

# Monitoring the Slopes of Linear Profiles

Junjia Zhu<sup>1</sup>,

Dennis K. J. Lin<sup>2</sup>

<sup>1</sup>Penn State Harrisburg,  
Middletown, Pennsylvania

<sup>2</sup>The Pennsylvania State  
University, University Park,  
Pennsylvania and Renmin  
University, Beijing, China

---

**ABSTRACT** Motivated by a vertical density profile problem, in this article we focus on monitoring the slopes of linear profiles. A Shewhart-type control chart for monitoring slopes of linear profiles is proposed. Both Phase I and Phase II applications are discussed. The performance of the proposed chart in Phase I applications is demonstrated using both real-life data in an illustrative example and simulated data in a probability of signal study. For Phase II applications, it is shown that the average run length (ARL) of the proposed control chart depends only on the shifts of slopes, whereas the ARL of the multivariate  $T^2$  chart depends on both the shifts of slopes and the correlation between the estimated slope and the intercept. When such a correlation is low, the proposed control chart has a better ARL performance than the  $T^2$  chart.

**KEYWORDS** profile monitoring, simple linear regression, statistical process control

---

## INTRODUCTION

In statistical process control (SPC) applications, the quality of a process or a product can often be characterized by a relationship (or a profile) between a response variable and one or more explanatory variables. For example, Kang and Albin (2000) presented two practical examples where it was necessary to use a profile to describe process or product's quality. Monitoring profile data is inevitable when the variation of the response variable  $Y$  cannot be adequately explained by the values of  $Y$  themselves.

Consider the vertical density profile (VDP) data set reported in Walker and Wright (2002; available online at <http://bus.utk.edu/stat/walker/VDP/Allstack.TXT>). For the VDP data, the density of the wood board ( $Y$ ) is measured by a profilometer that uses a laser device to take measurements at a series of fixed depths ( $x$ ) across the thickness of the board. For the inspection of the vertical density close to the wood board's surface, only the density close to the top is relevant (i.e., with the depth close to 0 in.). Truncating the VDP data by taking  $x \leq 0.02$  yields 24 linear profiles with each profile having  $n = 11$  pairs of  $x$  and  $Y$ 's. When a simple linear regression model is used, the intercept (interpreted as the value of  $Y$  when  $x$  is 0) of the truncated VDP data represents the wood board's surface density, and the slope (interpreted as the increase of  $Y$  when  $x$  increases by one unit) represents the speed of density increase when depth increases, which is an important characteristic to measure the quality of the wood board.

Address correspondence to Junjia Zhu, Department of Computer Science and Mathematical Sciences, W-256E Olmsted Building, Penn State Harrisburg, 777 W. Harrisburg Pike, Middletown, PA 17057. E-mail: [junjia@psu.edu](mailto:junjia@psu.edu)

The main interest here is thus to monitor the rate of density change with respect to depth (the slope), with various levels of wood board's surface density (the intercept). The truncated VDP profiles and their fitted linear equations are displayed in Figure 1. Note that using some existing control charts designed to monitor both the intercept and slope for this truncated VDP data set may be misleading, because monitoring the shift of slopes is the major concern.

The profile monitoring process using control charts, like many SPC applications, can be studied via two phases. In Phase I, the goal is to evaluate the stability of the process, find and remove any outliers with assignable causes, and estimate the in-control values of the process parameters. In Phase II, the goal is to monitor on-line data to quickly detect any shift in the process parameter from the baseline estimated in Phase I. The key in Phase I is to assess the probability of deciding that a process is unstable. In Phase II, the ARL (average run length) is often used as a criterion to compare competing methods.

Kang and Albin (2000) proposed a multivariate  $T^2$  control chart method that can be used to monitor both the intercept and slope for the linear profiles in Phase II. Kim et al. (2003) established a Phase II

exponentially weighted moving average (EWMA) control chart method for monitoring linear profile data where the  $x$ -values are centered and show that their method provides better ARL performance for detecting shifts in the regression model parameters than Kang and Albin's (2000) method. For Phase I analysis, Kang and Albin (2000) suggested that their Phase II methods be used in Phase I with estimates substituted for the values of unknown parameters. Kim et al. (2003) suggested replacing their phase II EWMA with Shewhart charts. Mahmoud and Woodall (2004) used a global  $F$  statistic based on an indicator variable technique to compare  $k$  regression lines.

This article proposes a control chart for efficiently monitoring slopes of linear profiles. The remainder of this article is organized as follows. In the next section, the statistical methodologies used in Phase II application are discussed, and a Shewhart-type control chart to monitor slopes of linear profiles is proposed. The in-control ARL and out-of-control ARL are then evaluated and compared in the following section. Next, the Phase I application of control chart is developed. The application of the proposed control chart in Phase I is illustrated by a real-life data set and simulated data sets. Finally, a summary and suggestions for future work are given in the last section.

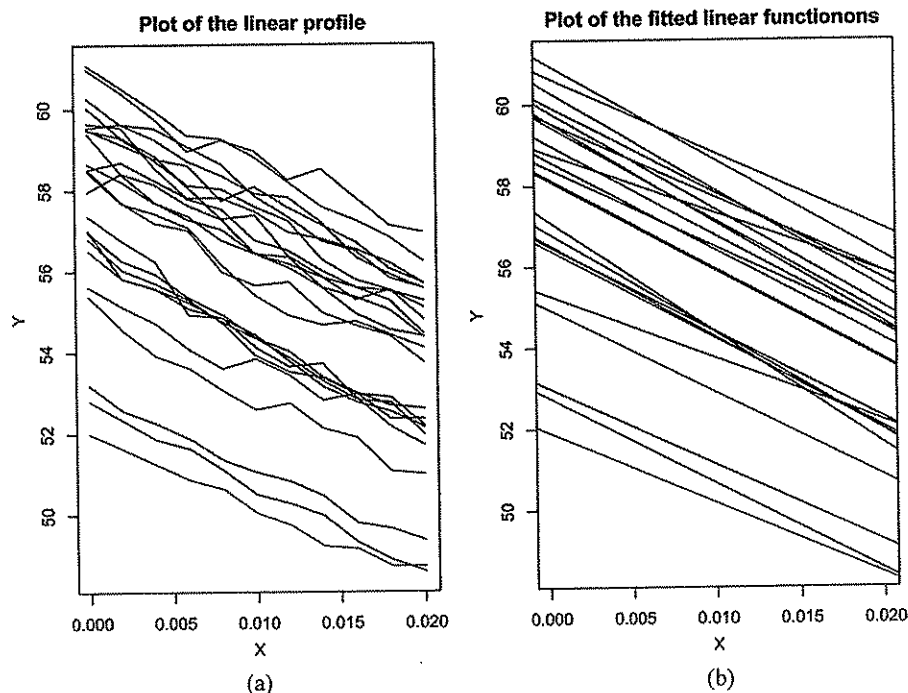


FIGURE 1 Plot of the truncated and fitted linear VDP data.

## PROBLEM FORMULATION VIA STATISTICAL METHODOLOGY

Suppose that for the  $j$ th profile, one has the observations  $(x_i, y_{ij})$ ,  $i = 1, 2, \dots, n$ . For the simplicity of presentation, we assume that the values of explanatory variable  $x$  are the same for all profiles. The underlying model for in-control profile is

$$Y_{ij} = A_0 + A_1 x_i + \varepsilon_{ij}, \quad i = 1, 2, \dots, n,$$

where  $\varepsilon_{ij}$  is distributed as independent normal with mean zero and variance  $\sigma^2$ . In Phase I the three parameters  $A_0$ ,  $A_1$ , and  $\sigma^2$  are to be estimated using a historical in-control profile data set, whereas in Phase II analysis, the parameters are known and on-line profiles are monitored. The quality engineer can usually obtain the in-control parameter value from two sources: (1) from the designer of the system (i.e., the "optimal" parameter value) or (2) from the parameter value estimated after Phase I study. The main interest is on monitoring the mean shift in  $A_1$ .

The ordinary least-squares (OLS) point estimate for  $A_1$  from profile  $j$  is  $a_{1j} = \frac{S_{xy(j)}}{S_{xx}}$ , where

$$S_{xy(j)} = \sum_{i=1}^n (y_{ij} - \bar{y}_j)(x_i - \bar{x}),$$

and

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2.$$

Furthermore, the sample statistic  $a_{1j}$  is a normally distributed random variable with mean  $A_1$  and variance  $\sigma_1^2 = \frac{\sigma^2}{S_{xx}}$ . Thus, the lower control limit (LCL) and upper control limit (UCL), based on the standard confidence interval concept, can be shown to be:

$$LCL = A_1 - z_{\alpha/2} * \sqrt{\frac{\sigma^2}{S_{xx}}}$$

and

$$UCL = A_1 + z_{\alpha/2} * \sqrt{\frac{\sigma^2}{S_{xx}}},$$

where  $z_{\alpha/2}$  is the  $100(1 - \alpha/2)$ th percentile of the standard normal distribution, and  $\alpha$  is a factor adjusting the in-control ARL.

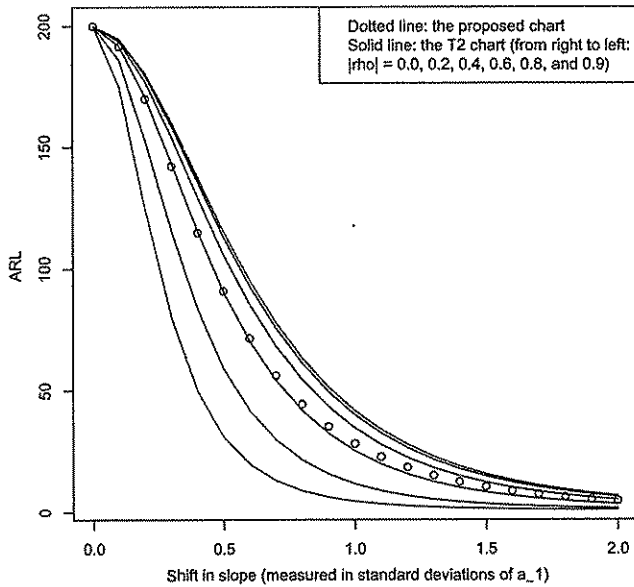
The estimate of the slope  $A_1$  remains the same when  $x$  or  $Y$  (or both) is centered; i.e., from  $x$  or  $Y$  subtract its corresponding mean  $\bar{x}$  or  $\bar{Y}$ . Thus, when monitoring the slope is the main concern, it is appropriate to center both  $x$  and  $Y$ . Because the fitted line of centered profile data will pass through the origin  $(0, 0)$ , the fitted line can be plotted in the same figure with the two lines of the lower and upper bounds for the slopes. The system engineer will be able to visually check whether any slope is out of control.

## COMPARISON ON AVERAGE RUN LENGTH PATTERNS

The run length of a control chart is defined as the count of the number of runs until the first out-of-control signal. For the Shewhart-type control charts, the run length is typically assumed as a geometric random variable. The parameter of the geometric distribution  $p$  is the probability of giving an out-of-control signal. The theoretical ARL of a control chart is the inverse of the statistical power of the control chart, when the system is not in control (i.e., there is a mean shift for at least one parameter) and equals  $1/\alpha$ , where  $\alpha$  is the system's Type I error rate when the system is in control.

ARL in Phase II analysis is a major criterion for the effectiveness of the control chart in detecting out-of-control signal when the process is not in control. In this section, we compare ARLs of the proposed control chart and the multivariate  $T^2$  control chart proposed by Kang and Albin (2000).

It can be shown that (see the Appendix A for detailed derivation) when the process only has shifts of slopes by an amount of  $\beta \frac{\sigma}{\sqrt{S_{xx}}}$ :  $(A_1 \rightarrow A_1 + \beta \frac{\sigma}{\sqrt{S_{xx}}})$ , the ARL of the proposed chart is  $1/p_{new}$ , where  $p_{new} = 1 - \Phi(z_{\alpha/2} - \beta) + \Phi(-z_{\alpha/2} - \beta)$  and  $\Phi$  is the cumulative distribution function (CDF) of the standard normal distribution; the ARL of a multivariate  $T^2$  chart is  $1/p_{T^2}$ , where  $p_{T^2} = 1 - F_{\chi^2_2(\lambda)}(\chi^2_{2,\alpha})$  with the noncentrality parameter  $\lambda = \frac{\beta^2}{1 - \rho_{a_0, a_1}^2}$ , and  $\rho_{a_0, a_1}$  is the correlation coefficient between the estimates of the intercept and the slope. In short, the ARL performance of the proposed chart depends only on the actual shift in the slope  $\beta$ , whereas the ARL performance for the multivariate  $T^2$  chart depends on the noncentrality parameter  $\lambda$ , which is a function of both  $\beta$  and the correlation  $\rho_{a_0, a_1}$ .



**FIGURE 2** ARL comparison between the proposed chart and the multivariate  $T^2$  chart.

Plots of the shifts in slope, measured in standard deviations of  $a_1 \left( \frac{\sigma}{\sqrt{S_{xx}}} \right)$ , against the corresponding ARLs for the proposed chart and multivariate  $T^2$  chart with various absolute values of  $\rho_{a_0, a_1}$  are shown in Figure 2. Figure 2 indicates that the proposed chart has better ARL performance than  $T^2$  chart when  $|\rho_{a_0, a_1}| < 0.6$ .

## PROPOSED PROCEDURES FOR PHASE I CONTROL CHART

In Phase I, a historical profile data set is analyzed. Three main objectives in Phase I analysis are to (1) evaluate the stability of the process, (2) detect and remove any outliers with assignable causes, and (3) estimate the in-control values of the process parameters after removing profile samples associated with any assignable causes.

Given a historical profile data set with  $k$  profiles, Kang and Albin (2000) estimated the parameters  $A_0$ ,  $A_1$ , and  $\sigma^2$  as

$$\hat{A}_0 = \frac{\sum_{j=1}^k a_{0j}}{k}, \quad \hat{A}_1 = \frac{\sum_{j=1}^k a_{1j}}{k},$$

where  $a_{0j} = \bar{Y}_j - a_{1j}\bar{x}$  is the OLS point estimate for  $A_0$  for profile  $j$ , and

$$\hat{\sigma}^2 = MSE = \frac{\sum_{j=1}^k MSE_j}{k},$$

where

$$MSE_j = \frac{\sum_{i=1}^n (y_{ij} - \hat{y}_{ij})^2}{n-2} = \frac{\sum_{i=1}^n (y_{ij} - a_{0j} - a_{1j}x_i)^2}{n-2}.$$

All three parameter estimators ( $\hat{A}_0$ ,  $\hat{A}_1$ , and  $\hat{\sigma}^2$ ) are unbiased estimators (see, for example, Kutner et al. 2005).

The Phase I Shewhart-type control chart limits for slope were developed by Mahmoud and Woodall (2004). It can be shown that the statistic  $\frac{a_{1j} - \hat{A}_1}{\hat{\sigma} \sqrt{\frac{k-1}{kS_{xx}}}}$  follows a  $t$  distribution with  $k(n-2)$  degrees of freedom.

Our proposed control chart in detecting and removing outliers and estimating the in-control values of the slopes of linear profiles can be performed for the following steps:

**Step 1:** Check assumptions:

- For each profile, make a scatterplot of  $Y$  against  $x$  to check linearity assumption visually. If replicate  $Y$  values are observed for the same level of  $x$ , then one may also perform a lack-of-fit test for linearity.
- For each profile, center both  $x$  and  $Y$ , build a simple linear regression model, and check the normality assumption (by a normal probability plot, for example) and independence assumption (by an autocorrelation plot, for example) of the residuals.

The assumption checking is to ensure the appropriateness of the proposed method. When any of these assumptions are violated, there are various remedies, such as removing the profile from the data set, or transformation (Kutner et al. 2005). This will not be further discussed here. We denote the number of remaining profiles after step 1 as  $k$ .

**Step 2:** Estimate the values of parameters: with

$$\hat{A}_1 = \frac{\sum_{j=1}^k a_{1j}}{k} \text{ and } \hat{\sigma}^2 = \frac{\sum_{j=1}^k MSE_j}{k}.$$

**Step 3:** Build the control chart for monitoring the slopes, where the centerline equals  $\hat{A}_1$ , the lower control limit (LCL) equals  $\hat{A}_1 - t_{k(n-2), \alpha/2} * \hat{\sigma} \sqrt{\frac{k-1}{kS_{xx}}}$ , and the upper control limit (UCL) equals  $\hat{A}_1 + t_{k(n-2), \alpha/2} * \hat{\sigma} \sqrt{\frac{k-1}{kS_{xx}}}$ . The  $t_{k(n-2), \alpha/2}$  is the  $100(1-\alpha/2)$ th percentile of the student's  $t$  distribution with  $k(n-2)$  degrees of freedom, and  $\alpha$  is a pre-defined constant used to control the in-control false-alarm rate.

**Step 4:** Any profile whose estimated slope falls outside the control limits is regarded as an outlier. If no outlier is detected, then one concludes that the process is stable and moves to the next step. If at least one outlier is detected, then one removes the profile with the largest deviance of slope from the centerline; i.e., remove the profile  $j$  with maximum  $|a_{1j} - \hat{A}_1|$  value. Repeat step 2 and step 3 on the remaining profile data set until no outliers are detected.

**Step 5:** Use the parameter estimates in step 2 through step 4 for Phase II applications.

### AN ILLUSTRATIVE EXAMPLE

Returning to the truncated VDP data set as plotted in Figure 1, where Figure 1(a) is the plot of the truncated profile, and Figure 1(b) is the plot of the fitted linear function for each profile, it can be seen that the truncated profiles are rather linear and the profiles are somewhat parallel with each other. Apparently, the variation of intercepts among profiles is much larger than the variation of slopes. Recall that the intercept of a truncated VDP profile is the wood board's surface density and the slope is the rate of change of the density ( $y$ ) with respect to the depth ( $x$ ). Because the main interest is to monitor the slope instead of the intercept, a large variation of the intercepts and a relatively small variation of the slopes (i.e., the parallel nature of the plotted profiles) gives the first impression that most of the profiles are in control. Thus, using the proposed control chart is more

appropriate than using some of the existing Phase I methods that are designed to monitor both intercept and slope where a large variation of the intercepts will generate a high rate of out-of-control signals.

In fact, a Phase I analysis on the 24 truncated VDP profiles was performed using the existing Phase I methods, including the multivariate  $T^2$  chart by Kang and Albin (2000), the three individual Shewhart-type control charts by Kim et al. (2003), and the global F-test by Mahmoud and Woodall (2004). It turns out that none of these three methods can handle this case appropriately. The multivariate  $T^2$  chart and the three Shewhart charts simply give out-of-control signals to every profile in the data set, and the global F-test only gives an out-of-control signal for the whole profile data set (and for almost every subset of these linear profiles) without indicating which profile is out of control. All three methods give signals that are not very informative for Phase I applications.

We next apply the proposed control chart to the truncated VDP data as follows.

**Step 1:** Figure 1 is used to check the linearity assumption visually. In addition, for each profile, a normal probability plot and an autocorrelation plot are generated (not shown here) for checking the normality assumption and the independence assumption of the residuals. No violation of these assumptions is detected.

**Steps 2 and 3:** Because we are monitoring the slopes only, a plot of fitted lines based on the

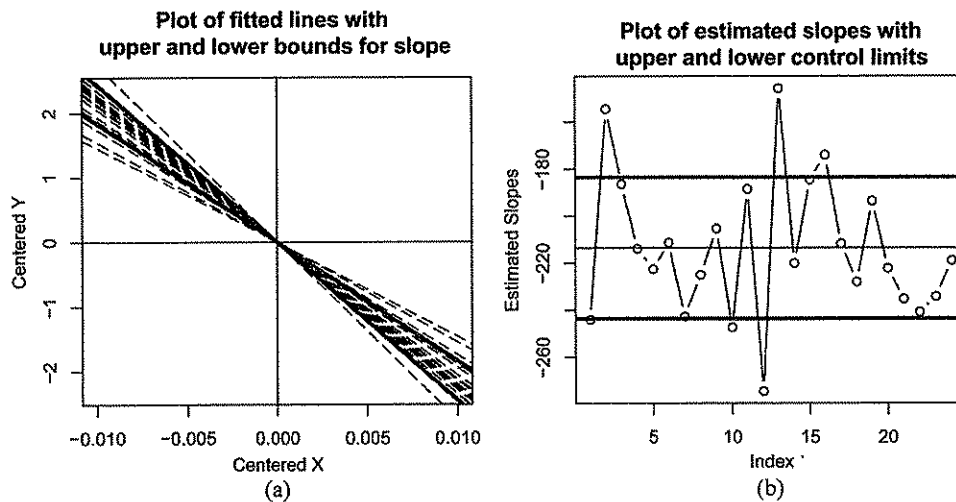


FIGURE 3 Control chart for slopes: based on the truncated VDP data.

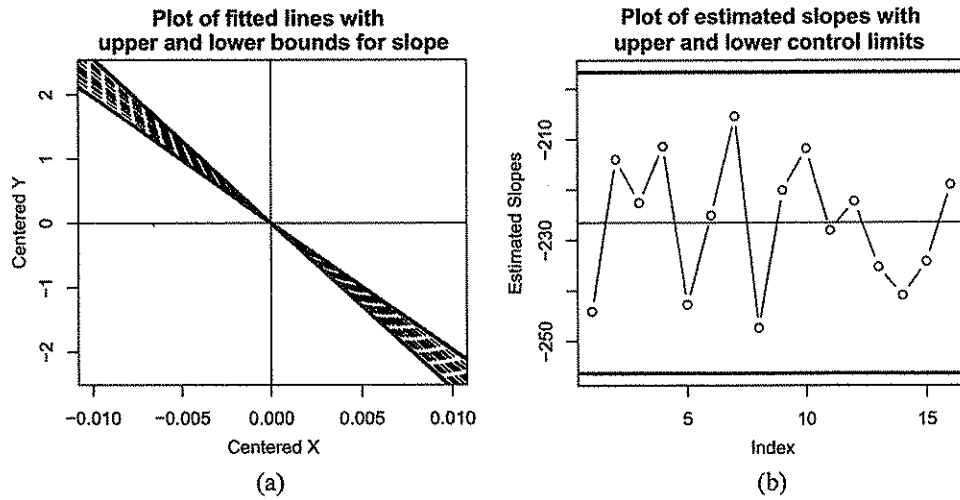


FIGURE 4 Phase I application of the proposed control chart for slopes: with outliers removed.

centered profiles with the upper and lower limits of the slope will be helpful in detecting outliers. This is shown in Figure 3(a). Phase I method is used to estimate the parameters ( $A_1$  and  $\sigma^2$ ). Based on the estimated parameters, the upper and lower control limits for the slopes are calculated and the proposed control chart for monitoring the slopes is built. This is shown in Figure 3(b), using  $\alpha = 0.005$  level.

**Step 4:** Figure 3(b) shows that 6 out of the 24 profiles are marked as out of control in terms of their slopes. Within the 6 out-of-control profiles, profile 13 is the one with largest deviance from the centerline, so it is removed from the data set. We use the remaining 23 linear profiles to estimate the parameters and rebuild the control chart. Steps 2 and 3 are thus repeated. After the eighth iteration, the procedure ends successfully with no outlier detected. There are 16 profiles left in the remaining data set. The sequence of outlier profile removal is (13, 2, 12, 16, 15, 3, 11, and 19).

**Step 5:** The plot of fitted lines based on the centered profiles with the upper and lower limits of the slope is shown in Figure 4(a). The final control chart is shown in Figure 4(b). We thus conclude that the process is in control and use the parameter estimates  $\hat{A}_1 = -226.52$  and  $\hat{\sigma}^2 = 0.050$  for the Phase II analysis on monitoring the on-line profiles.

Note: it is worth mentioning that we have also experimented with a different strategy in step 4 by

removing all outliers at one time. The resulting estimates are  $\hat{A}_1 = -221.76$  and  $\hat{\sigma}^2 = 0.049$ .

## A STUDY ON PROBABILITY OF SIGNAL

Probability of signal (POS) is defined as the probability that the control chart detects at least one out-of-control signal when being applied to a historical data set. The POS study is used to check the statistical power (or false-alarm error rate) of the control chart under various simulated out-of-control or in-control scenarios. In Phase I applications, the performance of control charts can be evaluated by POS studies.

As suggested by one referee, we compared the performance of the proposed control chart with the multivariate  $T^2$  chart proposed by Kang and Albin (2000) for monitoring the shift in slopes for linear profiles in Phase I in terms of the overall probability of a signal. The details of the POS studies are given in Appendix B. It was found that:

- when the system only has mean shifts of the slopes, the probability of out-of-control signal for the proposed chart will depend only on the actual shift in slope, whereas for the multivariate  $T^2$  chart, this will depend on both the shift in slope and the  $x$  values used in Phase I. Therefore, POS comparison between the proposed chart and the  $T^2$  chart is not possible, unless further assumption on the  $x$  values is imposed.

- when the system has a shift in the intercept only (which is regarded as in-control scenario in our research), the probability of an out-of-control signal (false alarm) using a multivariate  $T^2$  chart is much higher than that for the proposed chart.
- using  $\hat{\sigma}^2 = \text{median}(MSE_1, MSE_2, \dots, MSE_b)$  in step 2 of the proposed chart will yield an underestimated value of the variance and will increase the false-alarm rate and thus is not recommended (see Appendix B for a detailed discussion).

## CONCLUSION AND DISCUSSION

For profile data in general, each complete measurement can be described by a relationship between the response variable  $Y$  and one or more explanatory variables  $x_1, x_2, \dots, x_p$ . In many practical situations, the system engineer's main interest is focused on monitoring the *rate of change* for  $Y$  with respect to  $x$  (i.e., steepness or the slope of the linear relationship between  $Y$  and  $x$ ). The profiles are regarded as in control as long as they are parallel to each other (if all profiles are from a historical data set) or are parallel to a baseline profile (if all profiles are generated by on-line production). In addition, some processes generate profiles that have very large variation in intercepts but relatively small variation in slopes. The use of classical profile charts that are designed to monitor both intercept and slope can be misleading. Thus, developing a method for monitoring slopes only is necessary. In this article we study the simplest but most fundamental case: only one explanatory variable  $x$  exists, the relationship between  $Y$  and  $x$  is linear, and one is only interested in detecting the potential shift in the slopes. Both Phase I and Phase II applications are considered. Extension to a general linear regression model is rather straightforward.

Compared with a multivariate  $T^2$  control chart, the proposed control chart can significantly reduce the false-alarm rate when the system only has shifts of intercepts, for both Phase I and Phase II applications. The proposed chart is especially suitable in Phase I applications when the system in nature generates linear profiles with large variation in intercepts but relatively small variation in slopes. For such cases, the multivariate  $T^2$  chart has much higher probability in generating out-of-control signals.

In Phase II applications, given  $n, S_{xx}$ , and fixed  $\alpha$  level, the ARL performance of the proposed chart depends only on the actual shift in the slope,  $\beta$ . The comparison of ARL performances between the proposed control chart and the multivariate  $T^2$  chart will also depend on the correlation between the estimates of the intercept and the slope  $\rho_{a_0, a_1}$ . The proposed chart has better ARL performance than the  $T^2$  chart when  $|\rho_{a_0, a_1}| < 0.6$ .

Monitoring the linear profiles requires several assumptions, which includes the normality assumption of the error term ( $\varepsilon_{ij}$ ), linearity assumption between the true relationship between  $x$  and  $Y$ , the independence assumption among observations with each profile, and the independence assumption among the profiles. Violation of one or more of these assumptions may lead to an increase in the false-alarm rate. Future research dealing with the violation of assumptions may involve using non-linear regression, non parametric regression, or functional data analysis methods to build control charts. Moreover, because control chart for slopes is not very sensitive in detecting *small* shifts of slopes, future research may involve building EWMA or cumulative sum (CUSUM) control charts for monitoring slopes.

## ABOUT THE AUTHORS

Junjia Zhu is an assistant professor in the Department of Computer Science and Mathematical Sciences at the Penn State Harrisburg, Middletown. He received a Ph.D. in Statistics from the Pennsylvania State University, University Park. His research interests include statistical process control, linear models, categorical data analysis, and response surface methodology. He holds a B.S. degree in Industrial Foreign Trade from Hunan University (China) and an M.S. degree in Statistics from the University of Idaho.

Dennis K.J. Lin is a University Distinguished Professor of Supply Chain and Statistics at the Pennsylvania State University, University Park. He is an elected Fellow of the American Society for Quality, an elected Fellow of the American Statistical Association, and an elected member of the International Statistical Institute. He has published nearly 150 papers in the areas of design of experiments, response surface methodology, reliability, control charts, data mining, and statistical inference.

## REFERENCES

- Kang, L., Albin, S. L. (2000). On-line monitoring when the process yields a linear profile. *Journal of Quality Technology*, 32:418–426.
- Kim, K., Mahmoud, M. A., Woodall, W. H. (2003). On the monitoring of linear profiles. *Journal of Quality Technology*, 35:317–328.
- Kutner, M. H., Nachtsheim, C. J., Neter, J., Li, W. (2005). *Applied Linear Statistical Models*, 5th ed. New York: McGraw-Hill Irwin.
- Mahmoud, M. A., Woodall, W. H. (2004). Phase I Analysis of Linear Profiles with Calibration Applications. *Technometrics*, 46:377–391.
- Montgomery, D. C. (2005). *Introduction to Statistical Quality Control*, 5th ed. New York: John Wiley & Sons.
- Montgomery, D. C., Peck, E. A., Vining, G. G. (2001). *Introduction to Linear Regression Analysis*, 3rd ed. , New York: John Wiley & Sons.
- Murphy, B. J. (1987). Selecting out of control variables with the  $T^2$  multivariate quality control procedure. *The Statistician*, 36(5):571–581.
- Walker, E., Wright, S. P. (2002). Comparing curves using additive models. *Journal of Quality Technology*, 34:118–129.
- Woodall, W. H., Spitzner, D. J., Montgomery, D. C., Gupta, S. (2004). Using control charts to monitor process and product profiles. *Journal of Quality Technology*, 36:309–320.

## APPENDIX A: ARL DERIVATION WHEN THE PROCESS ONLY HAS SHIFTS OF SLOPES

### The Proposed Control Chart

Under the null hypothesis  $H_0$ : the process is in control [ $E(a_1) = A_1$ ], we have  $a_1 \sim N(A_1, \sigma^2 S_{xx}^{-1})$ . Under the alternative hypothesis  $H_A$ : there is a shift in slope  $\Delta A_1 = \beta \frac{\sigma}{\sqrt{S_{xx}}}$ , and assuming there is no change of  $\sigma^2$ , the distribution for the parameter estimate of slope  $a_1^*$  is  $a_1^* \sim N\left(A_1 + \beta \frac{\sigma}{\sqrt{S_{xx}}}, \sigma^2 S_{xx}^{-1}\right)$ .

Thus, we can calculate the statistical power for the proposed control chart for monitoring the slopes as:

$$\begin{aligned} p_{new} &= P(a_1^* > UCL \text{ or } a_1^* < LCL) \\ &= P\left(a_1^* > A_1 + z_{\alpha/2} \sqrt{\frac{\sigma^2}{S_{xx}}}\right) \\ &\quad + P\left(a_1^* < A_1 - z_{\alpha/2} \sqrt{\frac{\sigma^2}{S_{xx}}}\right) \\ &= P(Z > z_{\alpha/2} - \beta) + P(Z < -z_{\alpha/2} - \beta) \\ &= 1 - \Phi(z_{\alpha/2} - \beta) + \Phi(-z_{\alpha/2} - \beta) \end{aligned}$$

where  $Z$  is the standard normal random variable and  $\Phi$  is the cumulative distribution function (CDF) of the standard normal random variable. The ARL of the proposed chart for slope is  $1/p_{new}$ .

## The Multivariate $T^2$ Control Chart

Kang and Albin (2000) proposed a multivariate  $T^2$  control chart for Phase II applications in monitoring linear profiles. The test statistic is  $T^2 = (Z - \mu)' \Sigma^{-1} (Z - \mu)$ , where  $Z = (a_0, a_1)'$ ,  $\mu = (A_0, A_1)'$ , and  $\Sigma = \begin{bmatrix} \sigma_0^2 & \sigma_{01}^2 \\ \sigma_{01}^2 & \sigma_1^2 \end{bmatrix}$ . Note that for Phase II applications,

the proposed control chart tests for a change in slope away from an in-control value ( $H_0: E[a_1] = A_1$ ) whereas the  $T^2$  chart tests for a mean change in the joint distribution of both intercept ( $A_0$ ) and slope ( $A_1$ ) ( $H_0: E[Z] = \mu$ ), where  $Z$  and  $\mu$  are defined above.

Under the null hypothesis  $H_0$ : the process is in control [ $E(a_1) = A_1$ ], we have  $T^2 \sim \chi_2^2$ . Thus, for a given Type I error rate  $\alpha$ , the upper control limit is  $UCL = \chi_{2,\alpha}^2$ . Under the alternative hypothesis  $H_A$ : there is a shift in slope  $\Delta A_1 = \beta \frac{\sigma}{\sqrt{S_{xx}}}$ , and assuming there is no change of  $\sigma^2$ , the  $Z^* = (a_0, a_1)'$  has a true mean  $\mu_* = \left(A_0, A_1 + \beta \frac{\sigma}{\sqrt{S_{xx}}}\right)'$ ; by Murphy (1987) we can show that the test statistic

$$T^{*2} = (Z^* - \mu)' \Sigma^{-1} (Z^* - \mu) \sim \chi_2^2(\lambda),$$

where  $\chi_2^2(\lambda)$  is a non central chi-square distribution with 2 degrees of freedom and the non centrality parameter  $\lambda = (\mu^* - \mu)' \Sigma^{-1} (\mu^* - \mu)$ .

In our case, we have  $\mu^* - \mu = \left(0, \beta \frac{\sigma}{\sqrt{S_{xx}}}\right)'$ . We can calculate the statistical power for the multivariate  $T^2$  control chart for monitoring the slopes as:

$$\begin{aligned} P_{T^2} &= P(T^{*2} > UCL) \\ &= P(T^{*2} > \chi_{2,\alpha}^2) \\ &= 1 - F_{\chi_2^2(\lambda)}(\chi_{2,\alpha}^2) \end{aligned}$$

where  $F_{\chi_2^2(\lambda)}$  is the cumulative distribution function (CDF) of a non central chi-square distribution with 2 degrees of freedom and non centrality parameter  $\lambda$ .

It can be shown that the non centrality parameter  $\lambda$  is equal to:

$$\begin{aligned} \lambda &= \left(0, \beta \frac{\sigma}{\sqrt{S_{xx}}}\right) \Sigma^{-1} \left(0, \beta \frac{\sigma}{\sqrt{S_{xx}}}\right)' \\ &= \left(0, \beta \frac{\sigma}{\sqrt{S_{xx}}}\right) \frac{1}{\sigma^2} \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \left(0, \beta \frac{\sigma}{\sqrt{S_{xx}}}\right)' \\ &= \beta^2 \frac{\sum x_i^2}{S_{xx}} = \frac{\beta^2}{1 - \rho_{a_0, a_1}^2}, \end{aligned}$$



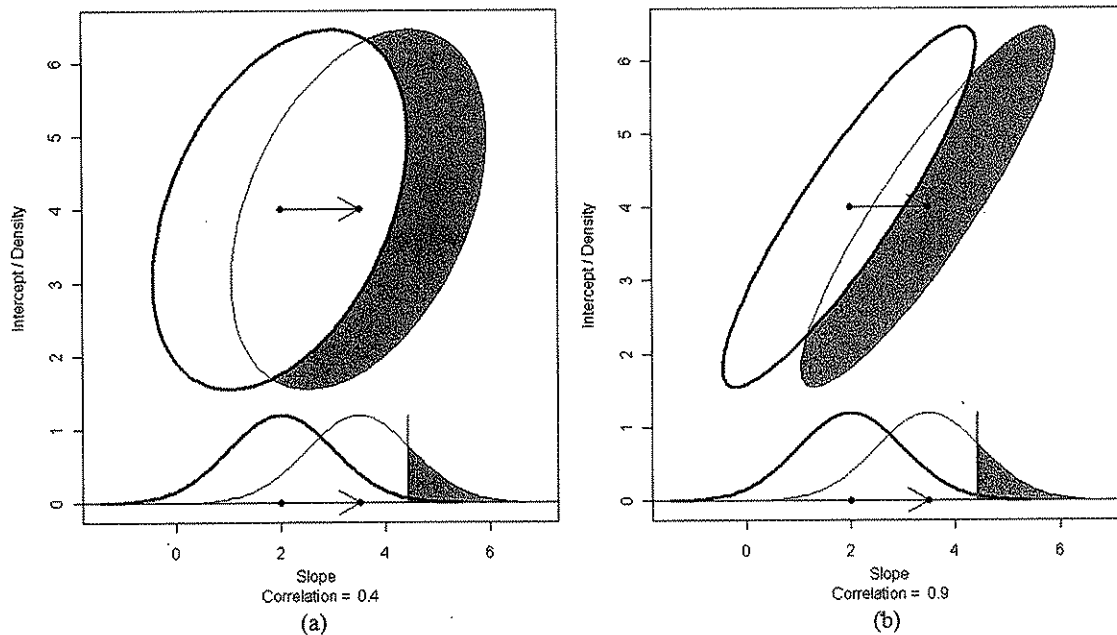
where  $\rho_{a_0, a_1}^2 = \frac{n\bar{x}^2}{\sum x_i^2} = \frac{n\bar{x}^2}{n\bar{x}^2 + S_{xx}}$  is the square of correlation coefficient between the estimates of intercept and the slope. The ARL of a multivariate  $T^2$  chart for slope is  $1/p_{T^2}$ .

Note that although the absolute value of  $\rho_{a_0, a_1}$  is affected by the explanatory variable  $x$  in general, it is actually affected by three aspects of  $x$ : the sample size  $n$ , the dispersion (variance) of the explanatory variable  $S_{xx}$ , and the location of the explanatory variable  $\bar{x}$ . For the ARL comparison between the proposed chart and  $T^2$  chart (i.e., Figure 2), we make the shift in slopes in the unit of standard deviations of the estimated slopes ( $A_1 \rightarrow A_1 + \beta \frac{\sigma}{\sqrt{S_{xx}}}$ ) and then compare the ARL performance of these two charts based on different values of  $\rho_{a_0, a_1}$ . In other words, we fix  $S_{xx}$  and then adjust  $n$  and  $\bar{x}$  to obtain different  $\rho_{a_0, a_1}$ . In actual practice, to study the performance of a  $T^2$  chart under various  $\rho_{a_0, a_1}$  values, it is probably easier to fix  $n$  and  $S_{xx}$  and then adjust  $\bar{x}$  to vary  $\rho_{a_0, a_1}$ . For example, in the POS study, we considered two data sets: seq(min=2, max=8, length=10) and seq(min=-3, max=3, length=10) as explained in Appendix B. The  $n$ 's and  $S_{xx}$ 's are the same for these two data sets, whereas  $\bar{x}$ 's are different.

## A Graphical Comparison Between the Proposed Chart and the $T^2$ Chart

Figure A1 is an illustration to demonstrate the ARL comparison between the proposed Shewhart-type control chart and the  $T^2$  chart. In Figure A1, two different values of  $\rho_{a_0, a_1}$  are used, 0.4 and 0.9, which represent low and high correlation between the estimated intercept and slope, correspondingly. The  $x$  axis is the value of the estimated slope and the  $y$  axis is the values for estimated intercept. It is assumed that there is no shift in intercept and the mean value of the slope shifts from 2 to 4.

In Figure A1 there are two ellipses; the ellipse on the left represents the 95% confidence region of the multivariate  $T^2$  chart under  $H_0$  (no shift in slopes), and the ellipse on the right represents the 95% confidence region of the  $T^2$  chart under  $H_A$ . The shaded area shows the reject region under each correlation. And the proportion of the shaded area to the whole 95% confidence region (the thin-line ellipse) of the  $T^2$  chart under  $H_A$  is the probability of detection (which equals to  $\frac{1}{ARL}$ ) when the process really has a shift in slope. It can be seen that the probability of detection under higher correlation is greater than that with lower correlation. This is consistent with our numerical results.



**FIGURE A1** Plot of the rejection region for the  $T^2$  chart under different correlations and the proposed control chart, when the process only has shift in slopes.

For the proposed Shewhart-type control chart, the distributions of the estimated slope for the profiles under  $H_0$  and  $H_A$  are plotted. In Figure A1, the density curve on the left is the null distribution, and the density curve on the right is the alternative distribution. The vertical line on the right tail of the null distribution is the critical value, and the shaded area is the reject region. It can be seen that the proposed control has the same reject region under different correlations. Again, this is consistent with our numerical findings.

## APPENDIX B: DETAILED STUDY IN PROBABILITY OF SIGNAL

We compared the performance of the proposed control chart with the multivariate  $T^2$  chart proposed by Kang and Albin (2000) for monitoring the shift in slopes for linear profiles in Phase I in terms of the overall probability of a signal. The underlying in-control profile is modeled as  $Y_{ij} = 3 + 2x + \epsilon_{ij}$ ,  $i = 1, 2, \dots, n$  and  $j = 1, 2, \dots, k$ , where  $n$  is the number of observations per profile and  $k$  is the number of profiles in the historical data set. In our study  $n = 10$  and  $k = 20$ . We set  $A_0 = 3$  and  $A_1 = 2$  here to be consistent with Kang and Albin (2000). The fixed  $x$  values were firstly set as  $\text{seq}(\text{min} = 2, 2, \text{max} = 8, \text{length} = 10)$  (i.e.,  $x_{\min} = 2, x_{\max} = 8$ , and 8 other evenly distributed grid points in between) and later changed to  $\text{seq}(\text{min} = -3, \text{max} = 3, \text{length} = 10)$ . Two different sets of  $x$  are used because the  $x$  values will affect the correlation between the estimated intercept and slope ( $\rho_{a_0, a_1}$ ) and thus will affect the POS for multivariate  $T^2$  chart.

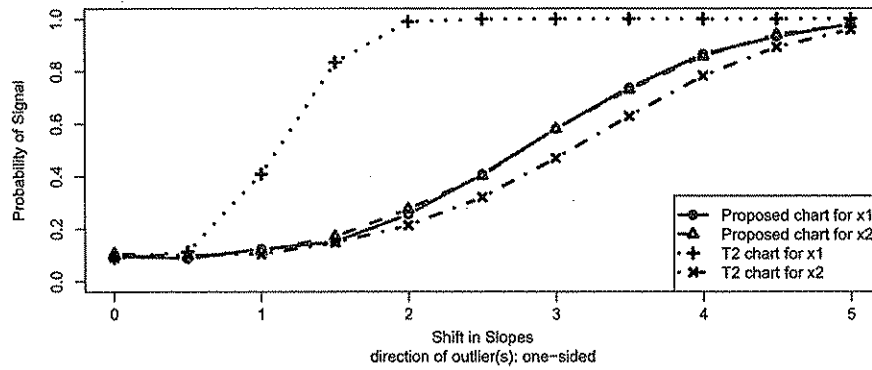
The pre defined  $\alpha$  rates for the proposed control chart and the multivariate  $T^2$  chart are both set equal to 0.005. Because we have total  $k = 20$  profiles in the historical data set, the overall false-alarm rate  $\alpha'$  for both charts can be calculated as  $1 - (1 - \alpha)^k \approx 0.1$ .

For  $k = 20$ , we considered fixed shifts of slopes in  $m$  out of the  $k$  individual profiles. We used  $m = 1, 2$ , and 5. The locations of the shifts are set as random, although the locations will not affect the performance of these two methods because for both methods each profile is considered as an independent entry. Shifts in the slopes are measured in units of  $\frac{\sigma}{\sqrt{S_{xx}}}$ . In addition, for  $m > 1$  we distinguish the direction of slope shift as "one-sided" and

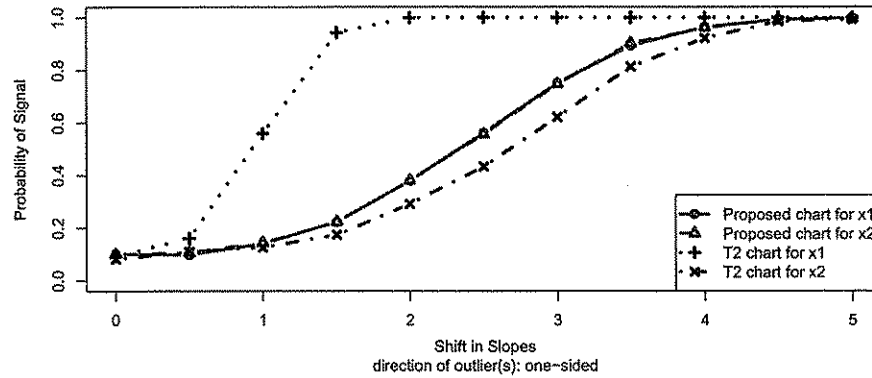
"both-sided," where one-sided shift means that all the slope shifts will follow the same direction, either all greater than the target value or all smaller than it, whereas both-sided shift means that each slope shift will be randomly assigned a direction, either greater than or smaller than the target value. Results show there is not much difference between these two directions in terms of the POS performance comparison between the two charts. Thus only one-sided shift results will be shown in this article.

One referee suggested replacing the mean estimates of the variance  $\hat{\sigma}^2$  in step 2 of the proposed chart by the median; i.e.,  $\hat{\sigma}^2 = \text{median}(MSE_1, MSE_2, \dots, MSE_k)$ . This approach was thoroughly investigated in our POS study. Our results show that when we use median estimates of the variance parameter in step 2, the POS of the proposed method is always greater than the nominal level; i.e., the Type I error rate is greater than  $\alpha$ . This will lead to smaller Type II error rates, and thus greater statistical power, in the out-of-control scenario. This is mainly because the sampling distribution of  $MSE_j$ s is right skewed, which implies that the median ( $MSE_1, MSE_2, \dots, MSE_k$ ) will be smaller than the mean  $\frac{\sum_{j=1}^k MSE_j}{k}$ . Using a median estimate will thus underestimate the variance and make the in-control range narrower. Consequently, the control chart is more likely to have an out-of-control signal in both in-control and out-of-control scenarios. Thus, it is recommended to use the mean, rather than the median estimate for the variance.

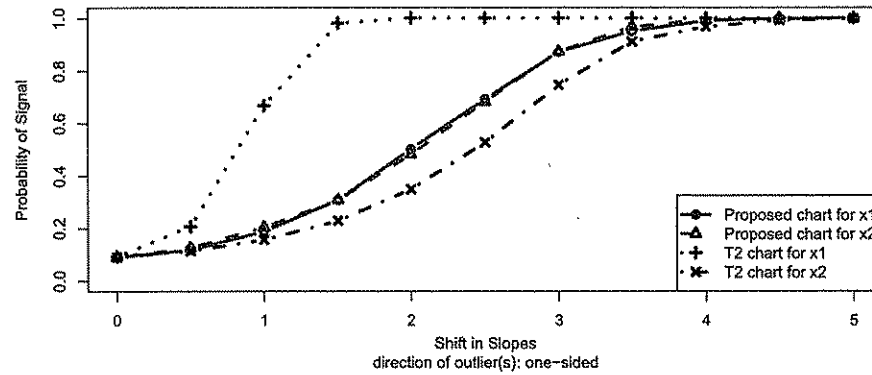
Figure B1 shows the simulated overall probability of generating at least one out-of-control signal for shifts in the slope from  $A_1$  to  $A_1 + \beta \frac{\sigma}{\sqrt{S_{xx}}}$ . The linear profiles are based on two sets of  $x$ s:  $x_1 = \text{seq}(\text{min} = 2, \text{max} = 8, \text{length} = 10)$  and  $x_2 = \text{seq}(\text{min} = -3, \text{max} = 3, \text{length} = 10)$ . The estimates of probabilities were obtained from 2000 simulations for  $\beta = 0, 0.5, 1, \dots, 5$ . The performance of the proposed chart is almost identical in both settings, whereas the performance of multivariate  $T^2$  chart varies. With the same in-control false-alarm rate, under  $x_1$  the multivariate  $T^2$  chart has higher probability in giving out-of-control signals than the proposed chart for real out-of-control cases ( $\beta > 0$ ) in all three scenarios ( $m = 1, 2$ , and 5). However, under  $x_2$  the proposed chart has better performance than the multivariate  $T^2$  chart.



(a) Number of Outliers  $m = 1$  Out of 20

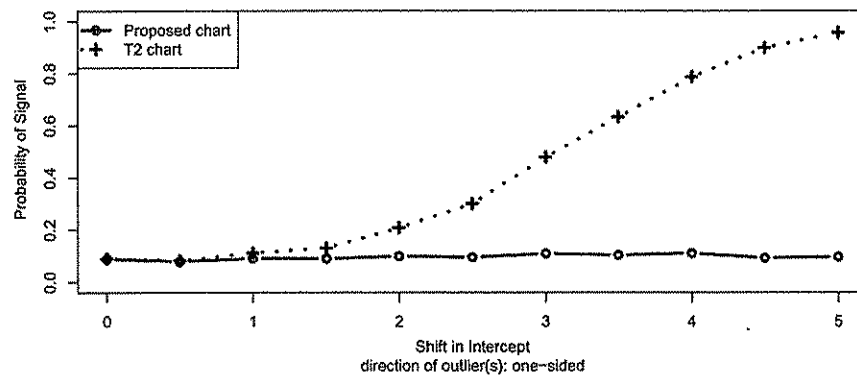


(b) Number of Outliers  $m = 2$  Out of 20



(c) Number of Outliers  $m = 5$  Out of 20

**FIGURE B1** Probability of out-of-control signal under slope shifts from  $A_1$  to  $A_1 + \beta \frac{\sigma}{\sqrt{S_{xx}}}$ , with  $x_1 = \text{seq}(\text{min}=2, \text{max}=8, \text{length}=10)$  and  $x_2 = \text{seq}(\text{min}=-3, \text{max}=3, \text{length}=10)$ .



**FIGURE B2** False-alarm rate under intercept shifts from  $A_0$  to  $A_0 + \lambda \frac{\sigma}{\sqrt{n}}$ , with  $x = \text{seq}(\text{min}=2, \text{max}=8, \text{length}=10)$ .

The POS of  $T^2$  charts depends on both the latitude of the shift and the correlation between the estimated intercept and slope  $\rho_{a_0, a_1}$ . In general, when the absolute value of  $\rho_{a_0, a_1}$  gets smaller, the multivariate  $T^2$  chart will have a smaller probability of generating out-of-control signals. The value of  $\rho_{a_0, a_1}$  is a function of  $\bar{x}$ ,  $S_{xx}$ , and  $n$ . Under  $x_1$  the  $\rho_{a_0, a_1}$  is 0.872, whereas under  $x_2$  the  $\rho_{a_0, a_1}$  is zero. This is why the POS performance of the  $T^2$  chart under  $x_1$  is much better than that under  $x_2$ . The relationship between the  $\rho_{a_0, a_1}$  and the POS for multivariate  $T^2$  chart in Phase I applications is similar to that in Phase II applications, which was discussed in Appendix A in detail.

Finally, the scenario that the system only has shifts in intercepts is tested. Shift in intercept is

measured in units of  $\frac{\sigma}{\sqrt{n}}$ ; i.e., shift from  $A_0$  to  $A_0 + \lambda \frac{\sigma}{\sqrt{n}}$ . Figure B2 shows the false-alarm rate for the proposed chart and the  $T^2$  chart: clearly the  $T^2$  chart has a much higher false-alarm rate than the proposed chart, especially when the shift in intercept gets large. This may be obvious because the proposed chart is specifically designed for detecting slope, whereas a  $T^2$  chart is for both intercept and slope. Thus, if shifts in intercept are not of interest, the use of a  $T^2$  chart may be misleading.

The probability of signal study is based on a Monte Carlo simulation program written in statistical software R. The computer code is available upon request.