Report 99-003

Two consistent nonparametric tests of the monotonicity of regression

Cheikh A.T. Diack

ISSN: 1389-2355

Two consistent nonparametric tests of the monotonicity of regression

Cheikh A. T. DIACK
EURANDOM, P.O. Box 513, 5600 Eindhoven,
The Netherlands. E-mail:diack@eurandom.tue.nl

January 22, 1999

Abstract

In this paper, we develop two consistent nonparametric tests of the monotonicity of a regression function. One is related to the so-called one-sided testing problem in multivariate analysis. It is analogous to Diack's test of convexity of a regression function. The other test is an adaptation of Schlee's idea in a nonparametric model with fixed design. The tests are consistent against all alternatives and have power against local misspecifications.

1 INTRODUCTION

Shape restrictions on a functional parameter such monotonicity arise in a variety of statistical models. Indeed, monotone functions have played an important role in data analysis.

Whenever monotonicity is avalid assumption, it is advantageous to impose monotonicity on a nonparametric local averaging estimators. However, when the true function is non-montone, this restriction may lead to erroneous inference. Thus there may be several competing models. Therefore, test of monotonicity provide a way in order to prevent from wrong conclusions.

In our best knowledge, only Schlee (1980) is concerned by the problem to test the monotonicity of a regression function.

Schlee's test uses, in a nonparametric model with random design a modified version of the kernel estimator. His test is based on the greatest discrepancy of an estimator of the first derivative of the regression function from zero. However, Schlee does not discuss asymptotic results of the test.

In this paper, we propose two consistent nonparametric tests of the monotonicity of a regression function. One is related to the so-called one-sided testing problem in multivariate analysis. It is analogous to Diack's test of convexity of a regression function. For this case, we will estimate the model by a quadratic spline estimator which allows us to formulate the monotonicity hypothesis in a very simple way. The other test is an adaptation of Schlee's idea in a nonparametric model with fixed design. As Schlee, we will use a kernel estimator (but different) of the first derivative of the regression function.

2 TEST VIA QUADRATIC SPLINE

The methodology is similar to that used by Diack(1999) to construct a convexity test of a regression function. The basic idea uses a connection between (1) a hypothesis test of monotonicity of a nonparametric regression function based on a quadratic splines estimator and (2) a hypothesis test for normal means constrained by linear inequalities.

We will consider the following nonparametric regression model:

$$y_{ij} = f(x_i) + \varepsilon_{ij}, i = 1, ..., r, j = 1, ..., n_i \quad x_i \in (0,1), i = 1, ..., r.$$

At each deterministic design point $x_i, (i = 1, ..., r), n_i$ measurements are taken. The probability measure assigning mass $\mu_i = n_i/n$ to the point $x_i (\sum \mu_i = 1)$ is referred to as the design and will be denoted by μ^n. We assume that the random errors ε_{ij} are uncorrelated and identically distributed with mean zero. Their variance σ^2 will be assumed unknown. Finally f is an unknown smooth regression function.
In what follows, we will assume some regularity conditions on f.

Consider the following class of functions:
For $l \in I\!N$ and $M > 0$, let

$$\mathcal{F}_{l,M} = \{f \in \mathcal{C}^{l+1}(0,1) : \sup_{0 \leq x \leq 1} | f^{((l+1)\wedge 3)}(x) | \leq M\}.$$

We intend to contruct a test of H_0 : "f is monotone" versus H_1 : "f is non-monotone."

It would be interesting to know how the test behaves under the local alternatives. So, we might consider a sequence of local alternatives H_{1n} : "$f_n = f_0 + h_n L$" where f_0 is a fixed function in the null hypothesis and L is known and lies in $\mathcal{F}_{l,M}$. This formulation includs the fixed alternative which corresponds to $h_n = 1$.

Throughout this paper, a testing problem with null hypothesis H_0 and alternative H_1 is denoted by $[H_0, H_1]$.

We will use a quadratic spline estimator and characterize monotonicity in a very simple way.

2.1 The Quadratic Spline Estimator

Let p be a positive continuous density on (0,1). We assume that

$$\min_{0 \leq x \leq 1} p(x) > 0.$$

Let $\eta_0 = 0 < \eta_1 < ... < \eta_{k+1} = 1$ be a subdivision of the interval $(0,1)$ by k distinct points defined by

$$\int_0^{\eta_i} p(x)\, dx = i/(k+1), \quad i = 0, ..., k+1. \tag{1}$$

Let $\delta_k = \max_{0 \leq i \leq k} (\eta_{i+1} - \eta_i)$.

For each fixed set of knots of the form (1), we define $\mathcal{S}(k,d)$ as the collection of all polynomial splines of order d (degree$\leq d-1$) having for knots $\eta_1 < ... < \eta_k$. The class $\mathcal{S}(k,d)$ of such splines is a linear space of functions of dimension $(k+d)$. A basis for this linear space is provided by the B-splines(see Schumaker 1981). Let $\{N_1, .., N_{k+d}\}$ denote the set of normalized B-splines associated with the following nondecreasing sequence $\{t_1, .., t_{k+2d}\}$:

$$\begin{cases} t_1 \leq t_2 \leq ... \leq t_d = 0 \\ t_{2d+k} \geq t_{2d+k-1} \geq ... \geq t_{d+k+1} = 1 \\ t_{d+l} = \eta_l \quad \text{for} \quad l = 1, .., k \end{cases}$$

The reader is referred to Schumaker (1981) for a discussion of these B-splines.

In what follows, we shall only work with the class of quadratic splines: $\mathcal{S}(k,3)$. It will be convenient to introduce the following notations:

$N(x) = (N_1(x),..,N_{k+3}(x))' \in \mathbb{R}^{k+3}$ and $F = (N(x_1),..,N(x_r))$; $(k+3) \times r$ matrix.

We will denote by $\widehat{f}_n$ the least squares spline estimator of f :

$$\widehat{f}_n(x) = \sum_{p=1}^{k+3} \widehat{\theta}_p N_p(x) \tag{2}$$

where

$$\widehat{\Theta} = (\hat{\theta}_1, ..., \hat{\theta}_{k+3})' = \arg \min_{\Theta \in \mathbb{R}^{k+3}} \sum_{i=1}^{r} \sum_{j=1}^{n_i} \left(y_{ij} - \sum_{p=1}^{k+3} \theta_p N_p(x_i) \right)^2 . \tag{3}$$

Let

$$\begin{aligned} \bar{y}_i &= \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij},\ \bar{\varepsilon}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \varepsilon_{ij}, \\ \bar{Y} &= (\bar{y}_1, .., \bar{y}_r)', \bar{\varepsilon} = (\bar{\varepsilon}_1, .., \bar{\varepsilon}_r)' \text{ and } f_\Delta = (f(x_1), .., f(x_r))' \end{aligned}$$

Let $\mathcal{D}(\mu^n)$ be the $r \times r$ diagonal matrix with diagonal elements $\mu_1, ..., \mu_r$, then, basic least squares arguments allow to prove that:

$$\widehat{\Theta} = M^{-1}(\mu^n) F \mathcal{D}(\mu^n) \bar{Y} \quad \text{with} \quad M(\mu^n) = \sum_{i=1}^{r} N(x_i) N'(x_i) \mu_i = F \mathcal{D}(\mu^n) F'.$$

Asymptotic properties of this estimator have been established in Argarwal and Studden(1980).

Note that the first moment of $\widehat{f}_n$ is given by

$$\mathbb{E}\widehat{f}_n(x) = N(x)' M^{-1}(\mu^n) F \mathcal{D}(\mu^n) f_\Delta.$$

Thus, if f is a quadratic spline function (that is to say there is Θ such that $f_\Delta = F'\Theta$) then $\widehat{f}_n$ is unbiased and $\mathbb{E}\widehat{\Theta} = \Theta$.

2.2 Monotonicity in $\mathcal{S}$(k,3)

Because, in the sequel, only non-decreasing functions will be considered, we will use the term non-decreasing function and monotone function interchangeably. Note that, if a function g is a quadratic spline, then its first derivative is a linear function between any pair of adjacent knots η_i and η_{i+1}, and it follows that g is a monotone function in the interval $\eta_i \leq x \leq \eta_{i+1}$ if and only if $g'(\eta_i)$ and $g'(\eta_{i+1})$ are both non negative (this property was used by Dierckx (1980) to define a monotone estimator). For a function g in the class $\mathcal{S}(k,3)$, we can write:

$$g(x) = \sum_{p=1}^{k+4} \theta_p N_p(x) \quad \text{with} \quad \Theta = (\theta_1, ..., \theta_{k+3})' \in \mathbb{R}^{k+3}.$$

$$\text{Then:} \qquad g'(\eta_l) = \sum_{p=1}^{k+3} \theta_p N'_p(\eta_l) = \sum_{p=1}^{k+3} \theta_p d_{p,l},$$

where the coefficients $d_{p,l}$ are easily calculated from the knots (see Dierckx 1980).

$$\begin{cases} d_{p,l} = 0 \text{ if } p \neq l+1 \text{ or } p \neq l+2 \\ d_{l+1,l} = \frac{-2}{t_{l+4}-t_{l+2}} \\ d_{l+2,l} = -d_{l+1} \end{cases} \quad \text{for} \quad l = 0, ..., k+1$$

Let $b_l = (0,0,...,0,-d_{l+1,l},-d_{l+2,l},0,0,...,0)' \in I\!R^{k+3}$and $\Theta = (\theta_1, ..., \theta_{k+3})'$, then

$$g'(\eta_l) = -b'_l\Theta.$$

Hence, we see that a quadratic spline g is a monotone function if and only if $b'_l\Theta\leq 0$ for all $l = 0, ..., k+1$. Therefore, a test for monotonicity in $\mathcal{S}(k, 3)$ can be written as $H_0^{'} : b'_l\Theta\leq 0$ for all $l = 0, ..., k+1$ versus $H_1^{'} : b'_l\Theta > 0$ for some $l \in \{0, ..., k+1\}$ and where Θ is the mean of the random vector $\widehat{\Theta}$. We will use below this property to construct our test.

2.3 Definition of the test

Let Y be a random vector distributed as $\mathcal{N}_q(\Theta, \Sigma_q)$ $(q \in I\!N, \quad q > 0)$ where Σ_q is a known nonsingular matrix.
We consider testing $H_0^{'} : b'_l\Theta\leq 0 \quad (l = 0, ..., p)$ against $H_1^{'} : b'_l\Theta > 0$ for some $l \in \{0, ..., p\}$ with $(p\leq q)$

In this paper we identify a hypothesis with the corresponding set of parameters. For example, we write $H_0^{'} = \{\Theta \in I\!R^q : b'_l\Theta\leq 0\}$.
It is easy to see that, the likelihood ratio test (LRT) rejects $H_0^{'}$ for large values of the test statistic

$$\bar{\chi}^2_{\Sigma_q}(p) = \inf_{x\in H_o^{'}} \| Y - x \|^2_{\Sigma_q^{-1}} .$$

It is worth noting that $H_o^{'}$ is a polyhedral cone and is thus closed and convex. Hence, for a given Y, this infinimum is attained at unique point denoted by $\Pi_{H_o^{'}}(Y)$ and represents the squared distance from Y to $H_o^{'}$.

We shall introduce some useful notations. Let B the q-by-p matrix defined by $B = (b_1|...|b_p)$. Let J be a subset (possibly empty) of $\{1, ..., p\}$ and let $\overline{J}$ be its complement. B_J will be the matrix consisting of those columns of B indexed by the elements of J. The matrix $B_{\overline{J}}$ is defined analogously. Finally $\#J$ will be the number of elememnts of J.

Therefore, using formula (6) in Diack (1999), we see that: the size-α likelihood ratio test with null hypothesis $H_o^{'}$ versus the alternative hypothesis $H_1^{'}$ is the test with reject the null hypothesis if

$$\bar{\chi}^2_{\Sigma_q}(p) \geq s^2_{\alpha,p}$$

where $s^2_{\alpha,p}$ is defined by

$$\sum_{q-p\leq j\leq q-1} \omega_j P\left(\chi^2_{q-j} \geq s^2_{\alpha,p}\right) = \alpha. \tag{4}$$

and where

$$\omega_j = \sum_{q-\#J=j} P\left[B'_{\overline{J}}Y - \left(B'_{\overline{J}}\Sigma_q B_J\right)(B'_J\Sigma_q B_J)^{-1}B'_J Y\leq 0\right] P\left[(B'_J\Sigma_q B_J)^{-1}B'_J Y \geq 0\right].$$

Moreover,

$$\omega_0 + \sum_{q-p \leq j \leq q-1} \omega_j = 1$$

where ω_0 is given by $\omega_0 = P\,(B'Y{\leq}0)$.

To calculate the probabilities in the right-hand side of (4), the values of ω_j are needed. However, even for moderate $q(q > 3)$, good closed form expressions for these level probabilities have not found. Thus approximations are of interest. For this, one may use Monte Carlo method, what we will do in simulation study.

The test statistic requires computing the projection $\Pi_{H'_0}(Y)$ of Y. However, a good closed-form solution has not found. Hence, this problem requires extensive numerical work to obtain solution. As Diack (1999), we propose an algorithm based on succesive projections which has been introduced by Dykstra (1983) (see also Boyle and Dykstra 1985). This algorithm determines the projection of a point X of any real Hilbert space onto the intersection $\mathcal{K}$ of convex sets $\mathcal{K}_j \quad (j = 1, ..., p)$ and it is meant for applications where projections onto the $\mathcal{K}_j$'s can be calculated relatively easily. Let $\mathcal{K}$ be a closed convex cone in $I\!R^q$. We suppose that $\mathcal{K}$ can be written as $\bigcap_{j=1}^{p} \mathcal{K}_j$ and each $\mathcal{K}_j$ is also convex cone. For all $X \in I\!R^q$, we denote by $X_{\mathcal{K}}^{\Gamma}$ the $\Gamma-$ projection onto $\mathcal{K}$, where Γ is a positive definite matrix. The algorithm consists of repeated cycles and every cycle contains p stages.
Let X_{mi}^{Γ} be the appproximation of $X_{\mathcal{K}}^{\Gamma}$ given by Dykstra's algorithm at the ith stage of mth cycle.
For all $m \in I\!N, \quad m > 0$, let use define $\bar{\chi}^2_{\Sigma_q}(p, m)$ by

$$\bar{\chi}^2_{\Sigma_q}(p,m) = \| Y - Y_{mp}^{\Sigma_q^{-1}} \|^2_{\Sigma_q^{-1}}$$

Using Boyle and Dykstra's results, Diack(1999) prove that $\bar{\chi}^2_{\Sigma_q}(p,m)$ converges almost surely to $\bar{\chi}^2_{\Sigma_q}(p)$ as m tends to infinity. Therefore, to implement the test, we will used $\bar{\chi}^2_{\Sigma_q}(p,m)$ instead of $\bar{\chi}^2_{\Sigma_q}(p)$.

Now, we are ready to define a monotonicity test.

Consider the problem $[H_o, H_1]$ where H_o means that the regression function f is monotone and H_1 is the unrestricted alternative.
Let $\widehat{\Theta}$ be the solution of the quadratic programming problem (3). The number of knots will be a function of the sample size $k = k_n$. To define $\widehat{\Theta}_{m,k_n+2}$ by

$$\widehat{\Theta}_{m,k_n+2} = \widehat{\Theta}^{\Sigma_n}_{m,k_n+2}, \quad (p = k_n + 2)$$

with $\Sigma_n^{-1} = \frac{\sigma^2}{n} M^{-1}(\mu^n)$ and where $\widehat{\Theta}^{\Sigma_n}_{m,k_n+2}$ given by the $(k_n+2)^{th}$ stage of the m^{th} cycle of Dykstra's algorithm.
Like this, we will defined our test of monotonicity by rejecting H_o when

$$\bar{\chi}^2_{\frac{\sigma^2}{n}M^{-1}(\mu^n)}(k_n+2, m) = \frac{n}{\sigma^2} \| \widehat{\Theta} - \widehat{\Theta}_{m,k_n+2} \|^2_{M(\mu^n)} \geq s^2_{\alpha,k_n+2}, \qquad (5)$$

where s^2_{α,k_n+2} is defined by (4).

In pratice, σ is unknown and we need to estimate it. This can be obtained in the case of the least squares estimator, using $\hat{\sigma}^2_n = \frac{1}{n-(k+4)} \sum_{i=1}^{r} (\bar{y}_i - \widehat{f}_n(x_i))^2$ or alternatively, any consistent estimator based on nonparametric regression techniques.
In what follows, we assume that μ^n converges to a design measure μ, where μ is an absolutely continuous measure. We denote by G_n and G the cumulative distribution

function of μ^n and μ respectively. The critical region of the test is

$$\Lambda_{n,m} = \{\bar{\chi}^2_{\frac{\hat{\sigma}^2_n}{n} M^{-1}(\mu^n)}(k_n + 2, m) = \frac{n}{\hat{\sigma}^2_n} \parallel \widehat{\Theta} - \widehat{\Theta}_{m,k_n+1} \parallel^2_{\Sigma_n} \geq s^2_{\alpha,k_n+2}\}.$$

The following theorem give a result about the size and the consistency of the test.

Theorem 1 *Let $f \in \mathcal{F}_{l,M}$ with $l \geq 2$. Let us consider the problem $[H_\circ, H_1]$. Then, under the following assumptions:*
(i) ε_{ij} $(i = 1, ..., r\ ;\ j = 1, ...n_i)$ i.i.d. with mean zero and finite variance σ^2.
(ii) $n_i \to +\infty$ and $\hat{\sigma}^2_n \to \sigma^2$ (with convergence in probability)
(iii) $\sup_{0 \leq x \leq 1} |G_n(x) - G(x)| = \circ(\delta_{k_n})$, as $k_n \longrightarrow +\infty$.
(iv) $\lim_{n \longrightarrow +\infty} rn\delta^6_{k_n} \left(\sup_{1 \leq i \leq r} \mu_i \right) = 0$
The test is asymptotically size α. More precisely,
$\limsup_{n \longrightarrow +\infty} \sup_{f \in H_\circ} \lim_{m \to +\infty} P_f(\Lambda_{n,m}) = \alpha$.
Moreover, if
(v) $\lim_{n \to +\infty} h^2_n n \delta^4_{k_n} = +\infty$
then, the test is consistent for all $f_n \in H_{1n}$ i.e.:
$\lim_{n \to +\infty} \lim_{m \to +\infty} P_{f_n}(\Lambda_{n,m}) = 1$.

The proof of **theorem 1** is in every respect similar to the proof of **theorem 5** in Diack (1999). The differences between the assumptions (iv) and (v) of this theorem and the corresponding hypotheses of **theorem 5** are due:

- For assumption (iv). Instead to use **lemma 7** and **lemma 8** in Diack (1999), we have to use the two following lemma which are respectively consequences of results in Schumaker (1981) and in Beatson (1982).

  Lemma 2 *There is a constant c such that for all $f \in \mathcal{F}_{l,M}$, there is a function S in $\mathcal{S}(k, 3)$ such that: $\sup_{0 \leq x \leq 1} |f(x) - S(x)| \leq c\delta^3_{k_n}$.*

  Lemma 3 *There is a constant c such that if $f \in \mathcal{F}_{l,M}$ is monotone, then there exists a monotone functions S in $\mathcal{S}(k, 3)$ such that: $\sup_{0 \leq x \leq 1} |f(x) - S(x)| \leq c\delta^3_{k_n}$.*

- For assumption (v) we have

$$\sup_{0 \leq j \leq k_n+1} \parallel b_j \parallel = \mathcal{O}(\inf_{0 \leq j \leq k_n+1} \parallel b_j \parallel) = \mathcal{O}\left(\delta^{-1}_{k_n}\right)$$

instead of

$$\sup_{0 \leq j \leq k_n+1} \parallel b_j \parallel = \mathcal{O}(\inf_{0 \leq j \leq k_n+1} \parallel b_j \parallel) = \mathcal{O}\left(\delta^{-2}_{k_n}\right)$$

as in the case of convexity (see Diack 1999).

Remark 4 *Theorem 1 gives the behavior of the test under the hypotheses H_{1n} : $f_n = f + h_n L$. The fixed alternative corresponds to $h_n = 1$. From **theorem 1**, the test statistic diverges to $+\infty$ under any fixed alternative to $H_\circ$. Moreover, the test* *has power to detect local alternatives of the type H_{1n} approaching the null at rate slower than $n^{-1/2}\delta_{k_n}^{-2}$.*

In the case of gaussian errors, assumption $n_i \to +\infty$ is not needed.

Assumption (iii) is the same as in Aggarwal and Studden (1980)

It is easy to see that assumption (iv) implies that $\lim_{n\to+\infty} k_n = +\infty$.

For a uniform design, i.e. $\mu_i = 1/r, i = 1,...,r$ then (iv) is equivalent to $\lim_{n\to+\infty} n\delta_{k_n}^6 = 0$.

3 SCHLEE'S TEST

To construct a monotonicity test of a regression function, Schlee (1980) uses, in a nonparametric model with random design, a modified version of the kernel estimator. Basically, his test statistic only need to estimate the first derivative of the regression function. However, computing the distribution of this statistic comes to compute the distribution of the supremum of this normalized estimator over an intervall. But this method imposes some theorical difficulties. To overcome the problem, he proposes a sequence of points from the interval and uses the theory of maximal deviation to obtain the distribution of the test statistic under the null hypothesis. However, this work does not discuss asymptotic results. Our goal in this section, is to fill this gap. Thus, we will prove the consistency of the test and show that it has power against local misspecifications. However, our framework will be in the following nonparametric model with fixed design:

$$y_i = f(x_i) + \varepsilon_i, \qquad i = 1, ..., n.$$

The deterministic design points x_i lie in $(0,1)$. We assume that the random errors ε_i are uncorrelated and identically distributed with mean zero. Their variance σ^2 will be assumed unknown. Finally f is an unknown and we will use a kernel estimator to approximate it.

3.1 Estimator

A classic estimator of a regression function in nonparametric model (with random design) is the well-known Watson-Nadaraya kernel estimator defined by:

$$\widehat{f}_n(x) = \frac{\sum_{i=1}^n y_i K(\frac{x-x_i}{h})}{\sum_{i=1}^n K(\frac{x-x_i}{h})} \tag{6}$$

where K is a kernel function and where $h = h_n$ is a sequence of positive bandwidths depending on n.

In contrast to the quotient type estimator (6), Gasser and Müller (1979) propose for a fixed design regression problem, the following kernel estimator:

$$\widehat{f}_n(x) = \frac{1}{h}\sum_{i=1}^n \left\{ \int_{s_{i-1}}^{s_i} K(\frac{x-u}{h})du \right\} y_i$$

where

$$0 = s_0 \leq s_1 \leq \cdots \leq s_n = 1 \text{ and } x_i \leq s_i \leq x_{i+1}.$$

For a differentiable function f, using a differentiable kernel, they adopt the following defintion for estimating the first derivative f'.

$$\widehat{f}_{n,1}(x) = \frac{1}{h^2}\sum_{i=1}^n \left\{ \int_{s_{i-1}}^{s_i} K'(\frac{x-u}{h})du \right\} y_i.$$

Some asymptotic properties, as weak and strong consistency are established by Gasser and Müller (1984).We will use $\widehat{f}_{n,1}$ to define a monotonicity test.

3.2 Assumptions

$(A1)$ f is m times continuously differentiable on $[0,1]$ with $m \geq 3$.
$(A2)$ support $K \subset [-\tau, \tau]$, with τ a positive real.
$(A3)$ $\int_{-\tau}^{\tau} K(x)dx = 1$.
$(A4)$ K' is Lipschitz-continuous of order γ with $0 < \gamma \leq 1$.
$(A5)$ K is of order $(1, m)$. The reader is referred to Gasser and Müller (1984) for a discussion about kernel of order (ν, k), where ν and k are positive reals.
$(A6)$ $\exists\, \delta > 1 : \max_j |s_j - s_{j-1}| = \mathcal{O}(1/n^{\delta})$.
Under $(A1)$ through $(A6)$, Gasser and Müller (1984) prove that:

$$E\widehat{f}_{n,1}(t) = \int_{-\tau}^{\tau} K(x) f^{'}(t - xh_n)dx + \mathcal{O}\left(1/nh_n\right) \tag{7}$$

$$\begin{aligned} cov(\widehat{f}_{n,1}(t_1), \widehat{f}_{n,1}(t_2)) &= (\sigma^2/nh_n^4) \int_0^1 K' \left(\frac{t_1 - u}{h_n}\right) K^{'} \left(\frac{t_2 - u}{h_n}\right) du && (8) \\ &\quad + \mathcal{O}\left(1/n^{\delta} h_n^3 + 1/n^{1+\gamma} h_n^{3+\gamma}\right) && (9) \end{aligned}$$

and

$$var(\widehat{f}_{n,1}(t)) = (\sigma^2/nh_n^3) \int_{-\tau}^{\tau} \left\{K^{'}(x)\right\}^2 dx + \mathcal{O}\left(1/n^{\delta} h_n^3 + 1/n^{1+\gamma} h_n^{3+\gamma}\right). \tag{10}$$

3.3 Test statistic

Consider the problem $[H_o, H_1]$. The idea of Schlee's test goes as follows. f differentiable and monotone is equivalent to : for all x, $f'(x) \geq 0$ or, in words:

$$\sup_x \{-f'(x)\} \leq 0.$$

Therefore, if $\widehat{f}_{n,1}$ is a consistent estimator of f', then $P\left(\sup_x \left\{-\widehat{f}_{n,1}(x)\right\} \leq 0\right)$ should be closed to 1 when f is monotone. To see that, it is natural to reject the null hypothesis of the test (that is the monotonicity of f) for the large values of $\sup_x \left\{-\widehat{f}_{n,1}(x)\right\}$. Questions concerning the determination of the distribution of the supremum over the interval $[0,1]$ are unresolved. Nevertheless, it is possible to claculate this distribution when the supremum is taken on a finite number of points (see Schlee 1980). Thus, we will discretize the intervall $[0,1]$.

Let $t_1, ..., t_{k_n}$ a increasing sequel of k_n points of $[0,1]$ such that

$$0 = t_1 < \cdots < t_{k_n} = 1.$$

We will assume that:

$$(A7) : t_{i+1} - t_i \geq 2\tau h_n, \quad i = 1, ..., k_n - 1. \tag{11}$$

Under $(A7)$ it is easy to see that $2\tau h_n(k_n - 1) \leq \sum_{i=1}^{k_n-1}(t_{i+1} - t_i) = 1$. Thus, if h_n converges to zero, we have $k_n h_n = \mathcal{O}(1)$. To determine the asymptotic distribution of $\sup_i \left\{-\widehat{f}_{n,1}(t_i)\right\}$ we will need the following results which about the maximal deviation theory. See Schlee (1980) for a proof.

Theorem 5 *Let $\{X_{nl}\}_l$ be a sequence of random variable with mean zero and unit variance such that*

$$X_{nl} = \sum_{i=1}^{n} Y_{nli},$$

$\{Y_{nli}\}$ independent random variables with a third absolute moment and which are absolutely continuous with respect to the Lebesgue measure.

The correlation coefficient $\rho_{nll'}$ of X_{nl} and $X_{nl'}$ has the properties

(i) $\lim_{n\to+\infty} \rho_{nll'} < 1$,

(ii) there is an integer ϖ such that $\lim_{n\to+\infty} \rho_{nll'} = 0$ if $|l - l'| > \varpi$.

(iii) Let be T_n a sequence of integer, $T_n > 0$, $\lim_{n\to+\infty} T_n = +\infty$. The third moments are assumed to satisfy

(iv) $\lim_{n\to+\infty} \left\{(\log T_n)^2 / T_n\right\} \sum_{i=1}^{T_n} \mathbb{E}X_{nl}^3 = 0$,

(v) $\lim_{n\to+\infty} \sum_{i=1}^{n} \mathbb{E}|Y_{nli}|^3 = 0$.

Then it is valid

$\lim_{n\to+\infty} P\left\{\max_{1\leq l\leq T_n} X_{nl} \leq a_n z + b_n\right\} = \exp\{-\exp(-z)\}$

whereas

$a_n = 1/\sqrt{2\log T_n}$, $b_n = \sqrt{2\log T_n} - \left(1/2\sqrt{2\log T_n}\right)(\log\log T_n + \log 4\pi)$.

The following results gives the asymptotic distribution of $\sup_i \left\{-\widehat{f}_{n,1}(t_i)\right\}$.

Theorem 6 *Under $(A1)$ through $(A7)$ and with the additional assumptions*

$\mathbb{E}|\varepsilon_i|^3 < +\infty$, $i = 1, ..., n$

$nh_n^3 \to +\infty$ *and* $nh_n^{2m+1} \to 0$,

We have:

$\lim_{n\to+\infty} P\ \frac{(nh_n^3)^{1/2}}{\sigma\left[\int_{-\tau}^{\tau}\{K'(x)\}^2 dx\right]^{1/2}} \max_{1\leq l\leq k_n} \left\{\widehat{f}_{n,1}(t_l) - f'(t_l)\right\} \leq a_n z + b_n\Big] = e^{-e^{-z}}$

whith

$a_n = 1/\sqrt{2\log k_n}$, $b_n = \sqrt{2\log k_n} - \left(1/2\sqrt{2\log k_n}\right)(\log\log k_n + \log 4\pi)$.

Proof. From (7) we can write for all $t \in [0,1]$

$$\mathbb{E}\widehat{f}_{n,1}(t) = \int_{-\tau}^{\tau} K(x) f'(t - xh_n)dx + \mathcal{O}(1/nh_n).$$

Using $(A3)$ we can rewrite this equality in the following form

$$\mathbb{E}\widehat{f}_{n,1}(t) = f'(t) + \int_{-\tau}^{\tau} K(x)\{f'(t - xh_n) - f'(t)\}\, dx + \mathcal{O}(1/nh_n).$$

Taylor series expansion gives

$$f'(t - xh_n) = f'(t) + \sum_{j=1}^{m-2} \frac{f^{(j+1)}(t)}{j\,!}(-xh_n)^j + \frac{f^{(m)}(t - xh_n\nu)}{(m-1)!}(-xh_n)^{m-1}$$

with $\nu = \nu(x, h_n, t) \in (0, 1)$. Since K is of order $(1, m)$, we have

$$\int_{-\tau}^{\tau} K(x)\left\{f'(t - xh_n) - f'(t)\right\} dx = \int_{-\tau}^{\tau} K(x)\frac{f^{(m)}(t - xh_n\nu)}{(m-2)!}(-xh_n)^{m-1} dx.$$

Since now, $f^{(m)}$ is continous,

$$\sup_{x\in[0,1]} \mid f^{(m)}(t) - f^{(m)}(t - xh_n) \mid \to 0.$$

Then

$$\sup_{(x,\nu)\in[0,1]\times(0,1)} \mid f^{(m)}(t) - f^{(m)}(t - xh_n\nu) \mid \to 0.$$

Hence,

$$\mathbb{E}\widehat{f}_{n,1}(t) = f'(t) + \frac{f^{(m)}(t)}{(m-1)!}h_n^{m-1}\int_{-\tau}^{\tau}(-x)^{m-1}K(x)dx + o(h_n^{m-1}) + \mathcal{O}(1/nh_n).$$

Therefore,

$$\sup_{t\in[0,1]} (nh_n^3)^{1/2} \mid \mathbb{E}\widehat{f}_{n,1}(t) - f'(t) \mid = \mathcal{O}\left((nh_n^{2m+1})^{1/2}\right). \tag{12}$$

On the other hand, using (8) we can see that

$$\frac{nh_n^3}{\sigma^2\int_{-\tau}^{\tau}\left\{K'(x)\right\}^2 dx} var(\widehat{f}_{n,1}(t)) \to 1. \tag{13}$$

Therefore, with (11) we see that

$$\max_{1\le l\le k_n} \frac{\widehat{f}_{n,1}(t_l) - \mathbb{E}\widehat{f}_{n,1}(t_l)}{\sqrt{var\left\{\widehat{f}_{n,1}(t_l)\right\}}}$$

has the same asymptotic distribution than

$$\max_{1\le l\le k_n} \frac{(nh_n^3)^{1/2}}{\sigma\left[\int_{-\tau}^{\tau}\left\{K'(x)\right\}^2 dx\right]^{1/2}}\left\{\widehat{f}_{n,1}(t_l) - f'(t_l)\right\}.$$

Let define X_{nl} by:

$$X_{nl} = \frac{\widehat{f}_{n,1}(t_l) - \mathbb{E}\widehat{f}_{n,1}(t_l)}{\sqrt{var\left\{\widehat{f}_{n,1}(t_l)\right\}}}$$

and

$$\omega_{nli} = \frac{1}{h_n^2}\frac{\left\{\int_{s_{i-1}}^{s_i} K'(\frac{t_l - x}{h_n})dx\right\}\varepsilon_i}{\sqrt{var\left\{\widehat{f}_{n,1}(t_l)\right\}}}.$$

We have:

$$X_{nl} = \sum_{i=1}^{n}\omega_{nli}.$$

with X_{nl} with mean zero and unit variance and $\{\omega_{nli}\}_i$ independent. To prove the theorem, it is enough to verify than the assumptions $(i)-(v)$ of **theorem 5** are satisfied. We have

$$\rho_{nll'} = \frac{cov\left\{\widehat{f}_{n,1}(t_l), \widehat{f}_{n,1}(t_{l'})\right\}}{\sqrt{var\left\{\widehat{f}_{n,1}(t_l)\right\} var\left\{\widehat{f}_{n,1}(t_{l'})\right\}}}.$$

From $(A7)$

$$| t_l - t_{l'} | \geq 2\tau h_n \quad l \neq l',$$

Then,

$$\int_0^1 K'\left\{\frac{t_l - u}{h_n}\right\} K'\left\{\frac{t_{l'} - u}{h_n}\right\} du = 0.$$

Thus, using (8) and (9) we get

$$cov\left\{\widehat{f}_{n,1}(t_l), \widehat{f}_{n,1}(t_{l'})\right\} = \mathcal{O}(1/n^6 h_n^3) + \mathcal{O}(1/n^{1+\gamma} h_n^{3+\gamma}).$$

We obtain, from (7) that

$$\lim_{n\to+\infty} \rho_{nll'} = 0 \qquad l \neq l'.$$

Thus, (i) and (ii) are satisfied. It is clear than (iii) is also satisfied since that $k_n > 0$ and $k_n \to +\infty$.
Besides,

$$| \omega_{nli} |^3 = \frac{1}{h_n^6} \frac{\left| \left\{ \int_{s_{i-1}}^{s_i} K'(\frac{t_l - x}{h_n}) dx \right\} \right|^3 | \varepsilon_i |^3}{\left[\sqrt{var\left\{\widehat{f}_{n,1}(t_l)\right\}}\right]^3}.$$

Now,

$$\int_{s_{i-1}}^{s_i} K'(\frac{t_l - x}{h_n}) dx = (s_i - s_{i-1}) K'(\frac{t_l - \phi_i}{h_n})$$

for some $s_{i-1} \leq \phi_i \leq s_i$. Using $(A6)$ we obtain

$$\left| \int_{s_{i-1}}^{s_i} K'(\frac{t_l - x}{h_n}) dx \right|^3 \leq \max_{x\in[-\tau,\tau]} | K'(x) |^3 | s_i - s_{i-1} |^3 = \mathcal{O}(1/n^3).$$

From (12) we have

$$\left[\sqrt{var\left\{\widehat{f}_{n,1}(t_l)\right\}}\right]^3 \sim \frac{\sigma^3 \left[\int_{-\tau}^{\tau} \{K'(x)\}^2 dx\right]^{3/2}}{(nh_n^3)^{3/2}}.$$

Therefore

$$\mathbb{E} | \omega_{nli} |^3 = \mathcal{O}\left(1/(nh_n)^{3/2}\right).$$

Then,

$$\sum_{i=1}^n \mathbb{E} | \omega_{nli} |^3 = \mathcal{O}\left(1/(nh_n^3)^{1/2}\right).$$

The assumption (v) is verified. It remains to verify the assumption (iv).
We have

$$(X_{nl})^3 = \left(\sum_{i=1}^{n} \omega_{nli} \right)^3 = \sum_{i,j,k} \omega_{nli}\omega_{nlj}\omega_{nlk}.$$

Since ε_i are independent and with mean zero, we obtain

$$\mathbb{E}(X_{nl})^3 = \sum_{i=1}^{n} \mathbb{E}\omega_{nli}^3.$$

Then,

$$\mid \mathbb{E}(X_{nl})^3 \mid = \mathcal{O}(1/(nh_n^3)^{1/2})$$

Hence,

$$\lim_{n \to +\infty} \mid \{(\log k_n)^2/k_n\} \sum_{l=1}^{k_n} \mathbb{E}X_{nl}^3 \mid \leq \lim_{n \to +\infty} (\log k_n)^2/(nh_n^3)^{1/2}.$$

Now, $k_n h_n = \mathcal{O}(1)$ and

$$(\log h_n)^2/(nh_n^3)^{1/2} = (2h_n^{1/2} \log h_n^{1/2})^2/(nh_n^5)^{1/2} \to 0.$$

The theorem follows. ■

We will define the monotonicity test as follows: we will reject the monotonicity of f at size α when

$$\frac{(nh_n^3)^{1/2}}{a_n \sigma \left[\int_{-\tau}^{\tau} \{K'(x)\}^2 dx \right]^{1/2}} \max_{1 \ l \ k_n} \left\{ -\widehat{f}_{n,1}(t_l) \right\} - \frac{b_n}{a_n} > -\log\{-\log(1-\alpha)\}$$

where a_n and b_n are the sequences of reals defined in **theorem 6.** To prove the consistency of the test we will use the following lemma which introduces the notion of uniformly subgaussian random variables. We shall say the random variables $z_1, ..., z_p$ are uniformly subgaussian random variables if, for some positive β, Γ, we have:

$$\mathbb{E}\{\exp \mid \beta z_i \mid^2\} \leq \Gamma < +\infty, \quad i = 1, ..., p.$$

For a proof of the following lemma, see Kuelbs (1978), **inequality 3.10.**

Lemma 7 *Let $z_1, ..., z_p$ be independent and centered random variables and $a_1, ..., a_p$ be a sequence of reals numbers. Assume that $z_1, ..., z_p$ are uniformily subgaussian with constants (β, Γ). Then there is a constant $\zeta > 0$ depending only on (β, Γ), such that for any positive t,*

$$P\left(\left| \sum_{i=1}^{p} a_i z_i \right| \geq t \right) \leq 2 \exp\left(-\frac{\zeta t^2}{\sum_{i=1}^{p} a_i^2} \right)$$

We are ready now for the main result of this section. Note that **theorem 8** gives a results about the power of the test. However, it is easy to see that the local convergence rate in which this test has power to detect local alternatives is the square root of the rate at which the test statistic converges to infinity (then $n^{-1/2}h_n^{-2}$, see **theorem**).

Theorem 8 *Under the assumptions of theorem 6, we assume that ε_i are uniformly subgaussian random variables and there is a sequence of positive reals δ_n such that $\delta_n \to 0$ and $nh_n^4\delta_n^2 \to +\infty$, Then the test has power to detect local misspecifications.*

Proof. *We have*

$$-\widehat{f}_{n,1}(t) + f'(t) = \left\{-\widehat{f}_{n,1}(t) + \mathbb{E}\widehat{f}_{n,1}(t)\right\} + \left\{-\mathbb{E}\widehat{f}_{n,1}(t) + f'(t)\right\}.$$

From (12),

$$\sup_{t\in[0,1]} (nh_n^3)^{1/2} \mid \mathbb{E}\widehat{f}_{n,1}(t) - f'(t) \mid = \mathcal{O}\left((nh_n^{2m+1})^{1/2}\right).$$

On the other hand

$$\sup_{t\in[0,1]} \left\{-\widehat{f}_{n,1}(t) + \mathbb{E}\widehat{f}_{n,1}(t)\right\} \leq \frac{1}{h_n^2}\sum_{i=1}^{n} \sup_{t\in[0,1]} \left\{\int_{s_{i-1}}^{s_i} K'(\frac{t_l - x}{h_n})dx\right\}\varepsilon_i.$$

Then, using lemma 7, we get

$$P\left[\sup_{t\in[0,1]} \left|-\widehat{f}_{n,1}(t) + \mathbb{E}\widehat{f}_{n,1}(t)\right| \geq \delta_n\right]$$
$$\leq 2\exp\left[-\frac{\zeta\delta_n^2 h_n^4}{\sum_{i=1}^{n}\left\{\sup_{t\in[0,1]}\int_{s_{i-1}}^{s_i} K'(\frac{t_l-x}{h_n})dx\right\}^2}\right] \leq 2\exp(-cn\delta_n^2 h_n^4).$$

Therefore

$$\sup_{t\in[0,1]} \mid -\widehat{f}_{n,1}(t) + \mathbb{E}\widehat{f}_{n,1}(t) \mid = \mathcal{O}_p(\delta_n).$$

Hence,

$$P_f(R_n) \geq P_f\left\{(nh_n^3)^{1/2} \max_{1\ l\ k_n} -f'(t_l) - (nh_n^3)^{1/2}\delta_n - (nh_n^{2m+1})^{1/2} - b_n \geq a_n q_\alpha\right\}$$

where R_n is the critical region of the test. Now, assume that f is non-decreasing, there is a positive real $\varsigma > 0$ and a point x_0 on $[0,1]$ such that $f'(x_0) = \varsigma$. Hence, since f' is continuous and the sequence of t_i is dense on $[0,1]$, we have,

$$\max_{1\leq l\leq k_n} -f'(t_l) > \varsigma.$$

On the other hand,

$$b_n = \mathcal{O}\left(\sqrt{-\log h_n}\right) \qquad \textit{and} \qquad \sqrt{-\log h_n}/(nh_n^3)^{1/2} \to 0.$$

Therefore,

$$(nh_n^3)^{1/2} \max_{1\leq l\leq k_n} -f'(t_l) - (nh_n^3)^{1/2}\delta_n - (nh_n^{2m+1})^{1/2} - b_n \to_p +\infty.$$

Furthermore, $a_n \to 0$ and the consistency of the test follows. ■

References

[1] G.G. Agarwal and W.J Studden. Asymptotic integrated mean square error using least squares and bias minimizing splines. *The Annals of Statistics*, 8(6):1307–1325, 1980.

[2] R.K. Beatson. Monotone and convex approximation by splines: error estimates and a curve fitting algorithm. *SIAM J. of Math. Analysis*, 19(4), 1982.

[3] J.P. Boyle and R.L. Dykstra. A method for finding projections onto the intersection of convex sets in hilbert spaces. *Lectures notes in mathematics*, 37:28–47, 1985.

[4] C.A.T. Diack. A consistent nonparametric test of the convexity of regression based on least squares splines. *Technical report*, 001(submitted), 1999.

[5] H. Dierckx. An algorithm for cubic spline fitting with convexity constraints. *Computing*, 24:349–371, 1980.

[6] J. Kuelbs. Some exponential moments of sums of independent radom variables. *Trans. Amer. Math. Soc.*, 240:145–162, 1983.

[7] W. Schlee. Nonparametric test of the monotony and convexity of regression. *Nonparametric Statistical Inference*, 2:823–836, 1980.

[8] L. Schumaker. *Spline function: Basic theory*. John Wiley, New York, 1981.