Supplementary Material

S-1 Additional discussion on the practical aspects

We proved in the paper that the choice of $\gamma_0^{(p)}$ is irrelevant for the asymptotic theory. In practice the choice of $\gamma_0^{(p)}$ could be related to prior information on a class of alternatives and thus help the practitioner to build a powerful test against that class. Concerning the set B_p, since $Q_n(\gamma) = Q_n(-\gamma)$ for any $\gamma \in \mathcal{S}^p$, one could restrict the set B_p to a half unit hypersphere like $\{\gamma \in \mathcal{S}^p : \gamma_1 \geq 0\}$. One could restrict B_p even more, and hence to speed up the optimization algorithms, when some prior information indicates a set of directions that would be able to detect alternatives.

Let us next discuss the influence of the choice of bandwidth. In our theory the choice of bandwidth does not appear in the asymptotic approximation of the size of the test. With finite samples, however, the law of $nh^{1/2}Q_n(\gamma)/\widehat{v}_n(\gamma)$ may change significantly with h even for a fixed γ, and hence a size correction is often necessary. We propose to make this correction using the simple wild bootstrap procedure described in the paper. Alternatively, one could look for more elaborate methods, as for instance those in Horowitz and Spokoiny (2001) or Gao and Gijbels (2008). Such theoretical investigations could likely be reproduced in our framework under suitable, though restrictive, assumptions. We argue that the aspects concerning the choice of bandwidth in a functional data framework, are quite challenging and hence deserve a separate investigation to be undertaken in future work.

For the sake of completeness, let us recall some notions related to the functional principal components (FPC), their estimation and the sample based $L^2[0,1]$ basis that could be obtained from them. The covariance operator Γ of X is defined by:

$$(\Gamma v)(t) = \int \sigma(t,s)v(s)ds, \qquad v \in L^2[0,1],$$

where X is supposed to satisfy the condition $\int \mathbb{E}(X^2(t))dt < \infty$ and $\sigma(t,s) = \mathbb{E}[\{X(t) - \mathbb{E}(X(t))\}\{X(s) - \mathbb{E}(X(s))\}]$ is supposed positive definite. Let $\lambda_1 \geq \lambda_2 \geq \cdots$ denote the ordered eigenvalues of Γ and let $\mathcal{R} = \{\psi_1, \psi_2, \cdots\}$ be the corresponding basis of eigenfunctions of Γ that are usually called the functional principal components (FPC). The FPCs represent the orthonormal basis of the Karhunen-Loève decomposition of X and provide optimal low-dimensional representations of X, with respect to the mean-squared error. See, for instance, Ramsay and Silverman (2005). In some cases where the law of X is given, the FPCs are available. However, most of the time this is not the case and the FPCs have to be estimated from the empirical covariance operator

$$(\widehat{\Gamma}v)(t) = \int \widehat{\sigma}(t,s)v(s)ds,$$

where $\widehat{\sigma}(t,s) = n^{-1}\sum_{i=1}^{n}\{X_i(t) - \overline{X}_n(t)\}\{X_i(s) - \overline{X}_n(s)\}]$ and $\overline{X}_n(t) = n^{-1}\sum_{i=1}^{n} X_i(t)$. Let $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \cdots \geq 0$ denote the eigenvalues of $\widehat{\Gamma}$ and let $\widehat{\psi}_1, \widehat{\psi}_2, \cdots$ be the corresponding basis of eigenfunctions, *i.e.,* the estimated FPCs.

S-2 Additional empirical evidence

First, in Figure 1 we plot the curve $\mu(t)$ which is the common curve shape for all individuals in the multiplicative effects model considered in section 4.1 in the paper.

Next, some additional experiments are provided to explore the possible influence of the algorithm on the statistical properties of the proposed test, the effect of the bandwidth and the penalization α_n, and to study the possible approximation by the asymptotic distribution of the test statistic. As in the main part of the manuscript, all the results will be provided on the basis of one thousand original samples.

At the end of this section extended versions of Tables 3, 4 and 6 in the main part of the manuscript are given, as well as the results of an heteroscedastic model.

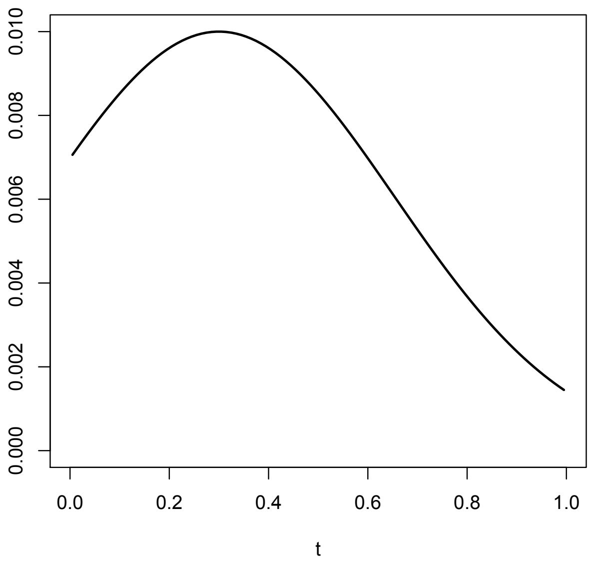

Figure 1: Function $\mu(t) = 0.01 \cdot \exp(-4 \cdot (t - 0.3)^2)$ representing the common curve shape under the multiplicative effects model with one dimension predictor.

S-2.1 The sequential algorithm

In Section 3.5 of the paper, a simplified algorithm for optimization in the hypersphere $\mathcal{S}^p$ is proposed which is based on $(p-1)$ one-dimensional optimizations. We shall show here that the statistical properties of the test with this simplified algorithm are similar to those obtained with a grid in the hypersphere. The same functional linear model studied in Section 4.2 of the paper will be considered. The procedure will be applied with the same values of the parameters used there which yielded the results in Tables 2 and 3. The dimension p is now taken to be 3 and 5. Table S.1 contains the percentages of rejections for the new test with the sequential algorithm and with a full-dimensional optimization algorithm based on a grid in the hypersphere $\mathcal{S}^p$. For the sequential algorithm a grid of 50 points was taken in each one-dimensional optimization. For the full-dimensional optimization, a grid of 1200 points was taken in the hypersphere when $p = 3$, while a grid of 3125 points was taken when $p = 5$. It can be observed that similar results are obtained with the two algorithms, while the one-dimensional algorithm is much less time consuming and it is still feasible for a large dimension p.

			Level = 10%		Level = 5%		Level = 1%	
Hypothesis	p	Algorithm	$n = 40$	$n = 100$	$n = 40$	$n = 100$	$n = 40$	$n = 100$
Null	3	Sequential	9.6	10.7	5.2	5.7	1.0	1.3
		Full-dim.	9.8	11.0	4.9	6.4	1.1	1.4
	5	Sequential	10.1	10.4	5.2	5.9	0.9	1.1
		Full-dim.	10.2	11.5	5.4	6.3	1.2	1.6
Alternative	3	Sequential	53.3	94.8	42.1	89.7	20.1	75.9
		Full-dim.	50.7	93.0	40.5	88.3	20.1	74.7
	5	Sequential	50.4	92.2	38.5	87.6	16.5	71.3
		Full-dim.	54.2	92.3	42.0	88.8	20.4	74.3

Table S.1. Percentages of rejections for the new test with the sequential algorithm and with a full-dimensional grid-based algorithm.

S-2.2 The bandwidth

To show the effect of the bandwidth on the test, the functional linear model studied in Section 4.2 is again considered, and the test is applied with the same values of the parameters used there. We shall fix the dimension p here, to be 3, while the bandwidth will be $h = c_h n^{-2/9}$ for different values of c_h.

Table S.2 contains the percentages of rejections under the null hypothesis and under the alternative, coming from the functional linear effect. The level is respected for all values of the bandwidth, while the power is not much affected by the bandwidth in the wide range of values from $c_h = 0.5$ to $c_h = 1.5$. A possible trend to a higher power for larger bandwidths can be derived within this range, which is explained by the smoothness of the functional linear alternative. On the other hand, more curved alternatives generally require a smaller bandwidth as was observed in similar smoothed testing methods. Either way, the test proposed here does not show a large effect coming from the bandwidth, and it should also be noted that, due to the nearest neighbor methodology, the choice of bandwidth does not depend on the covariate scale, so general rules like $h = c_h n^{-2/9}$, with c_h around o1, are applicable to the common models considered in the literature on lack-of-fit tests.

		Level = 10%		Level = 5%		Level = 1%	
Hypothesis	c_h	$n = 40$	$n = 100$	$n = 40$	$n = 100$	$n = 40$	$n = 100$
Null	0.5	11.8	11.1	6.0	6.2	1.3	1.1
	0.75	11.2	10.7	4.9	5.4	1.1	1.0
	1.0	9.6	10.7	5.2	5.7	1.0	1.3
	1.25	9.2	10.9	5.3	6.0	0.7	1.1
	1.5	9.3	9.9	5.1	5.5	0.8	1.2
Alternative	0.5	45.4	89.4	34.7	84.3	16.1	68.7
	0.75	48.1	92.2	37.4	86.6	18.7	72.6
	1.0	53.3	94.8	42.1	89.7	20.1	75.9
	1.25	54.0	95.8	42.0	91.5	21.0	78.2
	1.5	66.9	97.4	43.1	93.5	20.4	79.6

Table S.2. Percentages of rejections for the new test under the null hypothesis and the alternative, for different values of the bandwidth $h = c_h n^{-2/9}$.

S-2.3 The penalty amplitude α_n

The strength of the penalty is controlled through the value of α_n. This is one parameter that has to be chosen for the proposed test. We can say that there is no optimal choice for this parameter, but the decision will simply be based on the certainty about the best direction to detect the alternative. If the practitioner has a clear intuition that one direction will reveal a deviation from the null hypothesis, then this privileged direction should be protected by means of a large penalization α_n. On the other hand, if there is no clear reference for a privileged direction, then a low penalization would be the most natural option. Regarding the values for the penalization, since the test statistic is asymptotically standard normal, values from 2 to 5 provide a balance between the privileged direction and the direction maximizing the deviation.

In our simulation results, the power (percentages of rejections), obtained for a penalization α_n, takes values from the power obtained with the privileged direction ($\alpha_n = \infty$) to the power obtained with the maximizing direction ($\alpha_n = 0$). The directions we considered are the first eigenvector in the empirical FPC of the covariate, the second eigenvector and an un-informative direction with the same coefficients in all FPC components. The first eigenvector is a direction with very good power, the second has very poor power and the uninformative direction has moderate power, similar to the maximizing direction.

Table S.3 shows the percentages of rejections with the same model studied in Section 4.2 of the paper, under the null hypothesis and under the alternative given by the functional linear model, with the second eigenvector of the empirical FPC of the covariate as privileged direction, and different values of the penalization. All other parameters and configurations of the test are the same as used in Section 4.2 of the paper. We have chosen the worst direction as the privileged direction, in order to show that some values of the penalization protect against a bad choice of the privileged direction. At the same time, this kind of privileged direction allows one to distinguish more clearly the balancing effect of the penalization between the privileged direction and the direction maximizing the statistic.

We observe that, under the null hypothesis, the nominal levels are respected for any value of the penalization. Under the alternative, the value 0 of the penalization provides the power coming from the direction maximizing the statistic, while the value ∞ corresponds to the test based only on the second eigenvector as privileged direction, which is a test with no power at all. It should be noted that the power is preserved with the values in the range from 0 up to 4 or 5, which we recommend. Smaller values, such as 1, 2 or 3, lead to a test based on the maximizing direction, while values from 4 or 5 lead to a test mainly based on the privileged direction.

Model	α_n	Level = 10% $n = 40$	Level = 10% $n = 100$	Level = 5% $n = 40$	Level = 5% $n = 100$	Level = 1% $n = 40$	Level = 1% $n = 100$
Null	0.0	9.9	10.3	4.7	5.0	1.0	1.3
	0.5	9.6	10.3	4.9	5.2	1.1	1.3
	1.0	9.5	10.2	4.9	5.5	1.1	1.4
	1.5	9.2	9.7	4.9	5.5	1.1	1.5
	2.0	9.7	10.0	4.9	5.7	1.1	1.5
	2.5	9.7	10.7	4.7	5.6	1.1	1.3
	3.0	8.5	10.6	4.2	5.3	1.0	1.2
	3.5	8.9	11.5	4.5	5.6	1.1	1.3
	4.0	9.3	11.1	4.9	5.9	1.0	1.2
	4.5	9.3	10.7	4.8	5.6	1.0	1.2
	5.0	9.4	10.8	4.8	6.1	1.1	0.9
	7.0	9.2	10.9	4.7	6.0	0.8	1.1
	10.0	9.2	10.7	4.7	5.8	0.9	0.9
	∞	9.2	10.7	4.7	5.8	0.9	0.9
Alternative	0.0	65.1	98.6	51.9	96.2	23.9	85.4
	0.5	65.5	98.6	52.4	96.2	24.1	85.5
	1.0	66.0	98.6	52.9	96.2	24.4	85.6
	1.5	67.0	98.5	53.3	96.2	24.7	85.8
	2.0	68.2	98.6	54.3	96.4	25.2	86.4
	2.5	66.1	98.6	54.6	96.6	26.0	86.3
	3.0	56.8	96.1	51.5	96.5	26.3	86.6
	3.5	48.2	96.4	43.9	95.3	26.8	87.1
	4.0	39.7	93.3	36.4	92.2	27.2	87.0
	4.5	30.7	89.5	26.9	88.0	22.7	85.4
	5.0	23.6	83.3	19.3	81.4	15.7	79.7
	7.0	10.0	51.0	5.6	47.6	2.1	44.6
	10.0	8.6	21.5	4.0	16.4	0.5	12.6
	∞	8.6	10.9	4.0	5.2	0.5	1.3

Table S.3. Percentages of rejections for the new test under the null hypothesis and under the alternative (functional linear deviation), for different values of the penalization, α_n.

S-2.4 Approximation by the asymptotic distribution

Since the limit distribution of the test has been shown to be standard normal, one could wonder about the possibility of approximating the critical values of the test by the quantiles of the standard normal distribution.

The same model studied in Section 4.2 of the paper will be used under the null hypothesis, in the following experiment. In this way we shall check the accuracy of the standard normal approximation of the test. All the parameters and configurations of the test will be as in Section 4.2. The dimension p will be set to 3, while different values for the penalization α_n will be considered.

The penalization will play a crucial role in the approximation of the test. A large value of the penalization leads to a test based on the projections of the covariate on the privileged direction, while a small value of the penalization leads to a test based on the maximum over the set of directions. Large values of the penalization then lead to smaller values of the test statistic, while small values lead to larger test statistics.

Table S.4 below shows the percentages of rejections for the test, when the standard normal distribution is used for the approximation of the critical values. Different nominal levels, sample sizes and values of the penalization, α_n, are considered. Small values of the penalization (smaller than 2) will generally lead to percentages of rejections higher than the nominal level. Large values of the penalization will lead to percentages of rejections smaller than the nominal level. This fact is due to the negative correlation of the residuals which makes the test statistic based on a single direction negatively biased for small sample sizes. Since the approximation of the standard normal distribution would only be valid for a fine tuning of the penalization, we generally propose to use the bootstrap approximation that was shown to work for all models considered.

	Level = 10%		Level = 5%		Level = 1%	
α_n	$n = 40$	$n = 100$	$n = 40$	$n = 100$	$n = 40$	$n = 100$
0.0	14.6	21.7	8.8	15.3	4.1	6.5
0.5	14.4	21.2	8.8	15.1	4.1	6.4
1.0	13.8	19.8	8.5	14.2	3.8	6.4
1.5	12.4	18.0	8.0	13.6	3.7	6.4
2.0	10.7	15.2	7.6	12.1	3.7	5.9
2.5	8.0	12.2	6.6	10.1	3.4	5.5
3.0	5.5	7.9	5.3	7.1	2.9	4.6
3.5	2.9	5.0	2.8	4.6	2.1	3.7
4.0	2.0	3.3	1.9	2.9	1.4	2.3
4.5	1.2	2.2	1.1	1.7	0.6	1.0
5.0	1.0	1.9	0.9	1.4	0.4	0.7
7.0	0.8	1.7	0.7	1.2	0.2	0.5
∞	0.8	1.7	0.7	1.2	0.2	0.5

Table S.4. Percentages of rejections for the new test under the null hypothesis, with asymptotic distribution approximation, for different values of the penalization, α_n.

S-2.5 Complementary results to Section 4.2

In this section, we present some complementary results to those presented in section 4.2.First we provide extended versions of Tables 3, 4 and 6 in the main part of the manuscript.

p	Test	Level = 10% $n=40$	$n=100$	Level = 5% $n=40$	$n=100$	Level = 1% $n=40$	$n=100$
2	New	9.7	10.6	5.1	5.2	0.5	1.1
	KMSZ	8.0	8.7	2.7	3.7	0.3	0.3
3	New	9.6	10.7	5.2	5.7	1.0	1.3
	KMSZ	7.3	8.3	2.5	3.6	0.2	0.5
5	New	10.1	10.4	5.2	5.9	0.9	1.1
	KMSZ	7.4	8.5	2.6	3.9	0.5	0.8
7	New	10.9	11.2	5.5	5.1	1.2	0.9
	KMSZ	6.1	8.6	2.5	3.5	0.5	1.0
10	New	10.9	10.8	5.6	5.5	1.4	0.9
	KMSZ	7.0	8.9	2.6	4.2	0.3	0.7
15	New	11.2	10.8	7.5	5.8	1.9	1.3
	KMSZ	5.8	8.7	2.1	3.2	0.0	0.0
Random	New	10.5	11.1	4.9	5.6	1.2	0.8
	KMSZ	6.1	8.8	2.5	4.6	0.1	0.5

Table S.5. Percentages of rejections for the new test and Kokoszka *et al.* (2008)'s test under the null hypothesis (extended version of Table 3 in the main manuscript).

p	Test	Level = 10% $n = 40$	$n = 100$	Level = 5% $n = 40$	$n = 100$	Level = 1% $n = 40$	$n = 100$
2	New	62.7	97.5	45.8	94.1	21.6	80.9
	KMSZ	99.9	100.0	99.7	100.0	92.4	100.0
3	New	53.3	94.8	42.1	89.7	20.1	75.9
	KMSZ	99.6	100.0	97.1	100.0	74.3	100.0
5	New	50.4	92.2	38.5	87.6	16.5	71.3
	KMSZ	94.0	100.0	79.1	100.0	37.7	100.0
7	New	47.4	91.1	35.2	85.9	15.7	67.6
	KMSZ	79.6	100.0	59.0	100.0	19.5	100.0
10	New	43.1	86.1	31.5	79.5	13.4	60.0
	KMSZ	61.9	100.0	40.0	100.0	8.5	99.5
15	New	38.0	80.7	27.3	72.0	11.3	48.8
	KMSZ	38.9	99.9	19.8	98.8	1.9	91.1
Random	New	44.5	85.2	33.7	78.1	14.1	59.4
	KMSZ	67.4	100.0	44.7	100.0	12.7	98.4

Table S.6. Percentages of rejections for the new test and Kokoszka *et al.* (2008)'s test under the functional linear effect (extended version of Table 4 in the main manuscript).

p	Test	Level = 10% $n=40$	$n=100$	Level = 5% $n=40$	$n=100$	Level = 1% $n=40$	$n=100$
2	New	57.0	99.4	48.0	99.3	25.3	97.1
	KMSZ	23.7	25.7	14.5	17.3	3.6	6.8
3	New	73.0	99.8	61.0	99.8	26.9	99.3
	KMSZ	34.9	36.8	21.5	27.4	6.8	12.8
4	New	75.5	100.0	58.6	99.9	23.9	99.6
	KMSZ	39.2	44.1	26.6	32.3	8.9	17.1
5	New	71.6	100.0	52.3	100.0	18.6	99.8
	KMSZ	45.1	50.0	28.7	39.6	10.0	20.9
6	New	64.6	100.0	44.0	100.0	14.6	99.7
	KMSZ	39.8	48.0	25.5	36.8	9.2	17.9
7	New	57.2	100.0	36.1	100.0	10.6	99.4
	KMSZ	37.1	45.1	24.6	33.2	6.9	15.9
8	New	50.3	100.0	30.0	100.0	9.0	99.2
	KMSZ	33.3	44.6	20.5	30.8	5.2	13.9
9	New	45.5	100.0	26.6	99.9	7.6	99.0
	KMSZ	31.7	41.5	18.7	28.7	3.9	12.4
10	New	40.3	100.0	23.2	99.8	6.6	98.4
	KMSZ	29.0	39.8	16.1	26.9	3.1	11.1
Random	New	75.9	99.9	61.8	99.8	27.5	99.5
	KMSZ	34.8	39.5	21.9	28.7	7.5	14.3

Table S.7. Percentages of rejections for the new test and Kokoszka *et al.* (2008)'s test under the quadratic high-frequency effect model (extended version of Table 6 in the main manuscript).

Next we report the results obtained for the functional linear model with heteroscedastic error, given by

$$U_i(t) = \int_0^1 \zeta(s,t) X_i(s)\, ds + \sqrt{1/2 + X_i(t)^2}\epsilon_i(t), \qquad 1 \leq i \leq n$$

where X_i and ϵ_i are independent Brownian bridges and the kernel ψ was chosen to be $\zeta(s,t) = c \cdot \exp(t^2 + s^2)/2$, with $c = 0$ under the null and $c = 0.35$ under the alternative. It is the same functional linear model introduced in (4.1), but with heteroscedastic error. The new test and Kokoszka *et al.* (2008)'s test were applied with the same configuration used for the homoscedastic functional linear model.

Table S.8 contains the empirical powers under the null hypothesis for different values of the dimension p, and for a random p with 95% of explained variance, with different significance levels and sample sizes. We found that the new test respects the nominal level, while Kokoszka *et al.* (2008)'s test provides significant deviations from the nominal level. The detected over-rejection of their test is not reduced with increasing sample size. Note that their test assumes homoscedasticity. Table S.9 contains the percentages of rejections obtained under the alternative corresponding to a functional linear effect with heteroscedastic errors. Kokoszka *et al.* (2008)'s test is more powerful than the new test, particularly for small dimensions p and large nominal levels. This could be expected since their test is designed for detecting linear deviations. Some contribution to this higher power may also come from the excess of rejections of their test in this heteroscedastic setup.

		Level = 10%		Level = 5%		Level = 1%	
p	Test	$n = 40$	$n = 100$	$n = 40$	$n = 100$	$n = 40$	$n = 100$
2	New	8.8	10.8	4.5	5.4	0.5	1.3
	KMSZ	20.4	24.1	9.3	12.5	1.5	3.7
3	New	9.6	9.9	5.2	5.4	0.7	1.1
	KMSZ	18.0	21.5	9.3	12.0	1.3	3.3
5	New	9.7	10.4	4.8	4.8	0.7	1.0
	KMSZ	15.3	19.5	7.6	11.3	1.2	3.1
7	New	9.8	9.3	4.6	4.3	0.7	1.2
	KMSZ	13.0	18.3	6.8	9.7	0.9	2.0
10	New	11.0	10.2	5.5	5.0	1.5	0.9
	KMSZ	12.1	16.9	5.1	8.8	0.5	1.7
15	New	11.8	10.6	7.1	5.4	1.9	1.6
	KMSZ	8.7	14.0	2.5	7.4	0.2	0.7
Random	New	9.9	10.2	5.1	5.3	1.4	0.9
	KMSZ	12.0	16.8	5.6	8.6	0.5	1.5

Table S.8. Percentages of rejections for the new test and Kokoszka *et al.* (2008)'s test under the null hypothesis with heteroscedastic error.

p	Test	Level = 10% $n = 40$	$n = 100$	Level = 5% $n = 40$	$n = 100$	Level = 1% $n = 40$	$n = 100$
2	New	43.9	88.4	29.7	78.9	12.5	55.0
	KMSZ	99.8	100.0	99.0	100.0	90.1	100.0
3	New	38.7	81.0	27.9	72.3	9.9	49.5
	KMSZ	99.5	100.0	95.6	100.0	71.0	100.0
5	New	34.8	77.0	25.1	68.5	9.3	44.4
	KMSZ	93.2	100.0	79.4	100.0	37.4	100.0
7	New	34.8	75.0	22.9	65.8	7.8	39.2
	KMSZ	79.9	100.0	58.6	100.0	20.7	100.0
10	New	31.6	69.3	21.0	58.6	7.1	34.2
	KMSZ	61.6	100.0	39.3	100.0	10.5	98.6
15	New	27.2	61.6	18.8	48.6	5.5	27.5
	KMSZ	42.6	99.8	21.5	98.8	2.4	88.4
Random	New	31.9	68.5	22.4	57.8	7.5	32.9
	KMSZ	67.6	100.0	45.8	99.8	14.8	97.3

Table S.9. Percentages of rejections for the new test and Kokoszka *et al.* (2008)'s test under the functional linear effect with heteroscedastic error.

S-3 Technical lemmas and proofs

Below, $c, c_1, C, C_1, ...$ denote constants that may have different values from line to line. Moreover, for any integrable function ϕ defined on the real line, $\mathcal{F}[\phi]$ denotes its Fourier Transform, *i.e.*, $\mathcal{F}[\phi](t) = \int_{\mathbb{R}} \phi(x) \exp\{-2\pi i t x\} dx$.

The following result is due to Cover (1967) and is used in the proof of Lemma 3.3.

Lemma 3.1 *There are precisely $q(n,p)$ linearly inducible orderings of n points in general position in $\mathbb{R}^p$, where*

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

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 3.1 we obtain the simple upper bound for $q(n,p)$ when $n \ge 2p$, *i.e.,*

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

Lemma 3.2 *Let K be a density satisfying Assumption K-(a) and assume that $h \to 0$ and $nh \to \infty$. Let*

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

Constants c_1, c_2 then exist such that for sufficiently large n

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

Moreover, $S_n \to 1$.

Proof of Lemma 3.2. It is clear that $S_n - \widetilde{S}_n \to 0$ where

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

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

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

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

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

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. ■

Proof of Lemma 5.2. The bound for A_n is obvious. For C_n^2 note that

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

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

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

for any constant C that bounds from above $\mathbb{E}(\|U\|^2 \mid X)$, see Assumption D-(c). Finally, recall that K is bounded and 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 3.2 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 $\|z\| > 2M$, again using the Cauchy-Schwarz inequality and the triangle inequality and the first part of Lemma 3.2. For the bound of D_n, using the Cauchy-Schwarz inequality and the independence of Z_i and Z_j, we can write

$$\begin{aligned}
\mathbb{E}\sum_{i,j} h_{i,j}(Z_i,Z_j)f_i(Z_i)g_j(Z_j) &\leq \sum_{i,j}\frac{\mathbb{E}|\langle Z_i f_i(Z_i), Z_j g_j(Z_j)\rangle|}{n(n-1)hM^2}K_h(F_{\gamma,n}(\langle x_i,\gamma\rangle) - F_{\gamma,n}(\langle x_j,\gamma\rangle)) \\
&\leq \sum_{i,j}\frac{4C\mathbb{E}^{1/2}f_i^2(Z_i)\mathbb{E}^{1/2}g_j^2(Z_j)}{n(n-1)hM^2}K_h(F_{\gamma,n}(\langle x_i,\gamma\rangle) - F_{\gamma,n}(\langle x_j,\gamma\rangle)) \\
&\leq \frac{4C}{M^2}|||\mathcal{K}|||_2,
\end{aligned}$$

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

$$\mathcal{K}_{ij} = K\left((i-j)/nh\right)/[n(n-1)h], \quad i\neq j, \quad \text{and} \quad \mathcal{K}_{ii} = 0, \tag{S-3.5}$$

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 3.2, for any $u \in \mathbb{R}^n$,

$$
\begin{aligned}
\|\mathcal{K}u\|^2 &= \sum_{i=1}^n \left(\sum_{j=1, j\neq i}^n \frac{K_h\left((i-j)/nh\right)}{h\, n(n-1)} u_j \right)^2 \\
&\leq \sum_{i=1}^n \left(\sum_{j=1, j\neq i}^n \frac{K_h\left((i-j)/nh\right)}{h\, n(n-1)} \right) \sum_{j=1, j\neq i}^n \frac{K_h\left((i-j)/nh\right)}{h\, n(n-1)} u_j^2 \\
&\leq \|u\|^2 \left[\max_{1\leq i\leq n} \left(\sum_{j=1, j\neq i}^n \frac{K_h\left((i-j)/nh\right)}{h\, n(n-1)} \right) \right]^2 \\
&\leq cn^{-2}\|u\|^2, \qquad \text{(S-3.6)}
\end{aligned}
$$

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

Complements for the proof of Lemma 3.3. For the inverse of the variance estimator, for any $\gamma \in \mathcal{S}^p$, let us define

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

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

$$
\begin{aligned}
\mathbb{E}\left[\sup_\gamma \left| \widehat{v}^2_n(\gamma) - \widehat{v}^2_{N,n}(\gamma) \right| \mid X_1, \cdots, X_n \right] &\leq Ch^{-1}\mathbb{E}\left(\langle U_i, U_j\rangle^2 \mathbb{I}_{\{\langle U_i, U_j\rangle^2 > N\}} \right) \\
&\leq h^{-1}\mathbb{E}^{1/s}\left[\langle U_i, U_j\rangle^{2s} \right] \mathbb{P}^{(s-1)/s}\left[\langle U_i, U_j\rangle^{2s} > N^s \right] \leq h^{-1}\mathbb{E}^2\left[\|U_j\|^{2s} \right] N^{1-s}.
\end{aligned}
$$

Take $s = 4$, $N = n^{1/4}$ and deduce that the right bound in the last display tends to zero. Next, we apply the 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

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

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 as we did for $Q_{M,n}(\gamma)$, applying equation (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 $X_1, \cdots, X_n$. Deduce that

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

conditionally on $X_1, \cdots, X_n$ and unconditionally. It remains to note that Assumption D-(c) and Lemma 3.2 in the Supplementary Material guarantee that $\mathbb{E}[\widehat{v}^2_{N,n}(\gamma) \mid X_1, \cdots, X_n]$ stays away from zero. Conclude that $1/\widehat{v}^2_n(\gamma)$ is uniformly bounded in probability. ■

Proof of Theorem 3.5. By Lemma 3.4, if suffices to prove the asymptotic normality of the test statistic T_n defined with $\widehat{\gamma}_n = \gamma_0^{(p)}$. 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 distribution function of the sample $\langle x_1, \gamma_0^{(p)}\rangle, \cdots, \langle x_n, \gamma_0^{(p)}\rangle$,

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

and

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

Hence $Q_n(\gamma_0^{(p)}) = \sum_{i,j} W_{ij}$ and $\widehat{v}^2_n(\gamma_0^{(p)}) = 2n(n-1)h\sum_{i,j} W^2_{ij}$. 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 the matrix $(K_h((i-j)/nh)$ up to permutations of lines and columns. Following the notation of de Jong (1987), let

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

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

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

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 positive constants $\underline{c}$ and $\overline{c}$ exist such that

$$\frac{\underline{c}}{n^4h^2} K_{h,ij}^2(\gamma_0^{(p)}) \leq \sigma_{ij}^2 \leq \frac{\overline{c}}{n^4h^2} K_{h,ij}^2(\gamma_0^{(p)}). \tag{S-3.7}$$

Applying Lemma 3.2 with K replaced by K^2, one can deduce that for each i,

$$\begin{aligned}\frac{c_1}{n^3h} \leq \frac{\underline{c}}{n^4h^2} \min_{1\leq i\leq n} \sum_{j\neq i} K_h^2((i-j)/nh) \quad &\leq \quad \sum_{1\leq j\leq n, i\neq j} \sigma_{ij}^2 \\ &\leq \quad \frac{\overline{c}}{n^4h^2} \max_{1\leq i\leq n} \sum_{j\neq i} K_h^2((i-j)/nh) \leq \frac{c_2}{n^3h},\end{aligned} \tag{S-3.8}$$

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

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

It follows that

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

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

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

This shows that Condition 2 of Theorem 5.1 of de Jong holds true for 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, we use the upper bound of σ_{ij} in (S-3.7) to deduce that there exists a constant C (independent of 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, it should be noted that if Σ denotes the $n \times n$ matrix with generic element σ_{ij}, following the lines of equation (S-3.6) and using equation (S-3.7), for any $u \in \mathbb{R}^n$,

$$\begin{aligned} \|\Sigma u\|^2 &\leq \|u\|^2 \left[\max_{1\leq i\leq n} \left(\sum_{j=1, j\neq i}^{n} \sigma_{ij} \right) \right]^2 \\ &\leq c_1 \|u\|^2 \left[\max_{1\leq i\leq n} \left(\sum_{j=1, j\neq i}^{n} \frac{K_h\left((i-j)/nh\right)}{h\, n(n-1)} \right) \right]^2 \leq c_2 n^{-2} \|u\|^2, \qquad \text{(S-3.9)} \end{aligned}$$

for some constants $c_1, c_2 > 0$. It can be deduced that

$$\sigma(n)^{-2} \max_{1\leq i\leq n} \mu_i^2 \leq \frac{hn^2}{\underline{c}'} \frac{c_2}{n^2} \to 0,$$

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, it should be noted 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 calculation, it can be shown that the variance of

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

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

$$\frac{\widehat{v}_n^2(\gamma_0^{(p)})/n(n-1)h}{\sigma^2(n)} - 1 = o_{\mathbb{P}}(1) \tag{S-3.10}$$

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 (S-3.10). The proof is complete. ■

Proof of Theorem 3.6. The proof is based on inequality (3.6). Since $\mathbb{E}(\langle U_1, U_2 \rangle^2 \mid X_1, X_2) \ge \underline{\sigma}^2 + r_n^4 \langle \delta(X_1), \delta(X_2) \rangle^2$, clearly the variance estimate $\widehat{v}_n^2(\widetilde{\gamma})$ does not approach zero for all $\widetilde{\gamma}$. On the other hand, by Cauchy-Schwarz and the property of the spectral norm for matrices,

$$\begin{aligned} \widehat{v}_n^2(\widetilde{\gamma}) &\le \frac{2n/(n-1)}{n^2 h} \sum_{1 \le i,j \le n} \|\delta(X_i)\|^2 \|\delta(X_j)\|^2 K_h^2(F_{n,\widetilde{\gamma}}(\langle X_i, \widetilde{\gamma} \rangle) - F_{n,\widetilde{\gamma}}(\langle X_j, \widetilde{\gamma} \rangle)) \\ &\le |||\mathcal{K}_2|||_2 \sum_{1 \le i \le n} \|\delta(X_i)\|^4, \end{aligned} \tag{S-3.11}$$

where $\mathcal{K}_2$ is the matrix with entries $n^{-2}h^{-1}K_h^2(F_{n,\widetilde{\gamma}}(\langle X_i, \widetilde{\gamma} \rangle) - F_{n,\widetilde{\gamma}}(\langle X_j, \widetilde{\gamma} \rangle))$. From the arguments used in equation (S-3.9), $|||\mathcal{K}_2|||_2 = O_{\mathbb{P}}(n^{-1})$. 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 examine at the behavior of $Q_n(\widetilde{\gamma})$. By Lemma 3.1-(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 $V_{ni} = F_{n,\widetilde{\gamma}}(\langle X_i, \widetilde{\gamma}\rangle)$. We can write

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

Since $\widetilde{\gamma}$ is fixed, $Q_{0n}(\widetilde{\gamma}) = O_{\mathbb{P}}(n^{-1}h^{-1/2})$ (cf. proof of Theorem 3.5). Next, let us follow Guerre and Lavergne (2005), and denote by $\mathbb{E}_n$ the conditional expectation given $X_1, \cdots, X_n$ and define

$$
\overline{\delta_n}(X_i) = \frac{1}{n(n-1)h} \sum_{j=1,\, j\neq i}^{n} \delta(X_j) K_h(V_{ni} - V_{nj}), \qquad \overline{\delta} = (\delta(X_1), \cdots, \delta(X_n))'.
$$

Then the Marcinkiewicz-Zygmund inequality and the Cauchy-Schwarz and Jensen inequalities imply that

$$
\begin{aligned}
\mathbb{E}_n \left| \sum_{i=1}^{n} \langle U_i^0, \overline{\delta_n}(X_i)\rangle \right| &\leq c\, \mathbb{E}_n \left| \sum_{i=1}^{n} \left| \langle U_i^0, \overline{\delta_n}(X_i)\rangle \right|^2 \right|^{1/2} \leq c\, \mathbb{E}_n \left| \sum_{i=1}^{n} \|U_i^0\|^2 \|\overline{\delta_n}(X_i)\|^2 \right|^{1/2} \\
&\leq c \left\{ \sum_{i=1}^{n} \mathbb{E}_n \left(\|U_i^0\|^2 \right) \|\overline{\delta_n}(X_i)\|^2 \right\}^{1/2} \leq c\, C_2^{1/\nu} \left\{ \sum_{i=1}^{n} \|\overline{\delta_n}(X_i)\|^2 \right\}^{1/2} \\
&= c\, C_2^{1/\nu} \|\mathcal{K}_3 \overline{\delta}\| \leq c\, C_2^{1/\nu} n^{1/2} |||\mathcal{K}_3|||_2 \left\{ \frac{1}{n} \sum_{i=1}^{n} \|\delta(X_i)\|^2 \right\}^{1/2},
\end{aligned}
$$

for $\mathcal{K}_3$ a matrix with the same elements as the matrix $\mathcal{K}$ defined in equation (S-3.5) up to permutations of lines and columns, and C_2 and ν the constants in Assumption D, and c some constant independent of n. Since $|||\mathcal{K}|||_2 = |||\mathcal{K}_3|||_2 = O_{\mathbb{P}}(n^{-1})$, one can deduce that $Q_{1n}(\widetilde{\gamma}) = O_{\mathbb{P}}(n^{-1/2})$ conditionally on $X_1, \cdots, X_n$. Now, let us investigate $Q_{2n}(\widetilde{\gamma})$. With an inequality such as in equation (S-3.11) and the moment conditions on $\delta(\cdot)$ it is easy to bound $Q_{2n}(\widetilde{\gamma})$ in probability. It remains to show that it is bounded away from zero. Let $V_i = F_{\widetilde{\gamma}}(\langle X_i, \widetilde{\gamma}\rangle)$, so that $V_1, \cdots, V_n$ are independent uniform variables on $[0,1]$, and

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

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

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

We have

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

Next we 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 the Cauchy-Schwarz inequality, for some constant $C > 0$

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

where $\Delta_n = \sup_{1\leq i\leq n} |V_{ni} - V_i|$. Note that $\Delta_n \leq \sup_{t\in\mathbb{R}} |F_{n,\widetilde{\gamma}}(t) - F_{\widetilde{\gamma}}(t)| = O_{\mathbb{P}}(n^{-1/2})$. One can 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 $Q^{*}_{2n}(\widetilde{\gamma})$ does not approach zero. Let $\overline{\delta}(t,v) = \mathbb{E}[\delta(X_j)(t) \mid V_j = v]$ and note that $0 < \iint_{[0,1]\times[0,1]} |\overline{\delta}(t,v)|^2 dvdt < \infty$. By the Inverse Fourier Transform formula and repeated applications of Fubini's theorem we get

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

When $h \to 0$, by the Lebesgue dominated convergence theorem and the Plancherel theorem applied to the integral inside the square brackets,

$$
\mathbb{E}[Q_{2n}^{*}(\widetilde{\gamma})] \to \int_{[0,1]} \int_{[0,1]} |\overline{\delta}(t,v)|^2 dv dt.
$$

One can deduce that $\mathbb{P}[c^{-1} \leq Q_{2n}(\widetilde{\gamma}) \leq c] \to 1$ for some constant $c > 0$. Taking all the results together, we can conclude that for any $C > 0$, $\mathbb{P}[T_n \geq C] \to 1$. ■

Proof of Corollary 3.7. a) Let $\widehat{x}_{ik} = \int_{[0,1]} X_i(t) \widehat{\psi}_k(t) dt$, so that $\langle X_i, \gamma\rangle_n = \sum_{k=1}^{p} \widehat{x}_{ik}\gamma_k$. Note that $\widehat{x}_{ik}$, $1 \leq k \leq p$, $1 \leq i \leq n$ are measurable functions of $X_1, \cdots, X_n$. Now, let $\widehat{F}_{\gamma,n}$ denote the empirical distribution function of the sample $\langle X_1, \gamma\rangle_n, \cdots, \langle X_n, \gamma\rangle_n$. Note that the elements of the matrices $(K_h(\widehat{F}_{\gamma,n}(\langle X_i, \gamma\rangle_n) - \widehat{F}_{\gamma,n}(\langle X_j, \gamma\rangle_n)))$ and $(K\left((i-j)/nh\right))$ are the same up to permutations of lines and columns. Given that in the proofs of Lemma 3.3 and Theorem 3.5 the arguments were provided conditionally on $X_1, \cdots, X_n$, it is quite clear that the conclusion of Theorem 3.5 remains true if the $\langle X_i, \gamma\rangle$'s are everywhere replaced by the $\langle X_i, \gamma\rangle_n$.

b) Similarly, all but one of the arguments in the proof of Theorem 3.6 applies with the $\langle X_i, \gamma\rangle_n$'s. It only remains to investigate the counterpart of $Q_{2n}(\widetilde{\gamma})$ that was the leading term in $Q_n(\widetilde{\gamma})$. For this purpose, note that for any γ, $\langle X_i, \gamma\rangle_n = \langle X_i, \gamma\rangle + \langle X_i, \Delta_{n,\gamma}\rangle$ where

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

For an integral operator $(\Psi v)(t) = \int \psi(t,s)v(s)ds$ with $\int\int \psi^2(t,s)dtds < \infty$, consider the operator norm $\|\Psi\|_S$ defined by $\|\Psi\|_S^2 = \int\int \psi^2(t,s)dtds$. Under Assumption D-(a) and the moment assumption on $\|X\|$,

$$\|\widehat{\Gamma} - \Gamma\|_S = O_{\mathbb{P}}(1/\sqrt{n}),$$

see for instance Bosq (2000) or Horváth and Kokoszka (2012). Next, by the Cauchy-Schwarz inequality, Lemma 4.3 in Bosq (2000) or Lemma 2.3 in Horváth and Kokoszka (2012), and the fact that the spectral norm of the operator $\widehat{\Gamma} - \Gamma$ is less than or equal to $\|\widehat{\Gamma} - \Gamma\|_S$,

$$\int_{[0,1]} \Delta_{n,\gamma}^2(t)dt \leq \left[\sum_{k=1}^{p} \gamma_k^2\right] \sum_{k=1}^{p} \|\widehat{\psi}_k - \psi_k\|^2 \leq p\frac{8}{\varsigma_p^2}\|\widehat{\Gamma} - \Gamma\|_S^2,$$

where $\varsigma_p = \min_{1\leq j\leq p}(\lambda_j - \lambda_{j+1})$. Then the lower bound for the spacing between the eigenvalues implies

$$\sup_{\gamma\in\mathcal{S}^p} \int_{[0,1]} \Delta_{n,\gamma}^2(t)dt \leq cp^{2\eta+1}\|\widehat{\Gamma} - \Gamma\|_S^2,$$

for some constant $c > 0$. One can deduce that

$$\sup_{\gamma\in\mathcal{S}^p} \max_{1\leq i\leq n} |\langle X_i, \gamma\rangle_n - \langle X_i, \gamma\rangle| \leq \max_{1\leq i\leq n} \|X_i\|c^{1/2}p^{\eta+1/2}\|\widehat{\Gamma} - \Gamma\|_S = O_{\mathbb{P}}(p^{\eta+1/2}\ln n/\sqrt{n}),$$

where for the last equality we used the condition $\mathbb{E}[\exp(\varrho\|X\|)] < \infty$ to deduce that $\max_{1\leq i\leq n}\|X_i\| = O_{\mathbb{P}}(\ln n)$. Let $b_n \downarrow 0$ such that $b_n\sqrt{n}/[p^{\eta+1/2}\ln n] \to \infty$ and define the event

$$\mathcal{E}_n = \{\sup_{\gamma\in\mathcal{S}^p} \max_{1\leq i\leq n} |\langle X_i, \gamma\rangle_n - \langle X_i, \gamma\rangle| \leq b_n\}$$

so that $\mathbb{P}(\mathcal{E}_n^c) \to 0$. On the set $\mathcal{E}_n$, for any $\gamma \in \mathcal{S}^p$ and $t \in \mathbb{R}$ we can write

$$\widehat{F}_{\gamma,n}(t) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}_{\{\langle X_i,\gamma\rangle_n \le t\}} \le \frac{1}{n}\sum_{i=1}^n \mathbb{I}_{\{\langle X_i,\gamma\rangle \le t+b_n\}} = F_{\gamma,n}(t+b_n),$$

and similarly, $\widehat{F}_{\gamma,n}(t) \ge F_{\gamma,n}(t-b_n)$. One can deduce that on $\mathcal{E}_n$,

$$\begin{aligned}&\left|\widehat{F}_{\widetilde{\gamma},n}(\langle X_i,\widetilde{\gamma}\rangle_n) - \widehat{F}_{\widetilde{\gamma},n}(\langle X_j,\widetilde{\gamma}\rangle_n)\right| \\ &\quad \le \max\{|F_{\widetilde{\gamma},n}(\langle X_i,\widetilde{\gamma}\rangle + b_n) - F_{\widetilde{\gamma},n}(\langle X_j,\widetilde{\gamma}\rangle - b_n)|\,,\;\; |F_{\widetilde{\gamma},n}(\langle X_i,\widetilde{\gamma}\rangle - b_n) - F_{\widetilde{\gamma},n}(\langle X_j,\widetilde{\gamma}\rangle + b_n)|\}.\end{aligned}$$

On the other hand,

$$\begin{aligned}|F_{\widetilde{\gamma},n}(\langle X_i,\widetilde{\gamma}\rangle + b_n) - F_{\widetilde{\gamma},n}(\langle X_j,\widetilde{\gamma}\rangle - b_n)| &\le |F_{\widetilde{\gamma},n}(\langle X_i,\widetilde{\gamma}\rangle + b_n) - F_{\widetilde{\gamma}}(\langle X_i,\widetilde{\gamma}\rangle + b_n)| \\ &\quad + |F_{\widetilde{\gamma}}(\langle X_i,\widetilde{\gamma}\rangle + b_n) - F_{\widetilde{\gamma}}(\langle X_i,\widetilde{\gamma}\rangle - b_n)| + |F_{\widetilde{\gamma},n}(\langle X_i,\widetilde{\gamma}\rangle - b_n) - F_{\widetilde{\gamma}}(\langle X_i,\widetilde{\gamma}\rangle - b_n)| \\ &\le 2\sup_{t\in\mathbb{R}} |F_{\widetilde{\gamma},n}(t) - F_{\widetilde{\gamma}}(t)| + 2b_n \sup_{t\in\mathbb{R}} f_{\widetilde{\gamma}}(t) \\ &= O_{\mathbb{P}}(n^{-1/2} + b_n) = O_{\mathbb{P}}(b_n).\end{aligned}$$

From this and the Lipschitz condition on K, one can deduce that

$$\left|K_h(\widehat{F}_{\widetilde{\gamma},n}(\langle X_i,\widetilde{\gamma}\rangle_n) - \widehat{F}_{\widetilde{\gamma},n}(\langle X_j,\widetilde{\gamma}\rangle_n)) - K_h(F_{\widetilde{\gamma},n}(\langle X_i,\widetilde{\gamma}\rangle) - F_{\widetilde{\gamma},n}(\langle X_j,\widetilde{\gamma}\rangle))\right| = O_{\mathbb{P}}(b_n h^{-1}).$$

Let $\widehat{Q}_{2n}(\widetilde{\gamma})$ be defined like $Q_{2n}(\widetilde{\gamma})$ but with $\langle X_i,\widetilde{\gamma}\rangle$'s replaced by $\langle X_i,\widetilde{\gamma}\rangle_n$'s. One can deduce from above

$$\left|\widehat{Q}_{2n}(\widetilde{\gamma}) - Q_{2n}(\widetilde{\gamma})\right| \le O_{\mathbb{P}}(b_n h^{-2})\left[\frac{1}{n}\sum_{1\le i\le n}\|\delta(X_i)\|\right]^2 = o_{\mathbb{P}}(1),$$

provided $b_n\sqrt{n}/[p^{\eta+1/2}\ln n] \to \infty$ and $b_n = o(h^2)$. The conclusion follows. ■

Proof of Theorem 3.8. Consider the event $\mathcal{A}_n = \{\max_{1\leq i\leq n} \|U_i\| \leq M\}$ with $M = n^{1/4-a}$ for some small a. Assumption D-(a) guarantees $\mathbb{P}(\mathcal{A}_n^c) \to 0$. We define

$$h_{i,j}^b = \frac{\zeta_i \, \zeta_j}{n(n-1)h} C_{n,ij},$$

where

$$C_{n,ij} = \frac{\langle U_i \mathbb{I}_{\{\|U_i\|\leq M\}}, U_j \mathbb{I}_{\{\|U_j\|\leq M\}} \rangle}{M^2} K_h \left(F_{\gamma,n}(\langle x_i, \gamma \rangle) - F_{\gamma,n}(\langle x_j, \gamma \rangle) \right).$$

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

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

Note that for any $t > 0$

$$\mathbb{P}\left[\sup_{\gamma} \left| M^{-2} Q_n^b\left(\gamma\right) - Q_{M,n}^b\left(\gamma\right) \right| > t \mid U_1, X_1, \cdots, U_n, X_n \right] \leq \mathbb{P}(\mathcal{A}_n^c) \to 0. \qquad \text{(S-3.12)}$$

We define the quantities A_n^b, B_n^b, C_n^b and D_n^b as in (5.1)-(5.2) with $h_{i,j}$ replaced by $h_{i,j}^b$ and the expectations replaced by the conditional expectations given $(U_1, X_1), \cdots, (U_n, X_n)$. When the variables ζ_i are bounded, it is easy to check that the same upper bounds as in Lemma 5.2 could be derived on the event $\mathcal{A}_n$. When the variables ζ_i have an exponential moment, that is $\mathbb{E}(\exp\{c|\zeta|\}) < \infty$ for some $c > 0$, it suffice to redefine the event $\mathcal{A}_n = \{\max_{1\leq i\leq n} \|U_i\|\zeta_i \leq M\}$ with $M = n^{1/4-a} \ln n \leq n^{1/4-a'}$ for some small $0 < a' < a$. Thus, the upper bounds as in Lemma 5.2 could still be derived on this event $\mathcal{A}_n$. Moreover, the condition $\mathbb{E}(U \mid X) = 0$, a.s. was not required for deriving that bounds. Hence the arguments are valid both under the null hypothesis and the alternative hypotheses. Then equation (S-3.12) and the exponential inequality from Lemma 5.1 applied as in Lemma 3.3

yield, for any $C > 0$,

$$\mathbb{P}\left[\sup_{\gamma}\left|Q_n^b(\gamma)\right| > Cp\ln n/nh^{1/2} \mid U_1, X_1, \cdots, U_n, X_n\right] \to 0 \quad \text{in probability} . \tag{S-3.13}$$

The second part of Lemma 3.3 follows from similar arguments.

Next, let us consider the case where the null hypothesis H_0 holds true. Let $\widehat{\gamma}_n^b$ be a least favorable direction γ for H_0 obtained as in equation (3.4) using a bootstrap sample. We deduce that $\mathbb{P}(\widehat{\gamma}_n^b \neq \gamma_0^{(p)} \mid U_1, X_1, \cdots, U_n, X_n) \to 0$, in probability. It remains to reconsider the steps of Theorem 3.5 with $\widetilde{Z}_i$ replaced by ζ_i and $K_{h,ij}(\gamma_0^{(p)})$ by $K_{h,ij}^b(\gamma_0^{(p)}) = \langle U_i, U_j\rangle K_{h,ij}(\gamma_0^{(p)})$. Consequently W_{ij} becomes $W_{ij}^b = n^{-1}(n-1)^{-1}h^{-1}\zeta_i\zeta_j K_{h,ij}^b(\gamma_0^{(p)})$, σ_{ij}^2 is replaced by $(\sigma_{ij}^b)^2 = [K_{h,ij}^b(\gamma_0^{(p)})]^2/[n^2(n-1)^2h^2]$ and $\sigma(n)^2$ is now $\sigma^b(n)^2 = 2\sum_{i\neq j}(\sigma_{ij}^b)^2$. Let us define the set $\mathcal{E}_{1n} = \{\sigma^b(n)^2 \geq \underline{\sigma}^2\}$ where $\underline{\sigma}^2$ is the lower bound in Assumption D-(c)-(i). Since $\lim_n \mathbb{E}(\sigma^b(n)^2) \geq 2\underline{\sigma}^2$ and the variance of $\sigma^b(n)^2$ tends to zero, we deduce that $\mathbb{P}(\mathcal{E}_{1n}^c) \to 0$. Next, note that $\lim_n \mathbb{E}(\sigma^b(n)^2) \leq C$ where C is some constant that depends on C_2 and ν defined in Assumption D-(c)-(ii). Thus, if $\mathcal{E}_{2n} = \{\sigma^b(n)^2 \leq 2C\}$, since the variance of $\sigma^b(n)^2$ tends to zero, we have $\mathbb{P}(\mathcal{E}_{2n}^c) \to 0$. On $\mathcal{E}_{1n}\cap\mathcal{E}_{2n}$, Conditions 1 and 2 of Theorem 5.1 of de Jong (1987) are clearly satisfied, given $(U_1, X_1), \cdots, (U_n, X_n)$. For checking Condition 3, let $\mathcal{K}^b$ denote the matrix with generic element $\mathcal{K}_{ij}^b = K_{h,ij}^b(\gamma_0^{(p)})/[n(n-1)h]$ if $i \neq j$ and $\mathcal{K}_{ij}^b = 0$ otherwise. Recall that $\mathbb{E}_n$ stands for the conditional expectation given $X_1, \cdots, X_n$ and note that $\mathbb{E}_n(\|U_i\|\|U_j\|) \leq \mathbb{E}^{1/2}(\|U_i\|^2 \mid X_i)\mathbb{E}^{1/2}(\|U_j\|^2 \mid X_j) \leq C_2^{2/\nu}$. Using the conditional independence between any U_i and the rest of the sample, for any $w \in \mathbb{R}^n$

with $\|w\| = 1$,

$$\begin{aligned}
\mathbb{E}_n \left\|\mathcal{K}^b w\right\|^2 &\leq \frac{1}{h^2 n^2 (n-1)^2} \sum_{i=1}^n \mathbb{E}(\|U_i\|^2 \mid X_i) \mathbb{E}_n \left(\sum_{j=1, j\neq i}^n \|U_j\| K_{h,ij}(\gamma_0^{(p)}) |w_j| \right)^2 \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{[Cauchy-Schwarz inequality]} \\
&\leq \frac{C_2^{4/\nu}}{h^2 n^2 (n-1)^2} \sum_{i,j,k=1}^n K_{h,ij}(\gamma_0^{(p)}) K_{h,ik}(\gamma_0^{(p)}) |w_j w_k| \\
&\leq C_2^{4/\nu} K^2(0) \frac{1}{h^2 n (n-1)^2} \sum_{j,k=1}^n |w_j w_k| \\
&\leq \frac{C_3}{n^2} \frac{1}{nh^2}, \qquad\qquad\qquad\qquad \text{[Cauchy-Schwarz inequality]}
\end{aligned}$$

where $C_3 > 0$ is some constant. We deduce that $\mathbb{E}\left\|\mathcal{K}^b w\right\|^2 = o(n^{-2})$. Let $|||\mathcal{K}^b|||_2$ denote the spectral norm of $\mathcal{K}^b$ and define $\mathcal{E}_{3n} = \{|||\mathcal{K}^b|||_2 \leq 1/n\}$. We deduce from the above that $\mathbb{P}(\mathcal{E}_{3n}^c) \to 0$, and thus the conditional probability of $\mathcal{E}_{3n}^c$ given $(U_1, X_1), \cdots, (U_n, X_n)$ also tends to zero. Condition 3 in Theorem 5.1 of de Jong (1987) is clearly satisfied on $\mathcal{E}_{3n}$ and hence de Jong's CLT could be applied, given $(U_1, X_1), \cdots, (U_n, X_n)$, on the event $\mathcal{E}_n = \mathcal{E}_{1n} \cap \mathcal{E}_{2n} \cap \mathcal{E}_{3n}$ which has a probability tending to one. Finally, it remains to note that equation (S-3.10) holds on $\mathcal{E}_n$, given $(U_1, X_1), \cdots, (U_n, X_n)$. The arguments for the test statistic built with the estimated FPC basis (that is not changed in the bootstrap procedure) are similar.

Let us now consider the case where the null hypothesis does not hold true. Let $\widehat{v}_n^{b,2}(\gamma)$ be the variance estimator obtained by bootstrapping. Since $\zeta_i^2 \geq (\sqrt{5}-1)^2/4$ and $\langle U_i^b, U_j^b\rangle^2 > \langle U_i^b, U_j^b\rangle^2(\sqrt{5}-1)^4/16$, by the second part of Lemma 3.3,

$$\sup_\gamma \{1/\widehat{v}_n^{b,2}(\gamma)\} = O_{\mathbb{P}}(1).$$

From this and the bound in equation (S-3.13), we deduce $T_n^b = O_{\mathbb{P}}(p \ln n)$ given the original sample of (U, X), in probability. That means that for any $C > 0$, $\mathbb{P}(|T_n^b| > C \mid$

$U_1, X_1, \cdots, U_n, X_n) \to 0$, in probability. Consequently, $T_n^b/\alpha_n = o_{\mathbb{P}}(1)$ given the original sample of (U, X), in probability. In particular, this implies that for any $a \in (0,1)$, $z_{1-a,n}^b/\alpha_n = o_{\mathbb{P}}(1)$. On the other hand, we proved in Theorem 3.6 that T_n tend to infinity at the rate $O_{\mathbb{P}}(nh^{1/2}r_n^2)$, which implies that T_n/α_n tends to infinity, in probability. Thus, the second statement of Theorem 3.8 follows. Again, the arguments for the test statistic built with the estimated FPC basis are similar. ∎

Lemma A. Under the conditions of Lemma 3.1, for any $p \geq 1$ and $\gamma \in \mathcal{S}^p$,

$$\mathbb{E}(U \mid \langle X, \gamma\rangle) = \mathbb{E}(U \mid F_\gamma(\langle X, \gamma\rangle)) \;\; a.s.$$

Proof of Lemma A. It suffice to prove that for any U as in Lemma 3.1 and any Z a real-valued random variable with distribution function F, we have

$$\mathbb{E}(U \mid Z) = \mathbb{E}(U \mid F(Z)) \quad a.s. \tag{S-3.14}$$

For any random variable Z (not necessarily continuous) with distribution function F we have $\mathbb{P}(Q(F(Z)) \neq Z) = 0$ where $Q(t) = \inf\{y : F(y) \geq t\}, \forall 0 < t < 1$. (See, for instance Proposition 3, Chapter 1 in Shorack and Wellner (1986).) From this and the properties of the conditional expectations, for any bounded measurable function g we have

$$\begin{aligned}\mathbb{E}(g(Z)\mathbb{E}(U \mid Z)) &= \mathbb{E}(g(Z)U) \\ &= \mathbb{E}(g(Q(F(Z)))U) = \mathbb{E}(g(Q(F(Z)))\mathbb{E}(U \mid F(Z))) = \mathbb{E}(g(Z)\mathbb{E}(U \mid F(Z))).\end{aligned}$$

Since $\mathbb{E}(U \mid F(Z))$ is a measurable function of Z, the almost sure uniqueness of the conditional expectation implies the equality in equation (S-3.14). ∎

Now, let us provide some theoretical justification for the sequential numerical algorithm for searching an approximation of the direction $\widehat{\gamma}_n$ defined in equation (3.4). A justification is necessary only in the case of alternative hypotheses to ensure that a vector $\widetilde{\gamma}$ like in Theorem 3.6-(e) exists. No additional justification is required for the asymptotic results on the null hypothesis since Lemma 3.4 still holds with $\widehat{\gamma}_n$ replaced by the solution obtained through our sequential numerical algorithm.

It follows from the proof of Lemma 3.1-(B) that, if

$$\mathbb{P}\left(\mathbb{E}\left[U \mid \langle X, \psi_1\rangle, \cdots, \langle X, \psi_p\rangle\right] = 0\right) < 1,$$

then the set

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

has Lebesgue measure zero on the unit hypersphere $\mathcal{S}^p$ and is not dense. See also Lemma 2.1 in Patilea, Saumard and Sanchez (2012).

Let us next investigate what could happen when searching a direction in $\mathcal{S}^{p+1}$ using a direction in $\mathcal{S}^p$ and one-dimensional optimization. For a vector $v \in \mathbb{R}^p$, let $(v, 1) \in \mathbb{R}^{p+1}$ denote the vector obtained by adding an additional component equal to 1. Moreover, let $\mathbf{0}_p$ denote the null vector in $\mathbb{R}^p$. We prove in the following result that if $\gamma \notin \mathcal{A}_p$, then there exists at most a finite number of angles $\theta \in [0, \pi)$ such that

$$\cos\theta \cdot (\gamma, 0) + \sin\theta \cdot (\mathbf{0}_p, 1) \in \mathcal{A}_{p+1} \subset \mathcal{S}^{p+1}.$$

Moreover, if $\gamma \in \mathcal{A}_p$ and $\mathbb{E}(U \mid \langle X, \gamma\rangle, \langle X, \psi_{p+1}\rangle) \neq 0$ (for instance, this happens when $\mathbb{E}(U \mid \langle X, \psi_{p+1}\rangle) \neq 0$), one could draw the same conclusion on the cardinality of the set of θ such that $\cos\theta \cdot (\gamma, 0) + \sin\theta \cdot (\mathbf{0}_p, 1) \in \mathcal{A}_{p+1}$.

Lemma B. Assume that $\mathbb{E}\|U\| < \infty$, $\mathbb{E}(U) = 0$ and there exists $s > 0$ such that

$\mathbb{E}(\|U\| \exp(s\|X\|)) < \infty$. If either $\gamma \in \mathcal{S}^p \setminus \mathcal{A}_p$, or

$$\gamma \in \mathcal{A}_p \quad \text{and} \quad \mathbb{E}(U \mid \langle X, \gamma \rangle, \langle X, \psi_{p+1} \rangle) \neq 0,$$

then the set of $\theta \in [0, \pi)$ such that

$$\cos\theta \cdot (\gamma, 0) + \sin\theta \cdot (\mathbf{0}_p, 1) \in \mathcal{A}_{p+1} \subset \mathcal{S}^{p+1}$$

is empty or finite.

Proof of Lemma B. If $\gamma \notin \mathcal{A}_p$, then $\mathbb{E}(U \mid \langle X, \gamma \rangle) \neq 0$. In particular, we have $\mathbb{E}(U \mid \langle X, \gamma \rangle, \langle X, \psi_{p+1} \rangle) \neq 0$ and thus the arguments below apply for both cases in the statement of Lemma B. Now, if $\theta^* \in (0, \pi)$ is such that

$$\mathbb{E}(U \mid \cos\theta^* \cdot \langle X, \gamma \rangle + \sin\theta^* \cdot \langle X, \psi_{p+1} \rangle) = 0, \tag{S-3.15}$$

then, for any $b \in \mathbb{R}$, θ^* is a zero of the analytic function

$$\theta \mapsto \zeta_b(\theta) = \mathbb{E}\left[U \exp(b\{\cos\theta \cdot \langle X, \gamma \rangle + \sin\theta \cdot \langle X, \psi_{p+1} \rangle\})\right], \qquad \theta \in (0, 2\pi). \tag{S-3.16}$$

(Since X could be rescaled conveniently if necessary, we consider $\mathbb{E}(\|U\| \exp(s\|X\|)) < \infty$ for some $s \geq 1$. This guarantees that the expectation in the display (S-3.16) is well defined.) Since $\mathbb{E}(U \mid \langle X, \gamma \rangle, \langle X, \psi_{p+1} \rangle) \neq 0$, there exists b such that the analytic function $\zeta_b(\cdot)$ is not the null function. Otherwise, by the unicity of the Fourier Transform, necessarily $\mathbb{E}(U \mid \langle X, \gamma \rangle, \langle X, \psi_{p+1} \rangle) = 0$ a.s. Next, the zeros of a non-null analytic function in a bounded interval are necessarily in finite number. Hence, the set of θ^* satisfying equation (S-3.15) is necessarily finite. ■

ADDITIONAL REFERENCES

1. Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications.* Lecture Notes in Statistics (v. 149). Springer-Verlag, New-York.

2. Gao, J., and Gijbels, I. (2008). Bandwidth selection in nonparametric kernel testing. *J. Amer. Statist. Assoc.* **103**, 1584–1594.

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

4. Horowitz, J., and Spokoiny, V. (2001). An adaptive rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* **69**, 599–633.

5. Shorack, G.R., and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics.* John Wiley & Sons.