Minimax rates for nonparametric specification testing in regression models

Emmanuel Guerre, LSTA Paris 6 and CREST

Pascal Lavergne, INRA-ESR

February 1999, revised April 2000

Abstract

In the context of testing the specification of a nonlinear parametric regression function, we study the power of specification tests using the minimax approach. We determine the maximum rate at which a set of smooth local alternatives can approach the parametric model while ensuring consistency of a test uniformly against any alternative in this set. We show that a smooth nonparametric testing procedure has optimal minimax asymptotic properties for regular alternatives. As a by-product, we obtain the rate of the smoothing parameter that ensures optimality of the test. By contrast, many non-smooth tests, such as Bierens' (1982) integrated conditional moment test, have suboptimal minimax properties.

Keywords: Minimax approach, Specification testing.

JEL classification: Primary C52; Secondary C44.

Financial support from CREST and INRA is gratefully acknowledged. We thank the editor and the referees for comments that were helpful in improving our paper.
Corresponding address: INRA-ESR, BP 27, 31326 CASTANET-TOLOSAN Cedex FRANCE. E-mail addresses: eguerre@ccr.jussieu.fr, lavergne@toulouse.inra.fr.

1 Introduction

Specification analysis is a central topic in econometrics. In recent years, a substantial amount of work has focused on the search of specification tests that are consistent against a large spectrum of nonparametric alternatives. Bierens (1982) inaugurates this line of research by proposing an integrated conditional moment (ICM) test for checking the specification of a parametric regression model. His method relies on the empirical process of the residuals from the parametric model and has been further developped by Andrews (1997), Bierens (1990), Bierens and Ploberger (1997), Delgado (1993) and Stute (1997) among others. A competing approach compares parametric and smooth nonparametric regression estimators, see Fan and Li (1996a), Härdle and Mammen (1993), Hong and White (1995), Li and Wang (1998) and Zheng (1996) to mention just a few. Thus there now exists a large range of consistent specification tests for regression models, see Hart (1997) for an overview.

A focus of the literature concerns the power performances of the procedures derived from either approach. This has been mainly investigated by studying the tests' behavior under particular local alternatives, either theoretically or by means of Monte-Carlo experiments, see Hart (1997). A familiar approach consists in considering alternatives of the form

$$E\left(Y|X\right) = \mu(X, \theta_0) + \rho_n \delta(X), \tag{1.1}$$

where $\mu(X, \theta_0)$ is a member of the parametric model, $\delta(\cdot)$ is a specified function and ρ_n goes to zero as n tends to infinity. Non-smooth tests, as the ICM test, generally have nontrivial power against local alternatives of the form (1.1) when $\rho_n \propto n^{-1/2}$. This is not true for smooth nonparametric tests. However, it is possible to construct sequences of smooth alternatives different from (1.1), whose distance from the parametric model decreases more slowly than $n^{-1/2}$ and against which smooth tests are consistent, but against which the ICM test, among others, has trivial power, see e.g. Fan and Li (1996b). As we will show, no test can be uniformly consistent over a reasonable smoothness class of alternatives whose distance from the null hypothesis tends to zero as $n^{-1/2}$. Thus, there is a sense in which the set of alternatives (1.1) is too restrictive.

The main goal of this paper is to investigate the power properties of specification tests against local alternatives that belong to a specified smoothness class of functions. We adopt a minimax approach to determine the *minimax critical rate* of a test, that is, the maximum rate at which the alternatives can approach the null hypothesis and for which we can ensure a minimum predetermined power uniformly against the considered set of local alternatives. We prove that a large class of non-smooth tests, such as the ICM test, have high minimax critical rates. By contrast, a smooth test is shown to be *rate-optimal*, that is, it attains the lowest possible minimax critical rate for a specification test. As a by-product, we obtain the rate of the smoothing parameter that ensures rate-optimality of the test. By giving theoretical foundations for an optimal choice of the smoothing parameter, our analysis constitutes a first step towards a better understanding of its effects and the construction of practical procedures for its determination.

Horowitz and Spokoiny (1999) have recently proposed an adaptive test based on kernel estimation that automatically determine the smoothing parameter. They show that their test is rate-optimal, taking as given the minimax testing rate. By contrast, we focus here on the minimax rates for specification testing. Our results demonstrate for the first time the optimal minimax testing rates for the multiple regression model with heteroscedastic errors.

Our paper is organized as follows. In Section 2, we describe our framework and assumptions. In Section 3, we establish minimax rates for specification testing in regression models and we provide a testing procedure that is *rate-optimal* for regular alternatives. We then discuss implications of our results and we relate them to previous ones on minimax hypothesis testing. We also study the minimax properties of a class of non-smooth tests. Proofs of the main results are relegated to Section 4. Three appendices gather some auxiliary results.

2 Framework and assumptions

Let (X, Y) be a random variable in $\mathbb{R}^p \times \mathbb{R}$. We consider a parametric family $\mathcal{M}$ of regression functions $\mathcal{M} = \{\mu(., \theta)\ ;\ \theta \in \Theta\}, \Theta \subset \mathbb{R}^d$. The null hypothesis of interest is

$$H_0\ :\ m(.) \equiv \mathbb{E}[Y|X = .] \in \mathcal{M}\ .$$

Alternative hypotheses are characterized through the distance ρ of the regression function to the parametric model $\mathcal{M}$ and are denoted $H_1(\rho)$. Consider a test t_n depending on the sample size n with values in $\{0,1\}$, where $t_n = 1$ corresponds to the rejection of H_0, and whose level is $\alpha(t_n) = \sup_{H_0} I\!\!P_0\,(t_n = 1)$. In the minimax approach, the properties of the test are further characterized by its maximum type-II error over the considered alternatives, that is

$$\beta(t_n, \rho) = \sup_{H_1(\rho)} I\!\!P_m(t_n = 0).$$

In asymptotic settings, the approach should be adapted to deal with sequences of nested alternatives that get closer to the null as the sample size increases, that is, with $\rho_n \to 0$. Because we are evaluating the power properties of t_n through $\beta(t_n, \rho_n)$, we should focus on alternatives that maximizes the type-II error over $H_1(\rho_n)$. The specific alternatives that determine the minimax properties of a test are then among the "least favorable alternatives" and should lie at a distance ρ_n from the parametric specification. Evaluating the power performances of a test in this way comes to a prudent strategy where the practitioner wants to guard against the worst possible alternatives. This is of interest when the econometrician has not enough information about the possible misspecifications to consider specific alternatives such as (1.1).

Our subsequent analysis focuses on the rate of decay to zero of ρ_n and the corresponding properties of a test. For any consistent α-level test, there should exist a *critical rate* such that the test can discriminate local alternatives from the null hypothesis in the minimax sense if and only if ρ_n does not go to zero faster than this critical rate. This rate can in turn be optimized among all tests of given level and the resulting rate will be called the *minimax rate* of the testing problem. In Section 3, we derive a testing procedure as well as its critical rate. We then show that this test is *rate-optimal*, i.e. that it detects any alternative in $H_1(\tilde{\rho}_n)$ where $\tilde{\rho}_n$ is the minimax testing rate.

In the absence of other constraint, any α-level test of H_0 against $H_1(\rho)$ poorly behaves in the minimax sense, even when ρ is constant. Indeed, in the minimax approach, tests are evaluated uniformly over $H_1(\rho)$, which can contain very irregular functions that are not distinguishable from noise. However, an econometrician who sets up a parametric regression model and wants to test its specification has usually in mind alternatives that are regular anyway. Hence, as done in another context by Ingster (1993), we impose some regularity constraints on the nonparametric function of interest. For $s \in [0,1)$, let $C_p(L,s)$ be the set of measurable functions $m(\cdot)$ such that

$$|m(x) - m(y)| \leq L\|x-y\|^s \quad \forall x, y \ .$$

For $s \geq 1$, let $[s]$ be the greatest integer less than or equal to s, and let $C_p(L,s)$ be the set of functions $m(\cdot)$ differentiable up to order $[s]$, whose all derivatives of order $[s]$ belongs to $C_p(L,[s]-s)$.[1] We then consider the set of alternative hypotheses

$$H_1(\rho): \quad \inf_{\theta\in\Theta} \mathbb{E}\left(\mu(X,\theta) - m(X)\right)^2 \geq \rho^2 \ , \ m(.) \in C_p(L,s) \ ,$$

defined through the L_2-norm.

Our approach will involve the notion of "pseudo-true value" for the parameter θ, see White (1982) and Gourieroux, Monfort and Trognon (1984). We now describe some assumptions related to this pseudo-true value and the way it can be estimated.

Assumption M1 *For each $m(\cdot)$ in $C_p(L,s)$, there exists a unique $\theta^* = \theta^*_m$ such that*

$$\mathbb{E}\left(\mu(X,\theta^*) - m(X)\right)^2 = \inf_{\theta\in\Theta} \mathbb{E}\left(\mu(X,\theta) - m(X)\right)^2 \ .$$

For any sequence $\{m_n(.), n = 1, \ldots\}$ such that $\exists\ \theta$ in the interior of Θ with $\lim_{n\to+\infty} \mathbb{E}(m_n(X) - \mu(X,\theta))^2 = 0$, $\theta^*_{m_n}$ converges to θ.

Assumption M2 *i. For each $\theta \in \Theta$, $\mu(\cdot,\theta) \in C_p(L_{\mathcal{M}},s)$, $L_{\mathcal{M}} \leq L$, and $\mathbb{E}\mu^4(X,\theta) < \infty$.*

ii. For each x in $[0,1]^p$, $\mu(x,.)$ is twice continuously differentiable with respect to θ, with first and second order derivatives $\mu_\theta(\cdot,\cdot)$ and $\mu_{\theta\theta}(\cdot,\cdot)$ uniformly bounded in $x \in [0,1]^p$ and $\theta \in \Theta$.

iii. The matrix $\mathbb{E}\left[\dfrac{\partial\mu(X,\theta)}{\partial\theta}\dfrac{\partial\mu(X,\theta)}{\partial\theta^\top}\right]$ is invertible for all $\theta \in \Theta$.

iv. The set of gradient functions $\left\{\dfrac{\partial\mu(.,\theta)}{\partial\theta}\ ;\ \theta\in\Theta\right\}$ is compact in C_0, the set of continuous functions from $[0,1]^p$ to $\mathbb{R}^d$ equipped with the uniform norm.

[1] The Lipschitz condition could be replaced by an Hölder-type condition.

Assumption M3 $\sqrt{n}(\widehat{\theta}_n - \theta_m^*) = O_{\mathbb{P}_m}(1)$ *uniformly with respect to* $m(\cdot) \in C_p(L,s)$, *i.e.*

$$\forall \eta > 0, \exists \nu > 0 : \limsup_{n\to+\infty} \sup_{m(\cdot)\in C_p(L,s)} \mathbb{P}_m \left(\sqrt{n}\|\widehat{\theta}_n - \theta_m^*\| > \nu\right) \leq \eta \ .$$

When the parametric model is correctly specified, Assumption M1 is an identification condition because θ_m^* is then the true value of the parameter. Under misspecification, M1 defines the pseudo-true value of θ as the limit of the nonlinear least-squares estimator. Regularity conditions on the parametric model $\mathcal{M}$ are provided in Assumption M2. Similar assumptions are used by White (1982) to establish the $\sqrt{n}$-consistency of the nonlinear least-squares estimator of θ_m^*.[2] Assumption M2-i basically says that alternatives of interest belong to a larger smoothness class than the null parametric model.[3] In particular, they include alternatives of the type

$$\mu(.,\theta) + \delta(.) \ , \ \theta \in \Theta \ , \ \text{with } \delta(.) \in C_p(L - L_{\mathcal{M}}, s) \text{ and } \ \mathbb{E}\delta^2(X) \geq \rho^2 \ .$$

By Assumption M3, we allow for estimators different from nonlinear least-squares. These estimators have to be uniformly consistent with respect to the functions considered in the alternatives. Such a result is not usually shown in the literature. However, uniformity is essential for developing our minimax approach. Birgé and Massart (1993) have shown that Assumption M3 usually holds for nonlinear least-squares estimators. Consider for instance the simple univariate regression model where $\mu(X,\theta) = \theta X$ with θ in $[\underline{\theta}, \overline{\theta}]$. The pseudo-true value is then defined as $\theta_m^* = \mathbb{E}[Xm(X)] / \mathbb{E}(X^2)$. Assumptions M1 and M2 both reduce to standard assumptions. Moreover, the OLS estimator is such that

$$\widehat{\theta}_n - \theta^* = \left[(1/n)\sum_{i=1}^n X_i^2\right]^{-1} (1/n)\sum_{i=1}^n (m(X_i) - \theta^* X_i + \varepsilon_i) X_i \ .$$

Hence, Assumption M3 holds for $\widehat{\theta}_n$ when $\mathbb{E}X^4$ and $\mathbb{E}Y^4$ are finite, as the empirical mean of the numerator is centered, with a variance of order $O(1/n)$ uniformly in $m(\cdot)$.

As noted by Stone (1982), the minimax estimation rate of a nonparametric regression not only depends on its smoothness, but also upon the behavior of the density $f(.)$ of X. Hall and alii

[2] The main difference stands in the compactness of the set of first derivatives.

[3] This assumption does not prevent $\mathcal{M}$ to be a subset of a smoothness class with index $s_{\mathcal{M}} > s$.

(1997) show that optimal local polynomial estimates are unable to achieve the usual minimax rate when the density decreases too rapidly to zero at the boundaries of its support. Similar phenomena can appear in our testing framework. For instance, if the density of the regressors has unbounded support, it can be possible to find some sequences of functions $m(\cdot)$ in $H_1(\rho)$, with fixed ρ, against which any test has trivial power, see Appendix C for an illustration. Therefore, to avoid technicalities, we limit ourselves to explanatory variables X with bounded support.[4]

Assumption D *The density $f(.)$ of X has support $[0,1]^p$, with $0 < \mathrm{f} \leq f(x) \leq \mathrm{F} < +\infty$ for any x in $[0,1]^p$, and is continuous on $[0,1]^p$.*

3 Minimax rates for specification testing

3.1 Lower bounds for minimax testing rates

The following theorem states that if $\rho_n = o(\tilde{\rho}_n)$, where $\tilde{\rho}_n = n^{-\frac{2s}{p+4s}}$ if $s \geq p/4$ and $\tilde{\rho}_n = n^{-\frac{1}{4}}$ if $s < p/4$, any α-level test t_n of H_0 against alternatives of type $H_1(\rho_n)$ is comparable, in the minimax sense, to the test which chooses among the two hypothesis randomly with $\mathbb{P}(t_n = 0) = 1 - \alpha$. This is an impossibility result, which gives $\tilde{\rho}_n$ as a lower bound for the minimax testing rate. It formalizes the idea that alternatives of the type (1.1) can be too restrictive to study the local behavior of specification tests.

Assumption I *$\{(X_i, Y_i), i = 1, \ldots, n\}$ is an i.i.d. sample on (X,Y) from $\mathbb{R}^p \times \mathbb{R}$, $\mathbb{E}Y^4 < \infty$. For $\varepsilon = Y - m(Y|X)$, $\mathbb{E}_m \varepsilon^2 > 0$ and $\mathbb{E}_m \left[\varepsilon^4 | X = x\right] < \infty$, $\forall x \in [0,1]^p$.*

Theorem 1 *Let $\tilde{\rho}_n = n^{-\frac{2s}{p+4s}}$ if $s \geq p/4$ and $\tilde{\rho}_n = n^{-\frac{1}{4}}$ if $s < p/4$. Under Assumptions D, I, M1–M3, if each ε_i is $\mathcal{N}(0,1)$ conditionally upon X_i, for any test t_n of asymptotic level α,*

$$\beta(t_n, \rho_n) \geq 1 - \alpha + o(1) \quad \text{whenever } \rho_n = o(\tilde{\rho}_n).$$

[4] As pointed out by Bierens and Ploberger (1997), we can without loss of generality replace X by $\phi(X)$, where $\phi(\cdot)$ is bounded one-to-one smooth mapping.

The assumption of standard normal residuals, which is central to derive Theorem 1, can be relaxed as soon as regular distributions are considered. A common condition is to assume that the translation model associated with the residuals ε_i's is locally asymptotically normal (LAN), that is, the density $f_\varepsilon(\cdot)$ of these variables fulfils

$$\sum_{i=1}^{n}\left[\log f_\varepsilon\left(\varepsilon_i+\frac{u}{\sqrt{n}}\right)-\log f_\varepsilon(\varepsilon_i)\right]=uS_n-u^2I/2+o_{I\!P}(1)\ ,$$

where $I > 0$ is a constant, and S_n converge in distribution to $\mathcal{N}(0,I)$, see Ibragimov and Has'minskii (1981) for details. Under such a condition, Theorem 1 carries over at the price of some technicalities. Note that the LAN condition allows for the presence of heteroscedasticity. However, the minimax rate can change if the LAN condition does not hold. For instance, the lower bound of the minimax testing rate is likely to be improved if the residuals' distribution is not regular. Nevertheless, when the distribution of the residuals is unknwon, as is the case in the next section, it is clear that the lower bound of Theorem 1 is still valid.

3.2 Minimax testing rates and a rate-optimal test for regular alternatives

To determine minimax testing rates, we now build a specific specification test. A popular method in econometrics follows the Lagrange multiplier approach, see Godfrey (1988). This comes to estimate the model under the null hypothesis in the first place and to use this estimate as a basis for a test statistic in a second step. Here we first estimate θ and use the estimated parametric residuals $\widehat{U}_i = Y_i - \mu(X_i,\widehat{\theta}_n)$ to test H_0. To this purpose, we introduce a simple approximating family of functions, on which the parametric residuals will be projected. Let us define

$$I_k=\prod_{j=1}^{p}[k_jh,(k_j+1)h)\ ,$$

where the multivariate index $k=(k_1,\ldots,k_p)^\top\in\mathcal{K}\subset I\!N^p$ satisfies $0\le k_j\le K-1$ for $j=0,\ldots,p$, $K=K_n$ being an integer number and $h=1/K$ the associated binwidth. The bins I_k's define a partition of $[0,1]^p$, up to a negligible set. Let $N_k=\sum_{i=1}^{n} I\!I(X_i\in I_k)$ be the number of observations of the exogenous variables in bin I_k. Consider

$$\widehat{T}_n = \frac{1}{\sqrt{2}K^{p/2}} \sum_{k \in \mathcal{K}} \frac{I\!I[N_k > 1]}{N_k} \sum_{\{X_i, X_j\} \in I_k, i \neq j} \widehat{U}_i \widehat{U}_j$$

and

$$v_n^2 = (1/K^p) \sum_{k \in \mathcal{K}} \frac{I\!I(N_k > 1)}{N_k^2} \sum_{\{X_i, X_j\} \in I_k, i \neq j} \widehat{U}_i^2 \widehat{U}_j^2 \ .$$

The test is defined as $\tilde{t}_n = I\!I\left(v_n^{-1}\widehat{T}_n > z_\alpha\right)$, where v_n is the positive square-root of v_n^2 and z_α is the quantile of order $(1-\alpha)$ of the standard normal distribution.

Our test statistic is a simple histogram version of the kernel-based test of Zheng (1996). It can also be viewed as a modification of the familiar Pearson Chi-square statistic for goodness-of-fit for densities. Equivalently, it can be derived from Neyman (1937), since the considered indicator functions is an orthogonal system (with respect to any distribution for the X_i's). An advantage of this approach is to treat design density and conditional heteroskedasticity as nuisance parameters and then to avoid strong regularity assumptions on these functions.

Theorem 2 *Under Assumptions D, I and M1–M3,*

i. $\alpha(\tilde{t}_n) = \sup_{H_0} I\!P_m\left(v_n^{-1}\widehat{T}_n > z_\alpha\right) \to \alpha$, *whenever* $K \to \infty$ *and* $\frac{n}{K^p \log K^p} \to \infty$ *;*

ii. Assume $s > p/4$, *let* $\tilde{\rho}_n = n^{-\frac{2s}{p+4s}}$ *and* $K = [\tilde{\rho}_n^{-1/s}/\lambda]$, $\lambda > 0$. *Then, for any prescribed bound* β *in* $(0, 1-\alpha)$ *for the type-II error, there exists a constant* $\kappa > 0$ *such that*

$$\beta(\tilde{t}_n, \kappa\tilde{\rho}_n) = \sup_{H_1(\kappa\tilde{\rho}_n)} I\!P_m\left(v_n^{-1}\widehat{T}_n \leq z_\alpha\right) \leq \beta + o(1) \ .$$

As a consequence of Theorem 2, the test $\tilde{t}_n$ is an asymptotic α-level test of H_0, and has asymptotically a nontrivial minimax power against $H_1(\kappa\tilde{\rho}_n)$, for regular alternatives and κ large enough. This shows that our test can asymptotically distinguish any local alternative in this set, since β can be taken as small as desired. The quantity $\tilde{\rho}_n$ is thus an upper bound for the critical rate of our test.[5] Given Theorems 1 and 2, we obtain first that the lower bound $\tilde{\rho}_n$ on

[5]This does not mean that our test has trivial power against any alternative in $H_1(\rho_n)$ with $\rho_n = o(\tilde{\rho}_n)$, though it has trivial power against alternatives (1.1) with $\rho_n \propto n^{-1/2}$.

the minimax testing rate can be attained even when the residuals' distribution is unknown, and second that the test we have proposed is rate-optimal for regular alternatives.

Corollary 3 *Under Assumptions D, I and M1–M3 and if* $s > p/4$, $\tilde{\rho}_n = n^{-\frac{2s}{p+4s}}$ *is the minimax testing rate and* $\tilde{t}_n$ *is rate-optimal when* K *is chosen as in Theorem 2-ii.*

Our results give theoretical grounds for the choice of the binwidth in a specification testing framework. The testing optimal binwidth, ensuring that the test will be rate-optimal in the case of regular alternatives, i.e. for $s > p/4$, is

$$\tilde{h} \propto n^{-\frac{2}{p+4s}}.$$

For the same p and s, the optimal binwidth rate for testing the specification of a nonlinear parametric regression model is smaller than the optimal binwidth rate for minimax nonparametric estimation of the regression in the L_2-norm, which is $n^{-1/(p+2s)}$. Basically, choosing an optimal testing binwidth leads to balance a variance and a squared bias term, similar to the ones found in semiparametric estimation of $\mathbb{E}m^2(X)$. This implies some undersmoothing relative to optimal estimation of the regression function itself, as is the case in other semiparametric estimation problems, see e.g. Härdle and Tsybakov (1993), Powell and Stoker (1996). However, determining the optimal smoothing parameter for semiparametric estimation or testing are different problems in general.[6]

3.3 The case of irregular alternatives

The minimax testing rate generally depends on the relative standing of the smoothness indice s and the dimensionality of the model p. For irregular alternatives, i.e. $s \leq p/4$, the lower bound of Theorem 1 equals $n^{-1/4}$, and depends neither on the smoothness index nor on the dimension of the model. This rate corresponds to a baseline minimax testing rate when the residuals'

[6] In the white-noise model and alternatives defined through L_q norms, Lepski, Nemirovski and Spokoiny (1996) have shown that the minimax testing rate and the minimax estimation rate for the L_q norm coincide when q is even only.

unconditional variance σ^2 is known. Because the statistic

$$(1/n)\sum_{i=1}^{n}\widehat{U}_i^2 - \sigma^2 \tag{3.2}$$

estimates $E\left[Y-\mu(X,\theta^*)\right]^2-\sigma^2 = E\left[\mu(X,\theta^*)-m(X)\right]^2$ with rate of convergence equal to $\sqrt{n}$, a test that relies on this statistic detects any alternative in $H_1(\rho_n)$ with $\rho_n = n^{-1/4}$.

When σ^2 is unknown and the regression function is regular enough, σ^2 can be efficiently estimated with a $\sqrt{n}$-rate of convergence for regular alternatives, see e.g. Lavergne and Vuong (1996) and Newey (1994). The modified test statistic (3.2) where σ^2 is replaced by its efficient estimator then has a $\sqrt{n}$-degenerate behavior and our test statistic takes advantage of this degeneracy.[7] Unfortunately, this is not possible for irregular alternatives. Indeed, the average number of observations X_i's in each bin is of magnitude $O(nh^p) = O(n^{(4s-p)/(4s+p)})$, since the density is bounded away from 0 and infinity. Thus, for $s < p/4$, the average number of observations in each bin would go to 0 and this leads to a test statistic equal to zero with probability converging to 1.[8] Hence minimax rates for specification testing in regression models remain to be determined for irregular alternatives. It is likely that we would confront a problem where it is difficult to distinguish between signal and noise, so that the minimax testing rate depends on s. A related result in a different context can be found in Baraud, Huet and Laurent (1999).

3.4 Relations to other minimax testing rates

The main contribution of our work concerns the minimax rates for specification testing in regression models with multivariate random explanatory variables. It is interesting to compare our findings with those obtained in the continuous-time gaussian white noise model

$$dY_n(x) = m(x)dx + \frac{\sigma}{\sqrt{n}}dW(x)\ ,\ x \in [0,1]\ ,$$

[7]In the case of testing for a pure noise model, the specification test recently proposed by Dette and Munk (1998) is also based on (3.2), with σ^2 replaced by a inefficient difference-based estimator.

[8]Under the assumption that $m(\cdot)$ is bounded, our test can be applied when $s = p/4$ and it is rate-optimal, see Guerre and Lavergne (1999).

where $W(.)$ is a standard Brownian motion and the observations are $\{Y_n(x), x \in [0,1]\}$. Ingster (1993) has shown that the minimax testing rate of the null hypothesis $m(\cdot) \equiv 0$ is $n^{-\frac{2s}{1+4s}}$. Moreover, Brown and Low (1996) have established some equivalence results for statistical inference between this model and the univariate regression model with homoscedastic Gaussian errors. This suggests that previous results on testing the white noise model may be extended to regression models. But this is not generally the case. First, such asymptotic equivalence has its limits, as pointed out by Efromovich and Samarov (1996). To sum up, the available results imply that equivalence holds for $s > 1/2$, that nonequivalence holds $s < 1/4$ and $s = 1/2$, see Brown and Low (1998), while the other cases are undetermined. Second, at our knowledge, no work deals with extension of these results to multivariate settings. Third, the white noise model is not appropriate to deal with an unknown residual variance, because σ^2 is not a nuisance parameter in this model.[9] These are the reasons why we do not use previous results on the white noise model. Nevertheless, our work sheds some light on this issue. By explicitely dealing with the multivariate case, our results show that there is no equivalence between the (multivariate) white-noise model and the regression model with homoscedastic Gaussian errors when $s < p/4$.

3.5 Minimax critical rates of non-smooth tests

For smooth enough alternatives, i.e. when $s > p/4$, the minimax testing rate equals $n^{-\frac{s}{p+4s}}$ and approaches $n^{-1/2}$ from above when s grows to infinity. This means that in the minimax sense, it is impossible to detect alternatives that converges to the null at the "parametric rate" $1/\sqrt{n}$, even if the considered regression functions are infinitely differentiable.[10] This is the sense in which local alternatives of the type (1.1) are too restrictive. The minimax approach provides an alternative way of evaluating power properties of specification tests, and it seems interesting to study the minimax properties of non-smooth tests.

A well-known specification test in econometrics is the ICM test proposed by Bierens (1982)

[9] Two different values of σ^2 in the white noise model define measures that have disjoint supports.

[10] For estimation of a perfectly smooth signal in the white noise model, the minimax rate is $\sqrt{n/\log n}$, see e.g. Guerre and Tsybakov (1998).

and further developped by Bierens and Ploberger (1997). The ICM test statistic is

$$I_n = \int z^2(\xi)\, d\nu(\xi),$$

where $\nu(\cdot)$ is a measure on a compact set Ξ, $z(\xi) = (1/\sqrt{n}) \sum_{i=1}^n \widehat{U}_i w(X_i, \xi)$ with real-valued $w(X_i, \xi)$. Stinchcombe and White (1998) propose the more general statistic

$$I_{n,q} = \left[\int |z(\xi)|^q \, d\nu(\xi) \right]^{1/q}, \quad q \geq 1 \ .$$

Let $\widehat{t}_{n,q}$ be the test $\widehat{t}_{n,q} = \mathbb{1}(I_{n,q} > u_{\alpha,q})$, with $\lim_{n\to\infty} \alpha(\widehat{t}_{n,q}) = \alpha$.

Theorem 4 *Let $w(\cdot,\cdot)$ be bounded and such that $w(\cdot,\xi) \in C_p(\infty), \forall \xi \in \Xi$. Under Assumptions I, D, M1–M3, if each ε_i is $\mathcal{N}(0,1)$ conditionally upon X_i and $f(\cdot) \in C_p(\infty)$, then $\forall\, 1 \leq q < \infty$,*

$$\beta(\widehat{t}_{n,q}, \rho_n) = \sup_{H_1(\rho_n)} \mathbb{P}_m \left(I_{n,q} \leq u_{\alpha,q} \right) = 1 - \alpha + o(1), \quad \text{whenever } \rho_n = 0(n^{-a}),\ \forall a > 0.$$

The assumptions on $w(\cdot,\cdot)$ are justified by usual choices, such as $\exp(X'\xi)$ by Bierens (1990) or $(1+\exp(-X'\xi))^{-1}$ by White (1989). Furthermore, Stinchcombe and White (1998) show that considering $w(X,\xi) = G(X'\xi)$ with an analytic $G(\cdot)$ ensures desirable properties for the associated tests. Our result shows that such non-smooth tests are not rate-optimal in the minimax sense. It then follows that these tests cannot be asymptotically admissible against any alternative. This contrasts to the result obtained by Bierens and Ploberger (1997), who show that the ICM test is asymptotically admissible against specific alternatives of type (1.1). Moreover, the minimax properties of non-smooth tests are unsatisfactory, as their asymptotic minimax power is trivial against any sequence of alternatives $H_1(\rho_n)$ with ρ_n going to zero as a power of n.[11] We conjecture that similar results can be derived for other classes of tests, because empirical process based tests are basically identical to nonparametric smooth tests, with the major difference that the smoothing parameter is held fixed, see e.g. Eubank and Hart (1993) or Fan and Li (1996b).

[11] Without the assumption of an analytic $f(\cdot)$, our proof shows that a lower bound for the critical rate of such tests is $n^{-\frac{s}{p+2s}}$, which is greater than the minimax testing rate.

3.6 Directions for future research

It is likely that our results extend to the problem of testing econometric model defined by multiple moment conditions, as considered by Delgado, Dominguez and Lavergne (1998). Our testing methodology can also be easily adapted to deal with specific alternatives that are of interest for practitionners. Indeed, one could add some parametric components to the approximating family used to build our test statistic, such as polynomial ones. This would improve the power properties of the test against such specific local alternatives, without affecting its general minimax properties. A central direction for our future work is to develop data-driven techniques for choosing the smoothing parameter. Useful suggestions can be found in Hart's (1997) monograph and the references therein. This issue is addressed by Spokoiny (1996) for the white noise model and Horowitz and Spokoiny (1999) in the fixed-design regression model.

4 Proofs

4.1 Proof of Theorem 1

Some small alternatives

Let φ be any infinitely differentiable function from $[0,l]^p$ to $I\!R$ such that

$$\int \varphi(x)dx = 0 \text{ and } \int \varphi^4(x)dx < \infty .$$

Assume that l is large enough so that φ is in $C_p(L - L_{\mathcal{M}}, s)$. Let $h_n = (\lambda\rho_n)^{1/s}$, $\lambda > 0$ and define

$$I_{kl} = \prod_{j=1}^{p}[lk_j h_n, l(k_j+1)h_n) ,$$

for $k \in \mathcal{K}_n(l)$, i.e. $k \in I\!N^p$ with $0 \leq k_j \leq 1/(h_n l) - 1$. Then $I_{kl} \subset [0,1]^p$. Without loss of generality, we assume that $K_n(l) = 1/(h_n l)$ is an integer. Let

$$\varphi_k(x) = \frac{1}{h_n^{p/2}}\varphi\left(\frac{x - lkh_n}{h_n}\right) , \; k \in \mathcal{K}_n(l) .$$

The functions $\varphi_k(\cdot)$'s are orthogonal. Let $(B_k, k \in \mathcal{K})$ be any sequence with $|B_k| = 1 \; \forall k$, and

$$m_n(.) = \mu(., \theta_0) + \delta_n(.) \; , \; \delta_n(.) = \lambda\rho_n h_n^{p/2} \sum_{k\in\mathcal{K}_n(l)} B_k\varphi_k(.) \; , \tag{4.1}$$

where θ_0 is any inner point of Θ.

Lemma 1 *Under Assumptions D, M1, M2, $m_n(.)$ is in $H1(\rho_n)$ for λ and n large enough.*

Proof: i) $m_n(\cdot) \in C_p(L,s)$: For any $\beta \in \mathbb{N}^p$ with $\sum_{j=1}^p \beta_j = [s]$,

$$\begin{aligned}\left|\frac{\partial^{[s]}\delta_n(x)}{\partial y_1^{\beta_1}\dots\partial y_p^{\beta_p}} - \frac{\partial^{[s]}\delta_n(y)}{\partial y_1^{\beta_1}\dots\partial y_p^{\beta_p}}\right| &= \lambda\rho_n\left|\sum_{k\in\mathcal{K}_n(l)} B_k \frac{\partial^{[s]}\varphi_k(x)}{\partial x_1^{\beta_1}\dots\partial x_p^{\beta_p}} - \frac{\partial^{[s]}\varphi_k(y)}{\partial y_1^{\beta_1}\dots\partial y_p^{\beta_p}}\right| \\ &= \frac{\lambda\rho_n}{h_n^s}\left|\frac{\partial^{[s]}\varphi(x)}{\partial x_1^{\beta_1}\dots\partial x_p^{\beta_p}} - \frac{\partial^{[s]}\varphi(y)}{\partial y_1^{\beta_1}\dots\partial y_p^{\beta_p}}\right| \le (L - L_{\mathcal{M}})\|x-y\|^{[s]-s} ,\end{aligned}$$

because $\varphi_k(\cdot)$ is identically zero for all but one $k \in \mathcal{K}_n(l)$ and $\varphi(\cdot) \in C_p(L - L_{\mathcal{M}}, s)$.
ii) $m_n(\cdot)$ is distant from the null model: Let $\theta_n \equiv \theta^*_{m_n}$. Then

$$\begin{aligned}\mathbb{E}^{1/2}\left[m_n(X) - \mu(X,\theta_n)\right]^2 &\ge \mathbb{E}^{1/2}\delta_n^2(X) - \mathbb{E}^{1/2}\left[\mu(X,\theta_0) - \mu(X,\theta_n)\right]^2 \\ &\ge \left(\mathrm{f}\int \delta_n^2(x)dx\right)^{1/2} - O\left(\|\theta_n - \theta_0\|_2\right) ,\end{aligned} \tag{4.2}$$

by Assumptions D and M2, which gives that the gradient $\partial\mu(x,\theta)/\partial\theta$ is bounded. Now,

$$\int \delta_n^2(x)dx = (\lambda\rho_n)^2 h_n^p K_n^p(l) = (\lambda\rho_n)^2 l^{-p} . \tag{4.3}$$

As θ_n converges to θ_0, it is then an inner point of Θ. Therefore M1 yields, applying M2 and the Lebesgue dominated convergence theorem, that

$$\mathbb{E}\frac{\partial\mu(X,\theta_n)}{\partial\theta_n}\left[\mu(X,\theta_n) - m_n(X)\right] = 0 .$$

This leads to

$$\mathbb{E}\frac{\partial\mu(X,\theta_n)}{\partial\theta}\left[\mu(X,\theta_n) - \mu(X,\theta_0)\right] = \mathbb{E}\delta_n(X)\frac{\partial\mu(X,\theta_n)}{\partial\theta} .$$

A simple Taylor expansion, which holds by M2, yields

$$\theta_n - \theta_0 = \left(\mathbb{E}\frac{\partial\mu(X,\theta_0)}{\partial\theta}\frac{\partial\mu(X,\theta_0)}{\partial\theta^\top} + o(1)\right)^{-1}\mathbb{E}\delta_n(X)\frac{\partial\mu(X,\theta_n)}{\partial\theta} ,$$

so that

$$\|\theta_n - \theta\|_2 = O\left(\left\|\mathbb{E}\delta_n(X)\frac{\partial\mu(X,\theta_n)}{\partial\theta}\right\|_2\right) . \tag{4.4}$$

Because M2 implies that the functions $\left\{ \frac{\partial \mu(.,\theta)}{\partial \theta} f(.) \, ; \, \theta \in \Theta \right\}$ are equicontinuous, by Assumptions D and M2 and the Azrela-Ascoli theorem, see Rudin (1991), and as $\varphi(\cdot)$ has integral zero, we get

$$\begin{aligned}
& \mathbb{E}\delta_n(X)\frac{\partial \mu(X,\theta_n)}{\partial \theta} \\
= \; & \lambda \rho_n h_n^p \sum_{k \in \mathcal{K}_n(l)} B_k \int \left(\frac{\partial \mu(lkh_n + h_n u, \theta_n)}{\partial \theta} f(lkh_n + h_n u) - \frac{\partial \mu(lkh_n, \theta_n)}{\partial \theta} f(lkh_n) \right) \varphi(u) du \\
= \; & \lambda \rho_n h_n^p K_n^p(l) o(1) = \lambda \rho_n l^{-p} o(1) \; .
\end{aligned}$$

Combining this equality with (4.2)–(4.4) yields, for λ and n large enough,

$$\mathbb{E}^{1/2} \left[m_n(X) - \mu(X,\theta_n) \right]^2 \geq \lambda \rho_n l^{-p} \left(\mathrm{f}^{1/2} l^{p/2} - o(1) \right) \geq \rho_n \; .\Box$$

Main proof

We shall establish that for any test t_n

$$\sup_{m(.) \in H_0(\mathcal{M})} \mathbb{P}_m(t_n = 1) + \sup_{m(.) \in H_1(\mathcal{M},\rho_n)} \mathbb{P}_m(t_n = 0) \; \geq \; 1 + o(1) \; . \qquad (4.5)$$

Step 1: Choice of a Bayesian a priori measure. Let θ_0 be any inner point of Θ and denote Π_0 the associate Dirac mass. Consider i.i.d.s Rademacher B_k's independent of the observations, i.e. $\mathbb{P}(B_k = 1) = \mathbb{P}(B_k = -1) = 1/2$, and define Π_{1n} as the a priori distribution defined on $H_1(\mathcal{M}, \rho_n)$ by (4.1). Lemma 1 shows that the support of Π_{1n} is a subset of $H_1(\rho_n)$ and $\Pi_n = \Pi_0 + \Pi_{1n}$ is an a priori Bayesian measure over $H_0 \cup H_1(\rho_n)$. This gives the lower bound

$$\sup_{m(.) \in H_0} \mathbb{P}_m(t_n = 1) + \sup_{m(.) \in H_1(\rho_n)} \mathbb{P}_m(t_n = 0) \geq \int \mathbb{P}_m(t_n = 1) d\Pi_0(m) + \int \mathbb{P}_m(t_n = 0) d\Pi_{1n}(m) \; . \quad (4.6)$$

The r.h.s. of (4.6) is the Bayes error of the test t_n which is greater than the error of the optimal Bayesian test based on the likelihood ratio Z_n that we now introduce. Denote by $\mathcal{Y}$ and $\mathcal{X}$ the set of observations on Y and X respectively and let $p_m(\mathcal{Y}, \mathcal{X})$ be the density corresponding to the regression function $m(.)$. Define the a priori densities associated with the two hypotheses as

$$\boldsymbol{p}_0(\mathcal{Y},\mathcal{X}) = \int p_m(\mathcal{Y},\mathcal{X}) \, d\Pi_0(m) \quad \text{and} \quad \boldsymbol{p}_{1n}(\mathcal{Y},\mathcal{X}) = \int p_m(\mathcal{Y},\mathcal{X}) \, d\Pi_{1n}(m) \; .$$

Let $\boldsymbol{E}_0$ be the expectation under $\boldsymbol{p}_0$. The likelihood ratio of the optimal Bayesian test is

$$Z_n = \frac{\boldsymbol{p}_{1n}(\mathcal{Y},\mathcal{X})}{\boldsymbol{p}_0(\mathcal{Y},\mathcal{X})} = \frac{\boldsymbol{p}_{1n}(\mathcal{Y}|\mathcal{X})}{\boldsymbol{p}_0(\mathcal{Y}|\mathcal{X})}.$$

The optimal Bayesian test rejects H_0 if $Z_n \geq 1$ and its Bayesian error, see Lehman (1986), is

$$1 - \frac{1}{2}\int |\boldsymbol{p}_0(\mathcal{Y},\mathcal{X}) - \boldsymbol{p}_{1n}(\mathcal{Y},\mathcal{X})| \, d\mathcal{Y}\, d\mathcal{X} \quad = \quad 1 - \frac{1}{2}\mathbb{E}\boldsymbol{E}_0\left[|Z_n - 1||\mathcal{X}\right] ,$$

Then (4.6) implies that

$$\sup_{m(.)\in H_0} \mathbb{P}_m(t_n = 1) + \sup_{m(.)\in H_1(\rho_n)} \mathbb{P}_m(t_n = 0) \geq \liminf_{n\to+\infty} \mathbb{E}\left\{1 - \frac{1}{2}\boldsymbol{E}_0\left[|Z_n - 1||\mathcal{X}\right]\right\} + o(1) ,$$

and (4.5) holds if we can show that the limit in the r.h.s. is 1. We first note that $1 - \frac{1}{2}\boldsymbol{E}_0\left[|Z_n - 1||\mathcal{X}\right]$ is positive as a conditional Bayes testing error. Then the Fatou lemma implies that it is enough to show that $\boldsymbol{E}_0\left[|Z_n - 1||\mathcal{X}\right] \xrightarrow{\mathbb{P}} 0$, which is implied by $\boldsymbol{E}_0\left[(Z_n - 1)^2|\mathcal{X}\right] \xrightarrow{\mathbb{P}} 0$. But $\boldsymbol{E}_0\left[(Z_n - 1)^2|\mathcal{X}\right] = \boldsymbol{E}_0\left(Z_n^2|\mathcal{X}\right) - 1$ as $\boldsymbol{E}_0(Z_n|\mathcal{X}) = 1$. Hence, Inequality (4.5) holds if

$$\boldsymbol{E}_0\left(Z_n^2|\mathcal{X}\right) \xrightarrow{\mathbb{P}} 1 . \tag{4.7}$$

Step 2: Study of the likelihood ratio Z_n. On the one hand, the variables $\varepsilon_{i0} = Y_i - \mu(X_i,\theta_0)$, $i = 1,\ldots,n$, are standard normal variables under $\boldsymbol{p}_0$ and

$$\boldsymbol{p}_0(\mathcal{Y}|\mathcal{X}) = (2\pi)^{-n/2}\exp\left[-\sum_{i=1}^n \varepsilon_{i0}^2/2\right] .$$

On the other hand, given the definition of Π_{1n},

$$\begin{aligned}
\boldsymbol{p}_{1n}(\mathcal{Y}|\mathcal{X}) &= (2\pi)^{-n/2}\int\left\{\exp\left[-\frac{1}{2}\sum_{i=1}^n (Y_i - m_n(X_i))^2\right]\right\} d\Pi_{1n}(m) \\
&= (2\pi)^{-n/2}\int\left\{\exp\left(-\frac{1}{2}\sum_{i=1}^n \varepsilon_{i0}^2/2 - \frac{1}{2}\sum_{i=1}^n \delta_n^2(X_i) + \sum_{i=1}^n \varepsilon_{i0}\delta_n(X_i)\right)\right\} d\Pi_{1n}(m) \\
&= \boldsymbol{p}_0(\mathcal{Y}|\mathcal{X})\int\left\{\exp\left(-\frac{1}{2}\sum_{i=1}^n \delta_n^2(X_i) + \sum_{i=1}^n \varepsilon_{i0}\delta_n(X_i)\right)\right\} d\Pi_{1n}(m) .
\end{aligned}$$

The definition of the alternatives (4.1) gives

$$\sum_{i=1}^n \varepsilon_{i0}\delta_n(X_i) = \lambda\rho_n h_n^{p/2}\sum_{k\in\mathcal{K}(l)}\sum_{i=1}^n B_k\varepsilon_{i0}\varphi_k(X_i) \quad \text{and} \quad \sum_{i=1}^n \delta_n^2(X_i) = \lambda^2\rho_n^2 h_n^p \sum_{k\in\mathcal{K}(l)}\sum_{i=1}^n \varphi_k^2(X_i) ,$$

since $B_k^2 = 1$ and $\varphi_k(.)\varphi_{k'}(.) = 0$ for $k \neq k'$. This yields

$$\begin{aligned}
Z_n &= \exp\left(-\frac{\lambda^2\rho_n^2 h_n^p}{2}\sum_{k\in\mathcal{K}(l)}\sum_{i=1}^n \varphi_k^2(X_i)\right) \\
&\quad\times \prod_{k\in\mathcal{K}(l)}\frac{1}{2}\left[\exp\left(\lambda\rho_n h_n^{p/2}\sum_{i=1}^n \varepsilon_{i0}\varphi_k(X_i)\right) + \exp\left(-\lambda\rho_n h_n^{p/2}\sum_{i=1}^n \varepsilon_{i0}\varphi_k(X_i)\right)\right] .
\end{aligned}$$

Therefore,

$$
\begin{aligned}
Z_n^2 &= \exp\left(-\lambda^2\rho_n^2 h_n^p \sum_{k\in\mathcal{K}(l)}\sum_{i=1}^n \varphi_k^2(X_i)\right)\\
&\quad\times\prod_{k\in\mathcal{K}(l)}\frac{1}{4}\left[\exp\left(2\lambda\rho_n h_n^{p/2}\sum_{i=1}^n\varepsilon_{i0}\varphi_k(X_i)\right)+2+\exp\left(-2\lambda\rho_n h_n^{p/2}\sum_{i=1}^n\varepsilon_{i0}\varphi_k(X_i)\right)\right].
\end{aligned}
$$

Conditionally on $\mathcal{X}$, the variables $\sum_i \varepsilon_{i0}\varphi_k(X_i),\ k\in\mathcal{K}(l), k\in\mathcal{K}_n(l)$, are independent centered Gaussian with conditional variance given by $\sum_i \varphi_k^2(X_i)$. Using $\mathbb{E}\exp\mathcal{N}(0,\sigma^2)=\exp(\sigma^2/2)$, we get

$$
\begin{aligned}
\boldsymbol{E}_0\left(Z_n^2|\mathcal{X}\right) &= \prod_{k\in\mathcal{K}(l)}\exp\left(-\lambda^2\rho_n^2 h_n^p\sum_{k\in\mathcal{K}(l)}\sum_{i=1}^n\varphi_k^2(X_i)\right)\times\frac{1}{2}\left\{\exp\left(2\lambda^2\rho_n^2 h_n^p\sum_{i=1}^n\varphi_k^2(X_i)\right)+1\right\}\\
&= \prod_{k\in\mathcal{K}(l)}\cosh\left(\lambda^2\rho_n^2 h_n^p\sum_{i=1}^n\varphi_k^2(X_i)\right),
\end{aligned}
$$

where $\cosh(x)$ is the hyperbolic cosine function. As $\cosh(x)\le\exp(x^2)$ by a series expansion, this yields

$$
1\le\boldsymbol{E}_0\left(Z_n^2|\mathcal{X}\right)\le\exp\left[\sum_{k\in\mathcal{K}(l)}\left(\lambda^2\rho_n^2 h_n^p\sum_{i=1}^n\varphi_k^2(X_i)\right)^2\right]
$$

and (4.7) holds if

$$
\sum_{k\in\mathcal{K}(l)}\left(\rho_n^2 h_n^p\sum_i\varphi_k^2(X_i)\right)^2\xrightarrow{\mathbb{P}}0. \tag{4.8}
$$

Consider the expectation of this positive random variable. We have

$$
\mathbb{E}\left[\sum_{k\in\mathcal{K}(l)}\left(\rho_n^2 h_n^p\sum_{i=1}^n\varphi_k^2(X_i)\right)^2\right] = \rho_n^4 h_n^{2p}\sum_{k\in\mathcal{K}(l)}\left\{n\mathbb{E}[\varphi_k^4(X)]+n(n-1)\mathbb{E}^2[\varphi_k^2(X)]\right\}.
$$

Now the standard change of variables $x=lh_nk+h_nu$ and Assumption D yields

$$
\mathbb{E}\left[\varphi_k^4(X)\right]=\int h_n^{-2p}\varphi^4\left[(x/h_n)-lk\right]f(x)\,dx\le \mathrm{F}h_n^{-p}\int\varphi^4(u)\,du=O(h_n^{-p})
$$

and

$$
\mathbb{E}\left[\varphi_k^2(X)\right]=\int h_n^{-p}\varphi^2\left[(x/h_n)-lk\right]f(x)\,dx\le \mathrm{F}\int\varphi^2(u)\,du=O(1).
$$

As $h_n=O(1/K_n(l))=O(\rho_n^{1/s})$,

$$
\mathbb{E}\left[\sum_k\left(\rho_n^2 h_n^p\sum_i\varphi_k^2(X_i)\right)^2\right]=\left[n\rho_n^4+n^2\rho_n^4h_n^p\right]O(1)=\left[n\rho_n^4+n^2\rho_n^{(p+4s)/s}\right]O(1).
$$

We then consider the two following cases:

i. $\boldsymbol{s > p/4}$: $\rho_n = o(\tilde{\rho}_n) \Longrightarrow n\rho_n^4 = o\left(n^{(p-4s)/(p+4s)}\right) = o(1)$ and $n^2\rho_n^{(p+4s)/s} = o(1)$.

ii. $\boldsymbol{s \leq p/4}$: $\rho_n = o(\tilde{\rho}_n) = o\left(n^{-1/4}\right) \Longrightarrow n\rho_n^4 = o(1)$ and $n^2\rho_n^{(p+4s)/s} = o\left(n^{(4s-p)/4s}\right) = o(1)$.

Equation (4.8) follows and then (4.7). Step 1 shows that (4.5) holds and Theorem 1 is proved. □

4.2 Proof of Theorem 2

For random variables Z and Z', define $\mathbb{E}^k(Z) \equiv \mathbb{E}_m(Z|X \in I_k)$, $\text{Var}^k(Z) \equiv \text{Var}_m(Z|X \in I_k)$,

$$\langle Z, Z' \rangle_k \equiv \frac{\mathbb{I}[N_k > 1]}{N_k} \sum_{\{X_i, X_j\} \in I_k, i \neq j} Z_i {Z'}_j, \quad \forall k \in \mathcal{K}, \quad \text{and} \quad \langle Z, Z' \rangle \equiv \frac{1}{\sqrt{2} K^{p/2}} \sum_{k \in \mathcal{K}} \langle Z, Z' \rangle_k.$$

Let $\text{Proj}_{\mathcal{K}} Z \equiv \sum_k \mathbb{I}(x \in I_k) \mathbb{E}^k Z$ be the projection of Z onto the space of linear indicators $\mathbb{I}(x \in I_k)$, $k \in \mathcal{K}$. Key properties of this mapping are

$$\mathbb{E}[\text{Proj}_{\mathcal{K}} Z] = \sum_k \mathbb{P}(X \in I_k) \mathbb{E}^k Z = \mathbb{E}Z, \qquad \mathbb{E}\left[\text{Proj}_{\mathcal{K}}^2 Z\right] \leq \mathbb{E}Z^2,$$

as $\text{Proj}_{\mathcal{K}}$ is a projection mapping. We let $U^* = Y - \mu(X, \theta^*)$, $\varepsilon = Y - m(X)$, $\delta(X) = m(X) - \mu(X, \theta^*)$, $e(X) = \mu(X, \widehat{\theta}_n) - \mu(X, \theta^*)$ and $S_{\mathcal{K}} = (N_k, k \in \mathcal{K})^\top$. For simplicity, we assume that $K = \tilde{\rho}_n^{-1/s}/\lambda$ is integer. Finally, C_i, $i = 1, \ldots$, denote positive constants that may vary from line to line.

Preliminary results

Proposition 5 *Let* $v^2(K) = (1/K^p) \sum_{k \in \mathcal{K}} \mathbb{I}(N_k > 1) \frac{N_k - 1}{N_k} \left(\mathbb{E}^k U^{*2}\right)^2$. *Under Assumptions I, D and M1–M3,* $v^2(K)$ *is bounded from above and in probability from below uniformly in* $m(\cdot) \in C_p(L, s)$, *and* $v_n^2 - v^2(K) = o_{\mathbb{P}_m}(1)$ *whenever* $\frac{n}{K^p \log K^p} \to \infty$.

Proof of Proposition 5: By Assumption D, $\mathrm{f}h^p \leq \mathbb{P}(X \in I_k) \leq \mathrm{F}h^p$. Now, on the one hand,

$$\begin{aligned} v^2(K) &\leq (1/K^p) \sum_{k \in \mathcal{K}} \left(\mathbb{E}^k U^{*2}\right)^2 \leq (1/\mathrm{f}) \sum_{k \in \mathcal{K}} \mathbb{P}[X \in I_k] \left(\mathbb{E}^k U^{*2}\right)^2 = (1/\mathrm{f}) \mathbb{E}\left[\text{Proj}_{\mathcal{K}}^2 U^{*2}\right] \\ &\leq (1/\mathrm{f}) \mathbb{E}\left[U^{*4}\right] \leq (8/\mathrm{f}) \left[\mathbb{E}_m Y^4 + \mathbb{E}_m \mu^4(X, \theta_m^*)\right] < \infty. \end{aligned}$$

On the other hand, by Lemma 4, with probability going to one uniformly in $k \in \mathcal{K}$,

$$\begin{aligned} v^2(K) &\geq (1/2K^p) \sum_{k \in \mathcal{K}} \left(\mathbb{E}^k U^{*2}\right)^2 \geq (1/2\mathrm{F}) \sum_{k \in \mathcal{K}} \mathbb{P}[X \in I_k] \left(\mathbb{E}^k U^{*2}\right)^2 \\ &\geq (1/2\mathrm{F}) \mathbb{E}_m \left[\text{Proj}_{\mathcal{K}}^2 U^{*2}\right] \geq (1/2\mathrm{F}) \mathbb{E}_m^2 \left[U^{*2}\right] \geq (1/2\mathrm{F}) \mathbb{E}_m^2 \left[\varepsilon^{*2}\right] > 0. \end{aligned}$$

Let $v_n^{*2} = (1/K^p) \sum_{k \in \mathcal{K}} \langle U^{*2}, U^{*2} \rangle_k / N_k$. Then

$$\left| v_n^2 - v_n^{*2} \right| \leq (1/K^p) \sum_{k \in \mathcal{K}} \frac{\mathbb{I}(N_k > 1)}{N_k} \left| \langle \widehat{U}^2, \widehat{U}^2 \rangle_k - \langle U^{*2}, U^{*2} \rangle_k \right| . \tag{4.9}$$

But $\langle \widehat{U}^2, \widehat{U}^2 \rangle_k - \langle U^{*2}, U^{*2} \rangle_k = 4\langle U^{*2}, U^* e(X) \rangle_k + 2\langle U^{*2}, e^2(X) \rangle_k + 4\langle U^* e(X), U^* e(X) \rangle_k + 4\langle U^* e(X), e^2(X) \rangle_k$ $+ \langle e^2(X), e^2(X) \rangle_k$. By Assumptions M1–M3 , $|e(X_i)| = O_{\mathbb{P}_m}(1/\sqrt{n})$ uniformly in $m(\cdot)$ and i. Hence the dominant term in (4.9) is

$$(4/K^p) \sum_{k \in \mathcal{K}} \frac{\mathbb{I}(N_k > 1)}{N_k} |\langle U^{*2}, U^* e \rangle_k| = O_{\mathbb{P}_m}(1/\sqrt{n})(1/K^p) \sum_{k \in \mathcal{K}} \frac{\mathbb{I}(N_k > 1)}{N_k} \langle U^{*2}, |U^*| \rangle_k \ .$$

But, by Assumptions I and M1,

$$\begin{aligned}
& \mathbb{E}_m \left[(1/K^p) \sum_{k \in \mathcal{K}} \frac{\mathbb{I}(N_k > 1)}{N_k} \langle U^{*2}, |U^*| \rangle_k | S_K \right] \\
= & \ (1/K^p) \sum_{k \in \mathcal{K}} \frac{\mathbb{I}(N_k > 1)(N_k - 1)}{N_k} \mathbb{E}^k U^{*2} \mathbb{E}^k |U^*| \\
\leq & \ (1/\mathrm{f}) \sum_{k \in \mathcal{K}} \mathbb{P}[X \in I_k] \, \mathbb{E}^k U^{*2} \mathbb{E}^k |U^*| = (1/\mathrm{f}) \mathbb{E}_m \left[\mathrm{Proj}_{\mathcal{K}}^2 U^{*2} \mathrm{Proj}_{\mathcal{K}} |U^*| \right] \\
\leq & \ (1/\mathrm{f}) \mathbb{E}_m^{1/2} \left[U^{*4} \right] \mathbb{E}_m^{1/2} \left[U^{*2} \right] < \infty
\end{aligned}$$

This shows that $v_n^2 - v_n^{*2} = O_{\mathbb{P}_m}(1/\sqrt{n})$. Now $v_n^{*2} - v^2(K)$ is centered conditionally upon $S_{\mathcal{K}}$ and, by Lemma 4,

$$\begin{aligned}
\mathbb{E}_m \left[\left(v_n^{*2} - v^2(K) \right)^2 | S_{\mathcal{K}} \right] &= \mathrm{Var}_m \left[v_n^{*2} - v^2(K) | S_{\mathcal{K}} \right] = (1/K^{2p}) \sum_{k \in \mathcal{K}} \frac{\mathbb{I}(N_k > 1)}{N_k^4} \sum_{i \neq j} \mathrm{Var}^k U_i^{*2} U_j^{*2} \\
&\leq (1/K^{2p}) \sum_{k \in \mathcal{K}} \mathbb{I}(N_k > 1) \frac{N_k - 1}{N_k^3} \left(\mathbb{E}^k U^{*4} \right)^2 \\
&\leq O_{\mathbb{P}}(nh^p)^{-2} \sum_{k \in \mathcal{K}} \left(\mathbb{P}(X \in I_k) \mathbb{E}^k U^{*4} \right)^2 \leq O_{\mathbb{P}}(nh^p)^{-2} \mathbb{E}_m^2 U^{*4} \to 0 \ .\square
\end{aligned}$$

Let $T_n \equiv T_n(\theta^*)$, $A = \langle \delta(X), e(X) \rangle$, $B = \langle \varepsilon, e(X) \rangle$ and $R = \langle e(X), e(X) \rangle$. Then

$$\widehat{T}_n = T_n - 2\,(A + B) + R \ . \tag{4.10}$$

Proposition 6 *Under Assumptions D, I, M1—M3, R and B are both $O_{\mathbb{P}_m}(h^{p/2})$ uniformly for $m(\cdot)$ in $C_p(L,s)$, and $A = O_{\mathbb{P}_m}(\sqrt{nh^p}\mathbb{E}^{1/2}\delta^2(X))$ uniformly for $m(\cdot)$ in $C_p(L,s)$.*

Proof of Proposition 6: To simplify notations, we consider the case where $d = 1$. By Assumptions M1–M3 , $|e(X_i)| = O_{\mathbb{P}_m}(1/\sqrt{n})$ uniformly in $m(\cdot)$ and i. Thus

$$|R| = O_{\mathbb{P}_m}(nK^{p/2})^{-1} \sum_{k\in\mathcal{K}} N_k = O_{\mathbb{P}_m}(h^{p/2}),$$

uniformly for $m(\cdot)$ in $C_p(L,s)$. Under Assumptions M1 and M2, a standard Taylor expansion yields

$$e(X_i) = \left(\widehat{\theta}_n - \theta^*\right)' \mu_1(X_i) + \|\widehat{\theta}_n - \theta^*\|^2 \mu_2(X_i) \ , \tag{4.11}$$

where $\mu_1(X_i) = \mu_\theta(X_i,\theta^*)$ depends only on X_i and $\mu_2(X_i)$ depends on X_i and $\widehat{\theta}_n$. Therefore $B = \left(\widehat{\theta}_n - \theta^*\right)' B_1 + \|\widehat{\theta}_n - \theta^*\|^2 B_2$, where $B_1 = \langle \varepsilon, \mu_1(X)\rangle$ and $B_2 = \langle \varepsilon, \mu_2(X)\rangle$. Now $\mathbb{E}(B_1) = 0$ and

$$\begin{aligned}
\mathbb{E}(B_1^2) &= \frac{1}{2K^p} \sum_{k\in\mathcal{K}} \mathbb{E}\left[\frac{\mathbb{I}[N_k > 1]}{N_k^2} \sum_{\{X_i,X_j,X'_j\}\in I_k, i\neq j, i\neq j'} \varepsilon_i^2 \mu_1(X_j)\mu_1(X_{j'})\right] \\
&= \frac{O(1)}{2K^p} \sum_{k\in\mathcal{K}} \mathbb{E}\left[\frac{\mathbb{I}[N_k > 1]\,(N_k-1)^2}{N_k}\right] = O(nh^p),
\end{aligned}$$

using M2–i. Similarly,

$$\begin{aligned}
\mathbb{E}|B_2| &\leq \frac{O(1)}{\sqrt{2}K^{p/2}} \sum_{k\in\mathcal{K}} \frac{\mathbb{I}[N_k > 1]}{N_k} \sum_{\{X_i,X_j\}\in I_k, i\neq j} \mathbb{E}^k|\varepsilon_i| \\
&= \frac{O(1)}{\sqrt{2}K^{p/2}} \sum_{k\in\mathcal{K}} \mathbb{E}\left[\mathbb{I}[N_k > 1]\,(N_k-1)\right] = O(nh^{p/2}).
\end{aligned}$$

As $\sqrt{n}\left(\widehat{\theta}_n - \theta^*\right) = O_{\mathbb{P}_m}(1)$ uniformly in $m(\cdot)$, we obtain $B = O_{\mathbb{P}_m}(h^{p/2})$ uniformly in $m(\cdot)$.
From (4.11), $A = \left(\widehat{\theta}_n - \theta^*\right)' A_1 + \|\widehat{\theta}_n - \theta^*\|^2 A_2$, where $A_1 = \langle \delta(X), \mu_1(X)\rangle$ and $A_2 = \langle \delta(X), \mu_2(X)\rangle$. Now,

$$\mathbb{E}|A_1| \leq \frac{O(1)}{\sqrt{2}K^{p/2}} \sum_{k\in\mathcal{K}} \mathbb{E}(N_k - 1)\,\mathbb{I}[N_k > 1]\,\mathbb{E}^k|\delta(X)| \leq O(nh^{p/2})\mathbb{E}|\delta(X)| \leq O(nh^{p/2})\mathbb{E}^{1/2}\delta^2(X).$$

Similarly, $\mathbb{E}|A_2| = O(nh^{p/2})\mathbb{E}^{1/2}\delta^2(X)$. Since $\sqrt{n}\left(\widehat{\theta}_n - \theta^*\right) = O_{\mathbb{P}_m}(1)$ uniformly in $m(\cdot)$, we obtain $A = O_{\mathbb{P}_m}(\sqrt{nh^p}\,\mathbb{E}^{1/2}\delta^2(X))$ uniformly in $m(\cdot)$. □

Proposition 7 shows that projections on the set of indicator functions $\mathbb{I}(x \in I_k)$, $k \in \mathcal{K}$, can be used to approximate accurately enough the magnitude of the L_2-norm of $m(\cdot)$.

Proposition 7 *Under Assumption D,*

$$\mathbb{E}^{1/2}\left[\mathrm{Proj}_{\mathcal{K}}^2 m(X)\right] \geq C_1 \left(\mathbb{E}^{1/2} m^2(X) - h^s\right) \ ,$$

for any $m(\cdot) \in C_p(L,s)$ and h small enough, where $C_1 > 0$ depends only upon L, s and $f(\cdot)$.

A detailed proof is given in Appendix B, because it is new for multivariate random designs. It proceeds by proper modifications of the arguments used in Ingster (1993, pp. 253 sqq.) .

The following Proposition 8 gives some bounds for the unconditional mean and variance of T_n.

Proposition 8 *Let Assumptions D and I hold, and K be as in Theorem 2. Then, for any $m(\cdot) \in H_m(\kappa\tilde{\rho}_n)$ with $\kappa > \lambda^s$ and n large enough,*

$$\begin{array}{lll} \mathbb{E}_m T_n & \geq & C_2 n h^{p/2} \left(\mathbb{E}^{1/2} \delta^2(X) - h^s \right)^2 & \text{for some } C_2 > 0 \ , \\ \operatorname{Var}_m(T_n) & \leq & \mathbb{E}_m v^2(K) + C_3 n h^p \mathbb{E}_m \, \delta^2(X) + C_4 n \mathbb{E}_m^2 \, \delta^2(X) & \text{for some } C_3, C_4 > 0 \ . \end{array}$$

Proof of Proposition 8: Let $w_k = \langle U^*, U^* \rangle_k$. By Lemmas 2 and 3,

$$\begin{aligned} \mathbb{E}_m T_n &= \frac{1}{\sqrt{2}K^{p/2}} \sum_{k \in \mathcal{K}} \mathbb{E}_m \omega_k = \frac{1}{\sqrt{2}K^{p/2}} \sum_{k \in \mathcal{K}} \mathbb{E}\left[(N_k - 1) \mathbb{I}(N_k > 1) \right] \left(\mathbb{E}^k \delta(X) \right)^2 \\ &\geq \frac{n h^{p/2}}{2\sqrt{2}} \mathbb{E}\left[\operatorname{Proj}_{\mathcal{K}}^2 \delta(X) \right] \geq \frac{C_1}{2\sqrt{2}} n h^{p/2} \left(\mathbb{E}^{1/2} \delta^2(X) - h^s \right)^2 \ , \end{aligned}$$

for n large enough, using Proposition 7 and $\mathbb{E}^{1/2}\delta^2(X) - h^s > 0$ as $m(\cdot) \in H_1(\kappa\tilde{\rho}_n)$ with $\kappa > \lambda^s$. Because the ω_k's are uncorrelated given $S_{\mathcal{K}}$ by Lemma 2,

$$\operatorname{Var}_m(T_n) = \frac{1}{2K^p} \sum_{k \in \mathcal{K}} \mathbb{E}_m \left[\mathbb{I}(N_k > 1) \operatorname{Var}_m(\omega_k | S_{\mathcal{K}}) \right] + \frac{1}{2K^p} \operatorname{Var}_m \left[\sum_{k \in \mathcal{K}} \mathbb{I}(N_k > 1) \mathbb{E}_m \left(\omega_k \, | S_{\mathcal{K}} \right) \right] . \quad (4.12)$$

Using Lemmas 2 and 3, Assumption I and $\mathbb{P}(X \in I_k) \geq \underline{f} h^p$ uniformly in k, we get

$$\begin{aligned} \frac{1}{2K^p} \sum_{k \in \mathcal{K}} \mathbb{E}\left[\mathbb{I}(N_k > 1) \operatorname{Var}_m(\omega_k | S_{\mathcal{K}}) \right] &\leq \mathbb{E}_m v^2(K) + 2h^p \sum_{k \in \mathcal{K}} \mathbb{E} N_k \left(\mathbb{E}^k \delta(X) \right)^2 \left[\mathbb{E}^k \varepsilon^2 + \mathbb{E}^k \delta^2(X) \right] \\ &\leq \mathbb{E}_m v^2(K) + C_5 n h^p \mathbb{E}\left[\operatorname{Proj}_{\mathcal{K}}^2 \delta(X) \right] + C_6 n \mathbb{E}^2 \left[\operatorname{Proj}_{\mathcal{K}} \delta^2(X) \right] \ , \end{aligned}$$

$$\begin{aligned} & \frac{1}{2K^p} \operatorname{Var} \left(\sum_{k \in \mathcal{K}} \mathbb{I}(N_k > 1) \mathbb{E}_m \left[\omega_k \, | S_{\mathcal{K}} \right] \right) \\ & \leq \frac{1}{2K^p} \sum_k \left(\mathbb{E}^k \delta(X) \right)^4 \operatorname{Var}\left((N_k - 1) \mathbb{I}(N_k > 1) \right) \\ & \quad + \frac{1}{2K^p} \sum_{k \neq k'} \left(\mathbb{E}^k \delta(X) \right)^2 \left(\mathbb{E}^{k'} m(X) \right)^2 \operatorname{Cov}\left((N_k - 1) \mathbb{I}(N_k > 1), (N_{k'} - 1) \mathbb{I}(N_{k'} > 1) \right) \\ & \leq C_7 n \mathbb{E}^2 \left[\operatorname{Proj}_{\mathcal{K}}^2 \delta(X) \right] + C_8 n h^p \mathbb{E}^2 \left[\operatorname{Proj}_{\mathcal{K}}^2 \delta(X) \right] \ , \end{aligned}$$

where we use the properties of $\mathrm{Proj}_{\mathcal{K}}$. Combining inequalities, as $nh^p \to \infty$, we obtain

$$\begin{aligned}\mathrm{Var}\,(T_n) &\leq \mathbb{E}_m v^2(K) + C_5 nh^p \mathbb{E}\left[\mathrm{Proj}^2_{\mathcal{K}}\delta(X)\right] + C_6 n\mathbb{E}^2\left[\mathrm{Proj}_{\mathcal{K}}\delta^2(X)\right] \\ &\quad + C_7 n\mathbb{E}^2\left[\mathrm{Proj}^2_{\mathcal{K}}\delta(X)\right] + C_8 nh^p\mathbb{E}^2\left[\mathrm{Proj}^2_{\mathcal{K}}\delta(X)\right] \\ &\leq \mathbb{E}_m v^2(K) + C_3 nh^p \mathbb{E}\,\delta^2(X) + C_4 n\mathbb{E}^2\,\delta^2(X)\,.\square\end{aligned}$$

Main proof

Part *i.* From (4.10), Proposition 6 and as $A = 0$ under H_0, it suffices to show that $T_n/v_n \xrightarrow{d} N(0,1)$. Assume that some ordering (denoted by $\leq$) is given for the set $\mathcal{K}$ of indexes k. Let $J_1, \ldots, J_n$ be any (random) rearrangement of the indices $i = 1, \ldots, n$, such that $X_{J_i} \in I_k$ iff $\sum_{\ell<k} N_\ell < J_i \leq \sum_{\ell\leq k} N_\ell$. Let $\mathcal{F}_{n,k} = \left\{S_{\mathcal{K}}, Y_{J_i} : \sum_{\ell<k} N_\ell < J_i \leq \sum_{\ell\leq k} N_\ell\right\}$. Under H_0, $\left\{T_{n,k} = \sum_{k'\leq k}\omega_{k'}/\sqrt{2K^p}, \mathcal{F}_{n,k}\right\}$ is a zero-mean martingale array. It is then sufficient to show that

$$v_n^{-2}\sum_{k\in\mathcal{K}}\mathbb{E}_0\left[\omega_k^2/(2K^p)|\mathcal{F}_{n,k-1}\right] \xrightarrow{p} 1, \tag{4.13}$$

$$v_n^{-2}\sum_{k\in\mathcal{K}}\mathbb{E}_0\left[\omega_k^2/(2K^p)\,\mathbb{I}\left(\left|\omega_k/\sqrt{2K^p}\right| > \eta v_n\right)|\mathcal{F}_{n,k-1}\right] \xrightarrow{p} 0, \qquad \forall\eta > 0 \tag{4.14}$$

from Corollary 3.1 in Hall and Heyde (1980), see also the remarks following it. Now

$$\frac{1}{2K^p}\sum_{k\in\mathcal{K}}\mathbb{E}_0\left[\omega_k^2|\mathcal{F}_{n,k-1}\right] = \frac{1}{2K^p}\sum_{k\in\mathcal{K}}\mathbb{E}_0\left[\omega_k^2|S_{\mathcal{K}}\right] = \frac{1}{2K^p}\sum_{k\in\mathcal{K}}\frac{2(N_k-1)}{N_k}\mathbb{E}^k U^{*2} = v^2(K)$$

from Lemma 2, so that (4.13) follows from Proposition 5. Now (4.14) is implied by

$$v_n^{-4}\frac{1}{4K^{2p}}\sum_{k\in\mathcal{K}}\mathbb{E}_0\left[\omega_k^4|\mathcal{F}_{n,k-1}\right] \xrightarrow{p} 0.$$

By Assumption I, straightforward computations lead to

$$\frac{1}{K^p}\sum_{k\in\mathcal{K}}\mathbb{E}_0\left[\omega_k^4|\mathcal{F}_{n,k-1}\right] \leq \frac{4}{K^p}\sum_{k\in\mathcal{K}}\left(\mathbb{E}^k\varepsilon^4\right)^2 = O(1).$$

By Proposition 5, (4.14) follows. $\square$

Part *ii.* As v_n^2 is bounded in probability from below uniformly in $m(\cdot)$ from Proposition 5, (4.10) and Proposition 6 yields

$$\mathbb{P}_m\left(v_n^{-1}\widehat{T}_n \leq z_\alpha\right) \leq \mathbb{P}_m\left(T_n \leq z'_\alpha + 2M\sqrt{nh^p}\,\mathbb{E}^{1/2}\delta^2(X)\right) + o(1),$$

for any $M > 0$ and some $z'_\alpha > 0$, where the $o(1)$ is uniform in $m(\cdot)$. But

$$\begin{aligned}\mathbb{P}_m\left(T_n \leq z'_\alpha + 2M\sqrt{nh^p}\,\mathbb{E}^{1/2}\delta^2(X)\right) &= \mathbb{P}_m\left[-(T_n - \mathbb{E}_m T_n) \geq \mathbb{E}_m T_n - z'_\alpha - 2M\sqrt{nh^p}\,\mathbb{E}^{1/2}\delta^2(X)\right]^2 \\ &\leq \frac{\operatorname{Var}_m T_n}{\left[\mathbb{E}_m T_n - z'_\alpha - 2M\sqrt{nh^p}\,\mathbb{E}^{1/2}\delta^2(X)\right]^2},\end{aligned}$$

if $\mathbb{E}_m T_n - z'_\alpha - 2M\sqrt{nh^p}\,\mathbb{E}^{1/2}\delta^2(X) > 0$. It is then sufficient to show that κ can be chosen so that

$$\begin{aligned}\mathbb{E}_m T_n - z'_\alpha - 2M\sqrt{nh^p}\,\mathbb{E}^{1/2}\delta^2(X) &> 0, \\ \frac{\operatorname{Var}_m T_n^*}{\left[\mathbb{E}_m T_n^* - z'_\alpha - 2M\sqrt{nh^p}\,\mathbb{E}^{1/2}\delta^2(X)\right]^2} &\leq \beta,\end{aligned} \tag{4.15}$$

uniformly for $m(.)$ in $H_m(\kappa\tilde{\rho}_n)$. Proposition 8 gives that for any $m(.)$ in $H_m(\kappa\tilde{\rho}_n)$ and n large enough

$$\frac{\mathbb{E}_m T_n - z'_\alpha - 2M\sqrt{nh^p}\,\mathbb{E}^{1/2}\delta^2(X)}{nh^{p/2}\,\mathbb{E}\delta^2} \geq C_2\left[1 - \frac{\lambda^s}{\kappa}\right]^2 - \frac{z'_\alpha}{\kappa^2\lambda^{p/2}} - 2M\frac{1}{\kappa\sqrt{n}\tilde{\rho}_n},$$

and this lower bound is increasing in κ and positive for κ large enough. Proposition 8 also yields

$$\begin{aligned}\frac{\operatorname{Var}_m T_n}{\left(nh^{p/2}\,\mathbb{E}\delta^2(X)\right)^2} &\leq \frac{\mathbb{E}v^2(K) + C_3 nh^p\,\mathbb{E}\delta^2(X) + C_4 n\,\mathbb{E}^2\delta^2(X)}{n^2h^p\,\mathbb{E}^2\delta^2(X)} \\ &\leq \frac{\mathbb{E}v^2(K)}{\kappa^4\lambda^p} + \frac{C_3}{\kappa^2 n\tilde{\rho}_n^2} + \frac{C_4}{nh^p},\end{aligned}$$

and this upper bound is bounded for n large enough because of Proposition 5, and decreasing in κ. Hence (4.15) can be made smaller than β uniformly for $m(.)$ in $H_m(\kappa\tilde{\rho}_n)$ by taking κ large enough. □

4.3 Proof of Theorem 4

Without loss of generality, we consider the case of testing for a pure noise model, that is $\mathcal{M} = \{0\}$. Then

$$z(\xi) = z_0(\xi) + z_1(\xi) = (1/\sqrt{n})\sum_{i=1}^n \varepsilon_i w(X_i,\xi) + (1/\sqrt{n})\sum_{i=1}^n m(X_i)w(X_i,\xi).$$

Consider the a priori Π_{1n} defined in Theorem 1's proof, i.e. the measure defined by the random functions

$$m_n(\cdot) = \delta_n(\cdot) = \lambda\rho_n h_n^{p/2}\sum_k B_k\varphi_k(\cdot),$$

where $B_1, \ldots, B_{K_n}$ are independent Rademacher variables and $h_n = \lambda \rho_n^{1/s}$, and further assume that $\varphi(\cdot)$ has r-first zero moments. We have

$$\mathbb{E}_{\Pi_{1n}} z_1^2(\xi) = \frac{\lambda^2 \rho_n^2 h_n^p}{n} \sum_{i,j=1}^{n} \sum_{k \in \mathcal{K}} \mathbb{E}[w(X_i,\xi) w(X_j,\xi) \varphi_k(X_i) \varphi_k(X_j)]$$

uniformly in ξ. Now

$$\mathbb{E}\left[w^2(X_i,\xi)\varphi_k^2(X_i)\right] \le F \sup_{x \in [0,1]^p, \xi \in \Xi} w(x,\xi) \int \varphi^2(x)\, dx$$

and

$$\mathbb{E}[w(X_i,\xi)\varphi_k(X_i)] = h^{p/2} \int w(lk + hu, \xi) f(lk + hu) \varphi(u)\, du = O(h^{r+p/2}) \ .$$

Hence, we have uniformly in ξ

$$\mathbb{E}_{\Pi_{1n}} z_1^2(\xi) = \lambda^2 \rho_n^2 O(1) + \lambda^2 \rho_n^2 n h^{2r+p} O(1) \ .$$

Because r can be chosen as large as desired, $\mathbb{E}_{\Pi_{1n}} z_1^2(\xi) = o(1)$ whenever $\rho_n = O(n^{-a})$, for any $a > 0$. Under the same assumptions, $\mathbb{E}_{\Pi_{1n}} |z_1(\xi)|^q = o(1)$ for any $1 \le q < 2$ from Hölder inequality, and $\mathbb{E}_{\Pi_{1n}} |z_1(\xi)|^q = o(1)$ for any $2 < q < \infty$ from the Khinchin-Kahane inequality, see e.g. de la Peña and Giné (1999). Hence,

$$\mathbb{E}_{\Pi_{1n}} \int |z_1|^q(\xi)\ d\nu(\xi) = o(1) \ .$$

Thus,

$$\begin{aligned}
\sup_{H_1(\rho_n)} \mathbb{P}_m(I_{n,q} \le u_{\alpha,q}) &\ge \int \mathbb{P}_m(I_{n,q} \le u_{\alpha,q})\ d\Pi_{1n}(m) \\
&\ge \int \mathbb{P}_m\left(\left[\int |z_0(\xi)|^q\, d\nu(\xi)\right]^{1/q} \le u_{\alpha,q}\right) d\Pi_{1n}(m) + o(1) \\
&\ge \mathbb{P}_0(I_{n,q} \le u_{\alpha,q}) + o(1) = 1 - \alpha + o(1) \ .\square
\end{aligned}$$

Appendix A: Auxiliary results

Lemma 2 *Let $\omega_k = \langle U^*, U^* \rangle_k$. Under Assumptions I, for any $k \in \mathcal{K}$ such that $N_k > 1$,*

$$\begin{aligned}
\mathbb{E}_m[\omega_k | S_{\mathcal{K}}] &= (N_k - 1)\left(\mathbb{E}^k \delta(X)\right)^2 , \\
\mathrm{Var}_m[\omega_k | S_{\mathcal{K}}] &= \frac{2(N_k-1)}{N_k}\left(\mathbb{E}^k U^{*2}\right)^2 + \frac{4(N_k-1)(N_k-2)}{N_k}\left(\mathbb{E}^k \delta(X)\right)^2 \mathbb{E}^k U^{*2} \\
&\quad - \frac{2(N_k-1)(2N_k-3)}{N_k}\left(\mathbb{E}^k \delta(X)\right)^4 \ .
\end{aligned}$$

Moreover, the ω_k's are uncorrelated given $S_{\mathcal{K}}$.

Proof of Lemma 2: Conditionally upon $S_{\mathcal{K}}$, the X_i's are independent and identically distributed within each cell. The expression of the conditional expectation then follows from $\mathbb{E}^k U^* = \mathbb{E}^k \delta(X)$. The other claims are easily checked. □

Lemma 3 *Under Assumptions D and I, if $nh^p \to \infty$, then for n large enough,*

$$\begin{aligned}
\mathbb{E}[(N_k - 1)\,\mathbb{I}(N_k > 1)] &\geq \tfrac{n}{2}\mathbb{P}(X \in I_k) & \forall k \in \mathcal{K}\ ,\\
\operatorname{Var}\left[(N_k - 1)\,\mathbb{I}(N_k > 1)\right] &\leq 2n\mathbb{P}(X \in I_k) & \forall k \in \mathcal{K}\ ,\\
\operatorname{Cov}\left[(N_k - 1)\,\mathbb{I}(N_k > 1), (N_{k'} - 1)\,\mathbb{I}(N_{k'} > 1)\right] &\leq 2n\mathbb{P}(X \in I_k)\mathbb{P}(X \in I_{k'}) & \forall k \neq k' \in \mathcal{K}\ .
\end{aligned}$$

Proof of Lemma 3: Note that $(N_k - 1)\mathbb{I}(N_k > 1) = N_k - 1 + \mathbb{I}(N_k = 0)$. As $\mathbb{I}(N_k = 1)$ is a Bernoulli random variable, then by Assumptions D and I, for n large enough,

$$\begin{aligned}
\mathbb{E}[(N_k - 1)\,\mathbb{I}(N_k > 1)] &= n\mathbb{P}(X \in I_k) - 1 + (1 - \mathbb{P}(X \in I_k))^n \geq \frac{n}{2}\mathbb{P}(X \in I_k)\ ,\\
\operatorname{Var}\left[(N_k - 1)\,\mathbb{I}(N_k > 1)\right] &\leq n\mathbb{P}(X \in I_k)\left[1 - \mathbb{P}(X \in I_k)\right] + 1/4 - 2\mathbb{E}(N_k)\,\mathbb{P}(N_k = 0) \leq 2n\mathbb{P}(X \in I_k)\ .
\end{aligned}$$

The covariance equals

$$\operatorname{Cov}(N_k, N_{k'}) + \operatorname{Cov}\left(\mathbb{I}(N_k = 0), \mathbb{I}(N_{k'} = 0)\right) + \operatorname{Cov}\left(N_k, \mathbb{I}(N_{k'} = 0)\right) + \operatorname{Cov}\left(N_{k'}, \mathbb{I}(N_k = 0)\right)\ .$$

The first item is $-\mathbb{E}(N_k)\mathbb{E}(N_{k'})$ and the second item is

$$(1 - \mathbb{P}(X \in I_k) - \mathbb{P}(X \in I_{k'}))^n - (1 - \mathbb{P}(X \in I_k))^n\,(1 - \mathbb{P}(X \in I_{k'}))^n\ .$$

They are both negative. Moreover,

$$\operatorname{Cov}\left(N_k, \mathbb{I}(N_{k'} = 0)\right) = n\,(1 - \mathbb{P}(X \in I_{k'}))^{n-1}\,\mathbb{P}(X \in I_k)\mathbb{P}(X \in I_{k'}) \leq n\mathbb{P}(X \in I_k)\mathbb{P}(X \in I_{k'})\ .\square$$

Lemma 4 *Under Assumption I, if $\frac{n}{K^p \log K^p} \to \infty$,*

$$\mathbb{P}\left(\min_{k \in \mathcal{K}} \mathbb{I}(N_k > 1) = 1\right) \to 1 \qquad \text{and} \qquad \max_{k \in \mathcal{K}} \left|\frac{N_k}{\mathbb{E}N_k} - 1\right| = o_{\mathbb{P}}(1)\ .$$

Proof of Lemma 4: As N_k is a binomial random variable, the Bernstein inequality yields

$$I\!P\left[\left|\frac{N_k}{I\!EN_k}-1\right|\geq t\right]=I\!P\left[\left|\frac{N_k-I\!EN_k}{\sqrt{n}}\right|\geq\frac{tI\!EN_k}{\sqrt{n}}\right]\leq 2\exp\left[-\frac{t^2}{2\,(1+t/3)}I\!EN_k\right]\;,$$

for any $t>0$, see Shorack and Wellner (1986, p. 440). This yields

$$I\!P\left[\min_{k\in\mathcal{K}} 1\!I(N_k>1)=0\right]\leq\sum_{k\in\mathcal{K}} I\!P[N_k=0]\leq\sum_{k\in\mathcal{K}} I\!P\left[\left|\frac{N_k}{I\!EN_k}-1\right|\geq 1\right]\leq 2K^p\exp\left[-\frac{3}{8}\mathrm{f}\frac{n}{K^p}\right]\to 0\;,$$

as $I\!EN_{kK}\geq \mathrm{f}n/K^p$ under Assumption D, and

$$I\!P\left(\max_{K\in\mathcal{K}}\left|\frac{N_k}{I\!EN_k}-1\right|\geq t\right)\leq\sum_{K\in\mathcal{K}} I\!P\left[\left|\frac{N_k-I\!EN_k}{\sqrt{n}}\right|\geq\frac{tI\!EN_k}{\sqrt{n}}\right]\leq 2K^p\exp\left[-\frac{t^2}{2\,(1+t/3)}\mathrm{f}\frac{n}{K^p}\right]\to 0\;,$$

for any $t>0$, if $\frac{n}{K^p\log K^p}\to\infty$. □

Appendix B: Proof of Proposition 7

Step 1. Let $s'=[s+1]$, assume that $K=K_n$ is larger than s', and define

$$\kappa(0)=0\;,\;\kappa(1)=s'\;,\;\ldots\;,\;\kappa([K/s']-1)=([K/s']-1)s'\;,\;\kappa([K/s'])=K\;,$$

where [.] is the integer part. This gives, with $\ell=\ell_n=[K/s']$,

$$s'\leq\kappa(r+1)-\kappa(r)\leq 2s'\;,\;r=0,\ldots,\ell-1\;. \tag{B.1}$$

Let $\mathcal{Q}$ be the set of vectors whose generic element is q with p components in $\{\kappa(0),\ldots,\kappa(\ell-1)\}$, i.e.

$$q=(\kappa(r_{1,q}),\ldots,\kappa(r_{p,q}))^\top\;,\;r_{j,q}=0,\ldots,\ell-1\;,\;j=1,\ldots,p\;.$$

Consider the following subsets of $[0,1]^p$, which define a partition up a to negligible set:

$$\Delta_q(h)=\Delta_q=\prod_{j=1}^{p}[\kappa(r_{j,q})h,\kappa(r_{j,q}+1)h)\;,\;q\in\mathcal{Q}\;. \tag{B.2}$$

Let $P_{m,q}(.)$ be the Taylor expansion of order $[s]$ of $m(\cdot)$ around qh. Because $m(.)$ is in $C_p(L,s)$ and by definition of Δ_q, we get by (B.1) that $|m(x)-P_{m,q}(x)|\leq C_{s,L}h^s$ for any x in Δ_q for some constant $C_{s,L}$. If $P_m(.)$ is such that $P_m(.)=P_{m,q}(.)$ on Δ_q, we have

$$\|m-P_m\|_2^2\leq I\!E\left[\sum_{q\in\mathcal{Q}}C_{s,L}^2h^{2s}\,1\!I(X\in\Delta_q)\right]=C_{s,L}^2h^{2s}\;.$$

Assume that we have been able to establish that, for some constant $C_{s,f}$,

$$\|\mathrm{Proj}_{\mathcal{K}} P_m\|_2 \geq C_{s,f}\|P_m\|_2 \ . \tag{B.3}$$

Because $\mathrm{Proj}_{\mathcal{K}}$ is contractant, this would give the desired result, as

$$\begin{aligned} \|\mathrm{Proj}_{\mathcal{K}} m\|_2 &\geq \|\mathrm{Proj}_{\mathcal{K}} P_m\|_2 - \|\mathrm{Proj}_{\mathcal{K}}(m - P_m)\|_2 \geq \|\mathrm{Proj}_{\mathcal{K}} P_m\|_2 - \|m - P_m\|_2 \\ &\geq C_{s,f}\|(P_m - m) + m\|_2 - C_{s,L}h^s \geq C_{s,f}\|m\|_2 - (1 + C_{s,f})C_{s,L}h^s \ . \end{aligned}$$

Inequality (B.3) will follow by summation over $q \in \mathcal{Q}$ of inequalities of the type

$$\mathbb{E}\left[(\mathrm{Proj}_{\mathcal{K}} P(X))^2 \, 1\!\!1(X \in \Delta_q)\right] \geq C_{s,f}^2 \mathbb{E}\left[P^2(X) 1\!\!1(X \in \Delta_q)\right] \ , \tag{B.4}$$

for any polynomial functions $P(.)$ of degree $[s]$.

Step 2. Let us now give a matrix expression of (B.4). For any $\beta = (\beta_1, \ldots, \beta_p) \in \mathbb{N}^p$ with $\sum_{j=1}^p \beta_j \leq [s]$, let $x^{(\beta)} = \prod_{j=1}^p x_j^{\beta_j}$. Every polynomial functions of degree $[s]$ is completely determined by the coefficients $a = \left(a_\beta, \sum_{j=1}^p \beta_j \leq [s]\right)$ (with a suitable ordering for the index β in $\mathbb{N}^p$) such that

$$P(x) = \sum_{\beta, \sum \beta_j \leq [s]} a_\beta \left(\frac{x - qh}{h}\right)^{(\beta)} \ .$$

This gives, for x in Δ_q,

$$\mathrm{Proj}_{\mathcal{K}} P(x) = \sum_{I_k \subset \Delta_q} \sum_{\beta, \sum \beta_j \leq [s]} a_\beta \frac{1}{\mathbb{P}(X \in I_k)} \mathbb{E}\left[\left(\frac{X - qh}{h}\right)^{(\beta)} 1\!\!1(X \in I_k)\right] 1\!\!1(x \in I_k) \ .$$

Let $\nu_1 = \mathrm{Card}\{I_k \subset \Delta_q\}$, $\nu_2 = \mathrm{Card}\left\{\sum_{j=1}^p \beta_j \leq [s]\right\}$ and $B_q(h)$ be the $\nu_1 \times \nu_2$ matrix with typical element indexed by k and β

$$\frac{1}{\mathbb{P}(X \in I_k)} \mathbb{E}\left[\left(\frac{X - qh}{h}\right)^{(\beta)} 1\!\!1(X \in I_k)\right] \ , \ I_k \subset \Delta_q \ , \ \sum_{j=1}^p \beta_j \leq [s] \ .$$

Let $\Pi_q(h) = \mathrm{Diag}(\mathbb{P}(X \in I_k) \, , \, I_k \subset \Delta_q)$. Because the density $f(\cdot)$ is bounded from below and the $\Pi_q(h)$'s are diagonal, we have (for the standard ordering for positive symmetric matrices)

$$\Pi_q(h) >> f h^p \mathrm{Id} \ .$$

Hence the l.h.s. of (B.4) writes

$$\mathbb{E}\left[(\mathrm{Proj}_{\mathcal{K}} P(X))^2 \, 1\!\!1(X \in \Delta_q)\right] = a^\top B_q^\top(h)\Pi_q(h)B_q(h)a \geq f h^p a^\top B_q^\top(h)B_q(h)a \ .$$

Let $D_q(h)$ be the square ν_2 matrix with typical element, indexed by β and β',

$$\frac{1}{\mathbb{P}(X \in \Delta_q)} \mathbb{E}\left[\left(\frac{X-qh}{h}\right)^{(\beta+\beta')} \mathbb{I}(X \in \Delta_q)\right] \ , \ \sum_{j=1}^{p} \beta_j \leq [s] \ , \ \sum_{j=1}^{p} \beta'_j \leq [s] \ .$$

Since the density $f(.)$ is bounded from above, we have for the r.h.s. of (B.4)

$$\mathbb{E}\left[P^2(X)\mathbb{I}(X \in \Delta_q)\right] \leq \mathbb{P}(X \in \Delta_q) a^\top D_q(h) a \leq F(2s'h)^p a^\top D_q(h) a \ ,$$

using (B.1). Therefore, (B.4) holds as soon as, for any a, q, and h small enough,

$$a^\top D_q(h) a \leq C_{s,f} \ a^\top B_q^\top(h) B_q(h) a \ . \tag{B.5}$$

Step 3. We can limit ourselves to establish (B.5) for vectors a with norm 1 by homogeneity. This step works by showing that the matrices $D_q(h)$ and $B_q(h)$ converge (uniformly with respect to q) to some matrices D_q and B_q, B_q being of full rank for any q. Moreover the number of matrices B_q and D_q, $q \in \mathcal{Q}$, will be finite. If the B_q's are of full rank, a possible choice of $C_{s,f}$ in (B.5) is

$$C_{s,f} = \max_{q \in \mathcal{Q}} \sup\{a^\top D_q a \ : \ a^\top B_q^\top B_q a \leq 1\} + 1 \ .$$

Let us now determine the limits B_q. The entries of $B_q(h)$ are

$$\begin{aligned}
&\frac{1}{\mathbb{P}(X \in I_k)} \mathbb{E}\left[\left(\frac{X-qh}{h}\right)^{(\beta)} \mathbb{I}(X \in I_k)\right] \\
&= \frac{1}{\int_{[0,1]^p} f(kh+hu)\,du} \int_{[0,1]^p} (k-q+u)^{(\beta)} f(kh+hu)\,du \\
&= \frac{1}{f(kh)+o(1)} \int_{[0,1]^p} (k-q+u)^{(\beta)} (f(kh)+o(1))\,du \to \int_{[0,1]^p} (k-q+u)^{(\beta)} du \ ,
\end{aligned}$$

uniformly in k, q, since $f(.)$ is bounded away from 0 and uniformly continuous on $[0,1]^p$ by Assumption D. We now check that the number of limits B_q, q in $\mathcal{Q}$ is finite. The definitions (4.3) and (B.2) require that $I_k = kh + h[0,1)^p \subset \Delta_q = q + h[0,1)^p$, which implies that $k = (k_1, \ldots, k_p)^\top$ and $q = (\kappa(r_{1,q}), \ldots, \kappa(r_{p,q}))^\top$ are such that $\kappa(r_{j,q}) \leq k_j < \kappa(r_{j,q}+1)$, independently of h . Therefore,

$$0 \leq k_j - \kappa(r_{j,q}) < \kappa(r_{j,q}+1) - \kappa(r_{j,q}) \leq 2s' \ , \ j = 1, \ldots, p \ . \tag{B.6}$$

As $\sum_{j=1}^p \beta_j \leq [s]$, the number of B_q, q in $\mathcal{Q}$, is bounded by $(2s')^{[s]^p}$ independently of K. It can be similarly shown that the $D_q(h)$'s converge, uniformly in q, to some matrices D_q with entries

$$\int_{\prod_{j=1}^p [0, \kappa(r_{j,q}+1) - \kappa(r_{j,q}))} u^{(\beta+\beta')}\,du \ ,$$

which are also in finite number by (B.1) and (B.6).
To finish the proof, we need to check that all the B_q's are of full rank. To this purpose assume that there exists q in $\mathcal{Q}$ and $a = (a_\beta \ , \ \sum_{j=1}^p \beta_j \leq [s])$ with $B_q a = 0$, i.e. for all k such that $I_k \subset \Delta_q$,

$$\sum_{\beta, \sum_{j=1}^p \beta_j \leq [s]} a_\beta \int_{[0,1]^p} (k - q + u)^{(\beta)}\, du \quad = \quad \int_{k-q+[0,1]^p} \sum_{\beta, \sum_{j=1}^p \beta_j \leq [s]} a_\beta u^{(\beta)}\, du = 0 \ .$$

This implies that $P(x) = \sum_\beta a_\beta x^{(\beta)}$ of degree $[s]$ is such that,

$$\int_{\pi+[0,1]^p} P(u) du = 0 \ , \ 0 \leq \pi_j < s' \ , \ j = 1, \ldots, p \ , \qquad \text{(B.7)}$$

with $\pi = (\pi_1, \ldots \pi_p)^\top$ satisfying the conditions in (B.1) and (B.6). We now use an induction argument. Let $\mathcal{P}(p)$ be the proposition: if $P(x)$ of degree $[s]$, x in $[0,1]^p$, is such that (B.7) holds, then $P(.) = 0$. Note that $\mathcal{P}(1)$ holds, because (B.7) and the mean value theorem gives that $P(x(\pi)) = 0$ for some $x(\pi)$ in $]\pi, \pi+1[$, $\pi = 0, \ldots, s'$. Then the univariate polynomial function $P(.)$ of degree $[s]$ should have at least $[s]+1$ distinct roots, which is possible only if $P(.) \equiv 0$. We now show that $\mathcal{P}(p-1)$ implies $\mathcal{P}(p)$. Assume that $P(x)$ of degree $[s]$ with $x = (x_1, \ldots, x_p)^\top$ in $[0,1]^p$ is such that (B.7) holds. Define

$$x_{-1} = (x_2, \ldots, x_p)^\top \in [0,1]^{p-1} \ , \ P_{x_{-1}}(x_1) = P(x_1, x_{-1}) = P(x) \ .$$

Then (B.7) yields for any π_1 in $\mathbb{N}$ with $0 \leq \pi_1 < s'$,

$$\int_{u_{-1} \in \pi_{-1}+[0,1]^{p-1}} \left(\int_{\pi_1}^{\pi_1+1} P(u_1, u_{-1}) du_1 \right) du_{-1} = 0 \ , \ 0 \leq \pi_j < s' \ , \ j = 2, \ldots, p \ .$$

As a consequence, $\mathcal{P}(p-1)$ gives for any x_{-1} in $[0,1]^{p-1}$,

$$\int_{\pi_1}^{\pi_1+1} P(u_1, x_{-1}) du_1 = \int_{\pi_1}^{\pi_1+1} P_{x_{-1}}(u_1) du_1 = 0 \ , \ 0 \leq \pi_1 < s' \ .$$

Then $\mathcal{P}(1)$ shows that $P_{x_{-1}}(.) \equiv 0$ for any x_{-1} in $[0,1]^{p-1}$, which implies $\mathcal{P}(p)$. □

Appendix C

Proposition 9 *Assume $p = 1$ and $\mathcal{M} = \{0\}$. Let the c.d.f. of the design be $1 - x^{-\gamma}$, $x \geq 1$, $\gamma > 0$. If $2s > \gamma$, there exists a sequence $\{m_n(.)\}_{n \geq 1}$ of functions in $C_1(L, s)$ with $\mathbb{E}^{1/2} m_n^2(X) \geq \rho$, such that, for any α-level test t_n, $\liminf_{n \to +\infty} \mathbb{P}_{m_n}(t_n = 1) \geq 1 - \alpha$.*

Proof: Assume s is integer. Consider the $\Gamma(s+2)$ distribution c.d.f

$$I(x) = \frac{\mathbb{I}(x \geq 0)}{(s+1)!} \int_0^x t^{s+1} \exp(-t)\, dt \ ,$$

which admits s bounded continuous derivatives over $\mathbb{R}$. Let $m_n(x) = C(x-x_n)^s I(x-x_n)$, where $x_n = n^{2/\gamma}$ and C is a constant. Note that $m_n(x)$ vanishes if $x \leq x_n$. The binomial formula for derivatives yields

$$m_n^{(s)}(x) = C \sum_{k=0}^{s} I^{(k)}(x-x_n) \frac{(s!)^2}{(s_k)!(k!)^2} (x-x_n)^k \ .$$

Since the functions $(x-x_n)^k I^{(k)}(x-x_n)$, $k = 0, \ldots, s$, are bounded, $m(.)$ is in $C_1(L,s)$ for C small enough. Moreover,

$$\mathbb{E} m_n^2(X) = C^2 \gamma \int_{x_n}^{+\infty} I^2(x-x_n)(x-x_n)^{2s} x^{-\gamma-1} dx \ ,$$

and $\mathbb{E} m_n^2(X) = +\infty$ if $2s - \gamma \geq 0$, because $m_n^2(x) x^{-\gamma-1}$ is equivalent to $x^{2s-\gamma-1}$ when x grows. If $\sup X_i \leq x_n$, we have $m_n(X_i) = 0$, $i = 1, \ldots, n$, so that $Y_i = \sigma \varepsilon_i$, $i = 1, \ldots, n$. Hence,

$$\mathbb{P}_{m_n}(\tau_n = 0\, , \sup_{1 \leq i \leq n} X_i \leq x_n) = \mathbb{P}_0(\tau_n = 0\, , \sup_{1 \leq i \leq n} X_i \leq x_n) \ .$$

This leads to

$$\begin{aligned} \mathbb{P}_{m_n}(\tau_n = 1) &\geq \mathbb{P}_{m_n}(\tau_n = 1\, , \sup_{1 \leq i \leq n} X_i \leq x_n) = \mathbb{P}_0(\tau_n = 1\, , \sup_{1 \leq i \leq n} X_i \leq x_n) \\ &\geq \mathbb{P}_0(\tau_n = 1) - \mathbb{P}_0(\sup_{1 \leq i \leq n} X_i > x_n) \geq 1 - \alpha - n\mathbb{P}(X > x_n) = 1 - \alpha - n n^{-2} \ .\square \end{aligned}$$

REFERENCES

ANDREWS (1997): "A conditional Kolmogorov test," *Econometrica*, 65 (5), pp. 1097–1128.

BARAUD, Y., S. HUET and B. LAURENT (1999): "Adaptive tests of linear hypotheses by model selection," Ecole Normale Supérieure, Paris.

BIERENS, H.J. (1982): "Consistent model specification tests," *Journal of Econometrics*, 20, pp. 105–134.

BIERENS, H.J. (1990): "A consistent conditional moment test of functional form," *Econometrica*, 58(6), pp. 1443–1458.

BIERENS, H.J. and W. PLOBERGER (1997): "Asymptotic theory of integrated conditional moment tests," *Econometrica*, 65 (5), pp. 1129–1151.

BIRGE, L. and P. MASSART (1993): "Rates of convergence for minimum contrast estimators," *Probability Theory and Related Fields*, 97, pp. 113-150.

BROWN, L.D. and M.G. LOW (1996): "Asymptotic equivalence of nonparametric regression and white noise," *Annals of Statistics*, 24(6), pp. 2384–2389.

BROWN, L.D. and C.H. ZHANG (1998): "Asymptotic nonequivalence of nonparametric experiments when the smoothness index is 1/2," *Annals of Statistics*, 26(1), pp. 279–287.

DE LA PEÑA, V.H. and E. GINE (1999): *Decoupling.* New-York: Springer-Verlag.

DELGADO, M.A. (1993): "Testing the equality of nonparametric regression curves," *Statistics and Probability Letters*, 17, pp. 199–204.

DELGADO, M.A., M.A. DOMINGUEZ and P. LAVERGNE (1998): "Asymptotic and bootstrap specification tests for nonlinear in variables econometric models," Universidad Carlos III de Madrid.

DETTE, H. and A. MUNK (1998): "A simple goodness-of-fit test for linear models under a random design assumption," *Annals of the Institute of Statistical Mathematics*, 50 (2), pp. 253–275.

EFROMOVICH, S. and A. SAMAROV (1996): "Asymptotic equivalence of nonparametric regression and white noise has its limits," *Statistics and Probability Letters*, 28, pp. 143–145.

EUBANK, R.L. and J.D. HART (1993): "Commonality of cusum, von Neumann and smoothing-based goodness-of-fit tests," *Biometrika*, 80, pp. 89–98.

FAN, Y. and Q. LI (1996a): "Consistent model specification tests: omitted variables and semiparametric functional forms," *Econometrica*, 64 (4), pp. 865–890.

FAN, Y. and Q. LI (1996b):"Consistent model specification tests: nonparametric versus Bierens' tests," University of Windsor.

LI, Q. and S. WANG (1998): "A simple consistent bootstrap test for a parametric regression functional form," *Journal of Econometrics*, 87, pp. 145–165.

GODFREY, L.G. (1988): *Misspecification tests in econometrics*. Cambridge: Cambridge University Press.

GOURIEROUX, C., A. MONFORT and A. TROGNON (1984): "Pseudo-maximum likelihood methods: theory," *Econometrica*, 52, pp. 681–700.

GUERRE, E. and P. LAVERGNE (1999): "Minimax rates for nonparametric specification testing in regression models," LSTA, Université Paris 6.

GUERRE, E. and A. TSYBAKOV (1998): "Exact asymptotic minimax constants for the estimation of analytical functions in L_p," *Probability Theory and Related Fields*, 112, pp. 33–51.

HALL, P. and C.C. HEYDE (1980): *Martingale Limit Theory and Its Application*. New-York: Academic Press.

HALL, P., J.S. MARRON, M.H. NEUMANN and D.M. TITTERINGTON (1997): " Curve estimation when the design density is low," *Annals of Statistics*, 25(2), pp. 756–770.

HÄRDLE, W. and E. MAMMEN: "Comparing nonparametric versus parametric regression fits," *Annals of Statistics*, 21 (4), pp. 1926–1947.

HÄRDLE, W. and A.B. TSYBAKOV (1993): "How sensitive are average derivatives?" *Journal of Econometrics*, 58, 31–48.

HART, J.D. (1997): *Nonparametric Smoothing and Lack-of-Fit Tests*. New-York: Springer Verlag.

HONG, Y. and H. WHITE (1995): "Consistent specification testing via non-parametric series regressions," *Econometrica*, 63, pp. 1133–1160.

HOROWITZ, J.L. and V.G. SPOKOINY (1999): "An adaptive, rate-optimal test of a parametric model against a nonparametric alternative," University of Iowa.

IBRAGIMOV, I.A. and R.Z. HAS'MINSKII (1981): *Statistical estimation: asymptotic theory*. New-York: Springer-Verlag.

INGSTER, Y. I. (1993): "Asymptotically minimax hypothesis testing for nonparametric alternatives," Part I, II and III, *Mathematical Methods of Statistics*, 2, pp. 85-114, 171–189 and 249–268.

LAVERGNE, P. and Q.H. VUONG (1996): "Nonparametric selection of regressors: the nonnested case," *Econometrica*, 64 (1), pp. 207–219.

LEHMANN, E.L. (1986): *Testing statistical hypothesis*. New-York: Wiley & Sons (2nd edition).

LEPSKI, O., A. NEMIROVSKI and V. SPOKOINY (1996): "On estimation of non-smooth functionals," Weierstrass Institute for Applied Analysis and Stochastics.

NEWEY, W. K. (1994): "Series estimation of regression functionals," *Econometric Theory*, 10, pp. 1–28.

NEYMAN, J. (1937): "Smooth test for goodness of fit," *Skandinavisk Aktuarietidskrift*, 20, pp. 149-199.

POWELL, J.L. and T.M. STOKER (1996): "Optimal bandwidth choice for density-weighted averages," *Journal of Econometrics*, 75, pp. 291–316.

RUDIN, W. (1991): *Functional Analysis*. New-York: McGraw-Hill (2nd edition).

SPOKOINY, V.G. (1996): "Adaptive hypothesis testing using wavelets," *Annals of Statistics*, 24 (6), pp. 2477–2498.

STINCHCOMBE, M.B. and H. WHITE (1998): "Consistent specification testing with nuisance parameters present only under the alternative," *Econometric Theory*, 14(3), pp. 295–325.

STONE, C.J. (1982): "Optimal rates of convergence for nonparametric estimators," *Annals of Statistics*, 8 (6), pp. 1348–1360.

STUTE, W. (1997): "Nonparametric model checks for regression," *Annals of Statistics*, 25 (2), pp. 613–641.

WHITE, H. (1981): "Consequences and detection of misspecified nonlinear regression models," *Journal of the American Statistical Association*, 76, pp. 419–433.

WHITE, H. (1989): "An additional hidden unit test for neglected nonlinearity" *Proceedings of the International Joint Conference on Neural Networks*, vol. 2, pp. 451–455. New-York: IEEE Press.

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