Econometric Theory, **18**, 2002, 1139–1171. Printed in the United States of America.
DOI: 10.1017.S0266466602185069

OPTIMAL MINIMAX RATES FOR NONPARAMETRIC SPECIFICATION TESTING IN REGRESSION MODELS

EMMANUEL GUERRE
LSTA Université Paris VI

PASCAL LAVERGNE
INRA-ESR Toulouse

In the context of testing the specification of a nonlinear parametric regression function, we adopt a nonparametric minimax approach to determine the maximum rate at which a set of smooth alternatives can approach the null hypothesis while ensuring that a test can uniformly detect any alternative in this set with some predetermined power. We show that a smooth nonparametric test has optimal asymptotic minimax properties for regular alternatives. As a by-product, we obtain the rate of the smoothing parameter that ensures rate-optimality of the test. We show that, in contrast, a class of nonsmooth tests, which includes the integrated conditional moment test of Bierens (1982, *Journal of Econometrics* 20, 105–134), has suboptimal asymptotic minimax properties.

1. INTRODUCTION

Specification analysis is a central topic in econometrics. Recent work has focused on specification tests that are consistent against a large spectrum of nonparametric alternatives. Bierens (1982) inaugurates this line of research by proposing integrated conditional moment (ICM) tests for checking the specification of a parametric regression model. His method, which relies on the empirical process of the residuals from the parametric model, has been further developed by Andrews (1997), Bierens (1990), Bierens and Ploberger (1997), Delgado (1993), Stinchcombe and White (1998), and Stute (1997) among others. A competing approach compares parametric and smooth nonparametric regression estimates (see, among others, Fan and Li, 1996; Härdle and Mammen, 1993; Hong and White, 1995; Li and Wang, 1998; Zheng, 1996). Thus there now exists a large range of consistent specification tests for regression models. See Hart (1997) for a review.

The first author was with CREST when completing the first version of this paper. Financial support from CREST and INRA is gratefully acknowledged. We thank the editor and four referees for comments that were helpful in improving our paper. Address correspondence to: Pascal Lavergne, INRA-ESR, B.P. 27, 31326 Castanet-Tolosan Cedex, France; e-mail: lavergne@toulouse.inra.fr.

© 2002 Cambridge University Press 0266-4666/02 $9.50

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

A theme of this 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 (see, e.g., Hart, 1997). A familiar approach consists in considering a sequence of alternatives of the form

$$E(Y|X) = \mu(X, \theta_0) + r_n\, d(X), \tag{1.1}$$

where $\mu(X, \theta_0)$ is a member of the parametric model, $d(\cdot)$ is a specified function, and r_n goes to zero as the sample size n tends to infinity. It is generally found that smooth tests have trivial power against alternatives of the form (1.1) with $r_n \propto n^{-1/2}$, whereas nonsmooth tests such as ICM tests can detect such alternatives, thus suggesting that nonsmooth tests are more powerful. However, a reverse phenomenon can occur when considering different sequences of alternatives. Specifically, some alternatives that are more distant than $n^{-1/2}$ from the null hypothesis are detected by smooth tests but not by their competitors (see, e.g., Fan and Li, 2000). This shows that considering alternatives (1.1) is overly restrictive and can be misleading, as also argued by Horowitz and Spokoiny (2001).

In this paper, we adopt a nonparametric minimax approach, as detailed by Ingster (1993). Such an approach evaluates the power of a test uniformly over a set of alternatives $H_1(\rho_n)$ that lie at a distance ρ_n from the parametric model and that belong to a class of smooth functions with smoothness index s. The optimal minimax rate $\tilde{\rho}_n = \tilde{\rho}_n(s)$ is the fastest rate at which ρ_n can go to zero, whereas a test can uniformly detect any alternative in $H_1(\rho_n)$. Such a test is called rate optimal. Assuming that s is known, Ingster (1993) determines optimal minimax rates for goodness-of-fit testing of a uniform density and testing for white noise in the continuous-time Gaussian model. Considering s as an unknown nuisance parameter, the so-called adaptive framework, Spokoiny (1996) finds the optimal adaptive minimax rate $\tilde{\rho}_n^a$ in the latter testing problem. Assuming this rate applies in regression settings, Horowitz and Spokoiny (2001) propose a specification test that is asymptotically uniformly consistent against alternatives approaching the parametric model at rate $\tilde{\rho}_n^a$.

The main contribution of the present paper is to determine the optimal minimax rates for specification testing of a parametric nonlinear regression model with a multivariate random design and heteroskedasticity of unknown form. Following Ingster (1993), we assume that s, the regularity of the regression function, is known. Our results show that the optimal minimax rate $\tilde{\rho}_n$ for specification testing in regression models can differ from the optimal rate found in testing situations considered by Ingster (1993). We also provide a nonparametric smooth test that has power uniformly against alternatives approaching the null hypothesis at the optimal rate. This in turn yields the rate at which the smoothing parameter should go to zero to ensure rate-optimality of the test. Such a result constitutes a first step toward a better understanding of the

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

smoothing parameter's effect and the construction of practical procedures for its determination.

The paper is organized as follows. In Section 2, we describe our framework and assumptions. In Section 3, we establish optimal minimax rates for specification testing in regression models and provide a testing procedure that is *rate optimal* for alternatives that are regular enough. We also discuss the case of irregular alternatives. We finally illustrate the poor minimax properties of a class of ICM-type tests. Section 4 gives some concluding remarks in relation to the adaptive framework of Horowitz and Spokoiny (2001). Proofs of the main results appear in Section 5. Three appendixes gather some auxiliary results.

2. FRAMEWORK AND ASSUMPTIONS

Let (X,Y) be a random variable in $\mathbb{R}^p \times \mathbb{R}$ and assume that we have at hand observations on (X,Y) such that the following assumption applies.

Assumption I. $\{(X_i,Y_i), i = 1,\ldots,n\}$ is an independent and identically distributed (i.i.d.) sample on (X,Y) from $\mathbb{R}^p \times \mathbb{R}$. For $m(\cdot) \equiv \mathbb{E}(Y|X = \cdot)$, $\mathbb{E}m^4(X) \leq m_4 < \infty$ for some $m_4 > 0$. For $\varepsilon = Y - \mathbb{E}(Y|X)$, $\mathbb{E}\varepsilon^2 > 0$ and $\mathbb{E}\varepsilon^4 < \infty$.

Assumption I allows for heteroskedasticity of unknown form but restricts to regression functions with bounded fourth moments. In what follows, we acknowledge the dependence of the distribution of Y given X upon the regression function by denoting probabilities and expectation as $\mathbb{P}_m$ and $\mathbb{E}_m$, respectively.

We consider a parametric family $\mathcal{M}$ of regression functions $\mathcal{M} = \{\mu(\cdot,\theta); \theta \in \Theta\}$, $\Theta \subset \mathbb{R}^d$. The null hypothesis of interest is

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

To define the alternative hypothesis, the nonparametric minimax approach requires us to focus on some classes of smooth functions, as explained by Ingster (1993). For $s \in [0,1)$, let $C_p(L,s)$ be the Lipschitz class of maps $m(\cdot)$ from $\mathbb{R}^p$ to $\mathbb{R}$ such that

$$| m(x) - m(y) | \leq L\|x - y\|^s, \qquad L > 0, \qquad \forall x,y \in \mathbb{R}^p,$$

where $\|\cdot\|$ is a norm on $\mathbb{R}^p$. 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 $m(\cdot)$ almost everywhere differentiable up to order $[s]$, all of whose partial derivatives of order $[s]$ belong to $C_p(L, s - [s])$. We consider the alternative hypothesis

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

Here $H_1(\rho)$ is the set of regression functions in $C_p(L,s)$ at a distance ρ from the parametric model to be tested, with $\mathbb{E}m^4(X) \leq m_4 < \infty$ under Assumption I. For the following analysis, the latter restriction should hold uniformly

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

over the set of considered regression functions $m(\cdot)$. This assumption plays a role similar to the compactness of the parameter set in parametric estimation.

In the definition of the alternative hypothesis, the distance between the true regression function $m(\cdot)$ and the parametric model under consideration is closely related to the notion of "pseudo-true value" for the parameter θ (see White, 1981; 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.

(i) For each $\theta \in \Theta$, $\mu(\cdot,\theta) \in C_p(L_{\mathcal{M}}, s)$, $L_{\mathcal{M}} \leq L$, and $\mathbb{E}\mu^4(X,\theta) \leq \mu_4 < \infty$. There is an inner point θ_0 of Θ such that $\mathbb{E}\mu^4(X,\theta_0) < m_4$, for m_4 defined in Assumption I.

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

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

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

Assumption M1 (i) yields that the model $\mathcal{M}$ of interest is a subset of $C_p(L,s)$, a condition under which M1 (ii) implies that the parameter θ is identified under H_0. This assumption allows us to define the deviation of the regression function from the null hypothesis as

$$\delta(\cdot) \equiv \delta_m(\cdot) = m(\cdot) - \mu(\cdot,\theta_m^*). \tag{2.1}$$

Assumption M2.

(i) For each x, $\mu(x,\cdot)$ 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 and $\theta \in \Theta$.

(ii) The matrix $\mathbb{E}[(\partial\mu(X,\theta)/\partial\theta)(\partial\mu(X,\theta)/\partial\theta^\top)]$ is nonsingular for all $\theta \in \Theta$.

(iii) The set of gradient functions $\{(\partial\mu(\cdot,\theta)/\partial\theta);\ \theta \in \Theta\}$ is compact in C_0, the set of continuous functions from $\mathbb{R}^p$ to $\mathbb{R}^d$ equipped with the uniform norm.

Assumption M2 is similar to the assumption used by White (1981) to establish the $\sqrt{n}$-consistency of the nonlinear least-squares estimator of θ_m^*.[1]

Assumption M3. $\sqrt{n}(\hat{\theta}_n - \theta_m^*) = O_{\mathbb{P}_m}(1)$ uniformly with respect to $m(\cdot) \in C_p(L,s)$ with $\mathbb{E}m^4(X) \leq m_4 < \infty$, i.e.,

$$\forall\eta > 0, \exists\nu > 0 : \limsup_{n\to+\infty} \sup_{m(\cdot)\in C_p(L,s),\, \mathbb{E}m^4(X)\leq m_4} \mathbb{P}_m(\sqrt{n}\|\hat{\theta}_n - \theta_m^*\| > \nu) \leq \eta.$$

Assumption M3 deals with the existence of an $\sqrt{n}$-consistent estimator $\hat{\theta}_n$ of θ_m^*, uniformly with respect to $m(\cdot) \in C_p(L,s)$. Such a result is not usually shown in the literature. However, uniformity is essential for developing our minimax approach. Birgé and Massart (1993) show that Assumption M3 usually holds for approximate nonlinear least-squares estimators.[2] Consider, e.g., the simple univariate regression model where $\mu(X,\theta) = \theta X$, θ in $[\underline{\theta},\bar{\theta}]$. The

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

pseudo-true value is then defined as $\theta_m^* = I\!E[Xm(X)]/I\!E(X^2)$. Assumptions M1 and M2 hold provided X has bounded support and the ordinary least squares (OLS) estimator is such that

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

Hence, Assumption M3 holds for $\hat{\theta}_n$ under Assumption I when $I\!EX^4 < \infty$, as the empirical mean of the numerator is centered, with a variance of order $O(1/n)$ uniformly in $m(\cdot)$.

Consider a test $t_n \in \{0,1\}$ based on a sample of size n, where $t_n = 1$ corresponds to rejection of H_0. The behavior of the test under the null hypothesis is usually characterized by its level

$$\alpha(t_n) = \sup_{m(\cdot)\in H_0} I\!P_m(t_n = 1).$$

In our analysis, we focus on tests t_n with $\alpha(t_n) \leq \alpha + o(1)$ for some $\alpha > 0$. In Section 3.2, we consider a test $\tilde{t}_n$ with asymptotic type-I error α uniformly over H_0, i.e., such that $\sup_{m(\cdot)\in H_0} |I\!P_m(\tilde{t}_n = 1) - \alpha| \to 0$. In this case, $\alpha(\tilde{t}_n) \leq \alpha + o(1)$ also holds. In the minimax approach, the behavior of a test is evaluated uniformly against the alternative $H_1(\rho)$, i.e., through the minimax type-II error

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

The minimax power against $H_1(\rho)$ is then defined as $1 - \beta(t_n, \rho)$. A test with $\beta(t_n, \rho) = o(1)$ is said to be uniformly consistent against $H_1(\rho_n)$.

The definition of the optimal minimax rate for the testing problem relies on two conditions. First, the optimal minimax rate $\tilde{\rho}_n$ is such that no test has more than trivial minimax power against $H_1(\rho_n)$, for any ρ_n that goes faster to zero than $\tilde{\rho}_n$. Second, there exists a test $\tilde{t}_n$ that has a predetermined uniform power against alternatives approaching the null hypothesis at rate $\tilde{\rho}_n$. More formally, we have the following definition.

DEFINITION 1. *The optimal minimax (testing) rate $\tilde{\rho}_n$ is such that*

(i) For any test t_n with $\alpha(t_n) \leq \alpha + o(1)$, $\alpha > 0$,

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

(ii) There exists a test $\tilde{t}_n$ with $\alpha(\tilde{t}_n) \leq \alpha + o(1)$, $\alpha > 0$, such that for any prescribed bound β in $(0, 1-\alpha)$ for the minimax type-II error, there exists a constant $\kappa > 0$ ensuring

$$\beta(\tilde{t}_n, \kappa\tilde{\rho}_n) = \sup_{m(\cdot)\in H_1(\kappa\tilde{\rho}_n)} I\!P_m(\tilde{t}_n = 0) \leq \beta + o(1).$$

Such a test $\tilde{t}_n$ is called rate optimal.

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

As noted by Stone (1982), the minimax estimation rate of a nonparametric regression depends not only on its smoothness but also upon the behavior of the density $f(\cdot)$ of X at the boundary of its support. A similar phenomenon arises in our testing problem. For instance, if the density of the regressors has unbounded support, it is 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 whose density is bounded from above and below and has bounded support.[3]

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

3. OPTIMAL MINIMAX RATES FOR SPECIFICATION TESTING

As is usually done in this kind of analysis, we proceed in two stages corresponding to the two conditions of Definition 1. First, we find a testing rate $\tilde{\rho}_n$ below which alternatives in $H_1(\tilde{\rho}_n)$ cannot be uniformly detected. Second, we exhibit a nonparametric smooth test that has power uniformly against alternatives of order $\tilde{\rho}_n$.

3.1. Lower Bounds for Optimal Minimax Rates

The next result provides a lower bound $\tilde{\rho}_n$ for the optimal minimax rate corresponding to the smoothness index s. This result formally justifies that considering local alternatives (1.1) with $r_n \propto n^{-1/2}$ is not appropriate in the nonparametric minimax approach, because such alternatives are not in $H_1(\tilde{\rho}_n)$, as $n^{-1/2} = o(\tilde{\rho}_n)$ for any smoothness index s.

THEOREM 1. *Let $\tilde{\rho}_n = n^{-(2s/(p+4s))}$ if $s \geq p/4$ and $\tilde{\rho}_n = n^{-1/4}$ if $s < p/4$. Under Assumptions D, I, M1, and M2, if each ε_i is $\mathcal{N}(0,1)$ conditionally upon X_i, for any test t_n with $\alpha(t_n) \leq \alpha + o(1)$,*

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

To prove Theorem 1, it is enough to establish that $\alpha(t_n) + \beta(t_n, \rho_n) \geq 1 + o(1)$ for any test t_n with asymptotic level α. This is obtained by bounding the latter quantity from below via a proper choice of Bayesian a priori measures over subsets of H_0 and $H_1(\rho_n)$. Then, bounding the errors of the Bayesian likelihood-ratio test yields the result. Theorem 1's proof also shows that a uniformly consistent test of H_0 against $H_1(\kappa\tilde{\rho}_n)$, $\kappa > 0$, does not exist, though, as will be shown in Section 3.2, there exists a test that has a predetermined minimax power against $H_1(\kappa\tilde{\rho}_n)$.

The assumption of standard normal errors, which is used to derive Theorem 1, can be relaxed as soon as regular distributions are considered. A com-

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

mon condition is to assume that the translation model associated with the errors ε_i's is locally asymptotically normal (LAN), i.e., the density $f_i(\cdot) = f(\cdot|X_i)$ of the variables ε_i given X_i fulfills

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

where $I > 0$ is a constant and S_n converges in distribution to $\mathcal{N}(0, I)$ (see Ibragimov and Has'minskii, 1981). This condition also allows for the presence of heteroskedasticity. For instance, if $\varepsilon_i = \sigma(X_i)\eta_i$, where the η_i's are independent with density $f_\eta(\cdot)$ and are also independent of the X_i's, the LAN condition holds under standard regularity conditions on $f_\eta(\cdot)$ given $0 < \underline{\sigma} \leq \sigma(\cdot) \leq \bar{\sigma} < \infty$. Under the LAN condition, Theorem 1 carries over at the cost of some technicalities.[4] However, the assumption of Gaussian errors is only instrumental in our analysis. The next section shows how optimal minimax rates are determined for a general unknown error distribution.

3.2. Optimal Minimax Rates and a Rate-Optimal Test for Regular Alternatives

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

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

be a bin of $[0,1]^p$, where the multivariate index $k = (k_1, \ldots, k_p)^\top \in \mathcal{K} \subset I\!N^p$ satisfies $0 \leq k_j \leq K - 1$ 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, and the indicator functions $1\!I(x \in I_k)$ are therefore orthogonal. Then, following Neyman (1937) (see also Hart, 1997), a smooth test can be proposed by regressing the $\hat{U}_i$'s on the normalized variables $1\!I(X_i \in I_k)/\sqrt{N_k}$ for $k \in \mathcal{K}$, where $N_k = \sum_{i=1}^{n} 1\!I(X_i \in I_k)$ is the number of observations of the exogenous variables in bin I_k. If the $\hat{\tau}_k$'s are the corresponding estimated coefficients, a test can be based on

$$\sum_{k\in\mathcal{K}} \hat{\tau}_k^2 = \sum_{k\in\mathcal{K}}\left(\frac{1}{\sqrt{N_k}}\sum_{X_i\in I_k}\hat{U}_i\right)^2.$$

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

Such a test statistic can also be viewed as an estimator of $\mathbb{E}(\mu(X,\theta^*)-m(X))^2$ based on the regressogram method. However, this statistic is systematically biased under the null hypothesis, because it includes the squared estimated residuals $\hat{U}_i^2, i=1,\ldots,n$. To remove this systematic bias, we consider

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

and the simple estimator of the variance of $\hat{T}_n$

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

The test is defined as $\tilde{t}_n = \mathbb{1}(v_n^{-1}\hat{T}_n > z_\alpha)$, 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 is thus a simple regressogram version of the kernel-based test of Zheng (1996). This allows us to treat the design density and the conditional heteroskedasticity function as nuisance parameters and avoids unnecessary smoothness assumptions on these functions.

THEOREM 2. *Under Assumptions D, I, and M1–M3,*

(i) If $K\to\infty$ and $n/(K^p\log K^p)\to\infty$, the test $\tilde{t}_n$ is of asymptotic level α uniformly over H_0, i.e.,

$$\sup_{H_0} |\mathbb{P}_m(\tilde{t}_n=1)-\alpha| = \sup_{H_0} |\mathbb{P}_m(v_n^{-1}\hat{T}_n > z_\alpha)-\alpha| \to 0.$$

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

$$\beta(\tilde{t}_n,\kappa\tilde{\rho}_n) = \sup_{H_1(\kappa\tilde{\rho}_n)} \mathbb{P}_m(v_n^{-1}\hat{T}_n \le z_\alpha) \le \beta + o(1).$$

Theorem 2(i) indicates that $\tilde{t}_n$ has asymptotically a type-I error equal to α uniformly over H_0. Theorem 2(ii) shows that for $s>p/4$, $\tilde{t}_n$ has asymptotically minimax power $1-\beta$ against $H_1(\kappa\tilde{\rho}_n)$.[5] Note that β can be chosen as close to zero as desired by taking κ large enough.

Theorems 1 and 2 together establish the minimax optimality of the rate $\tilde{\rho}_n$ and the rate-optimality of the test $\tilde{t}_n$. This easily follows by checking the conditions of Definition 1. Condition (i) is fulfilled, as the lower bound of Theorem 1 cannot be improved if the conditional distribution of the ε_i's is unknown and lies in a set of densities including the normal.[6] Condition (ii) is fulfilled because of Theorem 2, which leaves this distribution unspecified.

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

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

As can be expected, the rate $\tilde{\rho}_n$ becomes slower when the dimension of X increases or when the smoothness index s decreases. When $p = 1$, the rate $\tilde{\rho}_n$ is similar to the one obtained for a test of $m(\cdot) = 0$ in the continuous-time Gaussian (CTG) model,

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

where $W(\cdot)$ is a standard Brownian motion (see Ingster, 1993). This model may be viewed as an ideal model: many optimality results valid in this context can be extended to the univariate regression model with homoskedastic Gaussian errors when the smoothness index of $m(\cdot)$ is such that $s > \frac{1}{2}$, thanks to an equivalence statement due to Brown and Low (1996). However, this equivalence does not hold for $s < \frac{1}{4}$ and $s = \frac{1}{2}$, as shown by Efromovich and Samarov (1996) and Brown and Zhang (1998). Moreover, it is not known if such an equivalence extends to a regression model with a multivariate random design, unknown variance, non-normality, or heteroskedasticity of the regression errors. For instance, the case of small smoothness indices, i.e., $s \leq p/4$, which is not treated in Corollary 3, seems to be specific to the regression model and is discussed subsequently. Horowitz and Spokoiny (2001), who do not assume that s is known, propose a test that is uniformly consistent against alternatives approaching the null hypothesis at rate $\tilde{\rho}_n(\ln\ln n)^{s/(4s+p)}$ when $s \geq \max(2, p/4)$, which for $p = 1$ is the optimal adaptive minimax rate for testing $m(\cdot) = 0$ in the CTG model according to Spokoiny (1996). Thus, the adaptive approach leads to an unavoidable but small loss in the optimal minimax rate.

Our results give theoretical grounds for the choice of the smoothing parameter in a specification testing framework. To understand how the binwidth is chosen to get a rate-optimal test, note that our results imply that

$$\frac{\hat{T}_n}{v_n^2} \geq O_{\mathbb{P}_m}(1)nh^{p/2}(\mathbb{E}^{1/2}\delta^2(X) - h^s)^2 \geq O_{\mathbb{P}_m}(1)nh^{p/2}(\kappa\rho_n - h^s)^2,$$

for any $m(\cdot) \in H_1(\kappa\rho_n)$ with $\kappa\rho_n \geq h_n^s$ and $nh^p/\ln h^p \to \infty$.[7] To bound the asymptotic minimax type-II error, one must force the lower bound of $\hat{T}_n$ to stay away from zero. Hence the smallest possible ρ_n has the same rate as h^s and the corresponding lower bound for $\hat{T}_n$ is an $O(nh^{(p+4s)/2})$. Therefore, for regular alternatives, i.e., for $s > p/4$, the optimal binwidth $\tilde{h}$ is such that $n\tilde{h}^{(p+4s)/2}$ has a nonvanishing finite limit; i.e.,

$$\tilde{h} \propto n^{-(2/(p+4s))}. \tag{3.1}$$

For the same p and s, the optimal binwidth rate for testing the specification of a nonlinear parametric regression model is faster than the optimal binwidth rate

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

for minimax nonparametric estimation of the regression function in the L_2-norm, which is $n^{-1/(p+2s)}$. Basically, choosing an optimal testing binwidth leads us to balance a variance term against 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 optimal smoothing parameters for semiparametric estimation and determining them for testing are typically different issues.[8]

3.3. The Case of Irregular Alternatives

The optimal minimax testing rate generally depends on the relative standing of the smoothness index 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 X. This contrasts with the result found by Ingster (1993) in the CTG model. The rate $n^{-1/4}$ corresponds to a baseline minimax testing rate when the variance function $\sigma^2(X) = \mathrm{Var}_m[\varepsilon|X]$ is known. Define

$$\hat{T}_n' = (1/n)\sum_{i=1}^{n}(\hat{U}_i^2 - \sigma^2(X_i)) \tag{3.2}$$

and observe that $\hat{T}_n'$ estimates $\mathbb{E}_m[Y - \mu(X,\theta^*)]^2 - \mathbb{E}\sigma^2(X) = \mathbb{E}[\mu(X,\theta^*) - m(X)]^2$ with rate of convergence $\sqrt{n}$. Therefore, it is easy to show that a test based on $\hat{T}_n'$ has asymptotically nontrivial minimax power against $H_1(\kappa n^{-1/4})$ for any $\kappa > 0$.[9]

The case where the variance function $\sigma^2(X)$ is unknown is more difficult to deal with. Even with homoskedastic errors, estimating σ^2 is problematic when $m(\cdot)$ is not smooth enough, as it is difficult to separate the signal $m(\cdot)$ from the noise ε. It is likely that the minimax testing rate depends upon s when $s \leq p/4$ and $\sigma^2(\cdot)$ is unknown. Note that choosing $h \propto n^{-1/p}$ in our testing procedure yields a test that uniformly detects with some predetermined power alternatives in $H_1(\kappa n^{-s/p})$, for κ large enough.[10] This implies that the optimal minimax rate is faster than or equal to $n^{-s/p}$ for irregular alternatives.

3.4. Minimax Properties of ICM-Type Tests

We now study the minimax behavior of some nonsmooth tests. Bierens and Ploberger (1997) have shown that ICM tests are asymptotically admissible against specific alternatives of the type (1.1). The nonparametric minimax approach provides an alternative way of evaluating power properties of such specification tests. Theorem 4, which follows, shows that ICM-type tests have asymptotically trivial minimax power against alternatives in $H_1(n^{-a})$ for any $a > 0$.

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

The ICM test statistic proposed by Bierens (1982), and further developed by Bierens and Ploberger (1997), is

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

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

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

Let $\hat{t}_{n,q}$ be the test $\hat{t}_{n,q} = \mathbb{1}(I_{n,q} > u_{\alpha,q})$, with $\alpha(\hat{t}_{n,q}) \leq \alpha + o(1)$. In what follows, $C_p(\infty)$ is the set of infinitely continuously differentiable functions from $\mathbb{R}^p$ to $\mathbb{R}$.

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, and M1–M3, if each ε_i is $\mathcal{N}(0,1)$ conditionally upon X_i and $f(\cdot) \in C_p(\infty)$, then $\forall 1 \leq q < \infty$,*

$$\beta(\hat{t}_{n,q}, \rho_n) = \sup_{m(\cdot) \in H_1(\rho_n)} \mathbb{P}_m(I_{n,q} \leq u_{\alpha,q})$$

$$\geq 1 - \alpha + o(1), \quad \textit{whenever } \rho_n = O(n^{-a}), \quad \forall a > 0.$$

Our assumptions on $w(\cdot,\cdot)$ are justified by usual choices, such as $\exp(X'\xi)$, see Bierens (1990), or $(1 + \exp(-X'\xi))^{-1}$, see White (1989). Theorem 4 relies upon a Bayesian approach similar to the one used in the proof of Theorem 1. We conjecture that similar results can be derived for other nonsmooth tests, because such 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; Fan and Li, 2000).

4. CONCLUSION

Our results illustrate the particular features of specification testing of nonlinear regression models under a multivariate random design. For regular alternatives, the optimal minimax rates, and also the optimal smoothing parameter, converge to zero faster than their analogues for estimation of the nonparametric regression. In particular, the optimal smoothing parameter for specification testing is derived from a different bias-variance trade-off than the one considered in regression estimation. For irregular alternatives, the optimal minimax rates can differ from those found in other testing situations, such as those considered by Ingster (1993). We also show that a class of ICM-type tests, in spite

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

of being admissible against alternatives (1.1) with $r_n \propto n^{-1/2}$, have trivial asymptotic minimax power against alternatives at distance n^{-a} from the null hypothesis for any $a > 0$. All these results are likely to extend to testing general conditional moment restrictions, as considered by Delgado, Dominguez, and Lavergne (2000).

An important direction for future research is the study of data-driven procedures for choosing the smoothing parameter. Some suggestions can be found in the monograph of Hart (1997) and the references therein. Our results explain why cross-validation and penalization procedures used in nonparametric regression estimation would lead to suboptimal tests. The search for adapted procedures is an important topic of recent work (see Baraud, Huet, and Laurent, forthcoming; Guerre and Lavergne, 2001; Guerre and Lieberman, 2000; Horowitz and Spokoiny, 2001; Spokoiny, 1996, 2001). A further step could be to compare the practical performances of the rate-optimal tests derived from our approach and the adaptive approach.

5. PROOFS

5.1. Proof of Theorem 1

Some small alternatives. For $l > 0$, let φ be any infinitely differentiable function from $\mathbb{R}^p$ to $\mathbb{R}$ with support $[0, l]^p$ such that

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

Assume that l is large enough so that φ is in $C_p((L - L_{\mathcal{M}})/2, s)$. For example, for a suitable constant C, one can choose

$$\varphi(x) = C \prod_{j=1}^{p} \left(\exp\left(\frac{-1}{x_j(l/2 - x_j)} \right) \mathbb{1}(x \in [0, l/2]^p) - \exp\left(\frac{-1}{(x_j - l/2)(l - x_j)} \right) \mathbb{1}(x \in [l/2, l]^p) \right).$$

Let $h_n = (\lambda \rho_n)^{1/s}$, $\lambda > 0$ and define

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

for $k \in \mathcal{K}_n(l)$, i.e., $k \in \mathbb{N}^p$ with $0 \le k_j \le 1/(h_n l) - 1$. 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

$$\varphi_k(x) = \frac{1}{h_n^{p/2}} \varphi\left(\frac{x - l k h_n}{h_n} \right), \qquad k \in \mathcal{K}_n(l). \tag{5.1}$$

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

The functions $\varphi_k(\cdot)$'s are orthogonal with disjoint supports I_{kl}. Let θ_0 be the inner point of Θ defined in Assumption M1, $\{B_k, k \in \mathcal{K}\}$ be any sequence with $|B_k| = 1\ \forall k$, and

$$m_n(\cdot) = \mu(\cdot, \theta_0) + \delta_n(\cdot), \qquad \delta_n(\cdot) = \lambda\rho_n h_n^{p/2} \sum_{k \in \mathcal{K}_n(l)} B_k \varphi_k(\cdot). \tag{5.2}$$

LEMMA 1. *Assume* $\rho_n \to 0$. *Under Assumptions D, M1, and M2,* $\mathbb{E}m_n^4(X) \le m_4$ *and* $m_n(\cdot)$ *is in* $H1(\rho_n)$ *for* λ *and n large enough.*

Proof.

(i) $\mathbb{E}m_n^4(X) \le m_4$ for n large enough, because $\mathbb{E}\mu^4(X, \theta_0) < m_4$ under Assumption M1 and, because the $\varphi_k(\cdot)$ have disjoint supports, $\mathbb{E}^{1/4}\delta_n^4(X) \le \sup_{x \in [0,1]^p} |\delta_n(x)| = O(\lambda\rho_n) \to 0$.

(ii) $m_n(\cdot) \in C_p(L, s)$: under Assumption M1 it is enough to show that $\delta_n(\cdot)$ is in $C_p(L - L_{\mathcal{M}}, s)$. For any $\beta \in \mathbb{N}^p$ with $\sum_{j=1}^p \beta_j = [s]$, we have

$$\frac{\partial^{[s]}\delta_n(x)}{\partial x_1^{\beta_1} \ldots \partial x_p^{\beta_p}} = \frac{\lambda\rho_n}{h_n^{[s]}} \sum_{k \in \mathcal{K}_n(l)} B_k \frac{\partial^{[s]}\varphi}{\partial x_1^{\beta_1} \ldots \partial x_p^{\beta_p}} \left(\frac{x - lkh_n}{h_n} \right).$$

Therefore, for any x and y that do not necessarily belong to the same bin I_{kl}, we get, using the definition of $\varphi(\cdot) \in C_p((L - L_{\mathcal{M}})/2, s)$, $h_n = (\lambda\rho_n)^{1/s}$, and $|B_k| = 1$,

$$\left| \frac{\partial^{[s]}\delta_n(x)}{\partial x_1^{\beta_1} \ldots \partial x_p^{\beta_p}} - \frac{\partial^{[s]}\delta_n(y)}{\partial x_1^{\beta_1} \ldots \partial x_p^{\beta_p}} \right| \le 2 \frac{L - L_{\mathcal{M}}}{2} \frac{\lambda\rho_n}{h_n^{[s]}} \| \frac{x - y}{h_n} \|^{s - [s]}$$

$$\le (L - L_{\mathcal{M}}) \| x - y \|^{s - [s]},$$

and $\delta_n(\cdot) \in C_p(L - L_{\mathcal{M}}, s)$ for any n and $\lambda > 0$.

(iii) $m_n(\cdot)$ is distant from the null model: let $\theta_n \equiv \theta_{m_n}^*$. Then

$$\mathbb{E}^{1/2}[m_n(X) - \mu(X, \theta_n)]^2 \ge \mathbb{E}^{1/2}\delta_n^2(X) - \mathbb{E}^{1/2}[\mu(X, \theta_0) - \mu(X, \theta_n)]^2$$

$$\ge \left(f \int \delta_n^2(x)\, dx \right)^{1/2} - O(\| \theta_n - \theta_0 \|), \tag{5.3}$$

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

$$\int \delta_n^2(x)\, dx = (\lambda\rho_n)^2 h_n^p K_n^p(l) \int \varphi^2(x)\, dx = (\lambda\rho_n)^2 l^{-p} \int \varphi^2(x)\, dx. \tag{5.4}$$

As θ_n converges to θ_0 by Assumption M1 because $\mathbb{E}(m_n(X) - \mu(X, \theta_0))^2 = \mathbb{E}\delta_n^2(X) \to 0$ as easily seen from step (i), θ_n is then an inner point of Θ. Therefore, from Assumption M2 and the Lebesgue dominated convergence theorem, Assumption M1 yields that

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

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

This leads to

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

A simple Taylor expansion, which holds by Assumption 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

$$\|\theta_n-\theta\|=O\left(\left\|\mathbb{E}\delta_n(X)\,\frac{\partial\mu(X,\theta_n)}{\partial\theta}\right\|\right). \tag{5.5}$$

The functions $\{(\partial\mu(\cdot,\theta)/\partial\theta)\,f(\cdot);\ \theta\in\Theta\}$ are equicontinuous from Assumptions D and M2 and the Arzela–Ascoli theorem (see Rudin, 1991). As $\varphi(\cdot)$ has integral zero,

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

Combining this equality with (5.3)–(5.5) yields, for λ and n large enough,

$$\mathbb{E}^{1/2}[m_n(X)-\mu(X,\theta_n)]^2\geq\lambda\rho_n l^{-p}((fl^p\int\varphi^2(x)\,dx)^{1/2}-o(1))\geq\rho_n. \qquad \blacksquare$$

Main proof. We shall establish that for any test t_n

$$\sup_{m(\cdot)\in H_0}\mathbb{P}_m(t_n=1)+\sup_{m(\cdot)\in H_1(\rho_n)}\mathbb{P}_m(t_n=0)\geq 1+o(1). \tag{5.6}$$

Step 1. Choice of a Bayesian a priori measure. As usual in the Bayesian setup, we consider now the regression function $m(\cdot)$ as a random variable and introduce some Bayesian a priori probabilities over H_0 and $H_1(\rho_n)$. Let θ_0 be the inner point of Θ defined in Assumption M1 and denote Π_0 the associate Dirac mass, i.e., $\Pi_0(m(\cdot)=\mu(\cdot,\theta_0))=1$. Consider i.i.d. Rademacher B_k's indepen-

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

dent of the observations, i.e., $\mathbb{P}(B_k = 1) = \mathbb{P}(B_k = -1) = \frac{1}{2}$. Let Π_{1n} be the a priori distribution defined on $H_1(\rho_n)$ by (5.2), i.e.,

$$\Pi_{1n}\left(m(\cdot) = \mu(\cdot, \theta_0) + \lambda\rho_n h_n^{p/2} \sum_{k\in\mathcal{K}_n(l)} b_k \varphi_k(\cdot)\right) = \prod_{k\in\mathcal{K}_n(l)} \mathbb{P}(B_k = b_k),$$

$$b_k \in \{1, -1\}.$$

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

$$\sup_{m(\cdot)\in H_0} \mathbb{P}_m(t_n = 1) + \sup_{m(\cdot)\in H_1(\rho_n)} \mathbb{P}_m(t_n = 0)$$

$$\geq \int \mathbb{P}_m(t_n = 1)\, d\Pi_0(m) + \int \mathbb{P}_m(t_n = 0)\, d\Pi_{1n}(m). \tag{5.7}$$

The right-hand side (r.h.s.) of (5.7) 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 $m(\cdot)$. Define the a priori densities associated with the two hypotheses as

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

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 Lehmann, 1986) is

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

where $\boldsymbol{E}_0$ is the expectation under $\boldsymbol{p}_0$. Then (5.7) implies that

$$\sup_{m(\cdot)\in H_0} \mathbb{P}_m(t_n = 1) + \sup_{m(\cdot)\in H_1(\rho_n)} \mathbb{P}_m(t_n = 0)$$

$$\geq \liminf_{n\to+\infty} \mathbb{E}\left\{1 - \frac{1}{2}\boldsymbol{E}_0[|Z_n - 1||\mathcal{X}]\right\} + o(1),$$

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

and (5.6) holds if we can show that the limit in the r.h.s. is 1. We first note that $1 - \frac{1}{2}\boldsymbol{E}_0[|Z_n - 1||\mathcal{X}]$ is positive as a conditional Bayes testing error. Then the Fatou lemma implies that it is enough to show that $\boldsymbol{E}_0[|Z_n - 1||\mathcal{X}] \xrightarrow{I\!P} 0$, which is implied by $\boldsymbol{E}_0[(Z_n - 1)^2|\mathcal{X}] \xrightarrow{I\!P} 0$. But $\boldsymbol{E}_0[(Z_n - 1)^2|\mathcal{X}] = \boldsymbol{E}_0(Z_n^2|\mathcal{X}) - 1$ as $\boldsymbol{E}_0(Z_n|\mathcal{X}) = 1$. Hence, inequality (5.6) holds if

$$\boldsymbol{E}_0(Z_n^2|\mathcal{X}) \xrightarrow{I\!P} 1. \tag{5.8}$$

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

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

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

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

The definition of the alternatives (5.2) gives

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

$$\sum_{i=1}^{n} \delta_n^2(X_i) = \lambda^2\rho_n^2 h_n^p \sum_{k\in\mathcal{K}(l)} \sum_{i=1}^{n} \varphi_k^2(X_i),$$

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

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

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

Therefore,

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

Conditionally on $\mathcal{X}$, the variables $\sum_i \varepsilon_{i0}\varphi_k(X_i)$, $k \in \mathcal{K}_n(l)$, 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

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

where $\cosh(x)$ is the hyperbolic cosine function. By a series expansion, $1 \le \cosh(x) \le \exp(x^2)$, and

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

Then (5.8) holds if

$$\sum_{k\in\mathcal{K}(l)}\left(\rho_n^2 h_n^p\sum_i \varphi_k^2(X_i)\right)^2 \xrightarrow{\mathbb{P}} 0. \tag{5.9}$$

Consider the expectation of this positive random variable. We have

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

Now the standard change of variables $x = lkh_n + uh_n$ and Assumption D yield

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

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

and

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

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

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

We then consider the two following cases.

(i) $s > p/4$: $\rho_n = o(\tilde{\rho}_n) = o(n^{-(2s/(p+4s))}) \Rightarrow n\rho_n^4 = o(n^{(p-4s)/(p+4s)}) = o(1)$ and $n^2\rho_n^{(p+4s)/s} = o(1)$.
(ii) $s \le p/4$: $\rho_n = o(\tilde{\rho}_n) = o(n^{-1/4}) \Rightarrow n\rho_n^4 = o(1)$ and $n^2\rho_n^{(p+4s)/s} = o(n^{(4s-p)/4s}) = o(1)$.

Equation (5.9) follows and then (5.8). Step 1 shows that (5.6) holds, and Theorem 1 is proved. ■

5.2. Proof of Theorem 2

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

$$\langle Z, Z'\rangle_k \equiv \frac{1\!\!1[N_k > 1]}{N_k} \sum_{\{X_i, X_j\}\in I_k, i\neq j} Z_i Z'_j \quad \forall k \in \mathcal{K} \quad \text{and}$$

$$\langle Z, Z'\rangle \equiv \frac{1}{\sqrt{2}K^{p/2}} \sum_{k\in\mathcal{K}} \langle Z, Z'\rangle_k.$$

Let $\mathrm{Proj}_{\mathcal{K}} Z \equiv \sum_k 1\!\!1(X \in I_k)\mathbb{E}^k Z$ be the projection of Z onto the space of linear indicators $1\!\!1(x \in I_k)$, $k \in \mathcal{K}$. Key properties of this mapping are

$$\mathbb{E}[\mathrm{Proj}_{\mathcal{K}} Z] = \sum_k \mathbb{P}(X \in I_k)\mathbb{E}^k Z = \mathbb{E}Z \quad \text{and} \quad \mathbb{E}[\mathrm{Proj}_{\mathcal{K}}^2 Z] \le \mathbb{E}Z^2.$$

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

Preliminary results.

PROPOSITION 5. *Let* $v^2(K) = (1/K^p)\sum_{k\in\mathcal{K}} 1\!\!1(N_k > 1)((N_k - 1)/N_k) \times (\mathbb{E}^k U^{*2})^2$. *Under Assumptions I, D, and M1–M3,* $v^2(K)$ *is bounded from above*

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

and in probability from below uniformly in $m(\cdot) \in C_p(L,s)$, *and* $v_n^2 - v^2(K) = o_{I\!P_m}(1)$ *uniformly in* $m(\cdot) \in C_p(L,s)$ *whenever* $n/(K^p \log K^p) \to \infty$.

Proof of Proposition 5. By Assumption D, $fh^p \le I\!P(X \in I_k) \le Fh^p$. Now, on the one hand,

$$\begin{aligned} v^2(K) &\le (1/K^p) \sum_{k \in \mathcal{K}} (I\!E^k U^{*2})^2 \le (1/f) \sum_{k \in \mathcal{K}} I\!P[X \in I_k](I\!E^k U^{*2})^2 \\ &= (1/f) I\!E[\mathrm{Proj}_{\mathcal{K}}^2 U^{*2}] \\ &\le (1/f) I\!E[U^{*4}] \le (8/f)[I\!E_m Y^4 + I\!E_m \mu^4(X, \theta_m^*)] < C. \end{aligned}$$

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

$$\begin{aligned} v^2(K) &\ge (1/2K^p) \sum_{k \in \mathcal{K}} (I\!E^k U^{*2})^2 \ge (1/2F) \sum_{k \in \mathcal{K}} I\!P[X \in I_k](I\!E^k U^{*2})^2 \\ &\ge (1/2F) I\!E_m[\mathrm{Proj}_{\mathcal{K}}^2 U^{*2}] \ge (1/2F) I\!E_m^2[U^{*2}] \ge (1/2F) I\!E_m^2[\varepsilon^2] > 0. \end{aligned}$$

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

$$|v_n^2 - v_n^{*2}| \le (1/K^p) \sum_{k \in \mathcal{K}} \frac{1\!I(N_k > 1)}{N_k} |\langle \hat{U}^2, \hat{U}^2 \rangle_k - \langle U^{*2}, U^{*2} \rangle_k|. \tag{5.10}$$

But $\langle \hat{U}^2, \hat{U}^2 \rangle_k - \langle U^{*2}, U^{*2} \rangle_k = -4\langle U^{*2}, U^* e(X) \rangle_k + 2\langle U^{*2}, e^2(X) \rangle_k + 4\langle U^* e(X), U^* e(X) \rangle_k - 4\langle U^* e(X), e^2(X) \rangle_k + \langle e^2(X), e^2(X) \rangle_k$. By Assumptions M1–M3, $|e(X_i)| = O_{I\!P_m}(1/\sqrt{n})$ uniformly in $m(\cdot)$ and i. Hence the dominant term in (5.10) is

$$\begin{aligned} &(4/K^p) \sum_{k \in \mathcal{K}} \frac{1\!I(N_k > 1)}{N_k} |\langle U^{*2}, U^* e(X) \rangle_k| \\ &\quad = O_{I\!P_m}(1/\sqrt{n})(1/K^p) \sum_{k \in \mathcal{K}} \frac{1\!I(N_k > 1)}{N_k} \langle U^{*2}, |U^*| \rangle_k. \end{aligned}$$

Now, by Assumptions I and M1,

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

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

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

$$\begin{aligned}
&\mathbb{E}_m[(v_n^{*2} - v^2(K))^2 | S_{\mathcal{K}}] = \mathrm{Var}_m[v_n^{*2} | S_{\mathcal{K}}] \\
&\quad = (1/K^{2p}) \sum_{k \in \mathcal{K}} 2 \frac{\mathbb{1}(N_k > 1) N_k (N_k - 1)}{N_k^4} \\
&\qquad \times [2(N_k - 2)(\mathbb{E}^k U^{*2})^2 \,\mathrm{Var}^k U^{*2} + (\mathbb{E}^k U^{*4})^2 - (\mathbb{E}^k U^{*2})^4] \\
&\quad \le (1/K^{2p}) \sum_{k \in \mathcal{K}} 2 \frac{\mathbb{1}(N_k > 1)}{N_k^2} [2N_k (\mathbb{E}^k U^{*2})^2 \mathbb{E}^k U^{*4} + (\mathbb{E}^k U^{*4})^2] \\
&\quad \le O_{\mathbb{P}}(nh^p)^{-2} \sum_{k \in \mathcal{K}} (\mathbb{P}(X \in I_k) \mathbb{E}^k U^{*4})^2 + O_{\mathbb{P}}(nh^p)^{-1} \\
&\qquad \times \sum_{k \in \mathcal{K}} \mathbb{P}(X \in I_k) (\mathbb{E}^k U^{*2})^2 \sum_{k' \in \mathcal{K}} \mathbb{P}(X \in I_{k'}) \mathbb{E}^{k'} U^{*4} \\
&\quad \le O_{\mathbb{P}}(nh^p)^{-2} \mathbb{E}^2 [\mathrm{Proj}_{\mathcal{K}} U^{*4}] + O_{\mathbb{P}}(nh^p)^{-1} \mathbb{E}[\mathrm{Proj}_{\mathcal{K}}^2 U^{*2}] \mathbb{E}[\mathrm{Proj}_{\mathcal{K}} U^{*4}] \\
&\quad \le O_{\mathbb{P}}(nh^p)^{-2} \mathbb{E}_m^2 U^{*4} + O_{\mathbb{P}}(nh^p)^{-1} \mathbb{E}_m^2 U^{*4} \to 0.
\end{aligned}$$

Hence $v_n^{*2} - v^{*2}(K) = O_{\mathbb{P}_m}(nh^p)^{-1/2}$ uniformly in $m(\cdot)$. ■

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

$$\hat{T}_n = T_n - 2(A + B) + R. \tag{5.11}$$

PROPOSITION 6. *Under Assumptions D, I, and 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

$$e(X_i) = (\hat{\theta}_n - \theta^*)' \mu_1(X_i) + \|\hat{\theta}_n - \theta^*\|^2 \mu_2(X_i), \tag{5.12}$$

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 $\hat{\theta}_n$. Therefore $B = (\hat{\theta}_n - \theta^*)' B_1 + \|\hat{\theta}_n - \theta^*\|^2 B_2$, 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

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

$$\mathbb{E}(B_1^2) = \frac{1}{2K^p} \sum_{k \in \mathcal{K}} \mathbb{E}\left[\frac{\mathbb{1}[N_k > 1]}{N_k^2} \sum_{\{X_i, X_j, X_{j'}\} \in I_k, i \neq j, i \neq j'} \varepsilon_i^2 \mu_1(X_j) \mu_1(X_{j'}) \right]$$

$$= \frac{O(1)}{2K^p} \sum_{k \in \mathcal{K}} \mathbb{E}\left[\frac{\mathbb{1}[N_k > 1](N_k - 1)^2}{N_k} \right] = O(nh^p),$$

using Assumption M2. Similarly,

$$\mathbb{E}|B_2| \leq \frac{O(1)}{\sqrt{2}K^{p/2}} \sum_{k \in \mathcal{K}} \frac{\mathbb{1}[N_k > 1]}{N_k} \sum_{\{X_i, X_j\} \in I_k, i \neq j} \mathbb{E}^k |\varepsilon_i|$$

$$= \frac{O(1)}{\sqrt{2}K^{p/2}} \sum_{k \in \mathcal{K}} \mathbb{E}[\mathbb{1}[N_k > 1](N_k - 1)] = O(nh^{p/2}).$$

As $\sqrt{n}(\hat{\theta}_n - \theta^*) = O_{\mathbb{P}_m}(1)$ uniformly in $m(\cdot)$, we obtain $B = O_{\mathbb{P}_m}(h^{p/2})$ uniformly in $m(\cdot)$. From (5.12), $A = (\hat{\theta}_n - \theta^*)' A_1 + \|\hat{\theta}_n - \theta^*\|^2 A_2$, where $A_1 = \langle \delta(X), \mu_1(X) \rangle$ and $A_2 = \langle \delta(X), \mu_2(X) \rangle$. Now,

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

$$\leq O(nh^{p/2})\mathbb{E}^{1/2}\delta^2(X).$$

Similarly, $\mathbb{E}|A_2| = O(nh^{p/2})\mathbb{E}^{1/2}\delta^2(X)$. Because $\sqrt{n}(\hat{\theta}_n - \theta^*) = O_{\mathbb{P}_m}(1)$ 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 $\mathbb{1}(x \in I_k)$, $k \in \mathcal{K}$, can be used to approximate accurately enough the magnitude of the L_2-norm of $m(\cdot)$.

PROPOSITION 7. *Under Assumption D,*

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

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

This result is new for multivariate random designs but follows from proper modifications of the arguments used in Ingster (1993, pp. 253 and following). A detailed proof is given in Appendix B.

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

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

PROPOSITION 8. *Under Assumptions D and I, if* $nh^p \to \infty$, *then, for any* $m(\cdot) \in H_m(\kappa\rho_n)$ *with* $\kappa\rho_n > h^s$ *and n large enough,*

$$\mathbb{E}_m T_n \geq C_2 nh^{p/2}(\mathbb{E}^{1/2}\delta^2(X) - h^s)^2 \qquad \text{for some } C_2 > 0,$$

$$\mathrm{Var}_m(T_n) \leq \mathbb{E}_m v^2(K) + C_3 nh^p \mathbb{E}_m \delta^2(X) + C_4 n\mathbb{E}_m^2 \delta^2(X) \qquad \text{for some } C_3, C_4 > 0.$$

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

$$\mathbb{E}_m T_n = \frac{1}{\sqrt{2}K^{p/2}} \sum_{k\in\mathcal{K}} \mathbb{E}_m \omega_k = \frac{1}{\sqrt{2}K^{p/2}} \sum_{k\in\mathcal{K}} \mathbb{E}[(N_k - 1)\mathbb{1}(N_k > 1)](\mathbb{E}^k \delta(X))^2$$

$$\geq \frac{nh^{p/2}}{2\sqrt{2}} \mathbb{E}[\mathrm{Proj}^2_{\mathcal{K}} \delta(X)] \geq \frac{C_1}{2\sqrt{2}} nh^{p/2}(\mathbb{E}^{1/2}\delta^2(X) - h^s)^2,$$

for n large enough, using Proposition 7 and $\mathbb{E}^{1/2}\delta^2(X) - h^s \geq 0$ as $m(\cdot) \in H_1(\kappa\rho_n)$ with $\kappa\rho_n > h^s$.

Because the ω_k's are uncorrelated given $S_{\mathcal{K}}$ by Lemma 2,

$$\mathrm{Var}_m(T_n) = \frac{1}{2K^p} \sum_{k\in\mathcal{K}} \mathbb{E}_m[\mathbb{1}(N_k > 1)\mathrm{Var}_m(\omega_k | S_{\mathcal{K}})]$$

$$+ \frac{1}{2K^p} \mathrm{Var}_m\left[\sum_{k\in\mathcal{K}} \mathbb{1}(N_k > 1)\mathbb{E}_m(\omega_k | S_{\mathcal{K}})\right].$$

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

$$\frac{1}{2K^p} \sum_{k\in\mathcal{K}} \mathbb{E}[\mathbb{1}(N_k > 1)\mathrm{Var}_m(\omega_k | S_{\mathcal{K}})]$$

$$\leq \mathbb{E}_m v^2(K) + 2h^p \sum_{k\in\mathcal{K}} \mathbb{E}N_k(\mathbb{E}^k \delta(X))^2[\mathbb{E}^k \varepsilon^2 + \mathbb{E}^k \delta^2(X)]$$

$$\leq \mathbb{E}_m v^2(K) + C_5 nh^p \mathbb{E}[\mathrm{Proj}^2_{\mathcal{K}} \delta(X)] + C_6 n\mathbb{E}^2[\mathrm{Proj}_{\mathcal{K}} \delta^2(X)],$$

$$\frac{1}{2K^p} \mathrm{Var}\left(\sum_{k\in\mathcal{K}} \mathbb{1}(N_k > 1)\mathbb{E}_m[\omega_k | S_{\mathcal{K}}]\right)$$

$$\leq \frac{1}{2K^p} \sum_k (\mathbb{E}^k \delta(X))^4 \mathrm{Var}((N_k - 1)\mathbb{1}(N_k > 1))$$

$$+ \frac{1}{2K^p} \sum_{k\neq k'} (\mathbb{E}^k \delta(X))^2 (\mathbb{E}^{k'} \delta(X))^2$$

$$\times \mathrm{Cov}((N_k - 1)\mathbb{1}(N_k > 1), (N_{k'} - 1)\mathbb{1}(N_{k'} > 1))$$

$$\leq C_7 n\mathbb{E}^2[\mathrm{Proj}^2_{\mathcal{K}} \delta(X)] + C_8 nh^p \mathbb{E}^2[\mathrm{Proj}^2_{\mathcal{K}} \delta(X)],$$

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

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

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

Main proof.

Part i. From (5.11) and Proposition 6 and as $A = 0$ under H_0, it suffices to show that $T_n/v_n \xrightarrow{d} \mathcal{N}(0,1)$. Assume that some ordering (denoted by $\le$) 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 $\sum_{\ell<k} N_\ell < J_i \le \sum_{\ell \le k} N_\ell$. Let $\mathcal{F}_{n,k} = \{S_{\mathcal{K}}, Y_{J_i} : \sum_{\ell<k} N_\ell < J_i \le \sum_{\ell \le k} N_\ell\}$. Under H_0, $\{T_{n,k} = \sum_{k' \le k} \omega_{k'} \big/ \sqrt{2K^p}, \mathcal{F}_{n,k}\}$, where $w_k = \langle U^*, U^* \rangle_k$, is a zero-mean martingale array. It is then sufficient to show that

$$v_n^{-2} \sum_{k \in \mathcal{K}} \mathbb{E}_0[\omega_k^2/(2K^p) | \mathcal{F}_{n,k-1}] \xrightarrow{p} 1, \tag{5.13}$$

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

$$\forall \eta > 0 \tag{5.14}$$

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

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

from Lemma 2, so that (5.13) follows from Proposition 5. Now (5.14) is implied by

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

By Assumption I, straightforward computations lead to

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

By Proposition 5, (5.14) follows. $\blacksquare$

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

Part ii. As v_n^2 is bounded from above uniformly in $m(\cdot)$ from Proposition 5, equation (5.11) and Proposition 6 yield

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

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

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

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

$$\begin{aligned}
&\mathbb{E}_m T_n - z'_\alpha - 2M\sqrt{nh^p}\mathbb{E}^{1/2}\delta^2(X) > 0, \\
&\frac{\mathrm{Var}_m T_n^*}{[\mathbb{E}_m T_n^* - z'_\alpha - 2M\sqrt{nh^p}\mathbb{E}^{1/2}\delta^2(X)]^2} \le \beta, \qquad \textbf{(5.15)}
\end{aligned}$$

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

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

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

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

and this upper bound is bounded because of Proposition 5 and decreasing in κ. Hence (5.15) can be made smaller than β uniformly for $m(\cdot)$ in $H_m(\kappa\tilde{\rho}_n)$ by taking κ large enough. ■

5.3. Proof of Theorem 4

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

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

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

Consider the a priori Π_{1n} defined in the proof of Theorem 1, 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, $h_n = \lambda \rho_n^{1/s}$, and $\varphi_k(\cdot)$ is defined by (5.1), and further assume that $\varphi(\cdot)$ is bounded with r-first zero moments. We have

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

uniformly in ξ. Now, uniformly in ξ and k,

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

and

$$\begin{aligned} I\!E[w(X_i, \xi) \varphi_k(X_i)] &= h_n^{p/2} \int w(lkh_n + uh_n, \xi) f(lkh_n + uh_n) \varphi(u)\, du \\ &= O(h_n^{r+p/2}). \end{aligned}$$

Hence, we have uniformly in ξ

$$I\!E_{\Pi_{1n}} z_1^2(\xi) = \lambda^2 \rho_n^2 O(1) + \lambda^2 \rho_n^2 n h^{2r+p} O(1).$$

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

$$I\!E_{\Pi_{1n}} \int |z_1(\xi)|^q\, d\nu(\xi) = o(1).$$

Thus

$$\begin{aligned} \sup_{H_1(\rho_n)} I\!P_m(I_{n,q} \le u_{\alpha,q}) &\ge \int I\!P_m(I_{n,q} \le u_{\alpha,q})\, d\Pi_{1n}(m) \\ &\ge \int I\!P_m\left(\left[\int |z_0(\xi)|^q\, d\nu(\xi) \right]^{1/q} \le u_{\alpha,q} \right) d\Pi_{1n}(m) \\ &\quad + o(1) \\ &\ge I\!P_0(I_{n,q} \le u_{\alpha,q}) + o(1) = 1 - \alpha + o(1). \end{aligned}$$

■

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

NOTES

1. The main difference lies in the compactness of the set of first derivatives.

2. Nonlinear least-squares estimators are adapted to our framework, but one could use estimators designed for specific purposes, as soon as they satisfy Assumption M3. See, e.g., Fan and Huang (2001).

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

4. In the expansion, the remainder term is zero with standard normal errors. Non-normality or heteroskedasticity induces a remainder term that must be studied via the Fatou lemma and some truncation arguments as done in Ibragimov and Has'minskii (1981) for efficient parametric estimation.

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 $r_n \propto n^{-1/2}$.

6. Suprema should then be considered over this set in Definition 1.

7. This follows from Propositions 5, 6, and 8 in Section 5.

8. In the CTG model and alternatives defined through L_q-norms, Lepski, Nemirovski, and Spokoiny (1999) show that the optimal minimax testing rate and the optimal minimax estimation rate for the L_q-norm coincide when q is even only.

9. In the case of testing for a pure noise model with homoskedastic errors and regular alternatives, the specification test proposed by Dette and Munk (1998) is also based on (3.2), with σ^2 replaced by an inefficient difference-based estimator.

10. This can be shown by adapting Proposition 8 to the case $h \propto n^{-1/p}$, as formally established in a previous version of this paper.

REFERENCES

Andrews, D.W.K. (1997) A conditional Kolmogorov test. *Econometrica* 65, 1097–1128.

Baraud, Y., S. Huet, & B. Laurent (forthcoming) Adaptive Tests of Linear Hypotheses by Model Selection. *Annals of Statistics*.

Bierens, H.J. (1982) Consistent model specification tests. *Journal of Econometrics* 20, 105–134.

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

Bierens, H.J. & W. Ploberger (1997) Asymptotic theory of integrated conditional moment tests. *Econometrica* 65, 1129–1151.

Birgé, L. & P. Massart (1993) Rates of convergence for minimum contrast estimators. *Probability Theory and Related Fields* 97, 113–150.

Brown, L.D. & M.G. Low (1996) Asymptotic equivalence of nonparametric regression and white noise. *Annals of Statistics* 24, 2384–2389.

Brown L.D. & C.H. Zhang (1998) Asymptotic nonequivalence of nonparametric experiments when the smoothness index is 1/2. *Annals of Statistics* 26, 279–287.

De La Peña, V.H. & E. Gine (1999) *Decoupling: From Dependence to Independence*. New York: Springer-Verlag.

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

Delgado, M.A., M.A. Dominguez, & P. Lavergne (2000) Consistent Tests of Conditional Moment Restrictions. Manuscript, Universidad Carlos III de Madrid.

Dette, H. & 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, 253–275.

Efromovich S. & A. Samarov (1996) Asymptotic equivalence of nonparametric regression and white noise has its limits. *Statistics and Probability Letters* 28, 143–145.

Eubank R.L. & J.D. Hart (1993) Commonality of cusum von Neumann and smoothing-based goodness-of-fit tests. *Biometrika* 80, 89–98.

Fan J. & L.S. Huang (2001) Goodness-of-fit tests for parametric regression models. *Journal of the American Statistical Association* 96, 640–652.

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

Fan, Y. & Q. Li (1996) Consistent model specification tests: Omitted variables and semiparametric functional forms. *Econometrica* 64, 865–890.
Fan, Y. & Q. Li (2000) Consistent model specification tests: Nonparametric versus Bierens' tests. *Econometric Theory* 16, 1016–1041.
Godfrey, L.G. (1988) *Misspecification Tests in Econometrics*. Cambridge: Cambridge University Press.
Gourieroux, C., A. Monfort, & A. Trognon (1984) Pseudo-maximum likelihood methods: Theory. *Econometrica* 52, 681–700.
Guerre, E. & P. Lavergne (2001) Rate-Optimal Data-Driven Specification Testing in Regression Models. Manuscript, LSTA Université Paris VI.
Guerre, E. & O. Lieberman (2000) α-Level Adaptive Testing in Nonparametric Regression via Selection Criteria. Manuscript, LSTA Université Paris VI.
Hall, P. & C.C. Heyde (1980) *Martingale Limit Theory and Its Application*. New York: Academic Press.
Härdle, W. & E. Mammen (1993) Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21, 1926–1947.
Härdle, W. & 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. & H. White (1995) Consistent specification testing via non-parametric series regressions. *Econometrica* 63, 1133–1160.
Horowitz, J.L. & V.G. Spokoiny (2001) An adaptive rate-optimal test of a parametric model against a nonparametric alternative. *Econometrica* 69, 599–631.
Ibragimov, I.A. & R.Z. Has'minskii (1981) *Statistical Estimation: Asymptotic Theory*. New York: Springer-Verlag.
Ingster, Y.I. (1993) Asymptotically minimax hypothesis testing for nonparametric alternatives, parts I, II, and III. *Mathematical Methods of Statistics* 2, 85–114, 171–189, and 249–268.
Lehmann, E.L. (1986) *Testing Statistical Hypothesis*, 2nd ed. New York: Wiley.
Lepski, O., A. Nemirovski, & V. Spokoiny (1999) On estimation of L_r norm of a regression function. *Probability Theory and Related Fields* 113, 221–253.
Li, Q. & S. Wang (1998) A simple consistent bootstrap test for a parametric regression functional form. *Journal of Econometrics* 87, 145–165.
Neyman, J. (1937) Smooth test for goodness of fit. *Skandinavisk Aktuarietidskrift* 20, 149–199.
Powell, J.L. & T.M. Stoker (1996) Optimal bandwidth choice for density-weighted averages. *Journal of Econometrics* 75, 291–316.
Rudin, W. (1991) *Functional Analysis*, 2nd ed. New York: McGraw-Hill.
Shorack, G.R. & J.A. Wellner (1986) *Empirical Processes with Application to Statistics*. New York: Wiley.
Spokoiny, V.G. (1996) Adaptive hypothesis testing using wavelets. *Annals of Statistics* 24, 2477–2498.
Spokoiny, V.G. (2001). Data-driven testing the fit of linear models. *Mathematical Methods of Statistics* 10, 465–497.
Stinchcombe, M.B. & H. White (1998) Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory* 14, 295–325.
Stone, C.J. (1982) Optimal rates of convergence for nonparametric estimators. *Annals of Statistics* 8, 1348–1360.
Stute, W. (1997) Nonparametric model checks for regression. *Annals of Statistics* 25, 613–641.
White, H. (1981) Consequences and detection of misspecified nonlinear regression models. *Journal of the American Statistical Association* 76, 419–433.
White, H. (1989) An additional hidden unit test for neglected nonlinearity. *Proceedings of the International Joint Conference on Neural Networks*, vol. 2, 451–455. New York: IEEE Press.
Zheng, X. (1996) A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics* 75, 263–289.

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

APPENDIX A: AUXILIARY RESULTS

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

$$\mathbb{E}_m[\omega_k | S_{\mathcal{K}}] = (N_k - 1)(\mathbb{E}^k \delta(X))^2,$$

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

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 i.i.d. 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,*

$$\mathbb{E}[(N_k - 1)\mathbb{1}(N_k > 1)] \geq \frac{n}{2} \mathbb{P}(X \in I_k) \quad \forall k \in \mathcal{K},$$

$$\mathrm{Var}[(N_k - 1)\mathbb{1}(N_k > 1)] \leq 2n\mathbb{P}(X \in I_k) \quad \forall k \in \mathcal{K},$$

$$\mathrm{Cov}[(N_k - 1)\mathbb{1}(N_k > 1), (N_{k'} - 1)\mathbb{1}(N_{k'} > 1)] \leq 2n\mathbb{P}(X \in I_k)\mathbb{P}(X \in I_{k'}) \quad \forall k \neq k' \in \mathcal{K}.$$

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

$$\mathbb{E}[(N_k - 1)\mathbb{1}(N_k > 1)] = n\mathbb{P}(X \in I_k) - 1 + (1 - \mathbb{P}(X \in I_k))^n \geq \frac{n}{2} \mathbb{P}(X \in I_k),$$

$$\mathrm{Var}[(N_k - 1)\mathbb{1}(N_k > 1)] \leq n\mathbb{P}(X \in I_k)[1 - \mathbb{P}(X \in I_k)] + \tfrac{1}{4} - 2\mathbb{E}(N_k)\mathbb{P}(N_k = 0) \leq 2n\mathbb{P}(X \in I_k).$$

The covariance equals

$$\mathrm{Cov}(N_k, N_{k'}) + \mathrm{Cov}(\mathbb{1}(N_k = 0), \mathbb{1}(N_{k'} = 0)) + \mathrm{Cov}(N_k, \mathbb{1}(N_{k'} = 0)) + \mathrm{Cov}(N_{k'}, \mathbb{1}(N_k = 0)).$$

The first item is $-\mathbb{E}(N_k)\mathbb{P}(X \in I_{k'})$, 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,

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

■

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

LEMMA 4. *Under Assumptions D and I, if* $n/(K^p \log K^p) \to \infty$,

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

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

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

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

$$\mathbb{P}\left[\min_{k\in\mathcal{K}} \mathbb{1}(N_k > 1) = 0\right] \leq \sum_{k\in\mathcal{K}} \mathbb{P}[N_k = 0] \leq \sum_{k\in\mathcal{K}} \mathbb{P}\left[\left|\frac{N_k - \mathbb{E}N_k}{\sqrt{n}}\right| \geq \frac{\mathbb{E}N_k}{\sqrt{n}}\right]$$

$$\leq 2K^p \exp\left[-\frac{3}{8} f\frac{n}{K^p}\right] \to 0,$$

as $\mathbb{E}N_k \geq fn/K^p$ under Assumption D, and $n/(K^p \log K^p) \to \infty$. Moreover, for any $t > 0$,

$$\mathbb{P}\left(\max_{K\in\mathcal{K}} \left|\frac{N_k}{\mathbb{E}N_k} - 1\right| \geq t\right) \leq \sum_{K\in\mathcal{K}} \mathbb{P}\left[\left|\frac{N_k - \mathbb{E}N_k}{\sqrt{n}}\right| \geq \frac{t\mathbb{E}N_k}{\sqrt{n}}\right]$$

$$\leq 2K^p \exp\left[-\frac{t^2}{2(1+t/3)} f\frac{n}{K^p}\right] \to 0. \qquad \blacksquare$$

APPENDIX B: PROOF OF PROPOSITION 7

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

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

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

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

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.,

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

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

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

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

Define $\|\delta\|_2^2 = I\!E\delta^2(X)$. Let $P_{m,q}(\cdot)$ be the Taylor expansion of order $[s]$ of $m(\cdot)$ around qh. Because $m(\cdot)$ 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 $P_m(\cdot)$ is such that $P_m(\cdot) = P_{m,q}(\cdot)$ on Δ_q, we have

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

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 contracting, 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

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

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

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

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

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

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

Let $\Pi_q(h) = \mathrm{Diag}(I\!P(X \in I_k), I_k \subset \Delta_q)$. 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)

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

Hence the left-hand side of (B.4) can be written as

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

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

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

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

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

$$I\!E[P^2(X) 1\!I(X \in \Delta_q)] \le I\!P(X \in \Delta_q) a^\top D_q(h) a \le F(2s'h)^p a^\top D_q(h) a,$$

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

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

Step 3. We can limit ourselves to establishing (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

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

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

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

uniformly in k, q, because $f(\cdot)$ 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. From their definitions, $I_k = kh + h[0,1)^p \subset \Delta_q = q + h[0,1)^p$ implies that $k = (k_1, \ldots, k_p)^\top$ and $q = (\kappa(r_{1,q}), \ldots, \kappa(r_{p,q}))^\top$ are such that $\kappa(r_{j,q}) \le k_j < \kappa(r_{j,q} + 1)$, independently of h. Therefore,

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

As $\sum_{j=1}^p \beta_j \le [s]$, 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 $a = (a_\beta, \sum_{j=1}^p \beta_j \le [s])$ with $B_q a = 0$; i.e., for all k such that $I_k \subset \Delta_q$,

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

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

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

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

with $\pi=(\pi_1,\dots,\pi_p)^\top$ 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 $P(\cdot)=0$. Note that $\mathcal{P}(1)$ holds, because (B.7) and the mean value theorem give that $P(x(\pi))=0$ for some $x(\pi)$ in $]\pi,\pi+1[$, $\pi=0,\dots,s'$. As the univariate polynomial function $P(\cdot)$ of degree $[s]$ should have at least $[s]+1$ distinct roots, this is possible only if $P(\cdot)\equiv 0$. We now show that $\mathcal{P}(p-1)$ implies $\mathcal{P}(p)$. Assume that $P(x)$ of degree $[s]$ with $x=(x_1,\dots,x_p)^\top$ in $[0,1]^p$ is such that (B.7) holds. Define

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

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

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

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

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

Then $\mathcal{P}(1)$ shows that $P_{x_{-1}}(\cdot)\equiv 0$ 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 cumulative distribution function (c.d.f.) of the design be $1-x^{-\gamma}$, $x\ge 1$, $\gamma>0$. If $2s>\gamma$, there exists a sequence $\{m_n(\cdot)\}_{n\ge1}$ of functions in $C_1(L,s)$ with $\mathbb{E}^{1/2}m_n^2(X)\ge\rho$, such that, for any α-level test t_n, $\liminf_{n\to+\infty}\mathbb{P}_{m_n}(t_n=1)\ge 1-\alpha$.*

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

$$I(x)=\frac{\mathbb{1}(x\ge0)}{(s+1)!}\int_0^x t^{s+1}\exp(-t)\,dt,$$

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

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press

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

Because the functions $(x - x_n)^k I^{(k)}(x - x_n)$, $k = 0, \ldots, s$, are bounded, $m(\cdot)$ 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,

$$\mathbb{P}_{m_n}\left(\tau_n = 0, \sup_{1\leq i\leq n} X_i \leq x_n\right) = \mathbb{P}_{H_0}\left(\tau_n = 0, \sup_{1\leq i\leq n} X_i \leq x_n\right).$$

This leads to

$$\begin{aligned}\mathbb{P}_{m_n}(\tau_n = 1) &\geq \mathbb{P}_{m_n}\left(\tau_n = 1, \sup_{1\leq i\leq n} X_i \leq x_n\right) = \mathbb{P}_{H_0}\left(\tau_n = 1, \sup_{1\leq i\leq n} X_i \leq x_n\right) \\ &\geq \mathbb{P}_{H_0}(\tau_n = 1) - \mathbb{P}\left(\sup_{1\leq i\leq n} X_i > x_n\right) \\ &\geq 1 - \alpha - n\mathbb{P}(X > x_n) = 1 - \alpha - nn^{-2}.\end{aligned}$$

■

https://doi.org/10.1017/S0266466602185069 Published online by Cambridge University Press