# TESTING THE SHAPE OF A REGRESSION CURVE

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

---

**Abstract**

This paper proposes a hypothesis testing procedure for nonparametric regression models based on least squares splines. We assume that the sample is a part of stationary sequence which satisfy a mild mixing property. The approach yields tests of monotonicity and convexity.

**Résumé**

Nous proposons une procédure de test pour la fonction de regression basée sur les splines de régression. Nous supposons que l'échantillon est une partie d'une suite stationnaire qui satisfait des conditions mélangeantes. Notre approche fournit un test de monotonie et de convexité.

Key words: Testing monotonicity, Convexity, Central limit theorem, B-splines, Mixing.

---

## 1 Introduction

In a variety of statistical models, a regression relationship can be assumed to be monotone or convex. A natural question is whether the available data support these assumptions. Therefore, testing monotonicity or convexity provides a way to prevent wrong conclusions. Some papers in statistics literature deal with nonparametric hypothesis tests for convexity or monotonicity of the regression function. Schlee (1982) proposes tests based on the greatest discrepancy between kernel type estimates of the derivatives of the response variable and zero. However, this paper lacks a discussion on consistency and conservativeness. Yatchew (1992) develops tests (with a semi-parametric model) based on comparing the nonparametric sum of squared residuals under monotonicity constraints, with the nonparametric sum of squared residuals without contraints. The Yatchew approach relies on sample splitting which results in a loss of efficiency. Yatchew and Bos (1997) avoid this drawback, essentially by doing an unrestricted nonparametric regression using the residuals from the restricted regression, then testing for significance. However, Yatchew and Bos' test does not have a good power asymptotically. Using a kernel type estimator, Bowman, Jones and Gijbels (1998) developed a test (of monotonicity) based on the size of a critical bandwidth. Bootstrapping is used to calculate the null distribution of the test statistics. The major drawback with this test is that its actual level is not guaranteed and its power can be low when there are flat parts in the regression function. Moreover, asymptotic theory is not provided. Besides, all these tests assume that the random variables in their models are independent.

We consider the following regression model

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

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

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

be its covariance sequence. Let $\sigma(Z_i, i \leq 0)$ and $\sigma(Z_i, i \geq j)$ be the $\sigma$-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 $\alpha$-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 assume that their spectral density is bounded away from zero and infinity.

To estimate the function $g$, we use a least squares spline estimator. If the degree of the polynomials is chosen properly, the first or second derivatives of these estimates are piecewise linear and lead to simple tests for positivity of these derivatives. To get the distribution of the test, we need to prove central limit theorems of the regression spline $\hat{g}(x)$ and its derivatives. We also provide results on the maximal deviation for some derivatives of $\hat{g}(x)$. In fact, these results are interesting in themselves and are formulated in Section 2. We discuss the construction and consistency of tests in Section 3. We also examine their local properties.

## 2 Main Results

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 $\eta_0 = 0 < \eta_1 < \ldots < \eta_{k+1} = 1$ 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 $d$ (degree $\leq d-1$) having a sequence of knots $\eta_1 < \ldots < \eta_k$. 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

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

We define $\delta_k$ by $\delta_k = \max_{0 \leq i \leq k}(\eta_{i+1} - \eta_i)$. Let $N(x)$ be the vector of $N_p(x), p = 1, ..., k+d$ and

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

Let $A^{(j)}(x)$ be a 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

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

where $\mathbb{Y} = (Y_1, ..., Y_n)'$. We can also write

$$\hat{g}^{(j)}(x) - \mathbb{E}\hat{g}^{(j)}(x) = \frac{1}{\sqrt{n}} A^{(j)}(x)' \mathbb{Z} \tag{2}$$

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

We need to specify some conditions. Here we assume that

$$\delta_k \sim k^{-1}. \tag{3}$$

Such an assumption is valid when the knots are generated by a positive continuous density on $[0,1]$. In the case where the design points $\{x_i\}_{i=1}^n$ are deterministic, we assume that

$$\sup_{x \in [0,1]} |H_n(x) - H(x)| = \circ\left(k^{-1}\right) \tag{4}$$

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

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

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 $\xi^{(j)}(x) = \hat{g}^{(j)}(x) - g^{(j)}(x) - b_d^{(j)}(x)$.

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 $k^{2j+1} = \circ(n)$ when $x$ is deterministic and $k^{2j+1} = \circ\left(n^{1/2}\right)$ when $x$ is random, $0 \leq j \leq d-2$. Assume that* (3) *and* (4) *hold, then for all $x \in [0,1]$*

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

*where $\Gamma$ is the $n \times n$ matrix with $(i,j)$th element $\Gamma_{ij} = \gamma_{i-j}$.*

A confidence band for $g^{(j)}(x)$ is easily obtained from Theorem 1. The next result is on the maximal deviation of $\hat{g}^{(d-2)}(x)$. It is worth noting that $\hat{g}^{(d-2)}$ is a linear function between any pair of adjacent knots $\eta_i$ and $\eta_{i+1}$, and it follows that

$$\sup_{x\in[0,1]} \hat{g}^{(d-2)}(x) = \max_{0 \leq i \leq k+1} \hat{g}^{(d-2)}(\eta_i) \text{ and } \inf_{x\in[0,1]} \hat{g}^{(d-2)}(x) = \min_{0 \leq i \leq k+1} \hat{g}^{(d-2)}(\eta_i). \quad (5)$$

Moreover $\hat{g}^{(d-2)}$ is non-negative if and only if it is non-negative on the knots. This is essential for our test procedure.

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

$$\mathbb{P}\left\{u_n\left(\max_{0 \leq i \leq k+1} \frac{\sqrt{n}\xi^{(d-2)}(\eta_i)}{\sqrt{A^{(d-2)}(\eta_i)' \Gamma A^{(d-2)}(\eta_i)}} - v_n\right) \leq x\right\} \to \exp\left(-\exp(-x)\right) \quad (6)$$

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

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

# 3 Inference

In this section we provide tests of monotonicity and convexity.

**Testing convexity:**

We consider the problem of testing whether the regression function is convex or not. The null hypothesis is $H_\circ$ : $g$ is convex and the alternative is $H_1$ : the null is false. The idea of our test procedure is as follows: the function $g$ is twice differentiable and convex if and only if for all $x, g^{(2)}(x) \geq 0$, or, in other words:

$$\sup_x \left\{-g^{(2)}(x)\right\} \leq 0.$$

Therefore, for any consistent estimator $\hat{g}$ of $g$, we can expect that $\mathbb{P}\left(\sup_x \left\{-\hat{g}^{(2)}(x)\right\} \leq 0\right)$ is close to 1 when $g$ is convex. Then, it is natural to reject the null hypothesis of the test (that is the convexity of $g$) for large values of $\sup_x \left\{-\hat{g}^{(2)}(x)\right\}$. We already mentioned that $\hat{g}^{(d-2)}$ is non-negative if and only if it is non negative on the knots. So, the distribution of $\sup_x \left\{-\hat{g}^{(d-2)}(x)\right\}$ is the same with the distribution of $\max\left\{\hat{g}^{(d-2)}(\eta_0), ..., \hat{g}^{(d-2)}\left(\eta_{k+1}\right)\right\}$. Therefore, using a cubic spline estimator ($d = 4$), we can construct a test of convexity based on Theorem 2 as follows. We reject the convexity of $g$ at level $\alpha$ when

$$u_n\left(\max_{0 \leq i \leq k+1} \frac{-\sqrt{n}\hat{g}^{(2)}(\eta_i)}{\sqrt{A^{(2)}(\eta_i)' \Gamma A^{(2)}(\eta_i)}} + v_n\right) \geq \log\left(-\frac{1}{\log(1-\alpha)}\right). \quad (7)$$

**Testing monotonicity:**

More precisely we consider testing whether the regression function is monotonically increasing. The testing procedure is an analogue of the convexity test. However we must use a quadratic spline estimator for $g$ $(d = 3)$. The null hypothesis is rejected at level $\alpha$ when

$$u_n\left(\max_{0\,\le\,i\,\le\,k+1}\frac{-\sqrt{n}\hat{g}^{(1)}\left(\eta_i\right)}{\sqrt{A^{(1)}\left(\eta_i\right)'\,\Gamma A^{(1)}\left(\eta_i\right)}}+v_n\right)\geq\log\left(-\frac{1}{\log\left(1-\alpha\right)}\right) \tag{8}$$

where $\hat{g}^{(1)}$ is the first derivative of the quadratic spline estimator of $g$.

Obviously, these testing procedures can be generalized to testing the non-negativity of the $(d-2)\,th$ derivative of $g$ by using the spline of order $d$.

In applications, the covariance matrix $\Gamma$ is unknown. Therefore, we must estimate it. The estimators which we shall use for $\Gamma_{ij} = \gamma_{|i-j|}$ are

$$\hat{\gamma}_h = \frac{1}{n}\sum_{i=1}^{n-h}\left(Z_i - \bar{Z}\right)\left(Z_{i+h} - \bar{Z}\right), h = 0, ..., n-1$$

where $\bar{Z}$ is the sample mean. The estimators $\hat{\gamma}_h, h = 0, ..., n-1$, 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).

**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 $g_0 \in H_\circ$) for a sequence of alternatives of the form

$$g_n\left(x\right) = g_0\left(x\right) + \tau_n\varphi\left(x\right),$$

where $g_n$ lies in the alternative hypothesis, $\varphi\left(.\right)$ is a known function and $\tau_n$ is a sequence of real variables converging to zero.

**Theorem 3** *We suppose that the assumptions of Theorem 1 hold and that*

$$\tau_n\left(\log n\right)^{1/2}n^{1/2}k^{(-2d+3)/2}\rightarrow+\infty. \tag{9}$$

*Then the test (convexity when $d = 4$ and monotonicity when $d = 3$) has a power equal to one under the above local alternatives.*

Our tests are asymptotically more powerful than the tests cited above. However, bootstrapping may improve considerably the power of the tests. Besides, it would be desirable to study their small sample behaviour through Monte Carlo simulations.

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

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

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

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

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

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

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

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

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

[12] 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.

# 4 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 $var\left(\xi^{(j)}(x)\,|x\right)$ instead of $var\left(\xi^{(j)}(x)\right)$. Hence we give the proofs for the deterministic case only.

Proof of Theorem 1: Reasoning as in proof of Theorem 1 in Barrow and Smith (1978), it is easy to see that $\mathbb{E}\xi^{(j)}(x) = \circ\left(k^{-d+j}\right)$. We can write $\xi^{(j)}(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} a_i^{(j)}(x)\,Z_i + \circ\left(k^{-d+j}\right)$. According to Corollary 2.1 in Peligrad (1996), it suffices to prove

$$\max_i \frac{\left|a_i^{(j)}(x)\right|}{\sqrt{nvar\left(\xi^{(j)}(x)\right)}} \to 0 \text{ as } n \to \infty \tag{10}$$

and

$$\sup_n \frac{1}{nvar\left(\xi^{(j)}(x)\right)} \sum_{i=1}^{n} \left\{a_i^{(j)}(x)\right\}^2 < \infty. \tag{11}$$

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

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

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

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

where $\lambda_{\min}\Gamma$ and $\lambda_{\max}\Gamma$ are, respectively, the smallest and largest eigenvalues of $\Gamma$. A classical result on Toeplitz matrices (see Grenander and Szego 1984) proves that $\lambda_{\min}\Gamma$ and $\lambda_{\max}\Gamma$ converge, respectively, to the minimum and the maximum of the spectral density of $Z$. 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 get $var\left(\xi^{(j)}(x)\right) \sim k^{2j+1}/n$. Hence (10) and (11) follow easily.

Proof of Theorem 2: We define $\beta_i$ 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

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

From Theorem 1 we know that $\beta_i$ 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

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

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

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

We can also write

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

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

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

Therefore

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

which proves Theorem 2.

Proof of Theorem 3: We define $m_i$ and $q_\alpha$ by

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

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

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

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

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

Besides, we have

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

and finally

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

Therefore the consistency under the local alternatives follows from te assumptions of Theorem 3.