Design Variable-Sampling Control Charts Using Covariate Information

Kai Yang and Peihua Qiu
Department of Biostatistics, University of Florida
Gainesville, FL 32610

Abstract

Statistical process control (SPC) charts are widely used in the manufacturing industry for monitoring the performance of sequential production processes over time. A common practice in using a control chart is to first collect samples and take measurements of certain quality variables from them at equally-spaced sampling times, and then make decisions about the process status by the chart based on the observed data. In some applications, however, the quality variables are associated with certain covariates, and it should improve the performance of an SPC chart if the covariate information can be used properly. Intuitively, if the covariate information indicates that the process under monitoring is likely to have a distributional shift soon based on the established relationship between the quality variables and the covariates, then it should benefit the process monitoring by collecting the next process observation sooner than usual. Motivated by this idea, we propose a general framework to design a variable-sampling control chart by using covariate information. Our proposed chart is self-starting and can well accommodate stationary short-range serial data correlation. It should be the first variable-sampling control chart in the literature that the sampling intervals are determined by the covariate information. Numerical studies show that the proposed method performs well in different cases considered.

Key Words: Auxiliary variables; Data correlation; Kernel estimation; Regression modelling; Sampling intervals; Self-starting; Statistical process control; Variable-sampling.

1 Introduction

For quality control and management, statistical process control (SPC) charts have been widely used in different applications for monitoring some quality variables of a sequential process (cf., Hawkins and Olwell 1998, Montgomery 2012; Qiu 2014). Conventional control charts can be roughly classified into the following four categories: Shewhart charts (Shewhart 1931), cumulative sum (CUSUM) charts (Page 1954), exponentially weighted moving average (EWMA) charts (Robert 1959), and change-point detection (CPD) charts (Hawkins et al. 2003). These conventional control charts are usually based on the assumptions that process observations at different time points are independent and identically distributed with a common parametric distribution (e.g., normal, Poisson) when the process is in-control (IC). Also, the observation times are usually equally spaced. In practice, some or all of these model assumptions could be violated, making them unreliable to use in such cases. To address this issue, some newer control charts have been developed for monitoring serially correlated data (e.g., Apley and Tsung 2002; Lee and Apley 2011; Qiu et al. 2020), dynamic processes (e.g., Qiu and Xiang 2014 and 2015) or processes with nonparametric distributions (e.g., Capizzi and Masarotto 2013; Chen et al. 2016; Qiu and Hawkins 2001). The control charts discussed above monitor a process based solely on the observed data of the quality variables. In many applications, however, the quality variables are associated with some covariates. For instance, product quality of a production line can be affected by workers' skills, equipment quality, and materials; crime rate in a city often depends on the economic and educational levels of the city population. For such applications, it should benefit process monitoring by using covariate information properly, which is the focus of this paper.

In the literature, there have been some existing control charts to use covariate information, which include the risk-adjusted control charts (e.g., Paynabar et al. 2012; Steiner and Jones 2010), the cause-selecting control charts (e.g., Asadzadeh et al. 2008), and the adaptive control chart (Yang and Qiu 2020). The risk-adjusted control charts try to adjust the impact of some important risk factors when monitoring a process, while the cause-selecting control charts try to distinguish shifts in different stages of a multi-stage production process. These charts take into account the covariate effect, but are sensitive to shifts in the

covariates no matter whether such shifts can actually cause shifts in the quality variables under monitoring. To address this issue, the adaptive control chart in Yang and Qiu (2020) uses the covariate information solely in the weighting parameter of an EWMA chart for monitoring the quality variables, so that the resulting EWMA chart is sensitive to shifts in the quality variables only while accommodating the covariate information. All these existing control charts, however, are designed for cases with equally spaced observation times. Intuitively, if the covariate information suggests that the quality variables under monitoring are likely to have a distributional shift soon based on the established relationship between the quality variables and the covariates, then it should benefit the process monitoring by collecting the next process observation sooner than usual. Based on this intuition, we develop a variable-sampling control chart in this paper, in which the variable sampling intervals are determined by the observed data of the covariates. As far as we know, this will be the first variable-sampling control chart whose sampling intervals are chosen using the covariate information.

In the SPC literature, there has been extensive discussion on process monitoring with variable sample sizes and/or variable sampling intervals (e.g., Arnold and Reynolds 2001; Prabhu et al. 1994; Reynolds et al. 1990; Runger and Pignatiello 1991; Zou et al. 2008). By a variable-sampling interval control chart, the next sampling interval would be chosen shorter if there is some indication of a shift in the process based on the observed data by the current time point, and longer otherwise. In that way, the resulting chart can detect a future shift faster than a conventional control chart. To keep the sampling scheme simple, traditionally the sampling interval is allowed to take two levels only. In the dynamic sampling scheme suggested by Li and Qiu (2014), the sampling interval is assumed to be a continuous function of the p-value of the related charting statistic, and it has been shown that such a dynamic sampling scheme would generally be more effective than the conventional 2-interval sampling scheme. It should be pointed out that the existing control charts with variable sampling intervals are usually designed for cases with independent process observations. As a comparison, our proposed new method allows process observations to be serially correlated.

The rest of the paper is organized as follows. Section 2 provides a detailed description

of the proposed method. Its numerical performance is investigated in Section 3 with some simulation studies. A demonstration of the new method in a real-data setup is provided in Section 4. Some remarks conclude the paper in Section 5. Some technical details and some extra numerical results are given in a supplementary file.

2 Proposed Method

Let Y be a quality variable of a process under monitoring and $\boldsymbol{X} = (X_1, \ldots, X_p)^T$ be a covariate vector of p dimensions. The observation of $(\boldsymbol{X}, Y)$ at time t is denoted as $(\boldsymbol{X}(t), Y(t))$. When the process is IC, $Y(t)$ and $\boldsymbol{X}(t)$ are assumed to follow the model

$$Y(t) = f\left(\boldsymbol{X}(t)\right) + \varepsilon(t), \text{ for } t \geq 0, \tag{1}$$

where $f(\cdot)$ is an unknown regression function and $\varepsilon(t)$ is a zero-mean random error. In model (1), the random error is allowed to be autocorrelated. For many SPC applications, it should be reasonable to assume that the autocorrelation is stationary and short-ranged. Namely, it is assumed that $\rho_y(\theta) = \text{Cov}\left(Y(t), Y(t+\theta)\right)$ depends on θ only, and $\rho_y(\theta) = 0$ when $\theta > \Theta$, where $\Theta > 0$ denotes the correlation range. Since observations of $\boldsymbol{X}$ are collected together with those for Y in the current research problem, both $\boldsymbol{X}$ and Y are treated as random variables in model (1) (cf., Cook and Weisberg 1999).

This paper focuses on developing a variable-sampling control chart for monitoring the quality variable Y by using helpful information in $\boldsymbol{X}$. The proposed method will be described in four parts. In the first part, we discuss how to compute initial estimates of some parameters of the IC distribution of $(\boldsymbol{X}, Y)$ from an IC dataset. Recursive updates of these estimates during online process monitoring and the corresponding self-starting control chart for monitoring a stationary, short-ranged, autocorrelated process is discussed in the second part. In the third part, the proposed variable-sampling control chart is described in detail. Finally, determination of some procedure parameters in the proposed chart is discussed in the fourth part.

2.1 Initial estimation of IC parameters

Let $\{(\boldsymbol{X}(t_i), Y(t_i)), i = 1, \ldots, m\}$ be an IC dataset collected before the online process monitoring, where $t_i \in [0, T]$ is the ith observation time, for each i. In this part, we mainly discuss how to estimate the IC quantities μ_y, $\rho_y(\theta)$ and $f(\cdot)$ from the IC dataset, where μ_y is the IC mean of Y. For μ_y and the IC variance $\rho_y(0)$, it is natural to estimate them by the sample mean $\widehat{\mu}_y^{(0)} = \sum_{i=1}^m Y(t_i)/m$ and the sample variance $\widehat{\rho}_y^{(0)}(0) = \sum_{i=1}^m [Y(t_i) - \widehat{\mu}_y^{(0)}]^2/(m-1)$. So, we focus on the estimation of $\rho_y(\theta)$, for $0 < \theta \leq \Theta$, and $f(\cdot)$ next.

To estimate the stationary covariance function $\rho_y(\theta)$, Hall et al. (1994) suggested the following estimator based on the local constant kernel smoothing procedure:

$$\widehat{\rho}_{y,H}^{(0)}(\theta) = \left[\sum_{i=1}^m \sum_{j=1}^m \widehat{Y}_{ij} K\left(\left(t_{ij} - \theta\right)/h\right)\right] \left[\sum_{i=1}^m \sum_{j=1}^m K\left(\left(t_{ij} - \theta\right)/h\right)\right]^{-1},$$

where $\widehat{Y}_{ij} = \{Y(t_i) - \widehat{\mu}_y^{(0)}\}\{Y(t_j) - \widehat{\mu}_y^{(0)}\}$, $t_{ij} = t_i - t_j$, $K(\cdot)$ is a kernel function, and $h > 0$ is a bandwidth. Under some regularity conditions, Hall et al. (1994) showed that the above kernel estimator had the good theoretical property that it could converge to the true covariance function $\rho_y(\theta)$ when m tended to infinity. In the local smoothing literature, however, it has been well discussed that the local constant kernel estimator, such as $\widehat{\rho}_{y,H}^{(0)}(\theta)$, could be biased in boundary regions (cf., Qiu 2005, Chapter 2), due to the fact that we can only use observations in a portion of a local neighborhood of a given point in such regions for computing the estimate at that point. To overcome this limitation, we suggest using the local linear kernel smoothing (LLKS) procedure to estimate the covariance function, which is discussed in detail in Section A of the supplementary file. Throughout this paper, the initial LLKS estimator of $\rho_y(\theta)$ is denoted as $\widehat{\rho}_y^{(0)}(\theta)$. Compared to the local constant kernel estimator, the LLKS estimator can correct the boundary problem automatically, which is the so-called "automatic boundary carpentry" property. See Hastie and Loader (1993) for a more detailed discussion.

Next, let us discuss initial estimation of the regression function $f(\cdot)$ in the IC model (1). For simplicity, we focus on cases when the relationship between Y and $\boldsymbol{X}$ is linear, i.e., $f(\boldsymbol{X}) = \beta_0 + \boldsymbol{X}^T\boldsymbol{\beta}$, where β_0 and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ are regression coefficients. Then, the

least squares (LS) estimates of these coefficients are

$$\widehat{\beta}_0^{(0)} = \frac{\gamma_2^{(0)}\eta_0^{(0)} - \boldsymbol{\gamma}_1^{(0)^T}\boldsymbol{\eta}_1^{(0)}}{m\gamma_2^{(0)} - \boldsymbol{\gamma}_1^{(0)^T}\boldsymbol{\gamma}_1^{(0)}}, \quad \text{and} \quad \widehat{\boldsymbol{\beta}}^{(0)} = \frac{m\boldsymbol{\eta}_1^{(0)} - \boldsymbol{\gamma}_1^{(0)}\eta_0^{(0)}}{m\gamma_2^{(0)} - \boldsymbol{\gamma}_1^{(0)^T}\boldsymbol{\gamma}_1^{(0)}}, \tag{2}$$

where $\boldsymbol{\gamma}_1^{(0)} = \sum_i \boldsymbol{X}(t_i)$, $\gamma_2^{(0)} = \sum_i \boldsymbol{X}(t_i)^T\boldsymbol{X}(t_i)$, $\eta_0^{(0)} = \sum_i Y(t_i)$ and $\boldsymbol{\eta}_1^{(0)} = \sum_i \boldsymbol{X}(t_i)Y(t_i)$. The initial estimate of $f(\boldsymbol{X})$ can be defined to be $\widehat{f}^{(0)}(\boldsymbol{X}) = \widehat{\beta}_0^{(0)} + \boldsymbol{X}^T\widehat{\boldsymbol{\beta}}^{(0)}$. This LS approach should work well when the dimensionality p is not large. In cases when p is large, variable selection methods by LASSO or other methods designed for handling high-dimensional data should be helpful (e.g., Tibshirani 1996; Zou and Hastie 2005). In some applications, it may not be appropriate to assume that $f(\cdot)$ is linear. In such cases, nonparametric regression techniques, including the local polynomial kernel smoothing methods and smoothing splines, should be helpful. For example, the LLKS procedure that is similar to the one described by (A.1) and (A.2) in Section A of the supplementary file can be used for estimating the nonparametric regression function $f(\cdot)$ when $p = 1$. When $p > 1$, estimation of a nonparametric function becomes more complicated because of the curse of dimensionality. In such cases, the generalized additive model could be a helpful tool (cf., Hastie and Tibshirani 1990).

2.2 Recursive updates of IC parameter estimates and self-starting process monitoring

For the initial estimates of the IC quantities to approximate the true IC quantities well, a large IC dataset is required, especially for estimates of $\rho_y(\theta)$ which are second moments of the IC process distribution. In SPC applications, however, a large IC dataset is often hard to collect. Consequently, control charts using these initial estimates would be unreliable. To address this issue, Hawkins (1987) suggested the self-starting idea for process monitoring, by which process observations collected during online process monitoring could be merged into the IC dataset if the process is declared to be IC at the current time point, and the estimates of the IC quantities could be updated using the expanded IC dataset. See Sullivan and Jones (2002) for a related discussion in multivariate SPC cases. To make our proposed new method feasible for different applications, the self-starting idea will be accommodated,

which is discussed in this part.

Let $\{(\boldsymbol{X}(t_n^*), Y(t_n^*)), n \geq 1\}$ be the process observations obtained during online process monitoring, where $\{t_n^*, n \geq 1\}$ are the observation times. We aim to develop a self-starting control chart for monitoring the serially correlated process observations $\{Y(t_n^*)\}$. Since the IC dataset $\{(\boldsymbol{X}(t_i), Y(t_i)), i = 1, \ldots, m\}$ will also be used for that purpose, let us introduce a new notation T_i for simplicity of presentation, where $T_i = t_i$ when $1 \leq i \leq m$ and $T_i = t_{i-m}^*$ when $i > m$. Then, $\{T_i\}$ contains observation times of both the IC dataset and the process observations obtained during online process monitoring. If the process is declared to be IC at t_n^*, then the estimates of μ_y and $\rho_y(\theta)$ can be updated by the following recursive formulas: for $n \geq 1$,

$$\begin{aligned}
\widehat{\mu}_y^{(n)} &= \frac{1}{m+n} Y(t_n^*) + \frac{m+n-1}{m+n} \widehat{\mu}_y^{(n-1)}, \\
\widehat{\rho}_y^{(n)}(0) &= \frac{1}{m+n-1} \left\{ Y(t_n^*) - \widehat{\mu}_y^{(n)} \right\}^2 + \frac{m+n-2}{m+n-1} \widehat{\rho}_y^{(n-1)}(0), \\
\widehat{\rho}_y^{(n)}(\theta) &= \frac{\varphi_2^{(n)} \psi_0^{(n)} - \varphi_1^{(n)} \psi_1^{(n)}}{\varphi_2^{(n)} \varphi_0^{(n)} - \varphi_1^{(n)} \varphi_1^{(n)}}, \quad \text{for } \ 0 < \theta \leq \Theta,
\end{aligned} \tag{3}$$

where $\varphi_k^{(n)}$ and $\psi_l^{(n)}$, for $k = 0, 1, 2$ and $l = 0, 1$, are defined recursively by the following formulas:

$$\begin{aligned}
\varphi_k^{(n)} &= \varphi_k^{(n-1)} + \sum_{i=1}^{m+n-1} (T_{i,m+n} - \theta)^k K\left(\frac{T_{i,m+n} - \theta}{h} \right) + \sum_{j=1}^{m+n} (T_{m+n,j} - \theta)^k K\left(\frac{T_{m+n,j} - \theta}{h} \right), \\
\psi_l^{(n)} &= \psi_l^{(n-1)} + \sum_{i=1}^{m+n-1} (T_{i,m+n} - \theta)^l \widehat{Y}_{i,m+n}^* K\left(\frac{T_{i,m+n} - \theta}{h} \right) + \sum_{j=1}^{m+n} (T_{m+n,j} - \theta)^l \widehat{Y}_{m+n,j}^* K\left(\frac{T_{m+n,j} - \theta}{h} \right),
\end{aligned}$$

$T_{ij} = T_i - T_j$, $\widehat{Y}_{ij}^* = \left(Y(T_i) - \widehat{\mu}_y^{(n)} \right) \left(Y(T_j) - \widehat{\mu}_y^{(n)} \right)$, for $1 \leq i, j \leq m+n$, and $\varphi_k^{(0)}$ and $\psi_l^{(0)}$ are defined in Section A of the supplementary file. The LS estimates of β_0 and $\boldsymbol{\beta}$ can be updated recursively as follows:

$$\widehat{\beta}_0^{(n)} = \frac{\gamma_2^{(n)} \eta_0^{(n)} - \boldsymbol{\gamma}_1^{(n)^T} \boldsymbol{\eta}_1^{(n)}}{m\gamma_2^{(n)} - \boldsymbol{\gamma}_1^{(n)^T} \boldsymbol{\gamma}_1^{(n)}}, \quad \text{and} \quad \widehat{\boldsymbol{\beta}}^{(n)} = \frac{m\boldsymbol{\eta}_1^{(n)} - \boldsymbol{\gamma}_1^{(n)} \eta_0^{(n)}}{m\gamma_2^{(n)} - \boldsymbol{\gamma}_1^{(n)^T} \boldsymbol{\gamma}_1^{(n)}}, \tag{4}$$

where

$$\begin{aligned}
\boldsymbol{\gamma}_1^{(n)} &= \boldsymbol{\gamma}_1^{(n-1)} + \boldsymbol{X}(t_n^*), \quad \gamma_2^{(n)} = \gamma_2^{(n-1)} + \boldsymbol{X}(t_n^*)^T \boldsymbol{X}(t_n^*), \\
\eta_0^{(n)} &= \eta_0^{(n-1)} + Y(t_n^*), \qquad \boldsymbol{\eta}_1^{(n)} = \boldsymbol{\eta}_1^{(n-1)} + \boldsymbol{X}(t_n^*) Y(t_n^*).
\end{aligned}$$

Then, the estimate of $f(\boldsymbol{X})$ can be updated by $\widehat{f}^{(n)}(\boldsymbol{X}) = \widehat{\beta}_0^{(n)} + \boldsymbol{X}^T \widehat{\boldsymbol{\beta}}^{(n)}$. It should be pointed out that recursive computation is critically important for a self-starting control chart, since the update of the IC parameter estimates needs to be implemented at each observation time during online process monitoring. Otherwise, the chart could be infeasible to use due to its heavy computational burden.

Conventional SPC charts are designed for cases when the process observations are independent of each other at different observation times. In reality, however, this assumption is rarely valid and serial data correlation is common. In the SPC literature, it has been well demonstrated that the conventional SPC charts would not be reliable to monitor processes with serially correlated data (cf., Apley and Tsung 2002, Qiu et al. 2020). To address this issue, we suggest using a data decorrelation and standardization algorithm based on the Cholesky decomposition of the covariance matrix of the data to remove the data correlation, as discussed in Qiu et al. (2020). The detailed description of this algorithm is given in Section B of the supplementary file. After using the data decorrelation and standardization algorithm, we obtain a sequence of the transformed process observations $\{Y^*(t_n^*), n \geq 1\}$. By some simple algebraic manipulations, it can be shown that $\{Y^*(t_n^*)\}$ are uncorrelated with mean 0 and variance 1 if $\widehat{\mu}_y^{(0)}$ and $\widehat{\mu}_y^{(n-1)}$ were the true IC process mean and $\widehat{\rho}_y^{(0)}(\theta)$ and $\widehat{\rho}_y^{(n-1)}(\theta)$ were the true IC process covariance function (cf., Li and Qiu 2016). So, it follows that $\{Y^*(t_n^*)\}$ are asymptotically uncorrelated with each other, and each has the asymptotic mean of 0 and the asymptotic variance of 1, since the estimated IC mean and IC covariance function are used in the above decorrelation algorithm.

After the data decorrelation and standardization, the following conventional EWMA chart can be considered for online process monitoring: for $n \geq 1$,

$$E_y(t_n^*) = \lambda Y^*(t_n^*) + (1-\lambda) E_y(t_{n-1}^*), \tag{5}$$

where $E_y(t_0^*) = 0$ and $\lambda \in (0, 1]$ is a weighting parameter. A signal of an upward mean shift is given by this chart if the charting statistic $E_y(t_n^*)$ defined in (5) exceeds a control limit $L > 0$. Control charts for detecting a downward or an arbitrary mean shift can be defined similarly. See Chapter 5 in Qiu (2014) for a more detailed discussion about the conventional

EWMA charts and their designs.

2.3 Proposed variable-sampling control chart

Our proposed variable-sampling control chart tries to determine the sampling intervals from the covariates in $\boldsymbol{X}$. From model (1), $\boldsymbol{X}$ affects the quality variable Y through $Z = f(\boldsymbol{X})$ in cases when the process is IC. Thus, the likelihood of a shift in Z could provide helpful information for detecting a shift in Y. Based on this intuition, we suggest determining the sampling intervals by the related information in Z. At the beginning of this section, it is assumed that the process under monitoring $\{Y(t), t \geq 0\}$ has a stationary and short-range autocorrelation. It is also assumed that the covariate process $\{\boldsymbol{X}(t), t \geq 0\}$ is a stationary process with a short-range dependence and the range of data dependence is the same as that of $\{Y(t), t \geq 0\}$. Then, it can be checked easily that $Z(t) = f(\boldsymbol{X}(t))$ is a stationary process as well, and its covariance function $\rho_z(\theta) = \text{Cov}\,(Z(t), Z(t+\theta))$ depends on θ only and is zero when $\theta > \Theta$.

To have a measurement of the likelihood that $Z(t)$ has a shift at the current time point t_n^*, we suggest using an EWMA chart described below. Let $\widehat{Z}(t_n^*) = \widehat{f}^{(n)}(\boldsymbol{X}(t_n^*))$, for $n \geq 1$, be the estimated covariate effect on the quality variable Y at t_n^*, where $\widehat{f}^{(n)}(\cdot)$ is defined immediately below (4) in Sections 2.2. Then, $\{\widehat{Z}(t_n^*)\}$ could be serially correlated. Thus, data decorrelation is needed before a conventional chart can be applied to them, which can be accomplished by the data decorrelation and standardization algorithm discussed in Section B of the supplementary file. More specifically, we can first calculate the sample mean $\widehat{\mu}_z^{(0)}$ and the sample variance $\widehat{\rho}_z^{(0)}(0)$ from the data $\{\widehat{Z}(t_i) = \widehat{f}^{(0)}(\boldsymbol{X}(t_i)), i = 1, \ldots, m\}$, where $\widehat{f}^{(0)}(\cdot)$ is the initial estimate of the regression function defined in Section 2.1. Then, after replacing $\widehat{Y}_{ij}$ by $\widehat{Z}_{ij} = \{\widehat{Z}(t_i) - \widehat{\mu}_z^{(0)}\}\{\widehat{Z}(t_j) - \widehat{\mu}_z^{(0)}\}$ in the LLKS procedure in (A.1)-(A.2) of the supplementary file, the initial estimate of the covariance function $\widehat{\rho}_z^{(0)}(\theta)$, for $\theta \in (0, \Theta]$, can be obtained. Similar to (3), recursive formulas can be derived for updating the estimates of μ_z and $\rho_z(\theta)$, and the updated estimates at time t_n^* are denoted as $\widehat{\mu}_z^{(n)}$ and $\widehat{\rho}_z^{(n)}(\theta)$, for $n \geq 1$. Finally, $\{\widehat{Z}(t_n^*)\}$ can be decorrelated and standardized using the algorithm discussed

in Section B of the supplementary file, after $Y(t_n^*)$, $\widehat{\mu}_y^{(n-1)}$ and $\widehat{\rho}_y^{(n-1)}(\theta)$ are replaced by $\widehat{Z}(t_n^*)$, $\widehat{\mu}_z^{(n-1)}$ and $\widehat{\rho}_z^{(n-1)}(\theta)$, respectively. The resulting decorrelated and standardized data are denoted as $\{Z^*(t_n^*), n \geq 1\}$.

After the above data decorrelation and standardization, the following EWMA chart can be applied to the decorrelated data $\{Z^*(t_n^*), n \geq 1\}$: for $n \geq 1$,

$$E_z(t_n^*) = \lambda Z^*(t_n^*) + (1-\lambda)E_z(t_{n-1}^*), \tag{6}$$

where $E_z(t_0^*) = 0$ and λ is the weighting parameter. Then, for a given value of λ, $E_z(t_n^*)$ can be used as a measure of the likelihood of a shift in $\{\widehat{Z}(t_n^*), n \geq 1\}$. From Model (1), $Y(t_n^*) = Z(t_n^*) + \varepsilon(t_n^*)$ when the related process is IC. Thus, $E_z(t_n^*)$ also measures the likelihood of a shift in $\{Y(t_n^*), n \geq 1\}$ due to the covariates. When $E_z(t_n^*)$ is positively larger, for instance, the likelihood of an upward shift in $\{Y(t_n^*), n \geq 1\}$ would be larger. However, the variability of $E_z(t_n^*)$ depends on λ. To get rid of such dependence, we can use the standardized version $\widetilde{E}_z(t_n^*) = \{\lambda[1-(1-\lambda)^{2n}]/(2-\lambda)\}^{-1/2}E_z(t_n^*)$, where $\lambda[1-(1-\lambda)^{2n}]/(2-\lambda)$ is the variance of $E_z(t_n^*)$ (cf., Qiu 2014, Chapter 5). Then, $\widetilde{E}_z(t_n^*)$ can be used for determining the sampling intervals, as described below.

At the current time point t_n^*, assume that the process under monitoring is declared to be IC. Then, we need to determine the next observation time t_{n+1}^* with the sampling interval $t_{n+1}^* - t_n^* =: d(\widetilde{E}_z(t_n^*))$, where $d(\cdot)$ is the variable-sampling interval function. In this paper, we suggest determining the sampling interval $t_{n+1}^* - t_n^*$ by $\widetilde{E}_z(t_n^*)$. So, $d(\widetilde{E}_z(t_n^*))$ is assumed to be a function of $\widetilde{E}_z(t_n^*)$. To choose the function $d(\cdot)$ properly, first note that a larger value of $\widetilde{E}_z(t_n^*)$ would indicate a larger chance to have an upward mean shift in the quality variable Y. Thus, $d(\widetilde{E}_z(t_n^*))$ should be chosen smaller. Namely, $d(\cdot)$ should be a decreasing function. In their dynamic sampling scheme, Li and Qiu (2014) suggested choosing $d(\cdot)$ from the quite flexible Box-Cox transformation family. However, that method cannot be used here, because the Box-Cox transformation is for transforming a positive variable, but $\widetilde{E}_z(t_n^*)$ could be negative. To overcome this difficulty, John and Draper (1980) suggested a modified version of the Box-Cox transformation which can be used with negative values. Based on that modified Box-Cox transformation, we suggest using the following variable-sampling interval

function:

$$d(x) = \phi\left(J_\alpha(x) + \kappa\right), \tag{7}$$

where $J_\alpha(x) = 1 + \text{sign}(x)\left[\{1/(1+|x|)\}^\alpha - 1\right]$ is the John-Draper function, and $\alpha \geq 0$, $\phi > 0$ and $\kappa \in (-\infty, \infty)$ are three parameters. Figure 1 displays several John-Draper functions when α changes among 0, 1, 2, 4, 6 and 8. From the figure, it can be seen that (i) $J_0(x) = 1$ for all x, (ii) when $\alpha > 0$, $J_\alpha(x)$ is a decreasing function with $J_\alpha(x) < 1$ when $x > 0$, (iii) the larger the value of α, the greater the difference between $J_\alpha(x)$ and 1, and (iv) the John-Draper functions are similar to each other when $\alpha \geq 6$. Selection of the parameters α, ϕ and κ will be discussed in Section 2.4.

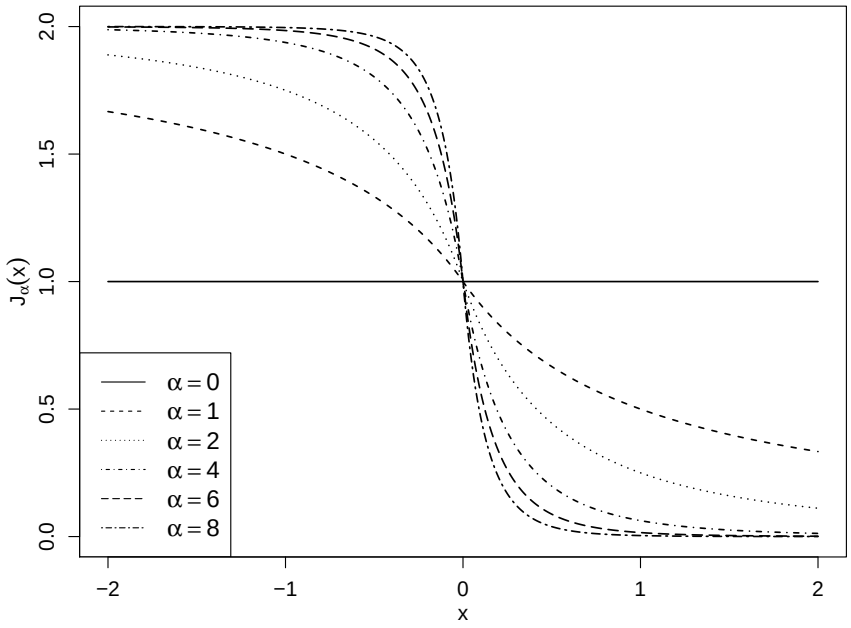

Figure 1: Several John-Draper functions $J_\alpha(x)$.

The performance of a control chart with a fixed sampling interval is usually evaluated by the IC average run length (ARL), denoted as ARL_0, which is defined to be the average number of time points from the beginning of process monitoring to the signal time when the process is IC, and the out-of-control (OC) ARL, denoted as ARL_1, which is the average number of time points from the occurrence of a shift to the signal time. For a variable-sampling control chart, the sampling times could be unequally spaced. In such cases, ARL_0 and ARL_1

cannot effectively measure its performance anymore, and the alternative performance measures are the IC average time to signal (ATS), defined to be the average time length from the start of online sequential monitoring to the signal time when the process is IC, and OC ATS, defined to be the average time length from the occurrence of a shift to the signal time. The IC and OC ATS values are denoted as ATS_0 and ATS_1, respectively. When the process is IC, a control chart with a larger ATS_0 value would have a smaller false alarm rate. When the process becomes OC, a control chart with a smaller ATS_1 value would have a better OC performance since it can detect a shift more quickly.

2.4 Design of the proposed variable-sampling control chart

To design a variable-sampling control chart, usually the ARL_0 value is pre-specified at a specific level (e.g., 200 or 370), and it is required that $\text{ATS}_0 = \text{ARL}_0$ so that control charts with a fixed sampling scheme and a variable sampling scheme have the same IC performance. In the proposed variable-sampling control chart (5)-(7), there are five parameters $(\lambda, L, \alpha, \kappa, \phi)$ to determine beforehand. Usually, the value of λ is pre-specified. Then, the control limit L can be determined by a bootstrap procedure (cf., Chatterjee and Qiu 2009) from the IC dataset such that the pre-specified ARL_0 level is reached by the control chart (5) with a fixed sampling interval (i.e., $d(x) \equiv 1$ in (7)). A detailed description of the bootstrap procedure is given in Section C of the supplementary file. If the values of α and κ are given, then ϕ can be chosen such that $\text{ATS}_0 = \text{ARL}_0$. Next, we discuss the determination of α. Figure 1 shows that the difference between $J_\alpha(x)$ and $J_0(x) = 1$ becomes larger when α increases. Thus, a large α value should be used for detecting shifts that are highly associated with the covariates. Since the difference among the John-Draper functions is very small when $\alpha \geq 6$, we suggest choosing $\alpha = 6$ in such cases. On the other hand, in cases when a future process shift is believed to have nothing to do with the covariates, $\widetilde{E}_z(t_i^*)$ would not contain much useful information for detecting the shift in Y. In such cases, it is expected that different values of α would not affect the performance of the variable-sampling control chart much. By combining the above two cases, we suggest choosing $\alpha = 6$ in practice.

Regarding the selection of κ, we present the following theoretical result to provide an insight on the impact of κ on the performance of the chart. Consider two scenarios with the parameters $(\lambda, \alpha, \kappa, \phi)$ and $(\lambda, \alpha, \kappa^*, \phi^*)$, respectively, for detecting a given upward mean shift in Y. In both cases, a given ATS_0 level is reached, and the ATS_1 values are denoted as ATS_1 and ATS_1^*. Then, we have the following result:

Theorem 1 *Assume that (i) the decorrelated and standardized data* $\{Y^*(t_n^*)\}$ *used in (5) are independent and identically distributed (i.i.d.) when the process is IC, and (ii)* $\{Z^*(t_n^*)\}$ *used in (6) are i.i.d. as well. Then, for detecting a step shift, we have*

$$ATS_1^* = \{1 - \phi^*(\kappa^* - \kappa)\}\, ATS_1 + \{\phi^*(\kappa^* - \kappa)\}\, ATS_1^{(0)}, \tag{8}$$

where $ATS_1^{(0)}$ *is the OC ATS of the control chart (5) with a constant sampling interval (i.e.,* $d(x) \equiv 1$ *in (7)) and the same values of* λ *and* ATS_0 *as those of the variable-sampling control charts.*

The proof of this theorem is given in the supplementary material.

In Theorem 1, it is assumed that the decorrelated and standardized data are i.i.d., which should be approximately true since the data decorrelation is essentially a linear transformation and thus the Central Limit Theorem would guarantee the decorrelated data to be asymptotically independent under some regularity conditions. To interpret the result in the theorem, without loss of generality, let us assume that $\kappa^* \geq \kappa$. In such cases, Expression (8) shows that ATS_1^* is a weighted average of ATS_1 and $\text{ATS}_1^{(0)}$. In cases when a future shift in Y is associated with the covariates, the variable-sampling control chart with the parameters $(\lambda, \alpha, \kappa, \phi)$ should have a better OC performance than the corresponding fixed-sampling chart. So, we have $\text{ATS}_1 \leq \text{ATS}_1^{(0)}$. Therefore, $\text{ATS}_1 \leq \text{ATS}_1^*$ (i.e., ATS_1^* is an increasing function of κ^*) in such cases, and consequently κ in the proposed variable-sampling control chart (5)-(7) should be chosen as small as possible. Recall that $d(x) = \phi(J_\alpha(x) + \kappa)$ is the variable-sampling interval function and $\phi > 0$. Thus, we must have $J_\alpha(x) + \kappa \geq 0$. So, the optimal choice for κ in such cases is $-J_\alpha(\varpi)$, where ϖ is the least upper bound of the set $\{\widetilde{E}_z(t_n^*), n \geq 1\}$. As discussed above, we suggest choosing $\alpha = 6$ in practice. Then, from

Figure 1, $J_6(x)$ is very close to 0 when $x \geq 1$. So, we suggest choosing $\kappa = 0$ unless we are certain that the standardized charting statistic $\widetilde{E}_z(t_n^*)$ is bounded by a number less than 1. In cases when a future shift in Y is not associated with the covariates, different values of κ would not make much difference on the performance of the proposed control chart, since $\widetilde{E}_z(t_n^*)$ does not contain any useful information in such cases. By combining the above two cases, we suggest choosing $\kappa = 0$ in practice.

3 Numerical Studies

In this section, we present some numerical results about the performance of the proposed variable-sampling control chart (5)-(7), denoted as NEW hereafter. First, let us provide a detailed description about the simulation setup. Assume that the IC observation times $\{t_i = i \times m/T, i = 1, \ldots, m\}$ are equally spaced in the interval $[0, T] = [0, 100]$, and the IC sample size m is assumed to be 1,000 if there is no further specification. The IC process observations $\{(\boldsymbol{X}(t_i), Y(t_i))\}$ are generated from the IC model:

$$Y(t_i) = \beta_0 + \boldsymbol{X}^T(t_i)\boldsymbol{\beta} + \varepsilon(t_i),$$

where $\beta_0 = 0$, $\boldsymbol{\beta} = (0.5, \ldots, 0.5)^T$, and the dimension p of $\boldsymbol{X}(t_i)$ is assumed to be 1, 3 or 5. In the above IC model, it is further assumed that the covariate vectors and the random errors follow the AR(1) models:

$$\begin{aligned}\boldsymbol{X}(t_i) =& \rho^{t_i - t_{i-1}}\boldsymbol{X}(t_{i-1}) + \{1 - \rho^{2(t_i - t_{i-1})}\}^{1/2}\boldsymbol{\zeta}_x(t_i),\\ \varepsilon(t_i) =& \ \rho^{t_i - t_{i-1}}\varepsilon(t_{i-1}) + \{1 - \rho^{2(t_i - t_{i-1})}\}^{1/2}\zeta(t_i),\end{aligned}$$

where $\{\boldsymbol{\zeta}_x(t_i)\}$ are i.i.d. with the $N_p(\boldsymbol{0}, \boldsymbol{I}_p)$ distribution, and $\{\zeta(t_i)\}$ are i.i.d. with the standard normal distribution $N(0, 1)$. Then, it is easy to check that the stationarity assumption is satisfied in the IC model. In the above AR(1) models, the parameter ρ is chosen to be 0, 0.25 and 0.5, and the related three cases are referred to as Cases (I)-(III), respectively. Case (IV) is the same as Case (II), except that the distributions $N_p(\boldsymbol{0}, \boldsymbol{I}_p)$ and $N(0, 1)$ are replaced by the centralized versions of $t_p(3, \boldsymbol{I}_p)$ and $t(3)$ distributions, respectively, where $t_p(3, \boldsymbol{I}_p)$ denotes a joint distribution of p i.i.d. random variables with the common distribution of $t(3)$.

Obviously, Cases (I)-(III) are for studying the impact of serial data correlation on the performance of the proposed chart, and Case (IV) is for studying the impact of the shape of the data distribution. Next, we consider two cases with more complicated serial data correlation. Because the sampling times may not be equally spaced when using a variable-sampling chart, the data generated from the traditional discrete-time ARMA models or Markov models may not be stationary either in such cases. To overcome this difficulty, we propose the following strategy for data generation. Let $\omega > 0$ be a basic time unit, which is the largest time unit that all observation times are its integer multiples. Then, the following two cases are considered in this section:

Case (V): $\boldsymbol{X}(i\omega) = 0.6\boldsymbol{X}((i-1)\omega) + 0.3\boldsymbol{X}((i-2)\omega) + 0.3\boldsymbol{\zeta}_x((i-1)\omega) + \boldsymbol{\zeta}_x(i\omega)$, $\varepsilon(i\omega) = 0.6\varepsilon((i-1)\omega) + 0.3\varepsilon((i-2)\omega) + 0.3\zeta((i-1)\omega) + \zeta(i\omega)$, where $\{\boldsymbol{\zeta}_x(i\omega)\}$ are i.i.d. with the multivariate central χ^2-distribution $\chi^2_p(5, \boldsymbol{I}_p)$ and $\{\zeta(i\omega)\}$ are i.i.d. with the $\chi^2(5)$ distribution.

Case (VI): $X_j(i\omega) = \xi_{x,j}((i-1)\omega) + \zeta_{x,j}(i\omega)$, for $j = 1, \ldots, p$, and $\varepsilon(i\omega) = \xi((i-1)\omega) + \zeta(i\omega)$, where $\{\xi_{x,j}(i\omega), i \geq 1\}$ is a two-state Markov process with the transition matrix $\begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix}$ and the two states $\{0, 1\}$, $\{\xi(i\omega)\}$ are generated from the same two-state Markov model, and $\{\zeta_{x,j}(i\omega)\}$ and $\{\zeta(i\omega)\}$ are both i.i.d. with the $N(0, 1)$ distribution.

In the simulation studies, the basic time unit ω is chosen to be 0.1. When using the proposed variable-sampling chart, the computed sampling interval value may not be an integer multiple of ω and a proper rounding of this value is thus needed.

After the occurrence of a shift, the process observations are assumed to follow the following OC model:

$$Y(t) = \beta_{0,\delta} + [\boldsymbol{X}^T(t)\boldsymbol{\beta}]_\delta + \varepsilon(t),$$

where $\beta_{0,\delta} = \beta_0 + \sigma_y\delta_0$, $[\boldsymbol{X}^T(t)\boldsymbol{\beta}]_\delta = \boldsymbol{X}^T(t)\boldsymbol{\beta} + \sigma_y\delta_x$, σ_y is the IC standard deviation of Y, δ_x and δ_0 denote the shifts in Y due to the covariates and other factors that are not included

in the model, respectively. For $\nu = 1, 2, 3$ and 4, the following three types of shifts will be considered:

Type (i): $\delta_0 = 0.25\nu$ and $\delta_x = 0$,

Type (ii): $\delta_0 = 0.5 \times 0.25\nu$ and $\delta_x = 0.5 \times 0.25\nu$,

Type (iii): $\delta_0 = 0$ and $\delta_x = 0.25\nu$.

Clearly, Type (i) shifts are not associated with the covariates, Type (iii) shifts are completely due to the covariates, and Type (ii) shifts are partially due to the covariates.

Throughout this section, the range of serial data correlation Θ is fixed at 5, unless further stated otherwise. The control limit L and the parameter ϕ in NEW are chosen by the bootstrap procedure with the bootstrap sample size of $B = 10,000$. The ATS values are computed based on 1,000 replicated simulations. The IC means, covariance functions and the regression function are all assumed unknown and estimated initially from an IC dataset. So, performance of NEW could depend on the IC dataset. To account for this dependence, the entire simulation described above, from generation of the IC dataset, estimation of the IC quantities, determination of L and ϕ, and computation of ATS, is repeated for 100 times. Then, the average of the 100 ATS values in a given case is used for estimating the true ATS value, and the standard error of the final estimate can be computed accordingly.

3.1 Impact of α, κ and λ

The parameters α, κ and λ need to be chosen properly before using the proposed chart NEW. In this part, let us first investigate the impact of α on the performance of NEW. To this end, we first consider cases when $\alpha = 0, 2, 4, 6, 8$, or 10, $p = 1$, $\lambda = 0.1$, $\kappa = 0$, and $\text{ATS}_0 = 200$. For detecting the three types of shifts in Case (II), the computed ATS_1 values are presented in Figure 2. From the figure, we have the following conclusions. 1) For detecting shifts of Types (ii) and (iii), the performance of NEW is better with a larger α, but it does not change much when $\alpha \geq 6$. 2) For detecting shifts that are not associated

with the covariates at all (i.e., Type (i) shifts), NEW with a constant sampling interval (i.e., $\alpha = 0$) performs the best, although its performance would not change much with a different α value. The computed ATS_1 values in Cases (I) and (III)-(VI) are presented in Figure S.1 of the supplementary file, from which similar conclusions can be made. This example confirms our suggested value of $\alpha = 6$ given in Section 2.4.

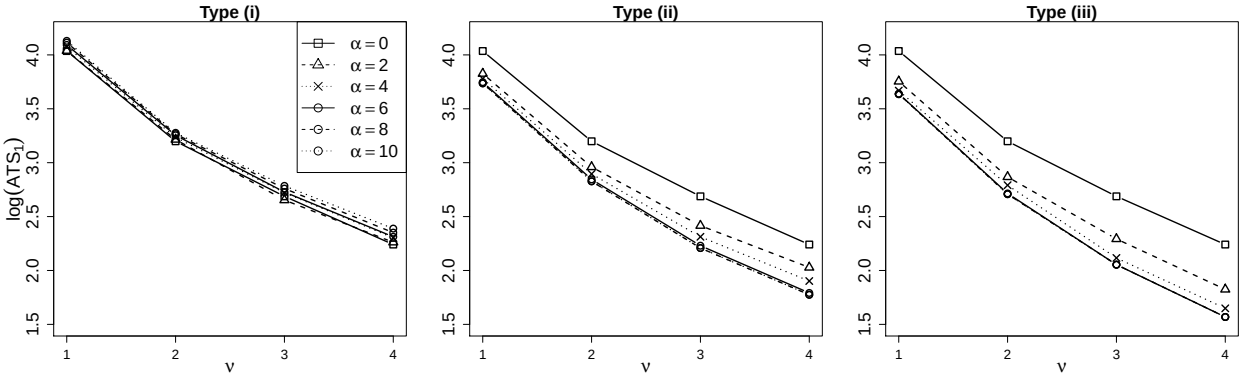

Figure 2: Calculated ATS_1 values of the proposed chart NEW for detecting the three types of shifts in Case (II) when $p = 1$, $\text{ATS}_0 = 200$, $\lambda = 0.1$ and $\kappa = 0$. In each plot, the y-axis is in the natural log scale.

Next, we study the impact of κ on the performance of NEW. In the setup of Figure 2, we fix α at 6 and consider six κ values in $\{0, 0.1, 0.2, 0.5, 1, 2\}$. The simulation results are shown in Figure 3. From the plot, it can be seen that 1) the performance of NEW improves when κ becomes closer to 0 for detecting shifts of Types (ii) and (iii), and 2) the calculated ATS_1 values of the chart using different κ values are almost the same for detecting shifts of Type (i). The corresponding results in Cases (I) and (III)-(VI) are presented in Figure S.2 of the supplementary file, and similar conclusions can be made. So, based on this example, our suggested value of $\kappa = 0$ given in Section 2.4 seems reasonable.

Besides the parameters α and κ, the OC performance of NEW would also depend on the weighting parameter λ. To investigate the impact of λ, in the previous example, we fix α at 6 and κ at 0, let λ change among $\{0.05, 0.1, 0.2\}$, and keep the other setups unchanged. The calculated ATS_1 of the chart for detecting the three types of shifts in Cases (I)-(VI) when $p = 1$ and $\text{ATS}_0 = 200$ are presented in Table 1. From the table, it seems that large values

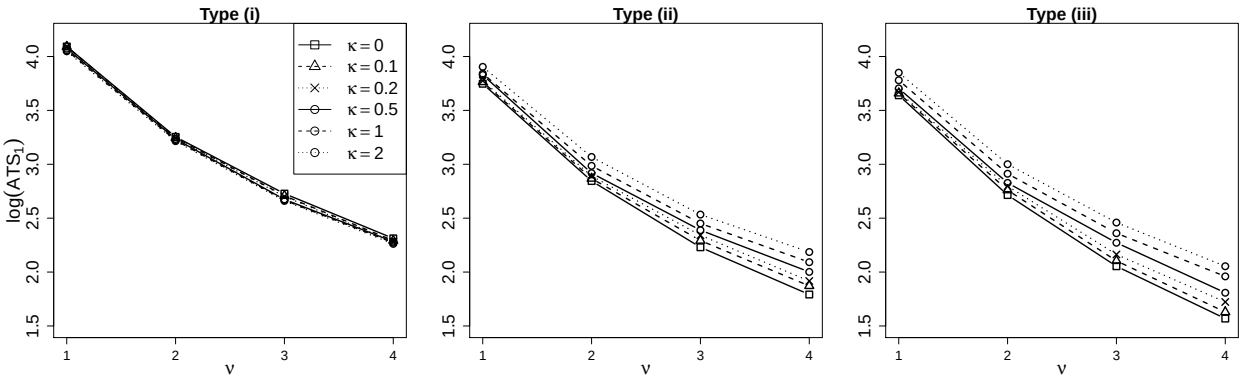

Figure 3: Calculated ATS_1 values of the proposed chart NEW for detecting the three types of shifts in Case (II) when $p = 1$, $\text{ATS}_0 = 200$, $\lambda = 0.1$ and $\alpha = 6$. In each plot, the y-axis is in the natural log scale.

of λ are good for detecting large shifts and small values of λ are good for detecting small shifts, which is generally true for an EWMA chart (cf., Qiu 2014, Chapter 5). By comparing the ATS_1 values in Cases (I)-(III), it can be seen that ATS_1 gets larger when the serial data correlation is stronger, which is intuitively reasonable.

3.2 Impact of m and Θ

Although the proposed chart NEW is a self-starting chart, enough IC data should be collected before online process monitoring to obtain a satisfactory performance. Otherwise, the initial estimates of certain IC quantities could have a large variability that can negatively affect the performance of the chart. Next, we study the impact of the IC data size m on the IC performance of NEW and then provide some practical guidelines on the selection of m. To this end, let m change between 300 and 1000 with a constant step of 100, p=1, 3, or 5, $\Theta = 5$, $\lambda = 0.1$, $\alpha = 6$, $\kappa = 0$, the nominal level of ATS_0 is fixed at 200, and other setups are the same as those in the example of Table 1. The computed ATS_0 values of NEW in Cases (I)-(VI) are presented in Figure 4, where the shaded region in each plot denotes ATS_0 values that are within 5% of the nominal value. From the plots in the figure, it can be seen that 1) when $m \geq 800$, the difference between the computed ATS_0 values and the nominal level of 200 is within 5% of the nominal level in all cases considered, 2) the IC performance is better

Table 1: Calculated ATS_1 values and their standard errors (in parentheses) of the chart NEW when $p = 1$, $\alpha = 6$, $\kappa = 0$ and $ATS_0 = 200$. Numbers in bold denote the smallest ATS_1 values when comparing different λ values.

		Type (i)			Type (ii)			Type (iii)		
Case	ν	$\lambda = 0.05$	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.05$	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.05$	$\lambda = 0.1$	$\lambda = 0.2$
(I)	1	**39.41**(0.57)	44.26(0.55)	52.60(0.70)	**14.74**(0.19)	14.89(0.24)	19.73(0.38)	10.93(0.15)	**10.61**(0.20)	14.40(0.37)
	2	**17.86**(0.21)	19.02(0.23)	21.16(0.26)	5.90(0.05)	**4.75**(0.04)	4.78(0.08)	3.99(0.04)	**3.04**(0.04)	3.12(0.07)
	3	**10.95**(0.11)	10.98(0.11)	11.52(0.14)	3.78(0.03)	2.89(0.02)	**2.56**(0.02)	2.30(0.02)	1.59(0.02)	**1.35**(0.02)
	4	7.81(0.09)	**7.62**(0.08)	7.66(0.08)	2.88(0.02)	2.22(0.01)	**1.92**(0.01)	1.57(0.02)	1.21(0.01)	**1.10**(0.01)
(II)	1	**52.56**(0.75)	59.90(0.98)	82.74(1.18)	**34.97**(0.76)	42.40(0.96)	57.34(1.04)	**30.28**(0.72)	38.08(0.89)	52.23(1.05)
	2	**23.95**(0.35)	25.97(0.32)	37.14(0.57)	**14.87**(0.35)	17.28(0.43)	24.51(0.48)	**12.61**(0.34)	15.11(0.43)	21.30(0.48)
	3	**14.24**(0.21)	15.30(0.20)	19.25(0.27)	**8.54**(0.19)	10.30(0.22)	13.02(0.28)	**6.94**(0.21)	7.84(0.23)	10.36(0.22)
	4	10.54(0.17)	**10.10**(0.11)	12.01(0.14)	6.33(0.12)	**6.03**(0.14)	8.20(0.16)	4.87(0.012)	**4.75**(0.13)	6.42(0.14)
(III)	1	**70.62**(1.16)	86.02(1.53)	105.92(2.22)	**48.38**(1.11)	58.39(1.07)	73.59(1.61)	**43.31**(1.06)	52.32(1.11)	67.15(1.34)
	2	**33.83**(0.55)	41.28(0.62)	54.61(1.47)	**21.47**(0.51)	25.17(0.55)	34.34(0.77)	**18.29**(0.49)	20.87(0.49)	28.05(0.64)
	3	**20.13**(0.33)	23.29(0.35)	30.77(0.99)	**12.57**(0.30)	14.33(0.31)	18.59(0.41)	**10.37**(0.29)	11.53(0.28)	14.40(0.35)
	4	**13.87**(0.25)	15.10(0.23)	18.54(0.46)	**8.52**(0.21)	9.29(0.19)	11.62(0.22)	**6.88**(0.18)	7.25(0.17)	8.93(0.16)
(IV)	1	105.92(2.22)	**73.90**(1.28)	107.36(2.07)	73.58(1.61)	**46.80**(1.91)	68.93(1.33)	67.15(1.34)	**41.38**(1.11)	61.00(1.55)
	2	54.61(1.46)	**29.89**(0.55)	52.33(1.41)	34.34(0.77)	**20.04**(0.59)	32.24(0.85)	28.05(0.64)	**16.35**(0.54)	27.34(0.80)
	3	30.76(0.99)	**16.17**(0.25)	20.98(0.61)	18.59(0.41)	**10.89**(0.35)	13.01(0.40)	14.40(0.35)	**8.71**(0.26)	10.29(0.28)
	4	18.54(0.46)	10.34(0.15)	**10.03**(0.14)	11.62(0.22)	6.94(0.18)	**6.19**(0.18)	8.92(0.16)	5.44(0.17)	**5.08**(0.16)
(V)	1	**52.89**(0.67)	65.35(0.95)	85.35(1.23)	**34.35**(0.58)	42.82(0.69)	59.85(1.01)	**30.19**(0.63)	38.08(0.75)	51.96(0.94)
	2	**23.83**(0.29)	28.64(0.33)	40.02(0.65)	**14.74**(0.29)	17.63(0.34)	25.85(0.53)	**12.30**(0.27)	14.62(0.32)	21.92(0.47)
	3	**14.30**(0.18)	15.85(0.15)	20.88(0.26)	**8.60**(0.16)	9.68(0.19)	14.15(0.31)	**6.95**(0.17)	7.76(0.19)	11.25(0.26)
	4	**10.04**(0.13)	10.64(0.10)	13.01(0.15)	**6.02**(0.11)	6.41(0.12)	8.71(0.16)	**4.59**(0.10)	4.90(0.11)	6.63(0.16)
(VI)	1	**40.20**(0.46)	45.24(0.55)	54.02(0.69)	**15.26**(0.19)	15.83(0.25)	21.12(0.41)	**11.56**(0.15)	11.58(0.22)	16.04(0.32)
	2	**17.52**(0.21)	19.08(0.24)	22.08(0.27)	6.14(0.05)	**5.16**(0.05)	5.54(0.09)	4.28(0.04)	**3.40**(0.04)	3.67(0.06)
	3	**10.80**(0.13)	11.22(0.12)	12.02(0.14)	3.95(0.03)	3.13(0.02)	**2.90**(0.02)	2.45(0.02)	1.78(0.02)	**1.65**(0.02)
	4	7.76(0.10)	**7.65**(0.09)	7.99(0.09)	2.99(0.02)	2.33(0.01)	**2.08**(0.01)	1.66(0.01)	1.16(0.01)	**1.06**(0.01)

when m is larger or p is smaller, and 3) by comparing the results in Cases (I)-(III), we can see that the performance is better when the serial data correlation is weaker. From these simulation results, it seems that it is necessary to have at least $m = 800$ IC observations before online process monitoring to have a reliable performance of NEW in cases when p is relatively small. This example also shows that the IC performance of NEW depends on the level of data correlation and the dimensionality p of the covariates. For instance, when the serial data correlation does not exist (i.e., Case (I)) and $p = 1$, the calculated ATS_0 value of NEW is within 5% of the nominal value when $m \geq 400$.

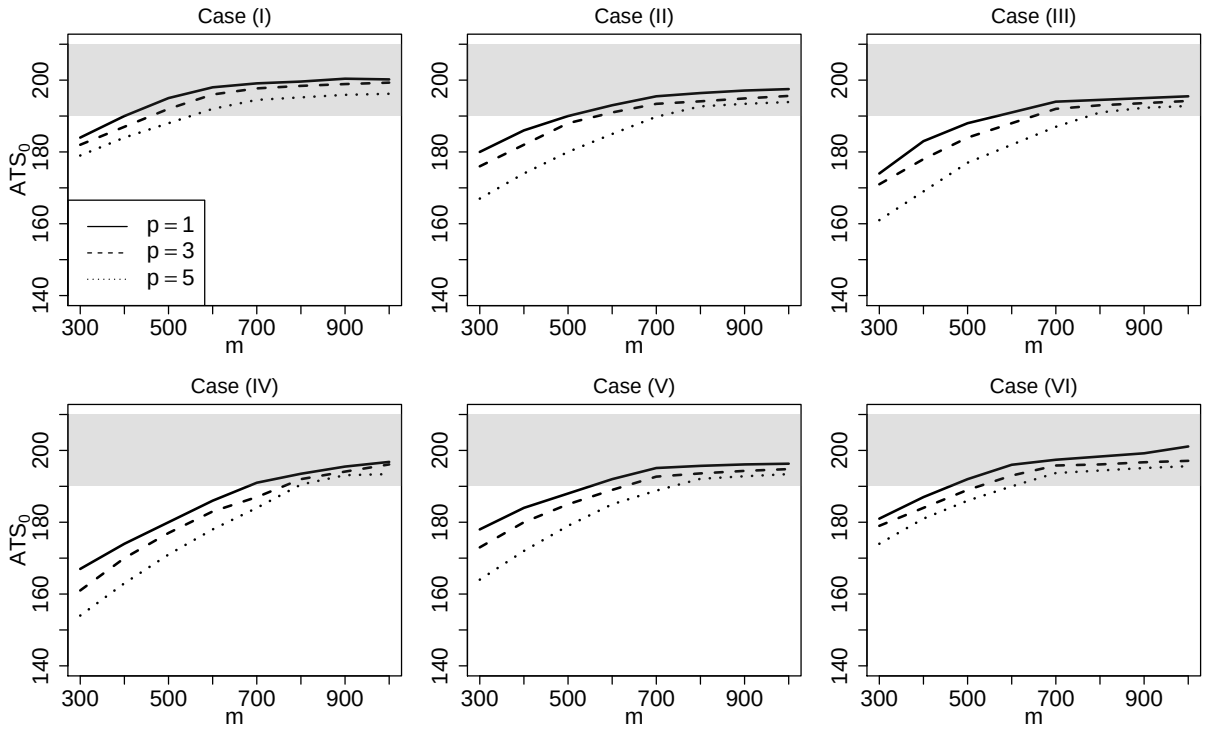

Figure 4: Calculated ATS_0 values of the chart NEW when $p = 1, 3$ or 5, $\Theta = 5$, $\lambda = 0.1$, $\alpha = 6$, $\kappa = 0$ and the nominal $\text{ATS}_0 = 200$. In each plot, the shaded area denotes those ATS_0 values that are within 5% of the nominal ATS_0 level.

In the proposed chart NEW, it is assumed that two observations are uncorrelated when the distance between their observation times is larger than Θ. So, the selected value of Θ could affect its performance. To study the impact of Θ, in the example of Figure 4, we fix m at 800, let Θ change among $\{1, 2, 3, 4, 5, 6, 8, 10\}$, and keep all the other setups. The calculated ATS_0 values of NEW are shown in Figure 5. From the figure, we find that 1)

the chart has a similar IC performance when $\Theta \geq 5$ in all cases considered, and 2) when $\Theta \geq 5$, the chart has a satisfactory IC performance in all cases considered, in the sense that its actual ATS_0 values are always within 5% of the nominal ATS_0 level. This example confirms that the chart NEW is reliable to use when Θ is chosen to be 5 or larger in the cases considered.

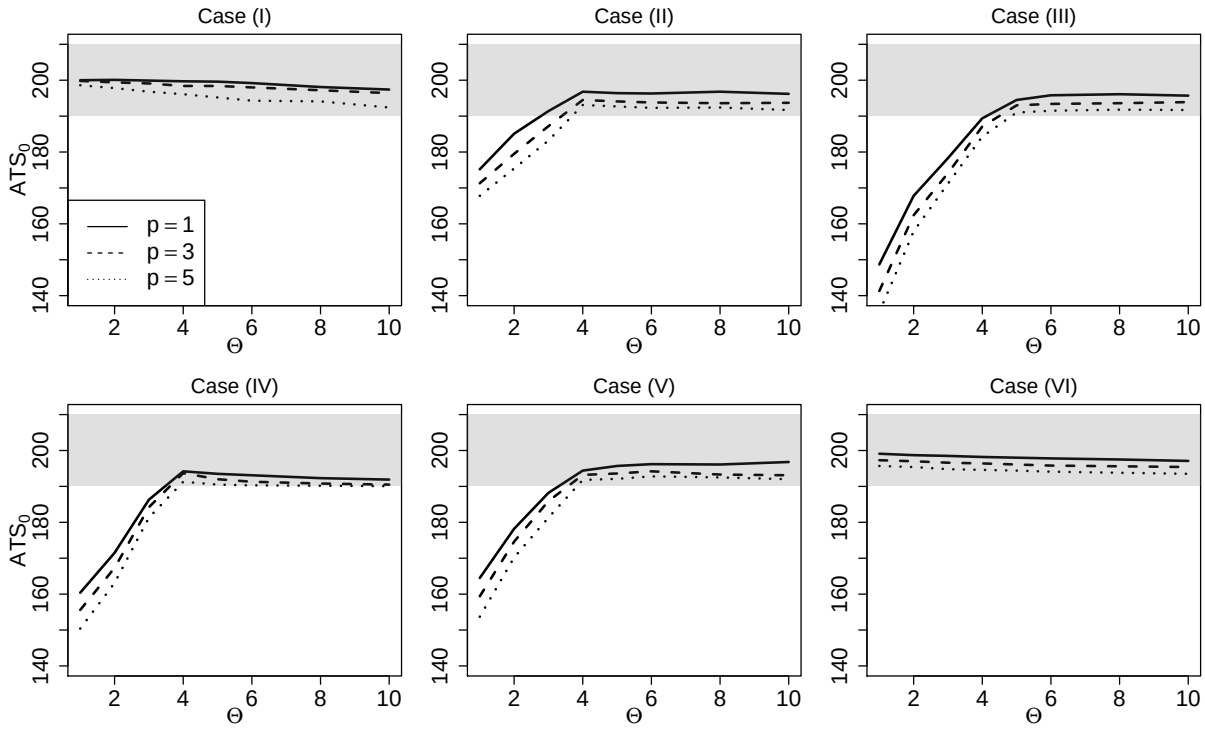

Figure 5: Calculated ATS_0 values of the chart NEW when $p = 1, 3$ or 5, $m = 800$, $\lambda = 0.1$, $\alpha = 6$, $\kappa = 0$ and the nominal $\text{ATS}_0 = 200$. In each plot, the shaded area denotes those ATS_0 values that are within 5% of the nominal ATS_0 level.

3.3 Comparison with alternative methods

In this part, we present some numerical results to compare the proposed variable-sampling chart NEW with some alternative methods. To this end, the following four alternative charts are considered in the comparison:

- In the proposed chart NEW, let $\alpha = 0$. The resulting chart then has a fixed sampling

interval, and thus does not use any covariate information. However, this chart is still self-starting and has all other features of NEW (e.g., data decorrelation during online process monitoring). It is denoted as WSS-WOC, representing "With Self-Starting but WithOut using Covariate information."

- In the proposed chart NEW, the self-starting feature is removed, but all other features are preserved. In such cases, the initial estimates of certain IC quantities would not be updated during online process monitoring. The resulting chart is denoted as WOSS-WC.

- In the chart WOSS-WC described above, set $\alpha = 0$. Then, the resulting chart does not have the self-starting feature and it does not use any covariate information either. This chart is denoted as WOSS-WOC.

- In the proposed chart NEW, consider using the conventional 2-interval sampling scheme. Let $0 < d_1 < d_2$ be two possible sampling interval levels. Then, the sampling interval is defined to be $d(\widetilde{E}_z(t_n^*)) = d_1$ if $\widetilde{E}_z(t_n^*) \in (L_z, \infty)$ and $d(\widetilde{E}_z(t_n^*)) = d_2$ otherwise. Here, we choose $d_1 = 0.1$ and $d_2 = 1.9$, as suggested in Reynolds et al. (1990). The parameter L_z is chosen from the IC data such that $\mathrm{ATS}_0 = \mathrm{ARL}_0$ by the bootstrap procedure with the bootstrap sample size of $B = 10,000$. This chart is denoted as WSS-WC2.

Although the above four alternative charts are all generated from the proposed chart NEW, they can represent various existing charts. For instance, the chart WSS-WOC is the conventional self-starting EWMA chart that is applied to the decorrelated data, the chart WOSS-WOC is the conventional EWMA chart (cf., (5)) applied to the decorrelated data, and the chart WSS-WC2 is a self-starting variable-sampling chart with the conventional 2-interval sampling scheme that is applied to the decorrelated data. The chart WOSS-WC is also included in the comparison for confirming the benefit to have the self-starting feature in the proposed chart NEW.

Comparison of IC performance. We first compare the IC performance of the five control charts discussed above. In the five control charts, the weighting parameter λ is chosen to

be 0.05, 0.1 or 0.2. In the charts NEW and WOSS-WC, the two parameters α and κ are chosen to be 6 and 0, respectively, as recommended above based on the related simulation results. For all charts, we let $p = 1$, $m = 800$, $\Theta = 5$, and the nominal ATS_0 level be 200. The control limits of the charts are determined by bootstrap procedures with the bootstrap sample size of $B = 10,000$. The calculated ATS_0 values and their standard errors are presented in Table 2. From the table, it can be seen that 1) the three self-starting charts NEW, WSS-WOC and WSS-WC2 have satisfactory IC performance in all cases considered, since their calculated ATS_0 values are close to the nominal level of 200 in these cases, 2) the two non-self-starting charts WOSS-WC and WOSS-WOC perform worse than the three self-starting charts NEW, WSS-WOC and WSS-WC2, and 3) by comparing the results in Cases (I)-(III), it can be seen that the calculated ATS_0 values are generally farther away from the nominal level when the serial data correlation is larger. This example shows that the self-starting feature is important for a reliable online process monitoring.

Comparison of OC performance. We then study the OC performance of the related control charts in Cases (I)-(VI). In each case, let $p = 1$, $m = 800$, $\Theta = 5$, the nominal ATS_0 level be 200, and the other setups be the same as those in Table 2. To make the comparison fair, the control limits of the five charts, the parameter ϕ in the charts NEW and WOSS-WC, and the parameter L_z in the chart WSS-WC2 have been adjusted such that their actual ATS_0 values are all equal to the nominal ATS_0 level of 200. Also, the optimal ATS_1 values of the five charts are compared here to make the comparison fair, where the optimal ATS_1 values are obtained by changing the value of the weighting parameter λ such that the related ATS_1 values are minimized. The results for detecting shifts of Types (i)-(iii) are presented in Figure 6. From the figure, we can have the following conclusions. 1) The three charts NEW, WSS-WC2 and WOSS-WC that use the covariate information perform better than the two charts WSS-WOC and WOSS-WOC that do not use the covariate information for detecting shifts of Types (ii) and (iii). 2) For detecting all types of shifts, NEW performs better than WOSS-WC while WSS-WOC performs better than WOSS-WOC. 3) For detecting shifts of Type (i), WSS-WOC performs slightly better than NEW and WSS-WC2. 4) NEW has a better performance than WSS-WC2 for detecting shifts of Types (ii) and (iii). Conclusion 1)

Table 2: Calculated ATS_0 values and their standard errors (in parentheses) of the five control charts when $p = 1$, $m = 800$, $\Theta = 5$ and the nominal ATS_0 level is 200.

Case	λ	NEW	WSS-WOC	WOSS-WC	WOSS-WOC	WSS-WC2
(I)	0.05	196.3 (2.7)	197.2 (2.8)	184.1 (2.2)	181.3 (2.3)	202.1 (2.9)
	0.1	199.6 (2.7)	199.8 (2.9)	185.6 (2.2)	182.4 (2.4)	203.4 (2.9)
	0.2	200.1 (2.9)	199.6 (2.9)	188.2 (2.4)	185.2 (2.4)	201.5 (2.8)
(II)	0.05	194.3 (2.6)	194.2 (2.6)	179.3 (2.1)	176.1 (2.2)	193.4 (2.5)
	0.1	196.4 (2.6)	197.2 (2.8)	182.1 (2.2)	178.4 (2.2)	195.8 (2.4)
	0.2	198.2 (2.7)	198.9 (2.8)	182.2 (2.2)	180.2 (2.3)	196.7 (2.6)
(III)	0.05	193.2 (2.6)	192.9 (2.5)	171.4 (2.0)	168.3 (1.9)	192.1 (2.4)
	0.1	194.5 (2.6)	194.8 (2.5)	176.7 (2.1)	172.1 (1.9)	194.8 (2.3)
	0.2	196.4 (2.6)	197.2 (2.6)	178.5 (2.1)	174.5 (2.1)	195.9 (2.4)
(IV)	0.05	192.1 (2.5)	193.4 (2.5)	168.1 (2.0)	165.2 (1.9)	192.8 (2.4)
	0.1	193.5 (2.6)	194.6 (2.6)	171.7 (2.0)	173.1 (1.9)	194.9 (2.4)
	0.2	195.4 (2.6)	196.8 (2.5)	175.3 (2.1)	174.4 (2.0)	195.1 (2.5)
(V)	0.05	193.1 (2.6)	192.8 (2.5)	165.9 (2.0)	164.3 (1.9)	191.3 (2.3)
	0.1	195.7 (2.7)	196.4 (2.5)	171.4 (2.1)	168.7 (2.0)	194.8 (2.4)
	0.2	196.2 (2.7)	196.5 (2.6)	173.2 (2.1)	169.1 (2.0)	196.2 (2.4)
(VI)	0.05	195.4 (2.6)	196.2 (2.6)	180.3 (2.2)	179.2 (2.2)	205.4 (3.0)
	0.1	198.3 (2.8)	198.6 (2.7)	182.1 (2.3)	181.1 (2.3)	204.2 (3.0)
	0.2	201.2 (2.9)	199.4 (2.7)	184.2 (2.3)	181.4 (2.2)	200.9 (2.8)

confirms that the covariate information is helpful for detecting shifts that are associated with the covariates. Conclusion 2) confirms that the self-starting feature is helpful for effectively detecting shifts during online process monitoring. Conclusion 3) says that NEW will lose some effectiveness in cases when the shifts are unassociated with the covariates, but the loss is small. Conclusion 4) confirms that it is beneficial to use a continuous, instead of the conventional 2-level, sampling interval function in NEW, which is consistent with the results in Li and Qiu (2014). Additionally, we compare the optimal OC performance of the five charts in Case (II) when $p = 3$ or 5 and the other setups are the same as those in Figure 6. The related results are shown in Figure S.3 of the supplementary file. From that figure, we still have the conclusion that NEW performs the best among the five charts in most cases considered.

4 A Real Data Application

In this section, we demonstrate the application of the proposed chart NEW by using a real dataset obtained from a white wine production process. This dataset can be downloaded from the UCI Machine Learning Repository with the link https://archive.ics.uci.edu/ml/datasets/Wine+Quality. The dataset contains a total of 4,898 observations of the wine *vinho verde* collected from May 2004 to February 2007 in Portugal. Portugal is a top ten wine exporting country with 3.17% of the market share in 2015, and this specific wine accounts for about 15% of the total Portuguese production (Cortez et al. 2009). Because of its huge market share, it is important to establish an online monitoring system to monitor some quality variables of the wine. In this paper, we focus on monitoring the wine alcohol (Y). Since the wine alcohol and the wine density are usually associated with each other, the wine density (X) is used as a covariate to help monitor the wine alcohol.

The 4,898 observations of the wine alcohol and wine density are shown in Figure 7. From the plots, it seems that the first 2,000 observations of each variable are quite stable. So, this part of the data is used as an IC dataset and the remaining observations are used for online monitoring. For the IC data, the p-values of the Box-Pierce test for testing the serial data

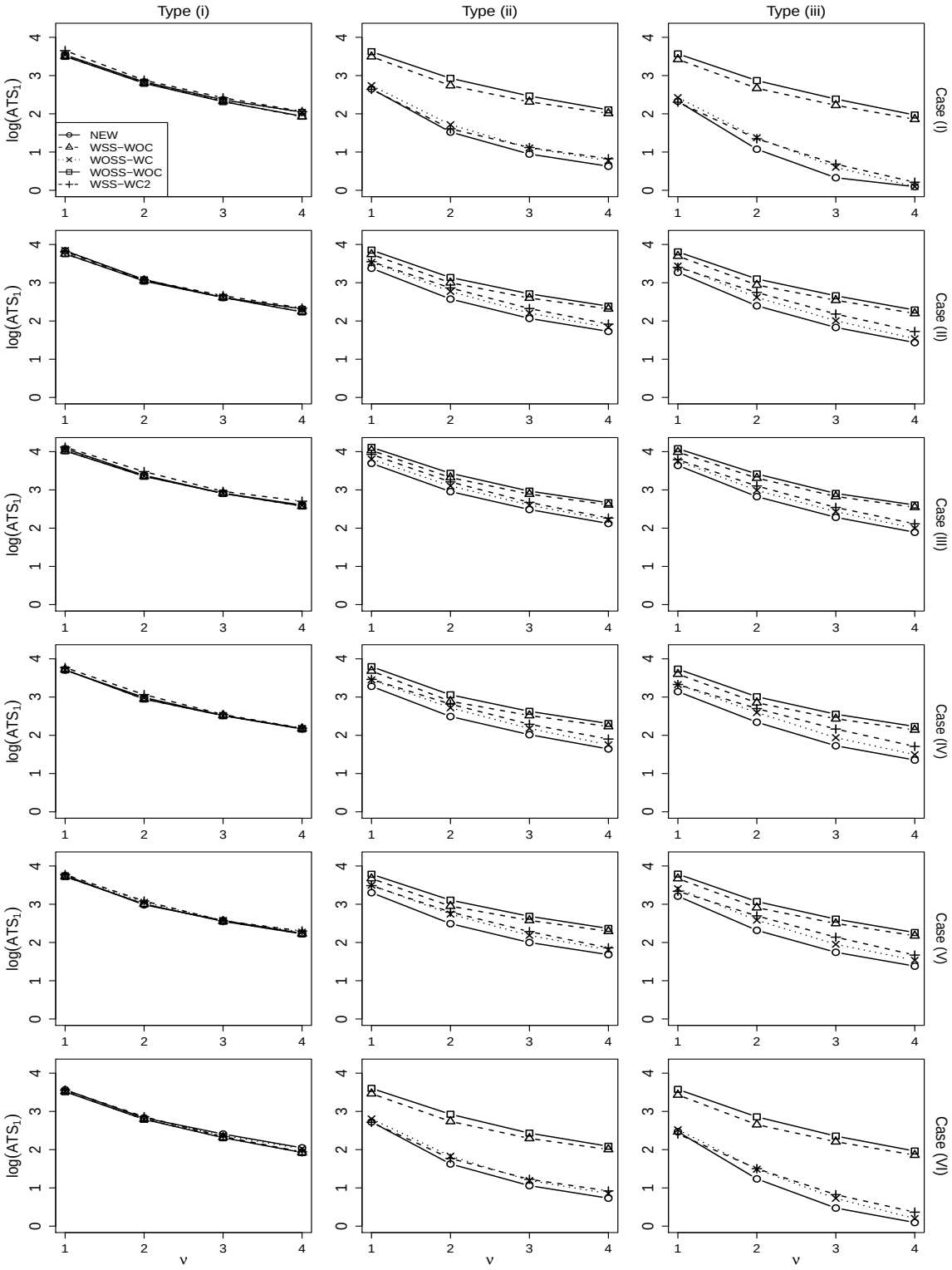

Figure 6: Calculated optimal ATS_1 values of the five charts for detecting the the shifts of Types (i)-(iii) when $p = 1$ and $\text{ATS}_0 = 200$. In each plot, the y-axis is in natural log scale.

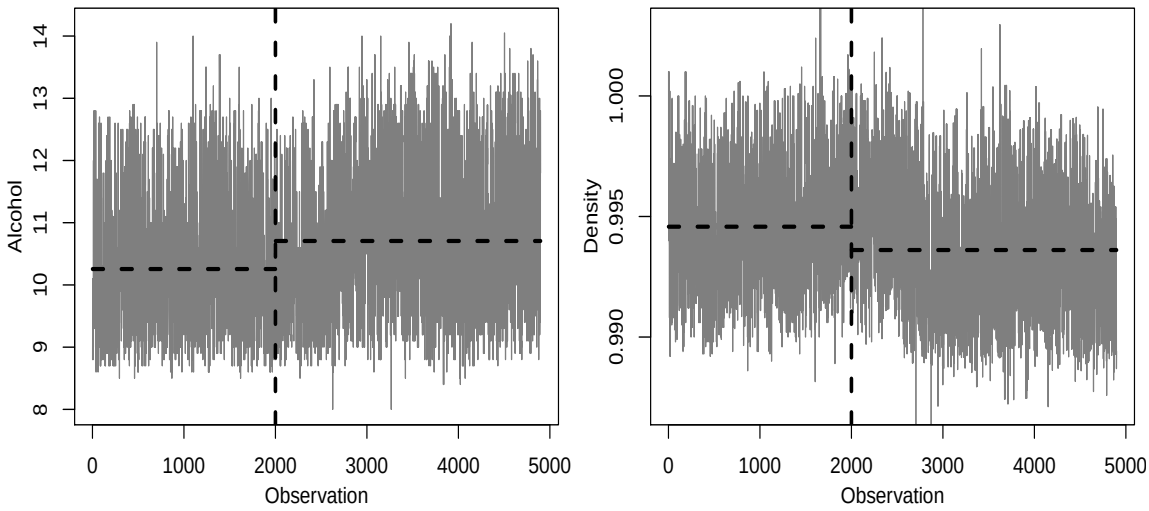

Figure 7: The 4,898 observations of the wine alcohol (left panel) and wine density (right panel). In each plot, the vertical dashed line separates the IC data and the data for process monitoring, and the two horizontal dashed lines denote the sample means of the two data subsets.

correlation in observations of Y and X are 2.2×10^{-16} and 2.2×10^{-16}, respectively, implying a statistically significant serial data correlation in the IC data. The Augmented Dickey-Fuller (ADF) test and the Phillips-Perron (PP) test for testing the stationarity of the serial data correlation in Y and X both give a p-value of 0.01, suggesting that it is reasonable to assume the serial data correlation in Y and X to be stationary. From the IC data, the initial estimates of the regression coefficients of model (1) when the regression function is assumed linear are $\widehat{\beta}_0^{\,(0)} = 313.82$, and $\widehat{\beta}^{(0)} = -305.22$. The negative value of $\widehat{\beta}^{(0)}$ indicates a negative association between the wine alcohol and the wine density. In this example, we are concerned about upward shifts in the wine alcohol. Thus, the upward version of the chart (5) will be considered here, as in all simulation examples discussed in Section 3.

Next, we discuss online process monitoring starting from the 2,001st observation. Without loss of generality, the beginning time of process monitoring is reset to be 0. The original observations in this data are collected at equally spaced time points. To illustrate the application of the proposed variable-sampling chart, it is assumed that the basic time unit ω is 0.1 and the computed sampling intervals are rounded to the nearest integer multiples of

ω, as in Cases (V) and (VI) of the simulation studies presented in Section 3. Thus, if the calculated sampling interval is 0.213 based on the data of X up to the 2,001st observation, then the actual sampling interval that we use would be 0.2. In the five charts, we fix $\Theta = 5$, $\text{ATS}_0 = 200$ and $\lambda = 0.05$. The parameters α and κ in the two charts NEW and WOSS-WC are chosen to be 6 and 0, respectively, as recommended in Section 3. The control limits of the five charts, the parameter ϕ in the charts NEW and WOSS-WC, and the parameter L_z in the chart WSS-WC2 are all determined from the IC data by the bootstrap procedures with the bootstrap sample size of 10,000. The charting statistics of the five charts are presented in Figure 8. From the figure, it can be seen that the first signals of the five charts NEW, WSS-WOC, WOSS-WC, WOSS-WOC and WSS-WC2 are at times 70.9, 97.0, 81.5, 97.0, 86.4, respectively. Therefore, the first signal of the proposed chart NEW is the earliest, and the first signals of the other two charts WOSS-WC and WSS-WC2 that use the covariate information are also earlier than those of WSS-WOC and WOSS-WOC that ignore the covariate information. This example confirms the benefit to use the covariate information in determining the sampling intervals for a variable-sampling control chart.

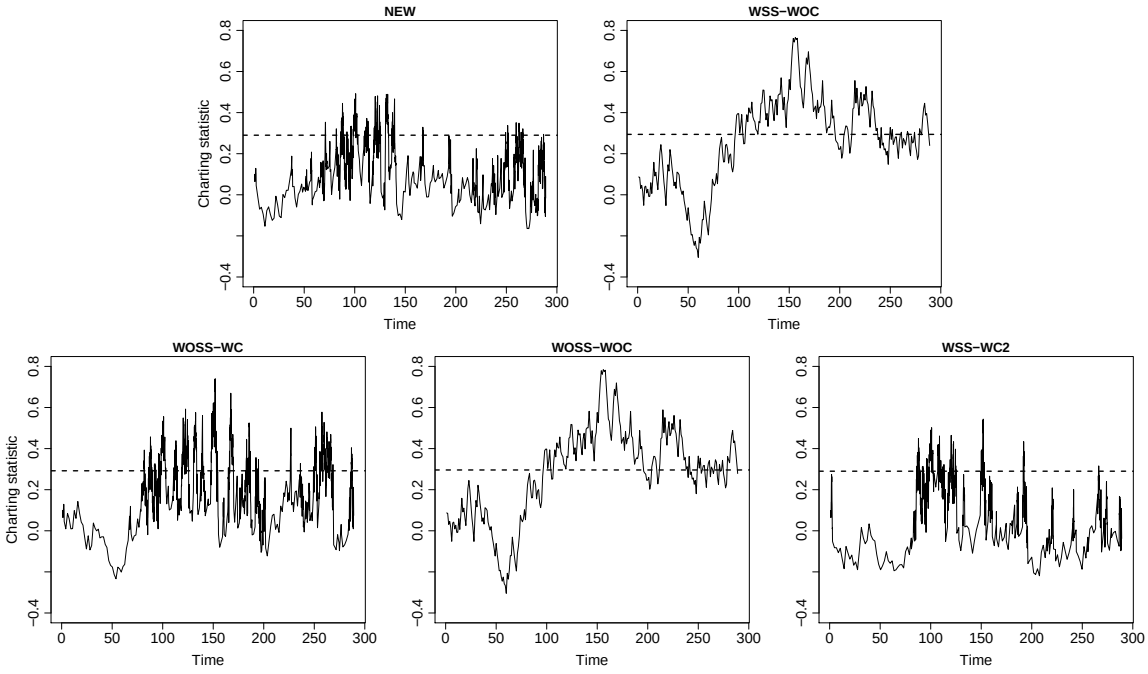

Figure 8: Charting statistics of the five charts for monitoring the wine alcohol data. The horizontal dashed line in each plot denotes the related control limit.

5 Concluding Remarks

In the previous sections, we have described a new variable-sampling charting scheme for monitoring quality variables using helpful information in covariates. One major feature of the new method is that it takes advantage of the covariate information when determining the sampling intervals. This method can accommodate stationary short-range serial data correlation, and is a self-starting monitoring procedure. Extensive simulation studies and a real data application have shown that it performs well in many different cases. It should be pointed out that there are still some issues about this method to address in the future research. For example, the current method is developed to detect upward mean shifts under the univariate EWMA framework. We believe that the idea presented in the paper is quite general and can be used in the univariate and/or multivariate CUSUM and CPD frameworks as well. Although the paper focuses solely on detecting mean shifts, we believe that the same idea could be used in charts for detecting variance shifts or other shifts. Also, it is assumed in the paper that the serial data correlation in process observations is stationary for the current method to work well. It is currently unknown to us whether this assumption can be weakened or lifted. In addition, the proposed method determines the sampling intervals based on the covariate information only. Actually, the information in the previous observations of the quality variable Y can also be helpful in determining the sampling intervals, as discussed in Qiu and Li (2014). It is still unknown to us whether it can further improve the performance of a variable-sampling control chart by using these two types of information in determining the sampling intervals. All these issues need to be studied carefully in the future research.

Acknowledgments: The authors thank the editors and three referees for their constructive comments and suggestions, which improved the quality of the paper greatly. This research is supported in part by the NSF grant of the number DMS-1914639.

Authors' Biographical Sketches

Kai Yang is currently a PhD student in the Department of Biostatistics at the University of Florida. His thesis research is mainly on spatio-temporal data modeling and monitoring under the supervision of Professor Peihua Qiu. He has published five papers on nonparametric estimation of the mean and variance/covariance structures of spatial data, and on spatio-temporal data monitoring. Inaddition to that topic, his thesis research also discusses effective process monitoring by using covariate information.

Peihua Qiu received his PhD in statistics from the Department ofStatistics at the University of Wisconsin - Madison in 1996. He worked as a senior research consulting statistician of the Biostatistics Center at the Ohio State University during 1996–1998. He then worked as an assistant professor (1998–2002), an associate professor (2002–2007), and a full professor (2007–2013) at the School of Statistics of the University of Minnesota. He is an elected fellow of the American Statistical Association, an elected fellow of the Institute of Mathematical Statistics, an elected member of the International Statistical Institute, a senior member of the American Society for Quality, and a lifetime member of the International Chinese Statistical Association. He served as an associate editor for *Journal of the American Statistical Association, Biometrics, Technometrics, Surgery*, and *Statistical Papers*, and guest co-editor for *Multimedia Tools and Applications*, and *Quality and Reliability Engineering International*. He was the editor-elect (2013) and editor (2014-2016) of *Technometrics*. He is currently an associate editor of *Quality Engineering*, and a Professor and the Founding Chair of the Department of Biostatistics at the University of Florida.

Peihua Qiu has made substantial contributions in the areas of jump regression analysis, image processing, statistical process control, survival analysis, and disease screening and surveillance. So far, he has published over 130 research papers in referred journals, many of which appeared in top journals, including *Technometrics, Journal of the American Statistical Association, Annals of Statistics, Annals of Applied Statistics, Journal of the Royal Statistical Society (Series B), Biometrika, Biometrics, IEEE Transactions on Pattern Analysis and Machine Intelligence*, and *IISE Transactions*. His research monograph titled *Image*

Processing and Jump Regression Analysis (2005, Wiley) won the inaugural Ziegel prize in 2007 for its contribution in bridging the gap between jump regression analysis in statistics and image processing in computer science. His second book titled *Introduction to Statistical Process Control* was published in 2014 by Chapman & Hall/CRC.

References

Apley D.W., and Tsung, F. (2002), "The autoregressive T^2 chart for monitoring univariate autocorrelated processes," *Journal of Quality Technology*, **34**, 80–96.

Asadzadeh S., Aghaie A., and Yang S. (2008), "Monitoring and diagnosing multistage processes: a review of cause selecting control charts," *Journal of Industrial and Systems Engineering*, **2**, 214–235.

Arnold, J.C., and Reynolds, M.R., Jr. (2001), "CUSUM control charts with variable sample sizes and sampling intervals," *Journal of Quality Technology*, **33**, 66–81.

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

Chatterjee, S., and Qiu, P. (2009), "Distribution-free cumulative sum control charts using bootstrap-based control limits," *The Annals of Applied Statistics*, **3**, 349–369.

Chen, N., Zi, X., and Zou, C. (2016), "A distribution-free multivariate control chart," *Technometrics*, **58**, 448–459.

Cook, R.D., and Weisberg, S. (1999), *Applied regression including computing and graphics*, New York: John Wiley & Sons.

Cortez, P., Cerdeira, A., Almeida, F., Matos, T., and Reis, J. (2009), "Modeling wine preferences by data mining from physicochemical properties," *Decision Support Systems*, **47**, 547–553.

Hastie, T., and Tibshirani, R. (1990), *Generalized additive models*, Boca Raton, FL: Chapman Hall/CRC.

Hastie, T., and Loader, C. (1993), "Local regression: automatic kernel carpentry," *Statistical Science*, **8**, 120–129.

Hall, P., Fisher, N.I., Hoffmann, B. (1994), "On the nonparametric estimation of covariance functions," *The Annals of Statistics*, **22**, 2115–2134.

Hawkins, D.M. (1987), "Self-starting CUSUM charts for location and scale," *Journal of the Royal Statistical Society, Series D (The Statistician)*, **36**, 299–316.

Hawkins, D.M., and Olwell, D.H. (1998), *Cumulative sum charts and charting for quality improvement*, New York: Springer-Verlag.

Hawkins, D., Qiu, P., and Kang, C.W. (2003), "The changepoint model for statistical process control," *Journal of Quality Technology*, **35**, 355–366.

John, J.A., and Draper, N.R. (1980), "An alternative family of transformations," *Journal of the Royal Statistical Society, Series C*, **29**, 190–197.

Lee, H. C., and Apley, D. W. (2011), "Improved design of robust exponentially weighted moving average control charts for autocorrelated processes," *Quality and Reliability Engineering International*, **27**, 337–352.

Li, Z., and Qiu, P. (2014), "Statistical process control using dynamic sampling," *Technometrics*, **56**, 325–335.

Li, J., and Qiu, P. (2016), "Nonparametric dynamic screening system for monitoring correlated longitudinal data," *IIE Transactions*, **48**, 772–786.

Montgomery, D.C. (2012), *Introduction to statistical quality control*, New York: John Wiley & Sons.

Page, E.S. (1954), "Continuous inspection schemes," *Biometrika*, **4**, 100–114.

Paynabar, K., Jin, J., and Yeh, A.B. (2012), "Phase I risk-adjusted control charts for monitoring surgical performance by considering categorical covariates," *Journal of Quality Technology*, **44**, 39–53.

Prabhu, S.S., Montgomery, D.C., and Runger, G.C. (1994), "A combined adaptive sample size and sampling interval X control scheme," *Journal of Quality Technology*, **26**, 164–176.

Qiu, P. (2005), *Image processing and jump regression analysis*, New York: John Wiley & Sons.

Qiu, P. (2014), *Introduction to statistical process control*, Boca Raton, FL: Chapman Hall/CRC.

Qiu, P., and Hawkins, D. (2001), "A rank based multivariate CUSUM procedure," *Technometrics*, **43**, 120–132.

Qiu, P., Li, W., and Li, J. (2020), "A new process control chart for monitoring short-range serially correlated data," *Technometrics*, **62**, 71–83.

Qiu, P., and Xiang, D. (2014), "Univariate dynamic screening system: an approach for identifying individuals with irregular longitudinal behavior," *Technometrics*, **56**, 248–260.

Qiu, P., and Xiang, D. (2015), "Surveillance of cardiovascular diseases using a multivariate dynamic screening system," *Statistics in Medicine*, **34**, 2204–2221.

Reynolds, M.R., Jr., Amin, R.W., and Arnold, J.C. (1990). "CUSUM charts with variable sampling intervals," *Technometrics*, **32**, 371–384.

Roberts, G. C. (1959), "Control chart tests based on geometric moving average," *Technometrics*, **1**, 239–250.

Runger, G.C., and Pignatiello, J.J., Jr. (1991), "Adaptive sampling for process control," *Journal of Quality Technology*, **23**, 135–155.

Shewhart, W.A. (1931), *The Economic control of the quality of manufactured production*, New York: Macmillan.

Steiner, S.H., and Jones, M. (2010), "Risk-adjusted survival time monitoring with an updating exponentially weighted moving average (EWMA) control chart," *Statistics in Medicine*, **29**, 444–454.

Sullivan, J.H., and Jones, L.A. (2002), "A self-starting control chart for multivariate individual observations," *Technometrics*, **44**, 24–33.

Tibshirani, R. (1996), "Regression Shrinkage and Selection via the LASSO," *Journal of the Royal Statistical Society, Series B*, **58**, 267–288.

Yang, K., and Qiu, P. (2020), "Adaptive process monitoring using covariate information," *Technometrics*, DOI: 10.1080/00401706.2020.1772115.

Zou, H., and Hastie, T. (2005), "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society, Series B*, **67**, 301–320.

Zou, C., Wang, Z., and Tsung, F. (2008), "Monitoring autocorrelated processes using variable sampling schemes at fixed-times," *Quality and Reliability Engineering International*, **24**, 55–69.