A Spatial Rank-Based Multivariate EWMA Control Chart

Changliang Zou,[1] Zhaojun Wang,[1] Fugee Tsung[2]

[1] *LPMC and Department of Statistics, School of Mathematical Sciences, Nankai University, Tianjin, China*

[2] *Department of Industrial Engineering and Logistics Management, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong*

Received 26 January 2011; revised 24 November 2011; accepted 8 December 2011
DOI 10.1002/nav.21475
Published online 10 February 2012 in Wiley Online Library (wileyonlinelibrary.com).

Abstract: Nonparametric control charts are useful in statistical process control when there is a lack of or limited knowledge about the underlying process distribution, especially when the process measurement is multivariate. This article develops a new multivariate self-starting methodology for monitoring location parameters. It is based on adapting the multivariate spatial rank to on-line sequential monitoring. The weighted version of the rank-based test is used to formulate the charting statistic by incorporating the exponentially weighted moving average control scheme. It is robust to non-normally distributed data, easy to construct, fast to compute and also very efficient in detecting multivariate process shifts, especially small or moderate shifts which occur when the process distribution is heavy-tailed or skewed. As it avoids the need for a lengthy data-gathering step before charting and it does not require knowledge of the underlying distribution, the proposed control chart is particularly useful in start-up or short-run situations. A real-data example from white wine production processes shows that it performs quite well. © 2012 Wiley Periodicals, Inc. Naval Research Logistics 59: 91–110, 2012

Keywords: distribution-free; nonparametric procedure; self-starting; spatial rank; multivariate EWMA; robustness; statistical process control

1. INTRODUCTION

In modern quality control, it is common to monitor several quality characteristics of a process simultaneously (Stoumbos et al. [30]). This is called multivariate statistical process control (MSPC) in the literature and it is the focus of this article. One of the tasks of MSPC is to detect the change in a multivariate process location vector of parameters $\boldsymbol{\theta}$ as quickly as possible. It is usually assumed that there are m_0 independent and identically distributed (i.i.d.) historical (reference) observations, $\mathbf{x}_{-m_0+1}, \ldots, \mathbf{x}_0 \in \mathbb{R}^p$, for some integer, $p \geq 1$, and the ith future observation, $\mathbf{x}_i$, is collected over time using the following multivariate location change-point model

$$\mathbf{x}_i \overset{\text{i.i.d.}}{\sim} \begin{cases} \boldsymbol{\mu}_0 + \boldsymbol{\Omega}\boldsymbol{\varepsilon}_i, & \text{for} \quad i = -m_0+1, \ldots, 0, 1, \ldots, \tau, \\ \boldsymbol{\mu}_1 + \boldsymbol{\Omega}\boldsymbol{\varepsilon}_i, & \text{for} \quad i = \tau+1, \ldots, \end{cases} \tag{1}$$

where τ is the unknown change point, $\boldsymbol{\mu}_0 \neq \boldsymbol{\mu}_1$, the p-vectors $\boldsymbol{\varepsilon}_i$ are independent, standardized and centered residuals, all having the same unknown distribution, $\boldsymbol{\Omega}$ is a full-rank $p \times p$ transformation matrix and $\boldsymbol{\Sigma} = \boldsymbol{\Omega}\boldsymbol{\Omega}^T > 0$ is a scatter matrix. In traditional parametric settings, it is usual to say that $\boldsymbol{\varepsilon}_i$ is "standardized" if $\text{cov}(\boldsymbol{\varepsilon}_i) = \mathbf{I}_p$, and it is "centered" if $E(\boldsymbol{\varepsilon}_i) = 0$.

Methods for accomplishing the monitoring task are usually based on the following quadratic formulation of the test statistics:

$$(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_0)^T \widehat{\boldsymbol{\Sigma}}_0^{-1} (\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_0), \tag{2}$$

where $\widehat{\boldsymbol{\mu}}_0$ and $\widehat{\boldsymbol{\Sigma}}_0$ are, respectively the mean vector and covariance matrix estimated from the IC reference sample of size m_0. It is often called a Shewhart χ^2 chart when we use the exact $(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ instead of $(\widehat{\boldsymbol{\mu}}_0, \widehat{\boldsymbol{\Sigma}}_0)$. In the literature, to accumulate information from past observations, many MSPC control charts are constructed in two steps. First, a sequence of multivariate vectors is constructed in the framework of a cumulative sum (CUSUM) or an exponentially weighted moving average (EWMA). Then, the charting statistic takes the quadratic form of the multivariate vectors in a similar way to (2), resulting in the multivariate CUSUM (Croisier [5]), the multivariate EWMA (MEWMA; Lowry et al. [15]) and the

Correspondence to: Fugee Tsung (season@ust.hk)

formulas of multivariate chart with change-point detection (Zamba and Hawkins [34]).

The above mentioned MSPC research is mostly based on a fundamental assumption that the process data have multinormal distributions. However, it is well recognized that in many applications, the underlying process distribution is unknown and not multinormal, so that the statistical properties of commonly used charts, which were designed to perform best under the normal distribution, could potentially be (highly) affected. Nonparametric or robust charts may be useful in such situations. In the last several years, univariate nonparametric control charts have attracted much attention from researchers and a nice overview of this topic was presented by Chakraborti et al. [2]. See Zou and Tsung [38], Hawkins and Deng [6] and the references therein for some recent development. Some effort has been devoted to robust multivariate SPC: Liu [14] proposed control schemes based on data-depth; Qiu and Hawkins [23, 24] suggested a computationally trivial nonparametric multivariate CUSUM procedure based on the antiranks of the measurement components. Stoumbos and Sullivan [31] recommended the classical MEWMA chart because it is robust in the sense that the in-control (IC) run-length distribution for a continuous non-normal process is quite close to the distribution for a multinormal process with the same control limit if the weighting parameter, λ, is small; finally Qiu [22] proposed a distribution-free multivariate CUSUM procedure based on log-linear modeling.

Alternatively, Zou and Tsung [39] developed a multivariate sign EWMA (MSEWMA) control chart for monitoring location parameters. The MSEWMA adapts the multivariate sign test proposed by Randles [28] to on-line sequential monitoring by incorporating the EWMA scheme, which results in a nonparametric counterpart of the MEWMA chart. It is easy to implement because only the multivariate median and the transformation matrix need to be specified from the reference dataset before monitoring just like in the MEWMA chart where only $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ need to be estimated rather than the entire distribution. As shown by Zou and Tsung [39], the MSEWMA is robust in attaining the IC average run length (ARL) and is also very efficient in detecting process shifts, especially small or moderate shifts which occur when the process distribution is heavy-tailed or skewed.

However, in many applications, m_0 is not large and in some cases it is in fact rather small. For example, Zou and Tsung [39] applied the MSEWMA chart to two real datasets: one is from an aluminium electrolytic capacitor manufacturing process; the other is the aluminum smelter example used by Qiu and Hawkins [23]. The number of IC historical observations used for calibrating the necessary parameters are only 170 and 95 for these two examples, respectively. In such situations, there would be considerable uncertainty in the parameter estimation, which in turn would distort the IC run-length distribution of the MSEWMA control chart. For instance, for the traditional univariate EWMA chart with $\lambda = 0.2$, 1500 observations are needed to achieve the desired level of IC performance (Jones [13]). Even if the control limit of the chart was adjusted properly to obtain the desired IC run-length behavior, its out-of-control (OC) run-length would still be severely compromised (cf., Jones [13]). This is essentially analogous to the estimated parameters problem in the context of parametric control charts (see Jensen et al. [12] for an overview). Zou and Tsung [39] used simulated examples to show that the performances of MSEWMA and MEWMA are similarly affected when m_0 is small and very large Phase I samples must be collected for both charts to perform as well as those with known parameters (at least a reference sample of size 4,000 is required to make MSEWMA or MEWMA attain the nominal IC ARL for a five-dimensional case like the aluminum smelter example). However, gathering such large reference samples may be costly in practice: it may require time consuming Phase-I analysis; many OC productions may be collected in such data-gathering processes; the engineers may spend lots of time to identify which productions are "good" based on their prior engineering knowledge and experience (see the white wine example given in Section 4). Besides, in many cases, it may not be feasible to wait for the accumulation of sufficiently large calibration samples because users usually want to monitor the process at the start-up stages.

In the situation where a sufficiently large reference dataset is unavailable, self-starting methods that handle sequential monitoring by simultaneously updating parameter estimates and checking for OC conditions would be of use (see, e.g., Quesenberry [25, 26], Zantek and Nestler [35]). The basic idea is to construct the so-called Q-statistics using some transformation so that they are i.i.d. with known probability distribution when the process is IC. The self-starting approach is also more appealing when many characteristics have to be controlled simultaneously, where even larger Phase I samples than those required in univariate monitoring settings must be collected to reduce the variability in the parameter estimates (c.f., Champ et al. [3]). The multivariate self-starting control charts have been developed accordingly, see Quesenberry [27], Sullivan and Jones [32], and Hawkins and Maboudou-Tchao [7]. However, they are all based on the multivariate normality assumption and thus they may not be robust and effective for heavy-tailed or skewed process distributions. In particular, Hawkins et al. [9] and Zamba and Hawkins [34] proposed approaches based on change-point models for on-line monitoring that can also be seen as self-starting methods. The key idea embodied in the these approaches is to utilize dynamic control limits which can be found by simulating a conditional probability expression. This method has been successfully extended by Zhou et al. [36], Zou and Tsung [38] and Hawkins and Deng [6] to univariate nonparametric control charts.

In this article, motivated by Zou and Tsung's [39] use of spatial sign and the self-starting control schemes, we develop a new multivariate self-starting methodology for monitoring location parameters. Instead of spatial sign, we use spatial ranks which will be shown to be more appropriate in a self-starting monitoring setting. An EWMA control chart is suggested. It has distribution-free properties over a broad class of population models in the sense that the IC run-length distribution is (or is always very close to) the nominal one when the same control limit designed for a multinormal distribution is used. As it avoids the need for a lengthy data-gathering step before charting (although it is generally necessary and advisable to collect a few preliminary stable observations by a Phase-I analysis) and it does not require knowledge of the underlying distribution, the proposed control chart is particularly useful in start-up or short-run situations. The remainder of this article is organized as follows: our proposed methodology is described in detail in Section 2; its numerical performance is thoroughly investigated in Section 3; we demonstrate the method in Section 4 using a real-data example from wine production processes; finally several remarks in Section 5 conclude the article. Some technical details are provided in the Appendix.

2. METHODOLOGY

Our proposed methodology is described in five parts. In Section 2.1, a brief introduction to the concept of spatial signs and ranks and the construction of the MSEWMA control chart is presented. In Sections 2.2–2.3, a self-starting multivariate nonparametric charting scheme combined with spatial rank tests is derived. Its computational issues are addressed in Section 2.4. Some practical guidelines for choosing the control limits are discussed in Section 2.5.

2.1. A Review of Spatial Signs, Ranks and MSEWMA

As is well known, the definitions of univariate signs and ranks are based on the ordering of the data. However, a natural ordering of the data points does not exist in the multivariate case. The multivariate concepts of a spatial sign and a spatial rank have been developed accordingly in the literature, see a recent book Oja [20] for a comprehensive introduction. Some key points are given in the following. In one dimension, the sign of an observation is basically its direction ($+1$ or -1) from the origin. In higher dimensions, in this spirit, the spatial sign function is defined as

$$U(\mathbf{x}) = \begin{cases} ||\mathbf{x}||^{-1}\mathbf{x}, & \mathbf{x} \neq \mathbf{0}, \\ \mathbf{0}, & \mathbf{x} = \mathbf{0}, \end{cases}$$

where $||\mathbf{x}|| = (\mathbf{x}^T\mathbf{x})^{1/2}$ is the Euclidean length of the vector $\mathbf{x}$. The function value is just a direction (a point in the unit p-sphere) whenever $\mathbf{x} \neq \mathbf{0}$. Applying the spatial sign function to the empirical distribution given by data points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ produces the so-called spatial sign vectors $\mathbf{u}_i = U(\mathbf{x}_i)$. Clearly, in the univariate case, it reduces to the classical sign statistic $\mathrm{sgn}(x_i)$, where $\mathrm{sgn}(\cdot)$ is the sign function.

Furthermore, in one dimension, the (observed) rank of $\mathbf{x}_i$ can be expressed as $R_i = \sum_{j=1}^{n} I(x_i \geq x_j)$ and the so-called centered rank is

$$R_i^* \equiv \frac{2}{n}\left[R_i - \frac{n+1}{2}\right] = \frac{2}{n}\sum_{j\neq i}^{n} I(x_i \geq x_j) - 1$$
$$= \frac{1}{n}\sum_{j=1}^{n} \mathrm{sgn}(x_i - x_j).$$

The spatial rank for the vector observation $\mathbf{x}_i$ can be defined in a similar manner, say the average of the spatial signs of pairwise differences,

$$\mathbf{r}_i = R_E(\mathbf{x}_i) = \frac{1}{n}\sum_{j=1}^{n} U(\mathbf{x}_i - \mathbf{x}_j),$$

where $R_E(\cdot)$ denotes the empirical spatial rank function. Intuitively speaking, $\mathbf{r}_i$ reflects the relative magnitudes of $\mathbf{x}_i$ in the data.

In the literature, the theoretical (population) spatial rank function of the vector $\mathbf{x}$ (with respect to distribution F) is defined accordingly by replacing the average with expectation in $R_E(\cdot)$, say

$$R_F(\mathbf{x}) = E_{\mathbf{y}}[U(\mathbf{x} - \mathbf{y})],$$

where $\mathbf{y}$ is distributed according to F. The notation $E_{\mathbf{y}}[\cdot]$ means that the expectation is taken with respect to the random vector $\mathbf{y}$. $R_F(\mathbf{x}_i)$ can also be regarded as the asymptotic version of $R_E(\mathbf{x}_i)$ ($n \to \infty$) only if $\mathbf{x}_j$'s are distributed as F. For a spherical distribution F, it can be shown that $R_F(\mathbf{x}) = q_F(r)\mathbf{u}$ (Theorem 4.3 in Oja [20]; see also (A.1) in the Appendix), where $r = ||\mathbf{x}||$, $\mathbf{u} = U(\mathbf{x})$, and $q_F(r)$ is a scalar function depending on F. Because $q_F(r)$ is a scalar, the direction of $R_F(\mathbf{x})$ is just $\mathbf{u}$ and thus gives the direction of $\mathbf{x}$ from the center of F (or the data cloud in the sample version). As $||\mathbf{u}|| = 1$, $q_F(r)$ can be regarded as the length of $R_F(\mathbf{x})$, describing the distance between this point and the center. Moreover, by noting

$$|q_F(r)|^2 = |q_F(r)|^2\mathbf{u}^T\mathbf{u} = ||R_F(\mathbf{x})||^2$$
$$\leq E_{\mathbf{y}}[U^T(\mathbf{x} - \mathbf{y})U(\mathbf{x} - \mathbf{y})] = 1,$$

we know $|q_F(r)| \leq 1$. Thus, the spatial rank lies in a unit p-sphere. For other distributions, the decomposition $R_F(\mathbf{x}) = q_F(r)\mathbf{u}$ may not hold, but the above geometrical illustration is essentially correct (Möttönen and Oja [18]). For

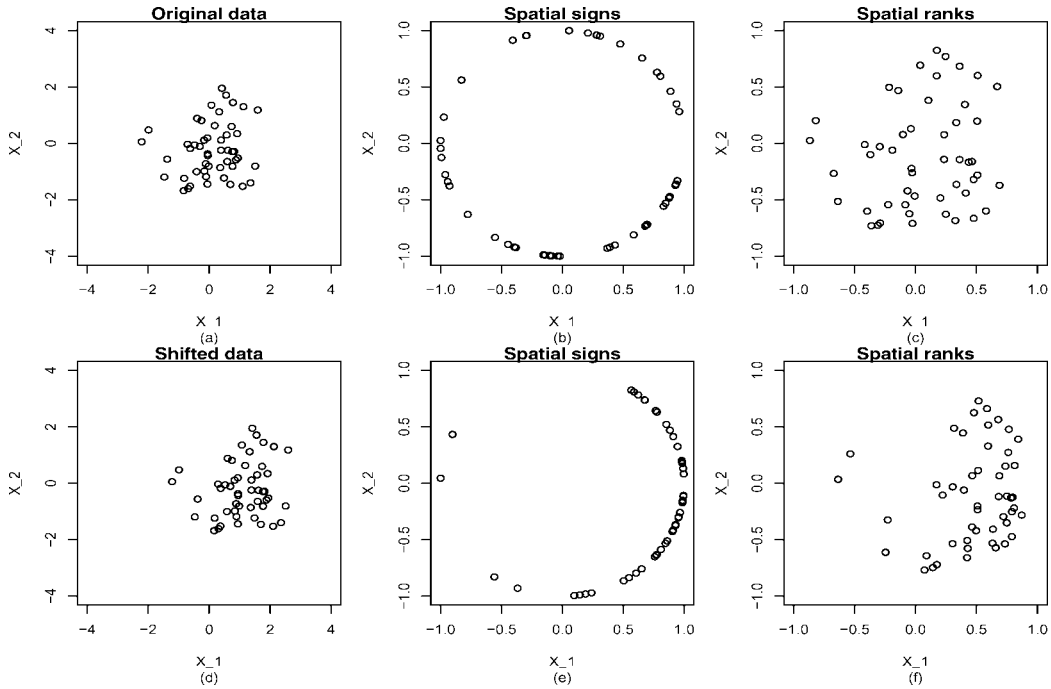

Figure 1. (a–c) The scatterplot for a random sample of size 50 from $N_2(0, \mathbf{I}_2)$ with scatterplots for corresponding spatial signs and ranks; (d–f) The scatterplot for the data in (a) with shifted first component (shifted by 1) with scatterplots for corresponding spatial signs and ranks.

illustration, Figs. 1a–c show the scatterplot for 50 bivariate observations from $N_2(0, \mathbf{I}_2)$ with the corresponding bivariate spatial signs and (theoretical) ranks. The data points are then shifted by 1 in the first component, and the corresponding behavior of the data, spatial signs and theoretical spatial ranks with F being still $N_2(0, \mathbf{I}_2)$ are shown in Figs. 1d–f, respectively.

The spatial signs and ranks have been used for testing and estimation in various multivariate location and scatter problems. A nice overview of this topic and related references have been prepared by Oja and Randles [19] and again Oja [20]. Similar to their univariate counterparts, the spatial signs or ranks based methods have been shown to be quite robust for various distributions since those methods use the direction of observations from the origin and/or the relative magnitudes rather than the original magnitudes of observations. Zou and Tsung [39] proposed an affine-invariant EWMA control chart, a MSEWMA chart, which adapts one of the most powerful and robust multivariate sign tests proposed by Randles [28] to tackle the on-line monitoring problem assuming m_0 is sufficiently large or equivalently the distribution of observations is known. Instead of assuming $\mathrm{cov}(\boldsymbol{\varepsilon}_i) = \mathbf{I}_p$ and $E(\boldsymbol{\varepsilon}_i) = 0$ in constructing the multivariate sign test, one reformulates the location model (1) through $\mathbf{x}_i = \boldsymbol{\theta} + \mathbf{T}^{-1}\boldsymbol{\varepsilon}_i$ in which $\boldsymbol{\varepsilon}_i$ is "standardized" if $\mathrm{cov}(U(\mathbf{T}(\mathbf{x}-\boldsymbol{\theta}))) = p^{-1}\mathbf{I}_p$, and it is "centered" if $E[U(\mathbf{T}(\mathbf{x}-\boldsymbol{\theta}))] = 0$. Note that $\mathrm{cov}(U(\boldsymbol{\varepsilon}_i)) = p^{-1}\mathbf{I}_p$ and $E[U(\boldsymbol{\varepsilon}_i)] = 0$ are roughly (from the viewpoint of matching the first two moments) saying that the direction of $\boldsymbol{\varepsilon}_i$ is spherically distributed (this is called spherical direction family; see Chapter 2 in Oja [20]), while the requirement of $\mathrm{cov}(\boldsymbol{\varepsilon}_i) = \mathbf{I}_p$ and $E[\boldsymbol{\varepsilon}_i] = 0$ means (not strictly) that the error $\boldsymbol{\varepsilon}_i$ is spherically distributed. Spherical direction family is a broader class of population than spherical distribution and thus the test which is developed and applicable under this distributional assumption is generally more robust than traditional tests designed for spherical error distributions (Randles [28]).

The basic idea of spatial sign based test is: under null hypothesis and mild conditions, the expectation of the (transformed) sign vector equals (or approximates) to zero, whereas it will be away from zero under alternatives (See Figs. 1b and 1e). To this end, in a MSEWMA chart, first we find the solution to the sample equations $(\widehat{\boldsymbol{\theta}}_0, \widehat{\mathbf{T}}_0)$ based on m_0 IC historical observations,

$$\frac{1}{m_0}\sum_{i=-m_0+1}^{0}\left(\frac{\mathbf{T}(\mathbf{x}_i-\boldsymbol{\theta})}{||\mathbf{T}(\mathbf{x}_i-\boldsymbol{\theta})||}\right)=0,$$

$$\frac{1}{m_0}\sum_{i=-m_0+1}^{0}\left(\frac{\mathbf{T}(\mathbf{x}_i-\boldsymbol{\theta})(\mathbf{x}_i-\boldsymbol{\theta})^T\mathbf{T}^T}{||\mathbf{T}(\mathbf{x}_i-\boldsymbol{\theta})||^2}\right)=\frac{1}{p}\mathbf{I}_p,$$

where $\mathbf{T}$ is a $p \times p$ upper triangular positive-definite matrix with a one in the upper left-hand element (see Tyler [33] or Randles [28] for detailed discussion). By supposing that m_0

is large enough, we may conveniently use an asymptotic version of $(\widehat{\boldsymbol{\theta}}_0, \widehat{\mathbf{T}}_0)$, say $(\boldsymbol{\theta}_0, \mathbf{T}_0)$. Then, we transform $\mathbf{x}_t$ to obtain the unit vector $\mathbf{v}_t$ through

$$\mathbf{v}_t = U(\mathbf{T}_0(\mathbf{x}_t - \boldsymbol{\theta}_0)).$$

Next, we define an EWMA sequence

$$\mathbf{w}_t = (1-\lambda)\mathbf{w}_{t-1} + \lambda \mathbf{v}_t,$$

where the initial vector, $\mathbf{w}_0$, is usually taken to be zero. Finally, the MSEWMA chart triggers a signal if

$$Q_t = \frac{2-\lambda}{\lambda} p \mathbf{w}_t^T \mathbf{w}_t > L,$$

where $L > 0$ is a control limit chosen to achieve a specific IC ARL.

Similar to the classical MEWMA chart, the MSEWMA requires estimating the IC parameters $(\widehat{\boldsymbol{\theta}}_0, \widehat{\mathbf{T}}_0)$ from historical sample of size m_0. As mentioned in the introduction, an estimation problem in using MSEWMA with insufficient reference data (m_0 is small) would result in rather unsatisfactory run-length performance. Following a similar idea to self-starting schemes, we could consider replacing the estimators $(\widehat{\boldsymbol{\theta}}_0, \widehat{\mathbf{T}}_0)$ with an updating version. For example, we could use $m + t - 1$ samples to estimate the parameters up to time point t. However, this seems computationally infeasible for MSEWMA because estimating $(\boldsymbol{\theta}_0, \mathbf{T}_0)$ involves a complicated iterative routine. Although computing power has improved dramatically and it is computationally trivial to perform a one-time estimation from the historical data by using some efficient algorithms (c.f., Hettmansperger and Randles [10]), for on-line process monitoring, repeated iterative procedures require a considerable amount of computing time at each sample point and worse still the computational effort grows linearly with t. Thus, such a method is obviously undesirable in practice. Note that the spatial-rank is automatically centered (see Section 2.2). There is no need to estimate the in-control location parameter by using spatial-ranks which would greatly facilitate the construction of control charts. Therefore, rather than spatial signs, we are interested in tackling problem (1) with spatial ranks instead, which well serves the self-starting purpose as we will detailly show in the next section.

2.2. A Theoretical Spatial Rank-Based EWMA Control Chart

Recall model (1) and the associated notation. In what follows, we elaborate on the individual observation model. The extension to the group case is presented later. Problem (1) is closely related to nonparametric statistical tests of hypotheses for the one-sample or multiple-sample location problems. Hence, to facilitate the derivation of the proposed charting statistic, we start by assuming that the underlying IC distribution $F(\mathbf{x}-\boldsymbol{\mu}_0)$ is completely known (equivalently m_0 is large enough), where $F(\cdot)$ represents a continuous p-dimensional distribution "located" at the vector $\boldsymbol{\mu}$.

Given a random observed vector $\mathbf{x} \sim F(\mathbf{x} - \boldsymbol{\mu})$, we want to test the null hypothesis, H_0, that $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ against the alternative hypothesis H_1 that $\boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. By definition, it is easy to see that under H_0,

$$\begin{aligned} E_{\mathbf{x}}E_{\mathbf{y}}[U(\mathbf{x}-\mathbf{y})] &= -E_{\mathbf{x}}E_{\mathbf{y}}[U(\mathbf{y}-\mathbf{x})] = -E_{\mathbf{y}}E_{\mathbf{x}}[U(\mathbf{y}-\mathbf{x})] \\ &= -E_{\mathbf{x}}E_{\mathbf{y}}[U(\mathbf{x}-\mathbf{y})], \end{aligned}$$

where the last equality comes from the fact that $\mathbf{x}$ and $\mathbf{y}$ are identically distributed (i.e., $F(\cdot - \boldsymbol{\mu}_0)$). Consequently, $E_{\mathbf{x}}[R_F(\mathbf{x})] = 0$ under H_0. However, under H_1 in which the location parameter of $\mathbf{x}$ is not $\boldsymbol{\mu}_0$, $E_{\mathbf{x}}[R_F(\mathbf{x})]$ is generally not zero (see Chapter 4 in Oja [20] for some discussion). Refer to Figs. 1c and 1f for graphical illustrations. Thus, the test statistic

$$R_F^T(\mathbf{x})\{\mathrm{cov}[R_F(\mathbf{x})]\}^{-1}R_F(\mathbf{x})$$

is a reasonable candidate for testing. When H_0 is true, this test statistic should be small. A large value leads to rejecting the null hypothesis. It can be easily verified that the test statistic is invariant under orthogonal transformation under H_0 in the sense that its value stays the same if we transform the data $\mathbf{x}$ to $\boldsymbol{\Gamma}\mathbf{x}$ for any orthogonal matrix $\boldsymbol{\Gamma}$. Unlike the classical Hotelling's T-square test statistic, it is unfortunately not affine invariant (if its value stays the same if the data $\mathbf{x}$ is transformed to $\mathbf{Dx}$ for any full-rank matrix $\mathbf{D}$ we call it an affine invariant test). This is rather unappealing because, for example, the P-value would depend on the chosen coordinate system.

An affine-invariant modification of the test can be achieved through

$$Q^{R_F} = R_F^T(\mathbf{Mx})\{\mathrm{cov}[R_F(\mathbf{Mx})]\}^{-1}R_F(\mathbf{Mx}), \qquad (3)$$

where $\mathbf{S} = (\mathbf{M}^T\mathbf{M})^{-1}$ is any scatter matrix (a symmetric $p \times p$ matrix $\mathbf{S}_{\mathbf{y}} \geq 0$ is a scatter matrix if it is affine equivariant in the sense that $\mathbf{S}_{\mathbf{Ax+b}} = \mathbf{A}\mathbf{S}_{\mathbf{x}}\mathbf{A}^T$ for all random vectors $\mathbf{x}$, all full-rank $p \times p$-matrices $\mathbf{A}$ and all p-vectors $\mathbf{b}$). The classical scatter matrix, namely the the covariance matrix $\mathrm{cov}(\mathbf{x}) = E[(\mathbf{x} - E(\mathbf{x}))(\mathbf{x} - E(\mathbf{x}))^T]$, say $\boldsymbol{\Sigma}$ in model (1), serves as the simplest one. In other words, $\mathbf{M} = \boldsymbol{\Omega}^{-1}$. We will use this simple transformation to construct our charting scheme. The transformation matrix $\mathbf{T}$ used in the MSEWMA scheme, which is known as "Tyler's transformation or inner standardization" (Tyler [33], Oja [20]), is another one of the most popular affine-equivalent transformations. However, as mentioned earlier, obtaining this type of transformation

requires iterative procedures and thus is not considered in our proposal.

A naive method that comes to mind for on-line monitoring is to use the current individual observation to construct the Q^{R_F} in (3). However, this would be very inefficient in the case of moderate or small changes since it completely ignores the past observations. As an alternative, we consider the EWMA-type control chart. Define an EWMA sequence

$$\mathbf{w}_t = (1-\lambda)\mathbf{w}_{t-1} + \lambda R_F(\mathbf{M}\mathbf{x}_t),$$

where $\mathbf{w}_0 = 0$. The charting statistic is given by

$$Q_t^{R_F} = \frac{2-\lambda}{\lambda}\mathbf{w}_t^T\{\mathrm{cov}[R_F(\mathbf{M}\mathbf{x})]\}^{-1}\mathbf{w}_t,$$

in which we use the fact that $\mathrm{cov}(\mathbf{w}_t) \approx \lambda\mathrm{cov}[R_F(\mathbf{M}\mathbf{x})]/(2-\lambda)$. We call this method the theoretical rank-based EWMA (TREWMA) control chart. After some derivations (please refer to the Appendix for the proofs), we have the following proposition which says that if the data points are rotated, or if they are reflected around a $p-1$ dimensional hyperplane, or if the scales of measurement are altered, the value of the charting statistic stays the same.

PROPOSITION 1: The IC run-length distribution of the TREWMA chart is affine and shift invariant.

The invariant property here is in the sense that for any $p \times p$ nonsingular matrix $\mathbf{D}$ and constant vector $\mathbf{b}$, the run-length distribution of the TREWMA chart stays the same if the IC observations are distributed as $\mathbf{D}\mathbf{x}+\mathbf{b}$. This property is intuitively appealing because it ensures that the performance of TREWMA is the same for any initial variance-covariance and location.

When the process is IC, the TREWMA shares a good feature with its parametric counterpart MEWMA and nonparametric counterpart MSEWMA. That is, under some specific error distributions, the $Q_t^{R_F}$ process is a Markov chain and the IC ARL of TREWMA can be calculated via the Markov chain approximation. We summarize these results into the following proposition.

PROPOSITION 2: Suppose that $\boldsymbol{\varepsilon}_t$ follows a spherical distribution, under the IC model,

1. The $Q_t^{R_F}$ process is a Markov chain.
2. The run-lengths of TREWMA can be approximately calibrated through a one-dimensional Markov chain approach.

The spherical assumption here is only used for theoretical derivation, but in practice we do not need to impose such a condition. This proposition is particularly useful in determining the control limit of the chart, because the IC ARL of TREWMA for a given spherical distribution can be calculated via the Markov chain model similar to what Runger and Prabhu [29] and Zou and Tsung [39] have done. However, this seems to be marginally interesting to us because TREWMA still requires a large number of IC historical datasets. However, due to its simplicity, it is straightforward to extend TREWMA to a more practical situation where m_0 is small or just larger than $p+1$. We will later see that Proposition 2 is still useful because the control limit of TREWMA serves as a good starting point for finding the control limits for its generalized form.

2.3. An Empirical Spatial Rank-Based EWMA Control Chart

Recall model (1) in which we only have m_0 reference samples before the start of monitoring rather than the IC distribution. In this situation, we may obtain an empirical spatial rank-based EWMA scheme by replacing the unknown quantities with the historical sample, say m_0+t-1 observations, at the time point t. Specifically, according to the definition of spatial rank introduced in Section 2.1, we apply the empirical spatial rank function to the tth future observations $\mathbf{x}_t$

$$R_E(\widehat{\mathbf{M}}_{t-1}\mathbf{x}_t) = \frac{1}{m_0+t-1}\sum_{j=-m_0+1}^{t-1} U(\widehat{\mathbf{M}}_{t-1}(\mathbf{x}_t-\mathbf{x}_j)), \tag{4}$$

where $\widehat{\mathbf{S}}_{t-1} = (\widehat{\mathbf{M}}_{t-1}^T\widehat{\mathbf{M}}_{t-1})^{-1}$ is just the sample covariance matrix based on m_0+t-1 observations,

$$\widehat{\mathbf{S}}_{t-1} = \frac{1}{m_0+t-1}\sum_{j=-m_0+1}^{t-1}(\mathbf{x}_j-\bar{\mathbf{x}}_{t-1})(\mathbf{x}_j-\bar{\mathbf{x}}_{t-1})^T,$$

and $\bar{\mathbf{x}}_{t-1}$ is the sample mean of those m_0+t-1 observations. Besides, we may see from the proof of Proposition 2 that under the IC and the spherical distribution assumption

$$\mathrm{cov}[R_F(\mathbf{M}\mathbf{x})] = E[||R_F(\mathbf{M}\mathbf{x}_t)||^2]\mathbf{I}_p/p.$$

Hence, to ease the computational effort and sequentially update the estimate of $E[||R_F(\mathbf{M}\mathbf{x}_t)||^2]$, we suggest using

$$\widehat{E}[||R_F(\mathbf{M}\mathbf{x}_t)||^2] \approx \left[\sum_{j=-m_0+1}^{0}||\tilde{R}_E(\widehat{\mathbf{M}}_0\mathbf{x}_j)||^2 + \sum_{j=1}^{t-1}||R_E(\widehat{\mathbf{M}}_{j-1}\mathbf{x}_j)||^2\right] /(m_0+t-1), \tag{5}$$

where

$$\tilde{R}_E(\widehat{\mathbf{M}}_0\mathbf{x}_j) = \frac{1}{m_0}\sum_{k=-m_0+1}^{0} U(\widehat{\mathbf{M}}_0(\mathbf{x}_j - \mathbf{x}_k)).$$

Obviously, (5) is a natural moment estimator which combines the information of m_0 historical observations and $t-1$ on-line observations up to time point t. Accordingly, $\widehat{\mathrm{cov}}[R_E(\widehat{\mathbf{M}}_{t-1}\mathbf{x}_t)] \approx \widehat{E}[||R_F(\mathbf{M}\mathbf{x}_t)||^2]\mathbf{I}_p/p$.

Based on the empirical quantities above, we propose the following empirical spatial rank EWMA charting statistic

$$Q_t^{R_E} = \frac{(2-\lambda)p}{\lambda\xi_t}||\mathbf{v}_t||^2, \tag{6}$$

where

$$\mathbf{v}_t = (1-\lambda)\mathbf{v}_{t-1} + \lambda R_E(\widehat{\mathbf{M}}_{t-1}\mathbf{x}_t),$$

$\mathbf{v}_0 = 0$ and $\xi_t \equiv \widehat{E}[||R_F(\mathbf{M}\mathbf{x}_t)||^2]$. Once $Q_t^{R_E}$ exceeds some control limit L, the control chart triggers a signal. Henceforth, this chart is referred to as the SREWMA chart for abbreviation.

Next, we present some asymptotic but insightful properties regarding the SREWMA charting statistic, which could justify the validity of its construction to a certain degree.

PROPOSITION 3: Under the IC model,

1. $\mathrm{cov}(R_E(\mathbf{M}\mathbf{x}_t), R_E(\mathbf{M}\mathbf{x}_k)) = 0$ for any $k \neq t$.
2. Suppose the assumption in Proposition 2 holds. As $\lambda \to 0$ and $\lambda t \to \infty$, $Q_t^{R_E} \xrightarrow{d} \chi_p^2$.

The first part of this proposition tells us that if $\widehat{\mathbf{M}}_t$'s are close enough to the true $\mathbf{M}$, $R_E(\widehat{\mathbf{M}}_{t-1}\mathbf{x}_t)$ are approximately mutually uncorrelated. As a result, our SREWMA sequence accumulates information as self-starting charts with exact Q-statistics in a general sense. The second part reveals that the marginal distributions of charting statistics are the same in an asymptotic viewpoint which allows us to use a fixed (not varying with the time point t) control limit L for SREWMA given the nominal IC ARL, m_0 and λ. Our simulation results shown in the next section concur with these asymptotic analysis that the IC run-length distributions of SREWMA charts behave similarly to those of EWMA charts with known IC parameters.

Before proceeding, we must determine the transformation matrix $\widehat{\mathbf{M}}_k$ used in SREWMA. Although any matrix that satisfies $\widehat{\mathbf{S}}_k = (\widehat{\mathbf{M}}_k^T\widehat{\mathbf{M}}_k)^{-1}$ will suffice for transforming the $\mathbf{x}_t$ to achieve the affine-invariant property, one choice is particularly attractive—the triangular Cholesky inverse root of $\widehat{\mathbf{S}}_k$. Take $\widehat{\mathbf{M}}_k^T$ as an upper triangular matrix. With this choice, the computation of SREWMA is extremely fast because of the rank-one update feature of the Cholesky decomposition.

To end this subsection, we note that when a group of g observations $\{\mathbf{x}_{i1}, \ldots, \mathbf{x}_{ig}\}$ are taken sequentially from the process at each time point, the SREWMA chart can be readily defined in a similar way to (6). It amounts to extending the empirical spatial rank to the group case at time point t,

$$R_{E,t}^g = \frac{1}{(m_0+t-1)g^2}\sum_{j=-m_0+1}^{t-1}\sum_{l=1}^{g}\sum_{k=1}^{g} U(\widehat{\mathbf{M}}_{t-1}(\mathbf{x}_{tk} - \mathbf{x}_{jl})).$$

2.4. Computational Issues

Fast implementation is important and some computational issues deserve our careful examination. For the proposed chart, computing the charting statistic $Q_t^{R_E}$ requires a considerable amount of computing time. In this part, we provide some updating formulas, which can greatly simplify the computation.

Write the total sum of squares at time t as

$$(m_0+t)\widehat{\mathbf{S}}_t = (m_0+t-1)\widehat{\mathbf{S}}_{t-1} + \alpha\boldsymbol{\beta}\boldsymbol{\beta}^T,$$

where $\alpha = (m_0+t-1)/(m_0+t)$ and $\boldsymbol{\beta} = (\mathbf{x}_t - \bar{\mathbf{x}}_{t-1})$. The Plackett updating formulas give a fast rank-one update of the inverses of the symmetric matrix $\widehat{\mathbf{S}}_t$,

$$[(m_0+t)\widehat{\mathbf{S}}_t]^{-1} = [(m_0+t-1)\widehat{\mathbf{S}}_{t-1}]^{-1} - \alpha\frac{\boldsymbol{\gamma}\boldsymbol{\gamma}^T}{1+\alpha\boldsymbol{\beta}^T\boldsymbol{\gamma}},$$

where $\boldsymbol{\gamma} = [(m_0+t-1)\widehat{\mathbf{S}}_{t-1}]^{-1}\boldsymbol{\beta}$. Furthermore, observing that $1 + \alpha\boldsymbol{\beta}^T\boldsymbol{\gamma} > 0$ (since $\widehat{\mathbf{S}}_{t-1}^{-1}$ is positive-definite), we can obtain $\widehat{\mathbf{M}}_t$ by the rank-one downdating Cholesky factorization. Such an algorithm is quite efficient and reliable in practice with a computational effort of $O(p^2)$. Subroutines or functions for this algorithm are available in most statistical software packages. In our numerical studies, we use the subroutine "DLDNCH" in the IMSL package of Visual Fortran 6.5.

In addition,

$$\xi_{t+1} = \frac{m_0+t-1}{m_0+t}\xi_t + ||R_E(\widehat{\mathbf{M}}_{t-1}\mathbf{x}_t)||^2/(m_0+t). \tag{7}$$

Therefore, the calculation of $Q_t^{R_E}$ can be operated in a "recursive" way. Of course, we should emphasize that, it is not exactly recursive because there is no recursive expression for calculating $R_E(\widehat{\mathbf{M}}_{t-1}\mathbf{x}_t)$. That is, the computational effort of calculating $R_E(\widehat{\mathbf{M}}_{t-1}\mathbf{x}_t)$ grows sequentially with time t. Computing power has improved dramatically and it is computationally trivial to implement the SREWMA chart by using the formulas given above.

2.5. Practical Guidelines

2.5.1. On Choosing the Smoothing Weight, λ

In general, a smaller λ leads to a quicker detection of smaller shifts (c.f., e.g., Lucas and Saccucci [16], Prabhu and Runger [21]). This applies also to SREWMA. From the simulation results shown in Section 3, unlike its parametric counterpart in which λ should be carefully chosen to balance the robustness to non-normality and the detection ability to various shift magnitudes (c.f., Stoumbos and Sullivan [31]), the SREWMA chart is quite robust under IC with any small weight. The exceptions are very skewed distributions and high dimensional cases (e.g., for a ten-dimensional multivariate gamma distribution with shape parameter 1, using $\lambda \geq 0.1$ would result in considerably large biases in IC ARLs; see Section 3 for details). In the very skewed situations, the SREWMA also requires a relatively smaller λ to achieve the nominal IC ARL as the parametric self-starting chart does. However the choice is much easier because the deteriorating effect of non-normality on its IC ARL is much slighter than that on its parametric counterparts. Based on our simulation results, we suggest choosing $\lambda \in [0.05, 0.1]$, which is a reasonable choice in practice, and using $\lambda \in [0.025, 0.05]$ when there is evidence that the underlying distribution is very skewed.

2.5.2. On the m_0

Although the SREWMA chart is a self-starting scheme in the sense that it can be implemented at the start-up of a process (as long as $m_0 \geq p+2$ to make the estimator $\widehat{\mathbf{M}}_0$ available), we believe that starting testing with an m_0 that is too small is not a good idea. An m_0 that is too small would result in a severe "masking-effect" if a short-run change occurs. Rather, we suggest that a practitioner should gather a modest number of observations through a Phase-I study to obtain at least an initial verification that the process is actually stable, and only then should the practitioner start the formal SREWMA chart. Such a stage of collecting a reasonable number of observations before monitoring is essentially similar to the so-called "warm-up" process introduced by Hawkins and Deng [6]. In the warm-up process, control charts update the parameter estimation or accumulate the information of IC processes but do not signal. We suggest collecting at least $m \geq 2p$ (of course the more the better) historical observations before monitoring. Our empirical results show that to obtain a satisfactory monitoring performance it may require at least 50-100 IC observations (say, $m_0+\tau \geq 50$; of course this number also depends on the dimensionality) before the change actually occurs.

2.5.3. On the Control Limits

Proposition 2 reveals that the control limits of TREWMA for spherical distributions can be obtained by using a Markov chain method. However, for the SREWMA, the Markov chain method is not feasible unless m_0 is very large. When m_0 is small, given $Q_i^{R_E}, i < t$, the conditional distribution of $Q_t^{R_E}$ for small t is considerably different from the steady-state conditional distribution. Hence, the control limits of SREWMA depend not only on λ and the nominal IC ARL, but also on the number of observations we have collected before the start of monitoring. In fact, this characteristic is shared by the procedures based on dynamic change-point detection and the self-starting schemes without formal i.i.d Q-statistics. See Zou et al. [40] and Zou and Tsung [38] for discussions. Certainly, this would not produce any difficulty in constructing the SREWMA chart but an extra parameter, m_0, should be added when searching for the control limit.

Table 1 provides the control limits of the SREWMA chart for various commonly used combinations of m_0, λ, p and IC ARL's obtained using simulations with a multinormal distribution. The simulation results to be shown in Section 3 demonstrate that the IC run-length performance of SREWMA is quite robust under various process distributions including very skewed distributions. Therefore, the control limits tabulated in Table 1 can be used for any continuous distribution. The control limits of TREWMA can be used as a good starting point for the searching procedure since the control limits for small values of m_0 do not differ much from those of TREWMA. Our empirical results show that for $m_0 \geq 200$, the control limits of SREWMA and TREWMA are almost the same for $\lambda \leq 0.1$ and $p \leq 20$. Hence, in those situations, we recommend directly using the control limits of TREWMA (which are exactly the same as those of MSEWMA and are given in Table 1 by Zou and Tsung [39] with the most commonly used parameters).

To end this section, we summarize the detailed steps for implementing the SREWMA as follows. The Fortran code for implementing the proposed scheme, including the procedures for finding the control limits, are available from the authors upon request.

The SREWMA Control Chart

1. Choose the desired IC ARL and smoothing parameter, λ. Determine the control limit, L, given m_0, λ and IC ARL.
2. Based on the historical sample $\{\mathbf{x}_{-m_0+1}, \ldots, \mathbf{x}_0\}$, calculate the sample covariance matrix $\widehat{\mathbf{S}}_0$ and obtain the corresponding $\widehat{\mathbf{M}}_0$ by Cholesky factorization. Then, compute $\sum_{j=-m_0+1}^{0} ||\tilde{R}_E(\widehat{\mathbf{M}}_0\mathbf{x}_j)||^2$ in (5).
3. Start monitoring the process and obtain product observations, $\mathbf{x}_t$, sequentially. For a new observation, compute $R_E(\widehat{\mathbf{M}}_{t-1}\mathbf{x}_t)$ in (4) and obtain the charting statistic, $Q_t^{R_E}$, given by (6).
4. If $Q_t^{R_E}$ exceeds the control limit, the control chart triggers an alarm; Otherwise, using the rank-one

Table 1. The control limits of the SREWMA chart for $ARL_0 = 200, 370$, and 500 under p-variate distributions with the standard multinormal distribution.

m_0	IC ARL	λ	$p=2$	$p=3$	$p=4$	$p=5$	$p=7$	$p=10$	$p=15$	$p=20$
10	200	0.1	8.172	10.139	11.934	13.604	16.682			
		0.05	7.182	9.077	10.811	12.452	15.396			
		0.025	5.832	7.566	9.219	10.707	13.470			
	370	0.1	9.399	11.433	13.375	15.121	18.336			
		0.05	8.590	10.602	12.546	14.229	17.322			
		0.025	7.308	9.255	11.070	12.687	15.675			
	500	0.1	10.006	12.088	14.062	15.851	19.149			
		0.05	9.297	11.343	13.299	15.046	18.235			
		0.025	8.019	10.051	11.957	13.646	16.700			
20	200	0.1	8.263	10.276	12.122	13.848	17.095	21.600	28.748	
		0.05	7.226	9.186	10.986	12.696	15.784	20.098	26.666	
		0.025	5.866	7.654	9.319	10.814	13.795	17.744	23.858	
	370	0.1	9.467	11.554	13.538	15.327	18.686	23.306	30.557	
		0.05	8.634	10.711	12.659	14.407	17.647	22.094	29.004	
		0.025	7.352	9.336	11.157	12.827	15.927	20.236	26.900	
	500	0.1	10.062	12.188	14.200	16.021	19.449	24.145	31.476	
		0.05	9.358	11.443	13.424	15.220	18.535	23.089	30.109	
		0.025	8.084	10.126	12.032	13.796	16.937	21.397	28.150	
40	200	0.1	8.297	10.328	12.197	13.992	17.307	21.957	29.207	36.355
		0.05	7.269	9.233	11.055	12.815	15.971	20.464	27.366	34.163
		0.025	5.894	7.729	9.400	10.958	13.970	18.081	24.589	30.910
	370	0.1	9.525	11.603	13.635	15.458	18.873	23.656	31.057	38.218
		0.05	8.674	10.767	12.740	14.547	17.835	22.470	29.622	36.501
		0.025	7.374	9.392	11.235	12.946	16.113	20.525	27.363	33.880
	500	0.1	10.128	12.244	14.275	16.151	19.612	24.455	31.935	39.110
		0.05	9.393	11.507	13.500	15.352	18.723	23.409	30.722	37.616
		0.025	8.120	10.173	12.107	13.908	17.100	21.657	28.649	35.257

downdating Cholesky factorization suggested in Section 2.4 to obtain the updating estimator of $\mathbf{M}$, say $\widehat{\mathbf{M}}_t$. Also, update $\widehat{E}[||R_F(\mathbf{M}\mathbf{x}_t)||^2]$ by (7). Then, continue to collect new observations.

3. PERFORMANCE ASSESSMENT

We present some simulation results in this section regarding the performance of the proposed SREWMA chart and compare it with some other procedures in the literature. All the results in this section are obtained from 10,000 replications unless indicated otherwise.

Comparing the SREWMA procedure with alternative nonparametric methods turned out to be difficult due to the lack of an obvious comparable method. This is because most of the approaches in the literature were designed for the cases where sufficient historical observations are available to accurately estimate the IC distribution of observations or some IC parameters. See Zou and Tsung [39] for some discussions and reviews. Thus, we consider the parametric self-starting multivariate EWMA (denoted as SSEWMA) control chart proposed by Hawkins and Maboudou-Tchao [7] which was designed under the normality assumption (Brown et al. [1]). Zamba and Hawkins's [34] change-point procedure does not require known parameters either, but we do not consider it here for the following two reasons: the difference in the way a charting statistic is constructed may distract our focus from the present robustness consideration and our simulation shows that the change-point detection procedure performs worse than the SSEWMA chart with a small smoothing parameter from the robustness viewpoint.

Following the robustness analyses of Stoumbos and Sullivan [31] and Zou and Tsung [39], we consider the following distributions: (i) multinormal; (ii) multivariate t with ζ degrees of freedom, denoted as $t_{p,\zeta}$; (iii) multivariate gamma with shape parameter ζ and scale parameter 1, denoted as $\text{Gam}_{p,\zeta}$. In addition, the following distribution is involved in the comparison: (iv) in each observed vector, the first $[p/2]$ measurement components are i.i.d. t distributed with ζ_1 degrees of freedom and the other $p - [p/2]$ measurement components are i.i.d. chi-square distributed with ζ_2 degrees of freedom. The reason for considering this distribution is that unlike (i)–(iii), its marginal distributions are not all the same. The number and variety of covariance matrices and shift directions are too large to allow a comprehensive, all-encompassing comparison. Our goal is to show the effectiveness, robustness and sensitivity of the SREWMA chart, and thus we only choose certain representative models for illustration. Specifically, for the first three distribution cases, the covariance matrix $\mathbf{\Sigma}_0 = (\sigma_{ij})$ is chosen to be $\sigma_{ii} = 1$ and

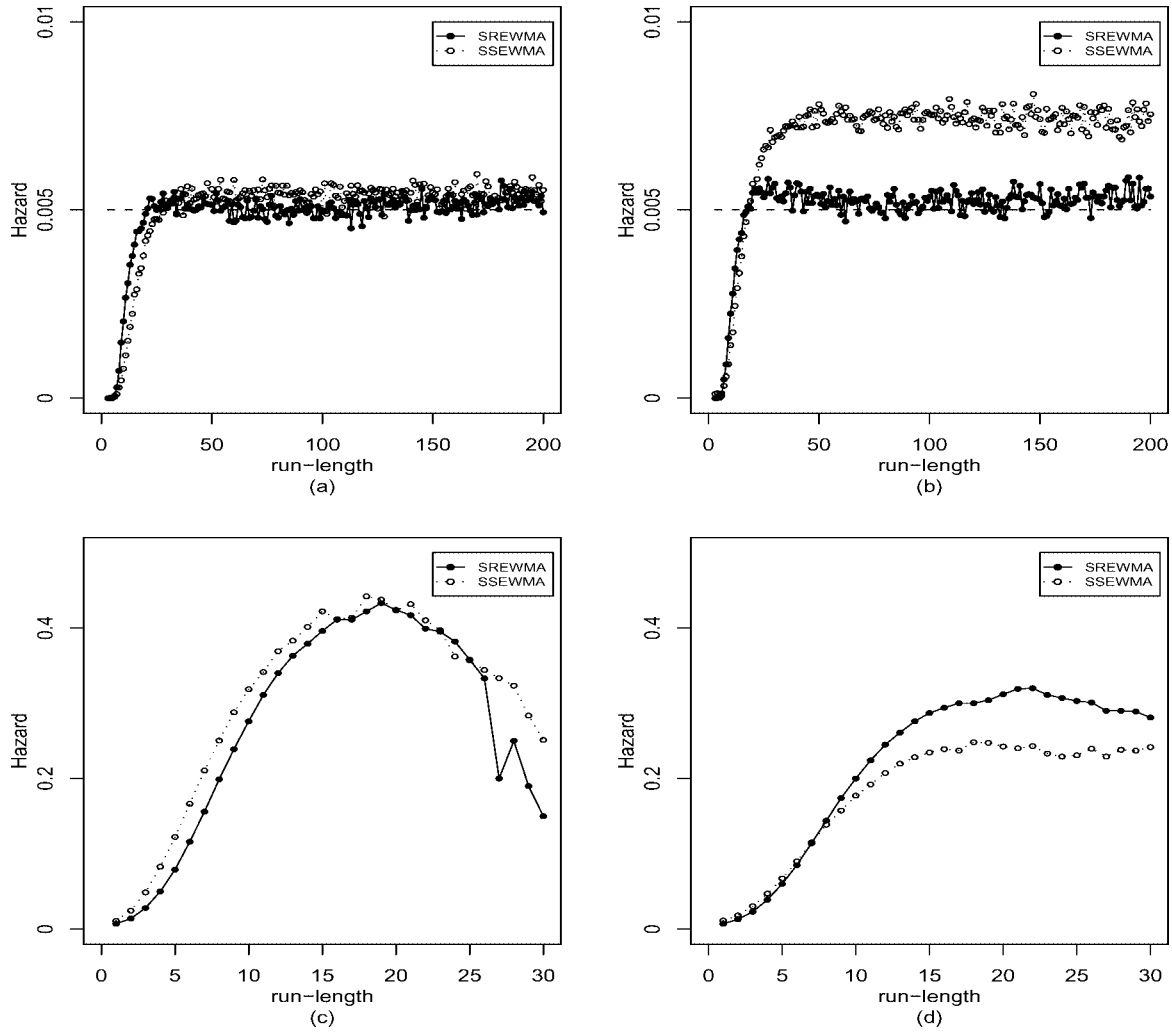

Figure 2. Hazard curves for SREWMA and SSEWMA with $\lambda = 0.05$, $p = 5$, $m_0 = 10$, and $\text{ARL}_0 = 200$: (a) IC multinormal; (b) IC multivariate t; (c) OC multinormal; (d) OC multivariate t.

$\sigma_{ij} = 0.5^{|i-j|}$, for $i, j = 1, 2, \ldots, p$. For brevity, a shift of size δ in only the first component is used, i.e., $\mathbf{x}_i + \delta\mathbf{e}_1$ with $\mathbf{e}_1 = (1, 0, \ldots, 0)^T$. We conducted some other simulations with various shift types and the results show that the general conclusions given below do not change.

3.1. The Run-Length Distribution of SREWMA

First, we study SREWMA's IC run-length distribution. As recognized in the literature, it is often insufficient to summarize run-length behavior by ARL, especially for self-starting control charts (cf., Jones [13]). As an alternative, here we use the hazard function $H_1(r)/H_2(r)$ recommended by Hawkins and Maboudou-Tchao [7], where $H_1(r)$ is the probability that the run-length equals r and $H_2(r)$ is the probability that the run-length equals r or a larger number. Note that if the run-length follows a geometric distribution, the corresponding hazard is a constant whose inverse is the ARL.

Figures 2a and 2b show the IC hazard functions of the SREWMA and SSEWMA charts for a multinormal distribution and a $t_{p,5}$ distribution, respectively, based on 250,000 replications, with $p = 5$, IC ARL $= 200$ (denoted as ARL_0), $\lambda = 0.05$ and $m_0 = 10$. From the plots, we can see that the IC hazard of SREWMA is initially around zero, then surges to around $0.005 = 1/200$ before stabilizing. Except for short run-lengths, the geometric distribution is an excellent fit of the IC run-length for the SREWMA chart, just as it is for the classical Shewhart chart with known parameters. These results reflect that the ARL is indeed a suitable summary of the IC run behavior of the SREWMA because from the ARL, any quantity such as a moment or a percentile, can be approximately calculated (the geometric distribution is fully characterized by its average). In comparison, the IC hazard

Table 2. ARL and SDRL values with multinormal distributions.

		SREWMA						SSEWMA					
		$\lambda = 0.1$		$\lambda = 0.05$				$\lambda = 0.1$		$\lambda = 0.05$			
		$\tau = 40$		$\tau = 40$		$\tau = 90$		$\tau = 40$		$\tau = 40$		$\tau = 90$	
	δ	ARL	SDRL	ARL	SDRL	ARL	SDRL	ARL	SDRL	ARL	SDRL	ARL	SDRL
$p = 5$	0.50	94.8	137	68.6	103	43.4	51.0	86.6	133	62.4	93.2	38.6	45.4
	0.75	36.5	65.11	26.3	34.0	19.9	13.1	35.1	67.4	25.1	34.0	18.4	11.9
	1.00	16.2	20.91	15.4	11.7	13.4	6.39	15.7	22.1	14.7	10.2	12.4	6.31
	1.50	8.18	3.79	9.52	4.08	8.67	3.31	7.22	3.67	8.59	3.90	7.70	3.19
	2.00	6.09	2.27	7.42	2.83	6.84	2.37	5.11	2.10	6.25	2.60	5.65	2.15
	3.00	4.62	1.46	5.80	2.02	5.36	1.69	3.37	1.22	4.26	1.63	3.83	1.36
	4.00	4.12	1.24	5.22	1.76	4.81	1.48	2.63	0.90	3.36	1.25	3.01	1.03
$p = 10$	0.50	116	164	81.7	119	56.1	74.2	103	145	73.0	101	50.6	65.0
	0.75	52.6	95.4	33.2	47.5	24.7	20.8	44.6	74.7	30.4	38.6	22.5	16.0
	1.00	21.9	38.4	18.2	14.9	15.7	7.91	19.4	27.7	17.2	13.1	14.7	7.71
	1.50	9.26	4.83	10.5	4.72	9.81	3.80	8.63	4.61	9.84	4.51	8.83	3.76
	2.00	6.61	2.60	7.97	3.14	7.49	2.65	5.89	2.48	7.14	2.92	6.46	2.50
	3.00	4.72	1.54	5.91	2.10	5.57	1.79	3.80	1.36	4.82	1.77	4.37	1.50
	4.00	4.02	1.23	5.09	1.74	4.79	1.48	2.93	0.99	3.79	1.36	3.40	1.11

of SSEWMA performs similar to that of SREWMA for the normal distribution (Fig. 2a), but is significantly larger than the inverse of the nominal IC ARL (0.005) for the $t_{p,5}$ distribution (Fig. 2b). This indicates that the false alarm rate of SSEWMA would be larger than the desire one if the control limit obtained under multivariate normal assumptions is used for a distribution with heavy-tails like $t_{p,5}$. Moreover, the sample mean and the sample standard deviations of the run-length (SDRL) of SREWMA are, respectively, 200 and 188 for (i) $N_p(\mathbf{0}, \mathbf{\Sigma})$ and 198 and 183 for (ii) $t_{p,5}$, which are quite close and further confirm that the SREWMA chart works well under this set of IC conditions. We conducted some other simulations with various combinations of λ, IC ARL, m_0 and p to check whether the above conclusions are true in other settings. The simulation results show that the SREWMA chart performs quite satisfactorily under these other settings as well.

The OC hazards shown in Figs. 2c and 2d are a shift of 1.5 standard deviations from the first component of the $N_p(\mathbf{0}, \mathbf{\Sigma})$ and $t_{p,5}$ data respectively, introduced after $\tau = 90$ IC observations. Any series in which a signal occurs before the $(\tau+1)^{\text{th}}$ observation is discarded (c.f., Hawkins and Olwell [8]). These hazard curves are far from constant, and indicate that the OC run-length distribution of SREWMA is tight and unimodal. The hazards of the two charts present similar pattern. The SREWMA seems to have larger hazards for the $t_{p,5}$ distribution, leading to quicker detection in short runs. In this setting, the ARL and SDRL are respectively 8.69 and 3.36 for the multinormal distribution and 9.84 and 4.12 for the multivariate t distribution. In both situations the SDRLs are much smaller than the ARL.

3.2. Comparisons Between SREWMA and SSEWMA

In this subsection, the SREWMA and SSEWMA charts are compared in terms of OC ARL. Because a similar conclusion holds for other cases, throughout this section, we only present the results when the $\text{ARL}_0 = 200$ for illustration. Results with other commonly used ARL_0s, such as 370 or 500, are available from the authors upon request. A low-dimensional case with $p = 5$ and a higher-dimensional case with $p = 10$ are involved for each distribution case. We fix $m_0 = 10$ and 20 for $p = 5$ and 10, respectively. This setting satisfies the requirements for starting SREWMA and SSEWMA ($m_0 \geq p + 2$). For an OC ARL comparison, we consider the steady-state ARL (SSARL). To evaluate the SSARL behavior of each chart, any series in which a signal occurs before the $(\tau+1)^{\text{th}}$ observation is discarded.

We first consider the multinormal distribution. The simulation results for the SREWMA and SSEWMA charts with $\lambda = 0.05$ and $\lambda = 0.1$ are presented in Table 2. Besides the ARLs, the corresponding SDRLs are also included in this table to give a broader picture of the run-length distribution. Apart from the parameters above, the performance of self-starting charts depends on the choice of τ. We consider $\tau = 40$ and 90. From this table, we observe that the SSEWMA chart has superior efficiency as we would expect, since the parametric hypothesis is the correct one in this case. The SREWMA chart also performs quite satisfactorily and the difference between SSEWMA and SREWMA is small, even when p is large. It should be pointed out that the efficiency of SSEWMA becomes even more superior when δ is quite large, say $\delta \geq 3$. For example, the SSEWMA has

Table 3. IC ARL values with multivariate t distributions $t_{p,\zeta}$.

		SREWMA			SSEWMA			
		λ			λ			
	ζ	0.1	0.05	0.025	0.1	0.05	0.025	0.01
$p = 5$	3	168 (0.16)	185 (0.07)	192 (0.04)	71.3 (0.64)	109 (0.45)	146 (0.27)	179 (0.11)
	4	178 (0.11)	193 (0.04)	198 (0.01)	93.6 (0.53)	133 (0.33)	168 (0.16)	188 (0.06)
	5	186 (0.07)	196 (0.02)	202 (0.01)	109 (0.45)	148 (0.26)	178 (0.11)	190 (0.05)
	7	188 (0.06)	196 (0.02)	201 (0.01)	134 (0.33)	170 (0.15)	187 (0.06)	196 (0.02)
	10	194 (0.03)	200 (0.00)	201 (0.01)	151 (0.24)	179 (0.10)	189 (0.05)	203 (0.01)
	15	196 (0.02)	200 (0.00)	200 (0.00)	173 (0.14)	189 (0.05)	197 (0.02)	198 (0.01)
$p = 10$	3	147 (0.26)	177 (0.12)	185 (0.08)	54.5 (0.73)	87.0 (0.56)	126 (0.37)	166 (0.17)
	4	165 (0.18)	185 (0.07)	193 (0.03)	71.7 (0.64)	111 (0.45)	149 (0.25)	182 (0.09)
	5	170 (0.15)	187 (0.06)	196 (0.02)	85.3 (0.57)	128 (0.36)	162 (0.19)	187 (0.06)
	7	181 (0.09)	193 (0.04)	197 (0.02)	107 (0.46)	149 (0.25)	176 (0.12)	194 (0.03)
	10	184 (0.08)	194 (0.03)	198 (0.02)	132 (0.34)	164 (0.18)	182 (0.09)	194 (0.03)
	15	194 (0.03)	201 (0.01)	199 (0.01)	150 (0.25)	177 (0.12)	189 (0.05)	196 (0.02)

The relative deviation of each IC ARL from the nominal one is in parentheses.

an ARL that is only 4% smaller than that of the SREWMA in detecting a shift size δ of 0.75, although the advantage increases to 25–30% when the shift size δ is 4.0. The phenomenon parallels the findings of Zou and Tsung [39] who compared the parametric and nonparametric non-self-starting charts.

Next, the multivariate t distribution and the multivariate gamma distribution are considered. We firstly evaluate the IC ARL values which are tabulated in Tables 3 and 4. For clearer comparisons, we also list the relative deviation of each IC ARL from the nominal one in parentheses, which is calculated as $|\text{ARL} - 200|/200$. As before, various cases with different combinations of dimensionality, λ and degrees of freedom ζ are considered. From the tables, we can see that the SREWMA is quite robust to the heavy-tailed and skewed distributions as long as λ is not too large (i.e., $\lambda \leq 0.1$). Its IC ARL is always quite close to the nominal one even for the extremely non-normal and high-dimensional distributions of $t_{10,3}$ and $\text{Gam}_{10,1}$. In comparison, the SSEWMA usually has a large bias in the IC ARL and the degradation becomes more pronounced as the dimensionality increases. For a very small ζ, even with $\lambda = 0.01$, the SSEWMA chart cannot maintain a desired IC ARL.

Now, we turn to Tables 5 and 6, which give OC ARL values with multivariate t observations with five degrees of freedom ($t_{p,5}$) and with multivariate gamma observations with three degrees of freedom ($\text{Gam}_{p,3}$), respectively. In these tables, we fix $m_0+\tau = 100$ for each comparison scenario for simplicity. For the SREWMA chart, the value of λ is chosen to be 0.05 or 0.025. For a fair comparison, the SSEWMA with $\lambda = 0.01$

Table 4. IC ARL values with multivariate gamma distributions $\text{Gam}_{p,\zeta}$.

		SREWMA			SSEWMA			
		λ			λ			
	ζ	0.1	0.05	0.025	0.1	0.05	0.025	0.01
$p = 5$	1	173 (0.13)	193 (0.03)	201 (0.01)	71.5 (0.64)	113 (0.44)	150 (0.25)	171 (0.14)
	2	185 (0.08)	198 (0.01)	204 (0.02)	103 (0.49)	148 (0.26)	176 (0.12)	188 (0.06)
	3	188 (0.06)	200 (0.00)	204 (0.02)	119 (0.40)	163 (0.18)	188 (0.06)	193 (0.04)
	5	192 (0.04)	200 (0.00)	204 (0.02)	143 (0.29)	173 (0.14)	194 (0.03)	196 (0.02)
	10	197 (0.02)	200 (0.00)	200 (0.00)	163 (0.19)	188 (0.06)	197 (0.02)	198 (0.01)
	15	199 (0.01)	200 (0.00)	200 (0.00)	171 (0.14)	189 (0.05)	198 (0.01)	199 (0.01)
$p = 10$	1	155 (0.23)	189 (0.06)	196 (0.02)	65.7 (0.67)	108 (0.46)	149 (0.26)	176 (0.12)
	2	168 (0.16)	191 (0.05)	197 (0.01)	93.6 (0.53)	137 (0.31)	170 (0.15)	187 (0.06)
	3	183 (0.09)	198 (0.01)	201 (0.01)	111 (0.45)	153 (0.24)	178 (0.11)	190 (0.05)
	5	185 (0.08)	198 (0.01)	198 (0.01)	133 (0.33)	169 (0.15)	186 (0.07)	194 (0.03)
	10	193 (0.04)	200 (0.00)	202 (0.01)	157 (0.22)	182 (0.09)	193 (0.03)	201 (0.01)
	15	195 (0.02)	201 (0.01)	201 (0.01)	171 (0.14)	187 (0.06)	196 (0.02)	199 (0.01)

The relative deviation of each IC ARL from the nominal one is in parentheses.

Table 5. ARL and SDRL values with multivariate $t_{p,5}$ observations.

		SREWMA		SSEWMA		SSEWMA-UA	
	δ	$\lambda = 0.05$	$\lambda = 0.025$	$\lambda = 0.01$	$\lambda = 0.005$	$\lambda = 0.05$	$\lambda = 0.025$
$p = 5$	0.00	197 (1.84)	202 (1.70)	190 (1.53)	198 (1.24)	200 (1.85)	200 (1.76)
	0.50	53.4 (0.68)	45.9 (0.45)	51.1 (0.46)	52.0 (0.41)	73.3 (0.97)	58.4 (0.72)
	0.75	23.9 (0.19)	24.7 (0.15)	30.8 (0.21)	33.8 (0.22)	33.7 (0.43)	29.7 (0.25)
	1.00	15.6 (0.08)	17.4 (0.09)	22.0 (0.12)	25.1 (0.14)	19.3 (0.14)	19.8 (0.12)
	1.50	9.85 (0.04)	11.6 (0.05)	14.5 (0.07)	16.4 (0.08)	10.8 (0.05)	12.3 (0.06)
	2.00	7.60 (0.03)	9.16 (0.04)	11.9 (0.05)	12.6 (0.06)	7.64 (0.03)	9.05 (0.04)
	3.00	5.78 (0.02)	7.08 (0.03)	7.54 (0.03)	8.75 (0.04)	5.04 (0.02)	5.99 (0.02)
	4.00	5.08 (0.02)	6.24 (0.02)	5.91 (0.02)	6.89 (0.03)	3.88 (0.01)	4.64 (0.02)
$p = 10$	0.00	187 (1.85)	196 (1.69)	190 (1.46)	196 (1.23)	200 (1.84)	200 (1.76)
	0.50	73.2 (1.00)	58.4 (0.66)	56.1 (0.49)	59.6 (0.46)	99.0 (1.27)	69.9 (0.84)
	0.75	32.3 (0.37)	30.5 (0.21)	35.4 (0.24)	39.9 (0.25)	47.4 (0.61)	36.8 (0.32)
	1.00	19.3 (0.12)	20.8 (0.11)	25.7 (0.14)	29.9 (0.16)	25.8 (0.27)	23.6 (0.15)
	1.50	11.4 (0.05)	13.3 (0.06)	17.0 (0.08)	20.0 (0.09)	13.4 (0.07)	14.3 (0.07)
	2.00	8.48 (0.03)	10.2 (0.04)	12.7 (0.06)	15.3 (0.07)	9.28 (0.04)	10.6 (0.04)
	3.00	6.14 (0.02)	7.56 (0.03)	8.78 (0.04)	10.8 (0.05)	5.99 (0.02)	7.02 (0.03)
	4.00	5.19 (0.02)	6.43 (0.02)	6.92 (0.03)	8.53 (0.04)	4.59 (0.02)	5.44 (0.02)

Standard errors are in parentheses.

or 0.005 is considered to ensure the nominal IC ARL can be attained. To appreciate the pros and cons of SREWMA with respect to SSEWMA under OC models, we consider an SSEWMA chart with the same value of λ as SREWMA but its control limit is adjusted to make the IC ARL equal to the nominal one. The corresponding ARL results with $\lambda = 0.05$ and 0.025 are given in the columns labeled with SSEWMA-UA (UA means unbiased-ARL). Note that such a charting scheme with adjustment is only for comparison use in our simulations but not applicable in practical applications since the error distribution is usually unknown as we claimed before.

The results in the two tables are similar: The SREWMA chart is more efficient in detecting the small and moderate shifts than the SSEWMA chart in the sense that even when the value of λ is much larger than that of the SSEWMA, the OC ARLs of the SREWMA are generally smaller than those of the SSEWMA in most of cases. When λ is small, the SSEWMA chart is robust to non-normality under the IC situation; however, its ability to detect moderate and large shifts is largely compromised. In particular, the SREWMA with $\lambda = 0.025$ performs almost uniformly better than the SSEWMA with $\lambda = 0.005$, and the difference is quite remarkable. In addition, it can be readily seen that the SSEWMA chart is

Table 6. ARL and SDRL values with multivariate $\text{Gam}_{p,3}$ observations.

		SREWMA		SSEWMA		SSEWMA-UA	
	δ	$\lambda = 0.05$	$\lambda = 0.025$	$\lambda = 0.01$	$\lambda = 0.005$	$\lambda = 0.05$	$\lambda = 0.025$
$p = 5$	0.00	200 (1.83)	203 (1.72)	193 (1.55)	199 (1.26)	200 (1.84)	200 (1.76)
	0.50	61.9 (0.77)	52.9 (0.58)	56.9 (0.54)	56.7 (0.49)	80.9 (1.08)	62.1 (0.72)
	0.75	28.7 (0.27)	28.4 (0.20)	33.9 (0.26)	36.5 (0.25)	39.6 (0.55)	33.9 (0.30)
	1.00	17.9 (0.11)	19.5 (0.11)	24.3 (0.15)	26.7 (0.16)	22.8 (0.20)	22.1 (0.14)
	1.50	10.8 (0.05)	12.7 (0.05)	15.6 (0.08)	17.7 (0.09)	12.1 (0.06)	13.4 (0.07)
	2.00	8.22 (0.03)	9.87 (0.04)	11.6 (0.06)	13.3 (0.07)	8.43 (0.04)	9.68 (0.04)
	3.00	6.12 (0.02)	7.50 (0.03)	7.97 (0.04)	9.26 (0.04)	5.37 (0.02)	6.48 (0.03)
	4.00	5.30 (0.02)	6.53 (0.02)	6.18 (0.03)	7.23 (0.03)	4.14 (0.01)	5.00 (0.02)
$p = 10$	0.00	198 (1.84)	201 (1.70)	187 (1.46)	196 (1.23)	200 (1.84)	200 (1.76)
	0.50	88.9 (1.09)	72.0 (0.79)	63.0 (0.60)	63.1 (0.52)	97.1 (1.25)	78.2 (0.97)
	0.75	44.1 (0.54)	37.9 (0.33)	39.6 (0.30)	43.9 (0.31)	52.2 (0.73)	41.3 (0.46)
	1.00	24.9 (0.24)	24.8 (0.15)	28.4 (0.17)	32.1 (0.19)	29.0 (0.32)	26.2 (0.18)
	1.50	13.4 (0.07)	15.3 (0.07)	18.6 (0.10)	21.8 (0.11)	14.2 (0.08)	15.6 (0.08)
	2.00	9.73 (0.04)	11.5 (0.05)	13.9 (0.07)	16.5 (0.08)	9.94 (0.05)	11.4 (0.05)
	3.00	6.77 (0.02)	8.25 (0.03)	9.5 (0.04)	11.5 (0.05)	6.33 (0.02)	7.52 (0.03)
	4.00	5.56 (0.02)	6.88 (0.02)	7.44 (0.03)	9.02 (0.04)	4.76 (0.02)	5.76 (0.02)

Standard errors are in parentheses.

Table 7. ARL and SDRL values with multivariate mixed-components distributions.

		SREWMA		SSEWMA		SSEWMA-UA	
	δ	$\lambda = 0.05$	$\lambda = 0.025$	$\lambda = 0.01$	$\lambda = 0.005$	$\lambda = 0.05$	$\lambda = 0.025$
$p = 5$	0.00	195 (1.78)	200 (1.72)	195 (1.54)	197 (1.39)	201 (1.88)	200 (1.77)
	0.50	74.7 (0.97)	61.8 (0.67)	62.3 (0.61)	60.8 (0.53)	89.9 (1.20)	70.3 (0.85)
	0.75	35.1 (0.38)	33.1 (0.26)	37.2 (0.28)	39.3 (0.27)	43.3 (0.56)	36.9 (0.35)
	1.00	21.1 (0.16)	22.3 (0.14)	26.7 (0.16)	29.3 (0.17)	24.7 (0.25)	23.7 (0.15)
	1.50	12.2 (0.06)	14.1 (0.06)	17.1 (0.09)	19.4 (0.10)	12.9 (0.07)	14.6 (0.07)
	2.00	9.03 (0.04)	10.8 (0.04)	12.8 (0.06)	14.6 (0.07)	8.99 (0.04)	10.5 (0.05)
	3.00	6.52 (0.02)	7.99 (0.03)	8.63 (0.04)	10.0 (0.05)	5.77 (0.02)	6.93 (0.03)
	4.00	5.53 (0.02)	6.81 (0.02)	6.77 (0.03)	7.88 (0.04)	4.39 (0.02)	5.33 (0.02)
$p = 10$	0.00	196 (1.91)	198 (1.75)	194 (1.42)	197 (1.18)	200 (1.86)	200 (1.75)
	0.50	99.7 (1.23)	80.2 (0.95)	69.8 (0.67)	70.4 (0.58)	105 (1.34)	81.2 (0.95)
	0.75	52.3 (0.68)	42.4 (0.38)	43.1 (0.32)	47.1 (0.31)	58.2 (0.81)	45.8 (0.47)
	1.00	28.4 (0.26)	27.5 (0.18)	31.4 (0.19)	35.7 (0.21)	32.5 (0.37)	28.8 (0.22)
	1.50	15.0 (0.08)	16.8 (0.08)	20.1 (0.10)	23.8 (0.12)	15.9 (0.11)	17.0 (0.09)
	2.00	10.7 (0.05)	12.6 (0.05)	15.2 (0.07)	18.0 (0.08)	10.7 (0.05)	12.3 (0.06)
	3.00	7.29 (0.03)	8.87 (0.03)	10.4 (0.05)	12.5 (0.05)	6.74 (0.03)	8.11 (0.03)
	4.00	5.88 (0.02)	7.27 (0.03)	8.05 (0.03)	9.77 (0.04)	5.08 (0.02)	6.18 (0.02)

Standard errors are in parentheses.

outperformed by the SREWMA chart with the same value of λ in detecting small or moderate magnitudes of shifts by a considerable margin, while the SSEWMA is more efficient in detecting large shifts, such as $\delta \geq 3.0$. This is understandable because the SREWMA, which is essentially based on ranks rather than distances, shares a similar drawback as those rank-based charts for univariate processes. That is, even though the shift is quite large, the ranks of the observations may not be able to grow larger.

Table 7 shows the ARL values of the SREWMA and SSEWMA charts in monitoring a shift in the first component of the multivariate mixed-components observations (iv). Clearly, with an appropriate value of λ, say 0.05 or 0.025, the SREWMA chart not only attains the desired IC ARL, but it also outperforms the SSEWMA chart in detecting small and moderate shifts, and in many cases the advantage is quite prominent. This demonstrates that the SREWMA chart is more sensitive to process shifts in non-normal observations, even for a distribution with different marginals, compared with the parametric self-starting chart.

In the foregoing examples, we fixed the value of change-point (τ) so that the number of IC observations collected before shifts is one hundred. Next, we study the performance of SREWMA by varying the values of change-point. To this end, the case with $p = 5$ and $m = 10$ is considered and the nominal IC ARL is fixed as 200. Figs. 3a and 3b summarize

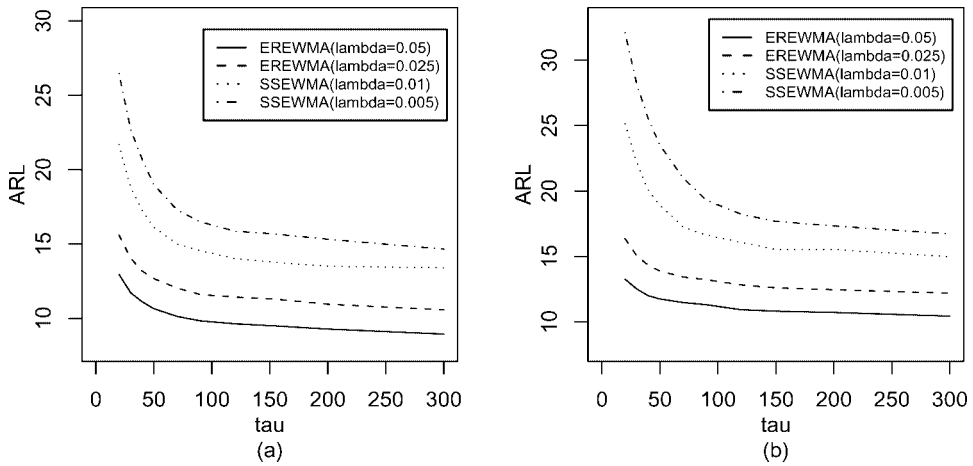

Figure 3. OC ARL curves with varying τ for SREWMA and SSEWMA and with $\delta = 1.5$, $p = 5$, $m_0 = 10$, and $\text{ARL}_0 = 200$: (a) OC multinormal; (b) OC multivariate t.

Table 8. ARL comparison between SREWMA and MSEWMA with $\lambda = 0.05$

		Normal		$t_{p,5}$		$\text{Gam}_{p,3}$	
		SREWMA	MSEWMA	SREWMA	MSEWMA	SREWMA	MSEWMA
$p = 5$	0.00	199 (1.84)	201 (1.85)	197 (1.84)	199 (1.83)	198 (1.83)	198 (1.78)
	0.25	64.9 (0.69)	68.4 (0.57)	76.2 (0.67)	72.9 (0.61)	80.3 (0.69)	82.9 (0.73)
	0.50	25.6 (0.15)	28.5 (0.17)	30.1 (0.19)	30.0 (0.19)	34.8 (0.23)	34.4 (0.22)
	0.75	15.7 (0.08)	17.3 (0.08)	18.6 (0.10)	18.8 (0.09)	20.4 (0.10)	20.1 (0.10)
	1.00	11.4 (0.05)	12.8 (0.05)	13.3 (0.06)	14.0 (0.06)	14.6 (0.06)	14.5 (0.06)
	1.50	7.69 (0.03)	9.02 (0.03)	8.81 (0.03)	9.85 (0.04)	9.47 (0.03)	9.90 (0.03)
	2.00	6.09 (0.02)	7.51 (0.02)	6.89 (0.02)	8.14 (0.03)	7.32 (0.02)	8.17 (0.03)
	3.00	4.79 (0.01)	6.40 (0.02)	5.28 (0.02)	6.82 (0.02)	5.52 (0.02)	6.87 (0.02)
	4.00	4.29 (0.01)	6.03 (0.02)	4.62 (0.01)	6.27 (0.02)	4.70 (0.01)	6.35 (0.02)
$p = 10$	0.00	198 (1.82)	201 (1.78)	197 (1.73)	200 (1.79)	197 (1.73)	199 (1.79)
	0.25	73.4 (0.63)	78.1 (0.67)	84.5 (0.74)	83.2 (0.73)	97.0 (0.90)	98.7 (0.91)
	0.50	29.8 (0.18)	32.2 (0.20)	34.4 (0.22)	34.7 (0.23)	43.6 (0.31)	44.5 (0.31)
	0.75	17.9 (0.09)	19.3 (0.09)	20.6 (0.11)	20.9 (0.11)	25.3 (0.14)	25.6 (0.14)
	1.00	12.8 (0.05)	13.9 (0.06)	14.3 (0.07)	15.1 (0.07)	17.6 (0.08)	17.9 (0.08)
	1.50	8.45 (0.03)	9.45 (0.03)	9.78 (0.04)	10.3 (0.04)	11.0 (0.04)	11.6 (0.04)
	2.00	6.52 (0.02)	7.55 (0.02)	6.76 (0.02)	8.20 (0.03)	8.14 (0.03)	8.96 (0.03)
	3.00	4.86 (0.01)	6.00 (0.02)	4.94 (0.01)	6.48 (0.02)	5.62 (0.02)	6.82 (0.02)
	4.00	4.24 (0.01)	5.46 (0.01)	4.56 (0.01)	5.80 (0.02)	4.98 (0.01)	6.02 (0.02)

Standard errors are in parentheses.

the ARL curves of the SREWMA and SSEWMA with a shift δ of 1.5 in the first component for the multinormal distribution and the $t_{5,5}$ distribution respectively. The conclusions drawn from these two figures are similar to those from Tables 3–6, but in addition we can observe the following: a control chart whose performance is worse would benefit a lot from an increase in τ. This finding holds for other magnitudes of δ, as shown through further simulations (not reported here).

We conducted other simulations with various correlation structures, p and IC ARL, to see if the above conclusions would change. These simulation results (not reported here) show that the SREWMA chart works well for other correlation structures as well as in terms of its OC ARL, and its performance continue to be good for other choices of p and IC ARL.

3.3. Comparisons Between SREWMA and MSEWMA

Although the major benefit of SREWMA is to be in self-starting situations, it is also important to compare the SREWMA and MSEWMA charts which can provide better understanding of the performance of SREWMA and the difference between spatial ranks and signs based charting schemes. As mentioned before, the MSEWMA chart cannot readily be extended to a self-starting version. To this end, we compare the two charts by assuming that m_0 is sufficiently large (equivalently, the IC parameters are known), in this case twenty thousand. In such situations, the SREWMA chart would essentially reduce to its theoretical counterpart (as the sequential update could hardly affect the charting statistic), TREWMA. Because a similar conclusion holds for other cases, we only present the results with $\lambda = 0.05$ for brevity. Again, the cases $p = 5$ and 10, are considered and the nominal IC ARL is fixed as 200. The control limits of both charts are the same (as justified by Proposition 2 and recommended in Section 2.5), being 12.646 and 20.288 for $p = 5$ and 10, respectively. Table 8 shows the ARL comparison between SREWMA and MSEWMA under the multivariate normal, $t_{p,5}$ and $\text{Gam}_{p,3}$ distributions. From this table, we can see that both charts can attain the nominal IC ARL for all the three distributions. For small shifts, the difference between the two charts is not significant. As the magnitude of shift increases, the SREWMA tends to perform better than MSEWMA, especially for $\delta \geq 2$. For example, the SREWMA generally has an ARL that is 15–20% smaller than that of the MSEWMA in detecting a shift size δ of 3.0. This phenomenon is partly due to the fact that the SREWMA chart incorporates more information on the relative magnitudes of vector observations than the MSEWMA chart which only uses the direction of observations from the origin. Of course, the SREWMA requires much more computational effort than MSEWMA.

4. A REAL-DATA APPLICATION

In this section, we demonstrate the proposed methodology by applying it to a real dateset from a white wine production process. The data set contains a total of 4898 observations, and is publicly available in the UC Irvine Machine Learning Repository (http://archive.ics.uci.edu/ml/datasets/Wine+Quality). The data were collected from May 2004

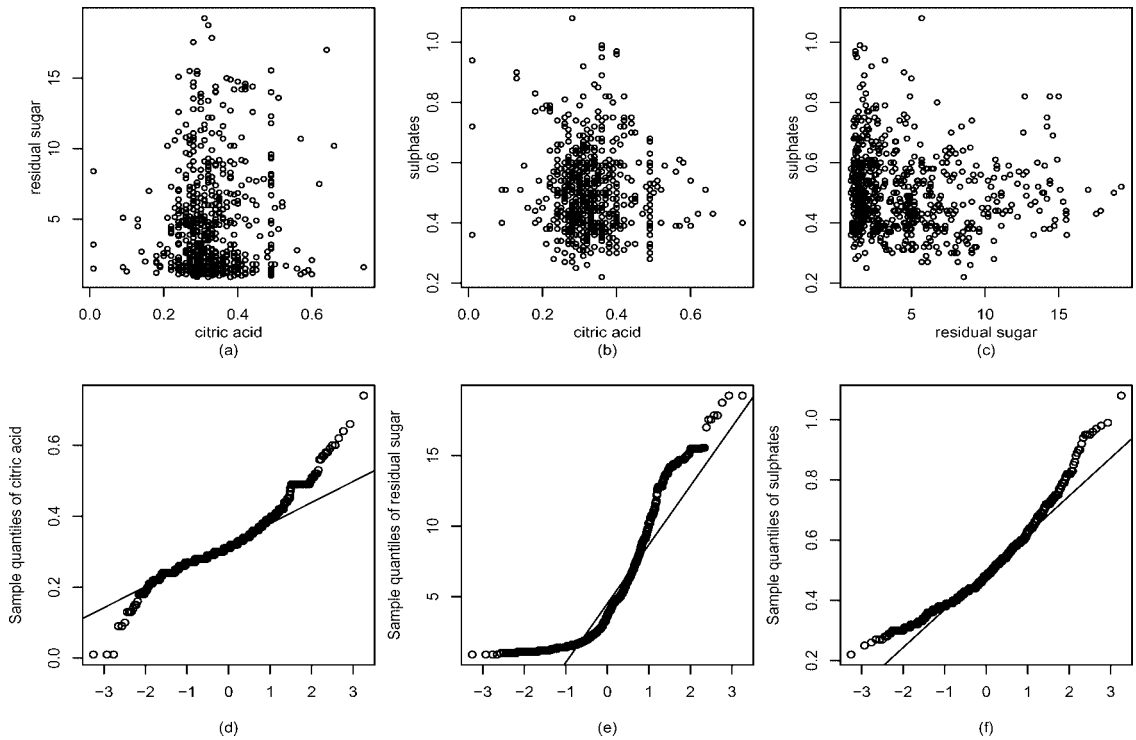

Figure 4. (a–c) The scatter plots of the white wine data; (d–f) the normal Q-Q plots for variables citric acid (x3), residual sugar (x4), and sulphates (x10), respectively.

to February 2007 using only protected designation of origin wine samples that were tested at an official certification entity, which is an interprofessional organization with the goal of improving the quality and marketing of Portuguese "Vinho Verde" wine. The data were recorded by a computerized system, which automatically manages the process of wine sample testing from producer requests to laboratory and sensory analysis. For each observation, there are eleven continuous measurements (based on physicochemical tests) including fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, and alcohol (denoted by x1, x2, . . . , x11, respectively). A categorical variable, quality, indicating the wine quality between 0 (very bad) and 10 (excellent) is also provided based on sensory analysis. The goal of this data analysis is mainly to model and monitor wine quality based on physicochemical tests. A more detailed discussion about this dataset is given by Cortez et al. [4] and the references therein.

As mentioned by Cortez et al. [4], it is desirable to set up an on-line detection system to monitor the production process of Vinho Verde wine to guarantee its quality. A natural method is to use a univariate control chart to monitor the (discrete) observations obtained from sensory analysis. However, those observations usually lag behind the real-time process and the collection exercise is rather expensive. So, it would be interesting to consider applying some multivariate control charts to those eleven continuous measurements collected automatically from physicochemical tests for early detection of abnormal production.

Under the SPC context of sequentially monitoring the wine production process, we assume that the standard quality level is 7 (LV7; as also suggested by Cortez et al. [4]). The sample correlation matrix of this data (not reported here) contains several large entries, which demonstrates that the variables have considerable interrelationships and consequently a multivariate control chart is likely to be more appropriate than a univariate control chart. Figs. 4a–4c show the scatter plots of the raw data for the three measurements, citric acid (x3), residual sugar (x4), and sulphates (x10), based on a total of 880 vectors belonging to this level. The joint distribution of each pair of variables are far from bivariate normal. The normal Q-Q plots for these three distributions are shown in Figs. 4d–f, which clearly indicate that the marginals are not normal either. The Shapiro-Wilk goodness-of-fit tests for normality conclude that none of these three variables is normally distributed (all the *P*-values are smaller than 1×10^{-5}). Mardia's [17] multivariate normality test is also performed and the *P*-value is about 4.63×10^{-9}. All these tests together with Figs. 4a–f suggest that the multivariate normality assumption is invalid and thus we could expect the nonparametric chart to be more robust and powerful than

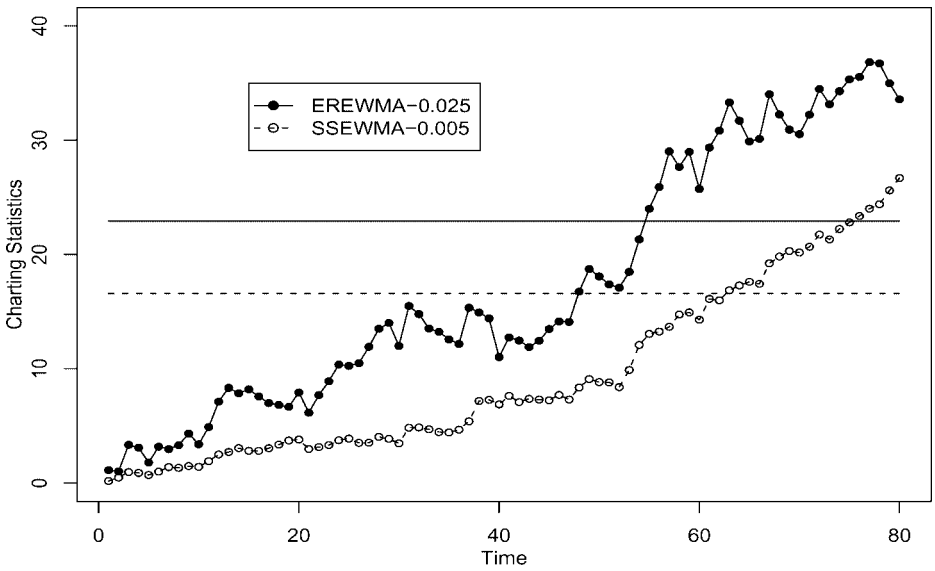

Figure 5. The SREWMA and SSEWMA control charts for monitoring the white wine production process, along with the solid and dashed horizontal lines indicating their control limits, respectively.

normal-based approaches for this dataset. It is also worth noting that although it is usually easy to collect observations from physicochemical tests, obtaining sufficiently large IC reference sample in this process is difficult: sensory tests rely mainly on human experts, and thus are rather time-consuming and expensive. Once a new technology in wine making is used or some improvements are made in the production process, one usually wants to monitor the process at the start-up stages in which only a small reference sample (through sensory tests) would be available. Therefore, self-starting control charts would be more desirable in this situation.

Next, we assume that we have $m_0 = 20$ historical observations from LV7 and initially monitored 30 observations from LV7 and then obtained the LV6 observations sequentially. Similar to Cortez et al. [4], the location parameter is of the greatest interest and thus we construct the SSEWMA and SREWMA control charts to monitor the wine quality. The IC ARL is fixed at 500, and the values of λ are chosen to be 0.025 and 0.005 for SREWMA and SSEWMA respectively to ensure their IC robustness to this non-normal data. Figure 5 shows the resulting SREWMA chart (solid curve connecting the dots) along with its control limit of 22.918 (the solid horizontal line). The corresponding SSEWMA (dashed curve connecting circles) is also presented in the figure, along with its control limit of 16.574 (dashed horizontal line). From the plot, it can be seen that the SREWMA chart exceeds its control limit from around the 55th observation (the 25th OC LV6 observation) onwards. This excursion suggests that a marked step-change has occurred as we intended. In comparison, the SSEWMA chart does not give any signal until the 63-rd observation, which is a delay of eight observations. Therefore, although requiring more computational effort, the SREWMA is a reasonable alternative for nonmultinormal processes if we take its efficiency, convenience and robustness into account. It is worth pointing out here that apart from quick detecting abnormal changes, isolating the shifted components or factors that are responsible for the change is also a fundamental task of MSPC. In this application, it would be interesting and helpful to determine which physicochemical factor(s) are responsible for the change of quality. The LASSO-based post-signal diagnostic method proposed by Zou and Qiu [37] could be used after the proposed SREWMA chart triggers a signal.

5. CONCLUDING REMARKS

In this article, we have proposed a multivariate nonparametric control scheme. This scheme integrates a powerful multivariate spatial rank test with EWMA process monitoring. It is a self-starting scheme which simultaneously updates process information and checks for OC conditions. Unlike parametric or nonparametric control schemes which assume the parameters or distributions are known before monitoring, this control scheme can be easily implemented even if the underlying process distribution is unknown and a lengthy data-gathering step can be avoided. Compared with parametric self-starting schemes, this scheme is not only much more robust in IC performance, but it is also generally more sensitive to the small and moderate shifts in location parameters for skewed and heavy-tailed multivariate distributions. In many cases, the improvement is quite remarkable. The drawback

of this scheme, which is common to almost all rank-based (univariate) nonparametric schemes, is that it is not as efficient as parametric schemes for very large shifts because it only uses the direction of observations from the origin and the relative magnitudes rather than the original magnitudes of observations. This disadvantage is mainly due to the trade-off between robustness and sensitivity.

Hawkins et al.'s [9] change-point scheme for on-line monitoring can also be seen as a self-starting method. Zhou et al. [36] and Hawkins and Deng [6] extended this strategy to the nonparametric univariate setting by utilizing the two-sample Mann-Whitney test statistics. In an ongoing effort, we are developing a control scheme that integrates sequential change-point detection with the two-sample spatial rank test, which might be expected to be more robust in detecting various magnitudes of shifts and in further alleviating the masking effect with a certain computational effort. Moreover, the current version of the proposed scheme is designed for detecting location shifts only. We believe that, after certain modifications, the proposed method should be able to monitor both the location and covariance structure (c.f., Huwang et al. [11]). It would also be of interest to study the performance of SREWMA in a high-dimensional monitoring environment and to investigate how to improve its efficiency by using some variable selection techniques (Zou and Qiu [37]).

APPENDIX: PROOFS

PROOF OF PROPOSITION 1: To prove this proposition is equivalent to showing that for any $p \times p$ nonsingular matrix $\mathbf{D}$ and constant vector $\mathbf{b}$, the charting statistics, $Q_t^{R_F}$, based on $\mathbf{x}_t$ and $\mathbf{y}_t = \mathbf{D}\mathbf{x}_t + \mathbf{b}$ are the same. In what follows, we use the subscripts "$\mathbf{y}$" to distinguish the corresponding statistics or parameters based on the sample $\mathbf{y}_t$, i.e., $Q_{t,\mathbf{y}}^{R_F}$.

First of all, by $\mathbf{S}_\mathbf{y} = \mathbf{D}\mathbf{S}\mathbf{D}^T$, $||\mathbf{M}(\mathbf{x}_t - \mathbf{x}_j)|| = ||\mathbf{M}_\mathbf{y}(\mathbf{y}_t - \mathbf{y}_j)||$. By the definition of $R_F(\cdot)$,

$$R_F(\mathbf{M}_\mathbf{y}\mathbf{y}_t) = \mathbf{M}_\mathbf{y}\mathbf{D}\mathbf{M}^{-1}R_F(\mathbf{M}\mathbf{x}_t).$$

Thus, it can be seen that

$$\begin{aligned}\mathrm{cov}[R_F(\mathbf{M}_\mathbf{y}\mathbf{y}_t)] &= (\mathbf{M}_\mathbf{y}\mathbf{D}\mathbf{M}^{-1})\mathrm{cov}[R_F(\mathbf{M}\mathbf{x}_t)](\mathbf{M}_\mathbf{y}\mathbf{D}\mathbf{M}^{-1})^T,\\ \mathbf{w}_{t,\mathbf{y}} &= \mathbf{M}_\mathbf{y}\mathbf{D}\mathbf{M}^{-1}\mathbf{w}_t.\end{aligned}$$

By the definition of $Q_t^{R_F}$, the result follows immediately. □

PROOF OF PROPOSITION 2: 1. The distribution of $\boldsymbol{\varepsilon}$ is spherically symmetrical around the origin if the density function $f(\boldsymbol{\varepsilon})$ of $\boldsymbol{\varepsilon}$ depends on $\boldsymbol{\varepsilon}$ through the modulus $||\boldsymbol{\varepsilon}||$. We can then write

$$f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}) = \exp\{-\rho(||\boldsymbol{\varepsilon}||)\},$$

for some function $\rho(\cdot)$. By the definition of model (1), we know that $\mathbf{M}(\mathbf{x}_t - \boldsymbol{\mu})$ has a spherically symmetrical distribution as well. It is easy to see that $R_F(\cdot)$ is a location-invariant transformation. Thus, we have

$$R_F(\mathbf{M}(\mathbf{x}_t - \boldsymbol{\mu})) = R_F(\mathbf{M}(\mathbf{x}_t)).$$

Thus, we may use $R_F(\mathbf{M}(\mathbf{x}_t - \boldsymbol{\mu}))$ to replace $R_F(\mathbf{M}(\mathbf{x}_t))$ in the construction of $Q_t^{R_F}$ (although $\boldsymbol{\mu}$ is unknown to us in practice, it does not matter in our theoretical analysis at all).

By Theorem 4.3 of Oja [20], we have

$$R_F(\mathbf{M}(\mathbf{x}_t - \boldsymbol{\mu})) = q_F(r_t)\mathbf{u}_t, \tag{A.1}$$

where $r_t = ||\boldsymbol{\varepsilon}_t||$, $\mathbf{u}_t = ||\boldsymbol{\varepsilon}_t||^{-1}\boldsymbol{\varepsilon}_t$ and $q_F(r)$ is a scalar function that depends on $\rho(\cdot)$ and r. The direction vector $\mathbf{u}_t$ is uniformly distributed on the p-dimensional unit sphere, $S(1)$, where $S(r)$ denotes the p-dimensional sphere of radius $r > 0$. Moreover, the radius r_t and direction $\mathbf{u}_t$ are independent, and $E(\mathbf{u}_t) = 0$ and $\mathrm{cov}(\mathbf{u}_t) = \mathbf{I}_p/p$. It is then easy to see that

$$\mathrm{cov}(R_F(\mathbf{M}(\mathbf{x}_t - \boldsymbol{\mu}))) = E_r\big[q_F^2(r)\big]\mathbf{I}_p/p,$$

where $E_r[\cdot]$ means that the expectation is taken with respect to the random variable $||\boldsymbol{\varepsilon}||$. Thus, $Q_t^{R_F}$ can be re-written as

$$Q_t^{R_F} = \frac{a}{E_r\big[q_F^2(r)\big]}||\mathbf{w}_t||^2,$$

where $a = (2-\lambda)p/\lambda$.

Next, we show that the distribution of $\mathbf{w}_t$ given $||\mathbf{w}_1||, \ldots, ||\mathbf{w}_t||$ is uniform on $S(||w_t||)$. From the distribution of $R_F(\mathbf{M}(\mathbf{x}_1 - \boldsymbol{\mu}))$ given by (A.1), this is true for $\mathbf{w}_1$. By the induction assumption, the distribution of $\mathbf{w}_{t-1}$ given $||\mathbf{w}_1||, \ldots, ||\mathbf{w}_{t-1}||$ is uniform on $S(||\mathbf{w}_{t-1}||)$. The conditional distribution of $\mathbf{w}_t$ given $||\mathbf{w}_1||, \ldots, ||\mathbf{w}_t||$ is thus the same as

$$(1-\lambda)||\mathbf{w}_{t-1}||\mathbf{u} + \lambda q_F(r_t)\mathbf{u}_t,$$

where $\mathbf{u}$ is on $S(1)$ and is independent of $\mathbf{u}_t$. It follows that the conditional distribution of $\mathbf{w}_t$ is uniform and that $\mathbf{w}_t$ given $||\mathbf{w}_1||, \ldots, ||\mathbf{w}_t||$ is on $S(||\mathbf{w}_t||)$.

Note that

$$\begin{aligned}&\Pr\big\{Q_t^{R_F} < l | Q_1^{R_F}, \ldots, Q_{t-1}^{R_F}\big\}\\ &\quad= \Pr\left\{\frac{a}{E_r\big[q_F^2(r)\big]}||\lambda R_F(\mathbf{M}(\mathbf{x}_t - \boldsymbol{\mu})) + (1-\lambda)\mathbf{w}_{t-1}||^2 \right.\\ &\qquad\left. < l \middle| Q_1^{R_F}, \ldots, Q_{t-1}^{R_F}\right\}\\ &\quad= \Pr\left\{\frac{a}{E_r\big[q_F^2(r)\big]}||(1-\lambda)||\mathbf{w}_{t-1}||\mathbf{u} + \lambda q_F(r_t)\mathbf{u}_t||^2 < l \middle| Q_{t-1}^{R_F}\right\}\end{aligned}$$

from which the result follows immediately.

2. By (i), the run-lengths of TREWMA can be approximated by mimicking the procedure used for MSEWMA which is detailed in the Appendix of Zou and Tsung [39]. The only difference lies in the calculation of transition probabilities in (A.2). By using again the fact that r_t and $\mathbf{u}_t$ are independent, this conditional probability distribution is just the distribution of a sum of two independent variables whose distributions are explicitly known to us. Hence, the details are omitted but are available from the authors upon request. □

PROOF OF PROPOSITION 3: 1. We only prove the result for the case where $k = t + 1$, because the proofs for other cases where $k > t$ are similar.

Note that $E[U(\mathbf{M}(\mathbf{x}_t - \mathbf{x}_j))] = 0$ for $j \neq t$. Denote $l = m_0 + t - 1$ and it is straightforward to see

$$\begin{aligned}
&\operatorname{cov}(R_E(\mathbf{M}_{t-1}\mathbf{x}_t), R_E(\mathbf{M}_t\mathbf{x}_{t+1})) \\
&= \operatorname{cov}\left(\frac{1}{l}\sum_{j=-m_0+1}^{t-1} U(\mathbf{M}(\mathbf{x}_t - \mathbf{x}_j)), \frac{1}{l+1}\sum_{j=-m_0+1}^{t} U(\mathbf{M}(\mathbf{x}_{t+1} - \mathbf{x}_j))\right) \\
&= \frac{1}{l(l+1)}\sum_{j=-m_0+1}^{t-1} \{E[U(\mathbf{M}(\mathbf{x}_t - \mathbf{x}_j))U(\mathbf{M}(\mathbf{x}_{t+1} - \mathbf{x}_j))] \\
&\quad + E[U(\mathbf{M}(\mathbf{x}_t - \mathbf{x}_j))U(\mathbf{M}(\mathbf{x}_{t+1} - \mathbf{x}_t))]\}.
\end{aligned}$$

By the assumption that the $\mathbf{x}_t$'s are i.i.d., we have for $j \leq t - 1$,

$$\begin{aligned}
E[U(\mathbf{M}(\mathbf{x}_t - \mathbf{x}_j))U(\mathbf{M}(\mathbf{x}_{t+1} - \mathbf{x}_t))] &= -E[U(\mathbf{M}(\mathbf{x}_j - \mathbf{x}_t))U(\mathbf{M}(\mathbf{x}_{t+1} - \mathbf{x}_t))] \\
&= -E[U(\mathbf{M}(\mathbf{x}_t - \mathbf{x}_j))U(\mathbf{M}(\mathbf{x}_{t+1} - \mathbf{x}_j))],
\end{aligned}$$

which clearly leads to $\operatorname{cov}(R_E(\mathbf{M}_{t-1}\mathbf{x}_t), R_E(\mathbf{M}_t\mathbf{x}_{t+1})) = 0$.

2. As $t \to \infty$,

$$R_E(\mathbf{M}_{t-1}\mathbf{x}_t) \xrightarrow{p} R_F(\mathbf{M}\mathbf{x}_t) \equiv \boldsymbol{\alpha}_t,$$

by Theorem 4.2 of Oja [20] and the continuous mapping theorem. Note that $E(\boldsymbol{\alpha}_t) = 0$ and $\operatorname{cov}(\boldsymbol{\alpha}_t) = p^{-1}E[||R_F(\mathbf{M}\mathbf{x}_t)||^2]\mathbf{I}_p$ under the spherical assumption.

Write $\mathbf{v}_t$ as $\mathbf{v}_t = \sum_{j=1}^{t} \lambda(1-\lambda)^{t-j}\boldsymbol{\alpha}_t(1 + o_p(1))$. By using the fact that the $\boldsymbol{\alpha}_t$'s are i.i.d., this proposition follows immediately from the Hajek-Sidak central limit theorem. □

ACKNOWLEDGMENTS

The authors thank the editor, associate editor, and two anonymous referees for their many helpful comments that have resulted in significant improvements in the article. This research was supported by the National Natural Science Foundation of China Grants 11001138, 11071128, 11131002, 11101306 and RGC Competitive Earmarked Research Grants 620010, RPC10EG15 and the RFDP of China Grant 20110031110002. Zou also thanks for the support of Nankai Young Grant 65010731 and the National Center for Theoretical Sciences, Math Division.

REFERENCES

[1] R.L. Brown, J. Durbin, and J.M. Evans, Techniques for testing the constancy of regression relationships over time, J R Stat Soc B 37 (1975), 149–192.
[2] S. Chakraborti, P. Van der Laan, and S.T. Bakir, Nonparametric control charts: an overview and some results, J Qual Technol 33 (2001), 304–315.
[3] C.W. Champ, L.A. Jones-Farmer, and S.E. Rigdon, Properties of the T^2 control chart when the parameters are estimated, Technometrics 47 (2005), 437–445.
[4] P. Cortez, A. Cerdeira, F. Almeida, T. Matos, and J. Reis, Modeling wine preferences by data mining from physicochemical properties, Decision Support Syst 47 (2009), 547–553.
[5] R.B. Croisier, Multivariate generalizations of cumulative sum quality-control schemes, Technometrics 30 (1988), 243–251.
[6] D.M. Hawkins and Q. Deng, A nonparametric change point control chart, J Qual Technol 42 (2010), 165–173.
[7] D.M. Hawkins and E.M. Maboudou-Tchao, Self-starting multivariate exponentially weighted moving average control charting, Technometrics 49 (2007), 199–209.
[8] D.M. Hawkins and D.H. Olwell, Cumulative sum charts and charting for quality improvement, Springer-Verlag, New York, 1998.
[9] D.M. Hawkins, P. Qiu, and C.W. Kang, The changepoint model for statistical process control, J Qual Technol 35 (2003), 355–366.
[10] T.P. Hettmansperger and R.H. Randles, A practical affine equivariant multivariate median, Biometrika 89 (2002), 851–860.
[11] L. Huwang, Y. Wang, A.B. Yeh, and Z. Chen, On the exponentially weighted moving variance, Nav Res Logistic 56 (2009), 659–668.
[12] W.A. Jensen, L.A. Jones, C.W. Champ, and W.H. Woodall, Effects of parameter estimation on control chart properties: A literature Review, J Qual Technol 38 (2006), 349–364.
[13] L.A. Jones, The statistical design of EWMA control charts with estimated parameters, J Qual Technol 34 (2002), 277–288.
[14] R. Liu, Control charts for multivariate processes, J Am Stat Assoc 90 (1995), 1380–1388.
[15] C.A. Lowry, W.H. Woodall, C.W. Champ, and S.E. Rigdon, Multivariate exponentially weighted moving average control chart, Technometrics 34 (1992), 46–53.
[16] J.M. Lucas and M.S. Saccucci, Exponentially weighted moving average control scheme properties and enhancements, Technometrics 32 (1990), 1–29.
[17] K.V. Mardia, Measures of multivariate skewness and kurtosis with applications, Biometrika 57 (1970), 519–530.
[18] J. Möttönen and H. Oja, Multivariate spatial sign and rank methods, J Nonparametric Stat 5 (1995), 201–213.
[19] H. Oja and R. Randles, Multivariate nonparametric tests, Stat Sci 19 (2004), 598–605.
[20] H. Oja, Multivariate nonparametric methods with R, Springer, New York, 2010.
[21] S.S. Prabhu and G.C. Runger, Designing a multivariate EWMA control chart, J Qual Technol 29 (1997), 8–15.
[22] P. Qiu, Distribution-free multivariate process control based on log-linear modeling, IIE Trans 40 (2008), 664–677.
[23] P. Qiu and D.M. Hawkins, A rank-based multivariate CUSUM procedure, Technometrics 43 (2001), 120–132.
[24] P. Qiu and D.M. Hawkins, A nonparametric multivariate CUSUM procedure for detecting shifts in all directions, J R Stat Soc Ser D 52 (2003), 151–164.
[25] C.P. Quesenberry, SPC Q charts for start-up processes and short or long runs, J Qual Technol 23 (1991), 213–224.
[26] C.P. Quesenberry, On properties of Q charts for variables, J Qual Technol 27 (1995), 184–203.
[27] C.P. Quesenberry, SPC methods for quality improvement, Wiley, New York, 1997.
[28] R.H. Randles, A simpler, affine invariant, multivariate, distribution-free sign test, J Am Stat Assoc 95 (2000), 1263–1268.
[29] G.C. Runger and S.S. Prabhu, A Markov chain model for the multivariate exponentially weighted moving averages control chart, J Am Stat Assoc 91 (1996), 1701–1706.
[30] Z.G. Stoumbos, M.R. Reynolds, T.P. Ryan, and W.H. Woodall, The state of statistical process control as we proceed into the 21st century, J Am Stat Assoc 95 (2000), 992–998.

[31] Z.G. Stoumbos and J.H. Sullivan, Robustness to non-normality of the multivariate EWMA control chart, J Qual Technol 34 (2002), 260–276.
[32] J.H. Sullivan and L.A. Jones, A self-starting control chart for multivariate individual observations, Technometrics 44 (2002), 24–33.
[33] D.E. Tyler, A distribution-free M-estimator of multivariate scatter, Ann Stat 15 (1987), 234–251.
[34] K.D. Zamba and D.M. Hawkins, A multivariate change-point for statistical process control, Technometrics 48 (2006), 539–549.
[35] P.F. Zantek and S.T. Nestler, Performance and properties of Q-statistic monitoring schemes, Nav Res Logistic 56 (2009), 279–292.
[36] C. Zhou, C. Zou, Y. Zhang, and Z. Wang, Nonparametric control chart based on change-point model, Stat Pap 50 (2009), 13–28.
[37] C. Zou and P. Qiu, Multivariate statistical process control using LASSO, J Am Stat Assoc 104 (2009), 1586–1596.
[38] C. Zou and F. Tsung, Likelihood ratio based distribution-free EWMA schemes, J Qual Technol 42 (2010), 174–196.
[39] C. Zou and F. Tsung, A multivariate sign EWMA control chart, Technometrics 53 (2011), 84–97.
[40] C. Zou, Y. Zhang, and Z. Wang, Control chart based on change-point model for monitoring linear profiles, IIE Trans 38 (2006), 1093–1103.