A new \(p\)-control chart with measurement error correction

Li-Pang Chen | Su-Fen Yang

Department of Statistics, National Chengchi University, Taipei, Taiwan

Correspondence
Su-Fen Yang, Department of Statistics, National Chengchi University, Taipei, Taiwan.
Email: yang@mail2.nccu.tw

Abstract
Control charts are important tools to monitor quality of products. One of useful applications is to monitor the proportion of non-conforming products. However, in practical applications, measurement error is ubiquitous and may occur due to false records or misclassification, which makes the observed proportion different from the underlying true proportion. It is also well-known that ignoring measurement error effects provides biases, and is expected that the resulting control charts may incur wrong detection. In this paper, we study this important problem and propose a valid method to correct for measurement error effects and obtain error-eliminated control chart for the proportion of non-conforming products. In addition, unlike traditional approaches, the corrected EWMA \(p\)-control chart provides asymmetric control limits and is flexible to handle the data with small sample size. Numerical results are conducted to justify the validity of the corrected EWMA \(p\)-control chart and verify the necessity of measurement error correction.

KEYWORDS
asymmetric control limits, average run length, error elimination; misclassification, monitoring, \(p\)-control charts

1 INTRODUCTION

Statistical process control (SPC) is an important tool in industrial statistics and is useful in monitoring the quality of products. The main interest is to develop control charts based on in-control (IC) statistic, and then use them to monitor and detect out-of-control (OC) process parameters. To do this, one may focus on location process or dispersion process. For example, as summarized in Qiu\(^1\), many classical methods have been developed under the parametric settings, including Shewhart charts, exponentially weighted moving average (EWMA) charts and cumulative sum (CUSUM) charts. Without imposing parametric assumptions, distribution free settings have been popular in recent years and relevant approaches have been developed in recent years, such as EWMA mean charts (e.g., Yang\(^2\)), variability monitoring (e.g., Yang and Arnold\(^3\), Yang and Wu\(^4\)) and likelihood ratio-based EWMA control charts (e.g., Zou and Tsung\(^5\)). In addition, non-parametric methods have been explored as well. For example, Chen et al.\(^6\) employed the kernel estimation method to construct the control region to monitor location and dispersion processes simultaneously.

In practice, noisy data usually exist and are inevitable in producing products. A typical phenomenon in manufacturer factories is measurement error, which reflects a fact that observed data are different from underlying true ones that are unobservable. In the presence of continuous random variables, some existing works have examined impacts of measurement error to different kinds of control charts, such as EWMA charts (e.g., Maravelakis et al.\(^7\), Asif et al.\(^8\), Nguyen et al.\(^9\), Tran et al.\(^10\)), Shewhart control charts (e.g., Nguyen et al.\(^9\), Linna and Woodall\(^11\)), multivariate process...
variability (e.g., Huwang and Hung\cite{12}), first-order autoregressive model (e.g., Shongwe et al.\cite{13}) and multivariate control charts (e.g., Linna et al.\cite{14}). However, while those works discussed measurement error effects on control limits, to the best of knowledge, none of them provided suitable strategies to correct for measurement error effects and adjust control limits.

In addition to continuous random variables, sometimes we are interested in the number/proportion of (non-)confirming products. Such a scenario refers binary random variables. Under the assumption of parametric distributions, the strategies of Shewhart charts have been adopted to construct $p$-control charts (e.g., Qiu,\cite{1} Section 3.3). Some relevant extensions have also been explored, such as quasi ARL-unbiased $p$-charts based on a heuristic method (e.g., Argoti and Carrión-García\cite{15}), the risk adjustment control charts for categorical random variables (e.g., Sparks\cite{16}) and the general framework of distribution free settings (e.g., Yang and Arnold,\cite{17} Aslam et al.\cite{18}). Some applications of $p$-control charts have also been discussed. For example, Shu and Wu\cite{19} applied $p$-control charts to monitor imprecise fraction of conforming items. Bourke\cite{20} extended the $np$ chart for detecting upward shifts in fraction defective. While the development of $p$-control chart has been widely discussed, however, there are several critical concerns in applications. First, similar to measurement error in continuous random variables, it is possible to encounter measurement error in binary random variables, which refers misclassification (e.g., Yi\cite{21}, Section 2.6; Chen and Yi\cite{22}). More specifically, when producing products in factories, sometimes they may be falsely detected as confirmed (or non-confirmed) due to imprecise measurement equipment or human-made mistakes. Because of such a misclassification, the observed status is different from what it should be. While measurement error in continuous random variables has been widely discussed, misclassification problem in SPC has been rarely discussed, and strategies for correction of measurement error effects are unavailable. The second concern is the small sample size in each monitoring time. It is known that, by the central limit theorem, the estimated proportion follows normal distributions when the sample size goes to the infinity. However, in the presence of small sample size, the sampling distribution is unknown, and thus, the construction of control charts is unknown.

Due to those concerns, in this paper, we focus on SPC for binary random variables subject to misclassification. We particularly focus on phase II scenario where the parameter is assumed to be known. Under small sample sizes, we employ EWMA charts to construct the asymmetric control charts. To deal with measurement error effects, we propose the ‘corrected’ proportion of non-confirming products, so that the corresponding control limits can be adjusted accordingly. To assess the performance of the corrected EWMA $p$-control chart, we compare the corrected EWMA $p$-control chart with the naive (uncorrected) EWMA $p$-control chart that simply adopt error-contaminated random variables. Numerical experiments show that the corrected EWMA $p$-control chart can precisely detect OC proportion.

The remainder is organized as follows. In Section 2, we introduce the data structure and the EWMA $p$-control chart to monitor non-confirming products. In addition, measurement error to binary random variables is introduced. In Section 3, we propose a valid estimation procedure to correct the measurement error effects and adopt the EWMA statistic to construct the asymmetric corrected EWMA $p$-control chart. Moreover, we present IC average run length ($ARL_0$) and OC average run length ($ARL_1$) to illustrate the OC detection performance. Empirical studies, including simulation results and real data analysis, are provided in Sections 4 and 5, respectively. We conclude the article with discussions in Section 6. The R programming code for the implementation is available in the Github: https://github.com/lchen723/SPC-ME-R-code.git.

2 | DATA STRUCTURE AND MISMEASUREMENT

In this section, we introduce the construction of the control chart with relevant notation, and discuss the issue of measurement error in binary random variables.

2.1 | Binary variables and construction of the $p$-control chart

Let $T$ denote the monitoring time. For each time $t = 1, \ldots, T$, there are $n$ subjects. Let $X_{it}$ for $i = 1, \ldots, n$ and $t = 1, \ldots, T$ denote the independent and identically distributed (i.i.d.) binary random variable with outcome 0 or 1, where $X_{it} = 1$ represents a non-conforming product, while $X_{it} = 0$ stands for a conforming product. The main interest is to monitor the proportion of non-conforming products for IC process.
Let $p_0 \triangleq P(X_{it} = 1)$ be the parameter of proportion of non-conforming products for IC process, and let $q_0 \triangleq 1 - p_0 = P(X_{it} = 0)$ denote the probability of conforming products for IC process. Under IC data with $n$ subjects for $T$ monitoring times, we define the sample proportion $\hat{p}_{0,t} \triangleq \frac{1}{n} \sum_{i=1}^{n} X_{it}$.

To build up the reliable control chart, one may employ the EWMA control chart, which is constructed based on a weighted average of IC sample proportion at the current time point. In addition, as commented in Yang,\(^2\) the EWMA chart is useful in monitoring since it is sensitive in detecting small shifts in process parameters.

Based on the binary random variable and the sample proportion $\hat{p}_{0,t}$, the in-control EWMA statistic is given by

$$\text{EWMA}_{0,t} = \lambda \hat{p}_{0,t} + (1 - \lambda) \text{EWMA}_{0,t-1},$$

where $\text{EWMA}_{0,0} = p_0$ with $p_0$ being the IC proportion, and $\lambda \in (0, 1]$ is a smoothing parameter. Moreover, when the monitoring time goes by, the OC non-conforming proportion is non-decreasing and is greater than IC non-conforming proportion. Thus, in applications, it is reasonable to set the lower control limit (LCL) to be zero, and the UCL is specified as

$$\text{UCL} = p_0 + L \sqrt{\frac{p_0(1 - p_0) \lambda(1 - (1 - \lambda)^2 t)}{n(2 - \lambda)}},$$

where $L$ is the coefficient of UCL. Now, and hereafter, we call this strategy the $\text{EWMA}_{p}$-control chart.

### 2.2 Misclassification

In practice, variables are often subject to mismeasurement. That is, instead of collecting unobserved variables $X_{it}$, sometimes we have only the observed version of $X_{it}$, denoted by $X_{it}^*$. The relationship between unobserved and observed variables $X_{it}, X_{it}^* \in \{0, 1\}$ can be characterized as $\pi_{kl} = P(X_{it}^* = k | X_{it} = l)$ for $k, l = 0, 1$. Specifically, $\pi_{00}$ and $\pi_{11}$, which are called classification probability, indicate that both $X_{it}$ and $X_{it}^*$ are either conforming or non-conforming products; on the other hand, $\pi_{01}$ or $\pi_{10}$ shows that the unobserved variable is opposite to the observed variable which is caused by measurement error. In this case, we call $\pi_{10}$ and $\pi_{01}$ as misclassification probability (e.g., Carroll et al.,\(^23\) Chen and Yi\(^22\)).

To further understand the concept of misclassification, we can think of $X_{it}$ as the true status of product with $X_{it} = 1$ being non-conforming and 0 otherwise, which is unknown. On the other hand, $X_{it}^*$ is understood as the observed status of product that is recorded by factory’s staffs. Therefore, $\pi_{01}$ (or $\pi_{10}$) is interpreted as the probability that a non-conforming (or conforming) product is falsely recorded as conforming (or non-conforming).

Let $p_0^* = P(X_{it}^* = 1)$ and $q_0^* = 1 - p_0^* = P(X_{it}^* = 0)$ denote the IC observed version of $p_0$ and $q_0$, respectively. To build up the relationship of $p_0^*$ and $p_0$, we apply the technique of law of total probability and obtain that

$$p_0^* = \pi_{11} p_0 + \pi_{10} q_0.$$

Similarly, $q_0^*$ and $q_0$ can be characterized as

$$q_0^* = \pi_{01} p_0 + \pi_{00} q_0.$$

The matrix form of Equations (3) and (4) is given by

$$\begin{pmatrix} p_0^* \\ q_0^* \end{pmatrix} = \Pi \begin{pmatrix} p_0 \\ q_0 \end{pmatrix},$$

where $\Pi = \begin{pmatrix} \pi_{11} & \pi_{10} \\ \pi_{01} & \pi_{00} \end{pmatrix}$ is the $2 \times 2$ (mis)classification matrix and the sum of elements in each column is equal to one, that is, $\pi_{11} + \pi_{01} = 1$ and $\pi_{10} + \pi_{00} = 1$. Following the similar discussion in Chen and Yi,\(^22\) we assume that $\Pi$ has the
spectral decomposition $\mathbf{II} = \mathbf{D}\mathbf{O}\mathbf{O}^{-1}$, where $\mathbf{D}$ is the diagonal matrix with diagonal elements being the eigenvalues of $\mathbf{II}$, and $\mathbf{O}$ is the corresponding matrix of eigenvectors.

Since our main target is to monitor $p_0$, however, from the equality (5), we observe that the observed version $p_0^*$ is no longer equal to $p_0$ if misclassification occurs, that is, $\pi_{10} \neq 0$ or $\pi_{01} \neq 0$. In particular, when $\pi_{10}$ or $\pi_{01}$ become larger, $p_0^*$ and $q_0^*$ are different from $p_0$ and $q_0$. Therefore, with the availability of $X_{it}^*$, the sample proportion $\tilde{p}_{0,i}^* = \frac{1}{n}\sum_{i=1}^{n} X_{it}^*$ and $\tilde{q}_{0,i}^* = 1 - \tilde{p}_{0,i}^*$, have biases for $p_0$ and $q_0$, and the corresponding control limits, determined by Equation (1) with $\tilde{p}_{0,i}$ replaced by $\tilde{p}_{0,i}^*$, may incur wrong detection.

### 2.3 Determination of classification matrix

While the probability of observed non-conforming or conforming products can be expressed as Equation (5), the (mis)classification matrix $\mathbf{II}$ is usually unknown because of involvement of the unobserved variable $X_{it}$. Therefore, to develop the method, we consider sensitivity analyses (e.g., Chen, 24 Chen and Yi 22), whose purpose is to specify different values of $\mathbf{II}$ to understand how mismeasurement effects may affect inference results and control charts, and it is usually employed when additional information is unavailable.

Since the specification of $\mathbf{II}$ may be not unique, in this paper, we consider the relative ratio (RR) to specify misclassification probabilities. Specifically, let the RR under $X_{it} = 1$ be

$$RR_1 = \frac{\pi_{11}}{1 - \pi_{11}}.$$  

(6)

where $RR_1$ is in an interval $[0, \infty)$. A larger $RR_1$ indicates a higher probability of correct classification. Then given $RR_1$, $\pi_{11}$ and $\pi_{01}$ can be derived as $\pi_{11} = \frac{RR_1}{1 + RR_1}$ and $\pi_{01} = \frac{1}{1 + RR_1}$, respectively. Similarly, under $X_{it} = 0$, the RR is given by

$$RR_0 = \frac{\pi_{00}}{1 - \pi_{00}}.$$  

(7)

Then for a specified $RR_0$, $\pi_{00}$ and $\pi_{10}$ are given by $\pi_{00} = \frac{RR_0}{1 + RR_0}$ and $\pi_{10} = \frac{1}{1 + RR_0}$, respectively. Therefore, $\mathbf{II}$ can be obtained.

### 3 METHODOLOGY

#### 3.1 Corrected control limits under expectation

Inspired by Equation (5), an intuitive approach is to take the inverse matrix $\mathbf{II}^{-1}$ to both sides of Equation (5), yielding

$$\mathbf{II}^{-1}\left(\begin{array}{c} p_0^* \\ q_0^* \end{array}\right) = \left(\begin{array}{c} p_0 \\ q_0 \end{array}\right).$$  

(8)

It suggests that $\mathbf{II}^{-1}$ is regarded as the term to correct for error effects in $p_0^*$ and $q_0^*$, and the left hand side of Equation (8) can reflect $p_0$ and $q_0$.

Specifically, from Equation (8), $p_0$ can be expressed by

$$p_0 = \frac{\pi_{00} p_0^* - \pi_{10} q_0^*}{\pi_{11} \pi_{00} - \pi_{10} \pi_{01}}$$

$$= \frac{(1 - \pi_{10}) p_0^* - \pi_{10} q_0^*}{(1 - \pi_{10})(1 - \pi_{01}) - \pi_{10} \pi_{01}}$$

$$= \frac{p_0^* - \pi_{10}}{1 - \pi_{10} - \pi_{01}},$$  

(9)
where the first equality is due to the manipulation of inverse matrix, the second equality is due to the property of $\Pi$ and the last step comes from the fact $p_0^* + q_0^* = 1$. Consequently, Equation (9) suggests that the ‘corrected’ proportion of non-conforming products, denoted as $p_{0}^{**}$, is defined as the right hand side of Equation (9). That is

$$p_{0}^{**} \triangleq \frac{p_0^* - \pi_{10}}{1 - \pi_{10} - \pi_{01}}.$$  \hspace{1cm} (10)

Thus, it suggests that the corrected random variable is given by

$$X_{it}^{**} \triangleq \frac{X_{it}^* - \pi_{10}}{1 - \pi_{10} - \pi_{01}}.$$  \hspace{1cm} (11)

Thus, the corrected sample proportion is denoted as $\hat{p}_{0,t}^{**} = \frac{1}{n} \sum_{i=1}^{n} X_{it}^{**}$.

### 3.2 An error-corrected asymmetric EWMA $p$-control chart

As discussed in Section 2.1, we can take the EWMA as the monitoring statistic and then develop the control chart. To address this, we apply Equation (1) with $\hat{p}_{0,t}$ replaced by $\hat{p}_{0,t}^{**}$ to obtain the corrected EWMA statistics, which is given by

$$\text{EWMA}_{0,t}^{**} = \lambda \hat{p}_{0,t}^{**} + (1 - \lambda)\text{EWMA}_{0,t-1}^{**}$$  \hspace{1cm} (12)

for $t = 1, \ldots, T$, where $\lambda \in (0, 1]$ is a smoothing parameter.

To construct the control chart, we require the expectation and variance of Equation (12) as stated in the following theorem.

**Theorem 3.1.** Under the monitoring statistic (12), we have

(a) $E\{\text{EWMA}_{0,t}^{**}\} = p_0$;

(b) $\text{var}\{\text{EWMA}_{0,t}^{**}\} = \frac{p_0^*(1-p_0^*)\lambda(1-(1-\lambda)^2)}{n(1-\pi_{10}-\pi_{01})^2(2-\lambda)}$.

Under Theorem 3.1, the corrected UCL and LCL are given by

$$\text{UCL}_{0,t}^{**} = p_{0}^{**} + L^{**} \sqrt{\frac{p_0^*(1-p_0^*)\lambda(1-(1-\lambda)^2)}{n(1-\pi_{10}-\pi_{01})^2(2-\lambda)}}$$ and $\text{LCL}_{0,t}^{**} = 0,$ \hspace{1cm} (13)

where $L^{**}$ is the corresponding coefficient. We call Equation (13) the corrected EWMA $p$-control chart.

**Remark 3.1.**

1. If the interest is $np$ charts, then simply multiplying $n$ to Equation (13) yields the desired result.
2. If sample sizes in different sub-group are different from each other, then the sample size $n$ in Equation (13) can be replaced by $n_k$ for the $k$th sub-group.

**Remark 3.2 (Naive EWMA statistic and $p$-control chart).** Under the observed random variable without error correction, we can also construct the EWMA chart. Specifically, let

$$\text{EWMA}_{0,t}^* = \lambda \hat{p}_{0,t}^* + (1 - \lambda)\text{EWMA}_{0,t-1}^*$$  \hspace{1cm} (14)
denote the naive EWMA statistic, and we define
\[
\text{UCL}^* = p^*_0 + L^* \sqrt{\frac{p^*_0 (1 - p^*_0)}{n(2 - \lambda)}} \{1 - (1 - \lambda)^{2t}\}, \quad \text{LCL}^* = 0
\]
as the naive EWMA \(p\)-control chart with \(L^*\) being the associated coefficient.

### 3.3 Examination of ARL

In the framework of SPC, to assess the performance of \(p\)-control charts, we usually examine the average run length (ARL). Let \(\text{ARL}_0\) denote the IC ARL, which reflects the mean of the run length when the process is IC. In addition, let \(\text{ARL}_1\) be the OC ARL, which gives the average number of samples collected from the time of shift occurrence to the time of signal.

The goal is to use Equation (13) to compute \(\text{ARL}_0\) and \(\text{ARL}_1\). Unlike the Shewhart chart where \(\text{ARL}_0\) and \(\text{ARL}_1\) can reflect Type I and II errors, respectively, the challenge is that the EWMA statistic has no such a property since it is dependent. In addition, the coefficient of control limit is usually unknown and the control limit is asymmetric. Therefore, to deal with those issues, we first fix \(\text{ARL}_0\) at a given level, and then calculate the coefficient of control limit \(L^{**}\). After that, based on the constructed EWMA chart, we compute \(\text{ARL}_1\). Detailed computational algorithms are placed in the following two sub-sections, and similar strategy can be applied to the naive EWMA \(p\)-control chart (15).

#### 3.3.1 Determination of critical values and in-control ARL

We first fix \(\text{ARL}_0\) whose typical choices include 200 or 370. In addition, under the error-prone data, we can use Equation (5) to calculate \(p^*_0\). To address measurement error effects, we employ Equation (10) to derive the corrected IC proportion \(p^{**}_0\).

Given a range \((a, b)\) with user-specific values \(a = 0.01\) and \(b = 10\) for \(L^{**}\), we aim to find the optimal value of \(L^{**}\) that satisfies a given \(\text{ARL}_0\). The strategy to determine \(L^{**}\) is the Monte Carlo method with repetition \(M\). Under the \(m\)th Monte Carlo step, we evaluate Equation (12) and UCL in Equation (13) for every \(t = 1, 2, ..., T\) based on a proportion determined by IC samples. Let \(t_{0,m}\) denote a run length, which is defined as a value \(t\) such that EWMA\^{**}_0, t \geq \text{UCL}^{**}\). Finally, under \(M\) repetitions in the Monte Carlo procedure, the estimated ARL under IC samples is given by \(\hat{\text{ARL}}_0 \triangleq \frac{1}{M} \sum_{m=1}^{M} t_{0,m}\), and thus, \(L^{**} \in (a, b)\) can be determined by minimizing \(|\text{ARL}_0 - \hat{\text{ARL}}_0| < 1\). The pseudo-code of computational procedure for critical value \(L^{**}\) and \(\text{ARL}_0\) is presented in Algorithm 1.

#### 3.3.2 Determination of out-of-control ARL

Let \(p_1\) denote the OC process proportion that is determined by re-scaled proportion \(p_0\). Moreover, in the presence of measurement error, we have the misclassified OC proportion \(p^{**}_1\), which can be characterized with respect to \(p_1\) by the similar relationship of Equation (5). To correct for measurement error effects and recover misclassified proportion to the true one, we again employ Equation (10) to define the adjusted OC proportion, and denote it as \(p^{**}_1\).

After adjusting the OC proportion, the next goal is to examine the constructed control chart and compute \(\text{ARL}_1\). Following the similar strategy in Section 3.3.1, we employ the Monte Carlo method to derive \(\text{ARL}_1\) based on the constructed EWMA control chart. For the \(m\)th step in the Monte Carlo procedure, we use Equation (12) to compute EWMA based on OC proportion for each monitor time \(t\). After that, let \(t_{1,m}\) denote an OC run length, which is given by a value \(t\) that satisfies EWMA\^{**}_1, t \geq \text{UCL}\^{**}\) with UCL\^{**} being determined in Algorithm 1. Consequently, under \(M\) repetitions in the Monte Carlo procedure, the estimated ARL under OC samples is given by \(\hat{\text{ARL}}_1 \triangleq \frac{1}{M} \sum_{m=1}^{M} t_{1,m}\). The pseudo-code of computation of \(\text{ARL}_1\) is presented in Algorithm 2.
Algorithm 1 Monte Carlo simulation to find coefficients of the control limit and ARL

Step 1: Given in-control $p_0$ and $\Pi$, $p_0^*$ is calculated by formula (4);

Step 2: Set $\lambda$, $n$, and a value of $ARL_0$;

Step 3: Set $a < L^{**} < b$ with $a = 0.01$ and $b = 10$ (say);

Step 4: Monte Carlo procedure:

for step $(m + 1)$ with $m = 1, 2, \ldots, M$ and set $M = 10001$ (say) do

Step 4.1: Let $EWMA_{0,0}^{**} = p_0^{**}$ and $t = 1$;

Step 4.2: Based on the observed and misclassified $X_{it}^*$, calculate

$$p_{0,t}^{**} = \frac{1}{n} \sum_{i=1}^{n} \frac{X_{it}^* - \pi_{0i}}{1 - \pi_{0i} - \pi_{01}};$$

- If $t = 1$, $EWMA_{0,1}^{**} = \lambda p_{0,1}^{**} + (1 - \lambda)EWMA_{0,0}^{**}$;
- If $t \neq 1$, $EWMA_{0,t}^{**} = \lambda p_{0,t}^{**} + (1 - \lambda)EWMA_{0,t-1}^{**}$.

Step 4.3: Set UCL$^{**}$ and LCL$^{**}$ as values in (9);

- If $EWMA_{0,t}^{**} \geq UCL^{**}$, then take $t_{0,m} \triangleq t$ as a run length and $m \leftarrow m + 1$.
  Go to Step 1;
- If $0 < EWMA_{0,t}^{**} \leq UCL^{**}$, then $t \leftarrow t + 1$. Go to Step 2.

end

Step 5: Calculate $\frac{1}{M} \sum_{m=1}^{M} t_{0,m}$ and take it as the estimate of the nominal $ARL_0$, which is denoted as $\tilde{ARL}_0$. Finally, determine $L^{**}$ by minimizing $|ARL_0 - \tilde{ARL}_0| < 1$

subject to $a < L^{**} < b$.

4 | SIMULATION STUDIES

4.1 | Simulation setup

Let $n$ denote the sample size in each monitoring time, where $n = 5, 10, 15$ and 20. Let $T$ be the monitoring time that is set as $T = 5000$. For $i = 1, \ldots, n$ and $t = 1, \ldots, T$, let the true IC data $X_{it}$ be generated from the Bernoulli distribution with the IC probability $p_0$ that is specified as 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5. Moreover, for the OC samples, the proportion $p_1$ is defined as $p_1 \triangleq (1 + \delta)p_0$ with $\delta$ being specified as 0.1 and 0.2.

For the misclassification model (5), we specify $\pi_{00} = \pi_{11} = \pi$ and $\pi_{10} = \pi_{01} = 1 - \pi$ with $\pi = 0.95$ and 0.99, yielding reliability ratios $RR_1 = RR_0 = 19$ or 99, respectively. Based on the misclassification model (5), we generate the error-prone binary random variable $X^*$. We denote $p_0^*$ and $p_1^*$ as the corresponding error-prone IC and OC probabilities, respectively.

To implement the corrected EWMA $p$-control chart, we adopt Equation (10) to calculate the corrected probability $p_0^{**}$, and then employ Equation (12) to construct the EWMA statistics with smoothing parameters being $\lambda = 0.05$ or 0.2. Finally, 10,000 simulations are run for all settings.
Algorithm 2 Calculation for ARL$_1$

Step 1: Given out-of-control $p_1^*$, $\lambda$, $n$, a nominal ARL$_0$, and UCL$^{**}$;

Step 2: Based on $X_{n}^*$ in OC, calculate $\hat{p}_{1,t}^{**} = \frac{1}{n} \sum_{i=1}^{n} \frac{X_{i}^* - \pi(\lambda)}{1 - \pi(\lambda)}$.

Step 3: Monte Carlo procedure:

for step $(m + 1)$ with $m = 1, 2, \ldots, M$ and set $M = 10001$ (say) do

Step 3.1: Let EWMA$_{1,0}^{**} = p_0^{**}$ and $t = 1$.

Step 3.2: calculate out-of-control statistics, denoted EWMA$^{**}_{1,t}$:

- If $t = 1$, EWMA$^{**}_{1,1} = \lambda \hat{p}_{1,1}^{**} + (1 - \lambda) \text{EWMA}^{**}_{1,0}$;
- If $t \neq 1$, EWMA$^{**}_{1,t} = \lambda \hat{p}_{1,t}^{**} + (1 - \lambda) \text{EWMA}^{**}_{1,t-1}$.

Step 3.3: Plot EWMA$^{**}_{1,t}$ in the chart.

- If EWMA$^{**}_{1,t} \geq \text{UCL}^{**}$, then take $t_{1,m} \triangleq t$ as a run length and $m \leftarrow m + 1$.
  Go to Step 3.1;
- If EWMA$^{**}_{1,t} \leq \text{UCL}^{**}$, then $t \leftarrow t + 1$. Go to Step 3.2.

end

Step 4: Calculate the estimated ARL$_1$, which is given by $\text{ARL}_1 = \frac{1}{M} \sum_{m=1}^{M} t_{1,m}$.

4.2 Simulation results

To assess the OC detection performance of the corrected EWMA $p$-control chart, we also investigate the naive method in Remark 3.2 and Equation (1) by assuming the existence of the true $X$. While the common choice of ARL$_0$ can be 370 or 200, here we only explore 370 because the result under ARL$_0 = 200$ has the similar pattern to that based on ARL$_0 = 370$.

We first present numerical results for coefficients of the control limit and UCLs based on given $\pi$ and $\lambda$ in Tables 1–4. We observe that UCL decreases when $n$ becomes large regardless of the methods. In the presence of misclassification, we observe that UCL$^*$ increases when misclassification is severe (i.e., $\pi$ becomes small); on the contrary, it is interesting to see that values of UCL$^{**}$ determined by the corrected EWMA $p$-control chart keep remained and are the same as UCL based on the true $X$. In addition, values of UCL$^*$ are greater than others, suggesting wider control limits determined by error-prone variables. On the other hand, with $\pi$ fixed, values of UCL will increase when $\lambda$ is increasing regardless of the implementation of methods. Regarding the results of coefficients of the control limit $L$, we observe that $L^{**}$ is smaller than $L^*$ for most settings, and values of $L^{**}$ have the same decreasing pattern as values of $\pi$ and $\lambda$ do.

After obtaining estimated control limits, we next assess the estimation results of ARL$_1$. Numerical results under ARL$_0 = 370$ are placed in Tables 5–8, where ARL$_1$ with superscripts $*$ and $**$ represent the naive and proposed methods, respectively, and the results for the true $X$ are recorded as ARL$_1$ without the superscript. We observe that all values decease when $n$, $p_1$ or $\delta$ increase. In general, we can see that values of ARL$_1^*$ are always greater than ARL$_1$ and ARL$_1^{**}$, and differences become large as $\delta$ and $n$ are small. It shows that the naive method is unsatisfactory to detect OC proportion, because smaller values of ARL$_1$ have better performance to detect OC (e.g., Maravelakis et al. 7). On the other hand, values of ARL$_1^{**}$ are close to ARL$_1$ under the true $X$. These results suggest that (1) the necessity of measurement error correction,
### Table 1: Simulation Results ARL₀ = 370, λ = 0.05 and π = 0.95

| n | \( p₀ \) | \( p₀' \) | \( p₀'' \) | \( p₀''' \) |
|---|---|---|---|---|
| 5 | \( L \) | 2.463 | 2.355 | 2.290 | 2.215 |
| | UCL | 0.088 | 0.151 | 0.209 | 0.265 |
| | \( L^* \) | 2.363 | 2.311 | 2.305 | 2.263 |
| | UCL* | 0.145 | 0.197 | 0.249 | 0.298 |
| | \( L^{**} \) | 1.645 | 1.833 | 1.902 | 1.955 |
| | UCL** | 0.088 | 0.151 | 0.209 | 0.265 |
| 10 | \( L \) | 2.374 | 2.295 | 2.270 | 2.258 |
| | UCL | 0.076 | 0.135 | 0.191 | 0.246 |
| | \( L^* \) | 2.316 | 2.285 | 2.261 | 2.241 |
| | UCL* | 0.129 | 0.180 | 0.229 | 0.278 |
| | \( L^{**} \) | 1.584 | 1.794 | 1.874 | 1.936 |
| | UCL** | 0.076 | 0.135 | 0.191 | 0.246 |
| 15 | \( L \) | 2.343 | 2.270 | 2.260 | 2.243 |
| | UCL | 0.071 | 0.128 | 0.183 | 0.237 |
| | \( L^* \) | 2.286 | 2.256 | 2.242 | 2.221 |
| | UCL* | 0.123 | 0.172 | 0.221 | 0.269 |
| | \( L^{**} \) | 1.562 | 1.772 | 1.871 | 1.921 |
| | UCL** | 0.071 | 0.128 | 0.183 | 0.237 |
| 20 | \( L \) | 2.323 | 2.267 | 2.247 | 2.236 |
| | UCL | 0.068 | 0.124 | 0.179 | 0.232 |
| | \( L^* \) | 2.284 | 2.250 | 2.235 | 2.220 |
| | UCL* | 0.119 | 0.168 | 0.216 | 0.263 |
| | \( L^{**} \) | 1.552 | 1.764 | 1.857 | 1.912 |
| | UCL** | 0.068 | 0.124 | 0.179 | 0.232 |

(Continues)

### Table 2: Simulation Results ARL₀ = 370, λ = 0.05 and π = 0.99

| n | \( p₀ \) | \( p₀' \) | \( p₀'' \) | \( p₀''' \) |
|---|---|---|---|---|
| 5 | \( L \) | 2.463 | 2.355 | 2.290 | 2.215 |
| | UCL | 0.088 | 0.151 | 0.209 | 0.265 |
| | \( L^* \) | 2.432 | 2.341 | 2.306 | 2.276 |
| | UCL* | 0.100 | 0.160 | 0.217 | 0.272 |
| | \( L^{**} \) | 2.228 | 2.228 | 2.205 | 2.215 |
| | UCL** | 0.088 | 0.151 | 0.209 | 0.265 |
| 10 | \( L \) | 2.374 | 2.295 | 2.270 | 2.258 |
| | UCL | 0.076 | 0.135 | 0.191 | 0.246 |
| | \( L^* \) | 2.361 | 2.295 | 2.257 | 2.250 |
| | UCL* | 0.087 | 0.144 | 0.199 | 0.252 |
| | \( L^{**} \) | 2.144 | 2.183 | 2.182 | 2.190 |
| | UCL** | 0.076 | 0.135 | 0.191 | 0.246 |

(Continues)
TABLE 2  (Continued)

| \( n \) | \( p_0 \) | \( p^{*} \) | \( p^{**} \) | \( \sigma \) | \( \lambda \) | \( \pi \) |
|---|---|---|---|---|---|---|
| 15 | L | UCL | L^* | UCL^* | L^** | UCL^** |
| 1.050 | 0.060 | 1.050 | 0.060 | 1.050 | 0.060 | 1.050 | 0.060 |
| 0.095 | 0.105 | 0.115 | 0.125 | 0.135 | 0.145 | 0.155 | 0.165 |
| 0.150 | 0.200 | 0.250 | 0.300 | 0.350 | 0.400 | 0.450 | 0.500 |
| 0.300 | 0.350 | 0.400 | 0.450 | 0.500 | 0.550 | 0.600 | 0.650 |
| 0.400 | 0.450 | 0.500 | 0.550 | 0.600 | 0.650 | 0.700 | 0.750 |
| 0.500 | 0.550 | 0.600 | 0.650 | 0.700 | 0.750 | 0.800 | 0.850 |
| 1.050 | 0.060 | 1.050 | 0.060 | 1.050 | 0.060 | 1.050 | 0.060 |
| 0.095 | 0.105 | 0.115 | 0.125 | 0.135 | 0.145 | 0.155 | 0.165 |
| 0.150 | 0.200 | 0.250 | 0.300 | 0.350 | 0.400 | 0.450 | 0.500 |
| 0.300 | 0.350 | 0.400 | 0.450 | 0.500 | 0.550 | 0.600 | 0.650 |
| 0.400 | 0.450 | 0.500 | 0.550 | 0.600 | 0.650 | 0.700 | 0.750 |
| 0.500 | 0.550 | 0.600 | 0.650 | 0.700 | 0.750 | 0.800 | 0.850 |

TABLE 3  Simulation results \( ARL_0 = 370, \lambda = 0.2 \) and \( \pi = 0.95 \)

| \( n \) | \( p_0 \) | \( p^{*} \) | \( p^{**} \) | \( \sigma \) | \( \lambda \) | \( \pi \) |
|---|---|---|---|---|---|---|
| 5 | L | UCL | L^* | UCL^* | L^** | UCL^** |
| 1.050 | 0.060 | 1.050 | 0.060 | 1.050 | 0.060 | 1.050 | 0.060 |
| 0.095 | 0.105 | 0.115 | 0.125 | 0.135 | 0.145 | 0.155 | 0.165 |
| 0.150 | 0.200 | 0.250 | 0.300 | 0.350 | 0.400 | 0.450 | 0.500 |
| 0.300 | 0.350 | 0.400 | 0.450 | 0.500 | 0.550 | 0.600 | 0.650 |
| 0.400 | 0.450 | 0.500 | 0.550 | 0.600 | 0.650 | 0.700 | 0.750 |
| 0.500 | 0.550 | 0.600 | 0.650 | 0.700 | 0.750 | 0.800 | 0.850 |
| 10 | L | UCL | L^* | UCL^* | L^** | UCL^** |
| 1.050 | 0.060 | 1.050 | 0.060 | 1.050 | 0.060 | 1.050 | 0.060 |
| 0.095 | 0.105 | 0.115 | 0.125 | 0.135 | 0.145 | 0.155 | 0.165 |
| 0.150 | 0.200 | 0.250 | 0.300 | 0.350 | 0.400 | 0.450 | 0.500 |
| 0.300 | 0.350 | 0.400 | 0.450 | 0.500 | 0.550 | 0.600 | 0.650 |
| 0.400 | 0.450 | 0.500 | 0.550 | 0.600 | 0.650 | 0.700 | 0.750 |
| 0.500 | 0.550 | 0.600 | 0.650 | 0.700 | 0.750 | 0.800 | 0.850 |
| 15 | L | UCL | L^* | UCL^* | L^** | UCL^** |
| 1.050 | 0.060 | 1.050 | 0.060 | 1.050 | 0.060 | 1.050 | 0.060 |
| 0.095 | 0.105 | 0.115 | 0.125 | 0.135 | 0.145 | 0.155 | 0.165 |
| 0.150 | 0.200 | 0.250 | 0.300 | 0.350 | 0.400 | 0.450 | 0.500 |
| 0.300 | 0.350 | 0.400 | 0.450 | 0.500 | 0.550 | 0.600 | 0.650 |
| 0.400 | 0.450 | 0.500 | 0.550 | 0.600 | 0.650 | 0.700 | 0.750 |
| 0.500 | 0.550 | 0.600 | 0.650 | 0.700 | 0.750 | 0.800 | 0.850 |
| 20 | L | UCL | L^* | UCL^* | L^** | UCL^** |
| 1.050 | 0.060 | 1.050 | 0.060 | 1.050 | 0.060 | 1.050 | 0.060 |
| 0.095 | 0.105 | 0.115 | 0.125 | 0.135 | 0.145 | 0.155 | 0.165 |
| 0.150 | 0.200 | 0.250 | 0.300 | 0.350 | 0.400 | 0.450 | 0.500 |
| 0.300 | 0.350 | 0.400 | 0.450 | 0.500 | 0.550 | 0.600 | 0.650 |
| 0.400 | 0.450 | 0.500 | 0.550 | 0.600 | 0.650 | 0.700 | 0.750 |
| 0.500 | 0.550 | 0.600 | 0.650 | 0.700 | 0.750 | 0.800 | 0.850 |
TABLE 4 Simulation results ARL$_{0}$ = 370, $\lambda$ = 0.2 and $\pi$ = 0.99

| $p_0$   | 0.050 | 0.100 | 0.150 | 0.200 | 0.250 | 0.300 | 0.350 | 0.400 | 0.450 | 0.500 |
|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| $p_0^*$ | 0.059 | 0.108 | 0.157 | 0.206 | 0.255 | 0.304 | 0.353 | 0.402 | 0.451 | 0.500 |
| $p_0^{**}$ | 0.050 | 0.100 | 0.150 | 0.200 | 0.250 | 0.300 | 0.350 | 0.400 | 0.450 | 0.500 |

| $n$ | 5 | 10 | 15 | 20 |
|-----|---|----|----|----|
| $L$ | 3.336 | 3.138 | 3.044 | 2.980 |
| UCL | 0.158 | 0.122 | 0.107 | 0.098 |
| $L^*$ | 3.257 | 3.095 | 3.001 | 2.980 |
| UCL* | 0.173 | 0.136 | 0.120 | 0.109 |
| $L^{**}$ | 3.032 | 2.842 | 2.842 | 2.701 |
| UCL** | 0.159 | 0.122 | 0.110 | 0.098 |

$\pi$ = 0.99

(2) the correction of measurement error effects recover the control limit to that determined by the true $X$, and makes the detection of OC be the same as that of true $X$. (3) the corrected EWMA $p$-control chart is successful to handle the scenario of small sample size and is robust in detecting OC proportion.

5 APPLICATION TO REAL DATA ANALYSIS

In this section, we apply the corrected EWMA $p$-control chart to the orange juice data, which can be found in the R package qcr.

The primary interest of this study is the production of 6-oz cardboard cans that frozen orange juice concentrate is packed in. These cans are formed on a machine by spinning them from cardboard stock and attaching a metal bottom panel. A can is then inspected to determine whether, when filled, the liquid could possible leak either on the side seam or around the bottom joint. If this situation occurs, then a can is considered non-conforming. The data were collected as 30 samples of 50 cans each at half-hour intervals over a three-shift period in which the machine was in continuous operation. From sample 15 used, a new batch of cardboard stock was put into production. Sample 23 was obtained when an inexperienced operator was temporarily assigned to the machine. After the first 30 samples, a machine adjustment was made. Then further 24 samples were taken from the process. Therefore, according to the data description, we take the IC monitoring time as $T = 24$ and assign the OC monitoring time as 30. For each IC monitoring time, the sample size is $n = 50$.

In applications, measurement error is ubiquitous in process monitoring. In addition, Pendrill and Puydarrieux et al. pointed out that risks arising from measurement are main concern in conformity assessment. In other words, from the orange juice data, it is possible that factory's staffs falsely record non-conforming (or conforming) cans to be conforming
TABLE 5 Simulation results for ARL based on Table 1

| δ  | n   | p₁  | p₁* | p₁** |
|----|-----|-----|-----|------|
| 0.1| 5   | ARL | 208.0 | 166.6 | 136.4 |
|    |     | p₁* | 251.9 | 198.4 | 160.6 |
|    |     | ARL** | 207.7 | 166.0 | 138.8 |
| 10 |     | ARL | 168.3 | 123.6 | 101.2 |
|    |     | p₁* | 186.7 | 145.2 | 121.2 |
|    |     | ARL** | 167.8 | 124.6 | 102.8 |
| 15 |     | ARL | 146.3 | 107.2 | 80.9 |
|    |     | p₁* | 164.6 | 120.9 | 99.9 |
|    |     | ARL** | 146.3 | 107.2 | 80.9 |
| 20 |     | ARL | 128.9 | 88.8 | 66.7 |
|    |     | p₁* | 156.1 | 115.9 | 94.8 |
|    |     | ARL** | 128.5 | 88.8 | 66.7 |
| 0.2| 5   | ARL | 128.7 | 89.2 | 66.7 |
|    |     | p₁* | 156.1 | 115.9 | 94.8 |
|    |     | ARL** | 128.5 | 88.8 | 66.7 |
| 10 |     | ARL | 151.6 | 116.0 | 83.5 |
|    |     | p₁* | 175.6 | 131.9 | 104.5 |
|    |     | ARL** | 151.6 | 116.0 | 83.5 |
| 15 |     | ARL | 91.8 | 58.4 | 34.3 |
|    |     | p₁* | 136.2 | 96.3 | 55.1 |
|    |     | ARL** | 91.8 | 58.4 | 34.3 |
| 20 |     | ARL | 73.0 | 44.9 | 23.5 |
|    |     | p₁* | 86.1 | 54.2 | 32.2 |
|    |     | ARL** | 73.0 | 44.9 | 23.5 |

(or non-conforming). As a result, it is reasonable to assume that the collected data are subject to measurement error, and the corresponding proportion is given by \( p^*_0 = 0.111 \).

Since measurement error correction is crucial, we now adopt Equation (5) to characterize measurement error and employ the corrected EWMA \( p \)-control chart (10) to correct for measurement error effects. Since \( \Pi \) in Equation (5) is unknown and additional information is unavailable, we employ sensitivity analyses to address different levels of measurement error effects. Specifically, we follow Section 4.1 to specify \( \pi_{00} = \pi_{11} = \pi \) and \( \pi_{10} = \pi_{01} = 1 - \pi \) with \( \pi = 0.95 \) and 0.99, which yield \( p^*_{00} = 0.068 \) and 0.103, respectively.

Now, for each fixed \( \pi = 0.95 \) or 0.99 and \( \lambda = 0.05 \) or 0.2, we adopt Equations (12) and (14) to compute the EWMA statistics, and apply Algorithm 1 to compute coefficients of control limits and UCL based on a given ARL\(_0\), yielding the corrected or naive EWMA \( p \)-control charts, respectively. For the choice of ARL\(_0\), we simply use 370 for exploration, and ARL\(_0\) = 200 gives the similar result.

We first report the estimation results for coefficients of control limits and UCL based on the naive and corrected EWMA \( p \)-control charts in Table 9. We observe that coefficients and UCL values based on the naive EWMA \( p \)-control chart are larger than those based on the corrected EWMA \( p \)-control chart regardless of choices of \( \lambda \), which indicates that the naive EWMA \( p \)-control chart gives wider control limits than those given by the corrected EWMA \( p \)-control chart. In addition, from the corrected EWMA \( p \)-control chart, we can see that both coefficients of control limits and UCL are decreasing when \( \pi \) becomes small, which reflects a phenomenon that the corrected control limits may be implicitly affected by the proportion of misclassification. Moreover, we also display naive and corrected EWMA \( p \)-control charts based on \( \lambda = 0.05 \).
and 0.2 in Figures 1 and 2, respectively, for visualization. We can see that all monitoring points are under the UCL, except for the third point that is known as the false alarm based on \( \pi = 0.95 \) in Figure 2.

Finally, based on the developed naive or corrected EWMA \( p \)-control charts, we further detect OC process parameters, and the associated control charts under \( \lambda = 0.05 \) or 0.2 and \( \pi = 0.95 \) or 0.99 are displayed in Figures 3 and 4, respectively. We observe that detected points under uncorrected or corrected EWMA \( p \)-control charts have similar pattern, and values of EWMA statistic becomes small as \( \pi \) is decreasing. In addition, the naive EWMA \( p \)-control chart seems to be sensitive to detect OC with the choice of \( \lambda \). In particular, when \( \lambda = 0.05 \), the second point of the naive EWMA \( p \)-control chart in Figure 3 is not detected as OC, while the corrected EWMA \( p \)-control chart can successfully detect OC regardless of the choices of \( \lambda \) or \( \pi \).

## 6 | DISCUSSION

SPC has been an important tool to monitor products. Some methods have been developed to construct control charts for binary and/or continuous random variables. In applications, however, measurement error exists due to imprecise operation systems or human-made mistakes, and ignoring measurement error effects may cause wrong detection. While many research works have been available to discuss measurement error effects on continuous random variables, however, to the best of our knowledge, little attention focused on binary random variables. Moreover, none of them provided suitable

| \( \delta \) | \( n \) | \( p_1 \) | 0.055 | 0.110 | 0.165 | 0.220 | 0.275 | 0.330 | 0.385 | 0.440 | 0.495 | 0.550 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.1 | 5 | ARL_1 | 209.4 | 165.7 | 137.2 | 119.5 | 103.7 | 89.1 | 80.3 | 69.7 | 61.5 | 54.2 |
| | | ARL_1* | 217.0 | 170.3 | 143.6 | 122.8 | 106.1 | 92.0 | 81.8 | 71.8 | 62.6 | 55.1 |
| | | ARL_1** | 206.5 | 165.2 | 136.6 | 120.2 | 104.1 | 88.8 | 79.9 | 69.7 | 61.6 | 53.8 |
| 10 | | ARL_1 | 169.3 | 124.5 | 101.4 | 82.8 | 69.9 | 60.3 | 51.0 | 44.8 | 37.9 | 33.3 |
| | | ARL_1* | 183.8 | 130.4 | 103.9 | 83.9 | 72.2 | 61.5 | 52.5 | 45.6 | 39.1 | 34.4 |
| | | ARL_1** | 166.3 | 126.0 | 101.7 | 82.4 | 69.6 | 60.1 | 50.7 | 44.5 | 37.9 | 33.3 |
| 15 | | ARL_1 | 146.2 | 102.5 | 80.9 | 65.3 | 54.4 | 45.6 | 38.8 | 33.5 | 28.1 | 24.5 |
| | | ARL_1* | 158.1 | 108.9 | 84.1 | 66.9 | 55.9 | 47.3 | 40.5 | 34.8 | 29.1 | 25.3 |
| | | ARL_1** | 145.6 | 103.2 | 80.3 | 65.0 | 54.0 | 45.7 | 39.0 | 33.6 | 28.3 | 24.6 |
| 20 | | ARL_1 | 130.0 | 89.2 | 67.3 | 53.4 | 44.7 | 37.1 | 31.6 | 26.7 | 22.9 | 19.6 |
| | | ARL_1* | 140.6 | 94.1 | 70.6 | 56.7 | 46.1 | 38.4 | 32.5 | 27.7 | 23.7 | 20.3 |
| | | ARL_1** | 129.6 | 87.9 | 67.4 | 53.3 | 44.8 | 36.8 | 31.6 | 26.9 | 23.0 | 19.7 |
| 0.2 | 5 | ARL_1 | 0.060 | 0.120 | 0.180 | 0.240 | 0.300 | 0.360 | 0.420 | 0.480 | 0.540 | 0.600 |
| | | ARL_1* | 0.069 | 0.128 | 0.186 | 0.245 | 0.304 | 0.363 | 0.422 | 0.480 | 0.539 | 0.598 |
| | | ARL_1** | 0.060 | 0.120 | 0.180 | 0.240 | 0.300 | 0.360 | 0.420 | 0.480 | 0.540 | 0.600 |
| 10 | | ARL_1 | 128.2 | 89.0 | 66.7 | 54.7 | 45.0 | 36.6 | 32.2 | 27.2 | 23.1 | 19.8 |
| | | ARL_1* | 138.3 | 93.1 | 71.4 | 57.0 | 46.7 | 37.7 | 33.0 | 27.9 | 23.8 | 20.4 |
| | | ARL_1** | 128.0 | 88.7 | 67.2 | 54.9 | 44.8 | 36.5 | 32.2 | 27.2 | 23.3 | 19.7 |
| 15 | | ARL_1 | 92.0 | 58.6 | 43.8 | 33.3 | 27.2 | 22.8 | 18.8 | 16.2 | 13.4 | 11.8 |
| | | ARL_1* | 103.3 | 62.6 | 45.4 | 34.3 | 28.4 | 23.5 | 19.4 | 16.6 | 13.9 | 12.2 |
| | | ARL_1** | 90.8 | 58.9 | 43.8 | 33.4 | 27.1 | 22.7 | 18.7 | 16.2 | 13.4 | 11.8 |
| 20 | | ARL_1 | 72.8 | 44.9 | 32.4 | 25.0 | 20.2 | 16.7 | 14.0 | 12.0 | 9.8 | 8.6 |
| | | ARL_1* | 81.8 | 48.3 | 34.3 | 26.0 | 20.9 | 17.3 | 14.5 | 12.4 | 10.2 | 8.8 |
| | | ARL_1** | 72.3 | 44.8 | 32.4 | 25.2 | 20.1 | 16.6 | 14.0 | 12.0 | 9.8 | 8.6 |
| 0.2 | 5 | ARL_1 | 61.1 | 37.0 | 26.1 | 19.9 | 16.4 | 13.3 | 11.2 | 9.5 | 8.0 | 6.8 |
| | | ARL_1* | 68.9 | 39.9 | 27.6 | 21.4 | 17.0 | 13.8 | 11.6 | 9.7 | 8.2 | 7.0 |
| | | ARL_1** | 61.1 | 37.0 | 26.1 | 19.9 | 16.4 | 13.3 | 11.2 | 9.5 | 8.0 | 6.8 |
strategies to adjust error effects. In this paper, we consider binary random variables that can be used to describe non-conforming product, and we allow binary random variables to be contaminated by misclassification. We propose the corrected proportion to adjust for measurement error effects, and then employ the corrected EWMA $p$-control chart to obtain reliable and asymmetric control chart under small sample size. Numerical results based on different settings verify the validity of the corrected EWMA $p$-control chart.

One of the concerns in misclassification model (5) is the determination of $\Pi$. In our development, we employ sensitivity analyses to explore the impact of measurement error effects by examining different levels of misclassification probabilities. In the framework of measurement error analysis, $\Pi$ can be estimated if auxiliary information is available. One of typical information is external validation data. Specifically, suppose that $\mathcal{M}$ with $|\mathcal{M}| = n$ is the subject set for the main study containing measurements $\{X_{it}^* : i \in \mathcal{M}, t = 1, \ldots, T\}$ and let $\mathcal{V}$ with $|\mathcal{V}| = m$ denote the subject set for the external validation study containing measurements $\{X_{it}, X_{it}^* : i \in \mathcal{V}, t = 1, \ldots, T\}$, where $\mathcal{M}$ and $\mathcal{V}$ do not overlap. Assume that the main study and the validation study share the same model (5). With the availability of external validation data $\mathcal{V}$, we have a $2 \times 2$ confusion table. Then the probability $\pi_{kl}$ for $k, l \in \{0, 1\}$ can be estimated by

\begin{equation}
\hat{\pi}_{kl} = \frac{\text{number of } \{X_{it}^* = k \text{ and } X_{it} = l\} \text{ for all } i \in \mathcal{V} \text{ and } t = 1, \ldots, T}{\text{number of } \{X_{it} = l\} \text{ for all } i \in \mathcal{V} \text{ and } t = 1, \ldots, T}. \tag{16}
\end{equation}

Therefore, $\Pi$ can be estimated by $\hat{\Pi} = [\hat{\pi}_{kl}]_{k,l \in \{0,1\}}$. 

| $\delta$ | $n$ | $p_{1}^*$ | $p_{1}^{**}$ | $p_{1}^*$ | $p_{1}^{**}$ | $p_{1}^*$ | $p_{1}^{**}$ | $p_{1}^*$ | $p_{1}^{**}$ | $p_{1}^*$ | $p_{1}^{**}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.1 | 5 | ARL | 242.9 | 209.1 | 184.0 | 164.8 | 152.7 | 119.0 | 108.3 | 98.0 | 88.5 |
| | | ARL* | 282.1 | 239.6 | 207.2 | 185.2 | 164.6 | 148.1 | 135.0 | 120.6 | 107.8 | 100.1 |
| | | ARL** | 245.9 | 209.6 | 185.5 | 164.2 | 147.5 | 133.0 | 120.2 | 106.9 | 96.7 | 85.5 |
| 10 | ARL | 213.2 | 171.9 | 142.9 | 124.2 | 107.0 | 92.9 | 81.6 | 70.4 | 62.0 | 53.0 |
| | | ARL* | 257.3 | 203.4 | 166.2 | 142.2 | 107.9 | 93.9 | 82.6 | 71.6 | 61.8 |
| | | ARL** | 213.3 | 172.0 | 144.6 | 124.5 | 107.3 | 93.0 | 81.1 | 70.9 | 61.6 | 53.1 |
| 15 | ARL | 190.4 | 147.4 | 120.9 | 101.6 | 85.0 | 73.0 | 62.5 | 53.3 | 45.3 | 38.6 |
| | | ARL* | 232.8 | 178.9 | 142.8 | 118.6 | 99.3 | 85.6 | 72.7 | 62.8 | 53.8 | 45.6 |
| | | ARL** | 190.1 | 147.9 | 120.3 | 100.3 | 85.3 | 72.8 | 62.5 | 53.3 | 45.2 | 38.6 |
| 20 | ARL | 175.0 | 130.6 | 104.2 | 85.1 | 71.5 | 59.3 | 50.3 | 42.8 | 36.2 | 29.9 |
| | | ARL* | 223.3 | 161.8 | 126.4 | 102.5 | 83.2 | 70.4 | 60.2 | 50.3 | 42.8 | 35.7 |
| | | ARL** | 174.1 | 130.7 | 104.4 | 85.1 | 71.0 | 59.7 | 50.2 | 42.8 | 35.8 | 30.4 |
| 0.2 | 5 | ARL | 166.8 | 128.4 | 102.4 | 85.4 | 70.6 | 59.6 | 50.4 | 43.1 | 36.7 | 31.1 |
| | | ARL* | 217.7 | 159.7 | 124.5 | 101.8 | 84.2 | 70.7 | 60.3 | 50.5 | 42.7 | 37.1 |
| | | ARL** | 169.7 | 128.6 | 104.3 | 84.3 | 70.2 | 59.6 | 50.4 | 42.4 | 36.3 | 30.1 |
| 10 | ARL | 132.2 | 91.4 | 68.0 | 53.7 | 43.0 | 34.8 | 29.0 | 23.9 | 20.0 | 16.6 |
| | | ARL* | 182.1 | 119.2 | 86.6 | 66.2 | 52.9 | 43.0 | 34.9 | 28.9 | 23.9 | 19.9 |
| | | ARL** | 131.1 | 90.5 | 68.2 | 53.8 | 42.9 | 34.9 | 28.9 | 23.9 | 19.9 | 16.7 |
| 15 | ARL | 108.9 | 70.9 | 51.8 | 39.5 | 31.0 | 25.1 | 20.5 | 16.9 | 13.8 | 11.5 |
| | | ARL* | 156.0 | 96.4 | 66.7 | 49.6 | 38.1 | 30.7 | 24.8 | 20.4 | 16.7 | 13.9 |
| | | ARL** | 108.9 | 70.7 | 51.7 | 39.2 | 30.9 | 25.1 | 20.4 | 16.8 | 13.8 | 11.5 |
| 20 | ARL | 93.4 | 58.1 | 41.3 | 30.8 | 24.4 | 19.4 | 15.7 | 13.0 | 10.7 | 8.9 |
| | | ARL* | 141.0 | 81.4 | 55.1 | 40.0 | 30.0 | 23.8 | 19.4 | 15.6 | 12.9 | 10.7 |
| | | ARL** | 93.6 | 58.1 | 41.3 | 30.9 | 24.4 | 19.4 | 15.8 | 13.1 | 10.8 | 9.0 |
TABLE 8  Simulation results for ARL based on Table 4

| $\delta$ | $n$ | ARL$_1$ | ARL$_{**}$ | ARL$^*$ | ARL$^{**}$ |
|----------|-----|---------|------------|--------|----------|
| 5        |     | 242.0   | 208.2      | 184.5  | 166.5    |
|          |     | 258.2   | 217.8      | 190.0  | 172.5    |
|          |     | 244.2   | 211.1      | 185.5  | 167.0    |
| 10       |     | 214.5   | 172.2      | 142.7  | 120.6    |
|          |     | 222.0   | 175.9      | 147.9  | 128.0    |
|          |     | 213.4   | 171.0      | 143.8  | 123.2    |
| 15       |     | 190.0   | 146.8      | 120.6  | 101.3    |
|          |     | 201.4   | 154.7      | 125.5  | 104.5    |
|          |     | 191.0   | 147.9      | 120.4  | 101.3    |
| 20       |     | 173.9   | 131.1      | 103.7  | 85.6     |
|          |     | 185.6   | 137.6      | 108.2  | 89.1     |
|          |     | 173.1   | 130.3      | 104.4  | 85.1     |
| 0.2      |     | 166.6   | 129.0      | 103.1  | 85.7     |
|          |     | 183.6   | 135.5      | 108.4  | 87.4     |
|          |     | 169.0   | 129.1      | 102.7  | 84.8     |
| 0.4      |     | 132.4   | 91.1       | 67.6   | 54.0     |
|          |     | 141.0   | 95.2       | 70.9   | 55.7     |
|          |     | 129.9   | 90.5       | 68.0   | 53.8     |
| 0.5      |     | 108.7   | 70.6       | 51.6   | 39.5     |
|          |     | 119.9   | 76.1       | 54.2   | 41.3     |
|          |     | 109.5   | 71.0       | 51.7   | 39.5     |
| 0.7      |     | 92.8    | 58.1       | 41.4   | 31.0     |
|          |     | 104.3   | 62.9       | 43.6   | 32.9     |
|          |     | 92.8    | 57.9       | 41.3   | 30.9     |

TABLE 9  Numerical results for the orange juice data

| $\lambda$ | Naive ($\pi = 1.00$) | Correct ($\pi = 0.95$) | Correct ($\pi = 0.99$) |
|-----------|---------------------|------------------------|------------------------|
|           | $L^*$ | UCL$^*$ | $L^{**}$ | UCL$^{**}$ | $L^{**}$ | UCL$^{**}$ |
| 0.05      | 2.222 | 0.126 | 2.013 | 0.080 | 2.183 | 0.118 |
| 0.20      | 2.753 | 0.151 | 2.019 | 0.101 | 2.616 | 0.142 |

There are some possible extensions for this project. First of all, in addition to the current strategy in Section 3.1 to address measurement error effects, in the framework of measurement error analysis, simulation and extrapolation (SIMEX) method (e.g., Chen [24]) is also a valid tool to measurement error effects. It is a worth exploration in depth because no relevant work has been available for SPC. Moreover, the corrected EWMA $p$-control chart can be naturally extended to other settings. For example, as developed by Yang and Arnold, [3] Yang and Arnold [27] and Yang and Wu, [4] to address distribution free continuous random variables and build up control chart, a common approach to translate the continuous random variables to binary ones. In the presence of measurement error, measurement error effects may affect the translated random variables and associated proportion. Therefore, the corrected EWMA $p$-control chart and the corrected proportion can be employed to deal with this issue. The other interesting is about the profile monitoring (e.g., Qiu, [1] Chapter 10), which aims to take auxiliary information as the covariates, and then build up a regression model whose responses are the main interest to be monitored. In the case of monitoring confirming product, we may consider logistic regression models,
FIGURE 1  The naive ($\pi = 1.00$) and corrected ($\pi = 0.99, 0.95$) EWMA $p$-control charts for the orange juice data under $\lambda = 0.05$ and IC samples.

FIGURE 2  The naive ($\pi = 1.00$) and corrected ($\pi = 0.99, 0.95$) EWMA $p$-control charts for the orange juice data under $\lambda = 0.20$ and IC samples.

FIGURE 3  The naive ($\pi = 1.00$) and corrected ($\pi = 0.99, 0.95$) EWMA $p$-control charts for the orange juice data under $\lambda = 0.05$ and OC samples. Red points are OC detection.
FIGURE 4  The naive ($\pi = 1.00$) and corrected ($\pi = 0.99, 0.95$) EWMA $p$-control charts for the orange juice data under $\lambda = 0.20$ and OC samples. Red points are OC detection.

and the similar strategy in Section 3.1 can be adopted to address this concern. Detailed and deep explorations can be our next research projects in the future.

DATA AVAILABILITY STATEMENT
Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

ORCID
Li-Pang Chen  https://orcid.org/0000-0001-5440-5036
Su-Fen Yang  https://orcid.org/0000-0002-6801-6645

REFERENCES
1. Qiu P. Introduction to Statistical Process Control. CRC Press; 2014.
2. Yang S-F. An improved distribution-free EWMA mean chart. Commun Stat Simul Comput. 2016;45:1410-1427.
3. Yang S-F, Arnold BC. A new approach for monitoring process variance. J Statist Comput Simul. 2016;86:2749-2765.
4. Yang S-F, Wu S. A double sampling scheme for process variability monitoring. Qual Reliab Eng Int. 2017;33:2193-2204.
5. Zou C, Tsung F. Likelihood ratio-based distribution-free EWMA control charts. J Qual Technol. 2010;42:174-196.
6. Chen L-P, Muzayyanah S, Yang S-F, Wang B, Jiang T-A. Monitoring process location and dispersion simultaneously using a control region. DYNA. 2022;97:71-78.
7. Maravelakis P, Panaretos J, Psarakis S. EWMA chart and measurement error. J Appl Stat. 2004;31:445-455.
8. Asif F, Khan S, Noor-ul-Amin M. Hybrid exponentially weighted moving average control chart with measurement error. Iran J Sci Technol, Trans A. 2020;44:801-811.
9. Nguyen HD, Tran KP, Celano G, Maravelakis PE, Castagliola P. On the effect of the measurement error on Shewhart and EWMA $t$ control charts. Int J Adv Manuf Technol. 2020;107:4317-4332.
10. Tran KD, Nguyen HD, Nguyen TH, Tran KP. Design of a variable sampling interval exponentially weighted moving average median control chart in presence of measurement error. Qual Reliab Eng Int. 2021;37:374-390.
11. Linna KW, Woodall WH. Effect of measurement error on Shewhart control charts. J Qual Technol. 2001;33:213-222.
12. Huwang L, Hung Y. Effect of measurement error on monitoring multivariate process variability. Stat Sin. 2007;17:749-760.
13. Shongwe SC, Malela-Majika J-C, Castagliola P. On monitoring the process mean of autocorrelated observations with measurement errors using the $u$-of-$u$ runs-rules scheme. Qual Reliab Eng Int. 2020;36:1144-1160.
14. Linna KW, Woodall WH, Busby KL. The performance of multivariate control charts in the presence of measurement error. J Qual Technol. 2001;33:349-355.
15. Argoti MA, Carrión-García A. A heuristic method for obtaining quasi ARL-unbiased $p$-charts. Qual Reliab Eng Int. 2019;35:47-61.
16. Sparks, R. Linking EWMA $p$ charts and the risk adjustment control charts. Qual Reliab Eng Int. 2017;33:617-636.
17. Yang S-F, Arnold BC. Monitoring process variance using an ARL-unbiased EWMA-$p$ control chart. Qual Reliab Eng Int. 2016;32:1227-1235.
18. Aslam M, Rao GS, AL-Marshadi AH, Jun C-H. A nonparametric HEWMA-$p$ control chart for variance in monitoring processes. Symmetry. 2019;11:356.
19. Shu M-H, Wu HC. Monitoring imprecise fraction of conforming items using $p$ control charts. J Appl Stat. 2010;37:1283-1297.
20. Bourke PD. The RL2 chart versus the np chart for detecting upward shifts in fraction defective. *J Appl Stat*. 2006;33:1-15.
21. Yi GY. *Statistical Analysis with Measurement Error and Misclassification: Strategy, Method and Application*. Springer; 2017.
22. Chen L-P, Yi GY. Analysis of noisy survival data with graphical proportional hazards measurement error models. *Biometrics*. 2021;77:956-969.
23. Carroll RJ, Ruppert D, Stefanski LA, Crainiceany CM. *Measurement Error in Nonlinear Models*. 2nd ed. Chapman & Hall; 2006.
24. Chen L-P. Semiparametric estimation for the transformation model with length-biased data and covariate measurement error. *J Statist Comput Simul*. 2020;90:420-442.
25. Pendrill LR. Using measurement uncertainty in decision-making and conformity assessment. *Metrologia*. 2014;51:S206-S218.
26. Puydarrieux S, Pou JM, Leblond L, et al. Role of measurement uncertainty in conformity assessment. In: 19th International Congress of Metrology. 2019;16003.

**AUTHOR BIOGRAPHIES**

**Li-Pang Chen** is an Assistant Professor of Statistics at the National Chengchi University in Taiwan. He received a PhD in Statistics from the University of Waterloo and worked as a postdoctoral fellow at the University of Western Ontario in Canada. His research interests include statistical learning, high-dimensional data analysis, measurement error analysis and their applications to survival analysis and statistical process control.

**Su-Fen Yang** is a Distinguished Professor and Chair of Statistics department at the National Chengchi University in Taiwan. She received a Ph.D. in Statistics from University of California, Riverside, CA, USA. She previously served as Chair of the Department of the commercial college at National Chengchi University. She is a member of the Quality Management Committee of the Ministry of Economic Affairs Bureau of Standards, Taiwan; Committee of Chinese Society for Quality; Associate Editor of Journal of Quality and JCIIE. She is the author of one book on Quality Management and of more than 100 peer-reviewed articles published in international journals. Her research interests are mainly in the statistical process control, quality engineering, and probability models.

**How to cite this article:** Chen L-P, Yang S-F. A new p-control chart with measurement error correction. *Qual Reliab Eng Int*. 2023;39:81–98. [https://doi.org/10.1002/qre.3219](https://doi.org/10.1002/qre.3219)