

Univariate Dynamic Screening System: An Approach For Identifying Individuals With Irregular Longitudinal Behavior

Peihua Qiu¹ and Dongdong Xiang²

¹School of Statistics, University of Minnesota

²School of Finance and Statistics, East China Normal University

Abstract

In our daily life, we often need to identify individuals whose longitudinal behavior is different from the behavior of those well-functioning individuals, so that some unpleasant consequences can be avoided. In many such applications, observations of a given individual are obtained sequentially, and it is desirable to have a screening system to give a signal of irregular behavior as soon as possible after that individual's longitudinal behavior starts to deviate from the regular behavior, so that some adjustments or interventions can be made in a timely manner. This paper proposes a dynamic screening system for that purpose in cases when the longitudinal behavior is univariate, using statistical process control and longitudinal data analysis techniques. Several different cases, including those with regularly spaced observation times, irregularly spaced observation times, and correlated observations, are discussed. Our proposed method is demonstrated using a real-data example about the SHARe Framingham Heart Study of the National Heart, Lung and Blood Institute. This paper has supplementary materials online.

Key Words: Correlation; Dynamic screening; Longitudinal data; Process monitoring; Process screening; Standardization; Statistical process control; Unequal sampling intervals.

1 Introduction

Most products (e.g., airplanes, cars, batteries of notebook computers) need to be checked regularly or occasionally about certain variables related to their quality and/or performance. If the observed values of the variables of a given product are significantly worse than the values of a typical well-functioning product of the same age, then some proper adjustments or interventions should be made to avoid any unpleasant consequences. This paper suggests a dynamic screening system for identifying such improperly functioning products effectively.

The motivating example of this research is the SHARe Framingham Heart Study of the National Heart, Lung and Blood Institute. In the study, 1055 patients were recruited, each patient was followed 7 times, and the total cholesterol level of each patient was recorded during each of the 7 clinic visits. Among the 1055 patients, 1028 of them did not have any strokes during the study, and the remaining 27 patients had at least one stroke. One major research goal of the study is to identify patients with irregular total cholesterol level patterns as early as possible, so that some medical treatments can be taken in a timely manner to avoid the occurrence of stroke. To this end, one traditional method is to construct pointwise confidence intervals for the mean total cholesterol level of the non-stroke patient population across different ages, using the longitudinal data of all non-stroke patients. Then, a patient can be identified as the one who has an irregular total cholesterol level pattern if his/her total cholesterol levels go beyond the confidence intervals. In the literature, there have been some discussions about the construction of such confidence intervals. See, for instance, Ma et al. (2012), Yao et al. (2005), and Zhao and Wu (2008).

The confidence interval methods mentioned above use a cross-sectional comparison approach. In the context of the total cholesterol level example, they compare patients' total cholesterol levels at the cross-section of a given age. One limitation of this approach is that it does not make use of the medical history of a given patient in the medical diagnostics. If the total cholesterol levels at different ages of a given patient are regarded as observations of a random process, then intuitively we should make use of all such observations of the process for medical diagnostics. If a patient's total cholesterol level is consistently above the mean total cholesterol level of the non-stroke patient population for an enough long time period, then that patient should still be identified as a person who has an irregular total cholesterol level pattern, even if his/her total cholesterol level at any given time point is within the confidence interval of the mean total cholesterol level of the non-stroke patient population. Further, it is critical that the diagnostic method should be dynamic in the sense that a decision about a patient can be made at any time during the observation process as long as all observations up to the current time point have provided enough evidence for the decision. That is because the major purpose to monitor a patient's total cholesterol level is to find any irregular pattern as soon as possible so that some necessary medical treatments can be given in a timely manner. The confidence interval methods mentioned in the previous paragraph usually do not have such a dynamic decision-making property, because they do not follow a patient sequentially over time. In the literature, some statistical process control (SPC) methodologies, including the

cumulative sum (CUSUM) and the exponentially weighted moving average (EWMA) charts, are designed for such purposes (e.g., Hawkins and Olwell 1998, Montgomery 2009, Qiu 2013). These control charts evaluate the performance of a process sequentially based on all observed data up to the current time point. However, they cannot be applied to the current problem directly for the following several reasons. First, most SPC methods are for monitoring a single process (e.g., a production line, a medical system). When the process performs satisfactorily, its observations follow a so-called in-control (IC) distribution and the process is said to be IC. The major purpose of the SPC control charts is to detect any distributional shift of the process observations as soon as possible so that the related process can be stopped promptly for adjustments and repairs. In the current problem, if each patient is regarded as a process, then there are many processes involved. To judge whether a patient's total cholesterol level is IC, we need to take the total cholesterol levels of all non-stroke patients into consideration. A conventional SPC control chart is not designed for that purpose. Second, in a typical SPC problem, if we are interested in monitoring the process mean, then the process mean is assumed to be a constant when the process is IC. In the current problem, however, the mean total cholesterol level of the non-stroke patients would change over time. Because of these fundamental differences between the conventional SPC problem and the current problem, the existing SPC methodologies cannot be applied directly to the current problem.

In this paper, we focus on cases when the variable to monitor (e.g., the total cholesterol level) is univariate and we are interested in monitoring its mean. In such cases, we propose a dynamic screening system (DySS) to dynamically identify individuals with irregular longitudinal patterns. In the DySS method, the regular longitudinal pattern is first estimated from the observed longitudinal data of a group of well-functioning individuals (e.g., the non-stroke patients in the stroke example). Then, the estimated regular longitudinal pattern is used for standardizing the observations of an individual to monitor. Finally, an SPC control chart is applied to the standardized observations of that individual for online monitoring of its longitudinal behavior.

It should be pointed out that the DySS problem described above looks like a profile monitoring problem in SPC. But, the two problems are actually completely different. In the profile monitoring problem, observations obtained from a sampled product are a profile that describes the relationship between a response variable and one or more explanatory variables, the profiles from different sampled products are usually ordered by observation times or spatial locations of the sampled products, and the major goal of the profile monitoring problem is for detecting any shift in the

mean and/or variance of the profiles over time/location (cf., e.g., Ding et al. 2006, Jensen and Birch 2009, Qiu et al. 2010). As a comparison, in the DySS problem, although observations of each individual look like a profile, each individual is actually treated as a separate process and we are mainly interested in monitoring the longitudinal behavior of each individual. Different individuals are often not ordered by time or location.

The remainder of the article is organized as follows. Our proposed DySS methodology is described in detail in Section 2. A simulation study for evaluating its numerical performance is presented in Section 3. Our proposed method is demonstrated using the motivating example about patients' total cholesterol levels in Section 4. Some concluding remarks are given in Section 5. Theoretical properties of the estimated IC mean function of the longitudinal response are discussed in a supplementary file, along with certain technical details and numerical results.

2 The DySS Method

Our proposed DySS method is described in four parts. In the first part, estimation of the regular longitudinal pattern from the observed longitudinal data of a group of well-functioning individuals is discussed. Then, an SPC chart is proposed in the second part for monitoring individuals dynamically, after their observations are standardized by the estimated regular longitudinal pattern. In this part, longitudinal observations are assumed to be independent and normally distributed. Situations when they are normally distributed and autocorrelated and the autocorrelation can be described by an AR(1) time series model are discussed in the third part. Finally, general cases when observations are autocorrelated with an arbitrary autocorrelation and when their distribution is non-normal are discussed in the fourth part.

2.1 Estimation of the regular longitudinal pattern

Assume that y is the variable that we are interested in monitoring, and observations of y for a group of m well-functioning individuals follow the model

$$y(t_{ij}) = \mu(t_{ij}) + \varepsilon(t_{ij}), \quad \text{for } i = 1, 2, \dots, m, j = 1, 2, \dots, J, \quad (1)$$

where t_{ij} is the j th observation time of the i th individual, $y(t_{ij})$ is the observed value of y at t_{ij} , $\mu(t_{ij})$ is the mean of $y(t_{ij})$, and $\varepsilon(t_{ij})$ is the error term. For simplicity, we assume that the

design space is $[0, 1]$. As in the literature of longitudinal data analysis (e.g., Li 2011), we further assume that the error term $\varepsilon(t)$, for any $t \in [0, 1]$, consists of two independent components, i.e., $\varepsilon(t) = \varepsilon_0(t) + \varepsilon_1(t)$, where $\varepsilon_0(\cdot)$ is a random process with mean 0 and covariance function $V_0(s, t)$, for any $s, t \in [0, 1]$, and $\varepsilon_1(\cdot)$ is a noise process satisfying the condition that $\varepsilon_1(s)$ and $\varepsilon_1(t)$ are independent for any $s, t \in [0, 1]$. In this decomposition, $\varepsilon_1(t)$ denotes the pure measurement error, and $\varepsilon_0(t)$ denotes all possible covariates that may affect y but are not included in model (1). In such cases, for any $s, t \in [0, 1]$, the covariance function of $\varepsilon(\cdot)$ is

$$V(s, t) = \text{Cov}(\varepsilon(s), \varepsilon(t)) = V_0(s, t) + \sigma_1^2(s)I(s = t), \quad \text{for any } s, t \in [0, 1], \quad (2)$$

where $\sigma_1^2(s) = \text{Var}(\varepsilon_1(s))$, and $I(s = t) = 1$ when $s = t$ and 0 otherwise. In model (1), observations of different individuals are assumed to be independent. As a sidenote, model (1) is similar to the nonparametric mixed-effects model discussed in Qiu et al. (2010). Their major difference is that the variance of the pure measurement error $\varepsilon_1(t)$ in (1) is allowed to vary over t , while such a variance is assumed to be a constant in Qiu et al. (2010). Also, J is assumed to be a constant in (1) for simplicity. Our method discussed in the paper can be generalized to cases when J depends on i without much difficulty.

For the longitudinal model specified by (1) and (2), Chen et al. (2005) and Pan et al. (2009) propose the following estimator of $\mu(t)$ based on the local p th-order polynomial kernel smoothing procedure:

$$\hat{\mu}(t; V) = \mathbf{e}_1' \left(\sum_{i=1}^m X_i' V_i^{-1} X_i \right)^{-1} \left(\sum_{i=1}^m X_i' V_i^{-1} \mathbf{y}_i \right), \quad (3)$$

where \mathbf{e}_1 is a $(p + 1)$ -dimensional vector with 1 at the first component and 0 anywhere else,

$$X_i = \begin{pmatrix} 1 & (t_{i1} - t) & \dots & (t_{i1} - t)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (t_{iJ} - t) & \dots & (t_{iJ} - t)^p \end{pmatrix}_{J \times (p+1)},$$

$V_i^{-1} = K_{ih}^{1/2}(t)(I_i \Sigma_i I_i)^{-1} K_{ih}^{1/2}(t)$, $K_{ih}(t) = \text{diag}\{K((t_{i1} - t)/h), \dots, K_h((t_{iJ} - t)/h)\}/h$, K is a density kernel function, $h > 0$ is a bandwidth, $I_i = \text{diag}\{I(|t_{i1} - t| \leq 1), \dots, I(|t_{iJ} - t| \leq 1)\}$, and Σ_i is the covariance matrix of $\mathbf{y}_i = (y(t_{i1}), y(t_{i2}), \dots, y(t_{iJ}))'$ which can be computed from $V(s, t)$ in (2). Throughout this paper, the inverse of a matrix refers to the Moore-Penrose generalized inverse, which always exists.

In practice, the error covariance function $V(s, t)$ is usually unknown, and it needs to be estimated from the observed data. To this end, we propose a method consisting of several steps,

described below. First, we compute an initial estimator of $\mu(t)$, denoted as $\tilde{\mu}(t)$, by (3) in which all Σ_i are replaced by the identity matrix. Obviously, $\tilde{\mu}(t)$ is just the local p th-order polynomial kernel estimator based on the assumption that the error terms in (1) at different time points are i.i.d. for each individual subject. Second, we define residuals

$$\tilde{\varepsilon}(t_{ij}) = y(t_{ij}) - \tilde{\mu}(t_{ij}), \quad \text{for } i = 1, 2, \dots, m, j = 1, 2, \dots, J.$$

Third, when $s \neq t$, we use the method originally proposed by Li (2011) to first estimate $V_0(s, t)$ in (2) by

$$\tilde{V}_0(s, t) = (A_1(s, t)V_{00}(s, t) - A_2(s, t)V_{10}(s, t) - A_3(s, t)V_{01}(s, t)) B^{-1}(s, t), \quad (4)$$

where $A_1(s, t) = S_{20}(s, t)S_{02}(s, t) - S_{11}^2(s, t)$, $A_2(s, t) = S_{10}(s, t)S_{02}(s, t) - S_{01}(s, t)S_{11}(s, t)$, $A_3(s, t) = S_{01}(s, t)S_{20}(s, t) - S_{10}(s, t)S_{11}(s, t)$, $B(s, t) = A_1(s, t)S_{00}(s, t) - A_2(s, t)S_{10}(s, t) - A_3(s, t)S_{01}(s, t)$,

$$S_{l_1 l_2}(s, t) = \frac{1}{mJ(J-1)} \sum_{i=1}^m \sum_{j=1}^J \sum_{j \neq j'}^J \left(\frac{t_{ij} - s}{h} \right)^{l_1} \left(\frac{t_{ij'} - t}{h} \right)^{l_2} K_h(t_{ij} - s) K_h(t_{ij'} - t),$$

$$V_{l_1 l_2}(s, t) = \frac{1}{mJ(J-1)} \sum_{i=1}^m \sum_{j=1}^J \sum_{j \neq j'}^J \tilde{\varepsilon}(t_{ij}) \tilde{\varepsilon}(t_{ij'}) \left(\frac{t_{ij} - s}{h} \right)^{l_1} \left(\frac{t_{ij'} - t}{h} \right)^{l_2} K_h(t_{ij} - s) K_h(t_{ij'} - t),$$

$K_h(t_{ij} - s) = K((t_{ij} - s)/h)/h$, and $l_1, l_2 = 0, 1, 2$. Note that the matrix $\tilde{V}_0(s, t)$ in (4) may not be semipositive definite. To address this issue, the adjustment procedure (15) in Li (2011) can effectively regularize it to be a well defined covariance matrix, which is described in the supplementary file. Then, the variance of $y(t)$, denoted as $\sigma_y^2(t) = V_0(t, t) + \sigma_1^2(t)$, can be regarded as the mean function of $\varepsilon^2(t)$, and it can be estimated from the new data $\{\tilde{\varepsilon}^2(t_{ij}), i = 1, 2, \dots, m, j = 1, 2, \dots, J\}$ by (3), in which \mathbf{y}_i are replaced by $(\tilde{\varepsilon}^2(t_{i1}), \dots, \tilde{\varepsilon}^2(t_{iJ}))'$ and Σ_i are replaced by the identity matrix. The resulting estimator is denoted as $\tilde{\sigma}_y^2(t)$. Finally, we define the estimator of $V(s, t)$ by

$$\tilde{V}(s, t) = \tilde{V}_0(s, t)I(s \neq t) + \tilde{\sigma}_y^2(t)I(s = t). \quad (5)$$

Consequently, the mean function $\mu(t)$ can be estimated by $\hat{\mu}(t; \tilde{V})$.

Note that, in the four steps described above for computing $\tilde{\mu}(t)$, $\tilde{V}_0(s, t)$, $\tilde{\sigma}_y^2(t)$, and $\hat{\mu}(t; \tilde{V})$, the bandwidth h could be chosen differently in each step. We did not distinguish them in the above description for simplicity. In all our numerical examples presented later, they are allowed to be different, and each of them is chosen separately and sequentially by the conventional cross-validation (CV) procedure (cf., the description in the supplementary file) as follows. First, the bandwidth for $\tilde{\mu}(t)$ is selected. Then, the bandwidths for $\tilde{V}_0(s, t)$ and $\tilde{\sigma}_y^2(t)$ are selected, respectively. Finally,

the bandwidth for $\hat{\mu}(t; \tilde{V})$ can be selected. Some theoretical properties of $\hat{\mu}(t; \tilde{V})$ are given in a proposition of the supplementary file, which shows that $\hat{\mu}(t; \tilde{V})$ is statistically consistent under some regularity conditions.

By the estimation procedure described above, we can compute the local polynomial kernel estimator $\hat{\mu}(t; \tilde{V})$ of the mean function $\mu(t)$ in cases when the error covariance function $V(s, t)$ is unknown. After $\mu(t)$ is estimated, the variance function $\sigma_y^2(t)$ can be estimated in the same way, except that the original observations $\{y(t_{ij})\}$ need to be replaced by

$$\hat{\varepsilon}^2(t_{ij}) = (y(t_{ij}) - \hat{\mu}(t_{ij}; \tilde{V}))^2, \quad \text{for } i = 1, 2, \dots, m, j = 1, 2, \dots, J.$$

The resulting estimator of $\sigma_y^2(t)$ is denoted as $\hat{\sigma}_y^2(t)$.

The models (1) and (2) are quite general and they include a variety of different situations as special cases. In certain applications, the variance structure of the observed data can be described by a simpler model. For instance, in the SPC literature, we often consider cases when the error terms at different time points are independent of each other or they follow a specific time series model. In the current setup, the first scenario corresponds to the case when $V_0(s, t) = 0$, for any $s, t \in [0, 1]$. In such cases, $\mu(t)$ and $\sigma_y^2(t)$ can be estimated in the same way as described above, except that we do not need to estimate $V_0(s, t)$.

2.2 SPC charts for dynamically identifying irregular individuals

In the previous part, we define the estimators $\hat{\mu}(t; \tilde{V})$ and $\hat{\sigma}_y^2(t)$ of the IC mean function $\mu(t)$ and the IC variance function $\sigma_y^2(t)$, based on the observed longitudinal data of m well-functioning individuals which are sometimes called *IC data* for simplicity. In this part, we propose SPC charts for sequentially monitoring the longitudinal behavior of a new individual, using the estimated regular pattern of the longitudinal response variable $y(t)$ that is described by $\hat{\mu}(t; \tilde{V})$ and $\hat{\sigma}_y^2(t)$.

Assume that the new individual's y values are observed at times t_1^*, t_2^*, \dots over the design interval $[0, 1]$. In cases when that individual's longitudinal behavior is IC, these y values should also follow the model (1). For convenience of discussion, we re-write that model as

$$y(t_j^*) = \mu(t_j^*) + \sigma_y(t_j^*)\epsilon(t_j^*), \quad \text{for } j = 1, 2, \dots \quad (6)$$

where $\sigma_y(t)\epsilon(t)$ here equals $\varepsilon(t)$ in model (1). Thus, $\epsilon(t)$ in model (6) has mean 0 and variance 1

at each t . For the y values of the new individual, let us define their standardized values by

$$\widehat{\epsilon}(t_j^*) = \frac{y(t_j^*) - \widehat{\mu}(t_j^*; \widetilde{V})}{\widehat{\sigma}_y(t_j^*)}, \quad \text{for } j = 1, 2, \dots \quad (7)$$

In the case when $\{y(t_j^*), j = 1, 2, \dots\}$ are independent of each other, and each of them has a normal distribution, then $\{\widehat{\epsilon}(t_j^*), j = 1, 2, \dots\}$ would be a sequence of asymptotically i.i.d. random variables with a common asymptotic distribution of $N(0, 1)$. If we further assume that $\{t_j^*, j = 1, 2, \dots\}$ are equally spaced, then $\{\widehat{\epsilon}(t_j^*), j = 1, 2, \dots\}$ can be monitored by a conventional control chart, such as a CUSUM or EWMA chart, as usual. For instance, to detect upward mean shifts in $\{\widehat{\epsilon}(t_j^*), j = 1, 2, \dots\}$, the charting statistic of the upward CUSUM chart can be defined by

$$C_j^+ = \max\left(0, C_{j-1}^+ + \widehat{\epsilon}(t_j^*) - k\right), \quad \text{for } j \geq 1, \quad (8)$$

where $C_0^+ = 0$, and $k > 0$ is the allowance constant. The chart gives a signal of an upward mean shift if

$$C_j^+ > \rho, \quad (9)$$

where $\rho > 0$ is a control limit. To evaluate the performance of the CUSUM chart (8)-(9), we can use the IC average run length (ARL), denoted as ARL_0 , which is the average number of time points from the beginning of process monitoring to the signal time under the condition that 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 under the condition that the process is OC. Usually, the allowance k is specified beforehand. It has been well demonstrated in the literature that large k values are good for detecting large shifts and small k values are good for detecting small shifts. If it is possible to know the shift size δ in the mean of $\widehat{\epsilon}$, then the optimal k is $\delta/2$ (cf., Moustakides 1986). The control limit ρ is chosen such that a pre-specified ARL_0 value is reached. Then, the chart performs better for detecting a given shift if its ARL_1 value is smaller. For some commonly used k and ARL_0 values, the corresponding ρ values can be found from Table 3.2 in Hawkins and Olwell (1998). They can also be computed easily by some software packages, such as the R-package `spc` (Knoth 2011).

In practice, however, the observation times $\{t_j^*, j = 1, 2, \dots\}$ may not be equally spaced, as in the total cholesterol level example discussed in Section 1. In such cases, ARL_0 and ARL_1 are obviously inappropriate for measuring the performance of the CUSUM chart (8)-(9), and we need to define new performance measures. To this end, let $\omega > 0$ be a basic time unit in a given application,

which is the largest time unit that all unequally spaced observation times are its integer multiples. Define

$$n_j^* = t_j^*/\omega, \quad \text{for } j = 0, 1, 2, \dots$$

where $n_0^* = t_0^* = 0$. Then, $t_j^* = n_j^*\omega$, for all j , and n_j^* is the j th observation time in the basic time unit. In cases when the new individual is IC and the CUSUM chart (8)-(9) gives a signal at the s th observation time, then n_s^* is a random variable measuring the time to a false signal. Its mean $E(n_s^*)$ measures the IC average time to the signal (ATS), denoted as ATS_0 . If the new individual has an upward mean shift starting at the τ th observation time and the CUSUM chart (8)-(9) gives a signal at the s th time point with $s \geq \tau$, then the mean of $n_s^* - n_\tau^*$ is called the OC ATS, denoted as ATS_1 . Then, to measure the performance of a control chart (e.g., the chart (8)-(9)), its ATS_0 value can be fixed at a certain level beforehand, and the chart performs better if its ATS_1 value is smaller when detecting a shift of a given size. It is obvious that the values of ATS_0 and ATS_1 are just constant multiples of the corresponding values of ARL_0 and ARL_1 in cases when the observation times are equally spaced. In such cases, the two sets of measures are equivalent. Also, when the observation times are unequally spaced, it is possible to use the IC average number of observations to signal (ANOS), denoted as $ANOS_0$, and the OC average number of observations to signal, denoted as $ANOS_1$, for measuring the performance of control charts. However, these quantities are also just constant multiples of ATS_0 and ATS_1 , in cases when the sampling distribution does not change over time. Furthermore, when the sampling distribution changes over time, $ANOS_0$ and $ANOS_1$ may not be able to measure the chart performance well. As a demonstration, let us consider two scenarios. In the first scenario, we collect observations at the following basic time units: 1, 3, 7, \dots , and a signal by a chart is given at the third observation time. In the second scenario, we collect observations at the following basic time units: 1, 30, 100, \dots , and a signal by another chart is also given at the third observation time. Assume that the processes are IC in both scenarios. Then, the IC performance of the two charts are the same if $ANOS_0$ is used as a performance measure. But, the second chart does not give a false signal until the 100th basic time unit, while the first chart gives a false signal at the 7th basic time unit. Obviously, $ANOS_0$ cannot tell this quite dramatic difference. For all these reasons, only ATS_0 and ATS_1 are used in the remaining part of the paper for measuring the performance of a control chart.

If the IC mean function $\mu(t)$ and the IC variance function $\sigma_y^2(t)$ are known, then the standard-

ized observations of the new individual can be defined by

$$\epsilon(t_j^*) = \frac{y(t_j^*) - \mu(t_j^*)}{\sigma_y(t_j^*)}, \quad \text{for } j = 1, 2, \dots \quad (10)$$

In such cases, as long as the distribution of the observation times $\{t_j^*, j = 1, 2, \dots\}$ is specified properly, for a given k value in (8) and a given ATS_0 value, we can easily compute the corresponding value of the control limit ρ by simulation such that the pre-specified ATS_0 value is reached. For instance, when the distribution of the observation times is specified by the sampling rate d , which is defined to be the number of observation times every 10 basic time units here, the computed ρ values in cases when $ATS_0 = 20, 50, 100, 150, 200$, $k = 0.1, 0.2, 0.5, 0.75, 1.0$, and $d = 2, 5, 10$ are presented in Table 1. From the table, it can be seen that ρ increases with ATS_0 and d , and decreases with k . In cases when $\mu(t)$ and $\sigma_y^2(t)$ are unknown, they need to be estimated from an IC dataset by $\hat{\mu}(t; \tilde{V})$ and $\hat{\sigma}_y^2(t)$, as discussed in Section 2.1. In such cases, the standardized observations of the new individual defined by (7) are only asymptotically $N(0, 1)$ distributed when the new individual is IC. When the IC sample size is moderate to large, we can still use the control limit ρ computed in the case of (10), which will be justified by some numerical results described in the next section.

Table 1: Control limits of the CUSUM chart (8)-(9) for known IC mean and variance functions and various combinations of ATS_0 , k , and d described in the text. For the entry with a “*” symbol, the actual ATS_0 value is more than 1% away from the assumed ATS_0 value. For all other entries, the actual ATS_0 values are within 1% of the assumed ATS_0 values.

d	k	$ATS_0 = 20$	$ATS_0 = 50$	$ATS_0 = 100$	$ATS_0 = 150$	$ATS_0 = 200$
2	0.1	0.929	1.844	2.820	3.552	4.171
	0.2	0.774	1.571	2.351	2.928	3.396
	0.5	0.382	0.986	1.493	1.822	2.088
	0.75	0.118	0.654	1.064	1.312	1.507
	1.0	0.001*	0.355	0.718	0.936	1.101
5	0.1	1.844	3.215	4.612	5.631	6.445
	0.2	1.571	2.664	3.713	4.381	5.005
	0.5	0.986	1.685	2.234	2.586	2.898
	0.75	0.654	1.200	1.612	1.869	2.072
	1.0	0.355	0.837	1.189	1.404	1.562
10	0.1	2.820	4.612	6.407	7.649	8.612
	0.2	2.351	3.713	4.937	5.744	6.408
	0.5	1.493	2.235	2.852	3.245	3.542
	0.75	1.061	1.610	2.039	2.310	2.509
	1.0	0.718	1.192	1.538	1.738	1.894

In the SPC literature, the concept of ATS has been proposed for measuring the performance of

a control chart with variable sampling intervals (VSIs). See, for instance, Reynolds et al. (1990). However, the VSI problem in SPC is completely different from the DySS problem discussed here. In the VSI problem, the next observation time is determined by the current and all past observations of the process; the next observation is collected sooner if there is more evidence of a shift based on all available process observations, and later otherwise. Therefore, the VSI scheme is an integrated part of process monitoring. As a comparison, in the DySS problem, observation times are often not specifically chosen for improving process monitoring. In the total cholesterol level example discussed in Section 1, for instance, patients often choose their clinic visit times based on their convenience or health conditions.

At the end of this subsection, we would like to make several remarks about the DySS problem and the CUSUM chart (8)-(9). First, the above description focuses on cases to detect upward shifts in the process mean function $\mu(t)$ only. Control charts for detecting downward mean shifts or arbitrary mean shifts can be developed in the same way, except that the upward charting statistic in (8) and the decision rule in (9) should be changed to the downward or two-sided version (cf., Qiu 2013, Chapter 4). Second, control charts for detecting shifts in the process variance function $\sigma_y^2(t)$, or shifts in both $\mu(t)$ and $\sigma_y^2(t)$, can be developed in a similar way, after (8) and (9) are replaced by appropriate charting statistics and the related decision rules. See, for instance, Gan (1995) and Yeh et al. (2004) for discussions about the corresponding problems in the conventional SPC setup. Third, although a CUSUM chart is discussed above, EWMA charts and charts based on change-point detection (CPD) can also be used in the DySS problem. Generally speaking, the CUSUM and EWMA charts have similar performance, and the CPD charts can provide estimates of the shift position at the time when they give signals although their computation is usually more extensive. See Qiu (2013) for a related discussion. Fourth, to use the proposed DySS method, the observation times $\{t_j^*\}$ of the new individual to monitor usually cannot be larger than the largest value of the observation times $\{t_{ij}\}$ in the IC data, to avoid extrapolation in the data standardization in (7). Or, it is inappropriate to use our DySS method to monitor the new individual outside the time range

$$[\min_{i,j} t_{ij}, \max_{i,j} t_{ij}].$$

From the description about the DySS method above, it can be seen that this methodology actually combines the cross-sectional comparison between the new individual to monitor and the m well-functioning individuals in the IC data with the sequential monitoring of the new individual in

a dynamic manner. Therefore, to use the method to monitor the longitudinal behavior of the new individual at time t , there should exist observations around t in the IC data to make the cross-sectional comparison possible. Fifth, although the simplest CUSUM chart (8)-(9) is used in this paper to demonstrate the DySS method, some existing research in the SPC literature to accommodate correlated data (e.g., Apley and Tsung 2002, Jiang 2004) or to detect dynamic mean shifts (e.g., Han and Tsung 2006, Shu and Jiang 2006) might also be considered here. In the numerical study presented in Section 3, the control charts by Han and Tsung (2006) and Shu and Jiang (2006) will be discussed.

2.3 Cases when the longitudinal observations follow AR(1) model

In the above discussion, we assume that the longitudinal observations $\{y(t_j^*), j = 1, 2, \dots\}$ are independent. In practice, they are often correlated. In the SPC literature, time series models, especially the AR(1) model, are commonly used for describing the autocorrelation in such cases (cf., Jiang et al. 2000, Lu and Reynolds 2001). In this section we discuss how to use the proposed DySS procedure in cases when the longitudinal observations follow an AR(1) model. In a conventional time series model, observation times are usually equally spaced. In the current DySS problem, however, they can be unequally spaced. In such cases, time series modeling is especially challenging (cf., Maller et al. 2008, Vityazev 1996), and our DySS approach with a more general time series model is not available yet, which is left for our future research.

Let us first discuss estimation of the regular longitudinal pattern from an IC dataset when the longitudinal data are correlated and follow an AR(1) model. Assume that there are m well-functioning individuals whose longitudinal observations follow the model (1) in which $\epsilon(t_{ij}) = \sigma_y(t_{ij})\epsilon(t_{ij})$ and $\epsilon(t_{ij})$ has mean 0 and variance 1 for each $t_{ij} \in [0, 1]$. Instead of assuming independence, in this subsection, we assume that $\epsilon(t_{ij})$ follow the AR(1) model

$$\epsilon(t_{ij}) = \phi\epsilon(t_{ij} - \omega) + e(t_{ij}), \quad \text{for } i = 1, \dots, m, j = 1, \dots, J, \quad (11)$$

where ϕ is a coefficient, ω is the basic time unit defined in Section 2.2, and $e(t)$ is a white noise process over $[0, 1]$. The model (11) is the conventional AR(1) model for cases with equally spaced observation times. It does not have a constant term on the right-hand-side because the mean of $\epsilon(t_{ij})$ is 0. In the current DySS problem, it is inconvenient to work with the model (11) directly because there may not be observations at the times $t_{ij} - \omega$ that are used in (11). To overcome

this difficulty, we can transform the AR(1) model (11) to the following time series model using the actual observation times $\{t_{ij}\}$:

$$\epsilon(t_{ij}) = \phi^{\Delta_{i,j-1}} \epsilon(t_{i,j-1}) + \tilde{\Theta}_{ij}(z) e(t_{ij}), \quad (12)$$

where $\Delta_{i,j-1} = (t_{ij} - t_{i,j-1})/\omega$, z is the lag operator used in time series modeling (e.g., $z\epsilon(t) = \epsilon(t - \omega)$), and

$$\tilde{\Theta}_{ij}(z) = 1 + \phi z + \dots + \phi^{\Delta_{i,j-1}-1} z^{\Delta_{i,j-1}-1}$$

is a lag polynomial. As in Sections 2.1 and 2.2, let $\hat{\mu}(t; \tilde{V})$ and $\hat{\sigma}_y^2(t)$ be the local p th-order polynomial kernel estimators of $\mu(t)$ and $\sigma_y^2(t)$, and

$$\hat{\epsilon}(t_{ij}) = \frac{y(t_{ij}) - \hat{\mu}(t_{ij}; \tilde{V})}{\hat{\sigma}_y(t_{ij})}, \quad \text{for } i = 1, 2, \dots, m, j = 1, 2, \dots, J.$$

Then, the least squares (LS) estimator of ϕ , denoted as $\hat{\phi}$, can be obtained by minimizing

$$\sum_{i=1}^m \sum_{j=2}^J \left\{ \hat{\epsilon}(t_{ij}) - \phi^{\Delta_{i,j-1}} \hat{\epsilon}(t_{i,j-1}) \right\}^2.$$

Next, we use the estimated IC longitudinal pattern described by $\hat{\mu}(t; \tilde{V})$, $\hat{\sigma}_y^2(t)$ and $\hat{\phi}$ to monitor a new individual's longitudinal pattern. Assume that the new individual's y observations follow the model (6), in which the error term $\epsilon(t_j^*)$ follows the AR(1) model

$$\epsilon(t_j^*) = \phi \epsilon(t_{j-1}^* - \omega) + e_j^*, \quad \text{for } j = 1, 2, \dots$$

where ϕ is the same coefficient as the one in (11). Then, similar to the relationship between models (11) and (12), the above AR(1) model implies that

$$\epsilon(t_j^*) = \phi^{\Delta_{j-1}^*} \epsilon(t_{j-1}^*) + e_j^*, \quad (13)$$

where $\Delta_{j-1}^* = (t_j^* - t_{j-1}^*)/\omega$, $e_j^* = \tilde{\Theta}_j^*(z) e(t_j^*)$, and

$$\tilde{\Theta}_j^*(z) = 1 + \phi z + \dots + \phi^{\Delta_{j-1}^*-1} z^{\Delta_{j-1}^*-1}.$$

It can be checked that, when the new individual is IC, $\{e_j^*, j = 1, 2, \dots\}$ are i.i.d. with mean 0 and variance $\sigma_{e^*}^2 = 1 - \phi^{2\Delta_{j-1}^*}$. In the SPC literature, it has been well demonstrated that, when $\epsilon(t_j^*)$ follows the time series model (13), to detect an upward mean shift in $\epsilon(t_j^*)$, we can just monitor $\{e_j^*, j = 1, 2, \dots\}$ (cf., e.g., Lu and Reynolds 2001). Based on all these results, to detect an upward

mean shift in the original response y , the charting statistic of the upward CUSUM chart can be defined by

$$C_j^+ = \max \left[0, C_{j-1}^+ + \left(\widehat{\epsilon}(t_j^*) - \widehat{\phi}^{\Delta_{j-1}^*} \widehat{\epsilon}(t_{j-1}^*) \right) / \sqrt{1 - \widehat{\phi}^{2\Delta_{j-1}^*}} - k \right], \quad \text{for } j \geq 2, \quad (14)$$

where $C_1^+ = 0$, $k > 0$ is an allowance constant, and $\widehat{\epsilon}(t_j^*) = (y(t_j^*) - \widehat{\mu}(t_j^*; \widetilde{V})) / \widehat{\sigma}_y(t_j^*)$. The chart gives a signal when

$$C_j^+ > \rho, \quad (15)$$

where $\rho > 0$ is a control limit chosen to achieve a given ATS_0 value.

At the end of this subsection, we would like to point out that there actually are two different ways to estimate $\mu(t)$, $\sigma_y^2(t)$, and ϕ from an IC data when the AR(1) model (12) is involved. Because the estimation procedure described in Section 2.1 does not require the error covariance function $V(s, t)$ to be known, the first way is to estimate $\mu(t)$ and $\sigma_y^2(t)$ by the four steps described in Section 2.1 without using any prior information about the error structure, and then estimate ϕ by the LS procedure described above. Alternatively, in the case when we are sure that the AR(1) model (12) is appropriate for describing the error structure, a parametric formula of $V(s, t)$ can be derived in terms of ϕ , denoted as $V_\phi(s, t)$. Then, $\mu(t)$ and $\sigma_y^2(t)$ can be estimated by $\widehat{\mu}(t; V_\phi)$ and $\widehat{\sigma}_{y,\phi}^2(t)$, respectively, where $\widehat{\sigma}_{y,\phi}^2(t)$ is the same as $\widehat{\sigma}_y^2(t)$ except that $\widehat{\mu}(t; V_\phi)$ instead of $\widehat{\mu}(t; \widetilde{V})$ is used in its construction. Then, ϕ can be estimated by the LS estimator defined by the solution of

$$\min_{\phi} \sum_{i=1}^m \sum_{j=2}^J \left\{ \widehat{\epsilon}_\phi(t_{ij}) - \phi^{\Delta_{i,j-1}} \widehat{\epsilon}_\phi(t_{i,j-1}) \right\}^2,$$

where

$$\widehat{\epsilon}_\phi(t_{ij}) = \frac{y(t_{ij}) - \widehat{\mu}(t_{ij}; V_\phi)}{\widehat{\sigma}_{y,\phi}(t_{ij})}, \quad \text{for } i = 1, 2, \dots, m, j = 1, 2, \dots, J.$$

We have checked that results by these two approaches are close to each other as long as the IC sample size is not small. In all simulation examples discussed in Section 3, the second approach is used, while the first approach is used in the real data example discussed in Section 4 because we are not sure whether the AR(1) model is appropriate at the beginning of data analysis.

2.4 Cases with arbitrary autocorrelation

In the previous two parts, we assume that the original observations $\{y(t_j^*), j = 1, 2, \dots\}$ of a new individual are either independent or correlated with the autocorrelation described by the

AR(1) model (11). Also, the observation distribution is assumed to be normal. In practice, all these assumptions could be violated. If one or more such model assumptions are violated, then the related control charts (8)-(9) and (14)-(15) may not be reliable because their actual ATS_0 values could be substantially different from the assumed ATS_0 values. See related discussions in Hawkins and Deng (2010), Qiu and Hawkins (2001), Qiu and Li (2011a,b), and Zou and Tsung (2010). In this subsection, we propose a numerical approach to compute the control limit ρ of the chart (8)-(9) from an IC data such that the chart is still appropriate to use in cases when within-subject observations are correlated with an arbitrary autocorrelation and when a parametric form of their distribution is unavailable.

Assume that there is an IC dataset consisting of a group of m well-functioning individuals whose observations of y follow the model (1). The data of the first m_1 individuals are used for obtaining estimators $\hat{\mu}(t; \tilde{V})$ and $\hat{\sigma}_y^2(t)$, as discussed in Section 2.1. Then, the control limit ρ of the chart (8)-(9) can be computed from the remaining $m_2 = m - m_1$ individuals using a block bootstrap procedure (e.g., Lahiri 2003) described below. First, we compute the standardized observations of the m_2 well-functioning individuals by

$$\hat{\epsilon}(t_{ij}) = \frac{y(t_{ij}) - \hat{\mu}(t_{ij}; \tilde{V})}{\hat{\sigma}_y(t_{ij})}, \quad \text{for } i = m_1 + 1, m_1 + 2, \dots, m, j = 1, 2, \dots, J.$$

Second, we randomly select B individuals with replacement from the m_2 well-functioning individuals, and use their standardized observations to compute the value of ρ by a numerical searching algorithm so that a given ATS_0 level is reached. Such a searching algorithm has been well described in Qiu (2008) and Qiu (2013, Section 4.2). Basically, we first give an initial value to ρ , and compute the actual ATS_0 value of the chart (8)-(9) from the standardized observations of the B resampled individuals. If the actual ATS_0 value is smaller than the given ATS_0 value, then the value of ρ is increased; otherwise, the value of ρ is decreased. Then, the actual ATS_0 value of the chart is computed again, using the updated value of ρ . This process is repeated until the given ATS_0 level is reached to a certain accuracy. In all examples of this paper, we choose $m_1 = m/2$.

It should be pointed out that, if it is reasonable to assume that the within-subject observations are independent, then we can also draw bootstrap samples in the conventional way from the set of residuals $\{\hat{\epsilon}(t_{ij}), i = m_1 + 1, m_1 + 2, \dots, m, j = 1, 2, \dots, J\}$. Namely, we can randomly select J residuals with replacement from this set as the standardized observations of a bootstrap resampled “individual”, and B bootstrap resampled individuals can be generated in that way for computing the value of ρ . When the within-subject observations are correlated, it has been demonstrated in

the literature that a block bootstrap procedure, such as the one described above, would be more appropriate to use (cf., Lahiri 2003).

3 Numerical Study

In this section, we present some simulation results to investigate the numerical performance of the proposed DySS procedure described in the previous section. In estimating the mean function $\mu(t)$ and the variance function $\sigma_y^2(t)$ of the longitudinal response $y(t)$ from an IC data by the local smoothing method described in Section 2.1, p is fixed at 1 (i.e., the local linear kernel smoothing is used), the kernel function is chosen to be the Epanechnikov kernel function $K(t) = 0.75(1-t^2)I(|t| \leq 1)$ that is widely used in the curve estimation literature, and all bandwidths are chosen by the CV procedure.

In the first example, it is assumed that the IC mean function $\mu(t)$ and the IC variance function $\sigma_y^2(t)$ are known to be

$$\mu(t) = 1 + 0.3t^{\frac{1}{2}}, \quad \sigma_y^2(t) = \mu^2(t), \quad \text{for } t \in [0, 1],$$

observations of a given individual at different time points are independent of each other, and the sampling rate of different individuals are the same to be d observations every 10 basic time units. In such cases, the standardized observations are defined in (10), and the control limit ρ of the CUSUM chart (8)-(9) can be easily computed by a numerical algorithm, after its allowance constant k and the ATS_0 value are pre-specified (cf., Table 1 and the related description in Subsection 2.2). Now, let us consider the mean shifts from $\mu(t)$ to

$$\mu_1(t) = \mu(t) + \delta, \quad \text{for } t \in [0, 1],$$

that occur at the initial time point, where δ is the shift size taking values from 0 to 2.0 with a step of 0.1. The basic time unit in this example is $w = 0.001$, and the sampling rate d is chosen to be 2, 5, or 10. In the CUSUM chart (8)-(9), its ATS_0 value is fixed at 100, and the allowance constant k is chosen to be 0.1, 0.2 or 0.5. The corresponding ρ values are computed to be 2.820, 2.351, and 1.493 when $d = 2$; 4.612, 3.713, and 2.234 when $d = 5$; and 6.407, 4.937, and 2.852 when $d = 10$. The ATS_1 values of the chart in cases when $d = 2, 5, 10$, computed based on 10,000 replicated simulations, are shown in Figure 1(a)-(c). From the plots, it can be seen that: (i) the chart with a smaller (larger) k performs better for detecting small (large) shifts, and (ii) the ATS_1 values tend

to be smaller when d is larger. The first result is generally true for CUSUM charts (cf., Hawkins and Olwell 1998), and the second result is intuitively reasonable because more observations are available for process monitoring when d is chosen larger.

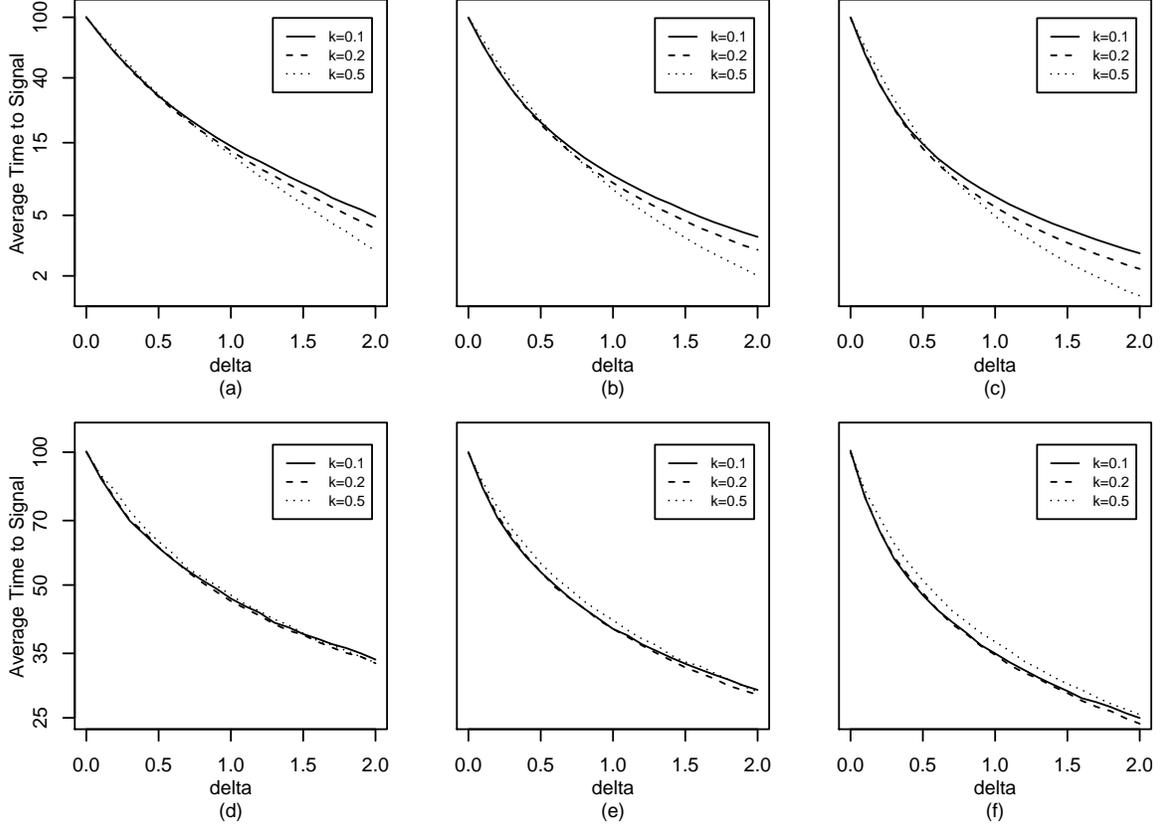


Figure 1: (a)-(c) ATS_1 values of the chart (8)-(9) in cases when the IC process mean function $\mu(t)$ and the process variance function $\sigma_y^2(t)$ are assumed known, $k = 0.1, 0.2, 0.5$, $ATS_0=100$, step shift size δ changes its value from 0 to 2 with a step of 0.1, and the sampling rate $d = 2$ (plot (a)), $d = 5$ (plot (b)), and $d = 10$ (plot (c)). (d)-(f) Corresponding results when the mean shift is a nonlinear drift from $\mu(t)$ to $\mu_1(t) = \mu(t) + \delta(1 - \exp(-10t))$, for $t \in [0, 1]$. The y -axes of all plots are in natural log scale.

In practice, a more realistic scenario would be that, once a shift occurs, the shift size would change over time. In the total cholesterol level example discussed in Section 1, if a patient's total cholesterol level begins to deviate from the regular pattern and no medical treatments are given in a timely fashion, then his/her total cholesterol level could deviate from the regular pattern more and more over time. To simulate such scenarios, we consider a nonlinear drift from $\mu(t)$ to

$$\mu_1(t) = \mu(t) + \delta(1 - \exp(-10t)), \quad \text{for } t \in [0, 1]$$

that occurs at the initial time point, where δ is a constant. When δ changes from 0 to 2.0 with a

step of 0.1 and other settings remain the same as those in Figure 1(a)-(c), the ATS_1 values of the chart (8)-(9) are shown in Figure 1(d)-(f). From the plots, similar conclusions to those from Figure 1(a)-(c) can be made here. We also performed simulations in cases when $w = 0.005$ and $ATS_0 = 20$. The results are included in a supplementary file, and they demonstrate a similar pattern to that in Figure 1.

Next, we consider a more realistic situation when the IC mean function $\mu(t)$ and the IC variance function $\sigma_y^2(t)$ are both unknown and they need to be estimated from an IC dataset. Assume that the IC dataset includes observations of m individuals. For each individual, the sampling rate is d , which is assumed to be the same as that of the new individual for online monitoring, and the basic time unit is $\omega = 0.001$. Then, when $d = 2$, each individual in the IC dataset has 200 observations in the design interval $[0, 1]$. After $\mu(t)$ and $\sigma_y^2(t)$ are estimated from the IC dataset, observations of the new individual are standardized by (7) for online monitoring. The standardized observations $\{\widehat{\epsilon}(t_j^*)\}$ are regarded as i.i.d. random variables with the common distribution $N(0, 1)$ when the new individual is IC. When ATS_0 , k and d are given, the control limit ρ can be computed as if $\mu(t)$ and $\sigma_y^2(t)$ are known (cf., Table 1). Then, the actual ATS_0 value of the chart (8)-(9) is computed using 10,000 simulations of the online monitoring. This entire process, starting from the generalization of the IC data to the computation of the actual ATS_0 value, is then repeated 100 times. The averaged actual ATS_0 values in cases when the nominal $ATS_0 = 100$, $k = 0.1, 0.2$ or 0.5 , $d = 2, 5$ or 10 , and $m = 5, 10$ or 20 are presented in Figure 2, along with the ATS_1 values of the chart for detecting step shifts of size $\delta = 0.25, 0.5, 0.75$, and 1.0 , occurring at the initial time point. These ATS_0 and ATS_1 values are also presented in Table S.1 of the supplementary file. From the figure and the table, it can be seen that: (i) the actual ATS_0 values are within 5% of the nominal ATS_0 value of 100 in all cases when $m \geq 10$, (ii) the actual ATS_0 values are closer to the nominal ATS_0 value when d is larger, (iii) the ATS_1 values are smaller if δ is larger or d is larger, (iv) the ATS_1 values do not depend on the m value much, especially when δ and d are large, and (v) the ATS_1 values are the smallest when $k = 0.1$ and δ is small (i.e., $\delta = 0.25$), or when $k = 0.2$ and δ is medium (e.g., $\delta = 0.5$ or 0.75), or when $k = 0.5$ and δ is large (i.e., $\delta = 1.0$). The corresponding results for detecting the drifts considered in Figure 1(d)-(f) are presented in Figure S.2 and Table S.2 of the supplementary file. Results in cases when $ATS_0 = 20$ are presented in Tables S.3 and S.4 of the supplementary file. Similar conclusions can be made from all these results.

In the above examples, the sampling rate d is chosen as 2, 5, and 10. In cases when $\omega = 0.001$,

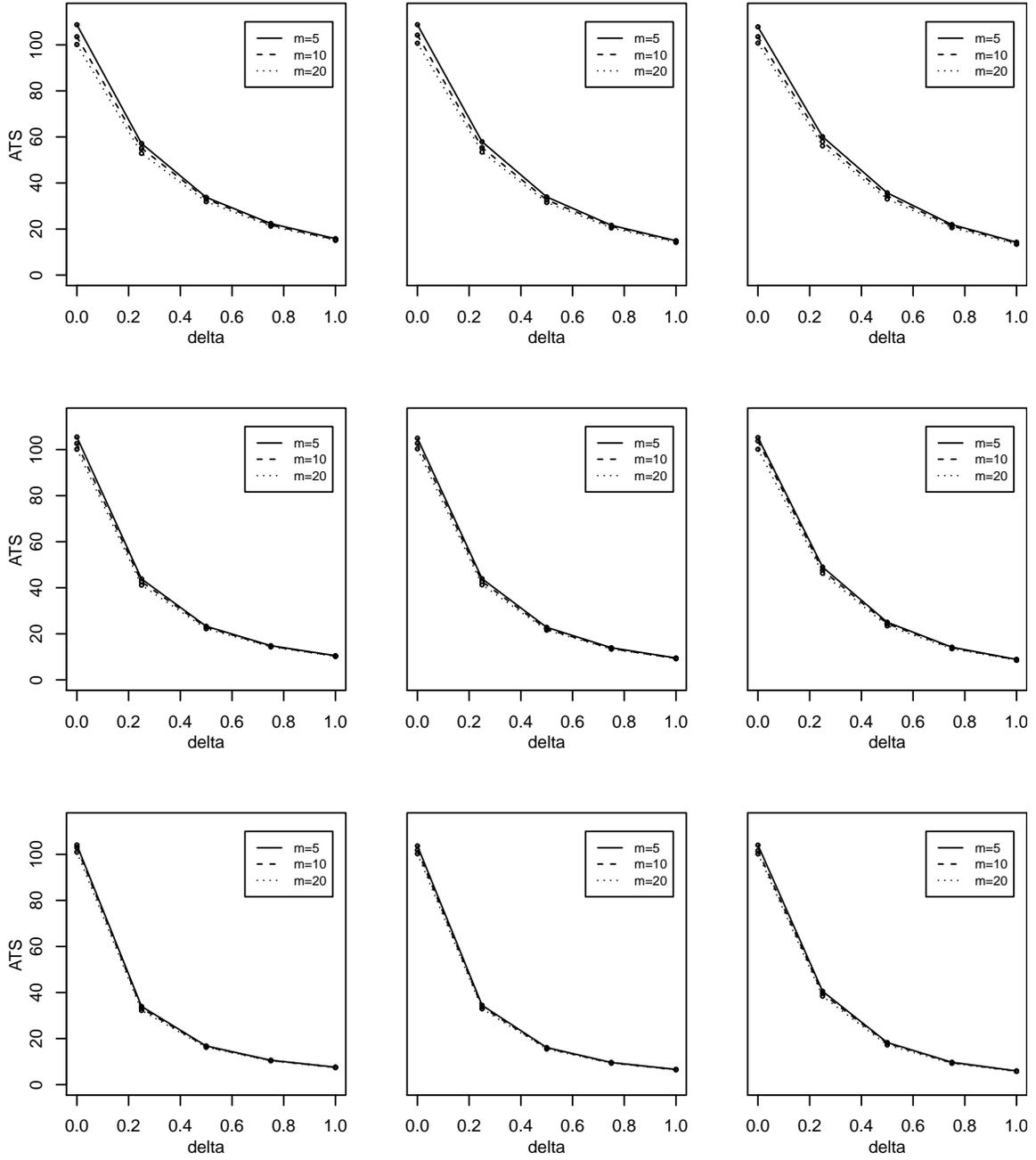


Figure 2: Actual ATS_0 and ATS_1 values of the chart (8)-(9), for detecting step mean shift of the size δ occurring at the initial time point, in cases when the IC mean function $\mu(t)$ and the IC variance function $\sigma_y^2(t)$ are estimated from an IC data with m individuals, $m = 5, 10$ or 20 , $k = 0.1$ (1st column), 0.2 (2nd column) or 0.5 (3rd column), $d = 2$ (1st row), 5 (2nd row) or 10 (3rd row), $\omega = 0.001$, and the nominal ATS_0 is 100 .

each individual would have 200, 500, and 1000 observations, respectively, in the design interval $[0, 1]$. In some applications, such as the total cholesterol level example discussed in Section 4, the individuals have fewer observations. Next, in the example of Figure 2, we choose d to be 0.2, 0.5, and 1, m to be 10, 20, and 40, and the other parameters to be unchanged. The corresponding results of the actual ATS_0 and ATS_1 values of the chart (8)-(9) for detecting the step shifts are shown in Table S.5 and Figure S.3 of the supplementary file. The results for detecting the drifts are shown in Table S.6 and Figure S.4 of the supplementary file. From these figures and tables, it can be seen that most conclusions made above from Figures 2 and S.2 are still valid here, except that (i) for a given m value, the actual ATS_0 values in this case seem a little farther away from the nominal ATS_0 value of 100, compared to the actual ATS_0 values shown in Figures 2 and S.2, (ii) it requires $m \geq 20$ in this case so that the actual ATS_0 values would be within about 5% of the nominal ATS_0 value, and (iii) the actual ATS_1 values in this case are generally larger than the corresponding values shown in Figures 2 and S.2. All these results are intuitively reasonable.

In the literature, there are some existing control charts designed specifically for detecting time-varying mean shifts, or mean drifts. One type of such control charts is based on the idea to design a chart adaptively using a dynamic prediction of the shift size. Such adaptive CUSUM or EWMA charts (cf., Capizzi and Masarotto 2003, Shu and Jiang 2006, Sparks 2000) can be used in our proposed DySS method in place of the conventional CUSUM chart (8)-(9). Another control chart designed for detecting time-varying mean shifts is the reference-free cuscore (RFCuscore) chart by Han and Tsung (2006). Next, we consider using the adaptive CUSUM chart by Shu and Jiang (2006) and the RFCuscore chart by Han and Tsung (2006) in our DySS method. By the adaptive CUSUM chart, the charting statistic C_j^+ in (8) needs to be changed to

$$C_{j,A}^+ = \max \left[0, C_{j-1,A}^+ + \left(\widehat{\epsilon}(t_j^*) - \widehat{\delta}_j^+ / 2 \right) / \rho(\widehat{\delta}_j^+ / 2) \right], \quad \text{for } j \geq 1,$$

where $C_{0,A}^+ = 0$, $\widehat{\delta}_j^+$ is the predicted shift size at time j defined by

$$\widehat{\delta}_j^+ = \max \left(\delta_{min}^+, (1 - \lambda)\widehat{\delta}_{j-1}^+ + \lambda\widehat{\epsilon}(t_j^*) \right), \quad \text{for } j \geq 1,$$

$\widehat{\delta}_0^+ = \delta_{min}^+$, δ_{min}^+ is a parameter denoting the minimum shift size, $\lambda \in (0, 1]$ is a weighting parameter, $\rho(k)$ is a control limit defined by

$$\rho(k) = [\log(1 + 2k^2 ARL_0 + 2.332k)] / 2k - 1.166,$$

and $ARL_0 = ATS_0 * d/10$ is the pre-specified IC ARL value. The chart gives a signal at time j when $C_{j,A}^+ > 1$. As in Shu and Jiang (2006), we choose $\lambda = 0.1$ in this procedure. The RFCuscore

chart gives a signal at time j if

$$\max_{1 \leq w \leq j} \left[\sum_{i=j-w+1}^j |\widehat{\epsilon}(t_i^*)| (\widehat{\epsilon}(t_i^*) - |\widehat{\epsilon}(t_i^*)|/2) \right] \geq c,$$

where c is a control limit chosen to reach a given ATS_0 level. In the example of Figures 2 and S.2, the actual ATS_0 and ATS_1 values of the adaptive CUSUM chart and the RFCuscore chart are computed in the same way as those of the chart (8)-(9), in cases when δ_{min}^+ is chosen to be 0.2, 0.4, and 1.0 and other parameters are chosen to be the same values as those in Figures 2 and S.2. These values are presented in Tables S.7-S.9 and Figures S.5-S.7 of the supplementary file. In cases when $m = 20$, $k = 0.2$ in the chart (8)-(9), and $\delta_{min}^+ = 0.4$ in the adaptive CUSUM chart, the computed actual ATS_0 and ATS_1 values of the three charts are shown in Figure S.8 in the supplementary file, for detecting both step shifts and drifts. From these results, it can be seen that (i) the conventional CUSUM chart (8)-(9) and the adaptive CUSUM chart perform almost identically in all cases considered, and (ii) the RFCuscore chart performs slightly worse in certain cases, especially when detecting drifts.

The supplementary file contains some results in cases when observations at different time points are correlated and their error terms follow the AR(1) model (11). These results show that the CUSUM chart (14)-(15), which is based on the standardized error terms, is indeed effective in detecting mean shifts in the original process observations. In cases when $\mu(t)$, $\sigma_y^2(t)$ and ϕ are unknown and they need to be estimated from an IC dataset, it seems that more IC data are needed to make the actual ATS_0 values of the chart (14)-(15) to be within 10% of the nominal ATS_0 value, compared to the results in Figure 2 when within-subject observations are independent. Finally, in the supplementary file, we also consider an example in which the within-subject observations are correlated, but the correlation does not follow the AR(1) model (11). The block bootstrap procedure discussed in Section 2.4 is used in designing our proposed DySS method. From the results in this example, it can be seen that the DySS method still performs well in such cases.

4 Application to the Total Cholesterol Level Example

In this section, we apply our proposed DySS method to the total cholesterol level example that is briefly described in Section 1. This example is a part of the SHARe Framingham Heart Study of the National Heart Lung and Blood Institute (cf., Cupples et al. 2007). In the data, there are

1028 non-stroke patients (i.e. $m = 1028$) and 27 stroke patients. In the study, each patient was followed 7 times (i.e., $J = 7$), and the total cholesterol level (in mg/100ml) was recorded at each time. Because the total cholesterol level, denoted as y , is an important risk factor of stroke, it is important to detect its irregular temporal pattern so that medical interventions can be applied in a timely manner and stroke can be avoided. Here, we demonstrate that our proposed DySS method can be used for identifying irregular total cholesterol level patterns in this example.

First, the observed data of the 1028 non-stroke patients are used as the IC data, from which we can estimate the IC mean function $\mu(t)$ and the IC variance function $\sigma_y^2(t)$. To this end, the covariance function $V(s, t)$ is first estimated by (5), and then $\mu(t)$ and $\sigma_y^2(t)$ are estimated by $\hat{\mu}(t; \tilde{V})$ and $\hat{\sigma}_y^2(t)$, respectively, as described in Section 2.1. The 95% pointwise confidence band of $\mu(t)$, defined to be $\hat{\mu}(t; \tilde{V}) \pm 1.96\hat{\sigma}_y(t)$, and $\hat{\mu}(t; \tilde{V})$ are presented in Figure 3 with the observed total cholesterol levels of the 27 stroke patients. From the figure, it can be seen that only 6 out of 27 stroke patients are detected to have irregular total cholesterol level patterns by this confidence interval approach because their observed total cholesterol levels exceed the upper confidence levels at least once. From the figure, it can also be seen that there are no stroke patients whose observed total cholesterol levels exceed the lower confidence levels.

After $\mu(t)$ and $\sigma_y^2(t)$ are estimated, we can compute the standardized residuals by

$$\hat{\epsilon}(t_{ij}) = \frac{y(t_{ij}) - \hat{\mu}(t_{ij}; \tilde{V})}{\hat{\sigma}_y(t_{ij})}, \quad \text{for } i = 1, \dots, m, j = 1, \dots, J.$$

To address the possible autocorrelation among $\{\hat{\epsilon}(t_{ij})\}$, we consider the AR(1) model (11) (see also (12)), and the coefficient ϕ in the model is estimated by the LS procedure described in Section 2.3 to be $\hat{\phi} = 0.9217$. The goodness-of-fit of the AR(1) model is studied as follows. First, the predicted values of the model (cf., (12))

$$\tilde{\epsilon}(t_{ij}) = \hat{\phi}^{\Delta_{i,j-1}} \hat{\epsilon}(t_{i,j-1})$$

are ordered from the smallest to the largest and then divided into g groups of the same size. Namely, the k th group includes all predicted values in the interval $I_k = [q_{k-1}, q_k)$, for $k = 1, 2, \dots, g$, where q_k is the (k/g) th quantile of the predicted values $\{\tilde{\epsilon}(t_{ij})\}$, for $k = 1, 2, \dots, g - 1$, $q_0 = -\infty$, and $q_g = \infty$. Let O_k be the proportion of all standardized residuals $\{\hat{\epsilon}(t_{ij})\}$ in the interval I_k , for $k = 1, 2, \dots, g$. Then, the following Pearson's χ^2 statistic:

$$X^2 = \sum_{k=1}^g \frac{(O_k - 1/g)^2}{1/g}$$

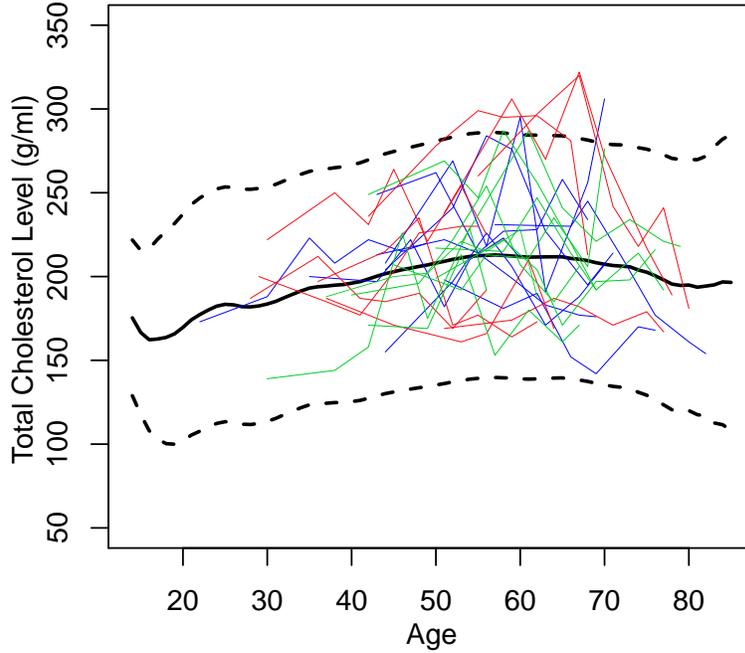


Figure 3: The 95% pointwise confidence band of the mean total cholesterol level $\mu(t)$ (dashed curves), the point estimator $\hat{\mu}(t; \tilde{V})$ of $\mu(t)$ (dark solid curve), and the observed total cholesterol levels of the 27 stroke patients (thin curves).

measures the discrepancy between the distribution of the standardized residuals and the distribution of their predicted values by the AR(1) model. If the AR(1) model fits the data well, then the distribution of X^2 should be approximately χ_{g-2}^2 , where the degree of freedom is $g - 2$ because there is one parameter (i.e., ϕ) in the AR(1) model that needs to be estimated beforehand. If we choose $g = 12$, then the observed value of X^2 is 0.2265, giving a p -value of about 1.0. We also tried other g values in the interval $[5, 50]$, and similarly large p -values were obtained. Therefore, the AR(1) model seems to fit the standardized residuals well.

After $\mu(t)$, $\sigma_y^2(t)$ and ϕ are all estimated, next we design a CUSUM chart to detect irregular longitudinal patterns of the total cholesterol levels of the 27 stroke patients. In this example, because we are mainly concerned about the upward shifts in patients' total cholesterol levels, only the upward CUSUM chart (14)-(15) is considered here. Since each stroke patient has an average of 2.3 observations every 10 years (cf., Figure 3), we choose $d = 2$. In the CUSUM chart, we choose $k = 0.1$ and $ATS_0 = 25$. In such cases, the control limit ρ is computed to be 0.927. The CUSUM charts for monitoring the 27 stroke patients are presented in Figure 4, from which we can see that 22 out of the 27 stroke patients are detected to have upward mean shifts. The 22 signal

times, computed from the starting point of each process monitoring, are listed in Table 2. The average signal time is 13.818 years. Compared to the confidence interval approach described at the beginning of this section, it can be seen that our DySS approach is more effective in detecting irregular longitudinal patterns of the total cholesterol levels.

Table 2: Signal times (years) of the 22 stroke patients by our proposed CUSUM chart (14)-(15).

Patient ID	Signal time	Patient ID	Signal time
2	12	16	19
3	13	17	12
4	12	18	8
5	22	19	12
6	11	20	16
7	8	21	8
9	23	22	9
11	24	23	11
12	8	24	16
13	7	25	26
15	8	27	19

5 Concluding Remarks

We have described the DySS screening system for dynamically identifying individuals with irregular longitudinal patterns. This approach tries to make use of all available information when monitoring the longitudinal pattern of an individual, including the information from an IC population of certain well-functioning individuals and all history data of the individual being monitored, using both longitudinal data analysis and SPC techniques. Our numerical examples have shown that it is effective in various cases. Although this paper focuses on the monitoring of the mean function of the longitudinal response, similar DySS screening systems can be developed for monitoring the variance function and other features of the response that may affect the longitudinal performance of the individuals in a specific application. This approach can also be generalized to cases when multiple responses are present.

In this paper, we have discussed cases when the observation times are regularly spaced, observation times are irregularly spaced, and the observations at different time points are correlated. The cases when observation times are irregularly spaced, the observations at different time points are correlated, and the correlation needs to be described by a parametric time series model are

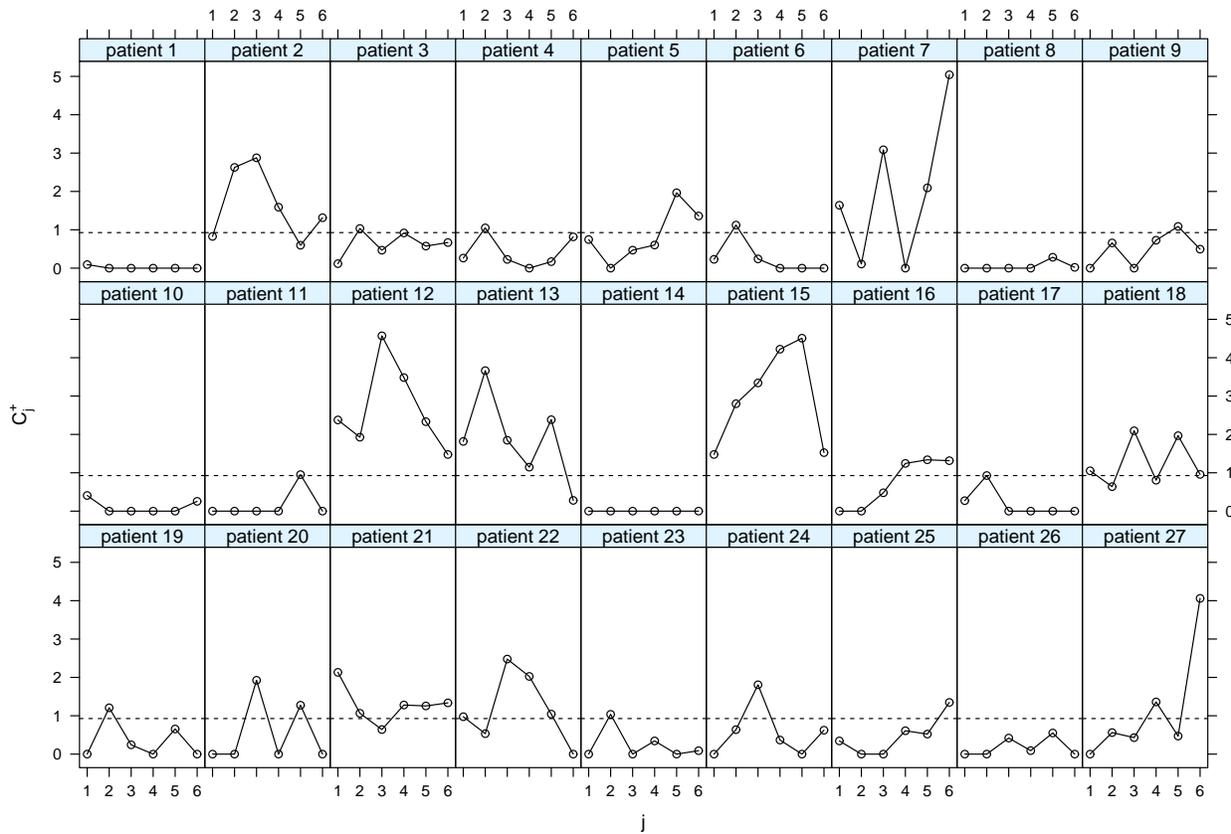


Figure 4: CUSUM charts (14)-(15) with $(k, h, ATS_0) = (0.1, 0.927, 25)$ for monitoring the 27 stroke patients. The dashed horizontal lines denote the control limit ρ .

especially challenging, because there is little existing discussion in the literature about time series modeling in such cases. In the paper, only the case when the AR(1) model (11) is appropriate is discussed. More future research is required to handle cases with more flexible parametric time series models.

Supplementary Materials

supplemental.pdf: This pdf file discusses some statistical properties of the estimator $\hat{\mu}(t; \tilde{V})$ defined in Section 2.1, some technical details about the adjustment procedure (15) in Li (2011) and the CV procedure discussed in Section 2.1, and some extra numerical results.

ComputerCodesAndData.zip: This zip file contains MATLAB source codes of our proposed DySS method and the stroke data used in the paper.

Acknowledgments: We thank the editor, the associate editor, and two referees for many constructive comments and suggestions, which greatly improved the quality of the paper. This research is supported in part by an NSF grant and two grants from Natural Sciences Foundation of China.

References

- Apley, D.W., and Tsung, F. (2002), “The autoregressive T^2 chart for monitoring univariate auto-correlated processes,” *Journal of Quality Technology*, **34**, 80–96.
- Brockwell, P., Davis, R., and Yang, Y. (2007), “Continuous-time Gaussian autoregression,” *Statistica Sinica*, **17**, 63–80.
- Capizzi, G., and Masarotto, G. (2003), “An adaptive exponentially weighted moving average control chart,” *Technometrics*, **45**, 199–207.
- Chen, K., and Jin, Z. (2005), “Local polynomial regression analysis of clustered data,” *Biometrika*, **92**, 59–74.
- Cupples, L.A. et al. (2007), The Framingham Heart Study 100K SNP genome-wide association study resource: overview of 17 phenotype working group reports,” *BMC Medical Genetics*, **8**, (Suppl 1): S1.
- Ding, Y., Zeng, L., and Zhou, S. (2006), “Phase I analysis for monitoring nonlinear profiles in manufacturing processes,” *Journal of Quality Technology*, **38**, 199–216.
- Gan, F.F. (1995), “Joint monitoring of process mean and variance using exponentially weighted moving average control charts,” *Technometrics*, **37**, 446–453.
- Han, D., and Tsung, F. (2006), “A reference-free cuscore chart for dynamic mean change detection and a unified framework for charting performance comparison,” *Journal of the American Statistical Association*, **101**, 368–386.
- Hawkins, D.M., and Deng, Q. (2010), “A nonparametric change-point control chart,” *Journal of Quality Technology*, **42**, 165–173.

- Hawkins, D.M., and Olwell, D.H. (1998), *Cumulative Sum Charts and Charting for Quality Improvement*, New York: Springer-Verlag.
- Jensen, W.A., and Birch, J.B. (2009), “Profile monitoring via nonlinear mixed models,” *Journal of Quality Technology*, **41**, 18–34.
- Jiang, W. (2004), “Multivariate control charts for monitoring autocorrelated processes,” *Journal of Quality Technology*, **36**, 367–379.
- Jiang, W., Tsui, K.L., and Woodall, W. (2000), “A new SPC monitoring method: the ARMA chart,” *Technometrics*, **42**, 399–410.
- Knoth, S. (2011), “spc: Statistical Process Control,” *R package version 0.4.0*, <http://CRAN.R-project.org/package=spc>.
- Lahiri, S.N. (2003), *Resampling Methods for Dependent Data*, New York: Springer.
- Li, Y. (2011), “Efficient semiparametric regression for longitudinal data with nonparametric covariance estimation,” *Biometrika*, **98**, 355–370.
- Lu, C.W., and Reynolds, M.R., Jr. (2001), “Control charts for monitoring an autocorrelated process,” *Journal of Quality Technology*, **33**, 316–334.
- Ma, S., Yang, L., and Carroll, R. (2012), “A simultaneous confidence band for sparse longitudinal regression,” *Statistica Sinica*, **22**, 95–122.
- Maller, R.A., Müller, G., and Szimayer, A. (2008), “GARCH modelling in continuous time for irregularly spaced time series data,” *Bernoulli*, **14**, 519–542.
- Montgomery, D. C. (2009), *Introduction To Statistical Quality Control (6th edition)*, New York: John Wiley & Sons.
- Moustakides, G.V. (1986), “Optimal stopping times for detecting changes in distributions,” *The Annals of Statistics*, **14**, 1379–1387.
- Pan, J., Ye, H., and Li, R. (2009), “Nonparametric regression of covariance structures in longitudinal studies,” *Technical Report*, **2009.46**, School of Mathematics, The University of Manchester, UK.
- Qiu, P. (2013), *Introduction to Statistical Process Control*, London: Chapman & Hall/CRC.

- Qiu, P. (2008), “Distribution-free multivariate process control based on log-linear modeling,” *IIE Transactions*, **40**, 664–677.
- Qiu, P., and Hawkins, D. (2001), “A rank based multivariate CUSUM procedure,” *Technometrics*, **43**, 120–132.
- Qiu, P., and Li, Z. (2011a), “On nonparametric statistical process control of univariate processes,” *Technometrics*, **53**, 390–405.
- Qiu, P., and Li, Z. (2011b), “Distribution-free monitoring of univariate processes,” *Statistics and Probability Letters*, **81**, 1833–1840.
- Qiu, P., Zou, C., and Wang, Z. (2010), “Nonparametric profile monitoring by mixed effects modeling (with discussions),” *Technometrics*, **52**, 265–277.
- Reynolds, M.R., Jr., Amin, R.W., and Arnold, J.C. (1990), “CUSUM charts with variable sampling intervals,” *Technometrics*, **32**, 371–384.
- Shu, L., and Jiang, W. (2006), “A Markov chain model for the adaptive cusum control chart,” *Journal of Quality Technology*, **38**, 135–147.
- Sparks, R.S. (2000), “CUSUM charts for signalling varying location shifts,” *Journal of Quality Technology*, **32**, 157–171.
- Vityazev, V.V. (1996), “Time series analysis of unequally spaced data: the statistical properties of the Schuster periodogram,” *Astronomical and Astrophysical Transactions*, **11**, 159–173.
- Yao, F., Müller, H.G., and Wang, J.L. (2005), “Functional data analysis for sparse longitudinal data,” *Journal of the American Statistical Association*, **100**, 577–590.
- Yeh, A.B., Lin, D.K.J., and Venkataramani, C. (2004), “Unified CUSUM charts for monitoring process mean and variability,” *Quality Technology and Quantitative Management*, **1**, 65–86.
- Zhao, Z., and Wu, W. (2008), “Confidence bands in nonparametric time series regression,” *Annals of Statistics*, **36**, 1854–1878.
- Zou, C., and Tsung, F. (2010), “Likelihood ratio-based distribution-free EWMA control charts,” *Journal of Quality Technology*, **42**, 1–23.