IIE Transactions

ISSN: 0740-817X (Print) 1545-8830 (Online) Journal homepage: http://www.tandfonline.com/loi/uiie20

Nonparametric Dynamic Screening System for Monitoring Correlated Longitudinal Data

Jun Li & Peihua Qiu

To cite this article: Jun Li & Peihua Qiu (2016): Nonparametric Dynamic Screening System for Monitoring Correlated Longitudinal Data, IIE Transactions

To link to this article: http://dx.doi.org/10.1080/0740817X.2016.1146423

Accepted author version posted online: 12 Feb 2016.

Nonparametric Dynamic Screening System for Monitoring Correlated Longitudinal Data

Jun Li[1] and Peihua Qiu[2]

[1]Department of Statistics, University of California at Riverside

[2]Department of Biostatistics, University of Florida

Abstract

In many applications, including disease early detection and prevention, and performance evaluation of airplanes and other durable products, we need to sequentially monitor the longitudinal pattern of certain performance variables of a subject. A signal should be given as soon as possible once the pattern becomes abnormal. Recently, a new statistical method called dynamic screening system (DySS) has been proposed to solve this problem. It is a combination of longitudinal data analysis and statistical process control. However, the current DySS method can only handle cases when observations are normally distributed and within-subject observations are independent or follow a specific time series model (e.g., AR(1) model). In this paper, we propose a new nonparametric DySS method which can handle cases when the observation distribution and the correlation among within-subject observations are arbitrary. Therefore, it broadens the application of the DySS method greatly. Numerical studies show that the new method works well in practice.

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

1 Introduction

In practice, we often need to detect significant difference between the longitudinal pattern of some performance variables of a given subject and the regular longitudinal pattern of some well-functioning subjects as soon as possible so that unpleasant consequences can be avoided. This dynamic screening (DS) problem is popular in our daily life. For instance, individual people's medical indices (e.g., blood pressure and cholesterol level) need to be checked regularly. If their observations are significantly worse than the values of a typical healthy person of the same age, then some proper treatments or interventions should be made to avoid stroke and other deadly cardiovascular diseases. This paper proposes a flexible and efficient method to solve the DS problem.

In the literature, there are two types of methods that are relevant to the DS problem. The first type belongs to the research area of *longitudinal data analysis (LDA)*. By an LDA method, we can construct confidence intervals for the performance variables at different time points based on an observed dataset of some well-functioning subjects. Then, a new subject can be identified as the one who has an irregular longitudinal pattern if its observed values of the performance variables fall outside the confidence intervals. Some existing methods for constructing such confidence intervals include Chen and Jin (2005), Li (2011), Liang and Zeger (1986), Ma et al. (2012), Wang (2003), Xiang et al (2013), and Zhao and Wu (2008). This confidence interval approach is inefficient in handling the DS problem because i) it does not make use of all historical data of the subject in question when making decisions about its performance at the current time point, and ii) it cannot monitor a subject sequentially over time. The second type of statistical methods relevant to the DS problem belongs to the research area of *statistical process control (SPC)*. By a SPC control chart, we can monitor each subject sequentially, and a signal will be given as soon as the chart detects a shift in the longitudinal pattern of the performance variables from an in-control (IC) status to an out-of-control (OC) status (cf., Montgomery 2009, Hawkins and Olwell 1998, Qiu 2014). However, a conventional SPC chart is designed for monitoring a single sequence of observations collected over time (a so-called single process here), and the distribution of the process observations is assumed unchanged when the process is IC to use such a chart. In the

DS problem, however, if the longitudinal pattern of the performance variables of each subject is regarded as a process, then there are many processes involved. Also, the observation distribution can change over time even for a well-functioning subject (e.g., the mean cholesterol level of a healthy person would change with age).

Recently, Qiu and Xiang (2014) suggested a so-called dynamic screening system (DySS) for solving the DS problem. That method combines certain strengths of the LDA and SPC methods. It consists of three main steps: i) a regular longitudinal pattern is estimated from an observed dataset (called IC dataset hereafter) of a number of well-functioning subjects, ii) observations of a new subject under study are standardized by the estimated regular longitudinal pattern, and iii) an SPC chart is used for monitoring the standardized observations of the new subject. In this method, the approach for estimating the regular longitudinal pattern in step one is completely nonparametric. But, in the remaining two steps, it assumes that observations are normally distributed, and within-subject observations are either independent or dependent following an AR(1) model. These assumptions greatly restrict its applications. In this paper, we propose a novel nonparametric DySS method for solving the DS problem without the restrictive assumptions mentioned above. The major idea behind this method is that observations of a new subject are decorrelated sequentially each time when a new observation is obtained based on the covariance function estimated from the IC dataset. These decorrelated data would also be asymptotically normally distributed.

The paper is organized as follows. Our proposed new methodology will be described in detail in Section 2. A numerical study is presented in Section 3. Then, the proposed method is demonstrated using a real-data example in Section 4. Some remarks conclude the article in Section 5. Some technical details are given in the appendix.

2 Methodology

Our proposed nonparametric DySS method consists of three main steps. In the first step, the regular longitudinal pattern of the performance variables needs to be estimated from an IC data. In this step, we still use the nonparametric modeling approach described in Qiu and Xiang (2014)

for analyzing the Phase I longitudinal data. That approach is briefly described in Subsection 2.1. Then, the remaining two steps for standardizing and monitoring the observations of a new subject are discussed in Subsection 2.2.

2.1 Phase I modeling of longitudinal data

Assume that we have m well-functioning subjects in the IC dataset. For the i-th ($i = 1, ..., m$) subject, the measurements are taken at times t_{i1}, ..., t_{in_i} in the time period $[0, T]$, and the corresponding measurements at those times are denoted by $y(t_{ij})$, for $j = 1, ..., n_i$. The following model has been used widely for modeling such longitudinal data in the literature:

$$y(t_{ij}) = \mu(t_{ij}) + \epsilon(t_{ij}), \quad i = 1, ..., m, \ j = 1, ..., n_i, \tag{1}$$

where $\mu(\cdot)$ is a smooth function and models the population mean curve, $\epsilon(\cdot)$ is the error term with the covariance function $V(s, t) = \text{cov}(\epsilon(s), \epsilon(t))$ for any $s, t \in [0, T]$.

Following the estimation procedure in Qiu and Xiang (2014), we can obtain the estimates of $\mu(t)$ and $V(s, t)$ by the algorithm below.

Step 1. Let Σ_i be the covariance matrix of $\boldsymbol{y}_i = (y(t_{i1}), y(t_{i2}), ..., y(t_{in_i}))^T$, for $i = 1, ...m$. In this step, we assume $\Sigma_i = I_{n_i}$, where I_{n_i} is the n_i-dimensional identity matrix. Define

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

and $K_{ih}(t) = \text{diag}(K((t_{i1} - t)/h), ...K((t_{in_i} - t)/h))/h$, where $K(\cdot)$ is a kernel function and $h > 0$ is a bandwidth. Let $J_i = \text{diag}(I_{\{|t_{i1}-t| \le h\}}, ..., I_{\{|t_{in_i}-t| \le h\}})$, where $I_{\{A\}}$ is the indicator function and takes the value of 1 if A is true and 0 otherwise. Then an initial pth-order local polynomial kernel estimator of $\mu(t)$ is given by

$$\tilde{\mu}(t) = \boldsymbol{e}_1^T \left(\sum_{i=1}^{m} X_i^T W_i X_i \right)^{-1} \left(\sum_{i=1}^{m} X_i^T W_i \boldsymbol{y}_i \right), \tag{2}$$

where $\boldsymbol{e}_1$ is a $(p + 1)$-dimensional vector with 1 at the first component and 0 anywhere else, $W_i = K_{ih}^{1/2}(t)(J_i \Sigma_i J_i)^{-1} K_{ih}^{1/2}(t)$ with $\Sigma_i = I_{n_i}$.

Step 2. Based on the initial estimate $\tilde{\mu}(t)$, calculate the residuals

$$\tilde{\epsilon}(t_{ij}) = y(t_{ij}) - \tilde{\mu}(t_{ij}), \quad i = 1, ..., m, \ j = 1, ..., n_i.$$

For $l_1, l_2 = 0, 1, 2$, define

$$S_{l_1 l_2}(s,t) = \frac{1}{Nh^2} \sum_{i=1}^{m} \sum_{j=1}^{n_i} \sum_{j' \neq j} \left(\frac{t_{ij} - s}{h}\right)^{l_1} \left(\frac{t_{ij'} - t}{h}\right)^{l_2} K\left(\frac{t_{ij} - s}{h}\right) K\left(\frac{t_{ij'} - t}{h}\right),$$

$$V_{l_1 l_2}(s,t) = \frac{1}{Nh^2} \sum_{i=1}^{m} \sum_{j=1}^{n_i} \sum_{j' \neq j} \tilde{\epsilon}(t_{ij})\tilde{\epsilon}(t_{ij'}) \left(\frac{t_{ij} - s}{h}\right)^{l_1} \left(\frac{t_{ij'} - t}{h}\right)^{l_2} K\left(\frac{t_{ij} - s}{h}\right) K\left(\frac{t_{ij'} - t}{h}\right),$$

where $N = \sum_{i=1}^{m} n_i(n_i - 1)$. Then, for $s \neq t$, an estimate of $V(s,t)$ is given by

$$\hat{V}(s,t) = \left(A_1(s,t)V_{00}(s,t) - A_2(s,t)V_{10}(s,t) - A_3(s,t)V_{01}(s,t)\right) B^{-1}(s,t),$$

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

Step 3. Since $V(t,t)$, the variance of $y(t)$, can be considered as the mean function of $\epsilon^2(t)$, it can be estimated similarly by using (2). An estimate of $V(t,t)$ is

$$\hat{V}(t,t) = \boldsymbol{e}_1^T \left(\sum_{i=1}^{m} X_i^T W_i X_i\right)^{-1} \left(\sum_{i=1}^{m} X_i^T W_i \tilde{\boldsymbol{\epsilon}}_i^2\right),$$

where $\tilde{\boldsymbol{\epsilon}}_i^2 = (\tilde{\epsilon}^2(t_{i1}), ..., \tilde{\epsilon}^2(t_{in_i}))^T$, and the W_i are the same as those used in Step 1.

Step 4. Since

$$\Sigma_i = \begin{pmatrix} V(t_{i1}, t_{i1}) & V(t_{i1}, t_{i2}) & \cdots & V(t_{i1}, t_{in_i}) \\ V(t_{i2}, t_{i1}) & V(t_{i2}, t_{i2}) & \cdots & V(t_{i2}, t_{in_i}) \\ \vdots & \vdots & \ddots & \vdots \\ V(t_{in_i}, t_{i1}) & V(t_{in_i}, t_{i2}) & \cdots & V(t_{in_i}, t_{in_i}) \end{pmatrix},$$

based on the estimate of $V(s,t)$ from Steps 2 and 3, we can obtain the estimate of Σ_i, denoted by $\hat{\Sigma}_i$. Then, an updated and final estimate of $\mu(t)$ is

$$\hat{\mu}(t) = \boldsymbol{e}_1^T \left(\sum_{i=1}^{m} X_i^T \hat{W}_i X_i\right)^{-1} \left(\sum_{i=1}^{m} X_i^T \hat{W}_i \boldsymbol{y}_i\right),$$

where $\hat{W}_i = K_{ih}^{1/2}(t)(J_i\hat{\Sigma}_i J_i)^{-1}K_{ih}^{1/2}(t)$.

In all our numerical examples presented later, we choose $p = 1$, $K(t) = 0.75(1 - t^2)I_{\{|t|\le 1\}}$, and the bandwidth h in the above four steps are selected separately by the conventional cross-validation method.

2.2 Phase-II monitoring of longitudinal data

In Phase II, the measurements of a new subject are sequentially collected. The task of Phase-II monitoring is to determine whether the new individual's measurements are following the estimated regular longitudinal pattern of the well-functioning subjects, described by the estimated mean function $\hat{\mu}(t)$ and the estimated covariance function $\hat{V}(s,t)$. If not, then a signal should be given as early as possible. Denote the times when the measurements are taken from the new subject by $\{t_j^*, j = 1, 2, ...\}$ and the corresponding measurements are $\{y(t_j^*), j = 1, 2, ...\}$.

For the monitoring purpose, Qiu and Xiang (2014) further assumes that $\{y(t_j^*), j = 1, 2, ...\}$ are independent of each other. Under this independence assumption, they define the following standardized values,

$$\hat{\epsilon}(t_j^*) = \frac{y(t_j^*) - \hat{\mu}(t_j^*)}{\hat{V}(t_j^*, t_j^*)}, \quad j = 1, 2, ..., \tag{3}$$

which are then asymptotically i.i.d. random variables. The $\hat{\epsilon}(t_j^*)$ are then used in a conventional control chart, such as a CUSUM or EWMA chart.

From the above description, we can see that, in order for the procedure in Qiu and Xiang (2014) to work, the independence among $\{y(t_j^*), j = 1, 2, ...\}$ is a necessary assumption. However, if the measurements from the new subject are IC, $\{y(t_j^*), j = 1, 2, ...\}$ should follow the same IC model as in (1). In other words, for $j = 1, 2, ...$,

$$y(t_j^*) = \mu(t_j^*) + \epsilon(t_j^*), \tag{4}$$

and the covariance matrix of the $\epsilon(t_j^*)$ is $V(s,t)$, which implies that the $y(t_j^*)$ are correlated. Under this situation, results from the procedure in Qiu and Xiang (2014) may not be reliable since the required independence assumption could be violated. In the following, we introduce a new method to standardize the data, and the resulting standardized values can be shown to be uncorrelated.

2.2.1 Standardization procedure

To facilitate the exposition, we treat $\mu(t)$ and $V(s,t)$ as known by using their respective estimates $\hat{\mu}(t)$ and $\hat{V}(s,t)$ from the Phase-I study described in the previous section. We start from the first measurement $y(t_1^*)$. We calculate $\epsilon(t_1^*) = y(t_1^*) - \mu(t_1^*)$, then the variance of $\epsilon(t_1^*)$ is $\sigma_{11} = V(t_1^*, t_1^*)$. Therefore, we can define its standardized value by

$$e^*(t_1^*) = \frac{\epsilon(t_1^*)}{\sqrt{\sigma_{11}}}.$$

For the second measurement $y(t_2^*)$, we can also calculate $\epsilon(t_2^*) = y(t_2^*) - \mu(t_2^*)$. Now, $\epsilon(t_2^*)$ is correlated with $\epsilon(t_1^*)$, and the covariance matrix of $\boldsymbol{\epsilon}_2 = (\epsilon(t_1^*), \epsilon(t_2^*))^T$ is given by $\Sigma_{22} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$, where $\sigma_{12} = \sigma_{21} = V(t_1^*, t_2^*)$, $\sigma_{22} = V(t_2^*, t_2^*)$. Based on the Cholesky decomposition, if we define $\Phi_2 = \begin{pmatrix} 1 & 0 \\ -\sigma_{12}\sigma_{11}^{-1} & 1 \end{pmatrix}$ and $D_2 = \begin{pmatrix} d_1^2 & 0 \\ 0 & d_2^2 \end{pmatrix} = \text{diag}(d_1^2, d_2^2)$, where $d_1^2 = \sigma_{11}$ and $d_2^2 = \sigma_{22} - \sigma_{21}\sigma_{11}^{-1}\sigma_{12}$, we have

$$\Phi_2 \Sigma_{22} \Phi_2^T = D_2.$$

This motivates us to define $\boldsymbol{e}_2 = (e(t_1^*), e(t_2^*))^T = \Phi_2 \boldsymbol{\epsilon}_2$, i.e.,

$$e(t_1^*) = \epsilon(t_1^*),$$
$$e(t_2^*) = -\sigma_{12}\sigma_{11}^{-1}\epsilon(t_1^*) + \epsilon(t_2^*).$$

Then $\text{cov}(\boldsymbol{e}_2) = D_2$, a diagonal matrix, which implies that $e(t_1^*)$ and $e(t_2^*)$ are uncorrelated. Noticing that $e(t_1^*)/d_1 = e^*(t_1^*)$, the standardized value we define at the first measurement, we can define the standardized value of the second measurement by

$$e^*(t_2^*) = \frac{e(t_2^*)}{d_2} = \frac{-\sigma_{12}\sigma_{11}^{-1}\epsilon(t_1^*) + \epsilon(t_2^*)}{d_2}.$$

Then $\{e^*(t_1^*), e^*(t_2^*)\}$ are uncorrelated with variance 1.

Similarly, for the third measurement $y(t_3^*)$, we calculate $\epsilon(t_3^*) = y(t_3^*) - \mu(t_3^*)$. The covariance matrix of $\boldsymbol{\epsilon}_3 = (\boldsymbol{\epsilon}_2^T, \epsilon(t_3^*))^T$ is given by $\Sigma_{33} = \begin{pmatrix} \Sigma_{22} & \boldsymbol{\sigma}_{23} \\ \boldsymbol{\sigma}_{23}^T & \sigma_{33} \end{pmatrix}$, where $\boldsymbol{\sigma}_{23} = \text{cov}(\boldsymbol{\epsilon}_2, \epsilon(t_3^*)) =$

$(V(t_1^*, t_3^*), V(t_2^*, t_3^*))^T$, $\sigma_{33} = \text{var}(\epsilon(t_3^*)) = V(t_3^*, t_3^*)$. It can be shown that the Cholesky decomposition of Σ_{33} leads to

$$\Phi_3 \Sigma_{33} \Phi_3^T = D_3.$$

where $\Phi_3 = \begin{pmatrix} \Phi_2 & \mathbf{0} \\ -\boldsymbol{\sigma}_{23}^T \Sigma_{22}^{-1} & 1 \end{pmatrix}$, and $D_3 = \text{diag}(d_1^2, d_2^2, d_3^2)$, $d_3^2 = \sigma_{33} - \boldsymbol{\sigma}_{23}^T \Sigma_{22}^{-1} \boldsymbol{\sigma}_{23}$. Therefore, if we define

$$e(t_3^*) = -\boldsymbol{\sigma}_{23}^T \Sigma_{22}^{-1} \boldsymbol{\epsilon}_2 + \epsilon(t_3^*),$$

then $\boldsymbol{e}_3 = (\boldsymbol{e}_2^T, e(t_3^*))^T = (e(t_1^*), e(t_2^*), e(t_3^*))^T = \Phi_3 \boldsymbol{\epsilon}_3$ and $\text{cov}(\boldsymbol{e}_3) = D_3$, which implies that $e(t_3^*)$ is uncorrelated with $e(t_1^*)$ and $e(t_2^*)$. The standardized value of the third measurement is then defined as

$$e^*(t_3^*) = \frac{e(t_3^*)}{d_3} = \frac{-\boldsymbol{\sigma}_{23}^T \Sigma_{22}^{-1} \boldsymbol{\epsilon}_2 + \epsilon(t_3^*)}{d_3},$$

which is uncorrelated with $\{e^*(t_1^*), e^*(t_2^*)\}$ and has variance 1.

Following the same procedure we can define the standardized value sequentially after a new measurement is collected from the new subject. More specifically, at the j-th measurement, we calculate $\epsilon(t_j^*) = y(t_j^*) - \mu(t_j^*)$. Then, the covariance matrix of $\boldsymbol{\epsilon}_j = (\boldsymbol{\epsilon}_{j-1}^T, \epsilon(t_j^*))^T$ is $\Sigma_{jj} = \begin{pmatrix} \Sigma_{j-1,j-1} & \boldsymbol{\sigma}_{j-1,j} \\ \boldsymbol{\sigma}_{j-1,j}^T & \sigma_{jj} \end{pmatrix}$, where $\boldsymbol{\epsilon}_{j-1}$ is the residual vector from the first $j-1$ measurements, $\Sigma_{j-1,j-1} = \text{cov}(\boldsymbol{\epsilon}_{j-1})$, $\boldsymbol{\sigma}_{j-1,j} = \text{cov}(\boldsymbol{\epsilon}_{j-1}, \epsilon(t_j^*)) = (V(t_1^*, t_j^*), ..., V(t_{j-1}^*, t_j^*))^T$, $\sigma_{jj} = \text{var}(\epsilon(t_j^*)) = V(t_j^*, t_j^*)$. It can be shown that the Cholesky decomposition of Σ_{jj} is given by

$$\Phi_j \Sigma_{jj} \Phi_j^T = D_j,$$

where $\Phi_j = \begin{pmatrix} \Phi_{j-1} & \mathbf{0} \\ -\boldsymbol{\sigma}_{j-1,j}^T \Sigma_{j-1,j-1}^{-1} & 1 \end{pmatrix}$, $D_j = \text{diag}(d_1^2, ..., d_j^2)$ with $d_j^2 = \sigma_{jj} - \boldsymbol{\sigma}_{j-1,j}^T \Sigma_{j-1,j-1}^{-1} \boldsymbol{\sigma}_{j-1,j}$, Φ_{j-1} and $\text{diag}(d_1^2, ..., d_{j-1}^2)$ are the Cholesky decomposition of $\Sigma_{j-1,j-1}$. Therefore, if we define

$$e(t_j^*) = -\boldsymbol{\sigma}_{j-1,j}^T \Sigma_{j-1,j-1}^{-1} \boldsymbol{\epsilon}_{j-1} + \epsilon(t_j^*),$$

then $\boldsymbol{e}_j = (e(t_1^*), ..., e(t_j^*))^T = (\boldsymbol{e}_{j-1}^T, e(t_j^*))^T = \begin{pmatrix} \Phi_{j-1} & \mathbf{0} \\ -\boldsymbol{\sigma}_{j-1,j}^T \Sigma_{j-1,j-1}^{-1} & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{\epsilon}_{j-1} \\ \epsilon(t_j^*) \end{pmatrix} = \Phi_j \boldsymbol{\epsilon}_j$ and $\text{cov}(\boldsymbol{e}_j) = D_j$. This implies that, if we define the standardized value of the j-th measurement by

$$e^*(t_j^*) = \frac{e(t_j^*)}{d_j} = \frac{-\boldsymbol{\sigma}_{j-1,j}^T \Sigma_{j-1,j-1}^{-1} \boldsymbol{\epsilon}_{j-1} + \epsilon(t_j^*)}{d_j},$$

$e^*(t_j^*)$ is uncorrelated with $\{e^*(t_1^*), ..., e^*(t_{j-1}^*)\}$ and has variance 1.

Based on the above standardization procedure, we are able to transform the correlated $\{y(t_j^*), j = 1, 2, ...\}$ to the uncorrelated $\{e^*(t_j^*), j = 1, 2, ...\}$. Because $e^*(t_j^*)$ is a linear combination of observations $y(t_1^*), y(t_2^*), ..., y(t_j^*)$, its distribution would be asymptotically normal. Therefore, the sequence $\{e^*(t_j^*), j = 1, 2, ...\}$ would be asymptotically i.i.d. with the common distribution N(0,1). In such cases, the regular CUSUM chart is reasonable to use. For example, to detect upward mean shifts in $\{y(t_j^*), j = 1, 2, ...\}$, we can define the CUSUM charting statistic by

$$S_j^+ = \max(0, S_{j-1}^+ + e^*(t_j^*) - k), \text{ for } j \geq 1, \tag{5}$$

where $S_0^+ = 0$ and $k > 0$ is the allowance constant. The corresponding CUSUM chart is to monitor S_j^+ and it triggers an alarm when S_j^+ exceeds some control limit l. To detect downward mean shifts, we can define the CUSUM charting statistic by

$$S_j^- = \min(0, S_{j-1}^- + e^*(t_j^*) + k), \text{ for } j \geq 1,$$

where $S_0^- = 0$ and $k > 0$ is the allowance constant. The corresponding CUSUM chart is to monitor S_j^- and it triggers an alarm when $S_j^- < -l$. To detect an arbitrary mean shift, we can combine S_j^+ and S_j^-, and the procedure triggers an alarm when $S_j^+ > l$ or $S_j^- < -l$. For simplicity, henceforth we use the CUSUM charting statistic in (5) as an example of our general CUSUM charting statistic.

Remark 1 *In the standardization procedure we propose above, at the j-th measurement, we need to calculate $\Sigma_{j-1,j-1}^{-1}$. To avoid calculating the inverse matrix directly, we suggest using the following recursive formula:*

$$\Sigma_{j-1,j-1}^{-1} = \begin{pmatrix} \Sigma_{j-2,j-2}^{-1} + \Sigma_{j-2,j-2}^{-1} \boldsymbol{\sigma}_{j-2,j-1} D_{j-1}^{-1} \boldsymbol{\sigma}_{j-2,j-1}^T \Sigma_{j-2,j-2}^{-1}, & -\Sigma_{j-2,j-2}^{-1} \boldsymbol{\sigma}_{j-2,j-1} D_{j-1}^{-1} \\ -D_{j-1}^{-1} \boldsymbol{\sigma}_{j-2,j-1}^T \Sigma_{j-2,j-2}^{-1}, & D_{j-1}^{-1} \end{pmatrix},$$

where $D_{j-1} = \sigma_{j-1,j-1} - \boldsymbol{\sigma}_{j-2,j-1}^T \Sigma_{j-2,j-2}^{-1} \boldsymbol{\sigma}_{j-2,j-1} = d_{j-1}^2$, *and* $\boldsymbol{\sigma}_{j-2,j-1} = \text{cov}(\boldsymbol{\epsilon}_{j-2}, \epsilon(t_{j-1}^*)) = (V(t_1^*, t_{j-1}^*), ..., V(t_{j-2}^*, t_{j-1}^*))^T$. *This result provides us a convenient way to calculate* $\Sigma_{j-1,j-1}^{-1}$ *recursively to simplify the computation.*

Remark 2 *Based on Remark 1, the computation associated with our proposed procedure should not be a big problem when* j *is relatively small. In practice, the* ATS_0 *value (see its definition in the next section) is often chosen to be a relatively small number (e.g., 50). In such cases, the control chart usually gives a signal at a time that is smaller than 3 or 4 times the chosen* ATS_0 *value. So, the computation and storage of* $\Sigma_{j-1,j-1}^{-1}$ *should not be a problem. However, in certain applications with very stable processes,* j *could be potentially large. In such situations, the computation and storage of* $\Sigma_{j-1,j-1}^{-1}$ *could be a problem because* $\Sigma_{j-1,j-1}^{-1}$ *is a large matrix, even after the recursive formula in Remark 1 is used. In such cases, we suggest the strategy briefly described below. First, we notice that the uncorrelated residuals* $e(t_j^*)$, *for* $j \geq 2$, *can also be computed recursively by*

$$e(t_j^*) = \epsilon(t_j^*) - \sum_{k=1}^{j-1} d_k^{-2} \sigma_{jk}^* e(t_k^*),$$

where $d_k^2 = \text{var}(e(t_k^*))$, $\sigma_{jk}^* = \text{cov}(\epsilon(t_j^*), e(t_k^*))$, *and the standardized residuals* $e^*(t_j^*) = e(t_j^*)/d_j$. *In practice, the correlation between* $\epsilon(t_i^*)$ *and* $\epsilon(t_j^*)$ *would become weaker when* t_i^* *and* t_j^* *are farther apart. Therefore, when* j *is large, we can set* σ_{jk}^* *to be* 0 *for those* t_k^* *that are much smaller than* t_j^*. *As a result, we only need to calculate the summation in the above formula over the* $e(t_k^*)$*'s that are relatively close to* t_j^*, *which makes the computation and data storage feasable and convenient.*

As we can see from the above, our proposed CUSUM chart is capable of dealing with any covariance structure. Next, we will show that, when the error term follows an AR(1) model, our proposed CUSUM chart is similar to the second CUSUM chart proposed in Qiu and Xiang (2014). More specifically, under the AR(1) model, the $\epsilon(t_{ij})$ $(j = 1, ..., n_i)$ in the model (1) are assumed to follow

$$\epsilon(t_{ij}) = \phi \epsilon(t_{ij} - \omega) + e(t_{ij}), \tag{6}$$

where ω is the basic time unit defined in Qiu and Xiang (2014), which is the largest time unit that the observation times $\{t_{ij}\}$ are all its integer multiples, $e(t_{ij})$ is a white noise process with mean

0 and variance σ_e^2. Under this AR(1) model, the standardization procedure in (3) is not efficient. Therefore, they proposed another CUSUM chart in order to take into account the AR(1) error structure. Their proposed CUSUM charting statistic is defined by

$$C_j^+ = \max\left[0, C_{j-1}^+ + \left(\epsilon^*(t_j^*) - \phi^{\Delta_{j-1,j}^*}\epsilon^*(t_{j-1}^*)\right) / \sqrt{1-\phi^{2\Delta_{j-1,j}^*}} - k\right], \text{ for } j \geq 2, \tag{7}$$

where $\epsilon^*(t_j^*) = (y(t_j^*) - \mu(t_j^*))/\sigma_y(t_j^*)$, $\sigma_y(t_j^*) = \sqrt{\sigma_e^2/(1-\phi^2)}$, and $\Delta_{j-1,j}^* = (t_j^* - t_{j-1}^*)/\omega$.

Proposition 1 *Under the AR(1) model in (6), $e^*(t_j^*)$ in our proposed CUSUM charting statistic in (5) is equal to $\left(\epsilon^*(t_j^*) - \phi^{\Delta_{j-1,j}^*}\epsilon^*(t_{j-1}^*)\right) / \sqrt{1-\phi^{2\Delta_{j-1,j}^*}}$ in the CUSUM charting statistic in (7), for $j \geq 2$.*

2.2.2 Determining the control limit

As pointed out in Qiu and Xiang (2014), in most longitudinal data studies, the observation times $\{t_j^*, j = 1, 2, ...\}$ may not be equally spaced. Under this situation, the average run length (*ARL*) commonly used in the SPC applications to evaluate the performance of control charts may not be appropriate. Instead, Qiu and Xiang (2014) proposed using the concept of average time to the signal (ATS) when evaluating the performance of control charts on longitudinal data. To calculate the ATS, the observation times t_j^* are expressed in terms of the basic time unit ω, by $n_j^* = t_j^*/\omega$. The ATS is then defined as the expected time to a signal in the basic time unit ω. In this paper, we also follow this idea and use ATS for evaluating the performance of our proposed control chart.

In the following we discuss how to determine the control limit l such that the IC ATS (denoted by ATS_0) of our CUSUM procedure in (5) is controlled at the desired level. From the development of our proposed procedure in Section 2.2.1, we know that $\{e^*(t_j^*), j = 1, 2, ...\}$ in (5) are uncorrelated. If it is reasonable to assume that the error term $\epsilon(\cdot)$ in (1) follows a normal distribution, then $\{e^*(t_j^*), j = 1, 2, ...\}$ are a sequence of i.i.d. standard normal random variables if the observations are in control. Therefore, determining the control limit l in our proposed CUSUM procedure can be achieved by simulating data from the standard normal distribution as $\{e^*(t_j^*), j = 1, 2, ...\}$ and finding l to obtain the desired ATS_0 for any given k through a bi-section search. The bi-section search algorithm runs as follows:

Step 1. For any control limit l, we simulate 10,000 IC sample paths. In each of the IC sample paths, $\{e^*(t_j^*), j = 1, 2, ...\}$ are simulated from the standard normal distribution. The corresponding ATS_0^l for the given l is determined by averaging out the time to a signal in the basic time unit ω from these 10,000 sample paths. Based on this approach, we first find l_1 such that $ATS_0^{l_1} < ATS_0$, and l_2 such that $ATS_0^{l_2} > ATS_0$.

Step 2. Find $ATS_0^{l_3}$ where l_3 is the midpoint of l_1 and l_2.

Step 3. If $ATS_0^{l_3} < ATS_0$, assign $l_1 = l_3$. If $ATS_0^{l_3} > ATS_0$, assign $l_2 = l_3$;

Step 4. Repeat Steps 2 and 3 until $ATS_0^{l_3}$ is sufficiently close to ATS_0;

Step 5. Use l_3 as the control limit.

As mentioned earlier, the observation times in most longitudinal data studies may not be equally spaced. Following Qiu and Xiang (2014), we specify the distribution of the observation times by the sampling rate d, which is defined to be the number of observation times every 10 basic time units here. The calculated control limits l using the above bi-section search algorithm for various values of d and k in our proposed CUSUM procedure (5) with $ATS_0 = 25$ or 50 are given in Table 1.

If the distribution of the error term $\epsilon(t)$ in (1) is unknown but the number of observations for each subject is large, we can still use the above method to determine the control limit, since the standardized residual, $e^*(t_j^*)$, is simply a linear combination of observations and its distribution is asymptotically normal. If the distribution of the error term $\epsilon(t)$ is unknown and the number of observations for each subject is not large, $e^*(t_j^*)$ might not be asymptotically normally distributed. In this case, we can resort to the bootstrap procedure for determining the control limit. More specifically, we assume that the IC dataset consists of m well-functioning subjects. We first use the observations from m_1 ($m_1 < m$) subjects in the IC dataset to obtain the estimates of the mean function and covariance function, $\hat{\mu}(t)$ and $\hat{V}(s,t)$, as discussed in Section 2.1. Based on these $\hat{\mu}(t)$ and $\hat{V}(s,t)$, we apply the standardization procedure described in Section 2.2.1 to the remaining $m - m_1$ subjects in the IC dataset and obtain the standardized residuals $\{e^*(t_{ij}), i = m_1 + 1, m_1 + 2, ..., m, j = 1, ..., n_i\}$.

Those residuals are asymptotically uncorrelated and can be used to approximate the distribution of $e^*(t_j^*)$ in our proposed CUSUM procedure in (5). Therefore, to determine the control limit l to achieve the desired ATS_0 for any given k and d, we can follow the same bi-section search algorithm as the above except that, in each of the IC sample paths, $\{e^*(t_j^*), j = 1, 2, ...\}$ are now drawn from the bootstrap resamples of $\{e^*(t_{ij}), i = m_1 + 1, m_1 + 2, ..., m, j = 1, ..., n_i\}$.

3 Numerical Study

In this section, we present some simulation studies to evaluate the performance of our proposed CUSUM procedure. Our first study considers cases when the IC mean function $\mu(t)$ and the IC covariance function $V(s,t)$ are both known. In particular, we assume that $\mu(t) = \sin(2\pi t)$. For $V(s,t)$, we consider two scenarios. In the first scenario, $V(s,t)$ is the covariance function of $\epsilon(t_{ij})$ generated from the following mixed effect model

$$\epsilon(t_{ij}) = \xi_{0,ij} + \sum_{l=1}^{3} \xi_{l,i}\phi_l(t_{ij}), \tag{8}$$

where $\xi_{0,ij}$ and the $\xi_{l,i}$ are independent random variables from $N(0, 0.3)$, and $\phi_1(t) = t^2 + 0.5$, $\phi_2(t) = \sin(3\pi t)$, $\phi_3(t) = \cos(3\pi t)$. The above mixed effect model was used in the simulation study of Li (2011). In the second scenario, $V(s,t)$ is the covariance function of $\epsilon(t_{ij})$ generated from the following ARMA(2,1) model

$$\epsilon(t_{ij}) = 0.5\epsilon(t_{ij} - \omega) + 0.2\epsilon(t_{ij} - 2\omega) + e(t_{ij}) + 0.2e(t_{ij} - \omega),$$

where ω is the basic time unit, $e(t_{ij})$ is the white noise from $N(0, 0.25)$.

As described earlier, the observation times for each new individual is specified by the sampling rate d. Throughout all the simulation studies presented in this section we choose the basic time unit ω to be 0.01, and d is chosen to be 2, 5, or 10. For our proposed CUSUM procedure (5), we fix the nominal ATS_0 value at 25 or 50, and the allowance constant k is chosen to be 0.1, 0.2 or 0.5.

Using the control limits reported in Table 1, we apply our proposed CUSUM procedure (5) to the new subject's observations generated from the IC mean function $\mu(t)$ and the IC covariance

function $V(s,t)$ described above (the two scenarios for generating $V(s,t)$ are denoted by Mixed Effect Model and ARMA(2,1), respectively), and obtain the time to the signal. This is repeated 1,000 times and the average of the 1,000 times to the signal is the simulated ATS_0 of our proposed CUSUM procedure. We repeat this simulation 100 times. Table 2 shows the average simulated ATS_0 of our proposed CUSUM procedure along with their corresponding standard errors (in the parentheses) under different settings over the 100 replications.

From Table 2, we can see that all the simulated ATS_0 can achieve the nominal level, which indicates that our standardization procedure can successfully decorrelate the data.

As a comparison, we also apply the three CUSUM procedures proposed by Qiu and Xiang (2014) to the new subject's observations generated in the same way as in Table 2. Their first CUSUM procedure assumes independence among the standardized values in (3), the second one assumes an AR(1) model for those standardized values, and the third one does not assume any particular error structure and uses block bootstrap to determine the control limit. When the true covariance structure of within-subject observations follows the Mixed Effect Model, the Qiu and Xiang's first two CUSUM procedures sometimes do not alarm. To make the calculation of ATS still feasible in this situation, we set $ATS = 200$ in those situations. The simulated ATS_0 values for different CUSUM procedures in Qiu and Xiang (2014) are reported in Tables 3-5. All the values are calculated based on 100 replicated simulations. As we can see from those tables, the simulated ATS_0 values for their first two procedures are quite off from the nominal levels. This is not surprising since it has been well demonstrated in the literature that the actual control chart performance would be quite different from the expected performance if the data correlation is not properly accommodated (cf., Kim et al. 2007, Runger and Willemain 1995). This example demonstrates that the first two methods by Qiu and Xiang (2014) may not be reliable to use in cases when the within-subject observations are correlated and the correlation is not properly accommodated, and our proposed method in this paper is reliable in such cases.

As expected, Qiu and Xiang's third CUSUM procedure can achieve the desired nominal ATS_0 as seen in Table 5, since this procedure does not assume any particular error structure. Besides being more computationally intensive than our proposed method, this procedure also requires that

the sampling scheme in Phase I should be exactly the same as the sampling scheme in Phase II. To show the importance of this sampling scheme requirement for Qiu and Xiang's third CUSUM procedure, we carry out another simulation using different sampling schemes for Phase I and Phase II. The sampling rate d for Phase I and Phase II are denoted by $d1$ and $d2$, respectively. The simulated ATS_0 for Qiu and Xiang's block bootstrap CUSUM procedure for different $d1$ and $d2$ are shown in Table 6. As we can see from the results, even when $d1$ and $d2$ are slightly different, the simulated ATS_0 can be quite off from the nominal level. In contrast, our proposed method does not have this sampling scheme requirement, and the sampling scheme in Phase I can be totally different from that in Phase II for our method. This makes our method more broadly applicable in many real applications, since the Phase-I data and Phase-II data in many situations might not follow the same sampling scheme (for example, see our real application in Section 4).

Next, we consider the situation when the IC mean function $\mu(t)$ and the IC covariance function $V(s,t)$ are both unknown and they need to be estimated from an IC dataset. We assume that we have $m = 1,000$ subjects in the IC dataset, their observations are generated from the IC mean function $\mu(t)$ and the IC covariance function $V(s,t)$ described above with the same sampling rate d as that of the new subject for online monitoring. The IC data usually have a fixed time frame. Without loss of generosity, we assume that the domain of the observation times $\{t_{ij}\}$ for the IC data is $[0,1]$. Therefore, the IC mean function and the IC covariance function estimated from the IC data are only appropriate for use within the domain $[0,1]$. As a consequence, when using those estimates in our proposed CUSUM procedure to monitor the observations from a new subject, even if our CUSUM procedure has not signaled when we reach the last observation of the individual within the domain $[0,1]$, we cannot continue to monitor. When this happens, we can only know that the actual time to the signal is greater than $1/\omega$, and we would not know the exact time to the signal. To make the calculation of ATS still feasible in this situation, we let the time to the signal to be $1/\omega$ when our CUSUM procedure has not signaled at the last observation of the individual within the domain $[0,1]$. In other words, the time to the signal in this simulation study is right truncated at $1/\omega$. Using this truncated time to the signal, we also adjust the control limits l in our CUSUM procedure accordingly for various d and k values and they are listed in Table 7.

We apply the estimation procedure described in Section 2.1 to the IC dataset, and obtain the estimated IC mean and IC covariance functions, $\hat{\mu}(t)$ and $\hat{V}(s,t)$. Throughout the simulation study, the bandwidths we use in our estimation procedure are $h = 0.1$ when $d = 2$, $h = 0.05$ when $d = 5$ and $h = 0.02$ when $d = 10$. Using those $\hat{\mu}(t)$ and $\hat{V}(s,t)$ in our proposed CUSUM procedure (5), we can do the online monitoring for the new subject. We first apply this CUSUM procedure using the control limits in Table 7 to the 1,000 new subjects generated from the IC mean function $\mu(t)$ and the IC covariance function $V(s,t)$, and obtain the average time to the signal from these 1,000 new subjects. The average time to the signal is the simulated ATS_0 of our proposed CUSUM procedure from one single IC dataset. We repeat this simulation with 100 different IC datasets. Table 8 shows the average simulated ATS_0 of our proposed CUSUM procedure along with their corresponding standard errors (in the parentheses) under different settings over the 100 IC datasets. As we can see from Table 8, all the simulated ATS_0 are within 10% of the nominal ATS_0 value. This indicates that the estimation procedure described in Section 2.1 provides reliable estimates of the IC mean and covariance functions.

We next apply our CUSUM procedure with $\hat{\mu}(t)$ and $\hat{V}(s,t)$ estimated from the IC dataset to the 1,000 new subjects whose observations are generated from the same IC covariance function $V(s,t)$ but different mean functions. Following Qiu and Xiang (2014), we consider two cases for the mean functions. In the first case, the mean function of the 1,000 new subjects is $\mu_1(t) = \mu(t) + \delta$, which represents a step change of size δ. In the second case, the mean function of the 1,000 new subjects is $\mu_2(t) = \mu(t) + \delta(1 - \exp(-10t))$, which represents a nonlinear drift from $\mu(t)$. In both cases, we choose δ to be 0.25, 0.5, 0.75, and 1. We apply our CUSUM procedure to those 1000 new subjects and obtain their average time to the signal. The average time to the signal is the simulated ATS_1 of our proposed CUSUM procedure from one single IC dataset. We repeat this simulation with 100 different IC datasets. Tables 9 and 10 show the average simulated ATS_1 values of our proposed CUSUM procedure along with their corresponding standard errors (in the parentheses) under different settings over the 100 IC datasets. As can be seen from those tables, as δ increases, the ATS_1 value of our proposed CUSUM procedure decreases quickly, which indicates that our CUSUM procedure has a good detection power.

It is easy to see that the performance of our proposed method depends on the accuracy of the estimated IC mean and IC covariance functions, $\hat{\mu}(t)$ and $\hat{V}(s,t)$. In general, the larger the IC sample size of m, the better those estimates. In our previous simulation setting, we set $m = 1000$. As we can see from the above results, our proposed procedure performs well based on this choice of m. To further investigate how large m has to be in order to ensure reliable performance of our proposed procedure, we run another simulation to evaluate the simulated ATS_0 of our procedure when $m = 500$, and the results are listed in Table 11. As we can see from the table, some of the results are slightly worse than those reported in Table 8, but the majority of them are still within 10% of the nominal ATS_0 value. This makes us to conclude that our proposed procedure can still perform well when $m = 500$.

In Section 2.2.2, when the distribution of the error term is unknown, we propose a bootstrap procedure to determine the control limit. In the following, we report some simulation study to evaluate the performance of our bootstrap procedure. More specifically, for each simulation we first generate an IC data set with $m = 1000$ subjects from the mixed effect model in (8) with $\xi_{0,ij}$ and the $\xi_{l,i}$ being independent random variables from the t distribution with 3 degrees of freedom and with variance 0.3. We then use the data from the first 800 subjects to obtain the estimates of the IC mean and covariance functions, $\hat{\mu}(t)$ and $\hat{V}(s,t)$. Based on these $\hat{\mu}(t)$ and $\hat{V}(s,t)$, we calculate the standardized residuals from the remaining 200 subjects. The control limit is then determined using the bootstrap resamples of those standardized residuals as described in Section 2.2.2. We apply our CUSUM procedure with this determined control limit to the 1000 new subjects generated from the mean function $\mu_1(t)$ or $\mu_2(t)$ (as in the previous simulation study) and the IC covariance functions $V(s,t)$. The average time to the signal from these 1000 new subjects is the simulated ATS. We repeat this simulation 100 times. Table 12 shows the average simulated ATS along with their standard errors (in the parentheses) under different settings over the 100 simulations. When $\delta = 0$ in $\mu_1(t)$ or $\mu_2(t)$, the new 1000 subjects are generated from the IC mean and covariance functions. Therefore the column corresponding to "$\delta = 0$" in Table 12 represents the simulated ATS_0. When $\delta \neq 0$, the values in Table 12 represents the simulated ATS_1. As we can see from the table, all the simulated ATS_0 are close to their nominal values, which indicates the validity of our bootstrap

procedure. When δ increases, the ATS_1 value decreases quickly, which indicates a good detection power of our bootstrap procedure.

4 Real application

In this section we use a data set from the SHARe Framingham Heart Study of the National Heart, Lung and Blood Institute to demonstrate the application of our proposed CUSUM procedure. The data set consists of the systolic blood pressure (mmHg) of 1028 non-stroke patients and 27 stroke patients. The systolic blood pressure of each patient was measured at 7 different times. Figure 1 shows the data from 10 randomly selected non-stroke patients (thin solid line) and 10 randomly selected stroke patients (dashed line). It seems that the stroke patients have higher systolic blood pressure than those non-stroke patients.

To develop a control chart that can monitor a patient's systolic blood pressure, we use the data of the 1028 non-stroke patients as the IC data, and then apply the estimation procedure described in Section 2.1 to the first 800 non-stroke patients to obtain the estimated IC mean function $\hat{\mu}(t)$ and the estimated IC covariance function $\hat{V}(s,t)$. In our estimation, we set the bandwidth to be $h = 0.15$. The estimated IC mean function $\hat{\mu}(t)$ is shown as the dark solid curve in Figure 1. Based on these $\hat{\mu}(t)$ and $\hat{V}(s,t)$, we apply our standardization procedure to the remaining 228 non-stroke patients to obtain the standardized residuals. The control limit is then determined by the bootstrap procedure described in Section 2.2.2. Since each stroke patient has an average of 2.3 observations every 10 years, we choose $d = 2$ and the basic time unit is a year. Following Qiu and Xiang (2014), we use $k = 0.1$ and $ATS_0 = 25$ in our proposed CUSUM procedure. The corresponding control limit l determined from our bootstrap procedure is 0.763. We then apply our CUSUM procedure with this control limit to the 27 stroke patients. The results of our CUSUM procedure for monitoring the 27 stroke patients are presented in Figure 2. As we can see from there, 23 out of the 27 stroke patients are detected to have upward mean shifts, and among those patients that our CUSUM procedure fails to signal, the charting statistics S_j^+ of Patients 18 and 22 are also very close to the control limit. The signal times of the 23 stroke patients are reported in Table 13, and

their average is 9.96 years.

We also consider applying the three CUSUM procedures proposed in Qiu and Xiang (2014) to this systolic blood pressure data set. We first run the Box-Pierce test of independence on the standardized values defined in (3) and the residuals after fitting an AR(1) model on those standardized values, respectively. Both tests yield significant *p*-values, indicating that the standardized values defined in (3) are significantly associated and the AR(1) model cannot describe the association in this data set. Therefore, to accommodate this arbitrary error structure, we can only apply the block bootstrap CUSUM procedure in Qiu and Xiang (2014) to the systolic blood pressure data. The results are presented in Figure 3. The control limit l of their block bootstrap CUSUM procedure is 4.675. As we can see from there, this CUSUM procedure can only detect 10 out of 27 stroke patients. The signal times of the 10 stroke patients are reported in Table 14, and they are much larger than those in Table 13 by our CUSUM procedure.

The reason why Qiu and Xiang's block bootstrap CUSUM procedure does not perform well might be explained by the following. As mentioned in our numerical study in Section 3, to ensure the block bootstrap CUSUM procedure to achieve the desired nominal ATS_0, the sampling scheme in Phase I has to be the same as the sampling scheme in Phase II. To investigate the sampling schemes in both non-stroke patients (Phase I) and stroke patients (Phase II), we plot the histograms of the times (in ages) when the 7 measurements of systolic blood pressure were taken from the non-stroke patients and stroke patients (see Figure 4). The average times when the 7 measurements were taken from the non-stroke patients and stroke patients are reported in Table 15. As seen from those results, the stroke patients tend to be observed at later ages, compared to the non-stroke patients, indicating that the sampling scheme for the non-stroke patients (Phase I) is indeed quite different from that for the stroke patients (Phase II). This might explain why the block bootstrap CUSUM procedure does not work well in this situation.

5 Concluding Remarks

In practice, there are many applications in which our main concern is to detect irregular longitudinal patterns of certain performance variables of a subject as soon as possible, so that some unpleasant consequences can be avoided. To solve this problem effectively, statisticians have developed a new statistical method called dynamic screening system (DySS). This method combines cross-sectional comparison with sequential monitoring, and it tries to make use of all available data about the subject in question, including those observed at the current time and all historical data. However, the current DySS method can only handle cases when observations are normally distributed and within-subject observations are independent or follow a specific time series model. In this paper, a novel nonparametric DySS method have been proposed which does not require restrictive assumptions on the observation distribution, and it allows arbitrary correlation among within-subject observations. A major feature of the method is the use of the Cholesky decomposition in data standardization (cf., Section 2.2.1). After this step, within-subject observations are decorrelated and the decorrelated observations are asymptotically normally distributed. It has been demonstrated that this new method performs well in various different cases.

In this paper, we only consider cases when observations are univariate. In many applications, multivariate performance variables are involved. In multivariate cases, the model for describing the IC data can be defined as follows. Let $\boldsymbol{y}(t_{ij}) = (y_1(t_{ij}), y_2(t_{ij})..., y_q(t_{ij}))^T$ denote the q-dimensional measurement collected at time t_{ij}. Then, the natural multivariate extension of the nonparametric model in (1) is

$$\boldsymbol{y}(t_{ij}) = \boldsymbol{\mu}(t_{ij}) + \boldsymbol{\epsilon}(t_{ij}), \quad i = 1, ..., m,\ j = 1, ..., n_i,$$

where $\boldsymbol{\mu}(\cdot) = (\mu_1(t_{ij}), \mu_2(t_{ij}), ..., \mu_3(t_{ij}))^T$ is the population mean function of $\boldsymbol{y}(t_{ij})$. Estimation of this multivariate model has been discussed in Xiang, Qiu and Pu (2013), and a multivariate EWMA chart has been proposed by Qiu and Xiang (2015) for solving the multivariate DS problem. As in univariate cases discussed by Qiu and Xiang (2014), this multivariate EWMA chart determines its control limit using the block bootstrap procedure. We believe that the proposed method in this paper can be generalized to multivariate cases and the resulting method can improve the performance

of the multivariate EWMA chart by Qiu and Xiang (2015), which will be investigated in our future research.

Acknowledgements

We thank the Department Editor, Professor Judy Jin, and the 3 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 DMS-1405698.

Appendix

Proof of Proposition 1 Instead of directly proving

$$e^*(t_j^*) = \frac{\left(\epsilon^*(t_j^*) - \phi^{\Delta^*_{j-1,j}}\epsilon^*(t_{j-1}^*)\right)}{\sqrt{1-\phi^{2\Delta^*_{j-1,j}}}},$$

we prove that for any $j \geq 2$,

$$\begin{aligned} &\boldsymbol{\sigma}_{j-1,j}^T \Sigma_{j-1,j-1}^{-1} \boldsymbol{\epsilon}_{j-1} = \phi^{\Delta^*_{j-1,j}} \epsilon(t_{j-1}^*), \\ &d_j^2 = \sigma_{jj} - \boldsymbol{\sigma}_{j-1,j}^T \Sigma_{j-1,j-1}^{-1} \boldsymbol{\sigma}_{j-1,j} = \sigma_e^2 (1-\phi^{2\Delta^*_{j-1,j}})/(1-\phi^2). \end{aligned} \tag{9}$$

It is easy to see that if the equations in (9) are true for any $j \geq 2$, our proposition holds for any $j \geq 2$.

We prove (9) by induction. We first show that (9) holds when $j = 2$. Under the AR(1) model in (6), when $j = 2$, we have $\sigma_{11} = \sigma_e^2/(1-\phi^2)$, and $\sigma_{12} = \sigma_e^2 \phi^{\Delta^*_{1,2}}/(1-\phi^2)$. Therefore, $\sigma_{12}\sigma_{11}^{-1}\epsilon(t_1^*) = \phi^{\Delta^*_{1,2}}\epsilon(t_1^*)$. Also, $d_2^2 = \sigma_{22} - \sigma_{21}\sigma_{11}^{-1}\sigma_{12} = \sigma_e^2(1-\phi^{2\Delta^*_{1,2}})/(1-\phi^2)$. Hence, (9) holds for $j = 2$.

Next we show that, given that the two equations in (9) hold for $j = k$, they hold for $j = k+1$. First we notice that

$$\Sigma_{k,k}^{-1} = \begin{pmatrix} \Sigma_{k-1,k-1}^{-1} + \Sigma_{k-1,k-1}^{-1}\boldsymbol{\sigma}_{k-1,k} D_k^{-1} \boldsymbol{\sigma}_{k-1,k}^T \Sigma_{k-1,k-1}^{-1} & -\Sigma_{k-1,k-1}^{-1}\boldsymbol{\sigma}_{k-1,k} D_k^{-1} \\ -D_k^{-1}\boldsymbol{\sigma}_{k-1,k}^T \Sigma_{k-1,k-1}^{-1} & D_k^{-1} \end{pmatrix}, \tag{10}$$

where $D_k = \sigma_{kk} - \boldsymbol{\sigma}_{k-1,k}^T \Sigma_{k-1,k-1}^{-1} \boldsymbol{\sigma}_{k-1,k} = d_k^2$. Also, under the AR(1) model in (6), $\boldsymbol{\sigma}_{k,k+1} = (\phi^{\Delta^*_{k,k+1}} \boldsymbol{\sigma}_{k-1,k}^T, \sigma_{k,k+1})^T$. Then,

$$
\begin{aligned}
&\boldsymbol{\sigma}_{k,k+1}^T \Sigma_{k,k}^{-1} \boldsymbol{\epsilon}_k \\
=&(\phi^{\Delta^*_{k,k+1}} \boldsymbol{\sigma}_{k-1,k}^T, \sigma_{k,k+1}) \begin{pmatrix} \Sigma_{k-1,k-1}^{-1} + d_k^{-2} \Sigma_{k-1,k-1}^{-1} \boldsymbol{\sigma}_{k-1,k} \boldsymbol{\sigma}_{k-1,k}^T \Sigma_{k-1,k-1}^{-1} & -d_k^{-2} \Sigma_{k-1,k-1}^{-1} \boldsymbol{\sigma}_{k-1,k} \\ -d_k^{-2} \boldsymbol{\sigma}_{k-1,k}^T \Sigma_{k-1,k-1}^{-1} & d_k^{-2} \end{pmatrix} \begin{pmatrix} \boldsymbol{\epsilon}_{k-1} \\ \epsilon(t_k^*) \end{pmatrix} \\
=&\left\{ \phi^{\Delta^*_{k,k+1}} \boldsymbol{\sigma}_{k-1,k}^T \left(\Sigma_{k-1,k-1}^{-1} + d_k^{-2} \Sigma_{k-1,k-1}^{-1} \boldsymbol{\sigma}_{k-1,k} \boldsymbol{\sigma}_{k-1,k}^T \Sigma_{k-1,k-1}^{-1} \right) - d_k^{-2} \sigma_{k,k+1} \boldsymbol{\sigma}_{k-1,k}^T \Sigma_{k-1,k-1}^{-1} \right\} \boldsymbol{\epsilon}_{k-1} \\
&+ \left\{ -\phi^{\Delta^*_{k,k+1}} d_k^{-2} \boldsymbol{\sigma}_{k-1,k}^T \Sigma_{k-1,k-1}^{-1} \boldsymbol{\sigma}_{k-1,k} + \sigma_{k,k+1} d_k^{-2} \right\} \epsilon(t_k^*) \qquad (11)
\end{aligned}
$$

Since the two equations in (9) hold for $j = k$, we have

$$
\begin{aligned}
&\boldsymbol{\sigma}_{k-1,k}^T \Sigma_{k-1,k-1}^{-1} \boldsymbol{\epsilon}_{k-1} = \phi^{\Delta^*_{k-1,k}} \epsilon(t_{k-1}^*), \\
&d_k^2 = \sigma_e^2 (1 - \phi^{2\Delta^*_{k-1,k}})/(1 - \phi^2), \\
&\boldsymbol{\sigma}_{k-1,k-1}^T \Sigma_{k-1,k-1}^{-1} \boldsymbol{\sigma}_{k-1,k} = \sigma_e^2 \phi^{2\Delta^*_{k-1,k}}/(1 - \phi^2), \\
&\sigma_{k,k+1} = \sigma_e^2 \phi^{\Delta^*_{k,k+1}}/(1 - \phi^2). \qquad (12)
\end{aligned}
$$

Plugging (12) into (11), we have

$$
\begin{aligned}
&\left\{ \phi^{\Delta^*_{k,k+1}} \boldsymbol{\sigma}_{k-1,k}^T \left(\Sigma_{k-1,k-1}^{-1} + d_k^{-2} \Sigma_{k-1,k-1}^{-1} \boldsymbol{\sigma}_{k-1,k} \boldsymbol{\sigma}_{k-1,k}^T \Sigma_{k-1,k-1}^{-1} \right) - d_k^{-2} \sigma_{k,k+1} \boldsymbol{\sigma}_{k-1,k}^T \Sigma_{k-1,k-1}^{-1} \right\} \boldsymbol{\epsilon}_{k-1} = 0, \\
&\left\{ -\phi^{\Delta^*_{k,k+1}} d_k^{-2} \boldsymbol{\sigma}_{k-1,k}^T \Sigma_{k-1,k-1}^{-1} \boldsymbol{\sigma}_{k-1,k} + \sigma_{k,k+1} d_k^{-2} \right\} \epsilon(t_k^*) = \phi^{\Delta^*_{k,k+1}} \epsilon(t_k^*).
\end{aligned}
$$

Therefore,

$$\boldsymbol{\sigma}_{k,k+1}^T \Sigma_{k,k}^{-1} \boldsymbol{\epsilon}_k = \phi^{\Delta^*_{k,k+1}} \epsilon(t_k^*).$$

Similarly,

$$
\begin{aligned}
&\boldsymbol{\sigma}_{k,k+1}^T \Sigma_{k,k}^{-1} \boldsymbol{\sigma}_{k,k+1} \\
=&(\phi^{\Delta^*_{k,k+1}} \boldsymbol{\sigma}_{k-1,k}^T, \sigma_{k,k+1}) \begin{pmatrix} \Sigma_{k-1}^{-1} + d_k^{-2} \Sigma_{k-1}^{-1} \boldsymbol{\sigma}_{k-1,k} \boldsymbol{\sigma}_{k-1,k}^T \Sigma_{k-1}^{-1} & -d_k^{-2} \Sigma_{k-1}^{-1} \boldsymbol{\sigma}_{k-1,k} \\ -d_k^{-2} \boldsymbol{\sigma}_{k-1,k}^T \Sigma_{k-1}^{-1} & d_k^{-2} \end{pmatrix} \begin{pmatrix} \phi^{\Delta^*_{k,k+1}} \boldsymbol{\sigma}_{k-1,k} \\ \sigma_{k,k+1} \end{pmatrix} \\
=&\phi^{2\Delta^*_{k,k+1}} \boldsymbol{\sigma}_{k-1,k}^T \left(\Sigma_{k-1}^{-1} + d_k^{-2} \Sigma_{k-1}^{-1} \boldsymbol{\sigma}_{k-1,k} \boldsymbol{\sigma}_{k-1,k}^T \Sigma_{k-1}^{-1} \right) \boldsymbol{\sigma}_{k-1,k} - 2\phi^{\Delta^*_{k,k+1}} d_k^{-2} \sigma_{k,k+1} \boldsymbol{\sigma}_{k-1,k}^T \Sigma_{k-1}^{-1} \boldsymbol{\sigma}_{k-1,k} \\
&+ d_k^{-2} \sigma_{k,k+1}^2,
\end{aligned}
$$

where $\Sigma_{k-1} = \Sigma_{k-1,k-1}$. Plugging (12) into the above expression, we have

$$\boldsymbol{\sigma}_{k,k+1}^T \Sigma_{k,k}^{-1} \boldsymbol{\sigma}_{k,k+1} = \sigma_e^2 \phi^{2\Delta^*_{k,k+1}} / (1 - \phi^2).$$

Therefore,

$$d_{k+1}^2 = \sigma_{k+1,k+1} - \boldsymbol{\sigma}_{k,k+1}^T \Sigma_{k,k}^{-1} \boldsymbol{\sigma}_{k,k+1} = \sigma_e^2 (1 - \phi^{2\Delta^*_{k,k+1}}) / (1 - \phi^2).$$

This completes our proof.

References

Chen, K., and Jin, Z. (2005), "Local polynomial regression analysis of clustered data," *Biometrika*, **92**, 59–74.

Hawkins, D.M., and Olwell, D.H. (1998), *Cumulative Sum Charts and Charting for Quality Improvement*, New York: Springer-Verlag.

Kim, S.H., Alexopoulos, C., Tsui, K.L., and Wilson, J.R. (2007), "A distribution-free tabular CUSUM chart for autocorrelated data," *IIE Transactions*, **39**, 317–330.

Li, Y. (2011), "Efficient semiparametric regression for longitudinal data with nonparametric covariance estimation," *Biometrika*, **98**, 355–370.

Liang, K.Y., and Zeger, S.L. (1986), "Longitudinal data analysis using generalized linear models," *Biometrika*, **73**, 13–22.

Ma, S., Yang, L., and Carroll, R. (2012), "A simultaneous confidence band for sparse longitudinal regression," *Statistica Sinica*, **22**, 95–122.

Montgomery, D. C. (2009), *Introduction To Statistical Quality Control (6th edition)*, New York: John Wiley & Sons.

Qiu, P. (2014), *Introduction to Statistical Process Control*, Boca Raton, FL: Chapman & Hall/CRC.

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.

Runger, G.C., and Willemain, T.R. (1995), "Model-based and model-free control of autocorrelated processes," *Journal of Quality Technology*, **27**, 283–292.

Wang, N. (2003), "Marginal nonparametric kernel regression accounting for within-subject correlation," *Biometrika*, **90**, 43–52.

Xiang, D., Qiu, P., and Pu, X. (2013), "Local polynomial regression analysis of multivariate longitudinal data," *Statistica Sinica*, **23**, 769–789.

Zhao, Z., and Wu, W. (2008), "Confidence bands in nonparametric time series regression," *Annals of Statistics*, **36**, 1854–1878.

ACCEPTED MANUSCRIPT

Table 1: Control limits l for different d and k values when observations are normally distributed.

	$ATS_0 = 25$			$ATS_0 = 50$		
k	$d = 2$	$d = 5$	$d = 10$	$d = 2$	$d = 5$	$d = 10$
0.1	0.969	2.031	3.125	1.750	3.125	4.563
0.2	0.828	1.734	2.609	1.500	2.625	3.688
0.5	0.431	1.106	1.625	0.938	1.645	2.227

Table 2: Simulated ATS_0 for different d and k values when $\mu(t)$ and $V(s,t)$ are both known.

	$ATS_0 = 25$					
	Mixed Effect Model			ARMA(2,1)		
k	$d = 2$	$d = 5$	$d = 10$	$d = 2$	$d = 5$	$d = 10$
0.1	24.898(0.074)	25.008(0.063)	24.924(0.077)	24.878(0.073)	25.009(0.071)	25.164(0.077)
0.2	25.341(0.073)	25.072(0.067)	25.073(0.073)	25.337(0.077)	25.126(0.074)	25.252(0.071)
0.5	25.340(0.077)	25.178(0.065)	24.578(0.074)	25.294 (0.078)	25.163(0.067)	24.648(0.080)
	$ATS_0 = 50$					
	Mixed Effect Model			ARMA(2,1)		
k	$d = 2$	$d = 5$	$d = 10$	$d = 2$	$d = 5$	$d = 10$
0.1	50.164(0.152)	49.475(0.139)	49.812(0.131)	49.943(0.145)	49.631(0.141)	50.022(0.134)
0.2	50.248(0.149)	50.210(0.134)	50.102(0.130)	50.021(0.142)	50.560(0.141)	50.471(0.143)
0.5	50.068(0.161)	49.905(0.143)	50.010(0.161)	49.757(0.155)	49.980(0.152)	50.126(0.143)

Table 3: Simulated ATS_0 of Qiu and Xiang's independent CUSUM procedure for different d and k values.

	$ATS_0 = 25$					
	Mixed Effect Model			ARMA(2,1)		
k	$d = 2$	$d = 5$	$d = 10$	$d = 2$	$d = 5$	$d = 10$
0.1	49.453(0.216)	52.089(0.191)	51.404(0.207)	29.023(0.091)	25.876(0.086)	21.269(0.063)
0.2	51.592(0.228)	55.358(0.184)	55.308(0.214)	29.905(0.098)	26.738(0.094)	21.690(0.063)
0.5	54.820(0.232)	65.618(0.226)	67.795(0.217)	31.082(0.107)	30.807(0.111)	24.920(0.080)
	$ATS_0 = 50$					
	Mixed Effect Model			ARMA(2,1)		
k	$d = 2$	$d = 5$	$d = 10$	$d = 2$	$d = 5$	$d = 10$
0.1	74.082(0.231)	67.936(0.224)	62.847(0.224)	49.582(0.157)	38.186(0.137)	29.178(0.089)
0.2	77.019(0.239)	71.428(0.224)	66.355(0.230)	50.590(0.162)	38.665(0.140)	28.741(0.087)
0.5	85.352(0.243)	84.325(0.242)	81.230(0.221)	54.574(0.161)	42.981(0.155)	31.724(0.098)

Table 4: Simulated ATS_0 of Qiu and Xiang's AR(1) CUSUM procedure for different d and k values.

	$ATS_0 = 25$					
	Mixed Effect Model			ARMA(2,1)		
k	$d = 2$	$d = 5$	$d = 10$	$d = 2$	$d = 5$	$d = 10$
0.1	50.344(0.187)	53.601(0.216)	65.479(0.245)	30.446(0.070)	28.220(0.071)	27.377(0.071)
0.2	51.940(0.193)	54.529(0.210)	69.392(0.232)	30.911(0.072)	28.492(0.069)	27.816(0.078)
0.5	53.093(0.199)	52.647(0.171)	53.772(0.238)	30.953(0.071)	29.008(0.075)	27.537(0.079)
	$ATS_0 = 50$					
	Mixed Effect Model			ARMA(2,1)		
k	$d = 2$	$d = 5$	$d = 10$	$d = 2$	$d = 5$	$d = 10$
0.1	80.790(0.225)	82.128(0.235)	88.887(0.273)	55.968(0.139)	52.672(0.123)	52.315(0.126)
0.2	83.145(0.226)	87.235(0.244)	99.761(0.243)	56.275(0.138)	54.080(0.145)	53.602(0.138)
0.5	88.859(0.244)	93.335(0.242)	108.614(0.248)	56.446(0.140)	55.519(0.157)	55.934(0.174)

Table 5: Simulated ATS_0 of Qiu and Xiang's block bootstrap CUSUM procedure for different d and k values.

	$ATS_0 = 25$					
	Mixed Effect Model			ARMA(2,1)		
k	$d = 2$	$d = 5$	$d = 10$	$d = 2$	$d = 5$	$d = 10$
0.1	25.052(0.239)	24.882(0.224)	25.114(0.249)	25.112(0.175)	25.296(0.206)	25.166(0.190)
0.2	25.187(0.227)	24.902(0.206)	24.923(0.264)	25.145(0.184)	25.389(0.193)	25.096(0.170)
0.5	25.633(0.195)	24.975(0.226)	25.145(0.248)	25.172(0.178)	25.244(0.195)	25.066(0.171)
	$ATS_0 = 50$					
	Mixed Effect Model			ARMA(2,1)		
k	$d = 2$	$d = 5$	$d = 10$	$d = 2$	$d = 5$	$d = 10$
0.1	50.457(0.302)	49.674(0.281)	50.039(0.301)	50.132(0.276)	50.372(0.284)	49.985(0.257)
0.2	50.503(0.328)	49.644(0.308)	49.836(0.334)	50.243(0.283)	50.363(0.278)	50.127(0.259)
0.5	50.500(0.307)	49.737(0.328)	49.981(0.301)	49.908(0.288)	50.469(0.280)	49.958(0.289)

Table 6: Simulated ATS_0 of Qiu and Xiang's block bootstrap CUSUM procedure when d differs in Phase I and Phase II.

	$ATS_0 = 25$					
	Mixed Effect Model			ARMA(2,1)		
k	$d1 = 2/d2 = 3$	$d1 = 5/d2 = 6$	$d1 = 5/d2 = 4$	$d1 = 2/d2 = 3$	$d1 = 5/d2 = 6$	$d1 = 5/d2 = 4$
0.1	19.497(0.185)	23.154(0.226)	27.755(0.245)	18.979(0.133)	22.134(0.150)	29.979(0.208)
0.2	19.390(0.180)	23.090(0.226)	27.721(0.245)	19.091(0.125)	22.249(0.159)	29.611(0.220)
0.5	19.717(0.156)	23.134(0.232)	27.719(0.240)	19.418(0.137)	22.575(0.163)	29.369(0.226)
	$ATS_0 = 50$					
	Mixed Effect Model			ARMA(2,1)		
k	$d1 = 2/d2 = 3$	$d1 = 5/d2 = 6$	$d1 = 5/d2 = 4$	$d1 = 2/d2 = 3$	$d1 = 5/d2 = 6$	$d1 = 5/d2 = 4$
0.1	41.461(0.244)	46.961(0.299)	54.452(0.301)	37.704(0.215)	44.161(0.224)	58.640(0.295)
0.2	41.855(0.271)	47.100(0.306)	54.006(0.293)	38.065(0.223)	44.331(0.232)	58.545(0.308)
0.5	42.393(0.299)	47.605(0.340)	53.150(0.322)	39.562(0.219)	44.953(0.237)	57.692(0.302)

Table 7: Control limits l for different d and k values when the time to the signal is truncated at 100.

	$ATS_0 = 25$			$ATS_0 = 50$		
k	$d = 2$	$d = 5$	$d = 10$	$d = 2$	$d = 5$	$d = 10$
0.1	0.991	2.039	3.149	1.938	3.375	4.937
0.2	0.828	1.750	2.633	1.688	2.875	3.984
0.5	0.431	1.109	1.648	1.092	1.820	2.406

Table 8: Simulated ATS_0 for different d and k values when $\mu(t)$ and $V(s,t)$ are both estimated.

	$ATS_0 = 25$					
	Mixed Effect Model			ARMA(2,1)		
k	$d = 2$	$d = 5$	$d = 10$	$d = 2$	$d = 5$	$d = 10$
0.1	26.109(0.089)	25.246(0.079)	25.030(0.077)	24.913(0.086)	24.237(0.071)	24.085(0.079)
0.2	26.053(0.085)	25.498(0.080)	25.199(0.078)	24.991(0.085)	24.836(0.078)	24.161(0.082)
0.5	26.212(0.084)	25.474(0.082)	24.911(0.078)	25.528(0.090)	26.467(0.090)	24.630(0.084)
	$ATS_0 = 50$					
	Mixed Effect Model			ARMA(2,1)		
k	$d = 2$	$d = 5$	$d = 10$	$d = 2$	$d = 5$	$d = 10$
0.1	50.770(0.125)	49.396(0.125)	49.535(0.116)	47.592(0.136)	46.388(0.111)	48.343(0.120)
0.2	51.694(0.126)	50.795(0.133)	49.611(0.116)	48.362(0.138)	47.338(0.109)	47.794(0.120)
0.5	52.421(0.134)	51.269(0.133)	49.575(0.121)	49.580(0.141)	48.623(0.115)	47.012(0.105)

Table 9: Simulated ATS_1 for different d and k values when $\mu_1(t) = \mu(t) + \delta$.

Mixed effect model						
d	ATS_0	k	$\delta = 0.25$	$\delta = 0.5$	$\delta = 0.75$	$\delta = 1$
2	25	0.1	19.355(0.070)	14.159(0.057)	10.422(0.045)	7.866(0.029)
		0.2	19.361(0.070)	14.146(0.056)	10.360(0.044)	7.780(0.029)
		0.5	19.588(0.075)	14.409(0.063)	10.487(0.047)	7.837(0.033)
	50	0.1	39.114(0.120)	28.743(0.102)	20.594(0.081)	14.745(0.065)
		0.2	40.077(0.131)	29.431(0.108)	20.902(0.085)	14.736(0.068)
		0.5	41.525(0.139)	30.960(0.115)	21.937(0.098)	15.234(0.073)
5	25	0.1	19.030(0.069)	14.077(0.057)	10.363(0.039)	7.685(0.031)
		0.2	19.334(0.068)	14.227(0.056)	10.367(0.043)	7.597(0.031)
		0.5	19.607(0.070)	14.582(0.061)	10.571(0.048)	7.591(0.036)
	50	0.1	38.338(0.121)	28.366(0.102)	20.407(0.082)	14.700(0.066)
		0.2	39.825(0.121)	29.630(0.106)	21.161(0.091)	14.883(0.071)
		0.5	41.496(0.133)	31.774(0.123)	22.942(0.101)	15.881(0.083)
10	25	0.1	19.188(0.065)	14.451(0.048)	10.800(0.041)	8.044(0.031)
		0.2	19.430(0.062)	14.593(0.051)	10.777(0.043)	7.913(0.033)
		0.5	19.616(0.063)	14.924(0.053)	11.065(0.048)	7.920(0.039)
	50	0.1	38.895(0.111)	29.082(0.099)	21.140(0.076)	15.400(0.062)
		0.2	39.472(0.109)	29.806(0.095)	21.602(0.082)	15.382(0.070)
		0.5	41.067(0.124)	32.347(0.111)	23.985(0.095)	17.008(0.081)
ARMA(2,1)						
d	ATS_0	k	$\delta = 0.25$	$\delta = 0.5$	$\delta = 0.75$	$\delta = 1$
2	25	0.1	16.934(0.059)	11.930(0.046)	8.724(0.032)	6.680(0.024)
		0.2	16.963(0.060)	11.875(0.044)	8.640(0.031)	6.598(0.024)
		0.5	17.344(0.063)	12.070(0.046)	8.698(0.032)	6.575(0.024)
	50	0.1	31.615(0.094)	21.160(0.070)	14.949(0.048)	11.093(0.036)
		0.2	32.359(0.097)	21.377(0.075)	14.853(0.050)	10.833(0.036)
		0.5	34.051(0.114)	22.382(0.087)	15.035(0.057)	10.550(0.043)
5	25	0.1	16.568(0.048)	11.699(0.032)	8.439(0.023)	6.250(0.019)
		0.2	16.952(0.048)	11.815(0.033)	8.402(0.026)	6.107(0.019)
		0.5	18.211(0.060)	12.468(0.040)	8.642(0.029)	6.069(0.021)
	50	0.1	30.621(0.090)	20.385(0.059)	14.311(0.041)	10.569(0.031)
		0.2	31.659(0.101)	20.856(0.061)	14.335(0.044)	10.316(0.033)
		0.5	33.929(0.106)	22.623(0.078)	14.997(0.055)	10.174(0.038)
10	25	0.1	17.107(0.053)	12.004(0.041)	8.716(0.024)	6.504(0.019)
		0.2	17.288(0.057)	12.063(0.042)	8.610(0.027)	6.278(0.022)
		0.5	17.601(0.063)	12.341(0.048)	8.646(0.035)	6.084(0.023)
	50	0.1	32.806(0.097)	21.437(0.063)	14.882(0.045)	10.947(0.032)
		0.2	33.629(0.102)	21.851(0.068)	14.784(0.047)	10.519(0.034)
		0.5	35.952(0.105)	24.273(0.091)	16.116(0.061)	10.897(0.043)

Table 10: Simulated ATS_1 for different d and k values when $\mu_2(t) = \mu(t) + \delta(1 - \exp(-10t))$.

Mixed effect model						
d	ATS_0	k	$\delta = 0.25$	$\delta = 0.5$	$\delta = 0.75$	$\delta = 1$
2	25	0.1	21.311(0.067)	17.566(0.054)	14.765(0.050)	12.663(0.041)
		0.2	21.326(0.069)	17.570(0.058)	14.761(0.053)	12.621(0.043)
		0.5	21.551(0.074)	17.856(0.066)	14.999(0.054)	12.793(0.046)
	50	0.1	41.225(0.111)	32.904(0.100)	26.425(0.077)	21.641(0.059)
		0.2	42.222(0.125)	33.753(0.103)	26.999(0.080)	21.940(0.063)
		0.5	43.761(0.128)	35.537(0.113)	28.495(0.088)	22.964(0.067)
5	25	0.1	20.470(0.069)	16.724(0.054)	13.846(0.042)	11.723(0.036)
		0.2	20.810(0.066)	16.990(0.054)	14.039(0.044)	11.829(0.035)
		0.5	21.134(0.072)	17.483(0.064)	14.552(0.050)	12.251(0.040)
	50	0.1	39.822(0.117)	31.404(0.095)	24.777(0.079)	19.895(0.060)
		0.2	41.311(0.121)	32.648(0.106)	25.696(0.084)	20.466(0.067)
		0.5	43.040(0.124)	34.990(0.116)	27.911(0.095)	22.175(0.078)
10	25	0.1	20.197(0.064)	16.348(0.050)	13.337(0.040)	11.044(0.032)
		0.2	20.514(0.063)	16.613(0.055)	13.548(0.044)	11.169(0.033)
		0.5	20.785(0.068)	17.228(0.057)	14.268(0.052)	11.867(0.043)
	50	0.1	39.873(0.112)	31.178(0.089)	24.216(0.072)	19.138(0.058)
		0.2	40.321(0.107)	31.711(0.089)	24.619(0.075)	19.302(0.061)
		0.5	42.059(0.125)	34.408(0.108)	27.425(0.089)	21.713(0.074)
ARMA(2,1)						
d	ATS_0	k	$\delta = 0.25$	$\delta = 0.5$	$\delta = 0.75$	$\delta = 1$
2	25	0.1	19.137(0.061)	15.534(0.044)	13.154(0.035)	11.496(0.030)
		0.2	19.198(0.061)	15.537(0.046)	13.117(0.037)	11.422(0.032)
		0.5	19.617(0.066)	15.836(0.049)	13.265(0.042)	11.462(0.035)
	50	0.1	34.452(0.093)	25.787(0.068)	20.588(0.049)	17.352(0.039)
		0.2	35.234(0.093)	26.193(0.070)	20.740(0.052)	17.297(0.039)
		0.5	37.030(0.113)	27.521(0.087)	21.386(0.057)	17.501(0.042)
5	25	0.1	18.515(0.048)	14.899(0.033)	12.574(0.027)	10.898(0.023)
		0.2	18.943(0.050)	15.162(0.037)	12.709(0.029)	10.950(0.024)
		0.5	20.346(0.060)	16.171(0.042)	13.383(0.034)	11.418(0.026)
	50	0.1	33.121(0.090)	24.488(0.057)	19.355(0.043)	16.169(0.031)
		0.2	34.199(0.103)	25.171(0.063)	19.624(0.046)	16.210(0.034)
		0.5	36.534(0.105)	27.246(0.079)	20.999(0.057)	16.944(0.040)
10	25	0.1	18.886(0.056)	14.962(0.042)	12.421(0.035)	10.677(0.029)
		0.2	19.118(0.061)	15.115(0.044)	12.488(0.038)	10.654(0.031)
		0.5	19.380(0.066)	15.519(0.050)	12.840(0.042)	10.929(0.035)
	50	0.1	35.144(0.097)	25.375(0.066)	19.634(0.049)	16.174(0.039)
		0.2	36.037(0.106)	25.937(0.075)	19.805(0.054)	16.055(0.040)
		0.5	38.278(0.111)	28.555(0.091)	21.763(0.068)	17.320(0.050)

Table 11: Simulated ATS_0 for different d and k values when $\mu(t)$ and $V(s,t)$ are both estimated with m=500.

	$ATS_0 = 25$					
	Mixed Effect Model			ARMA(2,1)		
k	$d = 2$	$d = 5$	$d = 10$	$d = 2$	$d = 5$	$d = 10$
0.1	24.698(0.091)	24.969(0.100)	24.708(0.090)	23.179(0.098)	24.113(0.083)	23.857(0.082)
0.2	24.696(0.088)	25.214(0.101)	24.772(0.089)	23.298(0.097)	24.690(0.086)	23.848(0.088)
0.5	24.959(0.090)	25.235(0.113)	24.415(0.099)	23.861(0.098)	26.276(0.084)	24.314(0.090)
	$ATS_0 = 50$					
	Mixed Effect Model			ARMA(2,1)		
k	$d = 2$	$d = 5$	$d = 10$	$d = 2$	$d = 5$	$d = 10$
0.1	48.274(0.125)	48.886(0.145)	49.068(0.151)	44.180(0.142)	46.220(0.133)	47.661(0.137)
0.2	49.172(0.128)	50.229(0.152)	48.900(0.149)	45.032(0.145)	47.173(0.138)	46.991(0.131)
0.5	50.271(0.128)	50.494(0.158)	48.610(0.153)	46.639(0.146)	48.304(0.147)	46.186(0.140)

Table 12: Simulated ATS for different d and k values when $\xi_{0,ij}$ and the $\xi_{l,i}$ in the mixed effect model are from the t distribution.

$\mu_1(t) = \mu(t) + \delta$							
d	ATS_0	k	$\delta = 0$	$\delta = 0.25$	$\delta = 0.5$	$\delta = 0.75$	$\delta = 1$
2	25	0.1	24.560(0.164)	17.197(0.119)	11.828(0.087)	8.338(0.062)	6.241(0.043)
		0.2	25.053(0.171)	17.561(0.130)	11.973(0.094)	8.339(0.066)	6.193(0.042)
		0.5	25.610(0.189)	18.199(0.147)	12.438(0.114)	8.530(0.074)	6.242(0.047)
	50	0.1	47.990(0.266)	34.750(0.272)	23.834(0.233)	16.339(0.180)	11.596(0.123)
		0.2	48.726(0.268)	35.945(0.279)	24.494(0.250)	16.382(0.187)	11.283(0.131)
		0.5	49.944(0.296)	38.861(0.304)	27.407(0.287)	17.906(0.234)	11.508(0.162)
5	25	0.1	24.585(0.158)	17.881(0.126)	12.475(0.104)	8.646(0.080)	6.098(0.063)
		0.2	24.916(0.154)	18.353(0.134)	12.763(0.110)	8.658(0.085)	5.923(0.064)
		0.5	25.498(0.176)	19.654(0.163)	14.017(0.150)	9.417(0.118)	6.120(0.082)
	50	0.1	48.763(0.248)	37.047(0.231)	25.962(0.200)	17.679(0.160)	12.373(0.124)
		0.2	48.871(0.239)	38.171(0.245)	27.137(0.233)	18.135(0.191)	12.166(0.141)
		0.5	49.631(0.240)	41.631(0.252)	32.349(0.273)	22.688(0.263)	14.541(0.210)
10	25	0.1	23.666(0.118)	17.840(0.103)	13.046(0.098)	9.355(0.090)	6.670(0.073)
		0.2	23.634(0.146)	18.160(0.127)	13.352(0.119)	9.434(0.108)	6.476(0.085)
		0.5	24.205(0.144)	19.773(0.135)	15.280(0.132)	11.049(0.127)	7.415(0.110)
	50	0.1	47.662(0.215)	36.868(0.222)	26.450(0.216)	18.560(0.193)	13.271(0.169)
		0.2	47.404(0.211)	38.003(0.215)	28.106(0.218)	19.585(0.212)	13.418(0.181)
		0.5	48.069(0.204)	42.011(0.215)	34.714(0.243)	26.676(0.272)	18.738(0.261)
$\mu_2(t) = \mu(t) + \delta(1 - \exp(-10t))$							
d	ATS_0	k	$\delta = 0$	$\delta = 0.25$	$\delta = 0.5$	$\delta = 0.75$	$\delta = 1$
2	25	0.1	24.560(0.164)	19.242(0.125)	15.442(0.094)	12.812(0.073)	11.006(0.061)
		0.2	25.053(0.171)	19.624(0.134)	15.683(0.102)	12.911(0.079)	11.024(0.065)
		0.5	25.610(0.189)	20.306(0.147)	16.293(0.123)	13.317(0.094)	11.223(0.076)
	50	0.1	47.990(0.266)	37.123(0.256)	28.261(0.225)	22.192(0.188)	18.100(0.156)
		0.2	48.726(0.268)	38.266(0.266)	29.122(0.233)	22.573(0.202)	18.146(0.165)
		0.5	49.944(0.296)	41.147(0.295)	32.300(0.276)	24.881(0.244)	19.337(0.208)
5	25	0.1	24.585(0.158)	19.460(0.126)	15.441(0.106)	12.513(0.085)	10.447(0.066)
		0.2	24.916(0.154)	19.958(0.133)	15.895(0.114)	12.790(0.091)	10.567(0.072)
		0.5	25.498(0.176)	21.241(0.159)	17.375(0.145)	14.159(0.126)	11.590(0.107)
	50	0.1	48.763(0.248)	38.596(0.211)	29.406(0.187)	22.610(0.155)	17.855(0.128)
		0.2	48.871(0.239)	39.628(0.223)	30.497(0.206)	23.290(0.170)	18.159(0.141)
		0.5	49.631(0.240)	43.024(0.246)	35.481(0.234)	28.182(0.229)	21.935(0.205)
10	25	0.1	23.666(0.118)	18.870(0.106)	14.903(0.100)	11.940(0.092)	9.848(0.080)
		0.2	23.634(0.146)	19.210(0.126)	15.338(0.115)	12.292(0.105)	10.027(0.091)
		0.5	24.205(0.144)	20.788(0.132)	17.448(0.118)	14.482(0.112)	11.908(0.109)
	50	0.1	47.662(0.215)	37.870(0.212)	28.797(0.197)	21.911(0.194)	17.023(0.195)
		0.2	47.404(0.211)	38.847(0.202)	30.160(0.190)	22.870(0.188)	17.518(0.189)
		0.5	48.069(0.204)	42.720(0.206)	36.338(0.211)	29.665(0.209)	23.402(0.206)

Table 13: Signal times (years) of the 23 stroke patients by our proposed CUSUM procedure.

Patient ID	Signal time	Patient ID	Signal time
1	16	13	7
2	12	14	8
3	0	16	26
4	23	17	0
5	0	19	8
6	15	20	24
7	0	21	16
8	0	23	0
9	19	24	12
10	0	25	19
11	0	26	12
12	12		

Table 14: Signal times (years) of the 10 stroke patients given by Qiu and Xiang's block bootstrap CUSUM chart.

Patient ID	Signal time	Patient ID	Signal time
3	20	11	8
5	7	12	24
7	24	13	19
8	15	14	13
10	28	23	19

Table 15: The average times (in ages) when each of the seven measurements was taken from the non-stroke patients and stroke patients.

Measurement	1	2	3	4	5	6	7
Non-stroke patients	35.37	43.21	47.62	51.06	54.72	58.75	61.61
Stroke patients	41.89	49.81	53.93	57.59	61.15	65.37	67.93

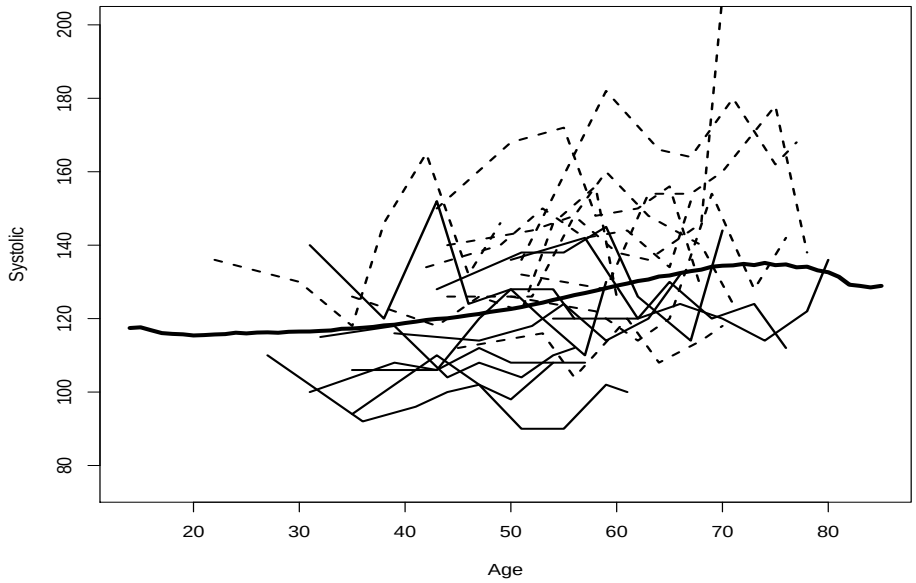

Figure 1: Systolic blood pressure (mmHg) of 10 randomly selected non-stroke patients (thin solid line) and 10 randomly selected stroke patients (dashed line). The dark solid curve is the estimated IC mean function $\hat{\mu}(t)$ using the 1028 non-stroke patients as the IC data.

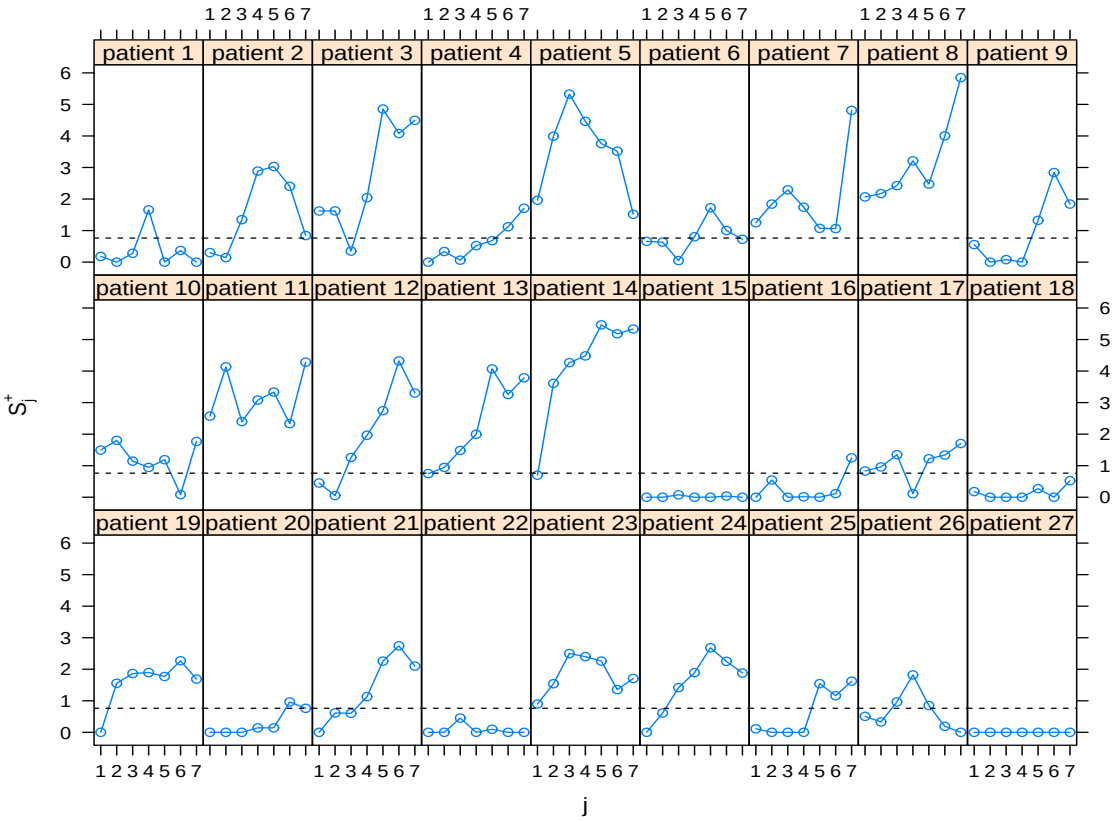

Figure 2: The proposed CUSUM chart for monitoring the 27 stroke patients. The dashed horizontal lines denote the control limit of $l = 0.763$.

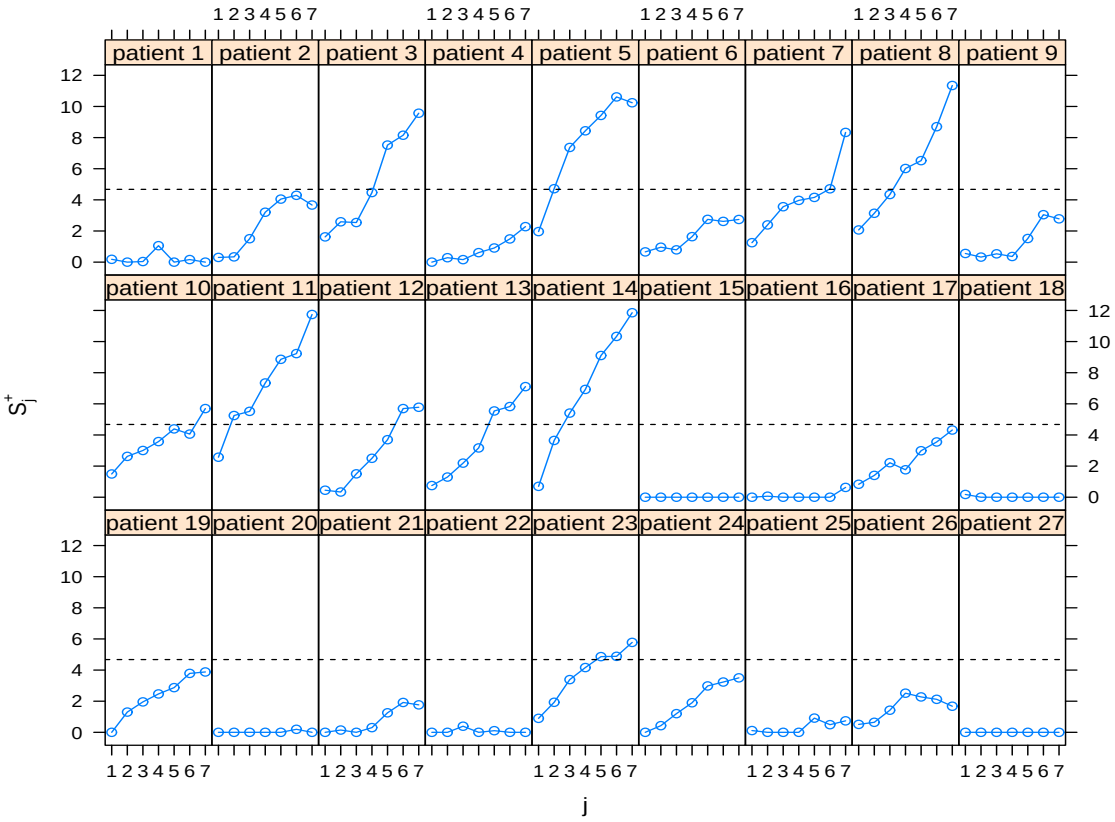

Figure 3: The CUSUM chart for monitoring the 27 stroke patients using the block bootstrap method proposed in Qiu and Xiang (2014). The dashed horizontal lines denote the control limit of $l = 4.675$.

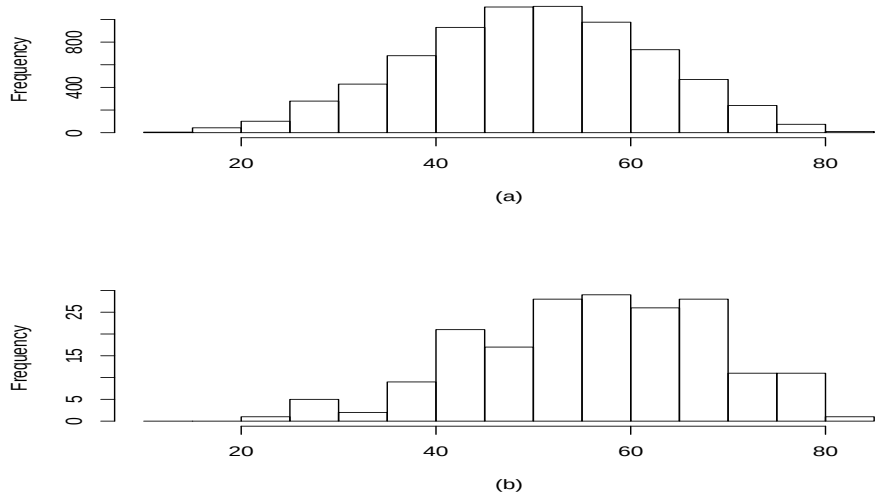

Figure 4: The histogram of the times (in ages) when the systolic blood pressure were taken from (a) the non-stroke patients ; (b) the stroke patients.