Statistical Process Control for Multistage Processes with Binary Outputs

Yanfen Shang[1], Fugee Tsung[2], and Changliang Zou[3*]

[1] *Department of Industrial Engineering, School of Management, Tianjin University, Tianjin, China*

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

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

Abstract

Statistical process control (SPC) including monitoring and diagnosis is very important and challenging for multistage processes with categorical data. This paper proposes a binary state space model (BSSM) for modeling multistage processes with binomial (binary) data and develops corresponding monitoring and diagnosis schemes by utilizing a hierarchical likelihood approach and directional information based on the BSSM. The proposed schemes not only provide an SPC solution that incorporates both interstage and intrastage correlations, but they also resolve the confounding issue in monitoring and diagnosis due to the cumulative effects from stage to stage. Our simulation results show that the proposed schemes consistently outperform the

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

existing χ^2 scheme in monitoring and diagnosing for binomial multistage processes. An aluminum electrolytic capacitor example from the manufacturing industry is used to illustrate the implementation of the proposed approach.
Keywords: Binary State Space Model; Categorical Data; Hierarchical Likelihood; Multistage Processes; Statistical Process Control

1 Introduction

Statistical process control has been demonstrated to be an important tool for monitoring process or product quality in the manufacturing and/or service industries. As modern technologies become increasingly sophisticated, most manufacturing operations, such as semiconductor manufacturing and automotive body assembly, comprise multiple stages. For example, print circuit board (PCB) manufacturing is a typical multistage process that involves exposure to black oxide, lay-up, hot press, cutting, drilling, and then inspection. Shi (2007) has provided an extensive review of the multistage process control problem with many industrial examples.

In many modern manufacturing and service environments, due to some restrictions, for instance the time, cost or intrinsic character of the variables, quantitative observations cannot be measured directly or collected promptly for on-line monitoring use. Instead, only qualitative or categorical observations can be obtained. For example, Lee and Yano (1988) studied a three-stage process for the light-emitting diode fabrication, including the die attach stage, the wire bond stage, and casting stage. The inspectors at each stage only reported either binary observations (e.g., pass/no pass) or nominal observations (e.g., good/scrap/rework), but no numerical measures on the inspected items. Another real example is the manufacturing of aluminium electrolytic capacitor (AEC), which is displayed in Figure 1. AECs are produced through several stages in the process, and in each stage, AECs or unfinished AECs are checked by sampling inspection based on the specifications of some variables, and the inspection output is "defective" or "good", so binary data are collected in each stage. Similarly, such a process in service industries also exists. For instance, the process in hotline center (Figure 2) is one real example from service industries. Customers' problems are solved or answered through several stages, and in each stage, the sampled staff members' actions or answers are inspected by the hotline center to guarantee their service quality. The inspection output is that the answer or action is "correct" or "not correct" (see the following motivating examples in next section for detailed illustration). There still remain several unique challenges in how to apply the SPC method to such multistage processes, including statistical modeling, monitoring and diagnosing. This is the focus of this paper.

Some research on modelling and analyzing multistage processes with categorical observations has been conducted in the literature. Among others, Kumar et al. (2006) provided a comprehensive review of modeling the yield of PCB manufacturing processes. Tang (1990) presented a model for a serial production line with random yield losses and uncertain demand. Barad and Braha (1996) addressed the problem of how to control the ordering of a single or a series of production runs when the various stages in the production process have binomial yields. In Jin et al. (2007), the relationship between the number of surface defects and the process settings was described with a Logistic regression model. However, all the methods mentioned above share a severe drawback that they lack the capability of representing variation propagation along the multiple stages. The assumption that the defective rates of successive stages are independent of each other is commonly invalid in practice. The malfunctioning of an upstream stage would usually increase the defective rate of the downstream stages. In the literature, one way to utilize the multistage information and capture the multistage relationships (e.g., the correlation structure between serially dependent stages) is to model the multistage process as a linear state space model (c.f., Basseville and Nikiforov 1993; Jin and Shi, 1999; Shi 2007). The extensive review on the quality control of multistage systems was presented by Shi and Zhou (2009). Although simple and popular, such a linear model and its corresponding estimation and testing approaches are apparently not appropriate for the multistage categorical observations since they require a fundamental assumption that the collected data is numerical. Ning et al. (2009) discussed the necessity to monitor multistage processes with categorical data, but they did not develop efficient SPC techniques.

With respect to the monitoring of categorical data, the conventional p and np charts are commonly used. In the last several years, this monitoring problem has attracted much attention from researchers and various types of charting schemes have been developed, such as Marcucci (1985), Gan (1993), Acosta-Mejia (1999), Bourke (1991, 2001), Steiner (1998), Reynolds and Stoumbos (1999, 2000), Wu and Spedding (2001), Somerville et al. (2002) and Gadre and Rattihalli (2005), etc. A nice overview on the topic of the control charts for attribute data was presented by Woodall (1997). All the above control charts for categorical data are developed for a single stage and cannot fully utilize the information from a multistage operation where the input of the current process stage may be related to the outputs of the early stages. Naturally, we are led to consider using some multivariate attribute control charts to deal with this problem, such as Patel (1973) and Lu et al. (1998), etc. However, the latest works on multistage process SPC with linear state space model (LSSM) have shown that the fully statistical model-based approaches usually cannot describe the relationship among stages explicitly due to the lack of engineering background and knowledge, and it is possible to develop much more efficient monitoring and diagnosis approaches based on some models that integrate both the statistical and engineering information. See Zhou et al. (2003), Zou et al. (2008), Li and Tsung (2009), and Jin and Tsung (2009) for detailed discussion.

In this study, in order to describe the quality linkage among multiple stages, we propose to model multistage processes with categorical data based on multivariate generalized linear mixed-effect modeling, which can effectively incorporate physical laws and engineering knowledge. A systematic monitoring and diagnosis framework based on the hierarchical likelihood is developed, which is able to fully incorporate directional information from the multistage model. Numerical studies and a real data example demonstrate that the effectiveness of the proposed schemes. The remainder of this paper is organized as follows: in Section 2, we introduce two examples from both the manufacturing and service industries that motivate this research. In Section 3, a new multistage process model is defined and the approach of its parameter estimation is also included. Following that, our methodology on monitoring and diagnosis is elaborated in Section 4. The performance of the proposed schemes are studied by simulation in Section 5. The motivating example is used to illustrate the implementation of the proposed approach step by step in Section 6. Finally, concluding remarks are offered to summarize the major contributions of the paper and to suggest issues for future research. Some technical details are provided in the Appendix.

2 Motivating examples

Here we present two examples taken from both the manufacturing and service industries to illustrate multistage processes with categorical data.

Example 1—A multistage process with binary data in the manufacturing industry: Figure 1 is the process flow of Aluminum Electrolytic Capacitors, which is a typical multistage process with binary output in an electronic manufacturing industry. As shown in Figure 1, the AEC are produced through those five stages, including clenching & rolling, soaking, assembly, cleaning & bushing and aging & classifying. The purpose of this process is to transform the raw materials (anode aluminum foil, cathode aluminum foil, guiding pin, electrolyte sheet, plastic cover, aluminum shell and plastic tube) into the AEC with precise specifications. The products after each stage will be inspected by sampling, and then the binary data (defective or good) are collected. A lot of types of defects may be induced through the manufacturing process, which can be classified into cosmetic defects, capacity defects, leakage current defects and dissipation factor defects. Several types of defects may be induced in each stage. For example, in the clenching & rolling stage, the possible defects include improper compound thickness, aluminum foil cracking, scored aluminum foil, ragged margins and so forth. In the soaking stage, dissipation factor defects may be found. Therefore, the quality character of AEC representing its appearance condition and functional performance condition at upstream stages will influence that of downstream stages, and then the different stages are correlated.

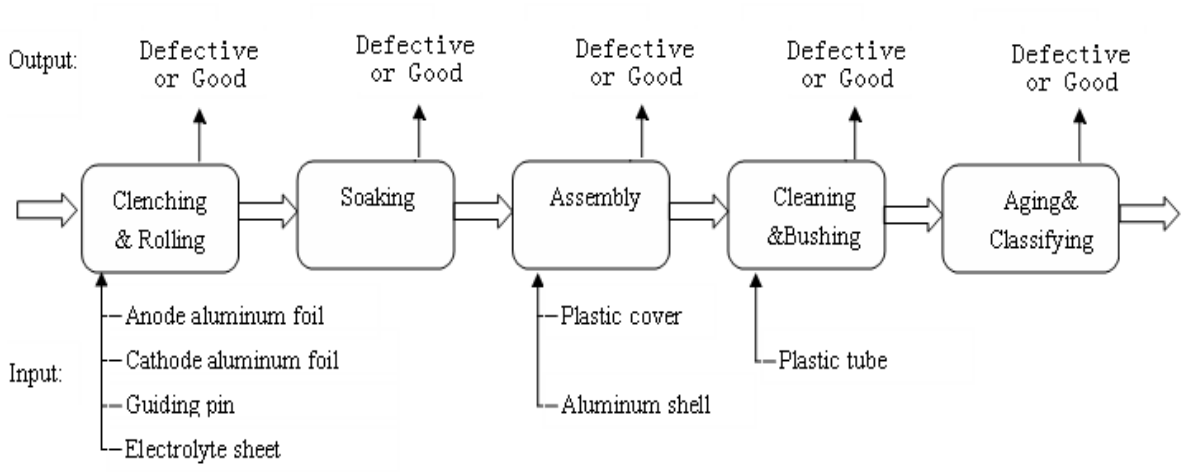


Figure 1: The manufacturing process for Aluminum Electrolytic Capacitors.

Example 2—A multistage process with binary data in the service industry: As shown in Figure 2, a process consisting of three stages represents the flow of dealing with a call in the hotline center. When a customer calls, the staff member will initially analyze the customer's problem and then find solutions in the database or from other sources, e.g., senior staff. After solutions are found, the staff member will present a solution to the customer. The hotline center aims to provide customers with a high-quality service, which means solving problems for customers in the shortest possible time. The service quality of staff members working in the call center is usually monitored by sampling and recording the errors the staff members make. Defects in this process include any error in the information and solutions provided by the staff. Therefore, the binary output in each stage is collected (correct or not correct as shown in Figure 2). Based on this process, it is clear that the service quality character (e.g., correctness of information or solutions) in the upstream stage will affect that of the downstream stage. This hotline center service process, similar to the above manufacturing process, is a typical multistage process with categorical observations. We aim to develop an effective model for this kind of process and to propose a general framework for on-line monitoring and diagnosis in order to control multistage processes by using binary outputs.

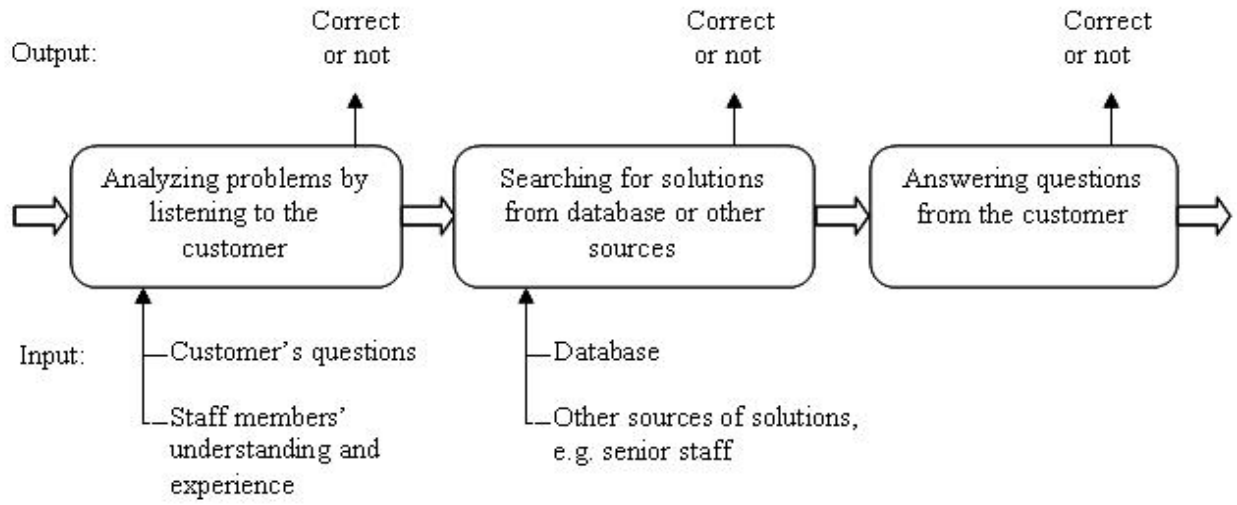


Figure 2: The typical service process in a hotline center

3 Modeling of Multistage Processes with Binomial data

Consider a common manufacturing process comprised of d stages. Without loss of generality, the stages are numbered in ascending order, such that if stage s precedes stage l, then $s < l$. Assume that a multistage process dataset (or a batch) contains m vector observations in the form

$$\boldsymbol{y}_j = (y_{1j}, y_{2j}, \cdots, y_{dj})^T, \ j = 1, , \ldots, m, \tag{1}$$

where $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$ are identically and independently distributed (i.i.d), and y_{ij} is a quality measurement observed from the jth sample at the ith stage and is assumed to be drawn from a binomial (or Bernoulli) distribution, denoted as $y_{ij} \sim \text{BIN}(n_i, p_{ij})$. Here, p_{ij} represents the defect rate of the sample j at stage i, say $p_{ij} = E(y_{ij})/n_i$, and n_i denotes the sample size at stage i. In many applications, the sample sizes at different stages are the same. However, we use n_i here to emphasize that different stages may be of different importance and accordingly the sample size may be different. As the LSSM has been shown by many researchers to be quite effective in describing multistage processes, we propose to use the following two-level BSSM to link y_{ij}

$$\begin{aligned} &\text{logit}(p_{ij}) = \alpha_i + \boldsymbol{x}_{ij}^T \boldsymbol{c}_i, \\ &\boldsymbol{x}_{ij} = \boldsymbol{\beta}_i + \mathbf{A}_i \boldsymbol{x}_{i-1j} + \boldsymbol{\omega}_{ij}, \quad i = 1, 2, \ldots, d, \quad j = 1, \ldots, m, \end{aligned} \tag{2}$$

where $\text{logit}(p_{ij}) = \log(p_{ij}/(1 - p_{ij}))$, α_is are intercept parameters and $\boldsymbol{x}_{ij}$, a q-dimensional vector, represents the quality characteristics at stage i, which cannot be directly (or quickly) be measured or observed.

The first level of the model is a generalized linear model (GLM) for binary or binomial data with a canonical link function logit$(\cdot)$, i.e., the Logistic model (c.f., McCullagh and Nelder 1989). It involves the fitting of the quality measurement to the quality characteristics, and $\boldsymbol{c}_i$, a q-dimensional vector, is used to relate the unobservable quality characteristics, $\boldsymbol{x}_{ij}$, to the observable quality output, y_{ij}, or the defect rate, p_{ij}. The second level of the model involves modeling the transfer of the quality characteristics from the previous stage to the present stage, in which $\mathbf{A}_i$ denotes the $(q \times q)$-dimensional transformation coefficient matrix of the quality characteristics from stage $i-1$ to stage i, and $\boldsymbol{\beta}_i$ is the q-dimensional intercept coefficient vector. The coefficients $\mathbf{A}_i$, $\boldsymbol{c}_i$ and $\boldsymbol{\beta}_i$ are either derived from the process/product design information, physical laws and engineering knowledge (see Zhou et al. 2003, Xiang and Tsung 2008 and the references therein) or estimated from the historical dataset. We will focus on the latter case in the next subsection. In addition, $\boldsymbol{\omega}_{ij}$s, q-dimensional vectors, which signify the unmodeled errors, are assumed to be independently and normally distributed. Similar to the assumptions commonly used in the LSSM (Zou and Tsung 2008), we assume the error $\boldsymbol{\omega}_{ij} \sim N(0, \boldsymbol{\Sigma}_{\omega_i})$, and the initial value $\boldsymbol{x}_{0j} \sim N(\boldsymbol{\alpha}_0, \boldsymbol{\Sigma}_0)$.

Note that the BSSM considered is essentially similar to the model proposed by Czado and Song (2008), which aims to model the longitudinal discrete response where the observation equation is specified in an additive form. Their model and associated estimation approach can be regarded as an extension of the nonlinear state space model investigated by Carlin et al. (1992) and Fahrmeir and Tutz (1994). Besides the fundamental difference in contexts, say the estimation problem considered in their paper and monitoring and diagnosis considered here, they developed a Markov chain Monte Carlo algorithm to carry out statistical inference, which is a computationally intensive method and thus is apparently infeasible in our on-line detection problem. This will be explained in more detail in Section 4.

Now, for the purpose of on-line monitoring and diagnosis, we propose a BSSM-based change-point model. Assume that $\boldsymbol{Y}^{(k)}$ is the quality measurement dataset collected at the k-th time point which contains $m \geq 1$ i.i.d vector observations $\{\boldsymbol{y}_1^{(k)}, \ldots, \boldsymbol{y}_m^{(k)}\}$, where $\boldsymbol{y}_j^{(k)}$ is defined in the same way as in (1) and the superscript "(k)" is used to distinguish the data set collected over time. Suppose that a fault happens in stage ζ after an unknown changepoint τ in the process. In light of (2), we consider the following change-point model: for $k = 1, 2, \ldots,$

$$\begin{aligned} &\text{logit}(p_{ij}^{(k)}) = \alpha_i + \boldsymbol{x}_{ij}^{(k)T}\boldsymbol{c}_i, \\ &\boldsymbol{x}_{ij}^{(k)} = \boldsymbol{\beta}_i + \mathbf{A}_i\boldsymbol{x}_{i-1j}^{(k)} + \boldsymbol{\omega}_{ij}^{(k)} + I_{\{i=\zeta, k\geq\tau\}}\boldsymbol{\delta}, \end{aligned} \quad \text{for} \quad i = 1, 2, \ldots, d, \quad j = 1, \ldots, m, \tag{3}$$

where $I_{\{\cdot\}}$ is the indicator function and $\boldsymbol{\delta}$, a q-dimensional vector, is the unknown mean shift magnitude. Based on this model, the major task of SPC for such a multistage process is to detect any shift ($\boldsymbol{\delta}$) as quickly as possible, using certain control charts, and to diagnose where the detected

shift starts (τ) and in which stage the shift occurs (ζ), so that proper adjustment to the process can be made in a timely fashion.

For constructing a Phase II control scheme, the in-control parameters, say $\mathbf{A}_i$, $\boldsymbol{c}_i$, α_i, $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_{\omega_i}$, are required. In practice, a set of in-control historical process data is gathered and analyzed in Phase I. Any unusual "patterns" in the data lead to adjustments and fine tuning of the process. Once all such assignable causes are accounted for, we are left with a clean set of data, gathered under stable operating conditions. Then, this dataset will be used for estimating certain in-control parameters of the process. In the remainder of this paper, we focus on how to estimate those parameters based on the assumption that the reference sample is statistically stable and on how to develop an effective monitoring and diagnosis scheme is based on the assumption that the in-control parameters are known, i.e., it is assumed that the in-control dataset used in Phase I is sufficiently large enough to estimate the parameters well. A Phase I study, aiming at identifying the outliers or changepoint in a reference dataset, will be presented in a separate paper.

In the Phase I estimation, to enable the model (2) estimated, observations on the quality information, say $\boldsymbol{x}_{ij}$ for $i = 0, \ldots, d$, are required. Such an assumption is valid in practice since we usually expend more measurement efforts on the Phase I analysis so that the underlying regression model can be estimated or calibrated accurately. Then, the initial parameters, $\boldsymbol{\alpha}_0$ and $\boldsymbol{\Sigma}_0$, can be determined directly by the sample mean and variance of $\boldsymbol{x}_{0j}$. The estimates of the coefficients, $\boldsymbol{\beta}_i$, $\mathbf{A}_i$ and $\boldsymbol{\Sigma}_{\omega_i}$, are obtained by multivariate linear regression of $\boldsymbol{x}_i$ on $\boldsymbol{x}_{i-1}$. The parameters $\boldsymbol{c}_i$ and α_i are estimated through a logistic regression of y_i on $\boldsymbol{x}_i$ by means of a standard iterative weighted least square (IWLS) procedure (McCullagh and Nelder 1989). This procedure is very popular and available in most of statistical softwares package, such as R, SAS, and S-plus. Certainly, it is natural to consider the estimation approaches through joint modeling of $\boldsymbol{x}_i$ and y_i from all the stages, although a detailed investigation is beyond the scope of this paper. This should be a subject of future research.

4 Monitoring and Diagnosis Schemes Using hierarchal likelihood (h-Likelihood)

In this section, we will present a general methodology of Phase II monitoring and diagnosis based on the proposed model (3) and h-likelihood estimation. In Section 4.1, we first reformulate the model (2) and derive the h-Likelihood to make inferences. In Section 4.2, we discuss the directional information from the multistage model and the corresponding estimation and testing approaches

are developed. Then, the analogous arguments are readily extendable to construct a Shewhart-type control scheme. Finally, practical guidelines regarding design and computational issues are addressed in Section 4.3.

4.1 Model Transformation and the h-Likelihood

Recall model (2) and the associated notation. Moreover, denote d-dimensional vectors, $\boldsymbol{\nu}_j = (\nu_{1j}, \ldots, \nu_{dj})^T$, and a matrix, $\boldsymbol{\Sigma} = (\rho^2_{ij})_{d\times d}$. Firstly, we rewrite the model (2) as the following equivalent form by replacing $\boldsymbol{x}_{ij}$ with $\boldsymbol{x}_{0j}$,

$$\begin{aligned} y_{ij}|p_{ij} \sim & \text{BIN}(n_i, p_{ij}) \\ \text{logit}(p_{ij}) = & \alpha_i + \sum_{l=1}^{i} \boldsymbol{\beta}_l^T \prod_{t=l+1}^{i} \mathbf{A}_t^T \boldsymbol{c}_i + \boldsymbol{x}_{0j}^T \prod_{t=1}^{i} \mathbf{A}_t^T \boldsymbol{c}_i \\ & + \sum_{l=1}^{i} \boldsymbol{\omega}_{lj}^T \prod_{t=l+1}^{i} \mathbf{A}_t^T \boldsymbol{c}_i, \quad \text{for} \quad i = 1, \ldots, d, \quad j = 1, \ldots, m, \end{aligned}$$

and then the above model can be written as the following,

$$\begin{aligned} & y_{ij}|p_{ij} \sim \text{BIN}(n_i, p_{ij}) \\ & \text{logit}(p_{ij}) = \mu_{0i} + \nu_{ij}, \quad \text{for} \quad i = 1, \ldots, d, \quad j = 1, \ldots, m, \end{aligned} \tag{4}$$

where $\boldsymbol{\nu}_j \overset{\text{i.i.d}}{\sim} N_d(\mathbf{0}, \boldsymbol{\Sigma})$, and

$$\begin{aligned} \mu_{0i} &= \alpha_i + \sum_{l=1}^{i} \boldsymbol{\beta}_l^T \prod_{t=l+1}^{i} \mathbf{A}_t^T \boldsymbol{c}_i + \boldsymbol{\alpha}_0^T \prod_{t=1}^{i} \mathbf{A}_t^T \boldsymbol{c}_i, \\ \rho^2_{ii} &= \boldsymbol{c}_i^T \Big[\Big(\prod_{t=1}^{i} \mathbf{A}_t^T \Big)^T \boldsymbol{\Sigma}_0 \prod_{t=1}^{i} \mathbf{A}_t^T + \sum_{l=1}^{i} \Big(\prod_{t=l+1}^{i} \mathbf{A}_t^T \Big)^T \boldsymbol{\Sigma}_{\omega_l} \prod_{t=l+1}^{i} \mathbf{A}_t^T \Big] \boldsymbol{c}_i, \\ \rho^2_{ir} = \rho^2_{ri} &= \boldsymbol{c}_i^T \Big[\Big(\prod_{t=1}^{i} \mathbf{A}_t^T \Big)^T \boldsymbol{\Sigma}_0 \prod_{t=1}^{r} \mathbf{A}_t^T + \sum_{l=1}^{i} \Big(\prod_{t=l+1}^{i} \mathbf{A}_t^T \Big)^T \boldsymbol{\Sigma}_{\omega_l} \prod_{t=l+1}^{r} \mathbf{A}_t^T \Big] \boldsymbol{c}_r \quad (i < r). \end{aligned}$$

Now, the BSSM is essentially equivalent to the multivariate binomial logit-normal model studied by Coull and Agresti (2000) and Rabe-Hesketh and Skrondal (2001). Also, this model can be regarded as a special case of the generalized linear mixed-effects model (GLMM; c.f., McCulloch and Searle 2001) with multivariate responses, called multivariate GLMM (MGLMM) proposed by Gueorguieva (2001) and further analyzed by Yun and Lee (2004).

In making inferences from such a MGLMM, the marginal likelihood, which is obtained as in the usual GLMM by integrating out the random effects, is always considered. Since such integrations have no explicit form, the commonly used methods are to either approximate the integration by Gauss-Hermit quadrature (GHQ) and Monte-Carlo EM (MCEM) methods (Gueorguieva 2001; Coull and Agresti 2000) or adopt some Bayesian inferential procedures (Czado and Song 2008). However, all these methods involve intensive computation. Yun and Lee (2004) and Lee et al. (2006) implement the h-Likelihood proposed by Lee and Nelder (1996), to make inferences from the MGLMM and compare it with the GHQ method for binary outcomes. They conclude that by using h-likelihood, neither time-consuming numerical integrations such as GHQ nor Monte-Carlo methods to implement the expectation step are necessary to produce good estimates. Due to its attractive features on computation, we follow the framework of h-likelihood in this paper. In addition, as the h-likelihood is quite general and inherits certain good properties of conventional inferences, it enables us to present a unified framework for monitoring and diagnosis of binary multistage processes.

Formally, for model (4), the hierarchical log-likelihood function is defined as follows:

$$h = \sum_{j=1}^{m} \{\log f(\boldsymbol{y}_j|\boldsymbol{\nu}_j) + \log f(\boldsymbol{\nu}_j)\}, \tag{5}$$

where $f(\boldsymbol{y}_j|\boldsymbol{\nu}_j)$ is the conditional density function of $\boldsymbol{y}_j$ given $\boldsymbol{\nu}_j$ with parameters $\boldsymbol{\mu}_0 = (\mu_{01}, \ldots, \mu_{0d})^T$,

$$\log f(\boldsymbol{y}_j|\boldsymbol{\nu}_j) = \sum_{i=1}^{d} \left[\log \begin{pmatrix} n_i \\ \\ y_{ij} \end{pmatrix} + y_{ij}(\mu_{0i} + \nu_{ij}) - n_i \log(1 + \exp\{\mu_{0i} + \nu_{ij}\}) \right],$$

and $f(\boldsymbol{\nu}_j)$ is the density function of random effect $\boldsymbol{\nu}_j$ with the parameter $\boldsymbol{\Sigma}$,

$$-\frac{1}{2}\log|2\pi\boldsymbol{\Sigma}| - \frac{1}{2}\boldsymbol{\nu}_j^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\nu}_j.$$

The h-likelihood defined above is an obvious generalization of Henderson's joint likelihood (c.f., Breslow and Clayton 1993; Davidian and Giltinan 1995). Given dispersion parameters $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_{w_i}$, we can have consistent estimates of $\boldsymbol{\alpha}_0$, α_i and $\boldsymbol{\beta}_i$ for $i = 1, \ldots, d$ by maximizing the h-likelihood function h, i.e., solving the simultaneous score equations of

$$\partial h/\partial \boldsymbol{\alpha} = 0, \quad \partial h/\partial \boldsymbol{\beta} = 0, \quad \partial h/\partial \boldsymbol{\nu} = 0. \tag{6}$$

The details of general estimation procedures and corresponding statistical software can be found in Lee et al. (2006) and hence are omitted here. We will elaborate on model (3), which is of our primary interest in the next section, and provide some useful formulations and algorithms to simplify the computation for the current special problem in the Appendix.

4.2 Directional Information and the Corresponding Test

Suppose that a process mean shift occurs at one stage, ζ, after τ observed datasets as represented by model (3), i.e., the unobservable quality characteristics, $\boldsymbol{x}_{\zeta j}^{(k)}, k = \tau+1, \tau+2, \ldots,$ shift. Then, the observable quality measurements, $y_{\zeta j}^{(k)}, \ldots, y_{dj}^{(k)}$, all may deviate from their expectations. In fact, the magnitude and type of such deviations depend on the parameters, $\mathbf{A}_i$, $\boldsymbol{c}_i$, α_i, $\boldsymbol{\Sigma}_0$, $\boldsymbol{\Sigma}_{w_i}$, the shift magnitude, $\boldsymbol{\delta}$, and the stage, ζ, of the shift occurrence. When these parameters are specified, the finite possible shift directions of the intermediate variable $\text{logit}(p_{ij}^{(k)})$, i.e., $\boldsymbol{\mu}$, can be determined. This a priori shift direction information is a particular characteristic in monitoring and diagnosing binary multistage processes. Note that the efficiency of control schemes that make use of the available shift information has been demonstrated by Jiang (2004) and Zou and Tsung (2008) in related but different contexts. We shall develop the control schemes based on the BSSM by taking the a priori knowledge of the mean shift direction into consideration.

To be more specific, similar to (4), we can see that the equivalent form of the out-of-control model (3) can be represented as follows: for $i = 1, \ldots, d, j = 1, \ldots, m,$ and $k > \tau$,

$$\begin{aligned} & y_{ij}^{(k)} | p_{ij}^{(k)} \sim \text{BIN}(n_i, p_{ij}^{(k)}) \\ & \text{logit}(p_{ij}^{(k)}) = \mu_{1i} + \nu_{ij}^{(k)}, \quad \mu_{1i} = \mu_{0i} + \boldsymbol{\delta}^T \boldsymbol{\psi}_{\zeta i}, \end{aligned} \tag{7}$$

where $\boldsymbol{\psi}_{\zeta i}$ is a q-dimensional vector

$$\boldsymbol{\psi}_{\zeta i} = \begin{cases} \prod_{t=\zeta+1}^{i} \mathbf{A}_t^T \boldsymbol{c}_i & \text{if} \quad i \geq \zeta \\ \mathbf{0} & \text{otherwise.} \end{cases}$$

Let $\Psi_\zeta = (\boldsymbol{\psi}_{\zeta 1}, \boldsymbol{\psi}_{\zeta 2}, \ldots, \boldsymbol{\psi}_{\zeta d})^T$. Therefore, for the multistage process with categorical data considered here, in which the mean shift occurs only in one of the d types of known directions, $\Psi_1, \Psi_2, \ldots, \Psi_d$, say $\zeta \in [1, d]$, the hypothesis test is given by,

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad \text{versus} \quad H_1 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 + \Psi_1 \boldsymbol{\delta} \quad \text{or} \quad \boldsymbol{\mu}_0 + \Psi_2 \boldsymbol{\delta} \quad \cdots \quad \text{or} \quad \boldsymbol{\mu}_0 + \Psi_d \boldsymbol{\delta}. \tag{8}$$

Similar to the generalized likelihood ratio test (GLRT) adopted in Jiang (2004) and Zou et al. (2008), we could use the generalized h-likelihood ratio test (GHLR) in the sense that we jointly maximize the h-likelihood over $\zeta \in \{1, \ldots, d\}$ and $\boldsymbol{\delta}$. This will result in rejecting the null hypothesis if, for given k, the test statistic,

$$T_L^{(k)} = \max_{1 \leq i \leq d} 2[h_i(\widehat{\boldsymbol{\delta}}_i; \mathbf{Y}^{(k)}, \widehat{\boldsymbol{\nu}}_i) - h_0(\boldsymbol{Y}^{(k)}, \widehat{\boldsymbol{\nu}}_0)], \tag{9}$$

is larger than a pre-specified critical value, where $h_0(\mathbf{Y}^{(k)}, \widehat{\boldsymbol{\nu}}_0)$ is the h-likelihood under H_0 conditions and $h_i(\widehat{\boldsymbol{\delta}}_i; \mathbf{Y}^{(k)}, \widehat{\boldsymbol{\nu}}_i)$ denotes the h-likelihood function under H_1 with Ψ_i. Here, $\widehat{\boldsymbol{\delta}}_i$ and $\widehat{\boldsymbol{\nu}}_i$ are the maximum h-likelihood estimators (MHLE) for h_i, and $\widehat{\boldsymbol{\nu}}_0$ is the MHLE for h_0. Note that since the in-control parameters $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}$ are both known under the Phase II setting, the fixed-effect parameter involved in maximizing h_i is just $\boldsymbol{\delta}$ and there are only random-effect parameters, $\boldsymbol{\nu}$, that need to be estimated for the h_0. This will greatly alleviate the computational effort and on-line implementation is feasible by using some appropriate algorithms. See Section 4.3 for details. Based on the GHLR (9), an estimator for ζ is immediately given by

$$\widehat{\zeta}_L^{(k)} = \arg\max_{1\le i\le d} 2[h_i(\widehat{\boldsymbol{\delta}}_i; \mathbf{Y}^{(k)}, \widehat{\boldsymbol{\nu}}_i) - h_0(\boldsymbol{Y}^{(k)}, \widehat{\boldsymbol{\nu}}_0)]. \tag{10}$$

By using $T_L^{(k)}$, we can directly construct a Shewhart-type control chart that triggers a signal if

$$T_L^{(k)} > C,$$

where $C > 0$ is a control limit chosen to achieve a specific in-control average run length (ARL), denoted as ARL_0. After the chart signals, $\widehat{\zeta}_L^{(k)}$ provides a direct diagnosis result at which stage a shift has occurred. It should be emphasized that using similar arguments to Appendix B in Lee and Nelder (1996), the expansion of $T_L^{(k)}$ leads to asymptotically equivalent Wald-type charting statistics:

$$T_W^{(k)} = \max_{1\le i\le d} \widehat{\boldsymbol{\delta}}_i^T \mathrm{Var}^{-1}(\widehat{\boldsymbol{\delta}}_i)\widehat{\boldsymbol{\delta}}_i, \quad k = 1, 2, \ldots, \tag{11}$$

and the corresponding diagnostic procedure will be

$$\widehat{\zeta}_W^{(k)} = \arg\max_{1\le i\le d} \widehat{\boldsymbol{\delta}}_i^T \mathrm{Var}^{-1}(\widehat{\boldsymbol{\delta}}_i)\widehat{\boldsymbol{\delta}}_i. \tag{12}$$

Since the explicit form of $\mathrm{Var}(\widehat{\boldsymbol{\delta}}_i)$ is not available, a simple but consistent estimator given in Appendix C is used instead. Note that when the multistage process with LSSM is considered, $T_L^{(k)}$ and T_W^k are exactly the same (c.f., Zou and Tsung 2008). However, for the BSSM, the small-sample properties of these two statistics are slightly different and we will examine the difference in Section 5. Our simulation results show that the $T_L^{(k)}$ and T_W^k have similar performance in detecting multistage shifts and the latter one performs slightly better in most of cases. It is also worth noting that using $T_W^{(k)}$ does not require computing the h-likelihood value under in-control conditions, say $h_0(\boldsymbol{Y}^{(k)}, \widehat{\boldsymbol{\nu}}_0)$, and thus it has a little advantage over using $T_L^{(k)}$ in computation effort. Therefore, we recommend $T_W^{(k)}$ in practical applications.

Note that Zou et al. (2008) establish some good asymptotic properties of the directional GLRT and corresponding estimators in the case that LSSM is considered. Although no theoretical result

will be presented here on the superiority of the directional GHLR chart under one of the special shift patterns (8), its efficiency can be expected to be superior to the testing approaches directly based on the observed vector, $\boldsymbol{Y}^{(k)}$, due to its exploration of more accurate information. Its superiority will be examined by simulation in Section 5.

4.3 Implementation of the Proposed Scheme

Although computing power constantly increases, for on-line process monitoring, which generally handles a large number of observations, fast implementation is important and some computational issues need our careful examination. We will detail the numerical algorithm to calculate the test statistic, $T_W^{(k)}$, and the implementation of the algorithm. Note that Lee et al. (2006) provide a general IWLS to obtain the MHLE, although some special characteristics in our problem would result in considerable reduction of computation effort and are hence worth pointing out more clearly here.

To obtain the $\widehat{\boldsymbol{\delta}}_i$ in (12), the score equations in (6) can be solved through the IWLS procedure with the augmented response variables (Lee and Nelder, 1996). To be specific, write $\boldsymbol{\xi} = (\boldsymbol{\delta}^T, \boldsymbol{\nu}^T)^T$ as parameters vectors, where $\boldsymbol{\nu}^T = (\boldsymbol{\nu}_1^T, \ldots, \boldsymbol{\nu}_m^T)$ and we suppress the subscript "i" in $\boldsymbol{\delta}$ and Ψ for notation convenience. We will show in Appendix A that $\widehat{\boldsymbol{\delta}}_i$ and $\widehat{\boldsymbol{\nu}}_i$ can be obtained by solving the following IWLS equation:

$$\boldsymbol{\Gamma}^T \boldsymbol{\Omega}^{-1} \boldsymbol{\Gamma} \boldsymbol{\xi} = \boldsymbol{\Gamma}^T \boldsymbol{\Omega}^{-1} \boldsymbol{z}, \tag{13}$$

where

$$\boldsymbol{\Gamma} = \begin{bmatrix} \Psi^*_{md\times q} & \mathbf{I}_{md} \\ \mathbf{0}_{md\times q} & \mathbf{I}_{md} \end{bmatrix}, \quad \boldsymbol{\Omega} = \mathrm{diag}\{\boldsymbol{\Omega}_0, \boldsymbol{\Omega}_1\},$$

$\mathbf{I}_{md}$ denotes the (md)-dimensional identity matrix, $\mathbf{0}_{md\times q}$ is the $(md \times q)$-dimensional matrix with all the elements zero, $\Psi^* = (\Psi^T, \Psi^T, \ldots, \Psi^T)^T$, $\boldsymbol{z} = (\boldsymbol{z}_0^T, \boldsymbol{z}_1^T)^T$ is the so-called GLM augmented response variables, and

$$\boldsymbol{z}_l = (\boldsymbol{z}_{l1}^T, \ldots, \boldsymbol{z}_{lm}^T)^T, \quad \boldsymbol{\Omega}_l = \mathrm{diag}\big\{\boldsymbol{\Omega}_{l1}, \ldots, \boldsymbol{\Omega}_{lm}\big\},$$

$$\boldsymbol{z}_{lj} = (z_{l1j}, \ldots, z_{ldj})^T, \quad \boldsymbol{\Omega}_{0j} = \mathrm{diag}\{\Omega_{01j}, \ldots, \Omega_{0dj}\},$$

$$z_{0ij} = \mathrm{logit}(p_{ij}) - \mu_{0i} + \frac{y_{ij} - n_i p_{ij}}{n_i p_{ij}(1-p_{ij})}, \quad \Omega_{0ij} = [n_i p_{ij}(1-p_{ij})]^{-1},$$

$$z_{1ij} = 0, \quad \boldsymbol{\Omega}_{1j} = \boldsymbol{\Sigma}, \quad \text{for} \quad l = 0, 1, \quad i = 1, \ldots, d, \quad j = 1, \ldots, m.$$

The IWLS for estimating $\boldsymbol{\xi}$ are described as follows:

1. Start with the given initial estimation values for $\boldsymbol{\xi}$, denoted as $\widehat{\boldsymbol{\xi}}_{(0)} = (\widehat{\boldsymbol{\delta}}_{(0)}, \widehat{\boldsymbol{\nu}}_{(0)}^T)^T$.

2. At the l-th iteration, for $l \geq 0$, calculate $\boldsymbol{z}_{(l)}$, $\boldsymbol{\Gamma}_{(l)}^T$ and $\boldsymbol{\Omega}_{(l)}$ based on $\widehat{\boldsymbol{\xi}}_{(l)}$.

3. Update the estimation of $\boldsymbol{\xi}$ via the following equation,

$$\widehat{\boldsymbol{\xi}}_{(l+1)} = (\boldsymbol{\Gamma}_{(l)}^T \boldsymbol{\Omega}_{(l)}^{-1} \boldsymbol{\Gamma}_{(l)})^{-1} \boldsymbol{\Gamma}_{(l)}^T \boldsymbol{\Omega}_{(l)}^{-1} \boldsymbol{z}_{(l)}. \tag{14}$$

4. Repeat Steps 2-3 until the following condition is satisfied:

$$\|\widehat{\boldsymbol{\xi}}_{(l)} - \widehat{\boldsymbol{\xi}}_{(l-1)})\|_1 \Big/ \|\widehat{\boldsymbol{\xi}}_{(l-1)}\|_1 \leq \epsilon,$$

   where ϵ is a pre-specified and small positive number (e.g., $\epsilon = 10^{-4}$) and $||\boldsymbol{\xi}||_1$ denotes the sum of absolute values of all elements of $\boldsymbol{\xi}$. Then, the algorithm stops at the l-th iteration.

Note that, in the last step, we use the relative error of the successive estimates of $\boldsymbol{\xi}$ in the convergence criterion. In fact, a part of estimates $\boldsymbol{\xi}$, such as $\boldsymbol{\delta}$, or $\boldsymbol{\nu}$ can also be used for this purpose. We use $\boldsymbol{\xi}$ here because our simulation shows that it gives good results in various cases. With respect to the choice of the initial values of $\widehat{\boldsymbol{\xi}}_{(0)}$, a simple but effective method is to set it to be the $\mathbf{0}$ for which we found the frequency of nonconvergence is negligible in all our simulation studies and the convergence is quite fast (usually only requiring 3-4 iterations). In addition, it seems that in each iteration, the IWLS equation(13) requires computing the inverse of a high-dimensional matrix, say $(\boldsymbol{\Gamma}_{(l)}^T \boldsymbol{\Omega}_{(l)}^{-1} \boldsymbol{\Gamma}_{(l)})^{-1}$. However, due to the special forms of $\boldsymbol{\Gamma}_{(l)}$ and $\boldsymbol{\Omega}_{(l)}$, we could accomplish the computation in the order of $O(d^3+m^2)$ rather than $O(m^3 d^3)$. The details are provided in Appendix B.

Based on the foregoing algorithm, it is easy to obtain the upper α percentile of $T_L^{(k)}$ or $T_W^{(k)}$ by simulation, say C, the control limit of the proposed Shewhart chart for $\text{ARL}_0 = 1/\alpha$. For instance, when $d = 10$, about 15 minutes is required to obtain C based on 10,000 simulations, using a Pentium 3.0MHz CPU. The Fortran code for implementing the proposed procedure is available from the authors upon request.

5 Performance Assessment

In this section, the monitoring performance and the diagnosis performance of our proposed schemes for a multistage process with binary data are investigated via Monte Carlo simulations. For ease

of exposition, only the univariate model is considered here,

$$\begin{aligned} &\text{logit}(p_{ij}) = \alpha_i + c_i x_{ij}, \\ &x_{ij} = \beta_i + a_i x_{i-1j} + \omega_{ij}, \end{aligned} \qquad \text{for} \quad i = 1, 2, \ldots, d,$$

where the initial value $x_{0j} \sim N(\alpha_0, \sigma_0^2)$, and the error $\omega_{ij} \sim N(0, \sigma_{\omega_i}^2)$.

Table 1: Charting performance comparison with parameters (a_i, c_i)=(1.0,1.0) for $d = 3$.

	ζ =1			$\zeta = 2$			$\zeta = 3$		
δ	$T_L^{(k)}$	$T_W^{(k)}$	χ^2	$T_L^{(k)}$	$T_W^{(k)}$	χ^2	$T_L^{(k)}$	$T_W^{(k)}$	χ^2
0.15	0.021	0.022	0.022	0.019	0.021	0.017	0.015	0.015	0.012
	(0.0014)	(0.0015)	(0.0015)	(0.0014)	(0.0014)	(0.0013)	(0.0012)	(0.0012)	(0.0011)
0.30	0.106	0.113	0.101	0.085	0.091	0.072	0.054	0.059	0.0451
	(0.0031)	(0.0032)	(0.0030)	(0.0028)	(0.0029)	(0.0026)	(0.0023)	(0.0024)	(0.0021)
0.45	0.330	0.341	0.304	0.265	0.275	0.220	0.160	0.174	0.1405
	(0.0047)	(0.0047)	(0.0046)	(0.0044)	(0.0045)	(0.0041)	(0.0037)	(0.0038)	(0.0035)
0.60	0.643	0.655	0.601	0.554	0.566	0.497	0.386	0.406	0.342
	(0.0048)	(0.0047)	(0.0049)	(0.0050)	(0.0050)	(0.0050)	(0.0049)	(0.0049)	(0.0047)
0.75	0.881	0.885	0.851	0.805	0.813	0.760	0.650	0.668	0.6048
	(0.0032)	(0.0032)	(0.0036)	(0.0040)	(0.0039)	(0.0043)	(0.0048)	(0.0047)	(0.0049)

Note: values in parentheses are the standard errors

5.1 Monitoring Performance Analysis

The monitoring performance of the proposed schemes based on GHLR $T_L^{(k)}$ and Wald-type $T_W^{(k)}$ are evaluated in this subsection in terms of the probability of correctly rejecting the null hypothesis, say the power. Since Shewhart-type control schemes are considered here, these power comparisons results are exactly consistent with the comparison in terms of ARL. Throughout this subsection, the in-control type I error, α, is fixed at 0.005 and the number of stages, $d = 3$ or 5, is considered.

Without loss of generality, we assume that the parameters $\alpha_0 = 0, \beta_i = 0, \sigma_0^2 = \sigma_{\omega i}^2 = 0.1$ and $\alpha_i = -2.2$ make the expectation of the defective rate p_{ij} approximately equal to 0.1. For the parameter n_i in binomial distribution, the value 30 is used for all i, and the parameter m is assumed to be 20. We consider a benchmark available in the literature, the χ^2 control scheme proposed by Patel (1973), which is one of the most representative approaches for monitoring multivariate binary data. Its charting statistic is described as follows:

$$G = m(\bar{\boldsymbol{y}} - \boldsymbol{\mu}_Y)^T \boldsymbol{\Sigma}_Y^{-1} (\bar{\boldsymbol{y}} - \boldsymbol{\mu}_Y),$$

where $\bar{\boldsymbol{y}} = m^{-1} \sum_{j=1}^{m} \boldsymbol{y}_j^{(k)}$, and $\boldsymbol{\mu}_Y$ and $\boldsymbol{\Sigma}_Y$ are the mean vector and covariance matrix of multivariate binary data, which are determined from the historical data.

As we know, charting performance is related to the shifting stage, ζ, where the shift initially occurs, and the mean shift magnitude δ. Here, we first consider $a_i = c_i = 1$ and the simulation results for $d = 3$ and $d = 5$ over 10,000 replications are shown in Tables 1 and 2, respectively. Similar parameter values are used by Xiang and Tsung (2008) and Zou and Tsung (2008). The shift magnitudes considered are 0.15, 0.30, 0.45, 0.60, 0.75, and 0.90. From these two tables, we can see that both of our proposed schemes (GHLR and Wald-type) are almost uniformly better than the χ^2 scheme in detecting any magnitude of shift δ and improvement is quite significant in most of cases. It is worth noting that the superiority of the proposed schemes to the χ^2 chart is more prominent in $d = 5$ cases and we can also expect that it will become more remarkable in higher dimensional problems because recognizing the directional information correctly allows focusing the detection power on a limited subspace with improved sensitivity. This finding is consistent with that of Zou and Tsung (2008) where the LSSM is considered. In some cases the difference between our proposed schemes and the χ^2 scheme is not significant. Certainly, it is not surprising to us that the χ^2 method is effective in certain cases because it utilizes the information of marginal distribution from each stage. However, as shown in this table, in some other cases, the proposed methods outperform the χ^2 method by a considerable margin. This can be understood that the proposed methods more efficiently exploit the information between stages by using BSSM. Besides, the Wald-type scheme performs slightly better than the GHLR scheme. A possible explanation for this phenomenon is that, in $T_W^{(k)}$, the $\widehat{\delta}_i$ is explicitly standardized by dividing its standard deviation, while the GHLR does this "internally", which may show that the first two moments of $[h_i(\widehat{\delta}_i; \mathbf{Y}^{(k)}, \widehat{\boldsymbol{\nu}}_i) - h_0(\boldsymbol{Y}^{(k)}, \widehat{\boldsymbol{\nu}}_0)]$ for different i are considerably different. As a consequence, the operator "max" loses its power to some extent. A similar finding for GLRT in a different context can be found in Zou et al. (2006).

Next, we demonstrate that the superiority of the proposed schemes to their counterparts still holds when the values of the parameters, a_i and c_i, vary. For illustration purposes, we consider the combinations of different values of a_i and c_i, each varying from 0.5 and 1.0. The performance of

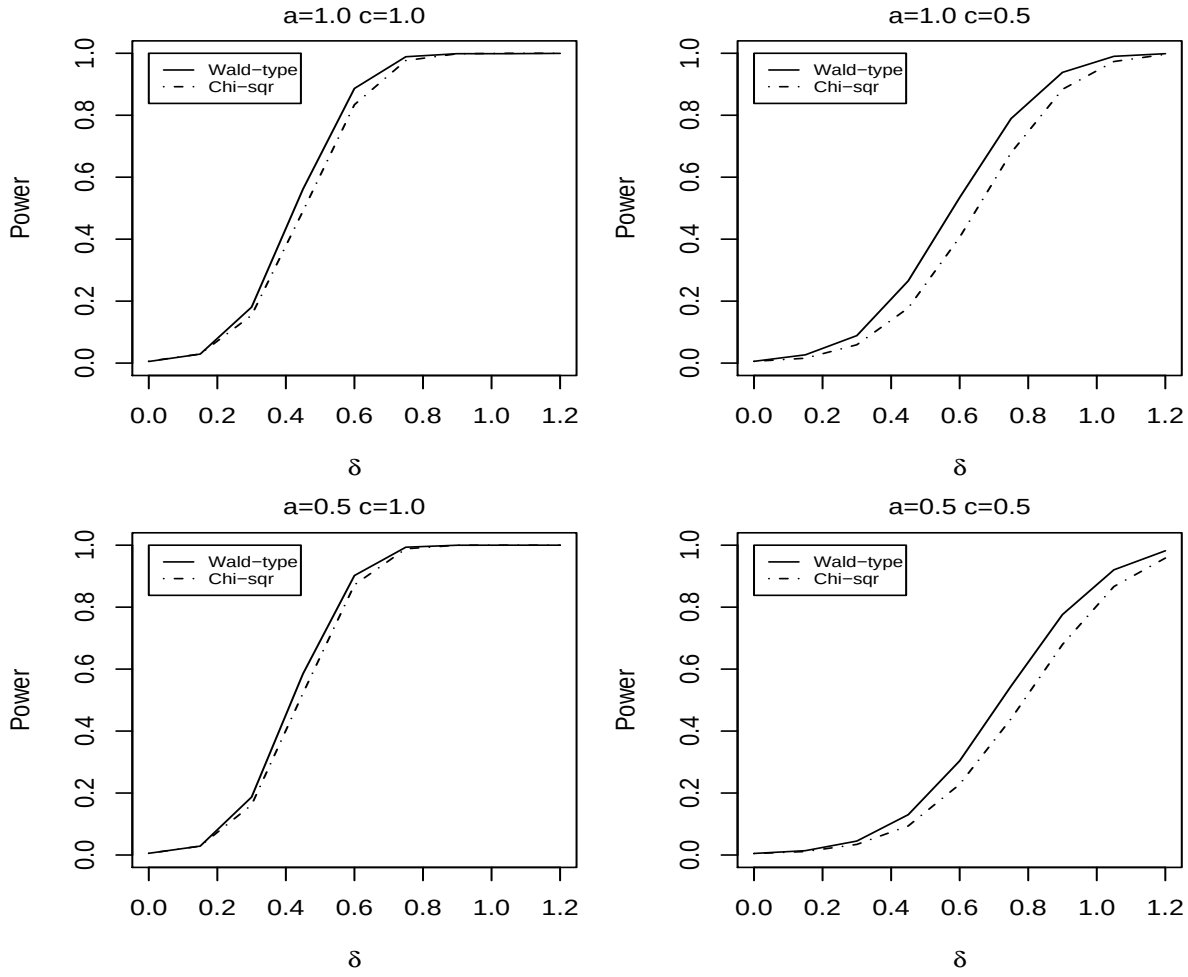


Figure 3: Power comparision with varying parameters for $\zeta = 3$ in a 5-stage process. The solid and dash-dotted lines represent the $T_W^{(k)}$ and χ^2 charts, respectively.

$T_L^{(k)}$ and $T_W^{(k)}$ are quite similar as shown in Tables 1 and 2. Thus, in what follows, we only consider the latter one. For the sake of brevity, only $d = 5$ results are shown here. The power curves for the given shifted stage $\zeta = 3$ against shift magnitude δ and the curves against ζ under the given magnitudes of mean shift δ are summarized in Figures 3 and 4, respectively. As shown in Figure 3, when the shift occurs in stage 3, for different shifts δ, the Wald-type scheme has larger power in signifying the shift than the χ^2 chart especially for middle shifts. Also as shown in Figure 4, when the shift δ is set as a value 0.4 or 0.6, the power of the Wald-type scheme is better than that of the χ^2 chart. In general, we can see that no matter how the values of (a_i, c_i) vary, the Wald-type scheme still performs uniformly better than the χ^2 chart does. In some situations, especially for $c = 0.5$ cases, the χ^2 chart is outperformed by a quite large margin. Based on the results in Tables

1 and 2, Figures 3 and 4 and other simulations for various multistage models (available from the authors), we conclude that the $T_W^{(k)}$ chart, by taking advantage of incorporating shift direction information from the BSSM, turns out to be almost always better than the χ^2 chart in detecting any process shift. In the following subsection, we assess the effectiveness of the corresponding diagnosis methods as well.

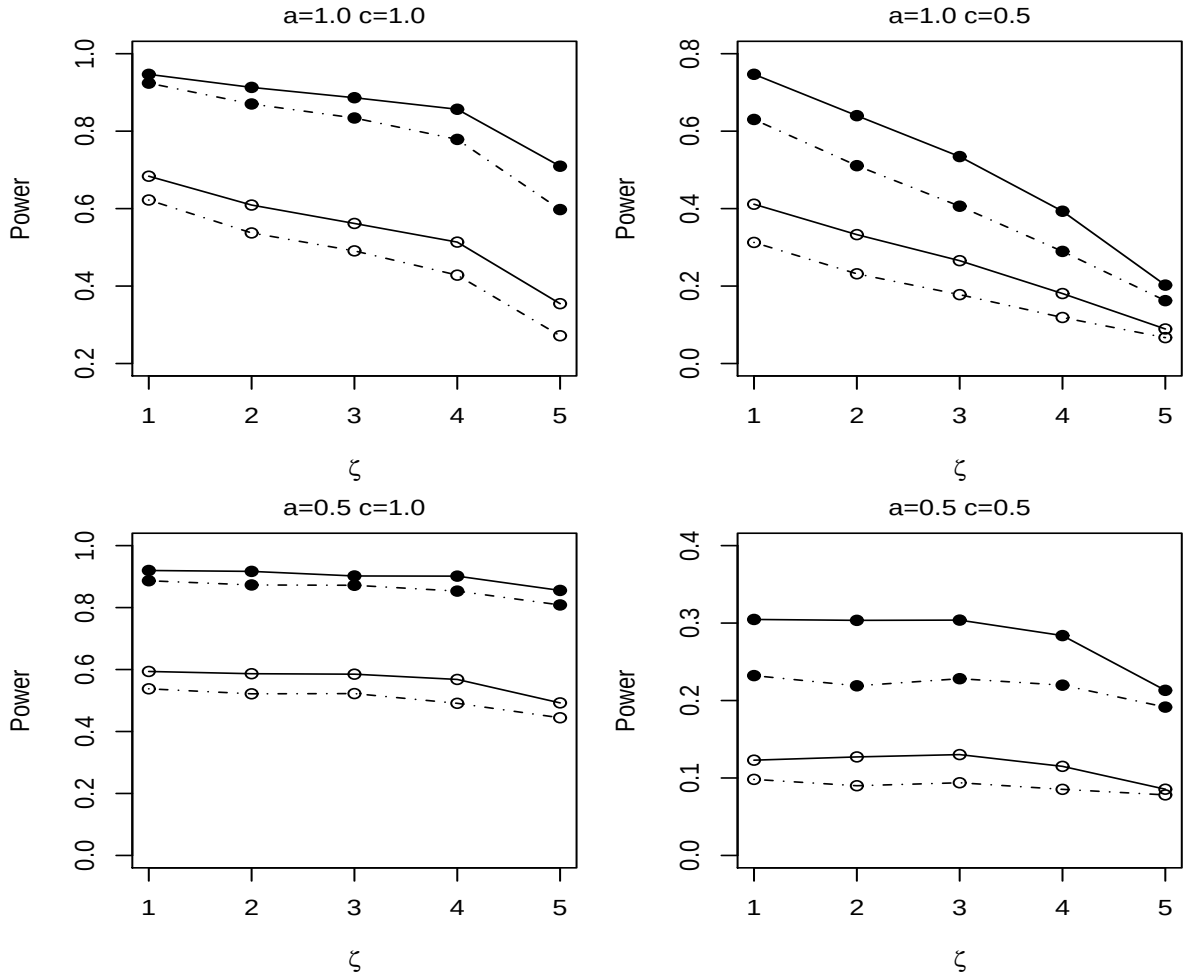


Figure 4: Power Comparision with varying parameters under given magnitudes of mean shift δ. The lines connecting the circles and dots represent the curves for $\delta = 0.45$ and $\delta = 0.6$ respectively. The solid and dash-dotted lines represent the $T_W^{(k)}$ and χ^2 charts, respectively.

5.2 Diagnosis Performance Analysis

In this subsection, we investigate the diagnosis performance of the proposed schemes for multistage process with binary data in diagnosing the shifting stage where the mean shift initially occurs. Although the diagnosis of the χ^2 scheme may be achieved by decomposing the statistic into several independent parts as Mason, Young and Tracy (1995) did for continuous data, it seems that the diagnosis based on decomposition would be misleading because it cannot correctly indicate the stage where the shift initially occurs. A comparable method that comes to mind is

$$\widehat{\zeta}_I = \arg\max_{1\le i\le d} \frac{m(\bar{y}_{(i)} - \mu_{(i)})^2}{\sigma^2_{(i)}},$$

where $\bar{y}_{(i)}$ denotes the i-th component of $\bar{\boldsymbol{y}}$, $\mu_{(i)}$ denotes the i-th component of $\boldsymbol{\mu}_Y$ and $\sigma^2_{(i)}$ denotes the (i,i)-th element of $\boldsymbol{\Sigma}_Y$. We will use this estimator as a benchmark to assess the diagnosis performance of $\widehat{\zeta}_W$. The results of $\widehat{\zeta}_L$ are not provided here because it has similar performance to $\widehat{\zeta}_L$ in the monitoring performance comparison.

The estimated probabilities, $\Pr(\widehat{\zeta} = \zeta)$ (denoted by P_0), for $d = 3$ and 5 are respectively presented in Tables 3 and 4. For $d = 5$, the probabilities, $\Pr(|\widehat{\zeta} - \zeta| \le 1)$ (denoted by P_1), are also provided, which provide certain indication of the precision of the two estimators. It can be clearly observed that $\widehat{\zeta}_W$ is much more accurate than $\widehat{\zeta}_I$ in most of cases except for $\zeta = d$. In the cases of $\zeta = d$, i.e., the last stage in the example, there is no shift propagation and thus $\widehat{\zeta}_I$ should be a consistent estimator. This is not the case in real situations and hence $\widehat{\zeta}_I$ is significantly outperformed by $\widehat{\zeta}_W$. Figure 5 shows the comparison results with varying parameters (a_i, c_i) in a five-stage process and we can see the foregoing conclusion still holds.

6 An Illustration of the Implementation Steps: The AEC Monitoring Case Revisited

In this section, we revisit one of the motivating examples, the AEC manufacturing process, to demonstrate the implementation of the proposed schemes based on multistage process with binary data. As we mentioned, this process involves five stages: clenching & rolling, soaking, assembly, cleaning & bushing and aging & classifying. The quality of unfinished AEC products that are called capacitor elements are inspected by sampling after each stage. Therefore, we use the multivariate BSSM in (2) to model this process.

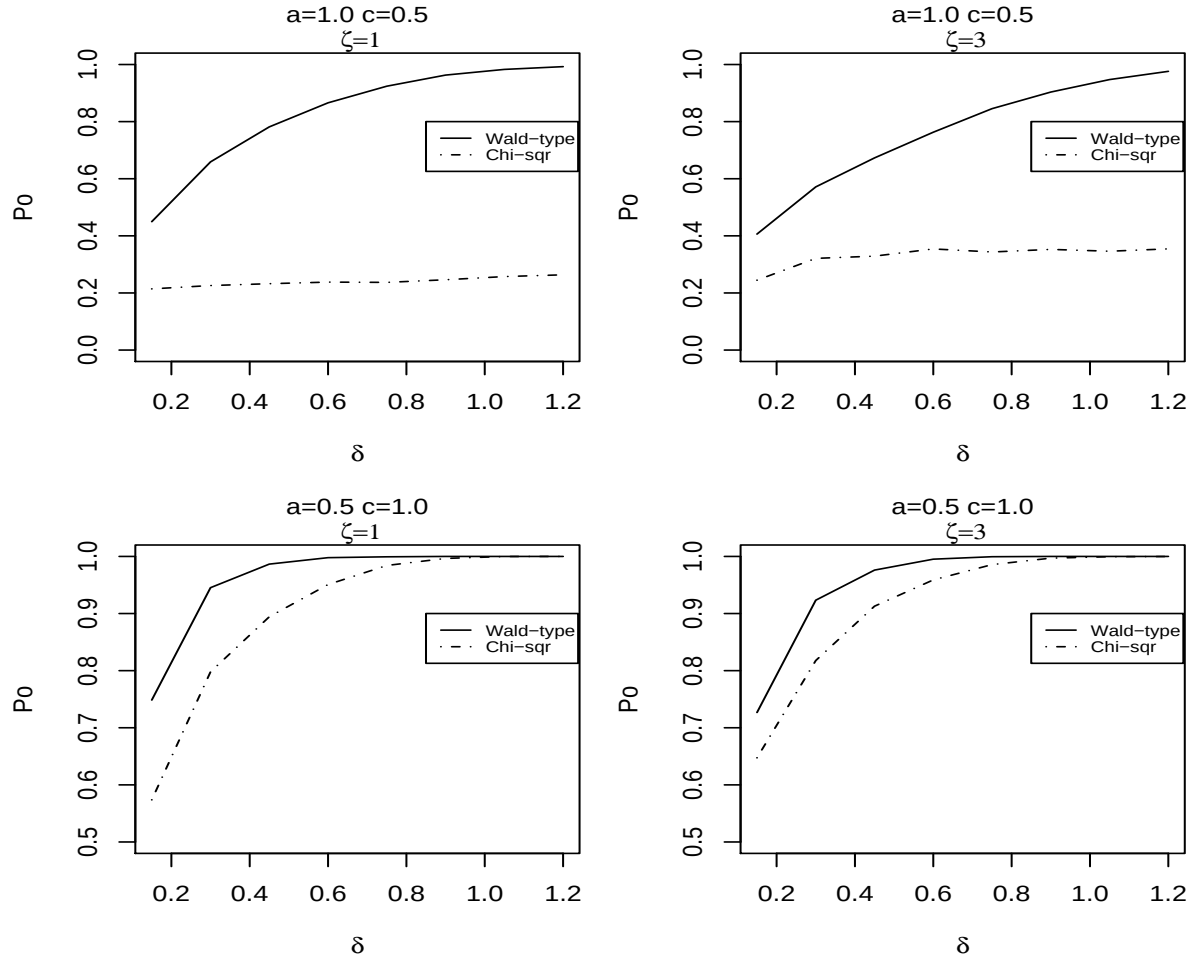


Figure 5: Diagnosis comparison with varying parameters for 5-stage process. The solid and dash-dotted lines represent the probabilities $\Pr(\widehat{\zeta}_W = \zeta)$ and $\Pr(\widehat{\zeta}_I = \zeta)$, respectively.

In this example, p_{ij} is the defect rate of the capacitor elements sample j at stage i. y_{ij} is the measurement of the number of defective capacitor elements in one sample with sample size n_i. $\boldsymbol{x}_{ij}$ is the quality characteristics in the jth sample at stage i, which is the sample mean of the measurements of capacitor elements. Due to the data availability in this factory, only one quality characteristic "compound thickness" is obtained in the clenching & rolling stage, and another quality characteristic "dissipation factor" is available in the soaking and aging stages. Although "compound thickness" is only measured in the first stage, the defects in "compound thickness" also affects the quality of capacitor elements in the next stages.

Therefore, without loss of generality, a three-stage multivariate model is used here to illustrate the estimation of the in-control model and monitoring and diagnosis of this process. The BSSM in

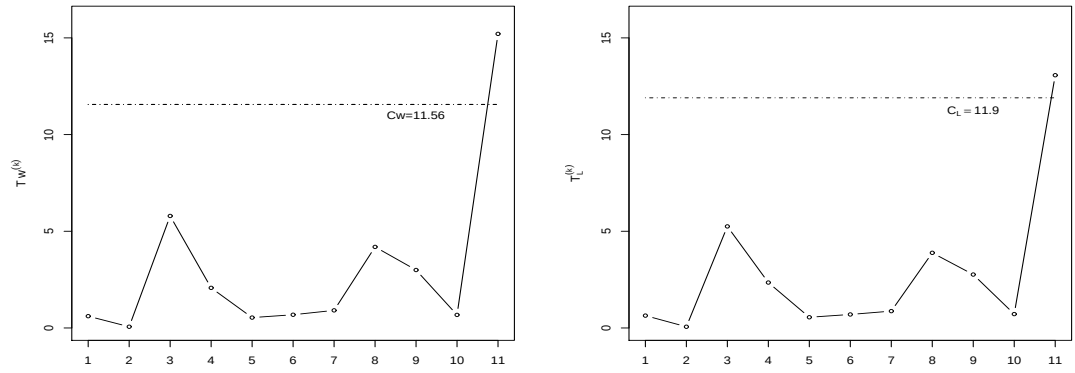
(2) for this process is described as follows:

$$\begin{aligned} &\text{logit}(p_{ij}) = \alpha_i + \boldsymbol{x}_{ij}^T \boldsymbol{c}_i, \\ &\boldsymbol{x}_{ij} = \boldsymbol{\beta}_i + \mathbf{A}_i \boldsymbol{x}_{i-1j} + \boldsymbol{\omega}_{ij}, \quad i = 1, 2, 3, \end{aligned} \tag{15}$$

where $\boldsymbol{x}_{ij}$ is the vector including x_{ij1} the sample mean of compound thickness and x_{ij2} the sample mean of dissipation factor of the capacitors, which are observable in the process design. Because the compound thickness is only measured in the first stage and the dissipation factor is appearing and measured in the next two stages, $x_{1j2} = 0$ and $x_{2j1} = x_{3j1} = x_{1j1}$. In addition, before the first stage, the above two quality characteristics do not exist, so we assume $\boldsymbol{x}_0 = 0$. Therefore, based on the historical observations $\boldsymbol{x}$ in each stage, the parameters in the second level equation of model (15), $\boldsymbol{\beta}_i$, $\mathbf{A}_i$ and $\boldsymbol{\Sigma}_{\omega_i}$, can be estimated except the arbitrarily given value "1.0" for the entries in the main diagonal of $\mathbf{A}_1$ and the entry in the bottom-right corner of $\mathbf{A}_2$. Similarly, based on the sample size $n_i = 100$, the measurements $\boldsymbol{y}$ and observed quality characteristics, $\boldsymbol{x}$, the parameters α_i and $\boldsymbol{c}_i$ can be estimated except the arbitrarily given value "1.0" for the second entry of $\boldsymbol{c}_1$, say c_{12},

$$\begin{aligned} &\boldsymbol{\alpha} = (-4.428, -4.403, -3.803)^T, \\ &\boldsymbol{c}_1 = \begin{pmatrix} -0.509 \\ 1.0 \end{pmatrix}, \quad \boldsymbol{c}_2 = \begin{pmatrix} -1.023 \\ 1.073 \end{pmatrix}, \quad \boldsymbol{c}_3 = \begin{pmatrix} -0.927 \\ 0.964 \end{pmatrix}, \\ &\boldsymbol{\beta}_1 = \begin{pmatrix} 0.3229 \\ 0.0 \end{pmatrix}, \quad \boldsymbol{\beta}_2 = \begin{pmatrix} 0 \\ 0.08574 \end{pmatrix}, \quad \boldsymbol{\beta}_3 = \begin{pmatrix} 0 \\ 0.004 \end{pmatrix}, \\ &\mathbf{A}_1 = \begin{pmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{pmatrix}, \quad \mathbf{A}_2 = \begin{pmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{pmatrix}, \quad \mathbf{A}_3 = \begin{pmatrix} 1.0 & 0.0 \\ 0.0 & 0.9 \end{pmatrix}, \\ &\boldsymbol{\Sigma}_{\omega_1} = \begin{pmatrix} 2.05 \times 10^{-6} & 0.0 \\ 0.0 & 0.0 \end{pmatrix}, \quad \boldsymbol{\Sigma}_{\omega_2} = \begin{pmatrix} 0.0 & 0.0 \\ 0.0 & 1.39 \times 10^{-3} \end{pmatrix}, \quad \boldsymbol{\Sigma}_{\omega_3} = \begin{pmatrix} 0.0 & 0.0 \\ 0.0 & 8.20 \times 10^{-5} \end{pmatrix}. \end{aligned}$$

Based on this process model, we simulate new observation matrix, $\boldsymbol{Y}^{(k)}$. In each matrix, we have 20 observation vectors $\boldsymbol{y}_j^{(k)}, j = 1, 2, \ldots, 20$. The first 10 observation matrix are from normal conditions and the latter are from the out-of-control condition with a mean shift of $\delta = 0.1$ at "dissipation factor" (about 3σ) in stage 2.



(a) (b)

Figure 6: Wald-type chart (a) and GHLR chart (b) for the AEC example

Here, we illustrate how to implement our proposed schemes (GHLR and Wald-type) to monitor the AEC manufacturing process and diagnose the stage in which the mean shift initially occurs:

1. Obtain shift direction information vectors, $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$:

$$\Psi_1 = \begin{pmatrix} -0.509 & 1.000 \\ -1.023 & 1.073 \\ -0.927 & 0.868 \end{pmatrix}, \quad \Psi_2 = \begin{pmatrix} 0.0 & 0.0 \\ -1.023 & 1.073 \\ -0.927 & 0.868 \end{pmatrix}, \quad \Psi_3 = \begin{pmatrix} 0.0 & 0.0 \\ 0.0 & 0.0 \\ -0.927 & 0.964 \end{pmatrix}$$

2. Obtain the control limits, C_L and C_W, for GHLR and Wald-type control schemes, respectively, by simulation given the desired in-control ARL. Here, we obtain $C_L = 11.90$ and $C_W = 11.56$ for ARL_0=200 ($\alpha = 0.005$), and then construct a control chart as in Figure 6.

3. Start the monitoring process and obtain new observations, y_{ij}. After obtaining the new observations, we calculate the control statistics by equations (9) and (11), and then plot these control statistics in control charts and compare them with control limits. In Figure 6, we can see that both the GHLR and Wald-type charts signal at the 11th time point.

4. Identify the shifting stage, ζ, by equations (10) and (12). Here, we easily estimate $\hat{\zeta} = 2$.

5. Identify and remove root causes after diagnosing the shifting stages, and then go back to step 1. Monitor the process continuously based on the revised control limits.

7 Conclusion and Considerations

In this paper, a binary state space model is proposed for modeling the multistage process with binary outputs. Statistical process control including monitoring and diagnosing is very important and challenging for multistage processes with categorical data. Therefore, based on BSSM for multistage processes with binary data, we develop two new control schemes, the GHLR scheme and the Wald-type scheme, by utilizing h-likelihood estimation and directional information. These two new schemes not only monitor the mean shift in multistage processes, but also diagnose where the mean shift initially occurs. According to simulation results in Section 5, our proposed schemes, the GHLR scheme and the Wald-type scheme, demonstrate higher power in detecting mean shifts in processes than does the existing χ^2 scheme that is often used for multivariate categorical data. In addition to the monitoring performance, the GHLR scheme and the Wald-type scheme can easily identify the shifting stage where the mean shift initially occurs. As shown in Section 5, under different process parameters, a_i and c_i, our proposed schemes consistently outperform the existing approach, the χ^2 scheme. The simulation results show that the GHLR scheme and the Wald-type scheme perform almost the same in monitoring and diagnosing. However, the Wald-type scheme seems to be a little superior to the GHLR scheme when detecting the mean shift in processes. Therefore, the Wald-type scheme is recommended in practice.

In future research, different types of control schemes from these Shewhart-type schemes can be investigated to sensitively detect the small mean shifts, such as the EWMA-type control scheme. Moreover, the multistage process with mixed types of data commonly exists in real industrial practice, in which some stages have numerical outputs and some stages have categorical outputs. One of the limitations is our proposed methods cannot be directly applied in such processes with mixed types of outputs. In addition, in some 100% inspection applications, a product is found to be defective in an early stage, it is desired to be removed from the process rather than going through the rest of the stages. Clearly, the proposed methods would not be efficient for such cases. One of our ongoing research aims to develop the model and control schemes for multistage processes with mixed types of data and for 100% inspection cases.

Acknowledgement

The authors would like to thank the Department 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 National Natural Science Foundation of China (Nos. 70931004, 70871087, 11001138, 11071128, 11131002, and 11101306). Zou also thanks for the support of RFDP of China

Grant No. 20110031110002 and National Center for Theoretical Sciences, Mathematics Division.

Appendix

Appendix A: Derivation of Eq. (14)

Denote $\boldsymbol{\eta} = (\boldsymbol{\eta}_0^T, \boldsymbol{\eta}_1^T)^T$, $\boldsymbol{u} = (\boldsymbol{u}_0^T, \boldsymbol{u}_1^T)^T$ and for $l = 0, 1$,

$$\boldsymbol{\eta}_l^T = (\boldsymbol{\eta}_{l1}^T, \ldots, \boldsymbol{\eta}_{lm}^T), \quad \boldsymbol{\eta}_{lj}^T = (\eta_{l1j}^T, \ldots, \eta_{ldj}^T),$$

$$\eta_{0ij} = \text{logit}(p_{ij}), \quad \eta_{1ij} = \nu_{ij}, \quad \boldsymbol{u}_0^T = (\boldsymbol{\mu}_0^T, \ldots, \boldsymbol{\mu}_0^T), \quad \boldsymbol{u}_1 = 0.$$

By a similar arguments in Section 6.2.3 of Lee et al. (2006), it is straightforward to see that the MHLE for $\boldsymbol{\xi}$ can be viewed as that for an augmented GLM with the response variables $(\boldsymbol{y}^T, \boldsymbol{\phi}^T)^T$ and the linear predictor

$$\boldsymbol{\eta} - \boldsymbol{u} = \boldsymbol{\Gamma}\boldsymbol{\xi},$$

where $\boldsymbol{y} = (\boldsymbol{y}_1^T, \ldots, \boldsymbol{y}_m^T)^T$ and $\boldsymbol{\phi}^T = (\boldsymbol{\phi}_1^T, \ldots, \boldsymbol{\phi}_m^T)$ is quasi data satisfying $E(\boldsymbol{\phi}_j) = \boldsymbol{\nu}_j$ and $\text{Var}(\boldsymbol{\phi}_j) = \boldsymbol{\Sigma}$. According to the standard procedure for the GLM estimation (McCullagh and Nelder 1989), we immediately have the IWLS equation (14), where we use the fact that

$$z_{0ij} = \eta_{0ij} - \mu_{0i} - (y_{ij} - n_i p_{ij}) \frac{\partial \eta_{0ij}}{\partial (n_i p_{ij})},$$

$$\boldsymbol{\Omega} = \boldsymbol{\Sigma}_a \mathbf{W}^{-1}, \quad \boldsymbol{\Sigma}_a = \text{diag}\{\mathbf{I}_{md}, \mathbf{I}_m \otimes \boldsymbol{\Sigma}\},$$

$$\mathbf{W} = \text{diag}\{\mathbf{W}_0, \mathbf{W}_1\}, \quad \mathbf{W}_l = \text{diag}\big\{\mathbf{W}_{l1}, \ldots, \mathbf{W}_{lm}\big\},$$

$$\mathbf{W}_{lj} = \text{diag}\{w_{l1j}, \ldots, w_{ldj}\}, \quad \text{for} \quad l = 0, 1,$$

$$w_{0ij} = \left(\frac{\partial (n_i p_{ij})}{\partial \eta_{0ij}}\right)^2 [\text{Var}(y_{ij}|p_{ij})]^{-1}, \quad w_{1ij} = 1.$$

Appendix B: Computational Simplification of (15)

Let $\mathbf{A} = \boldsymbol{\Gamma}^T \boldsymbol{\Omega}^{-1} \boldsymbol{\Gamma}$, which can be written as follows,

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}_{(q+md)\times(q+md)} = \begin{bmatrix} (\boldsymbol{\psi}^{*T}\mathbf{W}_0\boldsymbol{\psi}^*)_{q\times q} & (\boldsymbol{\psi}^{*T}\mathbf{W}_0)_{q\times md} \\ (\mathbf{W}_0\boldsymbol{\psi}^*)_{md\times q} & (\mathbf{W}_0 + \mathbf{I}_m \otimes \boldsymbol{\Sigma}^{-1})_{md\times md} \end{bmatrix}$$

where $\mathbf{A}_{22}$ can be further expressed as

$$\operatorname{diag}\{\boldsymbol{\Sigma}^{-1} + \mathbf{W}_{01}, \ldots, \boldsymbol{\Sigma}^{-1} + \mathbf{W}_{0m}\}.$$

Then, using the classical formula for the inverse of block matrix, we have

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}_{11\cdot 2}^{-1} & -\mathbf{A}_{11\cdot 2}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{A}_{11\cdot 2}^{-1} & \mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{A}_{11\cdot 2}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} + \mathbf{A}_{22}^{-1} \end{bmatrix},$$

where $\mathbf{A}_{11\cdot 2} = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}$. Hence, $\mathbf{A}^{-1}$ could be easily calculated as long as we obtain $\mathbf{A}_{22}^{-1}$ and $\mathbf{A}_{11\cdot 2}^{-1}$. Note that

$$\mathbf{A}_{22}^{-1} = \operatorname{diag}\{[\boldsymbol{\Sigma}^{-1} + \mathbf{W}_{01}]^{-1}, \ldots, [\boldsymbol{\Sigma}^{-1} + \mathbf{W}_{0m}]^{-1}\},$$

and considering that the matrix $\boldsymbol{\Sigma}^{-1}$ is fixed and can be determined before monitoring, the computation task here is trivial by virtue of the massive computing and data storage capabilities of modern computers.

Appendix C: The Estimator of $\operatorname{Var}(\widehat{\boldsymbol{\delta}}_i)$

By using the results in Section 3.3 of Lee and Nelder (1996) and the preceding matrix expressions in Appendices A and B, we have the following consistent estimate for $\operatorname{Var}(\widehat{\boldsymbol{\delta}}_i)$:

$$\widehat{\operatorname{Var}}(\hat{\boldsymbol{\delta}}) = A_{11\cdot 2}^{-1}.$$

Reference

Acosta-Mejia, C. A. (1999) Improved p Charts to Monitor Process Quality. *IIE Transactions*, **31**, 509–516.

Barad, M., and Braha, D. (1996) Control Limits for Multi-Stage Manufacturing Processes with Binomial Yield (Single and Multiple Production Runs). *The Journal of the Operational Research Society*, **47**, 98–112.

Basseville, M., and Nikiforov, I. (1993) *Detection of Abrupt changes: Theory and Application*, Prentice-Hall, Englewood Cliffs, NJ.

Bourke, P. D. (1991) Detecting a Shift in Fraction Nonconforming Using Run-Length Control Charts with 100% Inspection. *Journal of Quality Technology*, **23**, 225–238.

Bourke, P. D. (2001) The Geometric CUSUM chart with Sampling Inspection for Monitoring Fraction Defective. *Journal of Applied Statistics*, **28**, 951–972.

Breslow, N. E., and Clayton, D. G. (1993) Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association*, **88**, 9–25.

Carlin, B. P., Polson, N. G., and Stoffer, D. S. (1992) A Monte Carlo Approach to Nonnormal and Nonlinear State-Space Modeling. *Journal of the American Statistical Association*, **87**, 493–500.

Coull, B. A., and Agresti, A. (2000) Random Effects Modeling of Multiple Binomial Responses Using the Multivariate Binomial Logit-Normal Distribution. *Biometrics*, **56**, 73–80.

Czado, C., and Song, P. X. (2008) State Space Mixed Models for Longitudinal Observations with Binary and Binomial Responses. *Statistical Papers*, **49**, 691–714.

Davidian, M., and Giltinan, D. M. (1995) *Nonlinear Models for Repeated Measurement Data*, Chapman and Hall, London.

Fahrmeir, L., and Tutz, G. (1994) *Multivariate Statistical Modelling Based on Generalized Linear Models*, Springer-Verlag, New York.

Gadre, M. P., and Rattihalli, R. N. (2005) Unit and Group-Runs Chart to Identify Increases in Fraction Nonconforming. *Journal of Quality Technology*, **37**, 199–209.

Gan, F. F. (1993) An Optimal Design of CUSUM Control Charts for Binomial Counts. *Journal of Applied Statistics*, **20**, 445–460.

Gueorguieva, R. (2001) A Multivariate Generalized Linear Mixed Model for Joint Modelling of Clustered Outcomes in the Exponential Family. *Statistical Modelling*, **1**, 177–193.

Jiang, W. (2004) Multivariate Control Charts for Monitoring Autocorrelated Processes. *Journal of Quality Technology*, **36**, 367–379.

Jin, J., and Shi, J. (1999) State space modeling of sheet metal assembly for dimensional control. *ASME Transactions, Journal of Manufacturing Science and Engineering*, **121**, 756–762.

Jin, R., Li, J., and Shi, J. (2007) Quality Prediction And Control in Rolling Processes Using Logistic Regression. *Transactions of NAMRI/SME*, **35**, 113–120.

Jin, M., and Tsung, F. (2009) A Chart Allocation Strategy for Multistage Processes. *IIE Transactions*, **41**, 790–803.

Kumar, Y., Kennedy, K., and Gildersleeve, K. (2006) A Review of Yield Modeling Techniques for Semiconductor Manufacturing. *International Journal of Production Research*, **44**, 5019–5036.

Lee, Y., and Nelder, J. A. (1996) Hierarchical Generalized Linear Models (with discussion). *Journal of the Royal Statistical Society, Series B*, **58**, 619–678.

Lee, Y., Nelder, J. A., and Pawitan, Y. (2006) *Generalized Linear Models with Random Effects—Unified Analysis via H-Likelihood*, Boca Raton, Chapman & Hall/CRC, FL.

Lee, H. L., and Yano, C. A. (1988) Production Control in Multistage Systems with Variable Yield Losses. *Operations Research*, **36**, 269–278.

Li, Y., and Tsung, F. (2009) False Discovery Rate-Adjusted Charting Schemes for Multistage Process Monitoring and Fault Identification. *Technometrics*, **51**, 186–205.

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

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

Mason, R. L., Tracy, N. D., and Young, J. C. (1995) Decomposition of T^2 for Multivariate Control Chart Interpretation. *Journal of Quality Technology*, **27**, 99–108.

McCulloch, C. E., and Searle, S. R. (2001) *Generalized Linear and Mixed Models*, Wiley, New York.

McCullagh, P., and Nelder, J. A. (1989) *Generalized Linear Models*, Chapman and Hall, London.

Ning, X., Shang, Y., and Tsung, F. (2009) Statistical Process Control Techniques for Service Processes: a review. *6th International Conference on Service Systems and Service Management*, Xiamen, China, 927–931.

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

Rabe-Hesketh, S., and Skrondal, A. (2001) Parameterization of Multivariate Random Effects Models for Categorical Data. *Biometrics*, **57**, 1256–1264.

Reynolds, M. R., and Stoumbos, Z. G. (1999) A CUSUM Chart for Monitoring a Proportion when Inspecting Continuously. *Journal of Quality Technology*, **31**, 87–108.

Reynolds, M. R., and Stoumbos, Z. G. (2000) A General Approach to Modeling CUSUM Charts for a Proportion. *IIE Transactions*, **32**, 515–535.

Shi, J. (2007) *Stream of Variation Modeling and Analysis for Multistage Manufacturing Processes*, CRC Press, Boca Raton, FL.

Shi, J., and Zhou, S. (2009) Quality control and improvement for multistage systems: A survey. *IIE Transactions*, **41**, 744–753.

Somerville, S. E., Montgomery, D. C., and Runger, G. C. (2002) Filtering and Smoothing Methods for Mixed Particle Count Distributions. *International Journal of Production Research*, **40**, 2991–3013.

Steiner, S. H. (1998) Grouped Data Exponentially Weighted Moving Average Control Charts. *Applied Statistics*, **47**, 203–216.

Tang, C. S. (1990) The Impact of Uncertainty on a Production Line. *Management Science*, **36**, 1518–1531.

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

Wu, Z., and Spedding, T. A. (2001) A Synthetic Control Chart for Detecting Fraction Nonconforming Increases. *Journal of Quality Technology*, **33**, 104–111.

Xiang, L., and Tsung, F. (2008) Statistical Monitoring of Multistage Processes Based on Engineering Models. *IIE Transactions*, **40**, 957–970.

Yun, S., and Lee, Y. (2004) Comparison of Hierarchical and Marginal Likelihood Estimators for Binary Outcomes. *Computational Statistics and Data Analysis*, **45**, 639–650.

Zhou, S., Ding, Y., Chen, Y., and Shi, J. (2003) Diagnosability Study of Multistage Manufacturing Processes Based on Linear Mixed-Effects Models. *Technometrics*, **45**, 312–325.

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

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

Zou, C., Zhang, Y., and Wang, Z. (2006) Control Chart Based on Change-Point Model for Monitoring Linear Profiles. *IIE Transactions*, **38**, 1093–1103.

Table 2: Charting performance comparison with parameters (a_i, c_i)=(1.0,1.0) for $d = 5$.

δ	$T_L^{(k)}$	$T_W^{(k)}$	χ^2	$T_L^{(k)}$	$T_W^{(k)}$	χ^2
	shift in 1st stage			shift in 2nd stage		
0.15	0.042	0.042	0.039	0.036	0.037	0.035
	(0.0020)	(0.0020)	(0.0019)	(0.0019)	(0.0019)	(0.0018)
0.30	0.248	0.251	0.224	0.215	0.215	0.182
	(0.0043)	(0.0043)	(0.0042)	(0.0041)	(0.0041)	(0.0039)
0.45	0.677	0.684	0.622	0.608	0.609	0.537
	(0.0047)	(0.0047)	(0.0048)	(0.0049)	(0.0049)	(0.0050)
0.60	0.940	0.947	0.924	0.909	0.913	0.870
	(0.0024)	(0.0022)	(0.0026)	(0.0029)	(0.0028)	(0.0034)
0.75	0.967	0.969	0.994	0.984	0.987	0.984
	(0.0018)	(0.0017)	(0.0008)	(0.0013)	(0.0011)	(0.0013)
	shift in 3rd stage			shift in 5th stage		
0.15	0.023	0.024	0.022	0.019	0.019	0.016
	(0.0015)	(0.0015)	(0.0015)	(0.0013)	(0.0014)	(0.0013)
0.30	0.155	0.157	0.117	0.096	0.103	0.070
	(0.0036)	(0.0036)	(0.0032)	(0.0029)	(0.0030)	(0.0025)
0.45	0.512	0.514	0.428	0.342	0.355	0.272
	(0.0050)	(0.0050)	(0.0049)	(0.0047)	(0.0048)	(0.0044)
0.60	0.856	0.857	0.779	0.698	0.710	0.598
	(0.0035)	(0.0035)	(0.0042)	(0.0046)	(0.0045)	(0.0049)
0.75	0.983	0.983	0.962	0.929	0.934	0.872
	(0.0013)	(0.0013)	(0.0019)	(0.0026)	(0.0025)	(0.0033)

Note: values in parentheses are the standard errors

Table 3: Diagnosis performance comparison for $d = 3$ with various parameter combinations.

	$\zeta = 1$		$\zeta = 2$		$\zeta = 3$		$\zeta = 1$		$\zeta = 2$		$\zeta = 3$	
δ	$\widehat{\zeta}_W$	$\widehat{\zeta}_I$	$\widehat{\zeta}_W$	$\widehat{\zeta}_I$	$\widehat{\zeta}_W$	$\widehat{\zeta}_I$	$\widehat{\zeta}_W$	$\widehat{\zeta}_I$	$\widehat{\zeta}_W$	$\widehat{\zeta}_I$	$\widehat{\zeta}_W$	$\widehat{\zeta}_I$
	$(a_i, c_i) = (1.0,1.0)$						$(a_i, c_i) = (1.0,0.5)$					
0.15	0.623	0.351	0.574	0.549	0.619	0.753	0.485	0.318	0.446	0.428	0.477	0.538
0.30	0.846	0.421	0.756	0.509	0.833	0.929	0.643	0.308	0.563	0.540	0.614	0.727
0.45	0.921	0.434	0.841	0.547	0.901	0.959	0.731	0.327	0.650	0.546	0.731	0.835
0.60	0.962	0.443	0.925	0.568	0.951	0.982	0.808	0.350	0.726	0.546	0.777	0.918
0.75	0.983	0.459	0.963	0.574	0.969	0.988	0.861	0.360	0.779	0.555	0.831	0.955
	$(a_i, c_i) = (0.5,1.0)$						$(a_i, c_i) = (0.5,0.5)$					
0.15	0.727	0.562	0.672	0.684	0.692	0.692	0.573	0.410	0.433	0.507	0.472	0.456
0.30	0.900	0.744	0.872	0.785	0.888	0.919	0.726	0.525	0.675	0.679	0.694	0.675
0.45	0.964	0.835	0.939	0.865	0.958	0.966	0.847	0.664	0.794	0.736	0.820	0.840
0.60	0.983	0.890	0.971	0.918	0.980	0.988	0.908	0.740	0.865	0.800	0.892	0.911
0.75	0.991	0.934	0.987	0.948	0.991	0.996	0.935	0.791	0.908	0.850	0.936	0.948

Table 4: Diagnosis performance comparison with parameters (a_i, c_i)=(1.0,1.0) for $d = 5$.

	ζ =1				$\zeta = 3$				$\zeta = 5$			
	P_0		P_1		P_0		P_1		P_0		P_1	
δ	$\widehat{\zeta}_W$	$\widehat{\zeta}_{\chi^2}$	$\widehat{\zeta}_W$	$\widehat{\zeta}_{\chi^2}$	$\widehat{\zeta}_W$	$\widehat{\zeta}_{\chi^2}$	$\widehat{\zeta}_W$	$\widehat{\zeta}_{\chi^2}$	$\widehat{\zeta}_W$	$\widehat{\zeta}_{\chi^2}$	$\widehat{\zeta}_W$	$\widehat{\zeta}_{\chi^2}$
0.15	0.724	0.217	0.893	0.391	0.505	0.403	0.787	0.727	0.655	0.653	0.758	0.744
0.30	0.891	0.230	0.968	0.436	0.807	0.404	0.944	0.777	0.884	0.885	0.938	0.928
0.45	0.960	0.267	0.992	0.477	0.917	0.422	0.981	0.790	0.957	0.960	0.984	0.975
0.60	0.985	0.284	0.997	0.499	0.966	0.423	0.992	0.804	0.980	0.972	0.993	0.986
0.75	0.990	0.298	0.997	0.525	0.985	0.451	0.995	0.838	0.992	0.984	0.998	0.992