Shewhart-type Charts in Some Nonstandard Situations: Phase I and Phase II

Y. Yao[1], S. Chakraborti[21]
[1]Department of Information Systems, Statistics, and Management Science, The University of Alabama, Tuscaloosa, AL
[2]Department of Information Systems, Statistics, and Management Science, The University of Alabama, Tuscaloosa, AL

Abstract
Statistical process control (SPC) and monitoring techniques are often used in applications where multiple sources of variation are present, as for example in a variance components model. Control charts are useful in these situations and have been considered in some earlier works by Roes and Does (1995), Woodall and Thomas (1995) and others in the literature. We present refinements and adaptations of these charts based on the knowledge gained and the advancements made over the last several years on retrospective and prospective process monitoring. This work takes proper account of the effects of parameter estimation while designing and implementing the control charts. In the sequel, we calculate the corrected (adjusted) control limits in both Phase I and Phase II applications, two important phases of the overall SPC regime. In Phase I, the correct control limits are obtained using a multivariate t (normal) distribution. In Phase II, two types of corrected limits are provided, following the recent literature, one based on the unconditional perspective and the other based on the conditional perspective and the exceedance probability criterion (EPC). Results, tabulations and illustrations with data make the methodologies employable in practice. An R package is provided to help deployment of the new methodology.

Key Words: Shewhart control chart; Statistical process control; Moving range; Phase I; Phase II; Conditional run length

1. Introduction

Statistical Process Control (SPC) and monitoring methods are widely used in various industries. Many famous companies such as IBM, Phillips, Mercedes, GM and GE have successfully taken advantage of the technology and have revolutionized their product lines. Design of Experiments (DOE) is an important component of SPC where various mathematical models are used for the phenomenon under study and the resulting output is monitored with control charts. So control charts need to be developed for the type of model one uses, the parameter(s) of interest, along with the assumptions on the error term. Over twenty years ago, Roes and Does (1995), Woodall and Thomas (1995) and Park (1998), among others, considered various control charts including the Shewhart $\bar{X}$ charts (the simplest and the most popular control charts) for monitoring processes that follow some general linear models, with multiple sources of variation, that

[1] Yuhui Yao is a PhD student in the Department of ISM at the University of Alabama, Tuscaloosa, AL. His email address is yyao17@crimson.ua.edu

[2] Subha Chakraborti is a professor in the Department of ISM at the University of Alabama, Tuscaloosa, AL. His email address is schakrab@cba.ua.edu

are commonly used in DOE. These models reflect some processes under study, the sources and the types of variations of interest, more accurately. Such models are used even today. While these have been remarkable beginnings, we need to bring this technology to the twenty first century, in view of the knowledge gained by the researchers in process monitoring and control over the last couple of decades. This includes what are now known as the retrospective phase (Phase I) and the prospective phase (Phase II) of SPC, where the goals of process monitoring are somewhat different. In addition, there is a wealth of knowledge about the effects of parameter estimation, which is most common while setting up the control charts in practice, how it affects chart performance and how to adjust the control limits in order to achieve nominal in-control and reasonable out-of-control chart performance. In this paper, we consider these issues and provide correct Phase I and Phase II control limits for the Shewhart $\bar{X}$-chart for a model more general than the basic Shewhart model. Under this model there are two sources of common cause variation, the variation between and the variation within the batches. In the first section we describe the models, starting from the basic Shewhart model to the model under consideration, and provide the background for our methodology.

1.1 The Basic Shewhart Model

The basic Shewhart model (Montgomery, 2009) is given by

$$X_{tj} = \mu + E_{tj} \qquad (1)$$

where X_{tj} is the j^{th} observation, $j = 1, \dots, n$, in the t^{th} batch (subgroup), $t = 1, \dots, k$, μ is the process mean and E_{tj} is the random error that is assumed to follow a normal distribution with mean, 0 and an unknown variance, σ_e^2. This is the simplest and the most commonly studied model in SPC. The plotting statistic for the Shewhart $\bar{X}$-charts for this basic model is the subgroup mean $\bar{X}_{t.} = \frac{1}{n}\sum_{j=1}^{n} X_{tj}$ and the error variance σ_e^2 is estimated by the within batch variation $S_t^2 = \frac{1}{(n-1)}\sum_{j=1}^{n}(X_{tj} - \bar{X}_{t.})^2$. Over the last twenty plus years a great deal of knowledge has been acquired on how to monitor both the mean μ and the variance σ_e^2. We will expand on some of these advances a little later in the paper. For example, when μ is unspecified or is unknown, along with σ_e^2, the Shewhart $\bar{X}$-chart control limits for the mean thickness μ are given by $\bar{\bar{X}} \pm \frac{a}{\sqrt{n}}\sqrt{S_p^2}$ where $\bar{\bar{X}}$ and the S_p^2 are, respectively, the grand mean and the pooled variance of m Phase I sample variances, each sample of size n, collected when the process is deemed to be in-control. The charting constant a is determined so that the control chart has a specified nominal in-control average run length.

Roes and Does (1995), hereafter written as R&D, showed that there are interesting applications in practice where it is more reasonable to adopt a model beyond the basic Shewhart model. Their methodology was motivated by a case study with some data from Philips Semiconductors, a leading manufacturer of integrated circuits in Europe. The thickness of silicon wafers, used in the production process of integrated circuits was the product of measurement interest. A data set containing 30 batches with 5 observations in each batch from different grinder positions of wafer thickness is provided. The traditional Shewhart $\bar{X}$-chart and R-chart were employed demonstrating in-control behavior. However, deeper analysis uncovered batch to batch variation, attributed possibly to the grinder positions, which necessitated fitting a model with a term representing the batch effect. This is where the basic Shewhart model needed to be extended and R&D considered a linear model which includes a random batch component to accommodate the variation among the batches. So there is possible variation both between batches and within batches. This point was also introduced by Woodall and Thomas (1995). Thus, in summary, while the basic Shewhart model is useful and has been studied extensively in the literature, other models can be perhaps more applicable and useful in some applications. Hence it is important to consider process monitoring under more general models and we consider the model studied in R&D in the next section. Other, more sophisticated models can also be studied along the same lines and will be considered elsewhere.

1.2 A More General Model with Two Components for Common Cause Variability

R&D state that "the observations from the example suggest the following general model of a state of statistical control, when the fixed differences within a batch, as well as the random between-batch variation, both need to be considered as a part of the common-cause variation:

$$X_{tj} = \mu + w_j + B_t + E_{tj} \qquad (2)$$

where X_{tj} are the thickness measurements on the jth wafer, $j = 1, \ldots, n$, within the tth batch $t = 1,2, \ldots, k$. The $\mu + w_j$ term is the fixed mean level ($\sum w_j = 0$) for the jth wafer and B_t represents effect of the tth batch which is assumed to be random. Finally, E_{tj} models the random error for the jth wafer within the tth batch. It is assumed that the random variables B_t and E_{tj} are mutually stochastically independent, and both are normally distributed with means 0 and variance σ_b^2 and σ_e^2, respectively."

We consider Phase I and Phase II Shewhart $\bar{X}$-charts for the model in (2), which has two components of common cause variation; further generalization to other settings (and models for other designed experiments) such as the one in Woodall and Thomas (1995) and Park (1998) follow naturally and will be taken up elsewhere. In a similar manner, as in R&D, charts for certain contrasts of interest can be considered and adapted as well. Control charts for contrasts have been considered by several researchers, including Palm and De Amico (1995), Woodall and Gardner (1995), Blazek and Coleman (1995), and Hahn and Doganaksoy (1995) among others. We agree that the contrasts can be useful in process monitoring provided they have a meaningful interpretation. However, we don't pursue this line of work here. Note that other researchers have considered monitoring data from general models such as from a nested-random-effect model, which has two random effects. Such a model was used by Yashchin (1994) and a CUSUM chart was considered. Furthermore, note that the idea of Shewhart control charts for more general models has also been widely applied and extended in different areas such as manufacturing (Kim & Yum, 1999; Park, 2004) and the dairy industry (Ittzes, 2001). It will be interesting to apply our work to these situations and other types of control charts and this will be pursued elsewhere.

The proposed adaptations revolve around the central question of the charting constants that have been used for these charts until now. In a vast majority of the Shewhart $\bar{X}$-charts type charts, the traditional "3-sigma" limits have been used which raises questions about chart performance. The choice of 3 to be the charting constant does not conform to our current knowledge and understanding of retrospective and prospective process monitoring. Thus we find the "correct" charting constants and set up the corrected control limits. We do this for both in Phase I and in Phase II, using some well-accepted performance metrics, taking proper account of the effects of parameter estimation using some different perspectives. Our study and results provide both theoretical insights and practical results since the *correct* Phase I and Phase II analyses using control charts are important. To the best of our knowledge, this has not been considered in the literature. The Shewhart $\bar{X}$-charts are used in our work for simplicity and purposes of illustration. Other charts such as the CUSUM and the EWMA can be considered in a similar way. Our formulations and derivations can open the door for correctly monitoring the outputs of designed experiments with various linear models under the assumption of i.i.d. normal errors.

Consistent with the recent literature, we distinguish between Phase I and Phase II monitoring and set up both Phase I and Phase II control charts which are applicable in a situation as the one considered by R&D. Note that although some researchers suggest following the "3-sigma" rule, that is, using $c = 3$, this is not correct either in Phase I or II. This is because although the Shewhart Phase I and II limits have the same structure as Montgomery (2009) introduced, the charting constant is found using a different in-control performance metric in the two monitoring phases. In Phase I, the retrospective phase, one uses the false alarm probability (FAP), defined as the probability of at least one false alarm to design control charts, whereas in Phase II, the prospective phase, a popular choice is the in-control average run-length. These two criteria produce different constants, and therefore different control limits and charts as will be seen later. We delineate these two cases (Phase I and II) and treat them separately as follows.

2. Methodology

As noted earlier, R&D considered charts for the model in equation (2) for a given false nominal alarm rate and in addition to the unclear separation of Phases I and II, the effects of parameter estimation were not explicitly factored in the determination of the charting constants. This raises several performance issues of concern in light of the recent literature. First, it is now agreed that different in-control chart performance metrics are to be used in the two phases. Second, since the mean and the standard deviations are unknown, they need to be estimated from some in-control reference data from a Phase I analysis where Shewhart control charts of the same form are used on a trial basis (Montgomery, 2009) and the process is brought under control before any process monitoring based on control charts can start (Montgomery, 2009). Third, from a technical standpoint, when monitoring begins, the presence of the estimates of the mean and the variance in the control limits that the charting statistics are compared to, violates the independence of the so-called "signaling" events in both Phase I and II monitoring. As a consequence, the correct charting constant needs to be found after accounting for this dependence in both phases. In Phase I, this implies that one has to work with the joint distribution of the charting statistics. At the end of the Phase I analysis, when the in-control state is established, we have the parameter estimates to form the Phase II control charts. However, since the signaling events are dependent, the Phase II run-length distribution is not geometric (Quesenberry, 1993), the false alarm rate and the average run-length, two of the most popular chart performance measures, depend on the parameter estimators and thus are random variables. Thus these chart characteristics vary, sometimes very significantly, from estimator to estimator obtained from different Phase I samples. This variability, called the practitioner to practitioner variability, present challenges to design control charts with guaranteed in-control performance.

One recommendation to address this performance variation, which arises as a result of using estimators in the Phase II control limits, is to find the Phase II charting constant for a given nominal in-control average run-length, averaging over the distributions of the parameter estimators. This is the *unconditional* or the *marginal* approach (perspective). On this point, it is worth noting that in the discussion of the Roes and Does paper, Sullivan et al. (1995) actually noted, in their Point 3, that "There is much emphasis in the article, as well as in Roes, Does, and Schurink (1993), on obtaining control constants that will provide an exact specified probability of a false alarm. The probability obtained using the authors' control-limit constants appears to be the probability that a single future observation falls outside the control limits with no conditioning on the control limits actually obtained in an application. There is a complicating factor, however, which appears to have been overlooked. As Ghosh et al. (1981) and later Quesenberry (1993) pointed out, there is a lack of independence in the successive values of $\frac{(T_r - T)}{V_t}, t = k+1, k+2, \ldots$ This lack of independence, due to the common estimates of the mean and variance, implies that the run-length distribution can no longer be characterized by the probability of a signal. The run length is no longer geometric and, in general, the average run length is not the reciprocal of the signal probability as presented in their tables 3 and 4."

Thus, in our paper, we address these important concerns raised more than twenty years ago and find the proper and correct solutions. To the best of our knowledge and based on a thorough review of the literature, we are not aware of works addressing these concerns.

2.1 Phase I Shewhart $\bar{X}$ Chart for the Linear Model with Two Components for Common Cause Variability

Recall that the objective of a Phase I analysis (Chakraborti et al. 2009, Jones-Farmer et al. 2009, Capizzi and Masarotto, 2013) is to establish process control, on the basis of retrospective data that may be available, using statistical tools including control charts (applied possibly more than once, on a trial basis) and thus generating reference data from which knowledge about the process (such as the shape of the underlying distribution, etc.) is gained and parameter estimates are calculated. These estimates then are used in the construction of Phase II control charts and in prospective process monitoring. During a Phase I study, practitioners perform exploratory analysis to better understand the nature of the process. The more representative the parameter estimates are of the true process parameters, the better the performance of

Phase II control charts that use these parameter estimates. Thus, as an alternative approach, the wafer thickness data provided by R&D will be treated as Phase I data and the Shewhart Phase I $\bar{X}$-chart (Champ and Chou, 2003; Champ and Jones, 2004) will be applied with the proposed (correct) charting constants. The Phase I charts are designed for a given false alarm probability (FAP). The normal theory-based Phase I Shewhart $\bar{X}$ charts were studied by Champ and Jones (2004) and that work was recently extended by Yao et al. (2017). The FAP is the probability of at least one false alarm among the subgroups examined. Typical nominal values of FAP include 0.05 and 0.10.

We first determine the correct Phase I control limits for the $\bar{X}$ -chart for the model in equation (2). Phase I Shewhart $\bar{X}$ chart is given by the plotting statistic $\bar{X}_{t.}$ and the control limits

$$LCL = \bar{\bar{X}}_{..} - c_I^*\,\hat{\sigma}, CL = \bar{\bar{X}}_{..} \text{ and } UCL = \bar{\bar{X}}_{..} + c_I^*\,\hat{\sigma} \tag{3}$$

where $c_I^* = c^*(k, 1 - FAP_0)$ is the required Phase I charting constant, $\hat{\sigma} = s_b(X)/c_4 = \sqrt{\frac{1}{k-1}\sum_{t=1}^{k}(\bar{X}_{t.} - \bar{\bar{X}}_{..})^2}\,/c_4$ is based on the root-mean-square deviation recommended by recent research (see for example Mahmoud et al. 2010), and $c_4 = c_4(\nu) = \left(\frac{2}{\nu}\right)^{1/2}\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}$ is the unbiasing constant with $\nu = k - 1$.

In order to implement this chart, the charting constant c_I^* needs to be found for a given nominal FAP value, say FAP_0. To this end, note that the Phase I Shewhart $\bar{X}$ chart issues no signal for batch t if

$$\bar{\bar{X}}_{..} - c_I^*\,\hat{\sigma} \le \bar{X}_{t.} \le \bar{\bar{X}}_{..} + c_I^*\,\hat{\sigma}, \quad t = 1,2,\ldots.,k \tag{4}$$

This event, in equation (4), when there is no signal, is called a non-signaling event. The non-signaling event in equation (4) can be expressed as

$$-c_I^*\hat{\sigma} \le \bar{X}_{t.} - \bar{\bar{X}}_{..} \le c_I^*\hat{\sigma}, \quad t = 1,2,\ldots,k \tag{5}$$

Or, equivalently, as

$$-c_I^* \le \frac{\bar{X}_{t.} - \bar{\bar{X}}_{..}}{\hat{\sigma}} \le c_I^*, \quad t = 1,2,\ldots,k \tag{6}$$

Both the non-signaling events (and their complements, the signaling events) are dependent events since they share the estimators of the mean and the standard deviation. Hence the Phase I charting constant is given by

$$c_I^* = c_4\sqrt{\frac{k-1}{k}}\,l \tag{7}$$

where l is the solution of the equation

$$\int_{-l}^{l}\ldots\int_{-l}^{l} f_{MVT_k(\upsilon,\underline{P})}(u_1 \ldots u_k)du_1 \ldots du_k = 1 - FAP_0 \tag{8}$$

which follows from equation (6). Here $f_{MVT_k(\upsilon,\underline{P})}$ denotes the p.d.f. of a k dimensional t distribution (Johnson and Kotz, 1972) with $\upsilon = k - 1$ degrees of freedom and a kxk equal-correlation matrix $\underline{P}$ and $\underline{P}$ has off-diagonal elements all equal to $-\frac{1}{k-1}$.

Equivalently, it can be shown that one can also obtain l from the solution of

$$\int_0^{\infty}\left\{\int_{-l(y)}^{l(y)}\ldots\int_{-l(y)}^{l(y)} f_{MVN_k(\underline{0},\underline{P})}(u_1 \ldots u_k)\,du_1 \ldots du_k\right\} f_{\chi^2_{k-1}}(y)dy = 1 - FAP_0 \tag{9}$$

Here, $f_{MVN_k(\underline{0},\underline{P})}$ denotes the p.d.f. of the k-dimensional normal distribution with mean $\underline{0}$ and the kxk equal-correlation matrix $\underline{P}$ and $f_{\chi^2_{k-1}}$ denotes the p.d.f. of a chi-square distribution with $\upsilon = k - 1$ degrees of freedom. We call equation (8) as *the direct method*, and equation (9) as *the indirect method*, also known as the method under the unconditional perspective. Though both methods are seen to yield similar results, the direct method is much faster for computation. The unconditional perspective has been used in other situations in the literature and one advantage with this method is that it is possible to adopt other standard deviation estimators, such as the average moving range, or the average standard deviation (see Montgomery, 2009) whose distributions can be approximated by the square root of a chi-square (chi) distribution in this framework.

In table 1 we show the Phase I charting constants for the Shewhart $\bar{X}$-chart when the unbiased estimator of the process standard deviation is used based on the pooled variance S_b^2 given in equation (3).

Table 1: Two-sided charting constant c_I^* for a Shewhart Phase I $\bar{X}$-chart for the mean, for different numbers of Phase I batches and nominal FAP_0 values 0.05 and 0.10, respectively, when the unbiased standard deviation estimator is used. The charting constants are calculated by the direct method.

	FAP_0			FAP_0			FAP_0	
k	0.05	0.1	k	0.05	0.1	k	0.05	0.1
3	4.2601	2.9321	9	3.2732	2.8179	50	3.4217	3.1727
4	3.6679	2.7923	10	**3.2635**	2.8349	100	3.5569	3.3370
5	3.4793	**2.7664**	15	3.2644	2.9080	150	3.6423	3.4343
6	3.3692	2.7743	20	3.2862	2.9665	200	3.7044	3.5036
7	3.3173	2.7846	25	3.3139	3.0156	250	3.7536	3.5584
8	3.2872	2.8006	30	3.3384	3.0555	300	3.7937	3.6021

From table 1, it is noted that the charting constants are quite different from 3 and hence the popular "3-sigma" rule is not correct to be applied in the Shewhart Phase I $\bar{X}$-chart for the mean. Even for k = 300, the charting constant is 3.7936 and more remarkably the constant does not converge to 3.

Next we present an example, using the data from R&D, to illustrate the proposed Phase I Shewhart $\bar{X}$ chart.

2.2 Illustration of the Phase I Shewhart $\bar{X}$ Chart with the Correct Charting Constant

For the case data presented in R&D, the wafer thickness measurements are collected from the MPS-R600 grinder with 31 positions with a target thickness is 244 μm. The dataset has $k = 30$ batches and each batch has $n = 5$ observations which are taken from 5 of the 31 positions on the grinder. We treat these data as Phase I data for which the grand mean $\bar{\bar{X}}$ equals 245.1 and the estimator s_b equals 2.0367, which results in the unbiased estimator $\hat{\sigma}_0 = s_b/c_4 = 2.0544$, where $c_4 = 0.9914$ (Montgomery, 2009), for $k = 30$. The Phase I Shewhart $\bar{X}$-chart control limits are calculated using degrees of freedom $\nu = 30 - 1 = 29$ and a nominal FAP_0 of 0.05 and 0.1, respectively. The resulting charting constants are found from table 1 to be 3.3384 and 3.0555, respectively, which yield the pairs of lower and upper control limits: 238.2417 and 251.9583 as well as 238.8229 and 251.3771, respectively. The corresponding Phase I charts, including the control limits and the charting statistics, $\bar{X}_i$ are displayed in figure 1. In addition, the control limits (the solid lines) using the unadjusted charting constant $c_I^* = 3$, namely 238.9369 and 251.2631, are also shown in figure 1, as a reference.

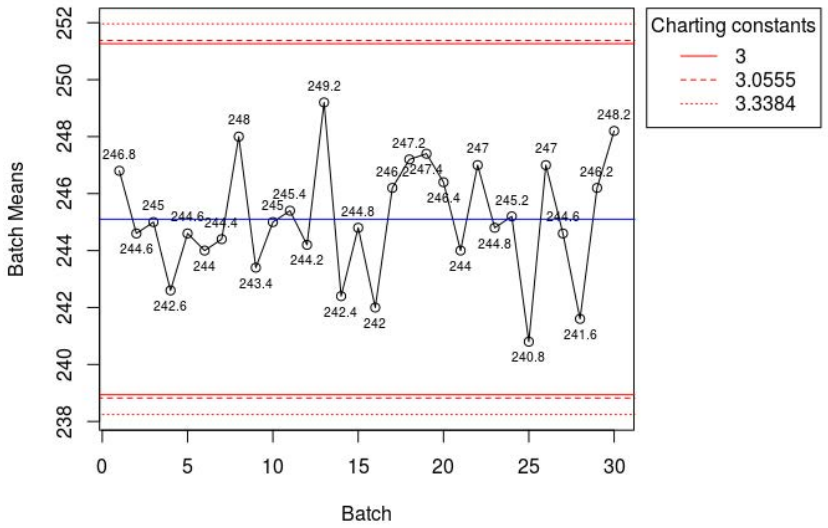

Figure 1: Phase I Shewhart $\bar{X}$ charts for the R&D data on the thickness of wafers where the blue solid line is the center line and the red solid lines are the control limits for the unadjusted charting constant 3. The dashed (dotted) lines are the corrected Phase I control limits for a nominal FAP_0 of 0.1 (0.05). The batch means are shown by the open circles with the circles connected by straight lines.

It is seen that both the corrected Phase I charts provide much support to the in-control state of the process, since no charting statistic plots outside the control limits and no patterns seem to be present. In fact, for these data, even the unadjusted control limits lead to the same conclusion. Accordingly, these data can be regarded as reference data and the estimates obtained for the process mean (245.1) and the standard deviation (2.0544) may be used in the construction of Phase II (prospective) control limits for process monitoring.

While the Phase I analysis helps understand the process better and the Phase I charts help establish process control in order to help collect the reference data, from which to calculate the parameter estimates, the prospective process monitoring, with new and independent incoming samples, starts in Phase II. The difference between Phase I and in Phase II charts is that while the available retrospective data is analyzed to establish control in Phase I, possibly on a trial and error basis, using control charts, in Phase II each incoming subgroup is monitored sequentially. Chakraborti (2000) analyzed the Shewhart $\bar{X}$ chart in four categories depending on which of the two parameters, the mean or the standard deviation of the underlying normal distribution, are known or unknown, here we only develop Shewhart $\bar{X}$ chart for the situation where both the mean and the standard deviation are unknown and are estimated from Phase I data. The remaining cases can be treated in a similar way.

Next we consider the Phase II Shewhart $\bar{X}$-chart for the mean for the model in equation (2).

2.3 Two-sided Phase II Shewhart $\bar{X}$ Chart for the Linear Model with Two Components for Common Cause Variability

The Phase II Shewhart $\bar{X}$ control chart for the mean is given by the plotting statistic $\bar{X}_{t.}$ and the control limits

$$LCL = \bar{\bar{X}}_{..} - c_{II}^{*}\,\hat{\sigma}, CL = \bar{\bar{X}}_{..} \text{ and } UCL = \bar{\bar{X}}_{..} + c_{II}^{*}\,\hat{\sigma} \qquad (10)$$

where $c_{II}^{*} = c_{II}^{*}(k, ARL_0)$ is the Phase II charting constant and $\bar{\bar{X}}_{..}$ and $\hat{\sigma} = s_b(X)/c_4$ are estimated by the mean and the standard deviation, respectively, in Phase I. Operationally, the main difference between the Phase I and Phase II control limits for the $\bar{X}$ chart are the charting constants, which boils down to the underlying chart performance metric that is used to construct the charts in each phase. Unlike in Phase I,

the charting constant c_{II}^* in Phase II is determined so that the in-control average run length ARL_0 is equal to some nominal value, such as 370. Other metrics of the in-control run length distribution, such as some percentile (for example, the median) can also be considered, but we do not pursue these here.

Note that using equation (10), the non-signaling event in Phase II is given by

$$\bar{X}_{..} - c_{II}^*\hat{\sigma} < \bar{X}_{s.} < \bar{X}_{..} + c_{II}^*\hat{\sigma}, \qquad s = k+1, k+2, \ldots \tag{11}$$

the Phase II run length distribution and the corresponding in-control average run length can be derived using the conditioning-unconditioning (CUC) method. This method recognizes that the parameter estimates are random variables and therefore the run-length distribution, as well as various performance characteristics (attributes of the run-length distribution) of the Phase II control chart are random variables, as functions of these estimates. As a result, one can consider the *unconditional* performance of the chart (for example, the unconditional run-length distribution, the unconditional average run-length, and so on) which basically averages the distributions of these characteristics over the distributions of the parameter estimates. To this end, a convenient starting point is the probability of a signal conditional on the Phase I parameter estimates, $\hat{\mu}$ and $\hat{\sigma}$. This is given by

$$\begin{aligned} P(|\bar{X}_{s.} - \bar{X}_{..}| \geq c_{II}^*\hat{\sigma}) &= 1 - P(|\bar{X}_{s.} - \bar{X}_{..}| < c_{II}^*\hat{\sigma}) \\ &= 1 - \left(\Phi\left(\frac{Z_1}{\sqrt{k}} + \frac{c_{II}^*}{c_4}\sqrt{\frac{Y}{k-1}} - \delta \right) - \Phi\left(\frac{Z_1}{\sqrt{k}} - \frac{c_{II}^*}{c_4}\sqrt{\frac{Y}{k-1}} - \delta \right) \right) \end{aligned} \tag{12}$$

where Φ denotes the c.d.f. of the standard normal distribution, Y follows a chi-square distribution with $k-1$ degrees of freedom associated with the Phase I variance estimator, Z_1 follows the standard normal distribution associated with the Phase I sample mean, Z_2 follows the standard normal distribution associated with the Phase II sample mean, and Y, Z_1 and Z_2 are mutually independent. The quantity $\delta = (\mu_2 - \mu)/\sigma$ denotes the mean shift in Phase II in terms of the standard deviation and μ_2 denotes the Phase II mean. If the process is in-control, $\mu_2 = \mu$ so that $\delta = 0$. Thus, the conditional probability of a signal when the process is in-control, the so-called conditional false alarm rate, denoted by $CFAR$, is given by

$$CFAR = P(|\bar{X}_{s.} - \bar{X}_{..}| \geq c_{II}^*\hat{\sigma}|IC) = 1 - \left(\Phi\left(\frac{Z_1}{\sqrt{k}} + \frac{c_{II}^*}{c_4}\sqrt{\frac{Y}{k-1}} \right) - \Phi\left(\frac{Z_1}{\sqrt{k}} - \frac{c_{II}^*}{c_4}\sqrt{\frac{Y}{k-1}} \right) \right) \tag{13}$$

Further, using the Probability Integral Transform (PIT), we can rewrite the $CFAR$ ($0 \leq CFAR \leq 1$) more conveniently as

$$\begin{aligned} CFAR &= 1 - \left(\Phi\left(\frac{\Phi^{-1}(U)}{\sqrt{k}} + \frac{c_{II}^*}{c_4}\sqrt{\frac{F_{\chi^2_{k-1}}^{-1}(V)}{k-1}} \right) - \Phi\left(\frac{\Phi^{-1}(U)}{\sqrt{k}} - \frac{c_{II}^*}{c_4}\sqrt{\frac{F_{\chi^2_{k-1}}^{-1}(V)}{k-1}} \right) \right) \\ &= 1 - p(U,V;k,c_{II}^*) = q(U,V;k,c_{II}^*) = 1 - p(U,V,k,c_{II}^*) = q(U,V), \end{aligned} \tag{14}$$

say, where Φ^{-1} denotes the quantile function of the standard normal distribution, $F_{\chi^2_{k-1}}^{-1}$ denotes the quantile function of a chi-square distribution with $k-1$ degrees of freedom, $p(U,V;k,c_{II}^*)$ is the probability of no signal under the in-control condition, and U and V are two independent and identically distributed uniform (0,1) random variables. Unless there is any confusion, $p(U,V;k,c_{II}^*)$ and $q(U,V;k,c_{II}^*)$ are written as $p(U,V)$ and $q(U,V)$, respectively. Note that the $CFAR$ is a random variable, being a function of the random variables Z_1 and Y (or U and V). The c.d.f. of the $CFAR$ is given by

$$F_{CFAR}(t;k,c_{II}^*) = P(CFAR \leq t) = P(1 - p(U,V) \leq t) = P(q(U,V) \leq t) \tag{15}$$

The notation $F_{CFAR}(t;k,c_{II}^*)$ will be written as $F_{CFAR}(t)$ until necessary.

Epprecht et al. (2015) derived the cumulative distribution function (cdf) of the $CFAR$ for the one-sided Phase II S charts for the Shewhart model and used it to find a lower prediction bound and the required Phase I sample size for the Shewhart model. Goedhart et al. (2017), for example, studied the $\bar{X}$-chart for the basic Shewhart model. We use a similar approach for the Shewhart $\bar{X}$-chart for the mean of our more

general model. Since the conditional in-control run-length distribution is geometric with the success probability equal to $CFAR$, various performance characteristics for the conditional run-length distribution can be obtained using properties of the geometric distribution. Using this fact, the Phase II charting constants are determined in the next section, under each of the unconditional and the conditional perspectives.

2.3.1 Phase II Shewhart $\bar{X}$ Charting Constant under the Unconditional Perspective

Chakraborti (2006) derived the Phase II charting constant for the Shewhart $\bar{X}$ chart with estimated parameters, for the basic Shewhart model, using the unconditional in-control average run length. This is called the *unconditional perspective*. Under this method, the charting constants are found by equating *the unconditional in-control average run length* (denoted ARL_{in}), which is the expectation of the *conditional* average run length $CARL_0$, which is ARL_{in} to a desired nominal value say ARL_0. The nominal values are taken to be 370 or 500, as in the known parameters case. Because the $CARL_0$ is the reciprocal of the $CFAR$

$$CARL_0 = CFAR^{-1} = [1 - p(U,V)]^{-1} = [q(U,V)]^{-1}, \quad 1 \le CARL_0 < \infty \tag{16}$$

Hence the *unconditional in-control average run length* ARL_{in} can be expressed as

$$ARL_{in} = E(CARL_0) = E(CFAR^{-1}) = \int_0^1 \int_0^1 [q(u,v)]^{-1} du dv, \quad 1 \le ARL_{in} < \infty \tag{17}$$

Note that in order for the expectation ARL_{in} to be finite, the probability $q(u,v)$ must belong to the left-open interval $(0,1]$. However, being a probability, by definition, $q(u,v)$ belongs to the closed interval $[0,1]$. The inclusion of the left end-point 0, which is possible under certain situations, as will be seen later, causes a singularity, which affects the computation of the ARL_{in}. Thus whenever $q(u,v)$ is equal to 0 or close to being equal to 0, which happens in some cases when the k (number of Phase I batches) is small, the double integral in equation (17) becomes unstable and hence the computation of the ARL_{in} becomes infeasible. To clarify our argument, we demonstrate this problem by using the following example. Suppose we follow the "3-sigma" rule for our nominal average run length $ARL_0 = 370$ and $ARL_0 = 500$, so the corresponding false alarm rate $FAR = 0.0027$ and $FAR = 0.002$. We will visualize this situation using the following figure.

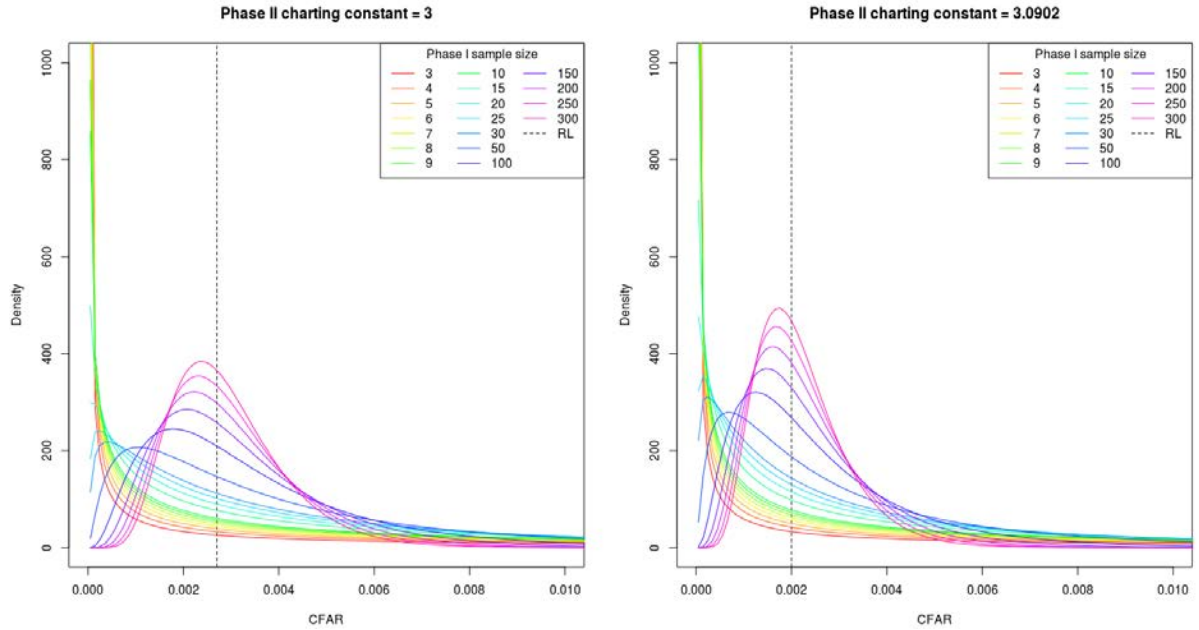

Figure 2: Density plots of midpoints of $CFAR$ for different numbers of Phase I batches with the bandwidth 0.0001, where 'RL' on the legend stands for the reference lines. For Phase II charting constant $c_{II}^* = 3$, $1/370 = 0.0027$ and, for Phase II charting constant, $1/500 = 0.002$.

We can observe that the density of $CFAR$ is seriously inflated in the interval close to 0 for small $k < 10$, so $CARL_0$, the inverse of $CFAR$, maybe not finite. That is the expectation ARL_{in} may be infinite or at least greater than our nominal average run length ARL_0. This singularity problem can be released by increasing numbers of Phase I batches or decreasing Phase I charting constants.

To correct the charting constant c_{II}^*, the new approach is based on solving the equation $ARL_{in} = ARL_0$. From equation (17), equivalently, one solves,

$$ARL_{in} - ARL_0 = \int_0^1 \int_0^1 [q(u,v)]^{-1} dudv - ARL_0 = \int_0^1 \int_0^1 [q(u,v;k,c_{II}^*)]^{-1} dudv - ARL_0 = 0 \quad (18)$$

using a search algorithm, such as the bisection method, for a given nominal ARL_0 and for a given value of k. This yields the unconditional Phase II charting constant. The double integral in (18) is solved using two methods, one, the adaptive integration (AI) method and two, the Monte Carlo (MC) method. The AI method cannot cover the whole domain in our computation. The domain covers from $[1.00\text{E}-11, \infty)$ and as the result from the AI method forms the upper bound to the true charting constant, the result from the MC method may be more reliable. For $k > 10$, both AI and MC methods produce similar results but the AI method is much more time-efficient. However, when k is 10 or less, the charting constants are a little suspect because of the larger discrepancies between the $E(CARL_0)$ and their nominal values, and also because of the large $SD(CARL_0)$ values. This discrepancy is caused by the variation in the distribution of the in-control average run length, given by $SD(CARL_0)$, which is exacerbated by the singularity problem for small k. As explained before, when k is small, the numerical results for the in-control average run length $E(CARL_0)$ are unstable, leading to a huge variation in the $SD(CARL_0)$ values. In order to reduce this variation further, to a manageable amount, the number of simulations was increased to 1-billion, but the results were found to be similar. Hence, we calculate and tabulate the charting constants for $k \geq 15$ in table 2 for less misunderstanding. The results shown in table 2 are obtained by the MC method.

Table 2: Two-sided Phase II Shewhart $\bar{X}$ charting constant c_{II}^*, based on the unconditional perspective, for different Phase I batch numbers k and nominal $ARL_0 = 370$ and 500, respectively, when the unbiased standard deviation estimator is used, along with the mean and standard deviation of $CARL_0$, that is $E(CARL_0)$ and $SD(CARL_0)$, respectively.

	$ARL_0 = 370$			$ARL_0 = 500$		
k	c_{II}^*	$E(CARL_0)$	$SD(CARL_0)$	c_{II}^*	$E(CARL_0)$	$SD(CARL_0)$
15	2.5571	370.4349	10381.82	2.6102	500.9479	19222.28
20	2.6665	370.2784	3633.735	2.7278	500.4716	6543.182
25	2.7330	370.1403	1869.674	2.7996	500.1107	3198.94
30	2.7776	370.1043	1228.349	2.8479	499.9972	2008.646
50	2.8669	369.8962	575.3241	2.9451	499.9165	869.4891
100	2.9337	369.8861	313.6814	3.0180	499.8732	457.6177
150	2.9558	369.8729	238.4686	3.0422	499.8916	345.055
200	2.9669	369.9874	199.8089	3.0543	499.9705	288.004
250	2.9735	370.002	175.2922	3.0615	499.9406	252.1142
300	2.9778	369.8974	157.9526	3.0663	499.9339	226.9487

For k greater or equal to 15, note that the charting constants yield target nominal in-control average run length ARL_0 equal to 370 or 500, respectively, consistently, on the basis of $E(CARL_0)$ and a smaller $SD(CARL_0)$ for increasing k. In fact, for k greater than or equal to 15, the charting constants seem reasonably acceptable.

Now we discuss the out-of-control (OOC) performance.

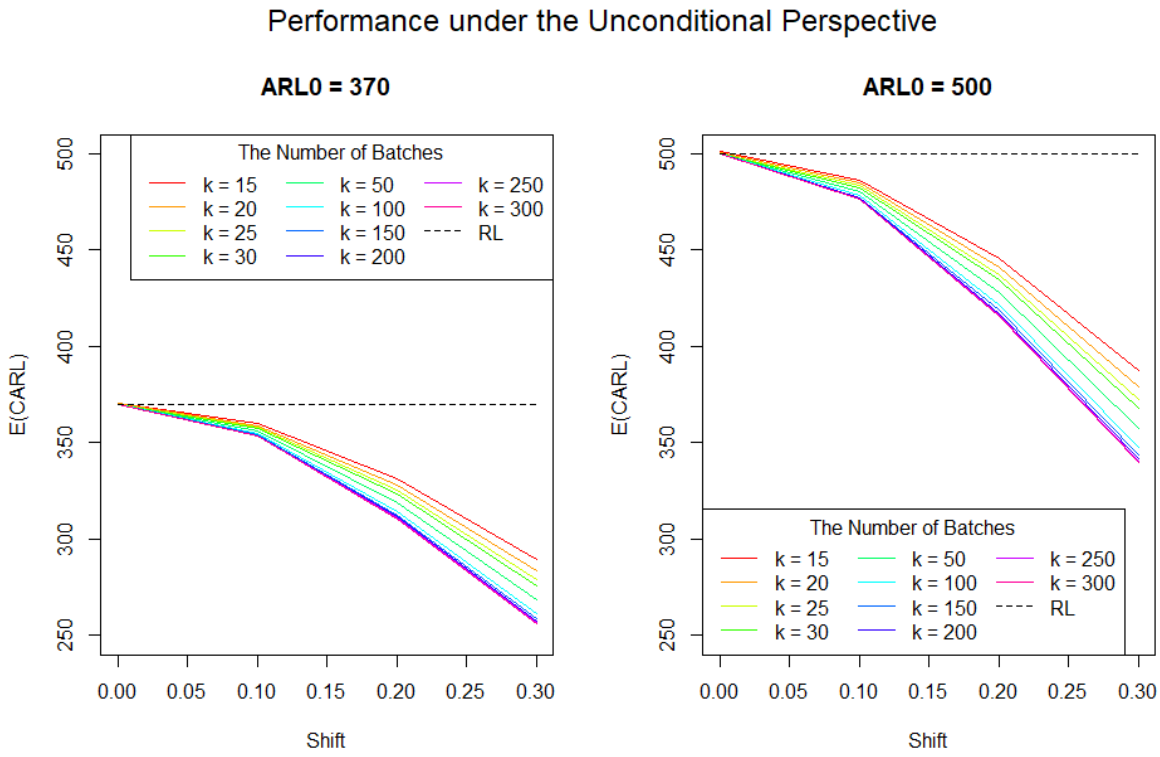

Figure 3: Two-sided Phase II Shewhart $\bar{X}$ control charts out-of-control average run length based on the unconditional perspective, for different Phase I batch numbers k, nominal $ARL_0 = 370$ and 500, and shifts $\delta = 0.1, 0.2$ and 0.3, respectively, where solid curves for different values of k have different colors as written on the legend. Note that the black dashed lines are the reference lines for different nominal $ARL_0 = 370$ and 500, respectively.

From figure 3, they demonstrate two facts: 1) OOC average run length decreases with increasing shifts, which is desirable. 2) The sensitivity of our control charts is improved by increasing the number of Phase I batches. From the view of efficiency of using k, CUC method is more suitable for the relative great number of Phase I batches.

2.3.2 Phase II Shewhart $\bar{X}$ Chart under the Conditional Perspective and the Exceedance Probability Criterion (EPC)

As $CARL_0$ is a random variable with a large variability, it is quite likely to be different (sometimes significantly) from its expectation, the ARL_0, as specified via the nominal value in the construction of the chart under the unconditional perspective. This may be problematic from a practical point of view. In the long run, the chart (charting constant) obtained under the unconditional perspective guarantees that the average chart performance reaches the target ARL_0 but it does not control the variability. A more recent alternative option recognizes the randomness of $CARL_0$ and considers various solutions. To this end, several authors including Albers et al. (2005), Epprecht et al. (2015) and Goedhart et al. (2017), have considered setting up the control chart limits so that the $CARL_0$ has a high probability of exceeding a given nominal value (a lower bound) such as 370. This formulation is related to the *exceedance probability criterion* (EPC) proposed by Gandy and Kvaloy (2013) in which one finds an upper bound (that can be tolerated in an application) to the random variable $CFAR_0$ with a high probability. Thus

$$P(CFAR \leq (1+\varepsilon)FAR_0) = P(CARL_0 \geq (1-\tilde{\varepsilon})ARL_0) \tag{19}$$

$$= P(CARL_0 \geq ARL_b) = P([1 - p(U,V)]^{-1} \geq (1 - \tilde{\varepsilon})ARL_0)$$
$$= P\left([q(U,V)]^{-1} \geq \left(\frac{1}{1+\varepsilon}\right)ARL_0\right) = 1 - p_0,$$

where p_0 $(0 < p_0 < 1)$ is a specified probability, typically a small value, such as 0.1 or 0.05, ε $(0 \leq \varepsilon < 1)$ is a "tolerance factor" allowing the user some flexibility for choosing a nominal conditional false alarm rate and $\tilde{\varepsilon} = \frac{\varepsilon}{1+\varepsilon}$ is the corresponding tolerance factor for a nominal conditional in-control average run length. The quantity $ARL_b = (1 - \tilde{\varepsilon})ARL_0 = \left(\frac{1}{1+\varepsilon}\right)ARL_0$ is a lower prediction bound to $CARL_0$ (equivalently, $(1 + \varepsilon)FAR_0$ is the upper prediction bound to $CFAR$). From equation (19), it is seen that ARL_b is the p_0-quantile of the distribution of $CARL_0$. So the lower prediction bound can be found from the c.d.f. of $CARL_0$.

Note that from a practical standpoint, we may be more interested in the lower prediction bound to $CARL_0$ (which is equivalent to an upper bound to $CFAR$), because this bound guarantees that the in-control chart performance based on the in-control ARL is greater than some specific desired value with a high probability (such as 0.9, 0.95). This may be more easily understood by the user.

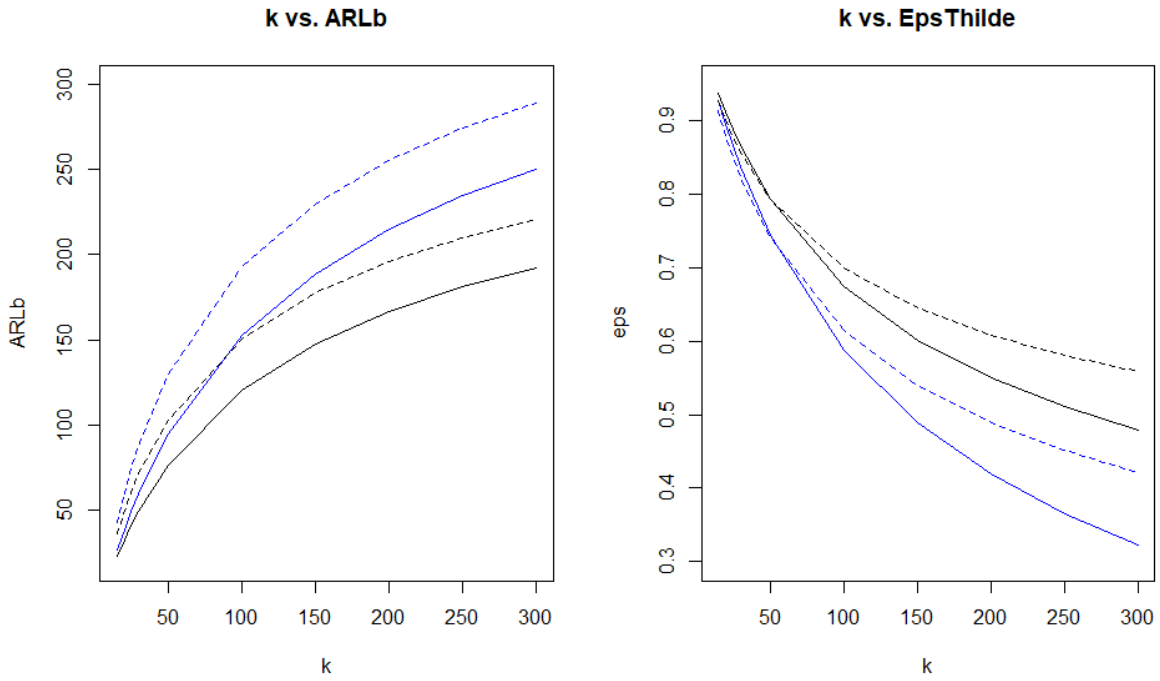

Figure 4: Lower prediction bound ARL_b and its corresponding tolerance factor $\tilde{\varepsilon}$ vs. sample size k for nominal $ARL_0 = 370$ $(c_{II}^* = 3, 3.0902)$, 500 $(c_{II}^* = 3, 3.0902)$, $p_0 = 0.05, 0.1,$ and various values of k. Black curves stand for $p_0 = 0.05$ and the blue curves stand for $p_0 = 0.1$. Solid curves stand for $c_{II}^* = 3$ and dashed curves stand for $c_{II}^* = 3.0902$.

For figure 4, we use the charting constant 3 and 3.0902, which yield an in-control average run length of 370 and 500, respectively, corresponding to the "3-sigma" rule in the known parameter case. This figure indicates that 1) The lower prediction bound ARL_b monotonically increases and the tolerance factor $\tilde{\varepsilon}$ monotonically decreases in k. 2) The higher charting constant leads to the lower prediction bound ARL_b

or the higher tolerance factor $\tilde{\varepsilon}$. 3) The higher nominal probability p_0 results in the higher lower prediction bound ARL_b or the lower tolerance factor $\tilde{\varepsilon}$. Also, even for our largest number of batches $k = 300$, the charting constants 3 and the nominal probability $p_0 = 0.1$ guarantee the lower prediction bound $ARL_b = 250.4748$ or the tolerance factor $\tilde{\varepsilon} = 0.3230$ which may be much lower than our expectation of nominal average run length ARL_0. That is, we may be not able to guarantee that the nominal average run length can be reached if we focus on the "3-sigma" rule expect for the extremely large number of batches which it may be not practical. Some researchers found out a new approach to give us a practical answer compared with the "3-sigma" rule. Before introducing the new approach, we will explain another evidence that many practitioners are interested in to solidify our motivation at this point.

2.3.2.1 Required Number of Phase I Batches

The number of Phase I batches k is an important factor, because in addition to the engineering reasons, k, along with the subgroup size n, it determines the Phase I sample size $m = kn$, that is the required amount of Phase I data. If data collection is expensive and/or time consuming, practitioners need to be very cautious about the Phase I sample size. Yet, sufficient data are needed to obtain reasonable chart performance. In the present setting, the Phase I subgroup size n is at least 1, so k is a lower bound to the Phase I sample size m. Similar to computing the lower prediction bound ARL_b, the number of Phase I batches k can be found from equation (19) when c_{II}^*, p_0 and ARL_b($\tilde{\varepsilon}$ and ARL_0) are given. In table 3 we show the required number of Phase I batches for different nominal significance levels, nominal average run lengths (ARL_0) and some tolerance factors ($\tilde{\varepsilon}$).

Table 3: Required number of Phase I batches k for the general model based on EPC, $\tilde{\varepsilon} = 0, 0.1, 0.2$ and a nominal $ARL_0 = 370(c_{II}^* = 3), 500(c_{II}^* = 3.0902)$with a nominal significance level $p_0 = 0.05, 0.1$, when the unbiased standard deviation estimator is used

p_0	ARL_0	$\tilde{\varepsilon} = 0$	$\tilde{\varepsilon} = 0.1$	$\tilde{\varepsilon} = 0.2$
0.05	370	1.1333E+08	11543	2591
	500	*	13190	2927
0.1	370	6.8833E+07	7053	1594
	500	*	8038	1795

* indicates that the solution is "infinite" as defined by R.

For example, consider $p_0 = 0.05$, $\tilde{\varepsilon} = 0.1$ and $ARL_0 = 370$ ($ARL_b = (1 - \tilde{\varepsilon})ARL_0 = 333$). If we apply the "3-sigma" rule, the Phase II charting constant c_{II}^* is equal to 3. Using equation (19), from table 3, k is found to be 11,543. If we keep the same setting but increase p_0 to 0.1, we find k decreases to 7,053. Then, if $\tilde{\varepsilon}$ is increased from 0.1 to 0.2, k is further decreased to 2,591. Also, for $p_0 = 0.05$ and $\tilde{\varepsilon} = 0.1$, if ARL_0 is increased from 370 to 500, which means that c_{II}^* is increased from 3 to 3.0902, k is increased to 13,190. In summary, the required number of Phase I batches decreases with increasing nominal p_0 and increasing tolerance factor $\tilde{\varepsilon}$. Our findings thus indicate that the required number of Phase I batches for the "3-sigma" rule are just too large and are not likely to be met in many practical applications. This raises the need for adjusting the control limits (charting constants) from a practical standpoint. In the next section find these adjusted charting constants for two well-known performance metrics.

2.3.2.2 Adjusted Charting Constants for a Guaranteed Conditional In-control Performance

As we have shown above, the traditional "3-sigma" rule should not be used under the EPC perspective unless one has a huge number of Phase I batches, numbering in hundreds of thousands. This is

most likely to be not the case in most practical situations. Instead, we propose finding the charting constant c_{II}^* under the EPC with a guaranteed lower bound of the in-control average run length. In table 4 we show the charting constants for $k \geq 25$. As we discussed, we have a serious singularity problem for smaller values of k and the problem is worse under the conditional perspective. Thus, for $k \leq 20$, the average run length is extremely large which is less useful in practice, so we do not show these charting constants hereafter.

Table 4: Two-sided Phase II Shewhart $\bar{X}$ charting constant c_{II}^*, based on the conditional perspective, for different Phase I batch numbers k, nominal $ARL_0 = 370$ and 500, the nominal probability $p_0 = 0.05$ and 0.1 and the tolerance factor $\tilde{\varepsilon} = 0$ and 0.1, respectively, when the unbiased standard deviation estimator is used, along with the mean and standard deviation of $CARL_0$, that is $E(CARL_0)$ and $SD(CARL_0)$, respectively.

			$\tilde{\varepsilon} = 0$			$\tilde{\varepsilon} = 0.1$		
ARL_0	p_0	k	c_{II}^*	$E(CARL_0)$	$SD(CARL_0)$	c_{II}^*	$E(CARL_0)$	$SD(CARL_0)$
370	0.05	25	3.9868	2.54E+06	9.77E+08	3.9442	1.63E+06	5.57E+08
		30	3.8707	2.42E+05	2.69E+07	3.8293	1.73E+05	1.67E+07
		50	3.6243	1.11E+04	6.67E+04	3.5854	9.03E+03	4.97E+04
		100	3.4098	2.29E+03	3.03E+03	3.3732	1.97E+03	2.51E+03
		150	3.3241	1.42E+03	1.22E+03	3.2884	1.24E+03	1.04E+03
		200	3.2755	1.11E+03	7.53E+02	3.2403	9.76E+02	6.44E+02
		250	3.2433	9.56E+02	5.47E+02	3.2085	8.42E+02	4.70E+02
		300	3.2201	8.60E+02	4.34E+02	3.1855	7.60E+02	3.74E+02
	0.1	25	3.7470	2.42E+05	4.50E+07	3.7069	1.69E+05	2.75E+07
		30	3.6617	4.84E+04	2.59E+06	3.6225	3.67E+04	1.70E+06
		50	3.4781	5.24E+03	2.28E+04	3.4408	4.37E+03	1.76E+04
		100	3.3160	1.56E+03	1.88E+03	3.2803	1.35E+03	1.58E+03
		150	3.2505	1.07E+03	8.71E+02	3.2156	9.39E+02	7.43E+02
		200	3.2133	8.83E+02	5.71E+02	3.1787	7.79E+02	4.91E+02
		250	3.1885	7.83E+02	4.32E+02	3.1542	6.93E+02	3.73E+02
		300	3.1706	7.21E+02	3.51E+02	3.1365	6.39E+02	3.04E+02
500	0.05	25	4.1065	9.46E+06	4.92E+09	4.0649	5.93E+06	2.79E+09
		30	3.9871	6.54E+05	1.06E+08	3.9467	4.59E+05	6.56E+07
		50	3.7335	2.00E+04	1.57E+05	3.6956	1.62E+04	1.16E+05
		100	3.5127	3.53E+03	5.19E+03	3.4770	3.03E+03	4.30E+03
		150	3.4244	2.10E+03	1.97E+03	3.3896	1.83E+03	1.66E+03
		200	3.3744	1.62E+03	1.18E+03	3.3400	1.42E+03	1.01E+03
		250	3.3412	1.37E+03	8.42E+02	3.3072	1.21E+03	7.24E+02
		300	3.3173	1.23E+03	6.60E+02	3.2835	1.08E+03	5.70E+02
	0.1	25	3.8595	6.96E+05	1.86E+08	3.8205	4.78E+05	1.13E+08
		30	3.7717	1.10E+05	8.70E+06	3.7335	8.21E+04	5.68E+06
		50	3.5829	8.91E+03	4.88E+04	3.5466	7.39E+03	3.73E+04
		100	3.4160	2.35E+03	3.13E+03	3.3813	2.03E+03	2.62E+03
		150	3.3486	1.56E+03	1.37E+03	3.3146	1.37E+03	1.17E+03
		200	3.3102	1.27E+03	8.79E+02	3.2766	1.12E+03	7.56E+02
		250	3.2847	1.11E+03	6.56E+02	3.2514	9.84E+02	5.67E+02
		300	3.2663	1.02E+03	5.29E+02	3.2331	9.02E+02	4.58E+02

For example, take $p_0 = 0.05, k = 25$ and $ARL_b = 370$ ($\tilde{\varepsilon} = 0$ and $ARL_0 = 370$). For this setting, $c_{II}^* = 3.9868$. However, for $p_0 = 0.01$ and the same values of the other parameters, $c_{II}^* = 3.7470$. If the nominal ARL_0 is increased to 500 we get $c_{II}^* = 4.1065$. If k is increased to 30, $c_{II}^* = 3.8707$. If $\tilde{\varepsilon}$ is increased to 0.1, $c_{II}^* = 3.8293$. Hence, the adjusted charting constant c_{II}^* decreases with increasing p_0, k, $\tilde{\varepsilon}$ and ARL_0. $E(CARL_0)$ and $SD(CARL_0)$ shows a similar trend. From table 4 two main points emerge. First, all the adjusted charting constants are different from 3 (or 3.0902) for a nominal $ARL_0 = 370$ (or 500). Second, even though the adjusted charting constants guarantee that the in-control average is at least equal to the lower bound with a high probability, the expectations of the in-control average run length values are so high, that it is almost impractical to be implemented, mainly due to the singularity problem that is seen in table 4. Thus, if we need to have a reasonable in-control average run length value under the EPC perspective, we need to either diminish the lower bound (or increase the tolerance factor) or the nominal significance level, or increase the number of Phase I batches if the number is not costly. In other words, we cannot guarantee that the in-control average run-length performance is satisfactory with a desired lower bound and a high probability for number of Phase I batches $k \leq 20$. Hence, the adjusted control charts are not recommended in practice for small batch numbers such as $k \leq 20$. For moderately large batch numbers such as $25 < k \leq 50$, we may need to lower the nominal significance level or decrease the desired lower bound for practical implementation.

OOC performance for the charts with the *EPC* adjusted constants are of interest. This is discussed below. The significance levels are chosen by following the traditional strategy of hypothesis testing but they can be any values between 0 and 1.

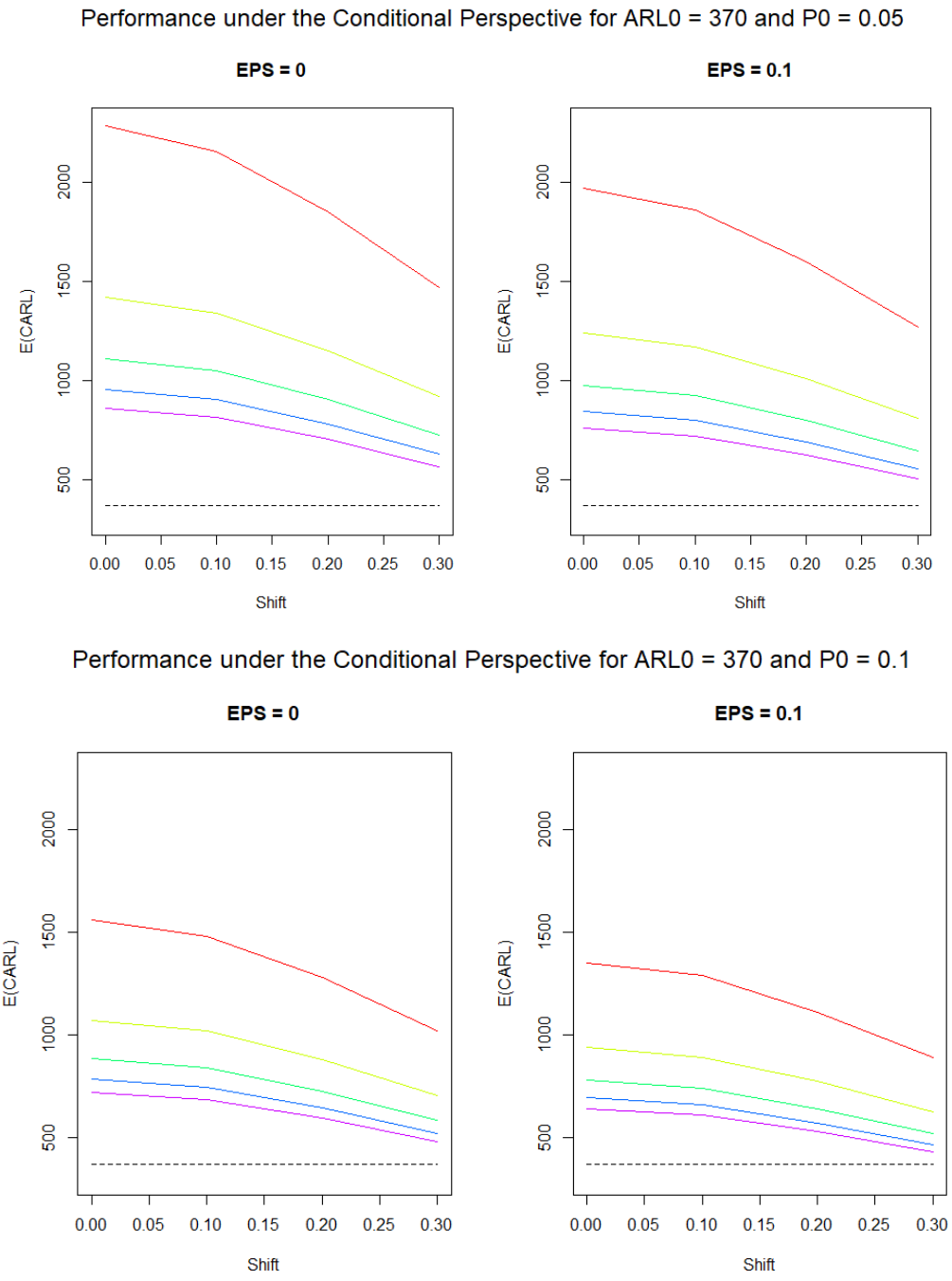

Figure 5: Out-of-control average run lengths of two-sided Phase II Shewhart $\bar{X}$ control charts based on the conditional perspective, for Phase I batch numbers $100 \leq k \leq 300$ and different parameters, where the red solid curves, the yellow solid curves, the green solid curves, the blue solid curves, the purple solid curves are for $k = 100,150,200,300$, respectively. Notice that the black dashed lines are the reference lines for different corresponding nominal average run length.

From figure 5, 1) it is clear that as expected, the OOC average run length decreases with increasing shifts. 2) The signals will be detected quicker with the higher number of Phase I batches 3) Also, we can see an interesting phenomenon here. The slopes will fall more gentle if users have the greater number of Phase I batches. In other words, the margin of detection of signals is lower, which contradicts what was

observed under the unconditional perspective. From the view of the efficiency of using k, the method under the conditional perspective is more suitable for the relative small number of Phase I batches, compared with the unconditional perspective.

We have shown two ways to finding the Phase II charting constants based on the unconditional and the conditional perspective based on the EPC, respectively. In the next section, we apply these charting constants in an illustrative example using some simulated data.

2.3.3 Phase II Example

As mentioned in the introduction, in this paper we find the "correct" charting constants and the "correct" process of building the Phase I and II $\bar{X}$ control charts for the model in equation (2). In Phase II, we use the parameter estimates obtained from our Phase I reference data: $\bar{\bar{X}} = 245.1$, $\hat{\sigma}_0 = 2.0544$ for $k = 30$. Note that the estimator $\hat{\sigma}_0$, is adjusted to be unbiased by using the relevant constant $c_4 = 0.9914$. The Phase II charting statistics (the batch means) of the next 20 batches are simulated in two stages. In the first stage, 10 observations (means of Phase II batches 1 through 10) are simulated from a normal distribution with a mean 245.1 and a standard deviation 2.0544. In the second stage, 10 more observations (means of Phase II batches 11 through 20) are simulated from a normal distribution with a mean 247.55 (with a 1% increase from 245.1) and the same standard deviation as in stage one. The batch means are shown in table 5.

Table 5: Simulated batch means for the Phase II example

Batch	*Batch Means*	*Batch*	*Batch Means*	*Batch*	*Batch Means*	*Batch*	*Batch Means*
1	246.303	6	241.365	11	247.312	16	249.229
2	246.558	7	246.395	12	251.285	17	245.73
3	244.875	8	244.533	13	248.312	18	246.87
4	244.168	9	244.516	14	248.62	19	249.853
5	246.345	10	243.211	15	246.009	20	248.165

The control charts are shown in figure 6. Because we have two types of charting constants (unconditional and the conditional) that may be used in Phase II, two pairs of control charts (limits) are shown. First, under the unconditional perspective, for a nominal $ARL_0 = 370$, the charting constant $c_{II}^* = 2.7776$ is found from table 2, which leads to the lower and upper control limits, 239.3938 and 250.8062, respectively. Second, given under the EPC perspective, the control limits are obtained as 237.1490 and 253.0510 for $ARL_0 = 370, p = 0.05, \tilde{\varepsilon} = 0$ and the charting constant $c_{II}^* = 3.8703$ from table 4. Also, shown are the reference control limits, 38.9369 and 251.2631, calculated with the charting constant 3, under the "3-sigma" rule.

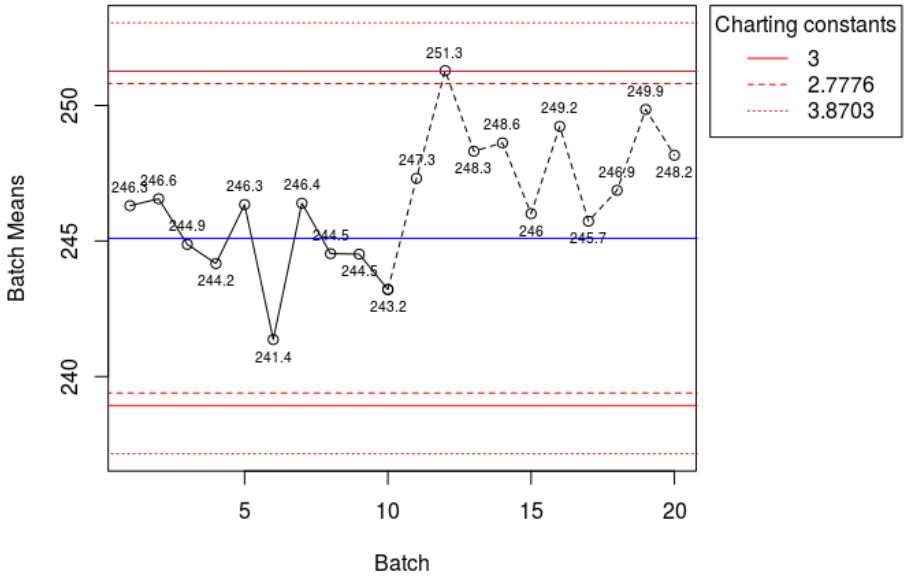

Figure 6: Phase II Shewhart $\bar{X}$ charts for simulated data

In figure 6, the blue solid line is the central line. The solid red lines are the control limits for the charting constant, 3. The dashed lines are the control limits for the charting constant, 2.7776. The dotted lines are the control limits for the charting constant, 3.8703 from table 4. The solid curve with dots (from batch 1 to batch 10) shows simulated batch means with no shift. The dashed curve with dots (from batch 11 to batch 20) show simulated batch means with 1% shift in the mean. Besides, the range of the control limits under the unconditional perspective is slightly smaller than the one of the traditional control limits which is much smaller than the one under the *EPC*. Also, we can observe that there is a signal under the unconditional perspective and the "3-sigma" rule at batch 12, but no signal occurs at the same place under the *EPC*. First of all, we discuss this event under the unconditional perspective and the "3-sigma" rule. Because the batch mean at batch 12 is much closer to the upper limit under the "3-sigma" rule compared to the one under the unconditional perspective, which the distance is 0.0214 compared to 0.4783, we consider that this event may have higher probability as a false alarm under the "3-sigma" rule. Second, we discuss this event under the unconditional perspective and the *EPC*. In the long run, the process is guaranteed to have ARL_0, 370, via the limits under the unconditional perspective, and the ARL_0 is much greater than 370 if the limits are based on *EPC*. In the short run, the limits based on the *EPC* will have fewer signals which may lead to fewer false alarm. Because we have no clue about how long is long enough for the runs and, in practice, whether we have enough budget to support this long run, if users need to keep fewer false alarms, the limits under *EPC* may be a better option.

3. Summary and Conclusions

In this paper we consider control charts for II monitoring the mean in both Phase I and Phase II for a linear model with two variance components. We obtain the adjusted control limits correcting for the effects of parameter estimation, in each phase, and using the proper performance metric, as recommended in the recent literature. Control charts in Phase I are to initialize and establish a process control. To achieve this purpose, charts are determined for a given false alarm probability, the charting statistics and its central tendency and dispersion as well as appropriate charting constants. Traditionally, the Shewhart charts

monitor the process "subgroup by subgroup" with a chunk of data, and the parameters come from this idea as well. In this case, the univariate normal distribution as a framework participates in the estimation and the judgement. In section 2.1, we monitor the process all at once, and extend the traditional Shewhart chart to our exact Shewhart chart under the multivariate normal distribution with a model that the "between-batch" variance is involved, comparing with the traditional ignorance of this variance. Many researchers upheld the "3-sigma" rule or other similar rule as the decision of the charting constants. However, according to our careful study, this rule may be not suitable in our case anymore. The charting constants shall be as table 1, whose parts are far from 3.

In Phase II, we monitor the process with the control limits developed with the estimated parameters in Phase I. In this Phase, the limitation of the traditional "3-sigma" rule, although popular, is shown here. In tables 2 and 4, our results are close to the traditional ones, such as 3 for $ARL_0 = 370$, only for very large number of Phase I batches under the unconditional perspective or the EPC. However, this may not be very practical in many applications and exact charting constants are needed. Under the unconditional perspective, all charting constants are less than 3 for $ARL_0 = 370$ in our study and increase as the number of Phase I batches increase, so if users keep using the traditional 3-sigma rule, this may cause the inaccuracy of their control charts with the relatively small number of Phase I batches in the long run. For example, for $k = 20$ and $ARL_0 = 370$, the adjusted Phase II charting constant equals to 2.6665, from table 2, which is less than 3. That means users will have lower FAR's if they chart their data with 3. That is to say that it may be not as "powerful" as we expected. We have fewer signals that may lead to a late decision in practice. This waste may be costly for many fields. In the short run, EPC shows us that the traditional method with the relatively small number of Phase I batches may cause more false alarms than what is expected. Take the number of Phase I batches $k = 20$, the nominal $ARL_0 = 370$, the nominal $p_0 = 0.1$ and the tolerance factor $\tilde{\varepsilon} = 0.1$. From table 4, the Phase II charting constant equals to 3.8296, which is much greater than 3. So the users will have more signals if they insist using 3 which will increase the false alarm rate. Also, we obtain several interesting constants: (1) the lower bound ARL_b. ARL_b guarantees that the Phase II chart runs at the appropriate performance level in the production process with a high probability. (2) The number of Phase I batches k. k is potentially related with the cost of examination in Phase I. (3) the adjusted charting constants c_{II}^* with affordable the number of Phase I batches k via EPC. c_{II}^* is the major output of our study in this study which shows evidences and doubts the "3-sigma" rule.

In conclusion, we show that monitoring the average output of a process described by a multiple components of variation model, with the Shewhart $\bar{X}$ chart, follows a structure similar to that in the case of the basic Shewhart model in Phase I and Phase II. In Phase I, the traditional "3-sigma" charts are not very useful and new charting constants are needed. In Phase II, the "3-sigma" charts can work if one has a very large Phase I dataset with a large number of batches, which may be available in some applications. But it is a completely different story if the number of Phase I batches available for prospective process monitoring is moderate to small. Here the correct charting constants must be calculated and used as shown here.

References

Albers, W., Kallenberg, W. C., & Nurdiati, S. (2005). Exceedance probabilities for parametric control charts. *Statistics, 39*(5), 429-443.

Blazek, L. W., & Coleman, D. E. (1995). [Shewhart-Type Charts in Nonstandard Situations]: Discussion. *Technometrics, 37*(1), 24-26.

Capizzi, G., & Masarotto, G. (2013). Phase I distribution-free analysis of univariate data. *Journal of Quality Technology, 45*(3), 273-284.

Chakraborti, S. (2000). Run length, average run length and false alarm rate of Shewhart X-bar chart: exact derivations by conditioning. *Communications in Statistics-Simulation and Computation, 29*(1), 61-81.

Chakraborti, S. (2006). Parameter estimation and design considerations in prospective applications of the X chart. *Journal of Applied Statistics, 33*(4), 439-459.

Chakraborti, S., Eryilmaz, S., & Human, S. W. (2009). A phase II nonparametric control chart based on precedence statistics with runs-type signaling rules. *Computational Statistics & Data Analysis, 53*(4), 1054-1065.

Champ, C. W., & Chou, S.-P. (2003). Comparison of standard and individual limits phase I Shewhart X, R, and S charts. *Quality and Reliability Engineering International, 19*(2), 161-170.

Champ, C. W., & Jones, L. (2004). Designing Phase I—X Charts with Small Sample Sizes. *Quality and Reliability Engineering International, 20*(5), 497-510.

Epprecht, E. K., Loureiro, L. D., & Chakraborti, S. (2015). Effect of the amount of phase I data on the phase II performance of S 2 and S control charts. *Journal of Quality Technology, 47*(2), 139-155.

Gandy, A., & Kvaloy, J. T. (2013). Guaranteed conditional performance of control charts via bootstrap methods. *Scandinavian Journal of Statistics, 40*(4), 647-668.

Goedhart, R., Schoonhoven, M., & Does, R. J. (2017). Guaranteed in-control performance for the Shewhart X and X control charts. *Journal of Quality Technology, 49*(2), 155-171.

Hahn, G. J., & Doganaksoy, N. (1995). [Shewhart-Type Charts in Nonstandard Situations]: Discussion. *Technometrics, 37*(1), 29-31.

Ittzes, A. (2001). Statistical process control with several variance components in the dairy industry. *Food Control, 12*(2), 119-125.

Johnson, N. L., & Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions.* New York: John Wiley & Sons. Inc.

Jones-Farmer, L., Jordan, V., & Champ, C. W. (2009). Distribution-Free Phase I Control Charts for Subgroup Location. *Journal of Quality Technology, 41*(3), 304-316.

Kim, K.-S., & Yum, B.-J. (1999). Control Charts for Random and Fixed Components of Variation in the Case of Fixed Wafer Locations and Measurement Positions. *IEEE Transactions on Semiconductor Manufacturing, 12*(2), 214-228.

Mahmoud, M. A., Henderson, G., Epprecht, E. K., & Woodall, W. H. (2010). Estimating the standard deviation in quality-control applications. *Journal of Quality Technology, 42*(4), 348-357.

Montgomery, D. C. (2009). *Introduction to statistical quality control.* New York: Wiley & Sons.

Palm, A. C., & Robert, D. L. (1995). [Shewhart-Type Charts in Nonstandard Situations]: Discussion. *Technometrics, 37*(1), 26-29.

Park, C. (1998). Design of and ewma charts in a variance components model. *Communications in Statistics-Theory and Methods, 27*(3), 659-672.

Park, S. (2004). Statistical Design of Experiments and Analysis on Gate Poly-Silicon Critical Dimension. *IEEE Transactions on Semiconductor Manufacturing, 17*(3), 362-374.

Quesenberry, C. P. (1993). The Effect of Sample Size on Estimated Limits for $\bar{X}$ and X Control Charts. *Journal of Quality Technology, 25*(4), 237-247.

Roes, K. C., & Does, R. J. (1995). Shewhart-Type Charts in Nonstandard Situations. *Technometrics, 37*(1), 15-24.

Sullivan, J. H., Woodall, W. H., & Gardner, M. M. (1995). [Shewhart-Type Charts in Nonstandard Situations]: Discussion. *Technometrics, 37*(1), 31-35.

Woodall, W. H., & Thomas, E. V. (1995). Statistical process control with several components of common cause variability. *IIE transactions, 27*(6), 757-764.

Yao, Y., Hilton, C. W., & Chakraborti, S. (2017). Designing Phase I Shewhart X-bar charts: Extended tables and software. *Quality and Reliability Engineering International, 33*(8), 2667-2672.

Yashchin, E. (1994). Monitoring Variance Components. *Technometrics, 36*(4), 379-393.