The Annals of Statistics
2019, Vol. 47, No. 1, 439–467
https://doi.org/10.1214/18-AOS1693

GOODNESS-OF-FIT TESTS FOR THE FUNCTIONAL LINEAR MODEL BASED ON RANDOMLY PROJECTED EMPIRICAL PROCESSES

BY JUAN A. CUESTA-ALBERTOS[*,1] EDUARDO GARCÍA-PORTUGUÉS[†,2,3], MANUEL FEBRERO-BANDE[‡,2] AND WENCESLAO GONZÁLEZ-MANTEIGA[‡,2]

University of Cantabria[*], *Carlos III University of Madrid*[†] *and University of Santiago de Compostela*[‡]

We consider marked empirical processes indexed by a randomly projected functional covariate to construct goodness-of-fit tests for the functional linear model with scalar response. The test statistics are built from continuous functionals over the projected process, resulting in computationally efficient tests that exhibit root-n convergence rates and circumvent the curse of dimensionality. The weak convergence of the empirical process is obtained conditionally on a random direction, whilst the almost surely equivalence between the testing for significance expressed on the original and on the projected functional covariate is proved. The computation of the test in practice involves calibration by wild bootstrap resampling and the combination of several p-values, arising from different projections, by means of the false discovery rate method. The finite sample properties of the tests are illustrated in a simulation study for a variety of linear models, underlying processes, and alternatives. The software provided implements the tests and allows the replication of simulations and data applications.

1. Introduction. The term "goodness-of-fit" was introduced at the beginning of the twentieth century by Karl Pearson and, since then, there have been an enormous amount of papers devoted to this topic: first, concentrated on fitting a model for one distribution function, and later, especially after the papers of Bickel and Rosenblatt (1973) and Durbin (1973), on more general models related with the regression function. Considering a regression model with random design

Received March 2017; revised February 2018.
[1]Supported by projects MTM2014-56235-C2-2-P and MTM2017-86061-C2-2-P from the Spanish Ministry of Economy, Industry and Competitiveness.
[2]Supported by projects MTM2013-41383-P and MTM2016-76969-P from the Spanish Ministry of Economy, Industry and Competitiveness, and the European Regional Development Fund; project 10MDS207015PR from Dirección Xeral de I + D, Xunta de Galicia; IAP network StUDyS from Belgian Science Policy.
[3]Supported in part by FPU Grant AP2010-0957 from the Spanish Ministry of Education and the Dynamical Systems Interdisciplinary Network, University of Copenhagen.
MSC2010 subject classifications. Primary 62G10, 62J05; secondary 62G09.
Key words and phrases. Empirical process, functional data, functional linear model, functional principal components, goodness-of-fit, random projections.

$Y = m(X) + \varepsilon$, the goal is to test the goodness-of-fit of a class of parametric regression functions $\mathcal{M}_\Theta := \{m_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^q\}$ to the data. This is the testing of

$$H_0 : m \in \mathcal{M}_\Theta \quad \text{vs.} \quad H_1 : m \notin \mathcal{M}_\Theta$$

in an omnibus way from a sample $\{(X_i, Y_i)\}_{i=1}^n$ from (X, Y). Here, $m(x) = \mathbb{E}[Y|X = x]$ is the regression function of Y over X, and ε is a random error centred such that $\mathbb{E}[\varepsilon|\mathbf{X}] = 0$. The literature of goodness-of-fit tests for the regression function is vast, and we refer to González-Manteiga and Crujeiras (2013) for an updated review of the topic.

Following the ideas on smoothing for testing the density function [Bickel and Rosenblatt (1973)], the pilot estimators usually considered for m were nonparametric, for example, the Nadaraya–Watson estimator [Nadaraja (1964), Watson (1964)]: $\hat{m}_h(x) := \sum_{i=1}^n W_{ni}(x) Y_i$, with $W_{ni}(x) := K((x - X_i)/h) / \sum_{j=1}^n K((x - X_j)/h)$, where K is a kernel function and h is a bandwidth parameter. Using these kinds of pilot estimators, statistical tests were given by $T_n = d(\hat{m}, m_{\hat{\boldsymbol{\theta}}})$, with d some functional distance and $\hat{\boldsymbol{\theta}}$ an estimator of $\boldsymbol{\theta}$ such that $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \mathcal{O}_{\mathbb{P}}(1)$ under H_0. Alternatively, following the paper by Durbin (1973) for testing about the distribution, the pilot estimator in the regression case was given by $I_n(x) := n^{-1} \sum_{i=1}^n \mathbb{1}_{\{X_i \le x\}} Y_i$, and the empirical estimation of the integrated regression function was then $I(x) := \mathbb{E}[\mathbb{1}_{\{X \le x\}} Y]$. Härdle and Mammen (1993), using $\hat{m}_h$, and Stute (1997), using I_n, are key references for these two approximations in the literature, and were only the beginning of more than two hundred papers published in the last two decades [González-Manteiga and Crujeiras (2013)].

More recently, there has been a growing interest in testing possible structures in a regression setting in the presence of functional covariates:

$$Y = m(\mathbf{X}) + \varepsilon, \tag{1}$$

with $\mathbf{X}$ a random element in a functional space, for example, in the Hilbert space $\mathcal{H} = L^2[0, 1]$, and Y a scalar response. This is the context of "Functional Data Analysis", which has received increasing attention in the last decade [see, e.g., Ramsay and Silverman (2005), Ferraty and Vieu (2006), and Horváth and Kokoszka (2012)] due to the practical need to analyse data generated by high-resolution measuring devices.

A simple null hypothesis H_0 considered in the literature for model (1) is $H_0 : m(\mathbf{X}) = c$, where $c \in \mathbb{R}$ is a fixed constant, that is, the testing of significance of the covariate $\mathbf{X}$ over Y. Following some of the ideas from Ferraty and Vieu (2006) on considering pseudometrics for performing smoothing with functional data, the test by Härdle and Mammen (1993) was adapted by Delsol, Ferraty and Vieu (2011a) as

$$T_{n,h}^{\mathrm{D}} := \int \bigl(\hat{m}_h(\mathbf{x}) - \bar{Y}\bigr)^2 \omega(\mathbf{x}) \, \mathrm{d}P_{\mathbf{X}}(\mathbf{x}) = d(\hat{m}_h, \bar{Y}),$$

$$\hat{m}_h(\mathbf{x}) := \sum_{i=1}^n \left[K\left(\frac{\bar{d}(\mathbf{x}, \mathbf{X}_i)}{h}\right) Y_i \Big/ \sum_{j=1}^n K\left(\frac{\bar{d}(\mathbf{X}_i, \mathbf{X}_j)}{h}\right) \right],$$

with $\bar{d}$ a functional pseudometric, K a kernel function adapted to this situation, h a bandwidth parameter, ω a weight function, and $P_{\mathbf{X}}$ the probability measure induced by $\mathbf{X}$ in $\mathcal{H}$. Testing H_0 has also been considered by Cardot et al. (2003) and Hilgert, Mas and Verzelen (2013), not in an omnibus way, but inside a Functional Linear Model (FLM): $m(\mathbf{X}) = \langle \mathbf{X}, \boldsymbol{\rho} \rangle$, where $\langle \cdot, \cdot \rangle$ represents the inner product in $\mathcal{H}$ and $\boldsymbol{\rho} \in \mathcal{H}$ is the FLM parameter. For both approximations, omnibus or not, there have also been other papers which consider the functional response case; see, for example, Chiou and Müller (2007), Kokoszka et al. (2008), and Bücher, Dette and Wieczorek (2011).

The generalization of the hypothesis $H_0 : m(\mathbf{X}) = c$ to the general case

$$H_0 : m \in \mathcal{M}_{\mathcal{P}} = \{m_{\boldsymbol{\rho}} : \boldsymbol{\rho} \in \mathcal{P}\} \quad \text{vs.} \quad H_1 : m \notin \mathcal{M}_{\mathcal{P}}, \tag{2}$$

where $\mathcal{P}$ can be of either finite or infinite dimension, has been the focus of very few papers, particularly in the context of omnibus goodness-of-fit tests. In Delsol, Ferraty and Vieu (2011b), a discussion is given, without theoretical results, for the extension of the testing of a more complex null hypothesis, such as an FLM. Only one paper is known to us in which the FLM hypothesis is analysed with theoretical results. In Patilea, Sánchez-Sellero and Saumard (2012), motivated by the smoothing test statistic considered by Zheng (1996) for finite dimensional covariates, a test based on

$$T_{n,h}^{\mathrm{P}} := \frac{1}{n(n-1)} \sum_{1 \le i \neq j \le n} \big(Y_i - \hat{m}_{H_0}(\mathbf{X}_i)\big)\big(Y_j - \hat{m}_{H_0}(\mathbf{X}_j)\big) \\ \times \frac{1}{h} K\left(\frac{F_{n,\mathbf{h}}(\langle \mathbf{X}_i, \mathbf{h}\rangle) - F_{n,\mathbf{h}}(\langle \mathbf{X}_j, \mathbf{h}\rangle)}{h}\right),$$

is employed for checking the null hypothesis of linearity with $\hat{m}_{H_0}(\mathbf{X}) := \langle \mathbf{X}, \hat{\boldsymbol{\rho}} \rangle$, $\hat{\boldsymbol{\rho}}$ a suitable estimator of $\boldsymbol{\rho}$, and $F_{n,\mathbf{h}}$ the empirical distribution function of $\{\langle \mathbf{X}_i, \mathbf{h}\rangle\}_{i=1}^n$. In the same spirit, Lavergne and Patilea (2008) developed a test for the finite dimensional context, and Patilea, Sánchez-Sellero and Saumard (2016) provided a test for functional response. From a different perspective, and motivated by the test given by Escanciano (2006) for finite dimensional predictors, García-Portugués, González-Manteiga and Febrero-Bande (2014) constructed a test from the marked empirical process $I_{n,\mathbf{h}}(x) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{\langle \mathbf{X}_i, \mathbf{h}\rangle \le x\}} Y_i$, with $x \in \mathbb{R}$, and $\mathbf{h} \in \mathcal{H}$. The test statistic averages the Cramér–von Mises norm of $I_{n,\mathbf{h}}$ over a finite-dimensional, estimation-driven space of random directions $\mathbf{h}$. Although this approach circumvents the technical difficulties that a marked empirical process indexed by $\mathbf{x} \in \mathcal{H}$ would represent [a possible functional extension of the process given in Stute (1997)], no results on the convergence of the statistic are available.

In this paper, we consider marked empirical processes indexed by random projections of the functional covariate. The motivation stems from the almost surely characterization of the null hypothesis (2) via a *projected hypothesis* that arises from the conditional expectation on the projected functional covariate. This allows, conditionally on a randomly chosen $\mathbf{h}$, the study of the weak convergence of the process $I_{n,\mathbf{h}}(x)$ for hypothesis testing with infinite-dimensional covariates and parameters. As a by-product, we obtain root-n goodness-of-fit tests that evade the curse of dimensionality and, contrary to smoothing-based tests, do not rely on a tuning parameter. In particular, we focus on the testing of the aforementioned hypothesis of functional linearity where, contrary to the finite dimensional situation, the functional estimator has a nontrivial effect on the limiting process and requires careful regularization. The test statistics are built by a continuous functional (Kolmogorov–Smirnov or Cramér–von Mises) over the empirical process and are effectively calibrated by a wild bootstrap on the residuals. To account for a higher power and less influence from $\mathbf{h}$, we consider a number K (not to be confused with a kernel function) of different random directions and merge the resulting p-values into a final p-value by means of the False Discovery Rate (FDR) of Benjamini and Yekutieli (2001). The empirical analysis reports a competitive performance of the test in practice, with a low impact of the choice of K above a certain bound, and an expedient computational complexity of $\mathcal{O}(n)$ that yields notable speed improvements over García-Portugués, González-Manteiga and Febrero-Bande (2014).

The rest of the paper is organized as follows. The characterization of the null hypothesis through the projected predictor is addressed in Section 2, together with an application for the testing of the null hypothesis $H_0 : m = m_0$ (Section 2.1). Section 3 is devoted to testing the composite hypothesis $H_0 : m \in \{\langle\cdot, \boldsymbol{\rho}\rangle : \boldsymbol{\rho} \in \mathcal{H}\}$. To that aim, the regularized estimator for $\boldsymbol{\rho}$ of Cardot, Mas and Sarda (2007), $\hat{\boldsymbol{\rho}}$, is reviewed in Section 3.1. The pointwise asymptotic distribution of the projected process is studied in Section 3.2, whereas Section 3.3 gives its weak convergence. Section 4 describes the implementation of the test and other practicalities. Section 5 illustrates the finite sample properties of the test through a simulation study and with some real data applications. Some final comments and possible extensions are given in Section 6. The Appendix presents the main proofs, whereas the Supplementary Material [Cuesta-Albertos et al. (2019)] contains the auxiliary lemmas and further results from the simulation study.

1.1. *General setting and notation.* Some of the general setting and notation considered in the paper are introduced now, while more specific notation will be introduced when required. The random variable (r.v.) $\mathbf{X}$ belongs to a separable Hilbert space $\mathcal{H}$ endowed with the inner product $\langle\cdot,\cdot\rangle$ and associated norm $\|\cdot\|$. The space $\mathcal{H}$ is a general real Hilbert space, but for simplicity, it can be regarded as $\mathcal{H} = L^2[0, 1]$. Y and $\mathbf{X}$ are assumed to be centred r.v.'s providing an independent and identically distributed (i.i.d.) sample $\{(\mathbf{X}_i, Y_i)\}_{i=1}^n \subset \mathcal{H} \times \mathbb{R}$. ε is a centred r.v. with variance σ_ε^2 that is independent from $\mathbf{X}$ (the independence between ε and $\mathbf{X}$ is a technical assumption required for proving Lemmas A.4 and A.5, while for the rest of the paper it suffices that $\mathbb{E}[\varepsilon|\mathbf{X}] = 0$). Given the $\mathcal{H}$-valued r.v. $\mathbf{X}$

and $\mathbf{h} \in \mathcal{H}$, we denote by $\mathbf{X}^{\mathbf{h}} := \langle \mathbf{X}, \mathbf{h} \rangle$ the projected $\mathbf{X}$ in the direction $\mathbf{h}$, by $F_{\mathbf{h}}$ the distribution function of $\mathbf{X}^{\mathbf{h}}$, and by $P_{\mathbf{X}}$ the probability measure of $\mathbf{X}$ in $\mathcal{H}$. Bold letters are used for vectors in $\mathcal{H}$ (mainly) or column vectors in $\mathbb{R}^p$ (whose transposition is denoted by $'$), and the type is clearly determined by the context. Capital letters represent r.v.'s defined on the same probability space (Ω, σ, ν) and $\sim$ denotes equality in distribution. Weak convergence is denoted by $\stackrel{\mathcal{L}}{\rightsquigarrow}$ and $D(\mathbb{R})$ represents the Skorohod's space of *càdlàg* functions defined on $\mathbb{R}$. Finally, we shall implicitly assume that the null hypotheses stated hold almost surely (a.s.).

2. Hypothesis projection. The pillar of the goodness-of-fit tests we present is the a.s. characterization of the null hypothesis (2), re-expressed as $H_0 : \mathbb{E}[Y - m_{\boldsymbol{\rho}}(\mathbf{X})|\mathbf{X}] = 0$ for some $\boldsymbol{\rho} \in \mathcal{P}$, by means of the associated *projected hypothesis on* $\mathbf{h} \in \mathcal{H}$, defined as $H_0^{\mathbf{h}} : \mathbb{E}[Y - m_{\boldsymbol{\rho}}(\mathbf{X})|\mathbf{X}^{\mathbf{h}}] = 0$. In the following, we identify $Y - m_{\boldsymbol{\rho}}(\mathbf{X})$ by Y for the sake of simplicity in notation. In this section, we give two necessary and sufficient conditions based on the projections of $\mathbf{X}$ such that $\mathbb{E}[Y|\mathbf{X}] = 0$ holds a.s.

The first condition only requires the integrability of Y, but the condition needs to be satisfied for every direction $\mathbf{h}$.

PROPOSITION 2.1. *Assume that* $\mathbb{E}[|Y|] < \infty$. *Then*

$$\mathbb{E}[Y|\mathbf{X}] = 0 \quad \text{a.s.} \quad \iff \quad \mathbb{E}\big[Y|\mathbf{X}^{\mathbf{h}}\big] = 0 \quad \text{a.s. for every } \mathbf{h} \in \mathcal{H}.$$

The second condition, more adequate for application, *somehow* generalizes Proposition 2.1, as it only needs to be satisfied for a randomly chosen $\mathbf{h}$. In exchange, it holds only under some additional conditions on the moments of $\mathbf{X}$ and Y. Before stating it, we need some preliminary results, the first taken from Cuesta-Albertos, Fraiman and Ransford (2007) and included here for the sake of completeness.

LEMMA 2.2 [Theorem 4.1 in Cuesta-Albertos, Fraiman and Ransford (2007)]. *Let* μ *be a nondegenerate Gaussian measure on* $\mathcal{H}$ *and* $\mathbf{X}_1, \mathbf{X}_2$ *be two* $\mathcal{H}$*-valued r.v.'s defined on* (Ω, σ, ν). *Assume that*:

(a) $m_k := \int \|\mathbf{X}_1\|^k \,\mathrm{d}\nu < \infty$, *for all* $k \geq 1$, *and* $\sum_{k=1}^{\infty} m_k^{-1/k} = \infty$.
(b) *The set* $\{\mathbf{h} \in \mathcal{H} : \mathbf{X}_1^{\mathbf{h}} \sim \mathbf{X}_2^{\mathbf{h}}\}$ *is of positive* μ*-measure.*

Then $\mathbf{X}_1 \sim \mathbf{X}_2$.

REMARK 2.2.1. The Gaussianity of μ in Lemma 2.2 is not strictly required. It can be replaced by assuming a certain smoothness condition on μ [see, for instance, Theorem 2.5 and Example 2.6 in Cuesta-Albertos et al. (2007)].

REMARK 2.2.2. Assumption (a) in Lemma 2.2 is not of a technical nature. According to Theorem 3.6 in Cuesta-Albertos, Fraiman and Ransford (2007), it becomes apparent that a similar condition is required. This assumption is satisfied if the tails of $P_{\mathbf{X}_1}$ are light enough or if $\mathbf{X}_1$ has a finite moment generating function in a neighbourhood of zero.

LEMMA 2.3. *If* $\mathbb{E}[Y^2] < \infty$ *and* $\mathbf{X}$ *satisfies* (a) *in Lemma* 2.2, *then* $l_k := \mathbb{E}[\|\mathbf{X}\|^k |Y|] < \infty$ *for all* $k \geq 1$, *and* $\sum_{k=1}^{\infty} l_k^{-1/k} = \infty$.

The second condition, and most important result in this section, is given as follows.

THEOREM 2.4. *Let* μ *be a nondegenerate Gaussian measure on* $\mathcal{H}$. *Assume that* $\mathbf{X}$ *satisfies* (a) *in Lemma* 2.2 *and that* $\mathbb{E}[Y^2] < \infty$. *If we denote* $\mathcal{H}_0 := \{\mathbf{h} \in \mathcal{H} : \mathbb{E}[Y|\mathbf{X}^{\mathbf{h}}] = 0 \text{ a.s.}\}$, *then*

$$\mathbb{E}[Y|\mathbf{X}] = 0 \quad \text{a.s.} \quad \iff \quad \mathcal{H}_0 \quad \text{has positive } \mu\text{-measure}.$$

COROLLARY 2.5. *Under the assumptions of the previous theorem*,

$$\mathbb{E}[Y|\mathbf{X}] = 0 \quad \text{a.s.} \quad \iff \quad \mu(\mathcal{H}_0) = 1.$$

According to this corollary, if we are interested in testing the simple null hypothesis $H_0 : \mathbb{E}[Y|\mathbf{X}] = 0$, then we can do so as follows: (i) select, at random with μ, a direction $\mathbf{h} \in \mathcal{H}$; (ii) conditionally on $\mathbf{h}$, test the projected null hypothesis $H_0^{\mathbf{h}} : \mathbb{E}[Y|\mathbf{X}^{\mathbf{h}}] = 0$. The rationale is simple yet powerful: if H_0 holds, then $H_0^{\mathbf{h}}$ also holds; if H_0 fails, then $H_0^{\mathbf{h}}$ also fails μ-a.s. In this case, with probability one, we have chosen a direction $\mathbf{h}$ for which $H_0^{\mathbf{h}}$ fails. Of course, the main advantage to testing $H_0^{\mathbf{h}}$ over testing H_0 directly is that in $H_0^{\mathbf{h}}$ the conditioning r.v. is real, which simplifies the problem substantially.

REMARK 2.5.1. The set of directions for which H_0 is not congruent with $H_0^{\mathbf{h}}$ has measure zero. A concrete example of this set is given as follows. Suppose we are interested in testing if a random p-vector $\mathbf{X}$ is Gaussian. By the Crámer–Wold device, $\mathbf{X}$ is Gaussian if and only if $\mathbf{a}'\mathbf{X}$ is Gaussian for any $\mathbf{a} \in \mathbb{R}^p$. However, by Theorem 3.6 in Cuesta-Albertos et al. (2007), it suffices that $\mathbf{h}'\mathbf{X}$ is Gaussian for a *single*, *randomly chosen* direction $\mathbf{h} \in \mathbb{R}^p$. Then the zero-measure set in which $H_0^{\mathbf{h}}$ and H_0 are incongruent (precisely, $H_0^{\mathbf{h}}$ holds, but H_0 does not) is the set of the *projection counterexamples* $\{\mathbf{a} \in \mathbb{R}^p : \mathbf{a}'\mathbf{X} \text{ is Gaussian}, \mathbf{X} \text{ is not Gaussian}\}$. For example, for $\mathbf{X} \sim (\text{Exp}(1), \mathcal{N}(0, 1))$, the set is $\{(0, \lambda) : \lambda \in \mathbb{R}\}$. Obviously, if $\mathbf{h} \in \mathbb{R}^2$ is chosen at random with a nondegenerate measure μ, it is impossible that $\mathbf{h}$ lies exactly on this line.

2.1. *Testing a simple null hypothesis*. An immediate application of Corollary 2.5 is the testing of the simple null hypothesis $H_0 : m = m_0$ via the empirical process I_n of Stute (1997). Recall that other testing alternatives can be considered on the projected covariate due to the μ-a.s. characterization. We refer to González-Manteiga and Crujeiras (2013) for a review of alternatives.

For a random sample $\{(\mathbf{X}_i, Y_i)\}_{i=1}^n$ from $(\mathbf{X}, Y)$, we consider the empirical process of the regression conditioned on the direction $\mathbf{h}$,

$$R_{n,\mathbf{h}}(x) := n^{1/2} I_{n,\mathbf{h}}(x) = n^{-1/2} \sum_{i=1}^{n} \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}} \leq x\}} Y_i, \qquad x \in \mathbb{R}.$$

Then the following result is trivially satisfied using Theorem 1.1 in Stute (1997).

COROLLARY 2.6. *Under $H_0^{\mathbf{h}}$ and $\mathbb{E}[Y^2] < \infty$, $R_{n,\mathbf{h}} \overset{\mathcal{L}}{\rightsquigarrow} \mathcal{G}_1$ in $D(\mathbb{R})$, with $\mathcal{G}_1$ a Gaussian process with zero mean and covariance function $K_1(s,t) := \int_{-\infty}^{s\wedge t} \mathbb{V}\mathrm{ar}[Y | \mathbf{X}^{\mathbf{h}} = u] \, \mathrm{d}F_{\mathbf{h}}(u)$.*

Different statistics for the testing of $H_0^{\mathbf{h}}$ can be built from continuous functionals on $R_{n,\mathbf{h}}(x)$. We shall cover this in more detail in Section 3.

EXAMPLE 2.7. Consider the FLM $Y = \langle \mathbf{X}, \boldsymbol{\rho} \rangle + \varepsilon$ in $\mathcal{H} = L^2[0,1]$, with $\mathbf{X}$ a Gaussian process with associated Karhunen–Loéve expansion (5) below, and ε independent from $\mathbf{X}$. Then $\mathbf{X}^{\mathbf{h}}$ and $\mathbf{X}^{\boldsymbol{\rho}}$ are centred Gaussians with variances $\sigma_{\mathbf{h}}^2$ and $\sigma_{\boldsymbol{\rho}}^2$, respectively, and $\mathbb{C}\mathrm{ov}[\mathbf{X}^{\mathbf{h}}, \mathbf{X}^{\boldsymbol{\rho}}] = \sum_{j=1}^{\infty} h_j \rho_j \lambda_j$, with $h_j := \langle \mathbf{h}, \mathbf{e}_j \rangle$, and $\rho_j := \langle \boldsymbol{\rho}, \mathbf{e}_j \rangle$. Hence,

$$\begin{aligned} K^1(s,t) &= \int_{-\infty}^{s\wedge t} \left(\frac{\sigma_{\boldsymbol{\rho}}^2 \sigma_{\mathbf{h}}^2 - (\sum_{j=1}^{\infty} h_j \rho_j \lambda_j)^2}{\sigma_{\mathbf{h}}^2} + \sigma_\varepsilon^2 \right) \phi(u/\sigma_{\mathbf{h}})/\sigma_{\mathbf{h}} \, \mathrm{d}u \\ &= \left(\frac{\sigma_{\boldsymbol{\rho}}^2 \sigma_{\mathbf{h}}^2 - (\sum_{j=1}^{\infty} h_j \rho_j \lambda_j)^2}{\sigma_{\mathbf{h}}^2} + \sigma_\varepsilon^2 \right) \Phi\big((s \wedge t)/\sigma_{\mathbf{h}}\big), \end{aligned}$$

where ϕ and Φ are the density and distribution functions of a $\mathcal{N}(0,1)$, respectively.

3. Testing the functional linear model. We focus now on testing the composite null hypothesis, expressed as

$$H_0 : m(\mathbf{X}) = \langle \mathbf{X}, \boldsymbol{\rho} \rangle = \mathbf{X}^{\boldsymbol{\rho}} \qquad \text{for some } \boldsymbol{\rho} \in \mathcal{H}. \tag{3}$$

According to Corollary 2.5, testing (3) is μ-a.s. equivalent to testing

$$H_0^{\mathbf{h}} : \mathbb{E}\big[(Y - \mathbf{X}^{\boldsymbol{\rho}}) | \mathbf{X}^{\mathbf{h}}\big] = 0 \qquad \text{for some } \boldsymbol{\rho} \in \mathcal{H},$$

where $\mathbf{h}$ is sampled from a nondegenerate Gaussian law μ. Again, we construct the associated empirical regression process indexed by the projected covariate following Stute (1997). Therefore, given an estimate $\hat{\boldsymbol{\rho}}$ of $\boldsymbol{\rho}$ under H_0, we consider

$$
\begin{aligned}
T_{n,\mathbf{h}}(x) &:= a_n \sum_{i=1}^{n} \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}} \le x\}}\big(Y_i - \mathbf{X}_i^{\hat{\boldsymbol{\rho}}}\big) \\
&= a_n\big(T_{n,\mathbf{h}}^1(x) + T_{n,\mathbf{h}}^2(x) + T_{n,\mathbf{h}}^3(x)\big),
\end{aligned}
\tag{4}
$$

where $a_n \to 0$ is a normalizing positive sequence to be determined later and

$$
\begin{aligned}
T_{n,\mathbf{h}}^1(x) &:= \sum_{i=1}^{n} \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}} \le x\}}\big(Y_i - \mathbf{X}_i^{\boldsymbol{\rho}}\big), \\
T_{n,\mathbf{h}}^2(x) &:= \sum_{i=1}^{n} \langle \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}} \le x\}}\mathbf{X}_i - \mathbb{E}[\mathbb{1}_{\{\mathbf{X}^{\mathbf{h}} \le x\}}\mathbf{X}], \boldsymbol{\rho} - \hat{\boldsymbol{\rho}}\rangle, \\
T_{n,\mathbf{h}}^3(x) &:= n\langle \mathbb{E}[\mathbb{1}_{\{\mathbf{X}^{\mathbf{h}} \le x\}}\mathbf{X}], \boldsymbol{\rho} - \hat{\boldsymbol{\rho}}\rangle.
\end{aligned}
$$

The selection of the right estimator $\hat{\boldsymbol{\rho}}$ has a crucial role in the weak convergence of $T_{n,\mathbf{h}}^3$, which requires a substantially more complex proof than that for the simple hypothesis. We consider the regularized estimate proposed in Sections 2 and 3 of Cardot, Mas and Sarda (2007) (subsequently denoted by CMS), whose construction is sketched here for the sake of the exposition of our results.

3.1. *Construction of the estimator of $\boldsymbol{\rho}$.* Consider the so-called Karhunen–Loéve expansion of $\mathbf{X}$:

$$
\mathbf{X} = \sum_{j=1}^{\infty} \lambda_j^{1/2} \xi_j \mathbf{e}_j. \tag{5}
$$

Here, $\{\mathbf{e}_j\}_{j=1}^{\infty}$ is a sequence of orthonormal eigenfunctions associated with the covariance operator of $\mathbf{X}$, $\Gamma\mathbf{z} := \mathbb{E}[(\mathbf{X} \otimes \mathbf{X})(\mathbf{z})]$, $\mathbf{z} \in \mathcal{H}$, and the ξ_j's are centred real r.v.'s (because $\mathbf{X}$ is centred) such that $\mathbb{E}[\xi_j \xi_{j'}] = \delta_{j,j'}$, where $\delta_{j,j'}$ is the Kronecker's delta. The Kronecker operator $\otimes$ is such that $(\mathbf{x} \otimes \mathbf{y})\mathbf{z} = \langle \mathbf{z}, \mathbf{x}\rangle \mathbf{y}$ for $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{H}$. We assume that the multiplicity of each eigenvalue is one, so $\lambda_1 > \lambda_2 > \cdots > 0$.

The functional coefficient $\boldsymbol{\rho}$ is determined by the equation $\Delta = \Gamma\boldsymbol{\rho}$, with Δ the cross-covariance operator of $\mathbf{X}$ and Y, $\Delta\mathbf{z} := \mathbb{E}[(\mathbf{X} \otimes Y)(\mathbf{z})]$, $\mathbf{z} \in \mathcal{H}$. To ensure the existence and uniqueness of a solution to $\Delta = \Gamma\boldsymbol{\rho}$, we require the next basic assumptions:

A1. $\mathbf{X}$ and Y satisfy $\sum_{j=1}^{\infty} \frac{1}{\lambda_j^2} \langle \mathbb{E}[\mathbf{X}Y], \mathbf{e}_j\rangle^2 < \infty$.

A2. The kernel of Γ is $\{\mathbf{0}\}$.

The estimation of $\boldsymbol{\rho}$ requires the inversion of $\Gamma_n := \frac{1}{n}\sum_{i=1}^n \mathbf{X}_i \otimes \mathbf{X}_i$, but since Γ_n is a.s. a finite rank operator, its inverse does not exist. CMS proposed a regularization yielding a family of continuous estimators for Γ^{-1}. Based on their Example 1, we define $\Gamma_n^\dagger$, an empirical finite rank estimate of Γ^{-1}:

$$\Gamma_n^\dagger := \sum_{i=1}^{k_n} \frac{1}{\hat{\lambda}_j} \hat{\mathbf{e}}_j \otimes \hat{\mathbf{e}}_j.$$

The construction of $\Gamma_n^\dagger$ (resp., the population version $\Gamma^\dagger := \sum_{i=1}^{k_n} \frac{1}{\lambda_j}\mathbf{e}_j \otimes \mathbf{e}_j$) is done by considering a sequence of thresholds $c_n \in (0, \lambda_1)$, $n \in \mathbb{N}$, with $c_n \to 0$. The procedure is as follows: (i) compute the Functional Principal Components (FPC) of $\mathbf{X}_1, \dots, \mathbf{X}_n$, that is, calculate the eigenvalues $\{\hat{\lambda}_j\}$ and eigenfunctions $\{\hat{\mathbf{e}}_j\}$ of Γ_n; (ii) define the sequences $\{\delta_j\}$, with $\delta_1 := \lambda_1 - \lambda_2$ and $\delta_j := \min(\lambda_j - \lambda_{j+1}, \lambda_{j-1} - \lambda_j)$ for $j > 1$, and set

$$k_n := \sup\{j \in \mathbb{N} : \lambda_j + \delta_j/2 \geq c_n\};$$

(iii) compute $\Gamma_n^\dagger$ (resp., $\Gamma^\dagger$) as the finite rank operator with the same eigenfunctions as Γ_n (resp., Γ) and associated eigenvalues equal to $\hat{\lambda}_j^{-1}$ (resp., λ_j^{-1}) if $j \leq k_n$ and 0 otherwise. The regularized estimator of $\boldsymbol{\rho}$ is then

$$\hat{\boldsymbol{\rho}} := \Gamma_n^\dagger \Delta_n = \frac{1}{n}\sum_{j=1}^{k_n}\sum_{i=1}^{n} \frac{\langle \mathbf{X}_i \otimes Y_i, \hat{\mathbf{e}}_j\rangle}{\hat{\lambda}_j}\hat{\mathbf{e}_j}. \tag{6}$$

Note that (6) is not readily computable in practice, since $\{\lambda_j\}$ is usually unknown (and hence, k_n). As in CMS, we consider the (random) finite rank

$$d_n := \sup\{j \in \mathbb{N} : \hat{\lambda}_j \geq c_n\}$$

as a replacement in practice for the deterministic k_n. As seen in Lemma A.2, $\nu[k_n = d_n] \to 1$. Hence the estimator (6) has the same asymptotic behaviour with either k_n or d_n. Therefore, we consider k_n in (6) due to the enhanced probabilistic tractability. The consideration of k_n in $\Gamma_n^\dagger$, instead of d_n, is the main difference between our definition of $\Gamma_n^\dagger$ and the proposal given from Example 1 in CMS.

The following assumptions allow us to obtain meaningful asymptotic convergences involving $\hat{\boldsymbol{\rho}}$:

A3. $\mathbb{E}[\|\mathbf{X}\|^2] < \infty$.
A4. $\sum_{l=1}^\infty |\langle \boldsymbol{\rho}, \mathbf{e}_l\rangle| < \infty$.
A5. For j large, $\lambda_j = \lambda(j)$, with $\lambda(\cdot)$ a convex positive function.
A6. $\frac{\lambda_n n^4}{\log n} = \mathcal{O}(1)$.
A7. $\inf\{|\langle \boldsymbol{\rho}, \mathbf{e}_{k_n}\rangle|, \frac{\lambda_{k_n}}{\sqrt{k_n}\log k_n}\} = \mathcal{O}(n^{-1/2})$.
A8. $\sup_j\{\max(\mathbb{E}[\xi_j^4], \mathbb{E}[|\xi_j|^5])\} \leq M < \infty$ for $M \geq 1$.
A9. There exist $C_1, C_2 > 0$ such that $C_1 n^{-1/2} < c_n < C_2 n^{-1/2}$ for every n.

A brief summary of these assumptions is given as follows. A3 is standard for obtaining asymptotic distributions, allows decomposition (5) and implies $\mathbb{E}[Y^2] < \infty$, which is required in Theorem 1.1 of Stute (1997). A4 and A5 are A.1 and A.2 in CMS. A6 is very similar to an assumption in the second part of Theorem 2 in CMS. A7 is the minimum requirement for controlling $\langle \mathbf{X}, \mathbf{L}_n \rangle$ (to be detailed in Section A.2) when Lemma 7 in CMS is used to prove Lemma A.7. A8 is a reinforcement of A.3 in CMS, where only fourth-order moments are used. This is because we handle inner products of $\hat{\boldsymbol{\rho}}$ times a nonindependent r.v., while in CMS the r.v. is not used to estimate $\boldsymbol{\rho}$. A9 is useful, mainly (but also see the final part of Lemma A.6) to control the behaviour of k_n. We show this fact in Proposition A.1, with a conclusion very close to assumption (8) in CMS and coinciding with one of the conditions of their Theorem 3 if $\lim_n t_{n,\mathbf{E}_{x,\mathbf{h}}} < \infty$ [the term $t_{n,\mathbf{E}_{x,\mathbf{h}}}$ is defined in (7) below]. Finally, we point out that in CMS the assumptions aim to control the behaviour of k_n while here we have sought to control the threshold c_n, as this can be modified by the statistician.

3.2. *Pointwise asymptotic distribution of $T_{n,\mathbf{h}}$.* Corollary 2.6 shows the weak convergence of $n^{-1/2}T^1_{n,\mathbf{h}}$. We analyse now the pointwise behaviour of $T^2_{n,\mathbf{h}}(x)$ and $T^3_{n,\mathbf{h}}(x)$ for a fixed $x \in \mathbb{R}$. We will show that $T^2_{n,\mathbf{h}}(x) = o_{\mathbb{P}}(n^{1/2})$ and that the rate of $T^3_{n,\mathbf{h}}(x)$ depends on the key normalizing sequence $\{t_{n,\mathbf{E}_{x,\mathbf{h}}}\}$, where

$$t_{n,\mathbf{x}} := \sqrt{\sum_{j=1}^{k_n} \frac{\langle \mathbf{x}, \mathbf{e}_j \rangle^2}{\lambda_j}} \quad \text{and} \quad \mathbf{E}_{x,\mathbf{h}} := \mathbb{E}[\mathbb{1}_{\{\mathbf{X}^{\mathbf{h}} \le x\}} \mathbf{X}]. \tag{7}$$

THEOREM 3.1. *Under $H_0^{\mathbf{h}}$ and A1–A9, and for a fixed $x \in \mathbb{R}$, it follows that*:

(a) $n^{-1/2} t^{-1}_{n,\mathbf{E}_{x,\mathbf{h}}} T^3_{n,\mathbf{h}}(x) \stackrel{\mathcal{L}}{\rightsquigarrow} \mathcal{N}(0, \sigma^2_\varepsilon)$.

(b) *If $\lim_n t_{n,\mathbf{E}_{x,\mathbf{h}}} = \infty$, then with $a_n = n^{-1/2} t^{-1}_{n,\mathbf{E}_{x,\mathbf{h}}}$ in (4), the asymptotic distribution of $T_{n,\mathbf{h}}(x)$ is $n^{-1/2} t^{-1}_{n,\mathbf{E}_{x,\mathbf{h}}} T^3_{n,\mathbf{h}}(x)$.*

(c) *If $\lim_n t_{n,\mathbf{E}_{x,\mathbf{h}}} < \infty$, then with $a_n = n^{-1/2}$ in (4), the asymptotic distribution of $T_{n,\mathbf{h}}(x)$ is $n^{-1/2}(T^1_{n,\mathbf{h}}(x) + T^3_{n,\mathbf{h}}(x))$.*

The behaviour of the sequence $\{t_{n,\mathbf{E}_{x,\mathbf{h}}}\}$, indexed by $n \in \mathbb{N}$ and with arbitrary $\mathbf{h} \in \mathcal{H}$ and $x \in \mathbb{R}$, is crucial for the convergence of $T_{n,\mathbf{h}}$. Since $\{t_{n,\mathbf{E}_{x,\mathbf{h}}}\}$ is nondecreasing, it has always a limit (finite or infinite). Its asymptotic behaviour is described next.

PROPOSITION 3.2. *The sequence $\{t_{n,\mathbf{E}_{x,\mathbf{h}}}\}$ has asymptotic order between $\mathcal{O}(1)$ and $\mathcal{O}(k_n^{1/2})$. In addition, if $\mathbf{X}$ is Gaussian and satisfies A3, then $\sigma^2_{\mathbf{h}} := \mathbb{V}\mathrm{ar}[\mathbf{X}^{\mathbf{h}}] < \infty$ and $\lim_n t_{n,\mathbf{E}_{x,\mathbf{h}}} = \phi(x/\sigma_{\mathbf{h}})$.*

3.3. *Weak convergence of $T_{n,\mathbf{h}}$ and the test statistics.* The result given in Theorem 3.1 holds for every $x \in \mathbb{R}$. For case (c) of Theorem 3.1 (where the estimation of $\boldsymbol{\rho}$ is not dominant) and under an additional assumption, the result can be generalized to functional weak convergence.

THEOREM 3.3. *Under $H_0^{\mathbf{h}}$, A1–A9, and (c) in Theorem 3.1, it follows that*:

(a) *The finite dimensional distributions of $T_{n,\mathbf{h}}$ converge to a multivariate Gaussian with covariance function $K_2(s,t) := K_1(s,t) + C(s,t) + C(t,s) + V(s,t)$, where*

$$C(s,t) := \int_{\{\mathbf{u}^{\mathbf{h}} \le s\}} \mathbb{V}\mathrm{ar}[Y|\mathbf{X}=\mathbf{u}]\langle \mathbf{E}_{t,\mathbf{h}}, \Gamma^{-1}\mathbf{u}\rangle \,\mathrm{d}P_{\mathbf{X}}(\mathbf{u}),$$

$$V(s,t) := \int \mathbb{V}\mathrm{ar}[Y|\mathbf{X}=\mathbf{u}]\langle \mathbf{E}_{s,\mathbf{h}}, \Gamma^{-1}\mathbf{u}\rangle\langle \mathbf{E}_{t,\mathbf{h}}, \Gamma^{-1}\mathbf{u}\rangle \,\mathrm{d}P_{\mathbf{X}}(\mathbf{u}).$$

(b) *If $\mathbb{E}[\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|^4] = \mathcal{O}(n^{-2})$, then $T_{n,\mathbf{h}} \stackrel{\mathcal{L}}{\rightsquigarrow} \mathcal{G}_2$ in $D(\mathbb{R})$, with $\mathcal{G}_2$ a Gaussian process with zero mean and covariance function K_2.*

REMARK 3.3.1. According to Theorem 1 in CMS, it is impossible for $\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}$ to converge to a nondegenerate random element in the topology of $\mathcal{H}$. To circumvent this issue and obtain the tightness of $T_{n,\mathbf{h}}$, we assume $\mathbb{E}[\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|^4] = \mathcal{O}(n^{-2})$, which implies $\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\| = \mathcal{O}_{\mathbb{P}}(n^{-1/2})$, and thus, a finite-dimensional parametric convergence rate for $\hat{\boldsymbol{\rho}}$. For instance, this happens when $\boldsymbol{\rho}$ is a linear combination of a finite number of the eigenfunctions of Γ. Notice that this is not needed for the convergence of the finite dimensional distributions of $T_{n,\mathbf{h}}$.

The next result gives the convergence of the Kolmogorov–Smirnov (KS) and Cramér–von Mises (CvM) statistics for testing the FLM.

COROLLARY 3.4. *Under the assumptions in Theorem 3.3 and $\mathbb{E}[\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|^4] = \mathcal{O}(n^{-2})$, if $\|T_{n,\mathbf{h}}\|_{\mathrm{KS}} := \sup_{x\in\mathbb{R}} |T_{n,\mathbf{h}}(x)|$ and $\|T_{n,\mathbf{h}}\|_{\mathrm{CvM}} := \int_{\mathbb{R}} T_{n,\mathbf{h}}(x)^2 \mathrm{d}F_{n,\mathbf{h}}(x)$, then*

$$\|T_{n,\mathbf{h}}\|_{\mathrm{KS}} \stackrel{\mathcal{L}}{\rightsquigarrow} \|\mathcal{G}_2\|_{\mathrm{KS}} \quad \textit{and} \quad \|T_{n,\mathbf{h}}\|_{\mathrm{CvM}} \stackrel{\mathcal{L}}{\rightsquigarrow} \int_{\mathbb{R}} \mathcal{G}_2(x)^2 \,\mathrm{d}F_{\mathbf{h}}(x).$$

REMARK 3.4.1. An alternative to (b) and Corollary 3.4 is to consider a deterministic discretization of the statistics, for which the convergence in law is trivial from (a). For example, if $\|T_{n,\mathbf{h}}\|_{\widetilde{\mathrm{KS}}} := \max_{k=1,\dots,G} |T_{n,\mathbf{h}}(x_k)|$ for a grid $\{x_1,\dots,x_G\}$, then $\|T_{n,\mathbf{h}}\|_{\widetilde{\mathrm{KS}}} \stackrel{\mathcal{L}}{\rightsquigarrow} \|\mathbf{Z}_2\|_{\widetilde{\mathrm{KS}}}$, where $\mathbf{Z}_2 \sim \mathcal{N}_G(\mathbf{0}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma}_{ij} = K_2(x_i, x_j)$.

4. Testing in practice. The major advantage of testing $H_0^{\mathbf{h}}$ over H_0 is that in $H_0^{\mathbf{h}}$ the conditioning r.v. is real. The potential drawbacks of this universal method are a possible loss of power and that the outcome of the test may vary for different projections. Both inconveniences can be alleviated by sampling several directions $\mathbf{h}_1, \ldots, \mathbf{h}_K$, testing the projected hypotheses $H_0^{\mathbf{h}_1}, \ldots, H_0^{\mathbf{h}_K}$, and selecting an appropriate way to mix the resulting p-values. For example, using the FDR method proposed in Benjamini and Yekutieli (2001) [see Section 2.2.2 of Cuesta-Albertos and Febrero-Bande (2010)], it is possible to control the final rejection rate to be *at most* α under H_0.

The drawing of random directions is clearly influential in the power of the test. For example, in the extreme case where the directions are orthogonal to the data, that is, $\mathbf{X}^{\mathbf{h}} = 0$, then $T_{n,\mathbf{h}}(x) = (n^{-1/2} \sum_{i=1}^{n} \hat{\varepsilon}_i) \mathbb{1}_{\{0 \le x\}}$ and $\|T_{n,\mathbf{h}}\|_{\mathrm{N}} = \|T_{n,\mathbf{h}}^{*b}\|_{\mathrm{N}} = 0$ under H_0. Therefore, Algorithm 4.2 would fail to calibrate the level of the test and potentially yield spurious results due to numerical inaccuracies in $\|T_{n,\mathbf{h}}^{*b}\|_{\mathrm{N}} \le \|T_{n,\mathbf{h}}\|_{\mathrm{N}}$. A data-driven compromise to avoid drawing directions in subspaces *almost* orthogonal to the data is the following: (i) compute the FPC of $\mathbf{X}_1, \ldots, \mathbf{X}_n$, that is, the eigenpairs $\{(\hat{\lambda}_j, \hat{\mathbf{e}}_j)\}$; (ii) choose $j_n := \min\{k = 1, \ldots, n-1 : (\sum_{j=1}^{k} \hat{\lambda}_j^2)/(\sum_{j=1}^{n-1} \hat{\lambda}_j^2) \ge r\}$ for a variance threshold r, for example, $r = 0.95$; (iii) generate the data-driven Gaussian process $\mathbf{h}_{j_n} := \sum_{j=1}^{j_n} \eta_j \hat{\mathbf{e}}_j$, with $\eta_j \sim \mathcal{N}(0, s_j^2)$ and s_j^2 the sample variance of the scores in the jth FPC. Without loss of generality, we will use this data-driven projecting process for drawing $\mathbf{h}$ in the rest of the paper (see the Supplementary Material [Cuesta-Albertos et al. (2019)] for the consideration of other data generating processes). Formally, the Gaussian measure μ associated with $\mathbf{h}_{j_n}$ does not respect the assumptions in Theorem 2.4, since it is degenerate (but recall that μ does not have to be independent from $\mathbf{X}$). A nondegenerate Gaussian process can be obtained as $\mathbf{h}_{j_n} + \mathcal{G}$, with $\mathcal{G}$ a Gaussian process tightly concentrated around zero, albeit employing $\mathbf{h}_{j_n}$ or $\mathbf{h}_{j_n} + \mathcal{G}$ has negligible effects in practice.

The testing procedure is described in the following generic algorithm.

ALGORITHM 4.1 (Testing procedure for H_0). Let T_n denote a test for checking $H_0^{\mathbf{h}}$ with $\mathbf{h}$ chosen by a nondegenerate Gaussian measure μ on $\mathcal{H}$:

(i) For $i = 1, \ldots, K$, set by p_i the p-value of $H_0^{\mathbf{h}_i}$ obtained with the test T_n.
(ii) Set the final p-value of H_0 as $\min_{i=1,\ldots,K} \frac{K}{i} p_{(i)}$, where $p_{(1)} \le \cdots \le p_{(K)}$.

The calibration of the test statistic for $H_0^{\mathbf{h}}$ is done by a wild bootstrap resampling. The next algorithm states the steps for testing the FLM. The particular case of the simple null hypothesis corresponds to $\boldsymbol{\rho} = \mathbf{0}$, so its calibration corresponds to setting $\hat{\boldsymbol{\rho}} = \hat{\boldsymbol{\rho}}^* = \mathbf{0}$ in the algorithm.

ALGORITHM 4.2 (Bootstrap calibration in FLM testing). Let $\{(\mathbf{X}_i, Y_i)\}_{i=1}^{n}$ be an i.i.d. sample from (1) and a given $\mathbf{h} \in \mathcal{H}$. To test (3), proceed as follows:

(i) Estimate $\boldsymbol{\rho}$ by FPC for a given d_n and obtain $\hat{\varepsilon}_i = Y_i - \langle \mathbf{X}_i, \hat{\boldsymbol{\rho}} \rangle$.
(ii) Compute $\|T_{n,\mathbf{h}}\|_{\mathrm{N}} = \|n^{-1/2} \sum_{i=1}^n \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}} \leq x\}} \hat{\varepsilon}_i\|_{\mathrm{N}}$ with N either KS or CvM.
(iii) *Bootstrap resampling*. For $b = 1, \dots, B$:

(a) Draw binary i.i.d. r.v.'s $V_1^*, \dots, V_n^*$ such that $\mathbb{P}\{V^* = (1-\sqrt{5})/2\} = (5+\sqrt{5})/10$ and $\mathbb{P}\{V^* = (1+\sqrt{5})/2\} = (5-\sqrt{5})/10$.
(b) Set $Y_i^* := \langle \mathbf{X}_i, \hat{\boldsymbol{\rho}} \rangle + \varepsilon_i^*$ using the bootstrap residuals $\varepsilon_i^* := V_i^* \hat{\varepsilon}_i$.
(c) Estimate $\boldsymbol{\rho}^*$ from $\{(\mathbf{X}_i, Y_i^*)\}_{i=1}^n$ by FPC using the same d_n used in (i).
(d) Obtain the estimated bootstrap residuals $\hat{\varepsilon}_i^* := Y_i^* - \langle \mathbf{X}_i, \hat{\boldsymbol{\rho}}^* \rangle$.
(e) Compute $\|T_{n,\mathbf{h}}^{*b}\|_{\mathrm{N}} := \|n^{-1/2} \sum_{i=1}^n \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}} \leq x\}} \hat{\varepsilon}_i^*\|_{\mathrm{N}}$.

(iv) Approximate the p-value by $\frac{1}{B} \sum_{b=1}^B \mathbb{1}_{\{\|T_{n,\mathbf{h}}^{*b}\|_{\mathrm{N}} \leq \|T_{n,\mathbf{h}}\|_{\mathrm{N}}\}}$.

Notice that the role of c_n is the selection of d_n in the estimation of $\boldsymbol{\rho}$. The selection of d_n can be done in a data-driven way by selecting from among a set of candidate d_n's the optimal one in terms of a model-selection criterion. For example, we consider the corrected Schwartz Information Criterion [McQuarrie (1999)], defined as $\mathrm{SICc}(d_n) := \ell(\hat{\boldsymbol{\rho}}_{d_n}) + \frac{\log(n) d_n}{n - d_n - 2}$, in order to overpenalize large d_n's that generate noisy estimates of $\boldsymbol{\rho}$, especially for low sample sizes. In the previous expression, $\ell(\hat{\boldsymbol{\rho}}_{d_n})$ represents the log-likelihood of the FLM for $\boldsymbol{\rho}$ estimated with d_n FPC's. Of course, this selection could be done in terms of the c_n's that determine the d_n's but, since the latter are directly related to the model complexity, its analysis is more convenient in practice. Note that steps (c) and (d) can be easily computed using the properties of the linear model; see Section 3.3 of García-Portugués, González-Manteiga and Febrero-Bande (2014).

The bootstrap process we are considering is given by (we consider $a_n = n^{-1/2}$)

$$\begin{aligned} T_{n,\mathbf{h}}^*(x) &:= n^{-1/2} \sum_{i=1}^n \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}} \leq x\}} \hat{\varepsilon}_i^* \\ &= n^{-1/2} \sum_{i=1}^n \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}} \leq x\}} \hat{\varepsilon}_i V_i^* + n^{-1/2} \sum_{i=1}^n \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}} \leq x\}} \mathbf{X}_i^{\hat{\boldsymbol{\rho}} - \hat{\boldsymbol{\rho}}^*}, \end{aligned}$$

which is estimating the distribution of

$$T_{n,\mathbf{h}}(x) = n^{-1/2} \sum_{i=1}^n \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}} \leq x\}} \hat{\varepsilon}_i + n^{-1/2} \sum_{i=1}^n \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}} \leq x\}} \mathbf{X}_i^{\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}}.$$

The bootstrap consistency could be obtained as an adaptation of Lemma A.1 of Stute, González Manteiga and Presedo Quindimil (1998) for the first term of $T_{n,\mathbf{h}}^*$, Lemma A.2 *ibid* for the second term, and using the decomposition of $\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}$ given in (11) in CMS.

5. Simulation study and data application. We illustrate the finite sample performance of the CvM and KS goodness-of-fit tests implemented using Algorithms 4.1 and 4.2 for the composite hypothesis. In order to examine the possible effects of different functional coefficients $\boldsymbol{\rho}$ and underlying processes for $\mathbf{X}$, we considered nine possible scenarios combining both factors. The detailed description of these scenarios is given in the Supplementary Material [Cuesta-Albertos et al. (2019)], while a coarse-grained graphical idea can be obtained from Figure 1.

The different data generating processes are encoded as follows. For the kth simulation scenario Sk, with functional coefficient $\boldsymbol{\rho}_k$, the deviation from H_0 is mea-

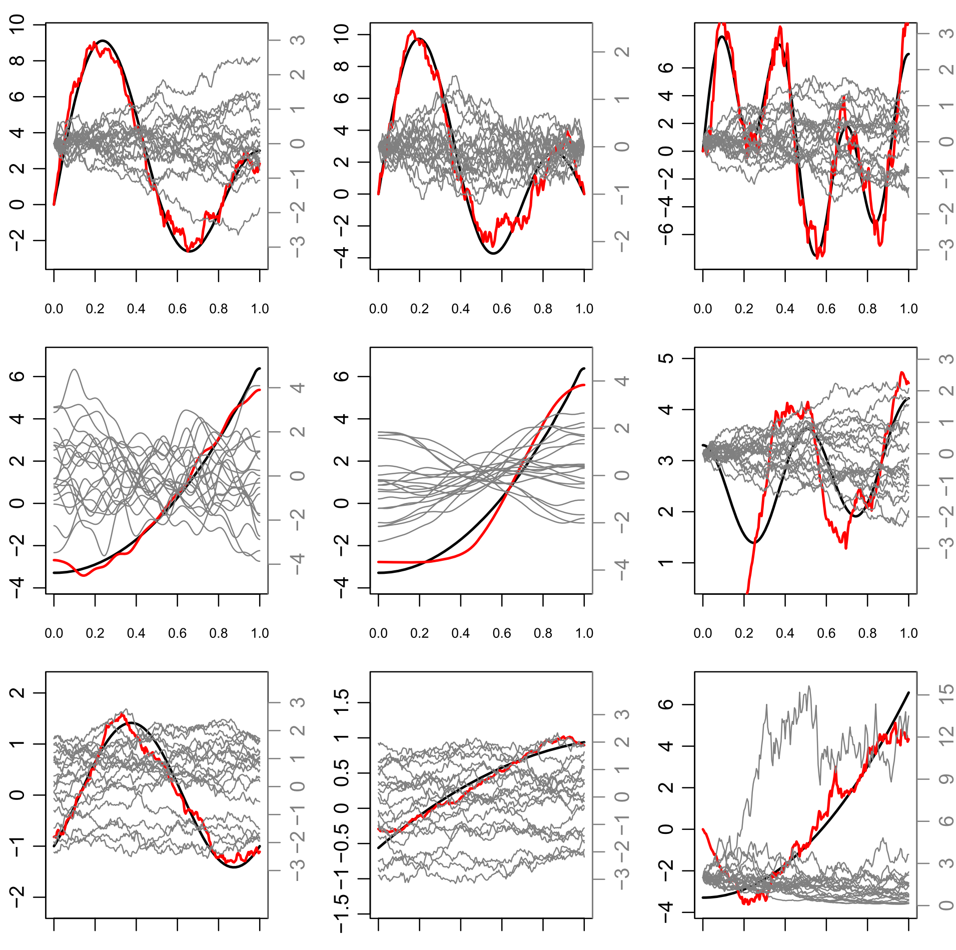

FIG. 1. *From left to right and up to down, functional coefficients* $\boldsymbol{\rho}$ (*black, right scale*) *and underlying processes* (*grey, left scale*) *for the nine different scenarios, labelled* $S1$ *to* $S9$. *Each graph shows a sample of* 100 *realizations of the functional covariate* $\mathbf{X}$ *and the estimate* $\hat{\boldsymbol{\rho}}$ (*red*) *with* d_n *selected by SICc*.

sured by a deviation coefficient δ_d, with $\delta_0 = 0$ and $\delta_d > 0$ for $d = 1, 2$. Then, with $H_{k,d}$ we denote the data generation from

$$Y = \langle \mathbf{X}, \boldsymbol{\rho}_k \rangle + \delta_d \Delta_{\eta_k}(\mathbf{X}) + \varepsilon,$$

where $\boldsymbol{\eta} := (1, 2, 1, 2, 2, 1, 2, 3, 3)'$ and the deviations from the linear model are constructed by including the nonlinear terms $\Delta_1(\mathbf{X}) := \|\mathbf{X}\|$, $\Delta_2(\mathbf{X}) := 25 \int_0^1 \int_0^1 \sin(2\pi ts)s(1-s)t(1-t)\mathbf{X}(s)\mathbf{X}(t)\,\mathrm{d}s\,\mathrm{d}t$, and $\Delta_3(\mathbf{X}) := \langle e^{-\mathbf{X}}, \mathbf{X}^2 \rangle$. The error ε is distributed as a $\mathcal{N}(0, \sigma^2)$, where σ^2 was chosen such that, under H_0, $R^2 = \frac{\mathbb{V}\mathrm{ar}[\langle \mathbf{X}, \boldsymbol{\rho} \rangle]}{\mathbb{V}\mathrm{ar}[\langle \mathbf{X}, \boldsymbol{\rho} \rangle] + \sigma^2} = 0.95$. The selection of d_n is done automatically by SICc throughout the section. The random directions are drawn from the data-driven Gaussian process described in Section 4 (see the Supplementary Material [Cuesta-Albertos et al. (2019)] for other data generating processes and their effects). The choice of the δ_d's is described in detail in the Supplementary Material.

We explore first the dependence of the tests with respect to the number of projections K. Figure 2 shows the empirical level for each scenario, based on $M = 10{,}000$ Monte Carlo trials and $B = 10{,}000$ bootstrap replicates. There is a clear L-shaped pattern in the empirical rejection rate curves, which is produced

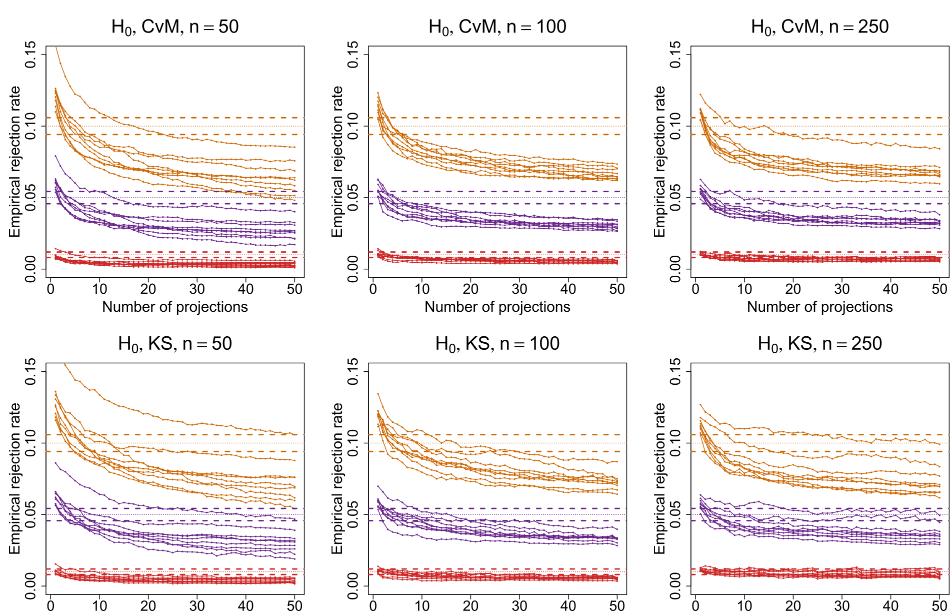

FIG. 2. *Empirical sizes of the CvM (upper row) and KS (lower row) tests for scenario Sk, $k = 1, \dots, 9$, depending on the number of projections $K = 1, \dots, 50$, and for sample sizes $n = 50, 100, 250$ (from left to right). The empirical sizes associated with the significance levels $\alpha = 0.01, 0.05, 0.10$ are coded in red, purple, and orange, respectively. Dashed thick lines represent the asymptotic 95% confidence interval for the proportion α obtained from M replicates.*

by the conservativeness of the FDR correction—under H_0, it ensures that the rejection rate is *at most* α—when dealing with the highly-dependent projected tests. For small K's ($K \approx 3$), both tests calibrate the three levels for different sample sizes reasonably well, with the main exception being $n = 50$ and $\alpha = 0.10$, for which the tests have a significant over-rejection of the null hypothesis. For moderate to large K's, the empirical rejection rates decrease and stabilize below α, resulting in a systematic violation of the confidence intervals. Figure 3 shows that the empirical powers with respect to K are almost constant or exhibit mild decrements, except for certain bumps at lower values of K that provide a significant power gain. Both facts point towards choosing the number of projections K to be relatively small, $K \in \{1, 2, 3, 4, 5\}$ and particularly $K = 3$, in order to make a reasonable compromise between correct calibration and power. In addition to the computational expediency that a small K yields, it also avoids requiring a large B to estimate properly the FDR p-values, provided that the FDR correction requires a finer precision in the discretization of the p-values for larger K (see the Supplementary Material [Cuesta-Albertos et al. (2019)]).

The tests based on the KS and CvM norms are compared with the test presented in García-Portugués, González-Manteiga and Febrero-Bande (2014) (denoted by PCvM), available in the R package `fda.usc` [Febrero-Bande and Oviedo de la Fuente (2017)], and whose test statistic can be regarded as the average of projected CvM statistics. The test was run with the same FPC estimation used in the new tests, the same number of components d_n, and $B = 10{,}000$ (considered also for CvM and KS). Table 1 presents the empirical rejection rates of the different simulation scenarios with $K = 1, 3, 5$ for KS and CvM tests. The results show two consistent patterns. First, in our simulation scenarios, the CvM test consistently dominates over the KS test, with only one exception: $H_{9,1}$ with $n = 50$ (see the Supplementary Material [Cuesta-Albertos et al. (2019)] for the latter). This is coherent with the fact that quadratic norms in goodness-of-fit tests are often more powerful than sup-norms [see, e.g., page 110 of D'Agostino and Stephens (1986) for the distribution case]. Second, PCvM tends to have a larger power than CvM for most of the situations, especially for small sample sizes and mild deviations. As an illustration, for $n = 50$, the average relative loss in the empirical power for CvM_3 with respect to PCvM is 12.7% ($d = 1$) and 4.6% ($d = 2$). For $n = 100$, the losses drop to 9.3% and 1.3%, respectively, and for $n = 250$, to 5.2% and 0.2%, respectively. The drop in performance for CvM with respect to PCvM is expected due to the construction of CvM, which opts for exploring a set of random directions instead of averaging uniformly distributed finite-dimensional directions, as PCvM does. This also yields one the strongest points of the CvM test, which is its relatively short running times, especially for large n. Not surprisingly, the number of evaluations performed for computing the CvM statistic is $\mathcal{O}(n)$, a notable reduction from PCvM's $\mathcal{O}((n^3 - n^2)/2)$. Also, the memory requirement for CvM is $\mathcal{O}(n)$, instead of PCvM's $\mathcal{O}((n^2 - n - 2)/2)$. The running times in Figure 4 is evidence of this improvement.

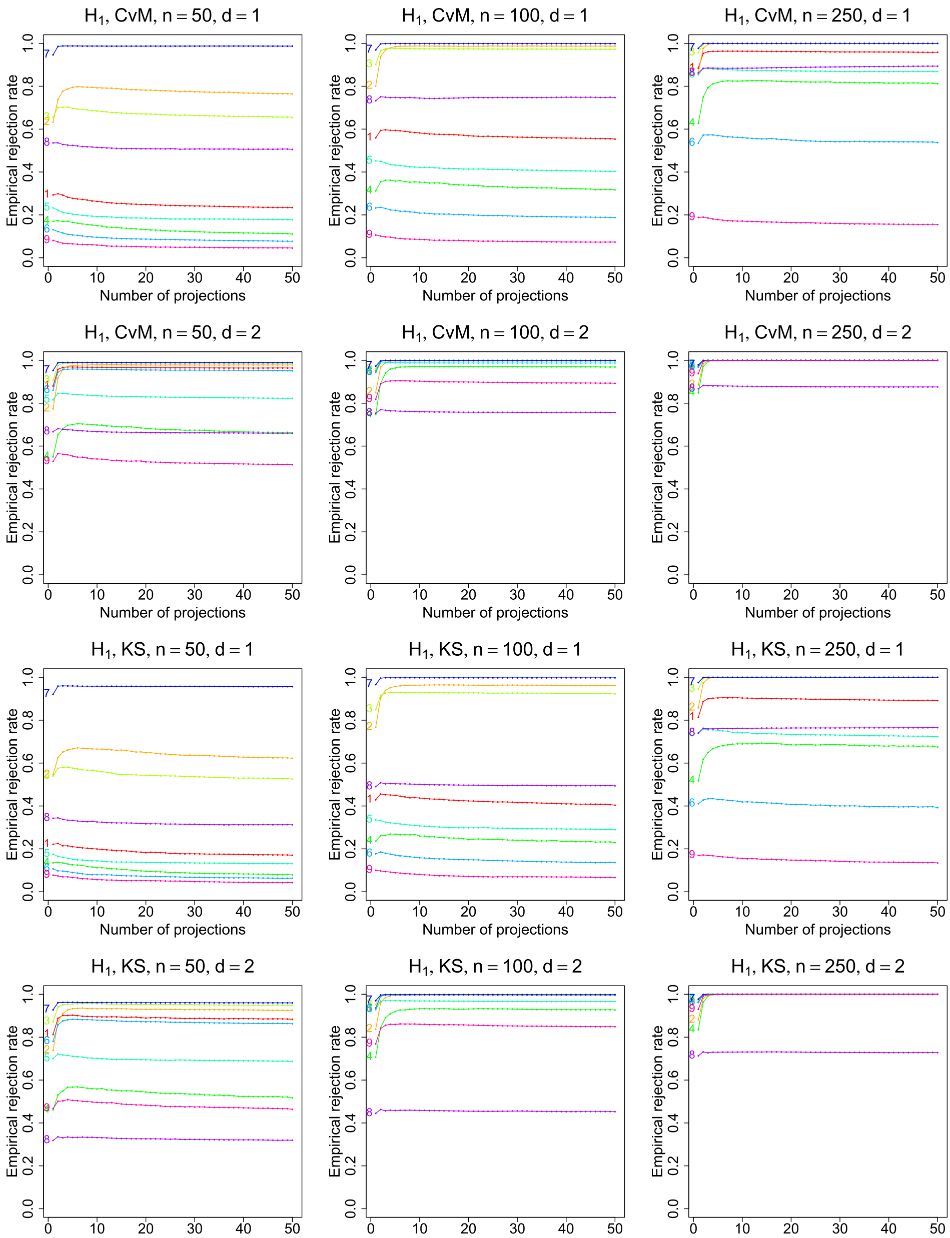

FIG. 3. *Empirical powers of the CvM (first two rows) and KS (last two) tests for scenario Sk*, $k = 1, \dots, 9$, *depending on the number of projections* $K = 1, \dots, 50$. *Odd rows correspond to the deviation index* $d = 1$, *while even account for* $d = 2$. *The significance level is* $\alpha = 0.05$ *and the sample sizes are* $n = 50, 100, 250$ (*columns, from left to right*).

TABLE 1

Empirical sizes and powers (in percentages) of the CvM, KS, and PCvM tests with $\alpha = 0.05$, sample sizes $n = 100, 250$, and estimation of $\boldsymbol{\rho}$ by data-driven FPC (d_n chosen by SICc). KS and CvM tests are shown with 1, 3, and 5 projections

	$n = 100$							$n = 250$						
$H_{k,\delta}$	CvM_1	CvM_3	CvM_5	KS_1	KS_3	KS_5	PCvM	CvM_1	CvM_3	CvM_5	KS_1	KS_3	KS_5	PCvM
$H_{1,0}$	5.1	3.9	3.5	5.5	4.6	4.2	4.8	5.4	3.9	3.6	5.6	4.2	3.9	4.9
$H_{2,0}$	5.4	4.6	4.2	5.6	5.1	4.6	3.6	5.8	4.8	4.3	6.1	5.4	4.9	4.7
$H_{3,0}$	6.2	4.9	4.5	7.0	6.0	5.2	5.7	5.6	4.1	3.8	5.8	4.7	4.2	5.3
$H_{4,0}$	5.9	4.4	4.1	5.9	5.0	4.8	4.6	6.3	5.2	4.8	6.4	5.9	5.6	4.9
$H_{5,0}$	5.5	4.0	3.6	6.0	4.3	4.0	4.9	4.9	4.2	3.8	5.0	4.0	3.5	5.0
$H_{6,0}$	5.4	4.3	3.9	6.0	4.9	4.5	5.2	5.6	4.3	4.0	6.0	5.0	4.8	4.8
$H_{7,0}$	5.5	3.9	3.7	6.0	4.7	4.0	5.1	5.4	4.1	3.8	5.5	4.7	4.1	5.2
$H_{8,0}$	5.1	3.5	3.3	5.3	3.7	3.4	4.9	5.3	3.9	3.6	5.4	4.3	3.9	5.1
$H_{9,0}$	6.3	4.8	4.3	6.1	4.9	4.5	6.1	5.6	4.4	4.1	5.7	4.8	4.1	5.9
$H_{1,1}$	56.0	59.4	58.3	42.9	45.0	43.7	69.9	88.4	96.3	96.3	81.4	90.3	90.3	98.4
$H_{2,1}$	80.1	98.5	98.7	76.7	95.7	96.3	99.2	86.5	100	100	85.8	100	100	100
$H_{3,1}$	90.2	97.6	97.4	85.0	93.0	92.8	99.2	95.6	100	100	94.5	100	100	100
$H_{4,1}$	31.2	35.7	35.3	23.6	26.8	26.0	43.6	62.7	81.8	82.5	51.8	67.7	68.9	88.6
$H_{5,1}$	45.2	43.1	42.1	33.5	31.8	30.6	49.9	85.4	87.9	87.4	73.8	75.3	74.1	91.5
$H_{6,1}$	23.3	22.2	20.8	17.7	17.0	15.7	27.9	53.5	57.0	56.1	41.1	43.0	41.9	66.9
$H_{7,1}$	96.9	99.9	99.9	96.6	99.8	99.8	99.9	97.7	100	100	97.5	100	100	100
$H_{8,1}$	73.3	74.8	74.5	49.0	50.3	50.1	74.7	86.2	88.3	88.4	74.1	76.0	76.2	87.7
$H_{9,1}$	10.6	9.2	8.6	10.0	8.9	8.1	12.1	18.8	17.9	17.2	17.0	16.4	15.5	22.3

TABLE 1
(*Continued*)

$H_{k,\delta}$	n = 100							n = 250						
	CvM$_1$	**CvM$_3$**	**CvM$_5$**	**KS$_1$**	**KS$_3$**	**KS$_5$**	**PCvM**	**CvM$_1$**	**CvM$_3$**	**CvM$_5$**	**KS$_1$**	**KS$_3$**	**KS$_5$**	**PCvM**
$H_{1,2}$	94.9	100	100	93.6	99.9	99.9	100	97.0	100	100	96.3	100	100	100
$H_{2,2}$	85.0	99.8	99.9	83.7	99.5	99.6	99.9	88.6	100	100	88.0	100	100	100
$H_{3,2}$	95.9	100	100	95.2	100	100	100	97.9	100	100	97.2	100	100	100
$H_{4,2}$	74.8	96.4	97.2	70.6	92.0	93.3	98.2	84.8	99.9	100	83.5	99.9	100	100
$H_{5,2}$	94.7	98.9	98.8	92.9	97.0	96.8	99.1	97.6	100	100	97.4	100	100	100
$H_{6,2}$	94.4	100	100	93.2	99.8	99.8	100	96.9	100	100	96.1	100	100	100
$H_{7,2}$	97.3	99.9	99.9	97.0	99.9	99.8	99.9	97.9	100	100	97.8	100	100	100
$H_{8,2}$	75.3	76.5	76.0	44.5	45.8	45.9	78.2	86.7	88.0	87.9	71.3	73.0	73.1	88.9
$H_{9,2}$	81.8	90.5	90.3	76.9	85.9	86.0	93.9	93.8	100	100	93.0	100	100	100

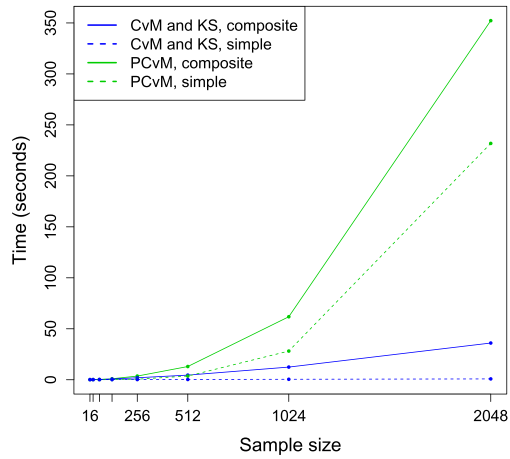

FIG. 4. *Running times for the CvM and KS tests (the computation of both tests is done in the same routine) and PCvM, for the composite and simple hypotheses. The tests were averaged over* $M = 100$ *trials and calibrated with* $B = 1000$. *The sample sizes are* $n = 2^k$, $k = 4, \dots, 11$, *and the number of projections considered is* $K = 3$. *Times were measured on a* 2.53 *GHz core. All the tests have a similar implementation in* `R` *that interfaces* `FORTRAN` *for the computation of the statistics.*

The new tests were also applied to the two data applications described in García-Portugués, González-Manteiga and Febrero-Bande (2014), yielding similar conclusions. Both datasets are provided in the library `fda.usc`. The first example uses the classical Tecator dataset, considered in Section 2.1.1 of Ferraty and Vieu (2006) as a motivating example for introducing nonlinear regression models. The dataset contains 215 spectrometric curves measuring the absorbance at wavelengths [850, 1050] of finely chopped meat samples. Covariates giving the fat, water and protein content of the meat are also available in the dataset. Typically, the goal is to predict the fat content of a meat sample using the spectrometric curve or any of its derivatives, and, for that, the FLM has been proposed as a candidate model. We test its adequacy for the dataset with the new goodness-of-fit tests proposed. The p-values obtained for $K = 3$ projections and $B = 10{,}000$ are 0.020 and 0.022 for CvM and KS, respectively. Using $\hat{\boldsymbol{\rho}}$ with d_n selected by SICc in PCvM gave a p-value of 0.006. Employing the first or second derivatives of the absorbance curves provided null p-values. In addition, the tests for $H_0 : \boldsymbol{\rho} = \mathbf{0}$ also had null p-values for all of tests. As a consequence, we conclude that, at level $\alpha = 0.05$, there is evidence against the FLM and there is a significant nonlinear relation between the fat content and the absorbance curves.

The second example mimics the classical dataset in Ramsay and Silverman (2005) on Canadian weather stations. The data are contains yearly profiles of temperature from 73 weather stations of the AEMET (Spanish Meteorological Agency; Spanish acronym) network and other meteorological variables, and the goal is to explain the mean of the wind speed at each location. Prior to its analysis, the dataset was preprocessed to remove the 5% most outlying curves using the Fraiman and Muniz (2001) depth. With the same settings as before, the CvM and KS tests for $K = 3$ projections provided p-values equal to 0.612 and 0.396, respectively, and PCvM gave a p-value $= 0.080$. In addition, the tests for $H_0 : \boldsymbol{\rho} = \mathbf{0}$ yielded null p-values. We conclude that, at level $\alpha = 0.05$, there is no evidence against the FLM and that the effect of the covariate on the response is significant and linear.

6. Discussion. We have presented a new way of building goodness-of-fit tests for regression models with functional covariates employing random projections. The methodology was illustrated using randomly projected empirical processes, which provided root-n consistent tests for testing functional linearity. The calibration of the tests was done by a wild bootstrap resampling and the FDR was used to combine K p-values coming from different projections to account for a higher power. The empirical analysis of the tests, conducted in a fully data-driven way, showed that, in our simulation scenarios, CvM yields higher powers than KS and that a selection of $K \in \{1, \dots, 5\}$, in particular $K = 3$, is a reasonable compromise between respecting size and increasing power. There is still a price to pay in terms of a moderate loss of power with respect to the PCvM test, which averages across a set of uniformly distributed finite-dimensional directions. However, the reduction in computational complexity of the new tests is more than notable.

We conclude the paper by sketching some promising extensions of the methodology for the testing of more complex models involving functional covariates:

(a) Testing the significance of the functional covariate of $(\mathbf{X}, \mathbf{W}) \in \mathcal{H} \times \mathbb{R}^q$ in the functional partially linear model [Aneiros-Pérez and Vieu (2006)] $Y = m(\mathbf{X}) + \mathbf{W}'\boldsymbol{\beta} + \varepsilon$. The process to be considered for a sample $\{(\mathbf{X}_i, \mathbf{W}_i, Y_i)\}_{i=1}^n$ and an estimator $\hat{\boldsymbol{\beta}}$ such that $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \mathcal{O}_{\mathbb{P}}(n^{-1/2})$ is $n^{-1/2} \sum_{i=1}^n \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}} \le x\}}(Y_i - \mathbf{W}_i'\hat{\boldsymbol{\beta}})$.

(b) Testing a functional quadratic regression model [Horváth and Reeder (2013)].

(c) Testing the significance of a functional linear model with functional response: $H_0 : \mathbb{E}[\mathbf{Y}|\mathbf{X}] = \mathbf{0}$, where now $(\mathbf{X}, \mathbf{Y}) \in \mathcal{H}_1 \times \mathcal{H}_2$ and the associated empirical process is $n^{-1/2} \sum_{i=1}^n \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}_1} \le x\}} \mathbf{Y}_i^{\mathbf{h}_2}$.

Software availability. The `R` package `rp.flm.test`, openly available at https://github.com/egarpor/rp.flm.test, contains the implementation of the tests and allows reproduction of the simulation study and data applications. The main function, `rp.flm.test`, has also been included in the `R` package `fda.usc` since version 1.3.1.

APPENDIX: PROOFS OF THE MAIN RESULTS

A.1. Hypothesis projection.

PROOF OF PROPOSITION 2.1. We denote by $\mathbf{X}_p$ both the vectors $(X_1, \dots, X_p)'$ and $(X_1, \dots, X_p, 0, \dots)'$ containing the first p coefficients of $\mathbf{X}$ in an orthonormal basis of $\mathcal{H}$. We prove first the result for the finite subspace of $\mathcal{H}$ spanned by the first p elements of the orthonormal basis. We need to show that

$$\begin{aligned} &\mathbb{E}[Y|\mathbf{X}_p] = 0 \qquad \text{a.s.} \\ &\qquad \iff \quad \mathbb{E}\big[Y|\langle \mathbf{X}_p, \mathbf{h}\rangle\big] = 0 \qquad \text{a.s. for every } \mathbf{h} \in \mathcal{H}. \end{aligned} \tag{8}$$

To prove this, we make use of Theorem 1 in Bierens (1982), which states that if $\mathbf{V}$ and $\mathbf{Z}$ are two $\mathbb{R}^p$-valued random vectors, then

$$\mathbb{E}[\mathbf{V}|\mathbf{Z}] = 0 \qquad \text{a.s.} \qquad \iff \qquad \mathbb{E}\big[\mathbf{V}e^{i\langle \mathbf{t}, \mathbf{Z}\rangle}\big] = 0 \qquad \text{for every } \mathbf{t} \in \mathbb{R}^p. \tag{9}$$

Assume that $\mathbb{E}[Y|\mathbf{X}_p] = 0$ and let $\mathbf{h} \in \mathcal{H}$. Since the σ-algebra generated by $\langle \mathbf{X}_p, \mathbf{h}\rangle$, $\sigma(\langle \mathbf{X}_p, \mathbf{h}\rangle)$, is contained in $\sigma(\mathbf{X}_p)$, we have that $\mathbb{E}[Y|\langle \mathbf{X}_p, \mathbf{h}\rangle] = \mathbb{E}[\mathbb{E}[Y|\mathbf{X}_p]|\langle \mathbf{X}_p, \mathbf{h}\rangle] = 0$ a.s., which shows the *if* part. To obtain the *only if* part, let $\mathbf{h} \in \mathcal{H}$, and compute $\mathbb{E}[Ye^{it\langle \mathbf{X}_p, \mathbf{h}\rangle}] = \mathbb{E}[\mathbb{E}[Y|\langle \mathbf{X}_p, \mathbf{h}\rangle]e^{it\langle \mathbf{X}_p, \mathbf{h}\rangle}] = 0$, for every $t \in \mathbb{R}$. Then (8) follows from (9).

Now we are in position to prove the result for $\mathcal{H}$. As before, the *if* implication follows from $\sigma(\langle \mathbf{X}, \mathbf{h}\rangle) \subset \sigma(\mathbf{X})$. To prove the *only if* implication, given $p \in \mathbb{N}$ and $\mathbf{h} \in \mathcal{H}$, since $\mathbf{h}_p \in \mathcal{H}$ and $\langle \mathbf{X}, \mathbf{h}_p\rangle = \langle \mathbf{X}_p, \mathbf{h}\rangle$, then $\sigma(\langle \mathbf{X}_p, \mathbf{h}\rangle) \subset \sigma(\langle \mathbf{X}, \mathbf{h}\rangle)$, and we have that the assumption implies that $\mathbb{E}[Y|\langle \mathbf{X}_p, \mathbf{h}\rangle] = 0$ a.s. Thus, from (8), we have that $\mathbb{E}[Y|\mathbf{X}_p] = 0$ a.s. for every p, and the result follows from the fact that $\sigma(\mathbf{X}_p) \uparrow \sigma(\mathbf{X})$ because of the integrability assumption on Y. □

PROOF OF LEMMA 2.3. From the properties of the conditional expectation, the Cauchy–Schwarz and Jensen inequalities, we have that

$$l_k = \mathbb{E}\big[\|\mathbf{X}\|^k \mathbb{E}\big[|Y||\mathbf{X}\big]\big] \le (m_{2k})^{1/2}\big(\mathbb{E}\big[Y^2\big]\big)^{1/2}.$$

Thus l_k is finite. By the convexity of the function $t \mapsto t^{(2k+1)/2k}$ and Jensen's inequality, $m_{2k}^{1/2k} \le m_{2k+1}^{1/(2k+1)}$. Hence, $\sum_{k=1}^{\infty} m_{2k}^{-1/2k} = \infty$. □

PROOF OF THEOREM 2.4. The *only if* part is trivial because $\sigma(\mathbf{X}^{\mathbf{h}}) \subset \sigma(\mathbf{X})$, and then $\mathbb{E}[Y|\mathbf{X}] = 0$ a.s. implies that $\mu(\mathcal{H}_0) = 1$. Concerning the *if* part, let us assume that $\mu(\mathcal{H}_0) > 0$. From the assumptions, we have that $\mathbb{E}[|Y||\mathbf{X}] < \infty$, and, if we take $\mathbf{h} \in \mathcal{H}_0$, then

$$\mathbb{E}[Y] = \mathbb{E}\big[\mathbb{E}\big[Y|\mathbf{X}^{\mathbf{h}}\big]\big] = 0. \tag{10}$$

Let us assume that $\mathbb{E}[Y|\mathbf{X}]$ is not zero a.s. Then the random variables

$$\Phi^+(\mathbf{X}) := \big(\mathbb{E}[Y|\mathbf{X}]\big)^+ = \max\big\{\mathbb{E}[Y|\mathbf{X}], 0\big\},$$

$$\Phi^-(\mathbf{X}) := \big(\mathbb{E}[Y|\mathbf{X}]\big)^- = \max\big\{-\mathbb{E}[Y|\mathbf{X}], 0\big\},$$

are integrable and positive with positive probability. Thus, (10) implies that

$$V := \int \Phi^+(\mathbf{x})\,\mathrm{d}P_{\mathbf{X}}(\mathbf{x}) = \int \Phi^-(\mathbf{x})\,\mathrm{d}P_{\mathbf{X}}(\mathbf{x}) > 0.$$

Consider now the probability measures ν_Φ^+ and ν_Φ^-, which are defined on $\mathcal{H}$ and whose Radon–Nikodym derivatives with respect to $P_{\mathbf{X}}$ are, respectively,

$$\frac{\mathrm{d}\nu_\Phi^+}{\mathrm{d}P_{\mathbf{X}}}(\mathbf{x}) := V^{-1}\Phi^+(\mathbf{x}) \quad \text{and} \quad \frac{\mathrm{d}\nu_\Phi^-}{\mathrm{d}P_{\mathbf{X}}}(\mathbf{x}) := V^{-1}\Phi^-(\mathbf{x}).$$

For $k \in \mathbb{N}$, the moments of ν_{Φ^+} verify that (analogously for Φ^-)

$$\int \|\mathbf{x}\|^k\,\mathrm{d}\nu_{\Phi^+}(\mathbf{x}) \le V^{-1}\int \|\mathbf{x}\|^k \mathbb{E}\big[|Y||\mathbf{X}=\mathbf{x}\big]\,\mathrm{d}P_{\mathbf{X}}(\mathbf{x}) = l_k,$$

and then, due to Lemma 2.3, they satisfy (a) in Lemma 2.2. Given $\mathbf{h} \in \mathcal{H}_0$, the r.v. $\mathbf{X}^{\mathbf{h}}$ is $\mathbf{X}$-measurable. Thus, a.s.

$$0 = \mathbb{E}\big[Y|\mathbf{X}^{\mathbf{h}}\big] = \mathbb{E}\big[\mathbb{E}[Y|\mathbf{X}]|\mathbf{X}^{\mathbf{h}}\big] = \mathbb{E}\big[\mathbb{E}[Y|\mathbf{X}]^+|\mathbf{X}^{\mathbf{h}}\big] - \mathbb{E}\big[\mathbb{E}[Y|\mathbf{X}]^-|\mathbf{X}^{\mathbf{h}}\big].$$

From here, it is easy to prove that the marginal distributions of ν_Φ^+ and ν_Φ^- on the one-dimensional subspace generated by $\mathbf{X}^{\mathbf{h}}$ coincide if $\mathbf{h} \in \mathcal{H}_0$. Since $\mathcal{H}_0$ has a positive μ-measure, from Lemma 2.2, we obtain that these probability measures indeed coincide and, as a consequence, $V^{-1}(\mathbb{E}[Y|\mathbf{X}])^+ = V^{-1}(\mathbb{E}[Y|\mathbf{X}])^-$ a.s., which trivially implies that $\mathbb{E}[Y|\mathbf{X}] = 0$ a.s. □

A.2. Testing the linear model.

PROOF OF THEOREM 3.1. We analyse the asymptotic distribution of the three terms separately by invoking some auxiliary lemmas. Their proofs are collected in the Supplementary Material [Cuesta-Albertos et al. (2019)].

The asymptotic distribution of $T_{n,\mathbf{h}}^1(x)$ follows from Corollary 2.6: $n^{-1/2} \times T_{n,\mathbf{h}}^1(x) \overset{\mathcal{L}}{\rightsquigarrow} \mathcal{N}(0, K_1(x,x))$. So, if $a_n = o(n^{-1/2})$, then $a_n T_{n,\mathbf{h}}^1(x) = o_{\mathbb{P}}(1)$. The following two lemmas give insights into the asymptotic behaviour of k_n and are required for the analysis of $T_{n,\mathbf{h}}^2$ and $T_{n,\mathbf{h}}^3$.

LEMMA A.1. *Under A6 and A9*, $k_n^3(\log k_n)^2 = o(n^{1/2})$.

LEMMA A.2. *Under A6 and A9, we have that* $\nu[d_n = k_n] \to 1$.

We employ the decomposition (11) from page 338 in CMS to arrive at

$$\hat{\boldsymbol{\rho}} - \boldsymbol{\rho} = \mathbf{L}_n + \mathbf{Y}_n + \mathbf{S}_n + \mathbf{R}_n, \tag{11}$$

where $\mathbf{L}_n := -\sum_{j=k_n+1}^{\infty}\langle\boldsymbol{\rho}, \mathbf{e}_j\rangle\mathbf{e}_j$, $\mathbf{Y}_n := \sum_{j=1}^{k_n}(\langle\boldsymbol{\rho}, \hat{\mathbf{e}}_j\rangle\hat{\mathbf{e}}_j - \langle\boldsymbol{\rho}, \mathbf{e}_j\rangle\mathbf{e}_j)$, $\mathbf{S}_n := (\Gamma_n^\dagger - \Gamma^\dagger)\mathbf{U}_n$, $\mathbf{R}_n := \Gamma^\dagger\mathbf{U}_n$, and $\mathbf{U}_n := \frac{1}{n}\sum_{i=1}^n \mathbf{X}_i \otimes \varepsilon_i$. The decomposition (11) in CMS contains an extra term, $\mathbf{T}_n$, which is null here because of our construction of $\Gamma_n^\dagger$.

We will profusely employ the notation

$$\overline{\mathbf{X}}_{x,\mathbf{h}} := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}} \le x\}} \mathbf{X}_i.$$

From (11), the term $T_{n,\mathbf{h}}^2(x)$ can be expressed as

$$T_{n,\mathbf{h}}^2(x) = n\langle \overline{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}}, \mathbf{L}_n + \mathbf{S}_n + \mathbf{Y}_n + \mathbf{R}_n\rangle.$$

As a consequence of the following lemmas, we have that $T_{n,\mathbf{h}}^2(x) = o_{\mathbb{P}}(n^{1/2})$.

LEMMA A.3. *Under* A3 *and* A4, $n^{1/2}\langle \overline{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}}, \mathbf{L}_n\rangle = o_{\mathbb{P}}(1)$.

LEMMA A.4. *Under* A6, A8, *and* A9, $n^{1/2}\langle \overline{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}}, \mathbf{R}_n\rangle = o_{\mathbb{P}}(1)$.

LEMMA A.5. *Under* A5, A6, A8, *and* A9, $n^{1/2}\langle \overline{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}}, \mathbf{S}_n\rangle = o_{\mathbb{P}}(1)$.

LEMMA A.6. *Under* A4, A6, A8, *and* A9, $n^{1/2}\langle \overline{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}}, \mathbf{Y}_n\rangle = o_{\mathbb{P}}(1)$.

The behaviour of the third term, yielding statement (a), is given by the next lemma.

LEMMA A.7. *Under* A3, A4, A6, A7, *and* A9, $n^{-1/2} t_{n,\mathbf{E}_{x,\mathbf{h}}}^{-1} T_{n,\mathbf{h}}^3(x) \overset{\mathcal{L}}{\rightsquigarrow} \mathcal{N}(0, \sigma_\varepsilon^2)$.

From the above results, $a_n T_{n,\mathbf{h}}^2(x) = o_{\mathbb{P}}(1)$ for cases (b) and (c), $T_{n,\mathbf{h}}^3$ is the dominant term in (b), and both $T_{n,\mathbf{h}}^1$ and $T_{n,\mathbf{h}}^3$ are dominant in (c). □

PROOF OF PROPOSITION 3.2. By the definition of $t_{n,\mathbf{E}_{x,\mathbf{h}}}$ and (5),

$$t_{n,\mathbf{E}_{x,\mathbf{h}}}^2 = \sum_{j=1}^{k_n} \frac{\mathbb{E}[\mathbb{1}_{\{\mathbf{X}^{\mathbf{h}} \le x\}}\langle \mathbf{X}, \mathbf{e}_j\rangle]^2}{\lambda_j} = \sum_{j=1}^{k_n} \mathbb{E}[\mathbb{1}_{\{\mathbf{X}^{\mathbf{h}} \le x\}}\xi_j]^2 \le \sum_{j=1}^{k_n} \mathbb{E}[\xi_j^2] = k_n.$$

We assume now that $\mathbf{X}$ is Gaussian. Obviously, the two-dimensional random vector $(\xi_j, \mathbf{X}^{\mathbf{h}})$, $j \in \mathbb{N}$, is centred normal. Moreover, the variance of ξ_j is one and, if $h_j = \langle \mathbf{h}, \mathbf{e}_j\rangle$, then $\sigma_{\mathbf{h}}^2 = \sum_{j=1}^\infty h_j^2 \lambda_j < \infty$ (due to $\sum_{j=1}^\infty h_j^2 < \infty$ and A3) and $\mathbb{C}\mathrm{ov}[\xi_j, \mathbf{X}^{\mathbf{h}}] = h_j \lambda_j^{1/2}$. Notice that, if $\mathbf{h} \neq \mathbf{0}$, $\sigma_{\mathbf{h}} > 0$ since $\lambda_j > 0$ for all $j \in \mathbb{N}$. Denoting by $\phi_{\mathbf{h}}(u, v)$ the joint density function of $(\xi_j, \mathbf{X}^{\mathbf{h}})$ and by $\phi_{\mathbf{h},2}(v)$ its second marginal, we have that

$$\begin{aligned}\mathbb{E}[\mathbb{1}_{\{\mathbf{X}^{\mathbf{h}}\le x\}}\xi_j] &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} u\mathbb{1}_{\{v\le x\}}\frac{\phi_{\mathbf{h}}(u,v)}{\phi_{\mathbf{h},2}(v)}\,\mathrm{d}u\right)\phi_{\mathbf{h},2}(v)\,\mathrm{d}v\\ &= \int_{-\infty}^{x}\mathbb{E}[\xi_j|\mathbf{X}^{\mathbf{h}}=v]\phi_{\mathbf{h},2}(v)\,\mathrm{d}v\\ &= \int_{-\infty}^{x}\frac{h_j\sqrt{\lambda_j}v}{\sigma_{\mathbf{h}}^2}\phi_{\mathbf{h},2}(v)\,\mathrm{d}v = -\frac{h_j\sqrt{\lambda_j}}{\sigma_{\mathbf{h}}}\phi(x/\sigma_{\mathbf{h}}).\end{aligned}$$

This, the initial development, and A3 give us that

$$t_{n,\mathbf{E}_{x,\mathbf{h}}}^2 = \frac{\phi^2(x/\sigma_{\mathbf{h}})}{\sigma_{\mathbf{h}}^2}\sum_{j=1}^{k_n} h_j^2\lambda_j \to \phi^2(x/\sigma_{\mathbf{h}}).$$

□

PROOF OF THEOREM 3.3. We first prove (a). The joint asymptotic normality of $(T_{n,\mathbf{h}}(x_1),\dots,T_{n,\mathbf{h}}(x_k))$ for $(x_1,\dots,x_k)\in\mathbb{R}^k$ follows by the Cramér–Wold device and the same arguments used in Lemma A.7. Also, in the proof of that lemma it is shown that $n^{1/2}\langle\mathbf{E}_{x,\mathbf{h}},\mathbf{L}_n+\mathbf{Y}_n+\mathbf{S}_n\rangle = o_{\mathbb{P}}(1)$. Then, due to (11) and (5),

$$\begin{aligned}T_{n,\mathbf{h}}(x) &= n^{-1/2}\big(T_{n,\mathbf{h}}^1(x)+T_{n,\mathbf{h}}^3(x)\big)\\ &= n^{-1/2}\sum_{i=1}^n \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}}\le x\}}\varepsilon_i + n^{-1/2}\sum_{i=1}^n\langle\mathbf{E}_{x,\mathbf{h}},\Gamma^{\dagger}\mathbf{X}_i\rangle\varepsilon_i + o_{\mathbb{P}}(1)\\ &= n^{-1/2}\sum_{i=1}^n\{A_x^i+B_x^i\}+o_{\mathbb{P}}(1),\end{aligned}$$

with $A_x^i := \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}}\le x\}}\varepsilon_i$, and $B_x^i := \sum_{j=1}^{k_n}\langle\mathbf{E}_{x,\mathbf{h}},\mathbf{e}_j\rangle\lambda_j^{-1/2}\xi_j^i\varepsilon_i$. Since the $\mathbf{X}_i$'s and ε_i's are i.i.d., and $\mathbb{E}[\varepsilon|\mathbf{X}]=0$,

$$\begin{aligned}&\mathbb{C}\mathrm{ov}\left[n^{-1/2}\sum_{i=1}^n\{A_s^i+B_s^i\}, n^{-1/2}\sum_{i'=1}^n\{A_t^{i'}+B_t^{i'}\}\right]\\ &\qquad = \mathbb{E}[A_s^1A_t^1]+\mathbb{E}[A_s^1B_t^1]+\mathbb{E}[B_s^1A_t^1]+\mathbb{E}[B_s^1B_t^1].\end{aligned}$$

Applying the tower property with the conditioning variables $\mathbf{X}^{\mathbf{h}}$ (first expectation) and $\mathbf{X}$ (second and third), it follows that

$$\begin{aligned}\mathbb{E}[A_s^1A_t^1] &= K_1(s,t),\\ \mathbb{E}[A_s^1B_t^1] &= \int_{\{\mathbf{x}^{\mathbf{h}}\le s\}}\mathbb{V}\mathrm{ar}[Y|\mathbf{X}=\mathbf{x}]\langle\mathbf{E}_{t,\mathbf{h}},\Gamma^{\dagger}\mathbf{x}\rangle\,\mathrm{d}P_{\mathbf{X}}(\mathbf{x}),\\ \mathbb{E}[B_s^1B_t^1] &= \int\mathbb{V}\mathrm{ar}[Y|\mathbf{X}=\mathbf{x}]\langle\mathbf{E}_{s,\mathbf{h}},\Gamma^{\dagger}\mathbf{x}\rangle\langle\mathbf{E}_{t,\mathbf{h}},\Gamma^{\dagger}\mathbf{x}\rangle\,\mathrm{d}P_{\mathbf{X}}(\mathbf{x}).\end{aligned}$$

Since $\Gamma^\dagger \to \Gamma^{-1}$ in the operator norm $\|\cdot\|_\infty$, Cauchy–Schwarz and $||(\Gamma^\dagger - \Gamma^{-1})\mathbf{x}|| \le \|\Gamma^\dagger - \Gamma^{-1}\|_\infty \|\mathbf{x}\|$ give that $\mathbb{E}[A_s^1 B_t^1] - C_1(s,t)$ and $\mathbb{E}[B_s^1 B_t^1] - C_2(s,t)$ converge to zero. The result then follows from Slutsky's theorem.

We now prove (b). The tightness of $n^{-1/2} T^1_{n,\mathbf{h}}$ is obtained using the same arguments as in Theorem 1.1 of Stute (1997). For the tightness of $n^{-1/2} T^3_{n,\mathbf{h}}$, define

$$\bar{T}^3_{n,\mathbf{h}}(u) := n\langle \mathbb{E}[\mathbb{1}_{\{U_\mathbf{h} \le u\}} \mathbf{X}], \boldsymbol{\rho} - \hat{\boldsymbol{\rho}}\rangle, \qquad U_\mathbf{h} := F_\mathbf{h}(\mathbf{X}^\mathbf{h}),$$

as the time-changed version of $\bar{T}^3_{n,\mathbf{h}}$ by $F_\mathbf{h}$, that is,

$$T^3_{n,\mathbf{h}}(x) = \bar{T}^3_{n,\mathbf{h}}(F_\mathbf{h}(x)).$$

Consider $0 \le u_1 < u < u_2 \le 1$ and the differences

$$n^{-1/2}(\bar{T}^3_{n,\mathbf{h}}(u) - \bar{T}^3_{n,\mathbf{h}}(u_1)) = n^{1/2}\langle \mathbb{E}[\mathbb{1}_{\{u_1 < U_\mathbf{h} \le u\}} \mathbf{X}], \boldsymbol{\rho} - \hat{\boldsymbol{\rho}}\rangle,$$
$$n^{-1/2}(\bar{T}^3_{n,\mathbf{h}}(u_2) - \bar{T}^3_{n,\mathbf{h}}(u)) = n^{1/2}\langle \mathbb{E}[\mathbb{1}_{\{u < U_\mathbf{h} \le u_2\}} \mathbf{X}], \boldsymbol{\rho} - \hat{\boldsymbol{\rho}}\rangle.$$

Then, by the Cauchy-Schwarz and Jensen inequalities,

$$\begin{aligned}
&\mathbb{E}[n^{-2}|\bar{T}^3_{n,\mathbf{h}}(u) - \bar{T}^3_{n,\mathbf{h}}(u_1)|^2 |\bar{T}^3_{n,\mathbf{h}}(u_2) - \bar{T}^3_{n,\mathbf{h}}(u)|^2] \\
&\quad \le n^2 \mathbb{E}[\|\mathbb{E}[\mathbb{1}_{\{u_1 < U_\mathbf{h} \le u\}} \mathbf{X}]\|^2 \|\mathbb{E}[\mathbb{1}_{\{u_1 < U_\mathbf{h} \le u\}} \mathbf{X}]\|^4 \|\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}\|^4] \\
&\quad = n^2 \mathbb{E}[\|\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}\|^4] \int \mathbb{E}[\mathbf{X}(t) \mathbb{1}_{\{u_1 < U_\mathbf{h} \le u\}}]^2 \,\mathrm{d}t \int \mathbb{E}[\mathbf{X}(t) \mathbb{1}_{\{u < U_\mathbf{h} \le u_2\}}]^2 \,\mathrm{d}t \\
&\quad \le n^2 \mathbb{E}[\|\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}\|^4] \int \mathbb{E}[\mathbf{X}^2(t) \mathbb{1}_{\{u_1 < U_\mathbf{h} \le u\}}] \,\mathrm{d}t \int \mathbb{E}[\mathbf{X}^2(t) \mathbb{1}_{\{u < U_\mathbf{h} \le u_2\}}] \,\mathrm{d}t \\
&\quad = n^2 \mathbb{E}[\|\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}\|^4][F(u) - F(u_1)][F(u_2) - F(u)] \\
&\quad \le n^2 \mathbb{E}[\|\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}\|^4][F(u_2) - F(u_1)]^2 \\
&\quad \le [G(u_2) - G(u_1)]^2,
\end{aligned}$$

where $F(u) := \int \mathbb{E}[\mathbf{X}^2(t) \mathbb{1}_{\{U_\mathbf{h} \le u\}}] \,\mathrm{d}t$ and $G(u) := \sup_n \{n^2 \mathbb{E}[\|\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}\|^4]\} F(u)$ are nondecreasing and continuous functions on [0, 1]. This corresponds to employing $\gamma = 2$ and $\alpha = 1$ in Theorem 15.6 of Billingsley (1968), which gives the weak convergence of $n^{-1/2} \bar{T}^3_{n,\mathbf{h}}$ in $D([0,1])$ and, as a consequence of the Continuous Mapping Theorem (CMT), $n^{-1/2} T^3_{n,\mathbf{h}} \overset{\mathcal{L}}{\rightsquigarrow} \mathcal{G}_2$ in $D(\mathbb{R})$.

Finally, we prove that $n^{-1/2} T^2_{n,\mathbf{h}} \overset{p}{\to} \mathbf{0}$. Note first that, by Cauchy–Schwarz,

$$\sup_{x \in \mathbb{R}} |n^{-1/2} T^2_{n,\mathbf{h}}(x)| \le \sup_{x \in \mathbb{R}} \|\overline{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}}\| n^{1/2} \|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|.$$

Assumption $\mathbb{E}[\|\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}\|^4] = \mathcal{O}(n^{-2})$ implies $\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\| = \mathcal{O}_\mathbb{P}(n^{-1/2})$. In addition, the weak law of large numbers in $\mathcal{H}$ [e.g., Hoffmann-Jørgensen and

Pisier (1976)] and A3 give $\overline{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}} \xrightarrow{p} \mathbf{0}$ in $\mathcal{H}$. Therefore, the CMT yields $\sup_{x\in\mathbb{R}} |n^{-1/2}T^2_{n,\mathbf{h}}(x)| \xrightarrow{p} 0$ and, as a consequence, $n^{-1/2}T^2_{n,\mathbf{h}} \xrightarrow{p} \mathbf{0}$ in $D(\mathbb{R})$. □

PROOF OF COROLLARY 3.4. $\|T_{n,\mathbf{h}}\|_{\mathrm{KS}} \overset{\mathcal{L}}{\rightsquigarrow} \|\mathcal{G}_2\|_{\mathrm{KS}}$ follows from the CMT. For the Cramér–von Mises norm, we use

$$\|T_{n,\mathbf{h}}\|_{\mathrm{CvM}} = \int_{\mathbb{R}} T_{n,\mathbf{h}}(x)^2 \mathrm{d}F_{\mathbf{h}}(x) + \int_{\mathbb{R}} T_{n,\mathbf{h}}(x)^2\, \mathrm{d}(F_{n,\mathbf{h}} - F_{\mathbf{h}})(x) \tag{12}$$

and $F_{n,\mathbf{h}} - F_{\mathbf{h}} \xrightarrow{p} \mathbf{0}$. By Slutsky's theorem, $(T_{n,\mathbf{h}}, F_{n,\mathbf{h}} - F_{\mathbf{h}}) \overset{\mathcal{L}}{\rightsquigarrow} (\mathcal{G}_2, \mathbf{0})$. Then, by the CMT, $\int_{\mathbb{R}} T_{n,\mathbf{h}}(x)^2 \mathrm{d}(F_{n,\mathbf{h}} - F_{\mathbf{h}})(x) \overset{\mathcal{L}}{\rightsquigarrow} \mathbf{0}$ and $\int_{\mathbb{R}} T_{n,\mathbf{h}}(x)^2\, \mathrm{d}F_{\mathbf{h}}(x) \overset{\mathcal{L}}{\rightsquigarrow} \int_{\mathbb{R}} \mathcal{G}_2(x)^2\, \mathrm{d}F_{\mathbf{h}}(x)$, which completes the proof. □

Acknowledgments. We thank one anonymous referee and an Associate Editor for detailed and useful comments that led to significant improvements on an earlier draft of the paper. We gratefully acknowledge the computational resources of the Supercomputing Center of Galicia (CESGA).

SUPPLEMENTARY MATERIAL

Supplement to "Goodness-of-fit tests for the functional linear model based on randomly projected empirical processes" (DOI: 10.1214/18-AOS1693SUPP; .pdf). Two extra Appendices contain the proofs of the technical lemmas and further results for the simulation study.

REFERENCES

ANEIROS-PÉREZ, G. and VIEU, P. (2006). Semi-functional partial linear regression. *Statist. Probab. Lett.* **76** 1102–1110. MR2269280

BENJAMINI, Y. and YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29** 1165–1188. MR1869245

BICKEL, P. J. and ROSENBLATT, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1** 1071–1095. MR0348906

BIERENS, H. J. (1982). Consistent model specification tests. *J. Econometrics* **20** 105–134. MR0685673

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York. MR0233396

BÜCHER, A., DETTE, H. and WIECZOREK, G. (2011). Testing model assumptions in functional regression models. *J. Multivariate Anal.* **102** 1472–1488. MR2819963

CARDOT, H., MAS, A. and SARDA, P. (2007). CLT in functional linear regression models. *Probab. Theory Related Fields* **138** 325–361. MR2299711

CARDOT, H., FERRATY, F., MAS, A. and SARDA, P. (2003). Testing hypotheses in the functional linear model. *Scand. J. Stat.* **30** 241–255. MR1965105

CHIOU, J.-M. and MÜLLER, H.-G. (2007). Diagnostics for functional regression via residual processes. *Comput. Statist. Data Anal.* **51** 4849–4863. MR2364544

CUESTA-ALBERTOS, J. A. and FEBRERO-BANDE, M. (2010). A simple multiway ANOVA for functional data. *TEST* **19** 537–557. MR2746001

CUESTA-ALBERTOS, J. A., FRAIMAN, R. and RANSFORD, T. (2007). A sharp form of the Cramér–Wold theorem. *J. Theoret. Probab.* **20** 201–209. MR2324526

CUESTA-ALBERTOS, J. A., DEL BARRIO, E., FRAIMAN, R. and MATRÁN, C. (2007). The random projection method in goodness of fit for functional data. *Comput. Statist. Data Anal.* **51** 4814–4831. MR2364542

CUESTA-ALBERTOS, J. A., GARCÍA-PORTUGUÉS, E., FEBRERO-BANDE, M. and GONZÁLEZ-MANTEIGA, W. (2019). Supplement to "Goodness-of-fit tests for the functional linear model based on randomly projected empirical processes." DOI:10.1214/18-AOS1693SUPP.

D'AGOSTINO, R. B. and STEPHENS, M. A., eds. (1986) Goodness-of-Fit Techniques. *Statistics: Textbooks and Monographs* **68**. Dekker, Inc., New York. MR0874534

DELSOL, L., FERRATY, F. and VIEU, P. (2011a). Structural test in regression on functional variables. *J. Multivariate Anal.* **102** 422–447. MR2755007

DELSOL, L., FERRATY, F. and VIEU, P. (2011b). Structural tests in regression on functional variable. In *Recent Advances in Functional Data Analysis and Related Topics* 77–83. Physica-Verlag/Springer, Heidelberg. MR2815564

DURBIN, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.* **1** 279–290. MR0359131

ESCANCIANO, J. C. (2006). A consistent diagnostic test for regression models using projections. *Econometric Theory* **22** 1030–1051. MR2328527

FEBRERO-BANDE, M. and OVIEDO DE LA FUENTE, M. (2017). *fda.usc: Functional Data Analysis and Utilities for Statistical Computing (fda.usc)*. R package version 1.3.1. Available at http://cran.r-project.org/web/packages/fda.usc/.

FERRATY, F. and VIEU, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer, New York. MR2229687

FRAIMAN, R. and MUNIZ, G. (2001). Trimmed means for functional data. *TEST* **10** 419–440. MR1881149

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. Statist.* **23** 761–778. MR3224655

GONZÁLEZ-MANTEIGA, W. and CRUJEIRAS, R. M. (2013). An updated review of goodness-of-fit tests for regression models. *TEST* **22** 361–411. MR3093195

HÄRDLE, W. and MAMMEN, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21** 1926–1947. MR1245774

HILGERT, N., MAS, A. and VERZELEN, N. (2013). Minimax adaptive tests for the functional linear model. *Ann. Statist.* **41** 838–869. MR3099123

HOFFMANN-JØRGENSEN, J. and PISIER, G. (1976). The law of large numbers and the central limit theorem in Banach spaces. *Ann. Probab.* **4** 587–599. MR0423451

HORVÁTH, L. and KOKOSZKA, P. (2012). *Inference for Functional Data with Applications*. Springer, New York. MR2920735

HORVÁTH, L. and REEDER, R. (2013). A test of significance in functional quadratic regression. *Bernoulli* **19** 2120–2151. MR3129046

KOKOSZKA, P., MASLOVA, I., SOJKA, J. and ZHU, L. (2008). Testing for lack of dependence in the functional linear model. *Canad. J. Statist.* **36** 207–222. MR2431682

LAVERGNE, P. and PATILEA, V. (2008). Breaking the curse of dimensionality in nonparametric testing. *J. Econometrics* **143** 103–122. MR2384435

MCQUARRIE, A. D. (1999). A small-sample correction for the Schwarz SIC model selection criterion. *Statist. Probab. Lett.* **44** 79–86. MR1706323

NADARAJA, È. A. (1964). On a regression estimate. *Teor. Veroyatn. Primen.* **9** 157–159. MR0166874

PATILEA, V., SÁNCHEZ-SELLERO, C. and SAUMARD, M. (2012). Projection-based nonparametric testing for functional covariate effect. Preprint. Available at arXiv:1205.5578.

PATILEA, V., SÁNCHEZ-SELLERO, C. and SAUMARD, M. (2016). Testing the predictor effect on a functional response. *J. Amer. Statist. Assoc.* **111** 1684–1695. MR3601727

RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. Springer, New York. MR2168993

STUTE, W. (1997). Nonparametric model checks for regression. *Ann. Statist.* **25** 613–641. MR1439316

STUTE, W., GONZÁLEZ MANTEIGA, W. and PRESEDO QUINDIMIL, M. (1998). Bootstrap approximations in model checks for regression. *J. Amer. Statist. Assoc.* **93** 141–149. MR1614600

WATSON, G. S. (1964). Smooth regression analysis. *Sankhya*, *Ser. A* **26** 359–372. MR0185765

ZHENG, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *J. Econometrics* **75** 263–289. MR1413644

J. A. CUESTA-ALBERTOS
DEPARTMENT OF MATHEMATICS, STATISTICS AND COMPUTER SCIENCE
UNIVERSITY OF CANTABRIA
39005, SANTANDER
SPAIN
E-MAIL: juan.cuesta@unican.es

E. GARCÍA-PORTUGUÉS
DEPARTMENT OF STATISTICS AND
UC3M-BS INSTITUTE OF FINANCIAL BIG DATA
CARLOS III UNIVERSITY OF MADRID
28911, LEGANÉS
SPAIN
E-MAIL: edgarcia@est-econ.uc3m.es

M. FEBRERO-BANDE
W. GONZÁLEZ-MANTEIGA
DEPARTMENT OF STATISTICS, MATHEMATICAL ANALYSIS AND OPTIMIZATION
UNIVERSITY OF SANTIAGO DE COMPOSTELA
15782, SANTIAGO DE COMPOSTELA
SPAIN
E-MAIL: manuel.febrero@usc.es
wenceslao.gonzalez@usc.es