Projection-based nonparametric goodness-of-fit testing with functional data

Valentin Patilea
(joint work with C. Sanchez-Sellero & M. Saumard)

CREST-Ensai, France

High dimensional and dependent functional data
University of Bristol, September 10-12, 2012

Outline

1. Introduction
   - The statistical problem
2. The principle: scalar responses case
3. The test statistic for scalar responses
4. Extending the principle: functional responses
5. Numerical illustrations
6. Conclusions

- The observations: independent copies of

$$(U, X, Z) \in \mathcal{H}_1 \times \mathcal{H}_2 \times \mathbb{R}^k$$

Typically

1. $\mathcal{H}_1 = \mathbb{R}$ and $\mathcal{H}_2 = L^2[0,1]$
2. or $\mathcal{H}_1 = \mathcal{H}_2 = L^2[0,1]$

- The observations: independent copies of

$$(U, X, Z) \in \mathcal{H}_1 \times \mathcal{H}_2 \times \mathbb{R}^k$$

Typically

1. $\mathcal{H}_1 = \mathbb{R}$ and $\mathcal{H}_2 = L^2[0,1]$
2. or $\mathcal{H}_1 = \mathcal{H}_2 = L^2[0,1]$

- The variable U could be observed or *estimated* from some model

- The observations: independent copies of

$$(U, X, Z) \in \mathcal{H}_1 \times \mathcal{H}_2 \times \mathbb{R}^k$$

Typically

1. $\mathcal{H}_1 = \mathbb{R}$ and $\mathcal{H}_2 = L^2[0, 1]$
2. or $\mathcal{H}_1 = \mathcal{H}_2 = L^2[0, 1]$

- The variable U could be observed or *estimated* from some model

- The problem: test

$$H_0 : \quad \mathbb{E}(U \mid X, Z) = 0 \quad a.s.$$

against the nonparametric alternative

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

Examples: testing the non-effect

- Independent samples of U, X and Z are observed
  - Scalar response U (functional covariate X, vector covariate Z)
  - Functional response U (functional covariate X, vector covariate Z)

Testing parametric functional regression models (1/2)

- Functional quadratic regression
  - samples of $Y \in \mathbb{R}$ and $X \in L^2[0,1]$ are observed

- For some *unknown* $c \in \mathbb{R}$ and $b_1 \in L^2[0,1]$, $b_2 \in L^2([0,1] \times [0,1])$

$$Y = c + \int_0^1 b_1(t)X(t)dt + \int_0^1 \int_0^1 b_2(t,s)X(t)X(s)dtds + U$$

- The functional quadratic model is correct *iff*

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

Testing parametric functional regression models (2/2)

- Functional generalized linear models
  - samples of $Y \in \mathbb{R}$, $X \in L^2[0,1]$ and $Z \in \mathbb{R}^k$ are observed

- For some given $g(\cdot)$ and *unknown* $c \in \mathbb{R}$, $\gamma \in \mathbb{R}^k$ and $b \in L^2[0,1]$

$$Y = g(c_0 + Z'\gamma + \langle b, X\rangle) + U$$

where

$$\langle b, X\rangle = \int_0^1 b(t)X(t)dt.$$

- Examples: $g(u) = u$, $g(u) = \exp(u)[1+\exp(u)]^{-1}$,...

- The parametric model with functional covariates is correct *iff*

$$\mathbb{E}(U \mid X, Z) = 0 \quad a.s.$$

Testing semiparametric functional regressions

- Semi-functional partial linear functional regression model
  - samples of $Y \in \mathbb{R}$, $Z \in \mathbb{R}^k$ and $X \in L^2[0,1]$ are observed
  - For some *unknown* $c \in \mathbb{R}$, $\gamma \in \mathbb{R}^k$ and $m(\cdot)$

$$Y = c + Z'\gamma + m(X) + U$$

- Before estimating the function $m(\cdot)$, one should check the effect of the functional covariate

- Testing the significance of the functional covariate:

$$\mathbb{E}(Y - c - Z'\gamma \mid X, Z) = 0 \quad a.s.$$

Outline

1 Introduction
2 The principle: scalar responses case
3 The test statistic for scalar responses
4 Extending the principle: functional responses
5 Numerical illustrations
6 Conclusions

Some notation

- Let fix $\mathcal{R} = \{\rho_1, \rho_2, \cdots\}$ a basis of functions in $L^2[0,1]$
- If $X \in L^2[0,1]$

$$X = \sum_{i \geq 1} x_j \rho_j$$

- For any positive integer p, let $\mathcal{S}^p = \{\gamma \in \mathbb{R}^p : \|\gamma\| = 1\}$
- If $\gamma = (\gamma_1, \cdots, \gamma_p)$, define

$$\langle X, \gamma \rangle = \sum_{j=1}^{p} x_j \gamma_j$$

- Let $X \in L^2[0,1]$, $Z \in \mathbb{R}^k$ and $U \in \mathbb{R}$ be random variables
- $\mathbb{E}|U| < \infty$ and $\mathbb{E}(U) = 0$

A fundamental Lemma

Lemma

(A) The following statements are equivalent:

1. $\mathbb{E}(U \mid X, Z) = 0$ *a.s.*
2. $\mathbb{E}(U \mid \langle X, \beta \rangle, Z) = 0$ *a.s.* $\forall \beta \in L^2[0,1]$ *with* $\|\beta\|_{L^2} = 1$
3. *for any integer* $p \geq 1$, $\mathbb{E}(U \mid \langle X, \gamma \rangle, Z) = 0$ *a.s.* $\forall \gamma \in \mathcal{S}^p$.
4. *for any integer* $p \geq 1$, $\mathbb{E}(U \mid X^{(p)}, Z) = 0$ *a.s.*

Lemma (Fundamental Lemma cont'd)

(B) Under some mild additional conditions if

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

then there exists a positive integer $p_0 \geq 1$ *such that for any integer* $p > p_0$*, the set*

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

has Lebesgue measure zero on the unit hypersphere $\mathcal{S}^p$*.*

Corollary

Let

- U *such that* $\mathbb{E}|U| < \infty$
- *For any* $p \geq 1$, $\gamma \in \mathbb{R}^p$, *let* $w_{p,\gamma}(t,z)$, $t \in \mathbb{R}$ *and* $z \in \mathbb{R}^k$, *be a real-valued function such that* $w_{p,\gamma}(\langle X, \gamma \rangle, Z) > 0$ *for all* $\|\gamma\| = 1$.

The following statements are equivalent:

1. *The null hypothesis* $H_0 : \mathbb{E}(U \mid X, Z) = 0$ *a.s. holds true.*
2. *for any* $p \geq 1$ *and any set* $B_p \subset \mathcal{S}^p$ *with strictly positive Lebesgue measure in on the unit hypersphere* $\mathcal{S}^p$,

$$\max_{\gamma \in B_p} \mathbb{E}\left[U \mathbb{E}\left(U | \langle X, \gamma \rangle, Z\right) w_{p,\gamma}(\langle X, \gamma \rangle, Z)\right] = 0. \tag{1}$$

- For any $\gamma \in \mathbb{R}^p$, let $f_\gamma(t,z)$ be the joint density of $\langle X, \gamma \rangle$ and Z
- Let

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

- For any $p \geq 1$, let $B_p \subset \mathcal{S}^p$ be a set with strictly positive Lebesgue measure in $\mathcal{S}^p$.
- By the Corollary, the null hypothesis $H_0 : \mathbb{E}(U \mid X, Z) = 0$ *a.s.* holds true if and only if

$$\forall p \geq 1, \quad \max_{\gamma \in B_p} Q(\gamma) = 0. \tag{2}$$

- Idea: build a sample approximation $Q_n(\gamma)$ of $Q(\gamma)$ and look for the worse direction γ by maximizing $Q_n(\gamma)$.

Outline

1. Introduction
2. The principle: scalar responses case
3. The test statistic for scalar responses
   - Behavior under the null
   - Behavior under the alternatives
4. Extending the principle: functional responses
5. Numerical illustrations
6. Conclusions

- For any $\gamma \in \mathcal{S}^p$, let

$$Q_n(\gamma) = \frac{1}{n(n-1)h^{k+1}} \sum_{1 \leq i \neq j \leq n} U_i U_j K\left(\langle X_i - X_j, \gamma \rangle / h\right) \widetilde{K}\left((Z_i - Z_j)/h\right),$$

where $K(\cdot)$ is a univariate kernel, $\widetilde{K}(\cdot)$ is a multivariate kernel and h a bandwidth.

- $Q_n(\gamma)$ is a sample based approximation of

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

- The least favorable direction γ for H_0 is defined as

$$\widehat{\gamma}_n = \arg\max_{\gamma \in B_p} \left[nh^{(k+1)/2} Q_n(\gamma)/\widehat{v}_n(\gamma) - \alpha_n \mathbb{I}_{\{\gamma \neq \gamma_0^{(p)}\}} \right], \qquad (3)$$

where

  - $\widehat{v}_n^2(\cdot)$ be as estimate of the variance of $nh^{1/2}Q_n(\cdot)$
  - $\gamma_0^{(p)}$ is an initial *guess* and $B_p \subset \mathcal{S}^p$ with strictly positive Lebesgue measure in $\mathcal{S}^p$ that contains $\gamma_0^{(p)}$
  - $\alpha_n \uparrow \infty$, $n \geq 1$ is a sequence that depends on the sample size and the rates of h and p

The test statistic

- Consider

$$T_n = nh^{(k+1)/2} \frac{Q_n(\widehat{\gamma}_n)}{\widehat{v}_n(\widehat{\gamma}_n)}$$

- An asymptotic α-level test is given by $\mathbb{I}\,(T_n \geq z_{1-a})$, where z_a is the $(1-a)$-th quantile of the standard normal distribution

- The variance could be estimated by

$$\widehat{v}_n^2\,(\gamma) = \frac{2}{n(n-1)h^{k+1}} \sum_{j \neq i} U_i^2 U_j^2 K^2\left(\langle X_i - X_j, \gamma \rangle / h\right) \widetilde{K}^2\left((Z_i - Z_j)/h\right)$$

Technical conditions (1/2)

(a) The random vectors $(U_1, X_1, Z_1), \ldots, (U_n, X_n, Z_n)$ are independent draws from the random vector $(U, X, Z) \in \mathbb{R} \times L^2[0,1] \times \mathbb{R}^k$ that satisfies $\mathbb{E}|U|^m < \infty$ for some $m > 11$.

(b) $\exists\, \underline{\sigma}^2$ and $\overline{\sigma}^2$ such that $0 < \underline{\sigma}^2 \leq Var(U \mid X, Z) \leq \overline{\sigma}^2 < \infty$ a.s.

(c) The sets $B_p \subset \mathcal{S}^p$, $p \geq 1$ are such that:

   (i) $\forall \gamma \in B_p$, $\langle X, \gamma \rangle$ and Z admit a joint density $f_\gamma(\cdot,\cdot)$ that satisfies some mild technical conditions;

   (ii) the initial 'guesses' $\gamma_0^{(p)} \in B_p$ satisfies the condition: $\exists C$ such that $f_{\gamma_0^{(p)}} \leq C$, $\forall p \geq 1$.

   (iii) $B_p \times 0_{p'-p} \subset B_{p'}$, $\forall 1 \leq p < p'$.

Technical conditions (2/2)

(a) The kernels K and $\widetilde{K}$ satisfy some mild conditions

(b) $h \to 0$ and $nh^{2(k+1)} / \ln^{\alpha} n \to \infty$ for some $\alpha > 1$.

(c) $p \geq 1$ depends on n: $\exists \lambda > 0$ such that $p \ln^{-\lambda} n$ is bounded.

The steps of the theory under the null hypothesis (1/3)

Lemma

Under the technical conditions and if H_0 holds true,

$$\sup_{\gamma \in B_p \subset \mathcal{S}^p} |Q_n(\gamma)| = O_{\mathbb{P}}(n^{-1} h^{-(k+1)/2} p^{3/2} \ln n).$$

The steps of the theory under the null hypothesis (1/3)

Lemma

Under the technical conditions and if H_0 holds true,

$$\sup_{\gamma \in B_p \subset \mathcal{S}^p} |Q_n(\gamma)| = O_{\mathbb{P}}(n^{-1} h^{-(k+1)/2} p^{3/2} \ln n).$$

Derived using concentration inequalities for degenerate $U-$processes

The steps of the theory under the null hypothesis (2/3)

Lemma

Under the technical conditions, for a positive sequence α_n, $n \geq 1$ such that $\alpha_n / \{p^{3/2} \ln n\} \to \infty$,

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

The steps of the theory under the null hypothesis (2/3)

Lemma

Under the technical conditions, for a positive sequence α_n, $n \geq 1$ such that $\alpha_n / \{p^{3/2} \ln n\} \to \infty$,

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

By definition,

$$nh^{(k+1)/2} Q_n(\gamma_0^{(p)}) / \widehat{v}_n(\gamma_0^{(p)}) \leq nh^{(k+1)/2} Q_n(\widehat{\gamma}_n) / \widehat{v}_n(\widehat{\gamma}_n) - \alpha_n \mathbb{I}(\widehat{\gamma}_n \neq \gamma_0^{(p)})$$

The steps of the theory under the null hypothesis (2/3)

Lemma

Under the technical conditions, for a positive sequence α_n, $n \geq 1$ such that $\alpha_n / \{p^{3/2} \ln n\} \to \infty$,

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

By definition,

$$nh^{(k+1)/2} Q_n(\gamma_0^{(p)}) / \widehat{v}_n(\gamma_0^{(p)}) \leq nh^{(k+1)/2} Q_n(\widehat{\gamma}_n) / \widehat{v}_n(\widehat{\gamma}_n) - \alpha_n \mathbb{I}(\widehat{\gamma}_n \neq \gamma_0^{(p)})$$

$$0 \leq \mathbb{I}(\widehat{\gamma}_n \neq \gamma_0^{(p)}) \leq \frac{nh^{(k+1)/2}}{\alpha_n} \left\{ Q_n(\widehat{\gamma}_n) / \widehat{v}_n(\widehat{\gamma}_n) - Q_n(\gamma_0^{(p)}) / \widehat{v}_n(\gamma_0^{(p)}) \right\} = o_{\mathbb{P}}(1)$$

The steps of the theory under the null hypothesis (3/3)

Theorem

If H_0 holds true, the test statistic T_n converges in law to a standard normal.

The steps of the theory under the null hypothesis (3/3)

Theorem

If H_0 holds true, the test statistic T_n converges in law to a standard normal.

Apply the CLT for the $U-$statistic

$$Q_n\left(\gamma_0^{(p)}\right) = \frac{1}{n(n-1)h^{k+1}} \sum_{1\leq i\neq j\leq n} U_i U_j K\left(\langle X_i - X_j, \gamma_0^{(p)}\rangle/h\right) \widetilde{K}\left((Z_i - Z_j)/h\right)$$

The steps of the theory under the null hypothesis (3/3)

Theorem

If H_0 holds true, the test statistic T_n converges in law to a standard normal.

Apply the CLT for the $U-$statistic

$$Q_n\left(\gamma_0^{(p)}\right) = \frac{1}{n(n-1)h^{k+1}} \sum_{1\leq i\neq j\leq n} U_i U_j K\left(\langle X_i - X_j, \gamma_0^{(p)}\rangle / h\right) \widetilde{K}\left((Z_i - Z_j)/h\right)$$

Control the spectral norm ($2-$norm) and the Frobenius norm of the zero-diagonal matrix with generic element

$$\frac{1}{n(n-1)h^{k+1}} K\left(\langle X_i - X_j, \gamma_0^{(p)}\rangle / h\right) \widetilde{K}\left((Z_i - Z_j)/h\right), \qquad i \neq j$$

The Omnibus test property

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

- Let some real-valued function $\delta(X, Z)$ such that $\mathbb{E}[\delta(X, Z)] = 0$ and $0 < \mathbb{E}[\delta^4(X, Z)] < \infty$, and some sequence of real numbers r_n that could decrease to zero (the case $r_n \equiv 1$ is also included).
- Consider the sequence of alternatives

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

- We show that such directional alternatives can be detected as soon as $r_n^2 n h^{(k+1)/2} / \alpha_n$ tends to infinity.

- However, in the functional data framework, to obtain the convenient standard normal critical values, we need $1/\alpha_n = o(p^{-3/2} \ln^{-1} n)$.

- Hence, the rate r_n at which the alternatives H_{1n} tend to the null hypothesis should satisfy $r_n^2 n h^{(k+1)/2} / \{p^{3/2} \ln n\} \to \infty$.

Testing the functional linear model

- Simplify and consider the case with no covariate Z
- The model we want to test is the functional linear model defined by

$$Y = a + \langle b, X \rangle + U,$$

where $b \in L^2[0,1]$ and $a \in \mathbb{R}$ are unknown parameters.

- The null hypothesis is

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

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

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

where $\overline{U}_n = n^{-1}\sum_{i=1}^n U_i$.

- Let $\widehat{U}_i = Y_i - \widehat{a} - \langle \widehat{b}, X_i \rangle$ be the residuals and let

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

where $\widehat{v}_n^2(\cdot; \widehat{a}, \widehat{b})$ is an estimate of the variance of $nh^{1/2}Q_n(\cdot; \widehat{a}, \widehat{b})$.

- Given $B_p \subset \mathcal{S}^p$, let

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

- The test statistic is then

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

- Suppose
  - $\|\widehat{b} - b\|_{L^2} = O_{\mathbb{P}}(n^{-\rho})$ for some $3/8 < \rho \leq 1/2$
  - The bandwidth h is such that $n^{1-2\zeta}h^{1/2} \to 0$ for some $3/8 < \zeta < \rho$.

- Under some stronger moment conditions, we show that an asymptotic α-level test is given by $\mathbb{I}\left(T_n \geq z_{1-a}\right)$, where z_a is the $(1-a)-$quantile of the standard normal distribution.

Consistency

- The alternatives of the functional linear model considered are

$$H_{1n}: \ Y_{in} = a + \langle b, X_i \rangle + r_n \delta(X_i) + U_i^0, \quad n \geq 1, \quad \text{with } \mathbb{E}(U_i^0 \mid X_i) = 0,$$

with $\delta(\cdot)$ an real-valued function such that $0 < \mathbb{E}[\delta^4(X)] < \infty$ and r_n, $n \geq 1$ a sequence of real numbers.

- Moreover, $\delta(\cdot)$ satisfies the orthogonality conditions

$$\mathbb{E}[\delta(X)] = 0 \qquad \text{and} \qquad \mathbb{E}[\delta(X)X] = 0.$$

- Assume that
  - (i) $r_n^2 n h^{1/2}/\alpha_n \to \infty$;
  - (ii) $r_n^{-1}\|\widehat{b} - b\|_{L^2} = o_{\mathbb{P}}(1)$;
  - (iii) $\alpha_n/\{p^{3/2}\ln n\} \to \infty$.

- Then the test based on T_n functional linear regression model with probability tending to 1 (under some mild conditions).

Outline

1 Introduction

2 The principle: scalar responses case

3 The test statistic for scalar responses

4 Extending the principle: functional responses

5 Numerical illustrations

6 Conclusions

The fundamental Lemma

Lemma

Let $U, X \in L^2[0,1]$ *be random functions and* $Z \in \mathbb{R}^k$ *be a random vector. Assume that* $\mathbb{E}\|U\| < \infty$ *and* $\mathbb{E}(U) = 0$.
(A) The following statements are equivalent:

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

(B) Under some mild additional conditions, if $\mathbb{P}[\mathbb{E}(U \mid X, Z) = 0] < 1$, *then there exists a positive integer* p_0 *such that for any integer* $p \geq p_0$, *the set*

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

has Lebesgue measure zero on the unit hypersphere $\mathcal{S}^p$.

The test statistic (1/2)

- For $\gamma \in \mathcal{S}^p$ define

$$Q_n(\gamma) = \frac{1}{n(n-1)h^{k+1}} \sum_{1 \le i \ne j \le n} \langle U_i, U_j \rangle K\left(\langle X_i - X_j, \gamma \rangle / h \right) \widetilde{K}\left((Z_i - Z_j)/h \right),$$

- Let

$$\widehat{v}_n^2(\gamma) = \frac{2}{n(n-1)h^{k+1}} \sum_{j \ne i} \langle U_i, U_j \rangle^2 K^2\left(\langle X_i - X_j, \gamma \rangle / h \right) \widetilde{K}^2\left((Z_i - Z_j)/h \right)$$

- Given $B_p \subset \mathcal{S}^p$, the least favorable direction γ for H_0 is defined by

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

The test statistic (2/2)

- The test statistic is

$$T_n = nh^{1/2} \frac{Q_n(\widehat{\gamma}_n)}{\widehat{v}_n(\widehat{\gamma}_n)} \ . \tag{4}$$

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

Outline

1. Introduction
2. The principle: scalar responses case
3. The test statistic for scalar responses
4. Extending the principle: functional responses
5. Numerical illustrations
6. Conclusions

The simulation design (1/2)

- X is a standard Brownian motion on the unit interval $[0, 1]$.
- Three scenarii for the distribution of U_i:
  - *Null hypothesis* U is $N(0, \sigma^2)$ where $\sigma = 1.219$, U independent of X.
  - *Linear alternative*

$$U_i = \langle b, X_i \rangle + \varepsilon_i$$

    where $b(t) = (\sin(2\pi t^3))^3$, and $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. $N(0, \sigma^2)$, where $\sigma = 1.219$, corresponding to a 10% signal-to-noise ratio that is, $E(\langle b, X \rangle^2)/(E(\langle b, X \rangle^2) + \sigma^2) = 0.1$.
  - *Quadratic alternative*

$$U_i = \int_0^1 \int_0^1 h(s, t) X(s) X(t) \, ds \, dt + \varepsilon_i$$

    where $h(s, t) = 0.6$, and $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. $N(0, \sigma^2)$, where $\sigma = 1$.

The simulation design (2/2)

- We use the Karhunen-Loève expansion of the Brownian motion X

$$X(t) = \sum_{j=1}^{\infty} x_j \frac{1}{(j-0.5)\pi} \sqrt{2} \sin\left((j-0.5)\pi t\right)$$

to build the basis $\mathcal{R} = \{\sqrt{2}\sin((j-0.5)\pi t) : j = 1, 2, \dots\}$

- 1000 samples of $(U_1, X_1), \dots, (U_n, X_n)$ of sizes $n = 100$ and $n = 200$
- $\alpha_n = 5$, Epanechnikov kernel K, several bandwidth values tested
- Critical values corrected by wild bootstrap

		$p=3$				$p=5$			
		$c_h=0.6$	$c_h=0.8$	$c_h=1.0$	CT	$c_h=0.6$	$c_h=0.8$	$c_h=1.0$	CT
$n=100$	$\widehat{\tau}_n^2(\gamma)$	5.9	4.8	4.7	5.8	5.6	5.6	5.1	6.6
	$\widehat{v}_n^2$	7.3	6.1	5.7	5.8	7.9	6.5	5.6	6.6
$n=200$	$\widehat{\tau}_n^2(\gamma)$	4.1	5.0	5.4	4.9	4.5	4.9	4.7	5.5
	$\widehat{v}_n^2$	5.7	5.9	5.6	4.9	5.7	5.8	5.5	5.5

Table: Percentage of rejections under H_0, nominal level 5% ('best' direction).

		$p=3$				$p=5$			
		$c_h=0.6$	$c_h=0.8$	$c_h=1.0$	CT	$c_h=0.6$	$c_h=0.8$	$c_h=1.0$	CT
$n=100$	$\widehat{\tau}_n^2(\gamma)$	4.9	4.9	5.0	5.8	5.0	4.9	4.5	6.6
	$\widehat{v}_n^2$	7.1	6.4	5.9	5.8	7.5	6.1	5.4	6.6
$n=200$	$\widehat{\tau}_n^2(\gamma)$	5.0	5.2	5.1	4.9	5.1	4.5	4.9	5.5
	$\widehat{v}_n^2$	6.3	6.2	6.0	4.9	5.9	5.5	5.4	5.5

Table: Percentage of rejections under H_0, nominal level 5% nominal level 5% ('worst' direction).

		$p=3$				$p=5$			
		$c_h=0.6$	$c_h=0.8$	$c_h=1.0$	CT	$c_h=0.6$	$c_h=0.8$	$c_h=1.0$	CT
$n=100$	$\widehat{\tau}_n^2(\gamma)$	47.5	47.8	43.6	79.1	47.7	48.6	43.3	72.4
	$\widehat{v}_n^2$	49.3	51.6	47.0	79.1	49.2	51.7	47.3	72.4
$n=200$	$\widehat{\tau}_n^2(\gamma)$	83.0	83.5	81.8	98.0	78.9	83.5	82.9	96.5
	$\widehat{v}_n^2$	84.4	85.7	84.2	98.0	81.1	84.7	85.7	96.5

Table: Percent. rejections under the linear altern., level 5% ('best' direction).

		$p=3$				$p=5$			
		$c_h=0.6$	$c_h=0.8$	$c_h=1.0$	CT	$c_h=0.6$	$c_h=0.8$	$c_h=1.0$	CT
$n=100$	$\widehat{\tau}_n^2(\gamma)$	24.0	21.7	17.9	79.1	28.5	22.7	17.1	72.4
	$\widehat{v}_n^2$	33.8	29.4	24.1	79.1	40.9	33.0	25.0	72.4
$n=200$	$\widehat{\tau}_n^2(\gamma)$	52.8	52.4	49.3	98.0	66.3	59.7	52.9	96.5
	$\widehat{v}_n^2$	59.2	58.6	55.2	98.0	72.5	65.8	59.3	96.5

Table: Percent. rejections under the linear altern., level 5% ('worst' direction).

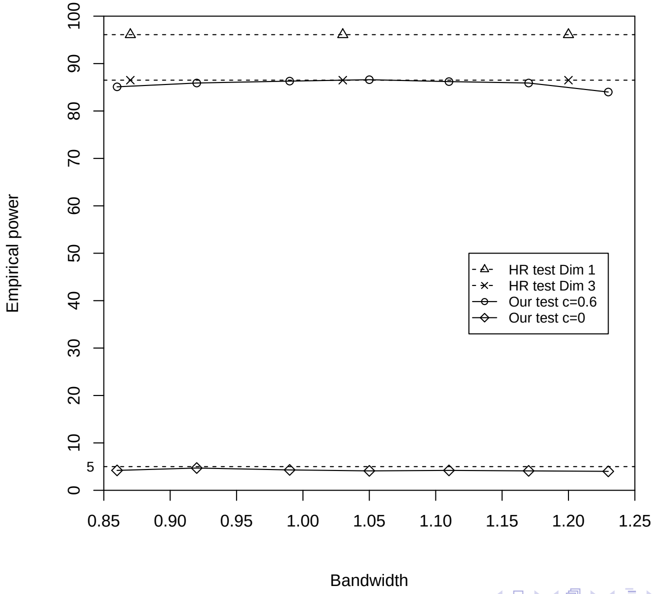
		$p=3$				$p=5$			
		$c_h=0.6$	$c_h=0.8$	$c_h=1.0$	CT	$c_h=0.6$	$c_h=0.8$	$c_h=1.0$	CT
$n=100$	$\widehat{\tau}_n^2(\gamma)$	19.1	21.9	23.1	8.6	18.2	21.7	23.1	8.1
	$\widehat{v}_n^2$	23.4	26.0	27.1	8.6	22.6	25.6	27.2	8.1
$n=200$	$\widehat{\tau}_n^2(\gamma)$	34.4	40.5	47.9	7.6	30.6	39.5	42.8	6.9
	$\widehat{v}_n^2$	39.6	44.3	47.9	7.6	36.6	42.9	47.4	6.9

Table: Percent. reject under the quadratic altern; $\gamma^{(p)} = (1, 0, \cdots)$.

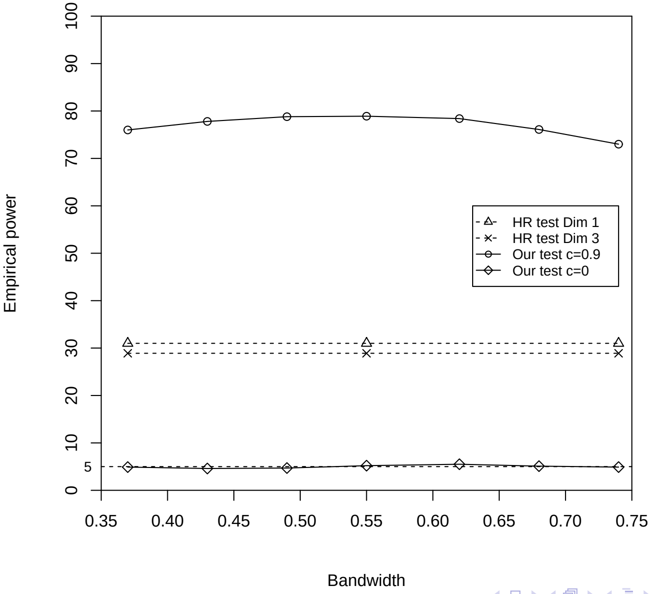
		$p=3$				$p=5$			
		$c_h=0.6$	$c_h=0.8$	$c_h=1.0$	CT	$c_h=0.6$	$c_h=0.8$	$c_h=1.0$	CT
$n=100$	$\widehat{\tau}_n^2(\gamma)$	12.9	11.9	10.6	8.6	18.4	13.6	12.4	8.1
	$\widehat{v}_n^2$	19.4	16.0	14.6	8.6	25.4	19.8	16.5	8.1
$n=200$	$\widehat{\tau}_n^2(\gamma)$	25.7	25.4	24.7	7.6	37.7	36.2	30.8	6.9
	$\widehat{v}_n^2$	30.4	29.4	28.0	7.6	42.1	39.1	33.7	6.9

Table: Percentage of rejections under the quadratic alternative; $\gamma^{(p)} = (0, 1, 0, \cdots)$.

Testing the linear model vs. quadratic alternative



Testing the linear model vs. cubic alternative



Testing the quadratic model vs. cubic alternative

- The simulated model

$$Y = a + \int_0^1 b(t)X(t)\,dt + \int_0^1 \int_0^1 h(s,t)X(s)X(t)\,ds\,dt + \delta_c(X) + U^0$$

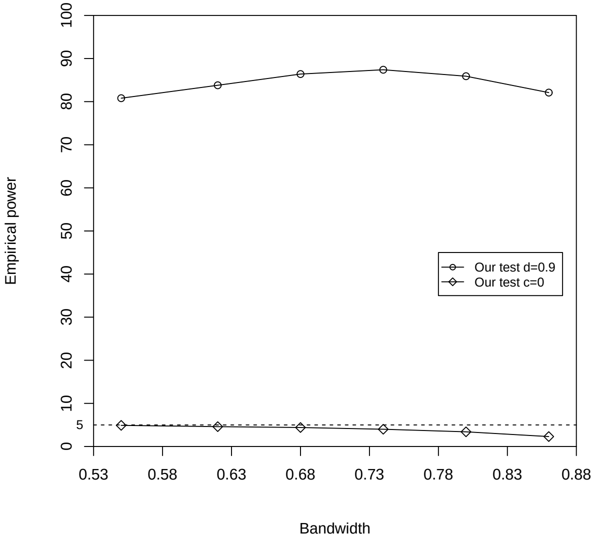
with $b(t) = 1$ for all $t \in [0,1]$, and $h(s,t) = 0.6$ for all $s, t \in [0,1]$

- The cubic alternative

$$\delta_c(X) = d\left(\int_0^1 \int_0^1 \int_0^1 X(s)X(t)X(z)\,ds\,dt\,dz - \int_0^1 X(t)dt\right)$$

where $d = 0$ under the null and $d = 0.9$ under the alternative.

Testing the quadratic model vs. cubic alternative



Real data application

- Use Tecator data set. The task is to predict the fat content of a meat sample on the basis of its near infrared absorbance spectrum.
- Test linear functional model and quadratic functional model
- Both models are rejected

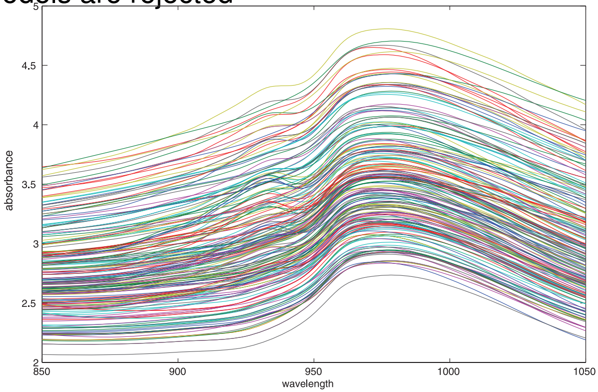


FIG. 1. *Sample of 204 absorbance spectra for meat specimens.*

		Linear model				Quadratic model			
	h	0.18	0.30	0.44	0.59	0.18	0.30	0.44	0.59
$p = 2$	$m = 1$	0.5	0.4	0.2	0.6	2.4	1.4	1.6	3.3
	$m = 2$	0.2	0.0	0.0	0.3	0.6	0.3	0.0	0.7
	$m = 3$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$p = 3$	$m = 1$	0.0	0.0	0.2	0.2	0.0	0.1	0.1	0.0
	$m = 2$	0.0	0.0	0.0	0.1	0.2	0.0	0.1	0.0
	$m = 3$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 1. p-values (in percentages) obtained by applying the new test to the Tecator data set.

Outline

1 Introduction

2 The principle: scalar responses case

3 The test statistic for scalar responses

4 Extending the principle: functional responses

5 Numerical illustrations

6 Conclusions

Conclusions and possible esxtensions

- A smoothing based goodness-of-fit test with hybrid (functional and finite-dimension) data
  - Test the effect on the covariates on the response
  - Check parametric functional regression (linear, quadratic, quantile,...)
- The asymptotic critical values are standard normal; wild bootstrap is needed for small sample sizes
- It detects nonparametric alternatives
- Mild conditions on the law of the covariates
- The response/errors could be heteroscedastic
- Several extensions are possible
  - dependent observations (once CLT and concentration inequalities for $U-$processes are available)
  - testing significance of covariates in functional nonparametric regression

Some References (1/2)

- CARDOT, H., GOIA, P., AND SARDA, P. (2004). Testing for no effect in functional linear regression models, some computational approaches. *Communications in Statistics - Simulation and Computation* **33**, 179–199.
- CHEN, D., HALL, P., MÜLLER, H.G. (2011). Single and multiple index functional regression models with nonparametric link. *Annals of Statistics* **39**, 1720–1747
- CRAMBES, C., KNEIP, A., AND SARDA, P. (2008). Smoothing splines estimators for functional linear regression. *Annals of Statistics* **37**, 35–72.
- DELSOL, L., FERRATY, F., AND VIEU, P. (2011). Structural test in regression on functional variables. *Journal of Multivariate Analysis* **102**, 422–447.
- FERRATY, F. (Ed.) (2011). *Recent Advances in Functional Data Analysis and Related Topics.* Springer-Verlag Berlin Heidelberg.
- GARCÍA-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]
- HALL, P., AND HOROWITZ, J.L. (2007). Methodology and convergence rates for functional linear regression. *Annals of Statistics* **35**, 70–91.

Some References (2/2)

- HORVÀTH, L., AND REEDER, R. (2011). A test of significance in functional quadratic regression. arXiv:1105.0014v1 [math.ST].
- MAJOR, P. (2006). An estimate on the supremum of a nice class of stochastic integrals and U-statistics. *Probability Theory and Related Fields* **134**, 489–537.
- MÜLLER, H.G. AND STADTMÜLLER, U. (2005). Generalized functional linear models. *Annals of Statistics* **33**, 774–805.
- RAMSAY, J., AND SILVERMAN, B.W. (2005). *Functional Data Analysis* (2nd ed.). Springer-Verlag, New York.
- YAO, F., AND MÜLLER, H.G. (2010). Functional quadratic regression. *Biometrika* **97**, 49–64.
- ZHENG, J.X. (1996). A consistent test of functional form via nonparametric estimation techniques. *J. Econometrics* **75**, 263–289.

Manuscripts available on arXiv

- PATILEA, V., SÁNCHEZ-SELLERO, C., AND SAUMARD, M. (2012). Projection-based nonparametric goodness-of-fit testing with functional covariates. arXiv:1205.5578 [math.ST]
- PATILEA, V., SÁNCHEZ-SELLERO, C., AND SAUMARD, M. (2012). Nonparametric testing for no-effect with functional responses and functional covariates. arXiv:1209.2085 [math.ST]