

# Monitoring Profiles Based on Nonparametric Regression Methods

**Changliang ZOU**

LPMC and Department of Statistics  
School of Mathematical Sciences  
Nankai University  
Tianjin, China  
(chlzou@yahoo.com.cn)

**Fugee TSUNG**

Department of Industrial Engineering  
and Logistics Management  
Hong Kong University of Science and Technology  
Clear Water Bay, Kowloon, Hong Kong  
(season@ust.hk)

**Zhaojun WANG**

LPMC and Department of Statistics  
School of Mathematical Sciences  
Nankai University  
Tianjin, China  
(zjwang@nankai.edu.cn)

The use of statistical process control (SPC) in monitoring and diagnosis of process and product quality profiles remains an important problem in various manufacturing industries. The SPC problem with a nonlinear profile is particularly challenging. This article proposes a novel scheme to monitor changes in both the regression relationship and the variation of the profile online. It integrates the multivariate exponentially weighted moving average procedure with the generalized likelihood ratio test based on nonparametric regression. The proposed scheme not only provides an effective SPC solution to handle nonlinear profiles, which are common in industrial practice, but it also resolves the latent problem in popular parametric monitoring methods of being unable to detect certain types of changes due to a misspecified, out-of-control model. Our simulation results demonstrate the effectiveness and efficiency of the proposed monitoring scheme. In addition, a systematic diagnostic approach is provided to locate the change point of the process and identify the type of change in the profile. Finally, a deep reactive ion-etching example from semiconductor manufacturing is used to illustrate the implementation of the proposed monitoring and diagnostic approach.

**KEY WORDS:** Exponentially weighted moving average; Generalized likelihood ratio test; Lack-of-fit test; Local linear smoother; Nonlinear profile; Statistical process control.

## 1. INTRODUCTION

Because of recent advances in sensor and information technologies, automatic data acquisition techniques are commonly used in industrial processes, and large amounts of data and information related to quality measurement have become available. Statistical process control (SPC) to monitor and control the quality of such data-rich processes is important, yet challenging. In many situations, the quality of such a process may be better characterized and summarized by the relationship between the response variable and one or more explanatory variables; that is, the focus should be on monitoring the profile that represents such a relationship, instead of on monitoring a single quality characteristic. An extensive discussion of research problems on this topic has been given by Woodall, Spitzner, Montgomery, and Gupta (2004).

Studies focusing on simple linear profiles have been particularly popular (see, e.g., Mestek, Pavlik, and Suchanek 1994; Stover and Brill 1998; Kang and Albin 2000; Kim, Mahmoud, and Woodall 2003; Mahmoud and Woodall 2004; Gupta, Montgomery, and Woodall 2006; Zou, Zhang, and Wang 2006; Mahmoud, Parker, Woodall, and Hawkins 2007). Zou, Tsung, and Wang (2007) recently extended the focus from a simple linear profile to a general linear profile that included both a polynomial regression and a multiple linear regression relationship by using a multivariate exponentially weighted moving average (MEWMA) scheme (Lowry, Woodall, Champ, and Rigdon

1992) for the transformations of estimated parameters. A recent review of the literature has been given by Woodall (2007).

In practice, however, there are many situations in which the profile cannot be represented well by a linear model. For example, the vertical-density profile (which apparently is nonlinear) has been analyzed and discussed by Walker and Wright (2002) and Woodall et al. (2004). Williams, Woodall, and Birch (2007) gave three general approaches to the formulation of  $T^2$  statistics based on nonlinear model estimation in the Phase I analysis. Colosimo and Pacella (2007) proposed methods for monitoring dimensional requirements on manufactured items, with a focus on monitoring roundness. Williams, Birch, Woodall, and Ferry (2007) used the nonlinear regression approach of Williams, Woodall, and Birch (2007) to monitor dose-response profiles used in high-throughput screening, using illustrative data provided by DuPont. A three-parameter logistic regression model was used to represent the profiles. A stamping tonnage profile is also a typical nonlinear profile that has been studied by Jin and Shi (1999) using dimension-reduction techniques. Lada, Lu, and Wilson (2002) and Ding, Zeng, and Zhou (2006) investigated a general category of nonlinear profiles by applying

dimension-reduction techniques, including wavelet and independent component analysis. How to apply SPC in monitoring and diagnosing such general profiles, including nonlinear profiles, remains a challenge, however.

In this article we focus on a study of the Phase II method for monitoring a general profile that can be well represented by a regression function, including a nonlinear regression model. To be specific, assume that for the  $j$ th random sample collected over time, we have the observations  $(\mathbf{X}_j, \mathbf{Y}_j)$ , where  $\mathbf{Y}_j$  is an  $n_j$ -variate response vector and  $\mathbf{X}_j$  is an  $n_j \times p$  regressor matrix. It is assumed that when the process is in statistical control, the underlying model is

$$y_{ij} = g(\mathbf{X}_j^{(i)}, \boldsymbol{\beta}) + \varepsilon_{ij}, \quad i = 1, \dots, n_j, j = 1, 2, \dots, \quad (1)$$

where  $\mathbf{X}_j^{(i)}$  denotes the  $i$ th row of  $\mathbf{X}_j$ ,  $\boldsymbol{\beta}$  is a  $q$ -dimensional parameter vector, and the  $\varepsilon_{ij}$ 's are iid normal random variables with mean 0 and variance  $\sigma_j^2$ . The function  $g$  can be either linear or nonlinear, and it has some degree of smoothness, such as continuity or the existence of derivatives.

In a nonlinear profile case, the parametric control schemes, such as the multicharts of Kim et al. (2003) and the MEWMA scheme of Zou et al. (2007), still may be applied by obtaining the least squares estimators of the parameters of each profile. This is usually accomplished by using the Gauss–Newton procedure and iterating until convergence (see Gallant 1987). Note that unlike linear regression, the exact small-sample distribution of parameter estimators in nonlinear regression is unobtainable. That said, these parameter-based charts still may be used, because the estimated parameters usually are asymptotically normally distributed (Seber and Wild 1989). Nevertheless, our limited simulations indicate that such a “rough” control scheme not only will deteriorate the properties of in-control (IC) average run length (ARL), but also will have a significant effect on the out-of-control (OC) performance. Moreover, when the process is OC, nonconvergence or slow convergence often occurs, resulting in extensive computational efforts or even a failure in monitoring and detection. This makes the practical implementation of parametric methods infeasible and inconvenient.

Besides the foregoing problems of nonlinear profiles, the parametric methods for monitoring model (1) with both linear and nonlinear regressions have an inherent problem; that is, they are based on the assumption that the IC and OC model have the same form, but only the parameters may be different. Although the IC model can be determined before online monitoring begins, the OC model often cannot be specified easily, especially when the IC curves have complicated forms, such as various nonlinear models. The parametric monitoring methods are generally powerful when matched with the specific OC model for which they were designed, but they can have very poor ARL performance with other types of OC models. This is related to the lack-of-fit tests in the statistical regression context. Some classical examples have been given by Hart (1997, chap. 5). In the literature, the fundamental method for testing the lack of fit that is free of any specific alternative model is based on the nonparametric regression approach (see, e.g., Azzalini and Bowman 1993; Hardle and Mammen 1993; Fan and Huang 2001; Horowitz and Spokoiny 2001; Hart 1997 for overviews and references). These works have motivated us to

tackle the smooth profile monitoring problem with the nonparametric regression approach. Williams, Birch, Woodall, and Ferry (2007) presented parametric nonlinear profile monitoring where replicates are observed, allowing the user to identify not only changes in the parameter values of the proposed nonlinear model, but also the lack of fit of the proposed parametric model.

The present work has two objectives: to deal with the nonlinear profile monitoring problem and to resolve the latent issue of popular parametric monitoring methods that are unable to detect certain types of changes due to misspecified OC models. Although there have been many collective efforts on hypothesis testing in nonparametric regression problems, online sequential detection of OC conditions in the general profile model (1) has several unique challenges, including (a) how to integrate an appropriate regression function nonparametric test with classical SPC techniques, such as EWMA and cumulative sum (CUSUM) procedures; (b) how to monitor an increase or decrease in the variation of the general profile, because the variance in the profile is also an important quality characteristic; (c) how to design a scheme that can be implemented relatively easily and conveniently; and (d) how to establish a systematic diagnostic approach when the control scheme triggers a change in the general profile. These remain challenging problems, and we address them in the remainder of the article.

In the next section we present our proposed control chart, which integrates the MEWMA procedure with the generalized likelihood ratio (GLR) test of Fan, Zhang, and Zhang (2001). We give both a diagnostic approach and a design guideline. In Section 3 we compare the performance of our proposed scheme with other methods through simulations. The results demonstrate the robustness and effectiveness of our proposed approaches. In Section 4 we use an industrial example from semiconductor manufacturing to illustrate the step-by-step implementation of the proposed approach. We conclude in Section 5 by summarizing its contributions and suggesting some future research issues. We detail several necessary derivations and proofs in the Appendix.

## 2. METHODOLOGY

### 2.1 Model and Assumption

To facilitate the presentation, we choose to use the one-dimensional covariant case of model (1); that is, the underlying model is

$$y_{ij} = g(x_{ij}) + \varepsilon_{ij}, \quad i = 1, \dots, n_j, j = 1, 2, \dots, \quad (2)$$

where  $x_{ij}$  denotes the value of the regressor for the  $i$ th observation in the  $j$ th profile. Note that in model (2) we suppress the parameter  $\boldsymbol{\beta}$ , because we focus on the nonparametric regression method. The  $n_j$ 's are taken to be equal (denoted as  $n$ ), and the explanatory variable,  $x_{1j}, x_{2j}, \dots, x_{nj}$ , is assumed to be fixed for different  $j$ 's (denoted as  $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$ ). In the nonparametric regression context, this case is the so-called “common fixed design.” This is usually the case in practical calibration applications in industrial manufacturing and also is consistent with the work of Kim et al. (2003) on simple linear regression, Zou et al. (2007) on general linear regression, and Williams, Woodall, and Birch (2007) on nonlinear regression.

Without loss of generality, we assume that  $x_1 \leq x_2 \leq \dots \leq x_n$  and that  $x_i$  varies in the interval  $[0, 1]$ . Otherwise, we can obtain this form through some appropriate linear transformations and permutation. This assumption facilitates the technical arguments and eases the exposition. In practice, making such a transformation in advance is not necessary.

Here we consider the Phase II case in which the IC regression function and variance, say  $g_0$  and  $\sigma_0^2$ , are assumed to be known; that is, it is assumed that the IC data set used in Phase I is sufficient to estimate the regression model well. Once the IC models are established as a baseline, in Phase II we would want to detect any change in the regression function and variance of the profile. Usually the change in the regression function indicates the existence of a special cause that leads to bias, dilatation, contraction, and skewness of the profile. An increase in the variation may be due to a coarse profile, and a decrease in the variation of the profile would indicate a possible improvement in the process (Woodall et al. 2004).

In later sections we propose using a single chart to monitor the regression function and variation simultaneously, so that the design and operation of the monitoring scheme can be simplified.

## 2.2 The Generalized Likelihood Ratio Test for Profile Data

To monitor a general profile model (2), the regression function,  $g$ , and the standard deviations,  $\sigma$ , should be controlled simultaneously. Our nonparametric monitoring approach for  $g$  is based on the GLR statistics proposed by Fan et al. (2001). Suppose that  $\{y_i, x_i\}_{i=1}^n$  is a random profile sampled from model (2). Consider the simple null hypothesis testing problem

$$\begin{aligned} H_0 : g &= g_0, & \sigma &= \sigma_0 & \longleftrightarrow \\ H_1 : g &\neq g_0, & \sigma &= \sigma_0. \end{aligned} \tag{3}$$

According to the normality assumption of the error, the logarithm of the likelihood function is given by

$$-n \ln(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - g(x_i))^2. \tag{4}$$

Then the maximum likelihood under  $H_0$  can be expressed by

$$l_0 = -n \ln(\sqrt{2\pi}\sigma_0) - \frac{1}{2\sigma_0^2} \sum_{i=1}^n (y_i - g_0(x_i))^2.$$

Fan et al. (2001) proposed replacing the unknown function  $g$  under  $H_1$  with a reasonable nonparametric estimator that leads to the logarithm of likelihood function  $H_1$ ,

$$l_1 = -n \ln(\sqrt{2\pi}\sigma_0) - \frac{1}{2\sigma_0^2} \sum_{i=1}^n (y_i - \widehat{g}(x_i))^2,$$

and the GLR test statistic

$$lr = -2(l_0 - l_1) = \frac{1}{\sigma_0^2} \left[ \sum_{i=1}^n (y_i - g_0(x_i))^2 - (y_i - \widehat{g}(x_i))^2 \right]. \tag{5}$$

Large values of  $lr$  lead to a rejection of the null hypothesis.

Fan et al. (2001) used the local linear smoother (see Fan and Gijbels 1996) to estimate,  $g$ . We also follow this approach. The local linear estimator for any given point,  $x$ , takes the form  $\widehat{g}(x) = \sum_{i=1}^n W_{ni}(x)y_i$ , where

$$\begin{aligned} W_{ni}(x) &= U_{ni}(x) / \sum_{j=1}^n U_{nj}(x); \\ U_{nj}(x) &= K_h(x_j - x)[m_{n2}(x) - (x_j - x)m_{n1}(x)]; \\ m_{nl}(x) &= \frac{1}{n} \sum_{j=1}^n (x_j - x)^l K_h(x_j - x), \quad l = 1, 2; \end{aligned}$$

and  $K_h(\cdot) = K(\cdot/h)/h$ , with  $K$  a symmetric probability density function and  $h$  a bandwidth. Here, to ease the exposition, we define an  $n \times n$  smoothing matrix,  $\mathbf{W}$ , as

$$\mathbf{W} = (\mathbf{W}_n(x_1), \mathbf{W}_n(x_2), \dots, \mathbf{W}_n(x_n))^T,$$

where  $\mathbf{W}_n(x_i) = (W_{n1}(x_i), W_{n2}(x_i), \dots, W_{nn}(x_i))^T$ . Then, the  $lr$  in (5) can be rewritten in vector matrix notation as

$$lr = \frac{1}{\sigma_0^2} [(\mathbf{Y} - \mathbf{G}_0)^\otimes - (\mathbf{Y} - \mathbf{WY})^\otimes],$$

where  $\mathbf{G}_0 = (g_0(x_1), g_0(x_2), \dots, g_0(x_n))^T$ ,  $\mathbf{Y} = (y_1, y_2, \dots, y_n)^T$ , and  $\mathbf{A}^\otimes$  means  $\mathbf{A}^T \mathbf{A}$ .

Fan et al. (2001) revealed the following Wilks phenomenon: Under some regular conditions, the asymptotic null distributions of (5) are independent of nuisance functions and approximately follow a scaled chi-squared distribution. Thus, we now can develop a control chart based on the  $lr$  statistics. But the dependence of the small-sample distribution of  $lr$  on  $g_0$  makes constructing the control chart for monitoring the general profile difficult and infeasible. To overcome this difficulty, we consider transforming each profile data set,  $\{y_i, x_i\}_{i=1}^n$ , to  $\{y_i - g_0(x_i), x_i\}_{i=1}^n$  first, because the function,  $g_0$ , is known in advance. As pointed out by Fan et al. (2001), now the testing problem (3) is equivalent to the problem  $H_0 : g = 0, \sigma = \sigma_0 \longleftrightarrow H_1 : g \neq 0, \sigma = \sigma_0$ . Write  $z_i = (y_i - g_0(x_i))/\sigma_0$  and  $\mathbf{Z} = (z_1, z_2, \dots, z_n)^T$ . Then the GLR statistics will be

$$lr_z = \mathbf{Z}^\otimes - (\mathbf{Z} - \mathbf{WZ})^\otimes = \mathbf{Z}^T \mathbf{VZ}^T, \tag{6}$$

where  $\mathbf{V} = \mathbf{W}^T + \mathbf{W} - \mathbf{W}^\otimes$ . The  $lr_z$  is something like the pseudolikelihood ratio test statistics based on residuals of Az-Zalini and Bowman (1993) that focus on checking the linearity of a regression relationship, although the parameters under  $H_0$  were unknown in that work. Proposition 1, which is a direct application of theorems 5 and 8 of Fan et al. (2001), shows the good properties of the GLR test statistics,  $lr_z$ .

*Proposition 1.* Suppose that the conditions presented in the Appendix A hold. Then the following hold:

a. Under  $H_0$ ,  $lr_z \xrightarrow{L} N(\mu_z, \sigma_z^2)$ , where

$$\mu_z = \frac{2}{h} \left( K(0) - \frac{1}{2} \int K^2(t) dt \right),$$

$$\sigma_z^2 = \frac{8}{h} \int \left( K(t) - \frac{1}{2} K * K(t) \right)^2 dt.$$

b. The GLR test has nontrivial power against contiguous alternative  $g(u) - g_0(u) = \Delta_n(u) \propto n^{-4/9}$ , for which  $\Delta_n(u)$  has a continuous second derivative.

Because the GLR test is indeed general, simple, and powerful for nonparametric testing problems based on the regression function estimation, next we integrate it with the MEWMA procedure for monitoring the general profile.

### 2.3 The Control Scheme for Monitoring a General Profile

With the  $j$ th random profile collected over time, we have the observation set  $\{x_i, y_i\}_{i=1}^n$ . Using the notation presented in the previous section, we write

$$\mathbf{Y}_j = (y_{j1}, y_{j2}, \dots, y_{jn})^T$$

and

$$\mathbf{Z}_j = (\mathbf{Y}_j - \mathbf{G}_0) / \sigma_0.$$

Because both the regression relationship and variance must be monitored, a nonparametric test for the  $\sigma_j$  also is needed. Note that under the conditions (A)–(D) in Appendix A, we have the following well-known fact (Fan et al. 2001; Hall and Marron 1990):

$$\widehat{\sigma}_j^2 = \frac{1}{n}(\mathbf{Z} - \mathbf{WZ})^\otimes = \sigma_j^2 + O_p(n^{-1/2}) + O_p((nh)^{-1}), \quad (7)$$

regardless of whether or not the regression function has changed. Thus we can use the statistic  $(\mathbf{Z} - \mathbf{WZ})^\otimes$  to conduct an appropriate test for a possible change in variance, in which a very large (small) value indicates an increase (decrease) in the variance of the process.

In adopting the GLR test and the foregoing test for variance for charting, we consider developing an EWMA-type chart, which is known to be more sensitive to small and moderate shifts compared with a Shewhart-type chart. First, we transform the  $\widehat{\sigma}_j^2$  to a normal random variable, that is,

$$\tilde{\sigma}_j = \Phi^{-1}\{\psi(n\widehat{\sigma}_j^2; \mathbf{I} - \mathbf{V})\},$$

where  $\Phi^{-1}(\cdot)$  is the inverse of the standard normal cumulative distribution function (cdf) and  $\psi(\cdot; \mathbf{A})$  is the cdf of random variable  $n\widehat{\sigma}_j^2$  when the process is IC. Note that this type of variance transformation with the use of an EWMA chart has been suggested by Quesenberry (1995) and Chen, Cheng, and Xie (2001). The advantage of this transformation is that the distribution of  $\tilde{\sigma}_j$  is symmetric, so that the EWMA control chart not only is relatively easy to construct, but also is sensitive to decreases in variance (see also the discussion in Zou et al. 2007).

Before proceeding, we need to determine the  $\psi(\cdot; \mathbf{A})$  cdf. Due to the fact that  $\mathbf{Z}_j$  is an  $n$ -variate standard multivariate normally distributed random vector,  $n\widehat{\sigma}_j^2 = (\mathbf{Z} - \mathbf{WZ})^\otimes = \mathbf{Z}^T(\mathbf{I} - \mathbf{V})\mathbf{Z}$  is a quadratic form of the normal random vector. Its distribution can be shown to be equivalent to the distribution of the linear combination of independent  $\chi_1^2$ -variates with coefficients given by the eigenvalues of  $(\mathbf{I} - \mathbf{V})$  (see Box 1954). Various algorithms exist for computing the distribution of such a linear combination (see, e.g., Imhof 1961); however, for our online detection purposes, the exact evaluation of this distribution is unnecessary and computationally burdensome. Here we use the

method that matches the first three moments of the distribution of  $n\widehat{\sigma}_j^2$  with those of a chi-squared distribution (see Johnson 1959; Imhof 1961), which has been shown to be quite accurate for approximating the distribution of the quadratic forms (see Imhof 1961; Azzalini and Bowman 1993; Young and Bowman 1995). Our numerical simulation also shows that this method not only eases the computation, but also is quite accurate and effective for our monitoring problem. The details of the method are given in Appendix B.

Next we denote  $\mathbf{U}_j$  as  $(\mathbf{Z}_j^T, \tilde{\sigma}_j)^T$ , which is an  $(n + 1)$ -variate random vector, and  $\Sigma = \begin{pmatrix} \mathbf{V} & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix}$ , which is a  $(n + 1)$ -dimensional symmetric matrix. The proposed EWMA charting statistic then can be defined as

$$\mathbf{E}_j = \lambda \mathbf{U}_j + (1 - \lambda)\mathbf{E}_{j-1}, \quad j = 1, 2, \dots, \quad (8)$$

where  $\mathbf{E}_0$  is a  $(n + 1)$ -dimensional starting vector and  $\lambda$  is a weight,  $(0 < \lambda \leq 1)$ , that regulates the magnitude of the smoothing. The chart signals if

$$Q_j = \mathbf{E}_j' \Sigma \mathbf{E}_j > L \frac{\lambda}{2 - \lambda}, \quad (9)$$

where  $L > 0$  is chosen to achieve a specified IC ARL. Herein after we call this nonparametric control scheme a NEWMA chart for brevity. The form of the NEWMA chart is analogous to the MEWMA charts used by Zou et al. (2007) for parametric monitoring of a linear profile; however, different issues arise in the present context where nonparametric regression is used.

One issue is that  $\mathbf{Z}_j$  and  $\tilde{\sigma}_j$  usually are dependent, but we have not considered their correlation in the matrix  $\Sigma$ . In fact, putting  $\mathbf{Z}_j$  and  $\tilde{\sigma}_j$  together makes the resulting  $\mathbf{U}_j$  an obscure joint distribution with normal marginal distributions. We may expect the proposed schemes to remain effective, however, because ignorance of the correlation of  $\mathbf{Z}_j$  and  $\tilde{\sigma}_j$  does not affect the testing of the regression relationship and the variance to any important extent, in agreement with the simulation results in the next section.

*Remark 1.* Here we discuss two possible extensions of the NEWMA chart to demonstrate its versatility. In practical applications, engineers may use the Phase I IC samples directly rather than fitting a linear or nonlinear regression model before starting the Phase II monitoring. In such a case, the local linear estimator of  $g_0$  based on the IC profile samples, say  $\hat{g}_0$ , may be used as a replacement for  $g_0$  in the NEWMA chart. There is no need to have equal design points for each IC profile sample, but the number and positions of the design points should be properly chosen to describe the regression function well. Moreover, the number of IC samples should be sufficiently large to result in a sufficient signal-to-noise ratio so that the properties of the IC ARL can be obtained. Note that a larger number of observations in each profile can be collected for Phase I analysis than for Phase II monitoring; that is, more measurement effort may be expended on the Phase I analysis, so that the underlying regression model can be estimated accurately and, correspondingly, the desired IC run-length behavior can be achieved, whereas a large  $n$  may not be necessary in Phase II because the OC condition usually can be captured effectively using relatively small or moderate  $n$  (see the simulation results in Sec. 3 and the example in Sec. 4). Based on extensive simulations and

theoretical properties of the local linear smoother (Fan 1993) in the Phase I analysis, we recommend using at least 40 IC profile samples, and the number of observations in each profile should not be fewer than 50. Of course, this is a general guide, and engineers need to take the engineering knowledge about a specific profile, such as the smoothness and variation of the profile curve, into consideration in practical applications. Determination of Phase I sample sizes is a topic where further research would be useful.

Another natural extension of NEWMA charts is in dealing with profiles with multidimensional regressors. For this case, a multivariate local linear regression estimator may be used in place of the univariate one, and the NEWMA chart still can be used. Technical details of the multivariate local linear regression estimator have been given by, for example, Ruppert and Wand (1994) and Fan and Gijbels (1996, sec. 7.8). In this article we make no attempt to evaluate the performance of these two extensions, but we believe that such evaluation certainly merits future research.

*Remark 2.* Note that we assume that the error variables,  $\varepsilon_{ij}$ , come from normal distributions. In fact, we use this assumption only to motivate our proposed method (the derivation of GLR); it is not necessary in asymptotic theory (Fan et al. 2001) and practical use. In fact, the error distribution needs to satisfy only  $E[\varepsilon_{ij}] = 0$  and  $E(|\varepsilon_{ij}|^4) < \infty$ . But without knowing the error distribution before monitoring, how we accurately approximate the  $\psi(\cdot; \mathbf{A})$  function and how we determine the control limit remain challenges. We do not address this issue further in this article, but believe that it certainly merits future research.

## 2.4 Guidelines for Design and Implementation

This section provides guidelines on how to design and implement the proposed scheme. Several practical issues are discussed, including the choices of the kernel function,  $K(\cdot)$ ; the bandwidth,  $h$ ; and the smoothing weight,  $\lambda$ ; and determination of the control limits,  $L$ .

*On Choosing the Smoothing Weight,  $\lambda$ .* First, the smoothing weight,  $\lambda$ , in (8) is taken to be .2 in our numerical study, consistent with the studies of Kim et al. (2003) and Zou et al. (2007). In general, a smaller  $\lambda$  leads to more rapid detection of smaller shifts (Lucas and Saccucci 1990). The starting vector,  $\mathbf{E}_0$ , is chosen to be the zero vector. In fact, the fast initial response (Lucas and Saccucci 1990) also can be extended to the NEWMA chart.

*On Choosing the Sample Size and Regressor Positions.* For the NEWMA chart to perform well, a relatively large sample size,  $n$ , of the profile is needed, especially with complicated profiles. This is because, due to its flexibility, the nonparametric smoother absorbs considerably more degrees of freedom compared with parametric approaches. But, this has become a less significant limitation with advances in electronic, sensor, and information technologies. New instruments can capture more information, making a large amount of data available at one time. In addition, we recommend that engineers carefully choose the regressors,  $x_i$ 's (i.e., the design point positions), in terms of the IC profile model. Because the shifts often tend to appear but are difficult to detect in the regions of the profile with more curvature, we suggest that a good design should

have a higher concentration of points at the  $x$ 's where  $g_0(x)$  has a sharp peak. This is borne out by the optimal design theory of Muller (1984).

*On Choosing the Kernel Function,  $K(\cdot)$ .* As introduced in Section 2.2, a local linear smoothing technique is used in this article. As noted by Fan et al. (2001), construction of the GLR test does not depend on the special structures of the smoothing procedure, although proofs of the theorems may need minor modification. For the kernel function, many kernels are available to meet the requirements, such as uniform, Epanechnikov, quadratic, and Gaussian kernels. We observe that the performance of the NEWMA chart is mostly unaffected by the choice of the kernel according to our simulations. For simplicity, we use the Epanechnikov kernel,

$$K_E(u) = \frac{3}{4}(1 - u^2)I(|u| \leq 1),$$

in our simulations in Section 3.

*On Choosing the Bandwidth,  $h$ .* In the context of nonparametric regression estimation, the optimal bandwidth,  $h$ , is usually determined by minimizing the asymptotic mean squared error of the estimator. Frequently used bandwidth selection techniques are data-driven methods, such as least squares cross-validation and generalized cross-validation (see Fan and Gijbels 1996; Hart 1997). The data-driven bandwidth methods that are well suited for producing visually smooth estimates of the underlying curves may not be generally appropriate for our online monitoring problem, however.

If an observed profile of data indeed comes from the IC model, then the optimal bandwidth for local fitting of the  $\mathbf{Z}$  should be close to 1 (according to our assumption that  $x_i \in [0, 1]$ ), and data-driven bandwidth selectors will lead to a large bandwidth. But the distributional properties of GLR statistics rely implicitly on the assumption that  $h \rightarrow 0$ . Moreover, we usually do not have specific information about the OC model, so we cannot choose an optimal  $h$  for the OC condition before we start Phase II monitoring. Roughly speaking, the size of the optimal bandwidth would be expected to be proportional to the smoothness of the underlying function. In other words, a very smooth difference between IC and OC regression models requires larger bandwidths compared with less smooth differences, all other factors being equal.

But we expect the value of  $h$  to be less important, because the performance of the chart indeed is insignificantly affected by  $h$ . The amount of smoothing applied will affect the power of the test, but in many simulation results (not reported here, but available from the authors), we see that the observed significance changes little over a wide range of values of  $h$ . These findings also have been found in many other studies in the context of the nonparametric lack-of-fit tests, such as those by Azzalini and Bowman (1993), Dette and Neumeyer (2001), and Zhang (2003), and others. Thus we recommend using the following empirical bandwidth formula:

$$h_E = c \times \left( \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{1/2} n^{-1/5}, \quad (10)$$

where  $\bar{x} = \sum_{i=1}^n x_i$  and  $c$  is a constant. Empirically,  $c$  can be any value in the interval [1.0, 2.0]. In the next section we use

$c = 1.0, 1.5,$  and  $2.0$  to conduct our simulations and present some further discussion. Note that here  $n^{-1/5}$  is the order of the classical optimal bandwidth for curve estimation and  $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$  is a measure of the sparseness of the design points, which also is involved in an asymptotic optimal  $h$  formula in the context of nonparametric estimation (see Fan 1993 for details).

*Remark 3.* It can be verified that using (10) to determine the bandwidth guarantees that the design of the NEWMA chart is affine-invariant to covariate  $x$ ; that is, the smoothing matrix,  $\mathbf{W}$ , remains unchanged if we do the same linear transformation for all  $x_i$ 's.

*On Determining the Control Limits,  $L$ .* Finally, there is a vital issue remaining to be addressed: how to determine the control limit,  $L$ , after all of the preceding functions and parameters have been chosen. Note that Zou et al. (2007) extended the work of Runger and Prabhu (1996) and calculated the IC and OC ARL of their proposed MEWMA chart through the Markov chain model. Although the NEWMA chart and the MEWMA chart have similar forms, it does not seem possible that the method for the MEWMA chart can be generalized to evaluate the ARL of the NEWMA chart, because the  $Q_j$  process cannot be modeled as a Markov chain. But the NEWMA chart is a single two-sided chart, and thus only one control limit,  $L$ , must be determined. As mentioned in Section 2.2, the  $L$  is free of the IC model,  $g_0(x)$ . Thus, once  $n, \mathbf{X}, \lambda, K(\cdot), h,$  and the desired IC ARL are fixed, the values of  $L$  can be determined using Monte Carlo simulations by generating a random series,  $\mathbf{Z}_i$ . Although the simulations involve calculations in the quadratic form, considering that the matrix  $\mathbf{\Sigma}$  is fixed, the computational task is trivial by virtue of the massive computing and data storage capabilities of modern computers. Determining the control limit of such a nonparametric scheme is slightly more time-consuming than determining that of a parametric approach (such as the MEWMA charts in Zou et al. 2007), but much more efficiency will be gained when their assumption on OC models is violated. (To conveniently determine  $L$ , a computer program is available from the authors on request.)

### 2.5 Diagnostic Aids in Profile Monitoring

In the practice of quality control, besides quickly detecting a process change, it is also critical to determine when the process change occurs and to identify the kind of change that occurs in a profile after an OC signal is triggered. A diagnostic aid to locating the change point in the process and to isolating the type of changes in the profile will help an engineer quickly and easily identify and eliminate the root cause of the problem. In this section we discuss the diagnosis of a general profile and provide operational steps for systematic diagnosis.

*Step 1: Identify the Location of the Change Point.* This is the first and critical step in our diagnostic procedure, because we need it to separate the OC profiles from the entire observed profile samples so that we can make an accurate inference about this type of change. Here we use the generalized maximum likelihood estimator of the change point statistic. We assume that an OC signal is triggered at the profile,  $k$ , by the NEWMA chart.

Our suggested estimator of the change point,  $\tau$ , of the step shift is given by

$$\hat{\tau} = \arg \max_{0 \leq t < k} \{lr(tn, kn)\}, \tag{11}$$

where  $lr(tn, kn)$  is the GLR statistic. The expressions of  $lr(tn, kn)$  and the involved deductions are given in Appendix C. The foregoing formulation parallels that of Zou et al. (2007), which is used to estimate the change point after the MEWMA chart gives a signal. The difference is that nonparametric smoothing is applied to maximally select the GLR in the case of the NEWMA chart, whereas a parametric maximum likelihood estimator was used by Zou et al. (2007). Next we present an asymptotic result on the consistency of the change point estimator (11), which ensures that the proposed estimator is asymptotically effective.

*Proposition 2.* Suppose there are  $k$  profile samples to be monitored and that the first  $\tau$  and the latter  $k - \tau$  samples are IC and OC. Assume that  $0 < \lim_{k \rightarrow \infty} \tau/k = \theta < 1$ . Suppose that conditions (A)–(D) in Appendix A hold and that  $nh^5 = O(1)$ . Consider the following two types of OC models:

- a.  $g(u) = g_0(u) + \Delta_n(u)$ , where  $\Delta_n(u)$  has a continuous second derivative and its rate satisfies  $nh \int_0^1 \Delta_n^2(u) f(u) du \rightarrow \infty$  as  $n \rightarrow \infty$
- b.  $\sigma = \delta\sigma_0$ , where the size of shift  $\delta$  satisfies  $n^{1/2} \times |\delta - 1| \rightarrow \infty$  as  $n \rightarrow \infty$ .

Then, under a or b, as  $k \rightarrow \infty$ , we have  $|\hat{\tau} - \tau| = O_p(1)$ .

Although this proposition is based on the fixed sample analysis and requires the assumption that the number of OC profiles goes to infinity (which is not quite reasonable in online monitoring), it indeed provides some theoretical support for using the estimator (11). In addition, note that the validity of this proposition also requires that  $n$  go to infinity. In practical applications, for any finite  $n$  (not too small, say larger than 20), the estimator (11) generally would be quite accurate if  $nh \int_0^1 \Delta_n^2(u) f(u) du$  or  $n^{1/2} |\delta - 1|$  were sufficiently large. Pignatiello and Samuel (2001) showed the estimator's good performance in conventional nonprofile monitoring, and our simulations for general profile monitoring also demonstrate its effectiveness.

*Step 2: Determine Whether the Variance Is Stable.* After obtaining the estimate of the change point, we first suggest testing whether the variance of the profile has changed based on  $(k - \hat{\tau})$  OC samples. This is because once the variation of the process has shifted, we usually cannot make an accurate inference on the regression relationship. By using an extension of the variance estimator introduced in Section 2.3 to the multiple-sample case, we may apply the following test statistic and acceptance region:

$$\begin{aligned} \psi_1^{-1} \left( \frac{\alpha}{2}; \mathbf{I} - \mathbf{V}, k - \hat{\tau} \right) &< \sum_{j=\hat{\tau}+1}^k \mathbf{Z}_j^T (\mathbf{I} - \mathbf{V}) \mathbf{Z}_j \\ &< \psi_1^{-1} \left( 1 - \frac{\alpha}{2}; \mathbf{I} - \mathbf{V}, k - \hat{\tau} \right), \end{aligned} \tag{12}$$

where  $\psi_1(\cdot; \mathbf{A}, l)$  is the cdf of the sum of  $l$  independent random variables  $\mathbf{Z}^T \mathbf{A} \mathbf{Z}$  and, correspondingly,  $\psi_1^{-1}(\alpha; \mathbf{A}, l)$  is the  $\alpha$  percentile of the  $\psi_1(\cdot; \mathbf{A}, l)$  distribution.  $\psi_1(\cdot; \mathbf{A}, l)$  can be

easily approximated by  $\psi(\cdot; \mathbf{A})$  and by using the additive property of chi-squared random variables. If the null hypothesis is rejected, then an increase in the variation corresponds to greater inaccuracies in the manufacturing process or increased measurement error, and a decrease in the variation about the curve indicates an improvement in the process.

*Step 3: Determine Whether the Regression Function Has Changed.* If the stability of the variance is ensured in step 2, we then need to determine whether the regression relationship in the profile has changed. Again, the GLR test can be adopted, and simple derivations lead to the following test statistic:

$$\frac{1}{(k - \hat{\tau})} \left( \sum_{j=\hat{\tau}+1}^k \mathbf{Z}_j \right)^T \mathbf{V} \left( \sum_{j=\hat{\tau}+1}^k \mathbf{Z}_j \right) > \psi^{-1}(1 - \alpha; \mathbf{V}), \quad (13)$$

where  $\psi^{-1}(\alpha; \mathbf{A})$  is the  $\alpha$  percentile of the distribution of quadratic form  $\mathbf{Z}^T \mathbf{A} \mathbf{Z}$ . If we accept the null hypothesis that there is no change in the regression function, we may conclude that a false alarm is triggered by the NEWMA chart. Otherwise, we can proceed to the next step.

*Step 4: Perform Further Diagnosis by Graphical Analysis.* At this point, the engineer may care about the following problems: which part of the regression curve changes; what is the OC regression function, and how great is the difference between the IC and OC models. It is obvious that using the traditional hypothesis testing method to get the answers seems inappropriate and infeasible. However, some simple methods, such as graphical analysis, may shed light on these problems. We suggest plotting the nonparametric smoothing curve of the average of  $(k - \hat{\tau})$  sample profiles, say  $\frac{1}{k - \hat{\tau}} \sum_{j=\hat{\tau}+1}^k \mathbf{W} \mathbf{Y}_j$ , and the IC profile model together.

Based on this, a priori engineering knowledge and experience then can be fused to offer a visual and practical interpretation of the foregoing problems. Such graphical analysis has been commonly adopted in various industrial applications with the emergence of massive computing and data storage capabilities. We illustrate this with a semiconductor manufacturing example in Section 4.

*Remark 4.* It is worth noting that the preceding test statistics in steps 2 and 3 do not really follow the given null distributions when the process is IC, because these statistics are obtained under the condition that the proposed NEWMA chart has triggered a signal and change point estimate has been obtained. This fact may not fundamentally change anything we do in the diagnostic analysis, but it perhaps should affect interpretation of the type I error. Moreover, the effectiveness of such a testing method depends on the accuracy of the estimate of the change point. Moreover, controlling the overall type I error seems intractable, because of the unknown dependency between the test statistics. Thus, although this testing diagnostic method indeed can work well in most cases, engineers still may need to engage their technical/engineering knowledge about the profile after obtaining guidance from the statistical diagnostic results.

*Remark 5.* In some cases, the difference between IC and OC models lies in the variation of the values of the parameters, not in the variation in the form of the model itself. These cases occur in some simple linear models as well, such as the simplest

linear regression (Kim et al. 2003) and the quadratic polynomial regression (Zou et al. 2007). For diagnosis in such cases, we may just add one more step after step 3 to construct a model-checking test; that is, we use the function with estimated parameters to replace the  $g_0(x)$  in  $\mathbf{Z}$  and still apply the GLR test (see Fan et al. 2001 for more details). If the null hypothesis is accepted (i.e., no variation in the form of the model), then the parametric tests of Zou et al. (2007, sec. 7.2) can be adopted to identify which parameter has changed. The performance of this method in profile diagnostics is beyond the scope of this article but should be a subject of future research.

### 3. CHARTING PERFORMANCE COMPARISONS

In this section we investigate the monitoring performance of the proposed NEWMA scheme through ARL comparisons. Although our proposed control chart can be used to monitor the general profile model (2), the literature seems to contain no other effective and comparable methods for such a model. Note that Williams, Woodall, and Birch (2007) proposed smoothing the individual profile at first and then using various metric-based statistics for measuring how “different” each individual profile is from a baseline profile. But their method focuses on the Phase I analysis and may not be appropriate for our comparisons. To evaluate the performance of the proposed NEWMA chart, we still follow two related approaches for comparisons. The first approach that comes to mind is a “naive” approach that treats  $\mathbf{Z}_j$  as a long multivariate vector and uses the MEWMA scheme to construct control charts. Despite its simplicity and convenience, this approach has two serious drawbacks, however: It completely ignores the continuity of the responses at neighboring covariate points, whereas too many dimensions ( $n$ ) may accumulate large stochastic noise and thus decrease the efficiency. The second approach is the parametric control scheme proposed by Zou et al. (2007), which obtains the estimates of parameters for each profile and then applies the MEWMA scheme for the transformation of the estimated parameters. For brevity, we designate these two schemes NM charts and PM charts.

For the sake of simplicity and consistency with the literature, the change point is assumed to be  $\tau = 0$ , and only the case of overall IC ARL = 200 is considered. Moreover, we restrict our study to the following equally spaced design points:

$$x_i = \frac{i - .5}{n}, \quad i = 1, \dots, n.$$

For NEWMA charts, the bandwidth,  $h$ , is chosen from (10), where  $c = 1.0, 1.5, \text{ and } 2.0$  are considered. All the ARL results in this section are evaluated with 10,000 simulations.

To illustrate the effectiveness of our proposed monitoring scheme, our simulations cover two IC models: a quadratic model and a nonlinear model. Note that the quadratic regression model falls within the domain of linear statistical models. For each model,  $n = 20$  and  $40$  are considered. The number and variety of OC models and combinations of parameters are too great to allow a comprehensive, all-encompassing comparison. Because our goal is to demonstrate the effectiveness, robustness, and sensitivity of the NEWMA chart, we choose only certain representative OC models for illustration.

Table 1. Parameters of three OC models for Scenario 1

	Model (I)				Model (II)					Model (III)		
	$\beta_0$	$\beta_1$	$\beta_2$	$\sigma$	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\sigma$	$\beta_1$	$\beta_2$	$\sigma$
(i)	1.0	2.0	3.1	1.0	.8	4.4	-3.0	4.0	1.0	.1	1.0	1.0
(ii)	1.0	2.1	3.1	1.0	.8	4.4	-3.0	4.1	1.0	.2	1.0	1.0
(iii)	1.1	2.1	3.1	1.0	.8	4.4	-3.2	4.1	1.0	.2	.8	1.0
(iv)	1.0	2.0	3.0	1.1	1.0	4.4	-3.2	4.1	1.0	.2	1.3	1.0
(v)	1.0	2.0	3.0	.7	.8	4.4	-3.0	4.0	1.1	.3	1.5	1.0
(vi)	1.1	2.1	3.1	1.1	.8	4.5	-3.0	4.0	1.1	.3	1.5	1.1

Scenario 1 (Quadratic IC model).

$$y_{ij} = 1 + 2x_i + 3x_i^2 + \epsilon_{ij}, \quad i = 1, \dots, n,$$

where  $\sigma_0^2 = 1$ . Three OC models are chosen, as follows:

- (I):  $y_{ij} = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_{ij}$
- (II):  $y_{ij} = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \epsilon_{ij}$
- (III):  $y_{ij} = 1 + 2x_i + 3x_i^2 + \beta_1 \sin(2\pi \beta_2 x_i) + \epsilon_{ij}$ .

These three OC models correspond to the following three cases: (a) The structure of the regression relationship does not change, and only the parameters shift; (b) The regression relationship changes to another linear model; and (c) The regression relationship changes to a nonlinear model. Some chosen parameters of the three OC models are listed in Table 1. Table 2 gives the OC ARLs of the NEWMA chart and those of NM and PM charts for detecting changes under OC models, and corresponding parameters are tabulated in Table 1. The control limits ( $L$ ) of these control schemes are tabulated in the last row of Table 2.

Under the OC model (I), the PM chart has superior efficiency, as we would expect, because the hypothesis is the correct one in this case (a quadratic polynomial vs. a quadratic polynomial). The NEWMA chart also offers satisfactory performance, and the difference between the NEWMA and PM charts is not significant, even when a relatively small bandwidth ( $c = 1.0$ ) is used. It is worth noting that the OC ARL is a decreasing function of the bandwidth under (I), which is consistent with our claim in Section 2.4 that more smooth statistics usually have better performance when the difference between the IC and OC models is quite smooth. Of course, the “smooth” method, whether the NEWMA chart or the PM chart, outperforms the NM chart, the one based to no smoothing in this case.

But in cases (i)–(iii) of model (II), the PM chart is dramatically outperformed by the NEWMA chart and is even less efficient than the NM chart. More seriously, in case (i) of model (II), the IC and OC ARL of the PM chart are equal, because when a quadratic is fitted to the data using least squares, the estimated coefficients each will be very close to those of the IC model. This phenomenon demonstrates the claimed disad-

Table 2. ARL comparisons of the NEWMA, NM, and PM charts for Scenario 1 (IC ARL = 200)

		$n = 20$					$n = 40$				
		NEWMA			NM	PM	NEWMA			NM	PM
		$c = 1.0$	$c = 1.5$	$c = 2.0$			$c = 1.0$	$c = 1.5$	$c = 2.0$		
(I)	(i)	150.1	144.7	140.0	171.3	137.1	120.6	115.3	110.1	162.3	103.1
	(ii)	66.6	60.8	56.7	98.2	54.2	37.9	34.0	31.8	79.4	29.2
	(iii)	20.7	18.3	17.5	33.1	16.5	11.2	10.3	9.8	22.4	9.1
	(iv)	30.5	27.4	27.0	43.7	26.5	17.5	16.2	15.5	30.5	14.8
	(v)	8.2	6.6	5.9	*	5.3	4.2	3.7	3.5	*	3.2
	(vi)	12.5	11.4	11.0	17.3	10.8	7.4	7.0	6.7	11.8	6.3
(II)	(i)	104.6	104.2	113.4	132.2	199.7	66.5	64.0	67.5	114.2	200.1
	(ii)	89.4	87.5	92.5	119.2	154.1	52.5	50.3	52.1	99.3	120.2
	(iii)	78.8	75.7	79.2	109.3	121.0	45.0	42.4	42.8	89.3	84.9
	(iv)	24.2	22.1	21.1	38.9	22.8	12.7	11.6	11.2	26.2	12.2
	(v)	24.5	23.7	22.6	35.8	25.7	13.9	13.0	12.8	24.4	14.0
	(vi)	22.4	20.6	20.3	32.3	22.5	12.6	12.0	11.6	21.7	12.6
(III)	(i)	106.8	103.1	102.3	137.2	119.6	69.9	66.7	66.2	122.0	81.3
	(ii)	37.8	35.6	34.8	60.1	47.1	20.0	18.4	18.0	43.1	25.0
	(iii)	35.8	32.2	31.3	57.7	31.5	19.0	16.9	16.5	40.6	16.3
	(iv)	37.3	36.2	38.2	58.1	51.8	19.5	18.3	18.6	41.7	27.2
	(v)	18.0	18.3	20.4	29.5	27.4	9.7	9.5	10.0	18.3	15.6
	(vi)	11.2	11.1	11.4	15.5	13.4	6.8	6.6	6.6	10.6	7.7
$L$		20.25	17.25	15.63	38.69	13.87	21.66	18.28	16.50	65.20	13.87

\*The NM chart cannot detect a decreasing shift in the variance.

Table 3. Parameters of three OC models for Scenario 2

	Model (I)			Model (II)			Model (III)		
	$\beta_1$	$\beta_2$	$\sigma$	$\beta_1$	$\beta_2$	$\sigma$	$\beta_1$	$\beta_2$	$\sigma$
(i)	1.0	1.3	1.0	.2	3.0	1.0	10.0	.1	1.0
(ii)	1.0	1.5	1.0	.3	3.0	1.0	10.0	.2	1.0
(iii)	1.1	1.0	1.0	.2	2.0	1.0	40.0	1.5	1.0
(iv)	1.3	1.0	1.0	.3	2.0	1.0	60.0	1.5	1.0
(v)	1.2	1.0	1.1	.2	4.0	1.1	50.0	1.2	1.1
(vi)	1.0	1.2	.7	.2	4.0	1.3	-3.0	4.0	1.1

vantage of parametric control charts, that is, their inefficiency when the OC model is misspecified. In fact, these two curves have very slight differences in the overall profile but have some deviation in the vicinity of the design points. The NEWMA chart can capture such details more effectively and trigger a faster alarm.

The third OC model is the IC quadratic model plus an oscillation of a sine function. The PM chart still can pick up the difference on most occasions, although the parametric hypothesis is clearly not the correct one. But it generally is outperformed by the NEWMA chart, except for case (iii), where the differences between the IC and OC model are indeed rather “flat.”

Scenario 2 (The nonlinear IC model). Next, we consider a nonlinear IC model,

$$y_{ij} = 1 - \exp(-x_i) + \epsilon_{ij}, \quad i = 1, \dots, n, j = 1, 2, \dots$$

Again, three OC models are considered:

(I):  $y_{ij} = 1 - \beta_1 \exp(-x_i^{\beta_2}) + \epsilon_{ij}$

(II):  $y_{ij} = 1 - \exp(-x_i) + \beta_1 \cos(\beta_2 \pi (x_i - .5)) + \epsilon_{ij}$

(III):  $y_{ij} = \frac{1}{1 + \beta_1 x_i^{\beta_2}} + \epsilon_{ij}$ .

The chosen parameters of the three OC models are tabulated in Table 3. As mentioned in Section 1, the PM control scheme is not appropriate for use in monitoring such a nonlinear model. Here we show only the OC ARL comparisons of the NEWMA and NM charts. Note that the control limits of NEWMA charts are equal to those in Table 2, because they do not depend on the choice of the IC model but only of  $n, h,$  and  $\mathbf{X}$ . From this table, we can see that no matter how the OC models and parameters vary, the NEWMA chart performs significantly better than the NM chart.

In general, compared with PM and NM charts, the proposed NEWMA chart has both robustness and sensitivity to the changes in regression functions and variance. Using the NEWMA chart can significantly alleviate the problem of the PM chart in misspecified OC models at the small cost of efficiency when indeed the OC model assumption is correct. Moreover, the NEWMA chart can conveniently and effectively tackle the problem of nonlinear profile monitoring.

Table 4. ARL comparisons of the NEWMA and NM charts for Scenario 2 (IC ARL = 200)

		$n = 20$				$n = 40$			
		NEWMA			NM	NEWMA			NM
		$c = 1.0$	$c = 1.5$	$c = 2.0$		$c = 1.0$	$c = 1.5$	$c = 2.0$	
(I)	(i)	144.5	139.9	134.2	168.3	110.3	103.1	100.6	155.0
	(ii)	99.1	93.0	86.2	131.2	63.8	58.7	55.1	113.3
	(iii)	113.7	109.1	102.2	143.0	79.5	72.7	67.4	129.1
	(iv)	19.9	17.8	16.6	31.8	10.6	9.8	9.4	21.0
	(v)	17.9	16.6	15.7	25.4	10.3	9.7	9.2	17.1
	(vi)	8.2	6.5	5.8	*	4.2	3.7	3.5	*
(II)	(i)	39.4	40.9	45.9	59.3	20.3	20.1	21.4	43.3
	(ii)	17.9	18.1	20.6	27.1	9.7	9.4	9.9	18.4
	(iii)	37.9	35.2	34.4	60.2	20.0	18.4	17.8	43.4
	(iv)	17.3	16.0	15.5	27.0	9.4	8.8	8.7	18.3
	(v)	17.4	17.8	19.0	24.1	9.8	9.9	10.3	16.1
	(vi)	5.4	5.0	4.9	8.5	3.5	3.3	3.3	6.0
(II)	(i)	69.2	61.9	58.4	99.4	39.1	34.9	33.2	80.1
	(ii)	40.2	36.7	33.9	64.8	21.5	19.4	18.1	46.5
	(iii)	11.8	10.9	10.6	17.9	6.9	6.4	6.2	12.1
	(iv)	15.4	14.6	13.9	23.7	8.4	8.0	7.7	15.7
	(v)	14.0	13.3	12.9	19.8	7.9	7.5	7.4	12.8
	(vi)	7.8	7.3	7.0	11.8	4.8	4.5	4.4	8.0
$L$		20.25	17.25	15.63	38.69	21.66	18.28	16.50	65.20

The choice of bandwidth has some effect on the OC ARLs, but using our recommended formula (10) can have generally satisfactory performance in all of the foregoing simulations. The different values of  $c \in [1.0, 2.0]$  offer similar ARL results and thus can be determined by engineers without too many considerations. Of course, if one has some a priori information about the OC model, then  $c$  can be chosen to make the NEWMA chart nearly optimal. A smaller  $c$  usually is more effective in detecting the sharp change in the local area and a larger  $c$  performs better when the difference between the IC and OC models is flat or smooth.

#### 4. ILLUSTRATIVE EXAMPLE: MONITORING NONLINEAR PROFILES IN A DEEP REACTIVE ION-ETCHING PROCESS

Here we apply the proposed monitoring scheme to a deep reactive ion-etching (DRIE) process in semiconductor manufacturing, which is very critical to the quality of the output wafer and requires careful control and monitoring on a run-to-run basis. In the DRIE process, the desired profile has smooth and vertical sidewalls, called the anisotropic profile. Ideally, the sidewalls of a trench are perpendicular to the bottom of the trench, with certain a degree of smoothness around the corners. Various shapes of profiles, such as positive and negative profiles, which are due to underetching and overetching, are considered unacceptable. More detailed discussion of the DRIE example has been given by, for example, Wang and Tsung (2007) and Zou et al. (2007).

In practice, monitoring the entire DRIE profile is not necessary, because the two sides of each profile are usually symmetrical. Here, similar to the approach of Zou et al. (2007), we focus on one side of the profile (e.g., the left side), and then rotate the side of the profile by 45 degrees along a reference point according to a prescribed chosen coordinate, as shown in Figure 1. We use the transformed nonlinear profile in Figure 1 to demonstrate the proposed monitoring scheme.

It can be seen that each side of the DRIE profile may be constituted by three key parts: the smooth and straight sidewall, corresponding to (i) in Figure 1; the smooth and curly corner, corresponding to (ii) in Figure 1; and the flat bottom, corresponding to (iii) in Figure 1. Here we select only the segments near the corner for our investigation, because engineers consider those segments to contain sufficient critical information to distinguish the out-of-control conditions in the process.

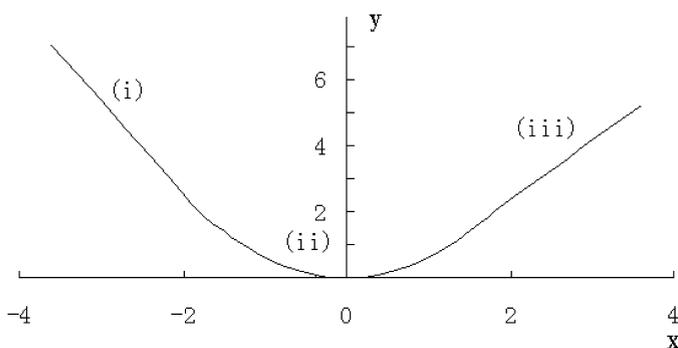


Figure 1. Modeling the DRIE profile.

Zou et al. (2007) found that the part (ii) profile may be adequately described by the following quadratic polynomial model:

$$y_{ij} = .62 \cdot x_i^2 + \varepsilon_{ij}, \quad i = 1, \dots, n, \quad (14)$$

and the data set was tabulated in the appendix of Zou et al. (2007), which contains five IC profile samples and nine OC profile samples. Here we apply our proposed NEWMA scheme to monitoring this data set. For each profile,  $x_i, i = 1, \dots, 11$ , are fixed as equally spaced values,  $-2.5, (.5), 2.5$ . Detailed implementation steps of the proposed scheme are as follows:

Step 1. Choose the kernel function  $K(\cdot)$ , bandwidth  $h$ , smoothing weight  $\lambda$ , and the desired IC ARL following the guidelines given in Section 2.4. Here we adopt the Epanechnikov kernel, and determine  $h$  by (10) with  $c = 1.5$ . According to the values of the design points,  $x_i$ 's, we can obtain  $L = 18.09$  given that  $ARL = 370$  and  $\lambda = .2$ . Consequently,  $L \frac{\lambda}{2-\lambda}$  will be 2.01. Then we can construct the NEWMA control chart as shown in Figure 2.

Step 2. Start monitoring the process, and sequentially compute the  $Z_j$  and  $E_j$  vectors. Then compute the plot statistic,  $Q_j$ , in (9) and compare it with the control limit,  $L \frac{\lambda}{2-\lambda}$ . In Figure 2 the NEWMA chart signals at the 14th sample, just as effectively as the parametric MEWMA chart of Zou et al. (2007) does.

Step 3. By computing the values of  $lr(jn, 14n)$  for  $j = 0, 1, \dots, 13$ , we can find that its maximum occurs at  $j = 6$  when  $lr(6n, 14n)$  is 18.60. In fact, the true change point is  $\tau = 5$ , which is precisely indicated by the parametric likelihood method given by Zou et al. (2007). In this case the nonparametric method provides less accurate but close diagnosis results.

Step 4. By computing the test statistics given in (12) and (13), we can obtain the corresponding  $p$  values of .209 and .002, which indicate that the variance is stable but there may be a shift in the regression function.

Step 5. Finally, we plot the local linear smoothing curve of the average of the last eight sample profiles and the IC profile model together, as shown in Figure 3. From this figure, we can see that the OC profile model presents an apparent negative trend, which may be due to overetching of the DRIE process.

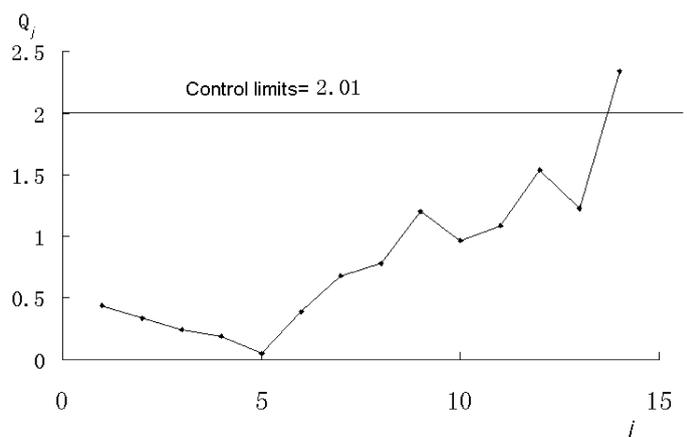


Figure 2. NEWMA chart for monitoring the example of Zou et al. (2007).

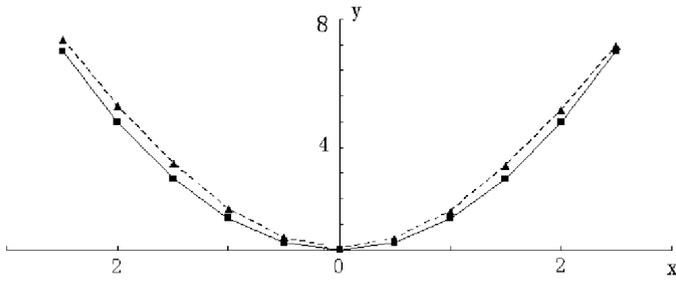


Figure 3. Graphical diagnosis of the example of Zou et al. (2007). The solid line represents the IC model; the dotted line, the estimated OC models.

Next, to demonstrate the effectiveness of our proposed approach, we consider monitoring the whole transformed DRIE profile (i.e., all three parts) shown in Figure 1. Obviously, these three parts cannot be readily viewed as a polynomial or a general linear model, and so existing parametric methods (as in Zou et al. 2007) cannot be applied directly. Thus we choose to apply the proposed NEWMA chart to monitor such a profile.

As discussed in Section 2, considering the efficiency of monitoring and diagnosing inferences in using the NEWMA chart, 35 design points,  $x_j$ 's, are chosen:  $-3.6, (.3), -1.8; -1.5, (.15), 1.5$ ; and  $1.8, (.3), 3.6$ . Note that the corner requires more design points (i.e., smaller-spaced values) than the other two parts of the profile. This setting is consistent with our previously mentioned principle on the choice of design points in Section 2.4. Here we collect dimensional readings (70 design points) by scanning electron microscopy from  $m = 18$  anisotropic profiles that are known to be IC (available from the authors on request). The local linear smoothing curve of the average of these 18 IC profiles is plotted in Figure 4. This curve can be considered an estimate ( $\hat{G}_0$ ) of the true IC profile model ( $G_0$ ), and thus we do not need to model this profile as some piecewise polynomial (or other nonlinear model) (see Remark 1). In addition, the estimate of variance is .409. From here, we can start the monitoring and diagnosing procedure.

Similar to the preceding monitoring case, here we use  $ARL = 370$ ,  $\lambda = .2$ , and  $c = 1.5$ , and thus the control limit is 2.30. For ease of illustration, we generate 6 IC profiles by adding random errors to the  $G_0$  estimated by those 18 anisotropic profiles. We also obtain three OC sample profiles, which are classified as inferior profiles based on engineering knowledge. In this example we artificially assume that we first monitor the six simulated IC samples and then obtain the three

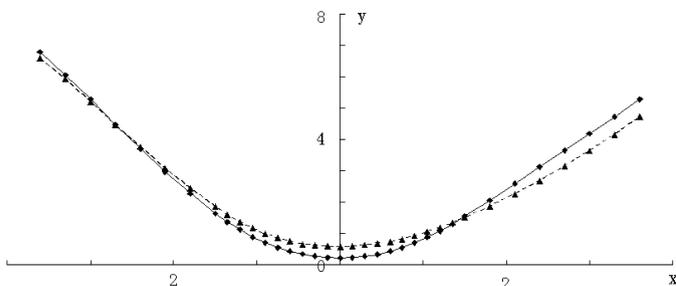


Figure 4. Graphical diagnosis of the new DRIE profile. The solid line represents the IC model; the dotted line, the estimated OC model.

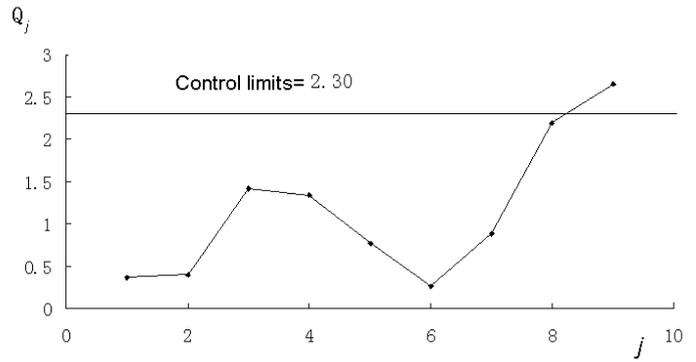


Figure 5. NEWMA chart for the new example.

OC profiles. Accordingly, we construct the NEWMA chart and plot the test statistics,  $Q_j$ , one by one, as shown in Figure 5.

We can see that the NEWMA chart signals after the third OC sample is monitored. Then, by looking at the values of  $lr(jn, 9n)$  for  $j = 0, 1, \dots, 8$ , we find that its maximum occurs at  $j = 6$  with  $lr(6n, 9n) = 41.09$ . This maximum accurately indicates the change point location,  $\tau$ , of the shift. In addition, by computing the test statistics given in (12) and (13), we can obtain the corresponding  $p$  values of .633 and  $10^{-6}$ , which indicate that the regression function has changed. Finally, the local linear smoothing curve of the average of these three sample profiles, coupled with the IC curve, are shown in Figure 4. From this figure, we can see that the OC profile model indicates an apparent positive trend possibly due to underetching of the DRIE process; thus we suggest that the ICP machine settings and the DRIE process conditions should be reexamined.

### 5. CONCLUSIONS AND CONSIDERATIONS

In this article, we have proposed a control scheme that integrates the EWMA procedure with the GLR test based on local linear regressions. It can be applied to monitoring a general profile that includes both linear and nonlinear regression models. The proposed scheme can be easily designed and constructed and has satisfactory performance. We have provided a systematic diagnostic approach to locating the change point of the process and identifying the type of changes in a profile. As demonstrated by the DRIE example, the proposed monitoring and diagnostic approach may be implemented in industrial practice as long as the quality of a process can be characterized by a general smooth regression profile.

Several issues not addressed here could be worthy topics of future research. First, note that this article and the previously published works on profile monitoring are based mainly on the assumption that the profiles are independent of one another and that the random errors associated with the measurements within a profile also are independent of one another. Engineering applications that give rise to profile data may lead to autocorrelated error terms (see Walker and Wright 2002 for discussion). A common source of autocorrelated errors is the spatial or serial manner in which the data are collected. Dependency of the observations not only affects the IC and OC properties of the NEWMA scheme, but also makes the selection of the bandwidth difficult (Opsomer, Wang, and Yang 2001). Therefore,

some necessary modifications that take the correlation structure within profiles into account are desirable. Second, as we showed in the simulations, the bandwidth,  $h$ , has some effect on the OC ARL performance, although not a large effect with our recommended choice. In the context of testing the regression model, some “data-driven” nonparametric methods have been proposed recently, including those of Horowitz and Spokoiny (2001) and Guerre and Lavergne (2005). One ongoing effort of ours is to develop a control scheme that integrates a sequential “data-driven” adaptive smoothing parameter selection method to make monitoring profiles nearly optimal. Finally, as shown by Fan et al. (2001), the GLR testing method is widely applicable in nonparametric lack-of-fit tests. The NEWMA scheme could be generalized monitor various more complex profile data models, such as the additive model, the single-index model, the semiparametric model, and the varying-coefficient model, whenever these models can closely represent industrial profiles.

ACKNOWLEDGMENTS

The authors thank the editor, associate editor, and two anonymous referees for their many helpful comments that have resulted in significant improvements in the article. This work was completed when Zou was a research assistant at the Hong Kong University of Science and Technology, whose hospitality is appreciated and acknowledged. This research was supported by RGC Competitive Earmarked Research Grant 620606, National Science Foundation of Tianjin Grant 07JCYBJC04300, and NNSF of China Grants 10771107 and 10711120448.

APPENDIX A: CONDITIONS

(A) There exists a positive density,  $f(u)$ , that is Lipschitz-continuous and bounded away from 0 such that the design point,  $x_i$ , satisfies

$$\int_0^{x_i} f(u) du = \frac{i}{n}, \quad i = 1, \dots, n.$$

(B)  $g_0(u)$  has a continuous second derivative.

(C) The function  $K(t)$  is symmetric and bounded. Furthermore, the functions  $t^3K(t)$  and  $t^3K'(t)$  are bounded, and  $\int t^4K(t) dt < \infty$ .

(D) The bandwidth  $h$  satisfies that  $h \rightarrow 0$ ,  $nh^{3/2} \rightarrow \infty$ , and  $nh^8 \rightarrow 0$ .

APPENDIX B: APPROXIMATE THE DISTRIBUTIONS OF QUADRATIC FORMS LIKE  $\mathbf{Z}^T \mathbf{A} \mathbf{Z}$

Recalling the assumption given in Section 2.2,  $\mathbf{Z}$  is a standard multivariate, normally distributed  $n$ -variate, and  $\mathbf{A}$  is a symmetric and semipositive matrix. By theorem 2.2 of Box (1954), the  $s$ th cumulant of  $\mathbf{Z}^T \mathbf{A} \mathbf{Z}$  is

$$\kappa_s = 2^{s-1}(s-1)! \text{tr}(\mathbf{A}^s).$$

This is used to match the first three moments of the distribution of  $\mathbf{Z}^T \mathbf{A} \mathbf{Z}$  with those of a chi-squared distribution; that is, given the mean, variance, and skewness of the distribution of  $\mathbf{Z}^T \mathbf{A} \mathbf{Z}$  (using the relationship between cumulant and moment), we find

$c_1, c_2$ , and  $c_3$  such that a  $c_1 \chi_{c_2}^2 + c_3$  distribution has the same mean, variance, and skewness. Simple calculations yield

$$c_1 = \sqrt{\text{tr}(\mathbf{A}^2)/\text{tr}(\mathbf{A}^3)},$$

$$c_2 = \text{tr}(\mathbf{A}^3),$$

and

$$c_3 = \text{tr}(\mathbf{A}) - \sqrt{\text{tr}(\mathbf{A}^2) \cdot \text{tr}(\mathbf{A}^3)}.$$

This  $c_1 \chi_{c_2}^2 + c_3$  distribution is then used as an approximate null distribution, to calculate a  $p$  value in the usual manner.

APPENDIX C: EXPRESSION OF  $l_r(tn, kn)$

Let  $g_i(\cdot)$  and  $\sigma_i^2$  denote the regression function and variance of the  $i$ th profile sample  $\{\mathbf{Y}_i, \mathbf{X}\}_{i=1}^n$ . Then, after  $k$  samples have been collected, we care about the following null hypothesis of no change:

$$H_0 : g_1 = g_2 = \dots = g_k = g_0$$

and

$$\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 = \sigma_0^2,$$

against the following alternative of one change:

$$H_0 : g_0 = g_1 = g_2 = \dots = g_{t^*} \neq g_{t^*+1} = \dots = g_k$$

or

$$\sigma_0^2 = \sigma_1^2 = \sigma_2^2 = \dots = \sigma_{t^*}^2 \neq \sigma_{t^*+1}^2 = \dots = \sigma_k^2,$$

where  $1 \leq t^* < k$ .

The logarithm of the likelihood function is given by

$$-\frac{1}{2} \sum_{j=1}^k \left[ n \ln(2\pi \sigma_j^2) + \frac{1}{\sigma_j^2} (\mathbf{Y}_j - \mathbf{G}_j)^\otimes \right],$$

where  $\mathbf{G}_i = (g_i(x_1), g_i(x_2), \dots, g_i(x_n))^T$ . If the data are collected under IC conditions (i.e., under the null hypothesis), then the value of the logarithm of the likelihood function is

$$l_0 = -\frac{1}{2} \sum_{j=1}^k \left[ n \ln(2\pi \sigma_0^2) + \frac{1}{\sigma_0^2} (\mathbf{Y}_j - \mathbf{G}_0)^\otimes \right].$$

Assuming that a change occurs after  $t$ , the corresponding logarithm of generalized likelihood is

$$l_1 = -\frac{1}{2} \sum_{j=1}^t \left[ n \ln(2\pi \sigma_0^2) + \frac{1}{\sigma_0^2} (\mathbf{Y}_j - \mathbf{G}_0)^\otimes \right] - \frac{(k-t)n}{2} \left[ \ln \left( \frac{2\pi}{(k-t)n} \sum_{j=t+1}^k (\mathbf{Y}_j - \widehat{\mathbf{G}})^\otimes \right) + 1 \right],$$

where  $\widehat{\mathbf{G}}$  is the local linear estimator based on the profile samples  $\{\mathbf{Y}_j, \mathbf{X}\}_{j=t+1}^k$ .

Because  $g_0$  and  $\sigma_0$  are known in advance, recalling the notation in Section 2.2, we can substitute the profile samples,  $\{\mathbf{Z}_j, \mathbf{X}\}_{j=1}^k, g_0 = 0$ , and  $\sigma_0^2 = 1$ , into the preceding expressions.

This will result in the following final expression of  $lr(tn, kn)$ :

$$\begin{aligned} lr(tn, kn) &= -2(l_0 - l_1) \\ &= \sum_{j=t+1}^k \mathbf{Z}_j^\otimes - (k-t)n \\ &\quad \times \left[ \ln \left( \frac{1}{(k-t)n} \sum_{j=t+1}^k (\mathbf{Z}_j - \mathbf{W}\bar{\mathbf{Z}}_{t,k})^\otimes \right) + 1 \right], \end{aligned} \quad (\text{A.1})$$

where  $\bar{\mathbf{Z}}_{t,k} = \frac{1}{k-t} \sum_{j=t+1}^k \mathbf{Z}_j$ . Using the classical binary segment procedure leads to the change point estimator (11).

#### APPENDIX D: PROOFS

##### Proof of Proposition 1

This proposition can be viewed as a corollary of theorems 5 and 7 of Fan et al. (2001). The conditions are the only difference. Note that the fixed design case is considered in this article, whereas Fan et al. (2001) focused on the random design. Condition (A) is imposed to replace condition (A1) of Fan et al. (2001). The technical arguments in the proof of theorem 5 of Fan et al. (2001) continue to hold.

Because  $g_0$  is known, the random part  $d_{1n}$  in theorem 5 of Fan et al. (2001) does not appear. Moreover,  $R_{n10}$ ,  $R_{n20}$ , and  $R_{n30}$  in the proof of theorem 5 of Fan et al. (2001) also equal 0, meaning that the conditions for the bandwidth  $h$  can be greatly relaxed to condition (D). Note that this condition gives a wide range [from  $O(n^{-1/3})$  to  $O(n^{-1/7})$ ] of bandwidths that include the optimal bandwidth,  $O(n^{-1/5})$  of nonparametric estimating regression function  $g(\cdot)$ .

##### Proof of Proposition 2

To prove part a, we first consider that  $1 \leq t < \tau$ . Assume that when  $k \rightarrow \infty$ ,  $l/k \rightarrow \theta_1$ , where  $\theta_1 \in (0, 1)$  and  $\theta_1 < \theta$ . By similar arguments as in the proofs of theorems 5 and 8 of Fan et al. (2001), it is easy to verify that

$$\begin{aligned} (k^{-1}h)lr_{\tau,k} &= nh(1-\theta) \\ &\quad \times \int_0^1 \Delta_n^2(u)f(u)du(1+o_p(1)) + O_p(nh^5) > 0. \end{aligned}$$

Thus it is sufficient to show that when  $n \rightarrow \infty$ ,  $k \rightarrow \infty$ , for any  $\tau < l < k$ ,

$$k^{-1}h(lr_{\tau,k} - lr_{l,k}) > 0 \quad \text{a.e.}$$

It then follows from expression (A.1) that

$$\begin{aligned} lr_{\tau,k} - lr_{l,k} &= - \sum_{j=t+1}^{\tau} \mathbf{Z}_j^\otimes + kn(1-\theta_1) \ln \\ &\quad \times \left( \frac{1}{(k-t)n} \sum_{j=t+1}^k (\mathbf{Z}_j - \mathbf{W}\bar{\mathbf{Z}}_{t,k})^\otimes \right) \end{aligned}$$

$$\begin{aligned} &- kn(1-\theta) \ln \left( \frac{1}{(k-\tau)n} \sum_{j=\tau+1}^k (\mathbf{Z}_j - \mathbf{W}\bar{\mathbf{Z}}_{\tau,k})^\otimes \right) \\ &\quad + kn(\theta - \theta_1) \\ &= \Lambda_1 + \Lambda_2 + \Lambda_3 + kn(\theta - \theta_1). \end{aligned}$$

Now we deal with  $\Lambda_2$ . Decompose  $\frac{1}{(k-t)n} \sum_{j=t+1}^k (\mathbf{Z}_j - \mathbf{W}\bar{\mathbf{Z}}_{t,k})^\otimes$  as follows:

$$\begin{aligned} &\frac{1}{(k-t)n} \sum_{j=t+1}^k (\mathbf{Z}_j - \mathbf{W}\bar{\mathbf{Z}}_{t,k})^\otimes \\ &= \frac{1}{(k-t)n} \left[ \sum_{j=t+1}^{\tau} (\mathbf{Z}_j - \mathbf{W}\bar{\mathbf{Z}}_{t,k})^\otimes + \sum_{j=\tau+1}^k (\mathbf{Z}_j - \mathbf{W}\bar{\mathbf{Z}}_{t,k})^\otimes \right] \\ &= \frac{1}{(k-t)n} \left[ \sum_{j=t+1}^{\tau} (\mathbf{Z}_j - \mathbf{W}\bar{\mathbf{Z}}_{t,\tau})^\otimes + \sum_{j=\tau+1}^k (\mathbf{Z}_j - \mathbf{W}\bar{\mathbf{Z}}_{\tau,k})^\otimes \right] \\ &\quad + \frac{1}{(k-t)n} \left[ \sum_{j=t+1}^{\tau} \left( \frac{1-\theta}{1-\theta_1} \mathbf{W}(\bar{\mathbf{Z}}_{t,\tau} - \bar{\mathbf{Z}}_{\tau,k}) \right)^\otimes \right. \\ &\quad \quad \left. + \sum_{j=\tau+1}^k \left( \frac{\theta-\theta_1}{1-\theta_1} \mathbf{W}(\bar{\mathbf{Z}}_{\tau,k} - \bar{\mathbf{Z}}_{t,\tau}) \right)^\otimes \right] \\ &\quad + 2 \frac{(1-\theta)(\theta-\theta_1)}{n(1-\theta_1)^2} \\ &\quad \times [(\bar{\mathbf{Z}}_{t,\tau} - \mathbf{W}\bar{\mathbf{Z}}_{t,\tau} + \mathbf{W}\bar{\mathbf{Z}}_{\tau,k} - \bar{\mathbf{Z}}_{\tau,k})^T \mathbf{W}(\bar{\mathbf{Z}}_{t,\tau} - \bar{\mathbf{Z}}_{\tau,k})] \\ &= \Lambda_{21} + \Lambda_{22} + \Lambda_{23}. \end{aligned}$$

Then simple calculations yield

$$\Lambda_{21} = 1 + O_p((nk)^{-1/2}) + O_p((nh)^{-1}) \quad (\text{A.2})$$

and

$$\Lambda_{22} = \frac{(1-\theta)(\theta-\theta_1)}{n(1-\theta_1)^2} (\mathbf{W}\mathbf{\Delta})^\otimes + O_p((knh)^{-1}), \quad (\text{A.3})$$

where  $\mathbf{\Delta} = (\Delta_n(x_1), \Delta_n(x_2), \dots, \Delta_n(x_n))^T$ . It follows from the Cauchy inequality and  $L_r$ -convergence properties of the local linear smoother (see, e.g., lemma 3 in Zhu and Xue 2006) that

$$\begin{aligned} \Lambda_{23} &\leq \frac{2(1-\theta)(\theta-\theta_1)}{(1-\theta_1)^2} \\ &\quad \times (n^{-1}(\mathbf{W}\mathbf{\Delta})^\otimes \cdot [O_p((knh)^{-1}) + O_p(h^4)])^{1/2}, \end{aligned}$$

which leads to  $\Lambda_{23} = o_p(\Lambda_{22})$ .

By using the fact that

$$\begin{aligned} &\frac{1}{(k-\tau)n} \sum_{j=\tau+1}^k (\mathbf{Z}_j - \mathbf{W}\bar{\mathbf{Z}}_{\tau,k})^\otimes \\ &= 1 + O_p((nk)^{-1/2}) + O_p((nh)^{-1}), \end{aligned}$$

we have

$$\begin{aligned} \Lambda_3 &= kn(1-\theta) \left( 1 - \frac{1}{(k-\tau)n} \sum_{j=\tau+1}^k (\mathbf{Z}_j - \mathbf{W}\bar{\mathbf{Z}}_{\tau,k})^\otimes \right. \\ &\quad \left. + [O_p((nk)^{-1/2}) + O_p((nh)^{-1})]^2 \right). \end{aligned}$$

Consequently, using Taylor expansion and combining equations (A.2) and (A.3), we have

$$k^{-1}(lr_{\tau,k} - lr_{l,k}) = k^{-1} \left( \Lambda_1 + \sum_{j=t+1}^{\tau} (\mathbf{Z}_j - \mathbf{W}\bar{\mathbf{Z}}_{t,\tau})^{\otimes} \right) + \frac{(1-\theta)(\theta-\theta_1)}{(1-\theta_1)} (\mathbf{W}\Delta)^{\otimes} + O((kh)^{-1}).$$

Using similar arguments as in the proof of theorem 5 of Fan et al. (2001), we can easily show that the term in the first bracket is of order  $O_p(k(\theta-\theta_1)h^{-1})$ ; thus we have

$$\begin{aligned} k^{-1}h(lr_{\tau,k} - lr_{l,k}) &= nh \frac{(1-\theta)(\theta-\theta_1)}{(1-\theta_1)} \\ &\times \int_0^1 \Delta_n^2(u) f(u) du (1 + O(h^4)) + O_p(k^{-1}). \end{aligned}$$

The alternative hypothesis implies that  $k^{-1}h(lr_{\tau,k} - lr_{l,k}) > 0$  almost everywhere, which completes the proof. The proof for case where  $\tau < t < k$  is analogous to the foregoing arguments and thus is omitted here.

For part b, by similar arguments as those in the proof of part a, the proof of consistency is straightforward and thus is omitted here. The details are available from authors on request.

*Remark A.1.* To facilitate the technical arguments and provide concise illustration, the alternative hypothesis considered here is free of  $k$ . In fact, for various contiguous alternatives with respect to  $k$  (Bhattacharya 1987; Gombay and Horvath 1996; Csorgo and Horvath 1997), the consistency of the foregoing change point estimator still holds when some stronger conditions are imposed.

[Received April 2007. Revised February 2008.]

## REFERENCES

- Azzalini, A., and Bowman, A. W. (1993), "On the Use of Nonparametric Regression for Checking Linear Relationships," *Journal of the Royal Statistical Society, Ser. B*, 55, 549–557.
- Bhattacharya, P. K. (1987), "Maximum Likelihood Estimation of a Change-Point in the Distribution of Independent Random Variables: General Multi-parameter Case," *Journal of Multivariate Analysis*, 23, 183–208.
- Box, G. E. P. (1954), "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, I: Effect of Inequality of Variance in the One-Way Classification," *The Annals of Mathematical Statistics*, 25, 290–302.
- Chen, G., Cheng, S. W., and Xie, H. (2001), "Monitoring Process Mean and Variability With One EWMA Chart," *Journal of Quality Technology*, 33, 223–233.
- Colosimo, B. M., and Pacella, M. (2007), "On the Use of Principle Component Analysis to Identify Systematic Patterns in Roundness Profiles," *Quality and Reliability Engineering International*, 23, 707–725.
- Csorgo, M., and Horvath, L. (1997), *Limit Theorems in Change-Point Analysis*, New York: Wiley.
- Dette, H., and Neumeyer, N. (2001), "Nonparametric Analysis of Covariance," *The Annals of Statistics*, 29, 1361–1400.
- Ding, Y., Zeng, L., and Zhou, S. (2006), "Phase I Analysis for Monitoring Non-linear Profiles in Manufacturing Processes," *Journal of Quality Technology*, 38, 199–216.
- Fan, J. (1993), "Local Linear Regression Smoothers and Their Minima Efficiencies," *The Annals of Statistics*, 21, 196–216.
- Fan, J., and Gijbels, I. (1996), *Local Polynomial Modeling and Its Applications*, London: Chapman & Hall.
- Fan, J., and Huang, L.-S. (2001), "Goodness-of-Fit Tests for Parametric Regression Models," *Journal of the American Statistical Association*, 96, 640–652.
- Fan, J., Zhang, C., and Zhang, J. (2001), "Generalized Likelihood Ratio Statistics and Wilks Phenomenon," *The Annals of Statistics*, 29, 153–193.
- Gallant, A. R. (1987), *Nonlinear Statistical Models*, New York: Wiley.
- Gombay, E., and Horvath, L. (1996), "Approximations for the Time and Change and the Power Function in Change-Point Models," *Journal of Statistical Planning and Inference*, 52, 43–66.
- Guerre, E., and Lavergne, P. (2005), "Data-Driven Rate-Optimal Specification Testing in Regression Models," *The Annals of Statistics*, 33, 840–870.
- Gupta, S., Montgomery, D. C., and Woodall, W. H. (2006), "Performance Evaluation of Two Methods for Online Monitoring of Linear Calibration Profiles," *International Journal of Production Research*, 44, 1927–1942.
- Hall, P., and Marron, J. S. (1990), "On Variance Estimation in Nonparametric Regression," *Biometrika*, 77, 415–419.
- Hardle, W., and Mammen, E. (1993), "Comparing Nonparametric versus Parametric Regression Fits," *The Annals of Statistics*, 21, 1926–1947.
- Hart, J. D. (1997), *Nonparametric Smoothing and Lack-of-Fit Tests*, New York: Springer.
- 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.
- Imhof, J. P. (1961), "Computing the Distribution of Quadratic Forms in Normal Variables," *Biometrika*, 48, 419–426.
- Jin, J., and Shi, J. (1999), "Feature-Preserving Data Compression of Stamping Tonnage Information Using Wavelet," *Technometrics*, 41, 327–339.
- Johnson, N. L. (1959), "On an Extension of the Connexion Between Poisson and  $\chi^2$  Distributions," *Biometrika*, 46, 352–363.
- Kang, L., and Albin, S. L. (2000), "On-Line Monitoring When the Process Yields a Linear Profile," *Journal of Quality Technology*, 32, 418–426.
- 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.
- Lowry, C. A., Woodall, W. H., Champ, C. W., and Rigdon, S. E. (1992), "Multivariate Exponentially Weighted Moving Average Control Chart," *Technometrics*, 34, 46–53.
- Lucas, J. M., and Saccucci, M. S. (1990), "Exponentially Weighted Moving Average Control Scheme Properties and Enhancements," *Technometrics*, 32, 1–29.
- Mahmoud, M. A., and Woodall, W. H. (2004), "Phase I Analysis of Linear Profiles With Calibration Applications," *Technometrics*, 46, 380–391.
- 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.
- Mestek, O., Pavlik, J., and Suchanek, M. (1994), "Multivariate Control Charts: Control Charts for Calibration Curves," *Journal of Analytical Chemistry*, 350, 344–351.
- Muller, H.-G. (1984), "Optimal Designs for Nonparametric Kernel Regression," *Statistics and Probability Letters*, 2, 285–290.
- Opsomer, J., Wang, Y., and Yang, Y. (2001), "Nonparametric Regression With Correlated Errors," *Statistical Science*, 16, 134–153.
- Pignatiello Jr., J. J., and Samuel, T. R. (2001), "Estimation of the Change Point of a Normal Process Mean in SPC Applications," *Journal of Quality Technology*, 33, 82–95.
- Quesenberry, C. P. (1995), "On Properties of  $Q$  Charts for Variables," *Journal of Quality Technology*, 27, 184–203.
- Runger, G. C., and Prabhu, S. S. (1996), "A Markov Chain Model for the Multivariate Exponentially Weighted Moving Averages Control Chart," *Journal of American Statistical Association*, 91, 1701–1706.
- Ruppert, D., and Wand, M. P. (1994), "Multivariate Locally Weighted Least Squares Regression," *The Annals of Statistics*, 22, 1346–1370.
- Seber, G. A. F., and Wild, C. J. (1989), *Nonlinear Regression*, New York: Wiley.
- Stover, F. S., and Brill, R. V. (1998), "Statistical Quality Control Applied to Ion Chromatography Calibrations," *Journal of Chromatography A*, 804, 37–43.
- Walker, E., and Wright, S. P. (2002), "Comparing Curves Using Additive Models," *Journal of Quality Technology*, 34, 118–129.
- Wang, K., and Tsung, F. (2007), "Run-to-Run Process Adjustment Using Categorical Observations," *Journal of Quality Technology*, 39, 312–325.
- 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. (2007), "Current Research on Profile Monitoring," *Revista Producção*, 17, 420–425.

- 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.
- Young, S. G., and Bowman, A. W. (1995), "Non-Parametric Analysis of Covariance," *Biometrics*, 51, 920–931.
- Zhang, C. (2003), "Calibrating the Degrees of Freedom for Automatic Data Smoothing and Effective Curve Checking," *Journal of the American Statistical Association*, 98, 609–628.
- Zhu, L., and Xue, L. (2006), "Empirical Likelihood Confidence Regions in a Partially Linear Single-Index Model," *Journal of the Royal Statistical Society, Ser. B*, 68, 549–570.
- Zou, C., Tsung, F., and Wang, Z. (2007), "Monitoring General Linear Profiles Using Multivariate EWMA Schemes," *Technometrics*, 49, 395–408.
- Zou, C., Zhang, Y., and Wang, Z. (2006), "Control Chart Based on Change-Point Model for Monitoring Linear Profiles," *IIE Transactions*, 38, 1093–1103.