國立交通大學

統計學研究所

碩士論文

使用剖面資料資訊提昇製程監控能力之研究

Enhancing the Power of Process Control by Utilizing Auxiliary Information in Profiles

研究生：黃香菱

指導教授：洪志真　博士

中華民國九十七年六月

使用剖面資料資訊提昇製程監控能力之研究

Enhancing the Power of Process Control by Utilizing Auxiliary Information in Profiles

研 究 生：黃香菱　　Student：Hsiang-Ling Huang

指導教授：洪志真　　Advisor：Dr. Jyh-Jen Horng Shiau

國 立 交 通 大 學

統計學研究所

碩 士 論 文

A Thesis

Submitted to Institute of Statistics

College of Science

National Chiao Tung University

in Partial Fulfillment of the Requirements

for the Degree of

Master

in

Statistics

June 2008

Hsinchu, Taiwan, Republic of China

中華民國九十七年六月

使用剖面資料資訊提昇製程監控能力之研究

研究生：黃香菱　　指導教授：洪志真博士

國立交通大學統計學研究所

摘要

傳統的統計製程管制(SPC)方法是直接監控有興趣的產品品質特性。由於現在科技發達，我們可以很容易的隨著時間點的改變紀錄一些與產品品質特性有相關的量測值，而這些量測值可以提供產品品質特性額外的資訊。舉例來說，我們可以在產品完成前，在不同的時間點上記錄品質特性的反應變數值。若我們把這些記錄值視為產品的剖面(profile)資料，則最終產品之品質特性即為剖面資料的最後一點。我們可以預期這些輔助的資訊可以提供我們更有效的製程監控方法。

本篇論文只探討第二階段的製程監控。我們提出的方法是用有母數、無母數迴歸的方法來配適產品的剖面資料，且利用配適出來之剖面的最後一點來對製程進行監控。我們證明利用了輔助資訊來對製程進行監控會有較好的偵測力。

關鍵字：剖面、正交多項式、無母數迴歸、區域多項式迴歸、高斯隨機過程

Enhancing the Power of Process Control by Utilizing Auxiliary Information in Profiles

Student: Hsiang-Ling Huang Advisor: Dr. Jyh-Jen Horng Shiau

Institute of Statistics

National Chiao Tung University

Abstract

Traditional control charting methods in statistical process control monitor the quality characteristic of the product or process of interest directly. Now thanks to well-developed technologies, we can easily record many other related data that may provide additional information about the quality characteristic at different time points. For example, we can record the quality response variable along the time before the end product is finished. If we regard the record of these values as the profile of a product item, then the quality characteristic of the end product is the endpoint of the profile. It is natural to expect that this auxiliary information would be able to help enhancing the control process.

In this study, we focus on Phase II monitoring. Our approach is to fit each profile by parametric or nonparametric regression methods, and then monitor the fitted endpoint value instead of the endpoint response. It can be shown that, with the additional profiles information, the fitted endpoint response has smaller variance than the endpoint response itself. Accordingly, when compared to the traditional approach, better detecting power can be achieved with the proposed approach.

Key Words: Profiles, orthogonal polynomials, nonparametric regression, local polynomial smoothing, Gaussian stochastic processes.

誌謝

這篇論文能順利的完成，首先要感謝洪志真教授，謝謝老師對於這篇論文辛勤的指導與批閱。於生活及求職上也要感謝洪志真老師與洪慧念老師適時的給予我建議及鼓勵。此外感謝博班家鈴學姐的幫忙及關心與同指導教授的怡玲及筱娟給予我的鼓勵及照顧，還有我的好室友珮琦、郃嵐的陪伴，及吃飯的好夥伴夙吟、姿蒨、佩芳，有著你們大家的參與，我的研究所的生涯才能過得如此多姿多彩。謝謝交大統計所，提供良好的學習環境，讓我抱著期待與學習的心到來，經過兩年的洗禮後，此刻我將懷著感謝與難忘的回憶離開這裡。

最後將此篇論文獻給我最敬愛的父母親，感謝他們讓我物質生活不虞匱乏，使我能將心力都放在課業上，謝謝他們辛苦的栽培。

黃 香 菱　　謹誌于

國立交通大學統計學研究所

中華民國九十七年六月

Contents

List of Tables

List of Figures

1 Introduction

Control charts, a primary tool in statistical process control (SPC), is powerful in monitoring the quality of processes or products and has been widely applied in industries and other fields since Shewhart introduced the technique in 1924. Standard control charting usually consists of two distinct phases, namely, Phase I and Phase II. The major goal of the Phase I is to quickly bring the process into a stable or in-control state and collect some in-control data so that reliable control limits can be established for effective Phase II on-line monitoring of future production.

In this study, we only focus on Phase II monitoring. The new approach we propose for control charting is different from the traditional approach of monitoring the quality characteristic of interest directly. Our approach is to record the quality characteristic along the time before the end product is finished. The record as such for a product item, considered as a function of time, is referred to as the profile of the product item hereafter. Then, the quality characteristic of the end product is the endpoint of the profile. For illustration, consider a bottle-filling example as follows. Suppose the quality characteristic of interest is the total amount of soda in the bottle and the filling process takes 5 seconds to fill the bottle with soda. Assume that the amount of soda inside the bottle can be recorded every 250 milliseconds. Then, the response of the end product, i.e., the endpoint of the profile, is the total amount of soda in the bottle.

While considering each of these records as a profile, we only focus on the endpoint of it. This is quite different from the usual profile monitoring, in which the quality characteristic monitored is the whole profile. The profile monitoring has been a popular research topic in SPC since Kang and Albin (2000) first noted that the quality of some processes can be better characterized by profiles.

Since our emphasis is on the endpoint of a profile, the rest of the profile can be considered as auxiliary information. Could we utilize this auxiliary information so as to monitor the end product more effectively? Or would it assist us to screen out substandard products more quickly?

Assume there are m profiles and each of them has n measurements. Denote y_{ij} as

the response of the jth profile at the time point x_i. Kang and Albin (2000) considered the following linear fixed-effect model for profiles obtained from an etching process in semiconductor manufacturing:

$$y_{ij} = A_0 + A_1 x_i + \varepsilon_{ij}, \quad i = 1, \cdots, n, \quad j = 1, \cdots, m, \tag{1}$$

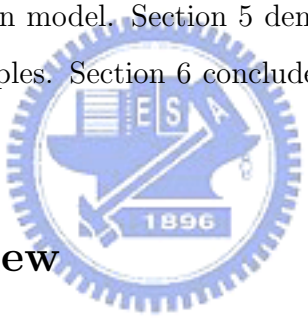
where the intercept A_0 and slope A_1 are fixed parameters. The random variables ε_{ij} are independent and identically distributed (i.i.d.) as a normal distribution with mean zero and variance σ^2. Kim et al. (2003) also studied this same problem and proposed an alternative method for linear profile monitoring. The model studied in Kang and Albin (2000) and Kim et al. (2003) is a fixed-effect model. However, under this model, some factors such as variations in temperature or pressure, characteristics of the measuring equipments, and other hard-to-control factors that may affect the response variable are all categorized as part of the random error ε_{ij}. For many applications, this simplified model may not be adequate because some of these hard-to-control factors may affect the values of the intercept and/or the slope of the linear profile. Shiau et al. (2006) argued that these hard-to-control factors should be considered as common causes of variation and proposed a random-effect linear model in order to cope with these common-cause variations. More specifically, now A_{0j} and A_{1j} are random variables instead of fixed parameters. With the same reasoning, we consider random-effect models in this study.

To utilize the auxiliary information contained in the profile, we fit each linear profile by a regression line and estimate the mean of the response at the endpoint by the fitted value. We then monitor this estimate instead of the original endpoint response. The reason behind this approach is simple: it is expected that this estimator would have smaller variance than that of the endpoint response, because we borrow some "strength" from the rest of the profile; and smaller variance may lead to better detecting power of the control chart. Our study shows that, with the new approach, better detecting power is indeed achieved for process monitoring.

We further extend this approach to nonlinear profiles. We believe the new approach would outperform the traditional approach in practice as well. However, the distribution of the parametric nonlinear regression estimator of the mean of the endpoint response

depends on the form of the parametric model, and is usually too complicated to compare with the endpoint response analytically. Thus we study the nonparametric regression approach. The nonparametric regression estimation method we started with was the kernel smoothing due to its simplicity. Unfortunately, the ordinary kernel estimate has a serious boundary-effect problem and the endpoint is a boundary point. We then decide to use the local polynomial smoothing because it is boundary-effect free. By monitoring the fitted endpoint value of the profile, we prove that better detecting power can be achieved.

The rest of paper is organized as follows. Section 2 reviews the linear profile with fixed/random effects, orthogonal polynomial regression as described in Montgomery et al. (2006), nonparametric regression estimation methods including kernel and local polynomial smoothing, and Gaussian stochastic processes. Section 3 studies the proposed approach with an orthogonal polynomial regression model. Section 4 studies our approach with a nonparametric regression model. Section 5 demonstrates the proposed methodology with two illustrative examples. Section 6 concludes the thesis with a brief summary and some remarks.



2 Literature Review

2.1 Linear profiles

2.1.1 Linear Profiles with Fixed Effects

Kang and Albin (2000) described the output quality characteristic of a process as a random variable Y that is a linear function (or profile) of an independent variable X. They modeled the linear profiles as in (1). For the jth profile, the least squares estimators α_{0j} and α_{1j} of A_0 and A_1 respectively are as follows:

$$a_{0j} = \bar{y}_j - a_{1j}\bar{x} \quad \text{and} \quad a_{1j} = \frac{S_{xy(j)}}{S_{xx}}, \tag{2}$$

where $\bar{y}_j = \sum_{i=1}^{n} y_{ij}/n$, $\bar{x} = \sum_{i=1}^{n} x_i/n$, $S_{xy(j)} = \sum_{i=1}^{n}(x_i - \bar{x})y_{ij}$, and $S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2$. In this model, the estimators a_{0j} and a_{1j} are not independent.

Kim et al. (2003) coded the x-values by centering so that the average coded-value is

zero. With the centered x-values, the least squares estimators of the intercept and slope are independent random variables. Details can be seen in Myers (1990, p. 11-15).

2.1.2 Linear Profiles with Random Effects

Shiau et al. (2006) considered representing linear profiles with random effects by the following model:

$$y_{ij} = A_{0j} + A_{1j}x_i + \varepsilon_{ij} \quad i = 1, \cdots, n, \quad j = 1, \cdots, m, \tag{3}$$

where the response variable y_{ij} refers to the ith measurement of the jth profile, random coefficients A_{0j}'s and A_{1j}'s are i.i.d. as $N(\alpha_0, \sigma_0^2)$ and $N(\alpha_1, \sigma_1^2)$ respectively, and error variables ϵ_{ij}'s are i.i.d. as $N(0, \sigma_e^2)$. Moreover, assume that A_{0j}, A_{1j}, and ε_{ij} are mutually independent.

Following Kim et al. (2003), Shiau et al. (2006) centered the x-values so that $\bar{x} = 0$ in model (3) for simplicity. To avoid introducing new notation, let the set points x_i's represent the centered x-values. Shiau et al. (2006) first fitted each profile by linear regression as follows. For the jth profile, by treating A_{0j} and A_{1j} as two fixed parameters, the least squares estimators $\hat{\alpha}_{0j}$ and $\hat{\alpha}_{1j}$ of A_{0j} and A_{1j} respectively are

$$\hat{\alpha}_{0j} = \bar{y}_j \quad \text{and} \quad \hat{\alpha}_{1j} = \frac{S_{xy(j)}}{S_{xx}},$$

where $\bar{y}_j = \sum_{i=1}^n y_{ij}/n$, $S_{xy(j)} = \sum_{i=1}^n x_i(y_{ij} - \bar{y}_j)/n = \sum_{i=1}^n x_i\, y_{ij}$, and $S_{xx} = \sum_{i=1}^n x_i^2/n$.

With centered x-values, they also showed that

$$(i)\, \hat{\alpha}_{01}, \cdots, \hat{\alpha}_{0m} \text{ are i.i.d. } N(\alpha_0, \sigma_0^2 + \tfrac{1}{n}\sigma_e^2),$$
$$(ii)\, \hat{\alpha}_{11}, \cdots, \hat{\alpha}_{1m} \text{ are i.i.d. } N(\alpha_1, \sigma_1^2 + \tfrac{1}{S_{xx}}\sigma_e^2),$$

and these statistics are mutually independent.

2.2 Orthogonal Polynomial Regression

2.2.1 General Statistical Properties

Using orthogonal polynomials to fit data is a popular method to avoid the ill-conditioning in computation. We review the orthogonal polynomials as described in Montgomery et

al. (2006). For a set of data $\{(x_i, y_i),\ i = 1, \cdots, n\}$, consider the orthogonal polynomial model as follows:

$$y_i = \alpha_0 P_0(x_i) + \alpha_1 P_1(x_i) + \alpha_2 P_2(x_i) + \cdots + \alpha_k P_k(x_i) + \varepsilon_i, \quad i = 1, \cdots, n,$$

where $P_r(\cdot)$ is an rth-degree orthogonal polynomial defined such that

$$\sum_{i=1}^{n} P_r(x_i) P_s(x_i) = 0, \quad r \neq s, \quad r, s = 0, 1, \cdots, k, \quad \text{and} \quad P_0(x_i) = 1, \text{ for } i = 1, \cdots, n.$$

Rewrite this model in vector form:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{y} = (y_1, \cdots, y_n)^T$, $\boldsymbol{\alpha} = (\alpha_0, \cdots, \alpha_k)^T$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n)^T$, and the design matrix

$$\boldsymbol{X} = \begin{bmatrix} P_0(x_1) & P_1(x_1) & \cdots & P_k(x_1) \\ P_0(x_2) & P_1(x_2) & \cdots & P_k(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ P_0(x_n) & P_1(x_n) & \cdots & P_k(x_n) \end{bmatrix}.$$

Then $\boldsymbol{X}$ has mutually orthogonal columns and

$$\boldsymbol{X}^T\boldsymbol{X} = \begin{bmatrix} \sum_{i=1}^n P_0^2(x_i) & 0 & \cdots & 0 \\ 0 & \sum_{i=1}^n P_1^2(x_i) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sum_{i=1}^n P_k^2(x_i) \end{bmatrix}.$$

The least squares estimator of $\boldsymbol{\alpha}$ is $\hat{\boldsymbol{\alpha}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} \equiv (\hat{\alpha}_0, \cdots, \hat{\alpha}_k)^T$, where

$$\hat{\alpha}_r = \frac{\sum_{i=1}^n P_r(x_i) y_i}{\sum_{i=1}^n P_r^2(x_i)}, \qquad r = 0, 1, \cdots, k. \tag{4}$$

2.2.2 Generating Orthogonal Polynomials with Equally-Spaced x-Values

When the levels of x are equally spaced, the orthogonal polynomials $P_r(x_i)$ can be easily constructed by the following:

$$\begin{aligned} P_0(x_i) &= 1, \\ P_1(x_i) &= \lambda_1(\tfrac{x_i - \bar{x}}{d}), \\ P_2(x_i) &= \lambda_2((\tfrac{x_i - \bar{x}}{d})^2 - (\tfrac{n^2-1}{12})), \\ P_3(x_i) &= \lambda_3((\tfrac{x_i - \bar{x}}{d})^3 - (\tfrac{x_i - \bar{x}}{d})(\tfrac{3n^2-7}{20})), \text{ and so on,} \end{aligned}$$

where d is the distance between the levels of x, n is the total number of levels, and $\{\lambda_r\}$ are constants such that the polynomials will have integer values. See Pearson (1996, Table 47) for more details. Orthogonal polynomials can be calculated directly by the following method (Seber, 2003, Ch7): if x-values are equally spaced, by transforming them to

$$x_i = i - \frac{1}{2}(n+1),$$

the orthogonal polynomials of degrees 1-3 are

$$\begin{aligned} P_0(x_i) &= 1, \\ P_1(x_i) &= \lambda_1 x_i, \\ P_2(x_i) &= \lambda_2(x_i^2 - \frac{1}{12}(n^2-1)), \\ P_3(x_i) &= \lambda_3(x_i^3 - \tfrac{1}{20}(3n^2-7)x_i). \end{aligned} \tag{5}$$

Again, the purpose of the multiples $\{\lambda_r\}$ here is to make the orthogonal polynomials to have integer values.

2.3 Fitting Data by Localized Least Squares

In general, the nonparametric regression model is defined as follows:

$$y_i = m(x_i) + \varepsilon_i,$$

where $m(x)$ is the regression function and ε_i's are i.i.d. random variables with $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = \sigma^2$. In the following, we give a brief review on kernel smoothing and local polynomial smoothing methods as described in Fan and Gijbels (1996), Schimek (2000), Simonoff (1996), Ramsay and Silverman (2005), and others.

2.3.1 Kernel Smoothing

The kernel estimator is the simplest and classic estimator obtained by locally averaging data. In other words, the kernel estimator at a given point x is a linear combination of local observations, i.e.,

$$\hat{m}(x) = \sum_{i=1}^{n} w_i(x) y_i \tag{6}$$

for some suitably defined weights $w_i(x)$. The most popular kernel estimator, the Nadaraya-Waston estimator, is constructed by using the weights

$$w_i(x) = \frac{K_h(x_i - x)}{\sum_{k=1}^{n} K_h(x_k - x)} , \tag{7}$$

where $K_h(\cdot) = K(\cdot/h)/h$ and K is a *kernel function* that is usually a symmetric probability density, and parameter h in the weight function is called the *bandwidth*. Small values of bandwidth imply that only observations close to x receive some weights, while large h means that a wide range of observations that are at a considerable distance from x are used for local averaging. Substituting (7) into (6), the Nadaraya-Waston estimator is

$$\hat{m}(x) = \frac{\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right) y_i}{\sum_{k=1}^{n} K\left(\frac{x_k - x}{h}\right)} . \tag{8}$$

2.3.2 Local Polynomial Smoothing

Suppose the regression function $m(x)$ can be approximated by

$$m(x) \approx \sum_{v=0}^{p} \frac{m^{(v)}(x_0)}{v!}(x - x_0)^v \equiv \sum_{v=0}^{p} \beta_v (x - x_0)^v$$

for x_0 in a neighborhood of x, by using Taylor's expansion. Then a pth *degree local polynomial regression estimator* at x_0 can be obtained by minimizing

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \{y_i - \sum_{v=0}^{p} \beta_v (x_i - x_0)^v\}^2 K_h(x_i - x_0) , \tag{9}$$

where $\boldsymbol{\beta} = (\beta_0, \cdots, \beta_p)^T$.

Let $\boldsymbol{X}$ be the design matrix

$$\boldsymbol{X} = \begin{pmatrix} 1 & (x_1 - x_0) & \cdots & (x_1 - x_0)^p \\ 1 & (x_2 - x_0) & \cdots & (x_2 - x_0)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (x_n - x_0) & \cdots & (x_n - x_0)^p \end{pmatrix} .$$

Further, let $\boldsymbol{W}$ be the $n \times n$ diagonal matrix of weights:

$$\boldsymbol{W} = \text{diag}\left(K_h(x_i - x_0)\right).$$

Denote $\boldsymbol{y} = (y_1, \cdots, y_n)^T$. Then the weighted least squares problem (9) can be written as

$$\min_{\boldsymbol{\beta}} \ (\boldsymbol{y} - \boldsymbol{X\beta})^T \boldsymbol{W} (\boldsymbol{y} - \boldsymbol{X\beta}).$$

The solution vector is

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{y}.$$

Let $\boldsymbol{S}_n = \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X}$, we can write the local polynomial regression estimator $\hat{\boldsymbol{\beta}}$ as

$$\hat{\boldsymbol{\beta}} = \boldsymbol{S}_n^{-1} \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{y},$$

where

$$\boldsymbol{S}_n = \begin{pmatrix} S_{n,0} & S_{n,1} & \cdots & S_{n,p} \\ S_{n,1} & S_{n,2} & \cdots & S_{n,p+1} \\ \vdots & \vdots & \ddots & \vdots \\ S_{n,p} & S_{n,p+1} & \cdots & S_{n,2p} \end{pmatrix}$$

with

$$S_{n,v} = \sum_{i=1}^{n} K_h(x_i - x_0)(x_i - x_0)^v. \tag{10}$$

2.3.3 Local Polynomial Smoothing with Equivalent Kernels

The estimator $\hat{\beta}_v$ of $m^{(v)}(x_0)/v!$ can be written as

$$\hat{\beta}_v = \boldsymbol{e}_{v+1}^T \hat{\boldsymbol{\beta}} = \boldsymbol{e}_{v+1}^T \boldsymbol{S}_n^{-1} \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{y} = \sum_{i=1}^{n} W_v^n \left(\frac{x_i - x_0}{h}\right) y_i, \quad v = 0, \cdots, p, \tag{11}$$

where the unit vector $\boldsymbol{e}_{v+1} = (0, \cdots, 0, 1, 0, \cdots, 0)^T$ with 1 on the $(v+1)$th, and $W_v^n(t) = \boldsymbol{e}_{v+1}^T \boldsymbol{S}_n^{-1} \left(1, th, \cdots, (th)^p\right)^T K(t)/h$. Then the estimator $\hat{\beta}_v$ is a type of kernel estimator except that the "kernel" W_v^n depends on the design points and x_0. This dependence explains why the local polynomial regression estimators adapt automatically to various designs and to the location x_0 (either in the interior or at the boundaries of the support).

It is easy to show that the kernel weight function W_v^n satisfies the following discrete moment conditions: for $0 \le v, q \le p$,

$$\sum_{i=1}^{n}(x_i - x_0)^q W_v^n\left(\frac{x_i - x_0}{h}\right)$$

$$= \boldsymbol{e}_{v+1}^T \boldsymbol{S}_n^{-1} \sum_{i=1}^{n}(x_i - x_0)^q \begin{pmatrix} 1 \\ x_i - x_0 \\ \vdots \\ (x_i - x_0)^p \end{pmatrix} K_h(x_i - x_0)$$

$$= \boldsymbol{e}_{v+1}^T \boldsymbol{S}_n^{-1} \boldsymbol{S}_n \boldsymbol{e}_{q+1} = \delta_{v,q}.$$

Consequently, the finite-sample bias is zero when using local polynomial fitting of degree p to estimate polynomials of degree up to p.

2.4 Gaussian Stochastic Processes

We introduce the covariance function G of the Gaussian stochastic process with mean zero as described in Wahba (1990). Let $\mathcal{T}$ be an index set, say, [0,1].

Definition 1 *A symmetric, real-valued function $G(s,t)$ with $s,t \in \mathcal{T}$ is said to be positive definite if, for any real $a_1, \cdots, a_n$ and $t_1, \cdots, t_n \in \mathcal{T}$,*

$$\sum_{i,j=1}^{n} a_i a_j G(t_i, t_j) \geq 0.$$

If $G(\cdot,\cdot)$ is positive definite, then we can define a family $\{Y(t), t \in \mathcal{T}\}$ of zero-mean Gaussian stochastic process with covariance function G, that is,

$$E\,Y(s)Y(t) = G(s,t), \qquad s,t \in \mathcal{T}.$$

Given a positive-definite function $G(\cdot,\cdot)$, we can associate with it a *reproducing kernel Hilbert space* (r.k.h.s.).

Definition 2 *A (real) reproducing kernel Hilbert space is a Hilbert space of real-valued functions on $\mathcal{T}$ with the property that, for each $t \in \mathcal{T}$, the evaluation functional L_t, which associates f with $f(t)$, $L_t f \mapsto f(t)$, is a bounded linear functional. The boundedness means that there exists an $M = M_t$ such that*

$$\mid L_t f \mid = \mid f(t) \mid \leq M \parallel f \parallel \quad \textit{for all} \quad f \quad \textit{in the r.k.h.s.},$$

where $\|\cdot\|$ *is the norm in the Hilbert space.*

If $\mathcal{H}$ is an r.k.h.s., then for each $t \in \mathcal{T}$ there exists, by the Riesz representation theorem, an element $\mathcal{R}_t$ in $\mathcal{H}$ with the property $L_t f =< R_t, f >= f(t)$, $f \in \mathcal{H}$, where R_t is called the representer of evaluation at t and $< \cdot, \cdot >$ is the inner product in $\mathcal{H}$.

Let $W_m[0,1] = \{f : f, f', \cdots, f^{m-1}$ are absolutely continuous and $f^{(m)} \in \mathcal{L}_2[0,1]\}$, $\mathcal{B}_m = \{f : f \in W_m[0,1],\ f^{(v)}(0) = 0,\ v = 0, \cdots, m-1\}$, and $W_m^0[0,1] = \{f : f \in \mathcal{B}_m,\ f, f', \cdots, f^{m-1}$ are absolutely continuous and $f^{(m)} \in \mathcal{L}_2[0,1]\}$.

Theorem (Taylor's theorem) *If* f *is a real-valued function on* $[0,1]$ *with* $m-1$ *continuous derivatives and* $f^{(m)} \in \mathcal{L}_2[0,1]$*, then we may write*

$$f(t) = \left\{ \sum_{v=0}^{m-1} \frac{t^v}{v!} f^{(v)}(0) \right\} + \left\{ \int_0^1 \frac{(t-u)_+^{(m-1)}}{(m-1)!} f^{(m)}(u) du \right\},$$

where $(x)_+ = x$ *for* $x \geq 0$ *and* $(x)_+ = 0$ *otherwise. If* $f \in \mathcal{B}_m$*, then*

$$\begin{aligned} f(t) &= \int_0^1 \frac{(t-u)_+^{m-1}}{(m-1)!} f^{(m)}(u) du \\ &= \int_0^1 G_m(t,u) f^{(m)}(u) du, \end{aligned}$$

where $G_m(t,u) = (t-u)_+^{m-1}/(m-1)!$.

Wahba (1990) proved that $W_m^0[0,1]$ is an r.k.h.s. with r.k. $G(s,t) = \int_0^1 G_m(t,u) G_m(s,u) du$. When $m = 2$, $W_2^0[0,1]$ is an r.k.h.s. with r.k.

$$G(s,t) = \int_0^1 G_2(t,u) G_2(s,u) du = \int_0^1 (t-u)_+ (s-u)_+ du = \int_0^{min(s,t)} (t-u)(s-u) du$$

$$= \begin{cases} \int_0^s (t-u)(s-u) du & \text{if } s < t \\ \int_0^t (t-u)(s-u) du & \text{if } s \geq t \end{cases} = \begin{cases} \frac{ts^2}{2} - \frac{s^3}{6}, & \text{if } s < t \\ \frac{st^2}{2} - \frac{t^3}{6}, & \text{if } s \geq t \end{cases}. \qquad (12)$$

The general form of r.k. is

$$G(s,t) = \begin{cases} \sum_{i=1}^{m} (-1)^{i+1} \frac{1}{(m+i-1)!(m-i)!} t^{m-i} s^{m+i-1} & \text{if } s < t \\ \sum_{i=1}^{m} (-1)^{i+1} \frac{1}{(m+i-1)!(m-i)!} s^{m-i} t^{m+i-1} & \text{if } s \geq t \end{cases}.$$

3 Methodologies: Parametric Regression Approach

3.1 Linear Profiles with Random Effects

Assume there are many profiles and each profile has n equally spaced measurements. Fit these profiles with the linear random-effect regression model (3). For the jth profile, by treating A_{ij} and A_{1j} as two fixed parameters, the least square estimators $\hat{\alpha}_{0j}$ and $\hat{\alpha}_{1j}$ of A_{ij} and A_{1j} respectively are in (2).

Proposition 1 *Under the linear random-effect model (3), we can show that*

$$\begin{aligned}
&(i)\, \hat{\alpha}_{01}, \cdots, \hat{\alpha}_{0m} \text{ are i.i.d. } N\left(\alpha_0, \sigma_0^2 + (\tfrac{1}{n} + \tfrac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2})\sigma_e^2\right), \\
&(ii)\, \hat{\alpha}_{11}, \cdots, \hat{\alpha}_{1m} \text{ are i.i.d. } N\left(\alpha_1, \sigma_1^2 + \tfrac{1}{\sum (x_i - \bar{x})^2}\sigma_e^2\right), \\
&(iii)\, Cov(\hat{\alpha}_{0j}, \hat{\alpha}_{1j}) = -\tfrac{\bar{x}}{\sum (x_i - \bar{x})^2}\sigma_e^2.
\end{aligned}$$

Then, the endpoint response Y_{nj} is normally distributed with

$$\begin{aligned}
E(Y_{nj}) &= E(A_{0j} + A_{1j}x_n + \varepsilon_{nj}) = \alpha_0 + \alpha_1 x_n, \\
Var(Y_{nj}) &= Var(A_{0j} + A_{1j}x_n + \varepsilon_{nj}) = \sigma_0^2 + \sigma_1^2 x_n^2 + \sigma_e^2 \,.
\end{aligned} \tag{13}$$

Let $\hat{Y}_{nj} \equiv \hat{\alpha}_{0j} + \hat{\alpha}_{1j}x_n$ be the fitted response value at the endpoint. By Proposition 1, $\hat{Y}_{nj}$ is normally distributed with

$$\begin{aligned}
E(\hat{Y}_{nj}) &= E(\hat{\alpha}_{0j} + \hat{\alpha}_{1j}x_n) = \alpha_0 + \alpha_1 x_n \\
Var(\hat{Y}_{nj}) &= Var(\hat{\alpha}_{0j} + \hat{\alpha}_{1j}x_n) = \sigma_0^2 + \sigma_1^2 x_n^2 + (\tfrac{1}{n} + \tfrac{(x_n - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2})\sigma_e^2 \,.
\end{aligned} \tag{14}$$

Comparing equations (13) and (14) above, we find that the variance of $\hat{Y}_{nj}$ is smaller than the variance of Y_{nj} as long as $(1/n + (x_n - \bar{x})^2 / \sum_{i=1}^n (x_i - \bar{x})^2) < 1$, a condition that usually holds for large n. It can be shown that smaller variance will lead to larger detecting power for the control chart in process monitoring.

Alternatively, model (3) can be re-expressed as the following mixed-effect model:

$$y_{ij} = \alpha_0 + \alpha_1 x_i + (A_{0j} - \alpha_0) + (A_{1j} - \alpha_1)x_i + \varepsilon_{ij},\ i = 1, \cdots, n,\ j = 1, 2, \cdots . \tag{15}$$

For the jth profile, by expressing (15) as

$$y_{ij} = \alpha_0 + \alpha_1 x_i + \varepsilon_{ij}^*,\ i = 1, \cdots, n,\ j = 1, 2, \cdots ,$$

where $\varepsilon^*_{ij} = (A_{0j} - \alpha_0) + (A_{1j} - \alpha_1)x_i + \varepsilon_{ij}$, we can obtain another set of estimators $\hat{\alpha}^*_{0j}$ and $\hat{\alpha}^*_{1j}$ for α_0 and α_1 by the weighted least squares estimation method. Let $\hat{Y}^*_{nj} = \hat{\alpha}^*_{0j} + \hat{\alpha}^*_{1j}x_n$. It is found by simulation that using $\hat{Y}^*_{nj}$ from the mixed-effect model (15) does not gain more detecting power than using $\hat{Y}_{nj}$. However, the form of $Var(\hat{Y}^*_{nj})$ is more complicated than that of $Var(\hat{Y}_{nj})$. Thus, we adopt monitoring $\hat{Y}_{nj}$ hereafter.

In the next subsection, we consider the case of nonlinear profiles that can be adequately fitted by polynomial regression models.

3.2 Polynomial Profiles with Random Effects

Assume there are many profiles and each profile has n equally spaced measurements. Fit these profiles with the following random-effect orthogonal polynomial regression model:

$$y_{ij} = A_{0j}P_0(x_i) + A_{1j}P_1(x_i) + \cdots + A_{pj}P_p(x_i) + \varepsilon_{ij}, \quad i = 1, \cdots, n, \quad j = 1, 2, \cdots \quad (16)$$

where $P_r(x_i)$ is an rth-degree orthogonal polynomial defined such that $\sum_{i=1}^{n} P_r(x_i)P_s(x_i) = 0 \;\; r \neq s, \; r, s = 0, \cdots, p$, and $P_0(x_i) = 1$. We assume that $A_{rj} \sim N(\alpha_r, \sigma_r^2), \; r = 0, \cdots, p$, are independent normal random variables and ε_{ij}'s are i.i.d. as $N(0, \sigma_e^2)$. Moreover, assume that A_{rj} and ε_{ij} are mutually independent for $r = 0, \cdots, p$.

For the jth profile, by treating $A_{0j}, A_{1j}, \cdots, A_{pj}$ as fixed parameters, we estimate them respectively by the least squares estimators $\hat{\alpha}_{0j}, \hat{\alpha}_{1j}, \cdots, \hat{\alpha}_{pj}$ given in equation (5). Let $Y_{nj} = A_{0j} + A_{1j}P_1(x_n) + \cdots + A_{pj}P_p(x_n) + \varepsilon_{nj}$ and $\hat{Y}_{nj} = \hat{\alpha}_{0j} + \hat{\alpha}_{1j}P_1(x_n) + \cdots + \hat{\alpha}_{pj}P_p(x_n)$.

Proposition 2 *Under the random-effect model* (16), *for each* $r = 0, 1, \cdots, p$, $\hat{\alpha}_{rj}$, $j = 1, 2, \cdots$, *are i.i.d. as* $N(\alpha_r, \sigma_r^2 + \frac{\sigma_e^2}{\sum_{i=1}^{n} P_r^2(x_i)})$ *and* $\hat{\alpha}_{rj}$, $r = 0, 1, \cdots, p$, $j = 1, 2, \cdots$, *are independent.*

Proposition 3 *Under the random-effect model* (16), Y_{nj}, $j = 1, 2, \cdots$, *are i.i.d. as* $N(\sum_{r=0}^{p} \alpha_r P_r(x_n), \sum_{r=0}^{p} \sigma_r^2 P_r^2(x_n) + \sigma_e^2)$

Proposition 4 *Under the random-effect model* (16), $\hat{Y}_{nj}$, $j = 1, 2, \cdots$, *are i.i.d. as* $N\left(\sum_{r=0}^{p} \alpha_r P_r(x_n), \sum_{r=0}^{p} [\sigma_r^2 P_r^2(x_n) + \frac{P_r^2(x_n)}{\sum_{i=1}^{n} P_r^2(x_i]} \sigma_e^2]\right)$.

The proofs of all the propositions in this thesis are given in Appendix A.

3.3 Phase II Method

In Phase II monitoring, we assume that the parameters $\alpha_0, \alpha_1, \cdots, \alpha_p$, $\sigma_0^2, \sigma_1^2, \cdots, \sigma_p^2$, and σ_e^2 are known. Set the overall false-alarm rate at α. Since the distributions of the monitoring statistics Y_{nj} and $\hat{Y}_{nj}$ are known, their control limits can be easily shown to be

$$LCL_{Y_n} = CL - Z_{\frac{\alpha}{2}}\eta_0, \quad UCL_{Y_n} = CL + Z_{\frac{\alpha}{2}}\eta_0,$$

$$LCL_{\hat{Y}_n} = CL - Z_{\frac{\alpha}{2}}\eta, \quad UCL_{\hat{Y}_n} = CL + Z_{\frac{\alpha}{2}}\eta,$$

where

$$\begin{aligned} CL &= \alpha_0 + \alpha_1 P_1(x_n) + \cdots + \alpha_p P_p(x_n)\,, \\ \eta_0 &= \left\{\sigma_0^2 + \sigma_1^2 P_1^2(x_n) + \cdots + \sigma_p^2 P_p^2(x_n) + \sigma_e^2\right\}^{1/2}, \\ \eta &= \left\{\sigma_0^2 + \sigma_1^2 P_1^2(x_n) + \cdots + \sigma_p^2 P_p^2(x_n) + \left[\frac{1}{n} + \frac{P_1^2(x_n)}{\sum_{i=1}^n P_1^2(x_i)} + \cdots + \frac{P_p^2(x_n)}{\sum_{i=1}^n P_p^2(x_i)}\right]\sigma_e^2\right\}^{1/2}, \end{aligned} \tag{17}$$

$Z_{\frac{\alpha}{2}}$ is the $(1-\alpha/2)$th quantile of the standard normal distribution. Y_n and $\hat{Y}_n$ charts have the same center line, $\alpha_0 + \alpha_1 P_1(x_n) + \cdots + \alpha_p P_p(x_n)$, but with different widths. The difference is in the standard deviation term. Similar to the case of linear profiles, when n is large, it is likely that the statistic $\hat{Y}_{nj}$ has smaller variance than Y_{nj}. We demonstrate this theoretically for the case of $p = 2$. If x-values are equally spaced, without loss of generality, we can transform them to $x_i = i - \frac{1}{2}(n+1)$ and calculate the orthogonal polynomials directly by (5). The results are as follows:

$$\begin{aligned} P_1^2(x_n) &= \frac{\lambda_1^2(n-1)^2}{4}, & \sum_{i=1}^n P_1^2(x_i) &= \frac{\lambda_1^2 n(n+1)(n-1)}{12}, \\ P_2^2(x_n) &= \frac{\lambda_2^2(n-2)^2(n-1)^2}{36}, & \sum_{i=1}^n P_2^2(x_i) &= \frac{\lambda_2^2 n(n^2-1)(n^2-4)}{180}, \end{aligned}$$

$$\left[\frac{1}{n} + \frac{P_1^2(x_n)}{\sum_{i=1}^n P_1^2(x_i)} + \frac{P_2^2(x_n)}{\sum_{i=1}^n P_2^2(x_i)}\right] = \frac{1}{n} + \frac{3(n-1)}{n(n+1)} + \frac{5(n-2)(n-1)}{n(n+1)(n+2)} \equiv C_n. \tag{18}$$

It is apparent that $C_n = O(1/n)$. We calculate C_n for various n and the result is shown in Table 1. From Table 1, we have $C_n \leq 1$ for all n and $C_n < 1$ for $n > 3$. From Figure 1, we can see the term C_n in (18) approaches zero for large n. Thus in the case of $p = 2$ and $n > 3$, we have $\eta < \eta_0$. Therefore $\hat{Y}_n$ chart has narrower control limits than Y_n chart. When n is larger, the variance of $\hat{Y}_{nj}$ becomes smaller. Since the normal distribution of $\hat{Y}_{nj}$ has smaller variance, the detecting power of $\hat{Y}_n$ chart is better than that of Y_n chart. The detecting power can be calculated by the following.

When the mean of Y_{nj} shifts with a size of $\delta\eta_0$, the probability for Y_{nj} still falling within the control limits is

$$\begin{aligned}\beta_0(\delta) &= P\left\{CL - Z_{\frac{\alpha}{2}}\eta_0 \leq Y_{nj} \leq CL + Z_{\frac{\alpha}{2}}\eta_0\right\} \\ &= P\left\{-\delta - Z_{\frac{\alpha}{2}} \leq \frac{Y_{nj} - (CL + \delta\eta_0)}{\eta_0} \leq -\delta + Z_{\frac{\alpha}{2}}\right\} \\ &= \Phi\left(-\delta + Z_{\frac{\alpha}{2}}\right) - \Phi\left(-\delta - Z_{\frac{\alpha}{2}}\right),\end{aligned}$$

where $\Phi(\cdot)$ is the c.d.f. of the standard normal distribution. If the mean of Y_{nj} shifts with a size of $\delta\eta_0$, then the mean of $\hat{Y}_{nj}$ will shifts with a size of $\delta'\eta$, where $\delta' = \delta\eta_0/\eta$. Then the probability that $\hat{Y}_{nj}$ falls within the control limits is

$$\begin{aligned}\beta(\delta) &= P\left\{CL - Z_{\frac{\alpha}{2}}\eta \leq \hat{Y}_{nj} \leq CL + Z_{\frac{\alpha}{2}}\eta\right\} \\ &= P\left\{-\delta' - Z_{\frac{\alpha}{2}} \leq \frac{\hat{Y}_{nj} - (CL + \delta'\eta)}{\eta} \leq -\delta' + Z_{\frac{\alpha}{2}}\right\} \\ &= \Phi\left(-\delta' + Z_{\frac{\alpha}{2}}\right) - \Phi\left(-\delta' - Z_{\frac{\alpha}{2}}\right).\end{aligned}$$

Because $\beta_0(\delta)$ and $\beta(\delta)$ are type II errors, the detecting powers of Y_{nj} and $\hat{Y}_{nj}$ are $1-\beta_0(\delta)$ and $1-\beta(\delta)$, respectively. Since $\eta_0 > \eta$, we have $\delta < \delta'$, which implies $1-\beta_0(\delta) < 1-\beta(\delta)$. Thus $\hat{Y}_n$ chart has better detecting power than Y_n chart.

The detecting power of Y_n chart, $1 - \beta_0(\delta)$, depends on δ only. However, the detecting power of $\hat{Y}_n$ chart, $1 - \beta(\delta)$, depends on δ, η_0, and η. Since there are many combinations of η_0 and η, it is impossible to list all the cases. For illustration, we consider the case of $p = 1$, the linear case, and various ratios of the component $\sigma_0^2+\sigma_1^2P_1^2(x_n)$ to the component σ_e^2. Because the advantage of the proposed method over the traditional method relies on the condition that the coefficient of σ_e^2 is less than 1, the ratio of the component $\sigma_0^2 + \sigma_1^2P_1(x_n)$ to the component σ_e^2 is important. If the magnitude of $\sigma_0^2 + \sigma_1^2P_1(x_n)$

dominates the variance term, then the enhancement of the detecting power is limited. In Figures 2-3, we demonstrate the enhancement for $n = 10, 20, 30, 40$ and the ratio $R = 4, 3, 2, 1, 2/3, 1/2, 2/5$. From Figures 2-3, we can see that the detecting power becomes larger as the ratio R gets smaller and/or the number of the set points n gets larger.

4 Methodologies: Nonparametric Regression Approach

4.1 Model Assumptions

Without loss of generality, let $\mathcal{T} = [0, 1]$. We assume that $\left\{Y(x),\, x \in \mathcal{T}\right\}$ is a Gaussian stochastic process with mean function $\mu(x)$ and covariance function $G(x, t)$, where $x, t \in \mathcal{T}$. For one profile, the nonparametric regression model is as follows:

$$y_i = Y(x_i) + \varepsilon_i, \quad i = 1, \cdots, n, \tag{19}$$

where ε_i's are i.i.d. as a normal distribution with mean zero and variance σ_e^2. Moreover, assume that $Y(x_i)$ and ε_i are mutually independent.

4.2 Monitoring Statistics and Their Distributions

If we use the Nadaraya-Waston estimator (8) to estimate $Y(x)$, we can get

$$\hat{Y}(x) = \sum_{i=1}^{n} w_i(x)\, y_i, \qquad \text{where } w_i(x) = \frac{K\left(\frac{x_i - x}{h}\right)}{\sum_{k=1}^{n} K\left(\frac{x_k - x}{h}\right)}.$$

It is well known that the Nadaraya-Waston estimator may estimate the boundaries badly, especially when h is large relative to the sampling rate. Since we focus on the boundary point, we must be cautious with the boundary effect. On the other hand, local polynomial smoothing is superior in the region of the boundaries; see, for example, the detailed discussions provided by Fan and Gijbels (1996). To avoid the boundary effect, we adopt the local polynomial smoothing method. By equation (11) with $p = 1$, the local linear

regression estimator $\hat{Y}(x)$ is $\hat{\beta}_0$ and this estimator can be explicitly expressed as

$$\hat{Y}(x) = \sum_{i=1}^{n} w_i(x)\, y_i, \qquad \text{where } w_i(x) = \frac{K\left(\frac{x_i - x}{h}\right)\left[S_{n,2} - (x_i - x)S_{n,1}\right]}{\sum_{k=1}^{n} K\left(\frac{x_k - x}{h}\right)\left[S_{n,2} - (x_k - x)S_{n,1}\right]}, \tag{20}$$

with $S_{n,j}$ as defined in (10). We use this estimator $\hat{Y}(x_n)$ with weight function $w_i(x_n)$ in (20) to obtain the fitted value $\hat{Y}_n$ of the endpoint of the profile.

Proposition 5 *$Y_n = Y(x_n) + \varepsilon_n$ follows a normal distribution with mean $\mu(x_n)$ and variance $G(x_n, x_n) + \sigma_e^2$.*

Proposition 6 *$\hat{Y}_n = \hat{Y}(x_n) = \sum_{i=1}^{n} w_i(x_n) y_i$ follows a normal distribution with mean $\sum_{i=1}^{n} w_i(x_n)\mu(x_i)$ and variance $\sum_{j=1}^{n}\sum_{i=1}^{n} w_i(x_n) w_j(x_n) G(x_i, x_j) + \sum_{i=1}^{n} w_i^2(x_n)\sigma_e^2$.*

4.3 Phase II method

In Phase II monitoring, for simplicity, we assume that the mean function $\mu(x)$ and covariance function $G(x,t)$ of the Gaussian stochastic process are known. Set the overall false-alarm rate at α. We can derive the control limits of Y_n chart and $\hat{Y}_n$ chart as follows:

$$LCL_{Y_n} = \mu(x_n) - Z_{\frac{\alpha}{2}}\eta_0^*, \qquad UCL_{Y_n} = \mu(x_n) + Z_{\frac{\alpha}{2}}\eta_0^*,$$

$$LCL_{\hat{Y}_n} = \sum_{i=1}^{n} w_i(x_n)\mu(x_i) - Z_{\frac{\alpha}{2}}\eta^*, \qquad UCL_{\hat{Y}_n} = \sum_{i=1}^{n} w_i(x_n)\mu(x_i) + Z_{\frac{\alpha}{2}}\eta^*,$$

where

$$\begin{aligned} \eta_0^* &= \left\{ G(x_n, x_n) + \sigma_e^2 \right\}^{1/2}, \\ \eta^* &= \left\{ \sum_{j=1}^{n}\sum_{i=1}^{n} w_i(x_n) w_j(x_n) G(x_i, x_j) + \sum_{i=1}^{n} w_i^2(x_n)\sigma_e^2 \right\}^{1/2}. \end{aligned} \tag{21}$$

In general, the center line $\sum_{i=1}^{n} w_i(x_n)\mu(x_i)$ of $\hat{Y}_n$ chart is different from the center line $\mu(x_n)$ of Y_n chart because the nonparametric method diminishes variance at the cost of getting some bias. However, we find that Y_n and $\hat{Y}_n$ charts have the same center line when the mean function of the Gaussian stochastic process, $\mu(\cdot)$, is a linear function and we use the local linear regression to fit the line.

Proposition 7 *If $\mu(\cdot)$ is a linear function and the local linear regression is used to smooth data, then $\sum_{i=1}^{n} w_i(x_n)\mu(x_i) = \mu(x_n)$.*

Proposition 7 indicates that Y_n and $\hat{Y}_n$ charts have the same center line if $\mu(\cdot)$ is linear. If the mean function of the Gaussian stochastic process is not a linear function, then the difference between $\sum_{i=1}^{n} w_i(x_n)\mu(x_i)$ and $\mu(x_n)$ is of size $O(h)$.

Proposition 8 *If $\mu(\cdot)$ is differentiable and $\mu'(\cdot)$ is bounded by some constant M, then the mean of $\hat{Y}_{nj}$ can be expressed as $\sum_{i=1}^{n} w_i(x_n)\mu(x_i) = \mu(x_n) + O(h)$.*

Unlike the mean of $\hat{Y}_{nj}$ and Y_{nj}, the variances of $\hat{Y}_{nj}$ and Y_{nj} are obviously different. In order to see the difference clearly, we derive the following proposition.

Proposition 9 *If $h \to 0$, $nh \to \infty$, $K(\cdot)$ is differentiable, and $K'(\cdot)$ is bounded by some constant C, then the variance of $\hat{Y}_n$ can be expressed as*

$$Var(\hat{Y}_n) = G(x_n, x_n) + O(h) + \left\{ \frac{1}{nh} \frac{\int_{-1}^{\min\left(1, \frac{1-x_n}{h}\right)} K^2(u)du}{\left[\int_{-1}^{\min\left(1, \frac{1-x_n}{h}\right)} K(u)du\right]^2} + O\left(\frac{1}{n^2h^2}\right) \right\} \sigma_e^2.$$

The two conditions that K is differentiable and $K'(x)$ is bounded by some constant C hold for commonly used kernel functions. From the Proposition 9, we can clearly see the difference between the variances of $\hat{Y}_{nj}$ and Y_{nj}. Recall that the variance of Y_{nj} is $G(x_n, x_n) + \sigma_e^2$. The difference in the variances of $\hat{Y}_{nj}$ and Y_{nj} includes an $O(h)$ term and the coefficients of σ_e^2. Note that the σ_e^2 term converges to zero as $nh \to \infty$. If we add an additional condition $nh^2 \to 0$, then $O(h)$ converges to zero more quickly than $1/nh$. Then, by Proposition 9, the variance of $\hat{Y}_{nj}$ is smaller than that of Y_{nj} asymptotically. When the mean function of the Gaussian stochastic process is a linear function, $\hat{Y}_n$ and Y_n charts have the same center line but different widths of the control limits. The control limits of $\hat{Y}_n$ chart are narrower than that of Y_n chart. Then, the detecting power of $\hat{Y}_n$

chart is better than that of Y_n chart. If the mean function of the Gaussian stochastic process is not a linear function, then the center lines of $\hat{Y}_n$ chart and Y_n chart may be slightly different. However, the bias of the center lines does not affect the result that $\hat{Y}_n$ has bigger detecting power than Y_n has. The detecting powers of $\hat{Y}_n$ and Y_n charts can be calculated as in the following.

When the mean of Y_{nj} shifts with a size of $\delta\eta_0^*$, the probability for Y_{nj} still falling within the control limits is as follows:

$$\begin{aligned}\beta_0^*(\delta) &= P\left(\mu(x_n) - Z_{\frac{\alpha}{2}}\eta_0^* \le Y_{nj} \le \mu(x_n) + Z_{\frac{\alpha}{2}}\eta_0^*\right) \\ &= P\left(-\delta - Z_{\frac{\alpha}{2}} \le \frac{Y_{nj} - \mu(x_n) - \delta\eta_0^*}{\eta_0^*} \le -\delta + Z_{\frac{\alpha}{2}}\right) \\ &= \Phi(-\delta + Z_{\frac{\alpha}{2}}) - \Phi(-\delta - Z_{\frac{\alpha}{2}}).\end{aligned}$$

Since $\beta_0^*(\delta)$ is the type II error for Y_{nj}, the detecting power of Y_{nj} is $1 - \beta_0^*(\delta)$.

If the mean of Y_{nj} shifts with a size of $\delta\eta_0^*$, then the mean of $\hat{Y}_{nj}$ shifts with a size of $\delta'\eta^*$, where $\delta' = \delta\eta_0^*/\eta^*$. Then the probability for $\hat{Y}_{nj}$ falling within the control limits is as follows:

$$\begin{aligned}\beta^*(\delta) &= P\left(\sum_{i=1}^n w_i(x_n)\mu(x_i) - Z_{\frac{\alpha}{2}}\eta^* \le \hat{Y}_{nj} \le \sum_{i=1}^n w_i(x_n)\mu(x_i) + Z_{\frac{\alpha}{2}}\eta^*\right) \\ &= P\left(-\delta' - Z_{\frac{\alpha}{2}} \le \frac{\hat{Y}_{nj} - \sum_{i=1}^n w_i(x_n)\mu(x_i) - \delta'\eta^*}{\eta^*} \le -\delta' + Z_{\frac{\alpha}{2}}\right) \\ &= \Phi(-\delta' + Z_{\frac{\alpha}{2}}) - \Phi(-\delta' - Z_{\frac{\alpha}{2}}).\end{aligned}$$

Since $\beta^*(\delta)$ is the type II error for $\hat{Y}_{nj}$, the detecting power of $\hat{Y}_{nj}$ is $1 - \beta^*(\delta)$.

Similar to the polynomial regression case, δ' is bigger than δ since η^* is smaller than η_0^* for large n. Since $\hat{Y}_{nj}$ shifts a larger size than Y_{nj} does, the detecting power of $\hat{Y}_n$ chart is bigger than that of Y_n chart. To describe the detecting power, we consider various ratios of the component $G(x_n, x_n)$ to the component σ_e^2. In Figures 4-5, we demonstrate the enhancement for $n = 20, 30, 40, 50$ and the ratio $R = 4, 3, 2, 1, 2/3, 1/2, 2/5$. Results are similar to the polynomial regression case.

5 Two Illustrative Examples

5.1 Linear Profile Example– Soda-Bottle Filling

In this subsection, we demonstrate the Soda-Bottle example. Assume the Soda-Bottle filling process takes 5 seconds to fill a bottle with soda and we record the weight of a bottle every 250 milliseconds. Then, we simulate 150 independent records of profiles and each consists of 20 measurements measured at $x_i = i/4$, $i = 1, \cdots, 20$. If the jth bottle, without soda, weights A_{0j} and that the rate of filling a bottle in every time unit is A_{1j}, we can assume the output profiles conform to the random-effect model in equation (3), where $\alpha_0 = 25$, $\alpha_1 = 25$, $\sigma_0^2 = 1$, $\sigma_1^2 = 1$, and $\sigma_e^2 = 4$. The simulated profiles and their fitted profiles are shown in Figure 6. The variance of Y_n is $\sigma_0^2+\sigma_1^2x_n^2+\sigma_e^2 = 30$, where $\sigma_0^2+\sigma_1^2x_n^2 = 26$ and $\sigma_e^2 = 4$; and the variance of $\hat{Y}_n$ is $\sigma_0^2+\sigma_1^2x_n^2+[1/n+(x_n-\bar{x})^2/\sum_{i=1}^n(x_i-\bar{x})^2]\sigma_e^2 = 26.7428571$, where $[1/n+(x_n-\bar{x})^2/\sum_{i=1}^n(x_i-\bar{x})^2]\sigma_e^2 = 0.7428571$. When the mean of the response at the endpoint shifts with a size of $\delta(\sigma_0^2+\sigma_1^2x_n^2+\sigma_e^2)^{1/2}$, we simulate 150 profiles and Figure 7-9 demonstrate the Y_n chart and $\hat{Y}_n$ chart respectively for $\delta = 1, 2, 3$. For $\delta = 1, 2, 3,$, Y_n chart signals $4, 19, 82$ times while $\hat{Y}_n$ signals $6, 31, 92$ times, respectively. $\hat{Y}_n$ chart signals more often than Y_n chart does.

5.2 Nonparametric Example

Assume the profiles are from a Gaussian stochastic process $\left\{Y(x),\, x \in \mathcal{T}\right\}$ with mean function $\mu(x) = 5e^{3x}$ and covariance function $G(x,t)$ defined in (12), where $\mathcal{T} = [0,1]$. Take n measurements at $x_i = (i-0.5)/n$. The ith simulated profile with measurement error is $y_i = Y(x_i) + \varepsilon_i$, where $\varepsilon_i \sim N(0,1), i = 1, \cdots, n$. Fit each simulated profile by the local linear regression estimator given in (20) with bandwidth $h = 0.2$. For $n = 20$, 150 simulated profiles and their fitted profiles are shown in Figure 10. The variance of Y_n is $G(x_n,x_n)+\sigma_e^2$, where $G(x_n,x_n) = 0.3333333$ and $\sigma_e^2 = 1$; and the variance of $\hat{Y}_n$ is $\sum_{i=1}^n\sum_{j=1}^n w_i(x_n)w_j(x_n)G(x_i,x_j)+\sum_{i=1}^n w_i^2(x_n)\sigma_e^2 = 0.3332085+0.7122667$. We can clearly see that $\hat{Y}_n$ and Y_n have about the same mean, but $\hat{Y}_n$ has smaller variance than Y_n does.

When the mean function of $\left\{Y(x),\, x \in \mathcal{T}\right\}$ shifts with a size of $\delta(G(x_n, x_n) + \sigma_e^2)^{1/2}$, the Y_n chart and $\hat{Y}_n$ chart are illustrated in Figures 11-13 for $\delta = 1, 2, 3$ respectively. For $\delta = 1, 2, 3$, Y_n chart signals $4, 22, 67$ times while $\hat{Y}_n$ signals $7, 34, 93$ times, respectively. $\hat{Y}_n$ chart signals more often than Y_n chart does.

The results are for the case of $n = 20$. If $n = 30$, $\sum_{i=1}^{n}\sum_{j=1}^{n} w_i(x_n)w_j(x_n)G(x_i, x_j) = 0.3331576$ and $\sum_{i=1}^{n} w_i^2(x_n)\sigma_e^2 = 0.543526$. We find that when n is larger, the variance of $\hat{Y}_n$ chart becomes smaller. This indicates that the larger the n is, the larger the detecting power $\hat{Y}_n$ chart has.

6 Conclusions

In this study, we focus on Phase II monitoring. Assuming the whole profile data of a product are available, we propose a monitoring scheme different from the traditional method that directly monitors the endpoint of the profile. Our method is to take advantage of the auxiliary information contained in the profile by fitting each profile with a regression function, and estimating the mean response at the endpoint by the fitted value. Take this estimator as the monitoring statistic instead of the original endpoint response. This approach is better than the traditional method since this estimator has smaller variance than the endpoint response when n is not too small. For the profile fitting, both parametric and nonparametric regression methods are considered. We show that better detecting power can be obtained by the new approach. If the profiles can be fitted adequately by polynomials, we will choose the orthogonal polynomial regression to fit these profiles; otherwise, we adopt the nonparametric regression method. How do we determine that the fitting is adequate? We can consider some goodness-of-fit tests. See Seber (2003, Ch4) for more details. In this study, we only consider the Phase II monitoring. Developing $\hat{Y}_n$ chart for Phase I analysis is a potential future research topic.

A Appendix: Proofs

Proof of Proposition 1:

(ii) Since $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$ and $\sum_{i=1}^{n}(x_i - \bar{x})x_i = \sum_{i=1}^{n}(x_i - \bar{x})^2$, we have

$$\begin{aligned}
\hat{\alpha}_{1j} &= \frac{S_{xy(j)}}{S_{xx}} = \frac{1}{S_{xx}}\sum_{i=1}^{n}(x_i - \bar{x})y_{ij} \\
&= \frac{1}{S_{xx}}\sum_{i=1}^{n}(x_i - \bar{x})(A_{0j} + A_{1j}x_i + \varepsilon_{ij}) \\
&= \frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2}[A_{0j}\sum_{i=1}^{n}(x_i - \bar{x}) + A_{1j}\sum_{i=1}^{n}(x_i - \bar{x})x_i + \sum_{i=1}^{n}(x_i - \bar{x})\varepsilon_{ij}] \\
&= A_{1j} + \frac{\sum_{i=1}^{n}(x_i - \bar{x})\varepsilon_{ij}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.
\end{aligned}$$

Thus,

$$E(\hat{\alpha}_{1j}) = \alpha_1 \text{ and}$$
$$Var(\hat{\alpha}_{1j}) = \sigma_1^2 + \frac{\sigma_e^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Therefore, $\hat{\alpha}_{1j} \sim N(\alpha_1, \sigma_1^2 + \frac{\sigma_e^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2})$.

(i) Since $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$ and $\sum_{i=1}^{n}(x_i - \bar{x})x_i = \sum_{i=1}^{n}(x_i - \bar{x})^2$, we have

$$\begin{aligned}
\hat{\alpha}_{0j} &= \bar{y}_j - \hat{\alpha}_{1j}\bar{x} \\
&= \frac{1}{n}\sum_{i=1}^{n} y_{ij} - (A_{1j} + \frac{\sum_{i=1}^{n}(x_i - \bar{x})\varepsilon_{ij}}{\sum_{i=1}^{n}(x_i - \bar{x})^2})\bar{x} \\
&= \frac{1}{n}\sum_{i=1}^{n}(A_{0j} + A_{1j}x_i + \varepsilon_{ij}) - (A_{1j} + \frac{\sum_{i=1}^{n}(x_i - \bar{x})\varepsilon_{ij}}{\sum_{i=1}^{n}(x_i - \bar{x})^2})\bar{x} \\
&= A_{0j} + \frac{1}{n}\sum_{i=1}^{n}\varepsilon_{ij} - \frac{\sum_{i=1}^{n}(x_i - \bar{x})\varepsilon_{ij}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\bar{x}.
\end{aligned}$$

Since $Cov\left(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_{ij}, \frac{\sum_{i=1}^{n}(x_i-\bar{x})\varepsilon_{ij}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)=0$, we have

$$\begin{aligned}
&E(\hat{\alpha}_{0j})=\alpha_0 \ \text{ and}\\
&Var(\hat{\alpha}_{0j})=\sigma_0^2+\frac{\sigma_e^2}{n}+\frac{\bar{x}^2\sigma_e^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}-2\bar{x}\,Cov\left(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_{ij}, \frac{\sum_{i=1}^{n}(x_i-\bar{x})\varepsilon_{ij}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)\\
&\qquad\qquad=\sigma_0^2+\frac{\sigma_e^2}{n}+\frac{\bar{x}^2\sigma_e^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}.
\end{aligned}$$

Therefore, $\hat{\alpha}_{0j}\sim N\left(\alpha_0, \sigma_0^2+(\frac{1}{n}+\frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2})\sigma_e^2\right)$.

(iii)

$$\begin{aligned}
Cov(\hat{\alpha}_{0j},\hat{\alpha}_{1j}) &= Cov(A_{0j}+\frac{1}{n}\sum_{i=1}^{n}\varepsilon_{ij}-\frac{\sum_{i=1}^{n}(x_i-\bar{x})\varepsilon_{ij}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\bar{x},\ A_{1j}+\frac{\sum_{i=1}^{n}(x_i-\bar{x})\varepsilon_{ij}}{\sum_{i=1}^{n}(x_i-\bar{x})^2})\\
&= Cov(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_{ij},\ \frac{\sum_{i=1}^{n}(x_i-\bar{x})\varepsilon_{ij}}{\sum_{i=1}^{n}(x_i-\bar{x})^2})-Cov(\frac{\sum_{i=1}^{n}(x_i-\bar{x})\varepsilon_{ij}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\bar{x},\ \frac{\sum_{i=1}^{n}(x_i-\bar{x})\varepsilon_{ij}}{\sum_{i=1}^{n}(x_i-\bar{x})^2})\\
&= -\bar{x}\,Var(\frac{\sum_{i=1}^{n}(x_i-\bar{x})\varepsilon_{ij}}{\sum_{i=1}^{n}(x_i-\bar{x})^2})\\
&= -\frac{\bar{x}\sigma_e^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}.
\end{aligned}$$

Proof of Proposition 2:

(1) Since $\sum_{i=1}^{n}P_r(x_i)=0,\ r=1,\cdots,p$, we have

$$\begin{aligned}
\hat{\alpha}_{0j}=\frac{\sum_{i=1}^{n}P_0(x_i)y_{ij}}{\sum_{i=1}^{n}P_0^2(x_i)} &= \frac{1}{n}\sum_{i=1}^{n}\left[A_{0j}+A_{1j}P_1(x_i)+\cdots+A_{pj}P_p(x_i)+\varepsilon_{ij}\right]\\
&= \frac{1}{n}\left[nA_{0j}+A_{1j}\sum_{i=1}^{n}P_1(x_i)+\cdots+A_{pj}\sum_{i=1}^{n}P_p(x_i)+\sum_{i=1}^{n}\varepsilon_{ij}\right]\\
&= A_{0j}+\frac{1}{n}\sum_{i=1}^{n}\varepsilon_{ij}\ .
\end{aligned}$$

Thus,

$$\begin{aligned}
&E(\hat{\alpha}_{0j})=E\left(A_{0j}+\frac{1}{n}\sum_{i=1}^{n}\varepsilon_{ij}\right)=\alpha_0 \ \text{ and}\\
&Var(\hat{\alpha}_{0j})=Var\left(A_{0j}+\frac{1}{n}\sum_{i=1}^{n}\varepsilon_{ij}\right)=\sigma_0^2+\frac{\sigma_e^2}{n}.
\end{aligned}$$

Therefore, $\hat{\alpha}_{0j} \sim N(\alpha_0, \sigma_0^2 + \frac{\sigma_e^2}{n})$.

(2) Since $\sum_{i=1}^{n} P_r(x_i) = 0$ and $\sum_{i=1}^{n} P_r(x_i)P_s(x_i) = 0$, for $r \neq s$, $r, s = 1, \cdots, p$, we have

$$\begin{aligned}
\hat{\alpha}_{rj} &= \frac{\sum_{i=1}^{n} P_r(x_i) y_{ij}}{\sum_{i=1}^{n} P_r^2(x_i)} \\
&= \frac{1}{\sum_{i=1}^{n} P_r^2(x_i)} \sum_{i=1}^{n} P_r(x_i) \left[A_{0j} + A_{1j}P_1(x_i) + \cdots + A_{rj}P_r(x_i) + \cdots + A_{pj}P_p(x_i) + \varepsilon_{ij} \right] \\
&= \frac{1}{\sum_{i=1}^{n} P_r^2(x_i)} \left(A_{0j} \sum_{i=1}^{n} P_r(x_i) + A_{1j} \sum_{i=1}^{n} P_r(x_i)P_1(x_i) + \cdots + \right. \\
&\qquad \left. A_{rj} \sum_{i=1}^{n} P_r^2(x_i) + \cdots + A_{pj} \sum_{i=1}^{n} P_r(x_i)P_p(x_i) + \sum_{i=1}^{n} P_r(x_i)\varepsilon_{ij} \right) \\
&= A_{rj} + \frac{\sum_{i=1}^{n} P_r(x_i)\varepsilon_{ij}}{\sum_{i=1}^{n} P_r^2(x_i)}.
\end{aligned}$$

Thus, $E(\hat{\alpha}_{rj}) = E\left(A_{rj} + \frac{\sum_{i=1}^{n} P_r(x_i)\varepsilon_{ij}}{\sum_{i=1}^{n} P_r^2(x_i)} \right) = \alpha_r$ and

$$Var(\hat{\alpha}_{rj}) = Var\left(A_{rj} + \frac{\sum_{i=1}^{n} P_r(x_i)\varepsilon_{ij}}{\sum_{i=1}^{n} P_r^2(x_i)} \right) = \sigma_r^2 + \frac{\sigma_e^2 \sum_{i=1}^{n} P_r^2(x_i)}{\left[\sum_{i=1}^{n} P_r^2(x_i) \right]^2} = \sigma_r^2 + \frac{\sigma_e^2}{\sum_{i=1}^{n} P_r^2(x_i)}$$

Therefore, $\hat{\alpha}_{rj} \sim N(\alpha_r, \sigma_r^2 + \frac{\sigma_e^2}{\sum_{i=1}^{n} P_r^2(x_i)})$.

To show that $\hat{\alpha}_r$ and $\hat{\alpha}_s$ are independent, for $r \neq s$, $r, s = 0, \cdots, p$,

$$Cov(\,\hat{\alpha}_{rj},\hat{\alpha}_{sj}\,) = Cov(\,A_{rj} + \frac{\sum_{i=1}^{n} P_r(x_i)\varepsilon_{ij}}{\sum_{i=1}^{n} P_r^2(x_i)}, A_{sj} + \frac{\sum_{i=1}^{n} P_s(x_i)\varepsilon_{ij}}{\sum_{i=1}^{n} P_s^2(x_i)})$$

$$= Cov(\,A_{rj}, A_{sj}\,) + Cov(\,A_{rj}, \frac{\sum_{i=1}^{n} P_s(x_i)\varepsilon_{ij}}{\sum_{i=1}^{n} P_s^2(x_i)}\,)$$

$$+Cov(\,\frac{\sum_{i=1}^{n} P_r(x_i)\varepsilon_{ij}}{\sum_{1}^{n} P_r^2(x_i)}, A_{sj}) + Cov(\,\frac{\sum_{i=1}^{n} P_r(x_i)\varepsilon_{ij}}{\sum_{i=1}^{n} P_r^2(x_i)}, \frac{\sum_{i=1}^{n} P_s(x_i)\varepsilon_{ij}}{\sum_{i=1}^{n} P_s^2(x_i)}\,)$$

$$= Cov(\,\frac{\sum_{i=1}^{n} P_r(x_i)\varepsilon_{ij}}{\sum_{i=1}^{n} P_r^2(x_i)}, \frac{\sum_{i=1}^{n} P_s(x_i)\varepsilon_{ij}}{\sum_{i=1}^{n} P_s^2(x_i)}\,)$$

$$= Cov(\,\frac{P_r(x_1)\varepsilon_{1j}}{\sum_{i=1}^{n} P_r^2(x_i)}, \frac{\sum_{i=1}^{n} P_s(x_i)\varepsilon_{ij}}{\sum_{i=1}^{n} P_s^2(x_i)}) + \cdots + Cov(\,\frac{P_r(x_n)\varepsilon_{nj}}{\sum_{i=1}^{n} P_r^2(x_i)}, \frac{\sum_{i=1}^{n} P_s(x_i)\varepsilon_{ij}}{\sum_{i=1}^{n} P_s^2(x_i)}\,)$$

$$= \frac{P_r(x_1)P_s(x_1)Var(\varepsilon_{1j})}{\sum_{i=1}^{n} P_r^2(x_i)\sum_{i=1}^{n} P_s^2(x_i)} + \cdots + \frac{P_r(x_n)P_s(x_n)Var(\varepsilon_{nj})}{\sum_{i=1}^{n} P_r^2(x_i)\sum_{i=1}^{n} P_s^2(x_i)}$$

$$= \frac{\sum_{i=1}^{n} P_r(x_i)P_s(x_i)\sigma_e^2}{\sum_{i=1}^{n} P_r^2(x_i)\sum_{i=1}^{n} P_s^2(x_i)}$$

$$= 0.$$

Proof of Proposition 3:

Trivial.

Proof of Proposition 4:

$$
\begin{aligned}
E(\hat{Y}_{nj}) &= E(\ \hat{\alpha}_{0j} + \hat{\alpha}_{1j}P_1(x_n) + \cdots + \hat{\alpha}_{pj}P_p(x_n)\ ) \\
&= E(\hat{\alpha}_{0j}) + E(\hat{\alpha}_{1j}P_1(x_n)) + \cdots + E(\hat{\alpha}_{pj}P_p(x_n)) \\
&= \alpha_0 + \alpha_1 P_1(x_n) + \cdots + \alpha_p P_p(x_n).
\end{aligned}
$$

$$
\begin{aligned}
Var(\hat{Y}_{nj}) &= Var(\ \hat{\alpha}_{0j} + \hat{\alpha}_{1j}P_1(x_n) + \cdots + \hat{\alpha}_{pj}P_p(x_n)\ ) \\
&= (\sigma_0^2 + \tfrac{\sigma_e^2}{n}) + (\sigma_1^2 + \tfrac{\sigma_e^2}{\sum_{i=1}^n P_1^2(x_i)})P_1^2(x_n) + \cdots + (\sigma_p^2 + \tfrac{\sigma_e^2}{\sum_{i=1}^n P_p^2(x_i)})P_p^2(x_n) \\
&= \sigma_0^2 + \sigma_1^2 P_1^2(x_n) + \cdots + \sigma_p^2 P_p^2(x_n) + \left[\frac{1}{n} + \frac{P_1^2(x_n)}{\sum_{i=1}^n P_1^2(x_i)} + \cdots + \frac{P_p^2(x_n)}{\sum_{i=1}^n P_p^2(x_i)}\right]\sigma_e^2 \ \equiv\ \eta^2.
\end{aligned}
$$

Thus, $\hat{Y}_{nj}$ are i.i.d. as $N(\ \alpha_0 + \alpha_1 P_1(x_n) + \cdots + \alpha_p P_p(x_n), \eta^2\ )$.

Proof of Proposition 5:

$E(\ Y_n\ ) = E(\ Y(x_n) + \varepsilon_n\ ) = \mu(x_n)$ and $Var(\ Y_n\ ) = Var(\ Y(x_n) + \varepsilon_n\ ) = G(x_n, x_n) + \sigma_e^2$.

Thus, $Y_n \sim N\left(\mu(x_n)\ ,\ G(x_n, x_n) + \sigma_e^2\right)$.

Proof of Proposition 6:

$$
\begin{aligned}
E(\hat{Y}_n) &= E\left(\sum_{i=1}^n w_i(x_n) y_i\right) = \sum_{i=1}^n w_i(x_n) E\left(Y(x_i) + \varepsilon_i\right) = \sum_{i=1}^n w_i(x_n)\mu(x_i). \\
Var(\hat{Y}_n) &= Var\left(\sum_{i=1}^n w_i(x_n) y_i\right) = Var\left(\sum_{i=1}^n w_i(x_n) Y(x_i) + \sum_{i=1}^n w_i(x_n)\varepsilon_i\right) \\
&= \sum_{j=1}^n \sum_{i=1}^n w_i(x_n) w_j(x_n) G(x_i, x_j) + \sum_{i=1}^n w_i^2(x_n)\sigma_e^2.
\end{aligned}
$$

Thus, $\hat{Y}_n \sim N\left(\sum_{i=1}^n w_i(x_n)\mu(x_i)\ ,\ \sum_{j=1}^n \sum_{i=1}^n w_i(x_n) w_j(x_n) G(x_i, x_j) + \sum_{i=1}^n w_i^2(x_n)\sigma_e^2\right)$.

Proof of Proposition 7:

Let $\ell_i = K_h(x_i - x_n)\left[S_{n,2} - (x_i - x_n)S_{n,1}\right]$.

Then, by (20),

$$
\sum_{i=1}^n w_i(x_n)\mu(x_i) = \frac{\sum_{i=1}^n \ell_i \mu(x_i)}{\sum_{i=1}^n \ell_i}.
$$

Assume that $\mu(x_i) = a + bx_i$, where a, b are real numbers. Then,

$$\frac{\sum_{i=1}^{n} \ell_i \mu(x_i)}{\sum_{i=1}^{n} \ell_i} = \frac{\sum_{i=1}^{n} \ell_i (a + bx_i)}{\sum_{i=1}^{n} \ell_i} = a + b\frac{\sum_{i=1}^{n} \ell_i x_i}{\sum_{i=1}^{n} \ell_i}.$$

We can show that $\sum_{i=1}^{n} \ell_i x_i = x_n \sum_{i=1}^{n} \ell_i$ by the following:

$$\begin{aligned}
\sum_{i=1}^{n} \ell_i x_i &= \sum_{i=1}^{n} K_h(x_i - x_n) \left[S_{n,2} - (x_i - x_n) S_{n,1} \right] x_i \\
&= S_{n,2} \sum_{i=1}^{n} K_h(x_i - x_n) x_i - S_{n,1} \sum_{i=1}^{n} K_h(x_i - x_n)(x_i - x_n) x_i \\
&= S_{n,2} \sum_{i=1}^{n} K_h(x_i - x_n)(x_i - x_n) - S_{n,1} \sum_{i=1}^{n} K_h(x_i - x_n)(x_i - x_n)(x_i - x_n) \\
&\quad + S_{n,2} \sum_{i=1}^{n} K_h(x_i - x_n) x_n - S_{n,1} \sum_{i=1}^{n} K_h(x_i - x_n)(x_i - x_n) x_n \\
&= S_{n,2} S_{n,1} - S_{n,1} S_{n,2} + x_n S_{n,2} S_{n,0} - x_n S_{n,1} S_{n,1} \\
&= x_n (S_{n,2} S_{n,0} - S_{n,1} S_{n,1}) \\
&= x_n \sum_{i=1}^{n} \ell_i \ .
\end{aligned}$$

The last equality is due to

$$\begin{aligned}
\sum_{i=1}^{n} \ell_i &= \sum_{i=1}^{n} K_h(x_i - x_n) \{ S_{n,2} - (x_i - x_n) S_{n,1} \} \\
&= \sum_{i=1}^{n} K_h(x_i - x_n) S_{n,2} - \sum_{i=1}^{n} K_h(x_i - x_n)(x_i - x_n) S_{n,1} \\
&= S_{n,0} S_{n,2} - S_{n,1}^2 \ .
\end{aligned}$$

Then, we get

$$\sum_{i=1}^{n} w_i(x_n) \mu(x_i) = a + bx_n = \mu(x_n).$$

The Mean Value Theorem for Integrals: *If f is a continuous function on interval I, then there exists a number $\xi \in I$ such that $\int_I f(x)dx = f(\xi)|I|$.*

Lipschitz Condition 1: *Let f be a real-valued function defined on a subset D of real numbers. We say that $f : D \subseteq R \to R$ satisfies a Lipschitz condition if there exists a*

constant $C \geq 0$ such that $|f(t) - f(x)| \leq C|t - x|$ for all t, x in D. Here, the smallest such C is called the Lipschitz constant of the function f.

Corollary 1: *If f is differentiable and $|f'(x)| \leq C$ for all $x \in (a, b)$, then f satisfies the Lipschitz condition*

$$|f(t) - f(x)| \leq C|t - x| \text{ for all } t, x \in (a, b),$$

where C is a constant.

Proof of Proposition 8:

For s in a small neighborhood of x_n, by Taylor's Series expansion, we have

$$\mu(s) = \mu(x_n) + (s - x_n)\frac{d\mu(s)}{ds}\bigg|_{s=x_n} + O(s - x_n).$$

For $|s - x_n| < h$, we have

$$|\mu(s) - \mu(x_n)| = O(h).$$

Since $\sum_{i=1}^{n} w_i(x_n) = 1$, we have

$$\sum_{i=1}^{n} w_i(x_n)[\mu(x_i) - \mu(x_n)] = O(h).$$

Lipschitz Condition 2: *Let f be a continuous and defined in a neighborhood of $(t_0, x_0) \in R^2$. We say that $f : R^2 \to R$ satisfies a Lipschitz condition if there exists a constant $C \geq 0$ such that $|f(t, x_1) - f(t, x_2)| \leq C|x_1 - x_2|$ for all (t, x_1) and (t, x_2) in a neighborhood of (t_0, x_0).*

Corollary 2: *If f is differentiable and $\left|\frac{\partial f(t,x)}{\partial x}\right|$ is continuous, then Lipschitz Condition 2 is automatic.*

For the above conditions and corollaries, see, for example, Marsden (1993) or Pugh (2002).

Lemma 1 *If G is differentiable in t and $\left|\frac{\partial G(s,t)}{\partial t}\right|$ is continuous, $|s-x_n|<h$, and $|t-x_n|<h$, then $\left|G(s,t)-G(x_n,x_n)\right|=O(h)$, where s,t, and $x_n\in[0,1]$.*

Proof:

For s,t in a small neighborhood of x_n, say $B(h)$, by Taylor's Series expansion, we have

$$G(s,t)=G(x_n,x_n)+(s-x_n)\left.\frac{\partial G(s,t)}{\partial s}\right|_{(s,t)=(x_n,x_n)}+(t-x_n)\left.\frac{\partial G(s,t)}{\partial t}\right|_{(s,t)=(x_n,x_n)}+R,$$

where R is the remainder. For $|s-x_n|<h$ and $|t-x_n|<h$, we have

$$|G(s,t)-G(x_n,x_n)|=O(h).$$

Lemma 2: *If K is differentiable on $(0,1)$ and $|K'(x)|$ is bounded by some constant C, then*

$$\frac{1}{n}\sum_{i=1}^{n}K\left(\frac{x_i-x_n}{h}\right)=\int_0^1 K\left(\frac{x-x_n}{h}\right)dx\ +O\left(\frac{1}{n}\right).$$

Proof:

Let $x_i=\dfrac{i-0.5}{n}$ and $I_i=\left[x_i-\frac{1}{2n}\ ,\ x_i+\frac{1}{2n}\right]$, $i=1,\cdots,n$. We calculate the absolute value $\left|\dfrac{1}{n}\sum_{i=1}^{n}K\left(\dfrac{x_i-x_n}{h}\right)-\int_0^1 K\left(\dfrac{x-x_n}{h}\right)dx\right|$ to see the difference. Because K is a continuous function, there exists a point $\xi_i\in I_i$ such that $\int_{I_i}K\left(\dfrac{x-x_n}{h}\right)dx=K\left(\dfrac{\xi_i-x_n}{h}\right)|I_i|$ for $i=1,\cdots,n$ by the Mean Value Theorem for Integrals. Since each $\mid I_j\mid=\dfrac{1}{n}$, the above difference can be rewritten as follows: $\left|\dfrac{1}{n}\sum_{i=1}^{n}K\left(\dfrac{x_i-x_n}{h}\right)-\dfrac{1}{n}\sum_{i=1}^{n}K\left(\dfrac{\xi_i-x_n}{h}\right)\right|$. Because K is a kernel function, $K(\frac{z-x_n}{h})$ is zero for the z-values outside the boundary interval $B(h)=[x_n-h,min(x_n+h,1)]$. Then

$$\begin{aligned}&\left|\frac{1}{n}\sum_{i=1}^{n}K\left(\frac{x_i-x_n}{h}\right)-\frac{1}{n}\sum_{i=1}^{n}K\left(\frac{\xi_i-x_n}{h}\right)\right|\\&=\frac{1}{n}\left|\sum_{x_i\in B(h)}K\left(\frac{x_i-x_n}{h}\right)-\sum_{x_i\in B(h)}K\left(\frac{\xi_i-x_n}{h}\right)\right|\\&\le\frac{1}{n}\sum_{x_i\in B(h)}\left|K\left(\frac{x_i-x_n}{h}\right)-K\left(\frac{\xi_i-x_n}{h}\right)\right|.\end{aligned}$$

By the given conditions, K satisfies the Lipschitz condition 1:

$$\left|K\left(\frac{x_i-x_n}{h}\right)-K\left(\frac{\xi_i-x_n}{h}\right)\right|\leq C\left|\frac{x_i-x_n}{h}-\frac{\xi_i-x_n}{h}\right|=C\left|\frac{x_i-\xi_i}{h}\right|\leq\frac{C}{nh}.$$

It follows that

$$\frac{1}{n}\sum_{x_i\in B(h)}\left|K\left(\frac{x_i-x_n}{h}\right)-K\left(\frac{\xi_i-x_n}{h}\right)\right|\leq\frac{1}{n}\lfloor nh\rfloor\frac{C}{nh}\ .$$

Thus, we have

$$\left|\frac{1}{n}\sum_{i=1}^{n}K\left(\frac{x_i-x_n}{h}\right)-\int_0^1K\left(\frac{x-x_n}{h}\right)dx\right|=O\left(\frac{1}{n}\right).$$

Proof of Proposition 9:

From Proposition 6, we know that

$$Var(\hat{Y}_n)=\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{K\left(\frac{x_i-x_n}{h}\right)}{\sum_{k=1}^{n}K\left(\frac{x_k-x_n}{h}\right)}\frac{K\left(\frac{x_j-x_n}{h}\right)}{\sum_{k=1}^{n}K\left(\frac{x_k-x_n}{h}\right)}G(x_i,x_j)+\sum_{i=1}^{n}\left(\frac{K\left(\frac{x_i-x_n}{h}\right)}{\sum_{i=1}^{n}K\left(\frac{x_i-x_n}{h}\right)}\right)^2\sigma_e^2.$$

Note that

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{K\left(\frac{x_i-x_n}{h}\right)}{\sum_{k=1}^{n}K\left(\frac{x_k-x_n}{h}\right)}\frac{K\left(\frac{x_j-x_n}{h}\right)}{\sum_{k=1}^{n}K\left(\frac{x_k-x_n}{h}\right)}=\frac{\sum_{i=1}^{n}K\left(\frac{x_i-x_n}{h}\right)\sum_{j=1}^{n}K\left(\frac{x_j-x_n}{h}\right)}{\left[\sum_{k=1}^{n}K\left(\frac{x_k-x_n}{h}\right)\right]^2}=1.$$

By Lemma 1,

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{K\left(\frac{x_i-x_n}{h}\right)}{\sum_{k=1}^{n}K\left(\frac{x_k-x_n}{h}\right)}\frac{K\left(\frac{x_j-x_n}{h}\right)}{\sum_{k=1}^{n}K\left(\frac{x_k-x_n}{h}\right)}[G(x_i,x_j)-G(x_n,x_n)]=O(h).$$

Thus,

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{K\left(\frac{x_i-x_n}{h}\right)}{\sum_{k=1}^{n}K\left(\frac{x_k-x_n}{h}\right)}\frac{K\left(\frac{x_j-x_n}{h}\right)}{\sum_{k=1}^{n}K\left(\frac{x_k-x_n}{h}\right)}G(x_i,x_j)=G(x_n,x_n)+O(h).$$

Using Lemma 2,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x_i - x_n}{h}\right) &= \int_0^1 K\left(\frac{x - x_n}{h}\right) dx \ + O\left(\frac{1}{n}\right) \\ &= \int_{x_n - h}^{\min(x_n + h, 1)} K\left(\frac{x - x_n}{h}\right) dx \ + O\left(\frac{1}{n}\right) \\ &= h\int_{-1}^{\min\left(1, \frac{1-x_n}{h}\right)} K(u)du + O\left(\frac{1}{n}\right) \quad ,\end{aligned}$$

then, the second term in $Var(\hat{Y}_n)$ can be re-expressed as follows:

$$\begin{aligned}\sum_{i=1}^{n}\left(\frac{K\left(\frac{x_i - x_n}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x_i - x_n}{h}\right)}\right)^2 \sigma_e^2 &= \frac{1}{\left[\frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x_i - x_n}{h}\right)\right]^2}\sum_{i=1}^{n}\left[\frac{1}{n}K\left(\frac{x_i - x_n}{h}\right)\right]^2\sigma_e^2 \\ &= \frac{1}{\left[h\int_{-1}^{\min\left(1, \frac{1-x_n}{h}\right)} K(u)du + O\left(\frac{1}{n}\right)\right]^2}\left(n^{-1}\right)\left[\frac{1}{n}\sum_{i=1}^{n} K^2\left(\frac{x_i - x_n}{h}\right)\right]\sigma_e^2 \\ &= \frac{n^{-1}\left[h\int_{-1}^{\min\left(1, \frac{1-x_n}{h}\right)} K^2(u)du + O\left(\frac{1}{n}\right)\right]}{\left[h\int_{-1}^{\min\left(1, \frac{1-x_n}{h}\right)} K(u)du\right]^2}\sigma_e^2 \\ &= \left\{\frac{1}{nh}\frac{\int_{-1}^{\min\left(1, \frac{1-x_n}{h}\right)} K^2(u)du}{\left[\int_{-1}^{\min\left(1, \frac{1-x_n}{h}\right)} K(u)du\right]^2} + O\left(\frac{1}{n^2h^2}\right)\right\}\sigma_e^2.\end{aligned}$$

Finally, we re-express $Var(\hat{Y}_n)$ by

$$Var(\hat{Y}_n) = G(x_n, x_n) + O(h) + \left\{\frac{1}{nh}\frac{\int_{-1}^{\min\left(1, \frac{1-x_n}{h}\right)} K^2(u)du}{\left[\int_{-1}^{\min\left(1, \frac{1-x_n}{h}\right)} K(u)du\right]^2} + O\left(\frac{1}{n^2h^2}\right)\right\}\sigma_e^2.$$

References

[1] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* London: Chapman & Hall.

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

[3] Kim, K., Mahmoud, M. A., and Woodall, W. H. (2003). "On The Monitoring of Linear Profiles". *Journal of Quality Technology* 35, 317-327.

[4] Marsden, J. E. (1993). *Elementary Classical Analysis.* New York: W.H. Freeman and Company.

[5] Montgomery, D. C., Peck, E. A., and Vining, G. G. (2006). *Introduction to Linear Regression Analysis.* 4th ed. New York: Wiley.

[6] Myers, R. H. (1990). *Classical and Modern Regression with Applications.* 2nd ed. Boston: PWS-KENT Publishing Company.

[7] Pearson, E. S. and Hartley, H. O. (1966). *Biometrika Tables for Statisticians, Volume one.* 3rd ed. London: Cambridge University Press.

[8] Pugh, C. C. (2002). *Real Mathematical Analysis.* New York: Springer.

[9] Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis.* 2nd ed. New York: Springer

[10] Schimek, M. G. (2000). *Smoothing and Regression: Approaches, Computation, and Application.* New York: John Wiley & Sons.

[11] Seber, L. (2003). *Linear Regression Analysis.* 2nd ed. New York: Wiley-Interscience.

[12] Shiau, J.-J. H., Lin, S.-H., and Chen, Y.-C. (2006). "Monitoring Linear Profiles Based on a Random-effect Model". Technicla Report. Institute of statistics, National Chiao Tung University.

[13] Simonoff , J. S. (1996). *Smoothing Methods in Statistics.* New York: Springer.

[14] Wahba, G. (1990). *Spline Models for Observational Data.* Philadelphia, Pennsylvania: SIAM.

C_1	C_2	C_3	C_4	C_5
1.00000	1.00000	1.00000	0.95000	0.88571
C_6	C_7	C_8	C_9	C_{10}
0.82143	0.76190	0.70833	0.66061	0.61819
C_{11}	C_{12}	C_{13}	C_{14}	C_{15}
0.58042	0.54670	0.51648	0.48929	0.46471
C_{16}	C_{17}	C_{18}	C_{19}	C_{20}
0.44240	0.42208	0.40351	0.38647	0.37078
C_{21}	C_{22}	C_{23}	C_{24}	C_{25}
0.35623	0.34289	0.33043	0.31885	0.30803
C_{26}	C_{27}	C_{28}	C_{29}	C_{30}
0.29792	0.28845	0.27956	0.27119	0.26331

Table 1: C_n values for $n = 1, \cdots, 30$

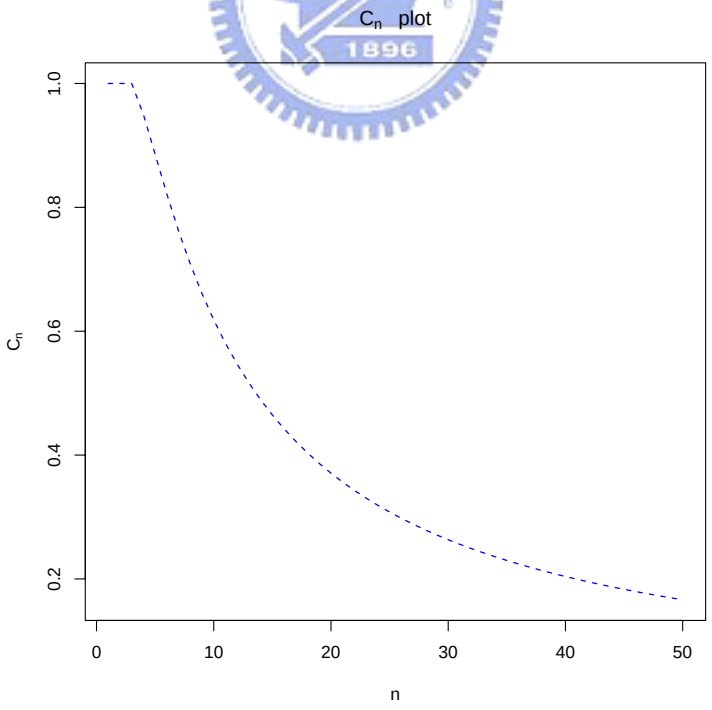


Figure 1: C_n plot for $n = 1, \cdots, 50$

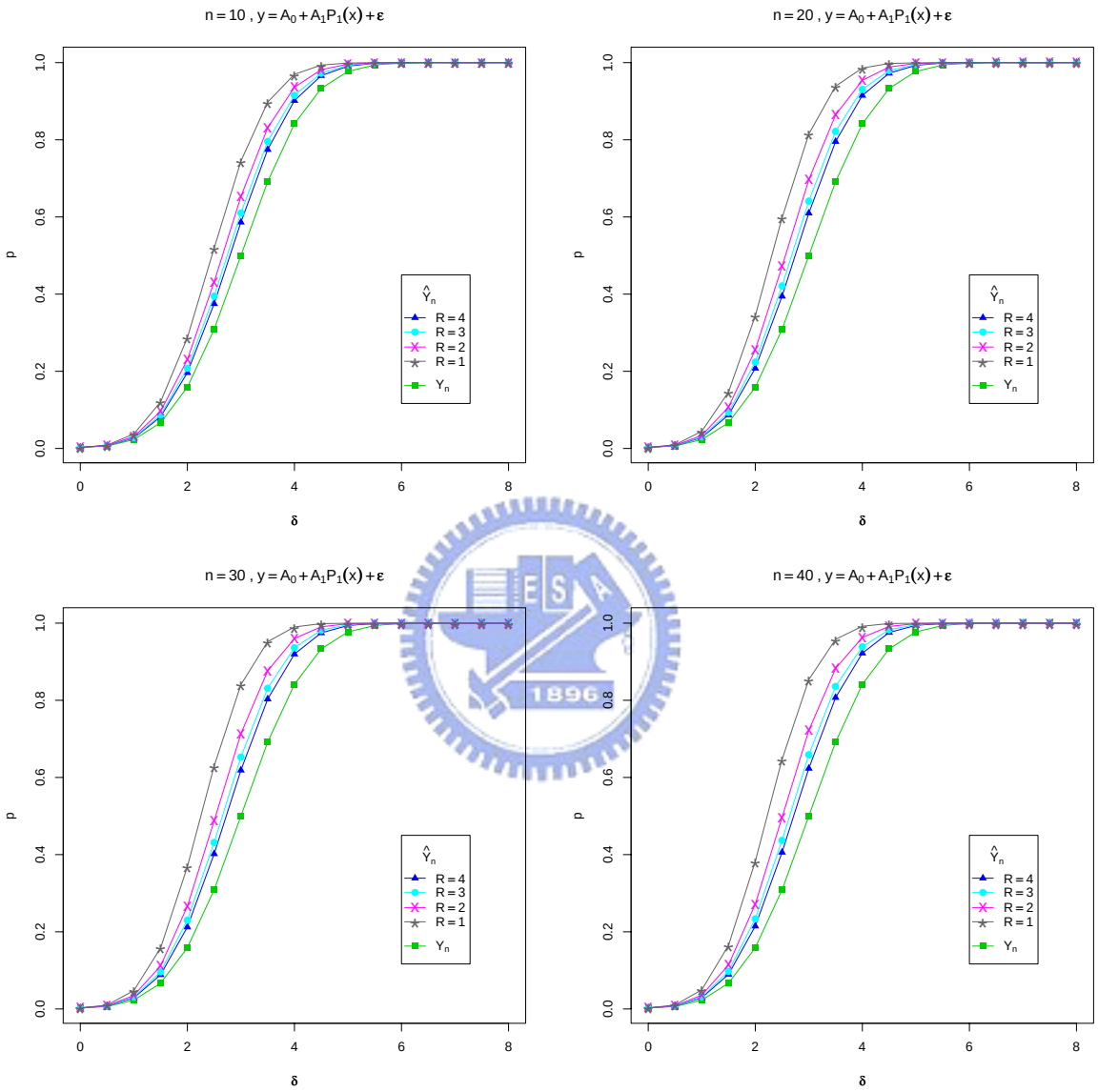


Figure 2: Detecting power of $\hat{Y}_n$ and Y_n charts, when linear regression is used to fit the liner profile. The ratio $R = (\,\sigma_0^2 + \sigma_1^2 P_1^2(x_n)\,)/\sigma_e^2 \geq 1$.

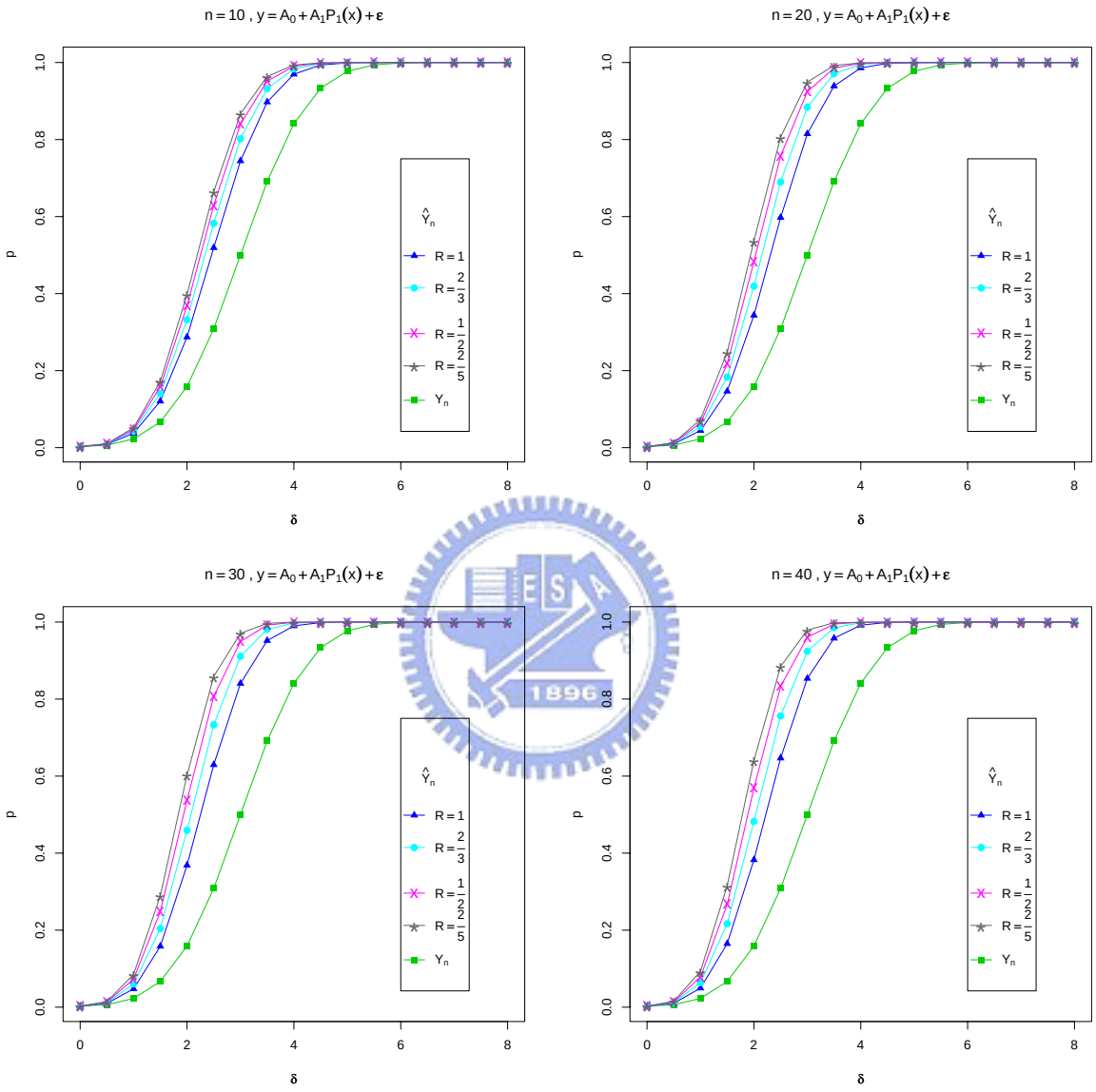


Figure 3: Detecting power of $\hat{Y}_n$ and Y_n charts, when linear regression is used to fit the liner profile. The ratio $R = (\sigma_0^2 + \sigma_1^2 P_1^2(x_n))/\sigma_e^2 \leq 1$.

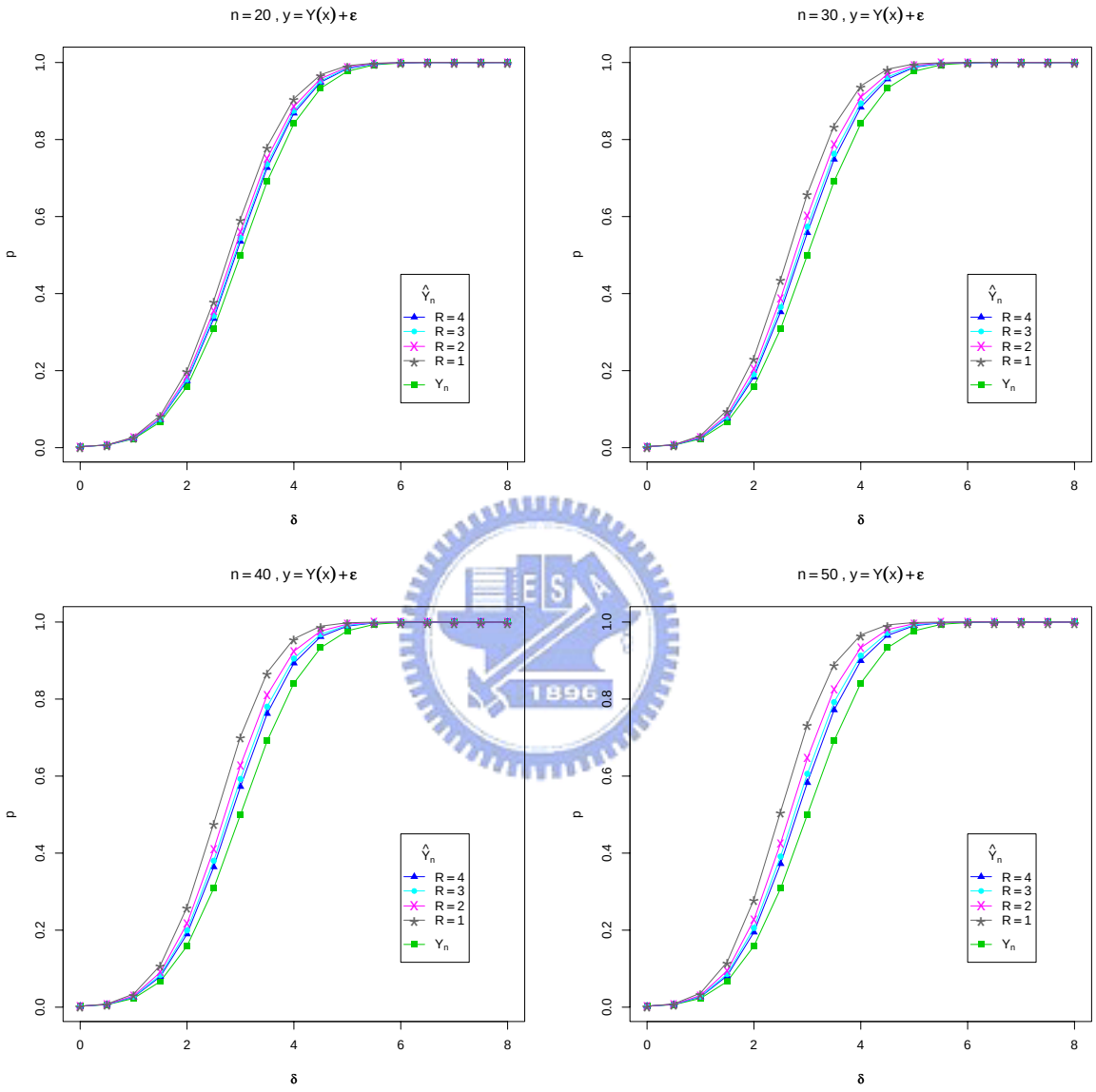


Figure 4: Detecting power of $\hat{Y}_n$ and Y_n charts, when linear regression is used to fit the profile. The ratio $R = G(x_n, x_n)/\sigma_e^2 \geq 1$.

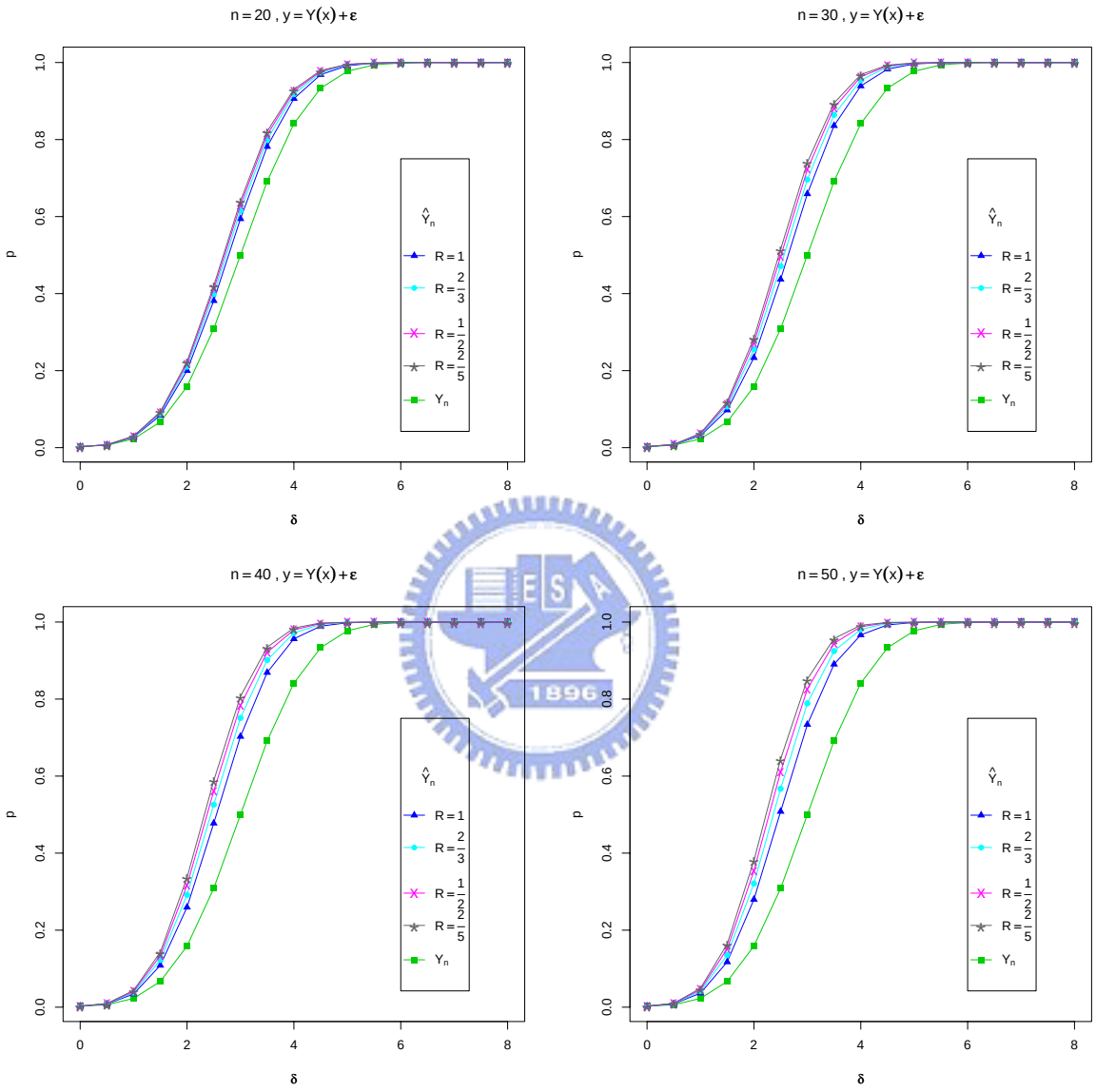


Figure 5: Detecting power of $\hat{Y}_n$ and Y_n charts, when linear regression is used to fit the profile. The ratio $R = G(x_n, x_n)/\sigma_e^2 \leq 1$.

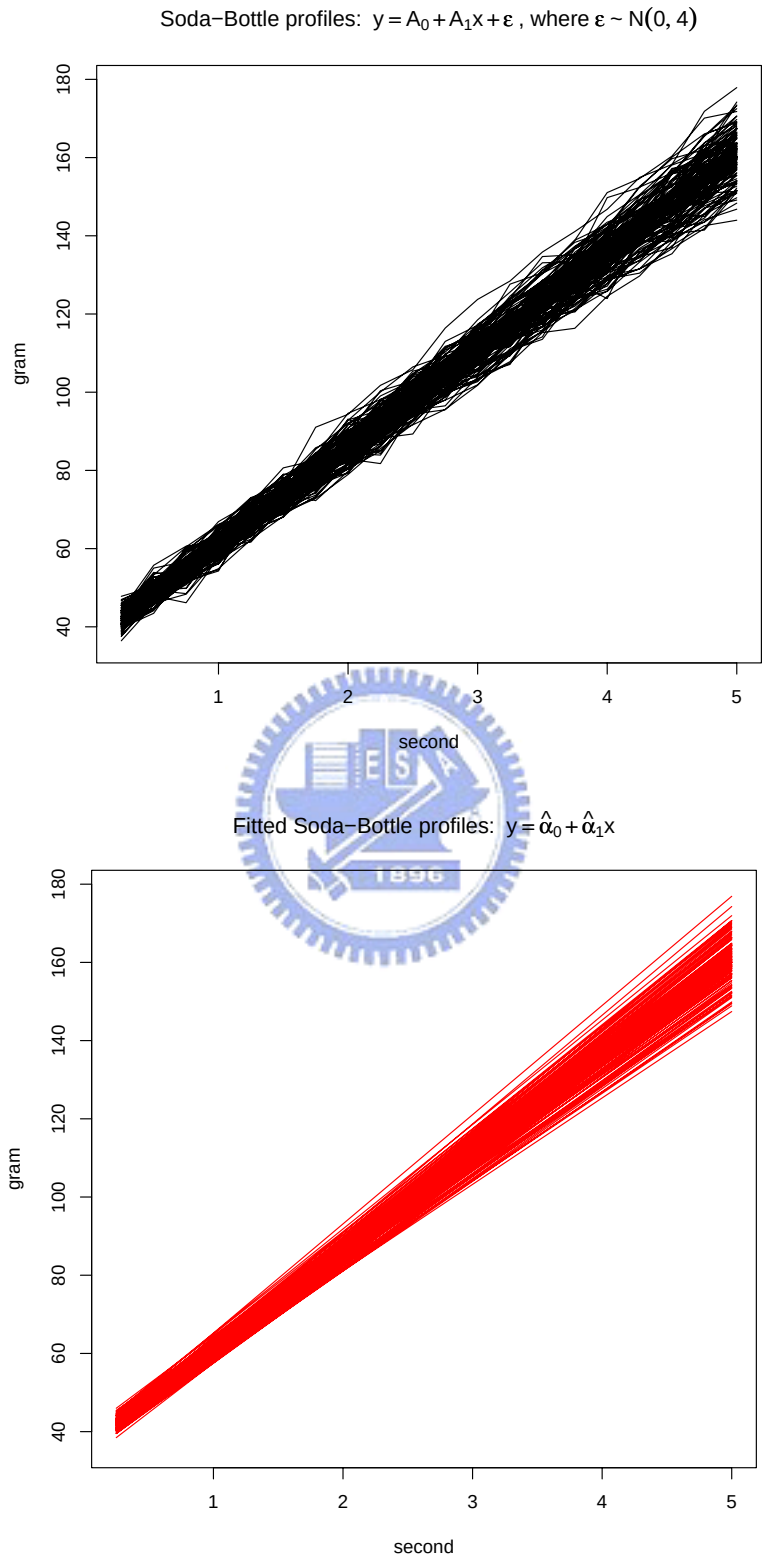


Figure 6: 150 simulated profiles and their fitted profiles for bottle-filling example.

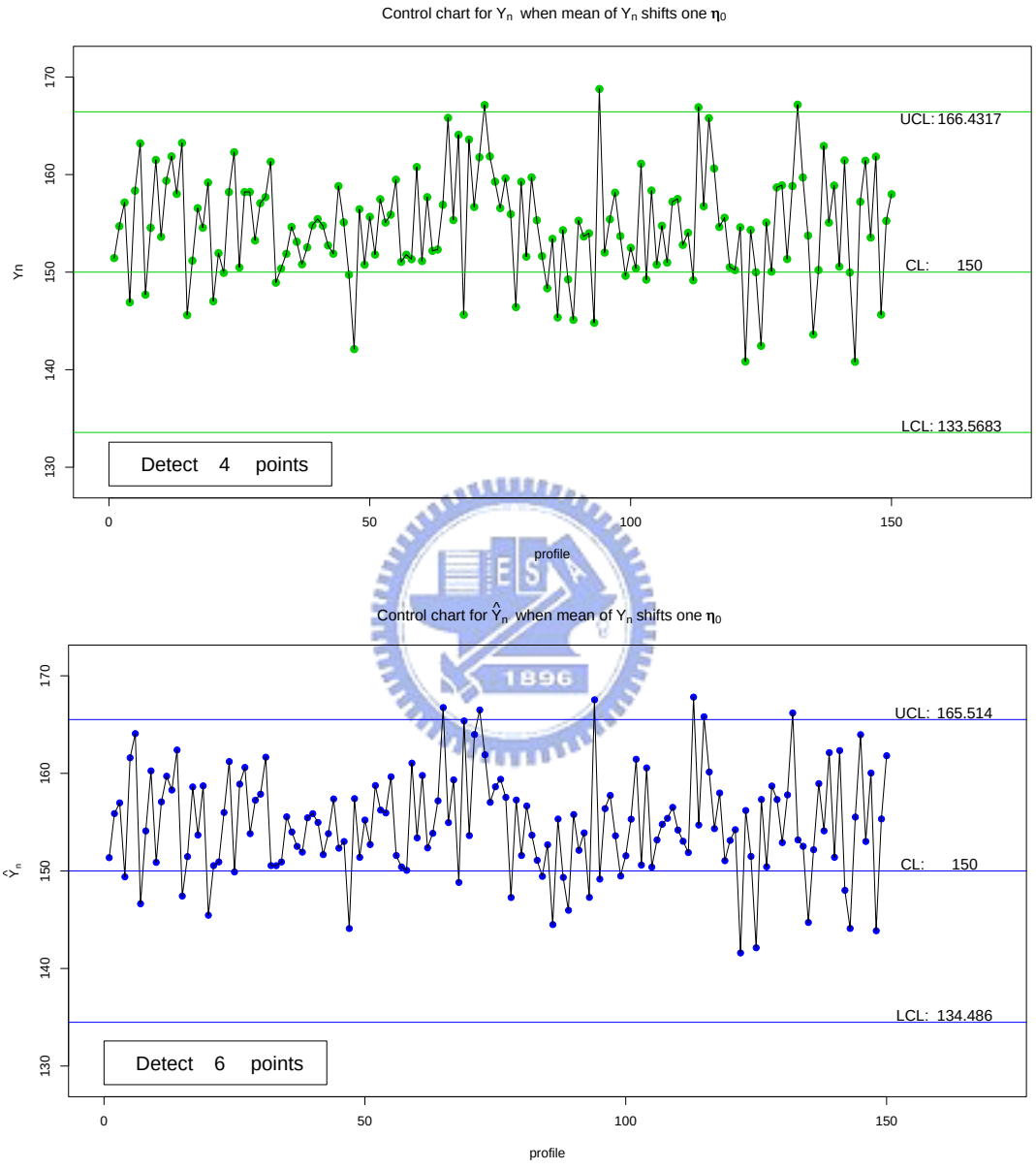


Figure 7: Y_n and $\hat{Y}_n$ charts with $\delta = 1$ for bottle-filling example.

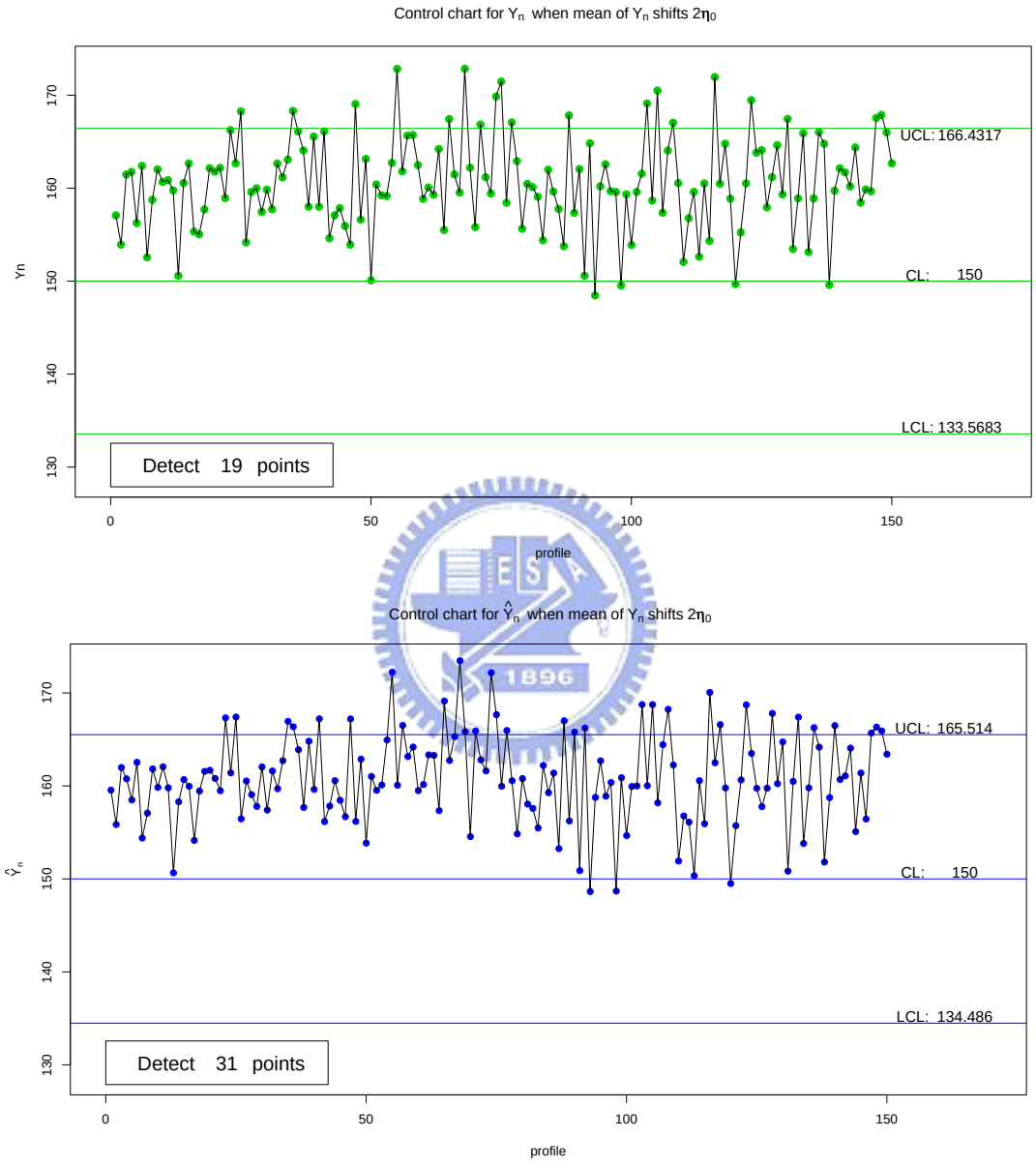


Figure 8: Y_n and $\hat{Y}_n$ charts with $\delta = 2$ for bottle-filling example.

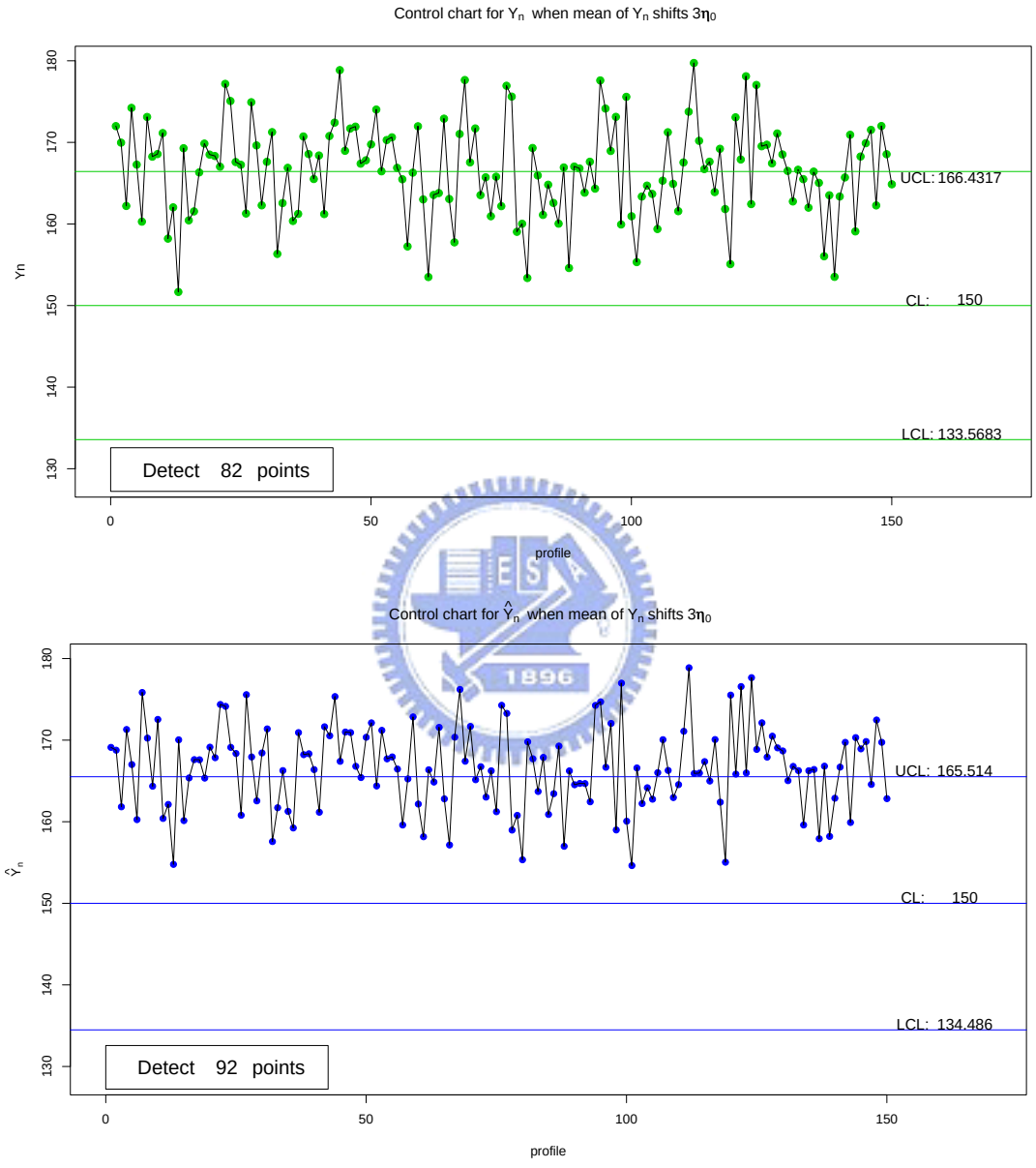


Figure 9: Y_n and $\hat{Y}_n$ charts with $\delta = 3$ for bottle-filling example.

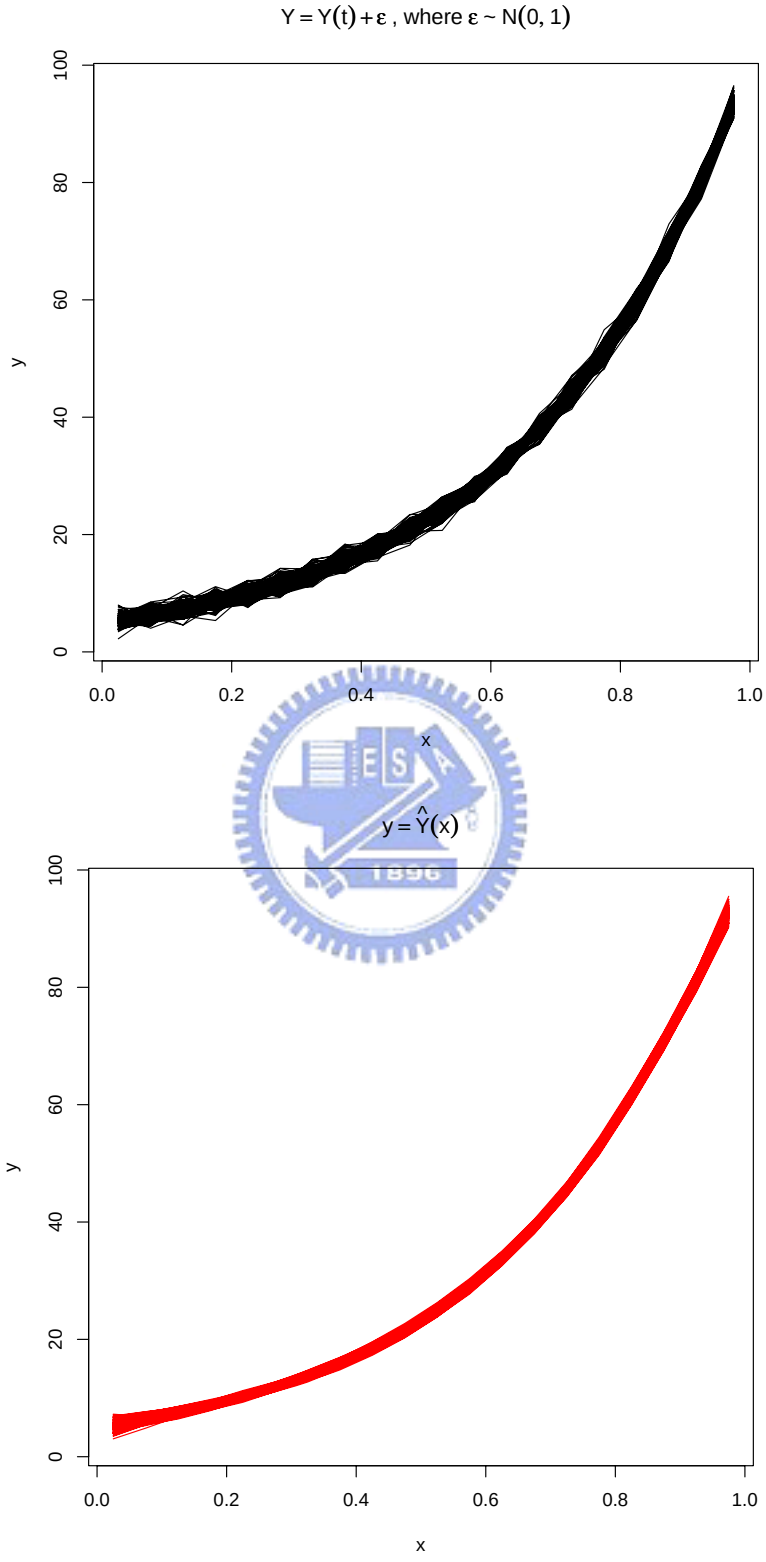


Figure 10: 150 simulated profiles and their fitted profiles for a nonlinear example.

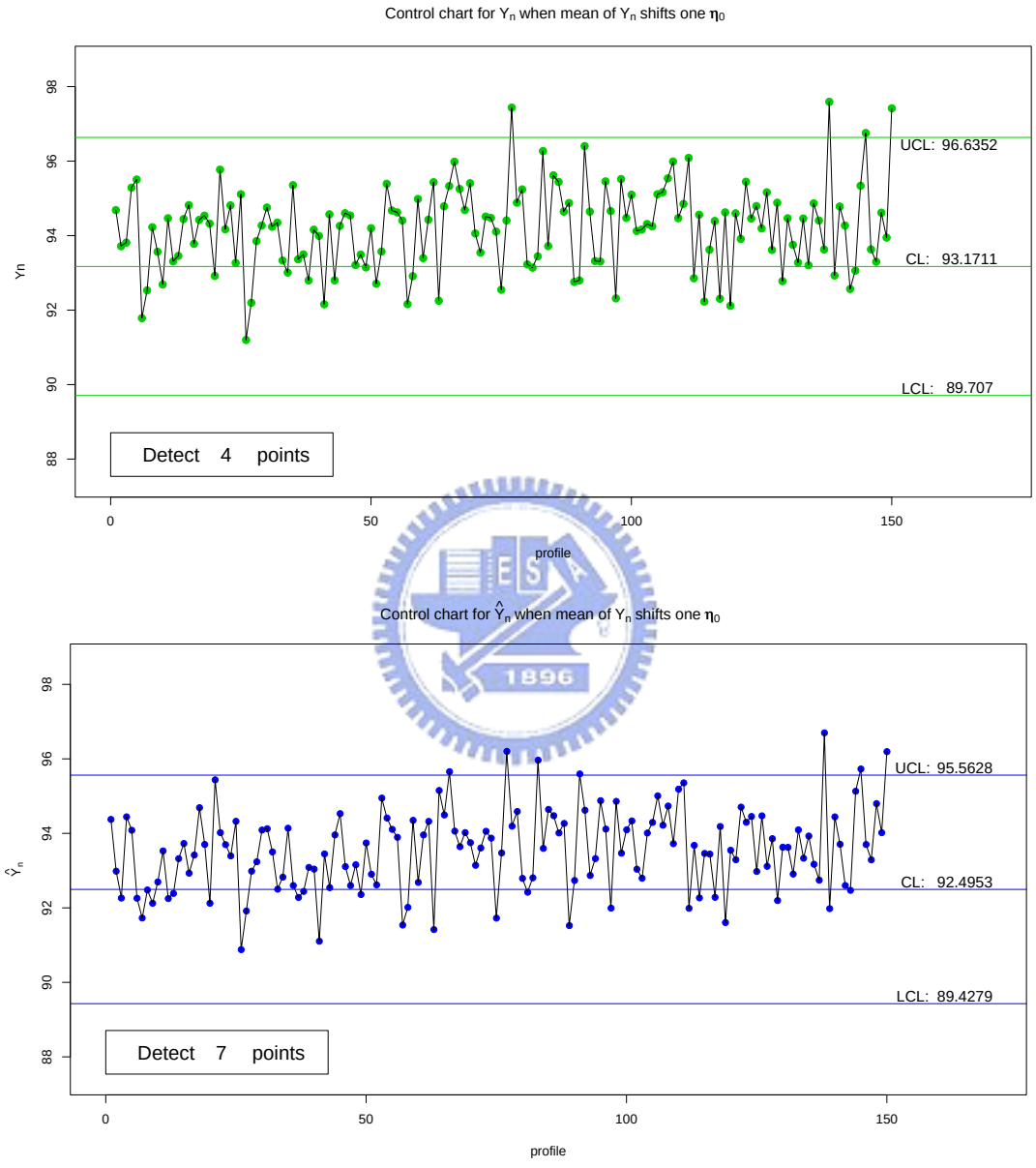


Figure 11: Y_n and $\hat{Y}_n$ charts with $\delta = 1$ for a nonlinear example.

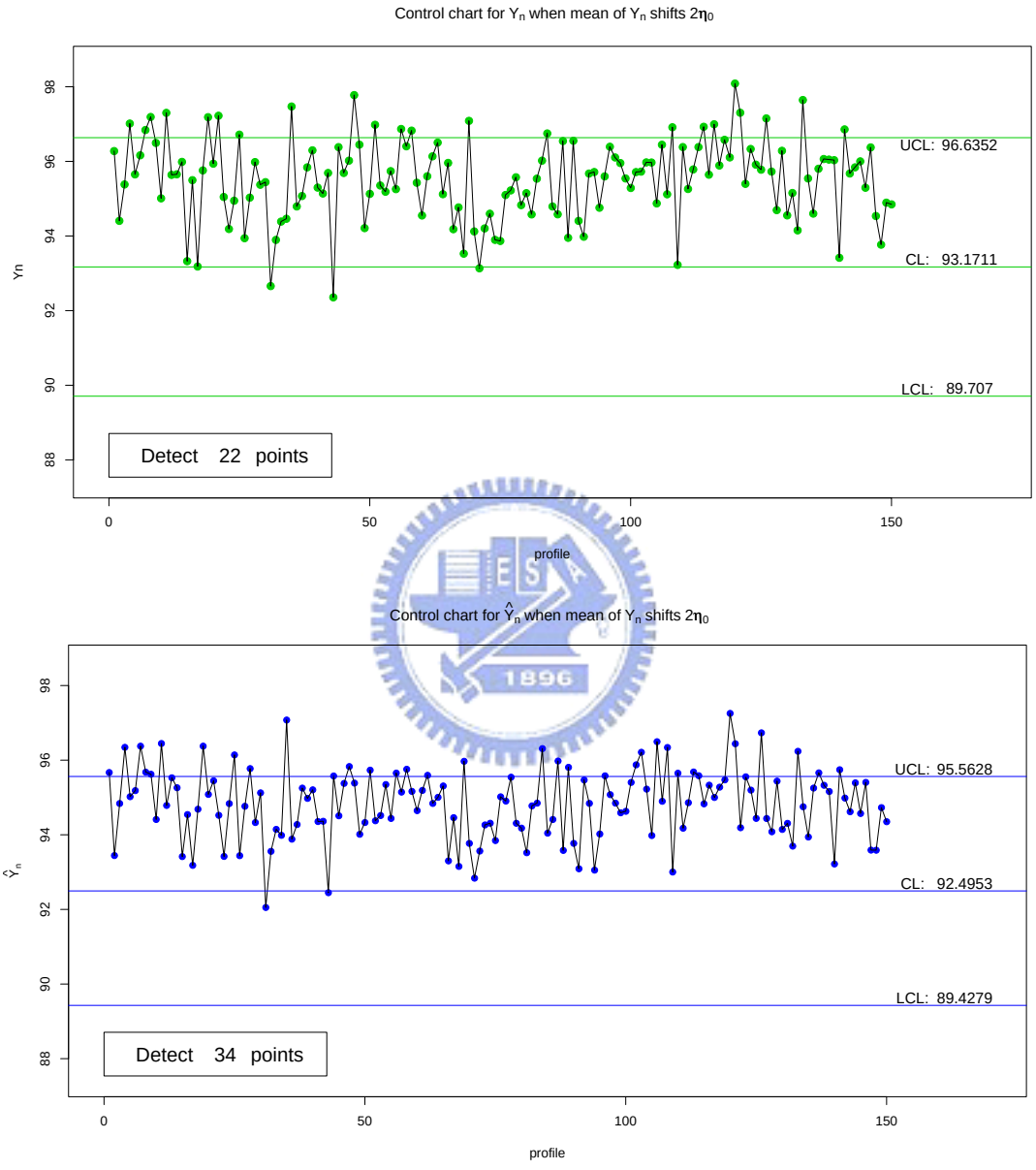


Figure 12: Y_n and $\hat{Y}_n$ charts with $\delta = 2$ for a nonlinear example.

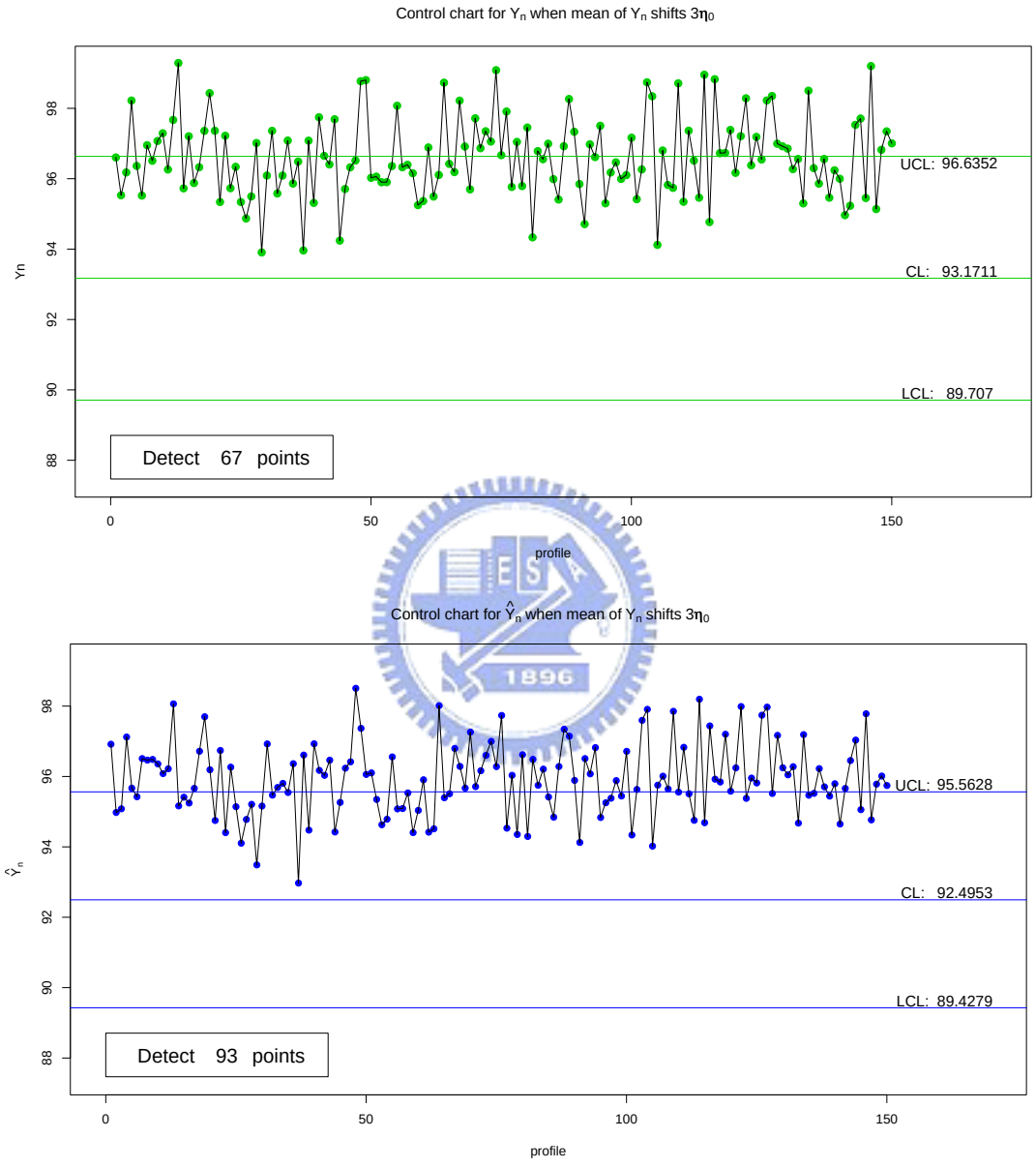


Figure 13: Y_n and $\hat{Y}_n$ charts with $\delta = 3$ for a nonlinear example.