 Vol 30: 241–255, 2003

Testing Hypotheses in the Functional Linear Model

HERVÉ CARDOT

Unité Biométrie et Intelligence Artificielle, INRA Toulouse

FRÉDÉRIC FERRATY

GRIMM, Université Toulouse Le Mirail

ANDRÉ MAS

CREST-INSEE et Laboratoire de Statistique et Probabilités, Université Paul Sabatier

PASCAL SARDA

Laboratoire de Statistique et Probabilités, Université Paul Sabatier

ABSTRACT. The functional linear model with scalar response is a regression model where the predictor is a random function defined on some compact set of $\mathbb{R}$ and the response is scalar. The response is modelled as $Y=\Psi(X)+\varepsilon$, where Ψ is some linear continuous operator defined on the space of square integrable functions and valued in $\mathbb{R}$. The random input X is independent from the noise ε. In this paper, we are interested in testing the null hypothesis of no effect, that is, the nullity of Ψ restricted to the Hilbert space generated by the random variable X. We introduce two test statistics based on the norm of the empirical cross-covariance operator of (X, Y). The first test statistic relies on a χ^2 approximation and we show the asymptotic normality of the second one under appropriate conditions on the covariance operator of X. The test procedures can be applied to check a given relationship between X and Y. The method is illustrated through a simulation study.

Key words: asymptotic normality, functional linear model, Hilbert space valued random variables, splines, tests

1. Introduction

In the recent literature, there has been an increasing interest in regression models for functional variables. Among these, there is the situation where the predictor is a random function and the response a scalar. This occurs, for instance, in chemometrics applications (see Frank & Friedman, 1993) and also in climatology (see Ramsay & Silverman, 1997). Several authors have pointed out the usefullness of functional methods that challenge standard regression tools such as partial least squares or principal component regression traditionally used, especially by chemometricians (see, for instance, Marx & Eilers, 1999).

In this context, a functional regression model that has been widely studied is the so-called *functional linear model* (FLM) defined as follows. Let $\{X(t), t \in \mathscr{C}\}$ be a square integrable random function defined on some compact set $\mathscr{C}$ of $\mathbb{R}$ and Y be a real random variable. The variable Y is modelled as

$$Y = \Psi(\{X(t), t \in \mathscr{C}\}) + \varepsilon,$$

where Ψ is a continuous linear operator and ε is a centred real random variable with variance σ^2 and independent from X. This model can also be written as

$$Y = \int_{\mathscr{C}} \psi(t)X(t)\,\mathrm{d}t + \varepsilon, \tag{1}$$

where ψ is a square integrable function defined on $\mathscr{C}$.

The problem of estimating the function ψ (or equivalently the operator Ψ) has been addressed in several papers. Hastie & Mallows (1993) introduce an estimator that minimizes a penalized least-squares criterion (see also Ramsay & Silverman, 1997). Estimators based on penalized splines have been studied by Marx & Eilers (1999) and Cardot *et al.* (1999a, b). Cardot *et al.* propose, in addition, an estimate based on a functional principal component regression and derive asymptotic results for both estimates.

An important problem in practical situations is to test if there is a connection between X and Y; that is, to test the nullity of operator Ψ (or of function ψ). More generally, we investigate in this paper the problem of testing the following hypothesis:

$$H_0: \ \psi = \psi_0, \tag{2}$$

against

$$H_1: \ \psi \neq \psi_0,$$

where ψ_0 is some known function defined on $\mathscr{C}$.

In Section 2, we introduce two test statistics based on the square norm of a normalized version of the cross-covariance operator Δ of the pair (X, Y). Indeed, we show that testing $\psi = 0$ in the closure of the space orthogonal to the null space of the covariance operator of X is equivalent to testing $\Delta = 0$. In Section 3, we illustrate our test procedure on some simulated data and compare the results obtained for both test statistics. We show (see appendix) that our second test statistic is asymptotically normal for an accurate estimator $\hat{\sigma}^2$ of σ^2 and under some conditions on the eigenvalues of the covariance operator Γ of X.

2. Hypothesis testing in the FLM

2.1. Notations

Suppose from now on that $\mathscr{C} = [0, 1]$ and let $\mathscr{H}$ be the separable Hilbert space of square integrable functions defined on $[0, 1]$. Let (X, Y) be a pair of random variables defined on the same probability space, with X valued in $\mathscr{H}$ and Y valued in $\mathbb{R}$. Let Ψ be the regression operator of Y on X, so that $Y = \Psi(\{X(t), t \in [0, 1]\}) + \varepsilon$, and assume that Ψ belongs to the space $\mathscr{H}'$ of continuous linear operator defined in $\mathscr{H}$ and valued in $\mathbb{R}$. Let $\langle \phi, \psi \rangle = \int_0^1 \phi(t)\psi(t)\,\mathrm{d}t$ be the usual inner product of functions ϕ and ψ in $\mathscr{H}$ and let $\|\phi\|$ denote the norm of a function ϕ on $\mathscr{H}$, induced by this inner product. By Riesz's representation theorem, one can identify the spaces $\mathscr{H}$ and $\mathscr{H}'$ and we will take in $\mathscr{H}'$ the norm induced by this identification; that is, for $T \in \mathscr{H}'$, $\|T\| = \|\tau\|$, where τ is the unique function in $\mathscr{H}$ such that $Tx = \langle \tau, x \rangle, x \in \mathscr{H}$.

With these notations, the FLM with scalar response (1) can be written as

$$Y = \langle \psi, X \rangle + \varepsilon, \tag{3}$$

where $\psi \in \mathscr{H}$.

We will assume in the following that X is centred (and then Y is centred) and has a finite second moment, that is, $E\|X\|^2 < \infty$. Then, the covariance operator Γ of X is defined as the operator on $\mathscr{H}$ such that

$$\Gamma x(t) = \int_0^1 E[X(t)X(s)]x(s)\,\mathrm{d}s, \quad x \in \mathscr{H}, \; t \in [0,1].$$

It is well known that the operator Γ is nuclear, self-adjoint and non-negative. Now the cross-covariance operator Δ of (X, Y) is defined as

$$\Delta x = \int_0^1 E[X(t)Y]x(t)\,\mathrm{d}t, \quad x \in \mathscr{H}.$$

The operators Γ, Δ and Ψ are linked by the relation

$$\Delta = \Psi\Gamma. \tag{4}$$

We will denote by $(\lambda_j), j = 1, \ldots$ the decreasing sequence of eigenvalues of operator Γ and by $(V_j), j = 1, \ldots$ a sequence of orthonormal eigenfunctions associated with these eigenvalues.

2.2. Changing the space of reference

Before describing the test procedures, let us make some remarks about the FLM defined in (1). At first, let us recall that, since Γ is self-adjoint, the Hilbert space $\mathscr{H}$ may be decomposed as

$$\mathscr{H} = \mathscr{N}(\Gamma) \oplus \overline{\mathrm{Im}(\Gamma)},$$

where $\mathscr{N}(\Gamma)$ is the null space of Γ, $\mathrm{Im}(\Gamma) = \{y \in \mathscr{H}, \; y = \Gamma x, \; x \in \mathscr{H}\}$ and $\overline{\mathrm{Im}(\Gamma)}$ is the closure in $\mathscr{H}$ of $\mathrm{Im}(\Gamma)$. Thus, ψ can be expanded as follows:

$$\psi = \psi_1 + \psi_2, \tag{5}$$

where $\psi_1 \in \mathscr{N}(\Gamma)$ and $\psi_2 \in \overline{\mathrm{Im}(\Gamma)}$ but only ψ_2 can be estimated when observing X since $E(\langle X, \psi_1\rangle^2) = \langle \Gamma\psi_1, \psi_1\rangle = 0$. This means that, in quadratic mean, the random variable $\langle X, \psi_1\rangle$ is equal to zero and thus model (1) is equivalent to

$$Y = \int_{\mathscr{C}} \psi_2(t)X(t)\,\mathrm{d}t + \varepsilon. \tag{6}$$

Thus, testing for no effect leads to test that ψ belongs to $\mathscr{N}(\Gamma)$ or equivalently to test that $\psi_2 = 0$ and the nullity of ψ can only be tested in the space $\overline{\mathrm{Im}(\Gamma)}$. At this point, we may distinguish two different situations:

- $\overline{\mathrm{Im}(\Gamma)}$ is a finite-dimensional space. Then, our framework is the parametric one; many procedures have been proposed in the literature.
- $\overline{\mathrm{Im}(\Gamma)}$ is an infinite-dimensional space. By construction, $H = \overline{\mathrm{Im}(\Gamma)} \subset \mathscr{H}$ is still a separable Hilbert space equipped with the inner product of $\mathscr{H}$.

In the following, we will only consider the second case and, taking into account the above remarks, suppose that ψ belongs to H. Equation (4) is still valid in H and then the eigenvalues of operator Γ restricted to H are strictly positive.

Remark. In the usual multivariate setting, similar situations in which the covariance matrix has null eigenvalues occur when one (or more) variable is a linear combination of the others. As a consequence, the design does not fill the whole space and only the part of the regressor that belongs to the span of the design variables can be identified. The usual procedures consist in selecting the 'good' explanatory variables such that the design matrix is a full rank matrix.

This is exactly what we are doing when considering an approximation of ψ in the space $H = \overline{Im(\Gamma)}$ instead of $\mathscr{H}$.

2.3. Test procedures

Suppose now that we have an i.i.d. sample $(X_i, Y_i), i = 1, \ldots, n$, drawn from (X, Y). The empirical covariance and cross-covariance operators are defined by

$$\Gamma_n x(t) = \frac{1}{n}\sum_{i=1}^{n}\langle X_i, x\rangle X_i(t), \quad x \in H, \;\; t \in [0, 1]$$

and

$$\Delta_n x = \frac{1}{n}\sum_{i=1}^{n}\langle X_i, x\rangle Y_i, \quad x \in H.$$

The eigenelements of Γ_n are denoted by $(\widehat{\lambda}_j, \widehat{V}_j), \; j = 1, \ldots, n$.

We will examine at first the case $\psi_0 = 0$ so that the null hypothesis is

$$H_0: \; \psi = 0.$$

Since the eigenvalues λ_j of the operator Γ restricted to H are non-null and that $\Delta = 0$, we have from relation (4), $\Psi\Gamma = 0$ and then

$$\Gamma\Psi^* = 0,$$

where Ψ^* is the adjoint operator of Ψ. We then have $\Psi^* = 0$ in H. Conversely, it is clear that under H_0 we have $\Delta = 0$. Then, testing $\Psi = 0$ is equivalent in that case to test

$$H_0: \; \Delta = 0. \tag{7}$$

In the following, we assume that the eigenvalues of Γ are distinct and we will build our test statistics on the above remark in that case and on the following theorems. Theorem 1 deals with the asymptotic behaviour of Δ_n under the null hypothesis. It is an application of the central limit theorem for independent Hilbert space valued random variables (see Grenander, 1963).

Theorem 1

Under the null hypothesis H_0, $\sqrt{n}\Delta_n$ converges in distribution to a gaussian centred random variable G_Δ with covariance operator $C = \sigma^2\Gamma$.

Theorem 2 is an application of the decomposition property for a gaussian random variable valued in H (see Grenander, 1963).

Theorem 2

The variable G_Δ admits the following decomposition:

$$G_\Delta = \sigma \sum_{j=1}^{+\infty} \eta_j \sqrt{\lambda_j} V_j, \; a.s.,$$

where $\eta_j, j \geq 1$ are i.i.d. centred real gaussian random variables with variance 1.

The natural way to test if the operator Δ is null is to look at its norm and then to base our test on $\|\sqrt{n}\Delta_n\|^2$, whose asymptotic distribution may be approximated by

$\|G_\Delta\|^2 = \sigma^2 \sum_{j=1}^{+\infty} \lambda_j \eta_j^2$ in view of Theorems 1 and 2. However, the distributions of such random variables are not explicitely known and these statistics depend on the parameters λ_j. For these reasons we introduce the operator A_n defined on H with matrix $\mathrm{diag}(1/\sqrt{\lambda_1}, \ldots, 1/\sqrt{\lambda_{p_n}}, 0, \ldots)$ in the basis $(V_j)_j$, that is to say $A_n(.) = \sum_{j=1}^{p_n} \lambda_j^{-1/2} \langle V_j, . \rangle V_j$, where (p_n) is a sequence of integers tending to infinity such that $p_n \leq n$. We also need the empirical counterpart of A_n denoted by $\widehat{A}_n(.) = \sum_{j=1}^{p_n} \hat{\lambda}_j^{-1/2} \langle \hat{V}_j, . \rangle \hat{V}_j$. Then, our first test statistic will be

$$D_n = \frac{1}{\hat{\sigma}_n^2} \left\| \sqrt{n} \Delta_n \hat{A}_n \right\|^2,$$

where $\hat{\sigma}^2$ is some estimator of the variance σ^2. It can be readily deduced from Theorems 1 and 2 that D_n follows approximatively, when n is large, a χ^2 distribution with p_n degrees of freedom. These kinds of approximations are widely used when testing statistical hypotheses with asymptotic gaussian distribution. Alternatively, one may consider the following test statistic:

$$T_n = \frac{1}{\sqrt{p_n}} \left(\frac{1}{\hat{\sigma}_n^2} \left\| \sqrt{n} \Delta_n \hat{A}_n \right\|^2 - p_n \right).$$

We will show in the appendix that under appropriate conditions on the sequence $(p_n)_n$ and on $\hat{\sigma}^2$ the test statistic T_n converges in distribution, under H_0, to a centred gaussian real random variable with variance 2. Then, the second test procedure is the following:

- Compute T_n for a sufficiently large n.
- Choose a level for the test, say α. Let q_α be the quantile of order α of a gaussian centred distribution with variance 1.
- If $|T_n| > \sqrt{2} q_\alpha$, reject H_0, otherwise accept it.

We will also prove in the appendix that this test is consistent (see Theorem 4). In Section 3, we will compare results for the two test statistics by means of a simulation study.

Now for the general case of testing $\psi = \psi_0$, where ψ_0 is some given function, we only need to consider the new real random variable $\widetilde{Y} = Y - \langle \psi_0, X \rangle$. The test is then performed as above, replacing Δ with the cross-covariance operator $\tilde{\Delta}$ of $(X, \tilde{Y})$.

3. A simulation study

In this section, we illustrate our test procedure on some simulated data. Let us first introduce the estimate $\hat{\sigma}^2$ of σ^2. This estimate is based on the residual sum of square. Indeed, let $\hat{\psi}$ be the penalized B-splines expansion estimator of the function ψ introduced in Cardot *et al.* (1999b). It is defined as

$$\hat{\psi} = \mathbf{B}'_{k,q} \hat{\boldsymbol{\theta}},$$

where $\hat{\boldsymbol{\theta}}$ is solution of the following minimization problem:

$$\min_{\boldsymbol{\theta} \in \mathbb{R}^{k+q}} \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \langle \mathbf{B}'_k \boldsymbol{\theta}, X_i \rangle \right)^2 + \rho \left\| \mathbf{B}_{k,q}^{('')'} \boldsymbol{\theta} \right\|^2, \tag{8}$$

where $\mathbf{B}_{k,q}$ (respectively, $\mathbf{B}_{k,q}^{('')}$) is the vector of B-splines (respectively, of second derivatives of B-splines) with degree q and k equispaced knots on $[0, 1]$ and ρ is some positive parameter of regularization. The value of this parameter ρ is chosen by minimizing a generalized cross-validation criterion (Cardot *et al.*, 1999b). The above estimator can be written as $\mathbf{SY}$ where $\mathbf{Y}$

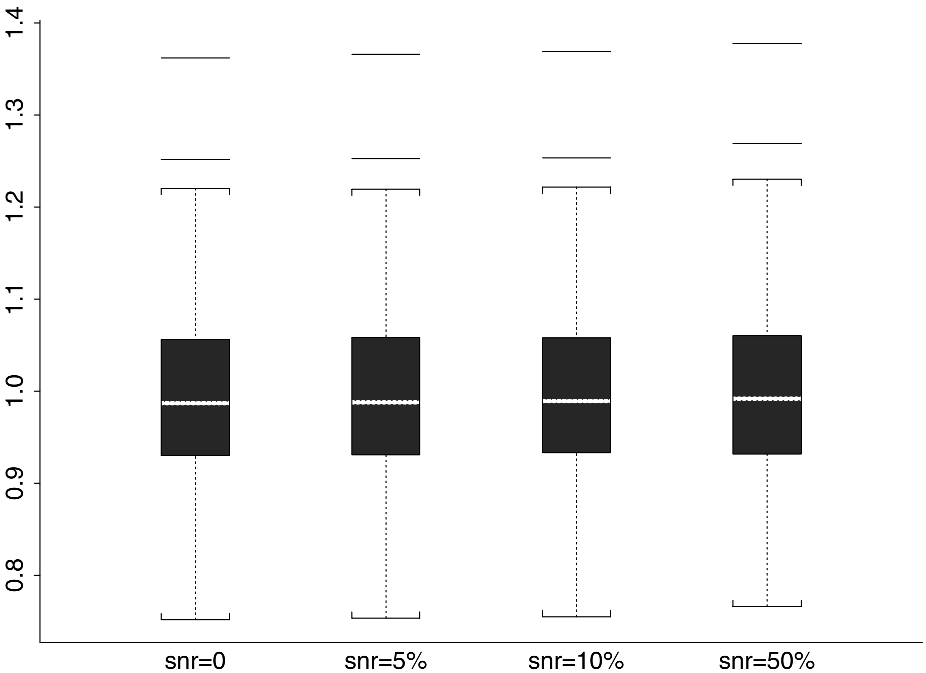

Fig. 1. Boxplot of $\hat{\sigma}^2/\sigma^2$ for different levels of signal-to-noise ratio and gaussian noise.

is the vector of Y_is and $\mathbf{S}$ is the 'hat' matrix. In analogy with the estimation of a variance in the classical linear regression model, we take as an estimate for σ^2

$$\hat{\sigma}^2 = \frac{1}{\mathrm{tr}(\mathbf{I}_n - \mathbf{S})} \sum_{i=1}^{n} \left(Y_i - \langle \mathbf{B}_k' \hat{\boldsymbol{\theta}}, X_i \rangle \right)^2,$$

where $\mathbf{I}_n$ is the identity matrix of size n. A similar kind of estimator of variance was introduced by Wahba (1983) in the setting of non-parametric regression. This estimator seems to be effective in approximating σ^2 in our Monte Carlo experiment as shown in Fig. 1 for gaussian noise (several different signal-to-noise ratios (*snr*) are defined). It also gives good estimates when the distribution of the noise is uniform but can give too large estimators when it follows a Student's t-distribution. This is a well-known consequence of the heavy tails of that distribution.

We have simulated $n_s = 800$ samples, each being composed of $n = 200$ independent realizations $(X_i, Y_i), i = 1, \ldots, n$, from model (6) in which $X(t)$ is a Brownian motion defined on $[0, 1]$, $\varepsilon = E[Y|X] - \Psi(X)$ has mean 0 and variance σ^2 and $\psi(t) = 0$ (Section 3.1) or $\psi(t) = \sin(2\pi t^3)^3$ (Section 3.2). The eigenelements of the covariance operator of X are known from Ash & Gardner (1975)

$$\lambda_j = \frac{1}{(j - 0.5)^2 \pi^2} \quad \text{and} \quad V_j(t) = \sqrt{2} \sin[(i - 0.5)\pi t], \;\; t \in [0, 1], \;\; j = 1, 2, \ldots$$

Thus, the eigenvalues are distinct and strictly positive, and then $H = \mathscr{H}$. Let us also notice that function $\psi(t) = \sin(2\pi t^3)^3$ cannot be decomposed as a finite combination of the basis functions $V_j(t)$ even if the approximation error:

$$\mathrm{Er}(q) = \left\| \psi - \sum_{j=1}^{q} \langle \psi, V_j \rangle V_j \right\|^2, \tag{9}$$

decreases rapidly as the number q of basis functions increase (see Fig. 2).

For simulation purposes, the Brownian random functions X_i and the function ψ were discretized by 100 design points equispaced in $[0, 1]$. The integral equation (6) is approximated by the trapezoidal method. The eigenelements of the covariance operator are also approximated by means of a quadrature rule. It allows us, in practical situations, to deal with non-equispaced design points (see, e.g. Castro *et al.*, 1986).

Let us define the snr as follows:

$$\mathrm{snr} = \frac{\mathrm{E}(\langle \psi, X \rangle^2)}{\mathrm{E}(\langle \psi, X \rangle^2) + \sigma^2},$$

and consider two different *snr*, 5 and 10 percent, when H_0 is false. These *snr* are controlled by means of the variance of the noise σ^2. Notice that when H_0 is true, *snr* equals 0.

3.1. Level of the test

The two test statistics D_n and T_n presented above are only based on an asymptotic study of their distribution. Thus, we have decided to compare their empirical levels with the nominal ones (i.e. the levels deduced from the asymptotic distribution). Then, we have compared the empirical distribution of the test statistics D_n with the appropriate χ^2 distribution and T_n with a $N(0, 2)$ distribution for different dimension values p_n, different

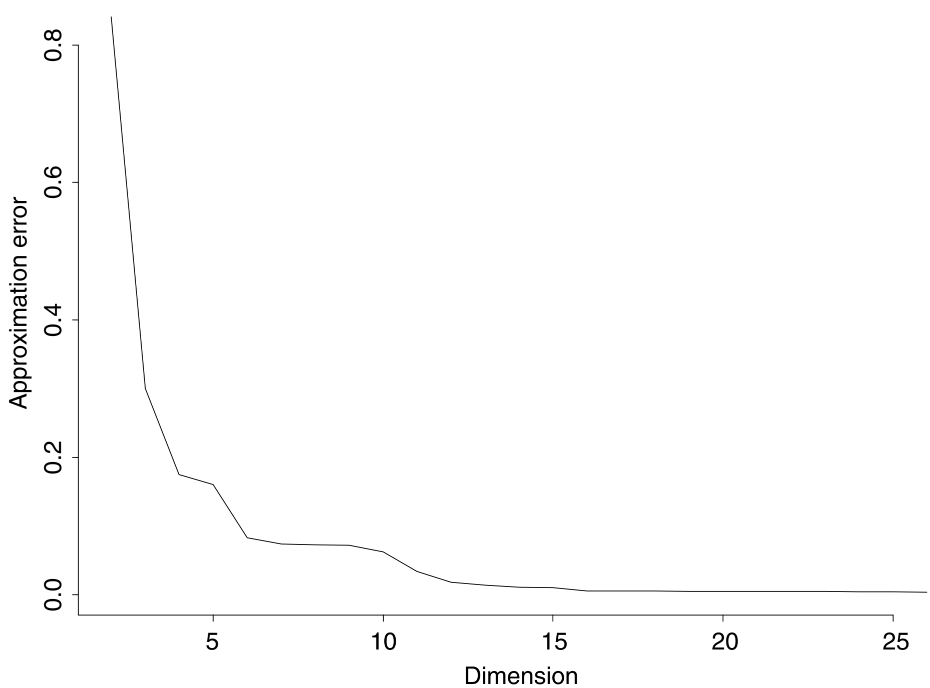

Fig. 2. Approximation error of function $\sin(2\pi t^3)^3$ for varying dimension q.

nominal levels and different distributions of the noise ε. Actually, the noise was generated with gaussian, uniform and Student's t-distributions. These results are shown in Table 1 for the gaussian noises, in Table 2 for the uniform noises and in Table 3 for the Student's t-noises.

In general, it seems that, for both test statistics, larger values of p_n lead to better approximations of the true nominal levels of the test. We noticed that the test based on T_n tends to underestimate the nominal levels which are large (20 and 10 percent) and overestimate the small ones (5 and 1 percent). On the other hand, the test based on D_n overestimates in a quasi-systematic way the true level of the tests. Finally, these two test statistics give rather good results for the different noise distributions, ensuring a certain amount of robustness for these statistics.

3.2. Power of the test

Then, we considered the model defined previously in which the function ψ is not null in order to evaluate and compare the power of both test statistics. This power function was

Table 1. *Comparison of the estimated levels (as percentage) for the two test statistics D_n and T_n and different dimension values of p_n when ε is* gaussian. *The nominal levels are based on the asymptotic distribution. The number of Monte Carlo experiments is $n_s = 800$ and each sample size is $n = 200$*

Nominal level (%)	D_n			T_n		
	$p_n = 5$	$p_n = 10$	$p_n = 20$	$p_n = 5$	$p_n = 10$	$p_n = 20$
20	18.1	21.7	20.0	15.0	18.5	18.8
10	11.4	12.1	10.6	8.5	9.4	8.5
5	6.0	6.1	5.2	6.0	5.3	4.6
1	1.6	0.9	1.5	3.5	2.2	2.0

Table 2. *Comparison of the estimated levels (as percentage) for the two test statistics D_n and T_n and different dimension values of p_n when noise distribution is* uniform. *The nominal levels are based on the asymptotic distribution. The number of Monte Carlo experiments is $n_s = 800$ and each sample size is $n = 200$*

Nominal level (%)	D_n			T_n		
	$p_n = 5$	$p_n = 10$	$p_n = 20$	$p_n = 5$	$p_n = 10$	$p_n = 20$
20	22.0	21.9	20.3	14.9	17.3	18.6
10	11.6	12.1	10.6	8.5	9.2	10.1
5	6.1	5.9	5.9	5.9	4.9	4.9
1	1.1	0.8	1.6	2.5	1.8	1.9

Table 3. *Comparison of the estimated levels (as percentage) for the two test statistics D_n and T_n and different dimension values of p_n when noise follows a* Student's *t-distribution with three degrees of freedom. The nominal levels are based on the asymptotic distribution. The number of Monte Carlo experiments is $n_s = 800$ and each sample size is $n = 200$*

Nominal level (%)	D_n			T_n		
	$p_n = 5$	$p_n = 10$	$p_n = 20$	$p_n = 5$	$p_n = 10$	$p_n = 20$
20	23.1	22.8	21.6	15.5	19.9	19.9
10	11.5	11.9	11.8	8.3	9.8	10.6
5	6.1	6.8	6.6	5.8	6.0	5.2
1	2.1	2.8	1.2	4.1	2.8	1.9

Table 4. *Comparison of the empirical power (as percentage) for the two test statistics D_n and T_n, different dimension values of p_n and different signal-to-noise ratios when the noise is* gaussian

snr (%)	Nominal level (%)	D_n $p_n = 5$	$p_n = 10$	$p_n = 20$	T_n $p_n = 5$	$p_n = 10$	$p_n = 20$
5	1	52.0	34.5	21.5	62.0	44.0	24.0
	5	73.0	63.0	42.0	72.0	62.5	38.0
	10	81.5	74.5	57.5	77.5	68.0	50.0
10	1	92.0	80.5	69.0	95.0	84.0	72.0
	5	96.0	93.0	86.5	96.0	92.5	85.5
	10	98.5	95.0	90.0	96.5	94.5	89.0

Table 5. *Comparison of the empirical power (as percentage) for the two test statistics D_n and T_n, different dimension values of p_n and different signal-to-noise ratios when the noise is* uniform

snr (%)	Nominal level (%)	D_n $p_n = 5$	$p_n = 10$	$p_n = 20$	T_n $p_n = 5$	$p_n = 10$	$p_n = 20$
5	1	48.8	37.9	22.8	61.0	44.8	25.4
	5	71.5	60.3	45.3	70.0	57.6	39.5
	10	80.0	70.3	58.9	74.4	64.2	50.5
10	1	89.4	80.0	65.8	93.6	84.5	69.5
	5	97.8	92.3	84.9	97.6	91.4	82.0
	10	99.0	96.3	90.9	98.4	94.1	86.5

Table 6. *Comparison of the empirical power (as percentage) for the two test statistics D_n and T_n, different dimension values of p_n and different signal-to-noise ratios when the noise has a* Student's *t distribution with three degrees of freedom*

snr (%)	Nominal level (%)	D_n $p_n = 5$	$p_n = 10$	$p_n = 20$	T_n $p_n = 5$	$p_n = 10$	$p_n = 20$
5	1	57.0	43.8	28.5	65.0	52.3	32.6
	5	73.4	64.0	51.8	73.6	62.5	48.2
	10	81.4	73.1	62.2	77.6	68.5	55.5
10	1	89.2	81.0	71.5	92.3	85.1	74.5
	5	95.9	91.4	84.5	95.6	90.7	82.9
	10	97.5	94.8	90.4	96.9	92.4	86.6

approximated empirically (see Tables 4–6) for different nominal levels of significance, levels of snr, dimensions p_n and distribution of the noise.

We notice that the power of the test decreases as the dimension p_n increases. Thus, tests are more effective to reject H_0 if it is false when p_n is small. Both tests give rather good results even if the variance of the signal is small compared to the variance of the noise. We also notice that for small nominal levels, the test based on T_n is more powerful. On the other hand, when nominal levels equal 5 or 10 percent then D_n should be preferred. However, this can be explained by the empirical study on the level in the previous section, which shows that the percentage of rejection is greater for T_n for small nominal levels and greater for D_n for nominal level greater than 5 percent. This leads us to think that the two tests have more or less the same power.

Nevertheless, an important matter is how the regression function ψ is expanded on the basis spanned by the eigenfunctions of Γ. More precisely, the null hypothesis may not be rejected with no real doubt whereas it is false just because the function ψ belongs to the orthogonal of the space spanned by the first eigenfunctions of Γ. Thus, we have to make a trade off between these two risks and this will lead us in practice to consider different values of p_n before deciding to accept or to reject the null hypothesis.

4. Concluding remarks

The test statistics proposed are based on the asymptotic distribution of the norm of the cross-covariance operator Δ and in view of the Monte Carlo experiment it is hard to say which test statistic (and then which approximation) is better. Actually, T_n seems to give empirical levels that are closer to the nominal ones when the levels are greater than 5 percent. Furthermore, for both statistics T_n and D_n, as p_n increases the power of the test decreases and, on the other hand, if p_n is too small then we may not accept H_0 whereas it is true if the projection of ψ onto the space spanned by the first eigenfunctions of Γ is null. Then, an important matter would be to have some knowledge on the asymptotic behaviour of the sequence (p_n). From a practical point of view, it should be noted that the determination of the sequence p_n is possible whenever additional assumptions are made about the rate of decay of the eigenvalues. This is done in a recent paper dealing with a functional test for the mean of a sample of random curves (see Mas, 2002).

Instead of using the approximations of Theorems 1–3, one may think of simulating the distribution of $\sqrt{n}\|\Delta_n\|^2$ under the null hypothesis by means of permutations that avoid the estimation of the eigenvalues of the covariance operator Γ as well as the estimation of the residual variance σ^2. This is the aim of a future work in preparation.

Among the possible extensions of the present work, one can test the belongingness of ψ to a given finite-dimensional space of $L^2_{[0,1]}$ using a pseudolikelihood ratio test statistic. One can also consider the test of a linear relationship against a non-parametric link between X and Y. Indeed, Ferraty & Vieu (2002) have introduced a kernel estimator of Ψ when it is not necessarily a linear operator. Then, a similar pseudolikelihood ratio test procedure as the one introduced by Azzalini & Bowman (1993) can be used in that case.

Acknowledgements

We gratefully thank the Editor and two referees for useful comments as well as D. Bosq and D. Tjøstheim for reading the manuscript in detail and making a number of valuable suggestions. We would also like to thank the members of the working group STAPH on Functional Statistics of the Department of Statistics of Université Paul Sabatier for helpful discussions.

References

Ash, R. B. & Gardner, M. F. (1975). *Topics in stochastic processes*. Academic Press, New York.

Azzalini, A. & Bowman, A. (1993). On the use of nonparametric regression for checking linear relationships. *J. Roy. Statist. Soc. Ser. B* **55**, 549–557.

Bosq, D. (1991). Modelization, non-parametric estimation and prediction for continuous time processes. In *Nonparametric functional estimation and related topics* (ed. G. Roussas). NATO, ASI Series, pp. 509–529. Kluwer academic, Dordrecht.

Cardot, H., Ferraty, F. & Sarda, P. (1999a). Functional linear model. *Statist. Probab. Lett.* **45**, 11–22.

Cardot, H., Ferraty, F. & Sarda, P. (1999b). Spline estimators for the functional linear model: consistency, applications and Splus implementation. Technical Report, Biometry and Artificial Intelligence, INRA Toulouse, France, 1-1999.

Castro, P., Lawton, W. & Sylvestre, E. (1986). Principal modes of variation for processes with continuous sample curves. *Technometrics* **28**, 329–337.
Dudley, R. M. (1968). Distances of probability measures and random variables. *Ann. Statist.* **39**, 1563–1572.
Ferraty, F. & Vieu, P. (2002). Functional nonparametric model and application to sprectrometric data. *Comput. Statist.,* to appear.
Frank, I. E. & Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics* **35**, 109–148.
Grenander, U. (1963). *Probabilities on algebraic structures*. Almquist and Wiksell, Stockholm.
Hastie, T. & Mallows, C. (1993). A discussion of 'A statistical view of some chemometrics regression tools' by I.E. Frank and J.H. Friedman. *Technometrics* **35**, 140–143.
Mas, A. (2002). Testing for the mean of random curves: from penalization to dimension selection. Preprint.
Marx, B. D. & Eilers P. H. (1999). Generalized linear regression on sampled signals and curves: a *P*-spline approach. *Technometrics* **41**, 1–13.
Ramsay, J. O. & Silverman, B. W. (1997). *Functional data analysis*. Springer-Verlag, Berlin.
Wahba, G. (1983). Bayesian 'confidence intervals' for the cross-validated smoothing spline. *J. Roy. Satist. Soc. Ser. B* **45**, 133–150.
Whitt, W. (1974). Preservation of rates of convergence under mappings. *Z. Wahr. verw. Gebiete* **29**, 39–44.

Received June 2000, in final form March 2002

Pascal Sarda, Laboratoire de Stastistique et Probabilités, Université Paul Sabatier, 31062 Toulouse cédex, France.
E-mail: pascal.sarda@lsp.ups-tlse.fr

Appendix 1: Asymptotic results

In order to prove the main results of this section we need the following assumptions:

$$(H.1) \qquad \frac{n\lambda_{p_n}^2}{\left(\sum_{j=1}^{p_n} a_j\right)^2} \underset{n\to\infty}{\longrightarrow} +\infty,$$

where

$$a_j = \begin{cases} 2\sqrt{2}/(\lambda_1 - \lambda_2), & j = 0, \\ 2\sqrt{2}/(\min(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1}), & j \neq 0; \end{cases}$$

$$(H.2) \qquad \sqrt{n}\left(\hat{\sigma}^2 - \sigma^2\right) \text{ is bounded in probability;}$$

$$(H.3) \qquad E\|X\|^4 \leq C_1 < +\infty.$$

Moreover, we will assume in the following that

$$\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_{p_n} > 0, \text{ a.s.}$$

Under the null hypothesis (H_0), we have the following main result.

Theorem 3
Assume that conditions $(H.2)$ *and* $(H.3)$ *hold. Then, there exists a sequence* (p_n) *satisfying* $(H.1)$ *and such that the test statistic* T_n *converges in distribution to a centred gaussian real random variable G with variance 2.*

We have moreover the following result.

Theorem 4
Assume that conditions (H.1)–(H.3) hold, then the test procedure based on T_n is consistent.

Appendix 2: Proof of Theorem 3

Let us introduce the variable T_n' defined as

$$T_n' = \frac{1}{\sqrt{p_n}}\left(\frac{1}{\sigma^2}\left\|\sqrt{n}\Delta_n\hat{A}_n\right\|^2 - p_n\right).$$

We have

$$\begin{aligned} T_n &= T_n - T_n' \\ &\quad + \frac{1}{\sigma^2\sqrt{p_n}}\left(\left\|\sqrt{n}\Delta_n\hat{A}_n\right\|^2 - \left\|\sqrt{n}\Delta_n A_n\right\|^2\right) \\ &\quad + \frac{1}{\sqrt{p_n}}\left(\frac{1}{\sigma^2}\left\|\sqrt{n}\Delta_n A_n\right\|^2 - p_n\right). \end{aligned} \tag{10}$$

The proof of Theorem 3 is based on the convergence in probability of the first two terms on the right-hand side of (10) and on the convergence in distribution to G of the third one. The proof of these three results are detailed in the following lemmas.

Lemma 1
Suppose that (H.1) and (H.2) hold. Then, if T_n' converges in distribution to some random variable T, T_n converges in distribution to T.

Proof. We have

$$\begin{aligned} T_n' - T_n &= \frac{1}{\sqrt{p_n}}\left\|\sqrt{n}\Delta_n\hat{A}_n\right\|^2\left(\frac{\hat{\sigma}^2 - \sigma^2}{\sigma^2\hat{\sigma}^2}\right) \\ &= \frac{1}{\sqrt{p_n}}\left(\frac{1}{\sigma^2}\left\|\sqrt{n}\Delta_n\hat{A}_n\right\|^2 - p_n\right)\left(\frac{\hat{\sigma}^2 - \sigma^2}{\hat{\sigma}^2}\right) + \sqrt{\frac{p_n}{n}}\sqrt{n}\left(\frac{\hat{\sigma}^2 - \sigma^2}{\hat{\sigma}^2}\right). \end{aligned}$$

Since $\sum_{j\geq 1}\lambda_j < +\infty$, we have for n sufficiently large $\lambda_{p_n} < p_n{}^{-1}$ and then

$$\sqrt{p_n n^{-1}} \;<\; \left(\lambda_{p_n}\sqrt{np_n}\right)^{-1}, \tag{11}$$

which gives us directly with $(H.1)$ and $(H.2)$, $T_n' - T_n$ converges in probability to 0 and then the result.

Lemma 2
Assume that (H.1) and (H.3) hold, then

$$\frac{1}{\sigma^2\sqrt{p_n}}\left(\left\|\sqrt{n}\Delta_n\hat{A}_n\right\|^2 - \left\|\sqrt{n}\Delta_n A_n\right\|^2\right) \underset{n\to\infty}{\longrightarrow} 0, \textit{ in probability}.$$

Proof. Let us define in H, the operator $\tilde{A}_n$ as $\tilde{A}_n(.) = \sum_{j=1}^{p_n}\hat{\lambda}_j^{-1}\langle V_j, .\rangle V_j$ and let Δ_n^* be the adjoint of Δ_n. Denoting by tr(.) the trace operator and by $\|.\|_\infty$ the supremum norm, we have

$$\frac{1}{\sigma^2\sqrt{p_n}}\left(\left\|\sqrt{n}\Delta_n\hat{A}_n\right\|^2-\left\|\sqrt{n}\Delta_nA_n\right\|^2\right)=\frac{1}{\sigma^2\sqrt{p_n}}\operatorname{tr}\left(n\Delta_n\Delta_n^*\left(\hat{A}_n^2-A_n^2\right)\right)$$
$$\leq\frac{1}{\sigma^2\sqrt{p_n}}\left\|\hat{A}_n^2-A_n^2\right\|_\infty\operatorname{tr}(n\Delta_n\Delta_n^*)$$
$$\leq\frac{\left\|\sqrt{n}\Delta_n\right\|^2}{\sigma^2\sqrt{p_n}}\left(\left\|\hat{A}_n^2-\tilde{A}_n^2\right\|_\infty+\left\|\tilde{A}_n^2-A_n^2\right\|_\infty\right).$$

We have using Lemma 3.1 in Bosq (1991)

$$\frac{\left\|\sqrt{n}\Delta_n\right\|^2}{\sigma^2\sqrt{p_n}}\left\|\hat{A}_n^2-\tilde{A}_n^2\right\|_\infty\leq\frac{2\left\|\sqrt{n}\Delta_n\right\|^2}{\sigma^2\hat{\lambda}_{p_n}\sqrt{p_n}}\left\|\Gamma-\Gamma_n\right\|_\infty\sum_{j=1}^{p_n}a_j,$$

whereas

$$\frac{\left\|\sqrt{n}\Delta_n\right\|^2}{\sigma^2\sqrt{p_n}}\left\|\tilde{A}_n^2-A_n^2\right\|_\infty\leq\frac{\left\|\sqrt{n}\Delta_n\right\|^2}{\sigma^2\hat{\lambda}_{p_n}\lambda_{p_n}\sqrt{p_n}}\sup_{1\leq j\leq p_n}\left|\hat{\lambda}_j-\lambda_j\right|$$
$$\leq\frac{\left\|\sqrt{n}\Delta_n\right\|^2}{\sigma^2\hat{\lambda}_{p_n}\lambda_{p_n}\sqrt{p_n}}\left\|\Gamma-\Gamma_n\right\|_\infty.$$

It is clear that under (H_0), $\left\|\sqrt{n}\Delta_n\right\|^2$ is bounded in probability. Let C_n be equal to $\sum_{j=1}^{p_n}a_j/(\hat{\lambda}_{p_n}\sqrt{p_n})$ and E_n be the space defined as

$$E_n=\left\{\lambda_{p_n}/2\leq\hat{\lambda}_{p_n}\leq 3\lambda_{p_n}/2\right\}.$$

Now, for $\epsilon>0$, we have

$$P\left(C_n\left\|\Gamma-\Gamma_n\right\|_\infty>\epsilon\right)\leq P\left(C_n\left\|\Gamma-\Gamma_n\right\|_\infty>\epsilon,E_n\right)+P(\bar{E}_n)$$
$$\leq\frac{4E\left\|\Gamma-\Gamma_n\right\|_\infty^2}{\lambda_{p_n}^2}\left(\frac{\left(\sum_{j=1}^{p_n}a_j\right)^2}{p_n\epsilon^2}+1\right)$$
$$\leq\frac{4E\left\|X\right\|_\infty^4}{n\lambda_{p_n}^2}\left(\frac{\left(\sum_{j=1}^{p_n}a_j\right)^2}{p_n\epsilon^2}+1\right).$$

Now, using assumptions $(H.1)$ and $(H.3)$, it is easy to get

$$\frac{\left\|\sqrt{n}\Delta_n\right\|^2}{\sigma^2\sqrt{p_n}}\left\|\hat{A}_n^2-\tilde{A}_n^2\right\|_\infty\underset{n\to\infty}{\longrightarrow}0\text{ in probability.}\tag{12}$$

With the same arguments we obtain

$$\frac{\left\|\sqrt{n}\Delta_n\right\|^2}{\sigma^2\sqrt{p_n}}\left\|\tilde{A}_n^2-A_n^2\right\|_\infty\underset{n\to\infty}{\longrightarrow}0\text{ in probability.}\tag{13}$$

The results (12) and (13) imply the proof of the lemma.

Lemma 3

There exists some increasing sequence $(p_n)_n$ satisfying $(H.1)$ such that

$$\frac{1}{\sqrt{p_n}}\left(\frac{\left\|\sqrt{n}\Delta_nA_n\right\|^2}{\sigma^2}-p_n\right)$$

converges in distribution to a centred Gaussian r.r.v. G with variance 2.

Proof. Let us note by π the Prohorov metric which induces the topology of weak convergence on the set of S-valued random variables where S is some separable metric space (see Dudley, 1968). Then, we have

$$\pi\left(\frac{1}{\sqrt{p_n}}\left(\frac{\|\sqrt{n}\Delta_n A_n\|^2}{\sigma^2}-p_n\right),G\right)\le \pi\left(\frac{1}{\sqrt{p_n}}\left(\frac{\|\sqrt{n}\Delta_n A_n\|^2}{\sigma^2}-p_n\right),\frac{1}{\sqrt{p_n}}\left(\frac{\|G_\Delta A_n\|^2}{\sigma^2}-p_n\right)\right)$$
$$+\,\pi\left(\frac{1}{\sqrt{p_n}}\left(\frac{\|G_\Delta A_n\|^2}{\sigma^2}-p_n\right),G\right).$$

Now to prove the lemma, it is enough to show that the two terms on the right-hand side of the above inequality tend to zero as n tends to infinity for some sequence $(p_n)_n$. We have by Theorems 1 and 2

$$\frac{1}{\sqrt{p_n}}\left(\frac{\|G_\Delta A_n\|^2}{\sigma^2}-p_n\right)=\frac{1}{\sqrt{p_n}}\sum_{j=1}^{p_n}\left(\eta_j^2-1\right),$$

so that we get directly by the central limit theorem

$$\pi\left(\frac{1}{\sqrt{p_n}}\left(\frac{\|G_\Delta A_n\|^2}{\sigma^2}-p_n\right),G\right)\underset{n\to\infty}{\longrightarrow}0.$$

Now,

$$\pi\left(\|\sqrt{n}\Delta_n A_n\|^2,\|G_\Delta A_n\|^2\right)=\pi\left(\mathrm{tr}\left(A_n(n\Delta_n^*\Delta_n)A_n\right),\mathrm{tr}\left(A_nG_\Delta^*G_\Delta A_n\right)\right)$$
$$=\pi\left(\Lambda_n(n\Delta_n^*\Delta_n),\Lambda_n(G_\Delta^*G_\Delta)\right),$$

where Λ_n is the continuous linear operator defined as $\Lambda_n T=\mathrm{tr}(A_nTA_n)$. Then, using the mapping Theorem 3.2 of Whitt (1974) and $\|\Lambda_n\|_\infty\le(1/\lambda_{p_n})$, we get

$$\pi\left(\frac{1}{\sqrt{p_n}}\left(\frac{\|\sqrt{n}\Delta_n A_n\|^2}{\sigma^2}-p_n\right),\frac{1}{\sqrt{p_n}}\left(\frac{\|G_\Delta A_n\|^2}{\sigma^2}-p_n\right)\right)$$
$$\le\frac{1}{\sigma^2\lambda_{p_n}\sqrt{p_n}}\pi(n\Delta_n^*\Delta_n,G_\Delta^*G_\Delta).$$

Now, using Theorem 1 and the continuous mapping theorem applied to the continuous mapping $h\Delta_n=\Delta_n^*\Delta_n$, we have

$$\pi(n\Delta_n^*\Delta_n,G_\Delta^*G_\Delta)=\delta_n\longrightarrow 0,$$

and one can find a sequence $(p_n)_n$ satisfying $(H.1)$ and such that

$$\frac{\delta_n}{\lambda_{p_n}\sqrt{p_n}}\longrightarrow 0. \tag{14}$$

Indeed, let us define for instance the sequence $(k_n)_n$ as

$$k_n=\max\left\{k:\frac{\delta_n}{\lambda_k}\le 1\right\},$$

and let $(p'_n)_n$ a sequence satisfying $(H.1)$. Now, taking for all n, p_n as

$$p_n=\min\{k_n,p'_n\},$$

it is easy to show that the sequence $(p_n)_n$ tends to infinity with for all n, $p_n\le n$ and satisfies $(H.1)$ and (14).

Appendix 3: Proof of Theorem 4

With Theorem 3 and condition (H.2), it is enough to show that, under the alternative hypothesis (H_1), we have for all ϵ

$$P\left(T_n' > \epsilon\right) \underset{n\to\infty}{\longrightarrow} 1.$$

We have

$$P\left(T_n' > \epsilon\right) = P\left(\frac{n}{\sigma^2 p_n}\left\|\Delta_n\hat{A}_n\right\|^2 > \frac{\epsilon}{\sqrt{p_n}} + 1\right).$$

We may clearly restrict ourselves to prove that

$$P\left(\frac{n}{p_n}\left\|\Delta_n\hat{A}_n\right\|^2 > 2\sigma^2\right) \underset{n\to\infty}{\longrightarrow} 1.$$

Now, since Δ is a non-null operator, it is possible to find j_0 such that $\Delta V_{j_0} \neq 0$. We then have

$$\begin{aligned}\left\|\Delta_n\hat{A}_n\right\|^2 &\geq \left(\Delta_n\hat{A}_n\hat{V}_{j_0}\right)^2\\ &= \frac{1}{\hat{\lambda}_{j_0}}\left(\Delta_n\hat{V}_{j_0}\right)^2.\end{aligned}$$

Now, the strong law of large numbers applied to Δ_n and the almost sure convergence of $\hat{V}_{j_0}$ and $\hat{\lambda}_{j_0}$ to V_{j_0} and λ_{j_0}, respectively (cf. Bosq, 1991), imply that

$$\frac{1}{\hat{\lambda}_{j_0}}\left(\Delta_n\hat{V}_{j_0}\right)^2 \underset{n\to\infty}{\longrightarrow} \frac{1}{\lambda_{j_0}}\left(\Delta_n V_{j_0}\right)^2, \text{ a.s.}$$

Then the result follows from (11) and (H.1) which imply that $n/p_n \longrightarrow +\infty$.