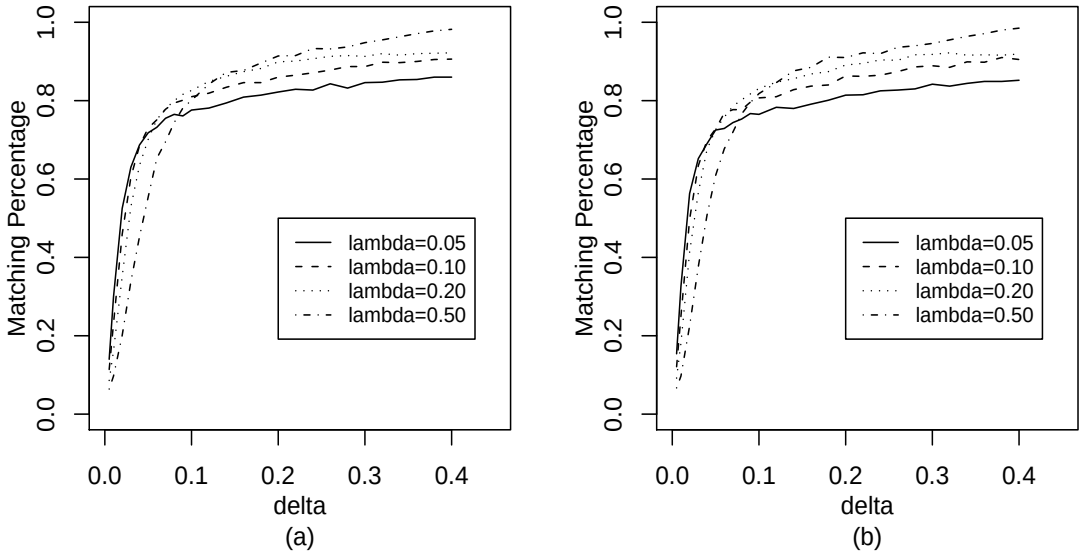

Figure 3: Matching probability curves for the diagnostic scheme with $N = 1,000$ and various values of λ for identifying the real shifts in: (a) $\beta_{(2,3)}$; (b) $\beta_{(4,5)}$.

Conclusion

This study has developed a new log-linear directional control chart for monitoring process shifts with high efficiency based on the integration of the log-likelihood of log-linear models and the EWMA scheme. As well, a post-signal diagnostic scheme for recognizing the

shift direction was formulated. The LLD chart was shown to work well with multivariate categorical processes, including multivariate binomial and multivariate multinomial distributions, and to incorporate factor interactions among the multiple categorical variables successfully. Such LLD charts are easy to construct and to implement in practice. Furthermore, the diagnostic method shows good performance in estimating fault directions. Practical guidelines for parameter settings have been provided along with an illustration of implementing the monitoring and diagnostic methodology.

Proceeding from this, future work will address monitoring schemes that can adapt themselves to shifts in more effects or coefficients with a reasonably simple computation. Corresponding diagnostic approaches that can identify the shift directions automatically will also be sought. In addition, instead of treating a multi-level factor as nominal data, the ordinal information of the factor levels (e.g., excellent-acceptable-unacceptable) may also be taken into account, and better performance of the constructed control chart could be expected.

Acknowledgement

The authors thank the editor and two anonymous referees for their many helpful comments that have resulted in significant improvements in the article. This research was supported by the RGC Competitive Earmarked Research Grants 620010 and NNSF of China Grants 11001138, 11071128, 11131002, and 11101306. Zou also thanks the support of Nankai Young Grant 65010731.

References

Agresti, A. (2002). *Categorical Data Analysis*, 2nd edition. Wiley, New York.

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd edition. Wiley, New York.

Bersimis, S.; Psarakis, S.; and Panaretos, J. (2007). "Multivariate Statistical Process Control Charts: An Overview". *Quality and Reliability Engineering International* 23, pp.517–543.

Bishop, Y. M. M.; Fienberg, S. E.; and Holland, P. W. (2007). *Discrete Multivariate Analysis*. Springer.

Chiu, J. and Kuo, T. (2008). "Attribute Control Chart for Multivariate Poisson Distribution". *Communications in Statistics: Theory and Methods* 37, pp.146–158.

Dahinden, C., Parmigiani, G., Emerick, M. C., and Bühlmann, P. (2007). "Penalized Likelihood for Sparse Contingency Tables with an Application to Full-Length cDNA Libraries". *BMC Bioinformatics* 8, 476.

Johnson, N. L.; Kotz, S.; and Balakrishnan, N. (1997). *Discrete Multivariate Distributions*. Wiley, New York.

Lowry, C. A. and Montgomery, D. C. (1995). "A Review of Multivariate Control Charts". *IIE Transactions* 27, pp.800–810.

Lowry, C. A.; Woodall, W. H.; Champ, C. W.; and Rigdon, S. E. (1992). "Multivariate Exponentially Weighted Moving Average Control Chart". *Technometrics* 34, pp.46–53.

Lu, X. S.; Xie, M.; Goh, T. N.; and Lai, C. D. (1998). "Control Charts for Multivariate Attribute Processes". *International Journal of Production Research* 36, pp.3477–3489.

Lucas, J. M. and Saccucci, M. S. (1990). "Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements". *Technometrics* 32, pp.1–29.

Marcucci, M. (1985). "Monitoring Multinomial Processes". *Journal of Quality Technology* 17, pp.86–91.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd edition. Chapman & Hall/CRC.

Montgomery, D. C. (2009). *Statistical Quality Control: A Modern Introduction*, 6th edition. Wiley.

Patel, H. I. (1973). "Quality Control Methods for Multivariate Binomial and Poisson Distributions". *Technometrics* 15, pp.103–112.

Qiu, P. (2008). "Distribution-Free Multivariate Process Control based on Log-linear Modeling". *IIE Transactions* 40, pp.664–677.

Ryan, A. G., Wells, L. J., and Woodall, W. H. (2011). "Methods for Monitoring Multiple Proportions When Inspecting Continuously". *Journal of Quality Technology* 43, pp.237–248.

Topalidou, E. and Psarakis, S. (2009). "Review of Multinomial and Multiattribute Quality Control Charts". *Quality and Reliability Engineering International* 25, pp.773–804.

Woodall, W. H. and Ncube, M. M. (1985). "Multivariate CUSUM Quality Control Procedures". *Technometrics* 27, pp.285–292.

Woodall, W. H. (1997). "Control Charts based on Attribute Data: Bibliography and Review". *Journal of Quality Technology* 29, pp.172–183.

Zou, C. and Tsung, F. (2008). "Directional MEWMA Schemes for Multistage Process Monitoring and Diagnosis". *Journal of Quality Technology* 40, pp.407–427.

Zou, C.; Tsung, F.; and Liu, Y. (2008). "A Change Point Approach for Phase I Analysis in Multistage Processes". *Technometrics* 50, pp.344–356.

Supplemental file for the paper titled "Directional Control Schemes for Multivariate Categorical Processes"

supplement.pdf

This pdf file provides certain technical details, including some derivations in Section *New Methodologies for Monitoring and Diagnosis*, the proof of the diagnostic consistency, and some other simulation results.

Appendix

A. Expressions of $G_{\mathrm{MB},k}$ and $G_{\mathrm{MM},(i)k}$

For the kth sample $\mathbf{n}_{\mathrm{MB},k}$ in a multivariate binomial process with p factors, the χ^2 charting statistic given by Patel (1973) is

$$G_{\mathrm{MB},k} = \frac{1}{N}\big(\mathbf{n}_{\mathrm{MB},k} - N\mathbf{p}^{(0)}_{\mathrm{MB}}\big)^T \boldsymbol{\Sigma}^{-1}_{\mathrm{MB}}\big(\mathbf{n}_{\mathrm{MB},k} - N\mathbf{p}^{(0)}_{\mathrm{MB}}\big).$$

Here $\boldsymbol{\Sigma}_{\mathrm{MB}}$ has the elements

$$\boldsymbol{\Sigma}_{\mathrm{MB}\,ij} = \begin{cases} p^{(0)}_{(i)}\big(1 - p^{(0)}_{(i)}\big) & \text{if } i = j \\ p^{(0)}_{(ij)} - p^{(0)}_{(i)}p^{(0)}_{(j)} & \text{if } i \neq j \end{cases},$$

where $p^{(0)}_{(ij)}$ is the IC probability of the factors C_i and C_j both taking Level 1.

For the kth sample in a multivariate multinomial process with p factors, if we focus on only the marginal sums $\mathbf{n}_{\mathrm{MM},(i)k}$ of the factor C_i, the chi-square charting statistic given by Marcucci (1985) is

$$G_{\mathrm{MM},(i)k} = \frac{1}{N}\big(\mathbf{n}_{\mathrm{MM},(i)k} - N\mathbf{p}^{(0)}_{\mathrm{MM},(i)}\big)^T \boldsymbol{\Sigma}^{-1}_{\mathrm{MM},(i)}\big(\mathbf{n}_{\mathrm{MM},(i)k} - N\mathbf{p}^{(0)}_{\mathrm{MM},(i)}\big),$$

where $\mathbf{\Sigma}_{\mathrm{MM},(i)}$ has the elements

$$\mathbf{\Sigma}_{\mathrm{MM},(i)\ uv} = \begin{cases} p^{(0)}_{(i,u)}\big(1 - p^{(0)}_{(i,u)}\big) & \text{if } u = v \\ -p^{(0)}_{(i,u)} p^{(0)}_{(i,v)} & \text{if } u \neq v \end{cases} \qquad u, v = 1, \ldots, h_i - 1.$$

B. An Example of Deriving the Design Matrix $\widetilde{\boldsymbol{X}}$

Take four factors C_1, C_2, C_3, and C_4 with 2, 2, 3 and 3 levels, respectively, for illustration. Let

$$\mathbf{1}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{1}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{J}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \boldsymbol{J}_3 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_2 \\ -\mathbf{1}_2^T \end{bmatrix}.$$

Note that the column sums of matrixes $\boldsymbol{J}_2$ and $\boldsymbol{J}_3$ are all zeros, which assures identifiability. For instance, the design submatrix corresponding to the main effect of C_3 is $\mathbf{1}_2 \otimes \mathbf{1}_2 \otimes \boldsymbol{J}_3 \otimes \mathbf{1}_3$, where $\otimes$ is Kronecker product operator. The design submatrix corresponding to the two-factor interaction effect of C_2C_4 is $\mathbf{1}_2 \otimes \boldsymbol{J}_2 \otimes \mathbf{1}_3 \otimes \boldsymbol{J}_3$, and the design submatrix corresponding to the three-factor interaction effect of $C_1C_3C_4$ is $\boldsymbol{J}_2 \otimes \mathbf{1}_2 \otimes \boldsymbol{J}_3 \otimes \boldsymbol{J}_3$. All the other design submatrixes can be constructed similarly. In a word, given $\mathbf{1}_2 \otimes \mathbf{1}_2 \otimes \mathbf{1}_3 \otimes \mathbf{1}_3$, the design submatrix corresponding to an effect is obtained by replacing $\mathbf{1}$ with $\boldsymbol{J}$ at all the positions where the factors are contained in this effect.

C. Derivation of the GLRT statistic for testing the hypothesis (7)

Denote the IC coefficient vector by $\widetilde{\boldsymbol{\beta}}^{(0)} = \big[\beta_0^{(0)}, \big(\boldsymbol{\beta}^{(0)}\big)^T\big]^T$, and the IC probability vector $\mathbf{p}^{(0)}$ satisfies

$$\mathbf{p}^{(0)} = \exp\Big(\big[\mathbf{1}, \boldsymbol{X}\big]\big[\beta_0^{(0)}, \big(\boldsymbol{\beta}^{(0)}\big)^T\big]^T\Big).$$

Consider the log-likelihood function for the log-linear model (5), which can be written from the probability mass function (PMF) of the multinomial distribution as

$$l\big(\widetilde{\boldsymbol{\beta}}\big) = \mathbf{n}^T \ln \mathbf{p} + \ln N! - (\ln \mathbf{n}!)^T \mathbf{1} = \mathbf{n}^T \widetilde{\boldsymbol{X}} \widetilde{\boldsymbol{\beta}} + \ln N! - (\ln \mathbf{n}!)^T \mathbf{1}.$$

Both the logarithm and the factorial operators on the column vectors operate on each of their entries.

Without loss of generality, assume that in the OC state only the ith ($i \in \{1, \ldots, g\}$) coefficient $\beta_i^{(0)}$ in $\boldsymbol{\beta}^{(0)}$ is incremented by an amount δ_i, and that all the other $\beta_j^{(0)}$ ($j \in$

$\{1,\ldots,h-1\}$ and $j \neq i$) remain unchanged. Here g is the number of coefficients corresponding to effects of the first q orders. The log-likelihood of the observation vector $\mathbf{n}$ in Phase II can be expressed as

$$l(\delta_i) = \mathbf{n}^T \left[\mathbf{1}, \boldsymbol{X}\right] \left[\beta_0^{(0)} + \alpha_i, \left(\boldsymbol{\beta}^{(0)} + \mathbf{d}_i\delta_i\right)^T\right]^T + \ln N! - (\ln \mathbf{n}!)^T\mathbf{1},$$

where α_i is the variation of $\beta_0^{(0)}$ induced by the constraint $\mathbf{1}^T\mathbf{p} = 1$. The MLE $\hat{\delta}_i$ of δ_i is actually the solution to $l'(\delta_i) = 0$. Solving this equation needs a numerical iteration such as that of Newton-Raphson. However, by some approximations, the MLE $\hat{\delta}_i$ of δ_i can be expressed as a simple form, which is shown below.

If only $\beta_i^{(0)}$ in $\boldsymbol{\beta}^{(0)}$ adds by a magnitude δ_i, the variation α_i of $\beta_0^{(0)}$ must satisfy the constraint

$$\mathbf{1}^T \exp\left(\mathbf{1}\beta_0^{(0)} + \boldsymbol{X}\boldsymbol{\beta}^{(0)} + \mathbf{1}\alpha_i + \mathbf{x}_i\delta_i\right) = 1, \tag{A.1}$$

and α_i is calculated as

$$\alpha_i = -\ln\left(\mathbf{1}^T \exp\left(\mathbf{1}\beta_0^{(0)} + \boldsymbol{X}\boldsymbol{\beta}^{(0)} + \mathbf{x}_i\delta_i\right)\right). \tag{A.2}$$

Therefore, $l(\delta_i)$ can be further rewritten as

$$\begin{aligned} l(\delta_i) = \mathbf{n}^T\big[\mathbf{1}\beta_0^{(0)} + \boldsymbol{X}\boldsymbol{\beta}^{(0)} - \mathbf{1}\ln\left(\mathbf{1}^T \exp\left(\mathbf{1}\beta_0^{(0)} + \boldsymbol{X}\boldsymbol{\beta}^{(0)} + \mathbf{x}_i\delta_i\right)\right) + \mathbf{x}_i\delta_i\big] + \\ \ln N! - (\ln \mathbf{n}!)^T\mathbf{1}. \end{aligned}$$

The first-order derivative of $l(\delta_i)$ with respect to δ_i is

$$\begin{aligned} s(\delta_i) &= \frac{\mathrm{d}l(\delta_i)}{\mathrm{d}\delta_i} \\ &= \mathbf{x}_i^T\mathbf{n} - N\mathbf{x}_i^T \exp\big[\mathbf{1}\beta_0^{(0)} + \boldsymbol{X}\boldsymbol{\beta}^{(0)} + \mathbf{x}_i\delta_i - \\ &\quad \mathbf{1}\ln\left(\mathbf{1}^T \exp\left(\mathbf{1}\beta_0^{(0)} + \boldsymbol{X}\boldsymbol{\beta}^{(0)} + \mathbf{x}_i\delta_i\right)\right)\big]. \end{aligned}$$

Let

$$\mathbf{k}(\delta_i) = \exp\left[\mathbf{1}\beta_0^{(0)} + \boldsymbol{X}\boldsymbol{\beta}^{(0)} + \mathbf{x}_i\delta_i - \mathbf{1}\ln\left(\mathbf{1}^T \exp\left(\mathbf{1}\beta_0^{(0)} + \boldsymbol{X}\boldsymbol{\beta}^{(0)} + \mathbf{x}_i\delta_i\right)\right)\right].$$

We can further formulate the second-order derivative $l(\delta_i)$ with respect to δ_i as

$$s'(\delta_i) = \frac{\mathrm{d}^2 l(\delta_i)}{\mathrm{d}\delta_i^2} = -N\mathbf{x}_i^T \mathrm{diag}\big(\mathbf{k}(\delta_i)\big)\mathbf{x}_i + N\mathbf{x}_i^T\mathbf{k}(\delta_i)\mathbf{k}^T(\delta_i)\mathbf{x}_i.$$

Clearly,

$$\begin{aligned} s(0) &= \mathbf{x}_i^T\mathbf{n} - N\mathbf{x}_i^T \exp\left(\mathbf{1}\beta_0^{(0)} + \boldsymbol{X}\boldsymbol{\beta}^{(0)}\right) = \mathbf{x}_i^T(\mathbf{n} - N\mathbf{p}^{(0)}), \\ s'(0) &= -N\mathbf{x}_i^T \mathrm{diag}\big(\mathbf{p}^{(0)}\big)\mathbf{x}_i + N\mathbf{x}_i^T\mathbf{p}^{(0)}(\mathbf{p}^{(0)})^T\mathbf{x}_i = -N\mathbf{x}_i^T\boldsymbol{\Sigma}^{(0)}\mathbf{x}_i. \end{aligned}$$

By performing the first-order Taylor expansion of $s(\delta_i)$ at $\delta_i = 0$, we have

$$s(\delta_i) \approx s(0) + s'(0)\delta_i.$$

Note that the MLE $\hat{\delta}_i$ of δ_i should be the solution to $s(\delta_i) = 0$, and therefore, the MLE $\hat{\delta}_i$ can be approximated as

$$\hat{\delta}_i \approx -\frac{s(0)}{s'(0)} = \frac{1}{N}\big(\mathbf{n} - N\mathbf{p}^{(0)}\big)^T\mathbf{x}_i\big(\mathbf{x}_i^T\boldsymbol{\Sigma}^{(0)}\mathbf{x}_i\big)^{-1}. \tag{A.3}$$

In the IC state, the log-likelihood function of the observation vector $\mathbf{n}$ in Phase II is

$$l_0 = \mathbf{n}^T\big[\mathbf{1}, \boldsymbol{X}\big]\big[\beta_0^{(0)}, \big(\boldsymbol{\beta}^{(0)}\big)^T\big]^T + \ln N! - (\ln \mathbf{n}!)^T\mathbf{1}.$$

The -2LRT statistic will then be $Q_i(\hat{\delta}_i) = 2(l(\hat{\delta}_i) - l_0)$. Here $Q_i(\hat{\delta}_i)$ also has a simple form by some approximations, which is shown below.

From (A.2), $Q_i(\delta_i)$ can be further rewritten as

$$\begin{aligned} Q_i(\delta_i) &= 2\mathbf{n}^T(\mathbf{1}\alpha_i + \mathbf{x}_i\delta_i) \\ &= -2N\ln\big(\mathbf{1}^T\exp\big(\mathbf{1}\beta_0^{(0)} + \boldsymbol{X}\boldsymbol{\beta}^{(0)} + \mathbf{x}_i\delta_i\big)\big) + 2\mathbf{n}^T\mathbf{x}_i\delta_i. \end{aligned}$$

By the second-order Taylor expansion of $Q_i(\delta_i)$ at $\delta_i = 0$, we have

$$Q_i(\delta_i) \approx Q_i(0) + Q_i'(0)\delta_i + \frac{1}{2}Q_i''(0)\delta_i^2. \tag{A.4}$$

In a similar way to the formulation of $s(0)$ and $s'(0)$, the following results hold:

$$\begin{aligned} &Q_i(0) = 0, \\ &Q_i'(0) = 2\mathbf{x}_i^T\big(\mathbf{n} - N\mathbf{p}^{(0)}\big), \\ &Q_i''(0) = -2N\mathbf{x}_i^T\boldsymbol{\Sigma}^{(0)}\mathbf{x}_i. \end{aligned}$$

By substituting these terms as well as $\hat{\delta}_i$ in (A.3) into (A.4) and some algebra, we obtain

$$Q_i(\hat{\delta}_i) \approx \frac{1}{N}\big(\mathbf{n} - N\mathbf{p}^{(0)}\big)^T\mathbf{x}_i\big(\mathbf{x}_i^T\boldsymbol{\Sigma}^{(0)}\mathbf{x}_i\big)^{-1}\mathbf{x}_i^T\big(\mathbf{n} - N\mathbf{p}^{(0)}\big).$$

Up to now, it has been assumed that at most one coefficient $\beta_i^{(0)}$ in $\boldsymbol{\beta}^{(0)}$ will deviate, but its location is unknown. Therefore, for each $\beta_i^{(0)}$ ($i \in \{1, \ldots, g\}$), calculate its corresponding -2LRT statistic $Q_i(\hat{\delta}_i)$ and then combine them into the following GLRT statistic

$$Q = \max_{i\in\{1,\ldots,g\}}\Big(\frac{1}{N}\big(\mathbf{n} - N\mathbf{p}^{(0)}\big)^T\mathbf{x}_i\big(\mathbf{x}_i^T\boldsymbol{\Sigma}^{(0)}\mathbf{x}_i\big)^{-1}\mathbf{x}_i^T\big(\mathbf{n} - N\mathbf{p}^{(0)}\big)\Big).$$

D. Proof of the consistency of $\arg\max_{j\in\{1,\ldots,g'\}} S_j$

Without loss of generality, we follow the assumption used in Appendix B that $\beta_i^{(0)}$ in $\boldsymbol{\beta}^{(0)}$ adds by a magnitude δ_i. In $\widetilde{\boldsymbol{\beta}}^{(1,i)}$, $\boldsymbol{\beta}^{(1,i)} = \boldsymbol{\beta}^{(0)} + \mathbf{d}_i\delta_i$ and $\beta_0^{(1,i)} = \beta_0^{(0)} + \alpha_i$, and the OC probability vector satisfies

$$\mathbf{p}^{(1,i)} = \exp\Big(\big[\mathbf{1}, \boldsymbol{X}\big]\big[\beta_0^{(1,i)}, \big(\boldsymbol{\beta}^{(1,i)}\big)^T\big]^T\Big).$$

To prove the consistency of this estimator is equivalent to showing that

$$\Pr\left\{\bigcup_{j\neq i}[S_j > S_i]\right\} \to 0 \quad \text{as} \quad N\to\infty.$$

By Bonferroni inequality, it further suffices to show that

$$\sum_{j\neq i}\Pr\{S_j > S_i\} \to 0 \quad \text{as} \quad N\to\infty.$$

Denote $z_j = S_j - S_i$. For any $\varepsilon > 0$,

$$\begin{aligned}&\sum_{j\neq i}\Pr\{z_j > \varepsilon\}\\ &\quad\le \sum_{j\neq i}\Pr\Big\{z_j - E(z_j) > \frac{\varepsilon}{2}\Big\} + \sum_{j\neq i}\Pr\Big\{E(z_j) > \frac{\varepsilon}{2}\Big\}\end{aligned} \tag{A.5}$$

Firstly, we handle the second term in (A.5). Note that the constraint (A.1) can also be rewritten as

$$\mathbf{1}^T\exp\big(\mathbf{1}\beta_0^{(1,i)} + \boldsymbol{X}\boldsymbol{\beta}^{(1,i)} - \mathbf{1}\alpha_i - \mathbf{x}_i\delta_i\big) = 1,$$

and correspondingly we have

$$\alpha_i = \ln\big(\mathbf{1}^T\exp\big(\mathbf{1}\beta_0^{(1,i)} + \boldsymbol{X}\boldsymbol{\beta}^{(1,i)} - \mathbf{x}_i\delta_i\big)\big).$$

Moreover, the IC probability vector

$$\begin{aligned}\mathbf{p}^{(0)} &= \exp\big(\mathbf{1}\beta_0^{(0)} + \boldsymbol{X}\boldsymbol{\beta}^{(0)}\big)\\ &= \exp\big(\mathbf{1}\beta_0^{(1,i)} + \boldsymbol{X}\boldsymbol{\beta}^{(1,i)} - \mathbf{1}\alpha_i - \mathbf{x}_i\delta_i\big)\\ &= \exp\big(\mathbf{1}\beta_0^{(1,i)} + \boldsymbol{X}\boldsymbol{\beta}^{(1,i)} - \mathbf{1}\ln\big(\mathbf{1}^T\exp\big(\mathbf{1}\beta_0^{(1,i)} + \boldsymbol{X}\boldsymbol{\beta}^{(1,i)} - \mathbf{x}_i\delta_i\big)\big) - \mathbf{x}_i\delta_i\big).\end{aligned}$$

Since diagnosis is considered in the OC state, we perform the first-order Taylor expansion of $\mathbf{p}^{(0)}$ at $\mathbf{p}^{(1,i)}$ (i.e., $\delta_i = 0$ in the above equation), and obtain

$$\begin{aligned}\mathbf{p}^{(0)} &\approx \exp\big(\mathbf{1}\beta_0^{(1,i)} + \boldsymbol{X}\boldsymbol{\beta}^{(1,i)}\big) - \big(\mathrm{diag}\big(\mathbf{p}^{(1,i)}\big)\mathbf{x}_i - \mathbf{p}^{(1,i)}(\mathbf{p}^{(1,i)})^T\mathbf{x}_i\big)\delta_i \\ &= \mathbf{p}^{(1,i)} - \boldsymbol{\Sigma}^{(1,i)}\mathbf{x}_i\delta_i.\end{aligned}$$

Since $\mathbf{n}$ is an observation vector of size N collected from $F\big(\widetilde{\boldsymbol{X}}; \widetilde{\boldsymbol{\beta}}^{(1,i)}\big)$, its expectation and covariance matrix satisfy

$$\mathrm{E}\big(\mathbf{n} - N\mathbf{p}^{(0)}\big) = N\mathbf{p}^{(1,i)} - N\mathbf{p}^{(0)} = N\boldsymbol{\Sigma}^{(1,i)}\mathbf{x}_i\delta_i,$$

$$\mathrm{Cov}\big(\mathbf{n} - N\mathbf{p}^{(0)}\big) = \mathrm{Cov}\big(\mathbf{n}\big) = N^2\boldsymbol{\Sigma}^{(1,i)}.$$

Let

$$\begin{aligned}&\mathbf{A}_i = \mathbf{x}_i\big(\mathbf{x}_i^T\boldsymbol{\Sigma}^{(1,i)}\mathbf{x}_i\big)^{-1}\mathbf{x}_i^T, \;\; i \in \{1, \ldots, g'\} \\ &\mathbf{A}_j = \mathbf{x}_j\big(\mathbf{x}_j^T\boldsymbol{\Sigma}^{(1,i)}\mathbf{x}_j\big)^{-1}\mathbf{x}_j^T, \;\; j \in \{1, \ldots, g'\} \;\text{ and }\; j \neq i.\end{aligned}$$

Taking the expectation of S_i and S_j, we have

$$\begin{aligned}\mathrm{E}S_i &= \frac{1}{N}\mathrm{tr}\big(\mathbf{A}_i\mathrm{Cov}\big(\mathbf{n} - N\mathbf{p}^{(0)}\big)\big) + \frac{1}{N}\mathrm{E}^T\big(\mathbf{n} - N\mathbf{p}^{(0)}\big)\mathbf{A}_i\mathrm{E}\big(\mathbf{n} - N\mathbf{p}^{(0)}\big) \\ &= N\mathrm{tr}\Big(\big(\mathbf{x}_i^T\boldsymbol{\Sigma}^{(1,i)}\mathbf{x}_i\big)^{-1}\big(\mathbf{x}_i^T\boldsymbol{\Sigma}^{(1,i)}\mathbf{x}_i\big)\Big) + N\mathbf{x}_i^T\boldsymbol{\Sigma}^{(1,i)}\mathbf{A}_i\boldsymbol{\Sigma}^{(1,i)}\mathbf{x}_i\delta_i^2 \\ &= N + N\mathbf{x}_i^T\boldsymbol{\Sigma}^{(1,i)}\mathbf{x}_i\delta_i^2, \\ \mathrm{E}S_j &= \frac{1}{N}\mathrm{tr}\big(\mathbf{A}_j\mathrm{Cov}\big(\mathbf{n} - N\mathbf{p}^{(0)}\big)\big) + \frac{1}{N}\mathrm{E}^T\big(\mathbf{n} - N\mathbf{p}^{(0)}\big)\mathbf{A}_j\mathrm{E}\big(\mathbf{n} - N\mathbf{p}^{(0)}\big) \\ &= N\mathrm{tr}\Big(\big(\mathbf{x}_j^T\boldsymbol{\Sigma}^{(1,i)}\mathbf{x}_j\big)^{-1}\big(\mathbf{x}_j^T\boldsymbol{\Sigma}^{(1,i)}\mathbf{x}_j\big)\Big) + N\mathbf{x}_i^T\boldsymbol{\Sigma}^{(1,i)}\mathbf{A}_j\boldsymbol{\Sigma}^{(1,i)}\mathbf{x}_i\delta_i^2 \\ &= N + N\big(\mathbf{x}_i^T\boldsymbol{\Sigma}^{(1,i)}\mathbf{x}_j\big)\big(\mathbf{x}_j^T\boldsymbol{\Sigma}^{(1,i)}\mathbf{x}_j\big)^{-1}\big(\mathbf{x}_j^T\boldsymbol{\Sigma}^{(1,i)}\mathbf{x}_i\big)\delta_i^2.\end{aligned}$$

From the Cauchy-Schwarz inequality, it follows clearly that

$$\mathrm{E}S_i > \mathrm{E}S_j,$$

say, $E(z_j) < 0$ and thus it immediately follows that the second term of (A.5) equals to zero.

On the other hand, it is not hard to verify that $z_j \to E(z_j)$ as $N \to \infty$ and $\mathrm{Var}(z_j) = O(N^{-1})$. Thus, by the Chebychev inequality, the first term of (A.5) tends to zero as $N \to \infty$. Taking all these results together, (A.5) tends to zero as $N \to \infty$, which establishes the consistency of diagnostic statistic $\arg\max_{j\in\{1,\ldots,g'\}} S_j$.

Table A1. OC ARL comparison between the LLD and MBE charts in cases of one-coefficient-shifts of the first two orders under other parameter settings

Case	δ	LLD		MBE		LLD		MBE		LLD		MBE	
		$\beta_{(3)}$				$\beta_{(5)}$				$\beta_{(1,4)}$			
$\lambda = 0.2$	0.02	122	(1.16)	116	(1.11)	165	(1.60)	142	(1.38)	91.2	(0.83)	180	(1.76)
$N = 1,000$	0.20	2.30	(0.01)	2.38	(0.01)	2.62	(0.01)	2.68	(0.01)	1.94	(0.01)	3.19	(0.01)
$\lambda = 0.1$	0.02	34.8	(0.26)	35.6	(0.27)	46.2	(0.36)	43.2	(0.34)	26.4	(0.17)	60.9	(0.52)
$N = 2,000$	0.20	2.09	(0.01)	2.14	(0.01)	2.33	(0.01)	2.36	(0.01)	1.82	(0.00)	2.68	(0.01)
$\lambda = 0.2$	0.02	54.7	(0.49)	53.8	(0.49)	78.4	(0.72)	69.9	(0.65)	39.2	(0.33)	97.7	(0.93)
$N = 2,000$	0.20	1.71	(0.00)	1.74	(0.01)	1.89	(0.01)	1.92	(0.01)	1.44	(0.00)	2.19	(0.01)
		$\beta_{(2,3)}$				$\beta_{(2,5)}$				$\beta_{(3,4)}$			
$\lambda = 0.2$	0.02	113	(1.08)	173	(1.69)	188	(1.84)	280	(2.75)	135	(1.30)	235	(2.30)
$N = 1,000$	0.20	2.24	(0.01)	3.08	(0.01)	2.78	(0.01)	5.60	(0.02)	2.34	(0.01)	4.16	(0.02)
$\lambda = 0.1$	0.02	34.0	(0.25)	57.7	(0.49)	53.9	(0.44)	123	(1.15)	38.3	(0.28)	88.6	(0.80)
$N = 2,000$	0.20	2.04	(0.01)	2.63	(0.01)	2.45	(0.01)	4.13	(0.01)	2.13	(0.01)	3.35	(0.01)
$\lambda = 0.2$	0.02	51.6	(0.46)	92.6	(0.86)	96.0	(0.90)	186	(1.81)	61.7	(0.57)	141	(1.40)
$N = 2,000$	0.20	1.68	(0.00)	2.15	(0.01)	1.99	(0.01)	3.45	(0.01)	1.75	(0.00)	2.76	(0.01)

NOTE: Standard errors are in parentheses.

Table A2. OC ARL comparison between the LLD and MME charts in cases of one-coefficient-shifts of the first two orders under other parameter settings

Case	δ	LLD		MME		LLD		MME		LLD		MME	
		$\beta_{(1)}$				$\beta_{(3_2)}$				$\beta_{(4_1)}$			
$\lambda = 0.2$	0.02	205	(2.04)	145	(1.42)	203	(2.01)	216	(2.16)	177	(1.71)	169	(1.61)
$N = 1,000$	0.20	2.89	(0.01)	2.60	(0.01)	2.89	(0.01)	2.89	(0.01)	2.65	(0.01)	2.58	(0.01)
$\lambda = 0.1$	0.02	60.1	(0.49)	41.5	(0.32)	78.1	(0.69)	77.3	(0.66)	52.6	(0.42)	50.3	(0.40)
$N = 2,000$	0.20	2.54	(0.01)	2.29	(0.01)	2.53	(0.01)	2.50	(0.01)	2.37	(0.01)	2.32	(0.01)
$\lambda = 0.2$	0.02	107	(1.02)	65.1	(0.60)	125	(1.20)	125	(1.19)	89.4	(0.83)	91.0	(0.86)
$N = 2,000$	0.20	2.06	(0.01)	1.85	(0.01)	2.06	(0.01)	2.01	(0.01)	1.90	(0.01)	1.90	(0.01)
		$\beta_{(1,2)}$				$\beta_{(1,3_2)}$				$\beta_{(2,3_1)}$			
$\lambda = 0.2$	0.02	149	(1.45)	165	(1.58)	211	(2.11)	286	(2.82)	167	(1.63)	216	(2.11)
$N = 1,000$	0.20	2.38	(0.01)	3.24	(0.01)	2.88	(0.01)	4.46	(0.02)	2.50	(0.01)	3.45	(0.01)
$\lambda = 0.1$	0.02	40.5	(0.30)	54.2	(0.44)	80.9	(0.71)	149	(1.42)	48.9	(0.39)	76.9	(0.68)
$N = 2,000$	0.20	2.16	(0.01)	2.73	(0.01)	2.52	(0.01)	3.51	(0.01)	2.25	(0.01)	2.87	(0.01)
$\lambda = 0.2$	0.02	68.3	(0.62)	86.9	(0.81)	131	(1.28)	211	(2.06)	81.6	(0.77)	124	(1.16)
$N = 2,000$	0.20	1.77	(0.00)	2.22	(0.01)	2.03	(0.01)	2.87	(0.01)	1.83	(0.00)	2.34	(0.01)
		$\beta_{(2,4_2)}$				$\beta_{(3_1,4_1)}$				$\beta_{(3_2,4_2)}$			
$\lambda = 0.2$	0.02	227	(2.26)	293	(2.89)	153	(1.46)	207	(2.07)	296	(2.95)	356	(3.55)
$N = 1,000$	0.20	3.07	(0.01)	5.14	(0.02)	2.41	(0.01)	3.31	(0.01)	5.03	(0.02)	32.5	(0.27)
$\lambda = 0.1$	0.02	90.3	(0.81)	172	(1.69)	45.4	(0.35)	73.5	(0.63)	180	(1.74)	318	(3.21)
$N = 2,000$	0.20	2.65	(0.01)	4.00	(0.01)	2.18	(0.01)	2.79	(0.01)	3.87	(0.01)	13.3	(0.07)
$\lambda = 0.2$	0.02	147	(1.44)	246	(2.43)	76.4	(0.71)	123	(1.17)	246	(2.45)	349	(3.49)
$N = 2,000$	0.20	2.14	(0.01)	3.35	(0.01)	1.78	(0.00)	2.29	(0.01)	3.21	(0.01)	15.5	(0.10)

NOTE: Standard errors are in parentheses.

Table A3. Observed matching probability in cases of one-coefficient-shifts of the first three orders under other parameter settings

Case	δ	$\beta_{(2)}$	$\beta_{(4)}$	$\beta_{(1,3)}$	$\beta_{(1,5)}$	$\beta_{(2,3)}$	$\beta_{(4,5)}$	$\beta_{(1,4,5)}$	$\beta_{(2,3,5)}$
	0.02	0.33	0.36	0.44	0.37	0.35	0.41	0.11	0.09
$\lambda = 0.2$	0.05	0.78	0.69	0.74	0.62	0.70	0.73	0.57	0.48
	0.20	0.96	0.87	0.88	0.81	0.90	0.89	0.90	0.88
	−0.02	0.40	0.36	0.44	0.35	0.37	0.40	0.12	0.11
$N = 1,000$	−0.05	0.78	0.67	0.72	0.62	0.69	0.70	0.58	0.49
	−0.20	0.95	0.88	0.87	0.82	0.89	0.89	0.89	0.88
	0.02	0.65	0.58	0.64	0.53	0.58	0.61	0.43	0.34
$\lambda = 0.1$	0.05	0.89	0.75	0.76	0.66	0.78	0.78	0.78	0.74
	0.20	0.96	0.88	0.88	0.81	0.89	0.88	0.90	0.88
	−0.02	0.66	0.56	0.62	0.52	0.59	0.61	0.43	0.35
$N = 2,000$	−0.05	0.88	0.74	0.75	0.66	0.76	0.76	0.77	0.73
	−0.20	0.96	0.88	0.88	0.81	0.90	0.89	0.90	0.88
	0.02	0.51	0.50	0.58	0.48	0.50	0.54	0.23	0.18
$\lambda = 0.2$	0.05	0.88	0.76	0.78	0.69	0.77	0.78	0.75	0.68
	0.20	0.97	0.90	0.91	0.85	0.92	0.91	0.92	0.90
	−0.02	0.54	0.48	0.57	0.46	0.49	0.53	0.24	0.20
$N = 2,000$	−0.05	0.87	0.75	0.76	0.67	0.76	0.77	0.75	0.68
	−0.20	0.97	0.90	0.89	0.87	0.91	0.90	0.93	0.90