Control Chart For Monitoring Nonparametric Profiles With Arbitrary Design

Peihua Qiu[1] and Changliang Zou[2]

[1]*School of Statistics, University of Minnesota, USA*

[2]*LPMC and Department of Statistics, Nankai University, China*

Abstract

Nonparametric profile monitoring (NPM) is for monitoring over time a functional relationship between a response variable and one or more explanatory variables when the relationship is too complicated to be specified parametrically. It is widely used in industry for the purpose of quality control of a process. Existing NPM approaches require a fundamental assumption that design points within a profile are deterministic (i.e., non-random) and they are unchanged from one profile to another. In practice, however, different profiles often have different design points, and in some cases they might even be random. NPM is particularly challenging in such cases because it is difficult to combine data in different profiles properly for data smoothing and for process monitoring. In this paper, we propose a novel exponentially weighted moving average (EWMA) control chart for handling this problem, based on local linear kernel smoothing. In the proposed chart, the exponential weights used in the EWMA scheme at different time points are integrated into a nonparametric smoothing procedure for smoothing individual profiles. Because of certain good properties of the charting statistic, this control chart is fast to compute, easy to implement, and efficient to detect profile shifts. Some numerical results show that it works well in applications.

Keywords: Bandwidth Selection; EWMA; Local Linear Kernel Smoothing; Nonparametric Regression; Profile Monitoring; Self-Starting; Statistical Process Control.

1 Introduction

In many applications, quality of a process is characterized by the functional relationship between a response variable and one or more explanatory variables. Profile monitoring is mainly for checking stability of this relationship (or profile) over time. In some calibration applications, the profile can be described adequately by a linear regression model, while in other applications more flexible models are necessary. This paper focuses on nonparametric profile monitoring (NPM) when the profile is too complicated to be specified parametrically.

In the literature, some existing references focus on linear profile monitoring. See, for instance, Kang and Albin (2000), Kim *et al.* (2003), Mahmoud and Woodall (2004), Zou *et al.* (2006, 2007b), and Mahmoud *et al.* (2007), among several others. Extensions to multiple and/or polynomial profile models are discussed by Zou *et al.* (2007a), and Kazemzadeh *et al.* (2008). Recently, nonlinear profile models have been considered by some statisticians, including Lada *et al.* (2002), Ding *et al.* (2006), Colosimo and Pacella (2007) and Williams *et al.* (2007a,b). NPM is discussed by Zou *et al.* (2008, 2009). For an overview on profile monitoring, see Woodall *et al.* (2004).

All the control charts mentioned above require the fundamental assumptions that design points within a profile are deterministic (i.e., non-random) and that they are the same from one profile to another. These assumptions are (approximately) valid in certain calibration applications of the manufacturing industry. In some other applications, however, they may be invalid. For instance, when data acquisition takes the *random design* scheme, design points within a profile would be i.i.d. random variables from a given distribution. Another commonly seen example occurs when observations within different profiles have missing values at different time points (e.g., the vertical-density profile (VDP) data considered in Walker and Wright 2002). Furthermore, we will demonstrate in this paper by both theoretical and empirical results that, even for applications where an equal design scheme (i.e., design points are the same from profile to profile and they are deterministic) is possible, one may get a better profile monitoring by using a random design scheme, as long as the two design schemes involve similar measurement effort.

In this article, we propose a novel control chart for handling the NPM problem when the profile design points are arbitrary. The proposed chart is based on local linear kernel smoothing of individual profile data and on the exponentially weighted moving average (EWMA) process control scheme as well. It incorporates the exponential weights used in the EWMA scheme at different time

points into the local linear kernel smoother. We show that this chart is effective in detecting profile shifts when profile design points are arbitrary. It is also fast to compute and easy to implement. This chart is described in detail in Section 2. Its numerical performance is investigated in Section 3. In Section 4, we demonstrate the method using a real-data example from the semiconductor manufacturing industry. Several remarks conclude the article in Section 5. Some technical details are provided in the Appendix.

2 Methodology

Our proposed methodology is described in seven parts. In Section 2.1, we briefly introduce the statistical process control (SPC) problem and the well known EWMA control chart. Then, in Section 2.2, an EWMA control chart accommodating nonparametric regression of individual profiles is introduced for monitoring nonparametric profiles with arbitrary design. Adaptive selection of its weighting parameter and bandwidth parameter, used in EWMA and nonparametric regression, are respectively discussed in Sections 2.3 and 2.4. Certain computational issues are addressed in Section 2.5. A self-starting version is given in Section 2.6. Finally, some practical guidelines regarding design and implementation of the proposed control chart are provided in Section 2.7.

2.1 Statistical process control and the EWMA control chart

SPC is for monitoring sequential processes (e.g., production lines in manufacturing industry) to make sure that they work stably. When the process works stably, it is in the in-control (IC) state, and it becomes out-of-control (OC) otherwise. In the literature, SPC is often divided into two phases. In Phase I, a set of process data is gathered and analyzed. Any unusual "patterns" in the data lead to adjustments and fine tuning of the process. Once all such assignable causes are accounted for, we are left with a clean set of data, gathered under stable operating conditions and illustrative of the actual process performance. This set is then used for estimating the IC distribution of the process measurements. In Phase II, the estimated IC measurement distribution from Phase I data is used, and the major goal of this phase is to detect any shift in the measurement distribution from the IC distribution after an unknown time point. Performance of a Phase II SPC procedure is often measured by the average run length (ARL), which is the average number of samples obtained at sequential time points that are needed for the procedure to signal a shift

in the measurement distribution. The IC ARL value of the procedure is usually controlled at a certain level, and the procedure performs better if its OC ARL is smaller when detecting a given shift, which is in parallel to the type-I and type-II error probabilities in hypothesis testing. In the literature, most SPC control charts are for Phase II process monitoring which is also the focus of the current paper.

Let $\{X_k, k = 1, 2, \ldots\}$ be the sequential, Phase II, univariate, process measurements. Then, the well-known EWMA control chart is based on the following sequence of statistics

$$S_k = (1 - \lambda)S_{k-1} + \lambda X_k, \quad \text{for } k = 1, 2, \ldots,$$

where $S_0 = 0$ and $\lambda \in [0, 1]$ is a weighting parameter. It signals a shift at the k-th time point if $S_k > L$, where L is a control limit chosen to achieve a given IC ARL value. Obviously, $S_k = \lambda X_k + \lambda(1 - \lambda)X_{k-1} + \cdots + \lambda(1 - \lambda)^{k-1}X_1$. So, S_k is a weighted average of all observations, more recent observations would receive more weight in the average, and the weight changes exponentially over time.

2.2 Monitoring nonparametric profiles when design points are arbitrary

In this paper, we are concerned about Phase II profile monitoring. At the k-th time point, the profile is assumed to follow the nonparametric model

$$y_{kj} = g\left(x_{kj}\right) + \varepsilon_{kj}, \quad j = 1, \ldots, n_k, \quad k = 1, 2, \ldots, \tag{1}$$

where $\{x_{kj}, y_{kj}\}_{j=1}^{n_k}$ are the k-th profile data, x_{kj} is the j-th design point in the k-th profile, g is a smooth nonparametric profile function, and ε_{kj}s are i.i.d. random errors with mean 0 and variance σ^2. Without loss of generality, we assume that $x_{kj} \in [0, 1]$, for all k and j. In cases when the design points $\boldsymbol{X}_k = \{x_{k1}, x_{k2}, \ldots, x_{kn_k}\}$ are unchanged from one profile to another, the nonparametric EWMA chart by Zou *et al.* (2008), which is called the NEWMA chart hereafter, first averages observed responses y_{kj}s across different profiles at each design point and then detects potential profile shifts using the generalized likelihood ratio (GLR) test statistic. This idea can not be applied to the current problem because the response is observed at different design points in different profiles in the current setup of the problem. A naive modification to Zou *et al.*'s method is to first obtain a nonparametric estimate of g from each profile data, and then predict response values using the estimated g at some points $\{z_1, z_2, \ldots, z_n\}$ in the design interval which are unchanged

from one profile to another. Then, the NEWMA chart can be applied to the predicted response values. However, this naive approach, which is called the NAEWMA chart hereafter, may not be efficient, due to the facts that only n_k observations are used for estimating g in the k-th profile, n_k could be very small, and thus the predicted response values could have large bias and variance. As an alternative, in this paper, we consider using the following weighted local likelihood at any point $z \in [0,1]$, which combines the exponential weighting scheme used in EWMA at different time points (i.e., through the term $(1-\lambda)^{t-k}$ in the expression below) with the local linear kernel smoothing procedure (cf., Fan and Gijbels 1996):

$$WL(a,b;z,\lambda,t) = \sum_{k=1}^{t}\sum_{j=1}^{n_k}[y_{kj} - a - b(x_{kj}-z)]^2 K_h\left(x_{kj}-z\right)(1-\lambda)^{t-k},$$

where t is the current time point for profile monitoring, $K_h(\cdot) = K(\cdot/h)/h$, K is a symmetric density kernel function, $\lambda \in [0,1]$ is a weighting parameter, and h is a bandwidth. Then, the local linear kernel estimator of $g(z)$, defined as the solution to a in the minimization problem $\min_{a,b} WL(a,b;z,\lambda,t)$, has the expression

$$\widehat{g}_{t,h,\lambda}(z) = \sum_{k=1}^{t}\sum_{j=1}^{n_k} U_{kj}^{(t,h,\lambda)}(z) y_{kj} \Big/ \sum_{k=1}^{t}\sum_{j=1}^{n_k} U_{kj}^{(t,h,\lambda)}(z), \tag{2}$$

where

$$\begin{aligned} U_{kj}^{(t,h,\lambda)}(z) &= (1-\lambda)^{t-k} K_h(x_{kj}-z)\left[m_2^{(t,h,\lambda)}(z) - (x_{kj}-z)m_1^{(t,h,\lambda)}(z)\right], \\ m_l^{(t,h,\lambda)}(z) &= \sum_{k=1}^{t}(1-\lambda)^{t-k}\sum_{j=1}^{n_k}(x_{kj}-z)^l K_h(x_{kj}-z), \quad l=0,1,2. \end{aligned} \tag{3}$$

From the expression of $WL(a,b;z,\lambda,t)$ above, we can see that this estimator makes use of all available observations up to the current (i.e., the t-th) time point, and different profiles are weighted as in a conventional EWMA chart (i.e., more recent profiles get more weight and the weight changes exponentially over time).

If the process under monitoring is IC up to the t-th time point, then $\widehat{g}_{t,h,\lambda}$ in (2) should be close to the IC profile function denoted as g_0. Therefore, a charting statistic for profile monitoring can be defined based on the difference between $\widehat{g}_{t,h,\lambda}$ and g_0. For simplicity, let us first assume that g_0 and the error variance σ^2 are both known. In such cases, a more convenient way to define the charting statistic is to use $\widehat{\xi}_{t,h,\lambda}(z)$, which is the estimator defined by (2), after y_{kj} are replaced by $\xi_{kj} = [y_{kj} - g_0(x_{kj})]/\sigma$, for all k and j. Then, $\widehat{\xi}_{t,h,\lambda}(z)$ should be uniformly close to 0 when the

process is IC up to the t-th time point. A natural charting statistic for profile monitoring is then defined by

$$T_{t,h,\lambda} = c_{0,t,\lambda} \int [\widehat{\xi}_{t,h,\lambda}(z)]^2 \Gamma_1(z)\, dz,$$

where

$$c_{t_0,t_1,\lambda} = a^2_{t_0,t_1,\lambda}/b_{t_0,t_1,\lambda}, \quad a_{t_0,t_1,\lambda} = \sum_{k=t_0+1}^{t_1} (1-\lambda)^{t_1-k} n_k, \quad b_{t_0,t_1,\lambda} = \sum_{k=t_0+1}^{t_1} (1-\lambda)^{2(t_1-k)} n_k,$$

and Γ_1 is some pre-specified positive density function. In the expression of $T_{t,h,\lambda}$, the scale parameter $c_{0,t,\lambda}$ is used for unifying its asymptotic variance (see Theorem 1 below and its proof in the Appendix). In practice, we can use the following approximation:

$$T_{t,h,\lambda} \approx \frac{c_{0,t,\lambda}}{n_0} \sum_{i=1}^{n_0} \left[\widehat{\xi}_{t,h,\lambda}(z_i)\right]^2, \tag{4}$$

where z_i, for $i = 1, \ldots, n_0$, are some pre-specified i.i.d. design points from Γ_1. Then, the control chart triggers a signal if $T_{t,h,\lambda} > L$, where $L > 0$ is a control limit chosen to achieve a specific IC ARL. Hereafter, this chart is referred to as the nonparametric profile control (NPC) chart.

It should be pointed out that it is computationally faster to use points z_i rather than the original design points x_{kj} in approximating the statistic $T_{t,h,\lambda}$. As shown in Section 2.5 below, $T_{t,h,\lambda}$ can be calculated in a recursive manner when z_i are used in the approximation, and it does not enjoy such a feature when x_{kj} are used. Further, from theoretical properties of $T_{t,h,\lambda}$ given in Theorem 2 below and certain empirical results presented in Section 3, selection of z_i and n_0 has little effect on the performance of the NPC chart as long as n_0 is not too small. See related discussion in Section 2.7 about practical guidelines on selection of certain procedure parameters.

As a remark, one may define $T_{t,h,\lambda}$ alternatively by

$$\frac{c_{0,t,\lambda}}{n_0} \sum_{i=1}^{n_0} [\widehat{g}_{t,h,\lambda}(z_i) - g_0(z_i)]^2. \tag{5}$$

Namely, we can first compute profile estimators $\widehat{g}_{t,h,\lambda}$ from the original data and then construct the control chart accordingly. It can be shown that (5) and (4) are asymptotically equivalent under some regularity conditions given in Appendix A. However, in finite-sample cases, properties of (5) depend on g_0. As a comparison, formula (4) transforms the testing problem of $H_0 : g = g_0$ versus $H_1 : g \neq g_0$ to the one of $H_0 : g = 0$ versus $H_1 : g \neq 0$. Therefore, *it is invariant to* g_0. Its IC distribution and all quantities related to this distribution do not depend on g_0 either. A direct

benefit of this property is that the control limit L can be simply searched from a process with zero IC profile and unity error standard deviation.

Next, we give some asymptotic properties of the charting statistic $T_{t,h,\lambda}$ which can justify the performance of the NPC chart to a certain degree and shed some light on practical design of the chart as well. The following theorem establishes the asymptotic null distribution of $T_{t,h,\lambda}$, in which design points x_{kj}s are assumed to be i.i.d. with a density Γ_2 in each IC profile.

Theorem 1 *Under conditions (C1)–(C5) and (C7) given in Appendix A, when the process is IC, we have*

$$\left(T_{t,h,\lambda} - \tilde{\mu}_h\right)/\tilde{\sigma}_h \xrightarrow{\mathcal{L}} N(0,1),$$

where

$$\tilde{\mu}_h = \frac{\int [K(u)]^2 du}{h} \int \frac{\Gamma_1(x)}{\Gamma_2(x)} dx, \quad \tilde{\sigma}_h^2 = \frac{2\int [K * K(u)]^2 du}{h} \int \frac{\Gamma_1^2(x)}{\Gamma_2^2(x)} dx.$$

The next theorem investigates the asymptotic behavior of $T_{t,h,\lambda}$ under the OC model

$$y_{kj} = \begin{cases} g_0(x_{kj}) + \varepsilon_{kj}, & \text{if} \quad 1 \le k \le \tau \\ g_1(x_{kj}) + \varepsilon_{kj}, & \text{if} \quad k > \tau, \end{cases} \tag{6}$$

where τ is an unknown shift time point, and $g_1(x) = g_0(x) + \delta(x)$ is the unknown OC profile function. Denote

$$\zeta_\delta = \frac{1}{\sigma^2} \int \left[\delta(u) + \frac{h^2 \eta_1}{2} \delta''(u)\right]^2 \Gamma_1(u) du, \quad \eta_1 = \int K(t) t^2 dt,$$
$$\zeta_1 = \int \delta^2(u) \Gamma_1(u) du, \quad \zeta_2 = \int [\delta''(u)]^2 \Gamma_1(u) du.$$

Theorem 2 *Under conditions (C1)-(C4), (C5'), (C6) and (C7) given in Appendix A and the extra condition that $\zeta_2 < M$ for some constant $M > 0$, we have*

(i) If $c_{0,t,\lambda} h \zeta_1 \to 0$, then $\left(T_{t,h,\lambda} - \tilde{\mu}_h - c_{0,t,\lambda}\zeta_\delta\right)/\tilde{\sigma}_h$ converges in distribution to $N(0,1)$.

(ii) If $\zeta_2 \to 0$, then $T_{t,h,\lambda}$ has a nontrivial power (i.e., the power will not converge to zero) when $\delta \propto c_{0,t,\lambda}^{-4/9}$ and $h = O(c_{0,t,\lambda}^{-2/9})$.

From Theorem 2, we notice that the asymptotic power of the test statistic $T_{t,h,\lambda}$ depends on δ and its second order derivative. The charting statistic of the NEWMA chart has similar leading

terms in its asymptotic expression. However, compared to NEWMA, $T_{t,h,\lambda}$ has the luxury to use a smaller bandwidth in local linear kernel smoothing because it uses observations from different profiles in its smoothing process, as described above. Therefore, the NPC chart based on $T_{t,h,\lambda}$ would be more effective when the profile shift has large curvature (i.e., δ'' is large) since small h would diminish the effect of $\frac{h^2\eta_1}{2}\delta''(u)$ in the above expression of ζ_δ. This explains why we can get a better profile monitoring by using a random design scheme, instead of an equal design scheme, when the curvature of δ is large. In Section 2.4, we will discuss how to select the bandwidth h adaptively in the NPC chart to accommodate different magnitudes of δ''.

2.3 Adaptive selection of the weighting parameter

It is well known that optimal selection of the weighting parameter λ used in EWMA charts depends on the target shift: small λ would be effective for detecting small shifts and large λ is effective for detecting large shifts. So, an EWMA chart with a given λ cannot have a "nearly minimum" ARL for both small and large shifts (cf., e.g., Lucas and Saccucci 1990). This assertion is also valid for our proposed control chart by noting the following result.

Proposition 1 *Under conditions (C1)-(C4) given in Appendix A and the extra conditions that* $h \to 0$, $n_k \to \infty$, $n_0 h^{\frac{3}{2}} \to \infty$, $n_k h^3 \to \infty$, *and* $n_k h^5 \to 0$, *if* $\frac{a^2_{\tau,t,\lambda}}{b_{0,t,\lambda}} h\zeta_1 \to 0$ *and* $\zeta_2 < M$ *for some constant* $M > 0$, *we have*

$$T_{t,h,\lambda} \overset{D}{\approx} \tilde{\mu}_h + h^{-\frac{1}{2}} w + \frac{a^2_{\tau,t,\lambda}}{b_{0,t,\lambda}} \zeta_\delta,$$

where $\overset{D}{\approx}$ *denotes asymptotic equality in distribution, and* w *is a normal random variable with mean zero and variance* $\tilde{\sigma}^2_h$.

The proof of this proposition is analogous to that of Theorem 2 given in Appendix B. So, it is omitted. For simplicity, let us discuss the case when $n_k = n$. In such a case, from expressions of $a_{\tau,t,\lambda}$ and $b_{0,t,\lambda}$, we have

$$\frac{a^2_{\tau,t,\lambda}}{b_{0,t,\lambda}} = \frac{2-\lambda}{\lambda[1-(1-\lambda)^{2t}]}[1-(1-\lambda)^{t-\tau}]^2.$$

By Proposition 1 and the above expression, intuitively, if ζ_δ is small, then it would require a large value of $t-\tau$ to signal and this also depends heavily on the factor $(2-\lambda)/\lambda$. When λ is chosen smaller, $(2-\lambda)/\lambda$ becomes larger. Consequently, the small shift would be detected quicker. On the

other hand, if ζ_δ is large, then we can expect the run length $t-\tau$ to be relatively small as long as λ is chosen relatively large. If λ is chosen small in such a case, then $1-(1-\lambda)^{t-\tau}$ would approach 1 too slow to detect shifts effectively.

Motivated by the AEWMA chart suggested by Capizzi and Masarotto (2003), in this section we suggest an adaptive procedure for choosing the weighting parameter λ. The underlying idea is to adapt weights used for past profiles to the goodness-of-fit of the current profile, so that the related chart can detect shifts of different sizes in a more efficient way. To be specific, let ψ be a score function used for determining the adaptive weights. Capizzi and Masarotto (2003) propose several candidates for ψ. For simplicity, we suggest using the following one:

$$\psi_{l_0,\lambda_0}(u) = \begin{cases} 1-(1-\lambda_0)l_0/u, & \text{if} \quad u \geq l_0 \\ \lambda_0, & \text{if} \quad u < l_0, \end{cases}$$

where $0 < \lambda_0 \leq 1$ and $l_0 > 0$ are two parameters, λ_0 defines the minimum weight, and l_0 is used for balancing detection ability of the control chart for large and small shifts. Apparently, a large (small) l_0 would generate a small (large) adaptive weight, making the control chart more sensitive to small (large) shifts. Further discussion on selection of λ_0 and l_0 will be given in Section 2.7. Then, the NPC chart with the adaptive weight, denoted as NPC-W, signals when

$$T_{t,h,\psi_{l_0,\lambda_0}(T^*_{t,h})} > L, \tag{7}$$

where the control limit $L > 0$ is chosen to achieve a specific IC ARL, and $T^*_{t,h}$ is defined in the same way as $T_{t,h,\lambda}$ except that only the current profile data $\{(x_{tj}, y_{tj}), j = 1, \ldots, n_t\}$ are used here. It is easy to check that $T^*_{t,h}$ is actually $T_{t,h,1}$; it is therefore easy to compute. So, the NPC-W chart essentially combines the EWMA and Shewhart procedures in a natural way. It is worth mentioning that implementation of this adaptive control chart doesn't require much extra computational effort, compared to that of the NPC chart, because recursive formulas given in Section 2.5 for computing $T_{t,h,\lambda}$ only require nonparametric regression of individual profiles. From numerical examples in Section 3, we can see that, after choosing λ_0 and l_0 properly, the NPC-W chart provides well balanced protection against various shifts.

2.4 Adaptive selection of the bandwidth parameter

Like many other smoothing-based tests, performance of the NPC chart depends on selection of the bandwidth parameter h. Optimal selection of h remains an open problem in this area, and it is

widely recognized that optimal h for nonparametric curve estimation is generally not optimal for testing (cf., e.g., Hart 1997). A uniformly most powerful test usually does not exist due to the fact that nonparametric regression functions have infinite dimensions. But the term $\frac{h^2\eta_1}{2}\delta''(\cdot)$ in Theorem 2 tells us that appropriate selection of h would improve the testing power. Intuitively, a smaller h would be more effective in detecting shifts with large curvature (i.e., large δ''), and a larger h would perform better when shifts are flat or smooth (i.e., small δ''). This intuition motivates us to use an adaptive selection procedure, briefly described below.

For the lack-of-fit testing problem, Horowitz and Spokoiny (2001) suggested choosing a single h based on the maximum of a studentized conditional moment test statistic over a sequence of smoothing parameters, and proved that the resulting test would have certain optimality properties. Because this method is easy to use and has good performance in various cases, we use it here for choosing h. Let $\mathcal{H}$ be a set of admissible smoothing parameter values defined to be the following geometric grid:

$$\mathcal{H} = \{h_j = h_{\max}\gamma^{-j} : h_j \geq h_{\min},\ j = 0, \ldots, J_n\}, \tag{8}$$

where $0 < h_{\min} < h_{\max}$ are the lower and upper bounds, and $\gamma > 1$ is a parameter. Clearly, the number of values in $\mathcal{H}$ is $J_n \leq \log_\gamma(h_{\max}/h_{\min})$. Then, the charting statistic of the NPC chart with adaptive bandwidth, denoted as NPC-B, becomes

$$\tilde{T}_{t,\mathcal{H},\lambda} = \max_{h\in\mathcal{H}} \frac{T_{t,h,\lambda} - \tilde{\mu}_h}{\tilde{\sigma}_h}, \tag{9}$$

where $\tilde{\mu}_h$ and $\tilde{\sigma}_h^2$ are respectively the asymptotic expectation and variance of $T_{t,h,\lambda}$, defined in Theorem 1. The next proposition establishes the consistency of $\tilde{T}_{t,\mathcal{H},\lambda}$ against smooth alternatives.

Proposition 2 *Under conditions (C1)-(C4), (C6), and the extra conditions that $h_{\min}$ and $h_{\max}$ both satisfy condition (C5), $\zeta_1 > M_1(c_{0,t,\lambda}^{-1}\ln\ln c_{0,t,\lambda})^{\frac{8}{9}}$, and $\zeta_2 < M_2$, where M_1 and M_2 are two positive constants, then $\tilde{T}_{t,\mathcal{H},\lambda}$ is consistent under model (6) in the sense that its power converges to 1 when t increases.*

From the proof of this proposition given in Appendix B, we can see that $\tilde{T}_{t,\mathcal{H},\lambda}$ would "automatically" maximize the asymptotic power function $c_{0,t,\lambda}h^{\frac{1}{2}}\zeta_\delta$. Thus, it adapts to different magnitudes of δ''; consequently, $\tilde{T}_{t,\mathcal{H},\lambda}$ would be more robust to various potential shifts, compared to $T_{t,\mathcal{H},\lambda}$ with a given bandwidth.

2.5 Some computational issues

Although computing power gets improved fast and it is computationally trivial to do nonparametric function estimation for individual profiles, for on-line process monitoring, which generally handles a large amount of profiles, fast implementation is important and some computational issues are worth our careful examination. For the proposed charts, computation of the test statistic $T_{t,h,\lambda}$ might be time-consuming, and it requires a substantial amount of storage for past profile observations as well. In this part, we provide updating formulas for computing the charting statistic, which will greatly simplify the computation and lessen the storage requirement. Let

$$\tilde{m}_l^{(t,h)}(z) = \sum_{j=1}^{n_k} (x_{tj} - z)^l K_h(x_{tj} - z), \quad l = 0, 1, 2,$$
$$\tilde{q}_l^{(t,h)}(z) = \sum_{j=1}^{n_k} (x_{tj} - z)^l K_h(x_{tj} - z) y_{tj}, \quad l = 0, 1.$$

Then, $m_l^{(t,h,\lambda)}(z)$ in (3) can be recursively updated by

$$m_l^{(t,h,\lambda)}(z) = (1-\lambda) m_l^{(t-1,h,\lambda)}(z) + \tilde{m}_l^{(t,h)}(z), \quad l = 0, 1, 2,$$

where $m_l^{(0,h,\lambda)}(z_i) = 0$, for $l = 0, 1, 2$. Let $q_l^{(t,h,\lambda)}(z)$, for $l = 0, 1$, be two working functions defined by the recursive formula

$$q_l^{(t,h,\lambda)}(z) = (1-\lambda) q_l^{(t-1,h,\lambda)}(z) + \tilde{q}_l^{(t,h)}(z), \quad l = 0, 1,$$

where $q_l^{(0,h,\lambda)}(z) = 0$, for $l = 0, 1$. Then, we have

$$\widehat{g}_{t,h,\lambda}(z) = \left[M^{(t,h,\lambda)}\right]^{-1} \Big\{ (1-\lambda)^2 M^{(t-1,h,\lambda)} \widehat{g}_{t-1,h,\lambda} + \left[\tilde{q}_0^{(t,h)} m_2^{(t,h,\lambda)} - \tilde{q}_1^{(t,h)} m_1^{(t,h,\lambda)}\right] + (1-\lambda) \left[q_0^{(t-1,h,\lambda)} \tilde{m}_2^{(t,h)} - q_1^{(t-1,h,\lambda)} \tilde{m}_1^{(t,h)}\right] \Big\}, \tag{10}$$

where $M^{(t,h,\lambda)}(z) = m_2^{(t,h,\lambda)}(z) m_0^{(t,h,\lambda)}(z) - [m_0^{(t,h,\lambda)}(z)]^2$. On the right hand side of the above equation, dependence on z in each function is not made explicit in notation for simplicity, which should not cause any confusion.

Using the above updating formulas, implementation of the NPC chart can be briefly described as follows. At time point t, we first compute quantities $\tilde{m}_l^{(t,h)}(z)$, for $l = 0, 1, 2$, and $\tilde{q}_l^{(t,h)}(z)$, for $l = 0, 1$, at n_0 pre-determined z locations (see related discussion in Sections 2.2 and 2.7 about selection of n_0 and $\{z_i, i = 1, 2, \ldots, n_0\}$). Then, $m_l^{(0,h,\lambda)}(z_i)$, for $l = 0, 1, 2$, and $q_l^{(0,h,\lambda)}(z_i)$, for

$l = 0, 1$, are updated by the above formulas. Finally, $\widehat{g}_{t,h,\lambda}(z)$ is computed from (10), and the test statistic $T_{t,h,\lambda}$ is computed from $\widehat{g}_{t,h,\lambda}(z)$, after replacing y_{kj} by ξ_{kj}. This algorithm only requires $O(n_0 n_k h)$ operations for monitoring each profile, which is in the same order as the computation involved in conventional local linear kernel smoothing. If n_k and n_0 are both large, we could further decrease the computation to the order of $O(n_k h)$, using the updating algorithm proposed by Seifert *et al.* (1994). See Fan and Marron (1994) for a similar algorithm. Clearly, using the proposed updating formulas, required computer storage does not grow with time t. In addition, compared to the NPC chart with fixed weight and bandwidth parameters, implementation of the NPC-W chart does not require much extra computational effort, and implementation of the NPC-B chart requires J_n times both computational effort and computer storage.

2.6 A self-starting version

The NPC chart makes explicit use of the IC regression function g_0 and the error variance σ^2 (cf., model (1)). In practice, both g_0 and σ^2 might be unknown. In such cases, they need to be estimated from an IC data set. If such IC data are of small to moderate size, then there would be considerable uncertainty in the estimates, which in turn would distort the IC run length distribution of the control chart. Even if the control limit of the chart is adjusted properly to attain a desired IC run length behavior, its OC run length would still be severely compromised (cf., e.g., Jones 2002). To avoid such problems, a large and thus costly collection of IC profile samples would be necessary (see Jensen *et al.* 2006 for related discussion). Zou *et al.* (2008) provide a general guideline on how many IC profile samples are necessary to obtain good run length behavior for the NEWMA chart, according to which at least forty IC profile samples with more than fifty observations in each profile sample are required to obtain satisfactory results in various cases. In this section, we present a self-starting version of the NPC chart, which can substantially reduce the required IC profile samples.

The basic idea of the self-starting version is to replace g_0 and σ^2, both of which are used in defining ξ_{kj}, with some appropriate estimators constructed from past profile data. If the chart does not give a signal of profile shift at time point t, then $g_0(x)$ can be estimated by the conventional local linear kernel estimator constructed from t historical profile samples, which is denoted as $\widehat{g}_0^{(t)}(x)$.

The variance σ^2 can be estimated recursively by

$$\widehat{\sigma}_t^2 = \left\{ \sum_{k=1}^{t-1} n_k \widehat{\sigma}_{t-1}^2 + \sum_{j=1}^{n_t} [y_{tj} - \widehat{g}_0^{(t-1)}(x_{tj})]^2 \right\} \Big/ \sum_{k=1}^{t} n_k.$$

Then, the self-starting version, denoted as NPC-S, is the control chart based on the charting statistic $\widehat{T}_{t,h,\lambda}$, which is constructed in the same way as $T_{t,h,\lambda}$ (cf., (4)) except that ξ_{kj} need to be replaced by $\tilde{\xi}_{kj} = [y_{kj} - \widehat{g}_0^{(k-1)}(x_{kj})]/\widehat{\sigma}_{k-1}$, for all k and j.

It is worth mentioning that, in practice, it is not necessary to update $\widehat{g}_0^{(t)}(x_{kj})$ and $\widehat{\sigma}_t^2$ after t is large enough. As a matter of fact, it is straightforward to show that, when the process is IC,

$$\begin{aligned} \widehat{g}_0^{(t)}(x) &= g_0(x) + O_p((N_t h)^{-\frac{1}{2}}) + O(h^4), \\ \widehat{\sigma}_t^2 &= \sigma^2 \left(1 + O_p(N_t^{-\frac{1}{2}}) + O_p((N_t h)^{-1}) \right), \end{aligned}$$

where $N_t = \sum_{k=1}^{t} n_k$. Thus, when t is sufficiently large, say $t \geq t_0$, the approximations of $\widehat{g}_0^{(t)}(x)$ and $\widehat{\sigma}_t^2$ to g_0 and σ^2 would be good enough, and we could simply use $\widehat{g}_0^{(t_0)}(x)$ and $\widehat{\sigma}_{t_0}^2$ for all profiles with $t \geq t_0$ in process monitoring. There are two benefits with this modification. First, it reduces much computation and storage requirement with very little loss of efficiency. Second, it may reduce the "masking-effect" (cf., Hawkins 1987) to certain extent. That is, when the potential shift occurs after time t_0, the estimates $\widehat{g}_0^{(t_0)}(x)$ and $\widehat{\sigma}_{t_0}^2$ would not be contaminated by the OC observations, which is not the case if these estimates are updated at every time point.

From the description in this and previous two subsections, it can be seen that the NPC-S chart can accommodate adaptive selection of the weight and bandwidth parameters. The resulting chart, denoted as NPC-SWB, should be able to offer a balanced protection against shifts of different magnitudes and adapt to the smoothness of the IC and OC profile functions as well. Formulation of the NPC-SWB chart can be readily obtained by incorporating (7) and (9) into $\widehat{T}_{t,h,\lambda}$; hence, it is not elaborated here. Its performance will be investigated in Section 3.

2.7 Certain practical guidelines

On choosing n_k and x_{kj}: In certain applications, design points x_{kj} are determined by the industrial process itself, and we can not do much in choosing them. In some others (cf., Zou *et al.* 2008), they need to be specified before process monitoring. As demonstrated by Theorem 2,

random design has some benefits, compared to the design in which different profiles share the same design points, because observations from different profiles would provide information about more details of the regression function g in the former case. So, if we can choose the design points, then random design would be a good choice. This amounts to determining a proper design distribution Γ_2, from which design points x_{kj} are generated for individual profiles. The number of design points n_k can also be random, although in many applications $n_k = n$ would be the most convenient scheme to use. The value of n can be chosen smaller than the one used in the NEWMA chart by Zou *et al.* (2008), because, in computing $T_{t,h,\lambda}$, roughly $c_{0,t,\lambda} = \frac{2-\lambda}{\lambda} n$ observations are actually used.

On choosing n_0 and z_is: Based on our numerical experience, selection of n_0 and z_is does not affect the performance of the NPC chart much, as long as n_0 is not too small and z_is cover all the key parts (e.g., peaks/valleys or oscillating regions) of g_0 well. In our numerical examples presented in Section 3, we find that results do not change much when $n_0 \geq 40$.

On choosing l_0 and λ_0 used in the NPC-W chart: As described in Section 2.3, λ_0 is the minimum weight used by the chart NPC-W, and l_0 is the parameter that controls the chance for the chart to use that minimum weight. The chart would use the minimum weight λ_0 if $T_{t,h,1} < l_0$. We suggest using the upper α_0 percentile of $T_{t,h,1}$ as the value of l_0, which could be obtained by simulation before profile monitoring. An appropriate method for determining α_0 is to set $\alpha_0 = c/ARL_0$, where $c > 1$ is a constant and ARL_0 is the desired IC ARL value. Since the reciprocal of ARL_0 can be regarded as a rough estimate of the false alarm probability, it would be reasonable to choose α_0 to be c/ARL_0 so that the chart NPC-W can achieve the given IC ARL value. With respect to λ_0, it should be chosen smaller than the commonly used value 0.2 in the EWMA literature (cf., Lucas and Saccucci 1990). Based on our numerical experience and the results in Capizzi and Masarotto (2003), we recommend using $\lambda_0 \in [0.05, 0.1]$ and $5 \leq c \leq 15$.

On choosing $\mathcal{H}$ used in the NPC-B chart: Theoretically speaking, the parameters γ, J_n, $h_{\max}$ and $h_{\min}$ should satisfy certain conditions to obtain the corresponding asymptotic results. See Appendix A for related discussion. Based on our simulations, we notice that the proposed control chart is actually quite robust to them, which is consistent with the findings in Horowitz and Spokoiny (2001). By both theoretical arguments and numerical studies, we recommend using the choices that $1 < \gamma < 2$, J_n could be 4, 5 or 6, $h_{\max} = M c_{0,t,\lambda}^{-1/7}$, and $h_j = h_{\max}\gamma^{-j}$, for $j = 1, \ldots, J_n$, where $0.5 \leq M \leq 2$ is a constant. Note that the recommended value $h_{\max} = M c_{0,t,\lambda}^{-1/7}$ is partially due to condition (C5) given in Appendix A.

On the NPC-S chart: As suggested by Hawkins *et al.* (2003), we recommend collecting three to ten IC profile samples before using the NPC-S chart. These preliminary profile samples are mainly for stabilizing the variation of $\widehat{T}_{t,h,\lambda}$.

3 A Simulation Study

We present some simulation results in this section regarding the numerical performance of the proposed NPC chart. Throughout this section, the kernel function is chosen to be the Epanechnikov kernel function $K(x) = 0.75(1 - x^2)I(-1 \le x \le 1)$. The IC ARL is fixed at 200. The error distribution is assumed to be Normal. For simplicity, we assume that $n_k = n = 20$ for all k, $x_{kj} \sim$ Uniform$(0, 1)$, for $j = 1, \ldots, n$, $z_i = (i - 0.5)/n_0$, for $i = 1, \ldots, n_0$, and $n_0 = 40$. All the ARL results in this section are obtained from 50,000 replications unless indicated otherwise. In addition, we focus on the steady-state OC ARL behavior of each chart (cf., Hawkins and Olwell 1998), and assume that $\tau = 30$ (cf., the OC model (6)). When computing the ARL values, any simulation runs in which a signal occurs before the $(\tau + 1)$-th profile are discarded. As a side note, our numerical results (not reported here to save some space) show that steady-states of the related charts considered in this section are reached when τ is as small as 10 in all cases considered.

To compare the NPC chart with alternative methods turns out to be difficult, due to lack of an obvious comparable method. One possible alternative method is the NEWMA chart proposed by Zou *et al.* (2008) in which design points are assumed to be equally spaced in each profile and they are unchanged from profile to profile. To make the procedures comparable, for the NEWMA chart, we assume that design points are $(i-0.5)/n$, for $i = 1, \ldots, n$, in each profile sample. Another possible method to compare is the naive modification of the NEWMA chart described in Section 2.2, which is called the NAEWMA chart below. By the NAEWMA chart, profile functions are first estimated from individual profile data, then response values are predicted from these estimated profile functions at some common points $\{z_1, z_2, \ldots, z_n\}$ in the design interval for different profiles, and finally the NEWMA chart is applied to the predicted response values. For this chart, we take $z_i = (i - 0.5)/n$, for $i = 1, \ldots, n$, as in the NEWMA chart.

To appreciate the benefits of different versions of the NPC chart, we investigate numerical performance of the NPC, NPC-W, NPC-B, NPC-S, and NPC-WBS charts separately. Note that, for charts NEWMA and NAEWMA, the bandwidth h and the weighting parameter λ should both be

pre-specified. Therefore, we first compare the charts NPC, NEWMA and NAEWMA in such cases. Following the recommendations by Zou *et al.* (2008), h is chosen to be either $h_1 = 1.5n^{-1/5}\sqrt{\text{Var}(x)}$ or $h_2 = 1.5[n(2-\lambda)/\lambda]^{-1/5}\sqrt{\text{Var}(x)}$. The smaller bandwidth h_2 is considered here because the actual number of observations used in the NPC chart at each time point is $c_{0,t,\lambda}$ which is roughly $n(2-\lambda)/\lambda$. In all three charts, λ is chosen to be 0.1 or 0.2. The IC model used is $g_0(x) = 1-\exp(-x)$, and the following two representative OC models are considered here:

$$\text{(I)} : g_1(x) = 1 - \exp(-x) + \theta x; \quad \text{(II)} : g_1(x) = 1 - \exp(-x) + \theta \sin(2\pi(x - 0.5)).$$

In case (I), $g_1(x) - g_0(x) = \theta x$ is a straight line; and $g_1(x) - g_0(x) = \theta \sin(2\pi(x - 0.5))$ oscillates a lot in case (II). Table 1 presents the OC ARL values of the three charts in various cases. Their control limits L are also included in the table.

From the table, we can have the following results. First, in case (I), larger h (i.e., h_1) yields better performance for both NPC and NEWMA charts, because $\delta(x) = g_1(x) - g_0(x)$ is straight in this case. For a given bandwidth, the NPC chart outperforms the NEWMA chart uniformly. This is because the effective number of observations used in the NPC chart at each time point is larger than that used in the NEWMA chart, and consequently the charting statistic of the NPC chart would converge to $c_{0,t,\lambda}\zeta_1$ faster (cf., Theorem 1). Second, in case (II) where $\delta(x)$ oscillates a lot, the NPC chart with the smaller h has better performance, which is intuitively reasonable. However, this is not the case for the NEWMA chart, because smaller bandwidth in the NEWMA chart would result in large bias in estimating the regression function and thus reduce its ability in detecting profile shift. This example confirms the fact that the NPC chart allows us to use a smaller bandwidth to better detect oscillating profile shifts (cf., related discussion at the end of Section 2.2). Third, in case (II), the NPC chart outperforms the NEWMA chart uniformly. Fourth, the NPC chart outperforms the NAEWMA chart by a quite large margin in most cases, and the NAEWMA chart also performs uniformly worse than the NEWMA chart. These results confirm our argument made in Section 2.2 that the naive chart NAEWMA would be inefficient because profile estimates constructed from individual profile data would have relatively large bias and variance. By the way, our simulations (not reported here) show that, when n_0 is chosen larger than 40, performance of the NPC chart would not change much.

Next we consider the NPC-W chart in which the weight parameter λ is adaptively chosen so that the chart would be robust to shift size. Its other parameters are chosen according to the practical guidelines given in Section 2.7. More specifically, we choose $\lambda_0 = 0.1$ and $\alpha_0 = 0.05$.

Table 1: OC ARL comparison of the NPC, NEWMA, and NAEWMA charts when IC ARL=200, $\lambda = 0.1$ or 0.2, and $n = 20$.

	θ	NPC		NEWMA		NAEWMA	
		h_1	h_2	h_1	h_2	h_1	h_2
				$\lambda = 0.1$			
	0.100	75.9 (.357)	87.4 (.486)	89.5 (.393)	102 (.452)	101 (.455)	103 (.464)
	0.200	27.6 (.106)	33.1 (.152)	32.7 (.118)	38.5 (.142)	39.7 (.154)	41.0 (.160)
	0.300	14.8 (.046)	17.4 (.066)	17.2 (.050)	19.7 (.059)	20.5 (.066)	20.9 (.067)
	0.400	9.85 (.026)	11.4 (.037)	11.2 (.027)	12.6 (.031)	13.3 (.036)	13.6 (.037)
(I)	0.600	5.90 (.013)	6.68 (.018)	6.64 (.013)	7.33 (.015)	7.64 (.016)	7.76 (.017)
	0.800	4.27 (.008)	4.80 (.011)	4.74 (.008)	5.19 (.009)	5.38 (.010)	5.44 (.010)
	1.200	2.82 (.005)	3.12 (.007)	3.12 (.005)	3.37 (.005)	3.46 (.006)	3.50 (.006)
	1.600	2.18 (.004)	2.37 (.004)	2.38 (.004)	2.56 (.004)	2.62 (.004)	2.65 (.004)
	0.100	68.9 (.313)	65.3 (.343)	72.4 (.316)	81.6 (.348)	94.2 (.419)	95.6 (.424)
	0.200	23.2 (.082)	22.2 (.090)	24.5 (.080)	27.6 (.092)	31.7 (.112)	32.2 (.114)
	0.300	12.4 (.035)	12.0 (.039)	13.0 (.034)	14.3 (.037)	16.0 (.044)	16.2 (.045)
(II)	0.400	8.29 (.020)	8.03 (.022)	8.76 (.019)	9.53 (.021)	10.5 (.024)	10.5 (.024)
	0.600	5.11 (.010)	4.96 (.012)	5.32 (.009)	5.75 (.010)	6.15 (.011)	6.20 (.011)
	0.800	3.73 (.007)	3.65 (.008)	3.89 (.006)	4.16 (.007)	4.42 (.008)	4.43 (.008)
	1.200	2.51 (.004)	2.42 (.005)	2.62 (.004)	2.78 (.004)	2.94 (.004)	2.95 (.004)
	1.600	1.94 (.003)	1.90 (.003)	2.04 (.003)	2.15 (.003)	2.25 (.003)	2.26 (.003)
L		9.49	13.05	14.39	19.02	12.88	13.32
				$\lambda = 0.2$			
	0.100	95.6 (.452)	108 (.501)	112 (.519)	125 (.595)	132 (.625)	135 (.648)
	0.200	34.6 (.148)	41.0 (.174)	43.2 (.188)	51.9 (.228)	60.0 (.269)	62.2 (.281)
	0.300	16.6 (.063)	19.6 (.076)	20.1 (.076)	23.7 (.089)	28.4 (.116)	29.6 (.120)
	0.400	10.0 (.031)	11.7 (.036)	11.7 (.036)	13.4 (.045)	16.1 (.056)	16.6 (.058)
(I)	0.600	5.39 (.013)	6.05 (.013)	6.03 (.013)	6.64 (.018)	7.69 (.020)	7.94 (.022)
	0.800	3.69 (.009)	4.09 (.009)	4.09 (.009)	4.44 (.009)	4.98 (.011)	5.15 (.011)
	1.200	2.35 (.004)	2.55 (.004)	2.56 (.004)	2.74 (.004)	3.00 (.005)	3.04 (.005)
	1.600	1.80 (.003)	1.93 (.003)	1.94 (.003)	2.06 (.003)	2.21 (.003)	2.26 (.003)
	0.100	93.1 (.461)	85.7 (.407)	93.9 (.429)	105 (.479)	133 (.630)	135 (.639)
	0.200	30.3 (.130)	27.6 (.112)	31.3 (.125)	35.4 (.148)	48.6 (.235)	55.3 (.242)
	0.300	13.8 (.049)	13.0 (.045)	14.2 (.049)	16.0 (.054)	22.5 (.081)	23.2 (.086)
(II)	0.400	8.31 (.022)	7.91 (.022)	8.57 (.022)	9.38 (.027)	12.3 (.036)	12.5 (.038)
	0.600	4.54 (.009)	4.37 (.009)	4.68 (.009)	5.00 (.009)	5.98 (.012)	6.05 (.013)
	0.800	3.17 (.004)	3.07 (.004)	3.26 (.004)	3.46 (.004)	4.00 (.007)	4.04 (.007)
	1.200	2.08 (.004)	2.01 (.004)	2.14 (.004)	2.25 (.004)	2.50 (.004)	2.53 (.004)
	1.600	1.61 (.003)	1.56 (.003)	1.66 (.003)	1.74 (.003)	1.91 (.003)	1.92 (.003)
L		10.47	13.09	15.41	19.10	15.00	15.51

NOTE: Standard errors are in parentheses.

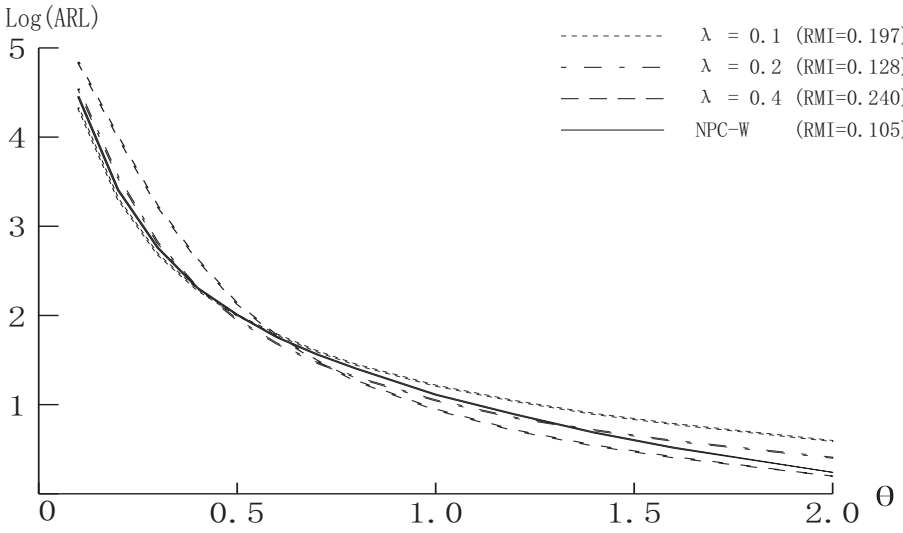

Figure 1: OC ARL comparison of the NPC-W chart and three NPC charts with $\lambda = 0.1, 0.2$ and 0.4.

Since we are mainly concerned about the robustness of the NPC-W chart to shift size, only the OC model (I) is considered here. The bandwidth h is chosen to be h_1, because it is more appropriate to use in this case than h_2, by Table 1. For comparison purpose, the OC ARL values of three NPC charts when $\lambda = 0.1, 0.2$ and 0.4 are also computed. To measure robustness of a chart T to shift size, the relative mean index (RMI) originally proposed by Han and Tsung (2006) is used, which is defined by

$$\text{RMI(T)} = \frac{1}{m}\sum_{i=1}^{m}\frac{\text{ARL}_{\theta_i}(\text{T}) - \text{MARL}_{\theta_i}}{\text{MARL}_{\theta_i}},$$

where $\text{ARL}_{\theta_i}(\text{T})$ is the OC ARL of T for detecting a shift of size θ_i, and MARL_{θ_i} is the smallest value among such OC ARL values of all charts considered. In this example, θ_i ranges from 0.1 to 2 with a step 0.1. Obviously, small RMI(T) implies that T has a robust performance in detecting shifts of various sizes. Figure 1 shows the OC ARL values (in log scale) and the RMI values of the four charts considered. It can be seen that performance of the three NPC charts depends heavily on their pre-specified λ values, as expected, and the NPC-W chart offers a balanced protection against various shift sizes. In terms of RMI, the NPC-W chart performs the best. After taking into account its convenient implementation, we think that the NPC-W chart is a valuable improvement of the NPC chart.

Table 2: OC ARL comparison of the NPC and NPC-B charts when IC ARL=200, $\lambda = 0.2$ and $n = 20$.

θ	NPC-B	NPC		
		$h = 0.6$	$h = 0.3$	$h = 0.15$
0.250	8.20 (.031)	7.84 (.022)	8.27 (.031)	10.8 (.036)
0.500	9.05 (.031)	8.95 (.027)	9.45 (.031)	12.2 (.040)
0.750	11.2 (.036)	11.4 (.036)	10.7 (.036)	13.9 (.040)
1.000	14.7 (.054)	16.2 (.058)	14.1 (.049)	15.9 (.058)
2.000	32.3 (.112)	86.3 (.398)	35.0 (.125)	26.9 (.103)
3.000	31.0 (.098)	49.0 (.224)	50.6 (.224)	24.4 (.094)
4.000	63.5 (.286)	166 (.814)	174 (.832)	48.0 (.206)
5.000	75.6 (.344)	120 (.577)	125 (.581)	81.4 (.376)
L	4.27	8.75	9.61	13.32

NOTE: Standard errors are in parentheses.

Next, we consider the NPC-B chart. By the guidelines in Section 2.7, we choose $\gamma = 1.4$, $h_{\max} = 1.0[(2 - \lambda/\lambda)n]^{-1/7}$, and $h_j = h_{\max}\gamma^{-j}$, for $j = 1, \ldots, 4$. We consider the following OC model $g_1(x) = 1 - \exp(-x) + 0.25\cos(\theta\pi(x - 0.5))$. By changing θ, this model can cover various cases with different smoothness of $\delta(\cdot)$. For comparison purposes, we also consider three NPC charts with bandwidth 0.6, 0.3, and 0.15, respectively. Their other parameters are chosen as in the example of Table 1. OC ARL values of related charts are shown in Table 2, with their control limits L listed in the bottom line. From the table, it can be seen that the NPC chart with a fixed bandwidth outperforms the NPC-B chart in certain ranges of θ, but they can also be much worse in other ranges of θ. As a comparison, the NPC-B chart is always close to the best chart in all cases, because it can adapt to the unknown smoothness of $\delta(x)$ and pick up an appropriate bandwidth accordingly from $\mathcal{H}$.

We now investigate the numerical performance of the NPC-S chart. First, we study its 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 2002). As an alternative, here we use the hazard function $H_1(r)/H_2(r)$ recommended by Hawkins and Maboudou-Tchao (2007), 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, then the corresponding hazard is a constant whose inverse is the ARL. In the example of Table 1 when $h = h_1$ and $L = 10.47$, which corresponds to an IC ARL of 200 when the IC model is

assumed known, the IC hazard function of the NPC-S chart based on 250,000 replications is shown in Figure 2. When computing the IC hazard function, we follow the suggestion given in Section 2.7 that process monitoring starts after five IC profiles are collected beforehand. From the plot, we can see that the IC hazard starts around 0.0065, then drops quickly to values around $0.005 = 1/200$, and gets stabilized at that level for good. This plot shows that, except for short run-lengths, the geometric distribution is an excellent fit to the IC run length of the NPC-S chart, which is consistent with the findings in Hawkins and Maboudou-Tchao (2007) about a self-starting chart for monitoring multivariate Normal processes. Furthermore, sample mean and sample standard deviation of the run lengths are 196 and 194, respectively, which are almost identical and which further confirms that the NPC-S chart works well under the IC condition. We conducted some other simulations with various combinations of n, h and Γ_2 to check whether the above conclusions are true in other settings. These simulation results, not reported here but available from the authors, show that the NPC-S chart has quite satisfactory performance in other cases as well, except certain extreme cases such as the ones when n is too small (e.g., $n \leq 5$).

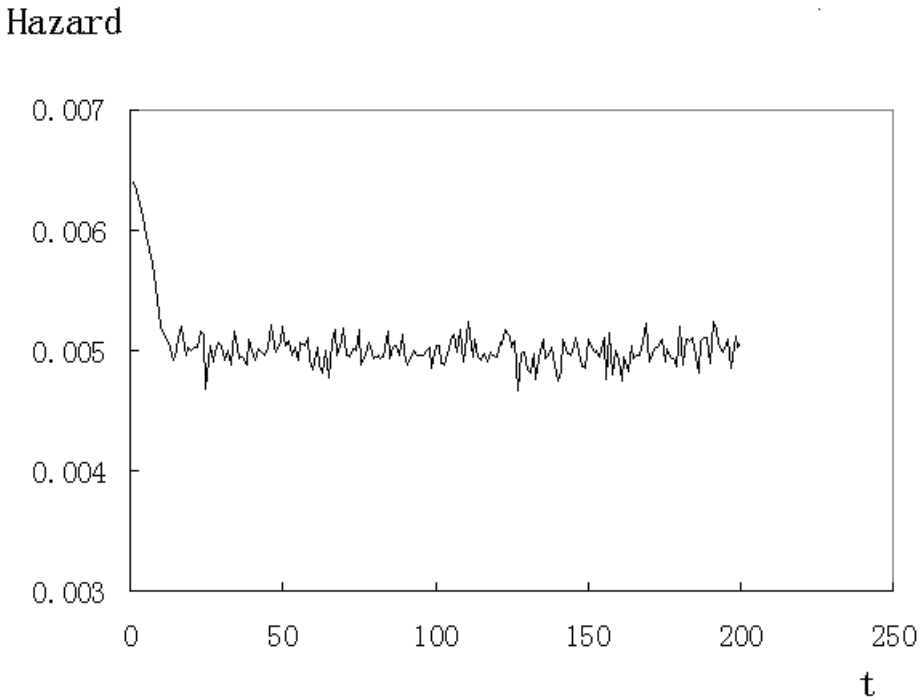

Figure 2: Hazard curve of the NPC-S chart.

Next, we examine the OC performance of the NPC-S chart. As demonstrated in the literature, OC performance of self-starting charts is generally affected by the shift time point (cf., e.g., Hawkins *et al.* 2003). In this example, we consider two shift times $\tau = 40$ and $\tau = 80$. The simulation

Table 3: OC ARL performance of the NPC-S and NPC-SWB charts when IC ARL=200, $\lambda = 0.2$ and $n = 20$.

		NPC-S				NPC-SWB	
	θ	$\tau = 40$		$\tau = 80$		$\tau = 40$	$\tau = 80$
		h_1	h_2	h_1	h_2		
	0.100	163 (.903)	176 (.939)	143 (.818)	161 (.899)	151 (.859)	132 (.859)
	0.200	101 (.738)	116 (.778)	68.2 (.505)	88.9 (.626)	82.5 (.626)	54.7 (.474)
	0.300	46.1 (.456)	60.4 (.523)	24.9 (.192)	34.5 (.268)	32.2 (.313)	19.6 (.130)
	0.400	17.5 (.188)	26.3 (.291)	11.7 (.054)	15.0 (.085)	14.0 (.098)	10.8 (.040)
(I)	0.600	6.01 (.018)	7.25 (.027)	5.66 (.018)	6.63 (.018)	6.20 (.018)	5.87 (.018)
	0.800	3.88 (.009)	4.37 (.009)	3.77 (.009)	4.22 (.009)	4.14 (.009)	3.95 (.009)
	1.200	2.38 (.004)	2.64 (.004)	2.38 (.004)	2.62 (.004)	2.38 (.004)	2.35 (.004)
	1.600	1.81 (.004)	1.95 (.004)	1.81 (.004)	1.96 (.004)	1.64 (.004)	1.64 (.004)
	0.100	156 (.872)	161 (.926)	146 (.823)	142 (.814)	150 (.836)	132 (.827)
	0.200	78.7 (.581)	82.2 (.631)	55.2 (.402)	54.5 (.398)	68.0 (.483)	48.1 (.367)
	0.300	26.1 (.206)	26.5 (.237)	18.3 (.103)	17.6 (.103)	22.9 (.165)	17.2 (.080)
(II)	0.400	10.9 (.063)	10.6 (.067)	9.44 (.036)	8.86 (.031)	10.7 (.040)	9.75 (.031)
	0.600	4.92 (.013)	4.68 (.013)	4.77 (.009)	4.56 (.009)	5.34 (.013)	5.10 (.013)
	0.800	3.30 (.004)	3.19 (.004)	3.24 (.004)	3.15 (.004)	3.62 (.009)	3.58 (.009)
	1.200	2.13 (.004)	2.05 (.004)	2.09 (.004)	2.03 (.004)	2.11 (.004)	2.10 (.004)
	1.600	1.60 (.004)	1.56 (.004)	1.61 (.004)	1.56 (.004)	1.43 (.004)	1.42 (.004)
	L	10.47	13.09	10.47	13.09	4.070	4.070

NOTE: Standard errors are in parentheses.

results in various cases considered in Table 1 are presented in Table 3. From the table, it can be seen that the NPC-S chart performs almost equally well for both values of τ when the shift size is large. For detecting small to moderate shifts, it generally performs better with a larger τ, because the updated parameter estimates would be more accurate in such a case under the IC condition, which is confirmed by the table. As a comparison, in Table 3, we also present the OC ARLs of the NPC-SWB chart, which is a combination of the self-starting chart and adaptive selection of the weight and bandwidth parameters. Its parameters are chosen to be the same as those used in the examples of Figure 1 and Table 2. From the table, we can see that the NPC-SWB chart outperforms both NPC-S charts using h_1 and h_2 in all cases except certain cases with moderate shifts. Thus, in practice, the NPC-SWB chart is recommended, if the extra computation involved is not a major concern (cf., the last paragraph of Section 2.5 for related discussion about computation).

4 A Real-Data Application

In this section, we demonstrate the proposed NPC chart by applying it to a dataset obtained from the semiconductor manufacturing industry for monitoring a deep reactive ion etching (DRIE) process which is critical to the output wafer quality and requires careful control and monitoring. In the DRIE process, the desired profile is the one with smooth and straight sidewalls and flat bottoms, and ideally the sidewalls of a trench are perpendicular to the bottom of the trench with a certain degree of smoothness around the corners (cf., the middle shape shown in Figure 3). Various other profile shapes, such as positive and negative ones (cf., the two left-hand-side and two right-hand-side shapes shown in Figure 3) due to underetching and overetching, are considered to be unacceptable. More detailed discussion about the DRIE example can be found in Rauf *et al.* (2002) and Zhou *et al.* (2004).

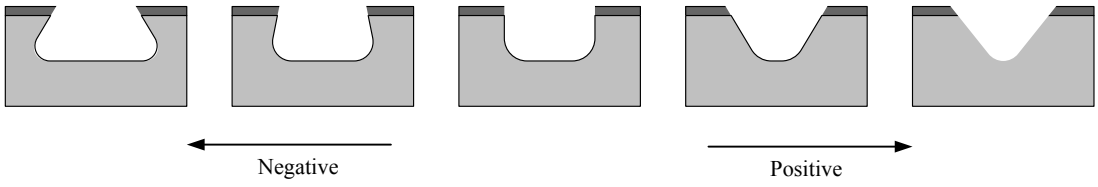

Figure 3: Illustrations of various etching profiles from a DRIE process.

The DRIE data considered here have 21 profiles. The original data include their images, like the ones shown in Figure 3. To monitor the DRIE process, we need to obtain samples from individual profiles, which can be acquired by the scanning electron microscope (SEM). Since the profiles are usually symmetric, we can focus on one half of each profile (e.g., the left half) for profile monitoring purposes. To make that part of the profile convenient to describe by a mathematical function, it is rotated by 45 degrees along a reference point in a pre-specified coordinate system, before dimensional readings of the profile are collected by SEM. Among the 21 profiles, based on engineering knowledge, the first 18 profiles are IC and the remaining 3 profiles are OC. Since the number of IC profiles is not large, the IC profile function g_0 and the error standard deviation σ may not be accurately estimated from the IC data. Therefore, we consider using the self-starting chart NPC-S in this example. As pointed out at the end of Section 2.2 and confirmed numerically in the example of Table 1, for applications such as the current one, a better profile monitoring can be

achieved by using a random design scheme, instead of a fixed design scheme for all profiles. Next, we illustrate how to design a reasonable random design scheme in this example and how to apply the NPC-S chart to the resulting data.

To design a random design scheme, we first need to specify the density Γ_1 of the design points. In this example, it seems that the corner part of the profile contains critical information regarding whether the profile is OC; it should receive enough attention. A reasonable design distribution that takes this into account is $x \sim \text{Normal}(0, 2.5)$, where the original point of x is located at the center of the corner. This random design ensures that most design points fall within [-4,4] that covers half of the bottom trench, and about 65% design points are located in [-1.5,1.5] that covers the corner part. For each profile, we fix $n = 20$, and dimensional readings are collected by SEM at n design points generated from Γ_1. Using electronic sensor and information technologies, this entire data acquisition process can be finished automatically by a computer. In the NPC-S chart, we fix the IC ARL at 200, n_0 at 40, and z_is to be equally spaced in [-3.5,3.5]. All other parameters of the NPC-S chart are chosen to be the same as those used in the example of Table 3. The control limit is computed to be $L = 16.07$ by simulation. Following the practical guidelines given in Section 2.7, we take the first 10 IC profiles as preliminary data, and profile monitoring starts at the 11th profile. The charting statistic $\widehat{T}_{t,h,\lambda}$, for $t = 11, \ldots, 21$, is shown in Figure 4, along with its control limit shown by the solid horizontal line. In that figure, we also present the NAEWMA chart and its control limit 22.32 by the dashed lines. Parameters of the NAEWMA chart are chosen to be $\lambda = 0.2$, IC ARL=200, $n = 20$, and $z_1, ...z_n$ are equally spaced in $[-3.5, 3.5]$. From the plot, it can be seen that the NPC-S chart gives a signal of profile shift at the 20th time point which corresponds to the 2nd OC profile. The NAEWMA chart does not give any signal, even after the 3rd OC profile is collected.

As a side note, it takes about 3.4/1000 seconds to compute all values of the charting statistic $\widehat{T}_{t,h,\lambda}$ that are plotted in Figure 4, by a Pentium 2.4MHz CPU. Therefore, the proposed procedure should be quite convenient to use for on-line automatic profile monitoring.

5 Summary and Concluding Remarks

In this paper, we propose a control chart for monitoring nonparametric profiles with arbitrary design. Our proposed control chart effectively combines the EWMA control chart and a nonpara-

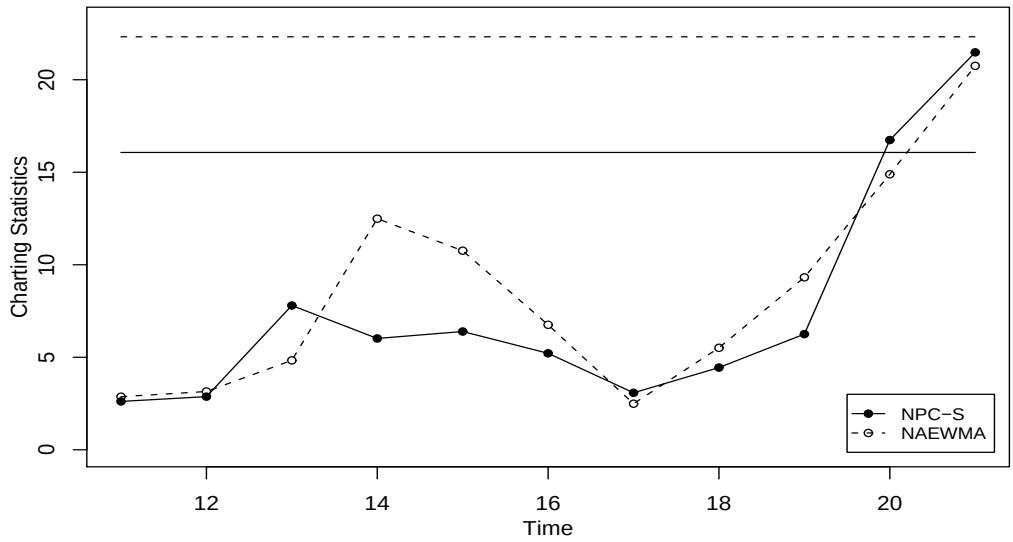

Figure 4: The NPC-S and NAEWMA control charts for monitoring the DRIE process. The solid and dashed horizontal lines indicate their control limits, respectively.

metric regression test. The proposal to adaptively choose the weight and bandwidth parameters further enhances the proposed chart. Moreover, a self-starting version is introduced for cases when the IC regression function and error variance are unknown. As demonstrated by the DRIE example, the proposed monitoring approach can be implemented conveniently in industrial applications. In addition, we show that a better monitoring performance can be obtained by using a random design instead of a fixed design. Numerical studies show that the proposed approach is effective in applications.

Our proposed control chart is under the assumptions that observations within and between individual profiles are independent of each other. In some applications, within-profile observations might be spatially or serially correlated, and between-profile observations might be auto-correlated (cf., Williams *et al.* 2007a,b, Zou *et al.* 2007a). It requires much future research to accommodate such correlations in nonparametric profile monitoring. In addition, sometimes we might be interested in monitoring multivariate relationship between a response variable and several predictors over time. At this moment, we are not aware of any existing research on this topic, and we leave it to our future research to generalize the proposed control chart discussed in this paper to multivariate cases.

Acknowledgments: The authors thank the editor, the associate editor, and two referees for many constructive suggestions and comments, which greatly improved the quality of the paper. This research is supported in part by an NSF grant.

Appendix: Technical details

Throughout the appendix, we use the following additional notations:

$$\alpha_{t,h,\lambda}(z) = \frac{1}{a_{0,t,\lambda}\Gamma_2(z)} \sum_{k=1}^{t}(1-\lambda)^{t-k} \sum_{j=1}^{n_k} K_h(x_{kj}-z)\varepsilon_{kj},$$

$$\beta_{t,h,\lambda}(z) = \frac{g_1''(z)}{2a_{0,t,\lambda}\Gamma_2(z)} \sum_{k=1}^{t}(1-\lambda)^{t-k} \sum_{j=1}^{n_k} (x_{kj}-z)^2 K_h(x_{kj}-z),$$

$$\phi_i(z) = \frac{1}{a_{0,t,\lambda}} \sum_{k=1}^{t}(1-\lambda)^{t-k} \sum_{j=1}^{n_k} (x_{kj}-z)^i K_h(x_{kj}-z)\varepsilon_{kj}, \quad i=0,1,$$

$$\phi_{i+2}(z) = \frac{1}{a_{\tau,t,\lambda}} \sum_{k=\tau+1}^{t}(1-\lambda)^{t-k} \sum_{j=1}^{n_k} (x_{kj}-z)^i K_h(x_{kj}-z) g_1(x_{kj}), \quad i=0,1,$$

$$d_{t_0,t_1,\lambda} = \sum_{k=t_0+1}^{t_1}(1-\lambda)^{4(t-k)} n_k, \quad e_{t_0,t_1,\lambda} = \sum_{k=t_0+1}^{t_1}(1-\lambda)^{4(t-k)} n_k^2.$$

Appendix A: Regularity Conditions Used In Section 2

(C1) Density functions Γ_1 and Γ_2 are Lipschitz continuous and bounded away from zero on [0,1].

(C2) $g_0(\cdot)$ and $g_1(\cdot)$ have continuous second order derivatives on [0,1].

(C3) The kernel function $K(u)$ is bounded and symmetric about 0 on [-1,1]. Furthermore, $u^3K(u)$ and $u^3K'(u)$ are both bounded, and $\int_{-1}^{1} u^4K(u)du < \infty$.

(C4) $E(|\varepsilon_{11}|^4) < \infty$.

(C5) n_0, h and $c_{0,t,\lambda}$ satisfy the conditions that $n_0 \to \infty$, $h \to 0$, $n_0h^{\frac{3}{2}} \to \infty$, $c_{0,t,\lambda} \to \infty$, $c_{0,t,\lambda}h^{\frac{3}{2}} \to \infty$ and $c_{0,t,\lambda}h^8 \to 0$.

(C5') n_0, h and $c_{0,t,\lambda}$ satisfy the conditions that $n_0 \to \infty$, $h \to 0$,

$n_0h^{\frac{3}{2}} \to \infty$, $c_{0,t,\lambda} \to \infty$, $c_{0,t,\lambda}h^3 \to \infty$ and $c_{0,t,\lambda}h^5 \to 0$.

(C6) $\frac{a_{\tau,t,\lambda}}{a_{0,t,\lambda}} - 1 = o(\min\{h^2, c_{0,t,\lambda}^{-\frac{1}{2}}\})$.

(C7) n_ks satisfy the condition that $\frac{\max_{1\le k\le t} n_k}{\min_{1\le k\le t} n_k}$ is bounded.

It is noted that conditions (C1)-(C4) are standard in nonparametric regression. (C5) and (C5') are the bandwidth conditions used in Theorem 1 and Theorem 2(i), respectively. Note that $\frac{(2-\lambda)}{\lambda}\min_{1\le k\le t} n_k \le c_{0,t,\lambda} \le \frac{(2-\lambda)}{\lambda}\max_{1\le k\le t} n_k$ for large t. Thus, if $\lambda \to 0$, we even do not require $n_k \to \infty$. The conditions listed here are much milder than those in Zou *et al.* (2008) where the number of design points in each profile should go to infinity. (C6) can be easily satisfied if t is large enough. (C7) implies that all n_ks are of the same order, which is common in practice.

Appendix B: Proofs

To prove the two theorems in Section 2, the following lemma is required.

Lemma 1 *For any $z \in [0,1]$, (i) under conditions in Theorem 1, we have*

$$\widehat{g}_{t,h,\lambda}(z) = \alpha_{t,h,\lambda}(z)(1 + o(h^{\frac{1}{2}}));$$

(ii) under conditions in Theorem 2, we have

$$\widehat{g}_{t,h,\lambda}(z) - g_1(z) = \alpha_{t,h,\lambda}(z)(1 + o(h^{\frac{1}{2}})) + \beta_{t,h,\lambda}(z)(1 + o_p(1)).$$

Proof We only prove the second equation because the first one can be proved in a similar way. For simplicity, we suppress the symbol "(t,h,λ)" in $m_i^{(t,h,\lambda)}(z)$, which should not cause any confusion. By some algebraic manipulations, it can be checked that

$$\begin{aligned}
\widehat{g}_{t,h,\lambda}(z) - g_1(z) &= a_{0,t,\lambda}m_0^{-1}(z)[\phi_0(z) + \phi_2(z)] + a_{0,t,\lambda}m_0^{-1}(z)m_1(z)[m_2(z) - m_1^2(z)m_0^{-1}(z)]^{-1} \\
&\quad \cdot\{m_0^{-1}(z)m_1(z)[\phi_0(z) + \phi_2(z)] - \phi_1(z) - \phi_3(z)\} - g_1(z) \\
&= a_{0,t,\lambda}m_0^{-1}(z)\phi_0(z) + a_{0,t,\lambda}m_0^{-1}(z)[\phi_2(z) - a_{0,t,\lambda}^{-1}m_0(z)g_1(z) - a_{0,t,\lambda}^{-1}m_1(z)g_1'(z)] \\
&\quad + m_0^{-1}(z)m_1(z)\{g_1'(z) + a_{0,t,\lambda}[m_2(z) - m_1^2(z)m_0^{-1}(z)]^{-1} \cdot \\
&\quad [m_0^{-1}(z)m_1(z)(\phi_0(z) + \phi_2(z)) - \phi_1(z) - \phi_3(z)]\} \\
&=: \Delta_1 + \Delta_2 + \Delta_3.
\end{aligned}$$

By Taylor expansions, it is straightforward that

$$\Delta_1 = \alpha_{0,t,h,\lambda}(z)\left(1 + O_p((c_{0,t,\lambda}h)^{-1/2}) + O(h)\right),$$
$$\Delta_2 = \beta_{t,h,\lambda}(z)\left(1 + O_p((c_{0,t,\lambda}h)^{-1/2}) + O(h)\right) + O\left(\frac{a_{\tau,t,\lambda}}{a_{0,t,\lambda}} - 1\right).$$

By the facts that

$$a_{0,t,\lambda}^{-1/2} m_1(z) = \int \Gamma_2(u)(u-z)K_h(u-z)du + O_p(c_{0,t,\lambda}^{-1/2}h^{1/2}) = O(h^2),$$
$$\phi_3(z) = g_1(z)\int \Gamma_2(u)(u-z)K_h(u-z)du + h^2\Gamma_2(z)g_1'(z)\eta_1 + O(h^3) + O_p(c_{0,t,\lambda}^{-1/2}h^{1/2}),$$
$$\phi_2(z) = g_1(z) + O(h), \qquad m_2(z) = O(h^2),$$

we have

$$\Delta_3 = O_p(h^3) + O_p(c_{0,t,\lambda}^{-1/2}h^{1/2}).$$

Combining all the above results, condition (C6), and the facts that $\alpha_{t,h,\lambda}(z) = O_p((c_{0,t,\lambda}h)^{-1/2})$ and $\beta_{t,h,\lambda}(z) = O_p(h^2)$, we can get result (ii) in the lemma. □

Proof of Theorem 1 Without loss of generality, we assume that $g_0 = 0$ (see related discussion after equation (5) in Section 2). By Lemma 1, we have

$$\begin{aligned}
T_{t,h,\lambda} &= \frac{c_{0,t,\lambda}}{n_0\sigma^2}\sum_{i=1}^{n_0}\left[\alpha_{t,h,\lambda}(z)\right]^2(1+o(h^{\frac{1}{2}})) \\
&= \frac{c_{0,t,\lambda}}{n_0}\sum_{i=1}^{n_0}\frac{1}{a_{0,t,\lambda}^2[\Gamma_2(z_i)]^2}\sum_{k=1}^{t}(1-\lambda)^{2(t-k)}\sum_{j=1}^{n_k}[K_h(x_{kj}-z_i)]^2\xi_{kj}^2(1+o(h^{\frac{1}{2}})) \\
&+ \frac{c_{0,t,\lambda}}{n_0}\sum_{i=1}^{n_0}\frac{1}{a_{0,t,\lambda}^2[\Gamma_2(z_i)]^2}\Bigg\{\sum_{k=1}^{t}(1-\lambda)^{2(t-k)}\sum_{j\neq l}[K_h(x_{kj}-z_i)][K_h(x_{kl}-z_i)]\xi_{kj}\xi_{kl} \\
&+ \sum_{k\neq k'}(1-\lambda)^{t-k}(1-\lambda)^{t-k'}\sum_{j,l}[K_h(x_{kj}-z_i)][K_h(x_{k'l}-z_i)]\xi_{kj}\xi_{k'l}\Bigg\}(1+o(h^{\frac{1}{2}})) \\
&=: (T_1+T_2)(1+o(h^{\frac{1}{2}}))
\end{aligned}$$

Note that, as $h \to 0$,

$$\begin{aligned}
T_1 &= \frac{c_{0,t,\lambda}}{a_{0,t,\lambda}^2}\sum_{k=1}^{t}(1-\lambda)^{2(t-k)}\sum_{j=1}^{n_k}\xi_{kj}^2\frac{1}{n_0}\sum_{i=1}^{n_0}\frac{1}{[\Gamma_2(z_i)]^2}[K_h(x_{kj}-z_i)]^2 \\
&= \frac{c_{0,t,\lambda}\eta_1}{ha_{0,t,\lambda}^2}\sum_{k=1}^{t}(1-\lambda)^{2(t-k)}\sum_{j=1}^{n_k}\xi_{kj}^2\frac{\Gamma_1(x_{kj})}{\Gamma^2(x_{kj})}(1+O(h)+O_p((n_0h)^{-\frac{1}{2}})).
\end{aligned}$$

It is easy to see that

$$E(T_1) = \tilde{\mu}_h + o(h^{-\frac{1}{2}}), \quad \mathrm{Var}(T_1) = \frac{d_{0,t,\lambda}}{b_{0,t,\lambda}^2 h^2}(1 + o(1)) = O((c_{0,t,\lambda}h^2)^{-1}),$$

where the last equation is from condition (C7). Thus, we have

$$T_1 = E(T_1) + O_p(\sqrt{\mathrm{Var}(T_1)}) = \frac{\eta_1}{h}\int \frac{\Gamma_1(u)}{\Gamma_2(u)}du + o_p(h^{-1/2}).$$

Similar to the manipulations for T_1, we have

$$\begin{aligned} T_2 = {} & \frac{c_{0,t,\lambda}}{a_{0,t,\lambda}^2 h}\Bigg\{ \sum_{k=1}^{t}(1-\lambda)^{2(t-k)} \sum_{j\neq l} \frac{\Gamma_1(x_{kj})}{[\Gamma_2(x_{kj})]^2} K * K((x_{kj} - x_{kl})/h)\xi_{kj}\xi_{kl} \\ & + \sum_{k\neq k'}(1-\lambda)^{t-k}(1-\lambda)^{t-k'} \sum_{j,l} \frac{\Gamma_1(x_{kj})}{[\Gamma_2(x_{kj})]^2} K * K((x_{kj} - x_{k'l})/h)\xi_{kj}\xi_{k'l} \Bigg\}(1 + O(h) + O_p((n_0h)^{-\frac{1}{2}})) \\ =: {} & (T_{21} + T_{22})(1 + O(h) + O_p((n_0h)^{-\frac{1}{2}})). \end{aligned}$$

Since $h^{1/2}(T_{21} + T_{22})$ can be written as a symmetric quadratic function of ξ_{kj}, for $j = 1, \ldots, n_k$ and $k = 1, \ldots$, with symmetric matrix $(\nu_{ij})_{N_t \times N_t}$ which has vanishing diagonal elements, here we can use Theorem 5.2 in de Jong (1987) to show the asymptotic normality of $h^{1/2}(T_{21} + T_{22})$. Obviously, the expectation of $T_{21} + T_{22}$ is zero. It can be checked that

$$\begin{aligned} \mathrm{Var}(h^{1/2}T_{21}) &= h\frac{e_{0,t,\lambda}}{b_{0,t,\lambda}^2}\tilde{\sigma}_h^2(1 + o(1)), \\ \mathrm{Var}(h^{1/2}T_{22}) &= h\left(1 - \frac{e_{0,t,\lambda}}{b_{0,t,\lambda}^2}\right)\tilde{\sigma}_h^2(1 + o(1)). \end{aligned}$$

Thus, the asymptotic variance of $h^{1/2}(T_{21} + T_{22})$ is $h\tilde{\sigma}_h^2$, after noting $\mathrm{Cov}(T_{21}, T_{22}) = 0$. Finally, by certain straightforward algebraic manipulations, we can verify that ν_{ij}s satisfy all the conditions given in Theorem 5.2 of de Jong (1987). Using this theorem and all the results above about T_1 and T_2, we have the result in Theorem 1. □

Proof of Theorem 2

(i). Without loss of generality, we assume that $g_0 = 0$. Thus, $g_1 = \delta$. By Lemma 1, we have

$$\begin{aligned} T_{t,h,\lambda} = {} & \frac{c_{0,t,\lambda}}{n_0\sigma^2}\sum_{i=1}^{n_0}\alpha_{t,h,\lambda}^2(z_i)(1 + o(h^{\frac{1}{2}})) + \frac{c_{0,t,\lambda}}{n_0\sigma^2}\sum_{i=1}^{n_0}[\delta(z_i) + \beta_{t,h,\lambda}(z_i)]^2(1 + o_p(1)) \\ & + \frac{2c_{0,t,\lambda}}{n_0\sigma^2}\sum_{i=1}^{n_0}\alpha_{t,h,\lambda}(z_i)\beta_{t,h,\lambda}(z_i)(1 + o_p(1)) + \frac{2c_{0,t,\lambda}}{n_0\sigma^2}\sum_{i=1}^{n_0}\alpha_{t,h,\lambda}(z_i)\delta(z_i)(1 + o_p(1)) \\ =: {} & T_1 + T_2 + (T_3 + T_4)(1 + o_p(1)). \end{aligned}$$

Obviously, T_1 is equivalent to $T_{t,h,\lambda}$ under the IC condition. It is straightforward to see that

$$\beta_{t,h,\lambda}(z) = \frac{h^2}{2}\delta''(z)\eta_1(1+o_p(1)).$$

By this result, we have $T_2 = c_{0,t,\lambda}\zeta_\delta(1+o_p(1))$, and

$$T_3 = \frac{h^2\eta_1 a_{0,t,\lambda}}{b_{0,t,\lambda}\sigma^2}\sum_{k=1}^{t}(1-\lambda)^{t-k}\sum_{j=1}^{n_k}\frac{\Gamma_1(x_{kj})}{\Gamma_2(x_{kj})}\varepsilon_{kj}\delta''(x_{kj})(1+o_p(1)).$$

Note that $\frac{1}{\sqrt{b_{0,t,\lambda}}}\sum_{k=1}^{t}(1-\lambda)^{t-k}\sum_{j=1}^{n_k}\frac{\Gamma_1(x_{kj})}{\Gamma_2(x_{kj})}\varepsilon_{kj}\delta''(x_{kj})$ is stochastically bounded. Thus, by condition (C5'), we have $T_3 = o_p(h^{-1/2})$. Similarly,

$$\begin{aligned} T_4 &= \frac{2a_{0,t,\lambda}}{b_{0,t,\lambda}\sigma^2}\sum_{k=1}^{t}(1-\lambda)^{t-k}\sum_{j=1}^{n_k}\frac{\Gamma_1(x_{kj})}{\Gamma_2(x_{kj})}\varepsilon_{kj}\delta(x_{kj}) \\ &= O_p((c_{0,t,\lambda}\int\delta^2(u)\Gamma_1(u)du)^{\frac{1}{2}}) = o_p(h^{-\frac{1}{2}}). \end{aligned}$$

By all these results and Theorem 1, we have result (i) in this theorem.

(ii). This result follows directly from result (i). □

Proof of Proposition 2

This proposition follows from Theorems 1 and 2 and from the proof of Theorem 4 in Horowitz and Spokoiny (2001). Here we just highlight some key steps, and the details are omitted. One important step is to derive the critical value of $\tilde{T}_{t,\mathcal{H},\lambda}$, denoted as C_α, for any given false alarm rate $0<\alpha<1$. By Lemmas 11 and 12 in Horowitz and Spokoiny (2001) and the proof of Theorem 1, we can show that, for $c_{0,t,\lambda}\to\infty$, $C_\alpha \le 2\sqrt{\ln\ln c_{0,t,\lambda} - \ln\alpha}$. Then, by Theorem 2, we have

$$T_{t,h,\lambda} = \frac{c_{0,t,\lambda}}{\sigma^2}[\zeta_1 + h^4\zeta_2 + 2h^2\zeta_3](1+o_p(1)) + \tilde{\mu}_h + O_p(h^{-\frac{1}{2}}), \tag{A.1}$$

where $\zeta_3 = \int\delta(u)\delta''(u)\Gamma_1(u)du$. When ζ_1 satisfies the condition $\zeta_1 > M_1(c_{0,t,\lambda}^{-1}\ln\ln c_{0,t,\lambda})^{\frac{8}{9}}$ for some sufficiently large M_1, by choosing h^* in the order of $(c_{0,t,\lambda}^{-1}\ln\ln c_{0,t,\lambda})^{\frac{2}{9}}$, the term $c_{0,t,\lambda}\zeta_1$ on the right hand side of (A.1) would dominate other terms, and it would be larger than $\tilde{\sigma}_h C_\alpha$ as well. Obviously, we can obtain such h^* by setting $h^* = h_{\max}\gamma^{-j_n}$, where j_n is the integer part of $\ln[kh_{\max}/(c_{0,t,\lambda}^{-1}\ln\ln c_{0,t,\lambda})^{\frac{2}{9}}]/\ln\gamma$ for some constant $k>0$. □

References

Capizzi, G., and Masarotto, G. (2003), "An Adaptive Exponentially Weighted Moving Average Control Chart," *Technometrics*, 45, 199–207.

Colosimo, B. M., and Pacella, M. (2007), "On the Use of Principal Component Analysis to Identify Systematic Patterns in Roundness Profiles," *Quality and Reliability Engineering International*, 23, 707–725.

De Jong, P. (1987), "A Central Limit Theorem for Generalized Quadratic Forms," *Probability Theory and Related Fields*, 75, 261–277.

Ding, Y., Zeng, L., and Zhou, S. (2006), "Phase I Analysis for Monitoring Nonlinear Profiles in Manufacturing Processes," *Journal of Quality Technology*, 38, 199–216.

Fan, J., and Gijbels, I. (1996), *Local Polynomial Modeling and Its Applications*, Chapman and Hall, London.

Fan, J., and Marron, S. (1994), "Fast Implementation of Nonparametric Curve Estimators," *Journal of Computational and Graphical Statistics*, 3, 35–56.

Fan, J., Zhang, C., and Zhang, J. (2001), "Generalized Likelihood Ratio Statistics and Wilks Phenomenon," *The Annals of Statistics*, 29, 153–193.

Han D., and Tsung, F. (2006), "A Reference-Free Cuscore Chart for Dynamic Mean Change Detection and a Unified Framework for Charting Performance Comparison," *Journal of American Statistical Association*, 101, 368–386.

Hart, J. D. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests*, Springer, New York.

Hawkins, D. M. (1987), "Self-Starting CUSUM charts for Location and Scale," *The Statistician*, 36, 299–316.

Hawkins, D. M., and Maboudou-Tchao, E. M. (2007), "Self-Starting Multivariate Exponentially Weighted Moving Average Control Charting," *Technometrics*, 49, 199–209.

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

Hawkins, D. M., Qiu, P., and Kang, C. W. (2003), "The Changepoint Model for Statistical Process Control," *Journal of Quality Technology*, 35, 355–366.

Horowitz, J. L., and Spokoiny, V. G. (2001), "An Adaptive, Rate-Optimal Test of a Parametric Mean-Regression Model Against a Nonparametric Alternative," *Econometrica*, 69, 599–631.

Jensen, W. A., Jones, L. A., Champ, C. W., and Woodall, W. H. (2006), "Effects of Parameter Estimation on Control Chart Properties: A Literature Review," *Journal of Quality Technology* 38, 349–364.

Jones, L. A. (2002), "The Statistical Design of EWMA Control Charts with Estimated Parameters," *Journal of Quality Technology*, 34, 277–288.

Kang, L., and Albin, S. L. (2000), "On-Line Monitoring When the Process Yields a Linear Profile," *Journal of Quality Technology*, 32, 418–426.

Kazemzadeh, R. B., Noorossana, R., and Amiri, A. (2008), "Phase I Monitoring of Polynomial Profiles," *Communications in Statistics: Theory and Methods*, 37, 1671–1686.

Kim, K., Mahmoud, M. A., and Woodall, W. H. (2003), "On the Monitoring of Linear Profiles," *Journal of Quality Technology*, 35, 317–328.

Lada, E. K., Lu, J. -C., and Wilson, J. R. (2002), "A Wavelet-Based Procedure for Process Fault Detection," *IEEE Transactions on Semiconductor Manufacturing*, 15, 79–90.

Lucas, J. M., and Saccucci, M. S. (1990), "Exponentially Weighted Moving Average Control Scheme Properties and Enhancements," *Technometrics*, 32, 1–29.

Mahmoud, M. A., Parker, P. A., Woodall, W. H., and Hawkins, D. M. (2007), "A Change Point Method for Linear Profile Data," *Quality and Reliability Engineering International*, 23, 247–268.

Mahmoud, M. A., and Woodall, W. H. (2004), "Phase I Analysis of Linear Profiles with Calibration Applications," *Technometrics*, 46, 380–391.

Rauf, S., Dauksher, W. J., Clemens, S. B., and Smith, K. H. (2002), "Model for a Multiple-Step Deep Si Etch Process," *Journal of Vacuum Science and Technology A*, 20, 1177–1190.

Seifert, B., Brockmann, M., Engel, J., and Gasser, T. (1994), "Fast Algorithms for Nonparametric Estimation," *Journal of Computational and Graphical Statistics*, 3, 192–213.

Walker, E., and Wright, S. P. (2002), "Comparing Curves Using Additive Models," *Journal of Quality Technology*, 34, 118–129.

Williams, J. D., Birch, J. B., Woodall. W. H., and Ferry, N. M. (2007a), "Statistical Monitoring of Heteroscedastic Dose-Response Profiles from High-Throughput Screening," *Journal of Agricultural, Biological, and Environmental Statistics*, 12, 216–235.

Williams, J. D., Woodall, W. H., and Birch, J. B. (2007b), "Statistical Monitoring of Nonlinear Product and Process Quality Profiles," *Quality and Reliability Engineering International*, 23, 925–941.

Woodall, W. H., Spitzner, D. J., Montgomery, D. C., and Gupta, S. (2004), "Using Control Charts to Monitor Process and Product Quality Profiles," *Journal of Quality Technology*, 36, 309–320.

Zhou, R., Zhang, H., Hao, Y., and Wang, Y. (2004), "Simulation of the Bosch Process with a String-Cell Hybrid Method," *Journal of Micromechanics and Microengineering*, 14, 851–858.

Zou, C., Qiu, P., and Hawkins, D. (2009), "Nonparametric control chart for monitoring profiles using the change point formulation," *Statistica Sinica*, 19, 1337–1357.

Zou, C., Tsung, F., and Wang, Z. (2007a), "Monitoring General Linear Profiles Using Multivariate EWMA Schemes," *Technometrics*, 49, 395–408.

Zou, C., Tsung, F., and Wang, Z. (2008), "Monitoring Profiles Based on Nonparametric Regression Methods," *Technometrics*, 50, 512–526.

Zou, C., Zhang, Y., and Wang, Z. (2006), "Control Chart Based on Change-Point Model for Monitoring Linear Profiles," *IIE Transactions*, 38, 1093–1103.

Zou, C., Zhou, C., Wang, Z., and Tsung, F. (2007b), "A Self-Starting Control Chart For Linear Profiles," *Journal of Quality Technology*, 39, 364-375.