MODEL CHECKS FOR REGRESSION UNDER α-MIXING

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

We study the local and global behaviors of regression splines under α-mixing. The asymptotic normality for the regression splines is established. We also prove a central limit theorem for integrated square error of least squares splines estimators. We investigate the limit distribution of the same functional when we substitute a constrained estimator for the regression function. In addition, results on the maximal deviation for some derivatives of the estimators are provided, which leads to the construction of goodness-of-fit-tests and testing the monotonicity or the convexity of the regression function. We prove that the tests are consistent and have power against some local alternatives.

Key words: G-O-F tests, Testing monotonicity, Convexity, Central limit theorem, B-splines, Mixing.

1 Introduction

For i.i.d. observations, the local and global properties of commonly used nonparametric estimators are well known so they allow good methods of statistical inference. However, much less is done in the case of dependent observations. Futhermore, in nonparametric curve estimation under mixing, usually only local properties are established.

We consider the problem of estimating a regression function when the design points are either deterministic or random and the errors are dependent. While it appears difficult to impose properties such as convexity or monotonicity on nonparametric local averaging estimators (for example kernel type estimators), this restriction is readily introduced by using spline estimators. The rate of convergence for such estimators are derived by Burman (1991) under mixing. More recently, Zhou et al.(1998) established asymptotic normality for regression splines when the errors in the model are uncorrelated. Our objective is to obtain local and global measures for the least squares spline as an estimate of the regression function. In particular, we generalize the result of Zhou et al.(1998) by establishing asymptotic normality for least squares spline with dependent errors. We also derive the central limit theorem for the integrated square error of the least squares spline estimator. In addition, we study the asymptotic distribution for the L_2 distance of the regression splines under convexity or monotonicity constraints. We also derive results on the maximal deviation for the first and second derivatives of the least squares spline estimator. We apply these results to validate an asymptotic goodness-of-fit test. We also propose tests of convexity and monotonicity of the regression function. The literature on nonparametric tests is extensive: specification tests are proposed by Hausman (1978), Bierens (1982, 1990), Lee (1988), Eubank and Spiegelman (1990), Wooldridge (1992), Yatchew (1992), Härdle and Mammen (1993), Hong and White(1995) and Yatchew and Bos (1997). Bickel and Rosenblatt (1973) and

Stoker (1989, 1991) propose tests of significance. Surveys in testing monotonicity and convexity include Schlee (1980), Yatchew (1992), Yatchew and Bos (1997), Diack and Thomas (1998), Bowman and al. (1998), Diack (1999, 2000) and Doveh and al.(1999). In contrast to our setting, all these authors assume that the random variables in their models are independent, whereas we assume this only when testing the shape of a regression function.

We consider the following regression model

$$Y_i = g(x_i) + Z_i, i = 1, \cdots, n. \tag{1}$$

The design points $\{x_i\}_{i=1}^n$ can be deterministic or random. Without loss of generality, we assume that $x_i \in [0,1]$. We also assume that $\{Z_k, k \in \mathbb{Z}\}$ is a strictly stationary sequence of real random variables with zero mean on a probability space $(\Omega, \mathcal{A}, P)$. Let

$$\gamma_k = \mathbb{E} Z_i Z_{i+k} \tag{2}$$

be its covariance sequence. Let $\sigma(Z_i, i \leq 0)$ and $\sigma(Z_i, i \geq j)$ be the σ-fields generated by $\{Z_i, i \leq 0\}$ and $\{Z_i, i \geq j\}$, respectively. We assume that the sequence $\{Z_k, k \in \mathbb{Z}\}$ is α-mixing, that is:

$$\alpha_j = \sup_{\substack{A \in \sigma(Z_i, i \leq 0) \\ B \in \sigma(Z_i, i \geq j)}} |P(AB) - P(A)P(B)| \to 0 \text{ as } j \to +\infty.$$

We also introduce the maximal coefficient of correlation

$$\rho_j^* = \sup_{\substack{A \in L_2(\sigma(Z_i, i \leq 0)) \\ B \in L_2(\sigma(Z_i, i \geq j))}} |\mathrm{corr}(A, B)| \to 0 \text{ as } j \to +\infty.$$

This paper is organised as follows: in Section 2 we define the constrained (e.g. monotone and convex) and unconstrained regression spline estimators. Section 3 presents results for asymptotic normality and limit theorems for the L_2 norms of the deviation of the estimate (constrained or unconstrained) from its expected value. We also derive asymptotic distributions for the maximum deviation of the first and second derivatives of the estimate. Section 4 discusses construction and consistency of tests. We also examine their local properties. Technical proofs are given in Section 5.

2 Estimators

To estimate the function g, we use a least squares spline estimator. Let $\eta_0 = 0 < \eta_1 < ... < \eta_{k+1} = 1$ be a subdivision of the interval $[0,1]$ by k distinct points. We define $S(k,d)$ as the collection of all polynomial splines of order d (i.e., degree $\leq d-1$) having a sequence of knots $\eta_1 < ... < \eta_k$. The class $S(k,d)$ of such splines is a linear space of functions with 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. The least squares spline estimator of g is defined by

$$\hat{g}(x) = \sum_{p=1}^{k+d} \hat{\theta}_p N_p(x),$$

where

$$\hat{\theta} = \left(\hat{\theta}_1, \cdots, \hat{\theta}_{k+d}\right)' = \arg\min_{\theta \in \mathbb{R}^{k+d}} \sum_{i=1}^{n} \left\{ Y_i - \sum_{p=1}^{k+d} \theta_p N_p(x_i) \right\}^2 . \tag{3}$$

Cubic spline functions are good for estimating the regression function under local convexity constraints. Indeed, beyond their common use in approximation problems, they allow a simple characterization of convexity as follows: if f is cubic spline, then its second derivative is a linear function between any pair of adjacent knots η_i and η_{i+1} and it follows that f is a convex function in the interval $\left[\eta_i, \eta_{i+1}\right]$ if and only if $f''(\eta_i)$ and $f''\left(\eta_{i+1}\right)$ are both non-negative (this property is used by Dierckx 1980 to define a convex estimator).
For a function f in the class $S(k, 4)$ of cubic splines, we can write:

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

Then,

$$f''(\eta_l) = \sum_{p=1}^{k+4} \theta_p N''_p(\eta_l) = \sum_{p=1}^{k+4} \theta_p d_{p,l},$$

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

$$\left\{ \begin{array}{r} d_{p,l} = 0 \quad \text{if} \quad p \leq l \quad \text{or} \quad p \geq l+4 \\ d_{l+1,l} = \frac{6}{(t_{l+5}-t_{l+2})(t_{l+5}-t_{l+3})} \\ d_{l+3,l} = \frac{6}{(t_{l+6}-t_{l+3})(t_{l+5}-t_{l+3})} \\ d_{l+2,l} = -(d_{l+3,l} + d_{l+1,l}). \end{array} \right. \quad \text{for} \quad l = 0, ..., k+1$$

Let $c_l = (0, 0, ..., 0, -d_{l+1,l}, -d_{l+2,l}, -d_{l+3,l}, 0, ..., 0)' \in \mathbb{R}^{k+4}$ and $\theta = (\theta_1, ..., \theta_{k+4})'$, then

$$f''(\eta_l) = -c'_l\theta.$$

Hence, the cubic spline f is a convex function if and only if $c_l'\theta \leq 0$ for all $l = 0, ..., k+1$. Therefore a convex estimator of the regression function can be defined by

$$\tilde{g}(x) = \sum_{p=1}^{k+4} \tilde{\theta}_p N_p(x),$$

where

$$\tilde{\theta} = \left(\tilde{\theta}_1, \cdots, \tilde{\theta}_{k+4}\right)' = \arg\min_{\substack{\theta \in \mathbb{R}^{k+4} \\ C'\theta \leq 0}} \sum_{i=1}^{n} \left\{ Y_i - \sum_{p=1}^{k+4} \theta_p N_p(x_i) \right\}^2 \tag{4}$$

and $C = (c_0|...|c_{k+1})$, a $(k+4) \times (k+2)$ matrix.
We apply the same reasonning as above and use the quadratic splines to build

a monotone estimator of the regression function. Obviously, this method can be adapted to estimate the regression function when the additional constraints are the non-negativity of the $(d-2)$th derivatives. We again denote this estimator by $\tilde{g}$ and the corresponding vector of parameters by $\tilde{\theta}$. In this case the matrix C is a $(k+d) \times (k+d-2)$ matrix.
Notice that the set of constraints, $\left\{x \in \mathbb{R}^{k+4} : C'x \leq 0\right\}$, is a polyhedral cone and so it is closed and convex. Hence, for a given $\mathbb{Y} = (Y_1, ..., Y_n)'$, the above nonlinear programming problem (4) has an unique solution. Let $N(x)$ be the vector of $N_p(x), p = 1, ..., k+d$ ($d = 4$ for convexity and $d = 3$ for monotonicity) and

$$F = (N(x_1), ..., N(x_n)) \text{ and } M_n = \frac{1}{n} \sum_{i=1}^{n} N(x_i) N(x_i)'.$$

It is easy to see that

$$\tilde{\theta} = \arg \min_{\substack{\theta \in \mathbb{R}^{k+d} \\ C'\theta \leq 0}} \left\| \hat{\theta} - \theta \right\|_{M_n}^2 \tag{5}$$

where for all $x \in \mathbb{R}^{k+d}$, $\|x\|_{M_n}^2 = x' M_n x$.

3 Limit Theorems

In this section we study some local and global properties of our estimates. The asymptotic distribution of the functional

$$\int \{\hat{g}(x) - g(x)\}^2 \, dx$$

is evaluated under appropriate conditions as the sample size $n \to \infty$. We also provide the limit distribution of the same functional when we substitute $\tilde{g}_c$ for g. We start with the asymptotic normality of the unconstrained estimator.

For any two sequences of positive real numbers $\{a_n\}$ and $\{b_n\}$, we write $a_n \sim b_n$ to mean that a_n/b_n stays bounded between two positive constants.

Let $A^{(j)}(x)$ be the vector in $\mathbb{R}^n$ defined by

$$A^{(j)}(x) = \left(a_1^{(j)}(x), ..., a_n^{(j)}(x)\right)' = \frac{1}{\sqrt{n}} F' M_n^{-1} N^{(j)}(x), \ 0 \leq j \leq d-2.$$

Basic least squares arguments prove that

$$\hat{\theta} = \frac{1}{n} M_n^{-1} F \mathbb{Y}. \tag{6}$$

Besides, we can write

$$\hat{g}^{(j)}(x) - \mathbb{E}\hat{g}^{(j)}(x) = \frac{1}{\sqrt{n}} A^{(j)}(x)' \mathbb{Z} \tag{7}$$

with $\mathbb{Z} = (Z_1, ..., Z_n)'$.

We need to specify some conditions. We assume that the sequence of knots is generated by $p(x)$, a positive continuous density on $[0, 1]$ such that

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

We set $\delta_k = \max_{0 \leq i \leq k} \left(\eta_{i+1} - \eta_i \right)$, then it is easy to see that

$$\delta_k \sim k^{-1}. \tag{8}$$

When the design points $\{x_i\}_{i=1}^n$ are deterministic, we assume that

$$\sup_{x \in [0,1]} |H_n(x) - H(x)| = o\left(k^{-1}\right) \tag{9}$$

where $H_n(x)$ is the empirical distribution function of $\{x_i\}_{i=1}^n$ and $H(x)$ is the limit distribution with positive density $h(x)$. Notice that when the sequence $\{x_i\}_{i=1}^n$ is sampled from a distribution $H(x)$, we obtain from the Glivenko-Cantelli Theorem

$$\sup_{x \in [0,1]} |H_n(x) - H(x)| = \mathcal{O}_p\left(n^{-1/2}\right).$$

We denote the $n \times n$ matrices with (i,j)th element $\Gamma_{ij} = \gamma_{|i-j|}$ and $\Gamma_{ij}^+ = \gamma_{i+j}$ by Γ and Γ^+ respectively. We assume that the spectral density of $\{Z_k, k \in \mathbb{Z}\}$ is bounded away from zero and infinity. A classical result on Toeplitz matrices (see Grenander and Szegö, 1984) proves that $2\pi\lambda_{\min}\Gamma$ and $2\pi\lambda_{\max}\Gamma$ (where $\lambda_{\min}\Gamma$ and $\lambda_{\max}\Gamma$ are the smallest and the largest eigenvalues of Γ respectively) converge, respectively, to the minimum and the maximum of the spectral density of Z. Hence, the assumption on the spectral density of $\{Z_k, k \in \mathbb{Z}\}$ guarantees that the eigenvalues of Γ are bounded away from zero and infinity.

For $x \in \left(\eta_i, \eta_{i+1}\right]$, we define the function b_d by

$$b_d(x) = -\frac{g^{(d)}(\eta_i)\left(\eta_i - \eta_{i+1}\right)^d}{d!} B_d\left(\frac{x - \eta_i}{\eta_i - \eta_{i+1}}\right)$$

where $B_d(.)$ is the dth Bernoulli polynomial (see Barrow and Smith 1978). We also set $\xi^{(j)}(x) = \hat{g}^{(j)}(x) - g^{(j)}(x) - b_d^{(j)}(x)$.

We first deal with the unconstrained estimator. Theorem 1 provides the asymptotic normality of $\xi^{(j)}(x)$ for fixed and random design.

Theorem 1 *Let* $g \in \mathcal{C}^d[0,1]$. *Suppose that* $k^{2j+1} = o(n)$ *when* x *is deterministic and* $k^{2j+1} = o\left(n^{1/2}\right)$ *when* x *random,* $0 \leq j \leq d-2$. *Assume that* (8) *and* (9) *hold and* $\lim_{n\to\infty} \rho_n^* < 1$, *then for all* $x \in [0,1]$

$$\frac{\sqrt{n}\xi^{(j)}(x)}{\sqrt{A^{(j)}(x)' \Gamma A^{(j)}(x)}} \rightsquigarrow \mathcal{N}(0,1).$$

Zhou et al.(1998) give a similar result for the case of uncorrelated errors and $j = 0$. Therefore Theorem 1 generalizes their result. A confidence band for $g^{(j)}(x)$ is easily obtained from Theorem 1.

Next, we give a result on the maximal deviation of $\hat{g}^{(d-2)}(x)$. Recall that $\hat{g}^{(d-2)}$ is a linear function between any pair of adjacent knots η_i and η_{i+1}, and it follows that

$$\sup_{x\in[0,1]} \hat{g}^{(d-2)}(x) = \max_{0 \leq i \leq k+1} \hat{g}^{(d-2)}(\eta_i) \text{ and } \inf_{x\in[0,1]} \hat{g}^{(d-2)}(x) = \min_{0 \leq i \leq k+1} \hat{g}^{(d-2)}(\eta_i).$$

Theorem 2 *Suppose that the assumptions of Theorem 1 hold. Then*

$$\mathbb{P}\left\{u_n\left(\max_{0\ \le\ i\ \le\ k+1}\frac{\sqrt{n}\xi^{(d-2)}\left(\eta_i\right)}{\sqrt{A^{(d-2)}\left(\eta_i\right)'\Gamma A^{(d-2)}\left(\eta_i\right)}}-v_n\right)\le x\right\}\to\exp\left(-\exp\left(-x\right)\right) \tag{10}$$

where $u_n=(2\log n)^{1/2}$ *and*

$$v_n=(2\log n)^{1/2}-\frac{1}{2}(2\log n)^{-1/2}(\log\log n+\log 4\pi).$$

A uniform confidence bound for $g^{(d-2)}(x)$ can be constructed from Theorem 2.

Now, we turn our attention to the L_2 distance of the constrained and unconstrained estimators from their expected values. We set

$$T=\int\left\{\hat{g}(x)-g(x)\right\}^2 h(x)\,dx \tag{11}$$

and

$$\tilde{T}=\int\left\{\tilde{g}(x)-\hat{g}(x)\right\}^2 h(x)\,dx. \tag{12}$$

Recall that $h(x)$ is the positive density of the limit distribution $H(x)$. It is convenient to introduce the following notations. We define the $(k+d)\times(k+d)$ matrix M_h by

$$M_h=\int N(x)\,N(x)'\,h(x)\,dx. \tag{13}$$

Let π_{pq} be the $(p,q)\,th$ element of the $n\times n$ matrix $F'M_n^{-1}M_hM_n^{-1}F$ and

$$\pi_0=\frac{1}{n}\mathrm{tr}\left(F'M_n^{-1}M_hM_n^{-1}F\right),\text{for }|p|=1,...,n-1:\pi_p=1/n\sum_{q=1}^{n-|p|}\pi_{|p|+q,q}. \tag{14}$$

We set $\Lambda=(\pi_1,...,\pi_{n-1})'$.

Theorem 3 *Assume that* $\sum_p^\infty\alpha_p^{1-2/\epsilon}<\infty$ *and* $\mathbb{E}|Z_1|^{2\epsilon}<\infty$ *for some* $\epsilon>2$. *We also assume that* $\{Z_t\}$ *is the two-sided moving average*

$$Z_t=\sum_{j=-\infty}^{+\infty}\psi_jX_{t-j},$$

where $X_t\sim IID(0,\sigma^2)$ *and* $\sum_{j=-\infty}^{+\infty}\left|j\psi_j\right|<+\infty$. *Then, under the assumptions of Theorem 1, if* $\mathbb{E}X_1^4=\eta\sigma^4<\infty$,

$$\frac{nT-\frac{1}{n}tr\left(F'M_n^{-1}M_hM_n^{-1}F\Gamma\right)-\frac{nB_{2d}}{(2d)!k^{2d}}\int\left\{g^{(d)}(x)/p^d(x)\right\}^2h(x)\,dx}{\sqrt{2\gamma_0\Lambda'\left(\Gamma+\Gamma^+\right)\Lambda+\pi_0^2\gamma_0^2+(\eta-3)\,\pi_0^2\gamma_0\sum_{p=0}^\infty\gamma_p+\left(\sum_{|p|<n}\pi_p\gamma_p\right)^2}}\rightsquigarrow\mathcal{N}(0,1).$$

The above theorem says that

$$IMSE \approx \frac{1}{n^2}\mathrm{tr}\left(F'M_n^{-1}M_hM_n^{-1}F\Gamma\right) + \frac{B_{2d}}{(2d)!k^{2d}}\int\left\{g^{(d)}(x)/p(x)^d\right\}^2 h(x)\,dx. \quad (15)$$

When the errors in the models are uncorrelated, $\Gamma = \gamma_0 I$ (with I the identity matrix) and $\frac{1}{n}tr\left(F'M_n^{-1}M_hM_n^{-1}F\Gamma\right) \sim k\gamma_0$. Therefore, (15) agrees with the results of Agarwal and Studden (1980).

In fact, Theorem 3 can be rewritten in terms of the parameter $\hat{\theta}$. Let $\bar{\theta}$ be the vector defined by

$$\bar{\theta} = \hat{\theta} - \mathbb{E}\hat{\theta}.$$

Theorem 4 *Under the assumptions of Theorem 3, we have*

$$\frac{n\left\|\bar{\theta}\right\|_{M_h} - \frac{1}{n}tr\left(F'M_n^{-1}M_hM_n^{-1}F\Gamma\right)}{\sqrt{2\gamma_0\Lambda'\left(\Gamma+\Gamma^+\right)\Lambda + \pi_0^2\gamma_0^2 + (\eta-3)\,\pi_0^2\gamma_0\sum_{p=0}^{\infty}\gamma_p + \left(\sum_{|p|<n}\pi_p\gamma_p\right)^2}} \rightsquigarrow \mathcal{N}(0,1).$$

Actually, we believe that Theorem 3 and 4 are new even for the case of uncorrelated errors when the variance is

$$2\gamma_0^2\sum_{p=0}^{n-1}\pi_p^2 + (\eta-1)\,\pi_0^2\gamma_0^2.$$

Next, we deal with the distribution of the functional defined by (12). However, we need stronger assumptions than those in Theorem 3. More precisely, we assume that the sequence $\{Z_t\}$ is a Gaussian white noise. Moreover, we only give the distribution of $\tilde{T}$ for the case in which the regression function is $g \equiv 0$. We denote the corresponding statistic by $\tilde{T}_0$. The question concerning the determination of the distribution of $\tilde{T}$ for any function g is very hard and is unsolved. We will approximate the distribution of $\tilde{T}_0$ with a mixture of chi-squared distributions. For that purpose, we shall measure the distance between these distributions by the following modification of the Mallows distance

$$d(\mu,\nu) = \inf_{X,Y}\left\{\mathbb{E}\left\|X-Y\right\|^2 \wedge 1 : \mathcal{L}(X)=\mu, \mathcal{L}(Y)=\nu\right\}.$$

Convergence in this metric is equivalent to weak convergence (cf Härdle and Mammen 1993).

Recall that the set of constraints is the polyhedral cone $\mathcal{C}[C]$ (thus closed and convex) determined by the $(k+d)\times(k+d-2)$ matrix $C = (c_0|...|c_{k+d-3})$ by

$$\mathcal{C}[C] = \left\{x \in \mathbb{R}^{k+d} : C'x \leq 0\right\}.$$

Let J be a subset (possibly empty) of $\{0,...,k+d-3\}$ and let $\overline{J}$ be its complement. C_J will be the matrix consisting of those columns of C indexed by the elements of J. The matrix $C_{\overline{J}}$ is defined analogously. We denote the cardinality of J by $\#J$.

Theorem 5 *Assume that $\{Z_t\}$ is a Gaussian white noise. Then, under the assumptions of Theorem 1,*

$$d\left\{\mathcal{L}\left(\frac{n\tilde{T}_0}{\gamma_0^2}\right), \mathcal{L}\left(\bar{\chi}^2\right)\right\} \to 0$$

where the random variable $\bar{\chi}^2$ is distributed as a mixture of chi-squared distributions, namely:

$$\mathbb{P}\left(\bar{\chi}^2 \geq s^2\right) = \omega_0 P\left(\chi_0^2 \geq s^2\right) + \sum_{2 \leq j \leq k+d-1} \omega_j P\left(\chi_{k+d-j}^2 \geq s^2\right) \tag{16}$$

$$\text{with } \omega_0 = \mathbb{P}\left(\hat{\theta} \in \mathcal{C}\left[C\right]\right) = \mathbb{P}\left(C'\hat{\theta} \leq 0\right),$$

$$\omega_j = \sum_{q-\#J=j} \mathbb{P}\left[C'_{\bar{J}}\hat{\theta} - \left(C'_{\bar{J}}\Sigma C_J\right)\left(C'_J\Sigma C_J\right)^{-1} C'_J\hat{\theta} \leq 0\right] \mathbb{P}\left[\left(C'_J\Sigma C_J\right)^{-1} C'_J\hat{\theta} \geq 0\right]$$

and where

$$\Sigma = \frac{n}{\gamma_0^2} M_n. \tag{17}$$

Moreover, $\omega_0 + \sum_{2 \leq j \leq k+d-1} \omega_j = 1.$

To calculate the probabilities in the right-hand side of (16), the values of ω_j are needed. However, even for moderate k ($k + d > 3$), closed form expressions for these level probabilities have not been found. Thus approximations are of interest. For this, one may use Monte Carlo method.

4 Inference

Theorems 1 and 2 provide convenient ways to obtain confidence bands for the estimates. However, it would be hard to obtain an explicit confidence band from Theorems 3 or 5. Nevertheless, we can use them to construct nonparametric tests. In this section, we present consistent nonparametric tests. We prove that the tests have asymptotic powers for some local alternatives.

4.1 Hypotheses and test statistics

Goodness-of-fit Tests: The null hypothesis is that $H_\circ : g = g_0$. Against an unrestricted alternative, it is natural to use the L_2 distance between the estimator $\hat{g}$ and g_0. Therefore, the statistic of the test is given by

$$T = \int \left\{\hat{g}(x) - g_0(x)\right\}^2 h(x)\, dx.$$

Using Theorem 3, we see that the null hypothsesis can be rejected at asymptotic level α if

$$nT \geq q_\alpha \sqrt{V} + \frac{1}{n} \mathrm{tr}\left(F' M_n^{-1} M_h M_n^{-1} F\Gamma\right) + \frac{nB_{2d}}{(2d)!k^{2d}} \int \left\{g_0^{(d)}(x)\right\}^2 h(x)\, dx \tag{18}$$

where

$$V = 2\gamma_0 \Lambda' \left(\Gamma + \Gamma^+\right) \Lambda + \pi_0^2 \gamma_0^2 + (\eta - 3)\, \pi_0^2 \gamma_0 \sum_{p=0}^{\infty} \gamma_p + \left(\sum_{|p|<n} \pi_p \gamma_p \right)^2 \tag{19}$$

and q_α is the upper 100α percentile of the standard normal distribution.
Specification test: Under some assumptions, the same cutoff point for the goodness-of-fit test may be used for testing composite hypotheses of the form $H_\circ : g = g_0\,(.,\beta)$ where $\beta \in \Theta$, is an unknown parameter. However, we must use the statistic T by substituting an estimate $\hat{\beta}$ for the unknown parameter β. We need the following assumption:

$$\sup_{t\in[0,1]} \left| g\,(t,\beta) - m\left(t,\hat{\beta}\right) \right| = \circ_p \left(\left\{ n^{-1} k \right\}^{1/2} \right).$$

Under some mild regularity conditions, estimators such as the least squares, generalized method for moments or the adaptive efficient weighted estimators satisfy the required assumption. Hence, the specification test has the same properties with the goodness-of-fit test.
Testing the shape: Our objective is to test the non-negativity of the $(d-2)$th derivative of the regression function. This corresponds to testing the convexity when $d = 4$ (with cubic spline functions). Testing whether the regression function is monotone, specifically non-decreasing, corresponds to $d = 3$. The test statistic can be based on the largest discrepancy between the estimate of the $(d-2)$th derivative of the regression function and zero. Thus, the null hypothesis can be rejected at asymptotic level α when

$$u_n \left(\max_{0\,\leq\, i\,\leq\, k+1} \frac{-\sqrt{n}\hat{g}^{(d-2)}\,(\eta_i)}{\sqrt{A^{(d-2)}\,(\eta_i)'\, \Gamma A^{(d-2)}\,(\eta_i)}} + v_n \right) \geq \log \left(-\frac{1}{\log\,(1-\alpha)} \right). \tag{20}$$

We denote this test by T_M.

Naturally, a testing procedure can also be based on the L_2 distance between the constrained and the unconstrained estimator of the regression function. When proving Theorem 5, we will see , that $\tilde{T}$ has the same limit distribution than $\left\| \hat{\theta} - \tilde{\theta} \right\|_{M_n}^2$. This latter statistic was used in Diack (1999) to test convexity. It can be shown (see the second corollary to Theorem 3.6 on page 2822 of Raubertas et al. 1986) that $\tilde{T}$ is asymptotically stochastically largest among all g in the null hypothesis when $g \equiv 0$, that is, when $\tilde{T} = \tilde{T}_0$. Therefore, following Diack (1999), we reject the null hypothesis at level α when

$$n\tilde{T}_0 \geq \gamma_0^2 s_{\alpha,k}^2 \tag{21}$$

where $s_{\alpha,k}^2$ is defined by

$$\sum_{2\,\leq\, j\,\leq\, k+d-1} \omega_j P\left(\chi_{k+d-j}^2 \geq s_{\alpha,p}^2\right) = \alpha. \tag{22}$$

The test statistic requires computing $\tilde{\theta}$ defined in (5). This is unsolved. Hence, this problem requires extensive numerical work to obtain a solution. We propose an algorithm based on successive 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 $K_j \quad (j = 1, ..., p)$ and it is meant for applications where projections onto the K_j's can be calculated relatively easily. Let $\mathcal{K}$ be the polyhedral cone defined by C. We see that $\mathcal{K}$ can be written as $\bigcap_{j=1}^{k+d-2} K_j$ where K_j is also a convex cone. For all $X \in \mathbb{R}^{k+d}$, we denote the M_n- projection onto $\mathcal{K}$ by $X_{\mathcal{K}}^{M_n}$. The algorithm consists of repeated cycles and every cycle contains $k + d - 2$ stages. Let $X_{mi}^{M_n}$ be the approximation of $X_{\mathcal{K}}^{M_n}$ given by Dykstra's algorithm at the ith stage of mth cycle. The following result (see Boyle and Dykstra, 1985) proves that the algorithm converges correctly.

Theorem 6 *For any* $(1 \leq i \leq k + d - 2)$, *the sequence* $\left\{X_{mi}^{M_n}\right\}$ *converges to* $X_{\mathcal{K}}^{M_n}$ *in the following sense:* $\left\|X_{mi}^{M_n} - X_{\mathcal{K}}^{M_n}\right\|_{\Gamma} \longrightarrow 0$ *as* $m \longrightarrow +\infty$.

Therefore, in practice, instead of using $\tilde{T}_0$ we use $\tilde{T}_{0,m}$ defined by

$$\tilde{T}_{0,m} = \int \left\{\tilde{g}_m(x) - \hat{g}(x)\right\}^2 h(x)\, dx \tag{23}$$

where $\tilde{g}_m$ is the spline function defined by the parameter $\tilde{\theta}_m$, which is the approximation of $\tilde{\theta}$ given by the last stage of the mth cycle of Dykstra's algorithm.

4.2 Asymptotic power

To make a local power calculation for the tests described above, we need to consider the behavior of different statistics (calculated under a fixed but unknown point $g_0 \in H_o$) for a sequence of alternatives of the form

$$g_n(x) = g_0(x) + \tau_n \varphi(x), \tag{24}$$

where g_n lies in the alternative hypothesis, $\varphi(.)$ is a known function and τ_n is a sequence of real variables converging to zero. We can see that $\tilde{T}$ is consistent against the local alternatives, approaching the null at rates slower than $n^{-1/2}k^{1/2}$. The reader is referred to Diack (1999) for a discussion about these results.

Theorem 7 *We suppose that the assumptions of Theorem 3 hold and that*

$$nk^{-1}\tau_n^2 \rightarrow +\infty. \tag{25}$$

Then T *has a power equal to one under the local alternatives. Besides, under the assumptions of Theorem 1, if*

$$\tau_n (\log n)^{1/2} n^{1/2} k^{(-2d+3)/2} \rightarrow +\infty \tag{26}$$

then T_M *has also a power equal to one under* (24).

Discussion: Using regression splines is advantageous when we want to impose properties such as monotonicity or convexity on nonparametric local averaging estimators. On this base, we have proposed goodness-of-fit test, monotonicity and convexity tests. The tests are consistent and have power against some local alternatives. It appears that the tests using the L_2 distance are asymptotically more powerful than those using the maximal deviation. However, the latter, may well be preferred for moderate sample sizes and some alternatives. Besides, the major drawback of the test $\tilde{T}$ is that the computation is not straightforward. On the other hand, an extensive study is necessary in order to relax the assumption of normality on the sequence $\{Z_t\}$.

In applications, the covariance matrix Γ is unknown. Therefore, we must estimate it. The estimators which we shall use for $\gamma_r, r \geq 0$ is

$$\hat{\gamma}_r = \frac{1}{n}\sum_{i=1}^{n-h}\left(Y_i - \bar{Y}\right)\left(Y_{i+r} - \bar{Y}\right), r = 0, ..., n-1,$$

where $\bar{Y}$ is the sample mean. The estimators $\hat{\gamma}_r, r = 0, ..., n-1$, have the desirable property that for each $n \geq 1$, the matrix $\hat{\Gamma}$ with elements $\hat{\Gamma}_{ij} = \hat{\gamma}_{|i-j|}$, is non-negative definite (cf. Brockwell and Davis 1991). However, plug-in $\hat{\Gamma}$ in order to estimate the variance does not guarantee that we have a consistent estimator. This is an open problem which is under study.

5 Proofs

The proofs of the theorems when x is deterministic and when x is random use similar arguments except for the fact that in the latter case, we must write for example $\text{var}\left(\xi^{(j)}(x)\,|x\right)$ instead of $\text{var}\left(\xi^{(j)}(x)\right)$. Hence we give the proofs for the deterministic case only.

Proof of Theorem 1: Reasoning as in the proof of Theorem 1 in Barrow and Smith (1978), it is easy to see that $\mathbb{E}\xi^{(j)}(x) = o\left(k^{-d+j}\right)$. We can write

$$\xi^{(j)}(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} a_i^{(j)}(x)\, Z_i + o\left(k^{-d+j}\right).$$

According to Corollary 2.1 in Peligrad (1996), it enough to show

$$\max_i \frac{\left|a_i^{(j)}(x)\right|}{\sqrt{n\text{var}\left(\xi^{(j)}(x)\right)}} \to 0 \text{ as } n \to \infty \tag{27}$$

and

$$\sup_n \frac{1}{n\text{var}\left(\xi^{(j)}(x)\right)}\sum_{i=1}^{n}\left\{a_i^{(j)}(x)\right\}^2 < \infty. \tag{28}$$

Straightforward calculations prove that $\text{var}\left(\xi^{(j)}(x)\right) = \frac{1}{n}A^{(j)}(x)'\,\Gamma A^{(j)}(x)$. This can be rewritten in the following form

$$\text{var}\left(\xi^{(j)}(x)\right) = \frac{1}{n^2}\text{tr}\left[\Gamma F' M_n^{-1} N^{(j)}(x)\, N^{(j)}(x)'\, M_n^{-1} F\right].$$

Now, using the Lemma 6.5 in Zhou et al. (1998), we get

$$\frac{\lambda_{\min}\Gamma}{n}\operatorname{tr}\left[N^{(j)}(x)\,N^{(j)}(x)'\,M_n^{-1}\right]\le \operatorname{var}\left(\xi^{(j)}(x)\right)\le \frac{\lambda_{\max}\Gamma}{n}\operatorname{tr}\left[N^{(j)}(x)\,N^{(j)}(x)'\,M_n^{-1}\right].$$

Agarwal and Studden (1980) prove that $\lambda_{\min}M_n^{-1}\sim k$ and $\lambda_{\max}M_n^{-1}\sim k$. We can also prove that for each x there is a p such that $\left|N_p^{(j)}(x)\right|\sim k^j$. Therefore using again Lemma 6.5 in Zhou et al. (1998), we have $\operatorname{var}\left(\xi^{(j)}(x)\right)\sim k^{2j+1}/n$. Hence (27) and (28) follow easily. □

Proof of Theorem 2: We define β_i and $\rho_{|i-j|}$ as

$$\beta_i=\frac{\sqrt{n}\xi^{(d-2)}(\eta_i)}{\sqrt{A^{(d-2)}(\eta_i)'\,\Gamma A^{(d-2)}(\eta_i)}}.$$

and

$$\rho_{|i-j|}=\left|\operatorname{corr}\left(\beta_i,\beta_j\right)\right|.$$

From Theorem 1 we know that β_i is asymptotically normally distributed. Therefore, according to Theorem 6.2.1 in Leadbetter et al. (1983) it suffices to prove that $\rho_n\log n\to 0$. We have

$$\operatorname{cov}\left\{\xi^{(d-2)}(\eta_i),\xi^{(d-2)}(\eta_j)\right\}=\frac{1}{n}A^{(d-2)}(\eta_i)'\,\Gamma A^{(d-2)}(\eta_j).$$

Using again Lemma 6.5 in Zhou et al. (1998) we obtain

$$\operatorname{cov}\left\{\xi^{(d-2)}(\eta_i),\xi^{(d-2)}(\eta_j)\right\}^2\le\frac{(\lambda_{\max}\Gamma)^2}{n^2}\left\{A^{(d-2)}(\eta_i)'\,A^{(d-2)}(\eta_j)\right\}^2.$$

On the other hand, we can write

$$A^{(d-2)}(\eta_i)'\,A^{(d-2)}(\eta_j)=\sum_{p,q}m_{pq}N_p^{(d-2)}(\eta_i)\,N_q^{(d-2)}(\eta_j),$$

where m_{pq} are the elements of the matrix M_n^{-1}. One can easily see that $N_p^{(d-2)}(\eta_i)=0$ if $p\le i$ or $p\ge i+d$, and otherwise we have $\left|N_p^{(d-2)}(\eta_i)\right|=\mathcal{O}\left(k^{d-2}\right)$. Moreover, we have $|m_{pq}|=\mathcal{O}\left(k\nu^{|p-q|}\right)$ for some $\nu\in(0,1)$ (see Lemma 6.3 in Zhou et al.1998). Now we take $j=i+n$ to obtain

$$\left|A^{(d-2)}(\eta_i)'\,A^{(d-2)}(\eta_j)\right|=O\left(k^{2(d-2)+1}\nu^{n-d+2}\right).$$

Therefore

$$\rho_n\le c_1\nu^{n-d+2},$$

which proves Theorem 2. □

Proof of Theorem 3: We can write $T=T_1+T_2+T_3$ where

$$T_1=\int\left\{\hat{g}(x)-\mathbb{E}\hat{g}(x)\right\}^2h(x)\,dx,$$

$$T_2 = \int \{g(x) - \mathbb{E}\hat{g}(x)\}^2 h(x)\, dx$$

and

$$T_3 = \int \{\hat{g}(x) - \mathbb{E}\hat{g}(x)\} \{\mathbb{E}\hat{g}(x) - g(x)\} h(x)\, dx.$$

From Theorem 3.1 in Agarwal and Studden (1980), we have

$$T_2 \approx \frac{B_{2d}}{(2d)!k^{2d}} \int \left\{ g^{(4)}(x) / p(x)^d \right\}^2 h(x)\, dx. \tag{29}$$

On the other hand, $\mathbb{E}T_3 = 0$ and $\mathrm{var}(T_3) = o(\mathrm{var}(T))$. Therefore, to prove Theorem 3, it is enough to prove that

$$T_1 \rightsquigarrow \mathcal{N}(U, V), \tag{30}$$

where $U = \frac{1}{n^2}\mathrm{tr}\left(F' M_n^{-1} M_h^{(j)} M_n^{-1} F\Gamma\right)$ and

$$V = 1/n^2 \left(2\gamma_0 \Lambda' \left(\Gamma + \Gamma^+\right) \Lambda + \pi_0^2 \gamma_0^2 + (\eta - 3)\, \pi_0^2 \gamma_0 \sum_{p=0}^{\infty} \gamma_p + \left(\sum_{|p|<n} \pi_p \gamma_p \right)^2 \right).$$

We have

$$T_1 = \frac{1}{n^2} (FZ)' M_n^{-1} M_h M_n^{-1} (FZ).$$

It follows easily that $\mathbb{E}T_1 = U$. Now, since $\{Z_k, k \in \mathbb{Z}\}$ is a strictly stationary sequence, straightforward calculations prove that

$$\mathrm{var}\left(T_1^{(j)}\right) = \mathrm{var}\left[\frac{1}{n} \sum_{|p|<n} \pi_p Z_0 Z_p\right],$$

where the π_p are defined by (14). Hence,

$$\mathrm{var}(T_1) = \frac{1}{n^2} \sum_{|p|<n} \sum_{|q|<n} \pi_p \pi_q \mathrm{cov}(Z_0 Z_p, Z_0 Z_q).$$

This can be rewritten in the following form

$$\mathrm{var}(T_1) = \frac{1}{n^2} \sum_{|p|<n} \sum_{|q|<n} \pi_p \pi_q \left\{ \mathbb{E}Z_0^2 Z_p Z_q - \gamma_p \gamma_q \right\}.$$

Now $Z_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j}$, $\{X_t\} \sim IID(0, \sigma^2)$ with $\mathbb{E}X_t^4 = \eta\sigma^4$. Hence, straigtforward calculations show that

$$\mathbb{E}Z_0^2 Z_p Z_q = (\eta - 3)\, \sigma^4 \sum_{i=-\infty}^{\infty} \psi_i^2 \psi_{i+p} \psi_{i+q} + \gamma_0 \gamma_{p-q} + 2\gamma_p \gamma_q.$$

It follows that

$$\operatorname{var}(nT_1) = \sum_{|p|<n} \sum_{|q|<n} \pi_p \pi_q \left\{ (\eta - 3)\, \sigma^4 \sum_{i=-\infty}^{\infty} \psi_i^2 \psi_{i+p} \psi_{i+q} + \gamma_0 \gamma_{p-q} + \gamma_p \gamma_q \right\}. \quad (31)$$

One can show easily the following equality

$$\sum_{|p|<n} \sum_{|q|<n} \pi_p \pi_q \gamma_0 \gamma_{p-q} = 2\gamma_0 \Lambda' \left(\Gamma + \Gamma^+ \right) \Lambda + \pi_0^2 \gamma_0^2.$$

The last term of (31) is equal to $\left(\sum_{|p|<n} \pi_p \gamma_p \right)^2$.

On the other hand,

$$|\pi_p - \pi_0| \le \begin{array}{c} \frac{1}{n} \left| \sum_{q=1}^{n-p} \sum_{r,s} N_r(x_q) \left(N_s(x_q) - N_s(x_{p+q}) \right) m_{rs}^* \right| \\ + \frac{1}{n} \left| \sum_{q=1}^{n-p} \sum_{r,s} N_r(x_q)\, N_s(x_q)\, m_{rs}^* \right| \end{array} \quad (32)$$

where m_{rs}^* is the (r,s)th element of the matrix $M_n^{-1} M_h M_n^{-1}$. Using equation (6.22) in Agarwal and Studden (1980) and Lemma 6.3 in Zhou et al.(1998), we see that $|m_{rs}^*| = \mathcal{O}\left(k \nu^{|r-s|} \right)$ with $\nu \in (0,1)$. Therefore, it is easy to see that the second term of the right hand side of (32) is $\mathcal{O}(kp/n)$. Besides,

$$|N_s(x_q) - N_s(x_{p+q})| \le |x_{p+q} - x_q| \sup_x |N'(x)|.$$

A classical result on B-splines proves that $\sup_x |N'(x)| = \mathcal{O}(k)$. Moreover, using (9), we see that $|x_{p+q} - x_q| = \circ\left(k^{-1} p \right)$. Finaly, we obtain

$$|\pi_p - \pi_0| = \mathcal{O}\left(\frac{kp}{n} \{ 1 + n\varepsilon_n \} \right) \quad (33)$$

with $\varepsilon_n \to 0$. It is worth noting that $\pi_0 \sim k$. Using this, we can write

$$\begin{aligned} \sum_{|p|<n} \sum_{|q|<n} \pi_p \pi_q \sum_{i=-\infty}^{\infty} \psi_i^2 \psi_{i+p} \psi_{i+q} \;=\; & \pi_0^2 \sum_{|p|<n} \sum_{|q|<n} \sum_{i=-\infty}^{\infty} \psi_i^2 \psi_{i+p} \psi_{i+q} \quad (34) \\ & + \pi_0 \sum_{|p|<n} \sum_{|q|<n} \sum_{i=-\infty}^{\infty} (\pi_q - \pi_0 + \pi_p - \pi_0)\, \psi_i^2 \psi_{i+p} \psi_{i+q} \\ & + \sum_{|p|<n} \sum_{|q|<n} \sum_{i=-\infty}^{\infty} (\pi_q - \pi_0)(\pi_p - \pi_0)\, \psi_i^2 \psi_{i+p} \psi_{i+q}. \end{aligned}$$

Interchanging the order of summation we find that

$$\lim_{n \to \infty} \pi_0^2 \sigma^4 \sum_{|p|<n} \sum_{|q|<n} \sum_{i=-\infty}^{\infty} \psi_i^2 \psi_{i+p} \psi_{i+q} = \pi_0^2 \gamma_0 \sum_{p=0}^{\infty} \gamma_p \sim k^2.$$

We will show that the two other terms of (34) are $\circ\left(k^2 \right)$. In what follows, we will denote the generic constants by $c_1, c_2, \ldots$ Using the absolute summability of $\{ j \psi_j \}$

we have

$$\begin{aligned}
&\pi_0 \left| \sum_{|p|<n} \sum_{|q|<n} \sum_{i=-\infty}^{\infty} (\pi_q - \pi_0 + \pi_p - \pi_0)\, \psi_i^2 \psi_{i+p} \psi_{i+q} \right| \\
&\le c_1 \pi_0 \frac{k}{n} \{1 + n\varepsilon_n\} \sum_{|p|<n} \sum_{|q|<n} \sum_{i=-\infty}^{\infty} (|p| + |q|) \left| \psi_i^2 \psi_{i+p} \psi_{i+q} \right| \\
&\le c_2 \pi_0 \frac{k}{n} \{1 + n\varepsilon_n\} \sum_{|p|<n} \sum_{i=-\infty}^{\infty} |p| \left| \psi_i^2 \psi_{i+p} \right| \sum_{q=-\infty}^{\infty} \left| \psi_{i+q} \right| \\
&\le c_3 \pi_0 \frac{k}{n} \{1 + n\varepsilon_n\} \sum_{p=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} (|p+i| + |i|) \left| \psi_i^2 \psi_{i+p} \right| \\
&\le c_4 \pi_0 \frac{k}{n} \{1 + n\varepsilon_n\} \sum_{p=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} (|p+i| + |i|) \left| \psi_i \psi_{i+p} \right| \\
&\le 2c_4 \pi_0 \frac{k}{n} \{1 + n\varepsilon_n\} \left(\sum_{i=-\infty}^{\infty} |i| \left| \psi_i \right| \right) \sum_{p=-\infty}^{\infty} \left| \psi_p \right| \\
= \;& o\left(k^2\right).
\end{aligned}$$

Reasoning as above, we see that

$$\begin{aligned}
&\left| \sum_{|p|<n} \sum_{|q|<n} \sum_{i=-\infty}^{\infty} (\pi_q - \pi_0)(\pi_p - \pi_0)\, \psi_i^2 \psi_{i+p} \psi_{i+q} \right| \\
&\le c_5 \frac{k^2}{n^2} \{1 + n\varepsilon_n\}^2 \sum_{|p|<n} \sum_{|q|<n} \sum_{i=-\infty}^{\infty} |p|\,|q| \left| \psi_i^2 \psi_{i+p} \psi_{i+q} \right| \\
&\le c_5 \frac{k^2}{n^2} \{1 + n\varepsilon_n\}^2 \sum_{i=-\infty}^{\infty} \left(\sum_{p=-\infty}^{\infty} |p| \left| \psi_i \psi_{i+p} \right| \right)^2 \\
&\le c_6 \frac{k^2}{n^2} \{1 + n\varepsilon_n\}^2 \sum_{i=-\infty}^{\infty} \psi_i^2 \left(\sum_{p=-\infty}^{\infty} (|p+i| + |i|) \left| \psi_{i+p} \right| \right)^2 \\
&\le c_7 \frac{k^2}{n^2} \{1 + n\varepsilon_n\}^2 \sum_{i=-\infty}^{\infty} \psi_i^2 \left(1 + i^2\right) \\
= \;& o\left(k^2\right).
\end{aligned}$$

Hence we have

$$\operatorname{var}(nT_1) = 2\gamma_0 \Lambda' \left(\Gamma + \Gamma^+\right) \Lambda + \pi_0^2 \gamma_0^2 + (\eta - 3)\, \pi_0^2 \gamma_0 \sum_{p=0}^{\infty} \gamma_p + \left(\sum_{|p|<n} \pi_p \gamma_p \right)^2 + o\left(k^2\right). \tag{35}$$

Next, we show that T_1 is gaussian. But we first show that $var\,(nT_1) \sim k^2$. Using Lemma 6.5 in Zhou et al.(1998), we see that we just need to show that $\|\Lambda\|^2 \sim k^2$. We have

$$\|\Lambda\|^2 = \sum_{p=0}^{n-1} \pi_p^2,$$

and

$$|\pi_p| = \frac{1}{n}\left|\sum_{q=1}^{n-p}\sum_{r,s} N_r(x_q) N_s(x_{p+q}) m_{rs}^*\right|$$
$$\leq \frac{c_8 k}{n}\sum_{q=1}^{n-p}\sum_{r,s} N_r(x_q) N_s(x_{p+q}) \nu^{|r-s|}$$

Noting that $N_r(x_q) = 0$ when $x_q \notin (t_r, t_{r+d})$ and since $|x_{p+q} - x_q| = \varepsilon_n k^{-1} p$, we have

$$|\pi_p| \leq \frac{c_9 k}{n}(n-p)\left(\nu^{\varepsilon_n k^{-1}}\right)^p.$$

Hence $\|\Lambda\|^2 \sim k^2$. Using the definition of π_{pq} in (14), we can write

$$T_1 = \frac{1}{n^2}\sum_{p=1}^{n}\sum_{q=1}^{n}\pi_{pq} Z_p Z_q = T_{1,1} + T_{1,2},$$

with

$$T_{1,1} = \frac{1}{n^2}\sum_{p=1}^{n}\pi_{pp} Z_p^2 + \sum_{p\neq q}\pi_{pq}\mathbb{E}Z_p Z_q$$

and

$$T_{1,2} = \frac{1}{n^2}\sum_{p\neq q}\pi_{pq}\left(Z_p Z_q - \mathbb{E}Z_p Z_q\right).$$

We have $\mathbb{E}T_{1,1} = \mathbb{E}T_1$. Besides,

$$\text{var}(T_{1,1}) = \frac{1}{n^4}\sum_{p,q}\pi_{pp}\pi_{qq}\text{cov}\left(Z_p^2, Z_q^2\right).$$

Using Lemma 4.1 in Burman (1991) and since $|\pi_{pp}| \leq k^2$ we have for some $\epsilon > 2$

$$var(T_{1,1}) \leq \frac{c_{10}k^2}{n^4}\sum_{p,q}\alpha_{|p-q|}^{1-2/\epsilon} \leq \frac{c_{10}k^2}{n^3}\sum_{p}\alpha_p^{1-2/\epsilon} = \circ\left(k^2\right).$$

Therefore it is enough to prove that $T_{1,2}$ is Gaussian. We have var$(T_{1,2}) \approx$var(T_1) According to Corollary 2.1 in Peligrad (1996), it is enough to prove

$$\max_{p,q}\frac{|\pi_{pq}|}{n^2\sqrt{var(T_1)}} \to 0 \text{ as } n \to \infty, \tag{36}$$

$$\sup_n \frac{1}{n^4 var(T_1)}\sum_{p,q}\pi_{pq}^2 < \infty, \tag{37}$$

and for every $\zeta > 0$

$$\frac{1}{\sigma_n^2}\sum_{p\neq q}\mathbb{E}\left(Z_p Z_q - \mathbb{E}Z_p Z_q\right)^2 \mathbb{I}\left(|Z_p Z_q - \mathbb{E}Z_p Z_q| > \zeta\sigma_n\right) \to 0 \text{ as } n \to \infty, \tag{38}$$

where $\sigma_n^2 = var\left(\sum_{p\neq q} Z_p Z_q - \mathbb{E}Z_p Z_q\right)$. (36) and (37) are trivial. Reasoning as above one can show that $\sigma_n^2 \geq n^3$ hence, (38) follows easily and Theorem 3 is proven. □

Proof of Theorem 4: We skip the proof of Theorem 4 since it is an immediate consequence of Theorem 3.

Proof of Theorem 5: We have

$$\tilde{T}_0 = \left(\hat{\theta} - \tilde{\theta}\right)' M_h \left(\hat{\theta} - \tilde{\theta}\right).$$

Recall that the regression function is assumed to be null. Now using relation (6.22) in Agarwal and Studden (1980), it is easily seen that $\tilde{T}_0$ has the same asymptotic distribution with the random variable

$$\left(\hat{\theta} - \tilde{\theta}\right)' M_n \left(\hat{\theta} - \tilde{\theta}\right).$$

Therefore, Theorem 5 is a straightforward consequence of the following variant of Theorem 3.1 in Shapiro (1985) (cf. also Diack, 1999):

Let X be a random vector distributed as $N_q(0, I_q)$, then,

$$P\left(\inf_{x\in\mathcal{C}[C]} \|X - x\|^2 \geq s^2\right) = \omega_0 P\left(\chi_0^2 \geq s^2\right) + \sum_{2\,\leq\, j\,\leq\, k+d-1} \omega_j P\left(\chi_{q-j}^2 \geq s^2\right) \quad (39)$$

with $\omega_0 = P\left(X \in \mathcal{C}[C]\right)$,

$\omega_j = \sum_{k+d-\#J=j} P\left(P_J(X) \in \Psi_J\right) P\left(C_J(C_J'C_J)^{-1}C_J'x \in \mathcal{C}^\circ[C]\right)$,

where

$$\Psi_J = \{x \in \mathbb{R}^{k+d} \ : C_J'x = 0, C_{\bar{J}}'x \leq 0\}.$$

$C^\circ[C]$ *being the polar cone of* $C[C]$. *Moreover,* $\omega_0 + \sum_{2\,\leq\, j\,\leq\, k+d-1} \omega_j = 1.$ □

Proof of Theorem 7: (25) follows quite readily from Theorem 3. It remains to prove (26). We define m_i and q_α by

$$m_i = \frac{\sqrt{n}\left(g_0^{(d-2)}(\eta_i) + \tau_n\varphi^{(d-2)}(\eta_i) + b_d^{(d-2)}(\eta_i)\right)}{\sqrt{A^{(d-2)}(\eta_i)'\,\Gamma A^{(d-2)}(\eta_i)}}, q_\alpha = \log\left(-\frac{1}{\log(1-\alpha)}\right).$$

Then the power of the test under the local alternatives is given by

$$\mathbb{P}\left[u_n\left(\max_i\{-\beta_i - m_i\} + v_n\right) \geq q_\alpha\right] \geq \mathbb{P}\left[u_n\left(-\max_i \beta_i + \max_i(-m_i) + v_n\right) \geq q_\alpha\right].$$

Hence to get a power equal to one it is enough to prove that $u_n \max_i(-m_i) \to +\infty$. Because g_n is non-convex and the η_i are dense in $[0,1]$, there is a positive real ϵ such that

$$\max_i\left(-g_0^{(d-2)}(\eta_i) - \tau_n\varphi(\eta_i)\right) > \epsilon.$$

Besides, we have

$$\max_{x\in[0,1]}\left|b_d^{(d-2)}(x)\right| = \mathcal{O}\left(k^{-d}\right),$$

and finally

$$\frac{\sqrt{n}}{\sqrt{A^{(d-2)}(\eta_i)'\,\Gamma A^{(d-2)}(\eta_i)}} \sim \frac{\sqrt{n}}{k^{(-2d+3)}/2}.$$

Therefore the consistency under the local alternatives 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] D.L. Barrow and P.W. Smith. Asymptotic properties of best $l_2[0,1]$ approximation by splines with variables knots. *Quarterly of applied mathematics*, pages 293–304, October 1978.

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

[4] P.J. Bickel and M. Rosenblatt. On some global measures of deviations of density function estimates. *The Annals of Statistics*, 1(6):1071–1095, 1973.

[5] H.J. Bierens. Consistent model specification tests. *Journal of Econometrics*, 20:105–134, 1982.

[6] H.J. Bierens. A consistent conditional moment test of functional form. *Econometrica*, 58:1443–58, 1990.

[7] M.C. Bowman, A.W. Jones and I. Gijbels. Testing monotonicity of regression. *Journal of computational and Graphical Statistics*, 7(4):489–500, 1998.

[8] 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.

[9] P.J. Brockwell and R.A. Davis. *Times Series: Theory and Methods, Second Edition*. Springer Series in Statistics, 1991.

[10] C.A.T. Diack. *Testing convexity.* Proceedings in Computational Statistics: Physica Verlag, 1996.

[11] C.A.T. Diack. A consistent nonparametric test of the convexity of regression based on least squares splines. *Technical Report of EURANDOM*, 001-99(submitted), 1999.

[12] C.A.T. Diack. Sur la convergence des tests de schlee et yatchew. *to appear in Canadian Journal of Statistics*, 28, 2000.

[13] C.A.T. Diack and C. Thomas-Agnan. A nonparametric test of the non-convexity of regression. *Journal of Nonparametric Statistics*, 49(4):335–362, 1998.

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

[15] A. Doveh, E. Shapiro and P.D. Feigin. Testing of monotonicity in regression models. *Prepints*, 1999.

[16] R.L. Dykstra. An algorithm for restricted least squares regression. *Journal of american statistical association*, 78(384), December 1983.

[17] R.L. Eubank and C.H. Spiegelman. Testing the goodness of fit of a linear model via nonparametric regression techniques. *American Statistical Association JASA, Theory and methods*, 85(410):387–392, June 1990.

[18] U. Grenander and G. Szego. *Toeplitz Forms and Their Applications.* Chelsea Publishing Company, New York, 1984.

[19] W. Hardle and E. Mammen. Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*, 21(4):1926–1947, 1993.

[20] J. Hausman. Specification tests in econometrics. *Econometrica*, 46:1251–72, 1978.

[21] Y Hong and H. White. Consistent specification testing via nonparametric series regression. *Econometrica*, 63:1133–1160, 1995.

[22] G. Leadbetter, M.R. Lindgren and H. Rootzen. *Extremes and related properties of random sequences and processes.* Springer-Verlag, 1983.

[23] B.J. Lee. *A nonparametric model specification test using a kernel regression method.* Ph.D. dissertation, University of Wisconsin-Madison, 1988.

[24] M. Peligrad. On the asymptotic normality of sequences of weak dependent random variables. *J. Theor. Prob*, 9(3):703–715, 1996.

[25] R.F. Raubertas, C.I.C. Lee, and E.V. Nordheim. Hypothesis tests for normal means constrained by linear inequalities. *Comm. Statist. Theor. Meth.*, 15:2809–2833, 1986.

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

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

[28] A. Shapiro. Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. *Biometrika*, 72(1):133–144, 1985.

[29] T. Stoker. Tests of additive derivative constraints. *Review of Economic Studies*, 56:535–552, 1989.

[30] T. Stoker. *Lectures on semiparametric econometrics.* Core Foundation Louvain-La-Neuve, 1991.

[31] J.M. Wooldridge. A test for functional form against nonparametric alternatives. *Econometrics Theory*, 8:452–475, 1992.

[32] A. Yatchew and L. Bos. Nonparametric regression and testing in economic models. *J. of Quantitative Economics*, 13:81–131, 1997.

[33] A.J. Yatchew. Nonparametric regression tests based on least squares. *Econometrics Theory*, 8:435–451, 1992.

[34] X. Zhou, S.Shen and D.A. Wolfe. Local asymptotics for regression splines and confidence regions. *The Annals of Statistics*, 26(5):1760–1782, 1998.