INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

Multivariate Randomness Statistics

By
Kilani Ghoudi

Thesis submitted to the graduate school
in partial fulfillement of the requirement
for the degree of
Doctor of Philosophy
in
Mathematics

at the
University of Ottawa

Acknowledgement

I wish to thank Professor David McDonald for his advice and constant encouragement throughout the development of this thesis.

I also wish to express my gratitude to the University of Ottawa and the Canadian International Development Agency for providing me with the financial support.

Abstract

During the startup phase of a production process while statistics on the product quality are being collected it is useful to establish that the process is under control. Small samples $\{n(i)\}_{i=1}^{q}$ are taken periodically for q periods. We shall assume each measurement is multivariate. A process is under control or on-target if all the observations are deemed to be independent and identically distributed. Let F^i represent the empirical distribution function of the i^{th} sample. Let $\bar{F}$ represent the empirical distribution function of all observations. Following Lehmann (1951) we propose statistics of the form

$$\sum_{i=1}^{q} \int_{-\infty}^{\infty} (F^i(s) - \bar{F}(s))^2 d\bar{F}(s). \tag{0.1}$$

The asymptotics of nonparametric q-sample Cramer-Von Mises statistics were studied in Kiefer (1959). The emphasis there, however, is on the case where $n(i) \to \infty$ while q stayed fixed. Here we study the asymptotics of a family of *randomness* statistics, that includes the above. These asymptotics are in the quality control situation (i.e $q \to \infty$ while $n(i)$ stay fixed).

Such statistics can be used in many situations; in fact one can use *randomness* statistics in any situation where the problem amounts to a test of homoscedasticity or homogeneity of a collection of observations. We give two such applications. First we show how such statistics can be used in nonparametric regression. Second we illustrate the application to retrospective quality control.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Testing for homoscedasticity has many applications in statistical analysis. For example regression analysis may be viewed as determining the set of parameters that makes the residuals homoscedastic. Also *off-line quality control* is equivalent to a test of homoscedaticity on past quality measurements. Assume for each $i = 1, \cdots, q$ a small sample $n(i)$ observations is collected. Let

- $N = \sum_{i=1}^{q} n(i)$,
- F^i is the empirical distribution function of the i^{th} sample,
- $\bar{F}$ is the empirical distribution function of all the samples taken together.

The problem is to test if these samples are homoscedastic. Lehmann (1951) considered the problem of testing the equality of the distributions of q sam-

ples. He proposed

$$\sum_{i=1}^{q} \int (F^i(s) - \bar{F}(s))^2 d\bar{F}(s).$$

Here we present the following statistics

$$S_q = \sum_{i=1}^{q} \int \cdots \int k_q(N, i, F^i(s_1, \ldots, s_d), \bar{F}(s_1, \ldots, s_d)) n(i) dF^i(s_1, \ldots, s_d).$$

These statistics are examples of the Crámer-Von Mises family of statistics. The asymptotic properties of these statistics were studied by Kiefer (1959). He considered the case where $n(i) \to \infty$ while q stayed fixed. McDonald (1991) considered the situation where $q \to \infty$ and $n(i)$ stays fixed for univariate observations. In Chapter 2 of this thesis a family of statistics called *randomness statistics* which contains the above statistics is introduced. The asymptotics of such a family of statistics constitutes the main theoretical result of this thesis. Note that these asymptotics deal with the case of multivariate observations and they are in the case where $q \to \infty$ while $n(i)$ stay fixed.

A typical application of the above statistics would be the case of a startup phase of a quality control procedure in which we do not dispose of the nominal control values nor of sufficient information on the distribution of the measured qualities. Usually we assume that the process is in control during this phase and we use the first measurements to collect statistics on the product qualities. The non-parametric statistics presented here allow us to do retrospective tests on the past measurements to ensure that the process was always in control. This application is presented in Chapter 4 of this work.

Chapter 3 presents another application of the above statistics to regression analysis.

Chapter 2

Asymptotics of multidimensional randomness statistics

2.1 Introduction

During the startup phase of a production process while statistics on the product quality are being collected it is useful to establish that the process is under control. Small samples $\{n(i)\}_{i=1}^{q}$ are taken periodically for q periods. We shall assume each measurement is an $\mathcal{R}^d$ random vector. A process is under control or on-target if all the observations are deemed to be independent and identically distributed. Testing if the process is on target is therefore equivalent to test if a collection of q small samples forms a sequence of i.i.d random vectors. Lehmann (1951) considered nonparametric tests of equality of the distributions of q samples. He proposed the following statistic (for

$d = 1$)

$$\sum_{i=1}^{q} \int_{-\infty}^{\infty} (F^i(s) - \bar{F}(s))^2 d\bar{F}(s) \tag{2.1}$$

where F^i represents the empirical distribution function of the i^{th} sample and $\bar{F}$ represents the empirical distribution function of all observations. Note that this statistic is a measure of distance between the overall empirical c.d.f and the empirical distributions of each of the q samples. The family of such statistics is known as the family of q-sample Cramer-Von Mises statistics. The asymptotics of nonparametric q-sample Cramer-Von Mises statistics were studied in Kiefer (1959). The emphasis there, however, is on the case where $n(i) \to \infty$ while q stays fixed. Here we study the asymptotics of a Cramer-Von Mises family of statistics when $q \to \infty$ while $n(i)$'s stay small. In the rest of this chapter we shall define a family of *randomness* statistics and then we shall establish the asymptotic normality of these statistics under some regularity assumptions.

2.2 Randomness Statistics

Let U be a random variable on $\mathcal{R}^d$ with distribution F_U. Then for any $\vec{t} \in \mathcal{R}^d$ with components t_k define the generalized distribution function $\vec{F}_U := \{G_a : a \in \mathcal{I}\}$ whose a^{th} component is of the form

$$G_a := E \prod_{k \in I_a} \chi_{(-\infty, t_k]} \prod_{k \in I_a^c} \chi_{[t_k, \infty)}(\vec{U})$$

where I_a is a subset of $\{1, 2, \cdots, d\}$ and I_a^c denotes the complement. The index a ranges over all 2^d subsets of $\{1, 2, \cdots, d\}$. Suppose, for each $i =$

$1, \cdots, q$, we have a sequence $\{\vec{U}_{ij} : j = 1, \cdots, n(i)\}$ of independent random vectors with values in $\mathcal{R}^d$ with distribution F_i and generalized distribution $\vec{F}_i$. Let $\vec{F} = \sum_{i=1}^{q} n(i)\vec{F}_i/N$ denote the mean generalized distribution. Let $\vec{U}_i := \{\vec{U}_{ij} : j = 1, \cdots, n(i)\}$. We may define the corresponding generalized empirical distributions of the i^{th} block:

$$F_a^i(\vec{t}) := \frac{1}{n(i)} \sum_{j=1}^{n(i)} \prod_{k \in I_a} \chi_{(-\infty, t_k]} \prod_{k \in I_a^c} \chi_{[t_k, \infty)}(\vec{U}_{ij}), \qquad \vec{F}^i = \{F_a^i\}_{a \in \mathcal{I}}$$

and of the total sample:

$$F_a^\bullet(\vec{t}) := \frac{1}{N} \sum_{i=1}^{q} \sum_{j=1}^{n(i)} \prod_{k \in I_a} \chi_{(-\infty, t_k]} \prod_{k \in I_a^c} \chi_{[t_k, \infty)}(\vec{U}_{ij}), \qquad \vec{F}^\bullet = \{F_a^\bullet\}_{a \in \mathcal{I}}.$$

For $\vec{t}_i = (t_{i1}, \cdots, t_{in(i)})$ where $t_{ij} \in \mathcal{R}^d$ and $\alpha \in \{1, \cdots, q\}$ define $\vec{S}_\alpha(\vec{t}_i) = \{\vec{F}_\alpha(t_{ij}) : j = 1, \cdots n(i)\}$ and $\vec{S}^\alpha(\vec{t}_i) = \{\vec{F}^\alpha(t_{ij}) : j = 1, \cdots n(i)\}$. The local structure of the i^{th} block is given by $\vec{S}^i(\vec{U}_i) = \{\vec{F}^i(U_{ij}) : j = 1, \cdots n(i)\}$. Define the global structure of the i^{th} block $\vec{S}^\bullet(\vec{U}_i) = \{\vec{F}^\bullet(U_{ij}) : j = 1, \cdots n(i)\}$. We also define $\vec{S}(\vec{U}_i) = \{\vec{F}(U_{ij}) : j = 1, \cdots n(i)\}$ and $\vec{S}_i(\vec{U}_i) = \{\vec{F}_i(U_{ij}) : j = 1, \cdots n(i)\}$. These vectors are all in the space $\mathcal{H}_i := \{h_{ja}; j \in \{1, 2, \cdots, n(i)\}, a \in \mathcal{I}\}$. We denote the norm of vectors $\vec{h}$ in $\mathcal{H}_i$ by

$$\|\vec{h}\|_i := [\sum_{j=1}^{n(i)} \sum_{a \in \mathcal{I}} h_{ja}^2]^{1/2}.$$

If $k(\vec{s}, \vec{t})$ is a function on $\mathcal{H}_i \times \mathcal{H}_i$ taking real values then we denote the derivative in the direction $\vec{v} \in \mathcal{H}_i$ by $\vec{v} \cdot \nabla_t k(\vec{s}, \vec{t})$. Furthermore we denote the Hessian applied to vectors $\vec{v}$ and $\vec{w}$ by $\vec{v} \cdot \nabla_{t,t} k(\vec{s}, \vec{t}) \cdot \vec{w}$. Mixed derivatives have the analogous notation.

Definition 2.1 *We say S_q is a multidimensional* randomness *statistic with kernel k_q if it is of the form*

$$S_q = \sum_{i=1}^{q} k_q(N, i, \vec{S}^i(\vec{U}_i), \vec{S}^\bullet(\vec{U}_i)).$$

We shall impose some of the following regularity conditions on the *randomness* statistic. For all positive integers $\{N, q, i\}$ and for all vectors $\vec{s}$ and $\vec{t}$ in $\mathcal{H}_i$ with components in $[0, 1]$:

C1 $E[k_q(N, i, \vec{S}^i(\vec{U}_i), \vec{S}(\vec{U}_i)) - Ek_q(N, i, \vec{S}^i(\vec{U}_i), \vec{S}(\vec{U}_i))]^4 \leq \kappa_3 n(i)^2$.

C2 $|\vec{v} \cdot \nabla_t k_q(N, i, \vec{t}, \vec{t})| \leq \sqrt{n(i)}\kappa_1 \|\vec{v}\|_i$ for all $\vec{v} \in \mathcal{H}_i$.

C3 $|\vec{v} \cdot \nabla_{t,t} k_q(N, i, \vec{s}, \vec{t}) \cdot \vec{w}| \leq \kappa_2 \|\vec{v}\|_i \|\vec{w}\|_i$ for all $\vec{v}$ and $\vec{w}$ in $\mathcal{H}_i$.

C4 $|\vec{v} \cdot \nabla_s k_q(N, i, \vec{s}, \vec{t})| \leq \sqrt{n(i)}\kappa_1 \|\vec{v}\|_i$ for all $\vec{v} \in \mathcal{H}_i$.

We remark that these *randomness* statistics extend the univariate *randomness* statistics discussed in McDonald (1991). Define the rankings $\vec{R}_i = \{R_{i(j)/(N+1)} : j = 1, \cdots n(i)\}$ of the $n(i)$ values U_{ij} of the i^{th} block when ranked among the N points of the q blocks. Here $R_{i(j)}$ denotes the rank of the j^{th} largest of the i^{th} block. We say S_q is a univariate regular *randomness* statistic with kernel k_q if it is of the form

$$S_q = \sum_{i=1}^{q} k_q(N, i, \vec{R}_i);$$

where k_q satisfies conditions C2 and C3. Note that condition C1 follows automatically if all F_i's are equal (for more detail see the proof of Corollary

2.2). A typical univariate *Randomness* statistic was defined in Chouinard and McDonald (1985) as follows:

$$\begin{aligned}\Re_1 &= \sum_{i=1}^{q}\int_{-\infty}^{\infty}\left(\frac{n(i)}{n(i)+1}F^i(s)-\frac{N}{N+1}F^{\bullet}(s)\right)^2 n(i)dF^i(s)\\ &= \sum_{i=1}^{q}\sum_{j=1}^{n(i)}(\frac{N}{N+1}F^{\bullet}(U_{ij})-\frac{j}{n(i)+1}F^i(U_{ij}))^2.\end{aligned}$$

Clearly $\Re_1$ is *small* if the sample is homoscedastic. Moreover $\Re_1$ is distribution free so the mean and variance can be explicitly calculated :

a)

$$E\frac{N}{N+1}F^{\bullet}(U_{ij})=\frac{j}{(n(i)+1)}$$

b)
$$E\Re_1=\frac{n(2n+1)}{6(n+1)}-\sum_{i=1}^{q}\frac{n(i)(2n(i)+1)}{6(n(i)+1)} \tag{2.2}$$

c)

$$\begin{aligned}Var\Re_1 &= \frac{1}{45(n+1)}\{n(q-1)+(2q-2n-5)\left(\sum_{i=1}^{q}\frac{1}{n(i)+1}\right)\\ &+ (n+2)\left(\sum_{i=1}^{q}\frac{1}{(n(i)+1)^2}\right)\\ &- \left(\sum_{i=1}^{q}\frac{1}{n(i)+1}\right)^2+q(-q+3)\}\end{aligned} \tag{2.3}$$

A natural multivariate (bivariate for simplicity) extension of this statistic can be given. In this case $\vec{F}(x,y) := (F(x,y), F(x,\infty)-F(x,y), F(\infty,y)-F(x,y), 1-F(x,\infty)-F(\infty,y)+F(x,y))$ and we define

$$\Re_2=\sum_{i=1}^{q}\sum_{j=1}^{n(i)}||\vec{F}^i(U_{ij})-\vec{F}^{\bullet}(U_{ij})||^2.$$

Again $\Re_2$ is *small* if the sample is homoscedastic but the distribution depends on the underlying distribution F. It is easy to check that $\Re_2$ is a *Randomness* statistic.

The following alternative multivariate (bivariate for simplicity) *Randomness* statistic can be useful in the quality control situation where the components of the observations should be uncorrelated. We remark that statistic generalizes the one given in Ghoudi (1990). Let $\vec{F}_X^\bullet$ be the empirical marginal distribution of the first component based on the observations in all q blocks. Let $\vec{F}_Y^\bullet$ be the corresponding empirical marginal distribution of the second component. Define $\vec{F}_{X,Y}^\bullet(x,y) := (\vec{F}_X^\bullet(x)\vec{F}_Y^\bullet(y), \vec{F}_X^\bullet(x)(1-\vec{F}_Y^\bullet(y)), \vec{F}_Y^\bullet(y)(1-\vec{F}_X^\bullet(x)), (1-\vec{F}_X^\bullet(x))(1-\vec{F}_Y^\bullet(y))$ and

$$\Re_3 = \sum_{i=1}^{q}\sum_{j=1}^{n(i)} ||\vec{F}^i(U_{ij}) - \vec{F}_{X,Y}^\bullet(U_{ij})||^2.$$

Clearly $\Re_3$ is *small* if the sample is homoscedastic and the components are independent. Moreover $\Re_3$ is distribution free so the mean and variance can be explicitly calculated (see Chapter 4 for more detail).

2.3 Asymptotic Normality

Next we propose to study the asymptotics of *randomness* statistics given in Definition 2.1. First consider a Taylor series expansion of the kernel $k_q(n,i,\vec{s},\vec{t})$ in the variable $\vec{t}$ around $\vec{S}(\vec{U}_i)$. We get $S_q = T_q + \Delta_q + L_q$ where

$$T_q = \sum_{i=1}^{q} k_q(N,i,\vec{S}^i(\vec{U}_i),\vec{S}(\vec{U}_i)), \tag{2.4}$$

$$\Delta_q = \sum_{i=1}^{q} (\vec{S}^{\bullet}(\vec{U}_i) - \vec{S}(\vec{U}_i))' \nabla_t k_q(N, i, \vec{S}^{i}(\vec{U}_i), \vec{S}(\vec{U}_i)) \tag{2.5}$$

and

$$\begin{aligned} L_q &= \frac{1}{2} \sum_{i=1}^{q} (\vec{S}^{\bullet}(\vec{U}_i) - \vec{S}(\vec{U}_i))' \nabla_{tt} k_q(N, i, \vec{S}^{i}(\vec{U}_i), \theta \vec{S}^{\bullet}(\vec{U}_i) + (1-\theta)\vec{S}(\vec{U}_i)) \\ &\quad (\vec{S}^{\bullet}(\vec{U}_i) - \vec{S}(\vec{U}_i)). \end{aligned} \tag{2.6}$$

where θ is some stochastic number in $(0, 1)$

Proposition 2.2 *Let $W_1, \cdots, W_n$ be i.i.d $\mathcal{R}^d$-random vectors with common distribution F then for all $a \in \mathcal{I}$ and for every $\epsilon > 0$ there exists a constant $C_{d,\epsilon}$ such that*

$$P\left\{ \sup_{t \in \mathcal{R}^d} |F_a^{\bullet}(t) - F_a(t)| \geq r \right\} \leq C_{d,\epsilon} \exp(-(2-\epsilon) n r^2).$$

Proof

This proposition is a consequence of Kiefer's Theorem in Kiefer (1960). In fact one needs only to notice that each $F_a^{\bullet}$ is the empirical distribution function of N *i.i.d* random vectors having F_a as their common distribution function. □

Corollary 2.3 *There exists a universal constant C which depends only on d such that for all i and α's*

i)

$$E\{ \sup_{\vec{t}_i \in (\mathcal{R}^d)^{n(i)}} \|(\vec{S}^{\alpha}(\vec{t}_i) - \vec{S}_{\alpha}(\vec{t}_i))\|_i^2 \} \leq \frac{Cn(i)}{n(\alpha)} \qquad \textit{and}$$

ii)

$$E\{\sup_{\vec{t}_i \in (\mathcal{R}^d)^{n(i)}} \|(\vec{S}^\alpha(\vec{t}_i) - \vec{S}_\alpha(\vec{t}_i))\|_i^4\} \leq \frac{Cn(i)^2}{n(\alpha)^2}.$$

Proof:

$$\begin{aligned}
& E\{\sup_{\vec{t}_i \in (\mathcal{R}^d)^{n(i)}} \|(\vec{S}^\alpha(\vec{t}_i) - \vec{S}_\alpha(\vec{t}_i))\|_i^2\} \\
&= E\{\sup_{\vec{t}_i \in (\mathcal{R}^d)^{n(i)}} \sum_{j=1}^{n(i)} \sum_{a \in \mathcal{I}} (F_a^\alpha(t_{ij}) - F_{\alpha a}(t_{ij}))^2\} \\
&\leq \sum_{a \in \mathcal{I}} E\{n(i) \sup_{t \in \mathcal{R}^d} |F_a^\alpha(t) - F_{\alpha a}(t)|^2\} \\
&\leq \frac{Cn(i)}{n(\alpha)} \qquad \text{by Proposition 2.2.}
\end{aligned}$$

For the second part of the corollary using Jensen's inequality we get

$$\begin{aligned}
& E\{\sup_{\vec{t}_i \in (\mathcal{R}^d)^{n(i)}} \|(\vec{S}^\alpha(\vec{t}_i) - \vec{S}_\alpha(\vec{t}_i))\|_i^4\} \\
&\leq E\{\sup_{\vec{t}_i \in (\mathcal{R}^d)^{n(i)}} n(i) 2^d \sum_{j=1}^{n(i)} \sum_{a \in \mathcal{I}} (F_a^\alpha(t_{ij}) - F_{\alpha a}(t_{ij}))^4\} \\
&\leq 2^d \sum_{a \in \mathcal{I}} E\{n(i)^2 \sup_{t \in \mathcal{R}^d} |F_a^\alpha(t) - F_{\alpha a}(t)|^4\} \\
&\leq \frac{Cn(i)^2}{n(\alpha)^2} \qquad \text{by Proposition 2.2.}
\end{aligned}$$

Proposition 2.4 *There exists a constant C that depend only on d, the dimension of the space, such that*

$$E\|(\vec{S}^\bullet(\vec{U}_i) - \vec{S}(\vec{U}_i))\|_i^2 \leq \frac{Cn(i)}{N}.$$

Proof:

In fact $\|(\vec{S}^{\bullet}(\vec{U}_i) - \vec{S}(\vec{U}_i))\|_i^2 = \sum_{j=1}^{n(i)} \sum_{a \in \mathcal{I}}^{2^d} (F_a^{\bullet}(U_{ij}) - G_a(U_{ij}))^2$ and from the definition of $F_a^{\bullet}$ it follows

$$\begin{aligned} E\{(F_a^{\bullet}(U_{ij}) - G_a(U_{ij}))^2\} &= E\{\sum_{l=1}^{q} \frac{n(l)}{N}[F_a^l(U_{ij} - F_{la}(U_{ij})]\}^2 \\ &= \sum_{l=1}^{q} \frac{n(l)^2}{N^2} E(F_a^l(U_{ij}) - F_{la}(U_{ij}))^2 \\ &\leq \sum_{l=1}^{q} \frac{n(l)}{N^2} E\{\sup_{t \in \mathcal{R}^d}\{n(l)(F_a^l(t) - F_{la}(t))^2\}\} \\ &\leq \frac{C}{N} \qquad \text{where } C \text{ depends only on } d. \end{aligned}$$

The second equality follows from the fact that if $l \neq h$ then at least one of them is not equal to i and hence $E\{(F_a^l(U_{ij}) - F_{la}(U_{ij}))(F_a^h(U_{ij}) - F_{ha}(U_{ij}))\} = 0$. The last inequality follows from Proposition 2.2 □

Proposition 2.5 *If k_q satisfies* C3 *then* $E|L_q| \leq C < \infty$.

Proof:

From C3 it follows that

$$\begin{aligned} E|L_q| &\leq \kappa_2 \sum_{i=1}^{q} E\|(\vec{S}^{\bullet}(\vec{U}_i) - \vec{S}(\vec{U}_i))\|_i^2 \\ &\leq C \qquad \text{by Proposition 2.4.} \end{aligned}$$ □

The next three propositions will give some insight about Δ_q but first we shall write Δ_q as the sum of U_q and W_q where

$$U_q = \sum_{i=1}^{q} \sum_{\substack{\alpha=1 \\ \alpha \neq i}}^{q} \frac{n(\alpha)}{N} (\vec{S}^{\alpha}(\vec{U}_i) - \vec{S}_{\alpha}(\vec{U}_i))' \nabla_t k_q(N, i, \vec{S}^{i}(\vec{U}_i), \vec{S}(\vec{U}_i)) \qquad (2.7)$$

and

$$W_q = \sum_{i=1}^{q} \frac{n(i)}{N} (\vec{S}^i(\vec{U}_i) - \vec{S}_i(\vec{U}_i))' \nabla_t k_q(N, i, \vec{S}^i(\vec{U}_i), \vec{S}(\vec{U}_i)). \tag{2.8}$$

Now note that $EU_q = 0$. Moreover, if we let

$$u_{i\alpha}(\vec{U}_i, \vec{U}_\alpha) = n(\alpha)(\vec{S}^\alpha(\vec{U}_i) - \vec{S}_\alpha(\vec{U}_i))' \nabla_t k_q(N, i, \vec{S}^i(\vec{U}_i), \vec{S}(\vec{U}_i))$$

then one can easily see that $E\{u_{i\alpha}(\vec{U}_i, \vec{U}_\alpha)|\sigma(\vec{U}_i)\} = E\{u_{i\alpha}(\vec{U}_i, \vec{U}_\alpha)\} = 0$. Also let

$$\hat{U}_q = \sum_{l=1}^{q} E\{U_q|\sigma(\vec{U}_l)\}$$

Proposition 2.6 *Let* $\hat{u}_{i\alpha}(\vec{U}_\alpha) = E\{u_{i\alpha}(\vec{U}_i, \vec{U}_\alpha)|\sigma(\vec{U}_\alpha)\}$ *then*

$$\hat{U}_q = \frac{1}{N} \sum_{i=1}^{q} \sum_{\substack{\alpha=1 \\ \alpha \neq i}}^{q} \hat{u}_{i\alpha}(\vec{U}_\alpha).$$

Proof:

Note that if $l \neq i$ and $l \neq \alpha$ then $E\{u_{i\alpha}(\vec{U}_i, \vec{U}_\alpha)|\sigma(U_l)\} = E\{u_{i\alpha}(\vec{U}_i, \vec{U}_\alpha)\} = 0$ and if $l = i \neq \alpha$ then $E\{u_{i\alpha}(\vec{U}_i, \vec{U}_\alpha)|\sigma(U_l)\} = 0$ by the above remark □.

Proposition 2.7 *If* k_q *satisfies condition C2 then*

$$Var(U_q - \hat{U}_q) \leq \frac{C}{N} \sum_{i=1}^{q} n(i)^2$$

Proof:

Let $\Psi_{i\alpha}(\vec{U}_i, \vec{U}_\alpha) = u_{i\alpha}(\vec{U}_i, \vec{U}_\alpha) - \hat{u}_{i\alpha}(\vec{U}_\alpha)$. From the definition of $\hat{u}$ it follows that $\Psi_{i\alpha}(\vec{U}_i, \vec{U}_\alpha)$ is $\sigma(\vec{U}_i, \vec{U}_\alpha)$ measurable and

$$E\{\Psi_{i\alpha}(\vec{U}_i, \vec{U}_\alpha)|\sigma(\vec{U}_i)\} = E\{\Psi_{i\alpha}(\vec{U}_i, \vec{U}_\alpha)|\sigma(\vec{U}_\alpha)\} = 0.$$

Since $U_q - \hat{U}_q = \frac{1}{N}\sum_{i=1}^{q}\sum_{\substack{\alpha=1\\ \alpha\neq i}}^{q}\Psi_{i\alpha}(\vec{U}_i,\vec{U}_\alpha)$ it follows that

$$Var(U_q-\hat{U}_q)=\frac{1}{N^2}\sum_{i=1}^{q}\sum_{\substack{\alpha=1\\ \alpha\neq i}}^{q}E\{\Psi_{i\alpha}(\vec{U}_i,\vec{U}_\alpha)\}^2+E\{\Psi_{i\alpha}(\vec{U}_i,\vec{U}_\alpha)\Psi_{\alpha i}(\vec{U}_\alpha,\vec{U}_i)\}.$$

First we see that

$$\begin{aligned}
E\{\Psi_{i\alpha}(\vec{U}_i,\vec{U}_\alpha)\}^2 &= E\{u_{i\alpha}(\vec{U}_i,\vec{U}_\alpha)-\hat{u}_{i\alpha}(\vec{U}_\alpha)\}^2\\
&\leq 2E\{u_{i\alpha}(\vec{U}_i,\vec{U}_\alpha)\}^2+2E\{\hat{u}_{i\alpha}(\vec{U}_\alpha)\}^2\\
&\leq 4E\{u_{i\alpha}(\vec{U}_i,\vec{U}_\alpha)\}^2 \qquad \text{by Jensen's inequality}\\
&\leq 4\kappa_1 n(i)n(\alpha)^2E\|\vec{S}^\alpha(\vec{U}_i)-\vec{S}_\alpha(\vec{U}_i)\|_i^2 \qquad \text{by Condition C2}\\
&\leq 4\kappa_1 n(i)n(\alpha)^2E\{\sup_{\vec{t}_i\in(\mathcal{R}^d)^{n(i)}}\|(\vec{S}^\alpha(\vec{t}_i)-\vec{S}_\alpha(\vec{t}_i))\|_i^2\}\\
&\leq Cn(i)^2n(\alpha) \qquad \text{by } i) \text{ in corollary 2.3.}
\end{aligned}$$

Moreover

$$E\{\Psi_{i\alpha}(\vec{U}_i,\vec{U}_\alpha)\Psi_{\alpha i}(\vec{U}_\alpha,\vec{U}_i)\}\leq\left[E\{\Psi_{i\alpha}(\vec{U}_i,\vec{U}_\alpha)\}^2E\{\Psi_{\alpha i}(\vec{U}_\alpha,\vec{U}_i)\}^2\right]^{\frac{1}{2}}\leq Cn(i)^2n(\alpha).$$

Therefore

$$Var(U_q-\hat{U}_q)\leq\frac{C}{N}\sum_{l=1}^{q}n(l)^2.$$

Proposition 2.8 *If k_q satisfies condition C2 then*

$$Var(W_q)\leq\frac{C}{N}\sum_{l=1}^{q}n(l)^2.$$

Proof:

$$W_q^2 \leq \sum_{i=1}^{q} \frac{n(i)}{N} \{(\vec{S}^i(\vec{U}_i) - \vec{S}_i(\vec{U}_i))' \nabla_t k_q(N, i, \vec{S}^i(\vec{U}_i), \vec{S}(\vec{U}_i))\}^2$$

$$\leq \kappa_1^2 \sum_{i=1}^{q} \frac{n(i)^2}{N} \|\vec{S}^i(\vec{U}_i) - \vec{S}_i(\vec{U}_i)\|_i^2 \qquad \text{by Condition C2.}$$

The first inequality follows from Jensen'n inequality. Now Corollary 2.3 gives

$$Var(W_q) \leq E(W_q)^2 \leq \frac{C}{N^2} \sum_{l=1}^{q} n(l)^2. \qquad \square$$

Theorem 2.9 *If S_q is a* Randomness *statistic satisfying conditions C1, C2 and C3 and* $\liminf_{q\to\infty} \frac{VarS_q}{q} > 0$ *and* $\lim_{q\to\infty} \sum_{i=1}^{q} n(i)^2/q^2 = 0$ *then* $(S_q - ES_q)/\sqrt{VarS_q} \Rightarrow Z(0,1)$ *where* $Z(0,1)$ *indicates a standard normal random variable.*

Proof:

If $\liminf_{q\to\infty} \frac{VarS_q}{q} > 0$ and $\lim_{q\to\infty} \sum_{i=1}^{q} n(i)^2/q^2 = 0$ then Propositions 2.5, 2.7 and 2.8 imply that $(S_q - ES_q)/\sqrt{Var(S_q)}$ is asymptotically equivalent to $(T_q - ET_q + \hat{U}_q)/\sqrt{Var(S_q)}$. One can easily verify that $\lim_{q\to\infty} Var(S_q)/Var(T_q + \hat{U}_q) = 1$ and hence we just need to check the Lyapounov conditions for $T_q - ET_q + \hat{U}_q = \sum_{i=1}^{q} \xi_i$ where after a change of the order of summation and a relabeling of i and α we get

$$\xi_i = k_q(n, i, \vec{S}^i(\vec{U}_i), \vec{S}(\vec{U}_i)) - Ek_q(n, i, \vec{S}^i(\vec{U}_i), \vec{S}(\vec{U}_i)) + \frac{1}{N} \sum_{\substack{\alpha=1 \\ \alpha\neq i}}^{q} \hat{u}_{\alpha i}(\vec{U}_i)$$

are independent random variables.

$$E\{\xi_i^4\} \leq 8E\{k_q(n, i, \vec{S}^i(\vec{U}_i), \vec{S}(\vec{U}_i)) - Ek_q(n, i, \vec{S}^i(\vec{U}_i), \vec{S}(\vec{U}_i))\}^4$$

$$
\begin{aligned}
&+\frac{8}{N^4}E\{\sum_{\substack{\alpha=1\\ \alpha\neq i}}^{q}\hat{u}_{\alpha i}(\vec{U}_i)\}^4 \\
&\leq Cn(i)^2+\frac{8}{N^4}E\{\sum_{\substack{\alpha=1\\ \alpha\neq i}}^{q}\hat{u}_{\alpha i}(\vec{U}_i)\}^4 \quad \text{by Condition C1} \\
&\leq Cn(i)^2+\frac{8}{N^4}E\{\sum_{\substack{\alpha=1\\ \alpha\neq i}}^{q}n(i)\sqrt{n(\alpha)}\|\vec{S}^i(\vec{U}_\alpha)-\vec{S}_i(\vec{U}_\alpha)\|_\alpha\}^4 \text{ by Condition C2} \\
&\leq Cn(i)^2+\frac{8}{N}E\{\sum_{\substack{\alpha=1\\ \alpha\neq i}}^{q}\frac{n(i)^4}{n(\alpha)}\|\vec{S}^i(\vec{U}_\alpha)-\vec{S}_i(\vec{U}_\alpha)\|_\alpha^4\} \text{ by Jensen's inequality} \\
&\leq Cn(i)^2+\frac{8}{N}\sum_{\substack{\alpha=1\\ \alpha\neq i}}^{q}\frac{n(i)^4}{n(\alpha)}\frac{Cn(\alpha)^2}{n(i)^2} \quad \text{by Corollary 2.3 (ii)} \\
&\leq Cn(i)^2.
\end{aligned}
$$

The above plus the hypothesis $\liminf_{q\to\infty}\frac{VarS_q}{q}>0$ and $\lim_{q\to\infty}\sum_{i=1}^{q}n(i)^2/q^2=0$ gives the Lyapounov condition for the fourth moment. □

Using the same proof we can have different hypotheses yielding the asymptotic normality.

Definition 2.1 *Alternate forms of condition C1 are*

C1a $E[k_q(N,i,\vec{S}(\vec{U}_i),\vec{S}(\vec{U}_i))-Ek_q(N,i,\vec{S}(\vec{U}_i),\vec{S}(\vec{U}_i))]^4=\kappa_3 n(i)^2.$

C1b $E[k_q(N,i,\vec{S}_i(\vec{U}_i),\vec{S}(\vec{U}_i))-Ek_q(N,i,\vec{S}_i(\vec{U}_i),\vec{S}(\vec{U}_i))]^4=\kappa_3 n(i)^2.$

Theorem 2.10 *If S_q is a* Randomness *statistic satisfying one of the following sets of conditions*

- *C1a, C2, C3 and C4 and the F_i are all equal to F*

- *C1b, C2, C3 and C4*

- *C2, C3, C4 and*

$$k_q(N,i,\vec{S}^i(\vec{U}_i),\vec{S}^\bullet(\vec{U}_i)) = \sum_{j=1}^{n(i)} k'_q(N,i,\vec{F}^i(\vec{U}_{ij}),\vec{F}^\bullet(\vec{U}_{ij}))$$

where k'_q is bounded

then if

$$\liminf_{q\to\infty} \frac{VarS_q}{q} > 0 \text{ and } \lim_{q\to\infty} \sum_{i=1}^{q} n(i)^2/q^2 = 0$$

then $(S_q - ES_q)/\sqrt{VarS_q} \Rightarrow Z(0,1)$ *where* $Z(0,1)$ *indicates a standard normal random variable.*

Proof:

In fact condition C1 is only used to establish that

$$E\{k_q(n,i,\vec{S}^i(\vec{U}_i),\vec{S}(\vec{U}_i)) - Ek_q(n,i,\vec{S}^i(\vec{U}_i),\vec{S}(\vec{U}_i))\}^4 \le Cn(i)^2 \qquad (2.9)$$

therefore to prove Theorem 2.10 we just need to verify inequality (2.9). First consider the conditions C1a and C4 and all the F_i's equal. A Taylor series development of $k_q(n,i,\vec{s},\vec{t})$ in the variable $\vec{s}$ around $\vec{S}(\vec{U}_i)$ gives

$$\begin{aligned} &E\{k_q(n,i,\vec{S}^i(\vec{U}_i),\vec{S}(\vec{U}_i)) - Ek_q(n,i,\vec{S}^i(\vec{U}_i),\vec{S}(\vec{U}_i))\}^4 \\ &\le 27E\{k_q(n,i,\vec{S}(\vec{U}_i),\vec{S}(\vec{U}_i)) - Ek_q(n,i,\vec{S}(\vec{U}_i),\vec{S}(\vec{U}_i))\}^4 \\ &\quad +54E\{(\vec{S}^i(\vec{U}_i) - \vec{S}(\vec{U}_i))'\nabla_s k_q(n,i,(\theta\vec{S}^i(\vec{U}_i) + (1-\theta)\vec{S}(\vec{U}_i),\vec{S}(\vec{U}_i))\}^4. \end{aligned}$$

Using Condition C1a and C4 the above gives

$$\begin{aligned} & E\{k_q(n,i,\vec{S}^{\bullet}(\vec{U}_i),\vec{S}(\vec{U}_i)) - Ek_q(n,i,\vec{S}^{\bullet}(\vec{U}_i),\vec{S}(\vec{U}_i))\}^4 \\ & \leq 27Cn(i)^2 + 54\kappa_3^4 n(i)^2 E\{\|\vec{S}^{\bullet}(\vec{U}_i) - \vec{S}(\vec{U}_i)\|_i^4\} \\ & \leq Cn(i)^2. \end{aligned}$$

The last inequality follows from Corollary 2.3(ii) and the fact that all F_i's are equal.

Next consider the conditions C1b and C4 and we see that a Taylor series development of $k_q(n,i,\vec{s},\vec{t})$ in the variable $\vec{s}$ around $\vec{S}_i(\vec{U}_i)$ gives

$$\begin{aligned} & E\{k_q(n,i,\vec{S}^{\bullet}(\vec{U}_i),\vec{S}(\vec{U}_i)) - Ek_q(n,i,\vec{S}^{\bullet}(\vec{U}_i),\vec{S}(\vec{U}_i))\}^4 \\ & \leq 27E\{k_q(n,i,\vec{S}_i(\vec{U}_i),\vec{S}(\vec{U}_i)) - Ek_q(n,i,\vec{S}_i(\vec{U}_i),\vec{S}(\vec{U}_i))\}^4 \\ & \quad +54E\{(\vec{S}^{\bullet}(\vec{U}_i) - \vec{S}_i(\vec{U}_i))'\nabla_s k_q(n,i,(\theta\vec{S}^{\bullet}(\vec{U}_i) + (1-\theta)\vec{S}_i(\vec{U}_i),\vec{S}(\vec{U}_i))\}^4. \end{aligned}$$

Using Condition C1b and C4 the above gives

$$\begin{aligned} & E\{k_q(n,i,\vec{S}^{\bullet}(\vec{U}_i),\vec{S}(\vec{U}_i)) - Ek_q(n,i,\vec{S}^{\bullet}(\vec{U}_i),\vec{S}(\vec{U}_i))\}^4 \\ & \leq Cn(i)^2 + 54\kappa_3^4 n(i)^2 E\{\|\vec{S}^{\bullet}(\vec{U}_i) - \vec{S}_i(\vec{U}_i)\|_i^4\} \\ & \leq Cn(i)^2 \qquad \text{by Corollary 2.3}(ii). \end{aligned}$$

Finally if $k_q(N,i,\vec{S}^{\bullet}(\vec{U}_i),\vec{S}^{\bullet}(\vec{U}_i)) = \sum_{j=1}^{n(i)} k'_q(N,i,\vec{F}^{\bullet}(\vec{U}_{ij}),\vec{F}^{\bullet}(\vec{U}_{ij}))$ where k'_q is bounded we show that C1b holds. We need only note that

$$k_q(n,i,\vec{S}_i(\vec{U}_i),\vec{S}(\vec{U}_i)) - Ek_q(n,i,\vec{S}_i(\vec{U}_i),\vec{S}(\vec{U}_i))$$

is the sum of $n(i)$ bounded independent and mean zero random variables so its fourth moment is less than $Cn(i)^2$. □

The main result in McDonald (1991) should have been stated as follows.

Corollary 2.2 *If S_q satisfies conditions C2 and C3 and if*

$$\lim_{q\to\infty}\sum_{i=1}^{q} n(i)^2/q^2 = 0, \quad \textit{and} \quad \liminf_{q\to\infty}\frac{VarS_q}{q} > 0$$

then $(S_q - ES_q)/\sqrt{VarS_q} \Rightarrow Z(0,1)$ where $Z(0,1)$ indicates a standard normal random variable.

The hypothesis $\limsup_{q\to\infty}\frac{N(q)}{q} < \infty$ was never used.

Proof: By Theorem 2.9 it suffices to verify condition C1. Using the smoothness of k_q and expanding around the point

$$\vec{p}_i := (\frac{1}{n(i)+1}, \frac{2}{n(i)+1}, \cdots, \frac{n(i)}{n(i)+1})$$

$$\begin{aligned} & k_q(N,i,\vec{U}_i) - Ek_q(N,i,\vec{U}_i) \\ &= \sum_{j=1}^{n(i)} \frac{\partial k_q}{\partial z_{ij}}(N,i,\vec{p}_i)(U_{i(j)} - \frac{j}{n(i)+1}) + (L_q - EL_q) \end{aligned}$$

where

$$L_q := \sum_{j=1}^{n(i)}\sum_{m=1}^{n(i)} \frac{1}{2}\frac{\partial^2 k_q}{\partial z_{ij}\partial z_{im}}(N,i,\theta\vec{U}_i+(1-\theta)\vec{p}_i)(U_{i(j)}-\frac{j}{n(i)+1})(U_{i(m)}-\frac{m}{n(i)+1}).$$

By hypothesis the absolute value of L_q is bounded by

$$\kappa_2\sum_{j=1}^{n(i)}(U_{i(j)} - \frac{j}{n(i)+1})^2.$$

Hence,

$$\begin{aligned} & E[k_q(N,i,\vec{U}_i) - Ek_q(N,i,\vec{U}_i)]^4 \\ \leq\ & 27E[\sum_{j=1}^{n(i)} \frac{\partial k_q}{\partial z_{ij}}(N,i,\vec{p}_i)(U_{i(j)} - \frac{j}{n(i)+1})]^4 + 27E(L_q^4) + 27(EL_q)^4) \end{aligned}$$

$$\leq 27n(i)^3 \sum_{j=1}^{n(i)} [\frac{\partial k_q}{\partial z_{ij}}(N,i,\vec{p}_i)E(U_{i(j)} - \frac{j}{n(i)+1})]^4 + 54E(L_q^4)$$

$$\leq 27n(i)^3 \sum_{j=1}^{n(i)} \kappa_1^4 E(U_{i(j)} - \frac{j}{n(i)+1})^4 + 54E(L_q^4)$$

using Jensen's inequality. Next,

$$EL_q^4 \leq E\kappa_2^4[\sum_{j=1}^{n(i)}(U_{i(j)} - \frac{j}{n(i)+1})^2]^4$$

$$\leq \kappa_2^4 n(i)^3 \sum_{j=1}^{n(i)} E(U_{i(j)} - \frac{j}{n(i)+1})^8$$

$$\leq \kappa_2^4 n(i)^3 \sum_{j=1}^{n(i)} E(U_{i(j)} - \frac{j}{n(i)+1})^4$$

again using Jensen's inequality and the fact that $|U_{i(j)} - j/(n(i)+1)| \leq 1$.

The fourth central moment of order statistics can be shown to be of order $1/n(i)^2$ so we conclude

$$E[k_q(N,i,\vec{U}_i) - Ek_q(N,i,\vec{U}_i)]^4 \leq \kappa n(i)^2$$

where κ is some constant depending only on κ_1 and κ_2. The result now follows. □

The corollary below follows as in McDonald (1991).

Corollary 2.11 *If,*

$$\liminf_{q\to\infty} \frac{1}{q}\sum_{i=1}^{q} 1\{n(i) \geq 2\} > 0, \ \lim_{q\to\infty} \sum_{i=1}^{q} n(i)^2/q^2 = 0,$$

then $(\Re_1 - E\Re_1))/\sqrt{Var(\Re_1)} \Rightarrow Z(0,1)$.

Chapter 3

Nonparametric regression

3.1 Literature review

Consider a multivariate linear regression model. Observations of a dependent vector $\mathbf{y} = (y_1, y_2, \cdots, y_d)$ and independent variables $\mathbf{x} = (x_1, x_2, \cdots, x_p)$ are indexed by $t = 1, \cdots, N$ and are governed by the model

$$\mathbf{y}_t = (\mathbf{x}_t)'(\beta) + \epsilon_t$$

with p unknown parameters $(\beta)' = (\beta_1, \beta_2, \cdots, \beta_p)'$. We assume throughout that the errors ϵ_t are independent identically distributed d dimensional vectors with unknown distribution F. Analysing the above model and determining an estimate of β is known as regression analysis. It is one of the most common routines in statistical analysis. In this analysis the least squares technique is commonly used for two reasons. First, it became a tradition and second it is easily implemented. Statistics for the least squares repose on the assumption of normal errors and it is now well understood that the least squares technique is very sensitive to outliers or departure from the

normal assumption. In fact Gauss introduced the normal distribution as the distribution that makes the least squares technique optimal. To overcome this weakness we need a robust procedures in the sense that the estimate is less sensitive to departure from normality and to the presence of outliers. First we shall present a brief review of the most common robust procedures in regression analysis. For an extensive literature review see Huber (1972, 1981) and Rousseeuw and Leroy (1987).

The notion of outliers is quite old; Bessel (1818) noted that three of his test samples show higher frequency of large errors than what a normal model would predict. He then decided to disregard those samples. In 1931 Pearson noted the high sensitivity to departure form normality of some standard procedures (tests of equality of variances). Because of the sensitivity of these tests Box (1953) introduced the term "*Robustness*". Edgeworth (1887) argued that least squares is sensitive to outliers because the residuals are squared so he proposed the *least absolute values regression estimator* which is determined by minimizing the sum of absolute value of the residuals. This is also known as the L_1 regression. L_1 estimates were known even before 1887. In fact Laplace used this idea in the case of one dimensional observations which gave him the sample median. He also introduced the double exponential or the Laplace distribution as the error distribution for which the median is optimal. Careful study (see Rousseeuw and Leroy 1987 for more details) of the L_1 regression estimate shows that this later is less sensitive than the least squares for outliers in the y direction but it is very sensitive

to outliers in the x direction, the so-called *leverage points.*

Generalizing the above idea, Huber (1973) introduced the use of the M-estimator for regression. He proposed minimizing $\sum_{i=1}^{n} \rho(r_i)$ where the r_i are the residuals and ρ is an even function having a unique minimum at zero. Because M-estimators are also sensitive to leverage points generalized M-estimators were introduced (see Rousseeuw and Leroy (1987)).

The next direction was that of L-statistics (Bickel (1973) and Koenker and Bassett (1978, 1982) . Bickel (1973) generalized the use of linear combinations of order statistics from the location problem (e.g. median, trimmed mean) to the one dimensional (i.e. $d = 1$) linear model. Koenker and Bassett (1978, 1982) and Bassett and Koenker (1982) introduced the regression quantile for linear model in the following way. They defined their estimate for β as the argument of $\min_\beta \sum_{i=1}^{N} \rho_\theta(Y_i - X_i\beta)$ where ρ_θ is the θ^{th} quantile of the residuals and $0 < \theta < 1$. In fact they characterized the regression quantiles as solutions of a family of linear programs. Gutenbrunner and Jurečkovà (1992) studied the properties of the dual solutions of the above linear programs. They called these solutions rank scores and they proved that these statistics can be used for regression.

The last direction is that of R-estimators (also for $d = 1$) was first introduced in the work of Jurečkovà (1971). Generalizing an idea of Adichie (1967) she proposed estimating the regression coefficients by taking $\hat{\beta}$ in the set $D_N = \{\beta : \sum_{j=1}^{p} |S_{Nj}(Y - X\beta)| = \min_\beta \sum_{j=1}^{p} |S_{Nj}(Y - X\beta)|\}$ where $S_{Nj}(Y - X\beta) = \sqrt{N} \sum_{i=1}^{N} (X_{ji} - \bar{X}_j) a_N(R_i(\beta))$ where $R_i(\beta)$ is the rank of

the i^{th} residual $Y_i - X_i\beta$ and $a_N(i) = \mathbf{E}\phi(U_N^{(i)})$ or $\phi(i/(N+1))$ and where ϕ is non-decreasing square integrable function. Second we cite the work of Jaeckel (1972) who introduced estimates of the regression coefficients based on a minimization of the dispersion of the residuals. He also proved that his estimates are asymptotically equivalent to those proposed by Jurečkovà (1971). In fact in his method, presented below, he considered measures of dispersion which are weighted sums of the residuals. His weights a_i are functions of the ranks of the residual Z_i's. The Wilcoxon score is the case where $a_i = R_i - (N+1)/2$ where R_i is the rank of Z_i among $Z_1, \ldots, Z_N$; see also Hettmansperger and McKean (1977). The estimate for β is the value that minimizes $D(Y - X\beta)$ where D is a measure of dispersion having the form $D(Z) = \sum_{i=1}^{N} a_i Z_i$.

All the above methods consider generalizing the Least Sum of Squares technique by changing the square of the residuals to some other measure of dispersion that is less sensitive to outliers. The first idea departing from this framework was due to Siegel (1982) who proposed a repeated median estimate defined as follows. Take all groups of p observations $(\mathbf{X}_{i_1}, \mathbf{Y}_{i_1}), \cdots, (\mathbf{X}_{i_p}, \mathbf{Y}_{i_p})$ which determine a unique parameter vector $\beta(i_1, \ldots, i_p)$ whose j^{th} coordinate is denoted $\beta_j(i_1, \ldots, i_p)$. The repeated median estimate $\hat{\beta}$ of the parameter vector β is given

$$\hat{\beta}_j = \underset{i_1}{\text{med}}\,(\underset{i_2}{\text{med}}\,(\cdots(\underset{i_p}{\text{med}}\,(\beta_j(i_1, \ldots, i_p))\cdots)$$

It can be shown that this method has a breakdown point of 50% where the breakdown point is the smallest fraction of contamination of the data

that causes the estimate to take values arbitrarily far away. It also means the smallest fraction of comtamination that makes the estimate have an infinite bias (see Rousseeuw and Leroy (1987) for a more precise definition). Rousseeuw and Leroy (1984) proposed changing the least sum of squares method not by changing the squares but by changing the sum. In fact they proposed the least median square technique which consists in minimizing the median of the squares of the residuals. They also showed that their method has a breakdown point of 50%.

3.2 Introduction

In many experimental designs there are replicate experiments for a fixed value of $\mathbf{x}_t$. In this case, group into blocks $i = 1, \cdots, q$ all the observations having the same independent variables $(x_i^1, x_i^2, \cdots, x_i^p)'$. Index the observations in block i by $j = 1, \cdots, n(i)$. The $n(i)$ observations in block i satisfy

$$\mathbf{y}_{ij} = (\mathbf{x}_{ij})'(\beta) + \epsilon_{ij}$$

where $(\mathbf{x}_{ij})' = (x_i^1, x_i^2, \cdots, x_i^p)'$. In practice even if there are no replicates we may simply pave the parameter space of the dependent variables with contiguous blocks indexed by $i = 1, \cdots, q$ such that within the same block the $(\mathbf{x}_{ij})'$ are close to each other.

For any choice of β we may calculate the residuals of the dependent vector. Denote the j^{th} residual in the i^{th} block by U_{ij}. The true (unknown) value for β, say β_0, makes the residuals i.i.d. hence the best fit or best choice of β is the one that makes these residuals homoscedastic. Suppose the block sizes

are $\{n(i)\}_{i=1}^{q}$ for a total of $N = \sum_{i=1}^{q} n(i)$ observations. Let F^i represent the empirical distribution function of the i^{th} block. Let $\overline{F}$ represent the empirical distribution function of all the observations. Consider the following Cramér-Von Mises statistic

$$\begin{aligned}\Re(\beta) &= \sum_{i=1}^{q} \int \cdots \int \left(F^i(s_1, \cdots, s_d) - \overline{F}(s_1, \cdots, s_d)\right)^2 n(i) dF^i(s_1, \cdots, s_d) \\ &= \sum_{i=1}^{q} \sum_{j=1}^{n(i)} \left(F^i(U_{ij}) - \overline{F}(U_{ij})\right)^2\end{aligned}$$

In principle the above statistic should be small when the residuals are homoscedastic since the empirical distribution of the residuals of each group would then have the same empirical distribution as the entire sample (see Chapter 2 for more details).

In Section 2 we give an efficient algorithm for finding the value $\hat{\beta}$ minimizing the statistic $\Re(\beta)$ as a function of β which compares well with the standard least squares estimator. Since the statistic $\Re(\beta)$ is based on empirical distributions it works well when the errors are not normal and even if outliers are present.

Having found $\hat{\beta}$ we calculate the N residuals. If $\hat{\beta}$ is close to β_0 these residuals should be homoscedastic. Following Hoeffding (1952), consider all $N!$ permutations of these residuals and suppose we recalculate the $\Re$ for each. Consequently $\Re(\hat{\beta})$ is one among $N!/(n(1)!n(2)! \cdots n(q)!)$ possible values for the $\Re$ statistic. By sampling from the permutation group we can estimate the percentiles of the $\Re$ statistic. If, in fact, homoscedasticity is violated one would expect that $\Re(\hat{\beta})$ lies in the upper percentiles of this bootstrap distri-

bution. If this proves to be the case we reject homoscedasticity and hence the regression model. Numerical studies in Section 4 show this procedure has good power for detecting deviations from the regression model.

When the sample size N increases it is typical that $q \to \infty$ while the $n(i)$ stay relatively small (we call this the quality control situation. For more detail see Chapter 4).

Theorem 2.9 gives the asymptotic normality of $\Re$, but to use this theorem we need to calculate $E\Re(\beta_0)$ and $\sqrt{Var(\Re(\beta_0))}$ which unfortunately depend on the distribution F of the error ϵ. Using Maple we can show

Proposition 3.1 *The expected value of* $\Re(\beta_0))$ *is*

$$\frac{1}{n^2}\sum_{i=1}^{q}(n-n(i))^2 + EF(\epsilon)[qn^2 - Nn + 6nq - 3\sum_{i=1}^{q}\frac{1}{n(i)} - 3N]$$
$$+ \quad EF^2(\epsilon)[n^2N - 3n^2q + 2n^2\sum_{i=1}^{q}\frac{1}{n(i)} - (n^2 - 2n - 2)N + 2n(n-2)q].$$

The asymptotic variance of $\Re(\beta_0)$ *is by equation 4.7*

If $d = 1$ then $F(\epsilon)$ is uniformly distributed on the unit interval so the above expressions simplify. But from now on we shall use the statistic $\Re_1$ defined in Chapter 2 whenever we are dealing with $d = 1$. Note the expected value and the variance of $\Re_1$ are given by (2.2) and (2.3) respectively. Using this we may give a confidence region for β_0. We know that with probability $1-\alpha$ the statistic $\Re(\beta_0)$ is less than $L_\alpha := E\Re + z_\alpha\sqrt{Var(\Re)}$. Hence the set of β such that $\Re(\beta) \leq L_\alpha$ is a $(1-\alpha)\cdot 100\%$ confidence region for $\Re(\beta_0)$. Since $\Re(\hat{\beta}) \leq \Re(\beta_0)$ it follows that if $\Re(\hat{\beta})$ exceeds L_α, homoscedasticity must be rejected. Numerical examples in the next section show the confidence region

is comparable to the region associated with the least squares estimates for a normal model. If the errors are Cauchy the least squares estimates fail miserably but this nonparametric technique continues to give good results.

When $d > 1$ the statistic $\Re(\beta_0)$ is asymptotically normal but is not a rank statistic and is not distribution free. We need to estimate $E\Re(\beta_0)$ and $\sqrt{Var(\Re(\beta_0))}$ in order to do as in the $d = 1$ case and construct a $(1-\alpha)\cdot 100\%$ confidence region for β_0. The mean and variance in Proposition 3.1 are given by functionals of the copula of the underlying distribution F of the error (for instance we need $EF(\epsilon), EF^2(\epsilon)$). It is natural to propose the U-statistics estimators

$$\frac{1}{N}\sum_{i=1}^{q}\overline{F}(U_{ij}), \qquad \frac{1}{N}\sum_{i=1}^{q}\overline{F}^2(U_{ij})$$

which are consistent by the Glivenko-Cantelli Lemma. Numerical examples in the next section show that even for the multivariate case the confidence region is comparable to the region associated with the least squares estimates for a normal model.

For functionals $\mathcal{F}$ which are more general than the $\Re$ statistic the bootstrap is the only possible way of giving a confidence region for β_0.

3.3 Properties

In this section we present some of the properties of the estimate $\hat{\beta}$.

3.3.1 Invariance properties

Proposition 3.2 *The estimate $\hat{\beta}$ is regression equivariant in the sense that for any $v \in \mathcal{R}^p$, $\hat{\beta}(\mathbf{X}, \mathbf{Y} + \mathbf{X}v) = \hat{\beta}(\mathbf{X}, \mathbf{Y}) + v$.*

Proof:

Denote $\mathbf{Z}(\mathbf{X}, \mathbf{Y}, \beta)$ the vector of residuals (i.e $\mathbf{Z}(\mathbf{X}, \mathbf{Y}, \beta) = \mathbf{Y} - \mathbf{X}\beta$) and $\Re(\mathbf{X}, \mathbf{Y}, \beta)$ the statistic $\Re(\beta)$. Now note that for any $\beta \in \mathcal{R}^p$, $\mathbf{Z}(\mathbf{X}, \mathbf{Y}, \beta) = \mathbf{Z}(\mathbf{X}, \mathbf{Y}+\mathbf{X}v, \beta+v)$. Hence $\Re(\mathbf{X}, \mathbf{Y}, \beta) = \Re(\mathbf{X}, \mathbf{Y}+\mathbf{X}v, \beta+v)$ and $\hat{\beta}(\mathbf{X}, \mathbf{Y}+\mathbf{X}v) = \hat{\beta}(\mathbf{X}, \mathbf{Y}) + v$ □

Proposition 3.3 *The estimate $\hat{\beta}$ is scale equivariant in the sense for any positive real c, $\hat{\beta}(\mathbf{X}, c\mathbf{Y}) = c\hat{\beta}(\mathbf{X}, \mathbf{Y})$.*

Proof:

Using the same notation as above, we see that $\mathbf{Z}(\mathbf{X}, c\mathbf{Y}, c\beta) = c\mathbf{Z}(\mathbf{X}, \mathbf{Y}, \beta)$. Since the empirical copula function (or even the generalized empirical copula function defined in Chapter 2) is unchanged if all the observations are multiplied by a positive constant we have $\Re(\mathbf{X}, c\mathbf{Y}, c\beta) = \Re(\mathbf{X}, \mathbf{Y}, \beta)$ □.

Proposition 3.4 *The estimate $\hat{\beta}$ is affine equivariant in the sense for any nonsingular matrix A; we have $\hat{\beta}(\mathbf{XA}, \mathbf{Y}) = \mathbf{A}^{-1}\hat{\beta}(\mathbf{X}, \mathbf{Y})$.*

Proof:

Note that $\mathbf{Z}(\mathbf{XA}, \mathbf{Y}, \mathbf{A}^{-1}\beta) = \mathbf{Z}(\mathbf{X}, \mathbf{Y}, \beta)$. Therefore $\Re(\mathbf{XA}, \mathbf{Y}, \mathbf{A}^{-1}\beta) = \Re(\mathbf{X}, \mathbf{Y}, \beta)$ and $\hat{\beta}(\mathbf{XA}, \mathbf{Y}) = \mathbf{A}^{-1}\hat{\beta}(\mathbf{X}, \mathbf{Y})$ □.

3.3.2 Existence

The statistic $\Re(\beta)$can take at maximum $\binom{n}{p}$ values. In fact this statistic changes values only at the values of β for which there exist p design points

that satisfy the linear system of equations $\mathbf{Y} = \mathbf{X}\beta$. This insures that there exists at least a β where $\Re(\beta)$attains its minimum.

3.3.3 Consistency

We show that for $d = 1$ if the errors have a continuous distribution function G with density g and if the design points X_i satisfy the conditions stated at the end of this section for any $\beta \neq \beta_0$ then

$$\lim_{q\to\infty} \frac{E\Re(\beta) - E\Re(\beta_0)}{\sqrt{q}} = \infty.$$

First, for any $\beta \in \mathcal{R}^p$ we have

$$E\Re_1(\beta) - E\Re_1(\beta_0) = \frac{2}{N+1}\sum_{i=1}^{q}\sum_{j=1}^{n(i)} \frac{j}{n(i)+1} E[R_{i(j)}(\beta_0) - R_{i(j)}(\beta)]$$

Now

$$\begin{aligned} &ER_{i(j)}(\beta) \\ &= j + \textstyle\sum_{l\neq i}^{q} n(l) \int_{-\infty}^{+\infty} \begin{pmatrix} n(i)-1 \\ j-1 \end{pmatrix} F_l(x)F_i(x)^{j-1}(1-F_i(x))^{n(i)-j} dF_i(x) \end{aligned}$$

Next let $F(x) = 1/N \sum_{i=1}^{q} n(i)F_i(x)$

$$\begin{aligned} &E\Re_1(\beta) - E\Re_1(\beta_0) \\ &= \textstyle\sum_{i=1}^{q} \frac{2}{(N+1)(n(i)+1)} \Big[\int_{-\infty}^{+\infty}[NF(x) - n(i)F(x)][(n(i)-1)F(x)+1]dF(x) \\ &\qquad - \textstyle\int_{-\infty}^{+\infty}[NF(x) - n(i)F_i(x)][(n(i)-1)_iF(x)+1]dF_i(x)\Big] \\ &\qquad = \textstyle\frac{2N}{N+1}\sum_{i=1}^{q}\int_{-\infty}^{+\infty}(F_i(x) - F(x))\frac{(n(i)-1)F_i(x)+1}{n(i)+1}dF_i(x) \\ &\qquad = \textstyle\sum_{i=1}^{q}\frac{2N(n(i)-1)}{(N+1)(n(i)+1)}\int_{-\infty}^{+\infty}(F_i(x) - F(x))F_i(x)dF_i(x) \end{aligned}$$

$$+\sum_{i=1}^{q}\frac{2N}{(N+1)(n(i)+1)}\int_{-\infty}^{+\infty}(F_i(x)-F(x))dF_i(x)$$

Integrating by parts we get

$$\begin{aligned}
&E\Re_1(\beta)-E\Re_1(\beta_0)\\
&=\sum_{i=1}^{q}\frac{N}{(N+1)(n(i)+1)}\Big[(n(i)-1)\Big(\int_{-\infty}^{+\infty}F_i(x)^2dF(x)-\int_{-\infty}^{+\infty}F_i(x)^2dF_i(x)\Big)\\
&\qquad+2\Big(\int_{-\infty}^{+\infty}F_i(x)dF_i(x)-\int_{-\infty}^{+\infty}F_i(x)dF(x)\Big)\Big]\\
&=\sum_{i=1}^{q}\frac{N}{(N+1)(n(i)+1)}\Big[(n(i)-1)\Big(\int_{-\infty}^{+\infty}F_i(x)^2dF(x)-\int_{-\infty}^{+\infty}F(x)^2dF(x)\Big)\\
&\qquad+2\Big(\int_{-\infty}^{+\infty}F(x)dF(x)-\int_{-\infty}^{+\infty}F_i(x)dF(x)\Big)\Big].
\end{aligned}$$

Now assume $n(i)=n(0)$ for all i's. From the definition of F we easily verify that

$$\sum_{i=1}^{q}\int_{-\infty}^{+\infty}F_i(x)dF(x)=\sum_{i=1}^{q}\int_{-\infty}^{+\infty}F(x)dF(x)$$

which reduces the difference of the expected values to

$$\begin{aligned}
E\Re_1(\beta)-E\Re_1(\beta_0) &= \sum_{i=1}^{q}\frac{N(n(0)-1)}{(N+1)(n(0)+1)}\Big[\int_{-\infty}^{+\infty}(F_i(x)^2-F(x)^2)dF(x)\Big]\\
&= \frac{N(n(0)-1)}{(N+1)(n(0)+1)}\Big[\int_{-\infty}^{+\infty}\sum_{i=1}^{q}(F_i(x)^2-F(x)^2)dF(x)\Big]\\
&= \frac{N(n(0)-1)}{(N+1)(n(0)+1)}\Big[\int_{-\infty}^{+\infty}\sum_{i=1}^{q}(F_i(x)-F(x))^2dF(x)\Big].
\end{aligned}$$

Note that for the regression problem $F_i(x)=G(x-X_i(\beta-\beta_0))$ therefore

$$\begin{aligned}
&E\Re_1(\beta)-E\Re_1(\beta_0)\\
&=\frac{N(n(0)-1)}{(N+1)(n(0)+1)}\int_{-\infty}^{+\infty}\sum_{i=1}^{q}\Big[G(x-X_i(\beta-\beta_0)-\frac{1}{q}\sum_{i=1}^{q}G(x-X_i(\beta-\beta_0))\Big]^2\\
&\qquad d(\frac{1}{q}\sum_{i=1}^{q}G(x-X_i(\beta-\beta_0))).
\end{aligned}$$

Now assume G has a density g and assume the sequence of measures that put weight $1/q$ at the points X_i converge weakly to some probability measure μ. then

$$\lim_{q\to\infty} \frac{E\Re_1(\beta) - E\Re_1(\beta_0)}{q}$$
$$= \int_{-\infty}^{+\infty} E[G(x - V(\beta - \beta_0)) - EG(x - V(\beta - \beta_0))]^2$$
$$Eg(x - V(\beta - \beta_0)))dx = h. \tag{3.1}$$

Where V is a random variable having probability measure μ and h is some constant depending of G, β and β_0..

Proposition 3.5 *If μ is such that h is strictly positive for all $\beta \neq \beta_0$. Then for any $\beta \neq \beta_0$*

$$\lim_{q\to\infty} P\{\beta \in CR\} = 0. \tag{3.2}$$

where CR is the confidence region given by the set of all β for which $\Re_1(\beta) \leq E\Re_1(\beta_0) + Z_\alpha\sqrt{Var(\Re_1(\beta_0))}$.

Proof:

First, $Var(\Re_1(\beta)) \leq CN$ follows form the representation of $\Re_1$ as the sum of q independent random variables (see the proof of Theorem 2.9 in Chapter 2 for more details). Next

$$\frac{N^2}{q^2} = \left(\sum_{i=1}^{q} \frac{n(i)}{q}\right)^2 \leq \frac{\sum_{i=1}^{q} n(i)^2}{q}.$$

Now

$$P\{\beta \in CR\}$$

$$
\begin{aligned}
&= P\{\Re_1(\beta) \leq E\Re_1(\beta_0) + Z_\alpha\sqrt{Var(\Re_1(\beta_0))}\} \\
&= P\{\Re_1(\beta) - E\Re_1(\beta) \leq E\Re_1(\beta_0) - E\Re_1(\beta) + Z_\alpha\sqrt{Var(\Re_1(\beta_0))}\} \\
&\leq P\{|\Re_1(\beta) - E\Re_1(\beta)| \geq E\Re_1(\beta) - E\Re_1(\beta_0) + Z_\alpha\sqrt{Var(\Re_1(\beta_0))}\} \\
&\leq \frac{Var(\Re_1(\beta))}{(E\Re(\beta) - E\Re_1(\beta_0) + Z_\alpha\sqrt{Var(\Re_1(\beta_0))})^2} \\
&\leq \frac{CN}{(hq)^2}.
\end{aligned}
$$

The second last inequality follows from Chebychev's inequality. Now the condition $\lim_{q\to\infty}\sum_{i=1}^{q} n(i)^2/q^2 = 0$ completes the proof. □

3.4 Algorithm

In this section we present an algorithm for computing the estimate $\hat{\beta}$ described above. The main idea of this algorithm is to perform a search for $\hat{\beta}$ in the set of β_J; $J = 1, \ldots, \binom{n}{p}$, defined below. Consider all possible subsamples of p different points and index them by J; $J = 1, \ldots, \binom{n}{p}$. Let β_J be the vector of coefficients of the regression surface passing through the p points of subsample J. The computation of such a β_J amounts to the solution of a linear system of p equations with p unknowns. The search procedure goes as follows: for each β_J we compute the corresponding residuals and then the corresponding statistics $S_q(\beta_J)$ which allows as to determine the argument $\hat{\beta}$ of the minimum of S_q and it provides us also with the curve of S_q as a function of β.

We note here that the number m of β's that should be examined is $\binom{n}{p}$ which varies rapidly in n and p. Rousseeuw and LeRoy (1984) used the same

basic idea for their algorithm to compute the **LMS** estimate. To overcome this rapidly growing number of search points, they proposed randomly selecting m subsamples where m is choosen in a way to insure a high probability of selecting a "good subsample". In the algorithm they are using such an m does not exceed 3000 for the extensive search option and 1500 for the quick search option. The values of m used in their program is given in Rousseeuw and LeRoy (1987) Table 2 page 199.

3.5 Numerical Results

We first consider the case $d = 1$ of univariate dependent variables. First we consider the model $y = 4x + \epsilon$ where ϵ is normal with mean 0 and variance 1. We make 5 observations $\{y_{ij} : j = 1, 2, \cdots, 5\}$ at $x_i = i; i = 1, 2, \cdots, 10$ The graph of $\Re(\beta)$ is given in Figure 3.1 and the minimum $\Re(\hat{\beta}) = 1.12111$ is obtained at $\hat{\beta} = 3.9033$.

Next we calculate the residuals $y_{ij} - \hat{\beta}x_i$. We then sample at random from the set of permutations of these residuals and then arrange these values in 10 blocks of 5 and recalculate $\Re$. The histogram of these resampled values of $\Re$ are given in Figure 3.2. We notice that the value $\Re(\hat{\beta})_0 = 1.12111$ obtained above is in the center of this histogram (at 33.6 percentile) so we conclude the linear model is compatible with our results. Finally the value $L_{0.05} = 1.88111$ is plotted on Figure 3.1 and the associated confidence interval is [3.647,4.096]. The corresponding least squares estimate and confidence interval are 3.892 and [3.782,4.001]. We repeated this experiment 50 times and the histogram

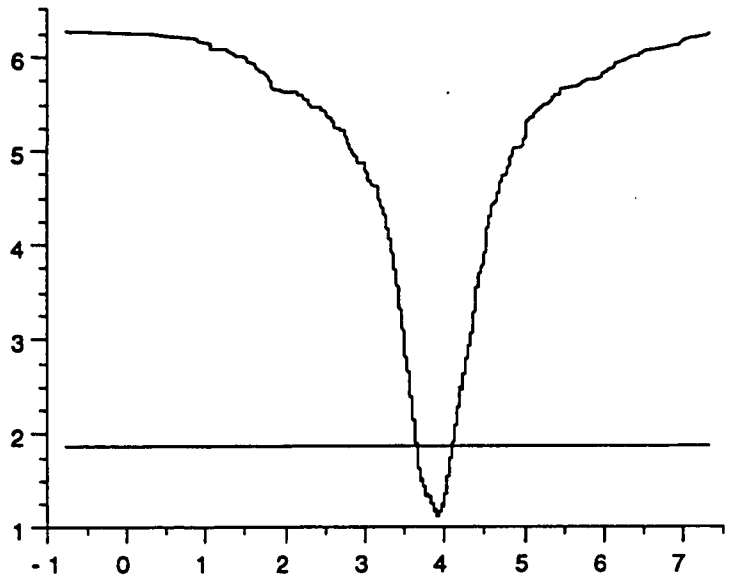

Figure 3.1: $\Re$ as a function of β

of the values of $\hat{\beta}$ is given in Figure 3.3. The histogram of the percentile of the observed value of $\Re(\hat{\beta})$ in the bootstrapped distribution of $\Re$ is given in Figure 3.4.

Suppose the above model is modified to $y = 4x+0.1x^2+\epsilon$ and we assumed that the model is still $y = 4x+\epsilon$. The graph of $\Re(\hat{\beta})$ is given in Figure 3.5 and the minimum is obtained at $\Re(\hat{\beta}) = 2.27444$ is obtained at $\hat{\beta} = 5.1516$. Next we calculate the residuals $y_{ij} - \hat{\beta}x_i$. Again we sample at random from the set of permutations of these residuals and then arrange these values in 10 blocks of 5 and recalculate $\Re$. The histogram of these resampled values of $\Re$ is given in Figure 3.6. We notice that the value $\Re(\hat{\beta}) = 2.27444$ obtained above is at the 99.3 percentile of this histogram. We conclude that the assumed linear

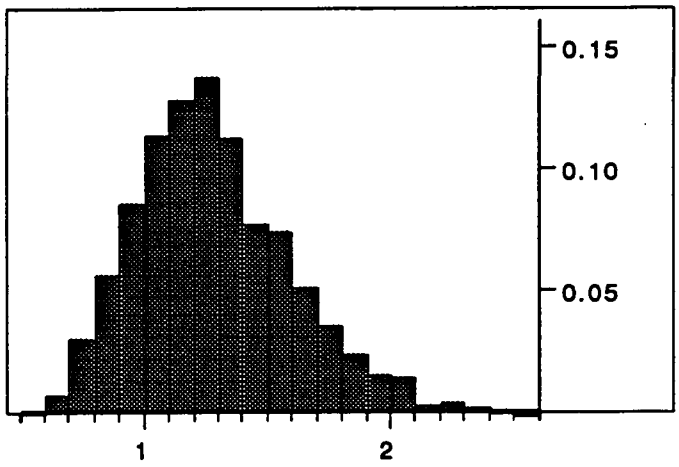

Figure 3.2: Histogram of the resampled values of $\Re$

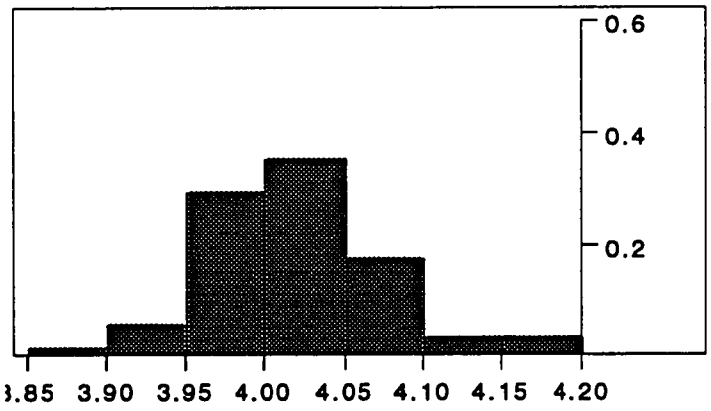

Figure 3.3: Histogram of the values of $\hat{\beta}$

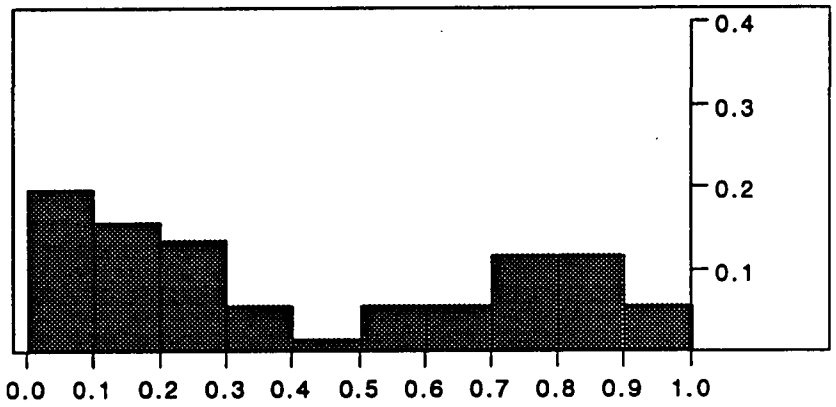

Figure 3.4: Histogram of the percentile of the observed value of $\Re(\hat{\beta})$

model $y = 4x + \epsilon$ must be rejected. Note that an application of the least square technique gives $\hat{\beta} = 5.1343$ a coefficient of determination of 0.9943 and the Shapiro-Wilk test of normality of the residuals is not significant *i.e.* the assumed linear model is accepted under these circumstances. We repeated the above 50 times. The result summarized in figures 3.7 and 3.8 shows that this is consistent fact not just a particular case.

Now modify the first model by changing the distribution of the error ϵ from a standard normal to a standard Cauchy distribution. The graph of $\Re(\beta)$ is given in Figure 3.9 and the minimum $\Re(\hat{\beta}) = 0.9478$ is obtained at $\hat{\beta} = 3.8478$. Again we calculate the residuals $y_{ij} - \hat{\beta}x_i$ and again we sample at random from the set of permutations of these residuals and then arrange these values in 10 blocks of 5 and recalculate $\Re$. The histogram of these resampled values of $\Re$ are given in Figure 3.10. We notice that the

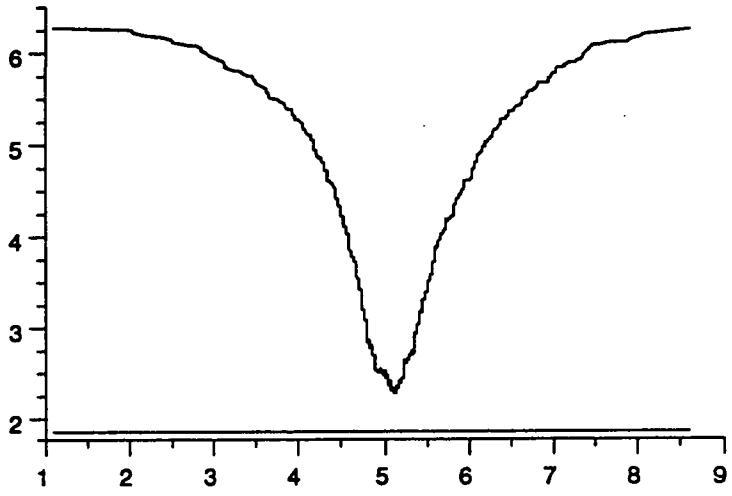

Figure 3.5: $\Re$ as a function of β

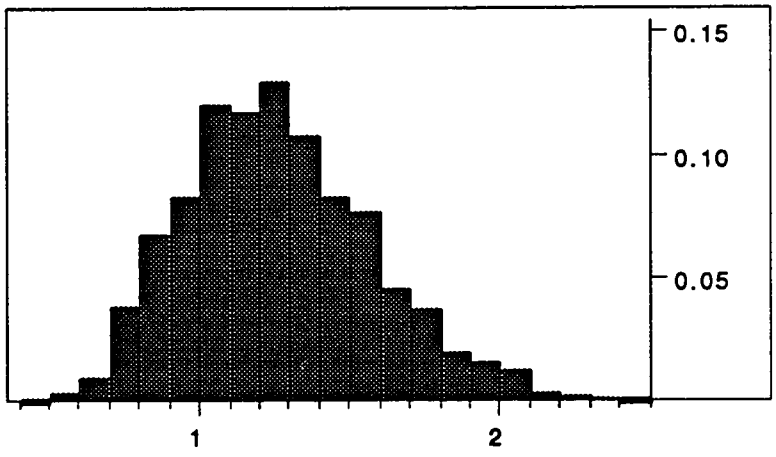

Figure 3.6: Histogram of the resampled values of $\Re$

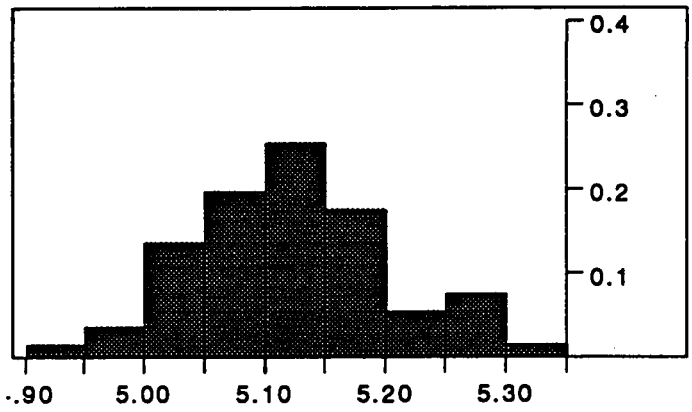

Figure 3.7: Histogram of the values of $\hat{\beta}$

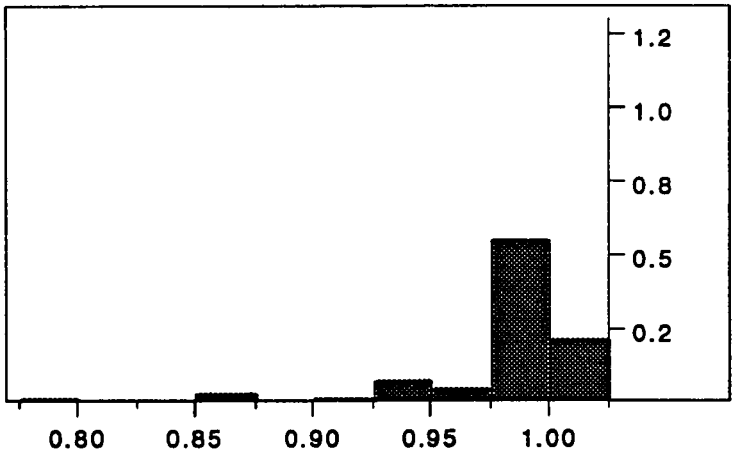

Figure 3.8: Histogram of the percentile of the observed value of $\Re(\hat{\beta})$

value $\Re(\hat{\beta}) = 0.9478$ obtained above is in the center of this histogram so we conclude the linear model is compatible with our results. Finally the value $L_{0.05} = 1.86111$ is plotted on Figure 3.9 and the associated confidence interval is [3.567,4.234]. The comparable least squares estimate for the slope is 3.289 and the associated confidence interval is [1.907,4.671]. We repeated the above experiment 50 times. Figure 3.11 gives the histogram of the observed values of $\hat{\beta}$ and shows the consistency of this value. It is clear that our nonparametric procedure is much more successful than the traditional least squares procedure which in some of the above experiments gave estimates like -30.

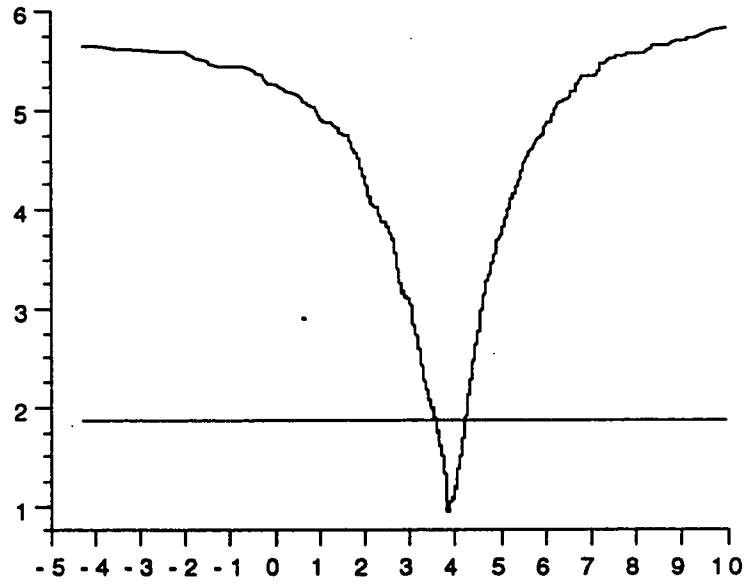

Figure 3.9: $\Re$ as a function of β

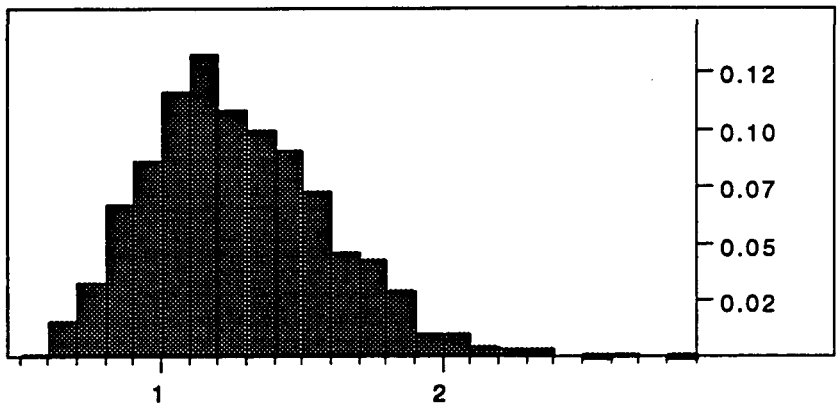

Figure 3.10: Histogram of the resampled values of $\Re$

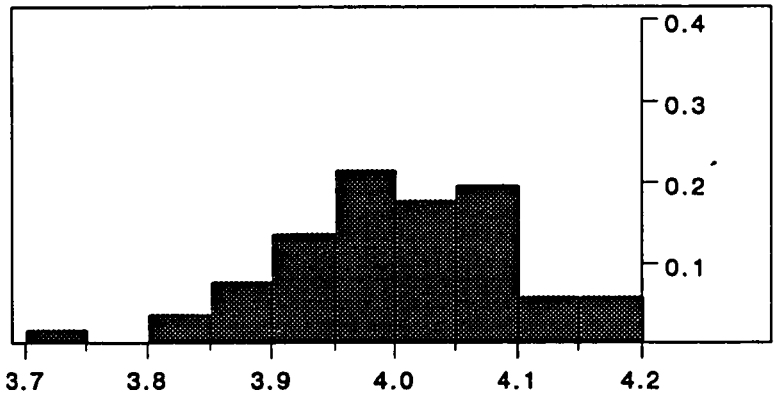

Figure 3.11: Histogram of the values of $\hat{\beta}$

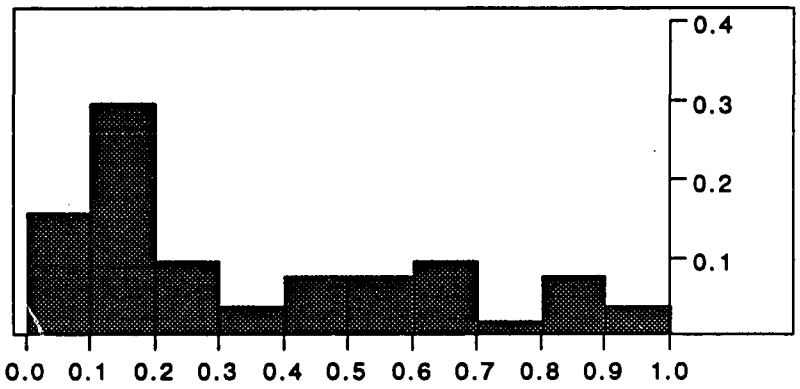

Figure 3.12: Histogram of the percentile of the observed value of $\Re(\hat{\beta})$

Now we consider the Lactic acid concentration data set given by Afifi and Azen (1979), see also Rousseeuw and Leroy (1987). This data give the true concentration (x_i) and the measured concentration (y_i) during the calibration af an instrument. It fits perfecty our framework in the sense that we have replicates for the same value of the design parameter (x). A plot of this data with the least squares, the least median squares and our estimate is given in Figure 3.13. It shows that the estimate given by our method is very comparable to both the least squares and the least median squares estimates. Now we shall introduce an outlier to the above set of observation; in fact we will deform the value of one of the observations. Figure 3.14 shows the new data set as well as the estimate given by the above three methods. This shows that our estimate is less sensitive to outliers.

We now consider the case of multivariate dependent variables ($d > 1$).

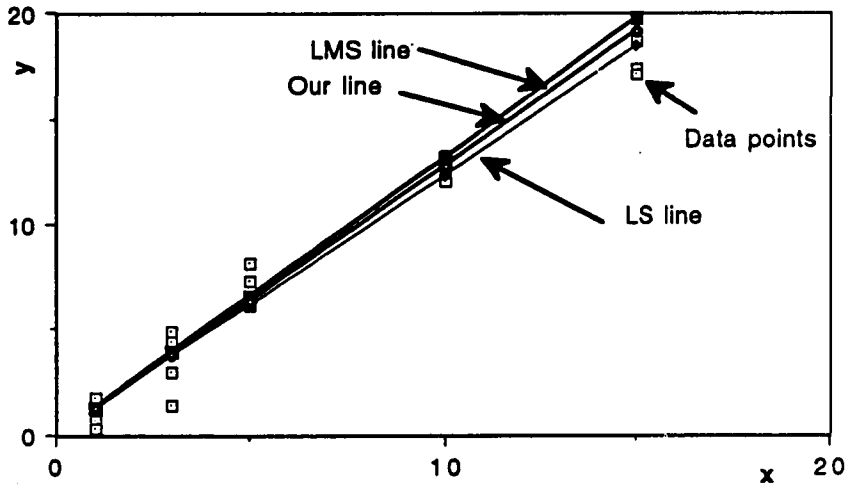

Figure 3.13: Lactic acid calibration data

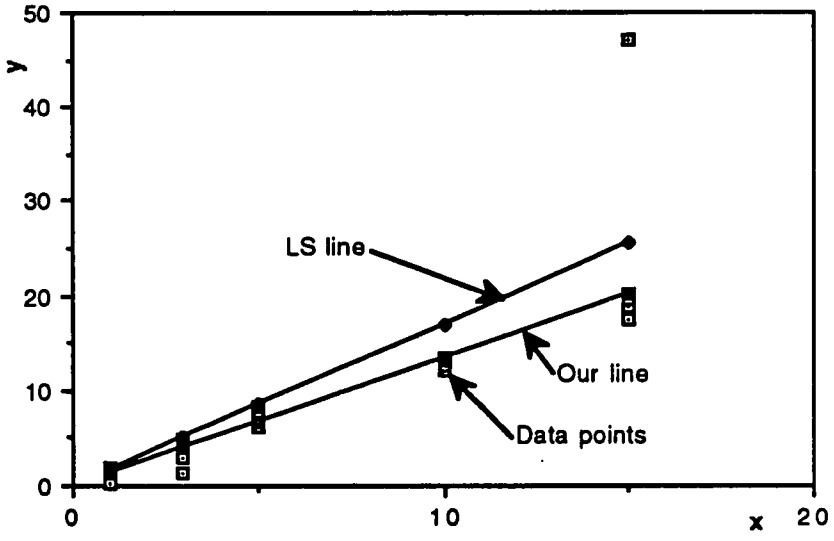

Figure 3.14: Modified lactic acid calibration data

Consider the model $(y(1), y(2)) = (4,1)x + (\epsilon(1), \epsilon(2))$ where the components of ϵ are normal with mean 0 and variance 1 and the correlation is 0.5. We make 5 observations $\{(y_{ij}(1), y_{ij}(2)) : j = 1, 2, \cdots, 5\}$ at $x_i = i; i = 1, 2, \cdots, 10$. The graph of $\Re(\beta)$ is given in Figure 3.15 and the minimum $\Re(\hat{\beta}) = 4.78$ is obtained at $\hat{\beta} = (4.07, 1.08)$. Next we calculate the residuals $(y_{ij}(1), y_{ij}(2)) - \hat{\beta}x_i$. We then sample at random from the set of permutations of these residuals and then arrange these values in 10 blocks of 5 and recalculate $\Re$. The histogram of these resampled values of $\Re$ are given in Figure 3.16. We notice that the value $\Re(\hat{\beta}) = 4.78$ obtained above is in the center of this histogram so we conclude the linear model is compatible with our results. Finally the value $L_{0.05} = 6.552$ is plotted on Figure 3.16. The comparable least squares estimate for the slope is (4.08,1.07).

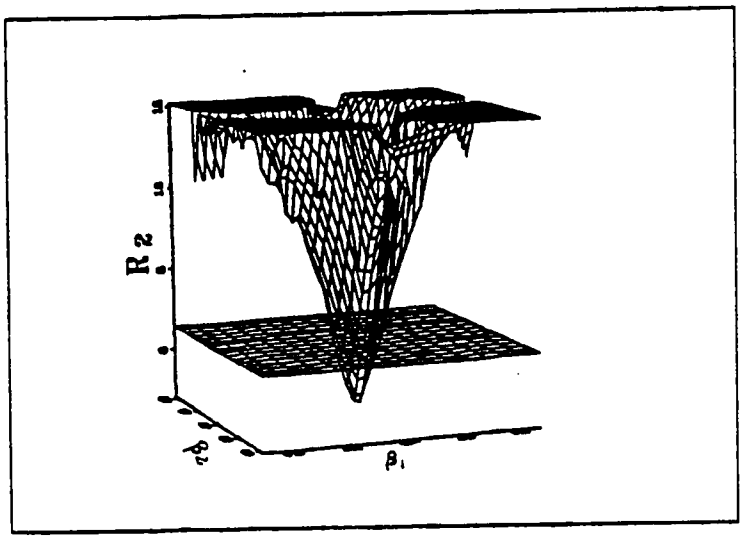

Figure 3.15: $\Re$ as a function of β

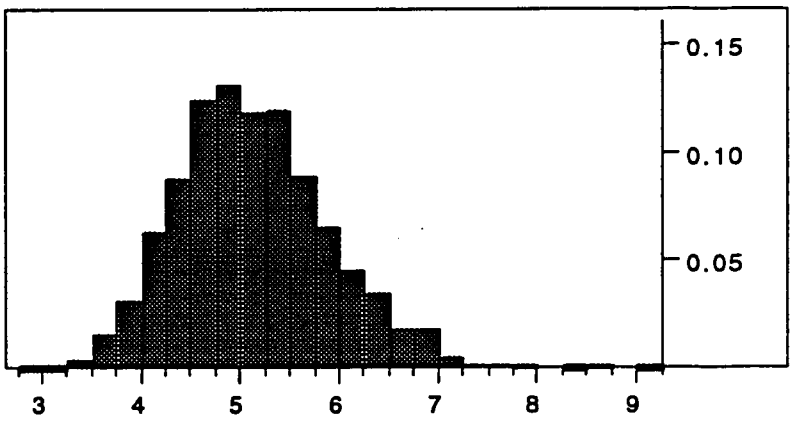

Figure 3.16: Histogram of the resampled values of $\Re$

Chapter 4

Multivariate quality control

4.1 Literature review

Installing an industrial quality control procedure requires three steps. We start with a nonparametric quality control procedure. After a period of operation without an *out of control* signal we decide to use the information collected to get precise values of the control limits. But before doing so we need to make sure that the information collected is representative of an *in control or on target* process. We, therefore, need to implement a retrospective, preferably nonparametric, quality control scheme. Finally, if the retrospective scheme does not signal, we use the new estimates of our process parameters to build sequential parametric quality control procedure. Next we shall give a quick review of the most common quality control procedures. We begin by dividing these procedures into posterior detection schemes and on line detection schemes.

4.1.1 Posterior detection

This class of methods deal with the first problem mentioned above and can be divided into two major subclasses which we describe in the following sections.

Parametric posterior detection

Hinkley (1970) , Darkhovskii (1976) and Darkhovskii and Brodskii (1980,1987) posed the problem in the following way.
Let $X_1, \ldots, X_N$ be a sequence of random variables and it is known that $X_1, \ldots, X_{n_0}$ have known distribution function F_0 and the remaining variables have known distribution function F_1 The problem consists of giving an estimate of the change point n_0 based on the sequence $X_1, \ldots, X_N$. The problem of testing if the sequence $X_1, \ldots, X_{n_0}$ contains a change was considered by Chernoff and Zacks (1964), Kander and Zacks (1966) and Gardner (1969).

Nonparametric Posterior detection

Sen and Srivastava (1975) presented a non-parametric method for testing whether the means of each variable in a sequence of independent random variables can be taken to be the same, against alternatives that a shift might have occurred after some point r. The same problem has been studied by Page (1954,1955) and Bhattacharya and Johnson (1968). Most of the papers on non-parametric posterior detection were focussed on a shift in the mean of the random variables.

4.1.2 Fastest detection of a change in distribution

In this section we will consider methods devoted to the detection of a change in distribution. A very wide literature exists on this subject and for more detail the reader may refer to Jandhyala (1985) who presented detailed literature review and a wide bibliography of this subject. As in the previous section we divide the methods into parametric and non-parametric and we start our description by the parametric ones.

Parametric Approach

The parametric approach itself can be divided into two subclasses. The first class uses a Bayesian approach and the second class groups the non-Bayesian approaches to the problem.

The Bayesian approach to the problem was first presented and solved by Shiryayev (1963, 1978). He posed the problem as stated below. Suppose one is able to sequentially observe a series of independent observations $X_1, X_2, \ldots$ whose distributions may change from a known distribution F_0 to a known distribution F_1 at some unknown point in time ν. Formally $X_1, X_2, \ldots$ are independent random variables such that $X_1, X_2, \ldots, X_{\nu-1}$ are independent identically distributed with distribution function F_0 and $X_\nu, X_{\nu+1}, \ldots$, are independent identically distributed with distribution F_1 and where ν is referred as the disorder time and is such that $1 \leq \nu \leq \infty$. He introduced a cost structure in which he assumed that one loses one unit (i.e inspection cost) if one stops before the change (i.e $\tau < \nu$) and one loses $c(\tau - \nu)$ units if one

stops after the change (i.e $\tau \geq \nu$). He also supposed that the disorder time ν has a geometric prior distribution. He proved that the optimal stopping rule is of the form $\tau^* = \inf\{n : \pi_n \geq A\}$, where π_n is the conditional probability that the change occurred given $(X_1, \ldots, X_n)$ and where A is a constant.

Next we cite the most known non-Bayesian approaches dealing with the disorder problem. The first optimality results in this context were presented by Lorden (1971). His results are based on the restriction that the stopping rule τ must satisfy $\mathbf{E}(\tau/\nu = \infty) \geq B$ (this denote the expectation of the stopping rule giving that no change has occured. It also controls the false alarm rate). His criteria of selecting the best stopping rule is $\{ess\sup \mathbf{E}(\tau - \nu + 1)^+ / X_1, \ldots, X_{\nu-1})\}$ which represents the speed with which the stopping time detects the change. He proved that a certain class of stopping rules is asymptotically (i.e $B \rightarrow \infty$) optimal and that Page's procedure (Page 1954) belongs to this class. A generalization of his results are presented by Moustakides (1986) .

Bojdecki (1979,1984) presented a probability maximizing approach for a somewhat generalized disorder problem. His aim was to find a stopping rule τ^* such that $P\{|\tau^* - \nu| \leq m\} = \sup\{P\{|\tau - \nu| \leq m\}; \tau \in \mathcal{T}\}$ where $\mathcal{T}$ is the set of all possible stopping rules. His generalization of the problem consists of considering the shift to an unknown distribution F_1 that belongs to a finite (or countable) family of distributions.

Pollak (1985) presented a non-Bayesian setting in which he derived a stopping rule τ that minimizes $\mathbf{E}(\tau/\tau \geq \nu)$ subject to the constraint $\mathbf{E}(\tau/\nu = \infty) \geq B$. He proved that this stopping time can be written as the limit of a sequence of Bayesian rules identical to those presented by Shiryayev (1963). He also presented an almost minimax stopping rule.

Nonparametric approach for the disruption problem

This set of methods is also divided in two subclasses; those using a Bayesian technique and those using non-Bayesian techniques.

First we cite the work of Zacks (1981) who used the same framework and the same cost structure as Shiryayev (1963) to solve a non-parametric version of the disorder problem. He considered the case of a change from an unknown binomial distribution with parameter θ to an unknown binomial distribution with parameter $\varphi > \theta$. He also assumed a prior $h(\theta, \varphi)$. He showed that the optimal stopping rule is of the form:

$$\tau^* = \inf\{n : \pi_n(\mathbf{X}_n) \geq b_n(\mathbf{X}_n)\},$$

where $\pi_n(\mathbf{X}_n)$ has the same definition as π_n above except that it depends on the past observations $\mathbf{X}_n = (X_1, \ldots, X_n)$ and where $b_n(\mathbf{X}_n)$ is a function of n and $\mathbf{X}_n$. Although the case considered is very simple the computation of $b_n(\mathbf{X}_n)$ is extremely difficult if not impossible.

Battacharya and Frierson (1981) designed a nonparametric but not completely sequential cumulative sums (Cusum) procedure to detect small disorders based on sequential ranks. The main result of their paper deals with

the asymptotic behavior of such procedure. McDonald (1990) presented a nonparametric cumulative sum procedure based also on sequential ranks designed to detect a change in the sampling distribution to a stochastically larger distribution. Gordon and Pollak (1991,1992) developed a nonparametric analogue of the Shiryayev procedure based on sequential ranks. Hackl and Ledolter (1989) presented a nonparametric technique using exponentially weighted moving averages(EWMA) on sequential ranks. Their sequential ranks were different than in McDonald (1990) and Battacharya (1981). They ranked the last observation among the g last observations where g is a fixed parameter for the procedure. Darkhovskii and Brodskii (1988) discussed a nonparametric method for the fastest detection of a change in the mean of a random sequence. They didn't base their analysis on the whole past but they only consider the last N observations (N is called memory size). They established that for a special choice of N their method is asymptotically optimal for a sequence of independent random variables.

4.1.3 Multivariate case

The literature dealing with the multivariate case is mostly devoted to practical rules such as Cusums and T-charts. Crosier (1988) proposed two different techniques for constructing multivariate Cusum procedures. The first technique consists in reducing each multivariate observation to a scalar and then constructing the cusum on the scalars. The second procedure consists in forming a cusum vector directly from the observations. The same ideas are also discussed by Woodall and Ncube (1985). Hotelling (1950) and Jackson

(1959,1979,1957) proposed some multidimensional T-charts to test if past observations are in control.

4.2 Introduction

Next we present a retrospective scheme based on the family of *randomness* statistics defined in Chapter 2. More precisely we shall consider two particular statistics of this family and based on these we shall construct tests for the *off-line quality control* problem. We note that the above statistics are defined for any dimension of the quality measurement therefore they can be used for univariate observation as well as multivariate observations. In the simulation study given in sections 2 and 3 we limit ourselves to bivariate observations, for simplicity.

Recall the following *randomness* statistics defined in Chapter 2.

$$\Re_2 = \sum_{i=1}^{q} \sum_{j=1}^{n(i)} \|\vec{F}^i(U_{ij}) - \vec{F}^\bullet(U_{ij})\|^2 \tag{4.1}$$

$$\Re_3 = \sum_{i=1}^{q} \sum_{j=1}^{n(i)} \|\vec{F}^i(U_{ij}) - \vec{F}^\bullet_{X,Y}(U_{ij})\|^2 \tag{4.2}$$

In a typical quality control situation the components X and Y of U are independent (i.e both X and Y consist of a nominal value plus a noise). An out of control situation induces a change in the joint distribution of X and Y. The most common change are shifts in the mean of these random variables, but in reality the processes are more complex and a loss of control may result in a correlation of these components or a more realistic mixture

of shifts, correlation and scaling. In the rest of this section we shall conduct a simulation study with the above statistics and we shall try to see their performances in detecting an out of control situation. This study will be decomposed into two parts. The first part uses the central limit theorem proved in Chapter 2 to construct a test of the *in-control* situation. We also note that a second approach using a permutation test, introduced in Chapter 3 with the statistics $\Re_2$ and $\Re_3$, can be applied. The method is exactly equivalent to the one used in Chapter 3. It consists in permuting the quality observations and testing if the sequence observed forms an unlikely permutation (i.e the value of the statistic $\Re_2$ or $\Re_3$ associated with this permutation falls in the upper tails of the of the distribution of $\Re_2$ or $\Re_3$ for all possible permutations). Since the method is essencially the same as in Chapter 3, here we put more emphesis on the use of the central limit theorem. Note that the expected values and the variances of $\Re_2$ and $\Re_3$ can be given explicitly in terms of some functional of the underlying distribution function. For example,

$$\begin{aligned} E\Re_2 \;=\; & \sum_{i=1}^{q} \frac{1}{n(i)n^2} \Big[(6nn(i) + n^2 n(i) - n(i)^2 n - 3n(i)^2 - 3n^2) E\{F_a(U)\} \\ & + (2n^2 + n(i)^2 n + 2n(i)^2 - 4nn(i) - n^2 n(i)) \\ & \times (E\{F_a(U)\}^2 + E\{F_b(U)\}^2 + E\{F_c(U)\}^2 + E\{F_d(U)\}^2) \\ & + (-n(i)^2 n - n(i)^2 + 2nn(i) - n^2 + n^2 n(i)) \\ & \times (E\{F_b(U)\} + E\{F_c(U)\} + E\{F_d(U)\}) \\ & + -2nn(i) + n(i)^2 + n^2 \Big]. \end{aligned} \tag{4.3}$$

Under the hypothesis of independent components, computation is much simpler and we have

$$E\Re_2 = \sum_{i=1}^{q} \frac{-10n(i)^2n+10n^2n(i)+7n(i)^2+7n^2-14nn(i)}{18n^2n(i)} \tag{4.4}$$

and

$$E\Re_3 = \sum_{i=1}^{q} \frac{10n(i)n^3(n-1)+7n(i)^2n^2-8n(i)n^2-8n^3n(i)^2+2n(i)^2+7n^4}{18n(i)n^4}. \tag{4.5}$$

The computation of the variances of $\Re_2$ and $\Re_3$ in terms of functionals of the distribution function is theoretically possible, but very messy. For the independent component hypothesis both $\Re_2$ and $\Re_3$ are distribution free, therefore one can easily estimate these variances using Monte Carlo simulation. Table (4.1) gives, for the sets of parameters (q and $n(i)$) we are interested in, the expected values $\Re_2$ and $\Re_3$ given by formulas (4.4) and (4.5) respectively, and estimate of the variances of $\Re_2$ and $\Re_3$, given by 1000 replications in a Monte Carlo simulation, under the hypothesis of independent components.

We recall the statistic $\Re$ introduced in Chapter 3

$$\Re = \sum_{i=1}^{q}\sum_{j=1}^{n(i)}(F^i(U_{ij}) - F^\bullet(U_{ij}))^2. \tag{4.6}$$

In fact $\Re$ represent the first piece in the four pieces constituting $\Re_2$. The asymptotic variance of $\Re$ is given by

$$Var\Re \simeq \sum_{i=1}^{q}\left[C_4 + \frac{C_3}{n(i)} + \frac{C_2}{n(i)^2} + \frac{C_1}{n(i)^2} + \frac{q}{N}\left(B_2 + \frac{2B_1}{n(i)}\right)\right] \tag{4.7}$$

where

$$C_4 = E\{F(X_1)F(X_2) - F(X_1 \wedge X_2)\}^2.$$

Expresions for C_3, C_2, C_1, B_2 and B_1 shall be given in the appendix. One can also find in this appendix the Maple program that computed these expressions.

Note the quantity inside the brackets in (refvar) is a variance and hence is non-negative. Also note that if $C_4 > 0$ then $Var\Re$ will be of order q provided $\sum_{i=1}^{q} I\{n(i) \geq n(0)\} = O(q)$ here $n(0)$ is such that the quantity inside the brackets in (4.7) is strictly positive for $n(i) \geq n(0)$. We conjecture that such $n(0) \leq 3$. The heuristics behind this conjecture is that, for equal $n(i)$'s for example, such expression is polynomial of degree 3 of $n(1)$ divided by $n(1)^3$ which is greater or equal to zero for all integers $n(1)$. Therefore it can only be zero at most in three integers two of which must be consecutive. Due to the nature of the statistic we see that if the $Var\Re$ is of order q for equal $n(i)$'s satisfying $n(i) \geq m$ it must also be of order q for equal $n(i)$'s satisfying $n(i) \geq m + 1$.

Remark

In one dimension (i.e if $d = 1$) one can see that $C_4 > 0$ and we can even give its value for all continous distributions. If $d > 1$ C_4 may be zero, moreover, we can exhibit a distribution function with uniform marginals for which $C_1 = \cdots = C_4 = B1 = B2 = 0$. For example for $d = 2$ this distribution function is defined on the unit square by $F(x, y) = \max\{0, x + y - 1\}$. It corresponds to the Fréchet lower bound (see Genest and MacKay (1986)).

q	$n(i)$	$E\Re_2$	$Var\Re_2$	$E\Re_3$	$Var\Re_3$
10	5	5.6300	0.8111	5.7837	0.8594
50	5	30.9571	4.2484	31.1123	4.5927
100	5	62.6230	9.0052	62.7783	9.4183
50	10	29.0897	4.3006	29.2229	4.3011

Table 4.1: Expected values and variances of $\Re_2$ and $\Re_3$

4.3 Use of the central limit theorem

In this section we consider quality measurements which are bivariate random variables with independent components under the *in control* hypothesis. We shall introduce a change of the distribution at some point in time and we shall see how the statistics $\Re_2$ and $\Re_3$ react to this change. Different changes, ranging from a shift in mean to a combination of shifts and correlation, shall be considered.

4.3.1 Shift in one direction

In this section we consider an out of control situation that can be summarized by a shift in the mean of one of the components. Precisely we shall carry the following simulation study. Assume we are observing quality measurements U_{ij} for $i = 1, \ldots, q$ and $j = 1, \ldots, n(i)$. We assume $U_{ij} = (X_{ij}, Y_{ij})$ where for $i = 1, \ldots, \tau$ X_{ij} and Y_{ij} are independent normal with mean 0 and variance 1 and for $i = \tau + 1, \ldots, q$ X_{ij} becomes a normal with mean s and variance 1. We only consider the case of equal $n(i)$'s. The simulation study consist in generating random variables according to the above scheme, and for each set

of $N = \sum_{i=1}^{q} n(i)$ we compute the values of the statistics $\Re_2$ and $\Re_3$. We say that the statistic $\Re_2$ (respectively $\Re_3$) detects an the out of control situation if the observed value of $\Re_2$ (respectively $\Re_3$) is bigger than $E\Re_2 + Z_\alpha\sqrt{Var\Re_2}$ (respectively $E\Re_3 + Z_\alpha\sqrt{Var\Re_3}$), where $E\Re_2$ and $Var\Re_2$ (respectively $E\Re_3$ and $Var\Re_3$) are given in Table (4.1). Here Z_α is $(1-\alpha)100$ quantile of the normal distribution and in fact we just consider $\alpha = 0.05$. Based on 1000 repetitions of the above procedure Table (4.2) gives the percentage of times the statistic $\Re_2$ (respectively $\Re_3$) detects an out of control situation. The last column of this table gives the probability that a Shewart chart detects an out of control situation. Note that this Shewart chart is the traditional $\bar{X}$ chart designed such that for a given q and $n(i)$'s and under the hypothesis of no change in distribution the probability of a false alarm is α (i.e the same as for $\Re_2$ and $\Re_3$).

It is worth noting here that both statistics are performing comparably to the parametric Shewart procedure. Is also important to remark that the conditions of the simulation were chosen to get the best performance of the Shewart chart (i.e normal observations with known mean and variance).

4.3.2 Shift in both directions

In this section we consider changes in distribution that consist of a shift of the mean of both components X and Y. We first restrict ourselves to equal shifts in both directions. In fact the setting of this experiment is exactly as in the preceding section with the exception that both X and Y get shifted by s after time τ. Table (4.3) gives a summary of a 1000 replication simulation study.

q	$n(i)$	τ	Shift	$\Re_3$	$\Re_2$	Shewart
100	5	100		5	5	5
100	5	80	0.5	16.2	17.2	22.2
100	5	80	1	71.0	72.6	91.7
100	5	80	1.5	99.4	99.0	100
100	5	80	2	100	100	100
50	10	50		5	5	5
50	10	30	0.5	45.6	45.6	60.9
50	10	30	0.7	84.3	84.4	95.5
50	10	30	1.0	99.9	100	100
50	10	30	1.5	100	100	100
50	10	30	2.0	100	100	100

Table 4.2: Percentage of time the statistics $\Re_3$ and $\Re_2$ detected a shift in the x coordinate

As in Table (4.2) the last column of Table (4.3) gives the exact probability that a Shewart $\bar{X}$ and $\bar{Y}$ chart, designed such that for a given q and $n(i)$'s the probability of false alarm is 0.05, detect an loss of control.

Comparing with the above results, one can see that the performance of both statistics is improving, in fact both performances are getting closer to the performance of the Shewart chart and even excelling it for small shifts.

4.3.3 Correlation

The change in distribution considered in this section is very particular. In fact after time τ we keep both the marginals of X and Y unchanged and we make X and Y correlated with a correlation coefficient ρ. Table (4.4) has the same structure as Table (4.2) and it summarize the simulation results for

q	$n(i)$	τ	Shift	$\Re_3$	$\Re_2$	Shewart
100	5	100	0.5	5	5	5
100	5	80	0.5	27.6	22.4	36.3
100	5	80	1	71.0	94.5	99.3
100	5	80	1.5	100	100	100
100	5	80	2	100	100	100
50	10	50		5	5	5
50	10	30	0.5	83.9	80.4	84
50	10	30	0.7	99.8	99.7	99.7
50	10	30	1.0	100	100	100
50	10	30	1.5	100	100	100
50	10	30	2.0	100	100	100

Table 4.3: Percentage of time the statistics $\Re_3$ and $\Re_2$ detected a shift in both coordinate x and y

the above experiment. In this case the Shewart chart is the same as in the shift in two directions case. It essentially behaves like under the no change in distribution hypothesis.

One immediate conclusion is that the statistic $\Re_3$ is out performing, by a large margin, both the statistic $\Re_2$ and the Shewart chart.

4.3.4 Combined shifts and correlation

In this section we consider a more general change in the distribution of the quality measurements. In fact after the change point X and Y become correlated with a correlation coefficient ρ and their means get shifted by small shift s. Table (4.5 gives the simulation results for this case. The Shewart chart is the same as in the shift in two directions case. It is only

q	$n(i)$	τ	Correlation	$\Re_3$	$\Re_2$
100	5	100		5.0	5.0
100	5	80	0.3	8.4	4.8
100	5	80	0.5	9.9	3.8
100	5	80	0.7	14.9	4.3
100	5	80	0.9	36.5	5.4
50	10	50		5.0	5.0
50	10	30	0.3	16.9	6.6
50	10	30	0.5	34.3	5.0
50	10	30	0.7	74.6	12.5
50	10	30	0.9	100	26.9

Table 4.4: Percentage of time the statistics $\Re_3$ and $\Re_2$ detected a correlation of the two coordinate x and y

sensitive to the shift and it behaves as if the only change is a shift in the mean (see Table 4.2).

Here we note that the statistic $\Re_3$ is collecting the effect of different changes and is therefore performing better than the parametric Shewart chart which only detects the presence of shifts.

4.4 Conclusion

It is worth mentioning, that the precision of these *randomness* statistics increases when the sample sizes $n(i)$'s increase. This is be explained by the fact that when the size, $n(i)$, of sample i increases, the empirical distribution function F^i gets closer to the true distribution function of this sample, say F_i and any deviation from the global empirical is easier to detect.

q	$n(i)$	τ	Shift	Correl	$\Re_3$	$\Re_2$
100	5	100			5	5
100	5	80	0.5	0.3	34	18.2
100	5	80	0.5	0.5	39.5	18
100	5	80	0.5	0.7	47.9	14.8
100	5	80	0.7	0.3	65.3	44.3
100	5	80	0.7	0.5	70.1	42.4
100	5	80	0.7	0.7	78.8	43.0
50	10	30	0.5	0.3	91.3	78.1
50	10	30	0.5	0.5	97.5	81.4
50	10	30	0.5	0.7	99.7	87.1
50	10	30	0.7	0.3	99.9	99.7
50	10	30	0.7	0.5	100	99.7

Table 4.5: Percentage of time the statistics $\Re_3$ and $\Re_2$ detected a mixture of shifts and correlation of the two coordinates x and y

Appendix A

Asymptotic variance of $\Re$

In this appendix we give the expressions of the constants C_4, C_3, C_2, C_1, B_2 and B_1 introduced in Chapter 4. Note that throughout this appendix X_1 and X_2 denote two independent identically distributed random vectors having distribution F. We let

$$p_1 = U(X_1, X_2), \qquad p_2 = U(X_2, X_1)$$

where $U(X_1, X_2) = 1$ if $X_2 \leq X_1$ and is equal to zero otherwise. Also

$$p_3 = F(X_1), \qquad p_4 = F(X_2), \qquad p_5 = F(X_1 \wedge X_2),$$

$$F_1 = E\{F(X_1)\},\ F_2 = E\{F(X_1)\}^2,\ F_3 = E\{F(X_1)\}^3 \text{ and } F_4 = E\{F(X_1)\}^4.$$

We have

$$\begin{aligned} C_3 \;=\; & E\big[-3p_2p_3^2 - 57F_2^2 + 52F_2F_1 - 13F_1^2 + 6p_1p_5 + 6p_2p_5 - 12p_5^2 \\ & -3p_1p_4^2 - 12p_1p_4p_5 + 2F_2 - 18p_4p_5 - 2F_1 + 16p_1p_3p_4^2 - p_3^2 + 3F_4 \\ & -12p_2p_3p_5 + 3p_2p_3 + 3p_1p_4 - 10p_1p_4p_3 - 10p_2p_4p_3 - 18p_3p_5 \\ & +16p_2p_3^2p_4 + 60p_4p_3p_5 - 6F_1^3 - p_4^2 + p_3 + 9p_5 + p_4\big] \end{aligned}$$

$$C_2 = E\Big[21p_2p_3^2 + 174F_2^2 - 192F_2F_1 + 19F_1^2 - 18p_1p_5 - 18p_2p_5 + 22p_5^2$$
$$+21p_1p_4^2 + 36p_1p_4p_5 - 4F_2 + 54p_4p_5 + 17F_1 + 3p_1 - 4p_1p_3 + 3p_2 - 64p_1p_3p_4^2$$
$$+7p_3^2 - 26F_4 + 36p_2p_3p_5 - 21p_2p_3 - 21p_1p_4 + 46p_1p_4p_3 + 46p_2p_4p_3 - 4p_2p_4$$
$$+54p_3p_5 - 64p_2p_3^2p_4 - 152p_4p_3p_5 + 62F_1^3 + 7p_4^2 - 7p_3 - 27p_5 - 7p_4\Big]$$

$$C_1 = E\Big[-18p_2p_3^2 - 120F_2^2 + 144F_2F_1 - 4F_1^2 + 12p_1p_5 + 12p_2p_5 - 12p_5^2$$
$$-18p_1p_4^2 - 24p_1p_4p_5 - 36p_4p_5 - 15F_1 - 3p_1 + 4p_1p_3 - 3p_2 + 48p_1p_3p_4^2 - 6p_3^2$$
$$+24F_4 - 24p_2p_3p_5 + 18p_2p_3 + 18p_1p_4 - 36p_1p_4p_3 - 36p_2p_4p_3 + 4p_2p_4$$
$$-36p_3p_5 + 48p_2p_3^2p_4 + 96p_4p_3p_5 - 60F_1^3 - 6p_4^2 + 6p_3 + 18p_5 + 6p_4\Big]$$

$$B_2 = E\Big[2p_2p_3^2 + 16F_2^2 - 28F_2F_1 + 12F_1^2 + 2p_1p_4^2 - 4F_2 + 10p_4p_5 + 4F_1$$
$$-2p_1p_3p_4^2 + 2p_3^2 - 2p_2p_3 - 2p_1p_4 + 2p_1p_4p_3 + 2p_2p_4p_3$$
$$+10p_3p_5 - 2p_2p_3^2p_4 - 12p_4p_3p_5 + 2p_4^2 - 2p_3 - 8p_5 - 2p_4\Big]$$

$$B_1 = E\Big[6p_1p_4 - 12F_2^2 - 6p_1p_4p_3 + 8p_4p_3p_5 - 6p_3p_5$$
$$-12F_1^2 + 24F_2F_1 + 2p_1p_3 - 4p_1p_4^2$$
$$-12F_1 + 8F_2 - 8p_4p_5 - 4p_4^2 + 8p_3 - 2p_1 + 6p_5 + 6p_4 + 4p_1p_3p_4^2 - 6p_3^2\Big]$$

In the rest of this appendix we give the Maple program used to compute the above coefficients. Note that the same program can be used to compute the exact variance, not just the asymptotic variance, but the result becomes lengthy.

Maple program computing the variance of R

```
#For the multi dimensional case
#This maple program computes the expected value And the variance of R
# We use a moment generating function for the random vector
# M(s,t) and we compute the partial derivative at (0,0)
# a and b are defined later f, g, k and h are used to simplify the computation.
printlevel:=-1;
M:=f(s,t)*g(s,t)^a*k(s,t)^b;
M1122:=simplify(diff(M,s,s,t,t));
M1122:=subs(diff(f(s,t),s,s,t,t)=f1122(s,t), diff(g(s,t),s,s,t,t)=g1122(s,t),
diff(k(s,t),s,s,t,t)=k1122(s,t), diff(k(s,t),s,s,t)=k112(s,t),
diff(k(s,t),s,t,t)=k122(s,t), diff(k(s,t),s,s)=k11(s,t), diff(k(s,t),t,t)=k22(s,t),
diff(k(s,t),s,t)=k12(s,t), diff(k(s,t),t,s)=k21(s,t), diff(k(s,t),t)=k2(s,t),
diff(k(s,t),s)=k1(s,t), diff(f(s,t),s,s,t)=f112(s,t), diff(g(s,t),s,s,t)=g112(s,t),
diff(f(s,t),s,t,t)=f122(s,t),diff(g(s,t),s,t,t)=g122(s,t),
diff(f(s,t),s,s)=f11(s,t),diff(g(s,t),s,s)=g11(s,t),
diff(f(s,t),s,t)=f12(s,t),diff(g(s,t),s,t)=g12(s,t),
diff(f(s,t),t,s)=f21(s,t),diff(g(s,t),t,s)=g21(s,t),
diff(f(s,t),t,t)=f22(s,t),diff(g(s,t),t,t)=g22(s,t),
diff(f(s,t),t)=f2(s,t),diff(g(s,t),t)=g2(s,t),
diff(f(s,t),s)=f1(s,t),diff(g(s,t),s)=g1(s,t),M1122);
printlevel:=-1;
t:=0; s:=0;a:=y-2;b:=n-y;
f(0,0):=1;
f1(0,0):=(n-y)*(1+p1); f2(0,0):=(n-y)*(1+p2);
f11(0,0):=(n-y)^2*(1+3*p1); f22(0,0):=(n-y)^2*(1+3*p2);
f12(0,0):=f1(0,0)*f2(0,0); f21(0,0):=f12(0,0);
f111(0,0):=expand(f11(0,0)*f1(0,0)); f111(0,0):=subs(p1^2=p1,f111(0,0));
f112(0,0):=expand(f11(0,0)*f2(0,0)); f121(0,0):=f112(0,0);
f122(0,0):=expand(f22(0,0)*f1(0,0));
f1122(0,0):=expand(f11(0,0)*f22(0,0));
g(0,0):=1;
g1(0,0):=(n-y)*p3; g2(0,0):=(n-y)*p4;
g11(0,0):=(n-y)^2*p3; g22(0,0):=(n-y)^2*p4;
g12(0,0):=(n-y)^2*p5; g21(0,0):=(n-y)^2*p5;
g111(0,0):=(n-y)^3*p3; g112(0,0):=(n-y)^3*p5;
g121(0,0):=(n-y)^3*p5; g122(0,0):=(n-y)^3*p5;
g1122(0,0):=(n-y)^4*p5; g1121(0,0):=(n-y)^4*p5;
k(0,0):=1;
k1(0,0):=-y*p3; k2(0,0):=-y*p4;
k11(0,0):=y^2*p3; k22(0,0):=y^2*p4;
k12(0,0):=y^2*p5; k21(0,0):=y^2*p5;
k111(0,0):=-y^3*p3; k112(0,0):=-y^3*p5;
k121(0,0):=-y^3*p5; k122(0,0):=-y^3*p5;
k1122(0,0):=y^4*p5; k1121(0,0):=y^4*p5;
printlevel:=-1; M1122:=collect(M1122,n);
printlevel:=1; Mf4:=coeff(M1122,n,4);
printlevel:=-1; s:='s';t:='t';
M11:=simplify(diff(M,s,s));
M11:=subs(diff(f(s,t),s,s,t,t)=f1122(s,t), diff(g(s,t),s,s,t,t)=g1122(s,t),
diff(f(s,t),s,s,t)=f112(s,t),diff(g(s,t),s,s,t)=g112(s,t),
```

```
diff(f(s,t),s,t,t)=f122(s,t),diff(g(s,t),s,t,t)=g122(s,t),
diff(f(s,t),s,s)=f11(s,t),diff(g(s,t),s,s)=g11(s,t),
diff(k(s,t),s,s,t,t)=k1122(s,t), diff(k(s,t),s,s,t)=k112(s,t),
diff(k(s,t),s,t,t)=k122(s,t), diff(k(s,t),s,s)=k11(s,t), diff(k(s,t),t,t)=k22(s,t),
diff(k(s,t),t)=k2(s,t), diff(k(s,t),s)=k1(s,t),
diff(f(s,t),s,t)=f12(s,t),diff(g(s,t),s,t)=g12(s,t),
diff(f(s,t),t,s)=f21(s,t),diff(g(s,t),t,s)=g21(s,t), diff(f(s,t),t,t)=f22(s,t),
diff(g(s,t),t,t)=g22(s,t), diff(f(s,t),t)=f2(s,t),diff(g(s,t),t)=g2(s,t),
diff(f(s,t),s)=f1(s,t),diff(g(s,t),s)=g1(s,t),M11);
t:=0;s:=0;printlevel:=-1;
M11:=collect(expand(M11),n); printlevel:=1;
M11:=subs(p3=h1,p1=h1,h1^2=F2,h1=F1,M11);
printlevel:=-1;
M11:=expand(M11^2);
NF4:=Mf4-coeff(M11,n,4);
NF4=normal(NF4);
C4:=collect(normal(NF4),y);
C4:=subs(p3^2*p4^2=F2^2,p3^2*p4=F1*F2,p4*p3^3=F1*F2,p3*p4^2=F1*F2,
p4^2*p3=F1*F2,p3*p4=F1^2,p4*p3=F1^2,C4);
C4:=subs(p1*p2=0,p1*p2*p3=0,p1*p2*p4=0,p1*p2*p5=0,p1*p2*p3^2=0,
p1*p2*p4^2=0,
p1*p2*p3*p4=0,p1*p2*p4*p3=0,C4);
printlevel:=1;
C4:=collect(C4,y);

printlevel:=-1;
M1122:='M1122';M11:='M11'; M:='M';
f1:='f1';f2:='f2';f11:='f11';f12:='f12';f111:='f111';
g1:='g1';g2:='g2';g11:='g11';g12:='g12';g111:='g111';
k1:='k1';k2:='k2';k11:='k11';k12:='k12';k111:='k111';
s:='s';t:='t'; a:='a';b:='b';
M:=f(s)*g(s)^a*k(s)^b;
M1111:=simplify(diff(M,s,s,s,s));
M1111:=subs(diff(f(s),s,s,s,s)=f1111(s), diff(g(s),s,s,s,s)=g1111(s),
diff(k(s),s,s,s,s)=k1111(s), diff(k(s),s,s,s)=k111(s),diff(k(s),s,s)=k11(s),
diff(k(s),s,s)=k11(s),diff(k(s),s)=k1(s), diff(f(s),s,s,s)=f111(s),diff(g(s),s,s,s)=g111(s),
diff(f(s),s,s)=f11(s),diff(g(s),s,s)=g11(s), diff(f(s),s)=f1(s),diff(g(s),s)=g1(s),M1111);
printlevel:=-1; t:=0; s:=0;a:=y-1;b:=n-y;
f(0):=1;
f1(0):=(n-y); f11(0):=(n-y)^2;
f111(0):=expand(f11(0)*f1(0)); f1111(0):=expand(f11(0)*f11(0));
g(0):=1;
g1(0):=(n-y)*p3; g11(0):=(n-y)^2*p3; g111(0):=(n-y)^3*p3;
g1111(0):=(n-y)^4*p3;
k(0):=1;
k1(0):=-y*p3; k11(0):=y^2*p3; k111(0):=-y^3*p3; k1111(0):=y^4*p3;
M1111:=collect(M1111,n);
Mf4:=coeff(M1111,n,4);
printlevel:=-1; s:='s';t:='t';
M11:=simplify(diff(M,s,s));
M11:=subs(diff(f(s),s,s)=f11(s),diff(g(s),s,s)=g11(s),
diff(k(s),s,s)=k11(s),diff(k(s),s)=k1(s),
diff(f(s),s,s)=f11(s),diff(g(s),s,s)=g11(s),
```

```
diff(f(s),s)=f1(s),diff(g(s),s)=g1(s),M11);
t:=0;s:=0;
M11:=collect(expand(M11),n);
M11:=subs(p3=h1,p1=h1,h1^2=F2,h1=F1,M11);
M11:=expand(M11^2);
NF4:=Mf4-coeff(M11,n,4);
C4a:=collect(NF4,y);
printlevel:=1;
C4a:=subs(p3^4=F4,P3^3=F3,P3^2=F2,p3=F1,C4a);
C4:=collect(normal(y*(y-1)*C4+y*C4a),y);
C44:=coeff(C4,y,4);
C43:=coeff(C4,y,3);
C42:=coeff(C4,y,2);
C41:=coeff(C4,y,1);
s:='s';t:='t';
M1122:='M1122';M11:='M11'; M:='M';
f1:='f1';f2:='f2';f11:='f11';f12:='f12';f111:='f111';
g1:='g1';g2:='g2';g11:='g11';g12:='g12';g111:='g111';
k1:='k1';k2:='k2';k11:='k11';k12:='k12';k111:='k111';
s:='s';t:='t'; a:='a';b:='b'; c:='c';
M:=u(s,t)*v(s,t);
printlevel:=-1;
M1122:=simplify(diff(M,s,s,t,t));
MM:=subs(diff(u(s,t),s,s,t,t)=U1122,
diff(v(s,t),s,s,t,t)=V1122,diff(v(s,t),s,s,t)=V112, diff(v(s,t),s,t,t)=V122,
diff(v(s,t),s,t)=V12, diff(v(s,t),s,s)=V11, diff(v(s,t),t,t)=V22,
diff(v(s,t),t)=V2, diff(v(s,t),s)=V1, diff(u(s,t),s,s,t)=U112,
diff(u(s,t),s,t,t)=U122, diff(u(s,t),s,t)=U12, diff(u(s,t),s,s)=U11, diff(u(s,t),t,t)=U22,
diff(u(s,t),t)=U2, diff(u(s,t),s)=U1, M1122);
printlevel:=-1;
u0(s,t):=f(s,t)*g(s,t)^a;  v0(s,t):=k(s,t)^b*k(s,t)^c;
u1(s,t):=diff(u0(s,t),s);  u2(s,t):=diff(u0(s,t),t);
u11(s,t):=diff(u1(s,t),s); u12(s,t):=diff(u1(s,t),t);  u22(s,t):=diff(u2(s,t),t);
u122(s,t):=diff(u12(s,t),t); u112(s,t):=diff(u11(s,t),t); u1122(s,t):=diff(u112(s,t),t);
v1(s,t):=diff(k(s,t)^b*h(s,t)^c,s);  v2(s,t):=diff(k(s,t)^b*h(s,t)^c,t);
v11(s,t):=diff(v1(s,t),s); v12(s,t):=diff(v1(s,t),t); v22(s,t):=diff(v2(s,t),t);
v122(s,t):=diff(v12(s,t),t); v112(s,t):=diff(v11(s,t),t); v1122(s,t):=diff(v112(s,t),t);
printlevel:=-1;
U1122:=subs(diff(f(s,t),s,s,t,t)=f1122(s,t), diff(f(s,t),s,s,t)=f112(s,t),
diff(f(s,t),s,t,t)=f122(s,t), diff(f(s,t),s,s)=f11(s,t),diff(f(s,t),t,t)=f22(s,t),
diff(f(s,t),s,t)=f12(s,t),diff(f(s,t),t)=f2(s,t), diff(f(s,t),s)=f1(s,t),
diff(g(s,t),s,s,t,t)=g1122(s,t), diff(g(s,t),s,s,t)=g112(s,t),diff(g(s,t),s,t,t)=g122(s,t),
diff(g(s,t),s,s)=g11(s,t),diff(g(s,t),t,t)=g22(s,t), diff(g(s,t),s,t)=g12(s,t),
diff(g(s,t),t)=g2(s,t), diff(g(s,t),s)=g1(s,t), u1122(s,t));
U112:=subs(diff(f(s,t),s,s,t)=f112(s,t), diff(f(s,t),s,s)=f11(s,t),diff(f(s,t),t,t)=f22(s,t),
diff(f(s,t),s,t)=f12(s,t),diff(f(s,t),t)=f2(s,t), diff(f(s,t),s)=f1(s,t),
diff(g(s,t),s,s,t)=g112(s,t), diff(g(s,t),s,s)=g11(s,t),diff(g(s,t),t,t)=g22(s,t),
diff(g(s,t),s,t)=g12(s,t),diff(g(s,t),t)=g2(s,t), diff(g(s,t),s)=g1(s,t), u112(s,t));
U122:=subs(diff(f(s,t),s,t,t)=f122(s,t), diff(f(s,t),s,s)=f11(s,t),diff(f(s,t),t,t)=f22(s,t),
diff(f(s,t),s,t)=f12(s,t),diff(f(s,t),t)=f2(s,t), diff(f(s,t),s)=f1(s,t),
diff(g(s,t),s,t,t)=g122(s,t), diff(g(s,t),s,s)=g11(s,t),diff(g(s,t),t,t)=g22(s,t),
diff(g(s,t),s,t)=g12(s,t),diff(g(s,t),t)=g2(s,t), diff(g(s,t),s)=g1(s,t), u122(s,t));
U12:=subs(diff(f(s,t),s,s)=f11(s,t),diff(f(s,t),t,t)=f22(s,t),diff(f(s,t),s,t)=f12(s,t),
```

```
diff(f(s,t),t)=f2(s,t), diff(f(s,t),s)=f1(s,t),
diff(g(s,t),s,s)=g11(s,t),diff(g(s,t),t,t)=g22(s,t),
diff(g(s,t),s,t)=g12(s,t),diff(g(s,t),t)=g2(s,t), diff(g(s,t),s)=g1(s,t), u12(s,t));
U11:=subs(diff(f(s,t),s,s)=f11(s,t),diff(f(s,t),t,t)=f22(s,t),
diff(f(s,t),s,t)=f12(s,t),diff(f(s,t),t)=f2(s,t), diff(f(s,t),s)=f1(s,t),
diff(g(s,t),s,s)=g11(s,t),diff(g(s,t),t,t)=g22(s,t),diff(g(s,t),s,t)=g12(s,t),
diff(g(s,t),t)=g2(s,t),diff(g(s,t),s)=g1(s,t), u11(s,t));
U22:=subs(diff(f(s,t),s,s)=f11(s,t),diff(f(s,t),t,t)=f22(s,t),
diff(f(s,t),s,t)=f12(s,t),diff(f(s,t),t)=f2(s,t),diff(f(s,t),s)=f1(s,t),
diff(g(s,t),s,s)=g11(s,t),diff(g(s,t),t,t)=g22(s,t),diff(g(s,t),s,t)=g12(s,t),diff(g(s,t),t)=g2(
s,t),
diff(g(s,t),s)=g1(s,t), u22(s,t));
U2:=subs(diff(f(s,t),t)=f2(s,t),diff(g(s,t),t)=g2(s,t),diff(g(s,t),s)=g1(s,t), u2(s,t));
U1:=subs(diff(f(s,t),s)=f1(s,t),diff(g(s,t),t)=g2(s,t),diff(g(s,t),s)=g1(s,t), u1(s,t));

V1122:=subs(diff(k(s,t),s,s,t,t)=k1122(s,t),diff(k(s,t),s,s,t)=k112(s,t),
diff(k(s,t),s,t,t)=k122(s,t),diff(k(s,t),s,s)=k11(s,t),diff(k(s,t),t,t)=k22(s,t),
diff(k(s,t),s,t)=k12(s,t),diff(k(s,t),t)=k2(s,t),diff(k(s,t),s)=k1(s,t),
diff(h(s,t),s,s,t,t)=h1122(s,t),diff(h(s,t),s,s,t)=h112(s,t),diff(h(s,t),s,t,t)=h122(s,t),
diff(h(s,t),s,s)=h11(s,t),diff(h(s,t),t,t)=h22(s,t),diff(h(s,t),s,t)=h12(s,t),diff(h(s,t),t)=h2(
s,t),
diff(h(s,t),s)=h1(s,t), v1122(s,t));
V112:=subs(diff(k(s,t),s,s,t)=k112(s,t),diff(k(s,t),s,s)=k11(s,t),diff(k(s,t),t,t)=k22(s,t),
diff(k(s,t),s,t)=k12(s,t),diff(k(s,t),t)=k2(s,t),diff(k(s,t),s)=k1(s,t),
diff(h(s,t),s,s,t)=h112(s,t),diff(h(s,t),s,s)=h11(s,t),diff(h(s,t),t,t)=h22(s,t),
diff(h(s,t),s,t)=h12(s,t),diff(h(s,t),t)=h2(s,t),diff(h(s,t),s)=h1(s,t), v112(s,t));
V122:=subs(diff(k(s,t),s,t,t)=k122(s,t),diff(k(s,t),s,s)=k11(s,t),diff(k(s,t),t,t)=k22(s,t),
diff(k(s,t),s,t)=k12(s,t),diff(k(s,t),t)=k2(s,t),diff(k(s,t),s)=k1(s,t),
diff(h(s,t),s,t,t)=h122(s,t),diff(h(s,t),s,s)=h11(s,t),diff(h(s,t),t,t)=h22(s,t),
diff(h(s,t),s,t)=h12(s,t),diff(h(s,t),t)=h2(s,t),diff(h(s,t),s)=h1(s,t), v122(s,t));
V12:=subs(diff(k(s,t),s,s)=k11(s,t),diff(k(s,t),t,t)=k22(s,t),
diff(k(s,t),s,t)=k12(s,t),diff(k(s,t),t)=k2(s,t),diff(k(s,t),s)=k1(s,t),
diff(h(s,t),s,s)=h11(s,t),diff(h(s,t),t,t)=h22(s,t),diff(h(s,t),s,t)=h12(s,t),diff(h(s,t),t)=h2(
s,t),
diff(h(s,t),s)=h1(s,t), v12(s,t));
V11:=subs(diff(k(s,t),s,s)=k11(s,t),diff(k(s,t),t,t)=k22(s,t),diff(k(s,t),s,t)=k12(s,t),
diff(k(s,t),t)=k2(s,t),diff(k(s,t),s)=k1(s,t),diff(h(s,t),s,s)=h11(s,t),diff(h(s,t),t,t)=h22(s,t)
,
diff(h(s,t),s,t)=h12(s,t),diff(h(s,t),t)=h2(s,t),diff(h(s,t),s)=h1(s,t), v11(s,t));
V22:=subs(diff(k(s,t),s,s)=k11(s,t),diff(k(s,t),t,t)=k22(s,t),
diff(k(s,t),s,t)=k12(s,t),diff(k(s,t),t)=k2(s,t),diff(k(s,t),s)=k1(s,t),
diff(h(s,t),s,s)=h11(s,t),diff(h(s,t),t,t)=h22(s,t),diff(h(s,t),s,t)=h12(s,t),diff(h(s,t),t)=h2(
s,t),
diff(h(s,t),s)=h1(s,t), v22(s,t));
V2:=subs(diff(k(s,t),t)=k2(s,t),diff(h(s,t),t)=h2(s,t),diff(h(s,t),s)=h1(s,t), v2(s,t));
V1:=subs(diff(k(s,t),s)=k1(s,t),diff(h(s,t),t)=h2(s,t),diff(h(s,t),s)=h1(s,t), v1(s,t));

printlevel:=-1;
t:=0; s:=0;
f(0,0):=1; f1(0,0):=(n-x)-x*p1; f2(0,0):=(n-y)-y*p2;
f11(0,0):=expand(f1(0,0)^2); f11(0,0):=subs(p1^2=p1,f11(0,0));
f22(0,0):=expand(f2(0,0)^2); f22(0,0):=subs(p2^2=p2,f22(0,0));
f12(0,0):=f1(0,0)*f2(0,0); f21(0,0):=f12(0,0);
```

```
f112(0,0):=expand(f11(0,0)*f2(0,0));
f122(0,0):=expand(f22(0,0)*f1(0,0));
f1122(0,0):=expand(f11(0,0)*f22(0,0));
g(0,0):=1; g1(0,0):=(n-x)*p3; g2(0,0):=-y*p4;
g11(0,0):=(n-x)^2*p3; g22(0,0):=y^2*p4;
g12(0,0):=-y*(n-x)*p5; g21(0,0):=-y*(n-x)*p5;
g112(0,0):=-y*(n-x)^2*p5; g121(0,0):=-y*(n-x)^2*p5;
g122(0,0):=y^2*(n-x)*p5; g1122(0,0):=y^2*(n-x)^2*p5;
k(0,0):=1;k2(0,0):=(n-y)*p4; k1(0,0):=-x*p3;
k22(0,0):=(n-y)^2*p4; k11(0,0):=x^2*p3; k12(0,0):=-x*(n-y)*p5;
k21(0,0):=-x*(n-y)*p5; k122(0,0):=-x*(n-y)^2*p5; k121(0,0):=x^2*(n-y)*p5;
k112(0,0):=x^2*(n-y)*p5; k1122(0,0):=x^2*(n-y)^2*p5;
h(0,0):=1; h1(0,0):=-x*p3; h2(0,0):=-y*p4; h11(0,0):=x^2*p3; h22(0,0):=y^2*p4;
h12(0,0):=x*y*p5; h21(0,0):=x*y*p5; h112(0,0):=-y*x^2*p5; h121(0,0):=-y*x^2*p5;
h122(0,0):=-y^2*x*p5; h1122(0,0):=x^2*y^2*p5;
printlevel:=-1;
a:=x-1;b:=y-1;c:=n-x-y; u(0,0):=1;v(0,0):=1;
M1122:='M1122';
u1122(s,t):='u1122(s,t)'; v1122(s,t):='v1122(s,t)';u122(s,t):='u122(s,t)';
v122(s,t):='v122(s,t)';u112(s,t):='u112(s,t)'; v112(s,t):='v112(s,t)';
u11(s,t):='u11(s,t)'; v11(s,t):='v11(s,t)'; u12(s,t):='u12(s,t)'; v12(s,t):='v12(s,t)';
u22(s,t):='u22(s,t)'; v22(s,t):='v22(s,t)';
MM1:=collect(U1122*v(s,t),n); MM1:=coeff(MM1,n,4);
MM2:=collect(V1122*u(s,t),n); MM2:=coeff(MM2,n,4);
MM3:=collect(2*U112*V2,n); MM3:=coeff(MM3,n,4);
MM4:=collect(2*V112*U2,n); MM4:=coeff(MM4,n,4);
MM5:=collect(U11*V22,n); MM5:=coeff(MM5,n,4);
MM6:=collect(V11*U22,n); MM6:=coeff(MM6,n,4);
MM7:=collect(2*U122*V1,n); MM7:=coeff(MM7,n,4);
MM8:=collect(2*V122*U1,n); MM8:=coeff(MM8,n,4);
MM9:=collect(4*U12*V12,n); MM9:=coeff(MM9,n,4);
Mf4:=MM1+MM2+MM3+MM4+MM5+MM6+MM7+MM8+MM9;
printlevel:=-1;
Mf4:=normal(Mf4);Mf4:=collect(Mf4,x);
printlevel:=-1;
MK1:=collect(U1122*v(s,t),n); MK1:=coeff(MK1,n,3);
MK2:=collect(V1122*u(s,t),n); MK2:=coeff(MK2,n,3);
MK3:=collect(2*U112*V2,n); MK3:=coeff(MK3,n,3);
MK4:=collect(2*V112*U2,n); MK4:=coeff(MK4,n,3);
MK5:=collect(U11*V22,n); MK5:=coeff(MK5,n,3);
MK6:=collect(V11*U22,n); MK6:=coeff(MK6,n,3);
MK7:=collect(2*U122*V1,n); MK7:=coeff(MK7,n,3);
MK8:=collect(2*V122*U1,n); MK8:=coeff(MK8,n,3);
MK9:=collect(4*U12*V12,n); MK9:=coeff(MK9,n,3);
Mf3:=MK1+MK2+MK3+MK4+MK5+MK6+MK7+MK8+MK9;
printlevel:=-1;
Mf3:=normal(Mf3);Mf3:=collect(Mf3,x);
s:='s';t:='t';
M11:=simplify(diff(M,s,s));printlevel:=-1;
M11:=subs(diff(u(s,t),s,s,t,t)=U1122, diff(v(s,t),s,s,t,t)=V1122,diff(v(s,t),s,s,t)=V112,
diff(v(s,t),s,t,t)=V122, diff(v(s,t),s,t)=V12, diff(v(s,t),s,s)=V11, diff(v(s,t),t,t)=V22,
diff(v(s,t),t)=V2, diff(v(s,t),s)=V1, diff(u(s,t),s,s,t)=U112,
diff(u(s,t),s,t,t)=U122, diff(u(s,t),s,t)=U12, diff(u(s,t),s,s)=U11, diff(u(s,t),t,t)=U22,
```

```
diff(u(s,t),t)=U2, diff(u(s,t),s)=U1, M11);
t:=0;s:=0;printlevel:=-1;
M11:=collect(normal(M11),n);
M11:=subs(p1=F1,p3=F1,F1^2=F2,M11);
s:='s';t:='t';
M22:=simplify(diff(M,t,t));printlevel:=-1;
M22:=subs(diff(u(s,t),s,s,t,t)=U1122, diff(v(s,t),s,s,t,t)=V1122,diff(v(s,t),s,s,t)=V112,
diff(v(s,t),s,t,t)=V122, diff(v(s,t),s,t)=V12, diff(v(s,t),s,s)=V11, diff(v(s,t),t,t)=V22,
diff(v(s,t),t)=V2, diff(v(s,t),s)=V1, diff(u(s,t),s,s,t)=U112,
diff(u(s,t),s,t,t)=U122, diff(u(s,t),s,t)=U12, diff(u(s,t),s,s)=U11, diff(u(s,t),t,t)=U22,
diff(u(s,t),t)=U2, diff(u(s,t),s)=U1, M22);
t:=0;s:=0;printlevel:=-1;
M22:=collect(normal(M22),n);
M22:=subs(p2=F1,p4=F1,F1^2=F2,M22);
MN:=collect(M11*M22,n);
MN3:=normal(coeff(MN,n,3));
MN4:=normal(coeff(MN,n,4));
printlevel:=1;
C4b:=collect(collect(normal(Mf4-MN4),x),y);
C4b:=subs(p3^2*p4^2=F2^2,p4^2*p3^2=F2^2,p3*p4^2=F1*F2,p4^2*p3=F1*F2,
p3^2*p4=F1*F2,p4*p3^2=F1*F2,p3*p4=F1^2,p4*p3=F1^2,p4^2=F2,p3^2=F2,
p3^3=F3,p4^3=F3,p3=F1,p4=F1,C4b);
C3:=collect(collect(normal(Mf3-MN3),x),y);
C3:=subs(p3^2*p4^2=F2^2,p4^2*p3^2=F2^2,
p3*p4^2=F1*F2,p4^2*p3=F1*F2,
p3^2*p4=F1*F2,p4*p3^2=F1*F2,p3*p4=F1^2,p4*p3=F1^2,C3);
C32:=coeff(coeff(C3,y,1),x,1);
C31:=coeff(coeff(C3,y,0),x,1);
quit;
```

Bibliography

[1] A.A.Afifi and S.P. Azen. (1979). *Statistical Analysis, A computer Oriented Approach.* Academic Press, New York.

[2] J.N. Adichie. (1967). Estimation of regression coefficients based on rank tests. *Annals of Mathematical Statistics* **38** 894–904.

[3] G. Bassett and R. Koenker. (1982). An empirical quantile function for linear models with iid errors. *J.A.S.A.* **77** 407–415.

[4] P.K Battacharya and D. Frierson. (1981). A nonparametric chart for detecting small disorders. *Annals of Statistics* **9** 544–554.

[5] P.K Battacharya and R.A Johnson. (1968). A nonparametric control for shift at an unknown time point. *Annals of Mathematical Statistics* **39** 1731–1743.

[6] P.J. Bickel. (1973). On some analogues to linear combination of order statistics in the linear model. *Annals of Statistics* **1** 597–616.

[7] T. Bojdecki. (1979). Probability maximizing approach to optimal stopping and its application to a disorder problem. *Stochastics* **3** 61–71.

[8] T. Bojdecki and J. Hosza. (1984). On a generalized disorder problem. *Stochastics* **18** 349–359.

[9] G.E.P. Box. (1953). Non-normality and tests on variances. *Biometrika* **40** 318–335.

[10] H. Chernoff and S. Zacks. (1966). Estimating the current mean of a normal distribution which is subject to changes in time. *Annals of Mathematical Statistics* **37** 1196–1210.

[11] Aline Chouinard and David McDonald. (1985). A characterization of non-homogeneous Poisson processes. *Stochastics* **15** 113–119.

[12] Ronald B. Crosier. (1988). Multivariate generalizations of cumulative sum quality-control schemes. *Technometrics* **30** 291–303.

[13] B.S. Darkhovskii. (1976). A nonparametric method for the posteriori detection of the disorder time of a sequence of independent random variables. *Theory of probability and its application* **21** 178–183.

[14] B.S. Darkhovskii and B.E.Brodskii. (1987). A nonparametric method for fastest detection of a change in the mean of a random sequence. *Theory of probability and applications* **32** 640–648.

[15] B.S. Darkhovskii and B.E. Brodskii. (1980). A posteriori detection of the disorder time of a random sequence. *Theory of probability and its application* **25** 624–628.

[16] L.A Gardner. (1969). On detecting changes in the mean of normal variates. *Annals of Mathematical Statistics* **40** 116–126.

[17] C. Genest and R.J. MacKay (1986). The joy of copulas: Bivariate distributions with uniform marginals. *The American Statistician* 40 280–283.

[18] K. Ghoudi. (1990). *Multivariate nonparametric quality control statistics.* Master's thesis, University of Ottawa, Ottawa, Ontario Canada.

[19] L. Gordon and M. Pollak. (1992). An efficient sequential nonparametric scheme for detecting a change of distribution. To be published.

[20] L. Gordon and M. Pollak. (1991). A robust surveillance scheme for stochastically ordered alternatives. To be published.

[21] C. Gutenbrunner and J. Jurečková. (1992). Regression rank scores and regression quantiles. *The Annals of statistics* **20** 305–330.

[22] Peter Hackl and Johannes Ledolter. March (1989). *A new nonparemetric quality control-technique.* Technical Report 160, Department of Statistics and Actuarial Science University of Iowa.

[23] T. P. Hettmansperger and J.W. McKean. (1977). A robust alternative based on ranks to the least squares in analyzing linear models. *Technometrics* **19** 275–284.

[24] Hinkley. (1970). Inference about the change point in a sequence of random variables. *Biometrika* **57** 1–17.

[25] W. Hoeffding. (1952). The large-sample power of tests based on permutations of observations. *The Annals of mathematical statistics* **23** 169–192.

[26] Harold Hotelling. (1950). A generalized T-test and measure of multivariate dispersion. *Proceeding of the second Berkeley synposium in mathematical statistics and probability* 23–41.

[27] P. Huber. (1972). Robust statistics: a review. *The Annals Of Mathematical Statistics* **43** 1041–1067.

[28] P.J. Huber. (1973). Robust regression: asymptotics, conjectures and monte carlo. *Annals of Statistics.* **1** 799–821.

[29] P.J. Huber. (1981). *Robust statistics.* John Wiley, New York.

[30] J.Edward Jackson. (1959). Quality control methods for several related variables. *Technometrics* **1** 359–377.

[31] J.Edward Jackson and R.H. Morris. (1957). An application of multivariate quality control to photographic processing. *Journal of the American Statistical Association* **52** 186–199.

[32] J.Edward Jackson and Govind S. Mudholkar. (1979). Control procedure for residuals associated with principal component analysis. *Technometrics* **21** 341–349.

[33] L.A. Jaeckel. (1972). Estimating regression coefficients by minimizing the dispersion of residuals. *The Annals of Mathematical Statistics* **43** 1449–1458).

[34] V.K. Jandhyala. October (1985). *Residual Processes For Regression Models with applications to detection of parameter change at unknown times.* PhD thesis, University of Western Ontario., London, Ontario Canada.

[35] J. Jurečková. (1971). Nonparametric estimate of regression coefficients. *The Annals of mathematical statistics* **42** 1328–1338.

[36] Z. Kander and S. Zacks. (1964). Test procedures for possible changes in parameters of statistical distributions occuring at unknown time points. *Annals of Mathematical Statistics* **35** 999–1018.

[37] J. Kiefer. (1959). K-sample analogues of the kolmogorov-smirnov and cramer-von mises tests. *Annals of Mathematical Statistics* **30** 420–447.

[38] J. Kiefer. (1960). On large deviations of the empiric df. of vector change variables and law of the iterated logarithm. *Annals of Mathematical Statistics* **31** 649–660.

[39] R. Koenker and G. Bassett. (1978). Regression quantile. *Econometrica.* **46** 33–50.

[40] R. Koenker and G. Bassett. (1982). Robust test for heteroscedasticity based on regression quantile. *Econometrica.* **50** 43–61.

[41] E.L Lehmann. (1951). Consistency an unbiasedness of certain nonparametric tests. *Annals of Mathematical Statistics* **22** 165–179.

[42] G. Lorden. (1971). Procedures for reacting to a change in distribution. *Annals of Mathematical Statistics* **42** 1897–1908.

[43] D. McDonald. (1990). A cusum procedure based on sequential ranks. *Naval Research Logistics* **37** 627–646.

[44] D. McDonald. (1991). On the asymptotics of randomness statistics. *Canadian Journal of Statistics* **19** 209–217.

[45] George Moustakides. (1986). On optimal stopping times for changes in distributions. *The Annals of Statistics* **14** 1379–1387.

[46] E.S. Page. (1954). Continuous inspection schemes. *Biometrika* **41** 100–115.

[47] E.S. Page. (1955). A test for change in a parameter occuring at unknown time point. *Biometrika* **42** 523–526.

[48] E.S. Pearson. (1931). The analysis of variance in cases of non normal variation. *Biometrika* **23** 114–133.

[49] Moshe Pollak. (1985). Optimal detection of a change in distribution. *The Annals of Statistics* **13** 206–227).

[50] P.J. Rousseeuw and A.M. Leroy. (1987). *Robust regression and outlier detection.* John Wiley, New York.

[51] A. Sen and M. Srivastava. (1975). On tests for detecting change in mean. *The Annals of Statistics* **3** 98–108.

[52] A.N. Shiryayev. (1963). On optimum methods in quickest detection problems. *Theory of probability and its applications* **8** 22–46.

[53] A.N. Shiryayev. (1978). *Optimal stopping rules.* Springer Verlag, New York, Heidelberg, Berlin.

[54] A.F Siegel. (1982). Robust regression using repeated medians. *Biometrika* **69** 242–244.

[55] William H. Woodall and Matoteng M. Ncube. (1985). Multivariate cusum quality control procedures. *Technometrics* **27** 285–292.

[56] S. Zacks and Z. Barzily. (1981). Bayes procedures for detecting a shift in the probability of success in a series of Bernoulli trials. *Journal of statistical planning and inference* **5** 107–119.