Methodol Comput Appl Probab
DOI 10.1007/s11009-008-9103-5

Goodness-of-Fit Tests for Continuous Regression

Alejandra Cabaña · Enrique M. Cabaña

Received: 1 March 2007 / Revised: 30 August 2008 /
Accepted: 5 September 2008

Abstract A family of transformations of the processes of accumulated residues of linear models is used to construct tests of fit of the models, consistent for any alternative, and focused on alternatives in the direction selected by the user. The resulting tests are asymptotically distribution free, both under the null hypothesis of fit, and under the selected alternatives. An interesting feature is that this distributions do not depend on (possible) parameter estimations.

Keywords Goodness-of-fit for linear models · Polynomial regression · Transformed accumulated residues processes

AMS 2000 Subject Classifications 62G10 · 62J02

1 Introduction

The transformation of experimental data in order to make noticeable certain deviations from pre-established models has been in use for at least a decade in the construction of goodness-of-fit tests. The theoretical ideas behind the transformations

A. Cabaña partially supported by Consejería de Educación y Cultura, Junta de Castilla y León, Grant PAPIJCL VA102A06 and by Grant TIN2005-08832-C03-02, MOISES-TA, Ministerio de Educación y Ciencia, Spain.

E. M. Cabaña partially supported by a grant from Programa de Desarrollo Tecnológico, Ministerio de Educación y Cultura, Uruguay.

A. Cabaña (✉)
Universidad de Valladolid, Valladolid, Spain
e-mail: acnigro@eio.uva.es

E. M. Cabaña
Universidad de la República, Montevideo, Uruguay

can be found in Cabaña and Cabaña (1997), and some examples of their use in Scavino (1999), Graneri (2003), Cabaña and Cabaña (2003, 2005).

The procedure consists in constructing a random function containing the same information as the data set and then transforming it in order to make visible some of the characteristics of the data which are of interest for the decisions under consideration.

Moreover, the asymptotic distribution of such transformed processes is normal. This has technical consequences, for instance, the limiting distributions of the functionals of these processes used as test statistics are easy to derive, and the use of Le Cam's Third Lemma to give a precise description of the behaviour of the resulting tests under *contiguous* models, is quite simple.

The use of Transformed Empirical Processes allowed the design of focused, consistent and asymptotically distribution free goodness-of-fit tests for i.i.d. samples. The present work is a first attempt on the application of the same sort of transformations to the process of *accumulated residues* in Linear Models, in order to check the validity of such a model.

The problem of assessing the validity of a linear model is quite old, starting from Nadaraya (1964) and Watson (1964). Most of the existing work deals with density estimation, and so, most of the results are not consistent against alternatives which approach the null hypothesis at rate $\sqrt{n}$. A fairly complete survey can be found in Stute (1997) where a statistic based on marked residuals is proposed, which is consistent against any alternative, although not focused. Stute's methods rely on certain martingale transforms introduced in Khmaladze (1981). The many applications in inference of Khmaladze transform rely on its isometric properties, but the use of a particular class of isometries (the L-isometries as named in Cabaña and Cabaña (1997) Section 6.1 and Section 7.1) endows the transformations with martingale properties that provide a powerful analytical way of proving the convergence of estimators.

The statistics based on transforming the accumulated residues we propose here (as well as the ones based in transforming the empirical process in previous papers on the subject) are consistent, and their asymptotic behaviour is always the same: it does not depend on the true (unknown) parameters of the model, and even whether parameter estimation is required or not, our transformations provide statistics with the same asymptotic laws. This last achievement imposes the use of particular isometries adapted to the estimation procedures, and consequently the transformations we apply not necessarily lead to martingales, so the asymptotic behaviour of our statistics has to be established under different (in fact weaker) assumptions than Khmaladze's (this is discussed in Section 3.1.6).

The general framework for random design models is described in Section 2. As a first introductory example of the use of the transformed residues process to focus the power on the selected alternatives, we describe in Section 3.1 how to test the simple hypothesis of fit to a completely determined model, in spite of its reduced practical interest.

In Section 3.2, we study the general problem of fit to a linear model, in Section 3.3 we describe the asymptotic properties of the test statistics, and Section 4 is devoted to their computation.

The particular example of polynomial regression is treated in Section 6, and the power of the proposed tests in front of certain alternatives is computed. An empirical description of powers illustrates the behaviour of our tests for finite samples. In particular we compare in Section 6.4 the power of our tests to that of Stute et al. (1998) through an example introduced in their paper.

Finally, Section 7 contains the proofs of the asymptotic results described in Section 3, and an appendix provides some technical properties about Legendre Polynomials expansions.

2 Statement of the Problem

Let us assume that a number of independent experiments can be performed at values $t_1, \ldots, t_n$ to be chosen of a parameter t in a given interval $[a, b]$. For each t_i, a random variable

$$Y_i = x(t_i) + \sigma Z_i \tag{1}$$

is obtained.

Each Y_i can be thought as an observation of an unknown *continuous* function $x(t)$ at $t = t_i$, with an additive random error σZ_i. The variables Z_i are assumed to be i.i.d. standard normal.

Let us introduce a parametric family

$$\mathcal{X} = \left\{ \sum_{i=0}^{p-1} \beta_i x_i : \boldsymbol{\beta} = (\beta_0, \ldots, \beta_{p-1})^{\text{tr}} \in \mathbf{R}^p \right\} \tag{2}$$

of linear combinations of p continuous functions $x_0, \ldots, x_{p-1}$. A typical example of such family are the polynomials of degree less or equal than $p-1$.

Our aim is to provide a test of the null hypothesis

$$\mathcal{H}_0 : x \in \mathcal{X}, \tag{3}$$

consistent for any alternative as the number n of observations goes to infinity and the set of points of observation become progressively dense (infill asymptotics).

It is well known that even for *omnibus* tests consistent under any fixed alternative, one cannot expect to have high power except for a finite number of alternatives (see Janssen 2000). Hence, it is important to have tests which are sensitive to *disturbing* alternatives.

Consequently, we shall design the test to be specially sensitive to the specific alternatives

$$x = x' + \beta^* x^*, x' \in \mathcal{X}, \beta^* \in \mathbf{R},$$

for $x^* \notin \mathcal{X}$ arbitrarily chosen by the user.

We shall assume that the functions x, x_j $(j = 0, 1, \ldots, p-1)$, x^*, are continuous in $[a, b]$ and the points of observation are an ordered sample $T_{n,1} < T_{n,2} < \cdots < T_{n,i} < \cdots < T_{n,n}$ of a continuous distribution function G in $[a, b]$.

With no loss of generality we assume in addition that the interval $[a, b]$ is $[0, 1]$, G is the uniform distribution in $[0, 1]$, and the functions $x_0, x_1, \ldots, x_{p-1}, x_p := x^*$ are an orthonormal set of functions in $L^2([0, 1])$ with the uniform measure (that is, $\int_0^1 x_j(t) x_k(t) dt = \delta_{j,k}$, where $\delta_{j,k} = \mathbf{1}_{\{j=k\}}$ denotes Kronecker delta).

If the observations are placed at random with other known continuous probability distribution G instead, we apply the *probability integral transformation* to write x in Eq. 1 as $x(t) = y(G(t))$ and let y play the role of x.

We show in Section 5 that for the *fixed design model* with equally spaced observation points the tests designed for uniformly distributed observations can be applied. Likewise, a similar reduction can be made for fixed but not equally spaced observations.

Then, under $\mathcal{H}_0$, the observed variables $\boldsymbol{Y}_n = (Y_{n,1}, Y_{n,2}, \ldots, Y_{n,n})^{\mathrm{tr}}$ follow the linear model

$$\boldsymbol{Y}_n = \boldsymbol{X}_n \boldsymbol{\beta} + \sigma \boldsymbol{Z}_n, \;\; \boldsymbol{\beta} \in \mathbf{R}^p, \sigma \in \mathbf{R}^+ \tag{4}$$

where $\boldsymbol{X}_n$ is the matrix with rows $(\boldsymbol{X}_n)_{i,\cdot} = \boldsymbol{x}(T_{n,i})$, $\boldsymbol{x} = (x_0, x_1 \ldots, x_{p-1})$ and $\boldsymbol{Z}_n$ is a random vector of i.i.d. standard normal components.

Weak limits can be established by means of strong convergences, by assuming that r is a given standard Wiener process on $[0, 1]$, and writing the components $Z_{n,1}, \ldots, Z_{n,p}$ of $\boldsymbol{Z}_n$ as

$$Z_{n,i} = \sqrt{n}\left(r\left(\frac{i}{n}\right) - r\left(\frac{i-1}{n}\right)\right). \tag{5}$$

This will render the convergence arguments easier.

3 The Goodness-of-Fit Test

A frequently used procedure for testing composite hypotheses for models depending on unknown parameters is to replace estimators instead of fixed parameter values in the related simple hypothesis test statistics. The Lilliefors tests for normality or exponentiality are typical examples of this.

The test we propose in this article follows this general rule:

- we solve the problem for known parameters, which is equivalent to assuming $p = 0$ and known variance, by introducing the accumulated residues process r_n and a suitably chosen transformation w_n of r_n (see Section 3.1), and
- we estimate the parameters, compute the estimated residues, and repeat the procedure used to test the fit to the model with fixed parameters letting the estimated residues play the role of the ordinary ones in Section 3.1.

Such procedures usually lead to suitable results, but there is some cost associated: the replacement of the true parameters by the estimated ones changes the actual and also the asymptotic distributions of the test statistics.

However, by a suitable selection of the transformation to be used in the definition of the *transformed accumulated residues process*, we get the same asymptotic distribution of the test statistics regardless the parameters are estimated or not. This is one of the main advantages of our proposal.

In addition, the test is designed for best discrimination of a particular sequence of alternatives, and the asymptotic distribution of the test statistic under these alternatives is also distribution free, depending only on the size of the alternatives as defined after Eq. 18.

3.1 The Model with Fixed Parameters

Testing $Y_{n,i} = x(T_{n,i}) + \sigma_f Z_{n,i}$ with known regression and variance is equivalent to testing Eq. 3 for model Eq. 1 with $p = 0$ in Eq. 2, and known variance $\sigma^2 = \sigma_f^2$, since $Y_{n,i}$ can be replaced by $Y_{n,i} - x(T_{n,i})$.

3.1.1 The Process of Accumulated Residues

In order to test the null hypothesis $\mathcal{H}_0$: "$Y_{n,i} = \sigma_f Z_{n,i}$" by means of the observations $Y_{n,i}$ corresponding to $T_{n,i}$, $i = 1, 2, \ldots, n$, let us compute the residues $e_{n,i} = \frac{Y_{n,i}}{\sigma_f} = Z_{n,i}$ and introduce the *normalized accumulated residues process*

$$r_n(t) = \frac{1}{\sqrt{n}} \sum_{Tn,i \leq t} e_{n,i}, \quad t \in [0, 1]. \tag{6}$$

Let F_n denote the empirical distribution function of the observation points $T_{n,i}$. Our assumption Eq. 5 implies $r_n(t) = r(F_n(t))$ and hence r_n converges uniformly to r a.s. because of the a.s. uniform continuity of this last process and Glivenko-Cantelli Law.

We then reject $\mathcal{H}_0$ when the behaviour of r_n differs significantly of that of a Wiener process.

3.1.2 The Transformed Accumulated Residues Process

Let us introduce a new process $w_n(t)$, $t \in [0, 1]$, that we call *Transformed Accumulated Residues Process* (TARP), associated to a normalized score function $a \in L^2([0, 1])$, $\int_0^1 a^2(t)dt = 1$ and to an isometry $\mathcal{T}$ of $L^2([0, 1])$ as follows:

$$w_n(t) = \int_0^1 \mathcal{T}(a\mathbf{1}_t)(s)dr_n(s) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathcal{T}(a\mathbf{1}_t)(T_{n,i})e_{n,i}. \tag{7}$$

A reader familiar with Transformed Empirical Processes (TEPs) requires no justification for the preceding definition, because of its close relationship with the definition of a TEP. Further justifications rely on the properties and applications of the TARPs, described in next sections.

The score function a will be chosen to focus the inference on the interesting alternatives. The isometry appears to be crucial in further applications, but plays no role in this simple case, and could be replaced by the identity map or any other surjective isometry.

In order to describe the asymptotic behaviour of w_n under $\mathcal{H}_0$, we may write it as $w_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathcal{T}(a\mathbf{1}_t)(U_{n,i})Z_{n,j(i)}$, where $\{U_{n,i}, i = 1, 2, \ldots, n\}$ are the i.i.d. uniform variables with order statistics $(T_{n,i} : i = 1, 2, \ldots, n)$, and $T_{n,j(i)} = U_{n,i}$. The random functions $\mathcal{T}(a\mathbf{1}_t)(U_{n,i})Z_{n,j(i)}$ are i.i.d. with expectation zero and covariances

$$\mathbf{E}\mathcal{T}(a\mathbf{1}_s)(U_{n,i})\mathcal{T}(a\mathbf{1}_t)(U_{n,i})e^2_{n,j(i)}$$
$$= \int_0^1 \mathcal{T}(a\mathbf{1}_s)(u)\mathcal{T}(a\mathbf{1}_t)(u)du = \int_0^{s\wedge t} a^2(u)du.$$

Let V denote the function $V(t) = \int_0^t a^2(u)du$ $(0 \leq t \leq 1)$. By the Central Limit Theorem, the vectors $(w_n(s_1), \ldots, w_n(s_k))$ of evaluations of w_n on an arbitrary finite set of points in $[0, 1]$ converge in law to the evaluations at the same points of a

V-Wiener process, that is, a centred Gaussian process with independent increments and variance function V. We shall say in short that w_n converges fi.di. to a V-Wiener process.

Let us notice that the process $w^{(V)}$ defined by means of the Wiener integral

$$w^{(V)}(t) = \int_0^1 \mathcal{T}(a\mathbf{1}_t)dr \tag{8}$$

is precisely a V-Wiener process. Therefore, the preceding observations can be summarized by expressing that w_n converges fi.di. to $w^{(V)}$.

3.1.3 The Behaviour of r_n and w_n Under Fixed Alternatives

Given a continuous function x^* with $L^2[0,1]$-norm equal one, let us denote $\boldsymbol{X}_n^* = (x^*(T_{n,1}), \dots, x^*(T_{n,n}))^{\text{tr}}$ and assume that the alternative $\mathcal{H}^*$: "$\boldsymbol{Y}_n = \beta^*\sigma_f \boldsymbol{X}_n^* + \sigma_f \boldsymbol{Z}_n$" holds.

The residues are $e_{n,i} = Z_{n,i} + \beta^* x^*(T_{n,i})$ and therefore $r_n(t)$ is the sum of $\frac{1}{\sqrt{n}}\sum_{T_{n,i}\le t} Z_{n,i}$, that converges in law to a Wiener process and is therefore stochastically bounded, plus $\sqrt{n}\beta^*$ times the average $\frac{1}{n}\sum_{i=1}^n \mathbf{1}_t(U_{n,i})x^*(U_{n,i})$ that converges a.s. to $\int \mathbf{1}_t(u)x^*(u)du = \int_0^t x^*(u)du$ by the Law of Large Numbers. Since the integral $\int_0^t x^*(u)du$ does not vanish for all t, this shows that r_n is able to distinguish consistently $\mathcal{H}_0$ from $\mathcal{H}^*$, because under $\mathcal{H}_0$, the process r_n is stochastically bounded, and under $\mathcal{H}^*$, $\lim_{n\to\infty}\sup_{0\le t\le 1}|r_n(t)| = \infty$.

As for the TARP, substituting the actual expression for the residues in Eq. 7 we get

$$w_n(t) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathcal{T}(a\mathbf{1}_t)(T_{n,i})Z_i + \sqrt{n}\beta^*\left[\frac{1}{n}\sum_{i=1}^n \mathcal{T}(a\mathbf{1}_t)(U_{n,i})x^*(U_{n,i})\right].$$

We have already established that the first term converges fi.di. to $w^{(V)}$ when n tends to infinity. The Law of Large Numbers implies that the bracket has the limit $\int_0^1 \mathcal{T}(a\mathbf{1}_t)(u)x^*(u)du = \int_0^t a(u)\mathcal{T}^{-1}x^*_{\mathcal{T}}(u)du$, where $x^*_{\mathcal{T}}$ is the projection of x^* on the range of $\mathcal{T}$.

Consequently, w_n also tends to infinity provided x^* is not orthogonal to the range of $\mathcal{T}$ and $a \neq 0$ on a subset of $[0,1]$ of Lebesgue measure 1.

3.1.4 Behaviour of the TARP Under Contiguous Alternatives

Let us assume that for each n the alternative

$$\mathcal{H}_n\text{: "}\boldsymbol{Y}_n = \frac{\delta\sigma_f}{\sqrt{n}}\boldsymbol{X}_n^* + \sigma_f \boldsymbol{Z}_n\text{"} \tag{9}$$

holds.

The arguments used in the previous section show that $r_n(t)$ tends to $r(t)$ $+\delta\int_0^t x^*(u)du$, and $w_n(t)$ converges fi.di. to $w^{(V)}(t) + \delta\int_0^t a(u)\mathcal{T}^{-1}x^*_{\mathcal{T}}(u)du$.

The optimum value of a in order to maximize the ratio between the drift $\delta\int_0^t a(u)\mathcal{T}^{-1}x^*_{\mathcal{T}}(u)du = \delta\int_0^1 \mathbf{1}_t(u)\frac{\mathcal{T}^{-1}x^*_{\mathcal{T}}(u)}{a(u)}dV(u)$ and the standard deviation $\sqrt{V(t)}$ of the random term - and hence make more noticeable the difference between the behaviour of w_n under $\mathcal{H}_0$ and under $\mathcal{H}_n$ - is $\hat{a} = \mathcal{T}^{-1}x^*_{\mathcal{T}}/\|x^*_{\mathcal{T}}\|$.

This is verified by using the Cauchy-Schwarz inequality to establish the bound

$$\left(\int_0^1 \mathbf{1}_t(u)\frac{\mathcal{T}^{-1}x_{\mathcal{T}}^*(u)}{a(u)}dV(u)\right)^2\frac{1}{V(t)} \leq \int_0^1\left(\frac{\mathcal{T}^{-1}x_{\mathcal{T}}^*(u)}{a(u)}\right)^2 dV(u)$$
$$= \int_0^1 (\mathcal{T}^{-1}x_{\mathcal{T}}^*(u))^2 du = \|x_{\mathcal{T}}^*\|^2$$

and verifying that the bound is reached for $a = \hat{a}$ and $t = 1$.

3.1.5 Performing the Test: The Rejection Region

We propose the use of a Cramér–von Mises test statistic, in order to decide if the TARP w_n behaves significantly different from a Wiener process: The null hypothesis is rejected for large values of the quadratic statistic of Watson type (see Watson 1961)

$$Q_n = \int_0^1\int_0^1\left(\int_0^1 c_{t_1,t_2}(s)dw_n(s)\right)^2 dV(t_1)dV(t_2)$$

constructed from the indicator function

$$c_{t_1,t_2}(s) = \mathbf{1}_{\{t_1<s<t_2\}} + \mathbf{1}_{\{t_2<t_1<s\}} + \mathbf{1}_{\{s<t_2<t_1\}}.$$

This notation has been introduced in Cabaña and Cabaña (2001).

Interchanging the integrals in the definition of Q_n, it reduces to the simpler form

$$Q_n = \int_0^1\int_0^1 \gamma(V(s), V(t))dw_n(s)dw_n(t), \tag{10}$$

where $\gamma(u, v) = (|u-v| - \frac{1}{2})^2 + \frac{1}{4}$, as shown in the above mentioned article.

The computation of our quadratic statistics is discussed in next section, related to the case of practical interest there considered.

3.1.6 On the Asymptotic Distribution of the Test Statistic Q_n

The asymptotic behaviour of Q_n is described by the following theorem. Its proof is deferred until Section 7, as well as some brief comments on Theorems 2 and 3.

Theorem 1

(i) *When $\mathcal{H}_0$ holds, the statistic Q_n defined in* Eq. 10 *converges in law to*

$$Q = \int_0^1\int_0^1 \gamma(V(s), V(t))dw^{(V)}(s)dw^{(V)}(t).$$

(ii) *When the alternatives* Eq. 9 *hold with x^* orthogonal to the range of $\mathcal{T}$ and $\|x^*\| = 1$, the statistic Q_n defined in* Eq. 10 *converges in law to*

$$Q + 2\delta\int_0^1\int_0^1 \gamma(V(s), V(t))dw^{(V)}(s)a(t)\mathcal{T}^{-1}x^*(t)dt$$
$$+ \delta^2\int_0^1\int_0^1 \gamma(V(s), V(t))a(s)\mathcal{T}^{-1}x^*(s)a(t)\mathcal{T}^{-1}x^*(t)dsdt. \tag{11}$$

(iii) *Let w be a standard Wiener process in $[0,1]$ such that $w^{(V)}(t) = w(V(t))$, and let $c_0 = 1$, and, for each $\nu = 1, 2, \ldots$, $c_\nu(u) = \sqrt{2}\cos 2\pi\nu u$ and $s_\nu(u) = \sqrt{2}\sin 2\pi\nu u$.*
Then the random variables

$$C_\nu = \int_0^1 c_\nu(t)dw(t),\ \nu = 0, 1, 2, \ldots$$

$$S_\nu = \int_0^1 s_\nu(t)dw(t),\ \nu = 1, 2, \ldots,$$

are i.i.d. standard normal, and

$$Q = \int_0^1 \int_0^1 \gamma(u,u)dw(u)dw(v) = \frac{C_0^2}{3} + \frac{1}{2}\sum_{\nu=1}^{\infty} \frac{C_\nu^2 + S_\nu^2}{\pi^2\nu^2}. \tag{12}$$

(iv) *If the test is focused on the alternatives Eq. 9, and hence $a = \mathcal{T}^{-1}x^*$, then Eq. 11 reduces to*

$$Q + \frac{2}{3}\delta C_0 + \frac{\delta^2}{3} = \frac{(C_0+\delta)^2}{3} + \frac{1}{2}\sum_{\nu=1}^{\infty} \frac{C_\nu^2 + S_\nu^2}{\pi^2\nu^2}.$$

3.2 Goodness of Fit to the Model with Unknown Parameters

3.2.1 The Accumulated Estimated Residues Process and its Limiting Law Under $\mathcal{H}_0$

Consider the problem of testing $\mathcal{H}_0$:"the model $\boldsymbol{Y}_n = \boldsymbol{X}_n\boldsymbol{\beta} + \sigma\boldsymbol{Z}_n$ for suitable $\boldsymbol{\beta} \in \mathbf{R}^p$ and $\sigma \in \mathbf{R}^+$ fits the data".

We estimate $\boldsymbol{\beta}$ as usual by

$$\hat{\boldsymbol{\beta}}_n = (\boldsymbol{X}_n^{\mathrm{tr}}\boldsymbol{X}_n)^{-1}\boldsymbol{X}_n^{\mathrm{tr}}\boldsymbol{Y}_n,$$

and σ^2 by the sum of squares of successive differences

$$\hat{\sigma}_n^2 = \frac{1}{2n-2}\sum_{i=2}^{n}(Y_{n,i} - Y_{n,i-1})^2 \tag{13}$$

because this estimator is less affected by departures from the regression specified by $\mathcal{H}_0$ than the widely used $\frac{1}{n}\|\boldsymbol{Y}_n - \boldsymbol{X}_n\hat{\boldsymbol{\beta}}_n\|^2$. In fact, as $n \to \infty$, $\max\{|T_{n,i} - T_{n,i-1}| : i = 2, 3, \ldots, n\} \to 0$ a.s. and this implies $\hat{\sigma}_n^2 \to \sigma^2$ a.s., provided Eq. 1 holds with a continuous regression function x, no matter if x suits the null hypothesis or not.

Then compute the vector $\hat{\boldsymbol{e}}_n = (\hat{e}_{n,1}, \ldots, \hat{e}_{n,n})^{\mathrm{tr}}$ of estimated residues

$$\hat{\boldsymbol{e}}_n = \frac{\boldsymbol{Y}_n - \boldsymbol{X}_n\hat{\boldsymbol{\beta}}_n}{\hat{\sigma}_n} = \frac{H_n\boldsymbol{Y}_n}{\hat{\sigma}_n},\quad H_n = I - \boldsymbol{X}_n(\boldsymbol{X}_n^{\mathrm{tr}}\boldsymbol{X}_n)^{-1}\boldsymbol{X}_n^{\mathrm{tr}}. \tag{14}$$

Under $\mathcal{H}_0$, Eq. 14 reduces to $\frac{H_n\boldsymbol{Z}_n}{\hat{\sigma}_n/\sigma}$, asymptotically equivalent to $H_n\boldsymbol{Z}_n$, which does not depend on the unknown parameters $\boldsymbol{\beta}$ and σ.

The *accumulated estimated residues process* is defined by replacing $e_{n,i}$ by $\hat{e}_{n,i}$ in Eq. 6:

$$\hat{r}_n(t) = \frac{1}{\sqrt{n}}\sum_{T_{n,i}\le t}\hat{e}_{n,i},\quad 0 \le t \le 1.$$

When n tends to infinity,

(i) $\frac{1}{\sqrt{n}}\sum_{T_{n,i}\le t} Z_{n,i} = r(F_n(t)) \to r(t)$, as noticed in Section 3.1,
(ii) $\frac{1}{\sqrt{n}}\boldsymbol{X}_n^{\mathrm{tr}}\boldsymbol{Z}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n(\boldsymbol{X}_n)_{i,\cdot}Z_{n,i} = \sum_{i=1}^n \boldsymbol{x}^{\mathrm{tr}}(T_{n,i})(r(\frac{i}{n}) - r(\frac{i-1}{n}))$ converges in probability to $\int_0^1 \boldsymbol{x}^{\mathrm{tr}}(s)dr(s)$,
(iii) $\frac{1}{n}\boldsymbol{X}_n^{\mathrm{tr}}\boldsymbol{X}_n \to \int_0^1 \boldsymbol{x}^{\mathrm{tr}}(t)\boldsymbol{x}(t)dt$ a.s., by the Law of Large Numbers; the limit equals the identity matrix I, because the functions x_i are chosen to be orthonormal, and
(iv) $\frac{1}{n}\sum_{T_{n,i}\le t}(\boldsymbol{X}_n)_{i,\cdot} = \frac{1}{n}\sum_{U_{n,i}\le t}\boldsymbol{x}(U_{n,i}) \to \int_0^t \boldsymbol{x}(s)ds$ a.s. also by the Law of Large Numbers.

Therefore, when $\mathcal{H}_0$ holds,

$$\hat{r}_n(t) = \frac{1}{\sqrt{n}}\sum_{T_{n,i}\le t}(H_n\boldsymbol{Z}_n)_i = \frac{1}{\sqrt{n}}\sum_{T_{n,i}\le t} Z_{n,i} - \frac{1}{n}\sum_{T_{n,i}\le t}(\boldsymbol{X}_n)_{i,\cdot}\left(\frac{1}{n}\boldsymbol{X}_n^{\mathrm{tr}}\boldsymbol{X}_n\right)^{-1}\frac{1}{\sqrt{n}}\boldsymbol{X}_n^{\mathrm{tr}}\boldsymbol{Z}_n$$

behaves asymptotically as

$$\rho(t) = r(t) - \int_0^t \boldsymbol{x}(s)ds\int_0^1 \boldsymbol{x}^{\mathrm{tr}}(s)dr(s) = r(t) - \sum_{j=0}^{p-1}\int_0^t x_j(s)ds\int_0^1 x_j(s)dr(s). \tag{15}$$

On the other hand, $\hat{\sigma}_n/\sigma \to 1$ a.s., so that the accumulated estimated residues process has the asymptotic law of Eq. 15.

3.2.2 Transforming the Accumulated Estimated Residues Process

The *Transformed Accumulated Estimated Residues Process (TAERP)* is defined for a given isometry $\mathcal{T}$ and a score function a such that $\int_0^1 a^2(s)ds = 1$, as

$$\hat{w}_n(t) = \int_0^1 \mathcal{T}(a\mathbf{1}_t)(s)d\hat{r}_n(s) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathcal{T}(a\mathbf{1}_t)(T_{n,i})\hat{e}_{n,i}. \tag{16}$$

From Eq. 15 follows that the finite dimensional distributions of the TAERP are asymptotically the ones of

$$\int_0^1 \mathcal{T}(a\mathbf{1}_t)(s)d\rho(s) = w^{(V)}(t) - \sum_{j=0}^{p-1}\int_0^1 \mathcal{T}(a\mathbf{1}_t)(s)x_j(s)ds\int_0^1 x_j(s)dr(s)$$

when the null hypothesis $\mathcal{H}_0$ holds.

In this case, as advanced at the beginning of this section, an adequate selection of the isometry becomes important:

Theorem 2 *When the range of $\mathcal{T}$ is orthogonal to $x_0, x_1, \ldots, x_{p-1}$, the asymptotic finite dimensional distributions of w_n are those of $w^{(V)}$, defined by Eq. 8, that is, the same limiting distributions obtained when parameters are known, instead of been estimated.*

Remark 1 We shall therefore choose $\mathcal{T}$ with range equal to the orthogonal complement of $x_0, x_1, \ldots, x_{p-1}$.

3.2.3 Consistency of the Test Under Fixed Alternatives

Suppose now that the observations follow the alternative model

$$\boldsymbol{Y}_n = \boldsymbol{X}_n\boldsymbol{\beta} + X'_n\beta_p + \sigma\boldsymbol{Z}_n, \quad \beta_p \neq 0, \tag{17}$$

for $X'_n = (x_p(T_{n,1}), x_p(T_{n,2}), \ldots, x_p(T_{n,n}))^{\text{tr}}$ where x_p is assumed with no loss of generality orthogonal to $x_0, x_1, \ldots, x_{p-1}$, and with norm one.

Under this model, $\hat{\boldsymbol{e}}_n = \frac{H_n\boldsymbol{Y}_n}{\hat{\sigma}_n/\sigma} = \frac{X'_n\beta_p/\sigma + H_n\boldsymbol{Z}_n}{\hat{\sigma}_n/\sigma}$ so that $\hat{r}_n(t) \to \infty$ as $n \to \infty$ for each t such that $\int_0^t x_p(s)ds \neq 0$, because when this holds $\frac{1}{\sqrt{n}}\sum_{T_{n,i}\leq t}\beta_p x(T_{n,i}) = \sqrt{n}(\beta_p\int_0^t x_p(s)ds + o(1))$ diverges. The same happens with $\hat{w}_n$ because x_p is in the range of $\mathcal{T}$.

Remark 2 As a consequence, any test that rejects $\mathcal{H}_0$ when $\hat{w}_n$ is large, is consistent.

3.2.4 Focusing the Power

Suppose now that we wish to focus the test to detect local alternatives including the function x_p with the properties specified in the previous section.

A contiguous sequence of such alternatives can be written as

$$\mathcal{H}_n(\delta, x_p) : \boldsymbol{Y}_n = \boldsymbol{X}_n\boldsymbol{\beta} + X'_n\frac{\delta\sigma}{\sqrt{n}} + \sigma\boldsymbol{Z}_n. \tag{18}$$

We call δ and x_p the *size* and *direction* of the sequence of alternatives.

Under the alternative model Eq. 18, $\hat{\boldsymbol{e}}_n = \frac{H_n\boldsymbol{Y}_n}{\hat{\sigma}_n} = \frac{X'_n\delta/\sqrt{n} + H_n\boldsymbol{Z}_n}{\hat{\sigma}_n/\sigma}$ so that $\hat{r}_n(t) = \frac{\sigma}{\hat{\sigma}_n}\left(\frac{1}{\sqrt{n}}\sum_{T_{n,i}\leq t}(H_n\boldsymbol{Z}_n)_i + \frac{\delta}{n}\sum_{T_{n,i}\leq t}x_p(T_{n,i})\right) \to \rho(t) + \delta\int_0^t x_p(s)ds$ and the fi.di. limit of $\hat{w}_n(t)$ is

$$\int_0^1 \mathcal{T}(a\mathbf{1}_t)(s)d\rho(s) + \delta\int_0^1 \mathcal{T}(a\mathbf{1}_t)(s)x_p(s)ds = w^{(V)}(t) + \delta\int_0^t a(s)\mathcal{T}^{-1}x_p(s)ds.$$

Remark 3 As in Section 3.1.2, we choose the score function a in order to maximize the rate between the drift $\delta\int_0^t a(s)\mathcal{T}^{-1}x_p(s)ds$ and the standard deviation $\sqrt{V(t)}$ of the random term, thus obtaining $\hat{a} = \mathcal{T}^{-1}x_p$.

Remark 4 A very simple way to construct the test of fit to the model with columns given by $x_0, x_1, \ldots, x_{p-1}$, focused on x_p is to complete $(x_j)_{j=0,1,\ldots,p}$ in order to get an orthonormal basis $(x_j)_{j=0,1,2,\ldots}$ of $L^2(0,1)$ and define $\mathcal{T} : x_j \mapsto x_{j+p}$. This leads to $\hat{a} = x_0$.

3.3 Asymptotic Critical Values and Power of the Tests with Optimum Scores

Theorem 3 *The limiting distribution of Q_n under $\mathcal{H}_n(\delta, x_p)$ is that of*

$$\int_0^1\int_0^1 \gamma(u,v)(dw(u) + \delta du)(dw(u) + \delta dv) = Q + \tfrac{2}{3}\delta w(1) + \tfrac{\delta^2}{3},$$

with $Q = \int_0^1\int_0^1 \gamma(u,v)dw(u)dw(u)$.

Remark 5 Consequently an approximate critical region of level α for large n is $Q_n > q(\alpha)$, where $q(\alpha)$ solves $\mathbf{P}\{Q > q(\alpha)\} = \alpha$. The resulting asymptotic power is

$$\Pi_Q(\delta) = \mathbf{P}\left\{Q + \frac{2}{3}\delta w(1) + \frac{\delta^2}{3} > q(\alpha)\right\}.$$

In our experience, the use of asymptotic critical levels for samples of moderate size leads to actual levels of significance smaller that the level α of design. The undesirable consequence is that the power obtained is less than the asymptotic one. This can be avoided by computing the critical value for the required n by means of a simulation.

Table 1 shows the asymptotic power of the test with rejection regions $Q_n > q(\alpha)$, for $\alpha = 5\%$ and $0 \le \delta \le 5$, corresponding to alternatives $\mathcal{H}_n(\delta, x_p)$ and optimum score function $\hat{a} = \mathcal{T}^{-1}x_p$. As a reference, the values of $\Pi_0(\delta) = \Phi(\Phi^{-1}(.025) + \delta) + \Phi(\Phi^{-1}0.025) - \delta)$ are also included in the table. The function $\Pi_0(\delta)$ is the asymptotic power of the test with rejection region $|w_n(1)| > \Phi^{-1}(.975)$, which is equivalent to the two-sided test based on the likelihood ratio statistic. These tests are not consistent, and it may be noticed that the loss in replacing Π_0 by the asymptotic power of the consistent test presented in this article, is not significant (see the last column in Table 1).

Remark 6 It is worth noticing that with the optimum selection of the score function, the asymptotic distribution of Q_n under the alternatives to which the test is focused is distribution free, that is, does not depend on the shape x^* of the alternative.

Remark 7 When the test is constructed with the score a and the alternative is actually in the direction of some unitary vector x in the range of $\mathcal{T}$, then the asymptotic laws of $\hat{w}_n$ and Q_n have biases $\delta \int_0^t a(s)\mathcal{T}^{-1}x(s)ds$ and $\delta^2 \int_0^1 \int_0^1 \gamma(V(s), V(t))x(s)x(t)dsdt$, respectively.

Table 1 Asymptotic powers of the tests with rejection regions $Q_n > q(.05)$ ($\Pi(\delta)$) and $|w_n(1)| > \Phi^{-1}(.975)$ ($\Pi_0(\delta)$), and relative difference in percent of both powers

δ	$\Pi(\delta)$	$\Pi_0(\delta)$	$100\frac{\Pi_0(\delta)-\Pi(\delta)}{\Pi_0(\delta)}$	δ	$\Pi(\delta)$	$\Pi_0(\delta)$	$100\frac{\Pi_0(\delta)-\Pi(\delta)}{\Pi_0(\delta)}$
0.0	0.050	0.050	0.00	2.6	0.737	0.739	0.32
0.2	0.055	0.055	0.03	2.8	0.797	0.800	0.28
0.4	0.068	0.069	0.13	3.0	0.849	0.851	0.24
0.6	0.092	0.092	0.24	3.2	0.891	0.893	0.19
0.8	0.126	0.126	0.32	3.4	0.924	0.925	0.15
1.0	0.169	0.170	0.38	3.6	0.948	0.950	0.12
1.2	0.223	0.224	0.42	3.8	0.966	0.967	0.09
1.4	0.287	0.288	0.44	4.0	0.979	0.979	0.06
1.6	0.358	0.360	0.45	4.2	0.987	0.987	0.04
1.8	0.435	0.437	0.44	4.4	0.992	0.993	0.03
2.0	0.514	0.516	0.42	4.6	0.996	0.996	0.02
2.2	0.592	0.595	0.40	4.8	0.998	0.998	0.01
2.4	0.668	0.670	0.36	5.0	0.999	0.999	0.01

4 Computing the Test Statistic Q_n by Means of Fourier Expansions

Let us assume that $x_0, x_1, \ldots, x_{p-1}, x_p, \ldots$ is a complete orthonormal set of functions in $L^2(0,1)$, and obtain the Fourier expansion of

$$\hat{w}_n(t) = \int_0^1 \mathcal{T}(a\mathbf{1}_t)(s)d\hat{r}_n(s) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathcal{T}(a\mathbf{1}_t)(T_{n,i})\hat{e}_{n,i},$$

with $a = x_0$ as suggested in Remark 4.

The function $x_0(s)\mathbf{1}_t(s)$ has the expansion $\sum_{j=0}^{\infty}\int_0^t x_0(r)x_j(r)drx_j(s)$, hence $\mathcal{T}(x_0\mathbf{1}_t)(s) = \sum_{j=0}^{\infty}\int_0^t x_0(r)x_j(r)drx_{j+p}(s)$. Therefore

$$\hat{w}_n(t) = \sum_{j=0}^{\infty}\int_0^t x_0(r)x_j(r)dr\bar{\epsilon}_{n,j+p},$$

with

$$\bar{\epsilon}_{n,j} = \frac{1}{\sqrt{n}}\sum_{i=1}^n x_j(T_{n,i})\hat{e}_{n,i}, \tag{19}$$

and hence $d\hat{w}_n(t) = \sum_{j=0}^{\infty} x_0(t)x_j(t)\bar{\epsilon}_{n,j+p}dt$, so that

$$Q_n = \int_0^1\int_0^1 \gamma(V(s),V(t))dw_n(s)dw_n(t) = \sum_{j,k} c_{j,k}\bar{\epsilon}_{n,j+p}\bar{\epsilon}_{n,k+p}, \tag{20}$$

is written as a quadratic form in the infinite vector of weighted means of residues Eq. 19, with coefficients

$$c_{j,k} = \int_0^1\int_0^1 \gamma(V(s),V(t))x_0(s)x_0(t)x_j(s)x_k(t)ds\,dt \tag{21}$$

that only depend on the orthonormal basis and the score function a.

In general, the coefficients of the quadratic form can be obtained by numerical computation. For practical purposes we replace Q_n by a finite approximation

$$Q_n^{\ell} = \sum_{j=0}^{\ell}\sum_{k=0}^{\ell} c_{j,k}\bar{\epsilon}_{n,j+p}\bar{\epsilon}_{n,k+p}$$

corresponding to a moderate value of ℓ. The choice of ℓ obeys only to numerical and not statistical reasons.

5 Goodness of Fit to a Fixed Design Model

We choose now the points of observation $t_{n,i}$ equally spaced over the interval [0, 1], namely $t_{n,i} = \frac{i}{n+1}$, $i = 1, 2, \ldots, n$.

The random matrix $\boldsymbol{X}_n$ on Eq. 4 is then replaced by the deterministic matrix $\boldsymbol{X}_n^*$ with rows

$$(\boldsymbol{X}_n^*)_{i,\cdot} = (x_0(t_{n,i}), \ldots, x_{p-1}(t_{n,i})).$$

In this case our statistics $Q_n^\ell = \sum_{j=0}^{\ell} \sum_{k=0}^{\ell} c_{j,k} \bar{\epsilon}_{n,j+p} \bar{\epsilon}_{n,k+p}$ computed from the empirical data through the linear combinations of residues

$$\bar{\epsilon}_{n,j} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_j(T_{n,i}) e_{n,i}, \text{ when the parameters are given, or}$$

$$\bar{\epsilon}_{n,j} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_j(T_{n,i}) \hat{e}_{n,i}, \text{ when the parameters are estimated,}$$

must be replaced by $Q_n^{\ell,*} = \sum_{j=0}^{\ell} \sum_{k=0}^{\ell} c_{j,k} \bar{\epsilon}^*_{n,j+p} \bar{\epsilon}^*_{n,k+p}$ where

$$\bar{\epsilon}^*_{n,j} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_j(t_{n,i}) e_{n,i}, \text{ or } \bar{\epsilon}^*_{n,j} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_j(t_{n,i}) \hat{e}^*_{n,i},$$

respectively. Let us notice that the new estimated residues differ from the old ones, even if the errors are assumed to be the same, because they are obtained from Eq. 14 with $\boldsymbol{X}_n^*$ substituted for $\boldsymbol{X}_n$.

We treat both cases separately, with the purpose of establishing that for each ℓ, $\text{plim}(Q_n^{\ell,*} - Q_n^\ell) = 0$, which ensures that the limiting law of both sequences of statistics is the same.

5.1 The Model with Known Parameters

As in Section 3.1, the residues $e_{n,i} = Z_{n,i} = \sqrt{n}\big(r(\frac{i}{n}) - r\big(\frac{i-1}{n}\big)\big)$ coincide with the standardized errors. We introduce now the notations $\xi_{n,j}$ for the sectionally constant random function that has the value $x_j(T_{n,i})$ on $\big(\frac{i-1}{n}, \frac{i}{n}\big)$, and $\xi^*_{n,j}$ for the sectionally constant function with the value $x_j(t_{n,i})$ on $\big(\frac{i-1}{n}, \frac{i}{n}\big)$.

Then

$$\bar{\epsilon}^*_{n,j} - \bar{\epsilon}_{n,j} = \sum_{i=1}^{n} (x_j(t_{n,i}) - x_j(T_{n,i})) \left(r\left(\frac{i}{n}\right) - r\left(\frac{i-1}{n}\right)\right)$$

can be thought as the stochastic integral of $\xi^*_{n,j} - \xi_{n,j}$ with respect to dr. That integral has expectation zero and variance $\mathbf{E} \int_0^1 (\xi n, j^*(t) - \xi_{n,j}(t))^2 dt = \frac{1}{n} \sum_{i=1}^{n} \mathbf{E}(x_j(t_{n,i}) - x_j(T_{n,i}))^2 \to 0$, so that $\text{plim}(\bar{\epsilon}^*_{n,j} - \bar{\epsilon}_{n,j}) = 0$ and hence, for each ℓ, $\text{plim}(Q_n^{\ell,*} - Q_n^\ell) = 0$.

5.2 The Model with Unknown Parameters

From the estimated residues

$$\hat{\boldsymbol{e}}^*_n = \frac{\boldsymbol{Y}_n - \boldsymbol{X}^*_n \hat{\boldsymbol{\beta}}^*_n}{\hat{\sigma}^*_n} = \frac{H^*_n \boldsymbol{Y}_n}{\hat{\sigma}^*_n} = \frac{H^*_n \boldsymbol{Z}_n}{\hat{\sigma}^*_n / \sigma}, \quad H^*_n = I - \boldsymbol{X}^*_n (\boldsymbol{X}^{*\text{tr}}_n \boldsymbol{X}^*_n)^{-1} \boldsymbol{X}^{*\text{tr}}_n,$$

we compute

$$\begin{aligned} \bar{\epsilon}^*_{n,j} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_j(t_{n,i}) \hat{e}^*_{n,i} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{x_j(t_{n,i})}{\hat{\sigma}^*_n / \sigma} (Z_{n,i} - (\boldsymbol{X}^*_n)_i (\boldsymbol{X}^{*\text{tr}}_n \boldsymbol{X}^*_n)^{-1} \boldsymbol{X}^{*\text{tr}}_n \boldsymbol{Z}_n). \end{aligned}$$

With the notation $\boldsymbol{\xi}_n^* = (\xi_{n,0}^*, \xi_{n,1}^*, \ldots, \xi_{n,p-1}^*)$, $\bar{\epsilon}_{n,j}^*$ is written as

$$\frac{\sigma}{\hat{\sigma}_n^*}\left(\int_0^1 \xi_j^* dr - \int_0^1 \xi_{n,j}^*(t)\boldsymbol{\xi}_n^*(t)dt\left(\int_0^1 (\boldsymbol{\xi}_n^*(t))^{\mathrm{tr}}\boldsymbol{\xi}_n^*(t)dt\right)^{-1}\int_0^1 (\boldsymbol{\xi}_n^*(t))^{\mathrm{tr}} dr(t)\right).$$

Since the continuity of the functions x_j implies that

(i) $\mathrm{plim}_{n\to\infty}\int_0^1 \xi_{n,j}^* dr = \int_0^1 x_j(t)dr(t)$,
(ii) $\lim_{n\to\infty}\int_0^1 \xi_{n,j}^*(t)\boldsymbol{\xi}_n^*(t)dt = (\delta_{0,j}, \delta_{1,j}, \ldots, \delta_{p-1,j}) = 0$ for $j \geq p$,
(iii) $\lim_{n\to\infty}\int_0^1 (\boldsymbol{\xi}_n^*(t))^{\mathrm{tr}}\boldsymbol{\xi}_n^*(t)dt = I$,
(iv) $\mathrm{plim}_{n\to\infty}\int_0^1 (\boldsymbol{\xi}_n^*(t))^{\mathrm{tr}} dr(t) = \int_0^1 (\boldsymbol{x}(t)^{\mathrm{tr}})dr(t)$

and $\hat{\sigma}^*$ given by Eq. 13 is likewise a consistent estimator of σ, then $\mathrm{plim}\bar{\epsilon}_{n,j}^* = \int_0^1 x_j(t)dr(t)$ for $j \geq p$.

On the other hand, as seen in Section 3.2.1 with arguments based on the properties of the stochastic integrals and the Law of Large Numbers, similar limits can be obtained for the formulas based on random matrices $\boldsymbol{X}_n$ and functions $\boldsymbol{\xi}_n$, and therefore the limit in probability of $\bar{\epsilon}_{n,j}$ for $j \geq p$ is the same Wiener integral $\int_0^1 x_j(t)dr(t)$.

This leads us to conclude again that $\mathrm{plim}(Q_n^{\ell,*} - Q_n^{\ell}) = 0$.

6 An Example: Polynomial Regression

We assume that each x_j is a polynomial of degree j. Since we impose that these functions constitute an orthonormal system, they are determined up to its sign to be the so called *modified Legendre Polynomials*.

6.1 The Legendre Polynomials

Consider the Hilbert space $L^2((0,1), d\lambda)$, where λ is the Lebesgue measure. The sequence of polynomials $p_n(t)$, $n = 0, 1, \ldots$ with degree equal to the index, orthonormal in $L^2((0,1), d\lambda)$, is obtained by a change of variable and normalization from the Legendre Polynomials P_n, given by

$$P_n(x) = \frac{1}{2^n n!}\frac{d^n}{dx^n}(x^2-1)^n.$$

The sequence P_n is orthogonal in $L^2((-1,1), \lambda)$ and

$$\langle P_m, P_n\rangle = \int_{-1}^1 P_m(s)P_n(s)ds = \frac{2\delta_{m,n}}{2n+1}, \tag{22}$$

with $\delta_{m,n} = \mathbf{1}_{\{m=n\}}$.

From Eq. 22, it follows that $p_n(t) = \sqrt{2n+1}P_n(2t-1)$ are the required orthonormal polynomials in $L^2([0,1], d\lambda)$. See for instance, Sansone (1959), Chapter III.

6.2 Testing if a Polynomial Regression of Order $p-1$ Fits the Data

In order to test $\mathcal{H}_0$ with $\mathcal{X}$ given by Eq. 2, $x_j = p_j$, focusing the power in the alternative of a polynomial model with degree p, we choose $a = x_0 = 1$, so that $V(t) = t$.

A general expression for the coefficients

$$c_{j,k} = \int_0^1 \int_0^1 \gamma(s,t) p_j(s) p_k(t) ds\, dt \tag{23}$$

is available (see Section 8, Theorem 4): they are all zero, except for $c_{0,0} = 1/3$, $c_{1,1} = 1/30$, and, for $j \geq 2$, $c_{j,j} = \frac{1}{(2j-1)(2j+3)}$, $c_{j-1,j+1} = c_{j+1,j-1} = -\frac{1}{4j+2}\sqrt{c_{j,j}}$.

Their values corresponding to small j and k are indicated in Table 2.

6.3 Empirical Determination of the Power in the Particular Case $p = 2$, with Equally Spaced Observations

Let us assume now that we wish to test the null hypothesis that a linear regression fits well the observations $\boldsymbol{Y}_n^{\text{tr}} = (Y_1, Y_2, \ldots, Y_n)$, and consider the alternative models

$$\boldsymbol{Y}_n = (\boldsymbol{x}_0, \boldsymbol{x}_1)\begin{pmatrix}\beta_0 \\ \beta_1\end{pmatrix} + \frac{\delta\sigma}{\sqrt{n}}\boldsymbol{x}_h + \sigma \boldsymbol{Z}_n,$$

$$\boldsymbol{x}_h = \left(p_h\left(\frac{1}{n+1}\right), p_h\left(\frac{2}{n+1}\right), \ldots, p_h\left(\frac{n}{n+1}\right)\right)^{\text{tr}} \tag{24}$$

for $h = 2, 3, 4, \ldots$

The power of the test based on $Q_n^{\ell,*}$ for different alternatives depends only on δ, h, ℓ and n, but not on the parameters β_0, β_1, σ. Therefore, an empirical study of the powers can be performed by simulating samples of the model with these parameters chosen arbitrarily, for instance, $\beta_0 = \beta_1 = 0$, $\sigma = 1$.

We have generated 20.000 samples following the null hypothesis $\delta = 0$ in order to compute the critical value of $Q_n^{\ell,*}$ for $n = 50$ and $\ell = 8, 20$. Then we have performed the test focused on the alternative corresponding to $h = 2$, for 5.000 samples of the same size, generated with $\delta = 0 : .25 : 10$ and $h = 2, 3, 4, 5, 6$.

The resulting powers are in Table 3. The change in ℓ makes no important differences. The powers corresponding to $\ell = 20$ are indicated graphically in Fig. 1.

Table 2 Values of $c_{j,k}$ for $j, k = 0, 1, \ldots, 7$

j	k=0	1	2	3	4	5	6	7
0	$\frac{1}{3}$	0	0	0	0	0	0	0
1	0	$\frac{1}{30}$	0	$-\frac{1}{10\sqrt{21}}$	0	0	0	0
2	0	0	$\frac{1}{21}$	0	$-\frac{1}{42\sqrt{5}}$	0	0	0
3	0	$-\frac{1}{10\sqrt{21}}$	0	$\frac{1}{45}$	0	$-\frac{1}{18\sqrt{77}}$	0	0
4	0	0	$-\frac{1}{42\sqrt{5}}$	0	$\frac{1}{77}$	0	$-\frac{1}{66\sqrt{13}}$	0
5	0	0	0	$-\frac{1}{18\sqrt{77}}$	0	$\frac{1}{117}$	0	$-\frac{1}{26\sqrt{165}}$
6	0	0	0	0	$-\frac{1}{66\sqrt{13}}$	0	$\frac{1}{165}$	0
7	0	0	0	0	0	$-\frac{1}{26\sqrt{165}}$	0	$\frac{1}{221}$

Table 3 Powers (rejections in 1000 trials based on 10000 replications) of the test with level 5% for linear regression focused on the addition of a quadratic term, for samples of the alternative model Eq. 24 with $\ell = 8$ and 20, and $h = 2, 3, 4, 5, 6$

δ	$\ell = 8$					$\ell = 20$				
	$h = 2$	3	4	5	6	$h = 2$	3	4	5	6
Sample size $n = 50$										
0	50	50	50	50	50	50	50	50	50	50
0.5	74	52	52	51	53	76	51	52	50	50
1	149	55	58	54	56	154	54	56	53	53
1.5	272	64	69	58	61	286	61	66	56	58
2	443	75	83	65	68	464	71	81	63	65
2.5	626	92	108	73	79	648	86	105	73	76
3	780	117	146	88	91	798	111	145	87	88
3.5	893	153	198	107	107	902	145	197	107	107
4	954	203	274	132	129	960	194	269	130	127
4.5	984	272	362	173	159	988	259	362	165	152
5	996	356	475	218	191	996	346	467	214	183
5.5	999	455	594	276	230	999	450	585	272	223
6	1000	566	708	350	279	1000	557	703	343	268
6.5	1000	673	803	432	335	1000	666	799	431	321
7	1000	769	879	520	394	1000	765	877	519	380
7.5	1000	849	937	611	460	1000	846	933	609	446
8	1000	907	972	700	528	1000	905	968	696	518
8.5	1000	947	988	781	599	1000	949	987	777	591
9	1000	970	995	845	668	1000	974	993	848	661
9.5	1000	986	998	898	736	1000	988	998	900	726
10	1000	992	999	936	795	1000	995	999	936	785
Sample size $n = 200$										
0	50	50	50	50	50	50	50	50	50	50
0.5	83	51	51	51	51	84	51	52	51	51
1	178	53	55	53	51	163	54	54	55	51
1.5	327	59	65	56	54	313	60	63	58	54
2	511	67	78	61	56	494	67	77	63	58
2.5	696	79	103	69	60	687	79	98	71	62
3	843	98	143	79	64	837	102	136	82	66
3.5	935	127	204	93	69	929	134	196	100	73
4	978	172	293	111	76	975	181	283	127	82
4.5	994	240	416	143	86	993	251	398	166	93
5	999	327	561	184	100	999	339	539	213	109
5.5	1000	439	701	238	118	1000	452	685	278	130
6	1000	565	824	310	143	1000	579	806	360	158
6.5	1000	692	912	402	173	1000	702	899	457	193
7	1000	800	960	510	212	1000	811	955	563	238
7.5	1000	885	985	616	268	1000	892	984	668	297
8	1000	941	995	719	332	1000	945	994	761	363
8.5	1000	973	999	812	411	1000	975	998	842	441
9	1000	990	1000	886	500	1000	991	1000	907	536
9.5	1000	997	1000	933	596	1000	996	1000	946	633
10	1000	999	1000	965	688	1000	998	1000	974	727

Note: Fig. 1 exhibits a power for the alternative $h = 3$ smaller than the one for $h = 4$. This apparent anomaly is explained because the relative asymptotic increments $\frac{\int_0^1 \gamma(u-v)x_{h-2}^2(u)du}{\int_0^1 \gamma(u-v)x_0^2(u)du} = 3\int_0^1 \gamma(u-v)x_{h-2}^2(u)du$ (see the final remark in Section 3.3) in the expectation of the statistic Q_n for a given δ and $h = 3, 4, 5, 6$ with respect to the increment for the same δ and $h = 2$ are respectively 10.0%, 14.2%, 6.7% and 3.9%, and hence, when the test is tuned to detect the contamination by a polynomial of order 2, it becomes easier to detect the contamination by a polynomial of degree four, than by a polynomial of degree three.

6.4 Embedding Particular Cases into the General Scheme. An Example

Let us apply the general scheme of Section 3.2.4, Remark 2 to an example of Stute, Thies and Zhu, taken from Stute et al. (1998):

The null hypothesis model is $Y_{n,i} = \theta T_{n,i} + \sigma Z_{n,i}$ where $T_{n,i}$ is the i-th order statistic of a sample of size n of the uniform distribution on $(0, 1)$ and $Z_{n,i}$, $i = 1, 2, \ldots, n$ are i.i.d. Normal(0,1).

Our test shall be focused on the alternatives $Y_{n,i} = \theta T_{n,i} + aT_{n,i}^2 + \sigma Z_i$. For that purpose we introduce the orthonormal system $x_0(t) = \sqrt{3}t = \frac{\sqrt{3}}{2}p_0(t) + \frac{1}{2}p_1(t)$, $x_1(t) = \sqrt{5}(4t^2 - 3t) = \frac{2}{3}p_2(t) + \frac{\sqrt{5}}{2\sqrt{3}}p_1(t) - \frac{\sqrt{5}}{6}p_0(t)$, $x_2(t) = \frac{\sqrt{5}}{\sqrt{3}}p_2(t) - \frac{1}{\sqrt{3}}p_1(t) + \frac{1}{3}p_0(t)$, $x_i = p_i$ $(i = 3, 4, 5, \ldots)$. The ingredients needed in order to compute the statistics statistic given by Eqs. 19, 20 and 21 are: The isometry $\mathcal{T}$ equal to the shift $\mathcal{T}x_i = x_{i+1}$, the score $a = \mathcal{T}^{-1}x_1 = x_0$ and consequently the function $V(t) = \int_0^t a^2(s)ds = t^3$.

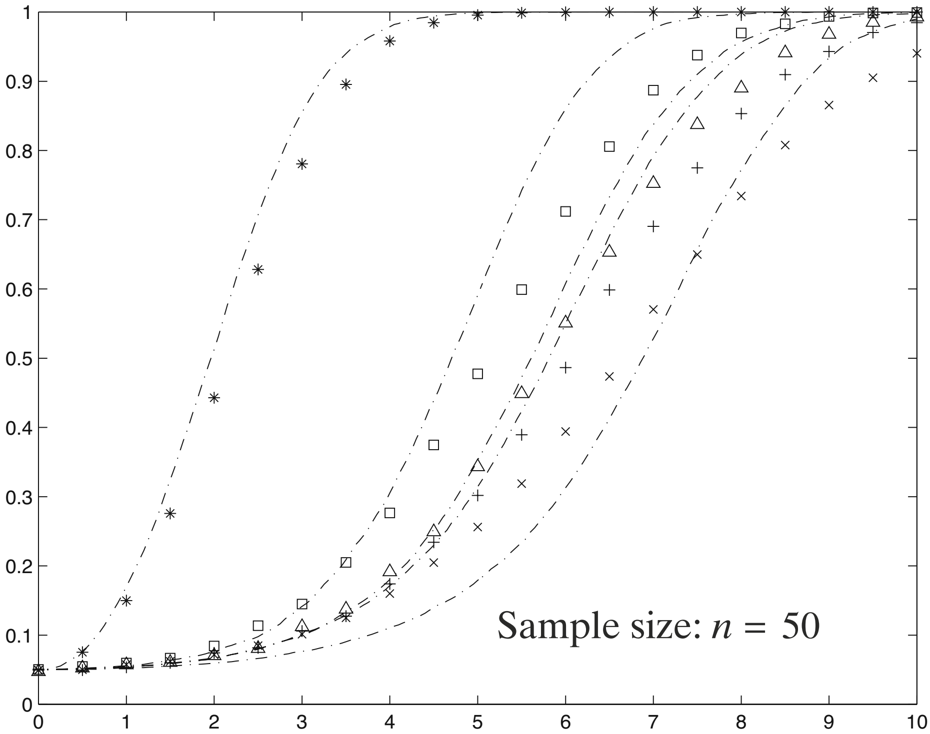
The coefficients Eq. 21 now are

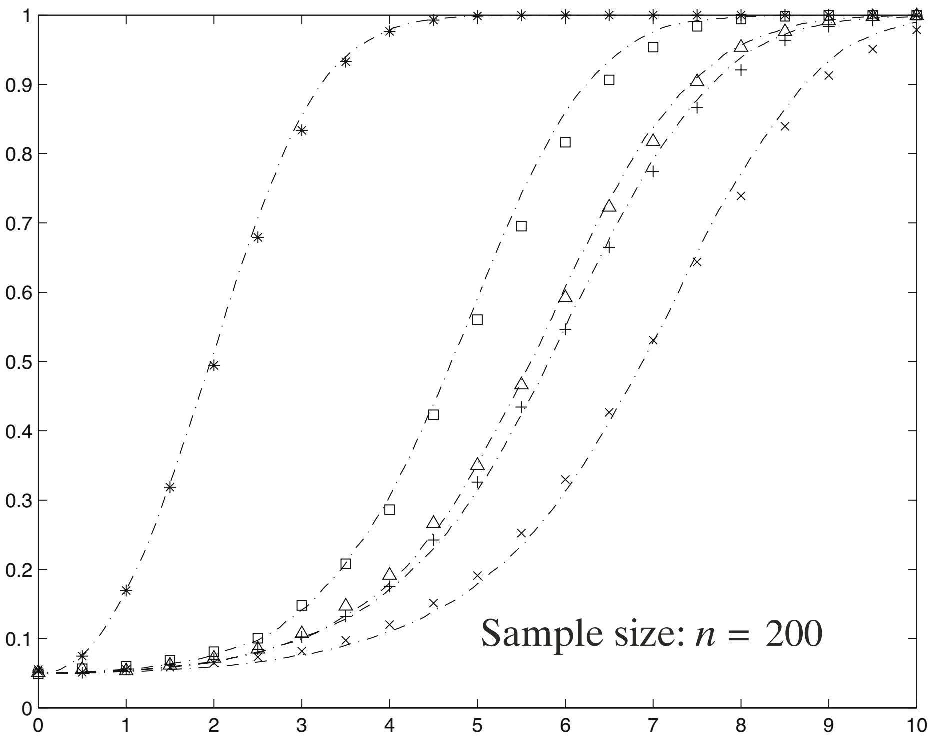
$$c_{j,k} = 3\int_0^1 \int_0^1 \gamma(s^3, t^3) st x_j(s) x_k(t) ds\, dt$$

and, since $sx_j(s)$ and $tx_k(t)$ are polynomials, then $c_{j,k}$ can be readily obtained after computing

$$\begin{aligned} b_{h,i} &= \int_0^1 \int_0^1 \gamma(s^3, t^3) s^h t^i ds dt = \int_0^1 \int_0^1 \left[(s^3 - t^3)^2 - |s^3 - t^3| + \frac{1}{2}\right] s^h t^i ds dt \\ &\quad -2\int_0^1 \left(\int_0^s (s^3 - t^3) t^i dt\right) s^h ds + \int_0^1 \left(\int_0^1 (s^3 - t^3) t^i dt\right) s^h ds \\ &= \frac{1}{(h+7)(i+1)} + \frac{1}{(h+1)(i+7)} - \frac{2}{(h+4)(i+4)} + \frac{1}{2(h+1)(i+1)} \\ &\quad -3\frac{h^2 + 5h + i^2 + 5i + 8}{(h+1)(h+4)(i+1)(i+4)(h+i+5)}. \end{aligned}$$

In fact, let B, C denote the (infinite) matrices with entries $b_{h,i}$, $c_{h,i}$ $(h, i = 0, 1, 2, \ldots)$, respectively, P the lower diagonal matrix with entries $p_{h,i}$, $h = 0, 1, 2, \ldots$,


Sample size: n = 50
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0 1 2 3 4 5 6 7 8 9 10


Sample size: n = 200
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0 1 2 3 4 5 6 7 8 9 10

◀ **Fig. 1** Powers of the test of fit for linear regression, focused on the addition of a quadratic term, for $\ell = 20$ and alternatives adding Legendre polynomials of degrees 2 (*asterisks*), 3 (*triangles*), 4 (*squares*), 5 (*plus signs*) and 6 (*multiplication signs*). The *dotted lines* indicate the asymptotic powers

$i = 0, \ldots, h$ equal to the coefficients of the Legendre polynomials $p_h(t) = \sum_{i=0}^{h} p_{h,i} t^i$, $\boldsymbol{r}(t)$ the vector of powers $\boldsymbol{r}(t) = (1, t, t^2.t^3, \ldots)^{\mathrm{tr}}$ and

$$M_0 = \begin{pmatrix} \sqrt{3}/2 & 1/2 & 0 \\ -\sqrt{5}/6 & \sqrt{5}/(2\sqrt{3}) & 2/3 \\ 1/3 & -1/\sqrt{3} & \sqrt{5}/\sqrt{3} \end{pmatrix}.$$

With these notations, the vector of Legendre polynomials can be computed as

$$\boldsymbol{p}(t) = (p_0(t), p_1(t), p_2(t), p_3(t), \ldots)^{\mathrm{tr}} = \boldsymbol{P}\boldsymbol{r}(t)$$

and the vector of functions of the orthonormal system adapted to our model is $\boldsymbol{x}(t) = (x_0(t), x_1(t), x_2(t), x_3(t), \ldots)^{\mathrm{tr}}$, given by

$$\boldsymbol{x}(t) = M\boldsymbol{p}(t), \text{ with } M = \begin{pmatrix} M_0 & 0 \\ 0 & I \end{pmatrix}$$

so that

$$C = 3\int_0^1 \int_0^1 \gamma(s^3, t^3) st\boldsymbol{x}(s)(\boldsymbol{x}(t))^{\mathrm{tr}} dsdt = 3MPI_1BI_1^{\mathrm{tr}}P^{\mathrm{tr}}M^{\mathrm{tr}}.$$

Our statistic Q_n is the quadratic form with matrix C evaluated in the vector $\bar{\boldsymbol{\epsilon}}$ of components $(\bar{\epsilon}_1, \bar{\epsilon}_2, \ldots)^{\mathrm{tr}}$ given by Eq. 19. Let $\bar{\boldsymbol{\eta}} = (\bar{\eta}_0, \bar{\eta}_1, \bar{\eta}_2, \ldots)^{\mathrm{tr}}$, $\bar{\boldsymbol{\eta}}_j = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} p_j(T_{n,i})\hat{e}_{n,i}$, so that $\bar{\boldsymbol{\epsilon}} = I_1 M\bar{\boldsymbol{\eta}}$ and hence $Q_n = \bar{\boldsymbol{\eta}}^{\mathrm{tr}}\tilde{C}\bar{\boldsymbol{\eta}}$ with

$$\tilde{C} = 3M^{\mathrm{tr}}I_1^{\mathrm{tr}}MPI_1BI_1^{\mathrm{tr}}P^{\mathrm{tr}}M^{\mathrm{tr}}I_1M. \tag{25}$$

The finite quadratic form $Q_n^{(\ell)}$ involving $(\bar{\epsilon}_1, \bar{\epsilon}_2, \ldots, \epsilon_{\ell+1})^{\mathrm{tr}}$ is obtained by replacing the infinite matrices in Eq. 25 by the square $\ell + 2$-matrices with rows and columns from 0 to $\ell + 1$, because of the particular (essentially upper-diagonal) shape of the factors at the right-hand side of B.

The empirical powers based on 10.000 replications are indicated in Table 4. The powers reported in Stute et al. (1998) for the test of Cramér - von Mises type

Table 4 Empirical powers of our test based on $Q_n^{(\ell)}$ and the test based on $\tilde{W}_n^2$ proposed in Stute et al. (1998) Section 2 for the null hypothesis $a = 0$, based on samples of size $n = 200$ of $Y_{n,i} = 5T_{n,i} + aT_{n,i}^2 + \sigma Z_i$ ($a = 0, 1,; \sigma^2 = 1, 2, 3$)

a	σ^2	$Q_n^{(8)}$-test		$\tilde{W}_n^2$-test	
		$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$
0	1	5.11	1.05	5.7	1.5
	2	4.92	1.01	5.2	0.8
	3	4.97	0.94	4.9	0.8
1	1	34.45	14.97	30.0	12.7
	2	19.8	6.29	19.8	7.5
	3	14.72	4.25	15.4	4.5
2	1	87.61	68.12	81.8	58.8
	2	60.41	33.08	52.2	28.8
	3	43.79	21.16	39.0	18.6

proposed in that article, are also indicated, for the sake of comparison. Since our test is tunned to the alternatives considered, it is not surprising that the powers of the $Q_n^{(\ell)}$-test be larger than the powers of the $\tilde{W}_n^2$-test.

7 Proofs of the Statements on the Asymptotic Laws of the Test Statistics

7.1 On Theorem 1

In order to prove Theorem 1 we establish first the following lemmas.

Lemma 1 *Let g denote a differentiable function in $L_V^2 = L^2((0,1), dV)$. Then the integrals $\int_0^1 g(t)dw_n(t)$ and $\int_0^1 g(t)dw^{(V)}(t)$ have expectation zero and variance $\|g\|_V^2 = \int_0^1 g^2(t)a^2(t)dt$.*

By applying integration by parts, the definition the TARP w_n in Eq. 7, then interchanging the integrals in t and u and finally interchanging the integral in t and the application of the isometry $\mathcal{T}$, we get

$$\begin{aligned}\int_0^1 g(t)dw_n(t) &= g(1)w_n(1) - \int_0^1 w_n(t)\dot{g}(t)dt \\ &= g(1)w_n(1) - \int_0^1\int_0^1 \mathcal{T}(a\mathbf{1}_t)(u)dr_n(u)\dot{g}(t)dt \\ &= g(1)w_n(1) - \int_0^1 \left(\int_0^1 \mathcal{T}(a\mathbf{1}_t)(u)\dot{g}(t)dt\right) dr_n(u) \\ &= g(1)w_n(1) - \int_0^1 \mathcal{T}(a(\cdot)(g(1)-g(\cdot)))(u)dr_n(u) = \int_0^1 \mathcal{T}(ag)(u)dr_n(u) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathcal{T}(ag)(T_{n,i})e_{n,i} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathcal{T}(ag)(U_{n,i})e_{n,j(i)}\end{aligned}$$

From the last expression we readily obtain $\mathbf{E}\int_0^1 g(s)dw_n(s) = 0$ and $\mathbf{E}(\int_0^1 g(s)dw_n(s))^2 = \int_0^1(\mathcal{T}(ag))^2du = \int_0^1 g^2dV = \|g\|_V^2$.

The same arguments apply to the integral with respect to $w^{(V)}$, and lead to

$$\int_0^1 g(t)dw^{(V)}(t) = \int_0^1 \mathcal{T}(ag)(u)dr(u).$$

This Wiener integral is a centred Gaussian variable with variance $\|\mathcal{T}(ag)\|^2 = \|ag\|^2 = \|g\|_V^2$ as stated.

Corollary 1.1 *For each n, the mappings $g \mapsto \int_0^1 g(V(t))dw_n(t)$ and $g \mapsto \int_0^1 g(V(t))dw^{(V)}(t)$ are isometries between $L^2([0,1], d\lambda)$ and centred random variables in $L^2(\Omega, d\mathbf{P})$.*

The statement follows from the Lemma because differentiable functions are dense in L_V^2.

Lemma 2 *For* $g \in L^2_V$,

$$\mathrm{plim}_{n\to\infty} \int_0^1 g(t)dw_n(t) = \int_0^1 g(t)dw^{(V)}(t).$$

It suffices to verify the statement for g continuously differentiable. In that case, after the computations made in the proof of Lemma 1, the required conclusion reads

$$\mathrm{plim}_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathcal{T}(ag)(T_{n,i})e_{n,i} = \int_0^1 \mathcal{T}(ag)(t)dr(t). \tag{26}$$

Since the function $\mathcal{T}(ag)$ is in L^2, given an arbitrary positive ε, we can write $\mathcal{T}(ag) = G + \Delta$, with G bounded and continuous, and $\|\Delta\|^2 < \varepsilon$.

The sectionally constant approximations to G and Δ defined by $G_n(t) = G(T_{n,[nt]+1})$ and $\Delta_n(t) = \Delta(T_{n,[nt]+1})$ are also introduced, so that the left-hand term in Eq. 26 can be expressed by means of $\int_0^1 (G_n(t) + \Delta_n(t))dr(t)$.

As in the proof of Lemma 1 we compute

$$\mathbf{E}\left(\int_0^1 \Delta_n(t)dr(t)\right)^2 = \mathbf{E}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \Delta(U_{n,i})e_{n,j(i)}\right)^2 = \|\Delta\|^2$$

and

$$\mathbf{E}\left(\int_0^1 \Delta(t)dr(t)\right)^2 = \|\Delta\|^2.$$

By applying the inequality $(a+b+c)^2 \le 3a^2 + 3b^2 + 3c^2$, we derive the estimate:

$$\begin{aligned}
&\mathbf{E}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathcal{T}(ag)(T_{n,i})e_{n,i} - \int_0^1 \mathcal{T}(ag)(t)dr(t)\right)^2 \\
&\le 3\mathbf{E}\left(\int_0^1 (G_n(t) - G(t))dr(t)\right)^2 + 3\mathbf{E}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \Delta(T_{n,i})e_{n,i}\right)^2 \\
&+ 3\mathbf{E}\left(\int_0^1 \Delta(t)dr(t)\right)^2 \le 3\int_0^1 \mathbf{E}(G_n(t) - G(t))^2 dt + 6\varepsilon \to 6\varepsilon \text{ as } n \to \infty
\end{aligned}$$

because of the continuity of G and the convergence to zero of the sequence $\sup_{i=1,2,\ldots,n} |T_{n,i} - i/(n+1)|$.

This bound for arbitrary ε establishes the required convergence in probability.

Lemma 3 *The integral operator* $f \mapsto \int_0^1 \gamma(\cdot, v) f(v)dv$ *on* $L^2([0,1], d\lambda)$ *with kernel* γ *admits the orthonormal system of eigenfunctions*

$$c_0 = 1, c_\nu(u) = \sqrt{2}\cos 2\pi\nu u, s_\nu(u) = \sqrt{2}\sin 2\pi\nu u, (\nu = 1, 2, \ldots)$$

with eigenvalues $\frac{1}{3}, \frac{1}{2(\pi\nu)^2}, \frac{1}{2(\pi\nu)^2}$.

Consequently, the diagonalization

$$\gamma(u, v) = \frac{1}{3} + \frac{1}{2}\sum_{\nu=1}^{\infty} \frac{c_\nu(u)c_\nu(v) + s_\nu(u)s_\nu(v)}{(\pi\nu)^2} \tag{27}$$

holds.

The Fourier coefficients of the function $g(t) = (|t| - \frac{1}{2})^2 + \frac{1}{4}$ on $(-1, 1)$ with the Lebesgue Measure λ, with respect to the orthonormal system $\frac{1}{\sqrt{2}}$, $\cos\pi\nu t$, $\sin\pi\nu t$, $\nu = 1, 2, 3, \ldots$ are $\frac{1}{\sqrt{2}}\int_{-1}^{1} g(t)dt = \frac{\sqrt{2}}{3}$, $\int_{-1}^{1} g(t)\cos\pi\nu t dt = 0$ for ν odd, $\int_{-1}^{1} g(t)\cos 2\pi\nu t dt = \frac{1}{(\pi\nu)^2}$ for $\nu = 1, 2, \ldots$, and $\int_{-1}^{1} g(t)\sin\pi\nu t dt = 0$ for all ν.
Consequently

$$\gamma(u, v) = g(u - v) = \frac{1}{3} + \sum_{\nu=1}^{\infty} \frac{1}{(\pi\nu)^2}\cos 2\pi\nu(u - v)$$

and the statement follows. □

Proof of Theorem 1 The diagonalization in Lemma 3 implies that Q_n can be written as the random series expansion with positive terms

$$\frac{1}{3}w_n^2(1) + \frac{1}{2}\sum_{\nu=1}^{\infty}\frac{1}{(\pi\nu)^2}\left[\left(\int_0^1 c_\nu(V(t))dw_n(t)\right)^2 + \left(\int_0^1 s_\nu(V(t))dw_n(t)\right)^2\right].$$

The partial sums of that series converge in probability to Q_n because the expectation of each squared integral is one and hence the expectation of the remainder $\frac{1}{2}\sum_{\nu=m+1}^{\infty}\frac{1}{\pi^2\nu^2}\left[\left(\int_0^1 c_\nu \circ V dw_n\right)^2 + \left(\int_0^1 s_\nu \circ V dw_n\right)^2\right]$ is $\sum_{\nu=m+1}^{\infty}\frac{1}{\pi^2\nu^2}$ and tends to zero uniformly in n as m goes to infinity. A similar argument shows that the series in Eq. 12 also converges in probability.

On the other hand, each integral in the expansion converges in probability to the integral of the same function with respect to $w^{(V)}$, by Lemma 2. This implies that the partial sums of the series with limit Q_n converge in probability to the corresponding partial sums of Eq. 12, and this suffices to establish (i).

The proofs of the remaining items are plain. The result in (iii) has been shown in Cabaña and Cabaña (2001). □

7.2 On Theorems 2 and 3

The discussion before the statement of Theorem 2 implies its validity. As for the proof of Theorem 3, the conclusions of Theorem 1 and the arguments in the lemmas preceding its proof apply with no essential changes.

8 Appendix. Legendre Polynomials Expansion of $\gamma(u,v) = \left(|u-v| - \frac{1}{2}\right)^2 + \frac{1}{4}$

Theorem 4 *The function $\gamma(u,v) = (|u-v| - \frac{1}{2})^2 + \frac{1}{4}$ admits the expansion in modified Legendre polynomials*

$$\gamma(u,v) = \tfrac{1}{3} + \tfrac{1}{30} p_1(u) p_1(v) + \sum_{j=2}^{\infty} \left[\tfrac{p_j(u) p_j(v)}{(2j-1)(2j+3)} - \tfrac{p_{j-1}(u) p_{j+1}(v) + p_{j+1}(u) p_{j-1}(v)}{(4j+2)\sqrt{(2j-1)(2j+3)}} \right].$$

In order to verify this, let us recall some well known properties of the Legendre polynomials: The following relations hold for all n:

$$P_n(1) = 1,$$

$$(n+1)x P_n(x) - (n+1) P_{n+1}(x) = (1-x^2) P'_n(x) = -nx P_n(x) + n P_{n-1}(x)$$

(the right-hand term has to be interpreted as zero for $n = 0$), and for $m \neq n$:

$$\int_x^1 P_m(y) P_n(y) dy = \frac{(1-x^2)[P_n(x) P'_m(x) - P_m(x) P'_n(x)]}{m(m+1) - n(n+1)}.$$

From these relations we obtain the following properties of the modified Legendre polynomials $p_j(u) = \sqrt{2j+1} P_j(2u-1)$:

$$p_j(1) = \sqrt{2j+1},$$

$$(j+1)(2u-1) \frac{p_j(u)}{\sqrt{2j+1}} - (j+1) \frac{p_{j+1}(u)}{\sqrt{2j+3}} = 2u(1-u) \frac{p'_j(u)}{\sqrt{2j+1}} \tag{28}$$

$$-j(2u-1) \frac{p_j(u)}{\sqrt{2j+1}} + j \frac{p_{j-1}(u)}{\sqrt{2j-1}} = 2u(1-u) \frac{p'_j(u)}{\sqrt{2j+1}} \tag{29}$$

and the particular case of the integral formula for $m = j, n = 0$:

$$\int_u^1 p_j(s) ds = \frac{u(1-u) p'_j(u)}{j(j+1)}. \tag{30}$$

Now multiply Eq. 28 by j, Eq. 29 by $j+1$ and add the resulting equations to get

$$j(j+1) \left[\frac{p_{j-1}(u)}{\sqrt{2j-1}} - \frac{p_{j+1}(u)}{\sqrt{2j+3}} \right] = 2\sqrt{2j+1} u(1-u) p'_j(u), \tag{31}$$

and also subtract the same equations to obtain the recurrence

$$(2j+1)(2u-1) \frac{p_j(u)}{\sqrt{2j+1}} = (j+1) \frac{p_{j+1}(u)}{\sqrt{2j+3}} + j \frac{p_{j-1}(u)}{\sqrt{2j-1}}. \tag{32}$$

For the proof of the Theorem, we establish separately three particular results, namely

$$j = 0 : \int_0^1 \gamma(u, v) dv = \frac{1}{3},$$

$$j = 1 : \int_0^1 \gamma(u, v) p_1(v) = \frac{1}{30} p_1(u) - \frac{p_3(u)}{10\sqrt{21}},$$

$$j = 2 : \int_0^1 \gamma(u, v) p_2(v) dv = \frac{p_2(u)}{21} - \frac{p_4(u)}{14\sqrt{45}}.$$

and the general case, for $j > 2$,

$$\int_0^1 \gamma(u, v) p_j(v) dv = \frac{p_j(u)}{(2j-1)(2j+3)} - \frac{p_{j-2}(u)}{(4j-2)\sqrt{(2j-3)(2j+1)}} - \frac{p_{j+2}(u)}{(4j+6)\sqrt{(2j+1)(2j+5)}}.$$

The particular formulas are obtained by direct integration, taking into account that $p_0(u) = 1$, $p_1(u) = \sqrt{3}(2u - 1)$, $p_2(u) = \sqrt{5}(6u^2 - 6u + 1)$, $p_3(u) = \sqrt{7}(20u^3 - 30u^2 + 12u - 1)$ and $p_4(u) = 3(70u^4 - 140u^3 + 90u^2 - 20u + 1)$. We omit the details.

The left-hand term of this general expression vanish for $u = 1$, since $\gamma(1, v)$ reduces to a polynomial of degree 2. As for the right-hand term, it is easily verified by replacing $p_j(1)$ by its value $\sqrt{2j+1}$, that also vanishes for $u = 1$. Therefore, in order to establish this general formula it suffices to verify the equality of the derivatives of both terms.

Using again that p_j is orthogonal to the polynomials of degree 2, it follows that the integral $\int_0^1 \gamma(u, v) p_j(v) dv$ reduces to $-\int_0^1 |u - v| p_j(v) dv$, that is,

$$\int_0^u (v - u) p - j(v) dv + \int_u^1 (u - v) p_j(v) dv = 2 \int_u^1 (u - v) p_j(v) dv,$$

with derivative

$$2 \int_u^1 p_j(v) dv = \frac{2u(1-u) p'_j(u)}{j(j+1)}$$

as follows from Eq. 30.

The identity to be verified is therefore

$$\frac{2u(1-u) p'_j(u)}{j(j+1)} = \frac{p'_j(u)}{(2j-1)(2j+3)} - \frac{p'_{j-2}(u)}{(4j-2)\sqrt{(2j-3)(2j+1)}} - \frac{p'_{j+2}(u)}{(4j+6)\sqrt{(2j+1)(2j+5)}},$$

which, after multiplying by $\sqrt{2j+1}u(1-u)$, applying Eq. 31, rearranging, reducing coefficients and abbreviating $p_j = \sqrt{2j+1}q_j$, leads to the equivalent equation

$$\begin{aligned} u(1-u)q_{j-1}(u) - u(1-u)q_{j+1}(u) \\ = \frac{3j^2-j-6}{4(2j+3)(2j-3)}q_{j-1}(u) - \frac{3j^2+7j-2}{4(2j+5)(2j-1)}q_{j+1}(u) \\ -\frac{(j-2)(j-1)}{4(2j-1)(2j-3)}q_{j-3}(u) + \frac{(j+2)(j+3)}{4(2j+3)(2j+5)}q_{j+3}(u). \end{aligned} \tag{33}$$

Apply now Eq. 32 to replace $\frac{(j+2)(j+3)}{4(2j+3)(2j+5)}q_{j+3}(u)$ by $(u^2-u+1/4)q_{j+1}(u) - \frac{(j+1)(2u-1)}{4(2j+3)}q_j(u) - \frac{(j+2)^2}{4(2j+3)(2j+5)}q_{j+1}(u)$. This transforms Eq. 33 into

$$\begin{aligned} u(1-u)q_{j-1}(u) = -\frac{j+1}{4(2j+3)(2j-1)}q_{j+1}(u) + \frac{3j^2-j-6}{4(2j+3)(2j-3)}q_{j-1}(u) \\ -\frac{(j-2)(j-1)}{4(2j-1)(2j-3)}q_{j-3}(u) - \frac{(j+1)(2u-1)}{4(2j+3)}q_j(u). \end{aligned}$$

The next two steps use again the recurrence Eq. 32: in the first one we replace $(j+1)q_{j+1}(u)$ by $(2j+1)(2u-1)q_j(u) - jq_{j-1}(u)$, and in the second one, $jq_j(u)$ by $(2j-1)(2u-1)q_{j-1} - (j-1)q_{j-2}(u)$. The equivalent identities to be verified, obtained after reducing the coefficients, are respectively

$$\begin{aligned} u(1-u)q_{j-1}(u) = -\frac{j(2u-1)}{4(2j-1)}q_j(u) \\ +\frac{3j^2-6j+2}{4(2j-1)(2j-3)}q_{j-1}(u) - \frac{(j-2)(j-1)}{4(2j-1)(2j-3)}q_{j-3}(u) \end{aligned}$$

and

$$(j-1)q_{j-1}(u) = (2j-3)(2u-1)q_{j-2}(u) - (j-2)q_{j-3}(u)$$

and this last one is Eq. 32 for $j-2$ instead of j.

References

Cabaña A, Cabaña EM (1997) Transformed empirical processes and modified Kolmogorov-Smirnov tests for multivariate distributions. Ann Stat 25:2388–2409

Cabaña A, Cabaña EM (2001) Goodness-of-fit tests based on quadratic functionals of transformed empirical processes. Statistics 35:171–189

Cabaña A, Cabaña EM (2003) Tests of normality based on transformed empirical processes. Method Comput Appl Probab 5:309–335

Cabaña A, Cabaña EM (2005) Goodness-of-fit to the exponential distribution, focused on Weibull alternatives. Commun Stat Simul Comput 34:711–724

Graneri J (2003) χ^2-type goodness of fit test based on transformed empirical processes for location and scale families. Commun Stat Theory Methods 32:2073–2087

Janssen A (2000) Global power functions of goodness of fit tests. Ann Stat 28:239–253

Khmaladze E (1981) A martingale approach in the theory of gooness-of-fi tests. Theory Probab Appl 26:240–257

Nadaraya EA (1964) On estimating regression.Theory Probab Appl 9:141–142

Sansone G (1959) Orthogonal functions. In: Courant R, Bers L, Stoker JJ (eds) Pure and applied mathematics, vol. IX. Dover Interscience, New York (re-published in 1991)

Scavino M (1999) Procesos empíricos transformados y pruebas de Kolmogorov-Smirnov con un parámetro estimado. Publicaciones Matemáticas del Uruguay 8:179–205
Stute W (1997) Nonparametric model checks for regression. Ann Stat 25:613–641
Stute W, Thies S, Zhu L-X (1998) Model checks for regression: an innovation process approach. Ann Stat 26:1916–1934
Watson GS (1961) Goodness-of-fit tests on a circle. Biometrika 48:109–114
Watson GS (1964) Smooth regression analysis. Sankhya A 26:359–372