

## CONTROL CHART FOR MONITORING NONPARAMETRIC PROFILES WITH ARBITRARY DESIGN

Peihua Qiu and Changliang Zou

*University of Minnesota and Nankai University*

*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 the assumption that design points within a profile are deterministic, and are unchanged from one profile to another. In practice, however, different profiles can have different design points and, in some cases, they are random. NPM is particularly challenging in such cases because it is difficult to properly combine data in different profiles purposes of data smoothing and process monitoring. In this paper, we propose an 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 procedure for smoothing individual profiles. Because of certain properties of the charting statistic, this control chart is fast to compute, easy to implement, and efficient in the detection of profile shifts. Some numerical results show that it works well.

*Key words and phrases:* 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 checks the 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, Mahmoud, and Woodall (2003), Mahmoud and Woodall (2004), Zou, Zhang, and Wang (2006), Zou et al. (2007), and Mahmoud et al. (2007), among several others. Extensions to

multiple and/or polynomial profile models are discussed by Zou, Tsung, and Wang (2007), and Kazemzadeh, Noorossana, and Amiri (2008). Recently, non-linear profile models have been considered by some people, including Lada, Lu, and Wilson (2002), Ding, Zeng, and Zhou (2006), Colosimo and Pacella (2007), Williams et al. (2007), and Williams, Woodall and Birch (2007). NPM is discussed by Zou, Tsung, and Wang (2008), and Zou, Qiu, and Hawkins (2009). For an overview on profile monitoring, see Woodall et al. (2004).

The control charts mentioned above require the assumptions that design points within a profile are deterministic, 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 are i.i.d. random variables from a given distribution. Another common 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 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.

Here 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. 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. The 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 semiconductor example. Several remarks conclude the article in Section 5. Some technical details are provided in the Appendix.

## 2. Methodology

The proposed methodology is described in stages. In Section 2.1, we briefly introduce the statistical process control (SPC) problem and the EWMA control chart. 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 and bandwidth

parameters, used in EWMA and nonparametric regression, are discussed in Sections 2.3 and 2.4, respectively. 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, this 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 and that is also our focus.

Let  $\{X_k, k = 1, 2, \dots\}$  be the sequential, Phase II, univariate, process measurements. Then, the EWMA control chart is based on  $S_k = (1 - \lambda)S_{k-1} + \lambda X_k$ , for  $k = 1, 2, \dots$ , 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,  $S_k = \lambda X_k + \lambda(1 - \lambda)X_{k-1} + \dots + \lambda(1 - \lambda)^{k-1}X_1$ . Thus,  $S_k$  is a weighted average of all observations, more recent observations receive more weight, and weights change exponentially over time.

**2.2. Monitoring nonparametric profiles when design points are arbitrary**

We are concerned with Phase II profile monitoring. At the  $k$ th time point, the profile is assumed to follow the nonparametric model

$$y_{kj} = g(x_{kj}) + \varepsilon_{kj}, \quad j = 1, \dots, n_k, \quad k = 1, 2, \dots, \tag{2.1}$$

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 the  $\varepsilon_{kj}$ 's are i.i.d. random errors with mean 0 and variance  $\sigma^2$ . Without loss of generality, we assume that  $x_{kj} \in [0, 1]$ , for all  $k$  and  $j$ . In cases for which the design points  $\mathbf{X}_k = \{x_{k1}, x_{k2}, \dots, x_{kn_k}\}$  are unchanged from one profile to another, the nonparametric EWMA chart of Zou, Tsung, and Wang (2008), called the NEWMA chart hereafter, first averages observed responses  $y_{kj}$ 's across different profiles at each design point, then detects potential profile shifts using the generalized likelihood ratio (GLR) test statistic. This idea cannot be applied when the response is observed at different design points in different profiles. A naive modification to Zou et al.'s method is to first obtain a nonparametric estimate of  $g$  from each profile data, and then to predict response values using the estimated  $g$  at some points  $\{z_1, z_2, \dots, z_n\}$  in the design interval that are unchanged from one profile to another. With this, the NEWMA chart can be applied to the predicted response values. However this naive approach, called the NAEWMA chart hereafter, may not be efficient since 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, we consider using a weighted local likelihood at any point  $z \in [0, 1]$  that combines the exponential weighting scheme used in EWMA at different time points with a local linear kernel smoothing procedure (cf., Fan and Gijbels (1996)). Thus,

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

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  $\min_{a,b} WL(a, b; z, \lambda, t)$ , is

$$\hat{g}_{t,h,\lambda}(z) = \frac{\sum_{k=1}^t \sum_{j=1}^{n_k} U_{kj}^{(t,h,\lambda)}(z) y_{kj}}{\sum_{k=1}^t \sum_{j=1}^{n_k} U_{kj}^{(t,h,\lambda)}(z)}, \quad (2.2)$$

where

$$U_{kj}^{(t,h,\lambda)}(z) = (1 - \lambda)^{t-k} K_h(x_{kj} - z) \left[ m_2^{(t,h,\lambda)}(z) - (x_{kj} - z) m_1^{(t,h,\lambda)}(z) \right],$$

$$m_l^{(t,h,\lambda)}(z) = \sum_{k=1}^t (1 - \lambda)^{t-k} \sum_{j=1}^{n_k} (x_{kj} - z)^l K_h(x_{kj} - z), \quad l = 0, 1, 2. \quad (2.3)$$

From the expression for  $WL(a, b; z, \lambda, t)$ , we can see that this estimator makes use of all available observations up to the  $t$ th time point, and different profiles are weighted as in a conventional EWMA chart.

If the process under monitoring is IC up to the  $t$ th time point, then  $\widehat{g}_{t,h,\lambda}$  in (2.2) should be close to the IC profile function, denoted as  $g_0$ . Thus, a charting statistic for profile monitoring can be defined based on the difference between  $\widehat{g}_{t,h,\lambda}$  and  $g_0$ . For simplicity, we first assume that  $g_0$  and the error variance  $\sigma^2$  are both known. In such cases, a more convenient way to define the charting statistic is to use  $\widehat{\xi}_{t,h,\lambda}(z)$ , the estimator defined by (2.2), after the  $y_{kj}$  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 given by

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

where

$$c_{0,t,\lambda} = \frac{a_{t_0,t_1,\lambda}^2}{b_{t_0,t_1,\lambda}}, \quad a_{t_0,t_1,\lambda} = \sum_{k=t_0+1}^{t_1} (1-\lambda)^{t_1-k} n_k,$$

$$b_{t_0,t_1,\lambda} = \sum_{k=t_0+1}^{t_1} (1-\lambda)^{2(t_1-k)} n_k,$$

and  $\Gamma_1$  is some pre-specified positive density function. In the expression for  $T_{t,h,\lambda}$ , 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:

$$T_{t,h,\lambda} \approx \frac{c_{0,t,\lambda}}{n_0} \sum_{i=1}^{n_0} [\widehat{\xi}_{t,h,\lambda}(z_i)]^2, \tag{2.4}$$

where  $z_i$ , for  $i = 1, \dots, n_0$ , are some pre-specified i.i.d. design points from  $\Gamma_1$ . 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 the  $z_i$  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 the  $z_i$  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 and certain empirical results presented in Section 3, selection of the  $z_i$  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

$$\frac{c_{0,t,\lambda}}{n_0} \sum_{i=1}^{n_0} [\hat{g}_{t,h,\lambda}(z_i) - g_0(z_i)]^2. \quad (2.5)$$

Namely, we can first compute profile estimators  $\hat{g}_{t,h,\lambda}$  from the original data and then construct the control chart accordingly. It can be shown that (2.5) and (2.4) are asymptotically equivalent under regularity conditions given in Appendix A. However, in finite-sample cases, properties of (2.5) depend on  $g_0$ . As a comparison, (2.4) 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.

We give some asymptotic properties of the charting statistic  $T_{t,h,\lambda}$  that can justify the performance of the NPC chart to a certain degree, and that shed some light on practical design of the chart as well. Our theorem establishes the asymptotic null distribution of  $T_{t,h,\lambda}$ , in which design points  $x_{kj}$  are assumed to be i.i.d. with a density  $\Gamma_2$  in each IC profile.

**Theorem 1.** *Under conditions (C1)–(C5) and (C7) given in Appendix A, when the process is IC,  $(T_{t,h,\lambda} - \tilde{\mu}_h) / \tilde{\sigma}_h \xrightarrow{\mathcal{L}} N(0, 1)$ , 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 result details the asymptotic behavior of  $T_{t,h,\lambda}$  under the OC model

$$y_{kj} = \begin{cases} g_0(x_{kj}) + \varepsilon_{kj}, & \text{if } 1 \leq k \leq \tau, \\ g_1(x_{kj}) + \varepsilon_{kj}, & \text{if } k > \tau, \end{cases} \quad (2.6)$$

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

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

$$\zeta_1 = \int \delta^2(u) \Gamma_1(u) du, \quad \zeta_2 = \int [\delta''(u)]^2 \Gamma_1(u) du.$$

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

- (i) if  $c_{0,t,\lambda}h\zeta_1 \rightarrow 0$ , then  $(T_{t,h,\lambda} - \tilde{\mu}_h - c_{0,t,\lambda}\zeta_\delta) / \tilde{\sigma}_h$  converges in distribution to  $N(0, 1)$ ;
- (ii) if  $\zeta_2 \rightarrow 0$ , 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  $\delta$  and its second order derivative. The charting statistic of the NEWMA chart has similar leading terms in its asymptotic expression. However, compared to NEWMA,  $T_{t,h,\lambda}$  can use a smaller bandwidth in local linear kernel smoothing because it uses observations from different profiles in its smoothing process. The NPC chart based on  $T_{t,h,\lambda}$  is more effective when the profile shift has large curvature (i.e.,  $\delta''$  is large), since small  $h$  would diminish the effect of  $(h^2\eta_1/2)\delta''(u)$  in the expression for  $\zeta_\delta$ . Thus we can get a better profile monitoring by using a random design scheme instead of an equal design scheme when the curvature of  $\delta$  is large. In Section 2.4, we discuss how to select the bandwidth  $h$  adaptively in the NPC chart to accommodate different magnitudes of  $\delta''$ .

### 2.3. Adaptive selection of the weighting parameter

It is well known that optimal selection of the weighting parameter  $\lambda$  used in EWMA charts depends on the target shift: small  $\lambda$  are effective for detecting small shifts and large  $\lambda$  are effective for detecting large shifts; an EWMA chart with a given  $\lambda$  cannot have a “nearly minimum” ARL for both small and large shifts (e.g., Lucas and Saccucci (1990)). This assertion also holds for our proposal.

**Proposition 1.** Under conditions (C1)–(C4) given in Appendix A,  $h \rightarrow 0$ ,  $n_k \rightarrow \infty$ ,  $n_0h^{3/2} \rightarrow \infty$ ,  $n_kh^3 \rightarrow \infty$ , and  $n_kh^5 \rightarrow 0$ , if  $(a_{\tau,t,\lambda}^2/b_{0,t,\lambda})h\zeta_1 \rightarrow 0$  and  $\zeta_2 < M$  for some constant  $M > 0$ , then

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

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

The proof of this proposition is analogous to that of Theorem 2 in Appendix B and is omitted. For simplicity, consider the case  $n_k = n$ . From expressions of  $a_{\tau,t,\lambda}$  and  $b_{0,t,\lambda}$ , we have

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

Then by Proposition 1, intuitively, if  $\zeta_\delta$  is small it would require a large value of  $t - \tau$  to signal, and this also depends heavily on the factor  $(2 - \lambda)/\lambda$ . When  $\lambda$  is chosen smaller,  $(2 - \lambda)/\lambda$  is larger. Consequently, the small shift would be detected quicker. On the other hand if  $\zeta_\delta$  is large, then we can expect the run length  $t - \tau$  to be relatively small as long as  $\lambda$  is chosen relatively large. If  $\lambda$  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), here we suggest an adaptive procedure for choosing the weighting parameter  $\lambda$ . 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 more efficiently. To be specific, let  $\psi$  be a score function used for determining the adaptive weights. Capizzi and Masarotto (2003) propose several candidates for  $\psi$ . For simplicity, we suggest using

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

where  $0 < \lambda_0 \leq 1$  and  $l_0 > 0$  are two parameters,  $\lambda_0$  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  $\lambda_0$  and  $l_0$  is given in Section 2.7. Then the NPC chart with the adaptive weight, denoted as NPC-W, signals when

$$T_{t,h,\psi_{l_0,\lambda_0}}(T_{t,h}^*) > L, \quad (2.7)$$

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  $\{(x_{tj}, y_{tj}), j = 1, \dots, n_t\}$  are used. It is easy to check that  $T_{t,h}^*$  is actually  $T_{t,h,1}$ ; it is therefore easy to compute. The NPC-W chart thus essentially combines the EWMA and Shewhart procedures in a natural way. It is worth mentioning that implementation does not 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 see that, after choosing  $\lambda_0$  and  $l_0$  properly, the NPC-W chart provides 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, and it is widely recognized that the optimal  $h$  for nonparametric curve estimation is generally not optimal for testing (e.g., Hart (1997)). A uniformly most powerful test usually does not exist due to the fact that nonparametric regression functions are of infinite dimension, but the term  $(h^2\eta_1/2)\delta''(\cdot)$  in Theorem 2 tells us that appropriate selection of  $h$  would improve testing power. Intuitively, a smaller  $h$  is more effective in detecting shifts with large curvature (i.e., large  $\delta''$ ), and a larger  $h$  would perform better when shifts are flat or smooth (i.e., small  $\delta''$ ). This motivates the adaptive selection procedure 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. Let the set of admissible smoothing parameter values be

$$\mathcal{H} = \{h_j = h_{\max}\gamma^{-j} : h_j \geq h_{\min}, j = 0, \dots, J_n\}, \tag{2.8}$$

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

$$\tilde{T}_{t,\mathcal{H},\lambda} = \max_{h \in \mathcal{H}} \frac{T_{t,h,\lambda} - \tilde{\mu}_h}{\tilde{\sigma}_h}, \tag{2.9}$$

where  $\tilde{\mu}_h$  and  $\tilde{\sigma}_h^2$  are, respectively, the asymptotic expectation and variance of  $T_{t,h,\lambda}$ , given 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), assuming that  $h_{\min}$  and  $h_{\max}$  both satisfy condition (C5),  $\zeta_1 > M_1(c_{0,t,\lambda}^{-1} \ln \ln c_{0,t,\lambda})^{8/9}$ , and  $\zeta_2 < M_2$ , where  $M_1$  and  $M_2$  are positive constants, then  $\tilde{T}_{t,\mathcal{H},\lambda}$  is consistent under model (2.6) in the sense that its power converges to 1 as  $t$  increases.*

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

### 2.5. Some computational issues

For on-line process monitoring, which generally handles a large number of profiles, fast implementation is important, and some computational issues are worth our careful examination. For the proposed charts, computation of  $T_{t,h,\lambda}$

might be time-consuming, and substantial amount of storage of past profile observations is required as well. Here we provide updating formulas for computing the charting statistic that greatly simplify the computation and lessen the storage requirement. Let

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

Then,  $m_l^{(t,h,\lambda)}(z)$  in (2.3) 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  $q_l^{(t,h,\lambda)}(z)$ ,  $l = 0, 1$ , 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

$$\begin{aligned}\hat{g}_{t,h,\lambda}(z) &= \left[ M^{(t,h,\lambda)} \right]^{-1} \left\{ (1 - \lambda)^2 M^{(t-1,h,\lambda)} \hat{g}_{t-1,h,\lambda} + \left[ \tilde{q}_0^{(t,h)} m_2^{(t,h,\lambda)} - \tilde{q}_1^{(t,h)} m_1^{(t,h,\lambda)} \right] \right. \\ &\quad \left. + (1 - \lambda) \left[ q_0^{(t-1,h,\lambda)} \tilde{m}_2^{(t,h)} - q_1^{(t-1,h,\lambda)} \tilde{m}_1^{(t,h)} \right] \right\},\end{aligned}\quad (2.10)$$

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 side of (2.10), dependence on  $z$  in each function is not made explicit in notation for simplicity.

Using the above updating formulas, implementation of the NPC chart can be briefly described as follows. At time point  $t$ , we compute  $\tilde{m}_l^{(t,h)}(z)$ , 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  $\{z_i, i = 1, 2, \dots, n_0\}$ ). 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. Finally,  $\hat{g}_{t,h,\lambda}(z)$  is computed from (2.10) and the test statistic  $T_{t,h,\lambda}$  is computed from  $\hat{g}_{t,h,\lambda}(z)$ , after replacing  $y_{kj}$  by  $\xi_{kj}$ . This algorithm only requires  $O(n_0 n_k h)$  operations for monitoring each profile, which is of 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  $\sigma^2$  (model (2.1)) and, in practice, both 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 is considerable uncertainty in the estimates, which in turn distorts 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 (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, Tsung, and Wang (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 them 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 that can substantially reduce the required IC profile samples.

The basic idea of the self-starting version is to replace  $g_0$  and  $\sigma^2$ , both of which are used in defining  $\xi_{kj}$ , 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, denoted as  $\hat{g}_0^{(t)}(x)$ . The variance  $\sigma^2$  can be estimated recursively by

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

Then, the self-starting version, denoted NPC-S, is the control chart based on the charting statistic  $\hat{T}_{t,h,\lambda}$  that is constructed in the same way as  $T_{t,h,\lambda}$  in (2.4), except that  $\xi_{kj}$  needs to be replaced by  $\tilde{\xi}_{kj} = [y_{kj} - \hat{g}_0^{(k-1)}(x_{kj})] / \hat{\sigma}_{k-1}$ , 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. It is straightforward to show that when the process is IC,

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

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  $\sigma^2$  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: it reduces computation and storage requirement with very little loss of efficiency; it may reduce the “masking-effect” (Hawkins (1987)) to a certain extent. For the last, when the potential shift occurs after time  $t_0$ , the estimates  $\widehat{g}_0^{(t_0)}(x)$  and  $\widehat{\sigma}_{t_0}^2$  are not contaminated by the OC observations, which is not the case if these estimates are updated at every time point.

It can be seen that the NPC-S chart accommodates adaptive selection of the weight and bandwidth parameters. The resulting chart, denoted NPC-SWB, offers balanced protection against shifts of different magnitudes, and adapts to the smoothness of the IC and OC profile functions. Formulation of the NPC-SWB chart can be readily obtained by incorporating (2.7) and (2.9) into  $\widehat{T}_{t,h,\lambda}$ . Performance is investigated in Section 3.

## 2.7. 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 cannot do much about choosing them. In some others (cf., Zou, Tsung, and Wang (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 provide information about more details of the regression function  $g$  in the former case. If available, random design is a good choice. This amounts to determining a proper design distribution  $\Gamma_2$ , 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 of Zou, Tsung, and Wang (2008) because, in computing  $T_{t,h,\lambda}$ , roughly  $c_{0,t,\lambda} = [(2 - \lambda)/\lambda]n$  observations are actually used.

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

**On choosing  $l_0$  and  $\lambda_0$  used in the NPC-W chart:** From Section 2.3,  $\lambda_0$  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 uses the minimum weight  $\lambda_0$  if  $T_{t,h,1} < l_0$ . We suggest using the upper  $\alpha_0$  percentile of  $T_{t,h,1}$  as the value of  $l_0$ ; this could be obtained by simulation before profile monitoring. An appropriate method for determining  $\alpha_0$  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 is reasonable to choose  $\alpha_0$  to be  $c/ARL_0$  so that the chart NPC-W can achieve the given IC ARL value. With respect to  $\lambda_0$ , 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:** The parameters  $\gamma$ ,  $J_n$ ,  $h_{\max}$ , and  $h_{\min}$  should satisfy certain conditions to bring in the corresponding asymptotic results, see Appendix A for related discussion. Based on simulations, we noticed that the proposed control chart was actually quite robust to them; this is consistent with the findings in Horowitz and Spokoiny (2001). Theoretical arguments and numerical studies suggest using  $1 < \gamma < 2$ ,  $J_n$  could be 4, 5 or 6,  $h_{\max} = Mc_{0,t,\lambda}^{-1/7}$ , and  $h_j = h_{\max}\gamma^{-j}$ , for  $j = 1, \dots, J_n$ , where  $0.5 \leq M \leq 2$  is a constant. Note that the recommended value  $h_{\max} = Mc_{0,t,\lambda}^{-1/7}$  is partially due to (C5) of Appendix A.

**On the NPC-S chart:** As suggested by Hawkins, Qiu, and Kang (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  $\hat{T}_{t,h,\lambda}$ .

### 3. Simulation Study

We present some simulation results about the proposed NPC chart. Throughout, the kernel function was chosen to be the Epanechnikov kernel function  $K(x) = 0.75(1 - x^2)I(-1 \leq x \leq 1)$ . The IC ARL was fixed at 200. The error distribution was assumed to be Normal. For simplicity, we assumed that  $n_k = n = 20$  for all  $k$ ,  $x_{kj} \sim \text{Uniform}(0, 1)$ , for  $j = 1, \dots, n$ ,  $z_i = (i - 0.5)/n_0$ , for  $i = 1, \dots, n_0$ , and  $n_0 = 40$ . All the ARL results in this section were obtained

from 50,000 replications, unless indicated otherwise. In addition, we focused on the steady-state OC ARL behavior of each chart (Hawkins and Olwell (1998)), and assumed that  $\tau = 30$  (cf., the OC model (2.6)). When computing the ARL values, any simulation run in which a signal occurred before the  $(\tau + 1)$ th profile was discarded. As a side note, our numerical results (not reported here to save space) show that steady-states of the related charts considered in this section were reached when  $\tau$  was as small as 10 in all cases considered.

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

We investigated 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  $\lambda$  should both be pre-specified. Therefore, we first compared the charts NPC, NEWMA, and NAEWMA in such cases. Following the recommendations of Zou, Tsung, and Wang (2008),  $h$  was 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  $h_2$  was considered because the actual number of observations used in the NPC chart at each time point was  $c_{0,t,\lambda}$ , or roughly  $n(2 - \lambda)/\lambda$ . In all three charts,  $\lambda$  was chosen to be 0.1 or 0.2. The IC model used was  $g_0(x) = 1 - \exp(-x)$ , and two representative OC models were considered:

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

In case (I),  $g_1(x) - g_0(x) = \theta x$  is a straight line, while  $g_1(x) - g_0(x) = \theta \sin(2\pi(x - 0.5))$  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 in case (I), larger  $h$  (i.e.,  $h_1$ ) yields better performance for both NPC and NEWMA charts. For a given bandwidth, the NPC chart outperforms the NEWMA chart uniformly, since the effective number of observations used in the NPC chart at each time point is larger than that used in the NEWMA chart. From the table in case (II), the NPC chart with the smaller  $h$  has better performance; however this is not the case for the NEWMA chart, since a smaller

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

$\theta$	NPC		NEWMA		NAEWMA	
	$h_1$	$h_2$	$h_1$	$h_2$	$h_1$	$h_2$
$\lambda = 0.1$						
(I) 0.100	75.90 (0.357)	87.40 (0.486)	89.50 (0.393)	102.00 (0.452)	101.00 (0.455)	103.00 (0.464)
0.200	27.60 (0.106)	33.10 (0.152)	32.70 (0.118)	38.50 (0.142)	39.70 (0.154)	41.00 (0.160)
0.300	14.80 (0.046)	17.40 (0.066)	17.20 (0.050)	19.70 (0.059)	20.50 (0.066)	20.90 (0.067)
0.400	9.85 (0.026)	11.40 (0.037)	11.20 (0.027)	12.60 (0.031)	13.30 (0.036)	13.60 (0.037)
0.600	5.90 (0.013)	6.68 (0.018)	6.64 (0.013)	7.33 (0.015)	7.64 (0.016)	7.76 (0.017)
0.800	4.27 (0.008)	4.80 (0.011)	4.74 (0.008)	5.19 (0.009)	5.38 (0.010)	5.44 (0.010)
1.200	2.82 (0.005)	3.12 (0.007)	3.12 (0.005)	3.37 (0.005)	3.46 (0.006)	3.50 (0.006)
1.600	2.18 (0.004)	2.37 (0.004)	2.38 (0.004)	2.56 (0.004)	2.62 (0.004)	2.65 (0.004)
(II) 0.100	68.90 (0.313)	65.30 (0.343)	72.40 (0.316)	81.60 (0.348)	94.20 (0.419)	95.60 (0.424)
0.200	23.20 (0.082)	22.20 (0.090)	24.50 (0.080)	27.60 (0.092)	31.70 (0.112)	32.20 (0.114)
0.300	12.40 (0.035)	12.00 (0.039)	13.00 (0.034)	14.30 (0.037)	16.00 (0.044)	16.20 (0.045)
0.400	8.29 (0.020)	8.03 (0.022)	8.76 (0.019)	9.53 (0.021)	10.50 (0.024)	10.50 (0.024)
0.600	5.11 (0.010)	4.96 (0.012)	5.32 (0.009)	5.75 (0.010)	6.15 (0.011)	6.20 (0.011)
0.800	3.73 (0.007)	3.65 (0.008)	3.89 (0.006)	4.16 (0.007)	4.42 (0.008)	4.43 (0.008)
1.200	2.51 (0.004)	2.42 (0.005)	2.62 (0.004)	2.78 (0.004)	2.94 (0.004)	2.95 (0.004)
1.600	1.94 (0.003)	1.90 (0.003)	2.04 (0.003)	2.15 (0.003)	2.25 (0.003)	2.26 (0.003)
$L$	9.49	13.05	14.39	19.02	12.88	13.32
$\lambda = 0.2$						
(I) 0.100	95.60 (0.452)	108.00 (0.501)	112.00 (0.519)	125.00 (0.595)	132.00 (0.625)	135.00 (0.648)
0.200	34.60 (0.148)	41.00 (0.174)	43.20 (0.188)	51.90 (0.228)	60.00 (0.269)	62.20 (0.281)
0.300	16.60 (0.063)	19.60 (0.076)	20.10 (0.076)	23.70 (0.089)	28.40 (0.116)	29.60 (0.120)
0.400	10.00 (0.031)	11.70 (0.036)	11.70 (0.036)	13.40 (0.045)	16.10 (0.056)	16.60 (0.058)
0.600	5.39 (0.013)	6.05 (0.013)	6.03 (0.013)	6.64 (0.018)	7.69 (0.020)	7.94 (0.022)
0.800	3.69 (0.009)	4.09 (0.009)	4.09 (0.009)	4.44 (0.009)	4.98 (0.011)	5.15 (0.011)
1.200	2.35 (0.004)	2.55 (0.004)	2.56 (0.004)	2.74 (0.004)	3.00 (0.005)	3.04 (0.005)
1.600	1.80 (0.003)	1.93 (0.003)	1.94 (0.003)	2.06 (0.003)	2.21 (0.003)	2.26 (0.003)
(II) 0.100	93.10 (0.461)	85.70 (0.407)	93.90 (0.429)	105.00 (0.479)	133.00 (0.630)	135.00 (0.639)
0.200	30.30 (0.130)	27.60 (0.112)	31.30 (0.125)	35.40 (0.148)	48.60 (0.235)	55.30 (0.242)
0.300	13.80 (0.049)	13.00 (0.045)	14.20 (0.049)	16.00 (0.054)	22.50 (0.081)	23.20 (0.086)
0.400	8.31 (0.022)	7.91 (0.022)	8.57 (0.022)	9.38 (0.027)	12.30 (0.036)	12.50 (0.038)
0.600	4.54 (0.009)	4.37 (0.009)	4.68 (0.009)	5.00 (0.009)	5.98 (0.012)	6.05 (0.013)
0.800	3.17 (0.004)	3.07 (0.004)	3.26 (0.004)	3.46 (0.004)	4.00 (0.007)	4.04 (0.007)
1.200	2.08 (0.004)	2.01 (0.004)	2.14 (0.004)	2.25 (0.004)	2.50 (0.004)	2.53 (0.004)
1.600	1.61 (0.003)	1.56 (0.003)	1.66 (0.003)	1.74 (0.003)	1.91 (0.003)	1.92 (0.003)
$L$	10.47	13.09	15.41	19.10	15.00	15.51

NOTE: Standard errors are in parentheses.

bandwidth in the NEWMA chart results in large bias in estimating the regression function and a reduced ability to detect profile shifts. Again in case (II), the NPC chart outperforms the NEWMA chart uniformly. The NPC chart outperformed the NAEWMA chart by a quite large margin in most cases, and the NAEWMA chart also performed uniformly worse than the NEWMA chart. Simulations (not reported here) showed that, when  $n_0$  was larger than 40, performance of the NPC

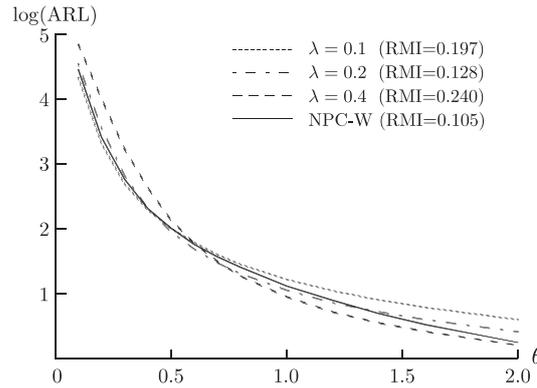


Figure 1. OC ARL comparison of the NPC-W chart and three NPC charts with  $\lambda = 0.1, 0.2,$  and  $0.4$ .

chart did not change much.

Next we consider the NPC-W chart in which the weight parameter  $\lambda$  was adaptively chosen so that the chart would be robust to shift size; the other parameters were chosen according to the practical guidelines given in Section 2.7 and, more specifically,  $\lambda_0 = 0.1$  and  $\alpha_0 = 0.05$ . Since we are mainly concerned about the robustness of the NPC-W chart to shift size, only the OC model (I) was considered here. The bandwidth  $h$  was chosen to be  $h_1$ . For comparison purposes, the OC ARL values of three NPC charts when  $\lambda = 0.1, 0.2,$  and  $0.4$  were also computed. To measure robustness of a chart T to shift size, the relative mean index (RMI) of Han and Tsung (2006) was used

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

where  $\text{ARL}_{\theta_i}(T)$  is the OC ARL of T for detecting a shift of size  $\theta_i$ , and  $\text{MARL}_{\theta_i}$  is the smallest value among such OC ARL values of all charts considered. Here  $\theta_i$  ranged from 0.1 to 2 with step size 0.1. Obviously, small  $\text{RMI}(T)$  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  $\lambda$  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. With its convenient implementation, the NPC-W chart looks to be a valuable improvement of the NPC chart.

Next, we considered the NPC-B chart. We chose  $\gamma = 1.4$ ,  $h_{\max} = 1.0[(2 - \lambda/\lambda)n]^{-1/7}$ , and  $h_j = h_{\max}\gamma^{-j}$ , for  $j = 1, \dots, 4$ , and took  $g_1(x) = 1 - \exp(-x) + 0.25 \cos(\theta\pi(x - 0.5))$ . By changing  $\theta$ , this model can cover various cases with

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

$\theta$	NPC-B	NPC		
		$h = 0.6$	$h = 0.3$	$h = 0.15$
0.250	8.20 (0.031)	7.84 (0.022)	8.27 (0.031)	10.8 (0.036)
0.500	9.05 (0.031)	8.95 (0.027)	9.45 (0.031)	12.2 (0.040)
0.750	11.20 (0.036)	11.40 (0.036)	10.70 (0.036)	13.9 (0.040)
1.000	14.70 (0.054)	16.20 (0.058)	14.10 (0.049)	15.9 (0.058)
2.000	32.30 (0.112)	86.30 (0.398)	35.00 (0.125)	26.9 (0.103)
3.000	31.00 (0.098)	49.00 (0.224)	50.60 (0.224)	24.4 (0.094)
4.000	63.50 (0.286)	166.00 (0.814)	174.00 (0.832)	48.0 (0.206)
5.000	75.60 (0.344)	120.00 (0.577)	125.00 (0.581)	81.4 (0.376)
$L$	4.27	8.75	9.61	13.32

NOTE: Standard errors are in parentheses.

different smoothness of  $\delta(\cdot)$ . For comparison purposes, we considered NPC charts with bandwidth 0.6, 0.3, and 0.15. Other parameters were chosen as in the example leading to 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  $\theta$ , but can be much worse in other ranges of  $\theta$ . The NPC-B chart is close to the best chart in all cases, since it adapts to the unknown smoothness of  $\delta(x)$ , and picks an appropriate bandwidth from  $\mathcal{H}$ .

We also investigated the numerical performance of the NPC-S chart, studying its IC run length distribution. It is often insufficient to summarize run length behavior by ARL, especially for self-starting control charts (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 is  $r$  and  $H_2(r)$  is the probability that the run length is  $r$  or more. For the example of Table 1 with  $h = h_1$  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 started monitoring after five IC profiles were collected. In the plot, the IC hazard starts around 0.0065, then drops quickly to values around  $0.005 = 1/200$ , and stabilizes. Except for short run-lengths, the geometric distribution is an excellent fit to the IC run length of the NPC-S chart, consistent with the findings in Hawkins and Maboudou-Tchao (2007) about a self-starting chart for monitoring multivariate Normal processes. The sample mean and sample standard deviation of the run lengths were 196 and 194, respectively, further confirming that the NPC-S chart works well under the IC condition. We conducted some other simulations with various combinations of

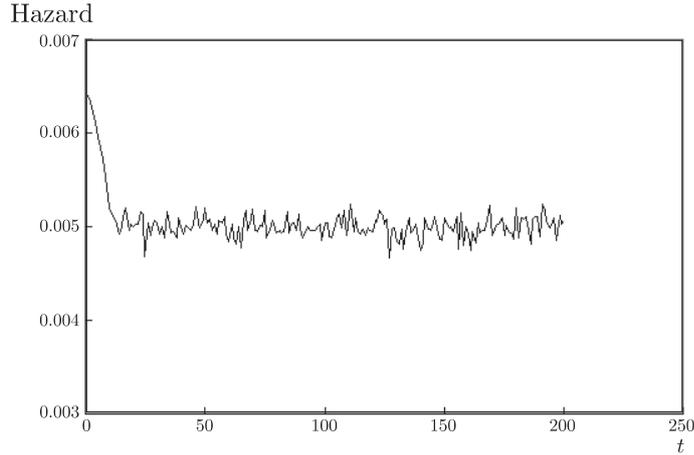


Figure 2. Hazard curve of the NPC-S chart.

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

$\theta$	NPC-S				NPC-SWB		
	$\tau = 40$		$\tau = 80$		$\tau = 40$	$\tau = 80$	
	$h_1$	$h_2$	$h_1$	$h_2$			
	0.100	163.00 (0.903)	176.00 (0.939)	143.00 (0.818)	161.00 (0.899)	151.00 (0.859)	132.00 (0.859)
	0.200	101.00 (0.738)	116.00 (0.778)	68.20 (0.505)	88.90 (0.626)	82.50 (0.626)	54.70 (0.474)
	0.300	46.10 (0.456)	60.40 (0.523)	24.90 (0.192)	34.50 (0.268)	32.20 (0.313)	19.60 (0.130)
	0.400	17.50 (0.188)	26.30 (0.291)	11.70 (0.054)	15.00 (0.085)	14.00 (0.098)	10.80 (0.040)
(I)	0.600	6.01 (0.018)	7.25 (0.027)	5.66 (0.018)	6.63 (0.018)	6.20 (0.018)	5.87 (0.018)
	0.800	3.88 (0.009)	4.37 (0.009)	3.77 (0.009)	4.22 (0.009)	4.14 (0.009)	3.95 (0.009)
	1.200	2.38 (0.004)	2.64 (0.004)	2.38 (0.004)	2.62 (0.004)	2.38 (0.004)	2.35 (0.004)
	1.600	1.81 (0.004)	1.95 (0.004)	1.81 (0.004)	1.96 (0.004)	1.64 (0.004)	1.64 (0.004)
	0.100	156.00 (0.872)	161.00 (0.926)	146.00 (0.823)	142.00 (0.814)	150.00 (0.836)	132.00 (0.827)
	0.200	78.70 (0.581)	82.20 (0.631)	55.20 (0.402)	54.50 (0.398)	68.00 (0.483)	48.10 (0.367)
	0.300	26.10 (0.206)	26.50 (0.237)	18.30 (0.103)	17.60 (0.103)	22.90 (0.165)	17.20 (0.080)
(II)	0.400	10.90 (0.063)	10.60 (0.067)	9.44 (0.036)	8.86 (0.031)	10.70 (0.040)	9.75 (0.031)
	0.600	4.92 (0.013)	4.68 (0.013)	4.77 (0.009)	4.56 (0.009)	5.34 (0.013)	5.10 (0.013)
	0.800	3.30 (0.004)	3.19 (0.004)	3.24 (0.004)	3.15 (0.004)	3.62 (0.009)	3.58 (0.009)
	1.200	2.13 (0.004)	2.05 (0.004)	2.09 (0.004)	2.03 (0.004)	2.11 (0.004)	2.10 (0.004)
	1.600	1.60 (0.004)	1.56 (0.004)	1.61 (0.004)	1.56 (0.004)	1.43 (0.004)	1.42 (0.004)
	$L$	10.47	13.09	10.47	13.09	4.070	4.070

NOTE: Standard errors are in parentheses.

$n$ ,  $h$ , and  $\Gamma_2$  to check whether the above conclusions held in other settings. These simulation results, not reported here but available from the authors, showed that the NPC-S chart performed satisfactorily in other cases as well, aside from cases in which  $n$  was too small (e.g.,  $n \leq 5$ ).

We examine the OC performance of the NPC-S chart as well. As demonstrated in the literature, the OC performance of self-starting charts is generally affected by the shift time point (e.g., Hawkins, Qiu, and Kang (2003)). We

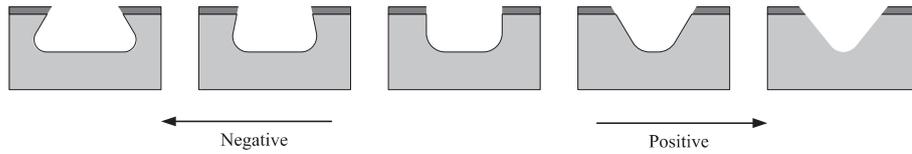


Figure 3. Illustrations of various etching profiles from a DRIE process.

consider shift times  $\tau = 40$  and  $\tau = 80$ . The simulation results in the 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 the two values of  $\tau$  when the shift size is large. For detecting small to moderate shifts, it generally performs better with a larger  $\tau$ , because the updated parameter estimates are more accurate in such a case under the IC condition, as confirmed by the table. As a comparison, in Table 3, we also present the OC ARLs of the NPC-SWB chart that is a combination of the self-starting chart and adaptive selection of the weight and bandwidth parameters. Its parameters were chosen to be 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.

#### 4. A Semiconductor Application

We applied the proposed NPC chart to a dataset obtained from the semiconductor manufacturing industry for monitoring a deep reactive ion etching (DRIE) process that is critical to the output wafer quality and which 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 (the two left-side and two right-side shapes 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).

The DRIE data have 21 profiles. The original data include images, like the ones shown in Figure 3. To monitor the DRIE process, one needs samples from individual profiles; these can be acquired by the scanning electron microscope (SEM). Using symmetry, we focused on the left half of each profile for monitoring purposes. Among the 21 profiles, and based on engineering knowledge, the first 18 were IC and the remaining 3 were OC. With these numbers, the IC profile

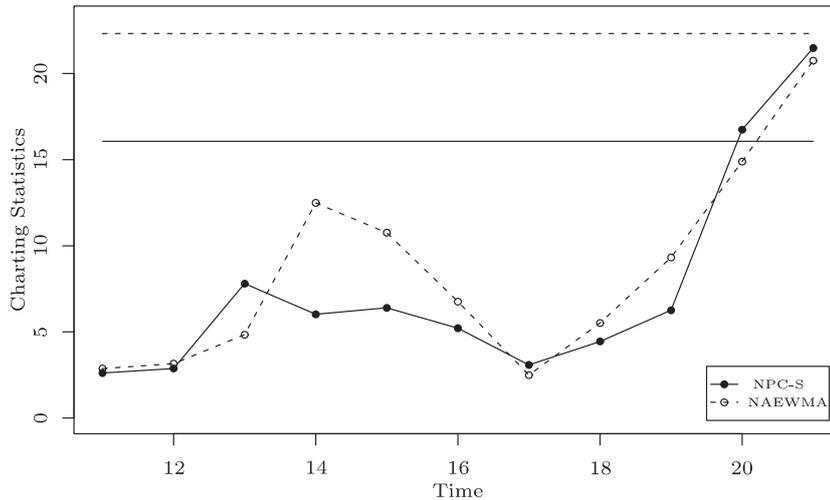


Figure 4. The NPC-S and NAEWMA control charts for monitoring the DRIE process. The solid and dashed horizontal lines indicate their control limits, respectively.

function  $g_0$  and the error standard deviation  $\sigma$  may not be accurately estimated from the IC data, so we used the self-starting chart NPC-S. For applications such as the current one, a better profile monitoring is anticipated when using a random design scheme. The corner part of the profile apparently contains the information regarding whether the profile is OC; after a rotation of  $45^\circ$ , we centered a normal distribution there so that half of the bottom trench was in its range and about 65% of the design points were in the corner part.

For each profile, we fixed  $n = 20$ , and dimensional readings were collected by SEM at the design points generated from the normal  $\Gamma_1$ . Using electronic sensor and information technologies, such a data acquisition process can be finished automatically by a computer. In the NPC-S chart, we fixed the IC ARL at 200,  $n_0$  at 40, and the  $z_i$ 's to be equally spaced over a comparable range. All other parameters of the NPC-S chart were chosen to be those used in the example of Table 3. The control limit was computed to be  $L = 16.07$  by simulation. We took the first 10 IC profiles as preliminary data, and profile monitoring started at the 11th profile. The charting statistic  $\hat{T}_{t,h,\lambda}$  values,  $t = 11, \dots, 21$ , are shown in Figure 4, along with the control limit. In that figure, we also give the NAEWMA chart and its control limit at 22.32. Parameters of the NAEWMA were  $\lambda = 0.2$ , IC ARL=200,  $n = 20$ , and  $z_1, \dots, z_n$  the same equally spaced points. In the plot, 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.

It took 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, suggesting convenience in on-line automatic profile monitoring.

**5. Summary and Concluding Remarks**

We have proposed a control chart for monitoring nonparametric profiles with arbitrary design. It effectively combines the EWMA control chart and a nonparametric 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 indicated 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.

Our proposed control chart operates under the assumption that observations within and between individual profiles are independent of each other. In some applications, within-profile observations are spatially or serially correlated, or between-profile observations are auto-correlated (cf., Williams et al. (2007), Williams, Woodall and Birch (2007), Zou et al. (2007)). More research is needed in these areas. Going further, sometimes we are interested in monitoring a multivariate relationship between a response variable and several predictors over time. We are not now aware of any existing research on this topic.

**Acknowledgements**

The authors thank the Editor, an associate editor, and two referees for many constructive suggestions and comments that greatly improved the quality of the paper. This research is supported in part by an NSF grant. Changliang Zou’s research is also supported in part by the grant No. 10771107 from NNSF of China.

**Appendix: Technical Details**

Throughout, we use the additional notations

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

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

$$\begin{aligned}\phi_i(z) &= \frac{1}{a_{0,t,\lambda}} \sum_{k=1}^t (1-\lambda)^{t-k} \sum_{j=1}^{n_k} (x_{kj} - z)^i K_h(x_{kj} - z) \varepsilon_{kj}, \quad i = 0, 1, \\ \phi_{i+2}(z) &= \frac{1}{a_{\tau,t,\lambda}} \sum_{k=\tau+1}^t (1-\lambda)^{t-k} \sum_{j=1}^{n_k} (x_{kj} - z)^i K_h(x_{kj} - z) g_1(x_{kj}), \quad i = 0, 1, \\ d_{t_0,t_1,\lambda} &= \sum_{k=t_0+1}^{t_1} (1-\lambda)^{4(t-k)} n_k, \quad e_{t_0,t_1,\lambda} = \sum_{k=t_0+1}^{t_1} (1-\lambda)^{4(t-k)} n_k^2.\end{aligned}$$

### Appendix A: Regularity Conditions Used In Section 2

- (C1) Density functions  $\Gamma_1$  and  $\Gamma_2$  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,  $u^3 K(u)$  and  $u^3 K'(u)$  are bounded, and  $\int_{-1}^1 u^4 K(u) du < \infty$ .
- (C4)  $E(|\varepsilon_{11}|^4) < \infty$ .
- (C5)  $n_0$ ,  $h$ , and  $c_{0,t,\lambda}$  satisfy  $n_0 \rightarrow \infty$ ,  $h \rightarrow 0$ ,  $n_0 h^{3/2} \rightarrow \infty$ ,  $c_{0,t,\lambda} \rightarrow \infty$ ,  $c_{0,t,\lambda} h^{3/2} \rightarrow \infty$  and  $c_{0,t,\lambda} h^8 \rightarrow 0$ .
- (C5')  $n_0$ ,  $h$ , and  $c_{0,t,\lambda}$  satisfy  $n_0 \rightarrow \infty$ ,  $h \rightarrow 0$ ,  $n_0 h^{3/2} \rightarrow \infty$ ,  $c_{0,t,\lambda} \rightarrow \infty$ ,  $c_{0,t,\lambda} h^3 \rightarrow \infty$  and  $c_{0,t,\lambda} h^5 \rightarrow 0$ .
- (C6)  $a_{\tau,t,\lambda}/a_{0,t,\lambda} - 1 = o(\min\{h^2, c_{0,t,\lambda}^{-1/2}\})$ .
- (C7) The  $n_k$ 's are such that  $\max_{1 \leq k \leq t} n_k / \min_{1 \leq k \leq t} n_k$  is bounded.

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  $[(2-\lambda)/\lambda] \min_{1 \leq k \leq t} n_k \leq c_{0,t,\lambda} \leq [(2-\lambda)/\lambda] \max_{1 \leq k \leq t} n_k$  for large  $t$ . Thus, if  $\lambda \rightarrow 0$ , we do not require  $n_k \rightarrow \infty$ . The conditions listed here are milder than those in Zou, Tsung, and Wang (2008) where the number of design points in each profile should go to infinity. (C6) is easily satisfied if  $t$  is large enough. (C7) implies that all  $n_k$ 's are of the same order, this is common in practice.

### Appendix B: Proofs

To prove the theorems of Section 2, a lemma is required.

**Lemma 1.** For any  $z \in [0, 1]$ , (i) under conditions in Theorem 1, we have  $\widehat{g}_{t,h,\lambda}(z) = \alpha_{t,h,\lambda}(z)(1 + o(h^{1/2}))$ ; (ii) under conditions in Theorem 2, we have  $\widehat{g}_{t,h,\lambda}(z) - g_1(z) = \alpha_{t,h,\lambda}(z)(1 + o(h^{1/2})) + \beta_{t,h,\lambda}(z)(1 + o_p(1))$ .

**Proof.** We only prove the second of these because the first one can be proved in a similar way. For simplicity, we suppress the symbol “ $(t, h, \lambda)$ ” in  $m_i^{(t,h,\lambda)}(z)$ . After some algebraic manipulations,

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

By Taylor expansions, it is straightforward that

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

Since

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

we have  $\Delta_3 = O_p(h^3) + O_p(c_{0,t,\lambda}^{-1/2} h^{1/2})$ . Combining the above results, (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 have (ii).

**Proof of Theorem 1.** Without loss of generality, take  $g_0 = 0$  (see the related discussion after (2.5) in Section 2). By Lemma 1,

$$\begin{aligned} T_{t,h,\lambda} &= \frac{c_{0,t,\lambda}}{n_0 \sigma^2} \sum_{i=1}^{n_0} [\alpha_{t,h,\lambda}(z)]^2 (1 + o(h^{1/2})) \\ &= \frac{c_{0,t,\lambda}}{n_0} \sum_{i=1}^{n_0} \frac{1}{a_{0,t,\lambda}^2 [\Gamma_2(z_i)]^2} \sum_{k=1}^t (1 - \lambda)^{2(t-k)} \sum_{j=1}^{n_k} [K_h(x_{kj} - z_i)]^2 \xi_{kj}^2 (1 + o(h^{1/2})) \\ & \quad + \frac{c_{0,t,\lambda}}{n_0} \sum_{i=1}^{n_0} \frac{1}{a_{0,t,\lambda}^2 [\Gamma_2(z_i)]^2} \left\{ \sum_{k=1}^t (1 - \lambda)^{2(t-k)} \sum_{j \neq l} [K_h(x_{kj} - z_i)] [K_h(x_{kl} - z_i)] \xi_{kj} \xi_{kl} \right\} \end{aligned}$$

$$\begin{aligned}
 & + \sum_{k \neq k'} (1-\lambda)^{t-k} (1-\lambda)^{t-k'} \sum_{j,l} [K_h(x_{kj} - z_i)] [K_h(x_{k'l} - z_i)] \xi_{kj} \xi_{k'l} \Big\} (1 + o(h^{1/2})) \\
 =: & (T_1 + T_2)(1 + o(h^{1/2})).
 \end{aligned}$$

Note that, as  $h \rightarrow 0$ ,

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

It is easy to see that

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

where the last equality is from (C7). Thus we have

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

Similar to the manipulations for  $T_1$ , we have

$$\begin{aligned}
 T_2 &= \frac{c_{0,t,\lambda}}{a_{0,t,\lambda}^2 h} \left\{ \sum_{k=1}^t (1-\lambda)^{2(t-k)} \sum_{j \neq l} \frac{\Gamma_1(x_{kj})}{[\Gamma_2(x_{kj})]^2} K * K\left(\frac{x_{kj} - x_{kl}}{h}\right) \xi_{kj} \xi_{kl} \right. \\
 & + \sum_{k \neq k'} (1-\lambda)^{t-k} (1-\lambda)^{t-k'} \sum_{j,l} \frac{\Gamma_1(x_{kj})}{[\Gamma_2(x_{kj})]^2} K * K\left(\frac{x_{kj} - x_{k'l}}{h}\right) \xi_{kj} \xi_{k'l} \Big\} \\
 & \times (1 + O(h) + O_p((n_0 h)^{-1/2})) \\
 =: & (T_{21} + T_{22})(1 + O(h) + O_p((n_0 h)^{-1/2})).
 \end{aligned}$$

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

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

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

**Proof of Theorem 2.**

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

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

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

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

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

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

Now (i) follows, and (ii) follows directly from (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, the details are omitted. One important step derives the critical value of  $\tilde{T}_{t,\mathcal{H},\lambda}$ , denoted as  $C_\alpha$ , 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  $c_{0,t,\lambda} \rightarrow \infty$ ,  $C_\alpha \leq 2\sqrt{\ln \ln c_{0,t,\lambda} - \ln \alpha}$ . Then, by Theorem 2, we have

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

where  $\zeta_3 = \int \delta(u) \delta''(u) \Gamma_1(u) du$ . When  $\zeta_1 > M_1 (c_{0,t,\lambda}^{-1} \ln \ln c_{0,t,\lambda})^{8/9}$  for some sufficiently large  $M_1$ , by choosing  $h^*$  in the order of  $(c_{0,t,\lambda}^{-1} \ln \ln c_{0,t,\lambda})^{2/9}$ , the term  $c_{0,t,\lambda} \zeta_1$  on the right side of (A.1) dominates other terms, and is larger than  $\tilde{\sigma}_h C_\alpha$  as well. We can obtain such  $h^*$  as  $h^* = h_{\max} \gamma^{-j_n}$ , where  $j_n$  is the integer part of  $\ln[k h_{\max} / (c_{0,t,\lambda}^{-1} \ln \ln c_{0,t,\lambda})^{2/9}] / \ln \gamma$  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. *Probab. Theory 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*. London: Chapman and Hall.
- Fan, J. and Marron, S. (1994). Fast implementation of nonparametric curve estimators. *Journal of Computational and Graphical Statistics* **3**, 35-56.
- Han, D. and Tsung, F. (2006). A reference-free cuscore chart for dynamic mean change detection and a unified framework for charting performance comparison. *J. Amer. Statist. Assoc.* **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*. Springer, New York.
- Hawkins, D. M., Qiu, P. and Kang, C. W. (2003). The changepoint model for statistical process control. *J. Quality Tech.* **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. *J. Quality Tech.* **38**, 349-364.

- Jones, L. A. (2002). The statistical design of EWMA control charts with estimated parameters. *J. Quality Tech.* **34**, 277-288.
- Kang, L. and Albin, S. L. (2000). On-line monitoring when the process yields a linear profile. *J. Quality Tech.* **32**, 418-426.
- Kazemzadeh, R. B., Noorossana, R. and Amiri, A. (2008). Phase I monitoring of polynomial profiles. *Comm. Statist. Theory Methods* **37**, 1671-1686.
- Kim, K., Mahmoud, M. A. and Woodall, W. H. (2003). On the monitoring of linear profiles. *J. Quality Tech.* **35**, 317-328.
- Lada, E. K., Lu, J.-C. and Wilson, J. R. (2002). A wavelet-based procedure for process fault detection. *IEEE Trans. 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. *J. Vacuum Science and Tech. A* **20**, 1177-1190.
- Seifert, B., Brockmann, M., Engel, J. and Gasser, T. (1994). Fast algorithms for nonparametric estimation. *J. Comput. Graph. Statist.* **3**, 192-213.
- Walker, E. and Wright, S. P. (2002). Comparing curves using additive models. *J. Quality Tech.* **34**, 118-129.
- Williams, J. D., Birch, J. B., Woodall, W. H. and Ferry, N. M. (2007). 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. (2007). 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. *J. Quality Tech.* **36**, 309-320.
- Zhou, R., Zhang, H., Hao, Y. and Wang, Y. (2004). Simulation of the bosch process with a string-cell hybrid method. *J. 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. *Statist. Sinica* **19**, 1337-1357.
- Zou, C., Tsung, F. and Wang, Z. (2007). 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. (2007). A self-starting control chart for linear profiles. *J. Quality Tech.* **39**, 364-375.

School of Statistics, University of Minnesota, Minneapolis, MN 55455, USA.

E-mail: qiu@stat.umn.edu

Department of Statistics, Nankai University, Tianjin, 300071, China.

E-mail: chlzhou@yahoo.com.cn

(Received November 2008; accepted August 2009)