



HAL
open science

Testing for lack-of-fit in functional regression models against general alternatives

Valentin Patilea, César Sánchez-Sellero

► To cite this version:

Valentin Patilea, César Sánchez-Sellero. Testing for lack-of-fit in functional regression models against general alternatives. *Journal of Statistical Planning and Inference*, 2020, 209, pp.229 - 251. <10.1016/j.jspi.2020.04.002>. <hal-03490504>

HAL Id: hal-03490504

<https://hal.science/hal-03490504v1>

Submitted on 22 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

Testing for Lack-of-Fit in Functional Regression Models Against General Alternatives

Valentin Patilea^{a,*}, César Sánchez-Sellero^b

^a*Univ Rennes, Ensai, CNRS, CREST-UMR 9194, F-35000 Rennes, France*

^b*Departamento de Estadística, Análisis Matemático y Optimización, Universidade de Santiago de Compostela, Spain*

Abstract

A lack-of-fit test for functional regression models is proposed. The test is based on the fact that checking the no-effect of a functional covariate is equivalent to checking the nullity of the conditional expectation of the error term given a sufficiently rich set of projections of that covariate. The idea then is to search the projection that is, in some sense, the least favorable for the null hypothesis. Finally, it remains to perform a nonparametric check of the nullity of the conditional expectation of the residuals of the regression given the selected least favorable projection. For the search of a least favorable projection and the nonparametric check we use a kernel-based approach. As a result, the test statistic is a quadratic form based on univariate kernel smoothing and the asymptotic critical values are given by the standard normal law. The test is able to detect general departures from the model. The error term of the regression could present heteroscedasticity of unknown form. The law of the functional covariate need not be known. The test could be implemented quite easily and performs well in simulations and real data applications.

Keywords: functional data regression, kernel smoothing, lack-of-fit testing

1. Introduction

Recently, there has been a large amount of work on functional data analysis. The monographs of Ramsay and Silverman (2005) and Horváth and Kokoszka (2012) provide a comprehensive landscape of the importance of statistical methods for functional data. See also Wang, Chiou and Müller (2016) for a recent review. Estimation and prediction with functional covariates has received substantial attention: see for example Ferraty and Vieu (2006), Cai and Hall (2006), Hall and Horowitz (2007), Crambes, Kneip and Sarda (2009), Yao and Müller

*Corresponding author

Email addresses: valentin.patilea@ensai.fr (Valentin Patilea),
cesar.sanchez@usc.es (César Sánchez-Sellero)

(2010) and the references therein. The lack-of-fit problem remains less explored,
10 especially general regression model checks against general alternatives.

To illustrate the problem we address in this paper, let $L^2[0, 1]$ be the space
of the square-integrable real-valued functions defined on the unit interval. For
any $X_1, X_2 \in L^2[0, 1]$, let $\langle X_1, X_2 \rangle = \int_0^1 X_1(t)X_2(t)dt$. Consider a functional
linear model $Y = a + \langle b, X \rangle + U$, where $U \in \mathbb{R}$ and $X \in L^2[0, 1]$ are random,
15 while $a \in \mathbb{R}$ and $b \in L^2[0, 1]$ are the unknown parameters of the model. Testing
for lack-of-fit of this regression model against general alternatives means to test
the null hypothesis

$$H_0 : \mathbb{E}(U|X) = 0 \quad \text{almost surely (a.s.)} \quad (1.1)$$

against the nonparametric alternative $\mathbb{P}[\mathbb{E}(U|X) = 0] < 1$.

There is a large literature on regression model checks against nonparametric
20 alternatives when X is a finite-dimension vector, see for instance Härdle and
Mammen (1993), Stute (1997), Hart (1997), Horowitz and Spokoiny (2001),
Guerre and Lavergne (2005). See González-Manteiga and Crujeiras (2013) for
a recent review.

An increasing interest for extending the regression model check approaches
25 to functional data is evidenced lately. Delsol, Ferraty and Vieu (2011) extended
the idea of Härdle and Mammen (1993) to a functional data framework. How-
ever, their results seem of little practical use since they are derived under strong
assumptions, for instance on the so-called small ball probabilities. Bücher, Dette
and Wieczorek (2011) proposed a test based on a stochastic process estimating
30 L^2 -distances between the response and the regression function, but restricted
their investigation to the functional linear model. Some more work was done
for testing for lack-of-fit in a functional linear model, see Cardot *et al.* (2003),
Cardot, Goia and Sarda (2004), Hilgert, Mas and Verzelen (2013), or for test-
ing the functional linear model against quadratic alternatives, see Horvath and
35 Reeder (2013). Though these approaches could be quite effective in some cases,
by construction, they are not designed to detect general departures from the
null hypothesis (1.1).

The test we introduce herein is based on the following dimension reduction
idea: condition (1.1) is equivalent to the nullity of the conditional expectation
40 of U given a sufficiently rich set of projections of X on elements of norm 1 from
finite-dimension subspaces of $L^2[0, 1]$. This remark was used by Lavergne and
Patilea (2008) in a finite dimension framework. Patilea, Sánchez-Sellero and
Saumard (2016) used a similar dimension reduction idea. Next, the idea is to
search in finite-dimension subspaces of $L^2[0, 1]$ for a least favorable element of
45 norm 1 and to check the nullity of the conditional expectation of U given the
scalar product between X and the selected least favorable direction. Our test
is able to detect nonparametric alternatives, including polynomial ones. The
conditional variance of U given X need not be constant and the expression of
this conditional variance need not be known. We do not require the law of the
50 covariate X to be given or to be of a certain type, as for instance Gaussian.
A related approach was considered recently by Cuesta-Albertos *et al.* (2019).

Extending the approach of Stute (1997), they propose to draw randomly an element of $L^2[0, 1]$, to project X in the direction of this element and to check the nullity of the conditional expectation of U given this random projection
 55 using the functionals of a marked empirical process. With finite samples, the random draw is also realized in finite-dimension subspaces of $L^2[0, 1]$.

The paper is organized as follows. In section 2 we derive the fundamental lemma for the dimension reduction idea. In section 3 the new test is introduced in a simplified setting where U is observed and the no-effect of X on U is being
 60 tested. Our statistic is a quadratic form, based on *univariate* kernel smoothing, that behaves asymptotically like a *standard normal* random variable under H_0 . We prove that, under mild assumptions, the induced test is consistent against *any* type of fixed alternatives and against sequences of directional alternatives approaching the null hypothesis at a suitable rate. Moreover, a wild bootstrap
 65 procedure is proposed as a mean to approximate the critical values with finite samples and its asymptotic validity is proved. In section 4 we apply our projection-based approach for nonparametric checks of the functional regression models. For the sake of clarity, we focus on the functional linear model. We still obtain standard normal critical values and consistency against nonpara-
 70 metric alternatives, fixed or approaching the null hypothesis. In section 5 an extensive empirical study with simulated data is reported. Finally, the new test is applied to check the lack-of-fit of the functional linear model and the functional quadratic model for the Tecator data set. Some concluding remarks are gathered in section 6. The assumptions and the main proofs are relegated to
 75 the Appendix. The remaining proofs, completed by additional technical proofs and details, and additional simulation results are provided in a Supplementary Material.

2. Dimension reduction in nonparametric testing

Let $\mathcal{R} = \{\phi_1, \phi_2, \dots\}$ be an arbitrarily fixed orthonormal basis of the function space $L^2[0, 1]$. The predictor process X can be then expanded into $X(t) = \sum_{j=1}^{\infty} \langle X, \phi_j \rangle \phi_j(t)$. The idea is to reduce dimension to a fixed positive integer p , so $X_{[p]} \in L^2[0, 1]$ will be the projection of X on the subspace generated by the first p elements of the basis \mathcal{R} , that is

$$X_{[p]}(t) = \sum_{j=1}^p \langle X, \phi_j \rangle \phi_j(t).$$

We identify $X_{[p]}$ with the random vector $(\langle X, \phi_1 \rangle, \dots, \langle X, \phi_p \rangle)$ and for any non-
 80 random vector $\gamma = (\gamma_1, \dots, \gamma_p)$ we write $\langle X, \gamma \rangle = \sum_{j=1}^p \gamma_j \langle X, \phi_j \rangle$.

Our approach relies on the following lemma, an extension to Hilbert space-valued conditioning random variables of Lemma 2.1 of Lavergne and Patilea (2008) and Theorem 1 in Bierens (1990). This requires some definitions and notations. A set in \mathbb{R}^p will be called *negligible* if it has Lebesgue measure zero
 85 and is nowhere dense. Let $\mathcal{S}^p = \{\gamma \in \mathbb{R}^p : \|\gamma\| = 1\}$ denote the unit hypersphere

in \mathbb{R}^p . A subset S of \mathcal{S}^p will be called *negligible* if the set $\{b\gamma : |b| \leq 1, \gamma \in S\}$ is negligible in \mathbb{R}^p . For the sake of simplicity, for building the test statistic and deriving the theoretical properties, we could assume, without loss of generality, a mean-zero covariate (predictor), that is $\mathbb{E}[X(t)] = 0, \forall t \in [0, 1]$, or equivalently
90 $\mathbb{E}[\langle X, \phi_j \rangle] = 0, \forall j \geq 1$. Let $\|\cdot\|_{L^2}$ denote the norm associated with the scalar product in $L^2[0, 1]$.

Lemma 2.1. *Let $X \in L^2[0, 1]$ and $U \in \mathbb{R}$ be random variables. Assume that $\mathbb{E}|U| < \infty$.*

(A) *The following statements are equivalent:*

- 95 1. $\mathbb{E}(U | X) = 0$ a.s.
2. for any integer $p \geq 1$, $\mathbb{E}(U | X_{[p]}) = 0$ a.s.
3. for any integer $p \geq 1$, $\mathbb{E}(U | \langle X, \gamma \rangle) = 0$ a.s. $\forall \gamma \in \mathcal{S}^p$.

(B) *Suppose in addition that there exists some $s > 0$ such that*

$$\mathbb{E}(|U| \exp\{s\|X\|_{L^2}\}) < \infty. \quad (2.1)$$

If $\mathbb{P}[\mathbb{E}(U | X) = 0] < 1$, then there exists a positive integer $p_0 \geq 1$ such that for
100 any integer $p > p_0$, the set $\{\gamma \in \mathcal{S}^p : \mathbb{E}(U | \langle X, \gamma \rangle) = 0 \text{ a.s.}\}$ is negligible.

Point (A) is a cornerstone for proving the behavior of our test under the null and the alternative hypothesis. García-Portugués, González-Manteiga and Febrero-Bande (2014) used this property to build a goodness-of-fit test for the functional linear model in the spirit of Stute (1997). Point (B) shows that in
105 practice it is not difficult to find directions γ able to reveal the failure of the null hypothesis (1.1). Under the additional assumption (2.1) such directions represent almost all the points in \mathcal{S}^p , provided p is sufficiently large. Condition (2.1) is guaranteed if, for instance, $\mathbb{E}(|U|^c) < \infty$ for some $c > 1$ and $\mathbb{E}(\exp\{\varrho \|X\|_{L^2}\}) < \infty$ for $\varrho = sc/(c-1)$. The latter condition is met if, for
110 instance, X is a mean-zero Gaussian process with $\sup_{t \in [0, 1]} \mathbb{E}[X^2(t)] < \infty$; see chapter A.2 in van der Vaart and Wellner (1996). Moreover, in general, moment restrictions on X are not restrictive for goodness-of-fit testing purposes. Indeed, if X does not satisfy (2.1), it suffices to transform X into some variable $W \in L^2[0, 1]$ such that W generates the same σ -field and W satisfies condition
115 (2.1).

The following reformulation of H_0 is a consequence of Lemma 2.1.

Corollary 2.2. *Consider a real-valued variable U such that condition (2.1) holds true. For any $p \geq 1$, let $B_p \subset \mathcal{S}^p$ be a set that is not negligible. For any $\gamma \in B_p$, let $t \mapsto w_p(\gamma, t) \in \mathbb{R}, t \in \mathbb{R}$, be a function such that $w_p(\gamma, \langle X, \gamma \rangle) > 0$.
120 The following statements are equivalent:*

1. the null hypothesis (1.1) holds true;
2. for any $p \geq 1$, $\max_{\gamma \in B_p} \mathbb{E}[U \mathbb{E}(U | \langle X, \gamma \rangle) w_p(\gamma, \langle X, \gamma \rangle)] = 0$.

3. Testing the effect of a functional covariate

We introduce a general approach for nonparametric testing of the effect of a functional covariate X on a real-valued outcome U satisfying $\mathbb{E}(U) = 0$. The variable U could be, for instance, the error term of a mean regression model, which means that our framework includes model check problems for a large panel of regression models. Hence the problem considered in this section is the test of the null hypothesis defined in equation (1.1).

Our approach is based on Corollary 2.2-(2) and *univariate* kernel smoothing. Hence we avoid the problem of smoothing in infinite-dimension; in particular we avoid the small ball function required in the kernel regression with functional covariates, see Ferraty and Vieu (2006), Delsol, Ferraty and Vieu (2011).

For any $p \geq 1$, let $B_p \subset \mathcal{S}^p$ be a set that is not negligible in \mathcal{S}^p . To avoid handling denominators close to zero, we set the weight function $\omega(\gamma, \cdot)$ in Corollary 2.2 equal to the density of $\langle X, \gamma \rangle$, denoted by $f_\gamma(\cdot)$, which is assumed to exist $\forall \gamma \in B_p$. For any $\gamma \in B_p$, let

$$Q(\gamma) = \mathbb{E}\{U \mathbb{E}[U \mid \langle X, \gamma \rangle] f_\gamma(\langle X, \gamma \rangle)\} = \mathbb{E}\{\mathbb{E}^2[U \mid \langle X, \gamma \rangle] f_\gamma(\langle X, \gamma \rangle)\}.$$

By Corollary 2.2, the null hypothesis (1.1) holds true if and only if

$$\forall p \geq 1, \quad \max_{\gamma \in B_p} Q(\gamma) = 0. \quad (3.1)$$

3.1. The test statistic

In view of equation (3.1), the goal is to estimate $Q(\gamma)$. Given an independent sample $(U_1, X_1), \dots, (U_n, X_n)$ from (U, X) , let

$$Q_n(\gamma) = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} U_i U_j \frac{1}{h} K_h(\langle X_i - X_j, \gamma \rangle), \quad \gamma \in \mathcal{S}^p,$$

where $K_h(\cdot) = K(\cdot/h)$, where $K(\cdot)$ is a kernel and h a bandwidth. In the case of finite dimension covariates, $Q_n(\gamma)$ is the statistic considered by Lavergne and Patilea (2008). For a fixed p , $Q_n(\gamma)$ is expected to be close to $Q(\gamma)$ uniformly in γ . One of our results shows that this uniform closeness remains valid when p grows with n . On the other hand, Lemma 2.1-(B) indicates that when H_0 fails, the maximum of $Q(\gamma)$ over γ is positive, provided p is large enough. A natural idea then would be to build a test statistic using the maximum of $Q_n(\gamma)$ with respect to γ . However, under H_0 , one expects $Q_n(\gamma)$ to be close to zero for any p and γ and thus the objective function of the maximization problem to be flat. To avoid this problem, we will choose a direction γ as the least favorable direction for the null hypothesis H_0 using a penalized criterion based on a standardized version of $Q_n(\gamma)$. Lavergne and Patilea (2008) considered this idea with finite dimension covariates using $Q_n(\gamma)$, see also Bierens (1990).

Let us fix some $b_0 = (b_{01}, b_{02}, \dots) \in \mathbb{R}^\infty$, with $\sum_{j=1}^\infty b_{0j}^2 < \infty$, that could be interpreted as an initial 'guess' of an unfavorable direction for H_0 . For any $p \geq 1$ such that $\sum_{j=1}^p b_{0j}^2 > 0$, let

$$\gamma_0^{(p)} = (b_{01}, \dots, b_{0p}) / \|(b_{01}, \dots, b_{0p})\|.$$

Let $\widehat{v}_n^2(\cdot)$ be an estimate of the variance of $nh^{1/2}Q_n(\cdot)$. Given $B_p \subset \mathcal{S}^p$ that
 150 contains $\gamma_0^{(p)}$, the least favorable direction γ for H_0 is defined as

$$\widehat{\gamma}_n = \arg \max_{\gamma \in B_p} \left[nh^{1/2}Q_n(\gamma)/\widehat{v}_n(\gamma) - \alpha_n \mathbb{I}_{\{\gamma \neq \gamma_0^{(p)}\}} \right], \quad (3.2)$$

where \mathbb{I}_A is the indicator function of a set A , and α_n , $n \geq 1$ is a sequence of
 positive numbers increasing to infinity. The rate of α_n will be made explicit
 below; it will depend on the sample size and the rates of h and p . The role
 of $\gamma_0^{(p)}$ will be explained in more detail in section 3.5.2. Let us note that the
 155 definition of $\widehat{\gamma}_n$ requires only finite dimension optimization.

The test statistic we consider is

$$T_n = nh^{1/2} \frac{Q_n(\widehat{\gamma}_n)}{\widehat{v}_n(\widehat{\gamma}_n)}. \quad (3.3)$$

We will prove that, with suitable rates of increase for α_n and p and decrease
 for h , the probability of the event $\{\widehat{\gamma}_n = \gamma_0^{(p)}\}$ tends to 1 under H_0 . Hence T_n
 behaves asymptotically like $nh^{1/2}Q_n(\gamma_0^{(p)})/\widehat{v}_n(\gamma_0^{(p)})$. If p were fixed, T_n would be
 160 asymptotically standard normal distributed under H_0 ; see for instance Guerre
 and Lavergne (2005). We will show that the asymptotic normality is preserved
 when p grows at a suitable rate. In particular this means that the choice of $\gamma_0^{(p)}$
 does not affect the asymptotic critical values. This choice will also not affect the
 consistency. In practice the choice of $\gamma_0^{(p)}$ could be related to prior information
 165 of the practitioner on a class of alternatives. See section 3.5 for a discussion on
 the implementation aspects.

3.2. Variance estimation

To find the direction $\widehat{\gamma}_n$ and to build the test statistics (3.3), an estimate
 of the variance of $nh^{1/2}Q_n(\gamma)$ is needed. The approach that is expected not to
 170 inflate the variance estimate under the alternatives and thus to guarantee better
 power would involve an estimate of the conditional variance of $nh^{1/2}Q_n(\gamma)$ given
 X_i 's which can be written as

$$\tau_n^2(\gamma) = \frac{2}{n(n-1)h} \sum_{j \neq i} \sigma_p^2(X_i) \sigma_p^2(X_j) K_h^2(\langle X_i - X_j, \gamma \rangle), \quad \gamma \in B_p, \quad (3.4)$$

where $\sigma_p^2(X) = \text{Var}[U | X_{[p]}]$. An estimator of $\tau_n^2(\gamma)$ can easily be obtained
 from an estimate of $\sigma_p^2(\cdot)$. In theory, a good solution would be a nonparametric
 175 estimate of the p -variate function $\sigma_p^2(\cdot)$, but this is of little practical value
 when p grows with n . A simplistic solution could be to replace $\sigma_p^2(X_i)$ by U_i^2
 and define

$$\widehat{\tau}_n^2(\gamma) = \frac{2}{n(n-1)h} \sum_{j \neq i} U_i^2 U_j^2 K_h^2(\langle X_i - X_j, \gamma \rangle). \quad (3.5)$$

Since we expect $\mathbb{P}(\widehat{\gamma}_n = \gamma_0^{(p)}) \rightarrow 1$ under H_0 , a first simple variance estimator we propose is

$$\widehat{v}_n^2(\gamma) = \widehat{v}_n^2(\gamma, \gamma_0^{(p)}) = \left(\widehat{\tau}_n^2(\gamma) + \widehat{\tau}_n^2(\gamma_0^{(p)}) \right) / 2. \quad (3.6)$$

Using $\gamma_0^{(p)}$ in the convex combination will simplify the control of the small values of $\widehat{v}_n^2(\gamma)$, in particular in the proof of the second part of Lemma 3.1.

The simplistic replacement of $\sigma_p^2(X_i)$ by U_i^2 inflates the variance estimator $\widehat{v}_n^2(\widehat{\gamma}_n)$ under the alternative hypothesis, and thus deteriorates the power of the test. To propose a compromise between using a nonparametric estimation of $\sigma_p^2(\cdot)$ or the squared U_i , let $\sigma_\gamma^2(\cdot) = \text{Var}(U \mid \langle X, \gamma \rangle = \cdot)$. Clearly, we have $\mathbb{E}(U^2 \mid \langle X, \gamma \rangle) \geq \text{Var}(U \mid \langle X, \gamma \rangle)$ and the inequality is strict under general alternative hypotheses. Hence a compromise for estimating the variance of $nh^{1/2}Q_n(\gamma)$ could be

$$\widehat{v}_n^2(\gamma) = \frac{2}{n(n-1)h} \sum_{j \neq i} \widehat{\sigma}_\gamma^2(\langle X_i, \gamma \rangle) \widehat{\sigma}_\gamma^2(\langle X_j, \gamma \rangle) K_h^2(\langle X_i - X_j, \gamma \rangle), \quad (3.7)$$

where $\widehat{\sigma}_\gamma^2(\cdot)$ is a nonparametric estimate of the univariate function $\sigma_\gamma^2(t) = \text{Var}(U \mid \langle X, \gamma \rangle = t)$, satisfying the conditions

$$\sup_{1 \leq i \leq n} \sup_{\gamma \in B_p} \widehat{\sigma}_\gamma^{-2}(\langle X_i, \gamma \rangle) = O_{\mathbb{P}}(1) \quad (3.8)$$

and there exists a sequence of sets I_n , $n \geq 1$, on the real line such that

$$\mathbb{P}\left(\langle X, \gamma_0^{(p)} \rangle \in I_n\right) \rightarrow 1, \quad \sup_{1 \leq i \leq n} \left| \frac{\widehat{\sigma}_{\gamma_0^{(p)}}^2(\langle X_i, \gamma_0^{(p)} \rangle)}{\sigma_{\gamma_0^{(p)}}^2(\langle X_i, \gamma_0^{(p)} \rangle)} - 1 \right| \mathbb{I}_{\{\langle X_i, \gamma_0^{(p)} \rangle \in I_n\}} = o_{\mathbb{P}}(1), \quad (3.9)$$

when p increases with the sample size. We will prove that both definitions, (3.6) and (3.7), for a variance estimator, guarantee the standard normal asymptotic critical values and the consistency of our test. In the Supplementary Material we provide an example of estimator satisfying conditions (3.8) and (3.9). In simulations, the variance estimator (3.7) produced a more powerful test.

3.3. Behavior under the null hypothesis

The first step is the study of the behavior of the process $Q_n(\gamma)/\widehat{v}_n(\gamma)$, $\gamma \in B_p$, under H_0 when p is allowed to increase with the sample size.

Lemma 3.1. *Under Assumptions D and K and if H_0 holds true,*

$$\sup_{\gamma \in B_p \subset S^p} |Q_n(\gamma)| = O_{\mathbb{P}}(n^{-1}h^{-1/2}p \ln n) \quad \text{and} \quad \sup_{\gamma \in B_p \subset S^p} [1/\widehat{v}_n^2(\gamma)] = O_{\mathbb{P}}(1),$$

with $\widehat{v}_n^2(\gamma)$ defined in (3.6), or with $\widehat{v}_n^2(\gamma)$ defined in (3.7) provided that the condition (3.8) also holds true.

Let us now describe the behavior of $\hat{\gamma}_n$ under H_0 .

Lemma 3.2. *Under Assumptions D, K, and condition (3.8) if the variance estimator $\hat{v}_n^2(\gamma)$ is the one defined in (3.7), for a positive sequence α_n , $n \geq 1$ such that $\alpha_n/\{p \ln n\} \rightarrow \infty$, we have $\mathbb{P}(\hat{\gamma}_n = \gamma_0^{(p)}) \rightarrow 1$ under H_0 .*

Under H_0 , α_n has to grow to infinity sufficiently fast to render the probability of the event $\{\hat{\gamma}_n = \gamma_0^{(p)}\}$ close to 1. We will see below that, for better detection of alternative hypotheses, α_n should grow as slow as possible. Indeed, smaller α_n will allow directions $\hat{\gamma}_n$ to be selected, that could be better suited than $\gamma_0^{(p)}$ for revealing the departure from H_0 . The rate of p is also involved in the search for a trade-off: larger p renders slower, the rate of uniform convergence to zero of $Q_n(\gamma)$, $\gamma \in B_p$, and hence requires larger α_n .

The following result provides the asymptotic critical values of our test.

Theorem 3.3. *Consider the conditions of Lemma 3.2, completed by the condition (3.9) if the variance estimator $\hat{v}_n^2(\gamma)$ is that defined in (3.7). If the hypothesis H_0 defined in (1.1) holds true, the statistic T_n converges in law to a standard normal. Consequently, the test given by $\mathbb{I}(T_n \geq z_{1-a})$, with z_a the $(1-a)$ -quantile of the standard normal distribution, has asymptotic level a .*

3.4. The behavior under the alternative hypotheses

First let us give a simple argument showing that by construction the test should be consistent. Consider $H_1 : \mathbb{P}[\mathbb{E}(U | X) = 0] < 1$. We can write

$$T_n \geq \max_{\gamma \in B_p} \frac{nh^{1/2}Q_n(\gamma)}{\hat{v}_n(\gamma)} - \alpha_n \geq \frac{nh^{1/2}Q_n(\tilde{\gamma})}{\hat{v}_n(\tilde{\gamma})} - \alpha_n, \quad \forall \tilde{\gamma} \in B_p \subset \mathcal{S}^p, \quad (3.10)$$

with $\hat{v}_n^2(\gamma)$ any of the variance estimators defined in (3.6) or (3.7). Since $\mathbb{E}(U^2 | \langle X, \gamma_0^{(p)} \rangle) \geq \text{Var}(U | \langle X, \gamma_0^{(p)} \rangle) \geq \underline{\sigma}^2$, it is clear that $\sup_{\gamma} \{1/\hat{v}_n(\gamma)\} = O_{\mathbb{P}}(1)$ for both variance estimates introduced above. On the other hand, from Lemma 2.1, there exists p_0 and $\tilde{\gamma} \in B_{p_0}$ such that the expectation of $Q_n(\tilde{\gamma})$ stays away from zero as the sample size grows to infinity and h decreases to zero. Finally, for any $p > p_0$ and any n and h , clearly $\max_{\gamma \in B_p} Q_n(\gamma) \geq Q_n(\tilde{\gamma})$, because $B_{p_0} \times 0_{p-p_0} \subset B_p$. All these facts show why our test is consistent against a nonparametric alternative like H_1 , provided $p \rightarrow \infty$.

To formalize the consistency result, let us fix some real-valued function $\delta(X)$ and a sequence of real numbers r_n that could decrease to zero, the case $r_n \equiv 1$ being also allowed. Consider the sequence of alternative hypotheses

$$H_{1n} : U = U^0 + r_n \delta(X), \quad n \geq 1, \quad \text{with } \mathbb{E}(U^0 | X) = 0. \quad (3.11)$$

Theorem 3.4. *Suppose that Assumption D holds true with U replaced by U^0 , Assumption K is satisfied, and so is the condition (3.8) if the variance estimator is the one defined in (3.7). Moreover,*

- (a) r_n , $n \geq 1$ is such that $r_n^2 nh^{1/2}/\alpha_n \rightarrow \infty$;

(b) $\mathbb{E}[\delta(X)] = 0$ and $0 < \mathbb{E}[\delta^4(X)] < \infty$.

The test based on T_n is then consistent against the sequence of alternatives H_{1n} if there exists $p \geq 1$ and $\tilde{\gamma} \in B_p$ such that $\mathbb{P}(\mathbb{E}[\delta(X) \mid \langle X, \tilde{\gamma} \rangle] = 0) < 1$ and one
240 of the following conditions is satisfied:

1. the density $f_{\tilde{\gamma}}$ is bounded;
2. the function $\mathbb{E}[\delta(X) \mid \langle X, \tilde{\gamma} \rangle = \cdot] f_{\tilde{\gamma}}(\cdot)$ is bounded;
3. the Fourier transform of $\mathbb{E}[\delta(X) \mid \langle X, \tilde{\gamma} \rangle = \cdot] f_{\tilde{\gamma}}(\cdot)$ is integrable on \mathbb{R} .

Theorem 3.4 states that directional alternatives H_{1n} can be detected as soon
245 as $r_n^2 n h^{1/2} / \alpha_n$ tends to infinity. This is exactly the same condition as in the finite dimension covariate case; see Lavergne and Patilea (2008). However, in the functional data framework, to obtain the standard normal critical values, we need $1/\alpha_n = o(p^{-1} \ln^{-1} n)$. Hence, the rate r_n at which the alternatives H_{1n} tend to the null hypothesis, should satisfy $r_n^2 n h^{1/2} / \{p \ln n\} \rightarrow \infty$. Finally, let us
250 recall that the existence of p and $\tilde{\gamma} \in B_p$ such that $\mathbb{P}[\mathbb{E}[\delta(X) \mid \langle X, \tilde{\gamma} \rangle] = 0] < 1$ is guaranteed by Lemma 2.1.

Finally, let us point out that for achieving consistency, it is clearly not necessary $\gamma_0^{(p)}$ to satisfy the condition $\mathbb{P}[\mathbb{E}[\delta(X) \mid \langle X, \gamma_0^{(p)} \rangle] = 0] < 1$. Indeed, under the alternative hypotheses, our procedure will automatically search for
255 a least favorable $\hat{\gamma}_n$ direction close to some $\tilde{\gamma}$ as in Theorem 3.4. However, a careful inspection of the proof of Theorem 3.4 reveals that in the case where the alternative hypotheses are given by a function $\delta(X) = d(\langle X, \gamma_0^{(p)} \rangle)$, with some univariate function $d(\cdot)$, our test could show even higher power in finite samples. Indeed, in such a case the penalty will likely induce a least favorable
260 $\hat{\gamma}_n$ close or equal to $\gamma_0^{(p)}$ and, in general, this will produce a test statistic with a larger positive mean and thus a higher probability of rejection.

3.5. Implementation aspects

Our lack-of-fit test requires the choice of several quantities: the basis \mathcal{R} in the space of X , the order p , the privileged direction $\gamma_0^{(p)}$ and the set B_p , the
265 penalty α_n and the bandwidth h . The theoretical results are based on mild restrictions on these quantities that allow for a large range of choices. However, suitable choices of these quantities will help the user to perform a powerful test with an accurate level. Let us provide some guidelines on how these quantities could be chosen in practice in order to achieve this. Simulations given on Section
270 5.2 and in the Supplementary Material are also helpful for a convenient choice of these quantities.

3.5.1. The choice of p

The theoretical results above are derived for a fixed basis \mathcal{R} and the assumptions that were used impose only very mild conditions on this basis. For the
275 implementation of the test, the statistician would likely prefer a basis allowing an accurate low-dimension representation of the covariate to be achieved, and hence small values of p .

A widely used basis is given by the eigenfunctions of the covariance operator \mathcal{K} that is defined by $(\mathcal{K}v)(t) = \int \mathcal{K}(t, s)v(s)ds$, $v \in L^2[0, 1]$, where X is supposed to satisfy the condition $\int \mathbb{E}(X^2(t))dt < \infty$ and $\mathcal{K}(t, s) = \mathbb{E}[\{X(t) - \mathbb{E}(X(t))\}\{X(s) - \mathbb{E}(X(s))\}]$ is supposed positive definite. Let $\theta_1 \geq \theta_2 \geq \dots$ denote the ordered eigenvalues of \mathcal{K} and let $\mathcal{R} = \{\phi_1, \phi_2, \dots\}$ be the corresponding basis of eigenfunctions that are usually called the functional principal components (FPC). The FPC represent the orthonormal basis of the Karhunen-Loève decomposition of X and provide optimal (with respect to the mean-squared error) low-dimension representations of X . See, for instance, Ramsay and Silverman (2005). In most of the applications the FPC are *unknown and are estimated* from

$$(\widehat{\mathcal{K}}v)(t) = \int_{[0,1]} \widehat{\mathcal{K}}(t, s)v(s)ds, \quad t \in [0, 1],$$

where

$$\widehat{\mathcal{K}}(t, s) = n^{-1} \sum_{i=1}^n \{X_i(t) - \bar{X}_n(t)\}\{X_i(s) - \bar{X}_n(s)\}$$

and $\bar{X}_n(t) = n^{-1} \sum_{i=1}^n X_i(t)$. Let $\hat{\theta}_1 \geq \hat{\theta}_2 \geq \dots \geq 0$ denote the eigenvalues of $\widehat{\mathcal{K}}$ and let $\hat{\phi}_1, \hat{\phi}_2, \dots$ be the corresponding basis of eigenfunctions, that is the estimated FPC. We adopt the usual identification condition and we suppose that for any j , $\langle \phi_j, \hat{\phi}_j \rangle \geq 0$. For any $\gamma = (\gamma_1, \dots, \gamma_p) \in \mathcal{S}^p$ let us define

$$\langle X_i, \gamma \rangle_n = \sum_{k=1}^p \gamma_k \int_{[0,1]} X_i(t) \hat{\phi}_k(t) dt.$$

Let \widehat{T}_n be the statistic defined in equations (3.2) and (3.3) after replacing the $\langle X_i, \gamma \rangle$ by the estimated versions $\langle X_i, \gamma \rangle_n$.

280 **Corollary 3.5.** *Let Assumptions D-(a,b) and K-(a,c) hold true. Moreover, $\mathcal{K}(t, s)$ is positive definite and there exist $C, \eta > 0$ such that $\theta_j - \theta_{j+1} \geq Cj^{-\eta}$, $\forall j \geq 1$. Assume that one of the following conditions is met:*

1. $nh^4/p^{2\eta+1} \ln^2 n \rightarrow \infty$;
2. $nh^2/p^{2\eta+1} \ln^2 n \rightarrow \infty$ and $\sup_p \sup_{\gamma \in B_p} f_\gamma < C$ for some constant C ;
- 285 3. $nh^{1+\alpha} \rightarrow \infty$ for some $\alpha \in (0, 1)$ and $\forall \varepsilon > 0, \exists \delta > 0$ such that $\forall t, s \in \mathbb{R}$, $|t - s| < \delta$ implies $\sup_p \sup_{\gamma \in B_p} |f_\gamma(t) - f_\gamma(s)| < \varepsilon$.

Then Theorems 3.3 and 3.4 remain valid with T_n replaced by \widehat{T}_n .

Conditions 1 to 3 in Corollary 3.5 represent three possible trade-offs between the range for h and regularity conditions on X and B_p . Condition 3 is met for
290 instance for a Gaussian covariate and $B_p = \{(\gamma_1, \dots, \gamma_p) : |\gamma_1| \geq c\}$ for some constant $c > 0$.

For consistency against general alternatives, p needs to grow to infinity. Meanwhile, larger p makes the optimization over B_p more difficult. Using the

FPC basis could be a good compromise to detect general alternatives with small p . Then, if $\theta_j \sim j^{-\lambda-1}$ for some $\lambda > 0$, an automatic choice for p could be given by $\min\{p : \sup_{j \geq p} \theta_j \leq C \ln^{-1} n\}$ for some constant C . This would result in a logarithmic rate for p . In practice θ_j should be replaced by the estimates $\widehat{\theta}_j$, but the rate of p will not change. Indeed, on one hand, $\sup_{j \geq 1} |\widehat{\theta}_j - \theta_j|$ is bounded by the Hilbert-Schmidt norm of $\widehat{\mathcal{K}} - \mathcal{K}$. See, for instance Cai and Hall (2006). On the other hand, this norm is of order $O_{\mathbb{P}}(n^{-1/2})$ under mild conditions. See, for instance, Horváth and Kokoszka, P. (2012), Theorem 2.5. Finally, it is easy to see that $\min\{p : \sup_{j \geq p} \theta_j \leq C \ln^{-1} n\}$ and $\min\{p : \sup_{j \geq p} \widehat{\theta}_j \leq C \ln^{-1} n\}$ have the same rate. In practice simple empirical rules work as well. For instance p could be the smallest value such that more than some fixed high percentage, say 90%, of the variance within the sample, is captured by the first p principal components.

3.5.2. The choice of $\gamma_0^{(p)}$ and the set of directions B_p

We have proven that the choice of the privileged direction $\gamma_0^{(p)}$ is irrelevant for the theory of the test. This privileged direction has a twofold purpose. On one hand, whatever the choice, it allows us to derive a test statistic with a simple standard normal asymptotic behavior under the null hypothesis. Meanwhile the test remains consistent. The privileged direction also has a practical purpose. In applications the practitioner may use $\gamma_0^{(p)}$ to emphasize different aspects of the model check, for instance to make the test powerful against specific alternatives. See also our comments following Theorem 3.4 above. Without such prior preference or information, $\gamma_0^{(p)} = (1, \dots, 1)/\sqrt{p}$ could be the choice by default. Concerning the set B_p , since $Q_n(\gamma) = Q_n(-\gamma)$ for any $\gamma \in \mathcal{S}^p$, one could restrict the set B_p to a half-unit hypersphere such as $\{\gamma \in \mathcal{S}^p : \gamma_1 \geq 0\}$. One could restrict B_p even more, and hence speed optimization algorithms, when some prior information indicates a set of directions suitable for detecting the alternative. Assumption D-(c)(i) also imposes some mild restrictions on B_p but in view of our extensive empirical investigation, we argue that in practice, the choice of B_p is not an issue; one can confidently let B_p be a whole half-hypersphere \mathcal{S}^p .

3.5.3. The choice of the penalty α_n

Unlike for Lavergne and Patilea (2008), we work here with a standardized version of $Q_n(\gamma)$ and this simplifies the choice of α_n . In theory, α_n has to grow as slowly as possible but faster than $p \ln n$. In practice, for sample sizes of hundreds and p up to 6 or 8, α_n needs not vary much with n and p . Larger α_n (for instance $\alpha_n = 10$) will likely result in taking $\widehat{\gamma}_n = \gamma_0^{(p)}$ and in this case the standard normal critical values could be accurate even for moderate samples sizes. Having $\widehat{\gamma}_n = \gamma_0^{(p)}$ might be reasonable when the practitioner judges $\gamma_0^{(p)}$ trustful for detecting alternatives. On the contrary, smaller α_n (for instance $\alpha_n = 1$) will likely lead to a value of the test statistic equal to the maximum value of $nh^{1/2}Q_n(\gamma)/\widehat{v}_n(\gamma)$ and hence in general the test $\mathbb{I}(T_n \geq z_{1-a})$ will be too liberal. Meanwhile, smaller α_n is preferable for detecting general alternatives.

On the basis of our detailed simulation investigation, we recommend $3 \leq \alpha_n \leq 5$ as a compromise, and a correction of the critical values using resampling, as explained in the empirical section below.

3.5.4. The choice of the bandwidth h

340 For an optimal bandwidth choice, one could follow Horowitz and Spokoiny (2001), Guerre and Lavergne (2005), and define a test based on the maximum of T_n over a grid of bandwidths. In view of Theorem 3.5 of Lavergne and Patilea (2008), such procedure would have some optimality property, uniformly over smoothness classes of alternatives (e.g., Hölder classes of functions), provided
 345 such functions depend only on some $\langle X, \gamma \rangle$ for a fixed finite-dimension γ . Under suitable assumptions, one could expect to extend the results to single-index functions depending on some $\langle X, b \rangle$ for some fixed $b \in L^2[0, 1]$. The optimal data-driven bandwidth choice for general classes of alternatives depending on X , remains a challenging problem that deserve a separate investigation. Already,
 350 in the empirical reported in the Supplementary Material, we show that the test performs reasonably well for a large range of bandwidths.

4. Testing the goodness-of-fit of functional regression models

Here we apply our projection-based methodology for testing the goodness-of-fit of the functional linear regression against general alternatives. To the best
 355 of our knowledge, our results are new in the functional regression framework.

Let U be a real-valued random variable and X be a random element of $L^2[0, 1]$. The model we want to test is the functional linear model

$$Y = a + \langle b, X \rangle + U, \quad a \in \mathbb{R}, b \in L^2[0, 1],$$

where a and b are unknown. The null hypothesis is $H_0 : \mathbb{E}(U|X) = 0$ a.s.

The observations are $(Y_1, X_1), \dots, (Y_n, X_n)$, independent copies of (Y, X) . Hence the error term U has to be estimated in a preliminary step from the estimates of the parameters a and b . We will investigate the behavior of our test
 360 statistic under the null and alternative hypotheses for a generic estimate of the slope with a suitable rate of convergence. Next, we will investigate the details in the standard case of the slope estimate based on the functional principal component analysis. For simplicity and to highlight the influence of each aspect on our goodness-of-fit test approach, in this section we consider again a fixed
 365 basis \mathcal{R} and we set \mathcal{R} to be the FPC basis. Clearly, the arguments used for Corollary 3.5 would still apply and a test built with the estimated FPC basis is readily available.

4.1. The test statistic and the behavior under the null hypothesis

Let $\widehat{b} \in L^2[0, 1]$ denote a generic estimator of the slope b and let

$$\widehat{a} = \bar{Y}_n - \int_0^1 \widehat{b}(t) \bar{X}_n(t) dt = a - \int_0^1 \{\widehat{b}(t) - b(t)\} \bar{X}_n(t) dt + \bar{U}_n,$$

with $\bar{U}_n = n^{-1} \sum_{i=1}^n U_i$. Let $\hat{U}_i = Y_i - \hat{a} - \langle \hat{b}, X_i \rangle$ be the i -th residual and

$$Q_n(\gamma; \hat{a}, \hat{b}) = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \hat{U}_i \hat{U}_j \frac{1}{h} K_h(\langle X_i - X_j, \gamma \rangle), \quad \gamma \in \mathcal{S}^p.$$

Since the test statistic only depends on the observed differences $X_i - X_j$, here
 370 it is not needed to impose X to have zero-mean. Let $\hat{v}_n^2(\cdot; \hat{a}, \hat{b})$ be an estimate
 of the variance of $nh^{1/2}Q_n(\cdot; \hat{a}, \hat{b})$ as in section 3.2. Given $\gamma_0^{(p)} \in B_p \subset \mathcal{S}^p$, B_p
 that is not negligible in \mathcal{S}^p , the least favorable direction γ for H_0 being defined
 as

$$\hat{\gamma}_n = \arg \max_{\gamma \in B_p} \left[nh^{1/2}Q_n(\gamma; \hat{a}, \hat{b}) / \hat{v}_n(\gamma; \hat{a}, \hat{b}) - \alpha_n \mathbb{I}_{\{\gamma \neq \gamma_0^{(p)}\}} \right]. \quad (4.1)$$

The test statistic is then

$$T_n = nh^{1/2} \frac{Q_n(\hat{\gamma}_n; \hat{a}, \hat{b})}{\hat{v}_n(\hat{\gamma}_n; \hat{a}, \hat{b})}. \quad (4.2)$$

375 To derive the behavior of the test under the null hypothesis, we show that

$$\sup_{\gamma \in \mathcal{S}^p} \{nh^{1/2}|Q_n(\gamma; \hat{a}, \hat{b}) - Q_n(\gamma)| + |1 - \hat{v}_n^2(\gamma; \hat{a}, \hat{b})\hat{v}_n^{-2}(\gamma)|\} = o_{\mathbb{P}}(1), \quad (4.3)$$

with $Q_n(\gamma)$ and $\hat{v}_n^2(\gamma)$ defined as in section 3; *i.e.*, that is we will bring the
 problem back to the case where the U_i are observed.

Lemma 4.1. *Assume that the conditions of Theorem 3.3 are met. Let $\hat{b} \in$
 $L^2[0, 1]$ be an estimator of b such that $\|\hat{b} - b\|_{L^2} = O_{\mathbb{P}}(n^{-\rho})$ for some $1/4 < \rho \leq$
 380 $1/2$. Moreover, suppose that $n^{1-2\zeta}h^{1/2} \rightarrow 0$ for some $1/4 < \zeta < \rho$. Then, under
 the hypothesis H_0 the uniform convergence in (4.3) holds true.*

At this stage it is worthwhile to note an important difference between the
 functional data framework and the finite-dimension case. In the latter case, the
 parameters of the regression model are usually estimated at the rate $O_{\mathbb{P}}(n^{-1/2})$
 385 and this immediately makes the equivalences (4.3) true. In the functional re-
 gression case, the rate of $\|\hat{b} - b\|_{L^2}$ depends on the regularities of the covariate
 and of the slope parameter and is in general less than $O_{\mathbb{P}}(n^{-1/2})$, see Hall and
 Horowitz (2007), Crambes, Kneip and Sarda (2009). To make the $\hat{U}_i - U_i$ suf-
 ficiently small and hence to preserve the standard normal critical values for T_n
 390 defined in (4.2), we pay a price on the bandwidth h : slower rates of $\|\hat{b} - b\|_{L^2}$
 will require faster decreases for h , and this will result in a loss of power against
 sequences of local alternatives.

The proof of the following result is a direct consequence of Lemma 4.1 and
 the arguments we used for Theorem 3.3, therefore, we shall omit it.

395 **Theorem 4.2.** *Under the conditions of Lemma 4.1 and if the hypothesis H_0
 holds true, the law of the statistic T_n is asymptotically standard normal. Con-
 sequently the test given by $\mathbb{I}(T_n \geq z_{1-a})$ has asymptotic level a .*

Next, let us consider the case where \widehat{b} is obtained using the standard functional principal component analysis approach; see, e.g., Ramsay and Silverman (2005) and Hall and Horowitz (2007). In this case

$$\widehat{b}(t) = \sum_{j=1}^m \widehat{b}_j \widehat{\phi}_j(t), \quad t \in [0, 1], \quad (4.4)$$

where $\widehat{b}_j = \widehat{\theta}_j^{-1} \widehat{g}_j$, $\widehat{g}_j = \langle \widehat{g}, \widehat{\phi}_j \rangle$ and

$$\widehat{g}(t) = n^{-1} \sum_{i=1}^n (Y_i - \bar{Y}_n)(X_i(t) - \bar{X}_n(t)) \quad (4.5)$$

with $\bar{Y}_n = n^{-1} \sum_{i=1}^n Y_i$. As in section 3.5, $\widehat{\theta}_1 \geq \widehat{\theta}_2 \geq \dots \geq 0$ denote the eigenvalues of $\widehat{\mathcal{K}}$ and $\widehat{\phi}_1, \widehat{\phi}_2, \dots$ is the corresponding basis of eigenfunctions. The truncation point m is a smoothing parameter in the estimation of b . Hall and Horowitz showed that under Assumption P and if $m \asymp n^{1/(\alpha+2\beta)}$,

$$\|\widehat{b} - b\|_{L^2}^2 = O_{\mathbb{P}}\left(n^{-\frac{2\beta-1}{\alpha+2\beta}}\right),$$

and this rate is optimal in a minimax sense. In this case $\rho = (2\beta - 1)/\{2(\alpha + 2\beta)\}$ and our condition $\rho > 1/4$, which guarantees a non empty range for the bandwidth, becomes $\beta > \frac{1}{2}\alpha + 1$, that is precisely the condition in Assumption P-(d) above and used by Hall and Horowitz (2007).

4.2. The behavior under the alternatives

The alternatives of the functional linear model we consider are of the form

$$H_{1n} : Y_{in} = a + \langle b, X_i \rangle + r_n \delta(X_i) + U_i^0, \quad \mathbb{E}(U_i^0 | X_i) = 0, \quad 1 \leq i \leq n, \quad (4.6)$$

with $\delta(\cdot)$ some fixed function and r_n , $n \geq 1$ a sequence of real numbers. Firstly, we have to analyze the behavior of the estimator \widehat{b} under these alternatives. To keep the paper at a reasonable length, hereafter we consider that \widehat{b} is the estimator obtained through the FPC approach. Moreover, we assume

$$0 < \mathbb{E}[\delta^4(X)] < \infty, \quad \mathbb{E}[\delta(X)] = 0 \quad \text{and} \quad \mathbb{E}[\delta(X)X] = 0. \quad (4.7)$$

The last two conditions are quite common in nonparametric testing, they allow us to focus on the performance of the test to detect departures from the model; see for instance equation (3.11) in Guerre and Lavergne (2005).

Lemma 4.3. *Assume that X_1, \dots, X_n are independent draws from X . Moreover, $\int_0^1 \mathbb{E}[X^4(t)]dt < \infty$ and condition (4.7) holds true. Let \widehat{b} (resp. \widehat{b}^0) be the estimator (4.4) obtained from data generated according to the model (4.6) with a bounded sequence $r_n \geq 0$, $n \geq 1$ (resp. with $r_n = 0$, $\forall n \geq 1$). Then*

$$\|\widehat{b}^0 - \widehat{b}\|_{L^2}^2 = O_{\mathbb{P}}(r_n^2 n^{-1}) \sum_{j=1}^m \widehat{\theta}_j^{-2}.$$

If in addition Assumption P hold true and $m \asymp n^{1/(\alpha+2\beta)}$, then

$$\int_0^1 \{\widehat{b}(t) - b(t)\}^2 dt = O_{\mathbb{P}} \left(n^{-\frac{2\beta-1}{\alpha+2\beta}} \right) + o_{\mathbb{P}}(r_n^2).$$

We will not need to investigate the rate of the estimator of a under the alternatives since by construction $\widehat{a} - a = -\int_0^1 \{\widehat{b}(t) - b(t)\} \overline{X}_n(t) dt + \overline{U}_n$. We can now analyze the behavior of the test statistics under the alternatives (4.6). The residuals \widehat{U}_i can be decomposed thus

$$\widehat{U}_i = U_i^0 + r_n \delta(X_i) - \langle \widehat{b} - b, X_i - \overline{X}_n \rangle - r_n \overline{\delta(X)}_n - \overline{U}_n^0. \quad (4.8)$$

Theorem 4.4. Consider the hypotheses (4.6) with $\delta(\cdot)$ satisfying (4.7). Let $\widehat{b} \in L^2[0, 1]$ be an estimator of b . Suppose the conditions of Theorem 4.2 are met with U replaced by U^0 . Moreover,

$$(i) r_n^2 n h^{1/2} / \alpha_n \rightarrow \infty; \quad (ii) r_n^{-1} \|\widehat{b} - b\|_{L^2} = o_{\mathbb{P}}(1); \quad (iii) \alpha_n / \{p \ln n\} \rightarrow \infty.$$

The test based on T_n defined in (4.2) then rejects the functional linear regression model with probability tending to 1, provided there exists $p \geq 1$ and $\tilde{\gamma} \in B_p$ such that one of conditions (1) to (3) of Theorem 3.4 holds true.

If Assumption P is met, condition (ii) of Theorem 4.4 indicates that the test could detect only alternatives H_{1n} that approach H_0 more slowly than $n^{-(2\beta-1)/\{2(\alpha+2\beta)\}}$. Meanwhile, to detect the fastest possible alternatives, the bandwidth should decrease to zero as slowly as allowed by condition (i), that is faster than $n^{-2(\alpha+1)/(\alpha+2\beta)}$ times a power of $\ln n$, provided the dimension p and α_n increase as fast as a power of $\ln n$ such that condition (iii) is met.

5. Empirical analysis

5.1. Bootstrap critical values

In the Supplementary Material we provide some empirical evidence about the effectiveness of the asymptotic approximation of our test statistics. The conclusion is that the accuracy of the theoretical standard normal critical values depend on the tuning of the sequence α_n . In order to improve the accuracy of the asymptotic critical values for finite samples, let us introduce a simple and effective bootstrap procedure that we extensively used in our simulation experiments.

Let $U_i^b = \zeta_i U_i$, $1 \leq i \leq n$, be the bootstrap sample where ζ_i are independent variables following the two-points distribution proposed by Mammen (1993) : $\zeta_i = -(\sqrt{5} - 1)/2$ with probability $(\sqrt{5} + 1)/(2\sqrt{5})$ and $\zeta_i = (\sqrt{5} + 1)/2$ with probability $(\sqrt{5} - 1)/(2\sqrt{5})$. A bootstrap statistic T_n^b is built from a bootstrap sample as was the original statistic T_n . Similarly, let \widehat{T}_n^b be the bootstrap statistic obtained with the estimated FPC basis. When the scheme is repeated many times, the bootstrap critical value $z_{1-a,n}^*$ at level a is the empirical $(1 - a)$ -th quantile of the bootstrapped test statistics. The asymptotic validity of

450 this bootstrap procedure is guaranteed by the next result that follows from the asymptotic results presented in the previous sections. The detailed justification of this is postponed to the Appendix.

Theorem 5.1. *If the conditions of of the Theorem 3.3 and Theorem 3.4 are met. Then,*

$$\sup_{x \in \mathbb{R}} |\mathbb{P}(T_n^b \leq x \mid U_1, X_1, \dots, U_n, X_n) - \mathbb{P}(T_n \leq x)| \rightarrow 0, \quad \text{in probability.}$$

The same statement remains true with \widehat{T}_n^b and \widehat{T}_n replacing T_n^b and T_n respectively, provided the conditions of Corollary 3.5 are met.

5.2. Simulation study

455 We present here some simulation results to illustrate the performance of the test both under the null and under the alternative. Some comparison with competing tests will be illustrated. Empirical evidence on the influence of some parameters of the test, such as the bandwidth h , the privileged direction and the penalization, is provided in the Supplementary Material.

460 Our first study will be focused on testing the covariate effect. A parametric test of covariate linear effect was proposed by Cardot *et al.* (2003). We will compare the performance of the new test proposed here with this parametric competitor, under the null and under linear and non-linear alternatives. The following models will be considered:

$$\begin{aligned} \text{No-effect:} & \quad Y = \sigma U^0 \\ \text{Linear effect:} & \quad Y = \langle b_1, X \rangle + \sigma U^0 \\ \text{Quadratic effect:} & \quad Y = \langle b_2, X \rangle^2 + U^0 \end{aligned}$$

465 where X is a standard Brownian motion on $[0, 1]$, U^0 is centered normal and U^0 and X are independent. The parameters are taken as $\sigma = 0.253$, $b_1(t) = (\sin(2\pi t^3))^3$ and $b_2(t) = 0.6$ for all $t \in [0, 1]$. The function b_1 was already considered by Cardot *et al.* (2003) and the value $\sigma = 0.253$ was chosen in order to obtain a signal-to noise ratio of 10% (see Cardot *et al.* (2003)).

470 Table 1 shows the percentages of rejection based on one thousand simulated samples obtained by the new test and Cardot's test when testing the covariate effect. For the new test, estimated functional principal components were used for the basis, together with an 'uninformative' privileged direction $\gamma_0^{(p)} = (1, \dots, 1)/\sqrt{p}$ and the penalization $\alpha_n = 3$. Several values of the dimension p were considered and their results are presented in Table 1. We used the Epanechnikov kernel and bandwidth $h = 1.4n^{-2/9}$. To search for the least favorable direction $\widehat{\gamma}_n$, we used the sequential algorithm proposed by Patilea, Sánchez-Sellero and Saumard (2016). This algorithm, that we explain in the Supplementary Material for the sake of completeness, involves only one-dimensional optimizations and preserves the consistency of the test. The sequential algorithm was applied to solve the optimization problem with a grid of 480 50 points in each dimension.

To estimate the conditional variance, both estimators (3.6) and (3.7) were considered. For the estimator (3.7), a kernel estimator of the errors' conditional variance was used, with uniform kernel and bandwidth $h_v = 0.5n^{-1/6}$. We observed a better power under the alternative with the estimator (3.7), so the results will be given with this estimator.

The wild bootstrap procedure proposed in Section 3.5 was used to approximate the critical values. For each sample, we used 199 bootstrap samples to compute the critical value for a nominal level of 5%.

As shown in Table 1, both the new test and Cardot's test respect the nominal level under the null hypothesis of no-effect. Under the linear alternative, Cardot's test is clearly more powerful than the new test. This was expected because their test is specially designed to detect this type of alternative. Under the quadratic alternative, the new test is powerful while Cardot's test does not detect this non-linear alternative. As expected, a larger sample size leads to a larger power. The dimension p has no much influence on the power results. In our experience, the power increases with the initial increments of the dimension p , while additional high dimensions are harmful for the power. Note also that the new test statistic is designed to overcome the effect of large dimensions. Other simpler tests would lead to a dramatic effect of the dimension on the power of the test.

Model	p	$n = 100$		$n = 200$	
		New	CFMS	New	CFMS
No-effect (Null)	3	6.0	5.8	6.3	4.9
	5	4.9	6.6	6.3	5.5
	7	4.7	6.0	6.2	4.9
	9	4.9	6.1	5.4	4.7
Linear effect	3	48.3	78.9	79.0	98.2
	5	51.3	72.0	82.9	96.8
	7	53.7	67.3	83.9	94.9
	9	53.8	63.0	84.8	92.6
Quadratic effect	3	29.9	8.4	54.7	7.7
	5	29.3	7.7	57.8	6.9
	7	29.2	7.7	56.0	6.8
	9	28.3	8.0	56.0	6.5

Table 1. Percentages of rejection by the new test and Cardot's test (entitled CFMS) for testing the covariate effect under the null hypothesis and under linear and quadratic alternatives. The nominal level was 5%.

The next experiment that we will present here consists of a lack-of-fit test of functional linear model. Under the null hypothesis, the simulated data will be generated following the linear effect. Note that this model was considered as an alternative in the test of effect. As a deviation from the functional linear model, we will now consider the functional quadratic model, that was defined

in Yao and Muller (2010) by

$$Y = \alpha + \int \beta(t)X^c(t) dt + \int \int h(s, t)X^c(s)X^c(t) dt + U$$

where $X^c = X - E(X)$ is the centered functional covariate. The first single integral represents the functional linear component, while the double integral is the quadratic deviation. Horvath and Reeder (2013) proposed a test to detect this type of quadratic alternative.

Table 2 shows the percentages of rejection of the new test and Horvath and Reeder's test, under the null hypothesis (functional linear model) and under the following quadratic and cubic models:

$$\text{Quadratic model: } Y = \int b_1(t)X^c(t) dt + \int \int h(s, t)X^c(s)X^c(t) dt + \sigma U^0$$

$$\text{Cubic model: } Y = \int b_1(t)X^c(t) dt + 0.3(\langle b_3, X \rangle^3 - \langle b_3, X \rangle) + \sigma U^0$$

where X and U^0 are generated as in the previous models, σ and b_1 also coincide with their values in previous models, $h(s, t) = 0.2$ for all $s, t \in [0, 1]$ and $b_3(t) = 1$ for all $t \in [0, 1]$. The basis, privileged direction, penalization, bandwidth and conditional variance will follow the same rules as in the test of effect. Again, the sequential algorithm is applied and the bootstrap approximation based on 199 replicates are used. The percentages of rejection are based on one thousand samples and the nominal level was 5%.

Model	p	$n = 100$		$n = 200$	
		New	HR	New	HR
Linear (Null)	3	4.8	10.6	4.2	6.6
	5	5.3	28.8	4.7	12.4
	7	4.7	70.3	4.4	28.9
	9	4.5	99.1	4.5	60.7
Quadratic	3	46.7	82.0	81.8	97.6
	5	49.8	82.1	83.6	95.5
	7	52.0	93.6	84.3	95.5
	9	53.0	99.7	83.8	97.6
Cubic	3	54.5	37.8	88.4	37.9
	5	54.5	55.0	88.6	46.3
	7	54.9	84.2	89.2	60.8
	9	57.3	99.6	90.0	82.0

Table 2. Percentages of rejection by the new test and Horvath and Reeder's test (entitled HR) for testing the lack-of fit of the functional linear model under the null hypothesis and under quadratic and cubic deviations. The nominal level was 5%.

Following Cardot *et al.* (2003), five dimensions were used to estimate the functional linear model. Several numbers of dimensions to detect the alternative

are considered in Table 2. Note that Horv ath and Reeder’s test also requires a dimension to estimate the possible quadratic deviation. However a large value of this dimension makes a dramatic effect on their level accuracy under the null. Observe that a chi-squared approximation is used there, which could be a very imprecise approximation for large dimensions. Meanwhile, the new test shows a good approximation of the nominal level under the null. The dimension p does not have much effect on the new test. Under the quadratic alternative, and ignoring large dimensions, Horv ath and Reeder’s test is more powerful, as expected. On the contrary, the new test is clearly more powerful for detecting the cubic deviation from the null. The conclusion is that the new test is required to detect general deviations from the null hypothesis.

5.3. Application to real data

We analyzed the data set collected by Tecator. The task is to predict the fat content of a meat sample on the basis of its near infrared absorbance spectrum. The data set, available at <http://lib.stat.cmu.edu/datasets/tecator>, contains 240 samples. For each sample of finely chopped pure meat, a 100 channel spectrum of absorbances was recorded using a Tecator Infratec Food and Feed Analyzer, a spectrometer that works in the wavelength range 850-1050 nm. These absorbances can be thought of as a discrete approximation of a continuous record $X(t)$. For each sample of meat, the fat content Y was measured by analytical chemistry.

Yao and M uller (2010) proposed a functional quadratic model to predict the fat content Y based on the absorbance spectrum X . Horv ath and Reeder (2013) applied their test to check whether the quadratic term is needed, versus the null hypothesis of a functional linear model. The conclusion was that the quadratic effect is significant.

		Linear model				Quadratic model			
c		0.8	1.2	1.6	2.0	0.8	1.2	1.6	2.0
$p = 2$	$m = 1$	0.1	0.3	0.4	0.8	0.5	1.4	1.7	4.1
	$m = 2$	0.2	0.1	0.2	0.3	0.5	0.3	0.4	0.9
	$m = 3$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$p = 3$	$m = 1$	0.0	0.3	0.4	0.5	0.2	0.4	0.3	0.1
	$m = 2$	0.1	0.1	0.2	0.3	0.1	0.1	0.1	0.1
	$m = 3$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 3. Percentages of rejection for the Tecator data with different bandwidths $h = c \cdot n^{-2/9}$, different m for model estimation and different p .

We applied the test proposed here, first to check the goodness-of-fit of the functional linear model, and next the goodness-of-fit of the functional quadratic model. Table 3 contains the p -values corresponding to our test for different values of the bandwidth, $h = c \cdot n^{-2/9}$, the parameter m for model estimation and the dimension p . The basis was estimated by FPC. The penalization was $\alpha_n = 3$. The p -values were approximated by means of 9999 bootstrap replicates. We can conclude that both the functional linear and the functional quadratic

models should be rejected for the Tecator data set. This conclusion confirms
 565 the empirical results of Chen, Hall and Müller (2011) who proposed an additive
 double index model. Indeed, the link functions estimated by Chen, Hall and
 Müller (2011) do not show respective linear and quadratic patterns.

6. Concluding remarks

A new method has been proposed to test the effect of a functional covariate
 570 on a scalar response, and also to test the lack-of-fit of a functional linear model.
 This method can be applied to lack-of-fit testing of many other functional re-
 gression models, such as for instance, the partial functional linear model, see
 Ramsay and Silverman (2005) chapter 10, and the generalized functional linear
 models introduced by Müller and Stadtmüller (2005).

575 Limit distribution of the test statistic under the null hypothesis has been
 shown to be standard normal. Consistency under alternatives converging to the
 null has been established, where the allowed rates of convergence are almost the
 same as in parametric model checks based on kernel smoothing with *univariate*
 covariate, see for instance Guerre and Lavergne (2005). It was also proved that
 580 our theoretical results, first derived for a fixed basis in $L^2[0, 1]$, remain true with
 an estimated basis.

A simulation study shows that the proposed bootstrap approximation per-
 forms well and that the test is able to detect general departures from the null
 hypothesis. Additional simulations, provided in the Supplementary material, il-
 585 lustrate the possible effect of some parameters necessary to implement the test,
 as the basis, the dimension p , the privileged direction, the penalization and the
 bandwidth. Some advice are then given in order to apply the test in practice.

In an illustration with the Tecator data set, the proposed test was useful to
 detect that two known models should be rejected, a conclusion that is consistent
 590 with the recent literature on this dataset.

7. Appendix

For real random variables, $A_n \asymp_{\mathbb{P}} B_n$ means that there exists a constant
 $C > 1$ such that $\mathbb{P}(1/C \leq A_n/B_n \leq C)$ goes to 1 when n grows. In the
 following C, C_1, c, c_1, \dots represent constants that may change from line to line.
 595 Below $\mathcal{F}[\cdot]$ denotes the Fourier transform, cf. Rudin (1987).

7.1. Assumptions

In the following we present the assumptions we use to derive the technical
 results and we provide some comments.

Assumption D.

600 (a) $(U_1, X_1), \dots, (U_n, X_n)$ is an independent sample of the random vector
 $(U, X) \in \mathbb{R} \times L^2[0, 1]$. $\mathbb{E}U^8 < \infty$ and $\exists \varrho > 0$ such that $\mathbb{E}(\exp\{\varrho \|X\|_{L^2}\}) < \infty$.

- (b) $\exists 0 < \underline{\sigma}^2 < \bar{\sigma}^2 < \infty$ and $\nu > 2$ such that $\underline{\sigma}^2 \leq \mathbb{E}(U^2 | X)$ and $\mathbb{E}(|U|^\nu | X) \leq \bar{\sigma}^2$ almost surely;
- (c) The sets $B_p \subset \mathcal{S}^p$, $p \geq 1$ from the definition of $\hat{\gamma}_n$ are such that:
- 605 (i) $\exists C_1$ (independent of p) such that $\forall p \geq 1$ and $\forall \gamma \in B_p$, the variable $\langle X, \gamma \rangle$ admits a density f_γ and $\int_{\mathbb{R}} f_\gamma^2 \leq C_1$;
- (ii) $\gamma_0^{(p)} \in B_p$ and $B_p \times 0_{p'-p} \subset B_{p'}$, $\forall 1 \leq p < p'$, where $0_{p'-p}$ denotes the null vector of dimension $p' - p$.
- (d) The initial ‘guess’ $b_0 = (b_{01}, b_{02}, \dots)$, with $\sum_{j=1}^{\infty} b_{0j}^2 < \infty$, is such that
- 610 (i) $\sum_{j=1}^{\infty} \langle X, \phi_j \rangle b_{0j}$ has a density f_0 such that $\int_{\mathbb{R}} f_0^2 < \infty$ and $\int_{\mathbb{R}} \{f_{\gamma_0^{(p)}} - f_0\}^2 \rightarrow 0$ when $p \rightarrow \infty$;
- (ii) $\exists c_f$ (independent of p) such that $\forall p \geq 1$ and $\forall \gamma \in B_p$, $\int_{\mathbb{R}} f_\gamma^2 \geq c_f > 0$;
- (iii) $\exists C_2$ (independent of p) such that $f_{\gamma_0^{(p)}} \leq C_2$, $\forall p \geq 1$.

615 **Assumption K.**

- (a) The kernel K is a symmetric integrable Lipschitz function of bounded variation and $\int_{\mathbb{R}} K = 1$.
- (b) $h \rightarrow 0$ and $nh^{1+\alpha} \rightarrow \infty$ for some $\alpha \in (0, 1)$.
- (c) $p \geq 1$ increases to infinity with n and there exists a constant $\lambda > 0$ such
- 620 that $p \ln^{-\lambda} n$ is bounded.

Let us comment on these assumptions. In Assumption K, continuity and bounded variation guarantee that K can be recovered by the inverse Fourier transform. See, for instance, Boas (1962). The Lipschitz property of K will help to control the increments of $Q_n(\gamma)$ when γ varies in \mathcal{S}^p . Moreover, this

625 mild property represent one of the three types of sufficient conditions that we propose for deriving the same asymptotic result with an estimated instead of a fixed basis \mathcal{R} in the definition of T_n , see section 3.5. In Assumption K-(c), it is also possible to let p grow faster than the logarithmic rate. However, in theory, this could induce a loss of power for our test. Concerning Assumption D, there

630 is a trade off between the moment conditions on U and the range of rates allowed for the bandwidth h : higher moments will be needed for wider ranges; see the proof of Lemma 3.1. On the other hand, faster decrease of h will be needed to smooth out the effect of slower convergence for the slope estimator when testing the lack of fit of the linear functional model. The condition $nh^{1+\alpha} \rightarrow \infty$

635 for some small $\alpha > 0$, is nearly optimal for this purpose. The upper and lower bounds on the conditional moments of U are mild conditions that allow the variance of $Q_n(\gamma)$ to be controlled. Finally, let us comment on Assumptions D-(c,d). On one hand, a key issue in the proof of Lemma 3.1 below, and some of the subsequent proofs, will be to control $\mathbb{E}[h^{-1}K_h(\langle X_1 - X_2, \gamma \rangle)]$ uniformly

640 in $\gamma \in B_p$ as p grows and h decreases to zero. A simple solution is to bound this quantity by a constant. Using the Fourier transform and Plancherel theorem, this is guaranteed if $\int_{\mathbb{R}} f_\gamma^2$ is uniformly bounded. Such integrability conditions, and even uniform boundedness, could be easily achieved for instance when the variables $\langle X, \phi_j \rangle$ are independent. It then suffices to fix some $j \geq 1$ such that

645 the density of $\langle X, \phi_j \rangle$ is bounded and some small c independent of p and to
take $B_p = \{(\gamma_1, \dots, \gamma_j, \dots, \gamma_p) \in \mathcal{S}^p : |\gamma_j| \geq c\}$. Elementary properties of the
convolution of densities then guarantee that f_γ are uniformly bounded $\forall p \geq 1$
and $\forall \gamma \in B_p$. On the other hand, we have to keep the variance estimate in the
denominator of the test statistic (3.3) away from zero. Given the way variance
650 estimators were defined, it suffices to ensure that $\mathbb{E}[h^{-1}K_h^2(\langle X_1 - X_2, \gamma_0^{(p)} \rangle)]$ is
bounded away from zero. Assumption D-(d)(i) provides mild sufficient condi-
tions for this. When one uses the variance estimator defined in (3.7), a similar
control of small values of the denominator appearing in (3.2) is also required
for $\mathbb{E}[h^{-1}K_h^2(\langle X_1 - X_2, \gamma \rangle)]$, uniformly with respect to $\gamma \in B_p$. This control is
655 realized using Assumption D-(d)(ii). See the proof of Lemma 3.1. Assumption
D-(d)(iii) will complete the sufficient conditions for deriving standard normal
critical values using the central limit theorem for U -statistics of Guerre and
Lavergne (2005, Lemma 2). Finally, the condition $B_p \times 0_{p'-p} \subset B_{p'}$, $\forall p < p'$,
is a mild technical condition that combined with Lemma 2.1-(A), greatly sim-
660 plifies the proof of the consistency of our test. Some practical guidance for the
choice of p , $\gamma_0^{(p)}$ and B_p is provided in section 3.5.

The following assumptions are standard conditions on the covariance oper-
ator \mathcal{K} and the slope parameter in the linear model, as could be found in Hall
and Horowitz (2007). Let $\xi_j = \langle X, \phi_j \rangle$, where ϕ_j are the FPC.

665 **Assumption P.**

- (a) For some constant $C > 1$, $\mathbb{E}[\xi_j - \mathbb{E}(\xi_j)]^4 \leq C\theta_j^2$ for all j .
- (b) The error U is independent of X , has zero mean and finite variance.
- (c) The eigenvalues θ_j of \mathcal{K} satisfy $\theta_j - \theta_{j+1} \geq C^{-1}j^{-\alpha-1}$, $\forall j \geq 1$.
- (d) The Fourier coefficients b_j satisfy $|b_j| \leq Cj^{-\beta}$ and $\alpha > 1$, $\frac{1}{2}\alpha + 1 < \beta$.

670 **7.2. Preliminary results**

The proofs of the results stated in this section are provided in the Supple-
mentary Material.

A tool used for proving Lemma 3.1 is an exponential bound for the tail
probability of U -statistics that we state below. Let Z_1, \dots, Z_n be independ-
ent, not necessarily identically distributed, random variables taking values in a
measurable space (\mathcal{Z}, Υ) . Let $h_{i,j}(\cdot, \cdot)$, $1 \leq i, j \leq n$ be bounded real-valued mea-
surable functions on \mathcal{Z}^2 such that $h_{i,j}(z_i, z_j) = h_{j,i}(z_j, z_i)$, $\mathbb{E}[h_{i,j}(z_i, Z_j)] = 0$,
 $\forall 1 \leq i, j \leq n$, $\forall z_i, z_j$. Define $A_n = \max_{i,j} \|h_{i,j}(\cdot, \cdot)\|_\infty$,

$$B_n^2 = \max_j \left\| \sum_i \mathbb{E}[h_{i,j}^2(Z_i, \cdot)] \right\|_\infty, \quad C_n^2 = \sum_{i,j} \mathbb{E}[h_{i,j}^2(Z_i, Z_j)],$$

(here and elsewhere $\|\cdot\|_\infty$ denote the sup norm) and

$$D_n = \sup \left\{ \mathbb{E} \sum_{i,j} h_{i,j}(Z_i, Z_j) f_i(Z_i) g_j(Z_j) : \mathbb{E} \sum_i f_i^2(Z_i) \leq 1, \mathbb{E} \sum_j g_j^2(Z_j) \leq 1 \right\}.$$

The following lemma is a version of Th. 3.3 in Giné, Latała and Zinn (2000).

Lemma 7.1. *There exists an universal constant $L < \infty$ (in particular, independent on n and the functions $h_{i,j}$) such that $\forall t > 0$*

$$\mathbb{P} \left\{ \left| \sum_{1 \leq i \neq j \leq n} h_{i,j}(Z_i, Z_j) \right| \geq t \right\} \leq L \exp \left[-\frac{1}{L} \min \left(\frac{t^2}{C_n^2}, \frac{t}{D_n}, \frac{t^{2/3}}{B_n^{2/3}}, \frac{t^{1/2}}{A_n^{1/2}} \right) \right].$$

Let $\gamma \in \mathcal{S}^p$ and let x_1, \dots, x_n be an arbitrary collection of non-random points in $L^2[0, 1]$. Consider $\tilde{Z}_1, \dots, \tilde{Z}_n$ independent real-valued variables such that for each $1 \leq i \leq n$, the law of \tilde{Z}_i is the conditional law of U_i given $X_i = x_i$. Given $\gamma \in \mathcal{S}^p$, we apply Lemma 7.1 with $h_{i,i} \equiv 0$ and

$$h_{i,j}(Z_i, Z_j) = \frac{Z_i Z_j}{n(n-1)hM^2} K_h(\langle x_i, \gamma \rangle - \langle x_j, \gamma \rangle), \quad 1 \leq i \neq j \leq n, \quad (7.1)$$

$Z_i = \tilde{Z}_i \mathbb{I}_{\{\tilde{Z}_i \leq M\}} - \mathbb{E}[\tilde{Z}_i \mathbb{I}_{\{\tilde{Z}_i \leq M\}}]$, $M > 0$ (M is allowed to increase with n). In particular, $\mathbb{E}[\tilde{Z}_i \mathbb{I}_{\{\tilde{Z}_i \leq M\}}]$ coincide with the values $\mathbb{E}[U_i \mathbb{I}_{\{U_i \leq M\}} \mid X_i = x_i]$.

For a squared matrix A , let $\|A\|_2 = \sup_{u \in \mathbb{R}^n, u \neq 0} \|Au\|/\|u\|$ and $\|A\|_F^2 = \text{tr}(A^\top A)$. The following lemmas provide upper bounds for the quantities A_n to D_n , given $x_1, \dots, x_n \in L^2[0, 1]$.

Lemma 7.2. *Let $\mathfrak{W}_n(\gamma)$ be the non random matrix with generic element*

$$\mathfrak{W}_{n,ij}(\gamma) = K_h(\langle x_i - x_j, \gamma \rangle) / [n(n-1)h], \quad i \neq j, \quad \text{and} \quad \mathfrak{W}_{n,ii} = 0.$$

Moreover, let $S_n(\gamma) = \sup_{t \in \mathbb{R}} n^{-1} h^{-1} \sum_{j=1}^n |K_h(t - \langle x_j, \gamma \rangle)|$. Under the conditions of Lemma 3.1, for $h_{i,j}$ defined in (7.1),

$$A_n = \frac{\|K\|_\infty}{n(n-1)h}, \quad B_n^2 \leq \frac{c\|K\|_\infty S_n(\gamma)}{n^3 h M^2}, \quad C_n^2 \leq \frac{c\|\mathfrak{W}_n(\gamma)\|_F^2}{M^4}, \quad D_n \leq \frac{cS_n(\gamma)}{nM^2},$$

for some constant c depending only on the upper bound of $\mathbb{E}(U^2 \mid X)$.

Let us define

$$S_n(\gamma; t) = \frac{1}{nh} \sum_{j=1}^n |K_h(t - \langle X_j, \gamma \rangle)|, \quad (7.2)$$

and let $W_n(\gamma; t)$ be the matrix with elements $W_{n,ii}(\gamma; t) = 0$

$$W_{n,ij}(\gamma; t) = K_h(\langle X_i - X_j, \gamma \rangle + t) / [n(n-1)h], \quad i \neq j. \quad (7.3)$$

Lemma 7.3. *Suppose that Assumptions D-(a) and K are met.*

1. Assume that $\sup_p \sup_{\gamma \in B_p} \int_{\mathbb{R}} f_\gamma^2 \leq C_1 < \infty$. Then, $\exists C'_1 > 0$ independent of p and h such that:

- (a) $\sup_{\gamma \in B_p} \sup_{t \in \mathbb{R}} \mathbb{E}[S_n(\gamma; t)] \leq \sqrt{C'_1/h}$;
- (b) $\sup_{\gamma \in B_p} \sup_{t \in \mathbb{R}} \mathbb{E} \{ \|W_n(\gamma; t)\|_F^2 \} \leq C'_1 / (n^2 h)$.

2. Assume that $f_\gamma(\cdot) \leq C_2$ for some $C_2 > 0$ independent of p and γ . Then, $\exists C'_2$ independent of p, h and γ such that $\sup_{t \in \mathbb{R}} \mathbb{E}[S_n(\gamma; t)] \leq C'_2$.
3. Assume that $\sup_p \sup_{\gamma \in B_p} \int_{\mathbb{R}} f_\gamma^2 \leq C_1$. Then

$$(a) \quad \sup_{\gamma \in B_p} \max_{1 \leq i \leq n} |S_n(\gamma; \langle X_i, \gamma \rangle) - \mathbb{E}[S_n(\gamma; \langle X_i, \gamma \rangle) \mid X_i]| = o_{\mathbb{P}}(h^{-1/2});$$

$$(b) \quad \sup_{\gamma \in B_p} \sup_{|t| \leq 1} \left| \|W_n(\gamma; t)\|_F^2 - \mathbb{E} \left\{ \|W_n(\gamma; t)\|_F^2 \right\} \right| = o_{\mathbb{P}}(1/(n^2 h)).$$

4. Assume that $f_{\gamma_0^{(p)}}(\cdot) \leq C_2$ for some $C_2 > 0$ independent of p . Then

$$\max_{1 \leq i \leq n} |S_n(\gamma_0^{(p)}; \langle X_i, \gamma_0^{(p)} \rangle) - \mathbb{E}[S_n(\gamma_0^{(p)}; \langle X_i, \gamma_0^{(p)} \rangle) \mid X_i]| = o_{\mathbb{P}}(1).$$

5. Assume that $\sup_p \sup_{\gamma \in B_p} f_\gamma(\cdot) \leq C$ for some $C > 0$. Then

$$\sup_{\gamma \in B_p} \sup_{|t| \leq \ln n} |S_n(\gamma; t) - \mathbb{E}[S_n(\gamma; t)]| = o_{\mathbb{P}}(1).$$

Let

$$v_n^2(\gamma_0^{(p)}) = \frac{2}{n(n-1)h} \sum_{j \neq i} \sigma_{\gamma_0^{(p)}}^2(\langle X_j, \gamma_0^{(p)} \rangle) \sigma_{\gamma_0^{(p)}}^2(\langle X_j, \gamma_0^{(p)} \rangle) K_h^2(\langle X_i - X_j, \gamma_0^{(p)} \rangle). \quad (7.4)$$

For proving Lemma 3.1, we also need the following result on the variance.

695 **Lemma 7.4.** *Let Assumptions D, K and hypothesis H_0 hold true. Then*

1. $v_n^2(\gamma_0^{(p)}) - \mathbb{E}[v_n^2(\gamma_0^{(p)})] = o_{\mathbb{P}}(1)$ and $v_n^2(\gamma_0^{(p)}) \asymp_{\mathbb{P}} 1$.
2. $\widehat{v}_n^2(\gamma_0^{(p)}) = v_n^2(\gamma_0^{(p)}) \{1 + o_{\mathbb{P}}(1)\}$, with $\widehat{v}_n^2(\gamma_0^{(p)})$ defined in equation (3.6);
3. $\widehat{v}_n^2(\gamma_0^{(p)}) = v_n^2(\gamma_0^{(p)}) \{1 + o_{\mathbb{P}}(1)\}$, with $\widehat{v}_n^2(\gamma_0^{(p)})$ defined in (3.7), provided condition (3.9) holds true;

The following facts are needed for proving Corollary 3.5. For any γ , let us write $\langle X_i, \gamma \rangle_n = \langle X_i, \gamma \rangle + \langle X_i, \Delta_{n, \gamma} \rangle$ where

$$\Delta_{n, \gamma}(t) = \sum_{k=1}^p \gamma_k [\widehat{\phi}_k(t) - \phi_k(t)], \quad t \in [0, 1].$$

Consider the operator norm $\|\mathcal{K}\|_S$ defined by $\|\mathcal{K}\|_S^2 = \int \int \sigma^2(t, s) dt ds$. Under Assumption D-(a) and the moment assumption on $\|X\|$, $\|\widehat{\mathcal{K}} - \mathcal{K}\|_S = O_{\mathbb{P}}(1/\sqrt{n})$, see for instance Bosq (2000). Next, by Cauchy-Schwarz inequality, Lemma 4.3 in Bosq (2000), and the fact that the spectral norm of the operator $\widehat{\mathcal{K}} - \mathcal{K}$ is smaller or equal to $\|\widehat{\mathcal{K}} - \mathcal{K}\|_S$,

$$\int_{[0,1]} \Delta_{n, \gamma}^2(t) dt \leq \left[\sum_{k=1}^p \gamma_k^2 \right] \sum_{k=1}^p \|\widehat{\phi}_k - \phi_k\|^2 \leq p \frac{8}{5^p} \|\widehat{\mathcal{K}} - \mathcal{K}\|_S^2,$$

where $\varsigma_p = \min_{1 \leq j \leq p} (\lambda_j - \lambda_{j+1})$. Then the lower bound for the spacing between the eigenvalues implies $\sup_{\gamma \in \mathcal{S}^p} \int_{[0,1]} \Delta_{n,\gamma}^2(t) dt \leq cp^{2\eta+1} \|\widehat{\mathcal{K}} - \mathcal{K}\|_S^2$, for some $c > 0$. Deduce $\sup_{\gamma \in \mathcal{S}^p} \max_{1 \leq i \leq n} |\langle X_i, \gamma \rangle_n - \langle X_i, \gamma \rangle|$ is bounded by

$$\max_{1 \leq i \leq n} \|X_i\|_{L^2} c^{1/2} p^{\eta+1/2} \|\widehat{\mathcal{K}} - \mathcal{K}\|_S = O_{\mathbb{P}}(p^{\eta+1/2} \ln n / \sqrt{n}).$$

700 For this, we used $\mathbb{E}[\exp(\varrho \|X\|_{L^2})] < \infty$ to deduce that $\max_{1 \leq i \leq n} \|X_i\|_{L^2} = O_{\mathbb{P}}(\ln n)$. Let $b_n \downarrow 0$ such that $b_n \sqrt{n} / [p^{\eta+1/2} \ln n] \rightarrow \infty$ and define the event $\mathcal{E}_{1n} = \{\sup_{\gamma \in \mathcal{S}^p} \max_{1 \leq i \leq n} |\langle X_i, \gamma \rangle_n - \langle X_i, \gamma \rangle| \leq b_n\}$ so that $\mathbb{P}(\mathcal{E}_{1n}^c) \rightarrow 0$.

Lemma 7.5. *Suppose that Assumptions D-(a,b) and K-(a,c) are met. Moreover, $\mathcal{K}(t, s)$ is positive definite and there exist $C, \eta > 0$ such that $\theta_j - \theta_{j+1} \geq Cj^{-\eta}$, $\forall j \geq 1$. Assume that one of the conditions 1 to 3 of Corollary 3.5 is met.*
705 *Then, for $l = 1$ and $l = 2$,*

$$\sup_{\gamma \in \mathcal{S}^p} \max_{1 \leq i \leq n} \frac{1}{nh} \sum_{j=1}^n |K_h^l(\langle X_i - X_j, \gamma \rangle_n) - K_h^l(\langle X_i - X_j, \gamma \rangle)| = o_{\mathbb{P}}(1), \quad (7.5)$$

$$\sup_{\gamma \in \mathcal{S}^p} \left| \frac{1}{n(n-1)h} \sum_{i \neq j} [K_h^2(\langle X_i - X_j, \gamma \rangle_n) - K_h^2(\langle X_i - X_j, \gamma \rangle)] \right| = o_{\mathbb{P}}(1). \quad (7.6)$$

The following result is needed for studying the functional linear model.

Lemma 7.6. *Suppose Assumptions D and K hold true and let*

$$U_{1n}(\gamma) = \frac{1}{n(n-1)h} \sum_{i \neq j} U_i K_{h,ij}(\gamma), \quad U_{2n}(\gamma) = \frac{1}{n(n-1)h} \sum_{i \neq j} U_i X_j K_{h,ij}(\gamma).$$

Then

$$\sup_{\gamma \in \tilde{B}_p} |U_{1n}(\gamma)| = O_{\mathbb{P}}\left(\frac{p^{1/2} \ln^{1/2} n}{h^{1/4} n^{1/2}}\right), \quad \sup_{\gamma \in \tilde{B}_p} \|U_{2n}(\gamma)\|_{L^2} = O_{\mathbb{P}}\left(\frac{p^{1/2} \ln^{3/2} n}{h^{1/4} n^{1/2}}\right).$$

7.3. Proofs

710 For simpler formulae and without loss of generality, we suppose $\varrho > 8/7$ in the exponential moment condition in Assumption D-(a), so that condition (2.1) holds with $s = 7\varrho/8 > 1$.

Proof of Lemma 2.1. (A). The implication (1) \Rightarrow (2) is obvious. To prove (2) \Rightarrow (1), note that $M_p = \mathbb{E}(U | X_{[p]})$ is a martingale. Doob's Martingale
715 Convergence Theorem implies that $M_p \rightarrow \mathbb{E}(U | X)$ with probability 1, thus necessarily $\mathbb{E}(U | X) = 0$. The equivalence (2) \Leftrightarrow (3) follows from Lemma 2.1-(A) of Lavergne and Patilea (2008) applied for each p .

(B). From (A)-(3), $\exists p_0 \geq 1$ such that $\mathbb{P}[\mathbb{E}(U | X_{[p_0]}) = 0] < 1$. Next, assume that p_0 is the smallest with the property $\mathbb{P}[\mathbb{E}(U | X_{[p_0]}) = 0] < 1$. Since

$\mathbb{E}(U \mid X_{[p_0]}) = \mathbb{E}[\mathbb{E}(U \mid X_{[p]}) \mid X_{[p_0]}]$, necessarily $\mathbb{P}[\mathbb{E}(U \mid X_{[p]}) = 0] < 1$, $\forall p > p_0$. Fix arbitrarily $p > p_0$ and note that by our assumptions, $\forall \gamma \in \mathcal{S}^p$, $\forall |b| \leq 1$,

$$\mathbb{E}(|U \exp\{b\langle X, \gamma \rangle\}|) \leq \mathbb{E}(|U| \exp\{|b|\langle X, \gamma \rangle\}) \leq \mathbb{E}(|U| \exp\{\|X\|_{L^2}\}) < \infty.$$

(Recall that, for any $\tilde{\gamma} \in \mathbb{R}^p$, by definition $\langle X, \tilde{\gamma} \rangle = \langle X_{[p]}, \tilde{\gamma} \rangle = \sum_{j=1}^p \tilde{\gamma}_j \langle X, \phi_j \rangle$.) Next, let $A_p = \{\gamma \in \mathcal{S}^p : \mathbb{E}(U \mid \langle X, \gamma \rangle) = 0 \text{ a.s.}\}$ and note that

$$\forall 0 < |b| \leq 1, \quad bA_p \subset \tilde{A}_p = \{\tilde{\gamma} \in \mathbb{R}^p : 0 < \|\tilde{\gamma}\| \leq 1, \mathbb{E}(U \exp\{\langle X, \tilde{\gamma} \rangle\}) = 0\}.$$

For (B) it suffice to show that \tilde{A}_p has Lebesgue measure zero and is nowhere dense in \mathbb{R}^p . For this purpose, we shall use the following property (proven in the Supplement): if W_1, W_2 are real variables with $\mathbb{E}(|W_1| \exp\{a|W_2|\}) < \infty$ for some $a > 0$, then $\mathbb{P}(\mathbb{E}(W_1 \mid W_2) = 0) < 1$ implies

$$\text{the set } \{|b| < a : \mathbb{E}(W_1 \exp\{bW_2\}) = 0\} \text{ is empty or finite.} \quad (7.7)$$

Now, let us fix some $1 < \zeta < s$. To investigate the properties of \tilde{A}_p , note that by (A), $\exists \tilde{\gamma} \in \mathcal{S}^p$ such that $\mathbb{P}(\mathbb{E}(U \mid \langle X, \tilde{\gamma} \rangle) = 0) < 1$. Then, property (7.7) applied with some $a = \zeta$, $W_1 = U$ and $W_2 = \langle X, \tilde{\gamma} \rangle$ implies that the set $\{|b| < \zeta : \mathbb{E}(U \exp\{b\langle X, \tilde{\gamma} \rangle\}) = 0\}$ is empty or finite. Deduce that there exists $v^* \in \mathbb{R}^p$, arbitrarily close to the origin, in particular satisfying $\|v^*\| < s - \zeta$, such that $\mathbb{E}(U \exp\{\langle X_{[p]}, v^* \rangle\}) \neq 0$. Next, we adapt the lines of the proof of Lemma 1 in Bierens (1990). Let $U^* = U \exp\{\langle X_{[p]}, v^* \rangle\}$ and $\xi_l = \langle X, \phi_l \rangle$. By construction, $\mathbb{P}(\mathbb{E}(U^* \mid \xi_1, \dots, \xi_l) = 0) < 1$, for $l = 1, \dots, p$. Define the sets

$$A_l^* = \{(t_1, \dots, t_l) \in \mathbb{R}^l : \|(t_1, \dots, t_l)\| < \zeta, \mathbb{E}(U^* e^{\xi_1 t_1 + \dots + \xi_l t_l}) = 0\},$$

$l = 1, \dots, p$. Since $|\xi_1 t_1 + \dots + \xi_l t_l + \langle X_{[p]}, v^* \rangle| < (\|v^*\| + \zeta)\|X\|_{L^2}$, deduce that the expectations in the definition of the sets A_l^* are well defined. Since the norm of v^* could be taken arbitrarily small such that $\tilde{A}_p \subset A_p^*$, it suffices to show that A_p^* is negligible, that is it has Lebesgue measure zero and is nowhere dense. For this purpose, first apply the property (7.7) with $a = \zeta$, $W_1 = U^*$ and $W_2 = \xi_1$ to deduce that the set A_1^* is empty or finite. Next, for $|t_1| < \zeta$, define the set

$$A_2^{**}(t_1) = \{|t_2| < \zeta : \mathbb{E}(U^* \exp\{\xi_1 t_1\} \exp\{\xi_2 t_2\}) = 0\}.$$

If $|t_1| < \zeta$ but $t_1 \notin A_1^*$, it means that $\mathbb{E}(U^* \exp\{\xi_1 t_1\}) \neq 0$. In particular this means that $\mathbb{P}(\mathbb{E}(U^* \exp\{\xi_1 t_1\} \mid \xi_2) = 0) < 1$. Then, for such $|t_1| < \zeta$ with $t_1 \notin A_1^*$, use again property (7.7) with $a = \zeta$, $W_1 = U^* \exp\{\xi_1 t_1\}$, $W_2 = \xi_2$. Deduce that, for $|t_1| < \zeta$ with $t_1 \notin A_1^*$, the set $A_2^{**}(t_1)$ is empty or finite. Next, note that

$$A_2^* \subset \{A_1^* \times (-\zeta, \zeta)\} \bigcup_{t_1 \in (-\zeta, \zeta) \setminus A_1^*} S_2(t_1), \quad \text{where } S_2(t_1) = \{t_1\} \times A_2^{**}(t_1).$$

Since the union of finitely many negligible sets is negligible, it suffices to show that the sets $A_1^* \times (-\zeta, \zeta)$ and $\bigcup_{t_1 \in (-\zeta, \zeta) \setminus A_1^*} S_2(t_1)$ are negligible. If A_1^* has zero Lebesgue measure, by Fubini's Theorem, its product with any interval $(-\zeta, \zeta)$ has also zero Lebesgue measure. Next, let us point out that if $A \subset \mathbb{R}^d$ is a nowhere dense set, the sets $(-\zeta, \zeta) \times A$ and $A \times (-\zeta, \zeta)$ are nowhere dense in \mathbb{R}^{d+1} . Indeed, we have $\text{Int}(A \times (-\zeta, \zeta)) = \text{Int}(A) \times \text{Int}([-\zeta, \zeta]) = \emptyset \times (-\zeta, \zeta) = \emptyset$. Deduce that $A_1^* \times (-\zeta, \zeta)$ is nowhere dense. (Here, $\text{Int}(A)$ denotes the interior of the set A while \bar{A} denotes its closure.) Concerning the set $\bigcup_{t_1 \in (-\zeta, \zeta) \setminus A_1^*} S_2(t_1)$, using Fubini's Theorem, the Lebesgue measure of this set is clearly equal to zero. Next, if by definition $S_2(t_1) = \emptyset$ for $t_1 \in A_1^*$, then it suffices to study a set under the form $S_2 = \bigcup_{t_1 \in (-\zeta, \zeta)} S_2(t_1)$ with each $S_2(t_1) \subset (-\zeta, \zeta)$ an empty or finite set. First, let us notice that the set $S_2^c = (-\zeta, \zeta)^2 \setminus S_2$ is open. Indeed, if $(\bar{t}_1, \bar{t}_2) \notin S_2$, then $\bar{t}_2 \notin S_2(\bar{t}_1)$. Given that $S_2(\bar{t}_1)$ is at most finite and the map $(t_1, t_2) \mapsto \mathbb{E}(U^* \exp\{\xi_1 \bar{t}_1\} \exp\{\xi_2 t_2\})$ is continuous, there exists $c > 0$ and $\delta > 0$ such that $|\mathbb{E}(U^* \exp\{\xi_1 t_1\} \exp\{\xi_2 t_2\})| > c$ for any (t_1, t_2) such that $|t_1 - \bar{t}_1|, |t_2 - \bar{t}_2| < \delta$. This means that for any $(\bar{t}_1, \bar{t}_2) \in S_2^c$, there exists a square neighborhood of (\bar{t}_1, \bar{t}_2) contained in S_2^c , and thus S_2^c is an open set. Next, let us notice that $(-\zeta, \zeta)^2 \subset \overline{S_2^c} \cap (-\zeta, \zeta)^2$. To check this it suffices to note that for any $(t_1, t_2) \in S_2$ there exists some c close to zero such that, for each t'_2 between t_2 and $t_2 + c$, $(t_1, t'_2) \notin S_2$ but $(t_1, t'_2) \in S_2^c$. Then, necessarily, $(t_1, t_2) \in \overline{S_2^c} \cap (-\zeta, \zeta)^2$ and thus the whole square $(-\zeta, \zeta)^2$ is contained in that closure. In particular this shows that S_2 is contained in the boundary of the open set S_2^c and is thus nowhere dense. Finally, repeat the same type of arguments with $l = 3, \dots, p$. \square

Proof of Corollary 2.2. By Lemma 2.1, if the null hypothesis holds true, $\forall p \geq 1$ and $\forall \gamma \in B_p$, $\mathbb{E}(U | \langle X, \gamma \rangle) = 0$ a.s. Thus $\mathbb{E}[U \mathbb{E}(U | \langle X, \gamma \rangle) w_p(\gamma, \langle X, \gamma \rangle)] = 0$. Conversely, first note that for any p, γ and any function $w_p(\cdot, \cdot)$,

$$\mathbb{E}[U \mathbb{E}(U | \langle X, \gamma \rangle) w_p(\gamma, \langle X, \gamma \rangle)] = \mathbb{E}[\mathbb{E}^2(U | \langle X, \gamma \rangle) w_p(\gamma, \langle X, \gamma \rangle)].$$

Next, for any $p \geq 1$, if $\max_{\gamma \in B_p} \mathbb{E}[U \mathbb{E}(U | \langle X, \gamma \rangle) w_p(\gamma, \langle X, \gamma \rangle)] = 0$, then necessarily $\mathbb{E}(U | \langle X, \gamma \rangle) = 0$ a.s., $\forall \gamma \in B_p$. Then, by Lemma 2.1-B and the fact that B_p is not negligible, necessarily $\mathbb{P}(\mathbb{E}(U | X) = 0) = 1$. \square

Proof of Lemma 3.1. By the Lipschitz property of K , $\forall \gamma, \gamma' \in \mathcal{S}^p$,

$$|K_h^l(\langle X_i - X_j, \gamma \rangle) - K_h^l(\langle X_i - X_j, \gamma' \rangle)| \leq C \|K\|_\infty^{l-1} h^{-1} \|\gamma - \gamma'\| \max_{1 \leq i \leq n} \|X_i\|_{L^2},$$

for $l = 1$ or $l = 2$ and some C independent of n, h, p and γ, γ' . Deduce that

$$|Q_n(\gamma) - Q_n(\gamma')| \leq C h^{-2} \|\gamma - \gamma'\| \max_{1 \leq i \leq n} \|X_i\|_{L^2}. \quad (7.8)$$

Next, fix some $a > 3$ and let $G_p \subset \mathbb{R}^p$ be a set of n^{ap} points on \mathcal{S}^p such that $\forall \gamma \in \mathcal{S}^p \exists \gamma' \in G_p$ with $\|\gamma - \gamma'\| \leq n^{-3}$. (To build such a grid, use

for instance evenly spaced points grids on the intervals $[0, \pi]$ and $[0, 2\pi]$ and spherical coordinates.) By the construction of G_p and equation (7.8),

$$\sup_{\gamma \in B_p \subset \mathcal{S}^p} |Q_n(\gamma)| \leq \sup_{\gamma \in G_p \subset \mathcal{S}^p} |Q_n(\gamma)| + o_{\mathbb{P}}(n^{-1}).$$

Fix some $\gamma \in G_p$. For any $c > 0$ and sample size n define the set of n -uples,

$$\mathcal{E}_n(c) = \{(x_1, \dots, x_n) : x_i \in L^2[0, 1], 1 \leq i \leq n, \max(S_n(\gamma), W_n(\gamma)) \leq ch^{-1/2}\}$$

with $S_n(\gamma)$ and $W_n(\gamma)$ the non-random quantities defined in Lemma 7.2. By Lemma 7.3, $\exists c'$ (independent of n, p, h and γ) such that the covariate observations belong to $\mathcal{E}_n(c')$ with probability tending to 1, *i.e.*, $\lim_n \mathbb{P}((X_1, \dots, X_n) \in \mathcal{E}_n(c')) = 1$. Now, let $M = n^{1/4-b}$ for some small b , define $\eta_i^M = U_i \mathbb{I}_{\{|U_i| \leq M\}} - \mathbb{E}(U_i \mathbb{I}_{\{|U_i| \leq M\}} | X_i)$ and consider

$$Q_{M,n}(\gamma) = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \eta_i^M \eta_j^M K_h(\langle X_i - X_j, \gamma \rangle). \quad (7.9)$$

Let x_1, \dots, x_n be a n -uple in $\mathcal{E}_n(c')$. Let $b_n = M^{-2} n^{-1} h^{-1/2} p \ln n$. By Lemma 7.1 and Boole's inequality, there exists universal constant L and C depending only c' and $\bar{\sigma}^2$ (independent of n, h, p) such that $\forall t > 0$,

$$\begin{aligned} \mathbb{P} \left(\sup_{\gamma \in G_p} |Q_{M,n}(\gamma)| > \frac{tp \ln n}{nh^{1/2}} \mid X_1 = x_1, \dots, X_n = x_n \right) \\ \leq \max\{L, 1\} \exp[ap \ln n - L^{-1} C t p \ln n]. \end{aligned}$$

The uniform rate of $Q_{M,n}(\gamma)$ follows. Moreover, using the moment conditions on U , we show in the Supplementary Material that $\sup_{\gamma \in \mathcal{S}^p} |Q_n(\gamma) - Q_{M,n}(\gamma)| = o_{\mathbb{P}}(n^{-1} h^{-1/2} p \ln n)$.

For the variance estimator, without loss of generality, suppose $K \geq 0$. By Assumption D-(b), $\exists C > 0$ such that $\mathbb{E}[U^2 \mathbb{I}_{\{|U| \leq C\}} | X] \geq \sigma^2/2$. Let

$$g_{\gamma}(Z_i, Z_j) = U_i^2 U_j^2 C^{-4} \mathbb{I}_{\{|U_i|, |U_j| \leq C\}} K_h^2(\langle X_i - X_j, \gamma \rangle).$$

By the construction of the variance estimator (3.6),

$$\hat{v}_n^2(\gamma) \geq \frac{1}{2} [U_n(g_{\gamma_0^{(p)}}) - \mathbb{E}[g_{\gamma_0^{(p)}}(Z_1, Z_2)]] + \frac{1}{2} \mathbb{E}[g_{\gamma_0^{(p)}}(Z_1, Z_2)].$$

First, we look for a lower bound for $\mathbb{E}[g_{\gamma_0^{(p)}}(Z_1, Z_2)]$. Note that

$$\mathbb{E}[g_{\gamma_0^{(p)}}(Z_1, Z_2)] \geq \frac{\sigma^4}{4h} \mathbb{E}[K_h(\langle X_1 - X_2, \gamma_0^{(p)} \rangle)] = \frac{\sigma^4 \sqrt{2\pi}}{4} \int_{\mathbb{R}} |\mathcal{F}[f_{\gamma_0^{(p)}}](t)|^2 \mathcal{F}[K](ht) dt.$$

Let f_0 be the density of $\sum_{j=1}^{\infty} \langle X, \phi_j \rangle b_{0j}$. Since $|\mathcal{F}[K]| \leq 1/\sqrt{2\pi}$, by Cauchy-Schwarz inequality and Plancherel Theorem, and Assumption D-(d)(i)

$$\begin{aligned}
& \left| \int_{\mathbb{R}} \{|\mathcal{F}[f_0](t)|^2 - |\mathcal{F}[f_{\gamma_0^{(p)}}](t)|^2\} \mathcal{F}[K](ht) dt \right| \leq \int_{\mathbb{R}} \left| |\mathcal{F}[f_0](t)|^2 - |\mathcal{F}[f_{\gamma_0^{(p)}}](t)|^2 \right| dt \\
& \leq \left[\int_{\mathbb{R}} \left| |\mathcal{F}[f_0](t)| + |\mathcal{F}[f_{\gamma_0^{(p)}}](t)| \right|^2 dt \right]^{1/2} \left[\int_{\mathbb{R}} \left| |\mathcal{F}[f_0](t)| - |\mathcal{F}[f_{\gamma_0^{(p)}}](t)| \right|^2 dt \right]^{1/2} \\
& \leq \left[2 \int_{\mathbb{R}} \{f_0^2(t) + f_{\gamma_0^{(p)}}^2(t)\} dt \right]^{1/2} \left[\int_{\mathbb{R}} |\mathcal{F}[f_0 - f_{\gamma_0^{(p)}}](t)|^2 dt \right]^{1/2} \\
& \leq C \left[\int_{\mathbb{R}} |f_0(t) - f_{\gamma_0^{(p)}}(t)|^2 dt \right]^{1/2} \rightarrow 0, \tag{7.10}
\end{aligned}$$

for some constant C . Moreover, by dominated convergence

$$\int_{\mathbb{R}} |\mathcal{F}[f_0](t)|^2 \mathcal{F}[K](ht) dt \rightarrow \int_{\mathbb{R}} |\mathcal{F}[f_0](t)|^2 dt > 0, \quad \text{when } h \rightarrow 0. \tag{7.11}$$

It follows that $0 < C \leq \mathbb{E}[g_{\gamma_0^{(p)}}(Z_1, Z_2)]$ for some C independent of n, h, p . Next, by the Hoeffding inequality for U -statistics, $U_n(g_{\gamma_0^{(p)}}) - \mathbb{E}[g_{\gamma_0^{(p)}}(Z_1, Z_2)] = o_{\mathbb{P}}(1)$; see Supplementary Material for the details. The statement for the first variance estimator follows. The statement for the second variance estimator is a direct consequence of Lemma 7.4-(2,3). \square

The proof of Lemma 3.2 and an example of nonparametric estimator satisfying the condition (3.9) are provided in the Supplementary Material.

Proof of Theorem 3.3. By Lemma 3.2, the probabilities of the events $\{Q_n(\widehat{\gamma}_n) = Q_n(\gamma_0^{(p)})\}$ and $\{\widehat{\tau}_n^2(\gamma_0^{(p)}) = \widehat{v}_n^2(\widehat{\gamma}_n)\}$, with $\widehat{\tau}_n^2(\cdot)$ and $\widehat{v}_n^2(\cdot)$ defined in (3.5) and (3.6), both converge to 1. Hence it suffices to derive the asymptotic normal distribution of $nh^{1/2}Q_n(\gamma_0^{(p)})/\widehat{v}_n^2(\gamma_0^{(p)})$ under H_0 , where $\widehat{v}_n^2(\cdot)$ is any of the variance estimators we propose. For this purpose we use the CLT in Lemma 2 of Guerre and Lavergne (2005). Let $W_n(\gamma)$ be the zero diagonal $n \times n$ -matrix with generic element $K_h(\langle X_i - X_j, \gamma \rangle)/(h n(n-1))$ for $i \neq j$. To apply CLT in Lemma 2 of Guerre and Lavergne (2005) it suffices to have

$$\|W_n(\gamma_0^{(p)})\|_2 = O_{\mathbb{P}}(n^{-1}) \quad \text{and} \quad nh^{1/2}\|W_n(\gamma_0^{(p)})\|_F \asymp_{\mathbb{P}} 1. \tag{7.12}$$

Since $\|W_n(\gamma_0^{(p)})\|_2 \leq [n(n-1)]^{-1/2} \max_i S_n(\gamma_0^{(p)}, \langle X_i, \gamma_0^{(p)} \rangle)$ with $S_n(\gamma, t)$ defined in equation (7.2), the first rate in (7.12) follows from Lemma 7.3-(2,4). By Lemma 7.3-(1b,3b), $nh^{1/2}\|W_n(\gamma_0^{(p)})\|_F = O_{\mathbb{P}}(1)$. To show that $nh^{1/2}\|W_n(\gamma_0^{(p)})\|_F$ is bounded away from zero in probability, recall that $\mathbb{E}[h^{-1}K_h(\langle X_1 - X_2, \gamma_0^{(p)} \rangle)] = \int_{\mathbb{R}} |\mathcal{F}[f_{\gamma_0^{(p)}}](t)|^2 \mathcal{F}[K](ht) dt$ and use equations (7.10) and (7.11) above. Now, $\forall n$ we can apply Lemma 2 of Guerre and Lavergne (2005) with the observations $(U_i, \langle X_i, \gamma_0^{(p)} \rangle)$. The conditional variance of $nh^{1/2}Q_n(\gamma_0^{(p)})$ given the $\langle X_i, \gamma_0^{(p)} \rangle$'s

is $v_n(\gamma_0^{(p)})$ defined in (7.4). Deduce that $\exists C$ depending only on $\underline{\sigma}^2$ and $\bar{\sigma}^2$ such that for n sufficiently large,

$$\begin{aligned} \sup_{z \in \mathbb{R}} \left| \mathbb{P} \left[nh^{1/2} Q_n(\gamma_0^{(p)}) / v_n(\gamma_0^{(p)}) \leq z \mid \langle X_1, \gamma_0^{(p)} \rangle, \dots, \langle X_n, \gamma_0^{(p)} \rangle \right] \right. \\ \left. - \mathbb{P} [N(0, 1) \leq z] \right| \leq C (\|W(\gamma_0^{(p)})\|_2 / \|W(\gamma_0^{(p)})\|_F)^{1/4} = h^{1/8} O_{\mathbb{P}}(1). \end{aligned}$$

Finally, to deduce the asymptotic law of our test statistic with any of the variance estimators, we propose it suffices to apply Lemma 7.4. \square

Proof of Theorem 3.4. The proof is based on inequality (3.10). Since $\mathbb{E}(U^2 \mid X) \geq \underline{\sigma}^2 + r_n^2 \delta^2(X)$, $\mathbb{E}(U \mid X) = r_n \delta(X)$, and $\text{Var}(U \mid \langle X, \tilde{\gamma} \rangle) \geq \underline{\sigma}^2 + r_n^2 \text{Var}(\delta(X) \mid \langle X, \tilde{\gamma} \rangle)$, clearly the variance estimate $\hat{v}_n^2(\tilde{\gamma})$ stays away from zero. It is easy to check that the moment conditions on $\delta^2(\cdot)$ guarantee that $\hat{v}_n^2(\tilde{\gamma})$ is also bounded in probability. Hence it suffices to look at the behavior of $Q_n(\gamma)$. By Lemma 2.1-(B) there exists p_0 and $\tilde{\gamma} \in B_{p_0} \subset \mathcal{S}^{p_0}$ (p_0 and $\tilde{\gamma}$ independent of n) such that $\mathbb{E}[\delta(X) \mid \langle X, \tilde{\gamma} \rangle] \neq 0$. Since $\max_{\gamma \in B_p} Q_n(\gamma) \geq Q_n(\tilde{\gamma})$, $\forall p \geq p_0$, it suffices to investigate the rate of $Q_n(\tilde{\gamma})$. We can write

$$\begin{aligned} Q_n(\tilde{\gamma}) &= \frac{1}{n(n-1)h} \sum_{i \neq j} U_i^0 U_j^0 K_{h,ij}(\tilde{\gamma}) + \frac{2r_n}{n(n-1)h} \sum_{i \neq j} U_i^0 \delta(X_j) K_{h,ij}(\tilde{\gamma}) \\ &\quad + \frac{r_n^2}{n(n-1)h} \sum_{i \neq j} \delta(X_i) \delta(X_j) K_{h,ij}(\tilde{\gamma}) \\ &=: Q_{0n}(\tilde{\gamma}) + 2r_n Q_{1n}(\tilde{\gamma}) + r_n^2 Q_{2n}(\tilde{\gamma}). \end{aligned} \tag{7.13}$$

Since $\tilde{\gamma}$ is fixed (and of finite dimension), $Q_{0n}(\tilde{\gamma}) = O_{\mathbb{P}}(n^{-1}h^{-1/2})$ (cf. proof of Theorem 3.3). Next we show that $Q_{1n}(\tilde{\gamma}) = O_{\mathbb{P}}(n^{-1/2})$. First we can use Condition (1) and the Marcinkiewicz-Zygmund inequality and proceed as in the proof of equations (25) and (28) in Lavergne and Patilea (2008). Alternatively, $Q_{1n}(\tilde{\gamma})$ can be decomposed as the sum of a degenerate U -statistic of order 2 with the rate $O_{\mathbb{P}}(h^{-1/2}n^{-1}) = O_{\mathbb{P}}(n^{-1/2})$ and

$$\frac{1}{n} \sum_{1 \leq i \leq n} U_i^0 \mathbb{E}[\delta(X_j) h^{-1} K_{h,ij}(\tilde{\gamma}) \mid X_i].$$

Hence it suffices to bound $v_n^2 = \mathbb{E}\{(U_i^0)^2 \mathbb{E}^2[\delta(X_j) h^{-1} K_{h,ij}(\tilde{\gamma}) \mid X_i]\}$. Condition (2) implies that the map $x \mapsto \mathbb{E}[\delta(X_j) h^{-1} K_h(\langle x - X_j, \tilde{\gamma} \rangle)]$ is bounded. This, combined with the bounded conditional variance of U_i^0 , yield $v_n^2 \leq c$ for some constant $c > 0$. If condition (3) is met, let $V_i = \langle X_i, \tilde{\gamma} \rangle$, $v = \langle x, \tilde{\gamma} \rangle$ and $\bar{\delta}(V_j) = \mathbb{E}[\delta(X_j) \mid V_j]$. Then using the inverse Fourier transform we have

$$\begin{aligned} \mathbb{E}[\delta(X_j) h^{-1} K_h(V_i - V_j) \mid X_i = x] &= \mathbb{E}[\delta(X_j) h^{-1} K_h(v - V_j)] \\ &= \mathbb{E} \left[\frac{\bar{\delta}(V_j)}{\sqrt{2\pi}} \int \exp\{it(v - V_j)\} \mathcal{F}[K](ht) dt \right] = \int_{\mathbb{R}} \exp\{itv\} \mathcal{F}[\bar{\delta}f_{\tilde{\gamma}}](t) \mathcal{F}[K](ht) dt. \end{aligned}$$

Use the fact that $\mathcal{F}[\bar{\delta}f_{\tilde{\gamma}}] \in L^1(\mathbb{R})$, the fact that $\mathcal{F}[K](ht) \rightarrow \mathcal{F}[K](0) = 1/\sqrt{2\pi}$ as $h \rightarrow 0$, $|\mathcal{F}[K](ht)| \leq \mathcal{F}[|K|](0)$, and Lebesgue dominated convergence theorem to deduce that the map $v \mapsto \mathbb{E}[\delta(X_j)h^{-1}K_h(v - V_j)]$ is bounded by a constant (and converges to $\bar{\delta}(v)f_{\tilde{\gamma}}(v)$). Hence v_n^2 is bounded by a constant. Deduce that with any of the conditions (1) to (3), $Q_{1n}(\tilde{\gamma}) = O_{\mathbb{P}}(n^{-1/2})$. Finally, it is easy to show that $\text{Var}[Q_{2n}(\tilde{\gamma})] \rightarrow 0$ (see, e.g., the proof of Lemma 7.4). It remains to study

$$\mathbb{E}[Q_{2n}(\tilde{\gamma})] = (2\pi)^{1/2} \int_{\mathbb{R}} |\mathcal{F}[\bar{\delta}f_{\tilde{\gamma}}]|^2(t) \mathcal{F}[K](ht) dt.$$

If condition (1) or (2) holds true, $\bar{\delta}f_{\tilde{\gamma}} \in L^2(\mathbb{R})$ and by the Plancherel theorem and the Lebesgue dominated convergence theorem, $\mathbb{E}[Q_{2n}(\tilde{\gamma})] \rightarrow \int_{\mathbb{R}} |\bar{\delta}f_{\tilde{\gamma}}|^2 > 0$. If condition (3) is met, $\mathcal{F}[\bar{\delta}f_{\tilde{\gamma}}] \in L^2(\mathbb{R})$. Hence $\bar{\delta}f_{\tilde{\gamma}} \in L^2(\mathbb{R})$ and we can continue with the same arguments. Deduce that with any of the Conditions (1) to (3), $Q_{2n}(\tilde{\gamma}) \asymp O_{\mathbb{P}}(1)$. Collecting the rates, we obtain the result. \square

Proof of Corollary 3.5. Let $X_{ij} = X_i - X_j$ and for $l = 1$ or $l = 2$ let $\Lambda_n^{(l)}(\gamma)$ be the $n \times n$ -matrix of generic element $\Lambda_{n,ij}^{(l)}(\gamma) = K_h^l(\langle X_{ij}, \gamma \rangle_n) - K_h^l(\langle X_{ij}, \gamma \rangle)$. First we consider the null hypothesis H_0 . For any $\gamma \in \mathcal{S}^p$, let

$$\Xi_n(\gamma) = \frac{1}{n(n-1)h} \sum_{1 \leq i \neq j \leq n} U_i U_j \Lambda_{n,ij}^{(1)}(\gamma).$$

Follow the steps of the proof of Lemma 3.1 with $Q_n(\gamma)$ replaced by $\Xi_n(\gamma)$ and consider the truncated version $\Xi_{n,M}(\gamma)$ analogously to $Q_{n,M}(\gamma)$ defined in equation (7.9). Apply Lemmas 7.1 and 7.2 and take into account the rates (7.5) and (7.6). Deduce that $\sup_{\gamma \in B_p} |\Xi_{n,M}(\gamma)| = o_{\mathbb{P}}(n^{-1}h^{-1/2}p \ln n)$. The tail probability of $\Xi_n(\gamma)$ is uniformly of rate $o_{\mathbb{P}}(n^{-1}h^{-1/2}p \ln n)$ by the same arguments we used for $Q_n(\gamma)$. To bound the difference between the variance estimators built as in (3.6) with $\langle \cdot, \cdot \rangle_n$ and $\langle \cdot, \cdot \rangle$, let us write

$$\sup_{\gamma \in B_p} \left| \frac{1}{n(n-1)h} \sum_{i \neq j} U_i^2 U_j^2 \Lambda_{n,ij}^{(2)}(\gamma) \right| \leq \sup_{\gamma \in B_p} \left\| \frac{\Lambda_n^{(2)}(\gamma)}{(n-1)h} \right\|_2 \frac{1}{n} \sum_{i=1}^n U_i^4 = o_{\mathbb{P}}(1),$$

where for the equality we used (7.5). Deduce that $\hat{T}_n - T_n = o_{\mathbb{P}}(1)$ under H_0 .

Under H_{1n} it remains to reconsider the terms $Q_{1n}(\tilde{\gamma})$ and $Q_{2n}(\tilde{\gamma})$ in the proof of Theorem 3.4, see equation (7.13). For the difference between $Q_{2n}(\tilde{\gamma})$ computed with $\langle \cdot, \cdot \rangle_n$ and $\langle \cdot, \cdot \rangle$, using again (7.5), we can write

$$\left| \frac{1}{n(n-1)h} \sum_{i \neq j} \delta(X_i) \delta(X_j) \Lambda_{n,ij}^{(1)}(\tilde{\gamma}) \right| \leq \left\| \frac{\Lambda_n^{(1)}(\tilde{\gamma})}{(n-1)h} \right\|_2 \frac{1}{n} \sum_{i=1}^n \delta^2(X_i) = o_{\mathbb{P}}(1).$$

Finally, consider the difference between $Q_{1n}(\tilde{\gamma})$ computed with $\langle \cdot, \cdot \rangle_n$ and $\langle \cdot, \cdot \rangle$. If condition 1 of Corollary 3.5 is met, use the Lipschitz property and deduce

795 that this difference is of rate $o_{\mathbb{P}}(n^{-1/2})$ under any of the conditions 1 to 3 of
Theorem 3.4. If condition 2 of Corollary 3.5 is met, necessarily $f_{\tilde{\gamma}}$ is bounded
and so we are under the condition 1 of Theorem 3.4. We can then follow the
lines of the proof of equation (28) and (29) of Lavergne and Patilea (2008) and
apply the Marcinkiewicz-Zygmund inequality, and use (7.5) with $\tilde{\gamma}$. Finally,
800 condition 3 of Corollary 3.5 is met, the density $f_{\tilde{\gamma}}$ is uniformly continuous and
thus necessarily bounded and so we are in the previous situation. The proof of
Corollary 3.5 is now complete. \square

Recall that without loss of generality, we can take $\mathbb{E}(X) = 0$. Moreover, for
simpler formulae, hereafter we write $\|\cdot\|$ instead of $\|\cdot\|_{L^2}$.

Proof of Lemma 4.1. By simple calculations, $\widehat{U}_i = U_i - \langle \widehat{b} - b, X_i - \bar{X}_n \rangle - \bar{U}_n$.
Moreover,

$$Q_n(\gamma; \widehat{a}, \widehat{b}) = Q_n(\gamma) - 2V_1(\gamma) - 2V_2(\gamma) + V_3(\gamma) + V_4(\gamma) + 2V_5(\gamma)$$

where

$$V_1 = \frac{\bar{U}_n \sum_{1 \leq i \neq j \leq n} U_i K_{h,ij}(\gamma)}{n(n-1)h}, \quad V_2 = \left\langle \widehat{b} - b, \frac{\sum_{i \neq j} U_i (X_j - \bar{X}_n) K_{h,ij}(\gamma)}{n(n-1)h} \right\rangle,$$

$$V_3 = \frac{1}{n(n-1)h} \sum_{i \neq j} \langle \widehat{b} - b, X_i - \bar{X}_n \rangle \langle \widehat{b} - b, X_j - \bar{X}_n \rangle K_{h,ij}(\gamma)$$

$$V_4 = \frac{\bar{U}_n^2 \sum_{i \neq j} K_{h,ij}(\gamma)}{n(n-1)h}, \quad V_5 = \bar{U}_n \left\langle \widehat{b} - b, \frac{\sum_{i \neq j} (X_j - \bar{X}_n) K_{h,ij}(\gamma)}{n(n-1)h} \right\rangle.$$

Below, we show that $\sup_{\gamma \in B_p \subset \mathcal{S}^p} nh^{1/2} |V_j| = o_{\mathbb{P}}(1)$, for $j = 1, \dots, 5$. First note
that, by the Fubini Theorem, $\mathbb{E}(\|\bar{X}_n\|^2) = n^{-1} \int_0^1 \mathbb{E}[X(t)^2] dt$ and so $\|\bar{X}_n\| =$
 $O_{\mathbb{P}}(n^{-1/2})$. For V_1 , since $\bar{U}_n = O_{\mathbb{P}}(n^{-1/2})$, by Lemma 7.6

$$\sup_{\gamma \in B_p \subset \mathcal{S}^p} nh^{1/2} |V_1| = nh^{1/2} O_{\mathbb{P}}(n^{-1/2}) O_{\mathbb{P}}(h^{-1/4} n^{-1/2} p^{1/2} \ln^{1/2} n) = o_{\mathbb{P}}(1).$$

Next,

$$V_2 = \left\langle \widehat{b} - b, \frac{1}{n(n-1)h} \sum_{i \neq j} U_i X_j K_{h,ij}(\gamma) \right\rangle - \frac{\langle \widehat{b} - b, \bar{X}_n \rangle}{n(n-1)h} \sum_{i \neq j} U_i K_{h,ij}(\gamma) = V_{21} - V_{22}$$

By the Cauchy-Schwarz inequality and Lemma 7.6,

$$\begin{aligned} nh^{1/2} \sup_{\gamma \in B_p} |V_{21}| &= nh^{1/2} O_{\mathbb{P}}(h^{-1/4} n^{-1/2} p^{1/2} \ln^{3/2} n) O_{\mathbb{P}}(\|\widehat{b} - b\|) \\ &= [n^{1/2-\zeta} h^{1/4}] n^{\zeta-\rho} p^{1/2} \ln^{3/2} n O_{\mathbb{P}}(1) = o_{\mathbb{P}}(1). \end{aligned}$$

By the Cauchy-Schwarz inequality, the rate of $\|\bar{X}_n\|$ and Lemma 7.6,

$$nh^{1/2} \sup_{\gamma \in B_p} |V_{22}| = nh^{1/2} O_{\mathbb{P}}(\|\hat{b} - b\|) O_{\mathbb{P}}(n^{-1/2}) O_{\mathbb{P}}(h^{-1/4} n^{-1/2} p^{1/2} \ln^{3/2} n).$$

805 Deduce that $\sup_{\gamma \in \mathcal{S}_p} nh^{1/2} |V_2| = o_{\mathbb{P}}(1)$. For V_3 take absolute values and use the Cauchy-Schwarz inequality to write

$$\begin{aligned} nh^{1/2} \sup_{\gamma \in B_p} |V_3| &\leq \sup_{\gamma \in B_p} \frac{nh^{1/2} \|\hat{b} - b\|^2}{n(n-1)h} \sum_{i \neq j} \{\|X_i - \bar{X}_n\|\} \{\|X_j - \bar{X}_n\|\} K_{h,ij}(\gamma) \\ &\leq nh^{1/2} \|\hat{b} - b\|^2 \left[\frac{1}{n} \sum_{i=1}^n \{\|X_i - \bar{X}_n\|\}^2 \right] \sup_{\gamma \in B_p} \|\widetilde{W}_n(\gamma)\|_F^2, \end{aligned}$$

where $\widetilde{W}_n(\gamma)$ is the matrix with generic element

$$\widetilde{W}_{n,ij}(\gamma) = K_h^{1/2}(\langle X_i - X_j, \gamma \rangle) / [n(n-1)h]^{1/2}, \quad i \neq j, \quad \text{and} \quad \widetilde{W}_{n,ii}(\gamma) = 0.$$

By easy adaptation of the arguments used for Lemma 7.3-(1b,3b), we have $\sup_{\gamma} \|\widetilde{W}_n(\gamma)\|_F^2 = o_{\mathbb{P}}(1)$. Thus $\sup_{\gamma} nh^{1/2} |V_3| = nh^{1/2} O_{\mathbb{P}}(\|\hat{b} - b\|^2) O_{\mathbb{P}}(1) = o_{\mathbb{P}}(1)$. Similar arguments apply for the uniform rates of V_4 and V_5 . \square

Proof of Lemma 4.3. Let \hat{g} (resp. \hat{g}^0) be the random function defined in (4.5) that one would obtain under the null (resp. alternative) hypothesis, *i.e.*, with covariates X_i and responses $a + \langle b, X_i \rangle + U_i^0$ (resp. $a + \langle b, X_i \rangle + \delta(X_i) + U_i^0$). We can write

$$\begin{aligned} \|\hat{g}^0 - \hat{b}\|^2 &= \sum_{j=1}^m (\hat{b}_j^0 - \hat{b}_j)^2 = \sum_{j=1}^m \hat{\theta}_j^{-2} |\langle \hat{g} - \hat{g}^0, \hat{\phi}_j \rangle|^2 \leq \sum_{j=1}^m \hat{\theta}_j^{-2} \|\hat{g} - \hat{g}^0\|^2 \|\hat{\phi}_j\|^2 \\ &= r_n^2 \int_0^1 \left(\frac{1}{n} \sum_{i=1}^n \delta(X_i) \{X_i(u) - \bar{X}_n(u)\} \right)^2 du \sum_{j=1}^m \hat{\theta}_j^{-2} =: r_n^2 \Gamma_n \sum_{j=1}^m \hat{\theta}_j^{-2}. \end{aligned}$$

By the Fubini Theorem applied twice and the Cauchy-Schwarz inequality,

$$\mathbb{E} \int_0^1 \left(\sum_{i=1}^n \delta(X_i) X_i(u) \right)^2 du = \int_0^1 \sum_{i=1}^n \mathbb{E}[\delta^2(X_i) X_i^2(u)] du \leq n \sqrt{\mathbb{E}[\delta^4(X)] \mathbb{E}[\|X\|^4]},$$

where for the equality we used the fact that $\mathbb{E}[X\delta(X)] = 0$. On the other hand, since $\mathbb{E}[\delta(X)] = 0$, $\overline{\delta(\bar{X})}_n = n^{-1} \sum_{i=1}^n \delta(X_i) = o_{\mathbb{P}}(1)$ by the law of large numbers. Recall that $\|\bar{X}_n\| = O_{\mathbb{P}}(n^{-1/2})$. Deduce

$$\int_0^1 \left(\frac{1}{n} \sum_{i=1}^n \delta(X_i) \bar{X}_n(u) \right)^2 du = \overline{\delta(\bar{X})}_n^2 \|\bar{X}_n\|^2 = o_{\mathbb{P}}(n^{-1}),$$

810 and finally that $\Gamma_n = O_{\mathbb{P}}(n^{-1})$. For the last part, Th. 1 of Hall and Horowitz (2007) provides the rate of $\int_0^1 \{\hat{g}^0(t) - b(t)\}^2 dt$. Next, Assumption P-(c) implies

$\theta_j \geq cj^{-\alpha}$ for some constant c and thus $\sum_1^m \theta_j^{-2} = O(n^{(2\alpha+1)/(\alpha+2\beta)}) = o(n)$ provided $m \asymp n^{1/(\alpha+2\beta)}$. Finally, deduce from the equations (5.6) to (5.9) of Hall and Horowitz (2007) that $\sum_{j=1}^m (\theta_j^{-2} - \widehat{\theta}_j^{-2}) = o_{\mathbb{P}}(n)$. \square

815 *Proof of Theorem 4.4.* It suffices to show that $Q_n(\widetilde{\gamma}, \widehat{a}, \widehat{b}) \asymp_{\mathbb{P}} r_n^2$ for some fixed p sufficiently large and $\widetilde{\gamma} \in B_p$. Decomposing \widehat{U}_i as in (4.8), $Q_n(\widetilde{\gamma}, \widehat{a}, \widehat{b})$ can be decomposed into 15 terms similar or identical to those analyzed in the proofs of Theorem 3.4 and Lemma 4.1. We omit the details. \square

Proof of Theorem 5.1. First consider the case of a fixed basis \mathcal{R} . Let $X_{ij} = X_i - X_j$. The idea is to show the asymptotic normality of T_n^b conditionally on the observed data adapting the steps used for the asymptotic normality for T_n . Consider the event $\mathcal{A}_n = \{\max_{1 \leq i \leq n} |U_i| \leq M\}$ with $M = n^{1/4-b}$ for some $b > 0$. Assumption D-(a) guarantees $\mathbb{P}(\mathcal{A}_n^c) \rightarrow 0$. Moreover, $\forall i, \mathbb{E}[U_i^2 \mathbb{I}_{\{|U_i| \leq M\}} | X_i] \geq \underline{\sigma}^2/2$ for M sufficiently large. Define

$$h_{i,j}^b = \frac{\zeta_i \zeta_j}{n(n-1)h} C_{n,ij}, \quad C_{n,ij} = \frac{U_i U_j \mathbb{I}_{\{|U_i|, |U_j| \leq M\}}}{M^2} K_h(\langle X_{ij}, \gamma \rangle).$$

Let $Q_n^b(\gamma)$ be the bootstrap version of $Q_n(\gamma)$, and let

$$Q_{M,n}^b(\gamma) = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} h_{i,j}^b, \quad \gamma \in \mathcal{S}^p.$$

Note that for any $t > 0$

$$\mathbb{P} \left[\sup_{\gamma} |M^{-2} Q_n^b(\gamma) - Q_{M,n}^b(\gamma)| > t \mid U_1, X_1, \dots, U_n, X_n \right] \leq \mathbb{P}(\mathcal{A}_n^c) \rightarrow 0. \quad (7.14)$$

Define the quantities A_n^b, B_n^b, C_n^b and D_n^b such as A_n to D_n but with $h_{i,j}$ replaced by $h_{i,j}^b$ and the expectations replaced by the conditional expectations given $(U_1, X_1), \dots, (U_n, X_n)$. It is easy to check that the same upper bounds as in Lemma 7.2, could be derived on the event \mathcal{A}_n . Then equation (7.14) and the exponential inequality from Lemma 7.1 applied (following the lines of the proof of Lemma 3.1) to $Q_{M,n}^b(\gamma)$ on the event \mathcal{A}_n yield,

$$\forall C > 0, \quad \mathbb{P}[\sup_{\gamma} |Q_n^b(\gamma)| > Cp \ln n / nh^{1/2} \mid U_1, X_1, \dots, U_n, X_n] \rightarrow 0,$$

820 in probability. The second part of Lemma 3.1 on the variance estimator follows by similar arguments. Deduce that $\mathbb{P}(\widehat{\gamma}_n \neq \gamma_0^{(p)} \mid U_1, X_1, \dots, U_n, X_n) \rightarrow 0$, in probability. It remains to reconsider the steps of Theorem 3.3 and to derive the rates in equation (7.12) for the matrix $W_n^b(\gamma_0^{(p)})$ with generic element $K_{h,ij}^b(\gamma_0^{(p)}) = U_i U_j K_h(\langle X_{ij}, \gamma_0^{(p)} \rangle)$. Suppose for simplicity, but without loss of
825 generality, $K \geq 0$. Let us note that

$$\sup_{j,k} \mathbb{E}[U_i^2 h^{-1} K_{h,ij}(\gamma_0^{(p)}) K_{h,ik}(\gamma_0^{(p)}) \mid X_j, X_k] \leq C < \infty. \quad (7.15)$$

Indeed, using the upper bound of $\mathbb{E}[U^2 \mid X]$ and Cauchy-Schwarz inequality,

$$\begin{aligned} & \mathbb{E}[U_i^2 h^{-1} K_{h,ij}(\gamma_0^{(p)}) K_{h,ik}(\gamma_0^{(p)}) \mid X_j, X_k] \\ & \leq \bar{\sigma}^2 \mathbb{E}^{1/2}[h^{-1} K_{h,ij}(\gamma_0^{(p)}) \mid X_j] \mathbb{E}^{1/2}[h^{-1} K_{h,ik}(\gamma_0^{(p)}) \mid X_k] \leq C, \end{aligned}$$

where for the last inequality we used Lemma 7.3-(2). Next, let \mathbb{E}_n denote the conditional expectation given X_1, \dots, X_n , note that $\mathbb{E}_n(|U_i U_j|) \leq \mathbb{E}^{1/2}(|U_i|^2 \mid X_i) \mathbb{E}^{1/2}(|U_j|^2 \mid X_j) \leq \bar{\sigma}^{2/\nu}$. Using the conditional independence between any U_i and the rest of the sample, for any $w \in \mathbb{R}^n$ with $\|w\| = 1$,

$$\begin{aligned} & \mathbb{E} \left\| W_n^b(\gamma_0^{(p)}) w \right\|^2 \leq \mathbb{E} \left[\sum_{i=1}^n \mathbb{E}(|U_i|^2 \mid X_i) \mathbb{E}_n \left(\sum_{j=1, j \neq i}^n |U_j| K_{h,ij}(\gamma_0^{(p)}) w_j \right)^2 \right] \\ & \leq \frac{\bar{\sigma}^{4/\nu}}{hn^2(n-1)^2} \mathbb{E} \left[\sum_{i,j,k=1}^n \mathbb{E} \left\{ h^{-1} K_{h,ij}(\gamma_0^{(p)}) K_{h,ik}(\gamma_0^{(p)}) \mid X_k, X_j \right\} |w_j w_k| \right] \\ & \leq \frac{C_3}{hn(n-1)^2} \sum_{j,k=1}^n |w_j w_k| \leq \frac{C_3}{(n-1)^2} \frac{1}{nh}, \end{aligned}$$

830 where C_3 is some constant. Deduce that $\|W_n^b(\gamma_0^{(p)})\|_2 = o_{\mathbb{P}}(n^{-1})$. On the other hand, by Assumption D-(b)

$$\underline{\sigma}^4 \mathbb{E}[\|W_n(\gamma_0^{(p)})\|_F^2] \leq \mathbb{E}[\|W_n^b(\gamma_0^{(p)})\|_F^2] \leq \bar{\sigma}^4 \mathbb{E}[\|W_n(\gamma_0^{(p)})\|_F^2]. \quad (7.16)$$

Moreover, by simple calculation, $\text{Var}(n^2 h \|W_n^b(\gamma_0^{(p)})\|_F^2) \rightarrow 0$. Hence it suffices to consider the upper and lower bounds of $nh^{1/2} \|W_n(\gamma_0^{(p)})\|_F$ and the result follows. In the case of an estimated basis, the equations (7.14) and (7.16) 835 remain valid. Moreover, by the arguments used in the proof of Lemma 7.5, the variance of $n^2 h \|W_n^b(\gamma_0^{(p)})\|_F^2$ still converges to zero and $\|W_n^b(\gamma_0^{(p)})\|_2$ is also of rate $o_{\mathbb{P}}(n^{-1})$, where now $W_n^b(\gamma_0^{(p)})$ is the matrix with generic element $K_{h,ij}^b(\gamma_0^{(p)}) = U_i U_j K_h(\langle X_{ij}, \gamma_0^{(p)} \rangle_n)$. Finally, it is easy to check that the arguments used above in this proof remain true if $U_i = U_i^0 + r_n \delta(X_i)$, that is in the 840 case of a sequence of alternatives. Now the proof is complete. \square

Acknowledgements

Valentin Patilea gratefully acknowledges support from the research program *New Challenges for New Data* of Genes, LCL and Fondation du Risque. César Sánchez-Sellero gratefully acknowledges support from the Spanish State Research Agency (AEI) through project MTM2016-76969-P, co-funded by the European Regional Development Fund (ERDF), and from Ensai. 845

References

- [1] BIERENS, H.J. (1990). A consistent conditional moment test of functional form. *Econometrica* **58**, 1443–1458.
- 850 [2] BOAS, R.P. JR. (1962). Inversion of Fourier and Laplace Transforms. *Amer. Math. Monthly* **69**(10), 955–960.
- [3] BOSQ, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*. Lecture Notes in Statistics 149. Springer-Verlag, New-York.
- 855 [4] BÜCHER, A., DETTE, H., AND WIECZOREK, G. (2011). Testing model assumptions in functional regression models. *J. Multivariate Anal.* **102**(10), 1472–1488.
- [5] CAI, T., AND HALL, P. (2006). Prediction in functional linear regression. *Ann. Statist.* **34**, 2159–2179.
- 860 [6] CARDOT, H., FERRATY, F., MAS, A., AND SARDA, P. (2003). Testing Hypotheses in the Functional Linear Model. *Scand. J. Stat.* **30**, 241–255.
- [7] CARDOT, H., GOIA, P., AND SARDA, P. (2004). Testing for no effect in functional linear regression models, some computational approaches. *Comm. Statist. Simulation Comput.* **33**, 179–199.
- 865 [8] CHEN, D., HALL, P., MÜLLER, H.G. (2011). Single and multiple index functional regression models with nonparametric link. *Ann. Statist.* **39**, 1720–1747.
- [9] CRAMBES, C., KNEIP, A., AND SARDA, P. (2009). Smoothing splines estimators for functional linear regression. *Ann. Statist.* **37**, 35–72.
- 870 [10] CUESTA-ALBERTOS, J. A., GARCÍA-PORTUGUÉS, E., FEBRERO-BANDE, M., AND GONZÁLEZ-MANTEIGA, W. (2019). Goodness-of-fit tests for the functional linear model based on randomly projected empirical processes. *Ann. Statist.* **47**, 439–467.
- [11] DELSOL, L., FERRATY, F., AND VIEU, P. (2011). Structural test in regression on functional variables. *J. Multivariate Anal.* **102**, 422–447.
- 875 [12] FERRATY, F., AND VIEU, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer, Berlin.
- [13] GARCÍA-PORTUGUÉS, E., GONZÁLEZ-MANTEIGA, W., AND FEBRERO-BANDE, M. (2014). A goodness-of-fit test for the functional linear model with scalar response. *J. Comput. Graph. Stat.* **23**(3), 761–778.
- 880 [14] GINÉ, E., LATAŁA, R., AND ZINN, J. (2000). Exponential and moment inequalities for U -statistics. In *High Dimensional II* 13–38. Progr. Probab. **47**. Birkhäuser, Boston.

- [15] GONZÁLEZ-MANTEIGA, W., AND CRUJEIRAS, R.M. (2013). An updated review of goodness-of-fit tests for regression models. *Test* **22(3)**, 361–411.
- 885 [16] GUERRE, E., AND LAVERGNE, P. (2005). Data-driven rate-optimal specification testing in regression models. *Ann. Statist.* **33**, 840–870.
- [17] HALL, P., AND HOROWITZ, J.L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* **35**, 70–91.
- [18] HÄRDLE, W., AND MAMMEN, E. (1993). Comparing nonparametric versus
890 parametric regression fits. *Ann. Statist.* **21**, 1296–1947.
- [19] HART, J.D. (1997). *Nonparametric smoothing and lack-of-fit tests*. Springer Series in Statistics. Springer-Verlag New York.
- [20] HILGERT, N., MAS, A., AND VERZELEN, N. (2013). Minimax adaptive tests for the functional linear model. *Ann. Statist.* **41(2)**, 838–869.
- 895 [21] HOROWITZ, J.L., AND SPOKOINY, V.G. (2001). An adaptive, rate-optimal test of a parametric model against a nonparametric alternative. *Econometrica* **69**, 599–631.
- [22] HORVÁTH, L., AND KOKOSZKA, P. (2012). *Inference for Functional Data with Applications*. Springer Series in Statistics. Springer, New-York.
- 900 [23] HORVÁTH, L., AND REEDER, R. (2013). A test of significance in functional quadratic regression. *Bernoulli* **19(5A)**, 2020–2051.
- [24] LAVERGNE, P. AND PATILEA, V. (2008). Breaking the curse of dimensionality in nonparametric testing. *J. Econometrics* **143** 103–122.
- [25] MAMMEN, E. (1993). Bootstrap and Wild Bootstrap for High Dimensional
905 Linear Models. *Ann. Statist.* **21** 255–285.
- [26] MÜLLER, H.G. AND STADTMÜLLER, U. (2005). Generalized functional linear models. *Ann. Statist.* **33** 774–805.
- [27] PATILEA, V., C. SÁNCHEZ-SELLERO, AND M. SAUMARD (2012). Projection-based nonparametric goodness-of-fit testing with functional co-
910 variates. arXiv:1205.5 578 [math.ST]
- [28] PATILEA, V., C. SÁNCHEZ-SELLERO, AND M. SAUMARD (2016). Testing the predictor effect on a functional response. *J. Amer. Statist. Assoc.* **111(516)**, 1684–1695.
- [29] RAMSAY, J., AND SILVERMAN, B.W. (2005). *Functional Data Analysis*,
915 2nd ed., Springer-Verlag, New York.
- [30] RUDIN, W. (1987). *Real and complex analysis*, 3rd ed., McGraw-Hill, Inc.
- [31] STUTE, W. (1997). Nonparametric models checks for regression. *Ann. Statist.* **25**, 613–641.

- [32] VAN DER VAART, A.D., AND WELLNER, J.A. (1996). *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New-York.
- [33] WANG, J.L., CHIOU, J.M., AND MÜLLER, H.G. (2016). Functional Data Analysis. *Annu. Rev. Stat. Appl.* **3(1)**, 257–295.
- [34] YAO, F., AND MÜLLER, H.G. (2010). Functional quadratic regression. *Biometrika* **97**, 49–64.