Austrian Journal of Statistics
January 2022, Volume 51, 16–34.
http://www.ajs.or.at/
doi:10.17713/ajs.v51i1.1047

Test for Linearity in Non-Parametric Regression Models

Djaballah-Djeddour Khedidja
MSTD Laboratory USTHB, Algeria

Tazerouti Moussa
University of Boumerdes, Algeria

Abstract

The problem of checking the linearity of a regression relationship is addressed. The test uses nonparametric estimation techniques. The null hypothesis is that the regression function is linear; it is tested against the non-specic alternatives hypotheses. This test is based on a Hermite transform characterization of conditional expectations. A statistical test is derived, the distribution of this statistic under the null hypothesis of linearity is determined. A power study using simulation shows the new statistic to be more sensitive to non-linearity.

Keywords: regression, non-linearity, Hermite coefficient, nonparametric regression, random design.

1. Introduction

Let (X, Y) be a pair of real-valued random variables. In many situations, the linear model is insufficient to explain the relationship between the response variable Y and its associated covariates X. A natural generalization is to model the mean nonparametrically in the covariates. Suppose $(X_i, Y_i)_{i=1,...,n}$ are an independent and identically distributed random variables as (X, Y), where Y is variable response and X is the covariates. Consider the following non parametric regression model

$$Y_i = \varphi(X_i) + \varepsilon_i, \quad i = 1, ..., n \tag{1}$$

where φ is assumed to be unknown. The function $\varphi : \mathbb{R} \to \mathbb{R}$ defined by $\varphi(x) = E(Y|X = x)$ is called the regression function of Y on X. The set of values $(X_1, ..., X_n)$ is called the design. The random design setting stands in contrast to the fixed design setting, where the covariates $X_1, ..., X_n$ are fixed (non-random), with only the responses $Y_1, ..., Y_n$ being treated as random. The covariance structure of the design points is completely known and need not be estimated. The residual ε_i are i.i.d. random variables with $E(\varepsilon_i) = 0$ and $var(\varepsilon_i) = \sigma^2$.

Nonparametric regression analysis relaxes the assumption of linearity, substituting the assumption of a smooth regression function. The cost of relaxing the assumption of linearity is much greater computation than in the case of ordinary least squares estimation and, in some instances, a more difficult-to-understand result. A variety of methods are available to estimate φ, based on non parametric regression models. These methods have been proposed to make the specification of the conditional mean function as flexible as possible. The standard

approaches include splines, wavelets, moving averages, running medians, local polynomials, regression trees, neural networks, and other methods like the kernel regression estimators.

Testing the unknown regression function appearing in a nonlinear model has not received much attention in the statistical literature, like the estimation of this non-parametric regression function. The most closely related articles in our literature is the test developed by Mohdeb and Mokkadem (1998) based on the Fourier coefficients. They consider a nonparametric regression with regular et deterministic design, in other words they assume that $X_i = \frac{i}{n}$, φ is a function from $[0,1]$ to $\mathbb{R}$ and the observations Y_i are given by

$$Y_i = \varphi\left(\frac{i}{n}\right) + \varepsilon_i \qquad i = 1, 2, ..., n \tag{2}$$

where ε_i are i.i.d. with mean zero and variance σ^2. They obtain the asymptotic behavior of their proposed test, that is the level and the asymptotic power of the test.

Pearson (1905), developed the test for linearity of regression expressed in terms of the correlation coefficient r and η the correlation ratio. Correlation ratio is a coefficient of non-linear association. In the case of linear relationship, the correlation ratio that is denoted by η becomes the correlation coefficient. In the case of non-linear relationship, the value of the correlation ratio is greater, and therefore the difference between the correlation ratio and the correlation coefficient refers to the degree of the extent of the non-linearity of relationship. The likelihood ratio test statistic is considered by Gallant (1975) for the hypothesis $H : \theta = \theta_0$ against $A : \theta \neq \theta_0$ using the nonlinear regression model $Y = \varphi(X, \theta) + \varepsilon$ with normal errors and unknown variance. A test of normality based on the Hermite polynomials was proposed by Bontemps and Meddahi (2005).

The tests for both the linearity hypothesis and the fit of a regression model have been proposed in some works. Actually, power and consistency have been proven frequently. Let us recall that, testing against a linear regression model literature is very extensive and still growing; we refer to González-Manteiga and Crujeiras (2013), which provided an excellent summary of existing procedures. The literature of these tests are wide, we point out Weihrather (1993) proposed a method for testing the quality of the fit of a linear regression model. Practically, the test statistic is based on a distance measurement between the adjustment of the linear model and the adjustment of the nonparametric model. The properties of the completed samples are studied by way of a simulation experiment with respect to the power of the test in special alternatives. Concerning the Härdle and Mammen (Hardle, Mammen *et al.* 1993) test assumes the non-parametric approach $Y = m(X) + e$ where the only available information is provided through the sample $\{(X_1, Y_1), ..., (X_n, Y_n)\}$. The terms e stands from a random error with a zero mean. Their goal is to test the hypothesis $H_0 : m(.)$ linear, using measurements of the gap between parametric and non-parametric approaches. This test is based on the integrated quadratic difference between parametric adjustment and nonparametric adjustments. It is worth noting that the power of the test has not studied in their work. We mention also the work of Stute and Manteiga (1996). This paper proposed statistics based on some distance between nonparametric and parametric estimation. It is carried out by a minimum-distance criterion, instead of maximum likelihood estimation. Eubank and Spiegelman (1990) procured the test by fitting a spline smoothing together with the residuals of linear regression. They investigated the use of nonparametric regression procedure to test the adequacy of a parametric linear model. The authors consider that the model in such setting are of dependent variable and from a sum of a linear part of a function in a known design points (x_i, $l(x_i)$), where l is unknown function up to an error (written as $y = a^t x + l(x) + e$). The function l is assumed to belong to a general class of functions. The tests are established from non-parametric regression adjustments to the residuals of linear regression. Simulation experiments involving a test based on fitting cubic smoothing splines to residues reveal that this test has good power properties against several alternatives. However, their test is limited by the assumption of normality on the error term. Azzalini and Bowman (1993) examined the problem of verifying the linearity of a regression relationship through the idea of smoothing a

residual plot. The authors apply the pseudo-likelihood ratio approach in the context of linear regression. The true regression function is estimated by non-parametric smoothing and then compared to an adjusted parametric model. Any deviation is assessed by a pseudo-likelihood ratio test. A power study has been examined, which shows that the new statistic is more sensitive to non-linearity compared to that of the Durbin-Watson statistic. Bierens (1982) introduced two coherent model specification tests. The first one is simple, but rather coarse; the second is more involved and laborious test. These tests are based on a characterization by Fourier transform of the conditional expectations. The author used a family of exponential functions to generate an infinite number of moment conditions that are necessary to assess the consistency of the conditional moment test. However, calculating the statistical test requires computing a maximum over an infinite set, which can impose a significant computational load in practice. To overcome the problem, the author proposed to draw randomly a sequence of elements of the infinite set and calculate the maximum. Zheng (1996) proposed a test that combines the idea of the conditional moment test and the methodology of nonparametric estimates. The author used the kernel method to construct a moment condition which can be used to distinguish between the hypotheses, null, and alternative. The test has an advantage over tests based on measuring the distance between parametric and non-parametric models. Actually, it imposes very few regularity conditions, beyond those generally required on nonlinear least squares. Most tests have the inconvenience of the inconsistency with deviations from the parametric model, the general alternatives, or an alternative with infinite dimensions.

The main purpose of our paper is to provide a new test for linearity in the regression model with random design. The approach is testing the linearity of φ from the data $(X_1, Y_1), ..., (X_n, Y_n)$, without estimating it. The statistic is based on the Hermite transform. We proceed as in Djeddour, Mokkadem, and Pelletier (2007) adapting the specificities.

The outline of this paper is as follows. Section 2 introduces the test construction. In Section 3 the main results are presented. In Section 4 we check the accuracy of the test on simulated data. Section 5 a conclusion has been drawn. An appendix provides main mathematical proofs in Section 6. The Section 7 is an appendix gives some definitions of Hermite polynomials.

2. A proposed test for linearity

2.1. Hermite polynomials

We use here the family of orthogonal polynomials on the real line. Generally the polynomial $p(x)$ is written in terms of the monomials x^j. This is known as the natural form of the polynomial. The trouble with the natural form is that the monomials are very highly correlated. The idea behind orthogonal polynomials is to select the basic polynomials $p_j(x)$ to be as different from each other as possible. Two polynomials p_i and p_j are said to be orthogonal if $p_i(X)$ and $p_j(X)$ are uncorrelated as X varies over some distribution.

1. Legendre polynomials are uncorrelated when X is uniform on $(-1, 1)$.
2. Chebyshev polynomials are uncorrelated when X is Beta $(1/2, 1/2)$ on $(-1, 1)$.
3. Laguerre polynomials are uncorrelated when X is gamma on $(0, \infty)$.
4. Hermite polynomials are uncorrelated when X is standard normal on $(-\infty, \infty)$.

There are many ways to approximate functions. However, polynomial approximation is, relatively straightforward for many purposes. The theorem of Weierstrass (Queffélec and Zuily 2007) state that a function, continuous in a finite closed interval, can be approximated with a preassigned accuracy by polynomials.

Remark 1. *The Hermite polynomials are thus orthogonal with respect to the standard normal probability density function* $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ *with mean zero 0 and variance 1.*

The Hermite transform has drawn significant attention, since it exhibits some important properties and high suitability for several applications in different research fields e.g. in astrophysics (Leonis 1980; Öztürk and Gülsu 2014).

Definition 2. Hermite polynomials are a series of polynomials. They are defined as:

$$H_n(x) = (-1)^n \exp(x^2/2)\frac{d^n \exp(x^2/2)}{dx^n}$$

$H_n(x)$ is a polynomial of degree n.

One of the remarkable properties of polynomials $H_n(x)$ is that the derivative of one of them is equal to the antecedent polynomial multiplied by a constant factor, ie:

$$\frac{dH_n(x)}{dx} = nH_{n-1}(x).$$

The other is a relation of recurrence linking three consecutive polynomials:

$$H_n(x) - xH_{n-1}(x) + (n-1)H_{n-2}(x) = 0.$$

The set of two relations is characteristic of polynomials Hermite; the only sequence of polynomials that satisfies the two equations is the sequence of Hermite polynomials.

Approximation of functions by polynomials is basic for a great many numerical techniques. Most numerical analysis texts include a treatment of polynomial approximation. There are many purposes for which polynomial approximation is in statistics. One of them is to model a nonlinear relationship between a response variable and an explanatory variable, as we will see in the sequel. Recall that if $E(\varphi(X)) = 0$ and $E(\varphi(X)^2) < \infty$ for $X \sim N(0,1)$, $\varphi(X)$ can be expanded in Hermite polynomials, that is,

$$\varphi(X) = \sum_{k=1}^{\infty} \frac{c_k}{k!} H_k(X). \tag{3}$$

and

$$c_k = E(\varphi(X)H_k(X)) \ , \ k \geq 1$$

Observe that the expansion (3) starts at $k = 1$, since

$$c_0 = E(\varphi(X)H_0(X)) = E(\varphi(X)) = 0$$

Denote by $k_0 \geq 1$ the Hermite rank of φ, namely the index of the first non-zero coefficient in the expansion (3). Formally $k_0 = \min\{k \geq 1, c_k \neq 0\}$.

Hermite transform is an integral transform, which uses Hermite polynomials $H_n(x)$ as kernels of the transform. The Hermite transform of a function $\varphi(x)$ is

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} H_n(x)\varphi(x)dx.$$

We can generalize the study in the case of p explanatory variables by using Hermite polynomial in $\mathbb{R}^p$. For $\alpha = (k_1, ..., k_n) \in Z^n_{\geq 0}$, we define the Hermite polynomial H_α by

$$H_\alpha(x) = H_{k_1}(x_1) \cdots H_{k_p}(x_p), \quad x = (x_1, ..., x_p) \in \mathbb{R}^p.$$

Because the collection of all Hermite polynomials H_k is an orthonormal basis for the Hilbert space $L^2(\gamma_1)$, we have that the collection of all Hermite polynomials H_α is an orthonormal basis for the Hilbert space $L^2(\gamma_p)$. For γ_p the standard Gaussian measure on $\mathbb{R}^p$, with mean $0 \in \mathbb{R}^p$ and covariance operator $I_{\mathbb{R}^p}$, the collection $\left\{H_\alpha : \alpha \in Z^n_{\geq 0}\right\}$ is an orthonormal basis for $L^2(\gamma_p)$. The disadvantage is the application of the moment formulas of Hermite polynomials which in this case are very painful to handle.

2.2. Test construction

In the following, we consider the regression with a single explanatory variable. This allows us to avoid too heavy calculations which do not enhance the subject. Let X real-valued random variable with the density of probability f on $\mathbb{R}$, assuming it exists and defined by $f(x) = \frac{1}{2\pi} \exp\left(-\frac{x^2}{2}\right)$ (see Remark 1). The problem is to construct a test of the hypothesis

$$H_0 : \varphi \text{ linear against the alternative} \quad H_1 : \ \varphi \text{ non-linear}$$

Let $H_k(x)$ be the Hermite polynomials and $c_k(\varphi)$ the Hermite coefficients of φ defined by:

$$c_k = \int_{-\infty}^{\infty} H_k(x)\varphi(x)f(x)dx \qquad \forall k \tag{4}$$

where φ is the true but unknown regression function. We assume that $f\varphi \in L^2(\mathbb{R})$ where the space $L^2(\mathbb{R})$ of square-integrable functions with respect to the Lebesgue measure on the real line are natural domains on which to define the Hermite transform and Hermite series.

The procedure we consider here is the following. In particular, we are interested in testing for non-linearity in regression models. Then under H_0, the process y_i is linear in mean conditional on x. The approach takes place for the hypotheses

$$H_0 : \varphi(x) = x^t\theta \ \text{ for some } \theta \in \mathbb{R}^2.$$

with x^t denoting the transpose of x. The alternative of interest is the negation of the null, that is,

$$H_1 : \varphi(x) \neq x^t\theta \text{ for all } \theta \in \mathbb{R}^2$$

In regression problems, the mean relationship, as the first-order quantity, is generally of much more interest than higher order properties such as constant variance or normality.

If φ is linear, we write it as

$$\varphi(x) = \alpha x + \beta$$

We establish the Hermite coefficients of φ according to (4); this gives

$$c_k = \ \alpha \int_{-\infty}^{\infty} xH_k(x)f(x)dx + \beta \int_{-\infty}^{\infty} H_k(x)f(x)dx. \tag{5}$$

We can write it as follows:

$$c_k = \alpha\gamma_k + \beta\delta_k$$

with

$$\begin{aligned}\gamma_k &= \int_{-\infty}^{\infty} xH_k(x)f(x)dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} H_1(x)H_k(x)e^{-\frac{x^2}{2}}dx = \begin{cases} 0 \text{ if } k \neq 1 \\ 1 \text{ if } k = 1 \end{cases}\end{aligned}$$

and

$$\begin{aligned}\delta_k &= \int_{-\infty}^{\infty} H_k(x)f(x)dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} H_k(x)e^{-\frac{x^2}{2}}dx \\ &= 0 \qquad \forall k\end{aligned}$$

Then test that φ is linear is equivalent to test if c_k is zero. In this setting, we test the null hypothesis

$$\sum_{k=1}^{\infty} |c_k|^2 = 0. \tag{6}$$

for $k \geq 2$. Let

$$\widehat{c}_k = \frac{1}{n}\sum_{j=1}^{n} H_k(X_j)Y_j \tag{7}$$

be the empirical estimators of c_k and let $m = m(n)$ be a sequence such that $m(n) \to \infty$ as $n \to \infty$ (this allows to consider only a pack of c_k in (6) eg $m = n/10$ or $m = n/2$ or $m = n$). We want to test (6), to this end we construct the statistic

$$\widetilde{T}_{m,n} = \sum_{k=1}^{m} |\widehat{c}_k|^2 \tag{8}$$

We reject H_0 when $\widetilde{T}_{m,n}$ is large.

3. Main results

To simplify matters and without loss of generality, we will assume σ to be known and equal to 1. If $\sigma \neq 1$ but known, this will not change the conclusions but it will weigh down the development of the calculations. It suffices to replace $E(\varepsilon_i^2)$ by σ^2 instead of 1. If σ is not known it is necessary to estimate it and to take into account its law to find the law of the statistic of test. It is not addressed in this work.

We consider first the case of regression of a response variable on a single covariate, with observed values $y = (y_i, ..., y_n)$ and $x = (x_1, .., x_n)$ respectively, expressed in the model (1) where φ is assumed to be unknown and where the ε_i are independent random variables with mean 0 and standard deviation 1. Our aim is to assess whether model (1) can be reduced to the simple linear form $y_i = \alpha x + \beta x_i + \varepsilon_i$ $(i = 1, ..., n)$ with least squares estimators of the intercept and slope parameter.

Proposition 3. *Under H_0, we have*

$$E\widehat{c}_k = 0 \text{ with } k \geq 2 \tag{9}$$

Proposition 4. *Under H_0, we show that*

$$var(\widehat{c}_k) = \frac{k!}{n} \times (\alpha^2(2k+1)+1) \tag{10}$$

We can write

$$\begin{aligned} E\left(|\widehat{c}_k|^2\right) &= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{j'=1}^{n} E\left(H_k(X_j)Y_jH_k(X_{j'})Y_{j'}\right) \\ &= \frac{1}{n}\sum_{j=1}^{n} E\left(H_k^2(X_j)Y_j^2\right) + \frac{1}{n^2}\sum_{j\neq j'}^{n} E\left(H_k(X_j)Y_jH_k(X_{j'})Y_{j'}\right) \end{aligned}$$

The empirical Hermite coefficients are defined as (7). Under H_0, this can be write as

$$\begin{aligned} \widehat{c}_k &= \frac{1}{n}\sum_{j=1}^{n}\left[H_k(X_j)\left(\alpha X_j + \varepsilon_j\right)\right] \\ &= \frac{\alpha}{n}\sum_{j=1}^{n} H_k(X_j)X_j + \frac{1}{n}\sum_{j=1}^{n} H_k(X_j)\varepsilon_j \\ &= \widehat{c}_{k1} + \widehat{c}_{k2} \end{aligned}$$

We prove the following theorems.

Proposition 5. *At fixed k, we demonstrate that*

$$\frac{\widehat{c}_{k1}}{\sqrt{\alpha^2 k!/n}} \sim N(0,1) \tag{11}$$

Proposition 6. *We show that the term $\widehat{c}_{k2} = \frac{1}{n}\sum_{j=1}^{n} H_k(X_j)\varepsilon_j$ converges to 0 in probability.*

By combining the result of proposition (5) and result of proposition (6), we conclude to the asymptotic normality of $\widehat{c}_k$ (at k fixed).

Proposition 7. *Under H_0, we have*

$$\begin{aligned} cov(\widehat{c}_k, \widehat{c}_l) &= \alpha^2 \times \frac{k!}{n} \text{ if } k = l+2 \\ &= var\widehat{c}_k \text{ if } k = l \\ &= 0 \text{ elsewhere} \end{aligned} \tag{12}$$

The variance matrix is written as:

$$n\Sigma = \begin{pmatrix} 2!(5\alpha^2+1) & 0 & \alpha^2 4! & 0 & 0 & \cdots & 0 \\ 0 & 3!(7\alpha^2+1) & 0 & \alpha^2 5! & 0 & \cdots & 0 \\ \alpha^2 4! & 0 & 4!(9\alpha^2+1) & 0 & \alpha^2 6! & \cdots & 0 \\ 0 & \alpha^2 5! & 0 & 5!(11\alpha^2+1) & 0 & \cdots & \vdots \\ 0 & 0 & \alpha^2 6! & 0 & 6!(13\alpha^2+1) & \cdots & \alpha^2 m! \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \alpha^2 m! & 0 & m!((2m+1)\alpha^2+1) \end{pmatrix}$$

It is positive definite because there is no almost sure affine relation between the components of the random vector. Since $\widehat{c}_k$ are correlated with each other, instead of taking $\widetilde{T}_{m,n}$ (defined by 8) as a test statistic, one can use the following test statistic

$$T_{m,n} = C_m^t \Sigma^{-1} C_m.$$

From what precedes, we can state the following theorem.

Theorem 8. *Under H_0, the test statistic $T_{m,n} \sim \chi^2(m-1)$, where m is fixed.*

Proof. The following proof characterizes the behavior of our statistics under the null hypothesis. Let $C_m = (\widehat{c}_2, ..., \widehat{c}_m)^t$. Concerning the asymptotic law of C_m; let $x_2, \ldots, x_m$, real fixed and be $V = (x_2 \ldots x_m)^t$. It must be shown that $V^t C_m = \sum_{k=2}^m x_k \widehat{c}_k$ converges in distribution to $N(0, V^t \Sigma V)$. We deduce that C_m converges in law towards $N(0, \Sigma)$, from there it comes that $T_n = C_m^t \Sigma^{-1} C_m$ converges in law towards $\chi^2(m-1)$.

Martingale theory is used to obtain a central limit theorem for C_m-statistics. By posing $\xi_{m,k} = x_k \widehat{c}_k$ $m = 2, 3, ...$ $k = 2, ..., m$. For a triangular array we can introduce the row-sums

$$S_m = \sum_{k=1}^{m} \xi_{m,k} \qquad m = 1, 2, ...$$

Furthermore

$$\sum_{k=2}^{m} E\left(\xi_{m,k}^2 1_{|\xi_{m,k}|>\varepsilon}\right) = E\left[x_k^2 \widehat{c}_k^2 1_{|x_k \widehat{c}_k|>m\varepsilon}\right]$$

As $E\left(V^t C_m\right) = V^t \Sigma^{-1} V < +\infty$, it's immediate that for any $\varepsilon > 0$,

$$\lim_{m \to +\infty} E\left[x_k^2 \widehat{c}_k^2 1_{|x_k \widehat{c}_k|>m\varepsilon}\right] = 0 \tag{13}$$

The martingale Central Limit theorem then assures that $\sum_{k=2}^m x_k \widehat{c}_k$converges in law to an $N(0, V^t \Sigma V)$. We deduce that C_m converges in distribution to $N_{m-1}(0, \Sigma)$. □

Recall that a sequence of random vectors $(U_n)_n$ of $\mathbb{R}^{m-1}$ converges in law towards a random vector U if and only if $x^t U_n$ converges in law towards $x^t U$ for all $x \in \mathbb{R}^{m-1}$ (from the characterization of the law convergence using the characteristic functions). To be able to use the test statistic $T_{m,n}$ for testing we must calculate it using sample and compare with the quantile of the distribution under H_0. Rejection of this null hypothesis generally leads to the belief of the existence of non-linear trend. Despite its mathematical convenience, there is no special reason to believe a simple linear trend function would be suitable to model the complex system. According to the central limit theorem, when m is "large" ($m > 100$), the law of a variable of χ^2, a sum of independent random variables, can be approximated by a normal law of expectation $(m-1)$ and of variance $2(m-1)$. Under the null hypothesis H_0, we deduce that for m large, $T_{m,n}$ will be normally distributed with mean zero and with a given variance. Under alternative hypothesis H_1 that the Hermite cœfficients $\widehat{c}_k \neq 0$, it is possible that for m large, $T_{m,n}$ will be normally distributed with a given variance which is a very complicated function. Despite the seemingly usually setting, the answer to this question is highly non-trivial.

4. Numerical results

4.1. Case studies

A necessary condition for an effective analysis of statistical data is that statistical models summarize the data with precision. Nonparametric regression does not specify the form of the regression function before examining the data. This theory might suggest that y depends on x, but it is unlikely to tell us that the relationship is linear.

This section features some simulation experiments of the test statistics which are performed out to assess the usefulness and the accuracy of the results obtained in Section 3. First we suppose that $y_i = \alpha x_i$ with $\alpha = 0.8$, we generate a random variable X according to the

Gaussian ditribution $N(0,1)$, we get $(y_1, y_2....y_n)$. In order to verify that the $\widehat{c}_k$ are Gaussian, we plot the histogram of a single coefficient namely $\widehat{c}_3$. For $k = 3$ fixed, that is

$$\widehat{c}_3 = \frac{1}{n}\sum_{j=1}^{n} H_3(X_j)Y_j$$

which follows a normal distribution with the size of the sample $n = 40000$, this is shown in Figure 1:

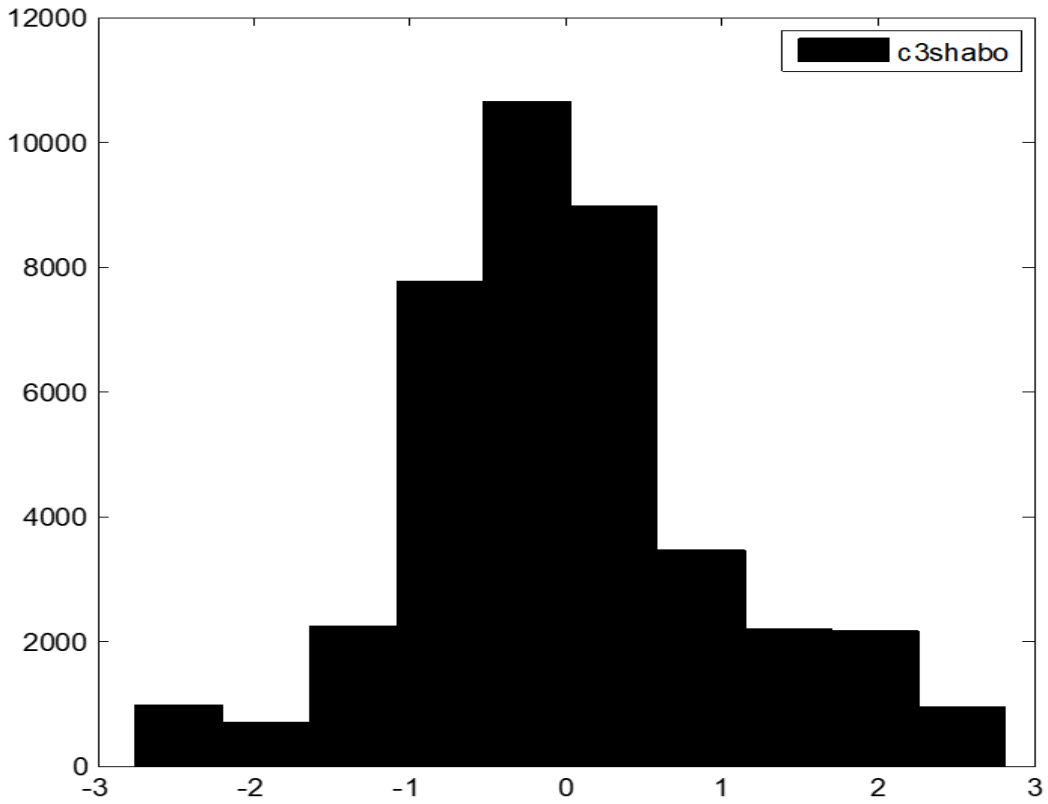

Figure 1: : Histogram of $\widehat{c}_3$ statistic

Figure 1 displays the histogramme of the statistic $\widehat{c}_3$, from which it is clear that there is convincing evidence of normality.

We present some results in order to glimpse a brief description of the test performance under H_0. The test statistic $T_{m,n}$ is calculated on a set of simulated data with different sizes n and m and for different values of α. Let

$$y = \alpha x \tag{14}$$

We follows the steps below:

1. we generate a random variable X which follows a law $N(0,1)$. We get $(y_1, y_2....y_n)$, for α fixed.
2. we calculate recursively the Hermite polynomials
3. we evalute $\widehat{C}_m$

Table 1 reports the test statistics calculated for different values of α, m and n, for the 5% critical value.

Table 1: Test statistic $T_{m,n}$ based on different m, n and α under linearity

		$m = 21$	$m = 33$	$m = 45$
α	n	$T_{m,n}$	$T_{m,n}$	$T_{m,n}$
-2	50	23.13	34.11	45.71
-1.5	70	22.88	34.42	45.88
-1	100	22.39	34.10	45.59
0.4	120	22.63	34.17	45.82
0.5	150	22.52	34.07	45.84
0.8	170	21.73	34.24	45.63
1.7	200	23.06	34.13	45.48
8	200	22.44	34.04	45.83

Reading the quantiles on the χ^2 table, give us: $\chi^2_{(m-1,0.05)} = 31.41$ for $m = 21$, $\chi^2_{(m-1,0.05)} = 46.19$ for $m = 33$ and $\chi^2_{(m-1,0.05)} = 60.48$ for $m = 45$. From the table we notice for all setting, that $T_{m,n} < \chi^2_{(m-1,0.05)}$, then we accept H_0 significantly at the level 5%; i.e φ linear, that's what we expected. The results in table 1 show that if the null is true then the test statistics are less the quantile of the chi-square distribution at $m - 1$ degrees of freedom, setting the risk at 5%.

In the afterpart, we investigated results test under H_1. Let

$$y = a \exp(bx) \tag{15}$$

For that we do a simulation that follows the steps below:

1. we generate a random variable X which follows a law $N(0, 1)$. We get $(y_1, y_2....y_n)$, for α, β fixed.
2. we calculate recursively the Hermite polynomials
3. we evalute $\widehat{C}_m$

Several values of the statistic $T_{m,n}$ are computed for different m, n and α.

Table 2: Test statistic $T_{m,n}$ based on different m, n and α under non-linearity

α	n	m	$T_{m,n}$	$\chi^2_{(m-1,0.05)}$
-1.5	50	21	33.61	31.41
-2	70	33	49.51	46.19
1.7	100	45	62.06	60.48
0.8	120	60	81.02	77.93
0.4	150	75	96.521	95.08
-1	170	88	112.32	109.77
0.5	200	99	135.72	122.11

From the table, we notice that $T_{m,n} > \chi^2_{(m-1,0.05)}$, so we reject H_0 significantly at the level 5%; i.e φ is non-linear, that's what we expected.

4.2. Simulation study

In this section, we demonstrate the performance of our test by some numerical examples. The first set of examples concerns the linear model. We applied the test in the form described in section 3. Moreover, in order to see the sensitivity of the results for the choice of n and m, we have applied the test several times, namely for different values of n and m. The results are

presented in Table 3. The test is evaluated with replications. The simulation are repeated 500 times for each setting and we calculate the average $\overline{T}_{m,n}$ at each setting and then compare the average $\overline{T}_{m,n}$ with the chi-square value. The test statistic $T_{m,n}$ is calculated on datasets with different size n and m and for different values of α accordind to (14). We get

Table 3: Test statistics under the linearity assumption

$\alpha = -2$	$m = 21$	$m = 33$	$m = 45$	$\alpha = 0.5$	$m = 21$	$m = 33$	$m = 45$
n	$\overline{T}_{m,n}$	$\overline{T}_{m,n}$	$\overline{T}_{m,n}$	n	$\overline{T}_{m,n}$	$\overline{T}_{m,n}$	$\overline{T}_{m,n}$
50	19.2533	32.1525	40.1096	50	19.1691	33.0432	44.5460
100	21.2145	34.1309	42.0944	100	21.3601	33.2094	44.1103
150	21.3175	33.2865	43.5910	150	22.2145	34.1309	46.0944
200	22.3490	34.8427	45.0174	200	23.340	36.2765	48.3591

$\alpha = 1.7$	$m = 21$	$m = 33$	$m = 45$
n	$\overline{T}_{m,n}$	$\overline{T}_{m,n}$	$\overline{T}_{m,n}$
50	21.6901	35.3385	43.6143
100	22.2601	34.2641	46.3710
150	23.9813	37.0461	48.9156
200	24.4398	38.1034	54.6018

According to the chi-square table we have:$\chi^2_{(m-1,0.05)} = 31.41$ for $m = 21$, $\chi^2_{(m-1,0.05)} = 46.19$ for $m = 33$ and $\chi^2_{(m-1,0.05)} = 60.48$ for $m = 45$. From the Table 2, we notice that $T_{m,n} < \chi^2_{(m-1,0.05)}$, so we accept H_0 significantly at the level 5%; i.e φ is linear, that's what we expected. In Table 3 we see that the test statistics of tests are rather sensitive for the choice of m.

Some additional numerical work has been carried out. The second set of examples concerns the non-linear model. The test is evaluated with replications. The simulation are repeated 500 times for each setting and we calculate the average $\overline{T}_{m,n}$ at each setting and then compare the average $\overline{T}_{m,n}$ with the chi-square value. The test statistic $T_{m,n}$ is calculated on datasets with different size n and m depending on the value of n and for different values of α accordind to (15).

Table 4: Test statistics under the non-linearity assumption

$\alpha = -2$	$m = 21$	$m = 33$	$m = 45$	$\alpha = 0.5$	$m = 21$	$m = 33$	$m = 45$
n	$\overline{T}_{m,n}$	$\overline{T}_{m,n}$	$\overline{T}_{m,n}$	n	$\overline{T}_{m,n}$	$\overline{T}_{m,n}$	$\overline{T}_{m,n}$
50	30.5320	45.7331	54.6814	50	29.6534	39.7614	60.5713
100	30.2657	42.2570	67.5193	100	30.0716	48.2309	59.0467
150	31.1375	48.6103	69.6392	150	32.7147	51.3572	63.8631
200	33.8342	66.3265	71.0361	200	44.4396	60.9123	67.632

$\alpha = 1.7$	$m = 21$	$m = 33$	$m = 45$
n	$\overline{T}_{m,n}$	$\overline{T}_{m,n}$	$\overline{T}_{m,n}$
50	26.8371	41	53.8745
100	31.4529	39	59.0925
150	43.9635	47	61.5420
200	50.0248	56	65.9562

According to the chi-square table we have:$\chi^2_{(m-1,0.05)} = 31.41$ for $m = 21$, $\chi^2_{(m-1,0.05)} = 46.19$ for $m = 33$ and $\chi^2_{(m-1,0.05)} = 60.48$ for $m = 45$. From the Table 2, we notice that often $T_{m,n} > \chi^2_{(m-1,0.05)}$, so we reject H_0 significantly at the level 5%; i.e φ is non-linear, that's what we expected.

The test was evaluated in terms of the power. For the power performance, we has select a

Table 5: The empirical size and empirical power of the test based on 500 independent replications (in %)

$n\|m$	$n=50\|m=21$		$n=100\|m=45$		$n=150\|m=75$		$n=200\|m=99$	
	size	power	size	power	size	power	size	power
$\alpha=-2$	1.2	10.9	2.4	32.2	2.7	45.6	2.6	40.4
$\alpha=0.5$	2.5	6.2	2.1	47.1	1.2	57.8	1.1	57.8
$\alpha=1.7$	1.8	21.3	1.2	54.6	2.1	41.2	1.8	66.2

simulation setting and generate 500 data sets under the alternative hypothesis and for each data calculate the test statistic. Then we calculate the number of times the null hypothesis is rejected.

For Type I error, we follow the same approach but the data is generated from the null hypothesis and we calculate the proportion of times the null hypothesis is rejected. We get

We call by empirical size, the percentage of falsely rejecting the null hypothesis H_0. On the other hand, the empirical power represents the percentage of rejection of H_0 when we arbitrary choose a false model. The results in Table 5 numerically confirm the results announced. We notice that the test is moderately powerful but theType I errorr is very often small. We deduce that the consistency of our procedure is numerically convincing, which is in accordance with we expected.

5. Conclusion

In this work, we developed a linearity test of a linear regression model using the use of Hermite polynomials. A nonparametric regression methodology with random design is performed. We have considered simple linear regression, but the procedure also applies to multiple regression, we emphasize that the results can be extended to multiple regression. For example, other models can be treated with several predictor variables which can have a general linear. This would seem to complicate the analysis and even the interpretations.

Hermite polynomials led the use of the Gaussian density function for the random explanatory variable X. The method does not work for categorical predictors because qualitative variables cannot have a Gaussian density which is continuous.

Extensions to this situation should be possible using other polynomials and will be explored in future research. These polynomials can be those of Legendre when X is uniform over $(-1,1)$, Chebyshev with X of Beta probability distribution $(1/2,1/2)$ over $(-1,1)$ or Laguerre are when X is of gamma distribution on $(0,\infty)$.

The proposed test was found to have reasonable properties in a simulation study. A small power study was carried out to compare the performances of the test. Some properties of the proposed test were discussed. Applications to simulated data suggested that the proposed test can improve the estimate of the function de regression.

Appendix 1: Proofs

Proof of Proposition 3

We have

$$\widehat{c}_k = \frac{1}{n}\sum_{j=1}^{n} H_k(X_j)Y_j \tag{16}$$

Under H_0,

$$\begin{aligned}\widehat{c}_k &= \frac{1}{n}\sum_{j=1}^{n} H_k(X_j)\left(\alpha X_j + \varepsilon_j\right) \\ &= \alpha \times \frac{1}{n}\sum_{j=1}^{n} X_j H_k(X_j + \frac{1}{n}\sum_{j=1}^{n} H_k(X_j)\varepsilon_j\end{aligned}$$

with $k \geq 2$

$$\begin{aligned}E\left|\widehat{c}_k\right| &= E\left(H_k(X_j)Y_j\right) \\ &= E\left[H_k(X_j)\left(\alpha X_j + \varepsilon_j\right)\right] \\ &= \alpha E\left(X_j H_k(X_j)\right) + \varepsilon_j\ H_k(X_j)) \\ &= \alpha E(H_1(X_j)H_k(X_j)) + E\varepsilon_j \times\ E(H_k(X_j) \\ &= 0\end{aligned}$$

Proof of Proposition 4

To describe the technical development, we write

$$\begin{aligned}E\left(|\widehat{c}_k|^2\right) &= \frac{1}{n^2}E\left(\sum_{j=1}^{n} H_k(X_j)Y_j \sum_{j'=1}^{n} H_k(X_{j'})Y_{j'}\right) \\ &= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{j'=1}^{n} E\left(H_k(X_j)Y_j H_k(X_{j'})Y_{j'}\right) \\ &= \frac{1}{n}\sum_{j=1}^{n} E\left(H_k^2(X_j)Y_j^2\right) + \frac{1}{n^2}\sum_{j\neq j'}^{n} E\left(H_k(X_j)Y_j H_k(X_{j'})Y_{j'}\right) \qquad (17)\end{aligned}$$

In other words we have under H_0

$$\begin{aligned}E\left(|\widehat{c}_k|^2\right) &= \frac{1}{n^2}\sum_{j=1}^{n} E\left(H_k^2(X_j)\left(\alpha X_j + \varepsilon_j\right)^2\right) + \frac{1}{n^2}\sum_{j\neq j'}^{n} E\left(H_k(X_j)\left(\alpha X_j + \varepsilon_j\right)H_k(X_{j'})\left(\alpha X_{j'} + \varepsilon_{j'}\right)\right) \\ &= \frac{1}{n}\sum_{j=1}^{n} E\left(H_k^2(X_j)\left(\alpha^2 X_j^2 + 2\alpha\varepsilon_j X_j + \varepsilon_j^2\right)\right) \\ &\quad + \frac{1}{n^2}\sum_{j\neq j'}^{n} E\left(H_k(X_j)H_k(X_{j'})\left(\varepsilon_j\varepsilon_{j'} + \alpha X_j\varepsilon_j + \alpha X_{j'}\varepsilon_{j'} + \alpha^2 X_j X_{j'}\right)\right) \qquad (18)\end{aligned}$$

To handle this problem, there are two parts.

The first term of (18) : $\frac{1}{n}\sum_{j=1}^{n} E\left(H_k^2(X_j)\left(\alpha^2 X_j^2 + 2\alpha\varepsilon_j X_j + \varepsilon_j^2\right)\right)$ Knowing that $E\varepsilon_n = 0$ $E\varepsilon_n^2 = 1$; we calculate $E\left(H_k^2(X_j)\left(\alpha^2 X_j^2 + 2\alpha\varepsilon_j X_j + \varepsilon_j^2\right)\right)$:

$$\begin{aligned}E\left(H_k^2(X_j)\left(\alpha^2 X_j^2 + 2\alpha\varepsilon_j X_j + \varepsilon_j^2\right)\right) &= \alpha^2 E\left[X_j^2 H_k^2(X_j)\right] + 2\alpha E\varepsilon_j \times E\left(X_j H_k^2(X_j)\right) + E\varepsilon_j^2 \times E\left(H_k^2(X_j)\right) \\ &= \alpha^2 E\left(X_j^2 H_k^2(X_j)\right) + E\left(H_k^2(X_j)\right) \\ &= \alpha^2 E\left(X_j^2 H_k^2(X_j)\right) + E\left(H_k^2(X_j)\right)\end{aligned}$$

We have $E\left(H_k^2(X_j)\right) = k!$, we calculate $E\left[X_j^2 H_k^2(X_j)\right]$. According to the formula 4.23 in (Declercq (1998)), we get

$$\begin{aligned} E\left(X_j^2 H_k^2(X_j)\right) &= E\left(H_1^2(X_j) H_k^2(X_j)\right) \\ &= k! \sum_{l=0}^{k} \binom{k}{l}\binom{1}{l}\binom{2l}{l} \end{aligned}$$

By posing $s_k = \sum_{l=0}^{1} \binom{k}{l}\binom{1}{l}\binom{2l}{l} = 2k+1$. We deduce

$$\begin{aligned} E\left(H_k^2(X_j)\left(\alpha^2 X_j^2 + 2\alpha\varepsilon_j X_j + \varepsilon_j^2\right)\right) &= \alpha^2 E\left[X_j^2 H_k^2(X_j)\right] + E\left[H_k^2(X_j)\right] \\ &= \alpha^2(2k+1)k! + k! \end{aligned}$$

and

$$E\left(|\widehat{c}_k|^2\right) = \frac{k!}{n}(\alpha^2(2k+1)+1) \tag{19}$$

The second term (18): $\frac{1}{n^2}\sum_{j\neq j'}^{n} E\left(H_k(X_j)H_k(X_{j'})\left(\varepsilon_j\varepsilon_{j'} + \alpha X_j\varepsilon_j + \alpha X_{j'}\varepsilon_{j'} + \alpha^2 X_j X_{j'}\right)\right)$
We use the fact that X_j and $X_{j'}$ are independent for $j \neq j'$

$$\left(\begin{array}{c} H_k(X_j)H_k(X_{j'})\left(\varepsilon_j\varepsilon_{j'} + \alpha X_j\varepsilon_j + \alpha X_{j'}\varepsilon_{j'} + \alpha^2 X_j X_{j'}\right) = \varepsilon_j\varepsilon_{j'} H_k(X_j)H_k(X_{j'}) \\ +\alpha X_j H_k(X_j)H_k(X_{j'})\varepsilon_j + \alpha X_{j'} H_k(X_j)H_k(X_{j'})\varepsilon_{j'} + \alpha^2 X_j X_{j'} H_k(X_j)H_k(X_{j'}) \end{array}\right)$$

First we calculate the expectation of $\left(H_k(X_j)H_k(X_{j'})\left(\varepsilon_j\varepsilon_{j'} + \alpha X_j\varepsilon_j + \alpha X_{j'}\varepsilon_{j'} + \alpha^2 X_j X_{j'}\right)\right)$:

$I)$ $E\varepsilon_j \times E\varepsilon_{j'} \times E\left(H_k(X_j)H_k(X_{j'})\right) = 0$

$II)\alpha$ $E\left(X_j H_k(X_j)H_k(X_{j'})\right) \times E\varepsilon_j = 0$

Idem for the term III)

$IV)$ $\alpha^2 E\left(X_{j'} H_k(X_j)\right) E\left(X_j H_k(X_{j'})\right) = 0$ since $k \neq 1$

We conclude that

$$E\left(H_k(X_j)H_k(X_{j'})\left(\varepsilon_j\varepsilon_{j'} + \alpha X_j\varepsilon_j + \alpha X_{j'}\varepsilon_{j'} + \alpha^2 X_j X_{j'}\right)\right) = 0 \tag{20}$$

Combining the first term of (17) and the second term (17) we have

$$var(\widehat{c}_k) = \frac{k!}{n}((2k+1)\alpha^2 + 1) \tag{21}$$

Proof of Proposition 7

We suppose that $k \neq l$ and under H_0, we can write

$$\begin{aligned} cov(\widehat{c}_k, \widehat{c}_l) &= E\left(\frac{1}{n}\sum_{j=1}^{n} H_k(X_j)Y_j \times \frac{1}{n}\sum_{j'=1}^{n} H_l(X_{j'})Y_{j'}\right) \\ &= \frac{1}{n}\sum_{j=1}^{n} E\left(H_k(X_j)H_l(X_j)Y_j^2\right) + \frac{1}{n^2}\sum_{j\neq j'}^{n} E\left(H_k(X_j)Y_j H_l(X_{j'})Y_{j'}\right) \end{aligned} \tag{22}$$

a-First term of (22): $\frac{1}{n}\sum_{j=1}^{n} E\left(H_k(X_j)H_l(X_j)Y_j^2\right)$

$$\begin{aligned} E\left(H_k(X_j)H_l(X_j)Y_j^2\right) &= E\left(\left[H_k(X_j)H_l(X_j)\right]\left(\alpha^2X_j^2+2\alpha\varepsilon_jX_j+\varepsilon_j^2\right)\right) \\ &= E\left(\left[\alpha^2X_j^2H_k(X_j)H_l(X_j)+2\alpha\varepsilon_jX_jH_k(X_j)H_l(X_j)+\varepsilon_j^2H_k(X_j)H_l(X_j)\right]\right) \\ &= \alpha^2E\left(X_j^2H_k(X_j)H_l(X_j)\right)+E\varepsilon_j^2\times E\left[H_k(X_j)H_l(X_j)\right] \\ &= \alpha^2E\left(X_j^2H_k(X_j)H_l(X_j)\right)+E\left(H_k(X_j)H_l(X_j)\right) \\ &= \alpha^2E\left[X_j^2H_k(X_j)H_l(X_j)\right] \end{aligned}$$

We know that if $k\neq l$ then $E\left[H_k(X_j)H_l(X_j)\right]=0$. We have $X_j^2=H_2(X_j)+1$, from where

$$\begin{aligned} E\left(X_j^2H_k(X_j)H_l(X_j)\right) &= E\left(\left(H_2(X_j)+1\right)H_k(X_j)H_l(X_j)\right) \\ &= E\left(H_2(X_j)H_k(X_j)H_l(X_j)\right)+E\left(H_k(X_j)H_l(X_j)\right) \\ &= E\left(H_2(X_j)H_k(X_j)H_l(X_j)\right) \end{aligned}$$

We have to calculate $E\left(H_2(X_j)H_k(X_j)H_l(X_j)\right)$. We use the result given in (Declercq (1998)): for $m\leq n\leq p$ that is

$$E\left(H_m(X_j)H_n(X_j)H_p(X_j)\right)=\binom{m}{k}\binom{n}{k}k!p! \tag{23}$$

and $p\leq m+n$, $k=\frac{1}{2}(m+n-p)$.
We set $m=2$; $2\leq l\leq k\Rightarrow u=\frac{1}{2}(2+l-k)$

$$\begin{aligned} E\left(H_2(X_j)H_k(X_j)H_l(X_j)\right) &= \binom{2}{u}\binom{l}{u}u!k! \\ &= \frac{2}{(2-u)!u!}\times\frac{l!}{(l-u)!u!}u!k! \\ &= \frac{2k!l!}{(2-u)!u!(l-u)!} \end{aligned}$$

Suppose $k\leq 2+l\Rightarrow u=\frac{1}{2}(2+l-k)\geq 0$. Two cases arise : $u=0$ and $u=1$.

- $u=0\Rightarrow k-l=2\Rightarrow k=l+2$

$$\begin{aligned} E\left(H_2(X_j)H_k(X_j)H_l(X_j)\right) &= C_2^0C_k^00!l!=\frac{2k!l!}{(2)!0!(l)!}=\frac{k!l!}{l!} \\ &= k! \end{aligned}$$

From where

$$\alpha^2E\left(X_j^2H_k(X_j)H_l(X_j)\right)=\alpha^2k!\text{ if }k=l+2 \tag{24}$$

- $u=1\Rightarrow\frac{1}{2}(2+k-l)=1\Rightarrow k-l=0\Rightarrow k=l$ we supposed $k\neq l$ since if $k=l$ it is the variance that is already calculated above.
- For $k\leq l$ we complete, by symmetry, the matrix of variance.

b-Second term (22): $\frac{1}{n^2}\sum_{j\neq j'}^{n} EH_k(X_j)Y_jH_l(X_{j'})Y_{j'}$

$$\begin{aligned} E\left(H_k(X_j)Y_j \times H_l(X_{j'})Y_{j'}\right) &= EH_k(X_j)H_l(X_{j'})\left(\varepsilon_j^2 + X_j\alpha\varepsilon_j + X_{j'}\alpha\varepsilon_j + X_jX_{j'}\alpha^2\right) \\ &= E\varepsilon_j^2 EH_k(X_j)H_l(X_{j'}) + \alpha E\varepsilon_j \times E\left(X_jH_k(X_j)H_l(X_{j'})\right) \\ &\quad +\alpha E\varepsilon_j \times E\left(X_{j'}H_k(X_j)H_l(X_{j'})\right) + \alpha^2 E\left(X_jX_{j'}H_k(X_j)H_l(X_{j'})\right) \\ &= \alpha^2 EX_jX_{j'}H_k(X_j)H_l(X_{j'}) \\ &= \alpha^2 EX_jH_k(X_j)EX_{j'}H_l(X_{j'}) \\ &= 0 \end{aligned}$$

Proof of Theorem 5

We use the theorem:

Let $(X_i)_i$ be a sequence of independent random variables, of the same law and square integrable (and not constant). Note $\mu = EX_1$, $\sigma^2 := varX_1$ with $\sigma > 0$ $S_n := \sum_{j=1}^n X_j$. Then

$$\frac{S_n - ES_n}{\sqrt{varS_n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} \sim N(0,1). \tag{25}$$

First we pose $Z_j = H_k(X_j)X_j = \widetilde{H}_k(X_j)$ are independent because the X_j are independent. We have

$$\widehat{c}_{k1} = \frac{\alpha}{n}\sum_{j=1}^n Z_j \tag{26}$$

The expectation and the variance of $\widehat{c}_{k1}$ are calculated:

- $E\widehat{c}_{k1} = \alpha EZ_1 = 0$. Indeed

$$\begin{aligned} \mu &= EZ_1 = E\left(X_jH_k(X_j)\right) \\ &= E\left(H_1(X_j)H_k(X_j)\right) = 0 \end{aligned}$$

- $\alpha^2 varZ_1 = \alpha^2 k!$. Since $k \geq 2$

$$\alpha^2 VarZ_1 = \alpha^2 E\left[H_1^2(X_j)H_k^2(X_j)\right] = \alpha^2 k! \tag{27}$$

So we deduce that

$$\widehat{c}_{k1} \sim N\left(0, \frac{\alpha^2 k!}{n}\right) \tag{28}$$

at k fixed.

Proof of Proposition 6

We have

$$\widehat{c}_{k2} = \frac{1}{n}\sum_{j=1}^n H_k(X_j)\varepsilon_j \tag{29}$$

The sequence $\{X_n, n \in N\}$ converges in probability to X if $\forall\varepsilon > 0$, $\lim_{n\to+\infty} P(\mid X_n - X \mid > \varepsilon) = 0$. Let X be a random variable of expectation μ and of finite variance σ^2. The inequality of Bienayme-Chebyshev is expressed as follows: for every positive real positive α, $P(\mid X_n - \mu \mid \geq \delta) \leq \frac{\sigma^2}{\delta^2}$. If we calculate, at fixed k

$$\begin{aligned} E(H_k(X_j)\varepsilon_j) &= E(H_k(X_j))E(\varepsilon_j) \\ &= 0 \end{aligned}$$

since $E(\varepsilon_j) = 0$, and

$$\begin{aligned} var(H_k(X_j)\varepsilon_j) &= E(H_k(X_j)\varepsilon_j)^2 \\ &= E(H_k^2(X_j)E(\varepsilon_j)^2 = k! \end{aligned}$$

$\Rightarrow$

$$var\left(\frac{1}{n}\sum_{j=1}^{n} H_k(X_j)\varepsilon_j\right) = \frac{1}{n} \times k! \tag{30}$$

$$P\left(\left|\frac{1}{n}\sum_{j=1}^{n} H_k(X_n)\varepsilon_n\right| \geq \delta\right) \leq \frac{k!}{n\delta^2} \to 0 \tag{31}$$

when $n \to \infty$ at fixed k.

$$\lim_{n\to+\infty} P\left(\left|\frac{1}{n}\sum_{j=1}^{n} H_k(X_n)\varepsilon_n\right| \geq \delta\right) = 0 \tag{32}$$

We conclude that $\frac{1}{n}\sum_{j=1}^{n} H_k(X_j)\varepsilon_j$ converges in probability to 0 when $n \to \infty$.

Appendix 2: The Hermite polynomials

Hermite polynomials are defined by : $H_0(x) = 1$ and according to Rodrigues' formula

$$H_n(x) = (-1)^n e^{\frac{x^2}{2}} \frac{d^n e^{-\frac{x^2}{2}}}{dx^n}$$

for $n = 1, 2, ...$
The equation $H_n(x) = 0$ has all its roots real. We have

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} H_n(x)dx = 0 \ \forall n.$$

except for $n = 0$,

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} H_n(x)H_m(x)dx = 0$$

for $n \neq m$, and

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} H_n^2(x)dx = n!$$

The system $\left\{\frac{e^{-\frac{x^2}{2}} H_k(x)}{\sqrt{2^n n!\sqrt{\pi}}}\right\}$ is orthonormed in $(-\infty, \infty)$, it is complete with respect to the functions square integrable i.e

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} H_n(x)H_m(x)dx = n!\delta_{n-m}. \tag{33}$$

A sequence of orthogonal polynomials in the space $L^2(\mathbb{R}, d\mu)$, where the measure μ is given by

$$d\mu(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Hermite polynomials also verify the following recurrence relation: $H_{n+1}(x) = xH_n(x) - nH_n^{'}(x)$ and $H_n^{'}(x) = nH_{n-1}(x)$. The only sequence of polynomials that satisfies the two equations is the sequence of Hermite polynomials. We thus find the first of these polynomials

$$H_0 = 1,\ H_1(x) = x,\ H_2(x) = x^2 - 1,\ H_3(x) = x^3 - 3x,\ ...$$

References

Azzalini A, Bowman A (1993). "On the Use of Nonparametric Regression for Checking Linear Relationships." *Journal of the Royal Statistical Society: Series B (Methodological)*, **55**(2), 549–557. doi:10.1111/j.2517-6161.1993.tb01923.

Bierens HJ (1982). "Consistent Model Specification Tests." *Journal of Econometrics*, **20**(1), 105–134. doi:10.1016/0304-4076(82)90105-1.

Bontemps C, Meddahi N (2005). "Testing Normality: A GMM Approach." *Journal of Econometrics*, **124**(1), 149–186. doi:10.1016/j.jeconom.2004.02.014.

Declercq D (1998). *Apport des Polynomes d'Hermite a la Modelisation Non Gaussienne et Tests Statistiques Associes*. Ph.D. thesis, Cergy-Pontoise.

Djeddour KH, Mokkadem A, Pelletier M (2007). "Test for Uniformity by Empirical Fourier Expansion." *Mathematical Methods of Statistics*, **16**(2), 124–141. doi:10.3103/S1066530707020044.

Eubank RL, Spiegelman CH (1990). "Testing the Goodness of Fit of a Linear Model via Nonparametric Regression Techniques." *Journal of the American Statistical Association*, **85**(410), 387–392. doi:10.2307/2289774.

Gallant AR (1975). "The Power of the Likelihood Ratio Test of Location in Nonlinear Regression Models." *Journal of the American Statistical Association*, **70**(349), 198–203. doi:10.2307/2285403.

González-Manteiga W, Crujeiras RM (2013). "An Updated Review of Goodness-of-fit Tests for Regression Models." *Test*, **22**(3), 361–411. doi:10.1007/s11749-013-0327-5.

Hardle W, Mammen E, *et al.* (1993). "Comparing Nonparametric versus Parametric Regression Fits." *The Annals of Statistics*, **21**(4), 1926–1947. doi:10.1214/aos/1176349403.

Leonis G (1980). "Hermite Polynomials and One-dimensional Restoration in Radio Astronomy." *Astronomy and Astrophysics*, **85**, 168–173.

Mohdeb Z, Mokkadem A (1998). "Tests d'Hypothèses sur les Coefficients de Fourier d'Une Fonction de Régression Non Linéaire." *Comptes Rendus de l'Académie des Sciences-Series I-Mathematics*, **326**(9), 1141–1144. doi:10.1016/S0764-4442(98)80077-3.

Öztürk Y, Gülsu M (2014). "An Approximation Algorithm for the Solution of the Lane–Emden Type Equations Arising in Astrophysics and Engineering Using Hermite Polynomials." *Computational and Applied Mathematics*, **33**(1), 131–145.

Pearson K (1905). *On the General Theory of Skew Correlation and Non-linear Regression*. 14. Dulau and Company.

Queffélec H, Zuily C (2007). *Analyse pour l'Agrégation: Cours et Exercices Corrigés*. Dunod.

Stute W, Manteiga WG (1996). "NN Goodness-of-fit Tests for Linear Models." *Journal of Statistical Planning and Inference*, **53**(1), 75–92. doi:10.1016/0378-3758(95)00144-1.

Weihrather G (1993). "Testing a Linear Regression Model against Nonparametric Alternatives." *Metrika*, **40**(1), 367–379. doi:10.1007/BF02613703.

Zheng JX (1996). "A Consistent Test of Functional Form via Nonparametric Estimation Techniques." *Journal of Econometrics*, **75**(2), 263–289.

Affiliation:

Djaballah-Djeddour Khadidja
University of science and technology Houari Boumediene
MSTD Laboratory of Mathematics
BP 32, El-Alia, 16111, Algeria.
E-mail: khajaballah@usthb.dz

Tazerouti Moussa
University of science and technology Houari Boumediene
Faculty of Mathematics
BP 32, El-Alia, 16111, Algeria.
And
University M'hamed Bougara of Boumerdes
Departement of Mathematics
Avenue de l'Indé}pendance 35000 Boumerdes
E-mail: tazerouti.moussa1@gmail.com

Austrian Journal of Statistics
published by the Austrian Society of Statistics
Volume 51
January 2022

http://www.ajs.or.at/
http://www.osg.or.at/

Submitted: 2019-10-21
Accepted: 2021-03-08