A link function specification test in the single functional index model

Lax Chan[1] · **Laurent Delsol**[2] · **Aldo Goia**[1]

Abstract

In this paper a test for specification in functional regression with scalar response that exploits semi-parametric principles is illustrated. Once the test statistics is defined, its asymptotic null distribution is derived under suitable conditions. The finite sample performances of the test are analyzed through a simulation study by using both the asymptotic p-value and some bootstrap approaches. To appreciate the potentialities of the method, an application to a spectrometric real dataset is performed.

Keywords Functional data · Single Index model · Goodness-of-fit test · Asymptotic normality

Mathematics Subject Classification Primary 62R10 · Secondary 62G08 · 62G10

1 Introduction

The problem of specifying the link function that models the dependence structure between two random elements is a very important task in many statistical regression analysis. That problem could be hard to treat when a functional predictor X, that is a random element taking values in a functional space, is used to explain the variability of a real random variable (r.v.) Y. In this case, in fact, the link is described through

✉ Aldo Goia
aldo.goia@uniupo.it

Lax Chan
lax.chan@uniupo.it

Laurent Delsol
laurent.delsol@univ-orleans.fr

1   Dipartimento di Studi per l'Economia e l'Impresa, Università del Piemonte Orientale, Via Perrone 18, 28100 Novara, Italy

2   Institut Denis Poisson, Université d'Orléans, Rue de Chartres, 45067 Orleans, France

a real valued operator r acting on a functional space and it is difficult to visualize it, and consequently, to select a coherent specification. This kind of model is known as functional regression model with scalar response:

$$Y = r[X] + \mathcal{E} \tag{1}$$

where $\mathcal{E}$ is a centered random error uncorrelated with X. It has appeared in many scientific domains and falls within the methodologies of the so-called functional statistics. For a review on this relatively recent branch of statistics, see for instance the monographes (Ferraty and Vieu 2006; Horvath and Kokoszka 2012) or (Ramsay and Silverman 2005), or the papers that appeared in recent special issues (see, for instance, Aneiros et al. 2022, 2019a, b).

The interest toward the checking of a specification of the regression operator for the model (1) has produced a rich line of research on structural testing procedures. Only to cite some examples, in Aneiros and Vieu (2013) the authors deal with testing linearity in semi-functional partially linear regression models, in Bücher et al. (2011) a test for the hypothesis of a specific parametric functional regression model is introduced, in Cuesta-Albertos et al. (2019) a goodness-of-fit tests for the functional linear model is illustrated, the paper (Delsol et al. 2011) tackles an omnibus goodness-of-fit tests in a full nonparametric framework, finally the work (Patilea et al. 2016) provides a test for the case of a functional response in the spirit of the smoothing test statistic considered by Sheng (1996).

A useful help in the context of the model specification can come from the semiparametric regression approaches. Thanks to a projective strategy, it allows to visualize a link function, combining the flexibility and the interpretability and, at the same time, to avoid some dimensionality problems that can occur in the full nonparametric context. Only to get a partial idea of the variety of techniques developed, one can see a general presentation in Härdle et al. (2004) and, as examples for what concerns the functional statistics, the papers Ferraty et al. (2013), Goia and Vieu (2015), Ling and Vieu (2021) and Novo et al. (2019) and references therein.

The present paper explores the possibility to build a specification test by exploiting the potential of the Single Functional Index Model (SFIM). This defines the relationship between X and Y through an unknown real link function g acting on a projection of the functional regressor along an unknown direction θ, constrained for identifiability. Formally, let X be a random element valued in a Hilbert space $\mathcal{H}$ of real functions defined over a compact interval $\mathcal{T}$, equipped with an inner product $\langle \cdot, \cdot \rangle$ and associated norm $\|\cdot\|$, then $r[X] = g(\langle \theta, X \rangle)$, where $g : \mathbb{R} \to \mathbb{R}$, $\theta \in \mathcal{H}$ with $\|\theta\|^2 = 1$ and $\theta(t) > 0$ for any fixed $t \in \mathcal{T}$ for identifiability (see e.g. Ait-Saïdi et al. 2008). So far, various techniques have been introduced to estimate g and θ from samples drawn from (X, Y) and the rate of convergence derived (see e.g. Jiang et al. 2020; Novo et al. 2019 or Shang 2020 for recent contributions).

The main interest in SFIM is that it makes possible to bring back an infinite dimensional problem to a one dimensional framework and to visualize an estimate $\widehat{g}$ of g, obtained from an observed dataset, that can suggest the nature of the link between X and Y. This allows to postulate a target specification g_0 for g, depending on some real parameter, and then to check if it is compatible with the observed dataset at a given

significance level. For instance, if the plot of $\widehat{g}$ exhibits a straightness shape, the linearity of the regression model should be investigated. As a consequence, once the link function is specified, the resulting model depends only on the functional parameter θ and hence it is full parametric, with some practical and theoretical benefits in the estimation step (for instance, a faster rate of convergence than in the semi-parmetric case, a full interpretability of the link, good estimates for rather small sample sizes, no smoothing parameters have to be introduced). By way of example, one can consider the prediction of the moisture value for 80 samples of corn by using the first derivatives of the corresponding Near-infrared (NIR) spectra measured over the wavelength range 1100–2498 nanometer (see the original spectra in the top panel of Fig. 1): an estimate of the link function g is plotted in the bottom panel of Fig. 1 and one may wonder if the hypothesis of a linear link is compatible with the empirical evidence.

The aim of this work is to define and operationalize a suitable specification test procedure and then analyze its performances: given the SFIM framework, one wants to test the null hypothesis that the link function g belongs to a family of possible parametric functions:

$$\mathcal{G}_0 = \left\{ g_0^\beta : \mathbb{R} \to \mathbb{R}, \beta = (\beta_0, \beta_1, \ldots, \beta_p) \in \mathbb{R}^{d+1} \right\} \quad d \geq 1 \text{ integer}$$

where the parameter β can be entirely specified or not, against the alternative that g is not an element of $\mathcal{G}_0$. The main idea is to exploit the so-called conditional moment test approach (see e.g. Newey 1985) based on the fact that, under the null hypothesis the quantity $\mathbb{E}[\mathcal{E}\mathbb{E}[\mathcal{E}|X]w(X)]$, where $w(X)$ is a suitable weight, is null, whereas it is strictly positive under the alternative one. Therefore, by using a kernel regression approach, a test statistic belonging to the family of U-statistic is derived and appropriately standardized, and, under suitable assumptions on the distribution of X (in particular, expressed in terms of small-ball probability), the involved kernel, the nature of the error, the behaviour of the estimator of the SFIM model used, it is proved that its asymptotic null distribution is the Gaussian one. Thanks to the latter statement, the p value of the test can be directly computed and an evaluation of the power of the test, based on the asymptotic result, can be evaluated by means of Monte Carlo experiments carried out on samples with finite sizes and under various experimental conditions, that is nature of the link, sample size, variability of the error model. In order to appreciate the robustness of the introduced test methodology, a comparison with the results obtained when some standard bootstrap procedures are used is made. Finally, to show how the test can be used for practical purposes, an application to the prediction via SFIM model for the spectrometric dataset is performed and discussed.

The outline of the paper is as follows: the theoretical background, the basic principle of the test and the test statistic are defined in Sect. 2. Theoretical properties of the test statistic are discussed in Sect. 3: in particular in Sects. 3.1 and 3.1.1 the null distribution of the test statistic is derived, whereas some remarks on consistency are provided in Sect. 3.1.2. The performances of the test are analyzed in Sect. 4 and the bootstrap approaches are described in Sect. 5. Finally the real world analysis is done in Sect. 6. For the sake of readability, all the detailed proofs of the theoretical results are postponed to the technical Appendix.
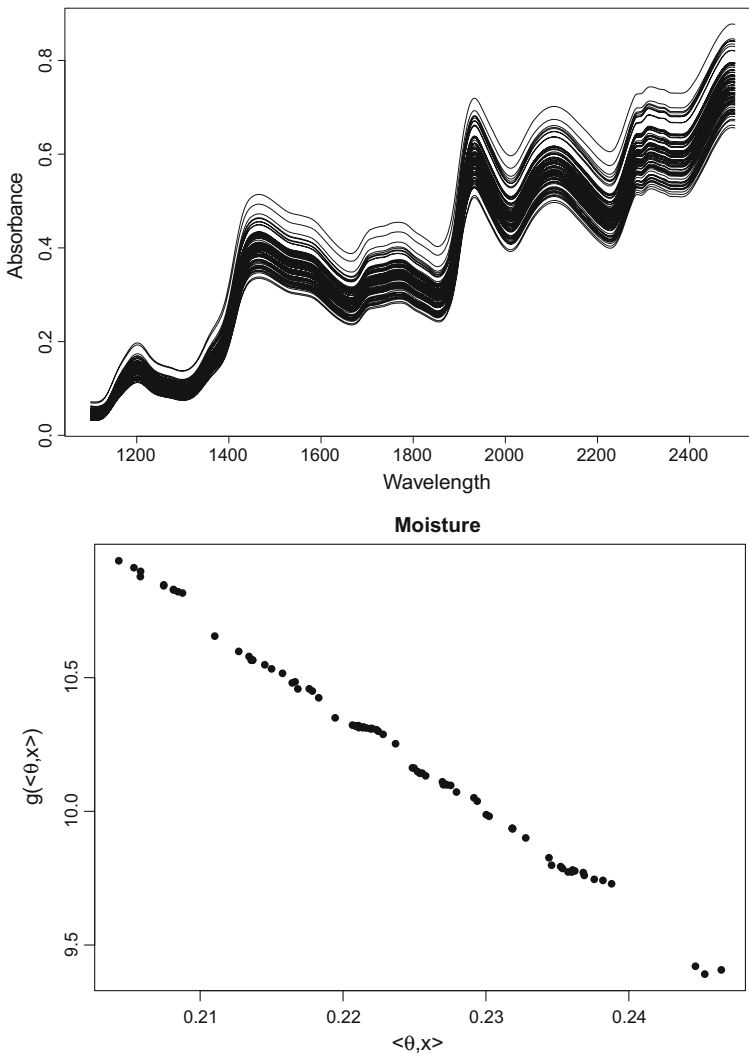
Fig. 1 Spectrometric curves (top) and estimate of the link function (bottom)

2 Notations and test definition

Consider the random element (X, Y) defined on a probability space and mapping on $\mathcal{H} \times \mathbb{R}$, where $\mathcal{H}$ is a Hilbert space of real functions defined over a compact interval $\mathcal{T}$. From now on, one takes $\mathcal{H} = \mathcal{L}^2_{[0,1]}$ the space of square integrable real functions defined over [0, 1], equipped with the natural inner product $\langle g, h \rangle = \int_0^1 g(s)h(s)ds$ and associated norm $\|g\|^2 = \langle g, g \rangle$.

Assume that the relation between Y and X is defined by the following SFIM:

$$Y = g(\langle X, \theta \rangle) + \mathcal{E} \tag{2}$$

where $g : \mathbb{R} \to \mathbb{R}$ is an unknown link function, $\theta \in \mathcal{H}$ is an unknown direction such that $\|\theta\|^2 = 1$ and $\theta(t) > 0$ for a fixed t, for identifiability, and $\mathcal{E}$ is a r.v. satisfying $\mathbb{E}[\mathcal{E}|X] = 0$ and $\mathbb{E}[\mathcal{E}^2|X] = \sigma^2$ (see e.g. Ait-Saïdi et al. 2008).

2.1 The test principle

Define $\mathcal{G}_0 = \{g_0^\beta : \mathbb{R} \to \mathbb{R}, \beta \in \mathbb{R}^{d+1}\}$, $d \geq 1$ integer, be a known function, measurable w.r.t. the σ-algebra generated by X, and depending on the parameter $\beta = (\beta_0, \beta_1, \ldots, \beta_d) \in \mathbb{R}^{d+1}$. Consider then the following hypothesis:

$$H_0 : g \in \mathcal{G}_0 \quad \text{vs.} \quad H_1 : g \in \mathcal{G}_1$$

where $\mathcal{G}_1$ is a set of real functions g_1^β such that $\mathcal{G}_1 \cap \mathcal{G}_0 = \emptyset$. If θ and β are fixed, the hypotheses are simple, otherwise complex. To fix the ideas, the above setting includes the possibility of testing the linearity of the regression by specifying $\mathcal{G}_0$ as the set of affine functions $g_0^\beta(u) = \beta_0 + \beta_1 u$, with $\beta_0, \beta_1 \in \mathbb{R}$, $\beta_1 \neq 0$. In practice, the specification of the model under the null hypothesis could be done by a direct inspection of the scatterplot between the observed values $\langle X, \theta \rangle$ and those of Y or by imposing an a priori model.

Consider:

$$\mathbb{E}\left[\left(g(\langle X, \theta \rangle) - g_0^\beta(\langle X, \theta \rangle) \right)^2 w(X) \right] \tag{3}$$

where $g_0 \in \mathcal{G}_0$, and w is a positive weight function. By the SFIM assumption (2), $\mathbb{E}[Y|X] = g(\langle X, \theta \rangle)$ and this implies that

$$g(\langle X, \theta \rangle) - g_0^\beta(\langle X, \theta \rangle) = \mathbb{E}[\mathcal{E}|X]$$

where $\mathcal{E} = Y - g_0^\beta(\langle X, \theta \rangle)$. Hence, (3) can be rewritten in the equivalent form:

$$\mathbb{E}\left[\mathbb{E}[\mathcal{E}|X]^2 w(X) \right] = \mathbb{E}[\mathcal{E}\mathbb{E}[\mathcal{E}|X]w(X)].$$

Under the null hypothesis the latter is null, because of $g(\langle X, \theta \rangle) = g_0(\langle X, \theta \rangle)$ a.s., whereas under the alternative it is strictly positive, because of $\mathbb{P}(g_0(\langle X, \theta \rangle) = g(\langle X, \theta \rangle)) < 1$.

Thanks to the above principle, it is possible to implement a test procedure starting from an empirical version of $\mathbb{E}[\mathcal{E}\mathbb{E}[\mathcal{E}|X]w(X)]$: the null hypothesis is rejected if it is significantly far from zero.

2.2 The test statistic

Consider a sample (X_i, Y_i), $i = 1, \ldots, n$, of i.i.d. replications of (X, Y) and suppose that θ and β are completely specified and equal to $\theta_\star$ and $\beta_\star$ respectively. Take a Nadaraya–Watson type nonparametric kernel estimate of $\mathbb{E}[\mathcal{E}|X]$ at the point X_i. Subsequently, an empirical version of $\mathbb{E}[\mathcal{E}\mathbb{E}[\mathcal{E}|X]w(X)]$ can be written as follows

$$\frac{1}{n} \sum_{i=1}^{n} \mathcal{E}_i \sum_{j=1, j\neq i}^{n} \mathcal{E}_j \frac{K(\delta(X_i, X_j)/h)}{\sum_{j\neq i} K(\delta(X_i, X_j)/h)} w(X_i) \tag{4}$$

where $\mathcal{E}_i = Y_i - g_0^{\beta_\star}(\langle X_i, \theta_\star \rangle)$, K is a kernel function, h is a bandwidth, which depends on n, and δ is a semi-metric.

In order to derive a convenient expression for the test statistic in the SFIM framework, some simplifications can be introduced. Firstly, as w can be chosen arbitrarily, if one can assume that the projection r.v. $\langle X, \theta_\star \rangle$ admits a strictly positive probability density function $f_{\theta_\star}$, then a possible choice is $w = f_{\theta_\star}$. Since it is unknown, consider the following cross-validated kernel estimate of $f_{\theta_\star}$ based on the same kernel, semi-metric and bandwidth as above:

$$f_{\theta_\star, n}(X_i) = \frac{1}{(n-1)h} \sum_{j=1, j\neq i}^{n} K\left(\frac{|\langle X_i - X_j, \theta_\star \rangle|}{h}\right).$$

Secondly, one can select the the projection semi-metric:

$$\delta(X_1, X_2) = |\langle X_1 - X_2, \theta_\star \rangle|.$$

By plugging these choices in (4), the following simplified expression of the test statistics follows:

$$Q_n(\theta_\star) = \frac{1}{n(n-1)h} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \mathcal{E}_i \mathcal{E}_j K_{ij}^{\theta_\star} \tag{5}$$

where $K_{ij}^{\theta_\star} = K(|\langle X_i - X_j, \theta_\star \rangle|/h)$.

Invoking similar arguments as in Sheng (1996) one can derive the following estimate for the variance of $n\sqrt{h}Q_n(\theta_\star)$:

$$v_n^2(\theta_\star) = \frac{2}{n(n-1)h} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \mathcal{E}_i^2 \mathcal{E}_j^2 \left(K_{ij}^{\theta_\star}\right)^2. \tag{6}$$

This allows to obtain the standardized test statistic version $n\sqrt{h}Q_n(\theta_\star)/v_n(\theta_\star)$ that can be used to test a simple null hypothesis.

When one deals with composite null hypothesis, θ and/or β are not specified and some estimates of them have to be introduced. In particular, let $\widehat{\theta}$ be an estimator of

θ, one defines $\widehat{\mathcal{E}}_i = Y_i - g_0^{\beta_\star}(\langle X_i, \widehat{\theta} \rangle)$ and the resulting test statistic can be written as follows:

$$Q_n(\widehat{\theta}) = \frac{1}{n(n-1)h} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \widehat{\mathcal{E}}_i \widehat{\mathcal{E}}_j K_{ij}^{\widehat{\theta}} \qquad (7)$$

where $K_{ij}^{\widehat{\theta}} = K\left(|\langle \widehat{\theta}, X_i - X_j \rangle|/h\right)$. To obtain a suitable estimator $\widehat{\theta}$ for θ it is possible to use, for instance, the first step of the approach proposed in Ferraty et al. (2013) which combines a spline approximation of the functional coefficient θ and the one-dimensional Nadaraya–Watson approach to estimate the link function in the SFIM (2).

In the case of the parameter β is also not specified and an estimate $\widehat{\theta}$ is available, one has to consider some estimators $\widehat{\beta}$ of β that can be obtained through a least square approach, by minimizing $\sum_{i=1}^{n} \{Y_i - g_0^{\beta}(\langle X_i, \widehat{\theta} \rangle)\}^2$. In particular, if g_0 is linear affine, one regresses directly the observations Y_is against the projections $\langle X_i, \widehat{\theta} \rangle$s. The studentized versions of these test statistics are obtained by plugging in (6) the estimates $\widehat{\theta}$ and $\widehat{\beta}$.

3 Asymptotic behaviour of the test statistics

To define the critical region of the test, one needs the derivation of the null distribution of the test statistics. To do this, one considers two scenarios, namely the simple null hypothesis, where both parameters β and θ are completely specified, and the complex one: they are discuss in Sects. 3.1 and 3.2 respectively. The study of the behaviour under some alternative hypothesis concludes this section (see Sect. 3.3).

3.1 Simple null hypothesis

Suppose that $g(\langle \theta, X \rangle) = g_0^{\beta_\star}(\langle \theta, X \rangle)$ and that the direction θ is specified and equals to $\theta_\star$. Consider the following sets of assumptions; in what follows $\mathcal{F}[\cdot]$ denotes the Fourier transform.

Assumptions on the sample

S-i. $(X_i, Y_i), i = 1, \ldots, n,$ of i.i.d. replications of (X, Y)

Assumptions on the model

M-i.   There exists the generating moments function of the error $\mathcal{E}$;

M-ii.   There exists $\underline{\sigma}^2$ and $\overline{\sigma}^2$ such that $0 < \underline{\sigma}^2 \leq Var(\mathcal{E}|X) \leq \overline{\sigma}^2 < \infty$ almost surely;

M-iii.1. There exist a constant $C_1 > 0$ such that

$$\frac{1}{C_1} \leq \mathbb{E}\left[f_{\theta_\star}(\langle \theta_\star, X \rangle) \right] \leq C_1.$$

M-iii.2. There exists $C_2 > 0, \epsilon > 0$ such that $\int_{|x| \leq \epsilon} |\mathcal{F}[f_{\theta_\star}]|^2(x)dx \geq C_2$.

M-iv. The function $g_0^{\beta_\star}$ is Lipschitz.

Assumptions on the Kernel function

K-i. The kernel K is a continuous density of bounded variation and with strictly positive Fourier transform on the real line.

K-ii. As n diverges, $h \to 0$ and $\dfrac{\ln n}{(nh^2)^\lambda} \to 0$, for some $\lambda \in (0, 1)$.

The main results are collected in the following theorem and its detailed proof is postponed in the Appendix for the sake of readability. The role of the assumptions in proving these results is discussed at the end of this section.

Theorem 1 Under the assumptions S, M and K and when the null hypothesis H_0 holds true, one has:

(i) $|Q_n(\theta_\star)| = O_{\mathbb{P}}\left(\dfrac{\ln n}{nh^{1/2}}\right)$,

(ii) $\dfrac{1}{v_n^2(\theta_\star)} = O_{\mathbb{P}}(1)$,

(iii) as n goes to infinity, $n\sqrt{h}\, Q_n(\theta_\star)/v_n(\theta_\star) \sim \mathcal{N}(0, 1)$.

According to statement (iii) in Theorem 1 if the null hypothesis H_0 holds true, the test statistics $T_n = n\sqrt{h}\, Q_n(\theta_\star)/v_n(\theta_\star)$ converges in law to a standard normal. Consequently, the test given by $\mathbb{I}_{\{T_n \geq z_{1-a}\}}$, with z_α the $(1 - \alpha)$-th quantile of the standard normal distribution, has an asymptotic level α.

To conclude the section, some remarks about the assumptions stated above are added. Firstly, one can point out that M-i., M-iv. and K-i. are standard hypothesis in the framework of nonparametric functional regression models (see for instance Ferraty and Vieu 2006). In particular, the bounded variation condition K-i. is satisfied by many well-known kernels, such as Gaussian or Epachnikov ones, and thanks to this condition, the class of kernel functions subsequently considered is Euclidean, a notion crucial for achieving the proofs and that is defined in the Appendix. In fact, the latter notion together with the regularity condition M.iv, imposed on the link function g_0, allow to utilise an existing exponential inequality due to Major (2006). To do this the conditional expectation of the squared errors and the squared kernel need to be investigated. The boundedness condition of the conditional variance (see M-ii.) allows one to bound the conditional errors below and above. The behaviour of the squared kernel is studied through the properties of its Fourier transform. The technical assumptions such as the positive Fourier transform on the real line K-i., the boundedness below and above of the density of the pdf $f_{\theta_\star}$ (see M-iii.1), the boundedness below of the restricted Fourier transform of the pdf $f_{\theta_\star}$ (see M-iii.2) allow one to establish the upper and lower bounds of the kernel and this verifies a condition required for the Major's exponential inequality. Note that the condition M-iii.2 is satisfied by many random variables, such as Gaussian or exponential. The existence of moment generating function of the error $\mathcal{E}$ (see M-i.), and the trade-off between the sample size n and the bandwidth h (see K-ii.) permit to establish the rate of convergence in statements (i) and (ii) of Theorem 1. Finally, the collection of hypotheses further allows one to operate in a similar setting as in Lavergne and Patilea (2008) to establish the asymptotic normality in statement (iii).

3.2 Complex null hypothesis

Suppose that β is fixed and θ is estimated by $\widehat{\theta}$. Then one has to investigate the behaviour of $Q_n(\widehat{\theta})$: to achieve the convergence results, the following set of extra assumptions is needed.

Extra assumptions on the estimate

E-i. Take $\widehat{\theta}$ belonging to a set Θ_n of possible directions of interest which is such that

$$\#\Theta_n = n^p, \, p > 0 \text{ and } \|\theta - \theta_\star\| \leq C\left(\frac{\ln n}{n}\right)^r \text{ for any } \theta \in \Theta_n$$

where $\#$ is the cardinality of a set, $p = p_n$ depends on n, and $r > 0$ depends on some regularity assumption on the regression operator in (1) (see e.g. Ferraty et al. 2013 or Novo et al. 2019).

E-ii. The bandwidth h satisfies: $n^{1-2r}h^{1/2} \to 0$.

Extra assumptions on the sample

S-ii. $\|X_i\|$ is bounded.

Extra assumptions on the model

M-iii.1.bis There exist a constant $C_1 > 0$ such that for any $\theta \in \Theta_n$

$$\frac{1}{C_1} \leq \mathbb{E}[f_\theta(\langle\theta, X\rangle)] \leq C_1.$$

M-iii.2.bis There exists $C_2 > 0, \epsilon > 0$ such that for any $\theta \in \Theta_n$, $\int_{|x|\leq\epsilon} |\mathcal{F}[f_\theta]|^2(x)dx \geq C_2$.

Extra assumptions on the Kernel function

K-iii. $p \geq 1$ increases to infinity with n and $p^{3/2}(\ln n)^{-\lambda}$ is bounded, for some constant $\lambda > 0$.

Return to the expression of $Q_n(\widehat{\theta})$ and define the following decomposition

$$Q_n(\widehat{\theta}) = \frac{1}{n(n-1)h} \sum_{i=1}^{n} \sum_{j=1,j\neq i}^{n} (Y_i - g_0(\langle X_i, \widehat{\theta}\rangle))(Y_j - g_0(\langle X_j, \widehat{\theta}\rangle))K_{ij}^{\widehat{\theta}}$$

$$= \frac{1}{n(n-1)h} \sum_{i=1}^{n} \sum_{j=1,j\neq i}^{n} \mathcal{E}_i\mathcal{E}_j K_{ij}^{\widehat{\theta}}+$$

$$+2\frac{1}{n(n-1)h} \sum_{i=1}^{n} \sum_{j=1,j\neq i}^{n} \mathcal{E}_i(g_0(\langle X_j, \theta_\star\rangle) - g_0(\langle X_j, \widehat{\theta}\rangle))K_{ij}^{\widehat{\theta}}+$$

$$+\frac{1}{n(n-1)h} \sum_{i=1}^{n} \sum_{j=1,j\neq i}^{n} (g_0(\langle X_i, \theta_\star\rangle) - g_0(\langle X_i, \widehat{\theta}\rangle))\times$$

$$\times \big(g_0(\langle X_j, \theta_\star \rangle) - g_0(\langle X_j, \widehat{\theta} \rangle) \big) K_{ij}^{\widehat{\theta}}$$
$$= \widehat{Q}_n^A(\theta) + 2\widehat{Q}_n^B(\theta) + \widehat{Q}_n^C(\theta)$$

The following result provides the asymptotic behaviour of each term in the previous decomposition and it demonstrates that the leading term is $\widehat{Q}_n^A(\theta)$ whilst the other two terms are negligible with respect to the first one.

Theorem 2 Under Assumptions S, M, K and E and the null hypothesis H_0 holds true, one has:

(i) $\ \ |\widehat{Q}_n^A(\theta)| = O_{\mathbb{P}}\left(\dfrac{p^{3/2} \ln n}{nh^{1/2}} \right),$

(ii) $\ \ |\widehat{Q}_n^B(\theta)| = o_{\mathbb{P}}\left(\dfrac{p^{3/2} \ln n}{nh^{1/2}} \right)$ and $|\widehat{Q}_n^C(\theta)| = o_{\mathbb{P}}\left(\dfrac{p^{3/2} \ln n}{nh^{1/2}} \right),$

(iii) as n goes to infinity, $n\sqrt{h}\, Q_n(\widehat{\theta}) \big/ v_n(\widehat{\theta}) \sim \mathcal{N}(0, 1).$

In what follows, a short discussion on the assumptions is reported. Firstly, note that since $\widehat{Q}_n^A(\theta)$ is the version of $Q_n(\theta)$ when one deals with complex null hypothesis, the argument to deduce the rate of convergence as in Theorem 2, statement (i), proceeds very similarly to the simple null hypothesis situation. Hence, the sets of conditions M and K on the models and on the kernel function respectively are still necessary but with certain modifications or additions with respect to the simple null hypothesis case, such as M-iii.1.bis and M-iii.2.bis. Extra assumption on the kernel, such as the trade off between the dimension p and the sample size n, allows one to establish the required result. More in detail, the fact that the kernel is a continuous density of bounded variation and the regularity condition on the link function g_0 (assumptions K-i. and M-iv. respectively) allow to invoke an inequality of Sherman (see Sherman 1994). Secondly, the assumption E-i. is rather standard to derive a uniform rate of convergence for the estimator of the link function in the SFIM (see, for instance, Novo et al. 2019). It imposes that the space of possible directions is finite but it can grow up to infinity as the sample size n increases and the direction of interest θ becomes closer to $\theta_\star$ and consequently also for $\widehat{\theta}$ as it is an element of Θ_n. The latter assumption, the boundedness of X (assumption S-ii.) and the extra condition on the trade-off between the sample size n and the bandwidth h (assumption E-ii.) combined with Sherman's inequality allow us to deduce the rate of convergence of the statement (ii) of Theorem 2.

When β is also not specified, a least square estimate $\widehat{\beta}$ can be used as mentioned previously. In that case, if β belongs to a compact subset of $\mathbb{R}^{d+1}$, if the used estimate achieves the rate n^{-r}, $0 < r \leq 1/2$, and if there exists a positive constant C such that for any fixed argument u, $|g_0^\beta(u) - g_0^{\beta'}(u)| \leq C\|\beta - \beta'\|$, the result in Theorem 2 remains valid (see the proof in the Appendix). This is verified, for instance, when g_0^β is a polynomial model.

3.3 Some remarks on consistency

Consider the following alternative hypothesis:

$$H_1 : g \in \mathcal{G}_1 = \left\{ g_1 = g_0^\beta + \gamma_n G, \ G : \mathbb{R} \to \mathbb{R}, \ G \notin \mathcal{G}_0 \right\} \tag{8}$$

where G is a smooth function and γ_n is a positive sequence tending to zero as $n \to +\infty$.

The following result provides the asymptotic behaviour of the test statistic under the considered alternative hypothesis. For the sake of simplicity the case in which the parameter β and the direction θ in the SFIM are completely specified (and equals to $\beta_\star$ and $\theta_\star$ respectively) is dealt with.

Theorem 3 Under the alternative hypothesis (8), the assumptions S, M and K of Theorem 1, if G is bounded variation and $f_{\theta_\star}$ is continuous and bounded, and $G f_{\theta_\star} \neq 0$ a.s., and if $\gamma_n^2 h^{-1/2}$ diverges as n goes to infinity, then the following statement holds:

$$n\sqrt{h}\, Q_n(\theta_\star)/v_n(\theta_\star) \longrightarrow +\infty \quad \text{in probability as } n \to +\infty. \tag{9}$$

The proof of this result is deferred in Appendix.

4 Simulation study

In this section the finite sample properties of the test proposed are explored by evaluating the empirical level and power under different experimental conditions. For each setting, the empirical power is computed as the proportion of times in which the test rejects the null hypothesis at the nominal level α (here $\alpha = 5\%$) over 1000 Monte Carlo replications. The critical region of the test is based on the Gaussian approximation of the null distribution provided in Theorem 2: one rejects the null hypothesis whenever the value of the studentized test statistics is greater than the quantile of order $1 - \alpha$ of the standard normal distribution. All the experiments are conducted using the software R.

The data used in all the simulations are generated according to the following SFIM model:

$$Y_i = g(\langle \theta, X_i \rangle) + \sigma \mathcal{E}_i \quad i = 1, \ldots, n$$

with $n = 50, 100, 200$, corresponding to small and medium sample sizes.

For any $i = 1, \ldots, n$, the functional covariate obeys to:

$$X_i(t) = 2a_i t^2 + b_i \sin(2\pi t) + c_i \cos(4\pi t) \quad t \in [0, 1]$$

where a_i, b_i, c_i are i.i.d. uniform r.v.s over $(-1, 1)$, so that the random curves are centered and bounded so that assumption S-ii. is satisfied; every trajectory is discretized over a grid of 100 equispaced design points. A sample of 30 of such a functional data is plotted in Fig. 2.
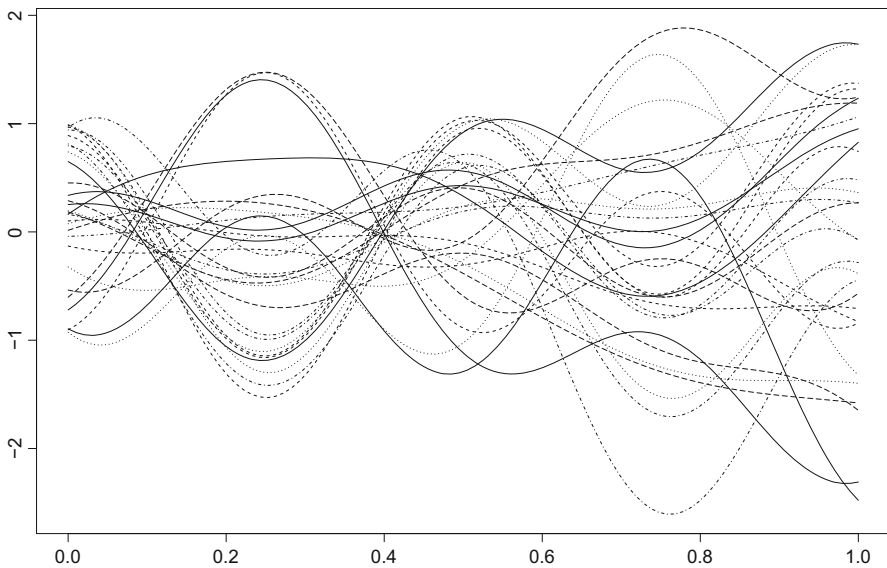
Fig. 2 A sample of 30 curves $X(t)$ randomly selected

For what concerns the functional coefficient, one uses the normalized direction:

$$\theta(t) = \sqrt{2}\sin(2\pi t) \qquad t \in [0, 1].$$

Due to the nature of the involved objects, the r.v.s $\langle \theta, X_i \rangle = \sqrt{2}(b_i/2 - a_i/\pi)$ are centered, symmetric, bounded, and with a strictly positive density f_θ over $\left(-(\pi + 2)/(\sqrt{2}\pi), (\pi + 2)/(\sqrt{2}\pi) \right)$ exhibiting a trapezoidal behavior over that interval, satisfying in this way the assumption M.iii.1. In practice, since one works with discretized curves, all the integrals $\langle \theta, X_i \rangle$ are approximated by summations.

About the error in the model, $\mathcal{E}_i$ are i.i.d. standard Gaussian r.v.s and, to control the signal-to-noise ratio, the variability coefficient σ is defined by $\sigma^2 = \rho^2 Var(g(\langle \theta, X \rangle))$, where the latter variance is estimated for each sample using the data. Here $\rho^2 = 0.2$ and 0.5 corresponding to a theoretical coefficient of determination R^2 of about 0.83 and 0.67, respectively. These choices guarantee that the assumptions M.i. and M.ii. hold.

In all the experiments some composite hypotheses, with both β and θ unknown, are tested. In particular, testing for linear and cubic link are analyzed in details (all the tested functions satisfy the assumption M.iv.). Hence, to operationalize the test procedure some estimate of β and θ and the evaluation of the bandwidth h are necessary. For what concerns β, the standard OLS approach is used, whereas to estimate θ one adopts the first step of the approach proposed in Ferraty et al. (2013). Here one uses cubic splines with 10 internal knots and the Epanechnikov Kernel

$$K(u) = \frac{3}{4}\left(1 - u^2\right) \qquad |u| \leq 1.$$

The same kernel is also used in evaluating the test statistic: since the bandwidth h is related to the estimate of f_θ, it is selected by the unbiased cross-validation approach that one uses in estimating that density. Due to the non parametric nature of the test statistic, the selection of the bandwidth could have an impact on the results of the simulation study. On the other hand, a systematic analysis of the effects on the performances of the test caused by changing the bandwidth goes beyond the scope of this work; therefore it is not performed.

4.1 Testing a linear affine link

Let $\beta_0, \beta_1 \in \mathbb{R}$, $\beta_1 \neq 0$, $u = \langle \theta, x \rangle$, θ unknown, $|u| < (\pi + 2)/\left(\sqrt{2}\pi\right)$ and consider the complex null hypothesis of a linear affine link:

$$H_0 : g(u) = \beta_0 + \beta_1 u$$

and the following alternatives:

$$H_1^{(1)}(\gamma) : g(u) = \beta_0 + \beta_1 u + \gamma \left(u\sqrt{2}\pi/(\pi + 2)\right)^2$$
$$H_1^{(2)}(\gamma) : g(u) = \beta_0 + \beta_1 u + \gamma \sin\left(u\sqrt{2}\pi^2/(\pi + 2)\right)$$
$$H_1^{(3)}(\gamma) : g(u) = \beta_0 + \beta_1 u + \gamma \cos\left(u\sqrt{2}\pi^2/(\pi + 2)\right)$$

where $\gamma > 0$, which controls the departure from the null hypothesis. To generate the models under the null and the alternative hypotheses are used $\beta_0 = 0$, $\beta_1 = 1$ so that $\mathbb{E}[g(U)] = 0$ and $\gamma = 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 1.2, 1.4$.

In order to appreciate the differences between the considered models, the shapes of the functions g under the null and the alternative hypotheses with $\gamma = 0.5$ are drawn in Fig. 3: one can note that the quadratic perturbation involved in $H_1^{(1)}$ interest mainly the tails of the distribution of the projected data $\langle \theta, x \rangle$, the sinus in $H_1^{(2)}$ the central parts of the two halves of the interval, whereas the cosinus in $H_1^{(3)}$ acts both on the central part and the tails of the interval. Hence one expects that the test performs rather well in the last case for any sample size n and ρ, also with γ close to zero. On the other hand, it is expected that the performances are not so good, at least for small samples and large ρ, for the first two alternatives $H_1^{(1)}$ and $H_1^{(2)}$ where a rather large sample size is necessary to detect small deviations from the linearity.

The estimated powers varying γ for the different considered scenarios are represented in Figs. 4, 5 and 6. First, note that, despite an asymptotic approximation for the null distribution is used, the empirical level is rather close to the theoretical one, also for a relatively small sample size. Second, as expected, for any sample size n and given ρ^2, the further one moves away from the linearity by increasing γ, the greater is the estimated power. In any case, the performances are very good when $n = 200$ and $\rho^2 = 0.2$, also with a relatively modest departure from the null hypothesis. Looking at more in detail, the graphs support the previous comments that relate the nature of
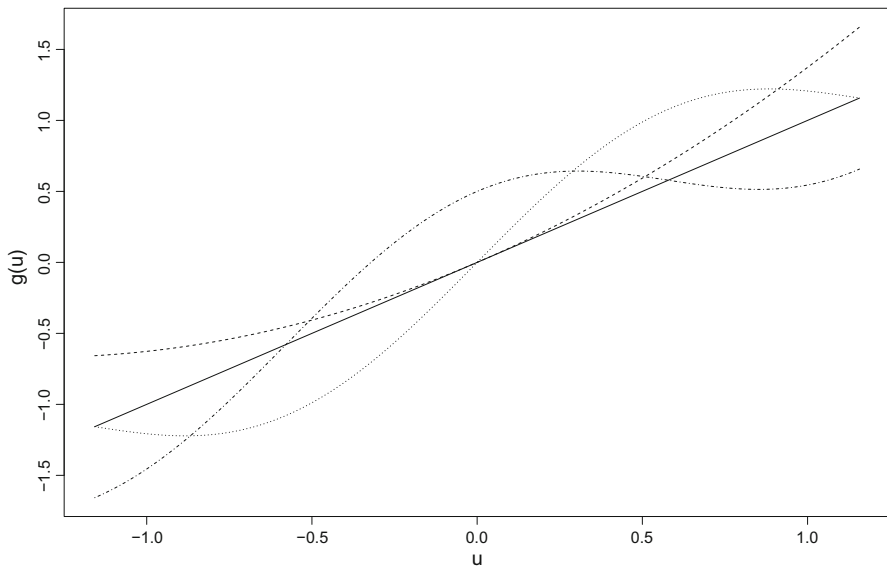
Fig. 3 Shapes of the link functions g under H_0 (solid), $H_1^{(1)}(0.5)$ (dahsed), $H_1^{(2)}(0.5)$ (dotted) and $H_1^{(3)}(0.5)$ (dotdash)
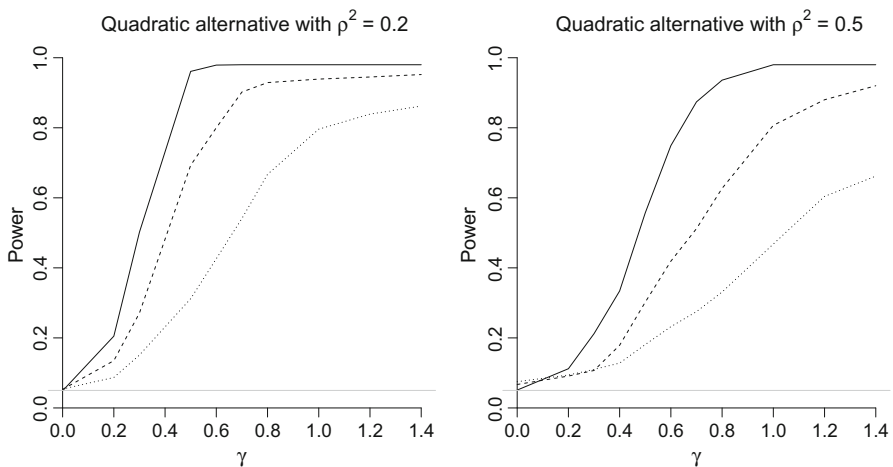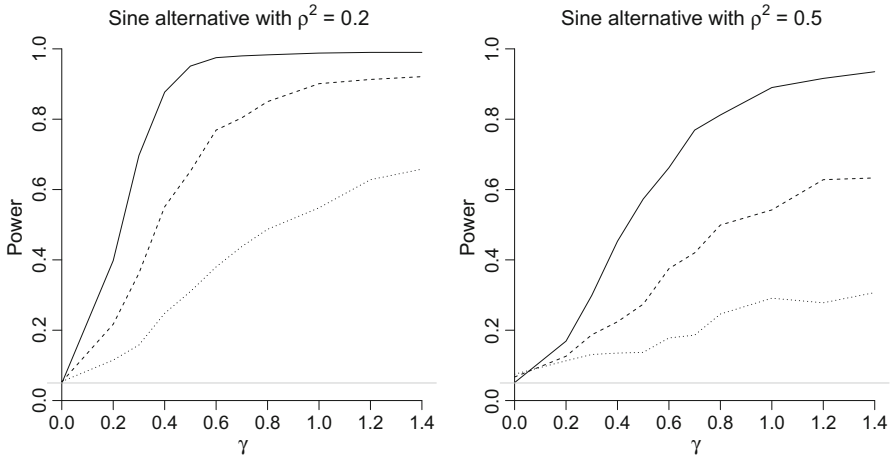

Fig. 4 Power curves for the alternatives $H_1^{(1)}(\gamma)$ varying γ and for $\rho^2 = 0.2$ (left panel) and $\rho^2 = 0.5$ (right panel). The grey line represents the nominal level, the dotted line $n = 50$, the dashed line $n = 100$, and the solid line $n = 200$

the link functions and the power behaviour. In particular the alternatives in the family indexed by $H_1^{(3)}$ produce the best result also for γ rather small, whereas it appears more complex to detect correctly $H_1^{(1)}$ and $H_1^{(2)}$, at least for small γ, small samples and a rather low signal-to-noise.

Fig. 5 Power curves for the alternatives $H_1^{(2)}(\gamma)$ varying γ and for $\rho^2 = 0.2$ (left panel) and $\rho^2 = 0.5$ (right panel). The grey line represents the nominal level, the dotted line $n = 50$, the dashed line $n = 100$, and the solid line $n = 200$
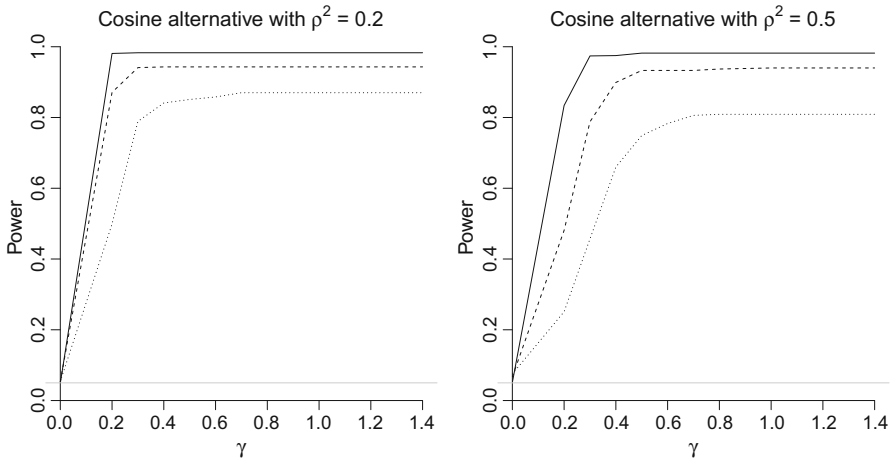
Fig. 6 Power curves for the alternatives $H_1^{(3)}(\gamma)$ varying γ and for $\rho^2 = 0.2$ (left panel) and $\rho^2 = 0.5$ (right panel). The grey line represents the nominal level, the dotted line $n = 50$, the dashed line $n = 100$, and the solid line $n = 200$

4.2 Testing a cubic link

Let $\beta_0, \beta_1 \in \mathbb{R}$, $\beta_1 \neq 0$, $u = \langle \theta, x \rangle$, θ unknown, $|u| < (\pi + 2)/\left(\sqrt{2}\pi\right)$. The second experiment considers the following null hypothesis:
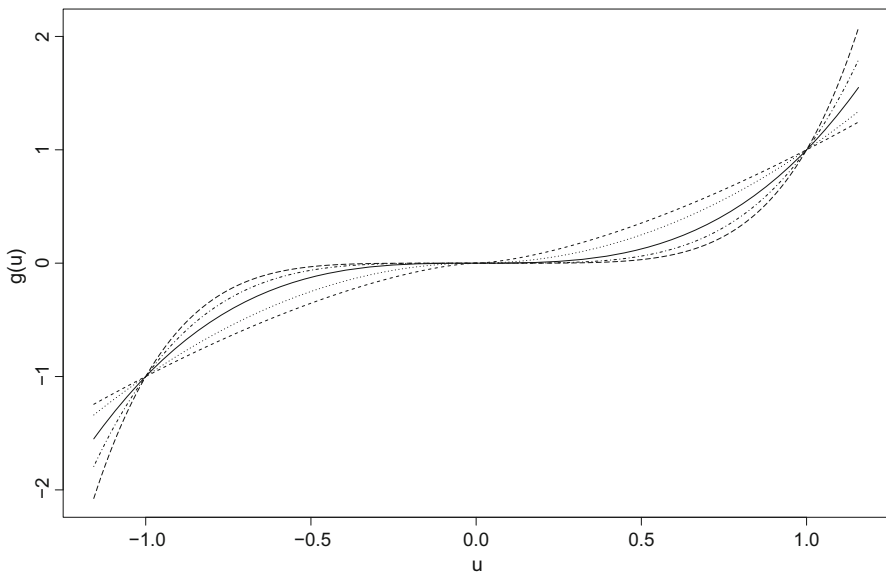
$$H_0 : g(u) = \beta_0 + \beta_1 u^3$$

Fig. 7 Shapes of the link functions g under H_0 (solid), $H_1^{(1)}(1.5)$ (dahsed), $H_1^{(2)}(2)$ (dotted), $H_1^{(3)}(4)$ (dotdash) and $H_1^{(3)}(5)$ (longdash)

against the following alternatives

$$H_1(\gamma) : g(u) = \beta_0 + \beta_1 \left(u^\gamma \mathbb{I}_{u>0} - |u|^\gamma \mathbb{I}_{u \leq 0} \right) \qquad \gamma \neq 3$$

where $\mathbb{I}_A$ is the indicator function of the set A, and $\gamma > 0$ measures the departure from the null hypothesis. To generate the models one uses $\beta_0 = 0$, $\beta_1 = 1$ and $\gamma = 1, 1.5, 2, 2.5, 3.5, 4, 4.5, 5$ (the behaviour of some link functions for some selected values for γ are drawn in Fig. 7).

Also in this second experiment, the obtained results, visualized in Fig. 8, are generally rather good. One can note that the estimated level is slightly higher than the nominal one, providing a liberal test in particular for relatively small sample size. For any fixed n and ρ^2, the estimated power increases coherently with the departure from the null hypothesis. In particular it emerges that it is possible to better discriminate situation with $\gamma < 3$ than $\gamma > 3$ where the test is slightly less efficient. Anyway, for rather large samples the results obtained are good also for values of γ close to zero and $\rho^2 = 0.5$. In general, the results corroborate the fact that the Gaussian approximation of the null distribution works reasonably well.

5 Some bootstrap procedures

Despite the use of the asymptotic null distribution in defining the critical region of the test has proven capable of producing good results, one can explore the possibility to
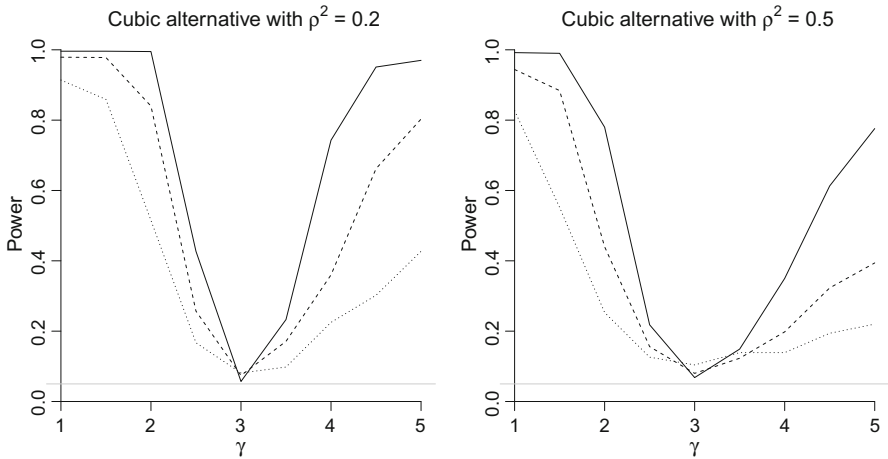
Fig. 8 Power curves for the alternatives $H_1(\gamma)$ varying γ and for $\rho^2 = 0.2$ (left panel) and $\rho^2 = 0.5$ (right panel). The grey line represents the nominal level, the dotted line $n = 50$, the dashed line $n = 100$, and the solid line $n = 200$

estimate the threshold of the critical region by using the quantiles calculated through some bootstrap algorithms. Since methods based on bootstrapping the pairs (X_i, Y_i) are not adapted, approaches based on boostrapping residuals are adopted.

The general procedure is described in the steps below:

1. Estimate β and θ and then compute the errors under the null hypothesis $\widehat{\mathcal{E}}_i$ and the value of the test statistic T_n;
2. Compute the bootstrap version $T_n^\star$ of the test statistics by using the bootstrapped errors $\widehat{\mathcal{E}}_i^\star$ (see below for details);
3. Repeat step 2. a large number B of times and compute the $(1 - \alpha)$-quantile $\tau_\alpha^\star$ of the distribution of $T_n^\star$;
4. Compare T_n with $\tau_\alpha^\star$: if the value of T_n is larger than $\tau_\alpha^\star$ reject the null hypothesis.

For what concerns step 2., both naive and wild bootstrap approaches are adopted:

(i) the naive bootstrap is based on a direct resampling with replacement of the estimated errors $\widehat{\mathcal{E}}_i$, $i = 1, \ldots n$;
(ii) the wild bootstrap errors are calculated as $\widehat{\mathcal{E}}_i^\star = \widehat{\mathcal{E}}_i \xi_i$, where ξ_i are i.i.d. and independent on (X_i, Y_i). In the experiment three different distributions for ξ_i are used:

   (a) the Rademacher distribution with equiprobable values $\{-1, 1\}$;
   (b) the distribution suggested by Mammen (1993) with values $\left(1 - \sqrt{5}\right)/2$ and $\left(1 + \sqrt{5}\right)/2$ with associated probability $\left(\sqrt{5} + 1\right)/\left(2\sqrt{5}\right)$ and $\left(\sqrt{5} - 1\right)/\left(2\sqrt{5}\right)$ respectively;
   (c) the standard Gaussian distribution.

Table 1 Estimated level and powers for testing quadratic specification under different experimental conditions when one uses the asymptotic null distribution (A), the naive bootstrap (N), the wild boostrap with Rademacher (R), Mammen (M) or Gaussian (G) distributions

Hypothesis	A	N	R	M	G
$H_1(1)$	0.841	0.839	0.848	0.825	0.829
$H_1(1.5)$	0.354	0.293	0.311	0.318	0.303
$H_1(2) : H_0$	0.064	0.051	0.047	0.052	0.047
$H_1(2.5)$	0.210	0.196	0.197	0.170	0.186
$H_1(3)$	0.655	0.616	0.599	0.605	0.584

To assess the performances of the test when the asympototic null distribution and the bootstrap approaches are employed, an experiment similar to the one presented in Sect. 4 is carried out, choosing $B = 1000$ in step 3. of the algorithm.

In particular let $\beta_0, \beta_1 \in \mathbb{R}$, $\beta_1 \neq 0$, $u = \langle \theta, x \rangle$, θ unknown, $|u| < (\pi + 2)/\left(\sqrt{2}\pi\right)$. Consider the null hypothesis of a quadratic link:

$$H_0 : g(u) = \beta_0 + \beta_1 u^2$$

against the alternatives

$$H_1(\gamma) : g(u) = \beta_0 + \beta_1 |u|^\gamma \qquad \gamma \neq 2$$

where $\gamma = 1, 1.5, 2.5, 3$. For what concerns the data generation process, one fixes $n = 100$, and $\beta_0 = 0$, $\beta_1 = 1$, and $\rho^2 = 0.2$.

The estimated powers varying γ (when $\gamma = 2$ one deals with the null hypothesis), on the base of 1000 MC replications, are collected in Table 1. The results obtained in the different cases are very similar: it emerges that the test based on the asymptotic distribution is slightly more liberal than all the tests based on the bootstrap approaches; for what concerns the estimated power it seems that the performances are very similar in all the analyzed cases. In conclusion, the test that uses the quantiles of the standard Gaussian performs quite well also when rather small samples are available.

6 Application to spectrometric data

An important task in domains like chemistry, medicine or food industry is to get the composition of a given substance. Since chemical analysis is rather expensive and require time, it is often preferred to estimate that composition by using spectrometric curves that can be easily obtained as the absorption of a reflected light for various wavelengths.

In this section one considers an example of such a modeling from food industry. The dataset (available at https://www.eigenvector.com/data) consists of 80 samples of corn to each of which corresponds both the values of moisture, oil, protein and starch

contents obtained by chemical analysis and the spectrometric curve measured by NIR spectrometers on the wavelength range 1100–2498 and discretized over an equispaced mesh of 700 points (the set of these curves is reproduced in Fig. 1). The aim is to model the chemical composition of the corn by using the spectrometric curves, or better, as it is done in similar context, their derivatives.

A systematic study on this dataset has been carried out in Delsol (2013) where no effect tests in functional regression were performed taking as explanatory variables the original spectrometric curves and their successive derivatives, from the first to the fourth order. According to the results in the cited paper, it emerges that the first derivative has a significant effect on moisture content, no significant effects are detected for oil and starch content, whereas the effect of the fourth derivative on the protein is not evident. On the other hand, the first derivative on the wavelength range 2010–2220 exhibits a significant effect on the protein content.

Starting from the latter evidences, one concentrates the attention on modelling by a SFIM the moisture content by using as covariate the first derivative on the whole wavelength range, and the protein content with covariate the first derivative on the range 2010–2220. The estimated links g and directions θ, obtained by using the same procedure adopted in Sect. 4 with cubic splines with 10 and 2 knots respectively, are depicted in Fig. 9. Both graphs suggest that the link functions g in the SFIMs could be specified linearly: therefore the test for linearity illustrated in Sect. 4.1 is performed.

In particular one wants to test the following model specification:

$$H_0 : Y = \beta_0 + \beta_1(\langle X, \theta \rangle) + \mathcal{E}$$

Since all the parameters involved (β_0, β_1 and θ) are unknown, they must be estimated from the data: the $\widehat{\theta}$s are the formerly estimated directions plotted in Fig. 9, whereas $\widehat{\beta_0}$ and $\widehat{\beta_1}$ are the OLS estimated for the model under the null hypothesis with covariate $\langle X, \widehat{\theta} \rangle$. For the sake of completeness, the estimates obtained are as follows: $\widehat{\beta_0} = 18.57$ and $\widehat{\beta_1} = -37.80$ when the response is the moisture content, $\widehat{\beta_0} = 13.74$ and $\widehat{\beta_1} = 20.91$ when the response is the protein content. The behaviour of the residuals of such a models, from which the test statistics are calculated, is shown in Fig. 10.

By using the asymptotic null distribution, the following results are achieved: the p-value for the model having as response the moisture content is 0.695 whereas the one when the response is the protein content equals 0.719. Hence, one can conclude that in both cases the hypothesis of linearity of the link function appears compatible with the empirical evidence.

Since one tests the linearity of the link functions, to complete the analysis, a comparison with the linearity test proposed by Garcia-Portugues et al. (2014) and available in the R package fda.usc is performed. To be coherent with what is done above, one uses the estimated method based on B-spline with the same number of elements of the basis considered in the SFIM. For what concerns the model with response the moisture content, the p-value is 0.217, whereas for the model having as response the protein content, the p-value equals 0.224. In both cases, there is no reason to reject of the hypothesis of linearity of the model: the latter results are then coherent with what emerged by the proposed test.
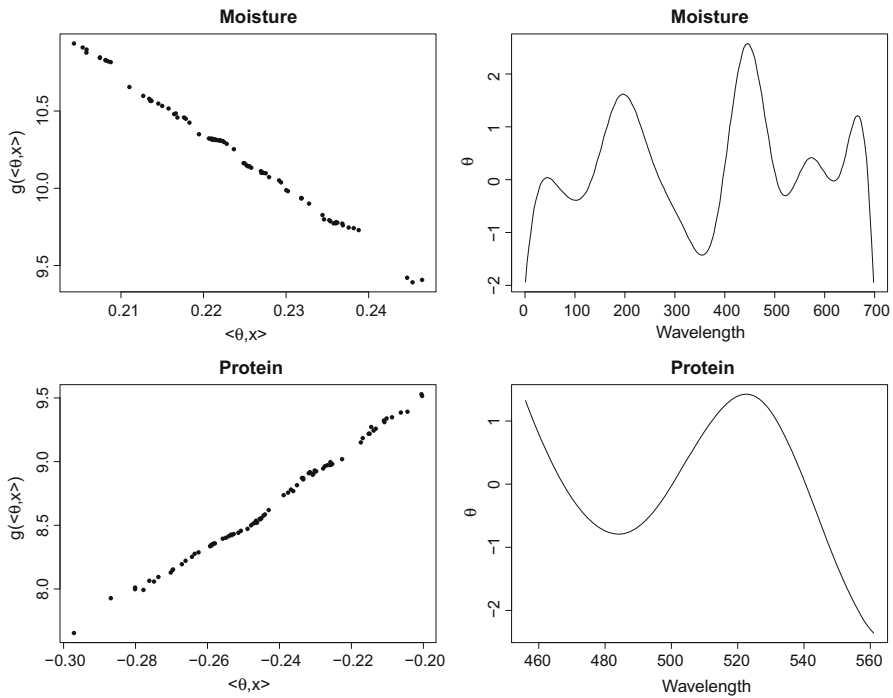
Fig. 9 Estimated link functions g and directions θ when the response is the moisture content (top) or the protein content (bottom)
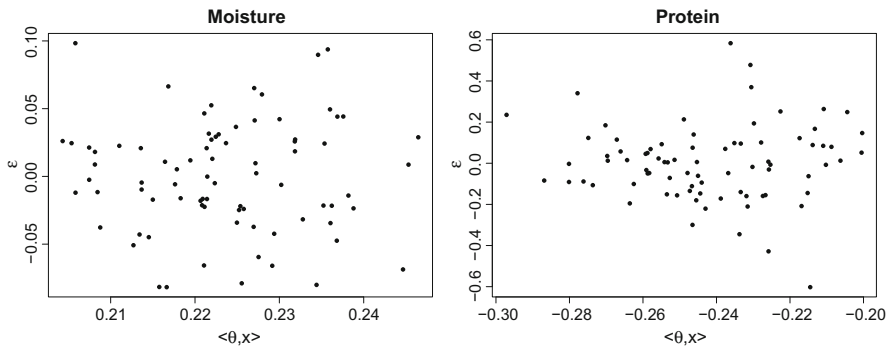

Fig. 10 Estimated errors under the null hypothesis when the response is the moisture content (left) or the protein content (right)

Declarations

Conflict of interest The authors declare that there is no conflict of interest.

Appendix A: Proofs

In this appendix the detailed proofs of theorems 1 and 2 are provided. The results are inspired by the ones in Patilea et al. (2019) and reported here for the sake of completeness and to make the work self-contained as much as possible.

To derive the main results, some properties of the Euclidean class of functions are considered; for more details, see Definition 2.7 and lemmas 2.13, 2.14 and 2.15 in Pakes and Pollard (1989).

Let introduce the following notation, where x denotes a fixed element in $\mathcal{H}$:

$$\mathcal{K}_\theta = \{u \longrightarrow K(\langle u - x, \theta \rangle / h), h > 0\}$$
$$\mathcal{K}_{\Theta_n} = \bigcup_{\theta \in \Theta_n} \mathcal{K}_\theta = \{u \longrightarrow K(\langle u - x, \theta \rangle / h), h > 0, \theta \in \Theta_n\}$$

It is possible to verify that the class of functions $\mathcal{K}_\theta$ is Euclidean for a constant envelope thanks to the bounded variation assumption K-i. together with the fact that $\langle X, \theta \rangle$ is a real-valued measurable map. Combining Lemma 22(ii) in Nolan and Pollard (1987) and Lemma 2.15 of Pakes and Pollard (1989) allow to reach the conclusion. Similarly, from Lemma 22(ii) in Nolan and Pollard (1987) the class $\mathcal{K}_{\Theta_n}$ is Euclidean for a constant envelope.

A.1: Proof of Theorem 1

Statement (i)

Define $\mathcal{E}_{M,i} = \mathcal{E}_i \mathbb{I}_{\{|\mathcal{E}_i| \leq M\}} - \mathbb{E}[\mathcal{E}_i \mathbb{I}_{\{|\mathcal{E}_i| \leq M\}} | X_i]$, where M depends on n and will be specified below, and take $\theta = \theta_\star$. Let

$$U_{M,n}(\theta) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \mathcal{E}_{M,i} \mathcal{E}_{M,j} K_{ij}^{\theta_\star} \tag{A1}$$

and $Q_{M,n}(\theta_\star) = U_{M,n}(\theta_\star)/h$. The class of all functions $\mathcal{K}_{\theta_\star}$ is Euclidean and by assumption M-iv. together with Lemma 2.13 of Pakes and Pollard (1989), $\mathcal{E}_{M,i}$ is Euclidean and similarly for $\mathcal{E}_{M,j}$. Since the product of classes of Euclidean functions is Euclidean (see Lemma 2.14(ii) of Pakes and Pollard 1989), then also the kernel of the U-statistics in (A1) is.

Therefore, Theorem 2 in Major (2006) can be invoked and for any $t > 0$ one has

$$\mathbb{P}\left(|U_{M,n}(\theta_\star)| \geq \frac{t h^{1/2} \ln n}{(n-1)}\right) \leq C_1 C_2 \exp\left(-C_3\left(\frac{t h^{1/2} \ln n}{M^2 \sigma_M}\right)\right) \tag{A2}$$

provided that

$$n\sigma_M^2 \geq \frac{t h^{1/2} \ln n}{M^2 \sigma_M} \geq C_4 \left[C_5 + \max\left(\frac{\ln C_2}{\ln n}, 0\right)\right]^{3/2} \ln\left(\frac{2}{\sigma_M}\right) \tag{A3}$$

where $C_1, \ldots, C_5$ are some positive constants and

$$\sigma_M^2 = \mathbb{E}\left[\left(\frac{\mathcal{E}_{M,1}}{M}\right)^2 \left(\frac{\mathcal{E}_{M,2}}{M}\right)^2 (K_{1,2}^{\theta_\star})^2\right].$$

One has to verify now condition (A3). To do this, apply the tower law to the definition of σ_M^2 to obtain:

$$\sigma_M^2 = \frac{1}{M^4} \mathbb{E}\left[(K_{1,2}^{\theta_\star})^2 \mathbb{E}\left[\mathcal{E}_{M,1}^2 | X_1\right] \mathbb{E}\left[\mathcal{E}_{M,2}^2 | X_2\right]\right].$$

The behaviour of $\mathbb{E}\left[(\mathcal{E}_{M,i})^2 | X_i\right]$, $i = 1, 2$ are investigated: thanks to assumption M-ii., it is bounded above and below by positive constants. It remains to study $\mathbb{E}\left[(K_{1,2}^{\theta_\star})^2\right]$.

Apply the Fourier inversion theorem to the kernel $(K_{1,2}^{\theta_\star})^2$ and since X_1, X_2 are independent then

$$\frac{1}{h}\mathbb{E}\left[(K_{1,2}^{\theta_\star})^2\right] = \frac{1}{2\pi}\mathbb{E}\left[\int_{\mathbb{R}} e^{it\langle X_1, \theta_\star\rangle} e^{-it\langle X_2, \theta_\star\rangle} \mathcal{F}[K(t)^2] dt\right]$$

$$= (2\pi)^{1/2} \int_{\mathbb{R}} |\mathcal{F}[f_{\theta_\star}(t)]|^2 \mathcal{F}[K(t)^2] dt. \tag{A4}$$

By Fubini theorem and then applying the Plancherel theorem together with assumption K-i., one gets

$$\int_{\mathbb{R}} |\mathcal{F}[f_{\theta_\star}(t)]|^2 \mathcal{F}[K(t)^2] dt \leq C \int_{\mathbb{R}} |\mathcal{F}[f_{\theta_\star}(t)]|^2 dt = C \int_{\mathbb{R}} f_{\theta_\star}(x)^2 dx$$

where C is a constant and the latter is bounded by a constant by assumption M-iii.1. On the other hand, (A4) is bounded from below: assumptions K-i. and M-iii.2. allow to write

$$\int_{\mathbb{R}} |\mathcal{F}[f_{\theta_\star}(t)]|^2 \mathcal{F}[K(t)^2] dt \geq C_3 \int_{\mathbb{R}} |\mathcal{F}[f_{\theta_\star}(t)]|^2 dt \geq C_3 \int_{|t| \leq \varepsilon} |\mathcal{F}[f_{\theta_\star}(t)]|^2 dt \geq C_4,$$

where C_3, C_4 are some constants.

Combining the above arguments, there exists a constant $C > 0$ such that $1/C \leq \sigma_M^2 M^4 / h \leq C$. Let

$$M^4 = \frac{nh}{(\ln n)^{1+\delta}} \to \infty$$

with $\delta > 0$ arbitrarily small, then $\sigma_M^2 \leq C(\ln n)^{1+\delta}/n$, and therefore σ_M^2 is of order $(\ln n)^{1+\delta}/n \to 0$. For any $t > 0$, using the fact that $1/C \leq \sigma_M^2 M^4 / h$, then the left hand side inequality in (A3) holds provided n is large enough.

Since $\sigma_M^2 M^4 / h \leq C$ there exists a positive constant C' such that

$$\frac{t h^{1/2} \ln n}{M^2 \sigma_M} \geq C' \ln n$$

which goes to infinity for sufficiently large t. Noting that $\ln(2/\sigma_M)/\ln n$ tends to a non-negative constant as $n \to \infty$, the right hand side inequality in (A3) holds.

Thanks to Major inequality (A2), one deduces that $|Q_{M,n}(\theta_\star)| = O_{\mathbb{P}}(\ln n/(nh^{1/2}))$ and the proof of Statement (i) is completed once one can prove that $|Q_n(\theta_\star) - Q_{M,n}(\theta_\star)| = o_{\mathbb{P}}(\ln n/(nh^{1/2}))$.

Note that $Q_n(\theta_\star) - Q_{M,n}(\theta_\star) = 2R_{1n} + R_{2n}$, where

$$R_{1n} = \frac{1}{n(n-1)h} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \mathcal{E}_{M,i} \asymp_j K_{ij}^{\theta_\star}$$

and

$$R_{2n} = \frac{1}{n(n-1)h} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \zeta_i \zeta_j K_{ij}^{\theta_\star},$$

where $\zeta_i = \mathcal{E}_i \mathbb{I}_{\{|\mathcal{E}_i| \geq M\}} - \mathbb{E}[\mathcal{E}_i \mathbb{I}_{\{|\mathcal{E}_i| \geq M\}} | X_i]$. By assumptions S-i. and K-i., one deduces that

$$
h\mathbb{E}[|R_{1n}|] = \mathbb{E}\left[\left|\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} \mathcal{E}_{M,i} \asymp_j K_{ij}^{\theta_\star}\right|\right]
$$
$$
\leq C\mathbb{E}[|\mathcal{E}_{M,1} \asymp_2|] \leq 2C\mathbb{E}[|\mathcal{E}_1|]\mathbb{E}[|\asymp_2|].
$$

Applying the Hölder inequality and Markov inequality, one obtains

$$
\mathbb{E}[|\asymp_2|] \leq 2\mathbb{E}[|\mathcal{E}_2|\mathbb{I}_{\{|\mathcal{E}_2|\geq M\}}]
$$
$$
\leq 2\mathbb{E}[|\mathcal{E}_2|^m]^{1/m}\mathbb{P}(|\mathcal{E}_2| \geq M)^{(m-1)/m}
$$
$$
\leq 2\mathbb{E}[|\mathcal{E}_2|^m]M^{1-m}.
$$

Using assumption K-ii. and with the choice of $m > 11$ as in assumption M-i., one has $M^{1-m} = o(h^{1/2}\ln n/n)$. Moreover, since $|R_{2n}|$ is of smaller order of $|R_{1n}|$. Invoke the Markov inequality and the proof is concluded.

Statement (ii)

In what follows, the dependence on $\theta_\star$ is dropped if there is not any confusion. Define

$$
v_n^2(\theta_\star) = \frac{1}{n(n-1)h}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} \mathcal{E}_i^2\mathcal{E}_j^2 K^2\left(\frac{|\langle X_i - X_j, \theta_\star\rangle|}{h}\right),
$$

and its the theoretical counterpart

$$
v^2 = \frac{1}{h}\mathbb{E}\left[\mathcal{E}_1^2\mathcal{E}_2^2 K^2\left(\frac{|\langle X_1 - X_2, \theta_\star\rangle|}{h}\right)\right].
$$

Define

$$
\kappa(Z_i, Z_j) = \mathcal{E}_i^2\mathcal{E}_j^2 K^2\left(\frac{|\langle X_i - X_j, \theta_\star\rangle|}{h}\right).
$$

and

$$
\kappa_1(x) = \mathbb{E}[\kappa(x, Z_2)] = \mathbb{E}\left[\mathcal{E}_x^2\mathcal{E}_2^2 K^2\left(\frac{|\langle x - X_2, \theta_\star\rangle|}{h}\right)\right]
$$
$$
\kappa_1(y) = \mathbb{E}[\kappa(Z_1, y)] = \mathbb{E}\left[\mathcal{E}_1^2\mathcal{E}_y^2 K^2\left(\frac{|\langle X_1 - y, \theta_\star\rangle|}{h}\right)\right]
$$
$$
\kappa(x, y) = \mathcal{E}_x^2\mathcal{E}_y^2 K^2\left(\frac{|\langle x - y, \theta_\star\rangle|}{h}\right).
$$

The centered Hoeffding decomposition allows to write:

$$h\left(v_n^2 - v^2\right) = \frac{2}{n}\sum_{i=1}^{n}\left(\kappa_1(X_i) - hv^2\right)$$

$$+ \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i,j=1}^{n}\left(\kappa(X_i,X_j) - \kappa_1(X_i) - \kappa_1(X_j) - hv^2\right)$$

$$= U_{1,n} + U_{2,n}$$

By using the same argument to prove that the kernel of (A1) is Euclidean, one can deduce that $U_{1,n}$ and $U_{2,n}$ are Euclidean. Consequently, one can use (Major, 2006, Theorem 2) to derive $\left|U_{2,n}\right| = O_{\mathbb{P}}\left(h^{1/2}\ln n/n\right)$ and (Vaart and Wellner 1996, Theorem 2.14.9) to obtain $\left|U_{1,n}\right| = O_{\mathbb{P}}\left(1/n^{1/2}\right)$. Gathering these rates of convergence and by assumption K-ii. it follows that the difference between the empirical version v_n^2 and its theoretical counterpart v^2 is negligible, that is $v_n^2 - v^2 = o_{\mathbb{P}}(1)$, and then the behaviour of v^2 is investigated.

By using the tower law and the independence of X_1, X_2, deduce that

$$v^2 = \frac{1}{h}\mathbb{E}\left[(K_{1,2}^{\theta_\star})^2 \mathbb{E}\left[\mathcal{E}_1^2|X_1\right]\mathbb{E}\left[\mathcal{E}_2^2|X_2\right]\right]$$

and thanks to assumption M-ii. one has:

$$\frac{\underline{\sigma}^4}{h}\mathbb{E}\left[(K_{1,2}^{\theta_\star})^2\right] \leq v^2 \leq \frac{\overline{\sigma}^4}{h}\mathbb{E}\left[(K_{1,2}^{\theta_\star})^2\right], \tag{A5}$$

Exploiting the same arguments used to study the behaviour of $\mathbb{E}\left[(K_{1,2}^{\theta_\star})^2\right]$ in statement (i), one gets that (A5) is bounded from above and below by positive constants. The proof is completed gathering the results.

Proof of Statement (iii)

Given the symmetric matrix $\mathcal{W}$ with entries

$$\mathbb{I}_{i\neq j}\frac{K_{ij}}{n(n-1)h}.$$

Denote by $Sp(\mathcal{W})$ the spectral radius and $\|\mathcal{W}\|$ the corresponding matrix norm. Given the random variables A_n, B_n, the notation $A_n \asymp B_n$ in probability means that there exists a constant $C > 0$ such that $\mathbb{P}(1/C \leq A_n/B_n \leq C)$ goes to 1 as n tends to infinity.

By Lemma 2(i) in Guerre and Lavergne (2005) $n\sqrt{h}\,Q_n/v_n$ converges to a standard normal conditionally to X_i if $Sp(\mathcal{W})/\|\mathcal{W}\|$ goes to zero in probability. In particular it is enough to show that $Sp(\mathcal{W}) = O_{\mathbb{P}}(1/n)$ and $nh^{1/2}\|\mathcal{W}\| \asymp 1$ in probability. The proof of these two results can be found in Lemma 6.2 in Lavergne and Patilea (2008) by using the assumptions K.i and K.ii.

A.2 Proof of Theorem 2

Consider first the term $\widehat{Q}_n^A(\widehat{\theta})$. Once the following result is proved:

$$\sup_{\theta \in \Theta_n} |Q_n(\theta)| = O_{\mathbb{P}}\left(\frac{p^{3/2}\ln n}{nh^{1/2}}\right)$$

the same rate of convergence for $\widehat{Q}_n^A(\widehat{\theta})$ is achieved, since $\widehat{\theta}$ is assumed to be an element of Θ_n and recall for $\widehat{Q}_n^A(\widehat{\theta})$, $\mathcal{E}_i = Y_i - g_0^\beta(\langle X_i, \theta \rangle)$. The argument to deduce that result proceeds almost identically to the simple hypothesis case by using assumptions M.iii.1.bis and M.iii.2.bis instead of M.iii.1 and M.iii.2 respectively, and assumption K.iii.

Define $\mathcal{E}_{M,i}$ as in the proof of Statement (i) of Theorem 1 and

$$U_{M,n}(\theta) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \mathcal{E}_{M,i}\mathcal{E}_{M,j} K_{ij}^{\theta} \tag{A6}$$

and $Q_{M,n}(\theta) = U_{M,n}(\theta)/h$. The class of all functions $\mathcal{K}_{\Theta_n}$ is Euclidean and, by using the same arguments to prove that (A1) is Euclidean, also the kernel of the U-statistics in (A6) is. Therefore, Theorem 2 in Major (2006) can be invoked but now also taking into consideration the properties of the space of directions Θ_n where the estimate $\widehat{\theta}$ is assumed to be an element in it, and for any $t > 0$ one has

$$\mathbb{P}\left(\sup_{\theta \in \Theta_n} |U_{M,n}(\theta)| \geq \frac{tp^{3/2}h^{1/2}\ln n}{(n-1)}\right) \leq C_1 C_2 \exp\left(-C_3\left(\frac{tp^{3/2}h^{1/2}\ln n}{M^2\sigma_M}\right)\right) \tag{A7}$$

provided that

$$n\sigma_M^2 \geq \frac{tp^{3/2}h^{1/2}\ln n}{M^2\sigma_M} \geq C_4\left[p + \max\left(\frac{\ln C_2}{\ln n}, 0\right)\right]^{3/2} \ln\left(\frac{2}{\sigma_M}\right) \tag{A8}$$

where $C_1, \ldots, C_4$ are some positive constants and

$$\sigma_M^2 = \sup_{\theta \in \Theta_n} \mathbb{E}\left[\left(\frac{\mathcal{E}_{M,1}}{M}\right)^2 \left(\frac{\mathcal{E}_{M,2}}{M}\right)^2 (K_{1,2}^{\theta})^2\right].$$

To verify condition (A8) one can evoke the same arguments used to prove Statement (i) of Theorem 1 with

$$M^4 = \frac{nh}{p^{3/2}(\ln n)^{1+\delta}} \to \infty$$

and $\delta > 0$ arbitrarily small. Thanks to Major inequality (A7), one concludes that $\sup_{\theta \in \Theta_n} |Q_{M,n}(\theta)| = O_{\mathbb{P}}(p^{3/2}\ln n/(nh^{1/2}))$.

It remains to prove that $\sup_{\theta \in \Theta_n} |Q_n(\theta) - Q_{M,n}(\theta)| = o_{\mathbb{P}}(p^{3/2} \ln n/(nh^{1/2}))$. To do this, one follows steps which are similar as in the simple hypothesis case and then they are omitted.

Turning the attention to the variance, by using assumptions M.iii.1.bis, M.iii.2.bis and K.iii., the following result is deduced

$$\sup_{\theta \in \Theta_n} \frac{1}{v_n^2(\theta)} = O_{\mathbb{P}}(1).$$

Since the proof proceeds almost identically to the simple hypothesis case, it is omitted. The main differences are in the Hoeffding decomposition where the results of the second order U-statistic is now $\sup_{\theta \in \Theta_n} |U_{2,n}| = O_{\mathbb{P}}(p^{3/2} h^{1/2} \ln n/n)$ and the first order U-statistics is now $\sup_{\theta \in \Theta_n} |U_{1,n}| = O_{\mathbb{P}}(p^{3/2}/n^{1/2})$.

The asymptotic normality is established by invoking Lemma 2(i) of Guerre and Lavergne (2005), in fact, using the same technicalities as in Statement (iii) in Theorem 1, the conditions in the aforementioned Lemma are satisfied.

In order to treat the terms $\widehat{Q}_n^B(\widehat{\theta})$ and $\widehat{Q}_n^C(\widehat{\theta})$ it is enough to investigate the behaviour of $\sup_{\theta \in \Theta_n} \widehat{Q}_n^B(\theta)$ and $\sup_{\theta \in \Theta_n} \widehat{Q}_n^C(\theta)$. Without loss of generality, in what follows C denotes any positive constant. Note preliminarily that by assumption E-i.,

$$\mathbb{E}[\|\theta - \theta_\star\| | X_1, \ldots, X_n] \le C \frac{(\ln n)^r}{n^r}.$$

Now consider the term $\widehat{Q}_n^B(\theta)$. Denote $W = (\mathcal{E}, X')'$ and $\kappa_{\theta,h}(W_i, W_j) = \mathcal{E}_i(g_0(\langle X_j, \theta_\star \rangle) - g_0(\langle X_j, \theta \rangle)) K_{ij}^\theta$. As a consequence of assumption M-iv. the class of functions g_0 is Euclidean (see Lemma 2.13 of Pakes and Pollard 1989). By assumption K-i., the kernel K is of bounded variation and by Lemma 22(ii) in Nolan and Pollard (1987) the class of all functions K is Euclidean. Since the product of two classes of Euclidean functions is Euclidean (see Lemma 2.14(ii) of Pakes and Pollard 1989) then $\kappa_{\theta,h}$ is Euclidean.

Apply the centered Hoeffding decomposition to $h\widehat{Q}_n^B(\theta)$ to obtain:

$$h\widehat{Q}_n^B(\theta) = \frac{2}{n} \sum_{i=1}^n \widetilde{\kappa}_1(W_i) + \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j \ne i, j=1}^n (\widetilde{\kappa}_{\theta,h} - \widetilde{\kappa}_1(W_j))$$
$$= U_n \widetilde{\kappa}_1 + U_n \widetilde{\kappa}_{\theta,h},$$

where $U_n \widetilde{\kappa}_1$ is the first order U-process associated with a kernel $\widetilde{\kappa}_1(W_i) = \mathbb{E}[\kappa_{\theta,h}(W_i, W_j)|W_i]$, whereas $U_n \widetilde{\kappa}_{\theta,h}$ is the second order U-process associated with the kernel $\widetilde{\kappa}_{\theta,h} = \kappa_{\theta,h} - \widetilde{\kappa}_1(W_i)$. Moreover, note that $\widetilde{\kappa}_1(W_j) = \mathbb{E}[\kappa_{\theta,h}(W_i, W_j)|W_j] = 0$.

One starts by investigating $U_n \widetilde{\kappa}_{\theta,h}$: from the argument above, $\kappa_{\theta,h}$ is Euclidean and by Lemma 5 in Sherman (1994) the corresponding class of functions by taking the conditional expectation is also Euclidean and conclude that $\widetilde{\kappa}_{\theta,h}$ is Euclidean as a consequence of Lemma 2.14(i) in Pakes and Pollard (1989). By the triangle inequality, one obtains $|\widetilde{\kappa}_{\theta,h}| \le |\kappa_{\theta,h}| + |\widetilde{\kappa}_1|$. First, $|\kappa_{\theta,h}|$ is investigated: by the definition and

the assumption M-iv. and the boundedness of the kernel:

$$\left|\kappa_{\theta,h}\right| \leq C|\mathcal{E}_i|\|\theta - \theta_\star\| \leq C|\mathcal{E}_i|\left(\frac{\ln n}{n}\right)^r.$$

Similarly, for what concerns the second addend one has

$$|\widetilde{\kappa}_1| \leq \mathbb{E}\left[\left|\mathcal{E}_i\big(g_0(\langle X_j, \theta_\star\rangle) - g_0(\langle X_j, \theta\rangle)\big)K_{ij}^\theta\right||W_i\right] \leq C|\mathcal{E}_i|\left(\frac{\ln n}{n}\right)^r. \qquad \text{(A9)}$$

Hence, by putting the results together, $\left|\widetilde{\kappa}_{\theta,h}\right| \leq C|\mathcal{E}_i|(\ln n/n)^r$.

The Main Corollary in Sherman (1994) states that

$$\mathbb{E}\left[\sup_{\theta \in \Theta_n} |nU_n\widetilde{\kappa}_{\theta,h}|\right] \leq C\mathbb{E}\left[\sup_{\theta \in \Theta_n}\left(\frac{1}{2n(2n-1)}\sum_{i=1}^n\sum_{j=1,j\neq i}\widetilde{\kappa}_{\theta,h}^2\right)^\alpha\right]^{1/2}$$

where $0 < \alpha < 1$. Plugging in the right hand side of the inequality the bound obtained for $|\widetilde{\kappa}_{\theta,h}|$, it can be controlled by $C(\ln n/n)^{r\alpha}\mathbb{E}\left[\left(|\mathcal{E}_i|^2\right)^\alpha\right]^{1/2}$. By concavity and Jensen's inequality, $\mathbb{E}\left[\left(|\mathcal{E}_i|^2\right)^\alpha\right] \leq (\mathbb{E}\left[|\mathcal{E}_i|^2\right])^\alpha$, where $\mathbb{E}[|\mathcal{E}_i|^2]$ is bounded. Gathering the previous results one obtains

$$\mathbb{E}\left[\sup_{\theta \in \Theta_n}\left|nU_n\widetilde{\kappa}_{\theta,h}\right|\right] \leq C\left(\frac{\ln n}{n}\right)^{r\alpha}.$$

Finally, by invoking the Markov inequality,

$$\sup_{\theta \in \Theta_n}\left|nh^{-1/2}U_n\widetilde{\kappa}_{\theta,h}\right| = O_{\mathbb{P}}\left(\frac{1}{\sqrt{h}}\left(\frac{\ln n}{n}\right)^{r\alpha}\right). \qquad \text{(A10)}$$

The behaviour $U_n\widetilde{\kappa}_1$ is now investigated. First, by using the fact that $\kappa_{\theta,h}$ is Euclidean, the corresponding class of functions by taking the conditional expectation is also Euclidean by (Sherman 1994, Lemma 5) and hence $\widetilde{\kappa}_1$ is. Then using the bound (A9) and the Main Corollary in Sherman (1994), one achieves

$$\mathbb{E}\left[\sup_{\theta \in \Theta_n} |n^{1/2}U_n\widetilde{\kappa}_1|\right] \leq C\left(\frac{\ln n}{n}\right)^{2r\alpha}\mathbb{E}\left[\left(|\mathcal{E}_i|^2\right)^\alpha\right].$$

Using similar arguments as above, one gets

$$\sup_{\theta \in \Theta_n}\left|nh^{-1/2}U_n\widetilde{\kappa}_1\right| = O_{\mathbb{P}}\left(\frac{1}{\sqrt{h}}\left(\frac{\ln n}{n}\right)^{r\alpha}\right). \qquad \text{(A11)}$$

By putting together (A10) and (A11), deduce that

$$\sup_{\theta \in \Theta_n} \left| nh^{1/2} \widehat{Q}_n^B(\theta) \right| = O_{\mathbb{P}} \left(\sqrt{h} \left(\frac{\ln n}{n} \right)^{r\alpha} \right) = o_{\mathbb{P}}(1).$$

Investigate now the behaviour of the term $\widehat{Q}_n^C(\theta)$. Denote its kernel by

$$\kappa_C = (g_0(\langle X_i, \theta_\star \rangle) - g_0(\langle X_i, \theta \rangle))(g_0(\langle X_j, \theta_\star \rangle) - g_0(\langle X_j, \theta \rangle)) K_{ij}^{\theta},$$

Arguing as above with the boundedness of X S-ii., Lipschitz condition M-iv., deduce that

$$|\kappa_C| \le C \|\theta - \theta_0\|^2 \le C \frac{(\ln n)^{2r}}{n^{2r}}.$$

Recalling that g_0 is Euclidean as well as the kernel K (see Nolan and Pollard, 1987, Lemma 22(ii)), then κ_C is Euclidean (see Pakes and Pollard, 1989 Lemma 2.14(ii)). Therefore, it is possible to appeal to the Main Corollary of Sherman (1994) and, as a consequence of the bound for $|\kappa_C|$

$$\mathbb{E} \left[\sup_{\theta \in \Theta_n} \left| n \widehat{Q}_n^C(\theta) \right| \right] \le C \left(\frac{(\ln n)^{4r}}{n^{4r}} \right)^{\frac{\alpha}{2}}.$$

By Markov inequality

$$\sup_{\theta \in \Theta_n} nh^{1/2} \left| \widehat{Q}_n^C(\theta) - \mathbb{E} \left[\widehat{Q}_n^C(\theta) \right] \right| = O_{\mathbb{P}} \left(h \left(\frac{(\ln n)^{4r}}{n^{4r} h^{1/\alpha}} \right)^{\alpha/2} \right) = o_{\mathbb{P}}(1)$$

Finally

$$\left| \mathbb{E} \left[\widehat{Q}_n^C(\theta) \right] \right| \le C \|\theta - \theta_\star\|^2 \mathbb{E} \left[\frac{1}{h} K_{ij}^{\widehat{\theta}} \right] = O \left(\left(\frac{\ln n}{n} \right)^{2r} \right) = o \left(\frac{p^{3/2} \ln n}{nh^{1/2}} \right).$$

Combining all the rates of convergence, the asymptotic normality follows directly.

In the case when β must be estimated by $\widehat{\beta}$, one can follows similar steps as developed before. In particular $Q_n(\widehat{\theta})$ can be developed as a sum of six terms; besides $\widehat{Q}_n^A$, $\widehat{Q}_n^B$ and $\widehat{Q}_n^C$ one has the following three extra-terms:

$$\widehat{Q}_n^D = \sum_{i=1}^n \sum_{j=1, j\neq i}^n \mathcal{E}_i \left\{ g_0^\beta(\langle \theta, X_j \rangle) - g_0^{\widehat{\beta}}(\langle \theta, X_j \rangle) \right\} K_{ij}^{\widehat{\theta}}$$

$$\widehat{Q}_n^E = \sum_{i=1}^n \sum_{j=1, j\neq i}^n \mathcal{E}_i \left\{ g_0^{\widehat{\beta}}(\langle \theta, X_j \rangle) - g_0^{\widehat{\beta}}(\langle \widehat{\theta}, X_j \rangle) \right\} K_{ij}^{\widehat{\theta}}$$

$$\widehat{Q}_n^F = \sum_{i=1}^n \sum_{j=1, j\neq i}^n \left\{ g_0^\beta(\langle \theta, X_i \rangle) - g_0^{\widehat{\beta}}(\langle \theta, X_i \rangle) \right\} \left\{ g_0^{\widehat{\beta}}(\langle \theta, X_j \rangle) - g_0^{\widehat{\beta}}(\langle \widehat{\theta}, X_j \rangle) \right\} K_{ij}^{\widehat{\theta}}.$$

The assumption S-ii., Lipschitzianity of g_0, the rate of convergence of $\widehat{\beta}$ and the fact that $|g_0^\beta(u) - g_0^{\beta'}(u)| \leq C\|\beta - \beta'\|$, allow to use the Euclidean property on the extra-terms to establish the claim.

A.3: Proof of Theorem 3

Firstly, note that Assumption M-ii. guarantees that the variance $v_n^2(\theta_\star)$ is bounded above and below. Secondly, consider then the following decomposition for $Q_n(\theta_\star)$:

$$
\begin{aligned}
Q_n(\theta_\star) &= \frac{1}{n(n-1)h} \sum_{i=1}^n \sum_{j=1, j\neq i}^n \left(Y_i - g_0^{\beta_\star}(\langle X_i, \theta_\star\rangle)\right)\left(Y_j - g_0^{\beta_\star}(\langle X_j, \theta_\star\rangle)\right) K_{ij}^\theta \\
&= \frac{1}{n(n-1)h} \sum_{i=1}^n \sum_{j=1, j\neq i}^n \mathcal{E}_i \mathcal{E}_j K_{ij}^{\theta_\star} + \\
&\quad + 2\frac{\gamma_n}{n(n-1)h} \sum_{i=1}^n \sum_{j=1, j\neq i}^n \mathcal{E}_i G\left(\langle X_j, \theta_\star\rangle\right) K_{ij}^{\theta_\star} + \\
&\quad + \frac{\gamma_n^2}{n(n-1)h} \sum_{i=1}^n \sum_{j=1, j\neq i}^n G(\langle X_i, \theta_\star\rangle) G\left(\langle X_j, \theta_\star\rangle\right) K_{ij}^{\theta_\star} \\
&= Q_n^a(\theta_\star) + 2Q_n^b(\theta_\star) + Q_n^c(\theta_\star).
\end{aligned}
$$

In what follows, if it is not strictly necessary, the dependency on $\theta_\star$ is dropped, since $\theta_\star$ is fixed.

Since Q_n^a is identical to Q_n and the assumptions are the same as in Theorem 1, one can invoke Major's inequality and proceed with similar steps as in to the proofs of statement (i) in Theorem 1. This leads to the asymptotic normality of $n\sqrt{h}Q_n^a/v_n$ and then $n\sqrt{h}Q_n^a$ is bounded in probability.

Concerning the term Q_n^b, first define a new kernel $\kappa_{ij}^b = \mathcal{E}_i G\left(\langle X_j, \theta_\star\rangle\right) K_{ij}^{\theta_\star}$, with the same argument used for the kernel $\leq_{\theta_\star, h}$ in the proof of Theorem 2, statement (ii), the kernel κ_{ij}^b is Euclidean and, by the bounded variation condition on the functions G and the kernel K, it follows that $\left|\kappa_{ij}^b\right| \leq C|\mathcal{E}_i|$. Hence, the main corollary of Sherman (1994) can be used on the U-statistic $\widetilde{Q}_n^b = hQ_n^b/\gamma_n$ and, together with the concavity property and hence the Jensen's inequality on the error $\mathcal{E}_i$ and by assumption M-i., one obtains $\mathbb{E}\left[\left|n\widetilde{Q}_n^b\right|\right] \leq C$. Finally, by the Markov inequality, $nh^{1/2}\left|Q_n^b\right| = O_\mathbb{P}\left(\gamma_n h^{-1/2}\right)$.

To study the behaviour of the last term $Q_n^c(\theta)$, define the new kernel $\kappa_{ij}^c = G(\langle X_i, \theta\rangle) G\left(\langle X_j, \theta\rangle\right) K_{ij}^\theta$. By the bounded variation condition on the functions G and K, it follows that $\left|\kappa_{ij}^c\right| \leq C$. Since the kernel κ_{ij}^c is Euclidean (see Pakes and Pollard, 1989) Lemma 2.14 (ii), it is therefore possible to utilise the main corollary of Sherman (1994) on the U-statistic $\widetilde{Q}_n^c = hQ_n^c/\gamma_n^2$ to derive $\mathbb{E}\left[\left|n\widetilde{Q}_n^c\right|\right] \leq C$. By the Markov inequality it holds $nh^{1/2}\left|Q_n^c - \mathbb{E}\left[Q_n^c\right]\right| = O_\mathbb{P}\left(\gamma_n^2 h^{-1/2}\right)$.

Let study now a lower bound for $\mathbb{E}[|hQ_n^c|]$. By definition

$$\frac{1}{\gamma_n^2}\mathbb{E}[Q_n^c] = \frac{1}{h}\mathbb{E}\left[G(\langle X_1, \theta_\star\rangle)G(\langle X_2, \theta_\star\rangle)K_{12}^{\theta_\star}\right]$$

$$= \frac{1}{h}\int_{\mathbb{R}^2}G(t_1)G(t_2)K\left(\frac{t_1 - t_2}{h}\right)f_{\theta_\star}(t_1)f_{\theta_\star}(t_2)dt_1dt_2 \tag{A12}$$

Let $u_1 = (t_1 - t_2)/h$ and $u_2 = t_2$, and set $M_{\theta_\star} = Gf_{\theta_\star}$. Since the kernel K has unit integral, G is bounded variation, and by the dominated convergence theorem, then one can treat (A12) as follows

$$\int_{\mathbb{R}^2}M_{\theta_\star}(u_1h + u_2)M_{\theta_\star}(u_2)K(u_1)du_1du_2$$

$$= \int_{\mathbb{R}^2}M_{\theta_\star}^2(u_2)K(u_1)du_1du_2 + \int_{\mathbb{R}^2}\left(M_{\theta_\star}(u_1h + u_2) - M_{\theta_\star}(u_2)\right)M_{\theta_\star}(u_2)K(u_1)du_1du_2$$

$$= \int_{\mathbb{R}}M_{\theta_\star}^2(x_2)dx_2\int_{\mathbb{R}}K(x_1)dx_1 + o(1) = \int_{\mathbb{R}}M_{\theta_\star}^2(x_2)dx_2 + o(1) \geq C + o(1)$$

where $C = \int_{\mathbb{R}}M_{\theta_\star}^2(x_2)dx_2 > 0$. One can conclude that for n large enough,

$$nh^{1/2}\mathbb{E}[|Q_n^c|] \geq C\gamma_n^2h^{1/2}n(1 + o(1)).$$

Consider now the following decomposition:

$$nh^{1/2}Q_n = nh^{1/2}Q_n^a + nh^{1/2}Q_n^b + nh^{1/2}(Q_n^c - \mathbb{E}[Q_n^c]) + nh^{1/2}\mathbb{E}[Q_n^c].$$

Thanks to the rates derived before one gets

$$\left|nh^{1/2}Q_n\right| = \left|nh^{1/2}\mathbb{E}[Q_n^c]\right|\left|O_{\mathbb{P}}\left(\frac{1}{\gamma_n^2nh^{1/2}}\right) + O_{\mathbb{P}}\left(\frac{1}{\gamma_nnh}\right) + O_{\mathbb{P}}\left(\frac{1}{nh}\right) + 1\right|$$

$$= \left|nh^{1/2}\mathbb{E}[Q_n^c]\right|(1 + o_{\mathbb{P}}(1)).$$

Since the lower bound for dominant term diverges as $n \to +\infty$, this concludes the proof.

References

Ait-Saïdi A, Ferraty F, Kassa R, Vieu P (2008) Cross-validated estimations in the single-functional index model. Statistics 42:475–494

Aneiros G, Vieu P (2013) Testing linearity in semi-parametric functional data analysis. Comput Stat 28:413–434

Aneiros G, Cao R, Fraiman R, Vieu P (2019a) Editorial for the special issue on functional data analysis and related topics. J Multivar Anal 170:1–2

Aneiros G, Cao R, Vieu P (2019b) Editorial on the special issue on functional data analysis and related topics. Comput Stat 34:447–450

Aneiros G, Horová I, Hušková M, Vieu P (2022) On functional data analysis and related topics. J Multivar Anal 189:104861

Bücher A, Dette H, Wieczorek G (2011) Testing model assumptions in functional regression models. J Multivar Anal 102:1472–1488

Cuesta-Albertos JA, García-Portugués E, Febrero-Bande M, González-Manteiga W (2019) Goodness-of-fit tests for the functional linear model based on randomly projected empirical processes. Ann Stat 47:439–467

Delsol L (2013) No effect tests in regression on functional variable and some applications to spectrometric studies. Comput Stat 28:1775–1811

Delsol L, Ferraty F, Vieu P (2011) Structural test in regression on functional variables. J Multivar Anal 102:422–447

Ferraty F, Vieu P (2006) Nonparametric functional data analysis. Springer, New York

Ferraty F, Goia A, Salinelli E, Vieu P (2013) Functional projection pursuit regression. TEST 22:293–320

Garcia-Portugues E, Gonzalez-Manteiga W, Febrero-Bande M (2014) A goodness-of-fit test for the functional linear model with scalar response. J Comput Graph Stat 23:761–778

Goia A, Vieu P (2015) A partitioned single functional index model. Comput Stat 30:673–692

Guerre E, Lavergne P (2005) Data-driven rate-optimal specification testing in regression models. Ann Stat 33:840–870

Härdle W, Müller N, Sperlich S, Werwatz A (2004) Nonparametric and semiparametric models. Springer, New York

Horvath L, Kokoszka P (2012) Inference for functional data with applications. Springer, New York

Jiang F, Baek S, Cao J, Ma Y (2020) A functional single-index model. Stat Sin 30:303–324

Lavergne P, Patilea V (2008) Breaking the curse of dimensionality in nonparametric testing. J Econ 143:103–122

Ling N, Vieu P (2021) On semiparametric regression in functional data analysis. Rev Comput Stat 13:e1538

Major P (2006) An estimate on the supremum of a nice class of stochastic integrals and U-statistics. Probab Theory Relat Fields 134:489–537

Mammen E (1993) Bootstrap and wild bootstrap for high dimensional linear models. Ann Stat 21:255–285

Newey WK (1985) Maximum likelihood specification testing and conditional moment tests. Econometrica 53:1047–1070

Nolan D, Pollard D (1987) U-processes: rates of convergence. Ann Stat 15:780–799

Novo S, Aneiros-Pérez G, Vieu P (2019) Automatic and location-adaptive estimation in functional single-index regression. J Nonparametr Stat 31:364–392

Pakes A, Pollard D (1989) Simulation and the asymptotics of optimization estimators. Econometrica 57:1027–1057

Patilea V, Sánchez-Sellero C, Saumard M (2016) Testing the predictor effect on a functional response. J Am Stat Assoc 111:1684–1695

Patilea V, Sánchez-Sellero C, Saumard M (2018) Projection-based nonparametric goodness-of-fit testing with functional covariates (Preprint)

Ramsay JO, Silverman BW (2005) Functional data analysis, 2nd edn. Springer, New York

Shang HL (2020) Estimation of a functional single index model with dependent errors and unknown error density. Commun Stat Simul Comput 49:3111–3133

Sherman RP (1994) Maximal inequalities for degenerate U-processes with applications to optimization estimators. Ann Stat 22:439–459

Zheng X (1996) A consistent test of functional form via nonparametric estimation technique. J Econ 75:263–289

van der Vaart AW, Wellner JA (1996) Weak convergence and empirical processes. Springer, New York