Nonparametric testing for no-effect with functional responses and functional covariates

Valentin Patilea[*]   César Sánchez-Sellero[†]   Matthieu Saumard[‡]

Abstract

This paper studies the problem of nonparametric testing for the no-effect of a random functional covariate on a functional response. That means testing whether the conditional expectation of the response given the covariate is almost surely zero or not without imposing any model relating response and covariate. The response and the covariate take values in possibly different separable Hilbert spaces. Hence the situations with scalar response or covariate will be particular cases. Our test is based on the remark that checking the no-effect of the functional covariate is equivalent to checking the nullity of the conditional expectation of the response given a sufficiently rich set of projections of the covariate. Such projections could be on elements from finite-dimension subspaces of the Hilbert space where the covariate takes values. Then, the idea is to search a finite-dimension element of norm 1 that is, in some sense, the least favorable for the null hypothesis. With at hand such a least favorable direction, it remains to check the nullity of the conditional expectation of the functional response given the scalar product between the covariate and the selected direction. We follow these steps using a nearest neighbors (NN) smoothing approach. As a result, our test statistic is a quadratic form involving univariate NN smoothing and the asymptotic critical values are given by the standard normal law. The test is able to detect nonparametric alternatives, not only linear ones. The responses could be heteroscedastic with conditional variance of unknown form. The law of the covariate does not need to be known. An empirical study with both simulated and real data is reported. The cases of functional response and functional or scalar covariate are considered. Our conclusion is that the test could be easily implemented and performs well in simulations and real data applications.

Keywords: functional data regression, nearest neighbors smoothing, nonparametric testing

MSC2000: Primary 62G10, 62G20 ; Secondary 62G08.

[*]CREST (Ensai) & IRMAR, France; patilea@ensai.fr
[†]Facultad de Matemáticas, Universidad de Santiago de Compostela, Spain; cesar.sanchez@usc.es
[‡]INSA-IRMAR, France; Matthieu.Saumard@insa-rennes.fr

1 Introduction

There has been substantial recent work on the methodology of regression analysis with functional data where predictors, responses, or both of them can be viewed as random functions. Functional data arise in many applications, the monograph of Ramsay and Silverman (2005) provides many compelling examples. In this paper we focus on the case where both the response and the predictor (or covariate) are random elements taking values in a space of functions. The functional linear model is the benchmark approach, see Chiou, Müller and Wang (2004), Yao, Müller and Wang (2005), Gabrys, Horváth and Kokoszka (2010) and the references therein. Recently, alternative nonparametric approaches have been considered; see Ferraty et al. (2011), Lian (2011), Ferraty, Van Keilegom and Vieu (2012).

An important step in the statistical modeling is the goodness-of-fit of the model considered, for instance the functional linear model. To our best knowledge only the papers of Chiou and Müller (2007) and Kokoszka *et al.* (2008) investigate the problem of goodness-of-fit. Chiou and Müller (2007) introduced diagnostics of the functional regression fit using plots of functional principal components scores (FPC) of the response and the covariate. They also used residuals versus fitted values FPC scores plots. (The FPC are the random coefficients in the Karhunen-Loève expansions.) It is easy to understand that such two-dimension plots could not capture all types of effects of the covariate on the response, such for instance the effect of the interactions of the covariate FPC. Kokoszka *et al.* (2008) used the response and covariate FPC scores to build a test statistic with χ^2 distribution under the null hypothesis of no linear effect. Again, by construction, the test of Kokoszka *et al.* cannot detect any nonlinear alternative. When little is known about the structure of the data, it is preferable to allow for flexible, nonparametric, alternatives for the goodness-of-fit test. Moreover, when proceeding to nonparametric estimation of the link between the response and the predictor, one should also check whether the predictor has an effect of the response or not.

Formally, the statistical issue we address in this paper could be formulated as follows. Consider a sample of independent copies $(U_1, X_1), \cdots, (U_n, X_n)$ of (U, X) where U and X takes values in some separable Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$. Without loss of generality we may suppose that U has zero expectation. The problem is to build a statistical test of the hypothesis of no-effect of U on X, that is

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

against the nonparametric alternative $\mathbb{P}[\mathbb{E}\left(U|X\right) = 0] < 1.$* Since $\mathcal{H}_1$ or $\mathcal{H}_2$ could be of finite dimension, for instance the real line, this framework covers all the common situations involving functional data. However, our focus of interest will be on the case functional response and functional covariate.

---

*See for instance Parthasarathy (1967) for the construction of the expectation and conditional expectation of a Hilbert-space valued random variable.

The goodness-of-fit or no-effect against *nonparametric alternatives* has been very little explored in functional data context. In the case of scalar response, Delsol, Ferraty and Vieu (2011) proposed a testing procedure adapted from the approach of Härdle and Mammen (1993). However, their procedure involves smoothing in the functional space and requires quite restrictive conditions which make it difficult to apply to real data situations. Patilea, Sánchez-Sellero and Saumard (2012) and García-Portugués, González-Manteiga and Febrero-Bande (2012) proposed alternative nonparametric goodness-of-fit tests for scalar response and functional covariate using one dimension projections of the covariate. Such projection-based methods are much less restrictive and performs well in applications. To our best knowledge, no nonparametric statistical test of no-effect or goodness-of-fit is available when both the response and the covariate are functional.

Our test is based on the remark that checking the no-effect of the functional covariate is equivalent to checking the nullity of the conditional expectation of the response given a sufficiently rich set of projections of the covariate. Such projections could be on elements of norm 1 from finite-dimension subspaces of the Hilbert space where the covariate takes values. Then, the idea is to search a finite-dimension element of norm 1 that is, in some sense, the least favorable for the null hypothesis. With at hand such a least favorable direction, it remains to check the nullity of the conditional expectation of the functional response given the scalar product between the covariate and the selected direction. Patilea, Sánchez-Sellero and Saumard (2012) used a similar idea with scalar responses. We follow these steps using a nearest neighbors (NN) smoothing approach. As a result, our new test statistic is a quadratic form involving univariate NN smoothing and the asymptotic critical values are given by the standard normal law. When the response is univariate, our statistic is related but different from the one introduced by Patilea, Sánchez-Sellero and Saumard (2012). By construction, the test is able to detect nonparametric alternatives. The responses could be heteroscedastic with conditional variance of unknown form. The law of the covariate does not need to be known.

The paper is organized as follows. In section 2 we introduce the main notation and we derive a fundamental lemma for our approach. This lemma states that checking condition (1.1) is equivalent to checking the nullity of the conditional expectation of U given a sufficiently rich set of projections of X on elements of norm 1 from finite-dimension subspaces of $\mathcal{H}_2$. In section 3 we introduce the test statistic for testing of no-effect of X on U when U is *observed*. Our statistic is a quadratic form, based on *univariate* NN smoothing, that behaves like a standard normal random variable under H_0. We prove that, under mild technical assumptions, the induced test is consistent against *any* type of fixed alternatives and against sequences of directional alternatives approaching the null hypothesis at a suitable rate. The allowed rates are almost the same as those obtained in parametric model checks based on smoothing with *univariate* covariate, see for instance Guerre and Lavergne (2005). Clearly, our test procedure applies also to the case where the sample of U is not observed and has to be estimated, for instance as the residual of a regression. Under suitable regularity conditions ensuring that the sample of U is estimated sufficiently accurate, the test statistic will still have standard normal critical values. To keep this paper at reasonable length, the extension of our methodology to the

case of estimated responses will be investigated elsewhere. In section 4.1 we propose a simple wild bootstrap procedure to approximate the critical values of our test statistic with small samples and we report the results of several simulation experiments. In particular we compare our test with the one proposed by Kokoszka *et al.* (2008). We conclude that the test could be easily implemented and performs well in applications. The proofs are relegated to the appendix.

2   A dimension reduction lemma

In order to simplify the presentation and without loss of generality, hereafter we focus on the case where the Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ are both equal to the space of square-integrable random functions defined on the unit interval.

Let us introduce some notation. For any $p \geq 1$, let $\mathcal{S}^p = \{\gamma \in \mathbb{R}^p : \|\gamma\| = 1\}$ denote the unit hypersphere in $\mathbb{R}^p$. Let $L^2[0,1]$ be the space of the square-integrable real-valued functions defined on the unit interval $\langle \cdot, \cdot \rangle$ denote the inner product in $L^2[0,1]$, that is for any $W_1, W_2 \in L^2[0,1]$

$$\langle W_1, W_2 \rangle = \int_0^1 W_1(t) W_2(t) dt.$$

Let $\| \cdot \|_{L^2}$ be the associated norm. Hereafter $\mathcal{R} = \{\rho_1, \rho_2, \cdots\}$ will be an arbitrarily fixed orthonormal basis of the function space $L^2[0,1]$, that is $\langle \rho_i, \rho_j \rangle = \delta_{ij}$. Then the response and the predictor processes can be expanded into

$$U(t) = \sum_{j=1}^{\infty} u_j \rho_j(t) \quad \text{and} \quad X(t) = \sum_{j=1}^{\infty} x_j \rho_j(t), \tag{2.2}$$

where the random coefficients u_j (resp. x_j) are given by $u_j = \langle U, \rho_j \rangle$ (resp. $x_j = \langle X, \rho_j \rangle$). For a fixed positive integer p and any $W \in L^2[0,1]$, $W^{(p)} \in L^2[0,1]$ will be the projection of X on the subspace generated by the first p elements of the basis $\mathcal{R}$, that is

$$W^{(p)}(t) = \sum_{j=1}^{p} w_j \rho_j(t).$$

By abuse we also identify $W^{(p)}$ with the $p-$dimension random vector $(w_1, \cdots, w_p)$. On the other hand, for any integer $p \geq 1$ and non random vector $\gamma = (\gamma_1, \cdots, \gamma_p) \in \mathbb{R}^p$, we identify γ with $\sum_{j=1}^{p} \gamma_j \rho_j(t) \in L^2[0,1]$ and hence we write $\langle W, \gamma \rangle = \langle W^{(p)}, \gamma \rangle = \sum_{i=1}^{p} x_j \gamma_j$. In the following we will also use $\beta = \sum_{j=1}^{\infty} b_j \rho_j(t)$ to denote a non random element of $L^2[0,1]$.

Our approach relies on the following lemma, an extension of Lemma 2.1 of Lavergne and Patilea (2008) and Theorem 1 in Bierens (1990) to Hilbert space-valued responses and conditioning random variables. For any $\gamma \in \mathcal{S}^p$, let F_γ denote the distribution function (d.f.) of the real-valued variable $\langle X, \gamma \rangle$, that is $F_\gamma(t) = \mathbb{P}(\langle X, \gamma \rangle \leq t)$, $\forall t \in \mathbb{R}$.

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

(A) The following statements are equivalent:

1. $\mathbb{E}(U \mid X) = 0$ *a.s.*

2. $\mathbb{E}\left[\langle U, \mathbb{E}\left(U \mid \langle X, \gamma \rangle\right)\rangle\right] = 0$ *a.s.* $\forall p \geq 1, \forall \gamma \in \mathcal{S}^p$.

3. $\mathbb{E}\left[\langle U, \mathbb{E}\left\{U \mid F_\gamma(\langle X, \gamma \rangle)\right\}\rangle\right] = 0$ *a.s.* $\forall p \geq 1, \forall \gamma \in \mathcal{S}^p$.

(B) Suppose in addition that for any positive real number s,

$$\mathbb{E}(\|U\| \exp\{s\|X\|\}) < \infty. \tag{2.3}$$

If $\mathbb{P}[\mathbb{E}(U \mid X) = 0] < 1$, then there exists a positive integer p_0 such that for any integer $p \geq p_0$, the set

$$\mathcal{A} = \{\gamma \in \mathcal{S}^p : \mathbb{E}(U \mid \langle X, \gamma \rangle) = 0 \ a.s.\} = \{\gamma \in \mathcal{S}^p : \mathbb{E}(U \mid F_\gamma(\langle X, \gamma \rangle)) = 0 \ a.s.\}$$

has Lebesgue measure zero on the unit hypersphere $\mathcal{S}^p$ and is not dense.

Point (A) is a cornerstone for proving the behavior of our test under the null and the alternative hypotheses. Point (B) shows that in applications it will not be difficult to find directions γ able to reveal the failure of the null hypothesis (1.1) since, under the very mild[†] conditions, such directions represent almost all the points on the unit hyperspheres $\mathcal{S}^p$, provided p is sufficiently large.

Let

$$Q(\gamma) = \mathbb{E}[\langle U, \mathbb{E}\{U \mid F_\gamma(\langle X, \gamma \rangle)\}\rangle] \tag{2.4}$$

The following new formulation of H_0 is a direct consequences of Lemma 2.1 above.

Corollary 2.2 *Consider a $L^2[0,1]-$valued random variable U such that $\mathbb{E}\|U\| < \infty$. The following statements are equivalent:*

1. *The null hypothesis (1.1) holds true.*

2. *for any $p \geq 1$ and any set $B_p \subset \mathcal{S}^p$ with strictly positive Lebesgue measure on the unit hypersphere $\mathcal{S}^p$,*

$$\forall p \geq 1, \quad \max_{\gamma \in B_p} Q(\gamma) = 0. \tag{2.5}$$

[†]If X does not satisfy condition (2.3), it suffices to transform X into some variable $W \in L^2[0,1]$ such that the $\sigma-$field generated by W is the same as the one generated by X and the variable W satisfies condition (2.3).

3   Testing the effect of a functional covariate

We introduce a general approach for nonparametric testing the no-effect of a functional covariate X on a functional random variable U based on the characterization (2.5) of the null hypothesis.

3.1   The test statistic

In view of equation (2.5), our goal is to estimate $Q(\gamma)$. With at hand a sample of (U, X), define

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

where $K_h(\cdot) = K(\cdot/h)$, $K(\cdot)$ is a kernel, h the bandwidth, and $F_{\gamma,n}$ is the empirical d.f. of the sample $\langle X_1, \gamma \rangle, \cdots, \langle X_n, \gamma \rangle$.[‡]

The statistic $Q_n(\gamma)$ is related to statistics considered by Fan and Li (1996) and Zheng (1996) for checks of parametric regressions for finite dimension data. See also Patilea, Sánchez-Sellero and Saumard (2012) for the extension of this type of statistics to testing the goodness-of-fit of functional linear model. The statistics considered by all these authors are based on a Nadaraya-Watson regression estimator. Here we use the nearest neighbor (NN) approach of Stute (1984) and hence our new statistic is more in the spirit of the one introduced by Stute and González Manteiga (1996) to test simple linear models with scalar outcome and covariate and homoscedastic error term. Herein we allow for heteroscedasticity of unknown form and hence, in the particular case where U and X are scalar, we extend the framework of Stute and González Manteiga (1996).

The idea of using projections of the covariates was also considered by Lavergne and Patilea (2008); see also Bierens (1990), Cuesta-Albertos *et al.* (2007), Cuesta-Albertos, Fraiman and Ransford (2007). The extension of the scope to functional responses seems to be new.

Under H_0, by the Central Limit Theorem (CLT) for degenerate U−statistics, for fixed p and $\gamma \in \mathcal{S}^p$, $nh^{1/2}Q_n(\gamma)$ has asymptotic centered normal distribution. Here we use the CLT in Theorem 5.1 in de Jong (1987). We will show de Jong CLT still applies and the asymptotic normal distribution is preserved even when p grows at a suitable rate with the sample size. On the other hand, Lemma 2.1-(B) indicates that if p is large enough, the maximum of $Q(\gamma)$ over γ stays away from zero under the alternative hypothesis and this will guarantee consistency against any departure from H_0.

The statistic $Q_n(\gamma)$ is expected to be close to $Q(\gamma)$ uniformly in γ, provided p increases suitably. Then a natural idea would be to build a test statistic using the maximum of $Q_n(\gamma)$ with respect to γ. However, like in the finite dimension covariate case, under H_0

[‡] Ties in the values $\langle X_i, \gamma \rangle$, $1 \le i \le n$, could be broken by comparing indices, that is if $\langle X_i, \gamma \rangle = \langle X_j, \gamma \rangle$, then we define $F_{\gamma,n}(\langle X_i, \gamma \rangle) < F_{\gamma,n}(\langle X_j, \gamma \rangle)$ if $i < j$. However, for simplicity in our assumptions below we will assume that the $\langle X_i, \gamma \rangle$'s have continuous distribution for all γ.

one expects $Q_n(\gamma)$ to converges to zero for any p and γ and thus the objective function of the maximization problem to be flat. Therefore we will choose a direction γ as the least favorable direction for the null hypothesis H_0 obtained from a penalized criterion based on a standardized version of $Q_n(\gamma)$; see also Lavergne and Patilea (2008) and Bierens (1990) for related approaches. More precisely, fix some $\gamma_0 \in L^2[0, 1]$ that could be interpreted as an initial *guess* of an unfavorable direction for H_0. Let b_{0j}, $j \geq 1$, be the coefficients in the expansion of γ_0 in the basis $\mathcal{R}$. For any given $p \geq 1$ such that $\sum_{j=1}^{p} b_{0j}^2 > 0$, let

$$\gamma_0^{(p)} = \frac{(b_{01}, \cdots, b_{0p})}{\|(b_{01}, \cdots, b_{0p})\|} \in \mathcal{S}^p .$$

Let

$$\widehat{v}_n^2(\gamma) = \frac{2}{n(n-1)h} \sum_{j \neq i} \langle U_i, U_j \rangle^2 K_h^2 \left(F_{\gamma,n}(\langle X_i, \gamma \rangle) - F_{\gamma,n}(\langle X_j, \gamma \rangle) \right), \quad \gamma \in \mathcal{S}^p, \qquad (3.1)$$

be an estimate of the variance of $nh^{1/2}Q_n(\cdot)$. Given $B_p \subset \mathcal{S}^p$ with positive Lebesgue measure in $\mathcal{S}^p$ that contains $\gamma_0^{(p)}$, the least favorable direction γ for H_0 is defined by

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

where $\mathbb{I}_A$ is the indicator function of a set A, and α_n, $n \geq 1$ is a sequence of positive real numbers decreasing to zero at an appropriate rate that depends on the rates of h and p and that will be made explicit below. Using a standardized version of $Q_n(\gamma)$ avoids scaling α_n according to the variability of the observations. Let us notice that the maximization used to define $\widehat{\gamma}_n \in B_p \subset \mathcal{S}^p$ is a finite dimension optimization problem. The choice of $\gamma_0^{(p)}$ will be shown to be theoretically irrelevant, it will not affect the asymptotic critical values and the consistency results. However, in practice the choice of $\gamma_0^{(p)}$ could be related to prior information of the practitioner on a class of alternatives. Since $Q_n(\gamma) = Q_n(-\gamma)$ for any $\gamma \in \mathcal{S}^p$, one could restrict the set B_p to a half unit hypersphere like $\{\gamma \in \mathcal{S}^p : \gamma_1 \geq 0\}$. One could restrict B_p even more, and hence to speed optimization algorithms, when some prior information indicates a set of directions that would be able to detect alternatives.

We will prove that with suitable rates of increase for α_n and p and decrease for h, the probability of the event $\{\widehat{\gamma}_n = \gamma_0^{(p)}\}$ tends to 1 under H_0. Hence $Q_n(\widehat{\gamma}_n)/\widehat{v}_n(\widehat{\gamma})$ behaves asymptotically like $Q_n(\gamma_0^{(p)})/\widehat{v}_n(\gamma_0^{(p)})$, even when p grows with the sample size. Therefore the test statistic we consider is

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

We will show that an asymptotic a-level test is given by $\mathbb{I}(T_n \geq z_{1-a})$, where z_{1-a} is the $(1-a)$-th quantile of the standard normal distribution.

3.2 Behavior under the null hypothesis

In order to derive the asymptotic behavior of the statistic T_n under null hypothesis, below we introduce a set of assumptions on the data (Assumption D), and on the kernel and the rates of h and p (Assumption K).

Assumption D

(a) *The random vectors $(U_1, X_1), \dots, (U_n, X_n)$ are independent draws from the random vector $(U, X) \in L^2[0,1] \times L^2[0,1]$ that satisfies $\mathbb{E}\|U\|^8 < \infty$.*

(b) *For any $p \geq 1$ and any $\gamma \in \mathcal{S}^p$, the d.f. F_γ is continuous.*

(c) *$\exists\, \underline{\sigma}^2,\, C_1, C_2 > 0$ and $\nu > 2$ such that:*

    (i) *$0 < \underline{\sigma}^2 \leq \mathbb{E}(\langle U_1, U_2\rangle^2 \mathbf{1}_{\{\langle U_1, U_2\rangle \leq C_1\}} \mid X_1, X_2)$ almost surely;*

    (ii) *$\mathbb{E}\left[\|U\|^\nu \mid X\right] \leq C_2$.*

(d) *For any $p \geq 1$, $\gamma_0^{(p)} \in B_p \subset \mathcal{S}^p$, B_p are open subsets of $\mathcal{S}^p$ and $B_p \times 0_{p'-p} \subset B_{p'}$, $\forall 1 \leq p < p'$ where $0_p \in \mathbb{R}^p$ denotes the null vector of dimension p.*

The continuity assumption required in (b) is a mild assumption that simplifies the NN smoothing. Condition (c) will allow to prove that the variance of the statistics $Q_n(\gamma)$ is bounded away from zero and infinity uniformly with respect to γ. The very mild conditions imposed on B_p simplify the proofs for the consistency. These conditions are satisfied for instance when B_p is a half unit hypersphere.

Assumption K

(a) *The kernel K is a continuous density on real line such that $K(x) = K(-x)$ and $K(\cdot)$ is non increasing on $[0, \infty)$.*

(b) *$h \to 0$ and $nh^2 \to \infty$.*

(c) *$p \geq 1$ increases to infinity with n and there exists a constant $\lambda > 0$ such that $p \ln^{-\lambda} n$ is bounded.*

The first step to derive a test statistic is the study of the behavior of the process $Q_n(\gamma)$, $\gamma \in B_p$, under H_0 when p is allowed to increase with the sample size. The following key lemma is crucially based on a powerful combinatorial result due to Cover (1967) on the number of possible orderings of $\langle X_1, \gamma\rangle, \cdots, \langle X_n, \gamma\rangle$ when γ belongs to the whole hypersphere $\mathcal{S}^p$, and on exponential inequalities for $U-$statistics.

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

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

Moreover, if $\widehat{v}_n^2(\gamma)$ is the estimate defined in equation (3.1),

$$\sup_{\gamma \in B_p \subset \mathcal{S}^p} \{1/\widehat{v}_n^2(\gamma)\} = O_{\mathbb{P}}(1).$$

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

Lemma 3.2 *Under Assumptions D, K, for a positive sequence α_n, $n \geq 1$ such that $\alpha_n p^{-1} \ln^{-1} n \to \infty$,*

$$\mathbb{P}(\widehat{\gamma}_n = \gamma_0^{(p)}) \to 1, \quad \text{under } H_0.$$

The proof of Lemma 3.2 is similar to the proof of Lemma 3.2 in Lavergne and Patilea (2008) and hence will be omitted. The following result shows that the asymptotic critical values of our test statistic are standard normal.

Theorem 3.3 *Under the conditions of Lemma 3.2 and if the hypothesis H_0 in (1.1) holds true, the test statistic T_n converges in law to a standard normal. Consequently, the test given by $\mathbb{I}(T_n \geq z_{1-a})$, with z_a the $(1-a)-$quantile of the standard normal distribution, has asymptotic level a.*

3.3 The behavior under the alternatives

Our test is consistent against the general alternative

$$H_1 : \mathbb{P}[\mathbb{E}(U \mid X) = 0] < 1,$$

that is the probability that the test statistic T_n is larger than any quantile z_{1-a}) tends to one under H_1. This could be rapidly understood from the following simple inequalities:

$$
\begin{aligned}
T_n &= \frac{nh^{1/2}Q_n(\widehat{\gamma}_n)}{\widehat{v}_n(\widehat{\gamma}_n)} \\
&= \max_{\gamma \in B_p} \left\{ nh^{1/2}Q_n(\gamma)/\widehat{v}_n(\gamma) - \alpha_n \mathbb{I}_{\{\gamma \neq \gamma_0^{(p)}\}} \right\} + \alpha_n \mathbb{I}_{\{\widehat{\gamma}_n \neq \gamma_0^{(p)}\}} \\
&\geq \max_{\gamma \in B_p} \frac{nh^{1/2}Q_n(\gamma)}{\widehat{v}_n(\gamma)} - \alpha_n \geq \frac{nh^{1/2}Q_n(\widetilde{\gamma})}{\widehat{v}_n(\widetilde{\gamma})} - \alpha_n, \quad \forall \widetilde{\gamma} \in B_p \subset \mathcal{S}^p, \qquad (3.4)
\end{aligned}
$$

with $\widehat{v}_n(\gamma)$ defined in (3.1). Since $Var(\langle U_1, U_2\rangle \mid X_1, X_2) \geq \underline{\sigma}^2$, it is clear that $1/\widehat{v}_n(\widetilde{\gamma}) = O_{\mathbb{P}}(1)$ for all $\widetilde{\gamma}$. On the other hand, from Lemma 2.1, there exists p_0 and $\widetilde{\gamma} \in B_{p_0}$ such that the expectation of $Q_n(\widetilde{\gamma})$ stays away from zero as the sample size grows to infinity and h decrease to zero. On the other hand, for any $p > p_0$ and any n and h, clearly $\max_{\gamma \in B_p} Q_n(\gamma) \geq Q_n(\widetilde{\gamma})$, because $B_{p_0} \times 0_{p-p_0} \subset B_p$. All these facts show why our test is omnibus, that is consistent against nonparametric alternatives, provided that $p \to \infty$.

To state the consistency result, let $\delta(X)$ be some $L^2[0,1]$-valued function such that $\mathbb{E}[\delta(X)] = 0$ and $0 < \mathbb{E}[\|\delta(X)\|^4] < \infty$, and let r_n, $n \geq 1$ be sequence of real numbers that decrease to zero or $r_n = 1$, $\forall n$. Consider the sequence of alternatives

$$H_{1n}: \ U = U^0 + r_n\delta(X), \quad n \geq 1, \quad \text{with } U^0 \in L^2[0,1], \ \mathbb{E}(U^0 \mid X) = 0.$$

We show below that such directional alternatives can be detected as soon as $r_n^2 n h^{1/2}/\alpha_n$ tends to infinity. This is exactly the condition one would obtain with scalar covariate; see Lavergne and Patilea (2008). However, in the functional data framework, to obtain the convenient standard normal critical values, we need $1/\alpha_n = o(p^{-1}\ln^{-1} n)$. Hence, the rate r_n at which the alternatives H_{1n} tend to the null hypothesis should satisfy $r_n^2 n h^{1/2}/\{p \ln n\} \to \infty$.

Theorem 3.4 *Suppose that*

(a) *Assumption D holds true with U replaced by U^0;*

(b) *Assumption K is satisfied and in addition $nh^4 \to \infty$ and there exists a constant C such that $|K(u) - K(v)| \leq C|u - v|$, $\forall u, v \in \mathbb{R}$;*

(c) *$\alpha_n/\{p \ln n\} \to \infty$ and r_n, $n \geq 1$ is such that $r_n^2 n h^{1/2}/\alpha_n \to \infty$;*

(d) *$\mathbb{E}[\delta(X)] = 0$ and $0 < \mathbb{E}[\|\delta(X)\|^4] < \infty$;*

(e) *there exists p and $\widetilde{\gamma} \in B_p \subset \mathcal{S}^p$ (independent of n) such that $\mathbb{E}[\delta(X) \mid \langle X, \widetilde{\gamma}\rangle] \neq 0$ and $\forall t \in [0,1]$, the Fourier Transform of $\overline{\delta}(t, \cdot) = \mathbb{E}[\delta(X)(t) \mid F_\gamma(\langle X, \gamma\rangle) = \cdot]$ is integrable;*

Then the test based on T_n is consistent against the sequence of alternatives H_{1n}.

The additional Lipschitz condition on the kernel $K(\cdot)$ and the restriction on the bandwidth range in Theorem 3.4-(b) are reasonable technical conditions that greatly reduce the complexity of the proof of the consistency. The existence of a vector $\widetilde{\gamma}$ such that $\mathbb{E}[\delta(X) \mid \langle X, \widetilde{\gamma}\rangle] \neq 0$ is guaranteed by Lemma 2.1-(B). In Theorem 3.4-(c) we impose a convenient mild technical condition on one of such vector.

4   Empirical study

A simulation study was carried out to assess the behavior of the proposed methods under the null and with different types of effects under the alternative. For comparison with the procedure proposed by Kokoszka *et al.* (2008), we considered a sample size $n = 40$. The critical values of our procedure were approximated by a wild bootstrap procedure as described below.

4.1   Bootstrap procedure

The bootstrap sample, denoted by U_i^b , $1 \leq i \leq n$, is obtained as: $U_i^b = Z_i U_i$, $1 \leq i \leq n$, where Z_i, $1 \leq i \leq n$ are independent random variables following the two-points distribution proposed by Mammen (1993), that is, $Z_i = -(\sqrt{5} - 1)/2$ with probability $(\sqrt{5} + 1)/(2\sqrt{5})$ and $Z_i = (\sqrt{5} + 1)/2$ with probability $(\sqrt{5} - 1)/(2\sqrt{5})$.

A bootstrap test statistic is built from a bootstrap sample as was the original test statistic. When this scheme is repeated many times, the bootstrap critical value $z_{1-a,n}^\star$ at level a is the empirical $(1 - a)$−th quantile of the bootstrapped test statistics. This critical value is then compared to the initial test statistic.

4.2   Simulation study

The first situation we considered was a functional linear model, given by

$$U_i(t) = \int_0^1 \psi(s, t) X_i(s) \, ds + \epsilon_i(t), \qquad 1 \leq i \leq n$$

where X_i and ϵ_i are independent Brownian bridges and ψ is square-integrable over $[0, 1) \times [0, 1)$. The kernel ψ was chosen to be $\psi(s, t) = c \cdot \exp(t^2 + s^2)/2$, with $c = 0$ under the null and $c = 0.3$ under the alternative.

The well-known Karhunen-Loeve decomposition of the Brownian bridge provides a good approximation of the covariate function. Thus, the orthonormal basis of eigenfunctions $\mathcal{R} = \{\sqrt{2} \sin(j\pi t) : 0 \leq t \leq 1, j = 1, 2 \ldots\}$ seems a good choice for our test statistic. Different possibilities for the privileged direction $\gamma_0^{(p)}$ were considered. The direction $\gamma_0^{(p)} = (1, 0, \ldots, 0) \in \mathcal{S}^p$ generally provides a powerful test. Here we present the results for an uninformative direction, with the same coefficients in all basic elements. For the penalization we used the value $\alpha_n = 1$, which provides a good trade-off between the privileged direction and the direction maximizing the standardized statistic.

To compute the statistic for each direction, we used the Epanechnikov kernel, $K(x) = (1 - x^2)\mathbb{I}_{\{|x| \leq 1\}}$. A grid of bandwidths was considered in order to explore the effect of the bandwidth on the power of the test.

The number of basic components was $p = 3$. For the optimization in the hypersphere $\mathcal{S}^p$, a grid of 1200 points was used. For each original sample, we used 499 bootstrap

11

samples to compute the critical value. One thousand original samples of size $n = 40$ were generated to approximate the percentages of rejection.

Figure 1 shows the empirical powers obtained for a grid of values of the bandwidth both under the null hypothesis of no-effect and under the functional linear alternative. We observe that the power is not very much affected by the bandwidth around a possibly optimal value. For purposes of comparison, the empirical power of the Kokoszka et al. (2008)'s test is also shown. These authors proposed a test of the functional linear effect, that is, a test specially designed to detect the alternative of a functional linear effect versus the no-effect. Our test provides similar or even better power than the Kokoszka *et al.*'s parametric test in their ideal framework. The level is quite well respected for any of the considered bandwidths.
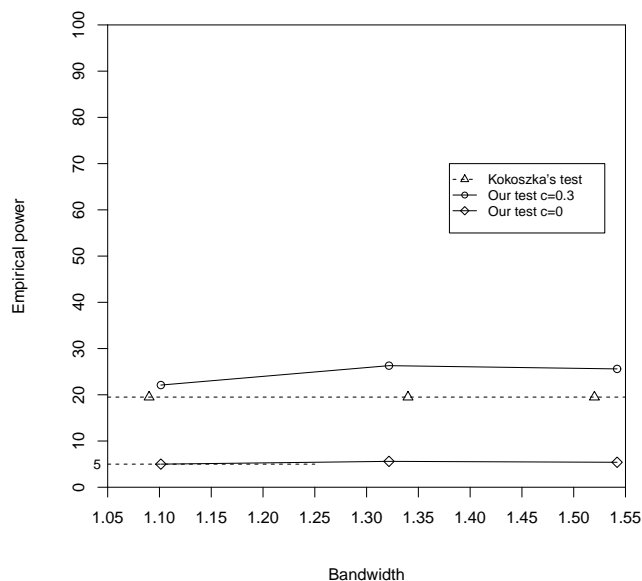

Figure 1: Testing the null-effect versus a functional linear alternative.

Another alternative was considered of the following type:

$$U_i(t) = \beta(t)X_i(t) + \epsilon_i(t), \qquad 1 \leq i \leq n$$

where X_i and ϵ_i are independent Brownian bridges (as in the previous situation) and β is a square-integrable function on $[0, 1]$. This is the so-called concurrent model studied in detail in Ramsay and Silverman (2005), where the covariate at time t, $X_i(t)$, only influences the response function at time t, $U_i(t)$. The function β was $\beta(t) = c \cdot \exp(-4(t - 0.3)^2)$, with $c = 0$ under the null and $c = 0.6$ under the alternative.

Figure 2 shows the power of our test under the concurrent alternative, in comparison with Kokoszka *et al.*'s test. In this case, Kokoszka *et al.*'s test is slightly more powerful

than ours. This is not necessarily surprising since the concurrent model is in a sense a degenerate functional linear model.
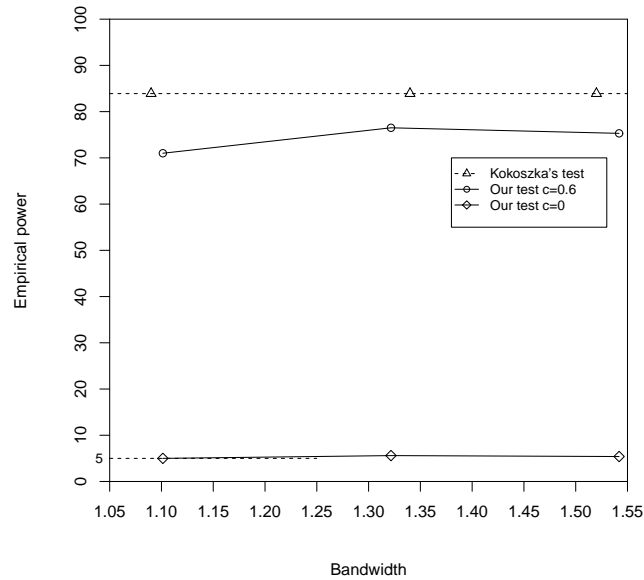

Figure 2: Testing the null-effect versus a concurrent model alternative.

A completely nonlinear alternative was also considered. In this case a quadratic model of this type was generated:

$$U_i(t) = H\left(X_i(t)\right) + \epsilon_i(t), \qquad 1 \le i \le n$$

where X_i and ϵ_i are independent Brownian motion and Brownian bridge, respectively, and $H_2(x) = x^2 - 1$ Since the covariate function is a Brownian motion, instead of the Brownian bridge of the previous situations, the basis was chosen as the orthonormal basis of eigenfunctions of the Brownian motion.

Figure 3 shows the percentages of rejections under the null and under this quadratic alternative for a range of bandwidths. The power of the Kokoszka et $al.$'s test is also plotted. As expected, Kokozska et $al.$'s test, which was designed to detect only linear effects, is not powerful under this quadratic alternative.
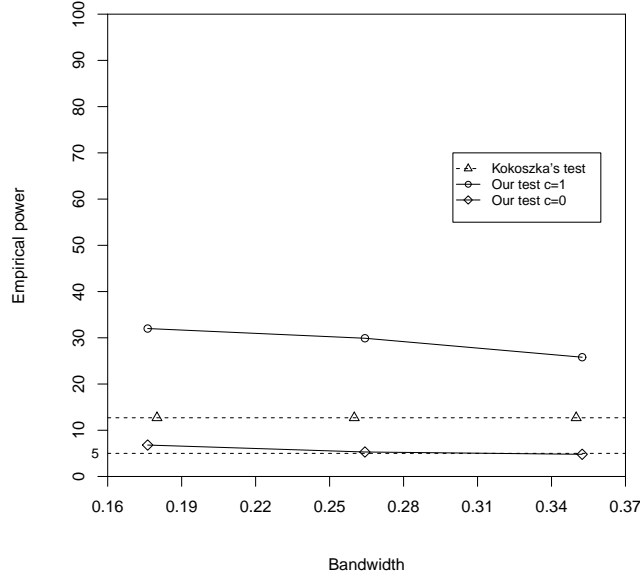
Figure 3: Testing the null-effect versus a quadratic alternative.

5   Appendix: technical proofs

In this section $c, c_1, C, C_1, \ldots$ denote constants that may have different values from line to line. Recall that if $X = \sum_{j=1}^{\infty} x_j \rho_j$, then $X^{(p)} = \sum_{j=1}^{p} x_j \rho_j$.

Proof of Lemma 2.1.   (A) We have

$$
\begin{aligned}
\mathbb{E}(U \mid X) = 0 \quad &\Leftrightarrow \quad \mathbb{E}(\langle U, \rho_j \rangle \mid X) = 0, \ \forall j \geq 1 \\
&\Leftrightarrow \quad \mathbb{E}(\langle U, \rho_j \rangle \mid X^{(p)}) = 0, \ \forall j \geq 1, \forall p \geq 1 \\
&\Leftrightarrow \quad \mathbb{E}(\langle U, \rho_j \rangle \mid \langle X, \gamma \rangle) = 0, \ \forall j \geq 1, \forall p \geq 1, \forall \gamma \in \mathcal{S}^p \\
&\Leftrightarrow \quad \mathbb{E}(U \mid \langle X, \gamma \rangle) = 0, \ \forall p \geq 1, \forall \gamma \in \mathcal{S}^p \\
&\Leftrightarrow \quad \mathbb{E}(U \mid F_\gamma(\langle X, \gamma \rangle)) = 0, \ \forall p \geq 1, \forall \gamma \in \mathcal{S}^p
\end{aligned}
$$

The first and the fourth equivalence in the last display are due to the fact that $\mathcal{R}$ is a basis in $L^2[0,1]$. Next, note that by Cauchy-Schwarz inequality $\forall j, \ \mathbb{E}|\langle U, \rho_j \rangle| \leq \mathbb{E}\|U\| < \infty$. Thus the second equivalence in the last display is guaranteed elementary properties of the conditional expectations and the Doob's Martingale Convergence Theorem, while the third equivalence is given by Lemma 2.1-(A) of Lavergne and Patilea (2008). For the last equivalence recall that for any random variable Y with d.f. F, $\mathbb{P}(F^{-1} \circ F(Y) \neq Y) = 0$ where $F^{-1}(t) = \{y : F(y) \geq t\}, \forall 0 < t < 1$; see for instance Proposition 3, Chapter 1 in Shorack and Wellner (1986). Deduce that $\mathbb{E}(U \mid \langle X, \gamma \rangle) = \mathbb{E}(U \mid F_\gamma(\langle X, \gamma \rangle))$. To

14

complete the proof of part (A) it suffices to note that

$$
\begin{aligned}
\mathbb{E}\left[\langle U, \mathbb{E}\left(U \mid \langle X, \gamma\rangle\right)\rangle\right] &= \mathbb{E}\left[\|\mathbb{E}(U \mid \langle X, \gamma\rangle)\|^2\right] \\
&= \mathbb{E}\left[\|\mathbb{E}(U \mid F_\gamma(\langle X, \gamma\rangle))\|^2\right] \\
&= \mathbb{E}\left[\langle U, \mathbb{E}\{U \mid F_\gamma(\langle X, \gamma\rangle)\}\rangle\right].
\end{aligned}
$$

(B) First note that $\mathcal{A} \subset \bigcap_{j \geq 1} \mathcal{A}_j$ where

$$
\mathcal{A}_j = \{\gamma \in \mathcal{S}^p : \mathbb{E}(\langle U, \rho_j\rangle \mid \langle X, \gamma\rangle) = 0 \text{ a.s. }\}.
$$

Now, if $\mathbb{P}[\mathbb{E}(U \mid X) = 0] < 1$, then there exists $j \geq 1$ such that $\mathbb{P}[\mathbb{E}(\langle U, \rho_j\rangle \mid X) = 0] < 1$. For any arbitrarily fixed $j \geq 1$, Lemma 2.1 in Patilea, Saumard and Sanchez (2012) allows to deduce that there exists $p_0 \geq 1$ such that, for any $p \geq p_0$, $\mathcal{A}_j$ has Lebesgue measure zero on $\mathcal{S}^p$ and is not dense. Since $\mathcal{A}$ is included in any $\mathcal{A}_j$, the conclusion follows. ∎

Lemma 5.1 *Let K be a density satisfying Assumption K-(a) and assume that $h \to 0$ and $nh \to \infty$. Let*

$$
S_{ni} = \frac{1}{(n-1)h} \sum_{1 \leq j \leq n,\, i \neq j} K\left(\frac{i-j}{nh}\right) \quad \text{and} \quad S_n = \frac{1}{n} \sum_{1 \leq i \leq n} S_{ni}.
$$

Then exists constants c_1, c_2 such that for sufficiently large n

$$
0 < c_1 \leq \min_{1 \leq i \leq n} S_{ni} \leq \max_{1 \leq i \leq n} S_{ni} \leq c_2 < \infty.
$$

Moreover, $S_n \to 1$.

Proof. Clearly that $S_n - \widetilde{S}_n \to 0$ where

$$
\widetilde{S}_n = \frac{1}{n^2 h} \sum_{1 \leq i,j \leq n} K\left(\frac{i-j}{nh}\right).
$$

If $[a]$ denote the integer part of any real number a, we can write

$$
\begin{aligned}
\widetilde{S}_n &= \int_{1/n}^{(n+1)/n} \int_{1/n}^{(n+1)/n} h^{-1} K\left(\frac{[ns]-[nt]}{nh}\right) ds\, dt \\
&= \int_{1/n}^{(n+1)/n} \int_{1/nh-t/h}^{1/h+1/nh-t/h} K\left(\frac{[nt+nzh]-[nt]}{nh}\right) dz\, dt \qquad [z=(s-t)/h] \\
&= \int_{1/n}^{(n+1)/n} \int_{1/nh-t/h}^{1/h+1/nh-t/h} K(z)\, dz\, dt + o(1) \\
&= \int_{-1/h}^{1/h} \int_{1/n-zh}^{1+1/n-zh} dt\, K(z)\, dz + o(1) \qquad \qquad [\text{Fubini}] \\
&\to 1,
\end{aligned}
$$

15

where the order $o(1)$ of the reminder on the right-hand side of the third equality could be obtained as a consequence of the fact K is symmetric and monotonic. Hence $S_n \to 1$. Similarly, we can write

$$
\begin{aligned}
\widetilde{S}_{ni} &= \int_{1/n}^{(n+1)/n} h^{-1} K\left(\frac{i - [nt]}{nh}\right) dt \\
&= \int_{(1-i)/nh}^{1/h + (1-i)/nh} K\left(\frac{i - [i + nzh]}{nh}\right) dz \qquad\qquad [z = (t - i/n)/h] \\
&= \int_{(1-i)/nh}^{1/h + (1-i)/nh} K(z)\, dz + o(1).
\end{aligned}
$$

Deduce that

$$
\int_0^1 K(z)dz + \underline{r}_{ni} \leq \widetilde{S}_{ni} \leq \int_{\mathbb{R}} K(z)dz + \overline{r}_{ni}
$$

where $\max_{1 \leq i \leq n}\{|\underline{r}_{ni}| + |\overline{r}_{ni}|\} = o(1)$. The result follows. ∎

One of the ingredients we will use for the proof of Lemma 3.1 is a moment inequality for $U-$statistics presented in Lemma 5.2 below and due to Giné, Latała and Zinn (2000). To state the result we will use, let us introduce some notation. Let $Z_1, \cdots, Z_n$ be independent random variables (not necessarily with the same distribution) taking values in a measurable space $(\mathcal{Z}, \Upsilon)$. Let $h_{i,j}(\cdot, \cdot)$, $1 \leq i, j \leq n$ be real-valued measurable functions on $\mathcal{Z}^2$ such that $h_{i,j}(z_i, z_j) = h_{j,i}(z_j, z_i)$ and $\mathbb{E}[h_{i,j}(z_i, Z_j)] = 0$, $\forall 1 \leq i, j \leq n$, $\forall z_i, z_j$. The functions $h_{i,j}$ could be different for different values of n. Define

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

and

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

The following result is simplified version of Theorem 3.3 in Giné, Latała and Zinn (2000).

Lemma 5.2 *There exist an universal constant $L < \infty$ (in particular, independent on n and the functions $h_{i,j}$) such that*

$$
\mathbb{P}\left\{ \sum_{1 \leq i \neq j \leq n} h_n(Z_i, Z_j) \geq t \right\} \leq L \exp\left[-\frac{1}{L} \min\left(\frac{t^2}{C_n^2}, \frac{t}{D_n}, \frac{t^{2/3}}{B_n^{2/3}}, \frac{t^{1/2}}{A_n^{1/2}} \right) \right], \qquad \forall t > 0.
$$

Let $\gamma \in \mathcal{S}^p$ and let $x_1, \cdots, x_n$ be an arbitrary collection of non-random points in $L^2[0, 1]$. Consider $\widetilde{Z}_1, \cdots, \widetilde{Z}_n$ independent random variables with values in $L^2[0, 1]$ such

that for each $1 \leq i \leq n$ the law of $\widetilde{Z}_i$ is the conditional law of U_i given $X_i = x_i$. We will apply Lemma 5.2 with $h_{i,i} \equiv 0$ and for $1 \leq i \neq j \leq n$

$$h_{i,j}(Z_i, Z_j) = \frac{\langle Z_i, Z_j \rangle}{n(n-1)hM^2} K_h \left(F_{\gamma,n}(\langle x_i, \gamma \rangle) - F_{\gamma,n}(\langle x_j, \gamma \rangle) \right), \tag{5.1}$$

where $Z_i = \widetilde{Z}_i \mathbf{1}_{\{\|\widetilde{Z}_i\| \leq M\}} - \mathbb{E}[\widetilde{Z}_i \mathbf{1}_{\{\|\widetilde{Z}_i\| \leq M\}}]$, $M > 0$ is some constant (that we will allow to increase with n).§ Here $F_{\gamma,n}$ is the empirical d.f. of the sample $\langle x_1, \gamma \rangle, \cdots, \langle x_n, \gamma \rangle$. The functions $h_{i,j}(\cdot, \cdot)$ vanish outside the rectangle $[-2M, 2M] \times [-2M, 2M]$. The following lemma provides upper bounds for the quantities A_n to D_n in this setup. The bounds are independent of the collection $x_1, \cdots, x_n \in L^2[0,1]$, and of $p \geq 1$ and $\gamma \in \mathcal{S}^p$.

Lemma 5.3 *Under the conditions of Lemma 3.1, for $h_{i,j}$ defined as in (5.1)*

$$A_n = \frac{\|K\|_\infty}{n(n-1)h}, \quad B_n^2 \leq \frac{c}{n^3 h M^2}, \quad C_n^2 \leq \frac{c}{n^2 h M^4} \quad and \quad D_n \leq \frac{c}{n M^2},$$

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

Proof. The bound for A_n is obvious. For C_n^2 note that

$$\mathbb{E}[h_{i,j}^2(Z_i, Z_j)] = \frac{M^{-4}}{n^2(n-1)^2 h} \mathbb{E} \left\{ \mathbb{E}[\langle Z_i, Z_j \rangle^2] h^{-1} K_h \left(F_{\gamma,n}(\langle x_i, \gamma \rangle) - F_{\gamma,n}(\langle x_j, \gamma \rangle) \right) \right\}.$$

By Cauchy-Schwarz inequality and triangle inequality and recalling that $\widetilde{Z}_i$ is distributed according to the conditional law of U_i given $X_i = x_i$,

$$\mathbb{E}\left[\langle Z_i, Z_j \rangle^2\right] \leq 16 \mathbb{E}\left[\|\widetilde{Z}_i\|^2\right] \mathbb{E}\left[\|\widetilde{Z}_j\|^2\right] \leq 16 C^2,$$

for any constant C that bounds from above $\mathbb{E}(\|U\|^2 \mid X)$, see Assumption D-(c). Finally, note that

$$\frac{1}{n(n-1)h} \sum_{1 \leq i \neq j \leq n} K_h \left(F_{\gamma,n}(\langle x_i, \gamma \rangle) - F_{\gamma,n}(\langle x_j, \gamma \rangle) \right) = \frac{1}{n(n-1)h} \sum_{1 \leq i \neq j \leq n} K \left(\frac{i-j}{nh} \right)$$

and apply the second part of Lemma 5.1 to derive the bound for C_n^2. To derive the bound for B_n^2 recall that $h_{i,j}(Z_j, z)$ vanishes for $|z| > 2M$, use again Cauchy-Schwarz inequality and triangle inequality and the first part of Lemma 5.1. For the bound of D_n, using Cauchy-Schwarz inequality and the independence of Z_i and Z_j, we can write

$$
\begin{aligned}
\mathbb{E}\sum_{i,j} h_{i,j}(Z_i, Z_j) f_i(Z_i) g_j(Z_j) &\leq \sum_{i,j} \frac{\mathbb{E}|\langle Z_i f_i(Z_i), Z_j g_j(Z_j) \rangle|}{n(n-1)hM^2} K_h(F_{\gamma,n}(\langle x_i, \gamma \rangle) - F_{\gamma,n}(\langle x_j, \gamma \rangle)) \\
&\leq \sum_{i,j} \frac{16 C^2 \mathbb{E}|f_i(Z_i)| \mathbb{E}|g_j(Z_j)|}{n(n-1)hM^2} K_h(F_{\gamma,n}(\langle x_i, \gamma \rangle) - F_{\gamma,n}(\langle x_j, \gamma \rangle)) \\
&\leq \frac{16 C^2}{M^2} \|\|\mathcal{K}\|\|_2,
\end{aligned}
$$

§Note that in particular the $\mathbb{E}[Z_i \mathbf{1}_{\{\|Z_i\| \leq M\}}]$ coincide with the values $\mathbb{E}[U_i \mathbf{1}_{\{\|U_i\| \leq M\}} \mid X_i = x_i]$.

17

where C is such that $\mathbb{E}(\|U\|^2 \mid X) \leq C$ and $\mathcal{K}$ is the matrix with elements

$$\mathcal{K}_{ij} = K\left((i-j)/nh\right)/[n(n-1)h], \quad i \neq j, \quad \text{and} \quad \mathcal{K}_{ii} = 0,$$

and $\||\mathcal{K}\||_2$ is the spectral norm of $\mathcal{K}$. By definition, $\||\mathcal{K}\||_2 = \sup_{u \in \mathbb{R}^n, u \neq 0} \|\mathcal{K}u\|/\|u\|$ and $|u'\mathcal{K}w| \leq \||\mathcal{K}\||_2\|u\|\|w\|$ for any $u, w \in \mathbb{R}^n$. By Lemma 5.1, for any $u \in \mathbb{R}^n$,

$$
\begin{aligned}
\|\mathcal{K}u\|^2 &= \sum_{i=1}^{n}\left(\sum_{j=1,j\neq i}^{n}\frac{K_h\left((i-j)/nh\right)}{h\,n(n-1)}u_j\right)^2 \\
&\leq \sum_{i=1}^{n}\left(\sum_{j=1,j\neq i}^{n}\frac{K_h\left((i-j)/nh\right)}{h\,n(n-1)}\right)\sum_{j=1,j\neq i}^{n}\frac{K_h\left((i-j)/nh\right)}{h\,n(n-1)}u_j^2 \\
&\leq \|u\|^2\left[\max_{1\leq i\leq n}\left(\sum_{j=1,j\neq i}^{n}\frac{K_h\left((i-j)/nh\right)}{h\,n(n-1)}\right)\right]^2 \\
&\leq cn^{-2}\|u\|^2,
\end{aligned}
\tag{5.2}
$$

for some constant $c > 0$. The bound for D_n follows immediately. ∎

Another ingredient is an upper bound for the number of different possible orderings in the sample $\langle X_1, \gamma\rangle, \cdots, \langle X_n, \gamma\rangle$ when γ belongs to the unit hypersphere in $\mathbb{R}^p$ (obviously the same number is obtained if γ is allowed to belong to the whole space $\mathbb{R}^p$). Let $x_1, \cdots, x_n$ a collection of n points in $\mathbb{R}^p$ and let π be a permutation of the set of integers $\{1, 2, \cdots, n\}$. Following Cover (1967), we shall say that $\gamma \in \mathcal{S}^p$ induces the ordering π if

$$\langle x_{\pi(1)}, \gamma\rangle < \langle x_{\pi(2)}, \gamma\rangle < \cdots < \langle x_{\pi(n)}, \gamma\rangle.$$

Conversely, the ordering π will be said to be linearly inducible if there exists such vector γ. The following result is due to Cover (1967).

Lemma 5.4 *There are precisely $q(n, p)$ linearly inducible orderings of n points in general position in $\mathbb{R}^p$, where*

$$q(n,p) = 2\sum_{k=0}^{p-1}S_{n,k} = 2\left[1 + \sum_{2\leq i\leq n-1}i + \sum_{2\leq i<j\leq n-1}ij + \cdots\right] \quad (p \text{ terms}),$$

where $S_{n,k}$ is the number of the $(n-2)!/(n-2-k)!k!$ possible products of numbers taken k at a time without repetition from the set $\{2, 3, \cdots, n-1\}$

By Lemma 5.4 we obtain a simple upper bound for $q(n, p)$ when $n \geq 2p$, that is

$$q(n,p) \leq 2[1 + n^2 + \cdots + n^p] \leq n^{p+1}.\tag{5.3}$$

Proof of Lemma 3.1. Fix M that depends on n in a way that will be specified below. Let

$$Q_{M,n}(\gamma) = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \langle U_{M,i}, U_{M,j} \rangle \frac{1}{h} K_h \left(F_{\gamma,n}(\langle X_i, \gamma \rangle) - F_{\gamma,n}(\langle X_j, \gamma \rangle) \right), \quad \gamma \in \mathcal{S}^p,$$

where $U_{M,i} = U_i \mathbf{1}_{\{\|U_i\| \leq M\}} - \mathbb{E}[U_i \mathbf{1}_{\{\|U_i\| \leq M\}}]$. We can write

$$\mathbb{P}\left(\sup_{\gamma \in \mathcal{S}^p} |Q_{M,n}(\gamma)| > \frac{tp \ln n}{nh^{1/2}} \right) = \mathbb{E}\left[\mathbb{P}\left(\sup_{\gamma \in \mathcal{S}^p} |Q_{M,n}(\gamma)| > \frac{tp \ln n}{nh^{1/2}} \mid X_1, \cdots, X_n \right) \right]$$

In view of Lemma 5.4, for any n, p, given $X_1, \cdots, X_n$ there exists a set $\mathcal{O}_{np} \subset \mathbb{R}^p$ with at most n^p elements, that depend on $X_1, \cdots, X_n$, such that

$$\sup_{\gamma \in \mathcal{S}^p} |Q_{M,n}(\gamma)| = \sup_{\gamma \in \mathcal{O}_{np}} |Q_{M,n}(\gamma)|.$$

Let $b_n = M^{-2} n^{-1} h^{-1/2} p \ln n$. By Lemmas 5.2 and 5.4 deduce that there exists an universal constant L such that for any $t > 0$,

$$\mathbb{P}\left(\sup_{\gamma \in \mathcal{S}^p} |Q_{M,n}(\gamma)| > \frac{tp \ln n}{nh^{1/2}} \mid X_1, \cdots, X_n \right) \leq \sum_{\gamma \in \mathcal{O}_{np}} \mathbb{P}\left(|M^{-2} Q_{M,n}(\gamma)| > tb_n \mid X_1, \cdots, X_n \right)$$

$$\leq \max\{L, 1\} \exp\left[(p+1) \ln n - \frac{1}{L} \min\left(\frac{(tb_n)^2}{C_n^2}, \frac{tb_n}{D_n}, \frac{(tb_n)^{2/3}}{B_n^{2/3}}, \frac{(tb_n)^{1/2}}{A_n^{1/2}} \right) \right].$$

Now, take $M = n^{1/4-a}$ for some (small) $a > 0$ and notice that the exponential bound in the last display is independent of $X_1, \cdots, X_n$ and tends to zero for any t. Deduce that

$$\sup_{\gamma \in \mathcal{S}^p} |Q_{M,n}(\gamma)| = O_{\mathbb{P}}(n^{-1} h^{-1/2} p \ln n)$$

Next we show that $\sup_{\gamma \in \mathcal{S}^p} |Q_n(\gamma) - Q_{M,n}(\gamma)| = o_{\mathbb{P}}(n^{-1} h^{-1/2} p \ln n)$. Let

$$R_{1n}(\gamma) = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \langle U_{M,i}, U_j - U_{M,j} \rangle \frac{1}{h} K_h \left(F_{\gamma,n}(\langle X_i, \gamma \rangle) - F_{\gamma,n}(\langle X_j, \gamma \rangle) \right), \quad \gamma \in \mathcal{S}^p,$$

and $R_{2n}(\gamma) = Q_n(\gamma) - Q_{M,n}(\gamma) - 2R_{1n}(\gamma)$. We have,

$$\mathbb{E} \sup_{\gamma} |R_{1n}(\gamma)| \leq Ch^{-1} \mathbb{E}\left(\|U_{M,i}\| \|U_j - U_{M,j}\| \right) \leq 2Ch^{-1} \mathbb{E}\left(\|U_i\| \right) \mathbb{E}\left(\|U_j - U_{M,j}\| \right).$$

By Hölder inequality and Chebyshev inequality

$$\mathbb{E}\left(\|U_j - U_{M,j}\| \right) \leq 2\mathbb{E}^{1/m}\left[\|U_j\|^m \right] \mathbb{P}^{(m-1)/m}\left[\|U_j\| > M \right] \leq 2\mathbb{E}\left[\|U_j\|^m \right] M^{1-m}.$$

Now, to deduce that $R_{1n}(\gamma)$ is uniformly negligible, it suffices to note that under Assumption K-(b), for $m > 7$ and a sufficiently small,

$$M^{1-m} = n^{(1-m)(1/4-a)} = o\left(n^{-1} h^{1/2} p \ln n \right).$$

19

Clearly, $\sup_{\gamma} |R_{2n}(\gamma)|$ is of smaller order than $\sup_{\gamma} |R_{1n}(\gamma)|$.

For the inverse of the variance estimator, for any $\gamma \in \mathcal{S}^p$ let us define

$$\widehat{v}^2_{N,n}(\gamma) = \frac{2}{n(n-1)h} \sum_{j \neq i} \langle U_i, U_j \rangle^2 \mathbf{1}_{\{\langle U_i, U_j \rangle^2 \leq N\}} K^2_h \left(F_{\gamma,n}(\langle X_i, \gamma \rangle) - F_{\gamma,n}(\langle X_j, \gamma \rangle) \right).$$

Using Hölder inequality, Chebyshev inequality and Cauchy-Schwarz inequality,

$$
\begin{aligned}
\mathbb{E} \sup_{\gamma} \left| \widehat{v}^2_n(\gamma) - \widehat{v}^2_{N,n}(\gamma) \right| &\leq Ch^{-1} \mathbb{E} \left(\langle U_i, U_j \rangle^2 \mathbf{1}_{\{\langle U_i, U_j \rangle^2 > N\}} \right) \\
&\leq h^{-1} \mathbb{E}^{1/s} \left[\langle U_i, U_j \rangle^{2s} \right] \mathbb{P}^{(s-1)/s} \left[\langle U_i, U_j \rangle^{2s} > N^s \right] \\
&\leq h^{-1} \mathbb{E}^2 \left[\|U_j\|^{2s} \right] N^{1-s}.
\end{aligned}
$$

Take $s = 4$, $N = n^{1/4}$ and deduce that the right bound in the last display tends to zero. On the other hand, we apply Hoeffding (1963) inequality for U−statistics to control the deviations of $\widehat{v}^2_{N,n}(\gamma) - \mathbb{E}[\widehat{v}^2_{N,n}(\gamma) \mid X_1, \cdots, X_n]$ conditionally on $X_1, \cdots, X_n$. For any fixed γ we have

$$\mathbb{P} \left(n^{1/2} h |\widehat{v}^2_{N,n}(\gamma) - \mathbb{E}[\widehat{v}^2_{N,n}(\gamma) \mid X_1, \cdots, X_n]| \geq t \mid X_1, \cdots, X_n \right)$$
$$\leq 2 \exp \left\{ -\frac{[n/2]n^{-1}t^2}{2[\tau^2 + K^2(0)Nn^{-1/2}t/3]} \right\}$$

where τ^2 is the variance of a term in the sum defining $h\widehat{v}^2_{N,n}(\gamma) - \mathbb{E}[h\widehat{v}^2_{N,n}(\gamma) \mid X_1, \cdots, X_n]$. Take $t = n^{1/2-c}h$ for some small $c > 0$ and note that $\tau^2 \leq C$ for some constant independent of γ and h. In the similar way we did for $Q_{M,n}(\gamma)$, applying Lemma 5.4, we obtain an exponential bound for the tail of $\widehat{v}^2_{N,n}(\gamma) - \mathbb{E}[\widehat{v}^2_{N,n}(\gamma) \mid X_1, \cdots, X_n]$ given $X_1, \cdots, X_n$ *uniformly* with respect to γ. This bound is independent of $X_1, \cdots, X_n$. Finally integrate out $X_1, \cdots, X_n$ and deduce that

$$\sup_{\gamma} |\widehat{v}^2_{N,n}(\gamma) - \mathbb{E}[\widehat{v}^2_{N,n}(\gamma) \mid X_1, \cdots, X_n]| = o_{\mathbb{P}}(1).$$

It remains to note that Assumption D-(c) and the first part of Lemma 5.1 guarantee that $\mathbb{E}[\widehat{v}^2_{N,n}(\gamma) \mid X_1, \cdots, X_n]$ stays away from zero. Gathering the results we conclude that $1/\widehat{v}^2_n(\gamma)$ is uniformly bounded in probability. Now the proof is complete. ∎

Proof of Theorem 3.3. By Lemma 3.2, if suffices to prove the asymptotic normality of the test statistic T_n defined with $\widehat{\gamma} = \gamma_0^{(p)}$. The proof of this asymptotic normality is based on the Central Limit Theorem 5.1 of de Jong (1987). We will apply the result of de Jong conditionally given the values of the covariate sample. Let $x_1, \cdots, x_n$ be an *arbitrary* collection of non-random points in $L^2[0,1]$. Consider $\widetilde{Z}_1, \cdots, \widetilde{Z}_n$ independent random variables with values in $L^2[0,1]$ such that for each i the law of $\widetilde{Z}_i$ is the conditional law of U_i given $X_i = x_i$. Let $F_{\gamma_0^{(p)},n}(\cdot)$ be the empirical d.f. of the sample $\langle x_1, \gamma_0^{(p)} \rangle, \cdots, \langle x_n, \gamma_0^{(p)} \rangle$,

$$K_{h,ij}(\gamma_0^{(p)}) = K_h \left(F_{\gamma_0^{(p)},n}(\langle x_i, \gamma_0^{(p)} \rangle) - F_{\gamma_0^{(p)},n}(\langle x_j, \gamma_0^{(p)} \rangle) \right)$$

20

and

$$W_{ij} = \frac{1}{n(n-1)}\langle \widetilde{Z}_i, \widetilde{Z}_j \rangle \frac{1}{h} K_{h,ij}(\gamma_0^{(p)}), \quad 1 \le i \ne j \le n, \qquad W_{ii} = 0, \quad 1 \le i \le n.$$

Hence $Q_n(\gamma_0^{(p)}) = \sum_{i,j} W_{ij}$ and $\widehat{v}_n^2(\gamma_0^{(p)}) = 2n(n-1)h \sum_{i,j} W_{ij}^2$. A crucial remark that is used several times in the following is that the elements of the matrix $(K_{h,ij}(\gamma_0^{(p)}))$ are the same as those of matrix $(K_h((i-j)/nh)$ up to permutations of lines and columns. Following the notation of de Jong (1987), let

$$\sigma_{ij}^2 = \mathbb{E}(W_{ij}^2) = \mathbb{E}[\langle U_i, U_j \rangle^2 \mid X_i = x_i, X_j = x_j] \frac{K_{h,ij}^2(\gamma_0^{(p)})}{n^2(n-1)^2 h^2}$$

and $\sigma^2(n) = 2 \sum_{i \ne j} \sigma_{ij}^2$. Since

$$\mathbb{E}[\langle U_i, U_j \rangle^2 \mid X_1, \cdots, X_n] = \mathbb{E}[\langle U_i, U_j \rangle^2 \mid X_i, X_j] \le \mathbb{E}[\|U_i\|^2 \mid X_i]\mathbb{E}[\|U_j\|^2 \mid X_j],$$

and $\mathbb{E}[\langle U_i, U_j \rangle^2 \mid X_i, X_j]$ is bounded away from zero by Assumption D-(c), deduce that there exist positive constants $\underline{c}$ and $\overline{c}$ such that

$$\frac{\underline{c}}{n^4 h^2} K_{h,ij}^2(\gamma_0^{(p)}) \le \sigma_{ij}^2 \le \frac{\overline{c}}{n^4 h^2} K_{h,ij}^2(\gamma_0^{(p)}). \tag{5.4}$$

Apply Lemma 5.1 with K replaced by K^2 and deduce that

$$\frac{c_1}{n^3 h} \le \max_{1 \le i \le n} \sum_{1 \le j \le n, i \ne j} \sigma_{ij}^2 = \max_{1 \le i \le n} \sum_{1 \le j \le n, i \ne j} K_h^2((i-j)/nh) \le \frac{c_2}{n^3 h},$$

for some constants c_1 and c_2. Moreover, there exist constants $\underline{c}'$ and $\overline{c}'$ such that

$$\underline{c}' n^{-2} h^{-1} \le \sigma(n)^2 \le \overline{c}' n^{-2} h^{-1}.$$

It follows that

$$\sigma(n)^{-2} \max_{1 \le i \le n} \sum_{j=1}^{n} \sigma_{ij}^2 = O(n^{-1}),$$

and thus Condition 1 in Theorem 5.1 of de Jong (1987) holds true as soon as $\kappa(n) = o(n^{1/2})$. For checking Condition 2 in Theorem 5.1 of de Jong (1987), let us use Hölder inequality with $p = \nu/2$ and $q = \nu/(\nu-2)$, with ν given by Assumption D-(c)-(ii), and Markov inequality to get, for some constant C,

$$\mathbb{E}[\sigma_{ij}^{-2} W_{ij}^2 \mathbf{1}_{\{\sigma_{ij}^{-1}|W_{ij}|>\kappa(n)\}}] \le \mathbb{E}^{2/\nu}[\sigma_{ij}^{-\nu}|W_{ij}|^\nu]\mathbb{P}^{(\nu-2)/\nu}[\sigma_{ij}^{-1}|W_{ij}| > \kappa(n)] \le C\kappa(n)^{-(\nu-2)/\nu}.$$

That shows that Condition 2 of Theorem 5.1 of de Jong holds true with any $\kappa(n)$ tending to infinity. Finally, let $\mu_1, \cdots, \mu_n$ denote the eigenvalues of the matrix (σ_{ij}). To check

Condition 3 of de Jong, use the upper bound of σ_{ij} in (5.4) to deduce that there exists a constant C (independent on n and i) such that

$$\sum_{j=1,j\neq i}^{n} \sigma_{ij} \leq \frac{C}{n^2 h} \sum_{j=1,j\neq i}^{n} K_{h,ij}(\gamma_0^{(p)}).$$

Next, note that if Σ denotes the $n \times n$ matrix with generic element σ_{ij}, following the lines of equation (5.2) and using equation (5.4), for any $u \in \mathbb{R}^n$,

$$
\begin{aligned}
\|\Sigma u\|^2 &\leq \|u\|^2 \left[\max_{1 \leq i \leq n} \left(\sum_{j=1,j\neq i}^{n} \sigma_{ij} \right) \right]^2 \\
&\leq c_1 \|u\|^2 \left[\max_{1 \leq i \leq n} \left(\sum_{j=1,j\neq i}^{n} \frac{K_h\left((i-j)/nh\right)}{h\,n(n-1)} \right) \right]^2 \\
&\leq c_2 n^{-2} \|u\|^2,
\end{aligned}
\tag{5.5}
$$

for some constants $c_1, c_2 > 0$. Deduce that

$$\sigma(n)^{-2} \max_{1 \leq i \leq n} \mu_i^2 \leq \frac{hn^2}{\underline{c}'} \frac{1}{n^2} \to 0,$$

and thus Condition 3 of de Jong (1987) holds true. To complete the proof of the asymptotic normality of the statistic $T_n = nh^{1/2} Q_n(\gamma_0^{(p)})/\widehat{v}_n(\gamma_0^{(p)})$ given the covariate values, note that

$$\sigma^2(n) = \mathbb{E}[Q_n^2(\gamma_0^{(p)}) \mid X_1 = x_1, \cdots, X_n = x_n] = \frac{\mathbb{E}[\widehat{v}_n^2(\gamma_0^{(p)}) \mid X_1 = x_1, \cdots, X_n = x_n]}{n(n-1)h}.$$

Moreover, by direct standard calculations it can be shown that the variance of

$$\frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \langle \widetilde{Z}_i, \widetilde{Z}_j \rangle^2 \frac{1}{h} K_{h,ij}^2(\gamma_0^{(p)})$$

is of rate $O(h^{-1}n^{-1}) = o(1)$. Deduce that

$$\frac{\widehat{v}_n^2(\gamma_0^{(p)})/n(n-1)h}{\sigma^2(n)} - 1 = o_{\mathbb{P}}(1) \tag{5.6}$$

given $X_1 = x_1, \cdots, X_n = x_n$. The asymptotic normality of T_n given $X_1 = x_1, \cdots, X_n = x_n$ is a consequence of Theorem 5.1 of de Jong and equation (5.6). The proof is complete. ■

Proof of Theorem 3.4. The proof is based on inequality (3.4). Since $\mathbb{E}(\langle U_1, U_2 \rangle^2 \mid X_1, X_2) \geq \underline{\sigma}^2 + r_n^4 \langle \delta(X_1), \delta(X_2) \rangle^2$, clearly the variance estimate $\widehat{v}_n^2(\widetilde{\gamma})$ stays away from

zero for all $\widetilde{\gamma}$. On the other hand, by Cauchy-Schwarz and the property of the spectral norm for matricis,

$$
\begin{aligned}
\widehat{v}_n^2(\widetilde{\gamma}) \;\leq\; & \frac{2n/(n-1)}{n^2 h} \sum_{1 \leq i,j \leq n} \|\delta(X_i)\|^2 \|\delta(X_j)\|^2 K_h^2(F_{n,\widetilde{\gamma}}(\langle X_i, \widetilde{\gamma}\rangle) - F_{n,\widetilde{\gamma}}(\langle X_j, \widetilde{\gamma}\rangle)) \\
\leq\; & \||\mathcal{K}_2\||_2 \sum_{1 \leq i \leq n} \|\delta(X_i)\|^4,
\end{aligned}
$$

where $\mathcal{K}_2$ is the matrix with entries $n^{-2} h^{-1} K_h^2(F_{n,\widetilde{\gamma}}(\langle X_i, \widetilde{\gamma}\rangle) - F_{n,\widetilde{\gamma}}(\langle X_j, \widetilde{\gamma}\rangle))$. By the arguments used in equation (5.5), $\||\mathcal{K}_2\||_2 = O_{\mathbb{P}}(n^{-1})$. This together with the finite fourth order moment condition for $\delta(\cdot)$ imply that $\widehat{v}_n^2(\widetilde{\gamma})$ is bounded in probability. Hence it suffices to look at the behavior of $Q_n(\widetilde{\gamma})$. By Lemma 2.1-(B) there exists p_0 and $\widetilde{\gamma} \in B_{p_0} \subset \mathcal{S}^{p_0}$ (p_0 and $\widetilde{\gamma}$ independent of n) such that $\mathbb{E}[\delta(X) \mid \langle X, \widetilde{\gamma}\rangle] \neq 0$. Hereafter, $\widetilde{\gamma}$ is supposed to have this property. Let $V_i = F_{\widetilde{\gamma}}(\langle X_i, \widetilde{\gamma}\rangle)$, so that $V_1, \cdots, V_n$ are independent uniform variables on $[0, 1]$. Next introduce $V_{ni} = F_{n,\widetilde{\gamma}}(\langle X_i, \widetilde{\gamma}\rangle)$ and

$$
\Delta_n = \sup_{1 \leq i \leq n} |V_{ni} - V_i| \leq \sup_{t \in \mathbb{R}} |F_{n,\widetilde{\gamma}}(t) - F_{\widetilde{\gamma}}(t)|.
$$

Note that for any $s \in \mathbb{R}$,

$$
\left| e^{isV_{ni}} - e^{isV_i} \right| \leq |s||V_{ni} - V_i| \leq |s|\Delta_n, \tag{5.7}
$$

and $\Delta_n = O_{\mathbb{P}}(n^{-1/2})$.

We can write

$$
\begin{aligned}
Q_n(\widetilde{\gamma}) \;=\; & \frac{1}{n(n-1)h} \sum_{i \neq j} \langle U_i^0, U_j^0 \rangle K_h(V_{ni} - V_{nj})) \\
& + \frac{2r_n}{n(n-1)h} \sum_{i \neq j} \langle U_i^0, \delta(X_j) \rangle K_h(V_{ni} - V_{nj}) \\
& + \frac{r_n^2}{n(n-1)h} \sum_{i \neq j} \langle \delta(X_i), \delta(X_j) \rangle K_h(V_{ni} - V_{nj}) \\
\;=:\; & Q_{0n}(\widetilde{\gamma}) + 2r_n Q_{1n}(\widetilde{\gamma}) + r_n^2 Q_{2n}(\widetilde{\gamma}).
\end{aligned}
$$

Since $\widetilde{\gamma}$ is fixed, $Q_{0n}(\widetilde{\gamma}) = O_{\mathbb{P}}(n^{-1} h^{-1/2})$ (cf. proof of Theorem 3.3). Let

$$
Q_{1n}^\star(\widetilde{\gamma}) = \frac{1}{n^2 h} \sum_{1 \leq i \neq j \leq n} \langle U_i^0, \delta(X_j) \rangle K_h(V_i - V_j).
$$

First we show that $Q_{1n}(\widetilde{\gamma}) - Q_{1n}^\star(\widetilde{\gamma}) = o_{\mathbb{P}}(1)$. If K satisfies a Lipschitz condition,

$$
|Q_{1n}(\widetilde{\gamma}) - Q_{1n}^\star(\widetilde{\gamma})| \leq \frac{C\Delta_n}{n^2 h^2} \sum_{1 \leq i \neq j \leq n} \|U_i^0\| \|\delta(X_j)\| = \frac{C\Delta_n}{h^2} O_{\mathbb{P}}(1) = o_{\mathbb{P}}(1).
$$

23

Next, the $U-$statistic $Q_{1n}^{\star}(\widetilde{\gamma})$ can be decomposed into a degenerate $U-$statistic of order 2 with the rate $O_{\mathbb{P}}(h^{-1}n^{-1}) = O_{\mathbb{P}}(n^{-1/2})$ and the sum average of centered variables

$$\frac{1}{n}\sum_{1\leq i\leq n}\langle U_i^0, \mathbb{E}[\delta(X_j)h^{-1}K_h(V_i-V_j)\mid V_i]\rangle.$$

Hence it suffice to bound the variance of the terms in the sum. We can write

$$
\begin{aligned}
v_n^2 &= \mathbb{E}\{\langle U_i^0, \mathbb{E}[\delta(X_j)h^{-1}K_h(V_i-V_j)\mid V_i]\rangle^2\}\\
&\leq \mathbb{E}\{\|U_i^0\|^2\|\mathbb{E}[\delta(X_j)h^{-1}K_h(V_i-V_j)\mid V_i]\|^2\}\\
&\leq c\mathbb{E}\{\|\mathbb{E}[\delta(X_j)h^{-1}K_h(V_i-V_j)\mid V_i]\|^2\},
\end{aligned}
$$

for some constant $c > 0$ larger than $\mathbb{E}\{\|U_i^0\|^2 \mid X_i\}$. Next we show that the map $v \mapsto \mathbb{E}[\delta(X_j)h^{-1}K_h(v-V_j)]$ is squared integrable. Let $\overline{\delta}(t,v) = \mathbb{E}[\delta(X_j)(t)\mid V_j = v]$ and note that $0 < \iint_{[0,1]\times[0,1]}|\overline{\delta}(t,v)|^2 dvdt < \infty$. Then using the Inverse Fourier Transform for K we have for any t

$$
\begin{aligned}
\mathbb{E}[\delta(X_j)(t)h^{-1}K_h(v-V_j)] &= \mathbb{E}\left[\overline{\delta}(t,V_j)\int \exp\{is(v-V_j)\}\mathcal{F}[K](hs)ds\right]\\
&= \int_{\mathbb{R}}\exp\{isv\}\mathcal{F}[\overline{\delta}(t,\cdot)](s)\mathcal{F}[K](hs)ds. \qquad (5.8)
\end{aligned}
$$

Take absolute value and deduce that

$$
\begin{aligned}
\int_{[0,1]\times[0,1]}\mathbb{E}^2[\delta(X_j)(t)h^{-1}K_h(v-V_j)]dvdt &\leq \int_{[0,1]\times[0,1]}\left(\int_{\mathbb{R}}\left|\mathcal{F}[\overline{\delta}(t,\cdot)](s)\right|ds\right)^2 dvdt\\
&\leq \int_{[0,1]\times[0,1]}\int_{\mathbb{R}}\left|\mathcal{F}[\overline{\delta}(t,\cdot)](s)\right|^2 dsdvdt\\
&= \int_{[0,1]\times[0,1]}\int_{\mathbb{R}}\left|\overline{\delta}(t,v)\right|^2 dvdt.
\end{aligned}
$$

Since the V_i's are uniformly distributed, we can deduce that v_n^2 is bounded and thus $Q_{1n}(\widetilde{\gamma}) = O_{\mathbb{P}}(n^{-1/2})$. Now, let

$$Q_{2n}'(\widetilde{\gamma}) = \frac{1}{n^2h}\sum_{1\leq i,j\leq n}\langle\delta(X_i),\delta(X_j)\rangle K_h(V_{ni}-V_{nj}),$$

$$Q_{2n}''(\widetilde{\gamma}) = \frac{1}{n^2h}\sum_{1\leq i,j\leq n}\langle\delta(X_i),\delta(X_j)\rangle K_h(V_i-V_j),$$

$$Q_{2n}^{\star}(\widetilde{\gamma}) = \frac{1}{n^2h}\sum_{1\leq i\neq j\leq n}\langle\delta(X_i),\delta(X_j)\rangle K_h(V_i-V_j).$$

24

It is easy to check that

$$Q'_{2n}(\widetilde{\gamma}) - \frac{n-1}{n}Q_{2n}(\widetilde{\gamma}) = Q''_{2n}(\widetilde{\gamma}) - \frac{n-1}{n}Q^\star_{2n}(\widetilde{\gamma}) = \frac{K(0)}{n^2 h}\sum_{i=1}^{n}\|\delta(X_i)\|^2 = O_{\mathbb{P}}(n^{-1}h^{-1}) = o_{\mathbb{P}}(1).$$

Next we have to show that $Q'_{2n}(\widetilde{\gamma}) - Q''_{2n}(\widetilde{\gamma}) = o_{\mathbb{P}}(1)$. If K satisfies a Lipschitz condition and $nh^4 \to \infty$, by Cauchy-Schwarz inequality, for some constant $C > 0$

$$|Q'_{2n}(\widetilde{\gamma}) - Q''_{2n}(\widetilde{\gamma})| \;\leq\; \frac{C\Delta_n}{h^2}\left[\frac{1}{n}\sum_{1\leq i\leq n}\|\delta(X_i)\|\right]^2 = o_{\mathbb{P}}(1).$$

Conclude that $Q_{2n}(\widetilde{\gamma}) - Q^*_{2n}(\widetilde{\gamma}) = o_{\mathbb{P}}(1)$, so that is suffices to investigate $Q^*_{2n}(\widetilde{\gamma})$. It is easy to show that the variance of $Q^*_{2n}(\widetilde{\gamma})$ tends to zero, so that it remains to show that the expectation of $Q^*_{2n}(\widetilde{\gamma})$ stay away from zero. From the representation (5.8) and repeated applications of Fubini's theorem we get

$$
\begin{aligned}
\mathbb{E}[Q^*_{2n}(\widetilde{\gamma})] &= \mathbb{E}\left[\langle\delta(X_i),\delta(X_j)\rangle h^{-1}K_h(V_i - V_j)\right]\\
&= \mathbb{E}(\langle\delta(X_i),\mathbb{E}[\delta(X_j)h^{-1}K_h(V_i - V_j) \mid X_i]\rangle)\\
&= \int_{[0,1]}\mathbb{E}\left(\delta(X_i)(t)\int_{\mathbb{R}}\exp\{isV_i\}\mathcal{F}[\overline{\delta}(t,\cdot)](-s)\mathcal{F}[K](hs)ds\right)dt\\
&= \int_{[0,1]}\int_{\mathbb{R}}\|\mathcal{F}[\overline{\delta}(t,\cdot)](s)\|^2\mathcal{F}[K](hs)dsdt.
\end{aligned}
$$

By Lebesgue dominated convergence theorem and Plancherel theorem,

$$\mathbb{E}[Q^*_{2n}(\widetilde{\gamma})] \to \int_{[0,1]}\int_{\mathbb{R}}|\overline{\delta}(t,v)|^2 dvdt.$$

Deduce that $\mathbb{P}[c^{-1} \leq Q_{2n}(\widetilde{\gamma}) \leq c] \to 1$ for some constant $c > 0$. Gathering the results conclude that for any $C > 0$, $\mathbb{P}[T_n \geq C] \to 1$. $\blacksquare$

REFERENCES

ANTOCH, J., PRCHAL, L., DE ROSA, M. AND SARDA, P. (2008) Functional linear regression with functional response: application to prediction of electricity consumption. *International Workshop on Functional and Operatorial Statistics 2008 Proceedings, Functional and operatorial statistics, Dabo-Niang and Ferraty (Eds.), Physica-Verlag, Springer.*

BIERENS, H.J. (1990). A consistent conditional moment test of functional form. *Econometrica* **58**, 1443–1458.

CHIOU, J-M., MÜLLER, H-G. (2007). Diagnostics for Functional Regression via Residual Processes. *Comput. Statist. Data Anal.* **15**, 4849–4863.

CHIOU, J-M., MÜLLER, H-G. AND WANG J-L. (2004). Functional response models. *Statist. Sinica* **14**, 659–677.

COVER, T.M. (1967). The number of linearly inducible orderings in $d-$space. *SIAM J. Appl. Math.* **15**, 434–439.

CRAMBES, C., AND MAS, A. (2009). Asymptotics of prediction in the functional linear regression with functional outputs. arXiv:0910.3070v3 [math.ST]

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.

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

CUEVAS, A., FEBRERO, M., FRAIMAN, R. (2002). Linear functional regression: The case of fixed design and functional response. *Canadian J. Statist.* **30**, 285–300.

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

DE JONG, P. (1987). A central limit theorem for generalized quadratic forms. *Probab. Theory Related Fields* **25**, 261–277.

FAN, Y., AND LI, Q. (1996). Consistent model specification tests: omitted variables and semiparametric functional forms. *Econometrica* **64**, 865–890.

FERRATY, F., LAKSACI, A., TADJ, A., AND VIEU, P. (2011). Kernel regression with functional response. *Electron. J. Stat.* **5**, 159–171.

FERRATY, F., VAN KEILEGOM, I., AND VIEU, P. (2012). Regression when both response and predictor are functions. *J. Multivariate Anal.* **109**, 10–28.

FERRATY, F., AND VIEU, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice.* Springer, Berlin.

GABRYS,R., HORVÁTH, L., AND KOKOSZKA, P. (2010). Tests for error correlation in the functional linear model. *J. Amer. Statist. Assoc.* **105**, 1113–1125.

GARCA-PORTUGUÉS, E., GONZÁLEZ-MANTEIGA, W., AND FEBRERO-BANDE, M. (2012) A goodness-of-fit test for the functional linear model with scalar response. arXiv:1205.6167v3 [stat.ME]

GINÉ, E., LATAŁA, R., AND ZINN, J. (2000). Exponential and moment inequalities for $U-$statistics. In *High Dimensional II* 13–38. Progr. Probab. **47**. Birkhäuser, Boston.

GUERRE, E., AND LAVERGNE, P. (2005). Data-driven rate-optimal specification testing in regression models. *Ann. Statist.* **33**, 840–870.

HÄRDLE, W., AND MAMMEN, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21**, 1296-1947.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13–30.

KOKOSZKA, P., MASLOVA, I., SOJKA, J., AND ZHU, L. (2008). Testing for lack of dependence in the functional linear model. *Canadian J. Statist.* **36**, 1–16.

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

LIAN, H. (2011). Convergence of functional $k-$nearest neighbor regression estimate with functional responses. *Electron. J. Statist.* **5**, 31–40.

MAMMEN, E. (1993). Bootstrap and Wild Bootstrap for High Dimensional Linear Models. *Ann. Statist.* **21**, 255–285.

PARTHASARATHY, K.R. (1967). *Probability measures on metric spaces.* A.M.S. New-York.

PATILEA, V., SÁNCHEZ-SELLERO, C., AND SAUMARD, M. (2012). Projection-based nonparametric goodness-of-fit testing with functional covariates. arXiv:1205.5578 [math.ST]

RAMSAY, J., AND SILVERMAN, B.W. (2005). *Functional Data Analysis* (2nd ed.). Springer-Verlag, New York.

SHORACK, G., AND WELLNER, J.A. (1986). *Empirical Processes with Applications in Statistics.* John Wiley & Sons, Inc.

STUTE, W. (1984). Asymptotic normality of nearest neighbor regression function estimates. *Ann. Statist.* **12**, 917–926.

STUTE, W., GONZÁLEZ MANTEIGA, W. (1996). NN goodness-of-fit tests for linear models. *J. Statist. Plann. Inference* **53**, 75–92.

YAO, F., AND MÜLLER, H.G., AND WANG, J-L. (2005). Functional linear regression analysis for longitudinal data. *Ann. Statist.* **33**, 2873–2903.

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