Statistica Sinica Preprint No: SS-2015-0337.R1	
Title	Nonparametric model checks of single-index assumptions
Manuscript ID	SS-2015-0337.R1
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202015.0337
Complete List of Authors	Samuel Maistre and Valentin Patilea
Corresponding Author	Samuel Maistre
E-mail	samuelmaistre@gmail.com
Notice: Accepted version subject to English editing.	

Statistica Sinica

NONPARAMETRIC MODEL CHECKS OF SINGLE-INDEX ASSUMPTIONS

Samuel Maistre[1,2,3] and Valentin Patilea[1]

[1] *CREST (Ensai)* [2] *Université de Lyon* [3] *Université de Strasbourg*

Abstract: Semiparametric single-index assumptions are widely used dimension reduction approaches that represent a convenient compromise between the parametric and fully nonparametric models for regressions or conditional laws. In a mean regression setup, the SIM assumption means that the conditional expectation of the response given the vector of covariates is the same as the conditional expectation of the response given a scalar projection of the covariate vector. In a conditional distribution modeling, under the SIM assumption the conditional law of a response given the covariate vector coincides with the conditional law given a linear combination of the covariates. In this paper, a novel kernel-based approach for testing SIM assumptions is introduced. The covariate vector needs not have a density and only the index estimated under the SIM assumption is used in kernel smoothing. Hence the effect of high-dimensional covariates is mitigated while asymptotic normality of the test statistic is obtained. Irrespective of the fixed dimension of the covariate vector, the new test detects local alternatives approaching the null hypothesis slower than $n^{-1/2}h^{-1/4}$, where h is the bandwidth used to build the test statistic and n is the sample size. A wild bootstrap procedure is proposed for finite sample corrections of the asymptotic critical values. The small sample performance of our test is illustrated through simulations.

Key words and phrases: Single-index regression, Conditional law, Lack-of-fit test, Kernel smoothing, $U-$statistics.

1. Introduction

Semiparametric single index models (SIM) are widely used tools for statistical modeling. The paradigm of such models is based on the assumption that the information contained in a vector of conditioning random variables is equivalent, in some sense, to the information contained in some index, that is usually a linear combination of the vector components. This assumption underlies most of the parametric models including covariates, but allows for more general semiparametric modeling. The most common semiparametric SIM are those for the mean regression. See Ichimura (1993), Härdle et al. (1993), see also Horowitz (2009) for a recent review. In such models, the index and the conditional mean given the index are unknown. SIM for quantile regression were considered recently, see Kong and Xia (2012). A more restrictive, but still of significant interest, class of models is obtained by imposing the single-index paradigm to the conditional distribution of response variable given a vector of covariates. In these cases the index and the conditional law of the response given the index are unknown. The famous Cox proportional hazard model, see Cox (1972), is a particular case of SIM for conditional laws. See Delecroix et al. (2003), Hall and Yao (2005), Chiang and Huang (2012) for more general situations.

The large amount of interest for SIM could be explained by the fact that the single-index assumption is very often the first intermediate step from a parametric framework towards a fully nonparametric paradigm. Then an important question is whether this dimension reduction compromise is good enough to capture the relevant information contained in the covariate vector. A possible way to answer is to build a statistical test of the single-index assumption against general alternatives. Several tests of the goodness-of-fit of single-index mean regression models have proposed in the literature. See Fan and Li (1996), Xia et al. (2004), Stute and Zhu (2005), Chen and Van Keilegom (2009), Xia (2009), Escanciano and Song (2010) and the references therein. The problem of testing SIM models for conditional distribution in full generality seems open.

In this paper we propose a new and quite simple kernel smoothing-based approach for testing single-index assumptions. We focus on mean regression and conditional law models. The approach is inspired by the remark that, up to some error in covariates, the single-index assumption check could be interpreted as a test of significance in nonparametric regression. Next, the single-index assumption could be conveniently reformulated as an equivalent unconditional moment condition. Finally, a kernel based test statistic could be used to test the unconditional moment condition. The smoothing based goodness-of-fit test approach allows to make the error in covariates negligible and thus to obtain a pivotal asymptotic law under the null hypothesis. Only the index estimated under the SIM assumption is used in kernel smoothing and this fact mitigates the effect of high-dimensional covariates. The covariate vector needs not have a density, discrete covariables are allowed, as long as the parameter defining the index is estimated sufficiently accurate. For the SIM considered below, one could expect the $O_{\mathbb{P}}(n^{-1/2})$ rate for estimators of the index parameter, even when some covariates are discrete. See, for instance, Xia (2006) and Chiang and Huang (2012). Meanwhile the asymptotical critical values are given by the quantiles of the normal law. Irrespective of the fixed dimension of the covariate vector, the new test detects local alternatives approaching the null hypothesis slower than $n^{-1/2}h^{-1/4}$, where h is the bandwidth used to build the test statistic.

The paper is organized as follows. In Section 2, we recall general considerations on single-index models. In Section 3, we present a general approach of testing nonparametric significance and in Section 4 we apply it to single-index hypotheses for mean regression as well as for conditional law. In Section 5 we introduce a wild bootstrap procedure to correct the asymptotic critical values with small samples and illustrate the performance of our test by an empirical study. The proofs are relegated to the appendix and some additional technical results are provided in a Supplementary Material.

2. Single-index models

Let $Y \in \mathbb{R}^d$, $d \geq 1$, denote the random response vector and let $X \in \mathbb{R}^p$, $p \geq 1$, be the random column vector of covariates. The data consists of independent copies of $(Y', X')'$. For mean regression the single-index assumption means that there exists a column parameter vector $\beta_0 \in \mathbb{R}^p$ such that

$$\mathbb{E}[Y \mid X] = \mathbb{E}[Y \mid X'\beta_0]. \tag{2.1}$$

Only the direction given by β_0 is identified, so that an additional identification condition accompanies the model assumption. The usual conditions are $\|\beta_0\| = 1$ and an arbitrary component is set positive, or an arbitrary component is set equal to 1. The scalar product $X'\beta_0$ is the so-called index. The direction β_0 and the nonparametric univariate regression $\mathbb{E}[Y \mid X'\beta_0]$ have to be estimated. See Hristache et al. (2001), Delecroix et al. (2006), Horowitz (2009), Xia et al. (2011) and the references therein for a panorama of the existing estimation procedures.

When applying the single-index paradigm to conditional laws of Y given X, one supposes

$$Y \perp X \mid X'\beta_0. \tag{2.2}$$

In this case the direction defined by β_0 and the conditional law of the response Y given the index $X'\beta_0$ have to be estimated. See Delecroix et al. (2003), Hall and Yao (2005) and Chiang and Huang (2012) for the available estimation approaches.

There are several model check approaches for SIM for mean regressions. Xia et al. (2004) use an empirical process-based statistic related to that of Stute et al. (1998). Fan and Li (1996) use a kernel smoothing-based quadratic form to a wide range of situations, including single-index. Our test statistics are somehow close to that of Fan and Li (1996). Chen and Van Keilegom (2009) use an empirical likelihood test for multi-dimensional Y in a parametric or semiparametric modeling, the single-index mean regression is presented as a particular case but without getting into the details.

In this paper we propose an alternative model check approach that is able to detect any departure from the single-index assumption, both for mean regressions and conditional law models. It is inspired by a general approach for testing nonparametric significance that is presented in the following section.

3. A general approach for testing nonparametric significance

Let $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ be a Hilbert space. The examples we have in mind corresponds to $\mathcal{H} = \mathbb{R}^d$, for some $d \geq 1$, or $\mathcal{H} = L^2[0,1]$. Consider $U \in \mathcal{H}$, $Z \in \mathbb{R}^q$ and $W \in \mathbb{R}^r$ and let (U_i, Z_i, W_i), $1 \leq i \leq n$ denote an independent sample of U, Z and W. Consider the problem of testing the equality

$$\mathbb{E}[U \mid Z, W] = 0 \qquad \text{a.s.} \tag{3.1}$$

against the nonparametric alternative $\mathbb{P}(\mathbb{E}[U \mid Z, W] = 0) < 1$. Several testing procedures against nonparametric alternatives, including the single-index assumptions check, lead to this type of problem. For instance, in the case of a mean regression single-index model, U could be proportional to the error term, Z could be the index $X'\beta_0$, while W should carry the remaining information contained in the covariate vector X. At this stage, one could of course use X instead of (Z, W). However, this split of the covariate vector prepares a major feature of our approach: kernel smoothing will only involve Z, for which a density is required, while no smoothing on W is used, and no density for W is required.

Let us introduce some notation: for any real-valued, univariate or multivariate function l, let $\mathcal{F}[l]$ denote the Fourier Transform of l. Let K be a multivariate kernel on $\mathbb{R}^q$ such that $\mathcal{F}[K] > 0$ and let $\phi(s) = \exp(-\|s\|^2/2)$, $\forall s \in \mathbb{R}^r$. The kernel K could be a multiplicative kernel with univariate kernels with positive Fourier Transform. Many univariate kernels have this property: gaussian, triangle, Student, logistic, etc.

Let $w(\cdot) > 0$ be some weight function, and for any $h > 0$, let

$$I(h) = \mathbb{E}\left[\langle U_1, U_2\rangle_{\mathcal{H}}\, w(Z_1)w(Z_2)h^{-q}K((Z_1 - Z_2)/h)\phi(W_1 - W_2)\right] \tag{3.2}$$

Our approach is based on the following key property: for any $h > 0$,

$$\mathbb{E}[U \mid Z, W] = 0 \;\; a.s. \quad \Leftrightarrow \quad I(h) = 0. \tag{3.3}$$

A formal proof of this statement is provided in Lemma A.1 the Appendix. To check condition (3.1) the idea is to build a sample based approximation of $I(h)$, to suitably normalize it and to let h to decrease to zero. See also Lavergne et al. (2015), Section 2.2, for a related approach. A convenient choice of $w(\cdot)$ could avoid handling denominators close to zero.

In many situations the sample of $Uw(Z)$ is not observed and has to be estimated inside the model. Then, an estimate of $I(h)$ is given by the $U-$statistic

$$I_n(h) = \frac{1}{n(n-1)h^q} \sum_{1\leq i\neq j\leq n} \left\langle \widehat{U_i w(Z_i)},\, \widehat{U_j w(Z_j)} \right\rangle_{\mathcal{H}} K_{ij}(h)\, \phi_{ij},$$

where

$$K_{ij}(h) = K((Z_i - Z_j)/h), \qquad \phi_{ij} = \exp(-\|W_i - W_j\|^2/2).$$

The variance of $I_n(h)$ could be estimated by $n^{-2}h^{-q}v_n^2(h)$ where

$$v_n^2(h) = \frac{2}{n(n-1)h^q} \sum_{1\leq i\neq j\leq n} \left\langle \widehat{U_i w(Z_i)},\, \widehat{U_j w(Z_j)} \right\rangle_{\mathcal{H}}^2 K_{ij}^2(h)\, \phi_{ij}^2.$$

Then the test statistic is

$$T_n = nh^{q/2}\frac{I_n(h)}{v_n(h)}.$$

Under mild technical conditions and provided that h converges to zero at a suitable rate, T_n converges in law to a standard normal distribution when condition (3.1) holds true. Hence, a one-sided test with standard normal critical values could be defined; see Lavergne et al. (2015). One could also show T_n tends to infinity in probability if $\mathbb{P}(\mathbb{E}[U \mid Z, W] = 0) < 1$. Making h to decrease to zero at suitable rate allows to render negligible the effect of the errors $\widehat{U_i w(Z_i)} - U_i w(Z_i)$. On the other hand, the test detects alternative hypotheses like

$$H_{1n}: \ \mathbb{E}(U \mid Z, W) = r_n \delta(Z, W), \quad n \geq 1, \tag{3.4}$$

as soon as $r_n^2 n h^{q/2} \to \infty$.

4. Single-index assumptions checks

In this section we extend the approach described in section (3) to test single-index assumptions like (2.1) and (2.2). In this case, with the notation from section 3,

$$q = 1, \ \ r = p - 1, \ \ Z = Z(\beta) \quad \text{and} \quad W = W(\beta)$$

where, for $\beta \in \mathcal{B} \subset \mathbb{R}^p$,

$$Z(\beta) = X'\beta \qquad \text{and} \qquad W(\beta) = X'\mathbf{A}(\beta)$$

with $\mathbf{A}(\beta)$ a $p \times (p-1)$ matrix with real entries such that the $p \times p$ matrix $(\beta \ \mathbf{A}(\beta))$ is orthogonal. The orthogonality is not necessary, invertibility suffices, but orthogonality is expected to lead to better finite sample properties for the tests. We assume that $\mathcal{B}$ is a set of vectors β that satisfy one of the two model identification conditions mentioned in Section 2 above.

An additional challenge will come from the fact that the sample of the covariates Z and W depend on estimator of the single-index direction β_0. Again, the kernel smoothing and a suitable choice of h allows to render this effect negligible and preserve a pivotal asymptotic law under the null hypothesis.

4.1. Testing SIM for mean regression

To simplify the presentation, we focus on the case of a univariate response $(d = 1)$. At the end, it will be quite clear how the case $d > 1$ could be handled. Let $\mathcal{H} = \mathbb{R}$, $Uw(Z) = U(\beta_0)w(Z;\beta_0)$ where

$$U(\beta)w(Z;\beta) = \{Y - \mathbb{E}[Y \mid Z(\beta)]\} f_\beta(Z(\beta)).$$

Here $f_\beta(\cdot)$ denotes the density of $X'\beta$ that is supposed to exist, at least for some β. Let

$$\widehat{U_i w(Z_i)}(\beta) = \frac{1}{n-1} \sum_{k \neq i} (Y_i - Y_k) \frac{1}{g} L_{ik}(\beta, g), \tag{4.1}$$

where L is a univariate kernel, $L_{ik}(\beta, g) = L((Z_i(\beta) - Z_k(\beta))/g)$ and g is a bandwidth converging to zero at some suitable rate described in a following section. Let $\hat{\beta}$ be some estimator of the index direction and consider

$$I_n^{\{m\}}(\hat{\beta}) = \frac{1}{n\,(n-1)\,h} \sum_{1 \le i \ne j \le n} \widehat{U_i w(Z_i)}(\hat{\beta}) \widehat{U_j w(Z_j)}(\hat{\beta}) K_{ij}(\hat{\beta}, h) \phi(W_i(\hat{\beta}) - W_j(\hat{\beta})),$$

where $K_{ij}(\hat{\beta}, h) = K((Z_i(\hat{\beta}) - Z_j(\hat{\beta}))/h)$. The variance of $I_n^{\{m\}}(\hat{\beta})$ could be estimated by

$$\hat{\omega}_n^{\{m\}}(\hat{\beta})^2 = \frac{2h^{-1}}{n\,(n-1)} \sum_{i \ne j} \left| \widehat{U_i w(Z_i)}(\hat{\beta}) \widehat{U_j w(Z_j)}(\hat{\beta}) \right|^2 K_{ij}^2(\hat{\beta}, h) \phi^2(W_i(\hat{\beta}) - W_j(\hat{\beta})).$$

The test statistic is then

$$T_n^{\{m\}}(\hat{\beta}) = nh^{1/2} \frac{I_n^{\{m\}}(\hat{\beta})}{\hat{\omega}_n^{\{m\}}(\hat{\beta})}.$$

Let us point out that only smoothing with the $X_i'\hat{\beta}$'s is required in order to build this statistic.

In section 4.3 we show that whenever $\hat{\beta} - \beta^* = O_{\mathbb{P}}(n^{-1/2})$, for some β^* that could depend on n,

$$I_n^{\{m\}}(\hat{\beta}) - I_n^{\{m\}}(\beta^*) = o_{\mathbb{P}}(I_n^{\{m\}}(\beta^*)) \quad \text{and} \quad \hat{\omega}_n^{\{m\}}(\hat{\beta}) - \hat{\omega}_n^{\{m\}}(\beta^*) = o_{\mathbb{P}}(\hat{\omega}_n^{\{m\}}(\beta^*)), \tag{4.2}$$

provided some mild technical conditions hold true. Under the null hypothesis (2.1) one expects to have $\beta^* = \beta_0$. Then $T_n^{\{m\}}(\hat{\beta})$ has an asymptotic standard normal law under the single-index assumption as soon as $T_n^{\{m\}}(\beta_0)$ is standard normal asymptotically distributed. Conditions for guaranteeing the asymptotic normality of $T_n^{\{m\}}(\beta_0)$ when (2.1) holds true are given in Lavergne et al. (2015).

When the SIM (2.1) is wrong, even asymptotically, in general a semiparametric estimator $\hat{\beta}$ converges at the rate $O_{\mathbb{P}}(n^{-1/2})$ to some *pseudo-true* value $\beta^* \in \mathcal{B}$ that depends on the estimation procedure; see Delecroix et al. (1999) for some general theoretical results. Then the asymptotic equivalence (4.2) and the results of Lavergne et al. (2015) imply that a test based on $T_n^{\{m\}}(\hat{\beta})$ would reject the null hypothesis with probability tending to 1, in just the way the test based on $T_n^{\{m\}}(\beta^*)$ would do. The case of Pitman alternatives requires a longer investigation since the conclusion depends on the estimation method and the properties of the deviation from the null hypothesis. Such a detailed investigation is beyond our present scope. Let us, however, briefly comment on the case where the index β_0 was estimated through a semiparametric least-squares procedure as introduced by Ichimura (1993). To estimate the mean regression single-index model, one defines a family of univariate regression functions $r_\beta(s) = \mathbb{E}[Y \mid Z(\beta) = s]$, $s \in \mathbb{R}$, $\beta \in \mathcal{B}$. The single-index model is valid if $\mathbb{E}(Y \mid X) = r_{\beta_0}(Z(\beta_0))$. Then β_0 is solution of the minimization problem

$$\min_{\beta\in\mathcal{B}} \mathbb{E}\left[\{Y - r_\beta(Z(\beta))\}^2\right].$$

In the semiparametric least-squares approach, one supposes that β_0 is the unique solution of this minimization problem, and defines $\hat\beta$ the semiparametric estimator as a minimum of a sample counterpart of $\mathbb{E}\left[\{Y - r_\beta(Z(\beta))\}^2\right]$ where the regression $r_\beta(\cdot)$ is replaced by a nonparametric estimate, for instance obtained by kernel smoothing. Now, consider the sequence of alternatives

$$Y = m(Z(\beta_0)) + r_n\delta(Z(\beta_0), W(\beta_0)) + \varepsilon, \quad n \geq 1,$$

where $\mathbb{E}(\varepsilon \mid X) = 0$ a.s., $m(\cdot)$ is some univariate function, $\delta(\cdot)$ is some function of the covariate vector, and r_n, $n \geq 1$ is some bounded sequence of real numbers. For illustration, assume that $\delta(X)$ satisfies the orthogonality conditions

$$\begin{aligned} \mathbb{E}[\delta(Z(\beta_0), W(\beta_0)) \mid Z(\beta_0)] &= 0 \\ \mathbb{E}[\delta(Z(\beta_0), W(\beta_0))m'(Z(\beta_0))\{X - \mathbb{E}[X \mid Z(\beta_0)\}] &= 0, \end{aligned} \tag{4.3}$$

where $m'(\cdot)$ denotes the derivative of the univariate function $m(\cdot)$. Then,

$$\forall n, \qquad \beta_0 = \arg\min_{\beta\in\mathcal{B}} \mathbb{E}\left[\{Y - r_\beta(Z(\beta))\}^2\right], \tag{4.4}$$

and, under suitable technical conditions and using the same type of arguments as used to study the rate of convergence of $\hat\beta$ when SIM is correct, it can be proved that $\hat\beta - \beta_0 = O_{\mathbb{P}}(n^{-1/2})$. (See the Appendix for a justification of the property (4.4) and a comment on the condition (4.3).) Hence, by our results in the sequel, $T_n^{\{m\}}(\hat\beta)$ behaves asymptotically as $T_n^{\{m\}}(\beta_0)$. More precisely, if $r_n^2 nh^{1/2} \to C$ with $0 \leq C < \infty$, $nh^{1/2}T_n^{\{m\}}(\hat\beta) \to \mathcal{N}(C\mu, 1)$ in law, where

$$\mu = \mathbb{E}\left[\int \delta(s, W_1)\, \delta(s, W_2)\, f_{\beta_0}^2(s)\, \pi(s \mid W_1)\, \pi(s \mid W_2)\, \phi(W_1 - W_2)\, ds\right] > 0,$$

where W_1, W_2 are independent copies of $W(\beta_0)$ and $\pi(\cdot \mid W(\beta_0) = w)$ denotes the density of $Z(\beta_0)$ given $W(\beta_0) = w$. If $r_n^2 n h^{1/2} \to \infty$, $nh^{1/2} T_n^{\{m\}}(\hat{\beta}) \to \infty$ in probability. See Theorem 1 in Lavergne et al. (2015). Thus our test detects such local alternatives as soon as $r_n^2 n h^{1/2} \to \infty$.

4.2. Testing SIM for the conditional law

To test the single-index condition (2.2) for the conditional law of an univariate Y given X, let $\mathcal{H} = L^2[0,1]$ and for each $t \in [0,1]$ and $\beta \in \mathcal{B}$, let

$$U(t;\beta)w(Z;\beta) = \{\mathbf{1}\{\Phi(Y) \le t\} - \mathbb{P}[\Phi(Y) \le t \mid Z(\beta)]\}\, f_\beta(Z(\beta)),$$

where Φ is some distribution function on the real line, for instance a normal distribution function or the marginal distribution function of Y. In the latter case, the marginal distribution could be estimated by the empirical distribution function. The case of multivariate Y could be also considered after obvious modifications and for the sake of simplicity will not be investigated herein.

For $\beta \in \mathcal{B}$ and $t \in [0,1]$, let

$$\widehat{U_i w(Z_i)}(\beta)(t) = \frac{1}{n-1} \sum_{k \ne i} (\mathbf{1}\{\Phi(Y_i) \le t\} - \mathbf{1}\{\Phi(Y_k) \le t\}) \frac{1}{g} L_{ik}(\beta, g). \qquad (4.5)$$

Next, define

$$I_n^{\{l\}}(\beta) = \frac{h^{-1}}{n\,(n-1)} \sum_{1 \le i \ne j \le n} \left\langle \widehat{U_i w(Z_i)}(\beta),\ \widehat{U_j w(Z_j)}(\beta) \right\rangle_{L^2} K_{ij}(\beta, h) \phi(W_i(\beta) - W_j(\beta)),$$

where for any $u(\cdot)$ and $v(\cdot)$ squared integrable functions defined on $[0,1]$, $\langle u, v\rangle_{L^2} = \int_0^1 u(t)v(t)dt$. The variance of $I_n^{\{l\}}(\beta)$ could be estimated by $n^{-2}h^{-1}\hat{\omega}_n^{\{l\}}(\beta)^2$ where

$$\hat{\omega}_n^{\{l\}}(\beta)^2 = \frac{2h^{-1}}{n\,(n-1)} \sum_{i \ne j} \left\langle \widehat{U_i w(Z_i)}(\beta),\ \widehat{U_j w(Z_j)}(\beta) \right\rangle_{L^2}^2 K_{ij}^2(\beta, h) \phi^2(W_i(\beta) - W_j(\beta)). \qquad (4.6)$$

Given $\widetilde{\beta}$ some estimator of β_0, the test statistic is

$$T_n^{\{l\}}(\tilde{\beta}) = nh^{1/2} \frac{I_n^{\{l\}}(\tilde{\beta})}{\hat{\omega}_n^{\{l\}}(\tilde{\beta})}.$$

In section 4.3 we show that, under suitable technical conditions, whenever $\widetilde{\beta} - \beta^\sharp = O_{\mathbb{P}}(n^{-1/2})$,

$$I_n^{\{l\}}(\widetilde{\beta}) - I_n^{\{l\}}(\beta^\sharp) = o_{\mathbb{P}}(I_n^{\{l\}}(\beta^\sharp)) \ \text{ and } \ \hat{\omega}_n^{\{l\}}(\widetilde{\beta}) - \hat{\omega}_n^{\{l\}}(\beta^\sharp) = o_{\mathbb{P}}(\hat{\omega}_n^{\{l\}}(\beta^\sharp)). \qquad (4.7)$$

Under the null hypothesis (2.2) one expects to have $\beta^\sharp = \beta_0$. Then the asymptotic normality of $T_n^{\{l\}}(\beta_0)$, proved in Proposition 2 below, implies that the asymptotic one-sided test based on $T_n^{\{l\}}(\tilde\beta)$ has standard normal critical values.

If the single-index assumption fails and the alternative is fixed, like in the case of mean regression, one expects $\widetilde\beta - \beta^\sharp = O_{\mathbb{P}}(n^{-1/2})$ for some *pseudo-true* value $\beta^\sharp \in \mathcal{B}$ that depends on the estimation procedure. Then $T_n^{\{l\}}(\tilde\beta)$ would detect the alternative with probability tending to 1. Concerning the case of local alternatives, let $\delta(X,t)$ and $r_n \to 0$ such that

$$\mathbb{P}[\Phi(Y) \le t \mid X] = \mathbb{P}[\Phi(Y) \le t \mid X'\beta_0] + r_n\delta(X,t), \qquad t \in [0,1],$$

is a conditional distribution function. Suitable orthogonality conditions for the function $\delta(X,t)$ would yield $\widetilde\beta - \beta_0 = O_{\mathbb{P}}(n^{-1/2})$ and hence $T_n^{\{l\}}(\widetilde\beta)$ allows to detect such local alternatives as soon as $r_n^2 nh^{1/2} \to \infty$.

4.3. Asymptotic results

Here we formally state the results that guarantee the asymptotic equivalences (4.2) and (4.7). Let $\widehat{U_i w(Z_i)}(\beta)$ be defined as in (4.1) or (4.5). Let $I_n(\beta)$ (resp. $\hat\omega_n(\beta)^2$) denote any of $I_n^{\{m\}}(\beta)$ or $I_n^{\{l\}}(\beta)$ (resp. $\hat\omega_n^{\{m\}}(\beta)^2$ or $\hat\omega_n^{\{l\}}(\beta)^2$).

Proposition 1. *Suppose the conditions in Assumption 1 in the Appendix are met. If β_n is an estimator such that $\beta_n - \bar\beta = O_{\mathbb{P}}(n^{-1/2})$, then*

$$I_n(\beta_n) - I_n(\bar\beta) = o_{\mathbb{P}}(I_n(\bar\beta)) \quad \text{and} \quad \hat\omega_n(\beta_n) - \hat\omega_n(\bar\beta) = o_{\mathbb{P}}(\hat\omega_n(\bar\beta)).$$

In particular, if $nh^{1/2}I_n(\bar\beta)/\hat\omega_n(\bar\beta)$ has a standard normal law under the single-index null hypothesis, the test defined by $nh^{1/2}I_n(\beta_n)/\hat\omega_n(\beta_n)$ has asymptotic standard normal critical values. Moreover, the test given by $nh^{1/2}I_n(\beta_n)/\hat\omega_n(\beta_n)$ detects local alternatives approaching the null hypothesis slower than $n^{-1/2}h^{-1/4}$ as soon as the test given by $nh^{1/2}I_n(\bar\beta)/\hat\omega_n(\bar\beta)$ does.

An usual question raised when using smoothing-based test statistics is the choice of the bandwidth h. The statistical literature includes some contributions on data-driven rate-optimal choices of the bandwidth for parametric mean-regression, see, for instance, Horowitz and Spokoiny (2001), Guerre and Lavergne (2005). Their extension to the present framework would require a theoretical investigation beyond the scope of this paper. Nevertheless, in the empirical section we provide some evidence on the effect of the bandwidth h.

As mentioned above, the asymptotic behavior of $nh^{1/2}I_n(\bar{\beta})/\hat{\omega}_n(\bar{\beta})$ in the case of mean regression was investigated by Lavergne et al. (2015). The case where $U_i w(Z_i)(\beta)$ is a stochastic process seems less explored and is hence considered in the following proposition.

Proposition 2. *Suppose the conditions in Assumption 1 in the Appendix are met and the null hypothesis (2.2) holds true. Consider β_n such that $\beta_n - \beta_0 = O_{\mathbb{P}}(n^{-1/2})$. Then $nh^{1/2}I_n^{\{l\}}(\beta_n)/\hat{\omega}_n^{\{l\}}(\beta_n) \to \mathcal{N}(0,1)$ in law under H_0, and*

$$\left[\hat{\omega}_n^{\{l\}}(\beta_0)\right]^2 \to \left[\omega^{\{l\}}(\beta_0)\right]^2 = 2\int K^2(u)\,du \times \int\int \Gamma^2(s,t)\,ds\,dt$$
$$\times\,\mathbb{E}\left[\int f_{\beta_0}^4(z)\,\phi^2\left(W_1(\beta_0) - W_2(\beta_0)\right)\pi_{\beta_0}\left(z \mid W_1(\beta_0)\right)\pi_{\beta_0}\left(z \mid W_2(\beta_0)\right)dz\right],$$

in probability, where $\pi_{\beta_0}(\cdot \mid w)$ is the conditional density of $Z(\beta_0)$ knowing that $W(\beta_0) = w$, and for $t, s \in [0,1]$,

$$\Gamma(s,t) = \mathbb{E}\left[\epsilon(s)\,\epsilon(t)\right], \qquad \epsilon(t) = \mathbf{1}\{\Phi(Y) \le t\} - \mathbb{P}[\Phi(Y) \le t \mid X'\beta_0].$$

5. Empirical evidence

For conditional mean, we simulate the data using the model

$$Y_i \;=\; X_i'\beta + 4\exp\{-\left(X_i'\beta\right)^2\} + \delta\|X_i\| + \sigma\varepsilon_i, \quad 1 \le i \le n, \tag{5.1}$$

where $X_i = (X_{i1}, \cdots, X_{ip})'$ follows a standard normal p-variate law, and the true value of the parameter is $\beta_0 = (1,1,0,\ldots,0)'$ and $\sigma = 0.3$. For the ε_i, we consider two cases: a standard univariate normal law independent of the X_i's and a centered log-normal heteroscedastic setup

$$\varepsilon_i = \left(\log\mathcal{N}(0,1) - \sqrt{\mathrm{e}}\right) \times \sqrt{(1+X_{i2}^2)/2}.$$

The model (5.1) was proposed by Xia et al. (2004) and investigated only in the case of a homoscedastic noise.

To estimate the parameter β, we consider the approach of Delecroix et al. (2006), that is we define

$$\tilde{\beta} = \arg\min_{\beta:\beta_1>0}\sum_{i=1}^n\left(Y_i - \frac{\sum_{k\neq i} Y_k \tilde{L}_{ik}(\beta)}{\sum_{k\neq i}\tilde{L}_{ik}(\beta)}\right)^2, \tag{5.2}$$

where

$$\tilde{L}_{ik}(\beta) = L\left((\widetilde{X}_i - \widetilde{X}_k)'\beta\right), \quad \widetilde{X}_i = \frac{n^{1/2} X_i}{\sqrt{\sum_{k=1}^n (X_k - \overline{X})^2}} \quad \text{and} \quad \overline{X} = n^{-1}\sum_{k=1}^n X_k.$$

The estimator is defined as $\hat{\beta} = \tilde{\beta}/\|\tilde{\beta}\|$ and the bandwidth g is equal to $\|\tilde{\beta}\|^{-1}$.

To improve the asymptotic critical values with small samples, we propose the following bootstrap procedure:

(i) Define

$$\hat{m}_i = \frac{\sum_{k\neq i} Y_k \tilde{L}_{ik}(\tilde{\beta})}{\sum_{k\neq i} \tilde{L}_{ik}(\tilde{\beta})}.$$

(ii) For $b \in \{1, \ldots, B\}$

- **(a)** let $Y_i^{*,b} = \hat{m}_i + \eta_i (Y_i - \hat{m}_i)$, where the η_is are independent variables with the two-point distribution

$$\mathbb{P}[\eta_i = (1-\sqrt{5})/2] = (5+\sqrt{5})/10\ , \ \mathbb{P}[\eta_i = (1+\sqrt{5})/2] = (5-\sqrt{5})/10.$$

- **(b)** define

$$\tilde{\beta}^{*,b} = \arg\min_{\beta:\beta_1>0} \sum_{i=1}^n \left(Y_i^{*,b} - \frac{\sum_{k\neq i} Y_k^{*,b} \tilde{L}_{ik}(\beta)}{\sum_{k\neq i} \tilde{L}_{ik}(\beta)} \right)^2$$

  and $\hat{\beta}^{*,b} = \tilde{\beta}^{*,b}/\|\tilde{\beta}^{*,b}\|$ and $g^{*,b} = \|\tilde{\beta}^{*,b}\|^{-1}$.

(iii) Define $T_n^{\{m\}*,b}$ as $T_n^{\{m\}}$ where the Y_is are replaced by the $Y_i^{*,b}$s, $\hat{\beta}$ by $\hat{\beta}^{*,b}$, and the bandwidth g by $g^{*,b}$. The bandwidth h does not change. Repeat Step (iii) B times. Compute the empirical quantiles of $T_n^{\{m\}*,b}$ using the B bootstrap values.

The justification of this bootstrap procedure has been made in Theorem 2 of Lavergne et al. (2015) in the case of significance testing. The same type of arguments, combined with the $\sqrt{n}$−convergence of $\tilde{\beta}^{*,b} - \tilde{\beta}$ given the original sample, could be used to justify the bootstrap procedure proposed above. However, presenting the detailed arguments is beyond our present scope.

In our experiments the bootstrap correction is used with $B = 499$ bootstrap samples. The level is fixed as $\alpha = 10\%$. We considered $L(\cdot) = K(\cdot)$ and equal to the standard gaussian density. With this choice no numerical problem occurred due to denominators too close to zero and therefore we did not consider any trimming in equation (5.2) and its bootstrap version.

First, we investigated the influence of the bandwidth h on the level. Several bandwidths were considered, that is $h = c \times n^{-2/9}$ with $c \in \{2^{k/2} : k = \pm 2, \pm 1, 0\}$. The results on empirical rejection rates for the model defined in equation (5.1) with $\delta = 0$ (that is on the null hypothesis) and $n = 100$ are presented in Figure 1. The results are based on 500 replications, with homoscedastic noise and $p = 2$, $p = 4$, and with heteroscedastic log-normal noise and $p = 4$. The normal critical values are quite inaccurate, while the bootstrap correction seems to overreject slightly, particularly for a large bandwidth h. For the third case with heteroscedastic noise, the test rejects too often. However, for larger sample sizes, this drawback is mitigated, as could be seen from the fourth plot in Figure 1 where we considered the heteroscedastic noise with $p = 4$ and $n = 200$.

Next, we studied the behavior of our statistic under the null hypothesis (500 replications) and several alternatives (250 replications) defined by some positive value of δ. We only considered the statistics with bandwidth factor $c = 1$ and compared it to the statistics introduced by Fan and Li (1996) with gaussian kernels, g_{FL} the estimated bandwidth and $h_{FL} = n^{-2/9}$ and Xia et al. (2004) based on $CVT_{\mathcal{D}}$ and $CVT^*_{\mathcal{D}}$. A referee brought our attention to the contribution in Xia (2009), where a general methodology for nonparametric check of semiparametric models is proposed. Xia's idea is to regress the residuals of the fitted model on single indices. In this way, in the case of single-index models, one searches for the second projective direction and build a test based in this. This idea is related, but quite different from ours. While in Xia's approach, that we call *Xia's SCV*, one searches for the worse direction to reveal the possible misspecification of the model, in some sense, our procedure averages over all projective directions. Hence, depending on the alternative, one would expect that in some cases Xia's procedure could outperform our test, while in others we could perform better. We also investigated the approach proposed in Xia (2009) using the code available at `http://www.stat.nus.edu.sg/~staxyc/SCV.m` with the same residuals as those we used in the other methods. The results are presented in Figure 2. Xia et al. (2004)'s test performs better for $p = 2$, while our test shows better performance for $p = 4$. It appears that the greater p is, the more advantageous it will be to use our test statistic. The alternative we considered being defined by a radial function, the performance of the test proposed by Xia (2009) is rather poor.

For conditional law, we simulate the data using the mixture model

$$Y_i \;=\; (1-\delta)\; X_i'\beta + \delta \left\| X_i \right\| + \varepsilon_i, \qquad 1 \le i \le n, \tag{5.3}$$

where $X_i = (X_{i1}, X_{i2})'$ has a standard normal bivariate law, $\varepsilon_i \sim \mathcal{N}(0,\, 0.25)$ and $\beta_0 = (1,1)'/\sqrt{2}$. We apply the test statistic $I_n^{\{l\}}$ based on the quantities $\widehat{U_i w(Z_i)}(\beta)(t)$ introduced in equation (4.5). Here the events $\{\Phi(Y_i) \le t\}$ are defined with $\Phi(\cdot)$ equal to the empirical distribution function of the Y_i's. In this case an event $\{\Phi(Y_i) \le t\}$ is determined by the rank of Y_i in the sample of the response variable. To estimate the index parameter β we use the approach of Chiang and Huang (2012). More precisely, let

$$\check{\beta} = \arg\min_{\beta:\beta_1>0} \sum_{i=1}^{n}\sum_{j=1}^{m} \Big(\mathbf{1}\left(Y_i \le Y_{(j)}\right) - \hat{G}_{i,\beta}\left(Y_{(j)}\right) \Big)^2,$$

where $Y_{(1)} \le \cdots \le Y_{(m)}$ are the m distinct ordered observations of Y,

$$\hat{G}_{i,\beta}\left(y\right) = \frac{\sum_{k\neq i} \mathbf{1}\left(Y_k \le y\right) \tilde{L}_{ik}\left(\beta\right)}{\sum_{k\neq i} \tilde{L}_{ik}\left(\beta\right)}$$

but $\hat{G}_{i,\beta}$ is set to 0 whenever its denominator is null.

For the bootstrap procedure of this test statistic, consider the following steps:

(i) Define

$$\tilde{G}_{i,\beta}\left(y\right) = \frac{\sum_{k=1}^{n} \mathbf{1}\left(Y_k \le y\right) \tilde{L}_{ik}\left(\beta\right)}{\sum_{k=1}^{n} \tilde{L}_{ik}\left(\beta\right)}$$

(ii) For $b \in \{1, \ldots,\, B\}$

- **(a)** let $Y_i^{*,b} = \tilde{G}_{i,\beta}^{-1}\left(\nu_i\right)$ where the ν_i's are independent variables $\mathcal{U}\left([0,\,1]\right)$ and $\tilde{G}_{i,\beta}^{-1}\left(u\right) = \inf\left\{ y : \tilde{G}_{i,\beta}\left(y\right) \ge u \right\}$.
- **(b)**

$$\check{\beta}^{*,b} = \arg\min_{\beta:\beta_1>0} \sum_{i=1}^{n}\sum_{j=1}^{m} \left(\mathbf{1}\left(Y_i^{*,b} \le Y_{(j)}^{*,b}\right) - \hat{G}_{i,\beta}\left(Y_{(j)}^{*,b}\right) \right)^2,$$

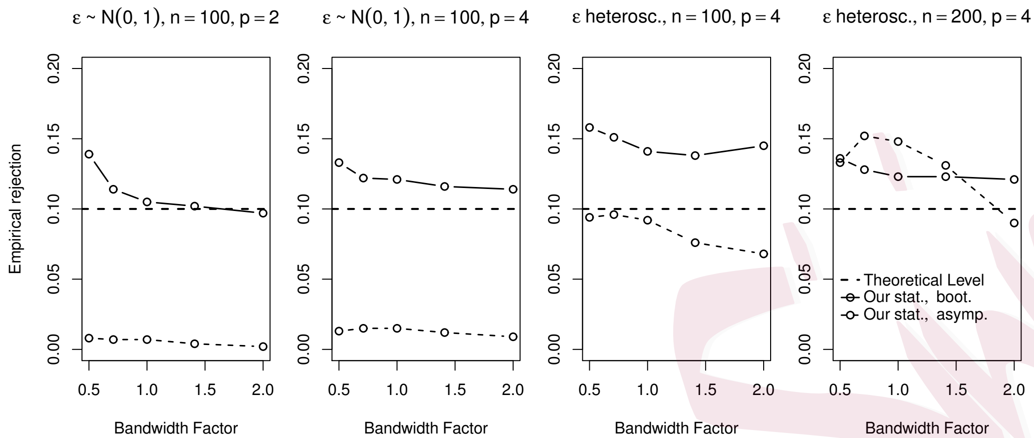

Figure 1: Empirical rejections under H_0 as a function of the bandwidth. ε heterosc. means $\varepsilon \sim (\log \mathcal{N}(0,1) - e) \times \sqrt{(1+X_2^2)/2}$.

(iii) Define $T_n^{\{l\}*,b}$ as $T_n^{\{l\}}$ where the Y_i's are replaced by the $Y_i^{*,b}$'s and $\breve{\beta}$ by $\breve{\beta}^{*,b}$. Repeat Step (iii) B times. Compute the empirical quantiles of $T_n^{\{l\}*,b}$ using the B bootstrap values.

Notice that $\tilde{G}_i(\cdot;\beta)$ is nothing but an estimation of the conditional c.d.f. of Y knowing $X'\beta = X_i'\beta$. Therefore, $Y_i^{*,b}$ is drawn by inverse transform sampling. This is equivalent to drawing the $Y_i^{*,b}$ from the law

$$P_{n,i}^* = \sum_{k=1}^n w_{ik}\delta_{Y_k}, \qquad w_{ik} = \frac{\tilde{L}_{ik}(\beta)}{\sum_{k=1}^n \tilde{L}_{ik}(\beta)},$$

where δ_{Y_k} denotes the Dirac mass at Y_k.

For different sample size $n = 50$, 100 or 200, we study the influence of bandwidth h on empirical rejection under H_0 on Figure 3, where $h = c \times n^{-2/9}$ with $c \in \{2^{k/2} : k = \pm 2, \pm 1, 0\}$, with 1000 replications and 199 bootstrap steps. The normal critical values produce empirical rejections close to zero (not reported), but the bootstrap correction is quite effective and results in a good level.

We also investigate the empirical rejection rate for different values of the proportion δ in the model (5.3). The results are presented on Figure 4. We used 1000 replications for $\delta = 0$, 500 replications otherwise, and 199 bootstrap steps.

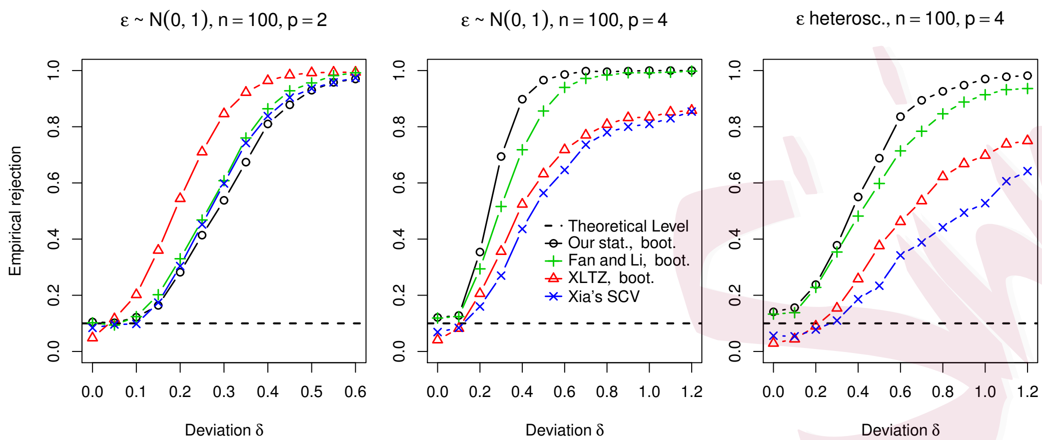

Figure 2: Power curves for model (5.1), $n = 100$

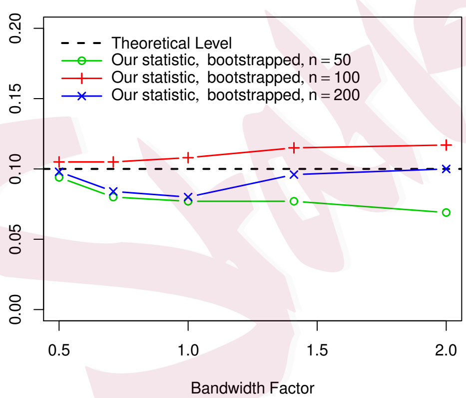

Figure 3: Empirical rejections under H_0, with bootstrap critical values and bandwidth $h = c \times n^{-2/9}$ with varying c.

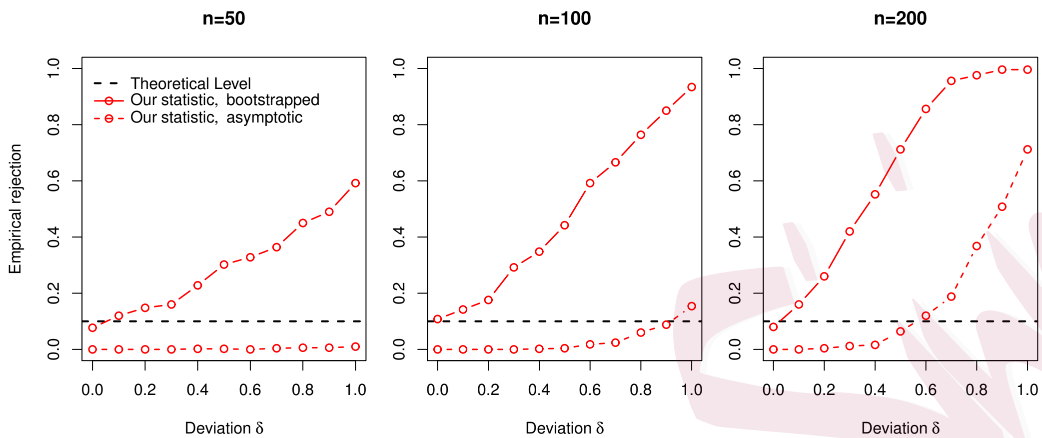

Figure 4: Power curves for model (5.3).

5.1 Real data application

The proposed approach for testing the single-index hypothesis for conditional law is applied to check the goodness-of-fit of the model on air quality data proposed by Chiang and Huang (2012). We try to explain the average ozone concentration from average wind speed (wind), maximum daily temperature (temp), and solar radiation level (solar) on 111 daily observations from May to September 1973 in New York metropolitan area. Based on a lasso procedure with variables wind, temp, solar, wind^2, temp^2, solar^2, wind$\times$temp, wind$\times$solar, and temp$\times$solar, Chiang and Huang (2012) remove variables solar, temp^2 and wind$\times$solar. Each of the 6 variables is standardized and the index estimation is given in Table 1. We use the test statistic with quartic kernel $L(u) = K(u) = (15/16)(1-u^2)^2\mathbf{1}(|u| \leq 1)$ and $h = n^{-2/9}$. It yields a $p-$value of 0.74 based on 199 bootstrap samples. Thus, on the basis of our test of the single-index assumption on the conditional law, the model proposed by Chiang and Huang (2012) is not rejected by the data.

6. Conclusions and furthers extensions

We have constructed new smoothing-based test procedures for SIM hypotheses for mean regression and for conditional law. Smoothing is only used on the estimated index, and the corresponding test statistics are asymptotically stan-

	Estimate
wind	0.434
temp	-0.639
wind^2	-0.320
solar^2	0.240
wind×temp	-0.473
temp×solar	-0.141

Table 1: Index estimate for the conditional law of ozone concentration.

dard normal. A quite effective wild bootstrap procedure allows to correct the critical values with small samples. Our approach also applies to the case of multivariate responses. See Chen and Van Keilegom (2009) for more general situations with multivariate responses where our test methodology applies. Moreover, our statistics directly generalize to test multiple index against fully nonparametric alternatives. It suffice to consider the general methodology presented in section 3 with q equal to the number of indices. Some other possible extensions that would require additional, though quite straightforward, investigation are the goodness-of-fit checks of index quantile regressions, see Kong and Xia (2012), and the functional index models, see Chen et al. (2011). Moreover, as suggested by a referee, like in the papers of Fan and Li (1996) and Xia (2009), the approach we introduce herein could be extended to a larger class of semiparametric models, as for instance, partially linear mean regression and varying coefficient models. Such extensions are left for future work.

7. Appendix: assumptions and proofs

7.1 Proof of the identity (3.3)

Let $I(h)$ be defined as in equation (3.2) with $\phi(s) = \exp(-\|s\|^2/2)$, $\forall s \in \mathbb{R}^r$. The following lemma justifies the identity (3.3). The proof is given in the Supplementary Material. See also Lavergne et al. (2015) for a related result.

Lemma A.1. *Let (U_1, Z_1, W_1) and (U_2, Z_2, W_2) be two independent draws of (U, Z, W). Let $K(\cdot)$ be a bounded, even, integrable function with positive, integrable Fourier transform. Assume $\mathbb{E}(\|Uw(Z)\|^2_{\mathcal{H}}) < \infty$, Then for any $h > 0$,*

$$\mathbb{E}\left[U \mid Z, W\right] = 0 \ \ a.s. \Leftrightarrow I(h) = 0.$$

Moreover, if $\mathbb{P}\left(\mathbb{E}\left[U \mid Z, W\right] = 0\right) < 1$, then $\inf_{h\in(0,1]} I(h) > 0$.

7.2 Assumptions and proofs of Propositions 1 and 2

In the following, let $\mathcal{H}$ be the real line or $L^2[0,1]$, the Hilbert space of squared integrable functions defined on $[0,1]$. The parameter set satisfies one of the following conditions: $\mathcal{B} \subset \{1\} \times \mathbb{R}^{d-1}$ or $\mathcal{B} \subset \{\|\gamma\|^{-1}\gamma : \gamma \in \mathbb{R}^d, \gamma_1 > 0\}$. For an observation $(Y_i, X_i')'$, $Y_i \in \mathbb{R}$ and $X_i \in \mathbb{R}^p$, let $Y_i(t) \equiv Y_i$ or $Y_i(t) = \mathbf{1}\{Y_i \leq \Phi^{-1}(t)\}$, and for any β in the parameter set $\mathcal{B} \subset \mathbb{R}^p$, let $r_i(t;\beta) = \mathbb{E}[Y_i(t) \mid Z_i(\beta)]$, $t \in [0,1]$. Thus, $Y_i(\cdot) \in \mathcal{H}$.

Let $(\epsilon_i(\cdot), X_i')'$, $1 \leq i \leq n$, be random variables such that $\epsilon_i(\cdot) \in \mathcal{H}$ and $X_i \in \mathbb{R}^p$. Let $\bar\beta$ be some element in the parameter set $\mathcal{B}$. Consider $r_i(t;\bar\beta)$ that depends only on $Z_i(\bar\beta) = X_i'\bar\beta$ and $\delta(X_i,t)$ be such that $\mathbb{E}[\delta(X_i,t) \mid Z_i(\bar\beta)] = 0$, $t \in [0,1]$. Define

$$Y_{ni}(t) = r_i(t;\bar\beta) + r_n\delta(X_i,t) + \epsilon_i(t), \qquad t \in [0,1], 1 \leq i \leq n,$$

where r_n, $n \geq 1$, is some bounded sequence of real numbers. In particular that means $\mathbb{E}[Y_{ni}(\cdot) \mid Z_i(\bar\beta)] = r_i(\cdot;\bar\beta)$. A null sequence (r_n) corresponds to the null hypothesis, while a sequence tending to zero corresponds to local alternatives.

Assumption 1. a) The random variables $(\epsilon_i(\cdot), X_i')'$, $1 \leq i \leq n$, are independent copies of $\epsilon(\cdot) \in \mathcal{H}$ and $X \in \mathbb{R}^p$. Moreover, $X'\bar\beta$ admits a bounded density $f_{\bar\beta}$.

b) $\mathbb{E}[\exp(\rho\|X_i\|)] < \infty$ for some $\rho > 0$ and $\mathbb{E}[\sup_t |r_i(t;\bar\beta) + \epsilon_i(t)|^a] < \infty$ for some $a > 8$. Moreover, $\mathbb{E}(\|\epsilon_i(\cdot)\|^2_{\mathcal{H}} \mid X_i)$ is bounded.

c) For any $t \in [0,1]$, the map $v \mapsto \mathbb{E}[Y_{ni}(t) \mid Z_i(\bar\beta) = v]$ is twice differentiable. The second derivative $r_i''(\cdot;\bar\beta)$ is uniformly Lipschitz (that is the Lipschitz constant independent of t) and uniformly bounded, while the first derivative satisfies $\mathbb{E}[\|r_i'(\cdot;\bar\beta)\|^4_{\mathcal{H}}] < \infty$.

d) The function $f_{\bar\beta}(\cdot)$ is Lipschitz.

e) The function $\delta(\cdot,\cdot)$ is bounded.

f) The kernels K and L are symmetric integrable functions, differentiable except at most a finite set of points and L' is Lipschitz continuous. Moreover, $\int_{\mathbb{R}} |L(t)|dt = \int |K(t)|dt = 1$ and $\int_{\mathbb{R}}(|L'(t)| + |K'(t)|)dt < \infty$. The map $v \mapsto |L'(v)|/v$ is bounded in a neighborhood of the origin, $v^2K(v) \to 0$ if $v \to \infty$, and $\int v^2\{|L(v)| + |K(v)|\}dv < \infty$. Moreover, the Fourier Transform $\mathcal{F}[K]$ is positive on the real line and integrable.

g) The bandwidths satisfy the conditions $g, h \to 0$, $h/g^2 \to 0$, $nh^{1/2}g^4 \to 0$,

$r_n^2 nh^{1/2} \to \infty$. Moreover, $g = n^{-\gamma}$ with $\gamma \in (1/5, 1/4)$ and thus $nh^2/\log^3 n \to \infty$.

Our conditions are related to those used by Lavergne et al. (2015) to derive the asymptotic results and justify the validity of their bootstrap. Like in Xia et al. (2004), we only require a density for the index $X'\bar{\beta}$, while Fan and Li (1996) and Xia (2009) impose a density for the covariate vector. Higher order kernels with positive Fourier transform could be also used, if more regularity on the regression function is imposed. This would result in reducing the bias of the nonparametric estimation of $Uw(Z)$, and hence relaxing the condition $nh^{1/2}g^4 \to 0$.

Proof of Proposition 1. First let us remark that for any $a_n \to \infty$,

$$\mathbb{P}\left(\max_{1\leq i\leq n} \|X_i\| > a_n \log n\right) \to 0 \quad \text{and} \quad \mathbb{P}\left(\|\beta_n - \bar{\beta}\| > a_n n^{-1/2}\right) \to 0. \tag{7.1}$$

Moreover, at least for β in a fixed but small enough neighborhood of $\bar{\beta}$, the matrix $\mathbf{A}(\beta)$ could be built such that the norm of each of the $p-1$ columns of $\mathbf{A}(\beta) - \mathbf{A}(\bar{\beta})$ is bounded by $c\|\beta - \bar{\beta}\|$ with c a constant independent of β. Indeed, one could consider and $(p-1)-$dimension orthogonal basis in the space of vectors orthogonal on $\bar{\beta}$. With a small enough neighborhood of $\bar{\beta}$, this orthogonal basis could be completed by any β close to $\bar{\beta}$ to form a basis in $\mathbb{R}^p$. Next one could use the Gram-Schmidt procedure to orthonormalize the basis, starting from β. Finally, use the last $p-1$ orthogonal vectors in the basis to build $\mathbf{A}(\beta)$. By construction, the norm of any column of $\mathbf{A}(\beta) - \mathbf{A}(\bar{\beta})$ is bounded by $c\|\beta - \bar{\beta}\|$ for some c depending only on $\bar{\beta}$ and its neighborhood and the initial $(p-1)-$dimension orthogonal basis orthogonal on $\bar{\beta}$. The details are provided in the Supplementary Material. All these facts show that we can reduce the parameter set to $\mathcal{B}_n$, $n \geq 1$, a sequence of balls centered at $\bar{\beta}$ of radius converging to zero. Consider the set of elementary events

$$\mathcal{E}_n = \left\{ \max_{1\leq i\leq n} \sup_{\beta\in\mathcal{B}_n} [\|Z_i(\beta) - Z_i(\bar{\beta})\| + \|W_i(\beta) - W_i(\bar{\beta})\|] \leq b_n \right\}, \tag{7.2}$$

where b_n is a sequence such that $b_n \downarrow 0$. The equation (7.1) indicates that the sequences $\mathcal{B}_n$ and b_n could be taken such that the radius of $\mathcal{B}_n$ converges to zero slower than $n^{-1/2}$ and faster than b_n, and $b_n n^{1/2}/\log n \to \infty$. Then $\mathbb{P}(\beta_n \in \mathcal{B}_n) \to 1$ and $\mathbb{P}(\mathcal{E}_n^c)$ decreases to zero faster than any negative power of n. Hence, in the following it will suffices to prove the statements on the set $\{\beta_n \in \mathcal{B}_n\} \cap \mathcal{E}_n$.

We will focus on $I_n(\beta_n)$ since the arguments for $\hat{\omega}(\beta_n)$ are similar and much simpler. Hereafter, by abuse, we write $Y_i(t)$ instead of $Y_{ni}(t)$ even when $r_n \neq 0$. To prove that $I_n(\beta_n) - I_n(\bar{\beta}) = o_{\mathbb{P}}(I_n(\bar{\beta}))$ we will show below that $I_n(\beta_n) - I_n(\bar{\beta}) = o_{\mathbb{P}}(n^{-1}h^{-1/2} + r_n^2)$. This shows that $I_n(\beta_n)$ is negligible compared to $I_n(\bar{\beta})$ both on the null and alternative hypotheses. Indeed, under the null hypothesis, $r_n \equiv 0$, $\bar{\beta} = \beta_0$ and $nh^{1/2} I_n(\beta_0)$ is asymptotically centered normal distributed, while on the alternative the $I_n(\bar{\beta})$ is driven by a term of order r_n^2.

In the following C, C',... denote constants that may have different values from line to line. Let us simplify notation and write $\widehat{V}_i(\beta) = \widehat{U_i w(Z_i)}(\beta)$ and

$$L_{ij}(\beta) = L_{ij}(\beta, g),\ K_{ij}(\beta) = K_{ij}(\beta, h),\ \phi_{ij}(\beta) = \phi(W_i(\beta) - W_j(\beta)). \tag{7.3}$$

Then,

$$\begin{aligned}
I_n(\beta) - I_n(\bar{\beta}) &= \frac{1}{n(n-1)h} \sum_{i \neq j} \Big[\Big\langle \widehat{V}_i(\beta),\ \widehat{V}_j(\beta) \Big\rangle_{\mathcal{H}} - \Big\langle \widehat{V}_i(\bar{\beta}),\ \widehat{V}_j(\bar{\beta}) \Big\rangle_{\mathcal{H}} \Big] K_{ij}(\bar{\beta}) \phi_{ij}(\bar{\beta}) \\
&+ \frac{1}{n(n-1)h} \sum_{i \neq j} \Big\langle \widehat{V}_i(\bar{\beta}),\ \widehat{V}_j(\bar{\beta}) \Big\rangle_{\mathcal{H}} \left[K_{ij}(\beta)\phi_{ij}(\beta) - K_{ij}(\bar{\beta})\phi_{ij}(\bar{\beta}) \right] \\
&+ \frac{1}{n(n-1)h} \sum_{i \neq j} \Big[\Big\langle \widehat{V}_i(\beta),\ \widehat{V}_j(\beta) \Big\rangle_{\mathcal{H}} - \Big\langle \widehat{V}_i(\bar{\beta}),\ \widehat{V}_j(\bar{\beta}) \Big\rangle_{\mathcal{H}} \Big] \\
&\qquad \times \left[K_{ij}(\beta)\phi_{ij}(\beta) - K_{ij}(\bar{\beta})\phi_{ij}(\bar{\beta}) \right] \\
&= D_{n1}(\beta) + D_{n2}(\beta) + D_{n3}(\beta).
\end{aligned}$$

We only investigate the rates of D_{n1} and D_{n2} since D_{n3} is uniformly smaller. Let

$$\begin{aligned}
D_{n1}(\beta) &= \frac{2}{n(n-1)h} \sum_{i \neq j} \Big\langle \widehat{V}_i(\beta) - \widehat{V}_i(\bar{\beta}),\ \widehat{V}_j(\bar{\beta}) \Big\rangle_{\mathcal{H}} K_{ij}(\bar{\beta}) \phi_{ij}(\bar{\beta}) \\
&+ \frac{1}{n(n-1)h} \sum_{i \neq j} \Big\langle \widehat{V}_i(\beta) - \widehat{V}_i(\bar{\beta}),\ \widehat{V}_j(\beta) - \widehat{V}_j(\bar{\beta}) \Big\rangle_{\mathcal{H}} K_{ij}(\bar{\beta}) \phi_{ij}(\bar{\beta}) \\
&= 2D_{n11}(\beta) + D_{n12}(\beta).
\end{aligned}$$

Moreover,

$$
\begin{aligned}
\widehat{V}_i(\bar\beta)(t) &= \frac{1}{n-1}\sum_{k\neq i}[Y_i(t)-Y_k(t)]\frac{1}{g}L_{ik}(\bar\beta)\\
&= [Y_i(t)-r_i(t;\bar\beta)]f_{\bar\beta}(X_i'\bar\beta) + [Y_i(t)-r_i(t;\bar\beta)]\left[\frac{1}{n-1}\sum_{k\neq i}\frac{1}{g}L_{ik}(\bar\beta) - f_{\bar\beta}(X_i'\bar\beta)\right]\\
&+ \frac{1}{n-1}\sum_{k\neq i}\{r_i(t;\bar\beta)-r_k(t;\bar\beta)\}\frac{1}{g}L_{ik}(\bar\beta) - \frac{1}{n-1}\sum_{k\neq i}\{Y_k(t)-r_k(t;\bar\beta)\}\frac{1}{g}L_{ik}(\bar\beta)\\
&= [Y_i(t)-r_i(t;\bar\beta)]f_{\bar\beta}(X_i'\bar\beta) + [Y_i(t)-r_i(t;\bar\beta)]R_{1,ni} + R_{2,ni}(t) - R_{3,ni}(t),
\end{aligned}
$$

and, by Lemma 2 $\sup_i|R_{1,ni}| = O_{\mathbb{P}}(g+n^{-1/2}g^{-1/2}\log^{1/2} n)$, and, Lemma 1 yields

$$
\sup_{1\leq i\leq n}\sup_{t\in[0,1]}|R_{3,ni}(t)| = O_{\mathbb{P}}(n^{-1/2}g^{-1/2}\log^{1/2} n).
$$

A representation of $R_{2,ni}(t)$ is provided in Lemma 3. On the other hand,

$$
\widehat{V}_i(\beta)(t) - \widehat{V}_i(\bar\beta)(t) \;=\; \frac{1}{n-1}\sum_{k\neq i}[Y_i(t)-Y_k(t)]\left[g^{-1}L_{ik}(\beta) - g^{-1}L_{ik}(\bar\beta)\right].
$$

Uniform bounds for D_{n1}.

The rate of D_{n11}. Since $Y_i(t) = r_i(t;\bar\beta) + r_n\delta(X_i,t) + \epsilon_i(t)$, with $\mathbb{E}[\epsilon_i(t)\mid X_i] = 0$, we have $D_{n11}(\beta) = D_{n111}(\beta) + R_{n11}(\beta)$ with

$$
\begin{aligned}
D_{n111}(\beta) &= \frac{1}{n(n-1)^2h}\sum_{i\neq j\neq k}\left\langle Y_i(\cdot)-Y_k(\cdot),\ Y_j(\cdot)-r_j(\cdot;\bar\beta)\right\rangle_{\mathcal{H}} f_{\bar\beta}(X_j'\bar\beta)\\
&\qquad\times\left[g^{-1}L_{ik}(\beta)-g^{-1}L_{ik}(\bar\beta)\right]K_{ij}(\bar\beta)\phi_{ij}(\bar\beta)\\
&= \frac{1}{n(n-1)^2h}\sum_{i\neq j\neq k}\langle Y_i(\cdot)-Y_k(\cdot),\ \epsilon_j(\cdot)\rangle_{\mathcal{H}} f_{\bar\beta}(X_j'\bar\beta)\\
&\qquad\times\left[g^{-1}L_{ik}(\beta)-g^{-1}L_{ik}(\bar\beta)\right]K_{ij}(\bar\beta)\phi_{ij}(\bar\beta),
\end{aligned}
$$

and $R_{n11}(\beta) = D_{n11}(\beta) - D_{n111}(\beta)$. We decompose

$$\begin{aligned} D_{n111}(\beta) &= \frac{1}{n(n-1)^2h}\sum_{i\neq j\neq k} \langle \epsilon_i(\cdot) - \epsilon_k(\cdot),\ \epsilon_j(\cdot)\rangle_{\mathcal{H}} f_{\bar\beta}(X_j'\bar\beta) \\ &\qquad\qquad \times \left[g^{-1}L_{ik}(\beta) - g^{-1}L_{ik}(\bar\beta)\right] K_{ij}(\bar\beta)\phi_{ij}(\bar\beta) \\ &\quad + \frac{1}{n(n-1)^2h}\sum_{i\neq j\neq k} \left\langle r_i(\cdot;\bar\beta) - r_k(\cdot;\bar\beta),\ \epsilon_j(\cdot)\right\rangle_{\mathcal{H}} f_{\bar\beta}(X_j'\bar\beta) \\ &\qquad\qquad \times \left[g^{-1}L_{ik}(\beta) - g^{-1}L_{ik}(\bar\beta)\right] K_{ij}(\bar\beta)\phi_{ij}(\bar\beta) \\ &\quad + \frac{r_n}{n(n-1)^2h}\sum_{i\neq j\neq k} \langle \delta(X_i,t) - \delta(X_k,t),\ \epsilon_j(\cdot)\rangle_{\mathcal{H}} f_{\bar\beta}(X_j'\bar\beta) \\ &\qquad\qquad \times \left[g^{-1}L_{ik}(\beta) - g^{-1}L_{ik}(\bar\beta)\right] K_{ij}(\bar\beta)\phi_{ij}(\bar\beta) \\ &= D_{n1111}(\beta) + D_{n1112}(\beta) + r_n D_{n1113}(\beta). \end{aligned}$$

The quantity $ghD_{n1111}(\beta)$ could be decomposed in a sum of degenerate $U-$process of order 3 and another one of order 2 indexed by β. To bound them we use the maximal inequality of Sherman (1994). Since $nh^2, ng^4 \to \infty$, deduce that the degenerate $U-$process of order 3 is of uniform rate

$$n^{-3/2}O_{\mathbb{P}}(h^{\alpha/2}\{b_n^2 g^{-1}\}^{\alpha/2}) = gh \times o_{\mathbb{P}}(n^{-1}h^{-1/2}),$$

over any sequence of balls centered at $\bar\beta$ with radius decreasing to zero faster than b_n, where b_n is a sequence such that $b_n n^{1/2}/\log n \to \infty$ and α could be a number in the interval $(0,1)$ arbitrarily close to 1. The details on how the maximal inequality of Sherman (1994) applies are provided below for deriving the uniform rate of D_{n21}. To bound the right-hand side term in that maximal inequality we use the fact that $\mathbb{E}(\|\epsilon\|_{\mathcal{H}}^2 \mid X)$ and $f_{\bar\beta}(X'\bar\beta)$ are bounded and the uniform bounds (S2.9), (S2.7) and (S2.8) from Lemma 4. Using very similar arguments, the degenerate $U-$process of order 2 in the decomposition of $ghD_{n1111}(\beta)$ could be shown to be of uniform rate

$$n^{-1}O_{\mathbb{P}}(h^{\alpha/2}\{b_n^2 g^{-1}\}^{\alpha/2}) = gh \times o_{\mathbb{P}}(n^{-1}h^{-1/2})$$

provided that $nh^2, ng^4 \to \infty$ and α is sufficiently close to 1. Next, for $ngD_{n1112}(\beta)$, that is centered, use the Hoeffding decomposition and the regularity of the function $v \mapsto \mathbb{E}[Y(t) \mid Z(\bar\beta) = v]$. For the degenerate $U-$processes of order 3 and 2 in the Hoeffding decomposition of $D_{n1112}(\beta)$ we apply the maximal inequality of

Sherman (1994) as previously. Deduce the respective uniform rates over $\mathcal{B}_n$

$$g^2 n^{-3/2} O_{\mathbb{P}}(h^{\alpha/2}\{b_n^2 g^{-1}\}^{\alpha/2}) = gh \times o_{\mathbb{P}}(n^{-1}h^{-1/2}),$$

$$g^2 n^{-1} O_{\mathbb{P}}(h^{\alpha/2}\{b_n^2 g^{-1}\}^{\alpha/2}) = gh \times o_{\mathbb{P}}(n^{-1}h^{-1/2}).$$

It remains the $U-$process of order 1. Using again the bounds from Lemma 4, deduce the uniform rate over $\mathcal{B}_n$

$$g^2 n^{-1/2} O_{\mathbb{P}}(h^{\alpha}\{b_n^2 g^{-1}\}^{\alpha/2}) = gh \times o_{\mathbb{P}}(n^{-1}h^{-1/2}).$$

Deduce $D_{n1112}(\beta_n) = o_{\mathbb{P}}(n^{-1}h^{-1/2})$. For $ghD_{n1113}(\beta)$ the arguments are similar, but without the g^2 factor, and yield the uniform rate

$$n^{-1/2} O_{\mathbb{P}}(h^{\alpha}\{b_n^2 g^{-1}\}^{\alpha/2}) = gh \times o_{\mathbb{P}}(n^{-1}h^{-1/2}) = o_{\mathbb{P}}(n^{-1/2}h^{-1/4}),$$

if $nh^2, ng^4 \to \infty$ and α is close to 1. Then $D_{n111}(\beta_n) = o_{\mathbb{P}}(n^{-1}h^{-1/2} + r_n^2)$.

For $R_{n11}(\beta)$ we can write

$$\begin{aligned} R_{n11}(\beta) \quad = \quad & \frac{1}{n(n-1)^2 h} \sum_{i \neq j} \left\langle Y_i(\cdot) - Y_j(\cdot),\ Y_j(\cdot) - r_j(t;\bar\beta) \right\rangle_{\mathcal{H}} f_{\bar\beta}(X_j'\bar\beta) \\ & \qquad \times \left[g^{-1} L_{ij}(\beta) - g^{-1} L_{ij}(\bar\beta) \right] K_{ij}(\bar\beta)\phi_{ij}(\bar\beta) \\ & + \frac{1}{n(n-1)^2 h} \sum_{i\neq j, i\neq k} \left\langle Y_i(\cdot) - Y_k(\cdot),\ Y_j(\cdot) - r_j(t;\bar\beta) \right\rangle_{\mathcal{H}} R_{1,nj} \\ & \qquad \times \left[g^{-1} L_{ik}(\beta) - g^{-1} L_{ik}(\bar\beta) \right] K_{ij}(\bar\beta)\phi_{ij}(\bar\beta) \\ & + \frac{1}{n(n-1)^2 h} \sum_{i\neq j, i\neq k} \langle Y_i(\cdot) - Y_k(\cdot),\ R_{2,nj}(\cdot) + R_{3,nj}(\cdot) \rangle_{\mathcal{H}} \\ & \qquad \times \left[g^{-1} L_{ik}(\beta) - g^{-1} L_{ik}(\bar\beta) \right] K_{ij}(\bar\beta)\phi_{ij}(\bar\beta) \\ = \quad & R_{n111}(\beta) + R_{n112}(\beta) + R_{n113}(\beta). \end{aligned}$$

We only investigate $R_{n111}(\beta)$, the terms $R_{n112}(\beta)$ and $R_{n113}(\beta)$ are uniformly smaller compared to $D_{n111}(\beta)$. We can write

$$\begin{aligned} R_{n111}(\beta) &= \frac{1}{n-1}\,\frac{1}{n(n-1)h}\sum_{i\neq j}\langle \epsilon_i(\cdot)-\epsilon_j(\cdot),\ \epsilon_j(\cdot)\rangle_{\mathcal{H}}\, f_{\bar\beta}(X_j'\bar\beta) \\ &\qquad\qquad \times\left[g^{-1}L_{ij}(\beta)-g^{-1}L_{ij}(\bar\beta)\right]K_{ij}(\bar\beta)\phi_{ij}(\bar\beta) \\ &\quad + \frac{1}{n-1}\,\frac{1}{n(n-1)h}\sum_{i\neq j}\left\langle r_i(\cdot;\bar\beta)-r_j(\cdot;\bar\beta),\ \epsilon_j(\cdot)\right\rangle_{\mathcal{H}} f_{\bar\beta}(X_j'\bar\beta) \\ &\qquad\qquad \times\left[g^{-1}L_{ij}(\beta)-g^{-1}L_{ij}(\bar\beta)\right]K_{ij}(\bar\beta)\phi_{ij}(\bar\beta) \\ &\quad + \frac{1}{n-1}\,\frac{r_n}{n(n-1)h}\sum_{i\neq j}\langle \delta(X_i,t)-\delta(X_j,t),\ \epsilon_j(\cdot)\rangle_{\mathcal{H}}\, f_{\bar\beta}(X_j'\bar\beta) \\ &\qquad\qquad \times\left[g^{-1}L_{ij}(\beta)-g^{-1}L_{ij}(\bar\beta)\right]K_{ij}(\bar\beta)\phi_{ij}(\bar\beta) \\ &= R_{n1111}(\beta)+R_{n1112}(\beta)+r_nR_{n1113}(\beta). \end{aligned}$$

The leading term in $R_{n1111}(\beta)$ is

$$\frac{1}{n-1}\frac{1}{n(n-1)h}\sum_{i\neq j}\|\epsilon_j(\cdot)\|^2_{\mathcal{H}}\, f_{\bar\beta}(X_j'\bar\beta)\left[\frac{1}{g}L_{ij}(\beta)-\frac{1}{g}L_{ij}(\bar\beta)\right]K_{ij}(\bar\beta)\phi_{ij}(\bar\beta).$$

By the boundedness of $\mathbb{E}[\|\epsilon_j(\cdot)\|^2_{\mathcal{H}}\mid X_j]$ and $f_{\bar\beta}(X_j'\bar\beta)$, and Lemma 4, deduce that $R_{n1111}(\beta_n)=o_{\mathbb{P}}(n^{-1})$. Gathering facts deduce $D_{n11}(\beta_n)=o_{\mathbb{P}}(n^{-1}h^{-1/2}+r_n^2)$.

The rate of D_{n12}. We have $\widehat V_i(\beta)(t)-\widehat V_i(\bar\beta)(t) = Y_i(t)\Delta_{1,ni}(\beta)+\Delta_{2,ni}(\beta)$ with $\Delta_{1,ni}(\beta)$ and $\Delta_{2,ni}(\beta)$ independent of t and

$$\sup_{1\le i\le n}\sup_{\beta\in\mathcal{B}_n}\{|\Delta_{1,ni}|+|\Delta_{2,ni}|\} = O_{\mathbb{P}}(n^{-1/2}g^{-1/2}\log^{1/2}n+b_n);$$

see Lemma 1. Replacing and taking absolute values, deduce

$$D_{n12}(\beta_n)=O_{\mathbb{P}}(n^{-1}g^{-1}\log n+n^{-1}\log^2 n)=o_{\mathbb{P}}(n^{-1}h^{-1/2}),$$

since $g^{-1}h^{1/2}\to 0$ and $h\log^4 n\to 0$. Gathering facts deduce that

$$D_{n1}(\beta_n)=D_{n11}(\beta_n)+D_{n12}(\beta_n)=o_{\mathbb{P}}(n^{-1}h^{-1/2}+r_n^2).$$

Uniform bounds for D_{n2}.

$$\begin{aligned}
D_{n2}(\beta) &= \frac{1}{n(n-1)h}\sum_{i\neq j}\Big\langle \widehat{V}_i(\bar\beta),\ \widehat{V}_j(\bar\beta)\Big\rangle_{\mathcal{H}}\left[K_{ij}(\beta)\phi_{ij}(\beta)-K_{ij}(\bar\beta)\phi_{ij}(\bar\beta)\right]\\
&= \frac{1}{n(n-1)h}\sum_{i\neq j}\Big\langle [Y_i(t)-r_i(t;\bar\beta)]f_{\bar\beta}(X_i'\bar\beta),\ [Y_j(t)-r_j(t;\bar\beta)]f_{\bar\beta}(X_j'\bar\beta)\Big\rangle_{\mathcal{H}}\\
&\qquad\qquad \times\left[K_{ij}(\beta)\phi_{ij}(\beta)-K_{ij}(\bar\beta)\phi_{ij}(\bar\beta)\right]+\text{terms of smaller rate}\\
&= D_{n21}(\beta)+\text{terms of smaller rate.}
\end{aligned}$$

Recall that by construction, $\mathbb{E}[Y_i(t)\mid X_i]=r_i(t;\bar\beta)+r_n\delta(X_i,t)$, so that

$$[Y_i(t)-r_i(t;\bar\beta)]f_{\bar\beta}(X_i'\bar\beta)=[\epsilon_i(t)+r_n\delta(X_i,t)]f_{\bar\beta}(X_i'\bar\beta),$$

with $\mathbb{E}[\epsilon_i(t)\mid X_i]=0$, $\forall t\in[0,1]$. Thus

$$\begin{aligned}
D_{n21}(\beta) &= \frac{1}{n(n-1)h}\sum_{i\neq j}\langle \epsilon_i(t),\ \epsilon_j(t)\rangle_{\mathcal{H}}\, f_{\bar\beta}(X_i'\bar\beta)f_{\bar\beta}(X_j'\bar\beta)\left[K_{ij}(\beta)\phi_{ij}(\beta)-K_{ij}(\bar\beta)\phi_{ij}(\bar\beta)\right]\\
&+\frac{2r_n}{n(n-1)h}\sum_{i\neq j}\langle \epsilon_i(t),\ \delta(X_j,t)\rangle_{\mathcal{H}}\, f_{\bar\beta}(X_i'\bar\beta)f_{\bar\beta}(X_j'\bar\beta)\left[K_{ij}(\beta)\phi_{ij}(\beta)-K_{ij}(\bar\beta)\phi_{ij}(\bar\beta)\right]\\
&+\frac{r_n^2}{n(n-1)h}\sum_{i\neq j}\langle \delta(X_i,t),\ \delta(X_j,t)\rangle_{\mathcal{H}}\, f_{\bar\beta}(X_i'\bar\beta)f_{\bar\beta}(X_j'\bar\beta)\left[K_{ij}(\beta)\phi_{ij}(\beta)-K_{ij}(\bar\beta)\phi_{ij}(\bar\beta)\right]\\
&= D_{n211}(\beta)+2r_nD_{n212}(\beta)+r_n^2D_{n213}(\beta).
\end{aligned}$$

The term $D_{n211}(\cdot)$ is a degenerate $U-$process of order 2, indexed by β. Consider $\mathcal{F}_n=\{h(\cdot,\cdot;\beta):\beta\in\mathcal{B}_n\}$ with

$$h((x_1,\epsilon_1),(x_2,\epsilon_2);\beta)=\langle\epsilon_1(\cdot),\ \epsilon_2(\cdot)\rangle_{\mathcal{H}}f_{\bar\beta}(x_1'\bar\beta)f_{\bar\beta}(x_2'\bar\beta)[K_{12}(\beta)\phi_{12}(\beta)-K_{12}(\bar\beta)\phi_{12}(\bar\beta)].$$

It is easy to see that $\mathcal{F}_n$ is a VC class, or Euclidean in the terminology of Sherman (1994), for a squared integrable envelope $H(\cdot)$, with some A and V independent of n. (Recall that the $\delta-$covering number of an Euclidean class of function is bounded by $A\delta^{-V}$.) Indeed, the functions $h(\cdot,\cdot;\beta)$ are indexed by the parameter β that occurs only in $K_{12}(\beta)\phi_{12}(\beta)$, all the other terms in the definition of $h(\cdot,\cdot;\beta)$ are fixed real-valued functions of the observations. Thus it suffices to use the bounded variation of the functions $K(\cdot)$ and $\phi(\cdot)$, and apply standard results from Nolan and Pollard (1987), Pakes and Pollard (1989) or van der Vaart (1998) to derive the polynomial covering number for the family $\mathcal{F}_n$, with constants A and

V independent of n. Since $\mathbb{E}(\|\epsilon_1(\cdot)\|^2_{\mathcal{H}} \mid X_1)$ and $f_{\bar\beta}(X_1'\bar\beta)$ are bounded, and the kernel K is bounded, by Lemma 4 deduce that $\mathbb{E}\left[\sup_{\beta\in\mathcal{B}_n} h(\cdot,\cdot;\beta)^2\right] \leq Ch^{1/2}b_n$ for some constant $C>0$ independent on n and $\bar\beta$. See Lemma 4. Applying the Main Corollary of Sherman (1994) with $k=2$, $p=1$, deduce that

$$\sup_\beta |hD_{n211}(\beta)| \leq \frac{C'}{n}(b_n h^{1/2})^{\alpha/2} = \frac{h^{1/2}}{n} \times (b_n n^{1/2})^{\alpha/2} \times O(\{nh^{-1+2/\alpha}\}^{-\alpha/4})$$

for $0<\alpha<1$. Since and α could be taken arbitrarily close to 1, in particular it could be in the interval $(2/3,1)$, and $b_n n^{1/2}$ could be any sequence diverging to infinity faster than $\log n$, and $nh^2\to\infty$, deduce that $D_{n211}(\beta_n) = o_{\mathbb{P}}(n^{-1}h^{-1/2})$. For the uniform rate of the centered $U-$process $D_{n212}(\cdot)$, use the Hoeffding decomposition. The degenerate $U-$process of order 2 in this decomposition is of uniform rate $o_{\mathbb{P}}(n^{-1/2})$. The degenerate $U-$process of order 1 in the decomposition, denoted by $D_{n212,1}(\beta)$, is defined by the equation

$$hD_{n212,1}(\beta) = \frac{1}{n}\sum_{i\neq j} \gamma_i(\beta;h) f_{\bar\beta}(X_i'\bar\beta)$$

where

$$\gamma_i(\beta;h) = \mathbb{E}\left\{\langle\epsilon_i(\cdot),\delta(X_j,\cdot)\rangle_{\mathcal{H}} f_{\bar\beta}(X_j'\bar\beta)\left[K_{ij}(\beta)\phi_{ij}(\beta) - K_{ij}(\bar\beta)\phi_{ij}(\bar\beta)\right] \mid X_i, \epsilon_i(\cdot)\right\}.$$

Since $f_{\bar\beta}$ and $\delta(X,\cdot)$ are supposed bounded, we have

$$|\gamma_i(\beta;h)| \leq C\|\epsilon_i(\cdot)\|_{\mathcal{H}}\mathbb{E}\left\{\left|K_{ij}(\beta)\phi_{ij}(\beta) - K_{ij}(\bar\beta)\phi_{ij}(\bar\beta)\right| \mid X_i\right\}$$

for some constant $C>0$ depending of the bounds of $f_{\bar\beta}(\cdot)$ and $\|\delta(\cdot,\cdot)\|_{\mathcal{H}}$. Now, by Lemma 4, and since $nh^2/\log^3 n\to\infty$ and $b_n^2 n$ could converge to infinity faster than $\log^2 n$ but slower than $\log^3 n$,

$$\begin{aligned}
&\mathbb{E}[\sup_{\beta\in\mathcal{B}_n} \gamma_i^2(\beta;h)] \\
&\quad\leq C\mathbb{E}\left[\mathbb{E}[\|\epsilon(\cdot)\|^2_{\mathcal{H}} \mid X] \sup_{\beta\in\mathcal{B}_n} \mathbb{E}^2\left\{\left|K_{ij}(\beta)\phi_{ij}(\beta) - K_{ij}(\bar\beta)\phi_{ij}(\bar\beta)\right| \mid X_i\right\}\right] \\
&\qquad\qquad = O(b_n^2 h) = o(h^2).
\end{aligned}$$

Let us point out that the rate could be improved if one tracks the dependence of the constants appearing in Sherman's result on the $\delta-$covering number of $\mathcal{F}_n$. This covering number decreases with n as the parameter set $\mathcal{B}_n$ shrinks to $\bar\beta$. For our purposes we do not need this refinement.

Since the functions $K(\cdot)$ and $\phi(\cdot)$ are functions of bounded variation, and the Euclidean property is preserved after taking conditional expectations, by the results of Nolan and Pollard (1987), Pakes and Pollard (1989) and Sherman (1994), the empirical process $hD_{n212,1}(\beta)$ is indexed by an Euclidean family of functions. By the maximal inequality of Sherman (1994), $D_{n212,1}(\beta_n) = o_{\mathbb{P}}(n^{-1/2})$. Gathering facts, $r_n D_{n212}(\beta_n)$ is negligible compared to r_n^2. By the same arguments, $D_{n213}(\beta_n) = o_{\mathbb{P}}(1)$. Conclude that $D_{n2}(\beta_n) = o_{\mathbb{P}}(n^{-1}h^{-1/2} + r_n^2)$. □

The proof of Proposition 2 is provided in the Supplementary Material.

Supplementary Material

The Supplementary Material contains the proofs of Lemma A.1 and Proposition 2, some details on equation (4.4) and on the proof of Proposition 1 and Lemmas 1–4.

Acknowledgements V. Patilea gratefully acknowledges support from the research program *New Challenges for New Data* of Genes, LCL and Fondation du Risque.

References

Chen, D., P. Hall, and H.-G. Müller (2011): Single and multiple index functional regression models with nonparametric link, *Ann. Statist.*, 39, 1720–1747.

Chen, S. X. and I. Van Keilegom (2009): A goodness-of-fit test for parametric and semiparametric models in multiresponse regression, *Bernoulli*, 15, 955–976.

Chiang, C.-T. and M.-Y. Huang (2012): New estimation and inference procedures for a single-index conditional distribution model. *J. Multivariate Anal.*, 111, 271–285.

Cox, D. R. (1972): Regression Models and Life-Tables, *J. R. Stat. Soc. Ser. B*, 34, 187–220.

Delecroix, M., W. Härdle, and M. Hristache (2003): Efficient estimation in conditional single-index regression, *J. Multivariate Anal.*, 86, 213–226.

Delecroix, M., M. Hristache, and V. Patilea (1999): Optimal Smoothing in Semiparametric Index Approximation of Regression Functions, CREST Working Paper 9952.

——— (2006): On semiparametric $M-$estimation in single-index regression. *J. Statist. Plan. Inference*, 136, 730–769.

Escanciano, J. C. and K. Song (2010): Testing single-index restrictions with a focus on average derivatives, *J. Econometrics*, 156, 377–391.

Fan, Y. and Q. Li (1996): Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms, *Econometrica*, 64, 865–90.

Guerre, E. and P. Lavergne (2005): Data-Driven Rate-Optimal Specification Testing in Regression Models, *Ann. Statist.*, 33, 840–870.

Hall, P. and Q. Yao (2005): Approximating conditional distribution functions using dimension reduction, *Ann. Statist.*, 33, 977–1454.

Härdle, W., P. Hall, and H. Ichimura (1993): Optimal smoothing in single-index models. *Ann. Statist.*, 21, 157–178.

Horowitz, J. L. (2009): *Semiparametric and Nonparametric Methods in Econometrics*, Springer Series in Statistics. Springer: New-York.

Horowitz, J. L. and V. G. Spokoiny (2001): An adaptive, rate-optimal test for a parametric mean-regression model against a nonparametric alternative, *Econometrica*, 69, 599–631.

Hristache, M., A. Juditsky, and V. Spokoiny (2001): Direct estimation of the index coefficient in a single-index model, *Ann. Statist.*, 29, 595–917.

Ichimura, H. (1993): Semiparametric least squares (SLS) and weighted SLS estimation of single-index models, *J. Econometrics*, 58, 71–120.

Kong, E. and Y. Xia (2012): A single-index quantile regression model and its estimation, *Econometric Theory*, 28, 730–768.

Lavergne, P., S. Maistre, and V. Patilea (2015): A significance test for covariates in nonparametric regression, *Electron. J. Statist.*, 9, 643–678.

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

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

Sherman, R. P. (1994): Maximal inequalities for degenerate U-processes with applications to optimization estimators, *Ann. Statist.*, 22, 439–459.

Stute, W., W. González-Manteiga, and M. P. Quindimil (1998): Bootstrap Approximations in Model Checks for Regression, *J. Amer. Statist. Assoc.*, 93, pp. 141–149.

Stute, W. and L.-X. Zhu (2005): Nonparametric Checks for Single-Index Models, *Ann.*

Statist., 33, pp. 1048–1083.

Xia, Y. (2006): Asymptotic distributions for two estimators of the single-index model, *Econometric Th.*, 22, 1112–1137.

Xia, Y. (2009): Model checking in regression via dimension reduction, *Biometrika*, 96, 133–148.

Xia, C., W. Härdle, and L. Zhu (2011): The EFM approach for single-index models, *Ann. Statist.*, 39, 1658–1688.

Xia, Y., W. K. Li, H. Tong, and D. Zhang (2004): A Goodness-of-Fit Test For Single-Index Models, *Statist. Sinica*, 14, 1–39.

Université de Strasbourg,

IRMA, UMR 7501,

7 rue René-Descartes,

67084 Strasbourg Cedex

FRANCE

E-mail: (maistre@math.unistra.fr)

CREST (Ensai)

FRANCE

E-mail: (patilea@ensai.fr)