Econometric Theory, **16**, 2000, 576–601. Printed in the United States of America.

NONPARAMETRIC SIGNIFICANCE TESTING

Pascal Lavergne
INRA-ESR

Quang Vuong
University of Southern California
and
INRA-ESR

A procedure for testing the significance of a subset of explanatory variables in a nonparametric regression is proposed. Our test statistic uses the kernel method. Under the null hypothesis of no effect of the variables under test, we show that our test statistic has an $nh^{p_2/2}$ standard normal limiting distribution, where p_2 is the dimension of the complete set of regressors. Our test is one-sided, consistent against all alternatives and detects local alternatives approaching the null at rate slower than $n^{-1/2}h^{-p_2/4}$. Our Monte-Carlo experiments indicate that it outperforms the test proposed by Fan and Li (1996, *Econometrica* 64, 865–890).

1. INTRODUCTION

In recent years, considerable work has been devoted to testing a parametric regression model against a semi- or a nonparametric alternative. An approach that has attracted a lot of attention relies on smoothing techniques and compares the parametric fit with a smooth nonparametric one. Examples include Cleveland and Devlin (1988), Eubank and Spiegelman (1990), Eubank and Hart (1993), Gozalo (1993), Härdle and Mammen (1993), Chen (1994), Horowitz and Härdle (1994), Hong and White (1995), and Zheng (1996), among others.[1]

In contrast, the issue of testing a nonparametric null against a nonparametric alternative has attracted less attention. A leading case where such a situation naturally arises is testing the significance of some explanatory variables in a regression function. Well-known procedures have been proposed in parametric settings, but their outcomes crucially depend on the choice of the parametric specification. When it is not desirable to adopt a finite parameterization, non-

Financial support from INRA and the National Science Foundation under grant SBR-9631212 is gratefully acknowledged. Earlier versions were presented at Université Toulouse 1, CentER—Tilburg, University Carlos III—Madrid, CREST—Paris, INRA—Jouy-en-Josas, ESEM 96—Istanbul, Séminaire Paris-Berlin 96, UC San Diego, the University of Southern California, and Texas Camp Econometrics. We thank participants, three referees, and the co-editor for helpful comments. Address correspondence to: Pascal Lavergne, INRA-ESR, BP 27, 31326 Castanet-Tolosan Cedex, France; e-mail: lavergne@toulouse.inra.fr.

  0266-4666/00 $9.50

https://doi.org/10.1017/S0266466600164059 Published online by Cambridge University Press

parametric regression provides a suitable alternative. A special case that has been investigated in several previous papers is the problem of testing for no effect. To our knowledge, the general case where the nonparametric null is nondegenerate has been considered only in a few published studies. Gozalo (1993) considers conditional moment tests that are made consistent against all alternatives through randomization. Yatchew (1992) considers the difference in residual sums of squares and uses sample splitting to circumvent its well-known $\sqrt{n}$-degeneracy in a nested situation, whereas Lavergne and Vuong (1996) treat the nonnested case.[2]

Our objective is to propose a testing procedure for the significance of a subset of explanatory continuous variables in a nonparametric regression, which circumvents the drawbacks of previously proposed procedures. Namely, our procedure does not use randomization but is nevertheless consistent against any deviation from the null hypothesis of no effect of the variables under test. Instead of using sample splitting or weighting, we deal with the $\sqrt{n}$-degeneracy issue to obtain a test statistic with a faster rate than $\sqrt{n}$. Our test statistic is based on the kernel method. We characterize its asymptotic distribution not only under the null hypothesis but also under a sequence of local alternatives. Our assumptions do not require normality or homoscedasticity of the regression errors and are not much more demanding on the bandwidths and on the considered functions than in nonparametric estimation. Though our test statistic is similar in spirit to that recently proposed by Fan and Li (1996a), we require less restrictive theoretical conditions on the smoothing parameters. As a result, our testing procedure does not require oversmoothing of the null regression model relative to the alternative one and hence puts both models on equal footing. In small samples, our simulation results show that our test statistic is nearly unbiased under the null hypothesis and leads to a test that is more powerful than Fan and Li's under a wide spectrum of alternatives.

The paper is organized as follows. In Section 2, we present our test statistic and we study its asymptotic properties under a sequence of local alternatives. We also show how our framework can accommodate the special case of testing the joint significance of all the regressors. Section 3 studies the small sample behavior of our testing procedure by means of Monte-Carlo experiments and compares it with the one proposed by Fan and Li (1996a). All the proofs are relegated to Section 4.

2. THEORETICAL RESULTS

Suppose (X_{2i}, Y_i), $i = 1, \ldots, n$ is a random sample from a $(p_2 + 1)$-variate distribution of (X_2, Y) and let $X_1 \subset X_2$ be a p_1-vector, $0 < p_1 < p_2$. Throughout we denote densities of X_1 and X_2 by $f_1(\cdot)$ and $f_2(\cdot)$. Let $E[Y|X_1] = r_1(X_1)$ and $E[Y|X_2] = r_2(X_2)$. The null hypothesis of interest is $H_0: r_1(X_1) = r_2(X_2)$ a.s., or equivalently, $H_0: E[u_1|X_2] = 0$ a.s., where $u_1 = Y - r_1(X_1)$. Our procedure can be viewed as a test of the unconditional moment restriction $E[u_1 \Psi(X_2)] = 0$,

https://doi.org/10.1017/S0266466600164059 Published online by Cambridge University Press

with $\Psi(X_2) = E(u_1|X_2)f_1^2(X_1)f_2(X_2)$. Indeed, this particular choice makes the test consistent against any alternative to H_0 as

$$E[u_1 E(u_1|X_2)f_1^2(X_1)f_2(X_2)] = E[E^2(u_1|X_2)f_1^2(X_1)f_2(X_2)]$$
$$= E[(r_2(X_2) - r_1(X_1))^2 f_1^2(X_1)f_2(X_2)].$$

Let K and L be two kernels on $\mathbb{R}^{p_2}$ and $\mathbb{R}^{p_1}$, respectively, and let h and g be two bandwidths. To test H_0, we consider

$$V_n = \frac{1}{n^{(4)}} \sum_a (Y_i - Y_k)(Y_j - Y_l) L_{nik} L_{njl} K_{nij}, \tag{2.1}$$

where $\sum_a$ denotes summation over the arrangements of m distinct elements $\{i_1, \ldots, i_m\}$ from $\{1,\ldots,n\}$ with $n^{(m)} = n!/(n-m)!$ the number of these arrangements, and where $L_{nik} \equiv g^{-p_1}L[(X_{1i} - X_{1k})/g]$ and $K_{nij} \equiv h^{-p_2}K[(X_{2i} - X_{2j})/h]$.

The statistic V_n is simple to compute and in particular does not require any trimming. It constitutes a natural basis for testing H_0, because it actually estimates $E[u_1 \Psi(X_2)]$. Indeed, assuming that $u_{1i} f_1(X_{1i})$ is observed, a sample analog of the latter is

$$V_{0n} = \frac{1}{n^{(2)}} \sum_a u_{1i} f_1(X_{1i}) u_{1j} f_1(X_{1j}) K_{nij}.$$

Fan and Li (1996a) obtain their statistic I_n from V_{0n} by replacing $u_{1i} f_1(X_{1i})$ by its leave-one-out kernel estimate $\hat{u}_{1i} \hat{f}_{1i}$. Although our test statistic resembles Fan and Li's, it was derived independently and differs from theirs by some important terms.[3] Specifically,

$$n^{(4)} V_n = n(n-1)^3 I_n - n^{(3)} V_{1n} - 2n^{(3)} V_{2n} + n^{(2)} V_{3n},$$

with

$$V_{1n} = \frac{1}{n^{(3)}} \sum_a (Y_i - Y_k)(Y_j - Y_k) L_{nik} L_{njk} K_{nij},$$

$$V_{2n} = \frac{1}{n^{(3)}} \sum_a (Y_i - Y_j)(Y_j - Y_k) L_{nij} L_{njk} K_{nij}$$
$$= \frac{1}{n^{(3)}} \sum_a (Y_i - Y_j)(n-1) \hat{u}_{1j} \hat{f}_{1j} L_{nij} K_{nij} + \frac{V_{3n}}{(n-2)},$$

and

$$V_{3n} = \frac{1}{n^{(2)}} \sum_a (Y_i - Y_j)^2 L_{nij}^2 K_{nij}.$$

In effect, our statistic V_n removes all "diagonal" terms from I_n, thus reducing the bias of the statistic without altering its properties as a test statistic for H_0.

https://doi.org/10.1017/S0266466600164059 Published online by Cambridge University Press

A similar idea has been recently proposed by Heffernan (1997) for unbiased estimation of central moments by U-statistics. In small samples, the bias reduction can be substantial, as our Monte-Carlo study of Section 3 shows. In practice, one may use the preceding formula instead of (2.1) to compute V_n.

To study the behavior of V_n under the null and some local alternative hypotheses simultaneously, we write

$$H_{1n}: r_2(X_2) = r_1(X_1) + \delta_n d(X_2), \quad \text{with } \delta_n \in [0,1].$$

We let $d(X_2) \equiv 0$ if $\delta_n = 0$. This general formulation allows us to include local alternatives, whose rates of convergence to H_0 are given by δ_n. We need the following definitions and assumptions.

DEFINITION 1.

(i) $\mathcal{U}^p$ *is the class of integrable uniformly continuous functions from* $\mathbb{R}^p$ *to* $\mathbb{R}$.
(ii) $\mathcal{D}^p_{m,q}$ *is the class of m-times differentiable functions from* $\mathbb{R}^p$ *to* $\mathbb{R}$, *with derivatives of order m that are uniformly Lipschitz continuous of order q.*

DEFINITION 2. $\mathcal{K}^p_m$, $m \geq 2$, *is the class of even integrable functions* K: $\mathbb{R}^p \to \mathbb{R}$ *with compact support satisfying* $\int K(s)\,ds = 1$ *and*

$$\int s_1^{\alpha_1} \ldots s_p^{\alpha_p} K(s)\,ds = 0 \quad \text{for } 0 < \sum_{i=1}^{p} \alpha_i \leq m-1, \alpha_i \geq 0 \quad \forall i.$$

Assumption 1. $\{(X_{2i}, Y_i), i = 1, \ldots, n\}$ is an independent and identically distributed (i.i.d.) sample from an absolutely continuous (with respect to Lebesgue measure) (p_2+1)-variate distribution, and $E[Y^8] < \infty$.

Assumption 2.

(i) $f_1(X_1)$ and $r_1(X_1)f_1(X_1)$ belong to $\mathcal{U}^{p_1} \cap \mathcal{D}^{p_1}_{m_1,q_1}$, $m_1 \geq 2$, and also $E(u_1^2|X_1) \times f_1(X_1)$ belongs to $\mathcal{U}^{p_1}$.
(ii) $f_2(X_2)$, $r_2(X_2)f_1(X_1)f_2(X_2)$, $E(u_1^2|X_2)f_1^2(X_1)f_2(X_2)$ and $E(u_1^4|X_2)f_1^4(X_1)f_2(X_2)$ belong to $\mathcal{U}^{p_2}$.
(iii) $K(\cdot) \in \mathcal{K}^{p_2}_2$ and $L(\cdot) \in \mathcal{K}^{p_1}_{m_1}$.

THEOREM 1. *Under Assumptions* 1 *and* 2, *if* $h \to 0$, $g \to 0$, $nh^{p_2} \to +\infty$, $ng^{p_1} \to \infty$, $h^{p_2}/g^{p_1} \to 0$, *and* $nh^{p_2/2}g^{2(m_1+q_1)} \to 0$, *then as* $n \to \infty$,

(i) $nh^{p_2/2}V_n \xrightarrow{d} N(C\mu, \omega^2)$ *if* $\delta_n^2 nh^{p_2/2} \to C < \infty$,
(ii) $nh^{p_2/2}V_n \xrightarrow{p} +\infty$ *if* $\delta_n^2 nh^{p_2/2} \to \infty$,

where $\mu = E[d^2(X_2)f_1^2(X_1)f_2(X_2)]$ *and* $\omega^2 = 2E[E^2(u_1^2|X_2)f_1^4(X_1)f_2(X_2)] \times \int K^2(s)\,ds$.

Remark 1. As shown in Section 4, V_n has the same behavior as V_{0n}. In general, V_{0n} is such that $\sqrt{n}[V_{0n} - E(V_{0n})]$ converges to a normal distribution $N(0,\tau^2)$, where τ^2 is the semiparametric efficiency bound for estimating

https://doi.org/10.1017/S0266466600164059 Published online by Cambridge University Press

$E[u_1\Psi(X_2)]$. But under H_0, we have both $E(V_{0n}) = 0$ and $\tau^2 = 0$. This degeneracy leads us to consider higher-order terms in the expansion of V_{0n}. For this we use a central limit theorem for degenerate U-statistics (see Hall, 1984a).[4]

Remark 2. Assumption 2 requires smoothness conditions on the underlying functions and kernels that are standard in nonparametric estimation. Functions of X_1 are assumed to be at least as smooth as functions of X_2. This is compatible with the nested situation under consideration. Instead, Fan and Li (1996a) requires similar smoothness of the constrained and unconstrained regression functions.

Remark 3. The generalization of our test to the situation where some of the X_1 are discrete with finite support is straightforward, as discrete variables neither create any bias nor change the variance of the nonparametric estimators. Our general results are not affected, where bandwidths only apply to continuous regressors. In particular, when all regressors X_1 are discrete and all regressors under test are continuous, our assumptions on the bandwidths reduce to the usual ones, i.e., $h \to 0$ and $nh^{p_2/2} \to +\infty$.

Remark 4. One of the main problems in obtaining asymptotic distributions of semiparametric estimators is the relative vanishing rates of the bias and variance terms from nonparametric estimation. For instance, Samarov (1993) notes that the bias term may dominate the variance term for his test statistic. Hall (1984a, 1984b) finds that the squared bias term of the integrated square error of kernel estimators is of order h^4. In the context of parametric specification testing, Hong (1993) and Gozalo (1995) find a bias term of order h^2 and propose a statistic that balances it with the variance term. In our context, the bias problem arises in each of the two smoothing steps: the nonparametric regression of Y on X_1 and the projection of the residual u_1 on X_2. The form of our statistic eliminates the bias in the second step, so that $E(V_{0n}) = 0$ under H_0. On the other hand, the bias from the first step is controlled through the "bias" condition $nh^{p_2/2}g^{2(m_1+q_1)} \to 0$, as in Fan and Li (1996a).

Remark 5. Though the theory is developed for a generic bandwidth (g or h) in each step, it is straightforward to extend it to a vanishing individual bandwidth for each regressor in each step.[5] In this case, one should replace g^{p_1} and h^{p_2} by $g_1 g_2 \dots g_{p_1}$ and $h_1 h_2 \dots h_{p_2}$, respectively. The "bias" condition becomes $n\prod_{i=1}^{p_2} h_i^{1/2}[\max_{i=1,\dots,p_1} g_i]^{2(m_1+q_1)} \to 0$.

Remark 6. Our assumptions on the bandwidths include the usual ones. The condition on the ratio h^{p_2}/g^{p_1} means that the variance of nonparametric estimators in the model with p_1 regressors is smaller than the variance of nonparametric estimators in the complete model. This seems reasonable in view of the higher sparsity of the data in high dimensional spaces, leading to the well-known "curse of dimensionality." In our testing framework, this condition can be better understood by considering individual bandwidths. In this case, it seems natural to

https://doi.org/10.1017/S0266466600164059 Published online by Cambridge University Press

use individual bandwidths for the regressors X_1 not under test that are identical between both steps, namely, $g_i = h_i$, $i = 1,\dots,p_1$, to avoid incorrect rejection of the null hypothesis. Then our ratio condition reduces to $\prod_{i=p_1+1}^{p_2} h_i \to 0$. This is no longer restrictive as vanishing individual bandwidths are obviously necessary to obtain a consistent test. Hence our "ratio" condition on the relative rates of the bandwidths seems to be minimal for testing the significance of continuous regressors. In contrast, when the regressors under test are discrete, $\prod_{i=p_1+1}^{p_2} h_i$ need not vanish, so that restricted and nonrestricted nonparametric estimates jointly determine the limit distribution of the test statistic, as studied by Lavergne (2000).

Fan and Li (1996a) require the stronger condition $h^{p_2}/g^{2p_1} \to 0$. As a result, Fan and Li's testing procedure excludes a large domain of bandwidths, including the optimal bandwidth rates for estimation $n^{-1/[p+2(m+q)]}$ when the dimension of X_1 is close to the dimension of X_2, and this for any degree of smoothness in the underlying regressions. For instance, this arises when $p_2 = 2$ and $p_1 = 1$ or when $p_2 = 3$ and $p_1 = 2$. In contrast, our testing procedure allows for a broader choice, including the optimal estimation rates when the constrained regression is sufficiently smooth, though these optimal estimation rates need not be optimal for testing purposes (see Guerre and Lavergne, 1999).

The asymptotic variance ω^2 can be written as

$$2E[(u_1 f_1(X_1))^2 E[(u_1 f_1(X_1))^2 | X_2] f_2(X_2)] \int K^2(s)\, ds.$$

It depends on the kernel through $\int K^2(s)\, ds$. This quantity can be minimized in the class of product nonnegative even kernels by choosing the Epanechnikov kernel (see Epanechnikov, 1969). Following (2.1), an estimator of ω^2 is

$$\omega_n^2 = \frac{2}{n^{(6)}} \sum_a (Y_i - Y_k)(Y_i - Y_{k'})(Y_j - Y_l)(Y_j - Y_{l'}) L_{nik} L_{nik'} L_{njl} L_{njl'} h^{p_2} K_{nij}^2.$$

An alternative estimator, which is computationally less demanding but more biased in small samples, is

$$\omega_n^2 = \frac{2}{n^{(2)}} \sum_a \hat{u}_{1i}^2 \hat{u}_{1j}^2 \hat{f}_{1i}^2 \hat{f}_{1j}^2 h^{p_2} K_{nij}^2, \tag{2.2}$$

where $\hat{u}_{1i} \hat{f}_{1i}$ is the kernel estimator of $u_{1i} f_1(X_{1i})$. The consistency of either form of ω_n^2 is shown using similar arguments as in the proof of Theorem 1 (see Sec. 4). Therefore, we can propose $nh^{p_2/2} V_n/\omega_n$ as a test statistic for H_0. From Theorem 1, by letting $\delta_n = 0$ or 1, this test statistic is asymptotically $N(0,1)$ under H_0 and diverges to $+\infty$ against any fixed alternative to H_0. The test is therefore a one-sided normal test. Moreover, the test has power to detect local alternatives H_{1n} approaching the null at rate slower than

https://doi.org/10.1017/S0266466600164059 Published online by Cambridge University Press

$(\sqrt{n}h^{p_2/4})^{-1}$. This rate agrees with that found in parametric specification testing procedures that use smoothing.[6]

Although Theorem 1 suggests that suitable critical values for our testing procedure can be obtained from the standard normal distribution, results from Eubank and LaRiccia (1993) and Härdle and Mammen (1993), among others, indicate that the normal approximation may not be adequate for small sample sizes. Indeed, our test statistic behaves like a weighted sum of chi-squares, in an asymptotic sense, and accordingly may approach normality slowly, especially for high dimensional settings. One alternative is to use a χ^2 approximation, as proposed by Hall (1983) and Buckley and Eagleson (1988) and used by Eubank and LaRiccia (1993) and Chen (1994) in the context of parametric specification testing. Although such a correction may help in high dimensions, it did not prove very useful in our limited Monte-Carlo experiments, where the normal approximation seems to work well. Another alternative is to use resampling techniques, such as the wild bootstrap considered by Härdle and Mammen (1993). The theoretical justification of such a technique in our context, and specifically the conditions under which it applies, is left for further research.

Last, it is possible to extend our procedure to the case where $p_1 = 0$, i.e., testing for no effect of all the regressors X_2. In this case the null hypothesis of interest is H_0^*: $r_2(X_2) = C \equiv E(Y)$ a.s. To test H_0^*, we can readily modify (2.1) to get

$$V_n^* = \frac{1}{n^{(4)}} \sum_a (Y_i - Y_k)(Y_j - Y_l) K_{nij}.$$

As before, we consider the local alternatives $r_2(X_2) = E(Y) + \delta_n d(X_2)$, with $\delta_n \in [0,1]$. Our Assumption 2 now reduces to the usual one in nonparametric estimation, namely, the following assumption.

Assumption 3.

(i) $f_2(X_2)$, $r_2(X_2) f_2(X_2)$, $\mathrm{Var}(Y|X_2) f_2(X_2)$, and $E((Y-c)^4|X_2) f_2(X_2)$ belong to $\mathcal{U}^{p_2}$.

(ii) $K \in \mathcal{K}_2^{p_2}$.

COROLLARY 1. *Under Assumptions* 1 *and* 3, *if* $h \to 0$, $nh^{p_2} \to +\infty$, *then as* $n \to \infty$,

(i) $nh^{p_2/2} V_n^* \xrightarrow{d} N(C\mu^*, \omega^{*2})$ *if* $\delta_n^2 nh^{p_2/2} \to C < \infty$,

(ii) $nh^{p_2/2} V_n^* \xrightarrow{p} +\infty$ *if* $\delta_n^2 nh^{p_2/2} \to \infty$,

where $\mu^* = E[d^2(X_2) f_2(X_2)]$ *and* $\omega^{*2} = 2E[\mathrm{Var}^2(Y|X_2) f_2(X_2)] \int K^2(s)\, ds$.

A consistent estimator of ω^{*2} is

$$\omega_n^{*2} = \frac{2}{n^{(6)}} \sum_a (Y_i - Y_k)(Y_i - Y_{k'})(Y_j - Y_l)(Y_j - Y_{l'}) h^{p_2} K_{nij}^2.$$

Alternatively, a simpler estimator analogous to (2.2) can be computed as

$$\omega_n^{*2} = \frac{2}{n^{(2)}} \sum_a (Y_i - \overline{Y})^2 (Y_j - \overline{Y})^2 h^{p_2} K_{nij}^2,$$

where $\overline{Y}$ is the empirical mean of the Y_i's. The latter is computationally less costly but more biased in small samples than the former. A consistent one-sided normal test for no effect of X_2 in the regression of Y can thus be based on $nh^{p_2/2} V_n^* / \omega_n^*$. As before, this test has power to detect local alternatives approaching the null at a rate slower than $(\sqrt{n} h^{p_2/4})^{-1}$.

Many other tests have been previously proposed for the special case of testing for no effect, as reviewed in Hart's (1997) monograph. Because the null is very simple in this case, it is possible to apply the empirical process approach and to derive omnibus tests (see, e.g., Buckley, 1991; Bierens, 1982, 1990). Alternatively, tests based on smoothing ideas have been considered (see, e.g., Eubank and Hart, 1993). In particular, it is possible to allow for data-driven smoothing parameters in such tests (see Barry and Hartigan, 1990) or to construct a test based on the smoothing parameter itself (see Eubank and Hart, 1992). However, with the exception of Bierens (1982, 1990), all these tests have been developed in the special case of a single regressor and homoscedastic errors. In addition, the limiting behavior of some of these tests is nonstandard.

The statistic V_n^* resembles Zheng's (1996) statistic for parametric specification testing in the case where the parametric model reduces to the constant regression, but removes all "diagonal" terms from the latter to make it unbiased under the null. Indeed, because the estimation of the smallest regression model is actually parametric, there is no bias corresponding to this stage. As the form of our statistic also eliminates the bias in the second stage, a notable feature of our statistic is that it is unbiased under the null hypothesis, i.e., $E(V_n^*) = 0$ under H_0. This is especially valuable in small samples.

3. MONTE-CARLO STUDY

In this section, we investigate the small sample behavior of our test and study its performances relative to Fan and Li's (1996a) test (hereafter, the FL test). We generate data through

$$Y = aX_1 + bX_1^3 + d(W) + U, \tag{3.1}$$

where X_1 and W are independent and distributed as $N(0,1)$ and U is independently distributed of the regressors as $N(0, \sigma^2)$. The null hypothesis corresponds to $d(W) \equiv 0$, and we consider different forms of alternatives as specified by $d(\cdot)$. We impose the restriction that $E[d(W)] = 0$, so that the nonparametric regression $r_1(X_1)$ remains the same whatever the data generating process. We set the parameters a, b, and σ^2 to -1, 1, and 4, respectively, so that the part of

https://doi.org/10.1017/S0266466600164059 Published online by Cambridge University Press

the variance of Y explained in its nonparametric regression on X_1 is moderate, i.e., 71%.

We consider small ($n = 100$) and moderate ($n = 200$) sample sizes and run 2,000 replications. We choose $K(\cdot)$ and $L(\cdot)$ as product kernels of the univariate Epanechnikov kernel with support $[-1,1]$, i.e., $L(u) = (3/4)(1 - u^2)\mathbb{I}[|u| \leq 1]$. As indicated in Remark 5, we can use individual bandwidths. The bandwidth parameter for the restricted model is chosen as $g = \hat{s}_{X_1} n^{-1/5}$, where $\hat{s}_{X_1}$ is the estimated standard deviation of X_1. This corresponds to the usual rule of thumb in kernel estimation (see, e.g., Härdle, 1991). For the unrestricted model, we keep the same smoothing parameter as in the restricted one for the first dimension, i.e., X_1, and choose the parameter for the second dimension, i.e., W, as $h_2 = c\hat{s}_W n^{-1/5}$, where $\hat{s}_W$ is the estimated standard deviation of W. Keeping the same bandwidth for regressors that are common to both models, in our case X_1, seems a natural choice in our testing framework. For the regressors under test, i.e., W, we apply the same rule of thumb with an additional varying constant c to investigate the sensitivity of our results to the smoothing parameter's choice.[7]

The design of the alternatives has been chosen to investigate the power of the competing tests with respect to the magnitude and the frequency of $d(\cdot)$. For the magnitude, we consider three linear alternatives of the form

$$d(W) = \alpha W,$$

with $\alpha = 0.5$, 1, and 2 corresponding, respectively, to DGP_1, DGP_2, and DGP_3. This allows us to compare the performances of the nonparametric tests to the standard Fisher test based on the true model (3.3). Alternatives corresponding to varying frequencies are defined through

$$d(W) = \sin(\delta\pi W),$$

with $\delta = 2, 1, \frac{2}{3}$, and $\frac{1}{2}$ corresponding respectively to DGP_4, DGP_5, DGP_6, and DGP_7. These departures from the null are of special interest, as it is known that smoothing tests of parametric specifications are sensitive to the frequency of the alternatives (see Eubank and Hart, 1993; Fan and Li, 1996b; Kuchibhatla and Hart, 1996; Hart, 1997). We expect that such a feature will hold for nonparametric significance tests.

Table 1 reports our Monte-Carlo results for the null hypothesis (DGP_0) and the linear alternatives. For each sample size ($n = 100, 200$), we let the constant c be 0.25, 0.5, 1, 2, 4. For each case, the first and second rows give the mean with standard deviation in parentheses of our test and the FL test, respectively. For computational reasons, we use the simplest, but biased, estimator of the variance (2.2). The third and fourth rows give empirical levels of rejections for our test and the FL test. The first figure corresponds to a 5% nominal level, whereas the second one corresponds to a 10% nominal level. For each sample size, the last row reports empirical rejection rates of the F-test for the same nominal levels.

TABLE 1. Null and linear alternatives

n	c	DGP_0		DGP_1		DGP_2		DGP_3	
100	0.25	−0.019	(0.876)	0.148	(0.909)	0.583	(0.949)	1.683	(0.894)
		−0.350	(0.959)	−0.159	(0.998)	0.340	(1.044)	1.620	(0.973)
		3.3%	8.7%	5.6%	11.8%	14.3%	25.3%	53.7%	69.4%
		2.1%	5.2%	3.7%	8.5%	11.5%	20.0%	51.3%	66.1%
	0.5	−0.007	(0.857)	0.231	(0.922)	0.838	(1.009)	2.353	(0.986)
		−0.462	(0.941)	−0.190	(1.014)	0.506	(1.112)	2.268	(1.075)
		4.1%	8.5%	7.8%	14.1%	21.5%	33.4%	77.0%	85.8%
		2.0%	4.8%	5.1%	8.7%	16.1%	24.4%	73.1%	82.3%
	1.0	0.010	(0.800)	0.335	(0.918)	1.170	(1.085)	3.241	(1.136)
		−0.620	(0.877)	−0.247	(1.010)	0.711	(1.197)	3.120	(1.238)
		3.6%	7.4%	9.0%	15.4%	31.3%	43.4%	91.9%	96.1%
		1.5%	3.3%	4.8%	8.4%	21.8%	30.3%	88.6%	92.9%
	2.0	0.018	(0.707)	0.450	(0.895)	1.556	(1.173)	4.279	(1.329)
		−0.852	(0.779)	−0.358	(0.989)	0.909	(1.299)	4.070	(1.450)
		2.4%	5.3%	9.8%	16.9%	42.6%	54.3%	98.0%	99.0%
		0.7%	1.6%	4.1%	6.4%	27.1%	35.1%	95.4%	97.4%
	4.0	0.018	(0.503)	0.497	(0.755)	1.714	(1.090)	4.686	(1.297)
		−1.144	(0.564)	−0.598	(0.842)	0.790	(1.215)	4.220	(1.433)
		1.2%	2.2%	7.3%	13.1%	48.2%	60.9%	99.1%	99.8%
		0.1%	0.2%	2.1%	3.0%	22.0%	31.4%	97.0%	98.0%
	F-test	5.1%	10.1%	69.5%	79.2%	99.7%	99.9%	100.0%	100.0%
200	0.25	0.007	(0.915)	0.301	(0.951)	1.080	(1.015)	3.018	(0.970)
		−0.308	(0.964)	0.010	(1.002)	0.855	(1.071)	2.974	(1.016)
		4.3%	9.6%	7.9%	15.6%	28.8%	42.6%	91.8%	95.6%
		2.4%	5.6%	5.7%	11.1%	23.7%	34.9%	90.3%	94.8%
	0.5	0.012	(0.901)	0.425	(0.974)	1.520	(1.109)	4.227	(1.134)
		−0.430	(0.951)	0.017	(1.029)	1.205	(1.174)	4.164	(1.189)
		5.4%	9.6%	11.8%	18.5%	44.2%	56.8%	98.7%	99.4%
		2.5%	5.5%	7.1%	12.3%	34.5%	45.8%	98.3%	99.1%
	1.0	0.015	(0.856)	0.596	(1.011)	2.129	(1.271)	5.898	(1.397)
		−0.604	(0.905)	0.024	(1.072)	1.685	(1.347)	5.804	(1.467)
		4.5%	8.2%	15.0%	23.7%	62.4%	72.6%	99.8%	99.9%
		1.5%	3.3%	7.5%	12.5%	48.2%	59.2%	99.7%	99.8%
	2.0	0.016	(0.763)	0.807	(1.057)	2.877	(1.481)	7.944	(1.725)
		−0.848	(0.809)	0.005	(1.123)	2.247	(1.573)	7.780	(1.816)
		2.9%	6.6%	19.2%	27.8%	79.5%	86.3%	100.0%	100.0%
		0.6%	1.4%	8.3%	13.1%	62.9%	71.9%	100.0%	100.0%
	4.0	0.019	(0.558)	0.949	(0.967)	3.364	(1.496)	9.245	(1.830)
		−1.153	(0.596)	−0.150	(1.030)	2.461	(1.593)	8.868	(1.935)
		0.9%	2.5%	20.8%	31.0%	87.5%	93.1%	100.0%	100.0%
		0.1%	0.1%	6.6%	9.9%	69.5%	77.0%	100.0%	100.0%
	F-test	5.2%	10.6%	93.1%	96.9%	100.0%	100.0%	100.0%	100.0%

The first column relates to the null hypothesis. First, the mean of our test statistic is very close to zero; i.e., our test statistic is nearly unbiased, irrespective of the smoothing parameter (see Remark 3). This is in sharp contrast with the FL test statistic, which is always negatively biased, up to −1.15. Second,

https://doi.org/10.1017/S0266466600164059 Published online by Cambridge University Press

the standard deviations of both test statistics are smaller than one. This is due partly to the fact that the simple variance estimator (2.2) always overestimates the variance. Although both tests exhibit empirical sizes that are smaller than the nominal ones, the FL test can be considerably undersized as a result of its strong negative bias. The size of our test is much closer to its nominal size, especially for bandwidths that are somewhat smaller than the rule of thumb. The empirical level as a function of the bandwidth h_2 displays an inverse U shape, as for very small bandwidths h_2, our statistic is identically zero.

Regarding the linear alternatives, we find that the FL test statistic is more variable than ours. Moreover, our test statistic has a higher mean than the FL one, which is due to the negative bias of the latter. This leads to a systematic higher empirical power for our test. As expected, power is increasing with the magnitude of the departure from the null, as measured by α. Our test can detect small linear alternatives such as DGP_1, unlike the FL test, which has close to trivial power in this situation. Furthermore, the power performance of our test can equal that of the Fisher test (see DGP_3), although the design is ideal for the latter. Our results also indicate that the highest power is attained for our test for the largest tried bandwidth, which is expected because the alternative is linear and the kernel smoother is a straight line for large bandwidths. However, using an infinite bandwidth should ultimately lead to a trivial power.

Table 2 has the same structure as Table 1 and reports results relative to the sinus alternatives. As in Table 1, our test statistic exhibits a larger mean and a smaller variance than the FL one and hence achieves higher power in all cases. The empirical power as a function of h_2 displays an inverse U shape for both tests. As shown in Figure 1, our test uniformly dominates the FL test for a large range of bandwidths. The maximum power of our test can be up to 50% higher. It is achieved for a bandwidth that increases with the smoothness of the alternative, as could be expected. Hence, our results suggest that the bandwidth should be adapted to the frequency of the alternative, namely, the higher the frequency, the smaller the bandwidth should be.

For comparative purposes, we also provide the empirical rejection rates of the F-test assuming a linear specification in W. The lowest frequency alternative DGP_7 is close to a linear specification in the range $[-1,1]$. Given that W is $N(0,1)$, the F-test therefore performs quite well, whereas our test has acceptable power up to 78%. For high frequency alternatives DGP_4 and DGP_5, the F-test has trivial power irrespective of sample size, whereas our test can attain an empirical power of 50% or 68%, respectively, for a moderate sample size of 200.

In summary, our test has better size and power than the FL test in all cases and seems to exhibit good properties for a wide range of nonlinear alternatives. Our Monte-Carlo study points out the importance of the bandwidth choice. There is clearly a trade-off between size and power. A better sized test seems to be achieved by slight undersmoothing relative to the rule of thumb, whereas better power is obtained in most cases by oversmoothing of the variable under test.

TABLE 2. Sinus alternatives

n	c	DGP_4		DGP_5		DGP_6		DGP_7	
100	0.25	0.563	(0.979)	0.564	(0.982)	0.571	(0.984)	0.540	(0.971)
		0.288	(1.068)	0.291	(1.072)	0.297	(1.077)	0.265	(1.065)
		15.0%	24.7%	15.0%	25.3%	15.4%	25.4%	14.4%	24.3%
		11.2%	19.9%	11.7%	20.2%	12.4%	20.4%	11.0%	19.1%
	0.5	0.705	(1.007)	0.798	(1.041)	0.815	(1.053)	0.773	(1.039)
		0.319	(1.100)	0.422	(1.136)	0.441	(1.152)	0.397	(1.139)
		18.2%	28.5%	22.0%	31.7%	22.0%	32.6%	20.6%	31.2%
		12.6%	20.1%	16.0%	23.3%	16.0%	23.5%	14.6%	22.6%
	1.0	0.598	(0.914)	1.010	(1.086)	1.093	(1.130)	1.055	(1.116)
		0.025	(1.000)	0.478	(1.185)	0.571	(1.234)	0.532	(1.222)
		13.2%	22.2%	26.6%	37.5%	29.9%	40.8%	28.6%	38.8%
		6.6%	11.6%	16.8%	23.8%	18.7%	26.9%	18.1%	26.2%
	2.0	−0.030	(0.705)	0.848	(0.944)	1.232	(1.124)	1.296	(1.155)
		−0.906	(0.779)	0.056	(1.033)	0.479	(1.228)	0.555	(1.265)
		2.2%	4.4%	18.1%	27.9%	31.9%	44.0%	35.2%	46.1%
		0.6%	1.0%	7.5%	12.5%	16.5%	23.0%	17.8%	26.0%
	4.0	0.050	(0.497)	0.046	(0.541)	0.713	(0.791)	1.071	(0.922)
		−1.107	(0.558)	−1.116	(0.602)	−0.384	(0.871)	0.013	(1.018)
		0.7%	2.3%	1.3%	2.8%	11.8%	20.6%	23.4%	36.3%
		0.2%	0.2%	0.2%	0.3%	2.5%	4.3%	6.5%	10.8%
	F-test	4.6%	9.9%	5.5%	9.9%	22.2%	32.1%	60.8%	72.3%
200	0.25	1.068	(1.054)	1.078	(1.058)	1.073	(1.047)	1.005	(1.033)
		0.813	(1.109)	0.825	(1.114)	0.821	(1.103)	0.750	(1.089)
		29.8%	41.6%	28.9%	42.2%	29.4%	42.1%	27.5%	39.8%
		23.6%	33.6%	23.1%	33.9%	22.9%	33.9%	21.3%	31.2%
	0.5	1.361	(1.148)	1.482	(1.172)	1.498	(1.165)	1.408	(1.142)
		0.996	(1.209)	1.126	(1.235)	1.143	(1.229)	1.051	(1.205)
		38.6%	50.5%	42.6%	54.5%	43.6%	55.0%	40.2%	52.6%
		28.9%	39.3%	32.3%	43.4%	32.9%	43.5%	29.7%	40.6%
	1.0	1.305	(1.113)	1.911	(1.298)	2.027	(1.339)	1.937	(1.313)
		0.759	(1.172)	1.401	(1.367)	1.525	(1.412)	1.433	(1.386)
		36.3%	47.8%	54.0%	66.6%	57.8%	70.0%	55.1%	66.2%
		21.5%	31.6%	39.5%	49.7%	43.4%	53.2%	41.0%	50.4%
	2.0	0.144	(0.798)	1.837	(1.238)	2.418	(1.460)	2.465	(1.482)
		−0.715	(0.847)	1.076	(1.304)	1.691	(1.540)	1.746	(1.566)
		4.7%	8.6%	52.4%	64.4%	67.8%	77.3%	69.0%	77.7%
		1.0%	2.1%	30.1%	40.8%	47.1%	56.5%	48.2%	58.3%
	4.0	0.088	(0.571)	0.362	(0.703)	1.722	(1.136)	2.299	(1.305)
		−1.078	(0.612)	−0.792	(0.745)	0.647	(1.200)	1.261	(1.382)
		1.5%	3.2%	5.9%	10.6%	47.4%	61.4%	67.1%	76.8%
		0.1%	0.2%	0.4%	1.1%	18.7%	26.5%	35.2%	44.7%
	F-test	5.1%	10.3%	5.3%	11.2%	34.0%	46.6%	85.6%	91.3%

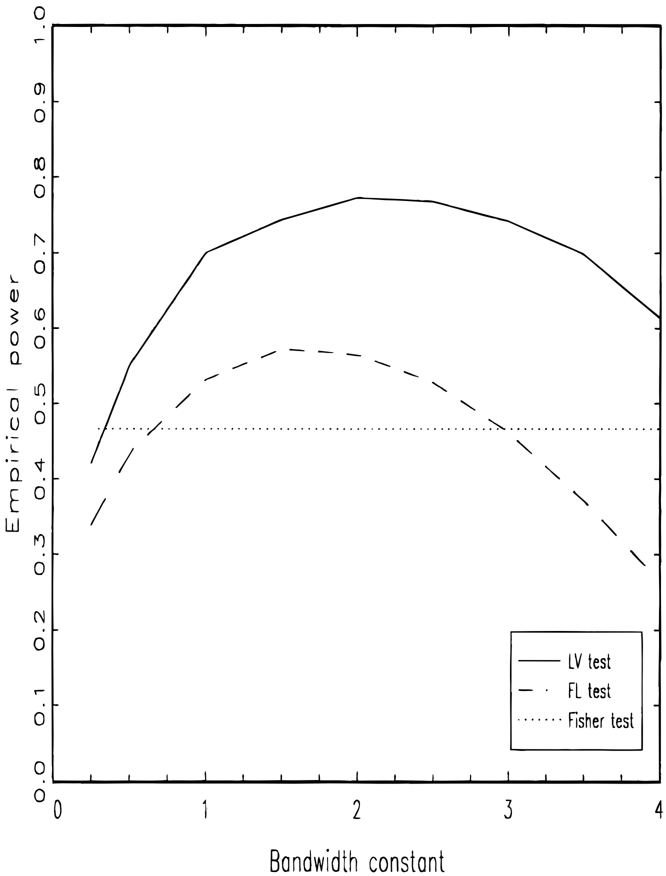

FIGURE 1. Empirical power of competing tests: DGP_6 with $n = 200$ at 10% level.

Our limited experiments suggest that the usual rule of thumb ($c = 1$) leads to an acceptable compromise between size and power.

Our procedure has been implemented within the XploRe environment by Pascal Lavergne and Gilles Teyssiere (see http://www.xplore-stat.de/help/lvtest.html).

4. PROOFS

Notations: In what follows, $f_i \equiv f_1(X_{1i})$, $f_{2i} \equiv f_2(X_{2i})$, $r_i \equiv r_1(X_{1i})$, $r_{2i} \equiv r_2(X_{2i})$, $u_i \equiv Y_i - r_i$, $u_{2i} \equiv Y_i - r_{2i}$, $d_i \equiv d(X_{2i})$, and Z_i stands for (Y_i, X_{2i}), $i = 0,1,\ldots,n$. Also $\mathbf{K} \equiv |K|$ and $\mathbf{L} \equiv |L|$ and i,j,k,l,i',j',k',l' refer to indices that are pairwise different unless stated otherwise. We let $\hat{f}_i = (n-1)^{-1}\sum_{k\neq i} L_{nik}$, and more generally for any index set I not containing i with cardinality $|I|$, $\hat{f}_i^I = (n-1-|I|)^{-1}\sum_{k\neq i,\, k\notin I} L_{nik}$.

4.1. Proof of Theorem 1

As $Y_i - Y_k = (u_i - u_k) + (r_i - r_k)$, and as $K(\cdot)$ is even, we have from (2.1)

$$V_n = \frac{1}{n^{(4)}} \sum_a (u_i - u_k)(u_j - u_l) L_{nik} L_{njl} K_{nij} + \frac{2}{n^{(4)}} \sum_a (u_i - u_k)(r_j - r_l) L_{nik} L_{njl} K_{nij} + \frac{1}{n^{(4)}} \sum_a (r_i - r_k)(r_j - r_l) L_{nik} L_{njl} K_{nij} = I_1 + 2I_2 + I_3,$$

where

$$\begin{aligned} I_1 &= \frac{n-2}{n-3} \frac{1}{n^{(2)}} \sum_a u_i u_j f_i f_j K_{nij} + \frac{2(n-2)}{n-3} \frac{1}{n^{(2)}} \sum_a u_i (\hat{f}_i^j - f_i) u_j f_j K_{nij} \\ &\quad + \frac{n-2}{n-3} \frac{1}{n^{(2)}} \sum_a u_i (\hat{f}_i^j - f_i) u_j (\hat{f}_j^i - f_j) K_{nij} - \frac{2}{n^{(3)}} \sum_a u_i f_i u_l L_{njl} K_{nij} \\ &\quad - \frac{2}{n^{(3)}} \sum_a u_i (\hat{f}_i^{j,l} - f_i) u_l L_{njl} K_{nij} + \frac{1}{n^{(4)}} \sum_a u_k u_l L_{nik} L_{njl} K_{nij} \\ &\quad - \frac{1}{n^{(4)}} \sum_a u_i u_j L_{nik} L_{njk} K_{nij} \\ &= \frac{n-2}{n-3} [V_{0n} + 2I_{1,1} + I_{1,2}] - 2I_{1,3} - 2I_{1,4} + I_{1,5} - I_{1,6}, \\ I_2 &= \frac{1}{n^{(3)}} \sum_a u_i f_i (r_j - r_l) L_{njl} K_{nij} + \frac{1}{n^{(3)}} \sum_a u_i (\hat{f}_i^{j,l} - f_i)(r_j - r_l) L_{njl} K_{nij} \\ &\quad - \frac{1}{n^{(4)}} \sum_a u_k (r_j - r_l) L_{nik} L_{njl} K_{nij} = I_{2,1} + I_{2,2} - I_{2,3}. \end{aligned}$$

Propositions 1–11 study each of the preceding terms. Collecting results, it follows that

$$\begin{aligned} nh^{p_2/2} V_{0n} &= A_n + \delta_n^2 n h^{p_2/2} \mu_n + \delta_n \sqrt{n} h^{p_2/2} O_p(1), \\ nh^{p_2/2} [I_1 - V_{0n}] &= \delta_n^2 n h^{p_2/2} o_p(1) + \delta_n \sqrt{n} h^{p_2/2} O_p(1) + o_p(1), \\ nh^{p_2/2} I_2 &= \delta_n^2 n h^{p_2/2} o_p(1) + \delta_n \sqrt{n} h^{p_2/2} o_p(1) + \delta_n n h^{p_2/2} g^{(m_1+q_1)} O_p(1) \\ &\quad + o_p(1), \\ nh^{p_2/2} I_3 &= o_p(1), \end{aligned}$$

where $A_n \xrightarrow{d} N(0, \omega^2)$. Therefore

$$nh^{p_2/2}[V_n - V_{0n}] = \delta_n^2 n h^{p_2/2} o_p(1) + \delta_n \sqrt{n} h^{p_2/2} O_p(1) + \delta_n n h^{p_2/2} g^{(m_1+q_1)} O_p(1) + o_p(1).$$

https://doi.org/10.1017/S0266466600164059 Published online by Cambridge University Press

In case (i), $nh^{p_2/2}[V_n - V_{0n}] = o_p(1)$ and $nh^{p_2/2}V_{0n} \xrightarrow{d} N(C\mu, \omega^2)$, as

$$\delta_n \sqrt{n} h^{p_2/2} = (\delta_n^2 nh^{p_2/2})^{1/2} h^{p_2/4} = o(1)$$

and

$$\delta_n nh^{p_2/2} g^{(m_1+q_1)} = (\delta_n^2 nh^{p_2/2})^{1/2} (nh^{p_2/2} g^{2(m_1+q_1)})^{1/2} = o(1).$$

In case (ii), $nh^{p_2/2}[V_n - V_{0n}] = o_p(\delta_n^2 nh^{p_2/2})$ and $nh^{p_2/2}V_{0n} = \delta_n^2 nh^{p_2/2}[\mu_n + o_p(1)]$, as

$$\delta_n \sqrt{n} h^{p_2/2} = (\delta_n^2 nh^{p_2/2}) \frac{h^{p_2/4}}{(\delta_n^2 nh^{p_2/2})^{1/2}} = o(\delta_n^2 nh^{p_2/2})$$

and

$$\delta_n nh^{p_2/2} g^{(m_1+q_1)} = (\delta_n^2 nh^{p_2/2}) \frac{(nh^{p_2/2} g^{2(m_1+q_1)})^{1/2}}{(\delta_n^2 nh^{p_2/2})^{1/2}} = o(\delta_n^2 nh^{p_2/2}). \qquad \blacksquare$$

4.1.1. Distribution of V_{0n}

PROPOSITION 1. $nh^{p_2/2}V_{0n} = A_n + \delta_n^2 nh^{p_2/2}\mu_n + \delta_n\sqrt{n}h^{p_2/2}B_n$, *where* $\mu_n \to \mu$, $A_n \xrightarrow{d} N(0,\omega^2)$, *and* $B_n \xrightarrow{d} 2N(0, \xi - \delta^2\mu^2)$, *with* $\delta = \lim_{n\to\infty}\delta_n$ *and* $\xi = E[u_1^2 d^2(X_2) f_1^4(X_1) f_2^2(X_2)]$.

Proof. Write $V_{0n} = U_{0n} + W_{0n} - \theta_n$, where $H_n(Z_i, Z_j) = u_i u_j f_i f_j K_{nij}$, $\theta_n = E[H_n(Z_1, Z_0)]$, $W_{0n} = (2/n)\sum_i E[H_n(Z_i, Z_0)|Z_i]$ and

$$U_{0n} = \binom{n}{2}^{-1} \sum_{i<j} \tilde{H}_n(Z_i, Z_j)$$
$$= \binom{n}{2}^{-1} \sum_{i<j} \{H_n(Z_i, Z_j) - E[H_n(Z_i, Z_0)|Z_i] - E[H_n(Z_0, Z_j)|Z_j] + \theta_n\}.$$

(i)

$$\theta_n = E[u_i f_i u_j f_j K_{nij}] = E[(u_{2i} + \delta_n d_i) f_i (u_{2j} + \delta_n d_j) f_j K_{nij}]$$
$$= \delta_n^2 E[d_i f_i d_j f_j K_{nij}] = \delta_n^2 \mu_n,$$

with $\mu_n \to \mu = E[d^2(X_2) f_1^2(X_1) f_2(X_2)]$, as $\delta_n d(X_2) f_1(X_1) f_2(X_2) \in \mathcal{U}^{p_2}$ and by Lemma 1.

(ii) Distribution of W_{0n}:

$$E[E^2(H_n(Z_i, Z_0)|Z_i)] = E[u_i^2 f_i^2 E^2(u_0 f_0 K_{ni0}|Z_i)]$$
$$= \delta_n^2 E[u_i^2 f_i^2 E^2(d_0 f_0 K_{ni0}|Z_i)] = \delta_n^2 \xi_n$$

with $\xi_n \to \xi = E[u_1^2 d^2(X_2) f_1^4(X_1) f_2^2(X_2)]$, as $\delta_n d(X_2) f_1(X_1) f_2(X_2) \in \mathcal{U}^{p_2}$ and by Lemma 1. Now $E|E[H_n(Z_i, Z_j)|Z_i]|^\nu = E|u_i^\nu f_i^\nu E^\nu[u_0 f_0 K_{ni0}|Z_i]| = O(1) = o(n^{\nu/2-1})$ for $2 < \nu \le 4$, as $E|Y^{2\nu}| < \infty$. Thus, by Theorem 7.1 of Hoeffding (1948),

$$\sqrt{n}[W_{0n} - 2\theta_n] \to 2\delta N(0, \xi - \delta^2\mu^2).$$

(iii) Distribution of U_{0n}: As $E[\tilde{H}_n(Z_i,Z_j)|Z_i]=0$, by Theorem 1 of Hall (1984a),

$$nh^{p_2/2}U_{0n} \xrightarrow{d} N(0,\omega^2) \quad \text{if} \quad \frac{E[\tilde{G}_n^2]+n^{-1}E[\tilde{H}_n^4]}{E^2[\tilde{H}_n^2]}=o(1),$$

where $\tilde{G}_n(Z_i,Z_j) = E[\tilde{H}_n(Z_i,Z_0)\tilde{H}_n(Z_j,Z_0)|Z_i,Z_j]$ and $\omega^2 = 2\lim_{n\to\infty} h^{p_2}E(\tilde{H}_n^2)$. By definition of $\tilde{H}_n(Z_i,Z_j)$, the preceding condition is equivalent to

$$\frac{E[G_n^2]+n^{-1}E[H_n^4]}{E^2[H_n^2]}=o(1), \tag{4.1}$$

where $G_n(Z_i,Z_j) = E[H_n(Z_i,Z_0)H_n(Z_j,Z_0)|Z_i,Z_j]$, and $\omega^2 = 2\lim_{n\to\infty} h^{p_2}E(H_n^2)$. Let $\sigma^2(X_2) \equiv E(u_1^2|X_2)$. As $\sigma^2(X_2)f_1^2(X_1)f_2(X_2) \in \mathcal{U}^{p_2}$, by Lemma 1,

$$h^{p_2}E[H_n^2(Z_i,Z_j)] = h^{p_2}E[\sigma^2(X_{2i})\sigma^2(X_{2j})f_i^2 f_j^2 K_{nij}^2] \to \omega^2/2,$$

where $\omega^2 = 2E[\sigma^4(X_2)f_1^4(X_1)f_2(X_2)]\int K^2(s)\,ds$. As $E(u_1^4|X_2)f_1^4(X_1)\times f_2(X_2) \in \mathcal{U}^{p_2}$, by Lemma 1,

$$E[H_n^4] = E[u_i^4 u_j^4 f_i^4 f_j^4 K_{nij}^4] = E[E(u_i^4|X_{2i})f_{1i}^4 E(u_j^4|X_{2j})f_{1j}^4 K_{nij}^4] = O(h^{-3p_2}).$$

As $G_n(Z_i,Z_j) = u_i f_i u_j f_j E[\sigma^2(X_{2,0})f_1^2(X_{1,0})K_{ni0}K_{nj0}|X_{2i},X_{2j}]$, we have by Lemma 1

$$\begin{aligned}
E[G_n^2] &= \int \sigma_{2i}^2 f_i^2 \sigma_{2j}^2 f_j^2 \left[\int \sigma_{2,0}^2 f_1^2(X_{1,0})K_{ni0}K_{nj0} f_2(X_{2,0})\,dX_{2,0}\right]^2 \\
&\quad \times f_{2i} f_{2j}\,dX_{2i}\,dX_{2j} \\
&= h^{-2p_2}\int \sigma_{2i}^2 f_i^2 \sigma_{2j}^2 f_j^2 \left[\int \sigma^2(X_{2i}-hs)f_1^2(X_{1i}-hs_1)K(s)\right. \\
&\qquad \left. \times K\left(s+\frac{X_{2j}-X_{2i}}{h}\right) f_2(X_{2i}-hs)\,ds\right]^2 \\
&\quad \times f_{2i} f_{2j}\,dX_{2i}\,dX_{2j} \\
&= h^{-p_2}\int \sigma_{2i}^2 f_i^2 \sigma^2(X_{2i}+ht)f_1^2(X_{1i}+ht_1) \\
&\quad \times \left[\int \sigma^2(X_{2i}-hs)f_1^2(X_{1i}-hs_1)K(s)K(s+t)f_2(X_{2i}-hs)\,ds\right]^2 \\
&\quad \times f_{2i} f_2(X_{2i}+ht)\,dX_{2i}\,dt \\
&= h^{-p_2}\int [\sigma^2(X_2)]^4 f_1^8(X_1)f_2^4(X_2)\,dX_2 \int \left[\int K(s)K(s+t)\,ds\right]^2 \\
&\quad \times dt + o(h^{-p_2}) \\
&= O(h^{-p_2}),
\end{aligned}$$

where s_1 and t_1 denote the first p_1 elements of s and t. Thus condition (4.1) holds as $h\to 0$ and $nh^{p_2}\to\infty$. Collecting results, Proposition 1 follows. ■

https://doi.org/10.1017/S0266466600164059 Published online by Cambridge University Press

4.1.2. U-Statistics Let $U_n = (1/n^{(m)})\sum_a H_n(Z_{i_1},\ldots,Z_{i_m})$ be an arbitrary U-statistic, where the Z_i's are i.i.d. but H_n is not necessarily symmetric. Then,

$$E(U_n^2) = \left(\frac{1}{n^{(m)}}\right)^2 \sum_{c=0}^{m} \frac{n^{(2m-c)}}{c!} \sum_{|\Delta_1|=c=|\Delta_2|}^{(c)} I(\Delta_1,\Delta_2)$$

$$= \sum_{c=0}^{m} O(n^{-c}) \sum_{|\Delta_1|=c=|\Delta_2|}^{(c)} I(\Delta_1,\Delta_2), \qquad \textbf{(4.2)}$$

where $\sum^{(c)}$ denotes summation over sets Δ_1 and Δ_2 of (ordered) positions of length c,

$$I(\Delta_1,\Delta_2) = E[H_n(Z_{i_1},\ldots,Z_{i_m})H_n(Z_{j_1},\ldots,Z_{j_m})],$$

and the i's in position Δ_1 coincide with the j's in position Δ_2 and are pairwise distinct otherwise. Note that this formula corrects equation (A.1) in Stute (1991). In what follows, we let $\xi_c = \sum^{(c)} I(\Delta_1,\Delta_2)$ and intensively use (4.2) to bound $E(U_n^2)$. Indeed, if Z_c denotes the vector of common Z_i's, we have by conditioning on Z_c

$$I^2(\Delta_1,\Delta_2) = E^2[E[H_n(Z_{i_1},\ldots,Z_{i_m})|Z_c]E[H_n(Z_{j_1},\ldots,Z_{j_m})|Z_c]]$$

$$\leq E[E^2[H_n(Z_{i_1},\ldots,Z_{i_m})|Z_c]]E[E^2[H_n(Z_{j_1},\ldots,Z_{j_m})|Z_c]]$$

by Cauchy–Schwartz inequality.

PROPOSITION 2. $nh^{p_2/2}I_{1,3} = \delta_n\sqrt{n}h^{p_2/2}O_p(1) + o_p(1)$.

Proof. $I_{1,3}$ is a U-statistic with kernel $H_n(Z_i,Z_j,Z_l) = u_i f_i u_l L_{njl} K_{nij}$. To use (4.2), we need to compute the corresponding ξ_c, $c = 0,1,2,3$.

(i) $\xi_2 = O(g^{-p_1})$. Indeed we have

$$E(H_n|Z_i,Z_j) = u_i f_i K_{nij} E(u_l L_{njl}|Z_j) = 0,$$

$$E(H_n|Z_i,Z_l) = u_i f_i u_l E(L_{njl}K_{nij}|Z_i,Z_l),$$

$$E(H_n|Z_j,Z_l) = u_l L_{njl} E(u_i f_i K_{nij}|Z_j) = \delta_n u_l L_{njl} E(d_i f_i K_{nij}|Z_j).$$

Then, using $\mathbf{K} \equiv |K|$ and $\mathbf{L} \equiv |L|$, we have, by successive applications of Lemma 1,

$$E[E^2(H_n|Z_i,Z_l)] = E[u_i^2 f_i^2 u_l^2 E(L_{njl}K_{nij}|Z_i,Z_l)E(L_{nj'l}K_{nij'}|Z_i,Z_l)]$$

$$= O(g^{-p_1})E[u_i^2 f_i^2 u_l^2 E(\mathbf{L}_{njl}\mathbf{K}_{nij}|Z_i,Z_l)E(\mathbf{K}_{nij'}|Z_i,Z_l)]$$

$$= O(g^{-p_1})E[u_i^2 f_i^2 u_l^2 \mathbf{L}_{njl}\mathbf{K}_{nij} f_2(X_{2i})] = O(g^{-p_1}),$$

$$E[E^2(H_n|Z_j,Z_l)] = \delta_n^2 E[u_l^2 L_{njl}^2 E^2(d_i f_i K_{nij}|Z_j)]$$

$$= \delta_n^2 E[u_l^2 L_{njl}^2 d_j^2 f_j^2 f_2^2(X_{2j})]$$

$$= \delta_n^2 O(g^{-p_1})E[u_l^2 \mathbf{L}_{njl} d_j^2 f_j^2 f_2^2(X_{2j})] = O(g^{-p_1}).$$

https://doi.org/10.1017/S0266466600164059 Published online by Cambridge University Press

(ii) $\xi_1 = O(\delta_n^2)$. Indeed, $E(H_n|Z_i) = E(H_n|Z_j) = 0$ and $E(H_n|Z_l) = \delta_n u_l E(d_i f_i L_{njl} K_{nij}|Z_l)$. Then

$$E[E^2(H_n|Z_l)] = \delta_n^2 E[u_l^2 E^2(d_i f_i L_{njl} K_{nij}|Z_l)]$$
$$= O(\delta_n^2) E[u_l^2 E^2(L_{njl} d_j f_j f_{2j}|Z_l)] = O(\delta_n^2).$$

(iii) $E[H_n] = 0$. Thus $\xi_0 = 0$.
(iv) $\xi_3 = O(g^{-p_1} h^{-p_2})$, as $E[H_n^2]$ equals

$$E[u_i^2 u_l^2 f_i^2 L_{njl}^2 K_{nij}^2] = O(g^{-p_1} h^{-p_2}) E[u_i^2 u_l^2 f_i^2 \mathbf{L}_{njl} \mathbf{K}_{nij}] = O(g^{-p_1} h^{-p_2}).$$

Collecting results, $E(nh^{p_2/2} I_{1,3})^2 = \delta_n^2 n h^{p_2} O(1) + h^{p_2}/g^{p_1} O(1) + O(ng^{p_1})^{-1}$. ■

PROPOSITION 3. $nh^{p_2/2} I_{1,5} = o_p(1)$.

Proof. $I_{1,5}$ is a U-statistic with kernel $H_n(Z_i, Z_j, Z_k, Z_l) = u_k u_l L_{nik} L_{njl} K_{nij}$.

(i) $\xi_3 = O(g^{-2p_1})$. Indeed we have

$$E(H_n|Z_i, Z_j, Z_k) = u_k L_{nik} K_{nij} E(u_l L_{njl}|Z_j) = 0,$$

$$E(H_n|Z_i, Z_j, Z_l) = u_l L_{njl} K_{nij} E(u_k L_{nik}|Z_i) = 0,$$

$$E(H_n|Z_i, Z_k, Z_l) = u_k u_l L_{nik} E(L_{njl} K_{nij}|Z_i, Z_l),$$

$$E(H_n|Z_j, Z_k, Z_l) = u_k u_l L_{njl} E(L_{nik} K_{nij}|Z_j, Z_k).$$

Then, we have, by successive applications of Lemma 1,

$$E[E^2(H_n|Z_i, Z_k, Z_l)] = E[u_k^2 u_l^2 L_{nik}^2 E(L_{njl} K_{nij}|Z_i, Z_l) E(L_{nj'l} K_{nij'}|Z_i, Z_l)]$$
$$= O(g^{-2p_1}) E[u_k^2 u_l^2 \mathbf{L}_{nik} E(\mathbf{L}_{njl} \mathbf{K}_{nij}|Z_i, Z_l) E(\mathbf{K}_{nij'}|Z_i, Z_l)]$$
$$= O(g^{-2p_1}) E[u_k^2 u_l^2 \mathbf{L}_{nik} \mathbf{L}_{njl} \mathbf{K}_{nij} f_{2i}] = O(g^{-2p_1}),$$

$$E[E^2(H_n|Z_j, Z_k, Z_l)] = E[u_k^2 u_l^2 L_{njl}^2 E^2(L_{nik} K_{nij}|Z_j, Z_k)] = O(g^{-2p_1}).$$

(ii) $\xi_2 = O(g^{-p_1})$. Indeed we have $E(H_n|Z_i, Z_j) = E(H_n|Z_i, Z_k) = E(H_n|Z_i, Z_l) = E(H_n|Z_j, Z_k) = E(H_n|Z_j, Z_l) = 0$ and $E(H_n|Z_k, Z_l) = u_k u_l E(L_{nik} L_{njl} K_{nij}|Z_k, Z_l)$, and

$$E[E^2(H_n|Z_k, Z_l)] = E[u_k^2 u_l^2 E(L_{nik} L_{njl} K_{nij}|Z_k, Z_l) E(L_{ni'k} L_{nj'l} K_{nij'}|Z_k, Z_l)]$$
$$= O(g^{-p_1}) E[u_k^2 u_l^2 E(\mathbf{L}_{nik} \mathbf{L}_{njl} \mathbf{K}_{nij}|Z_k, Z_l) E(\mathbf{L}_{nj'l} \mathbf{K}_{nij'}|Z_k, Z_l)]$$
$$= O(g^{-p_1}) E[u_k^2 u_l^2 \mathbf{L}_{nik} \mathbf{L}_{njl} \mathbf{K}_{nij} f_l^2] = O(g^{-p_1}).$$

(iii) $\xi_1 = 0$.
(iv) $E[H_n] = 0$. Thus $\xi_0 = 0$.
(v) $\xi_4 = O(g^{-2p_1} h^{-p_2})$, as $E[H_n^2]$ equals

$$E[u_k^2 u_l^2 L_{nik}^2 L_{njl}^2 K_{nij}^2] = O(g^{-2p_1} h^{-p_2}) E[u_k^2 u_l^2 \mathbf{L}_{nik} \mathbf{L}_{njl} \mathbf{K}_{nij}] = O(g^{-2p_1} h^{-p_2}).$$

Collecting results $E(nh^{p_2/2} I_{1,5})^2 = (h^{p_2}/g^{p_1})[O(1) + O(ng^{p_1})^{-1}] + O(ng^{p_1})^{-2}$. ■

https://doi.org/10.1017/S0266466600164059 Published online by Cambridge University Press

PROPOSITION 4. $nj^{p_2/2} I_{1,6} = \delta_n^2 n h^{p_2/2} o_p(1) + o_p(1)$.

Proof. $(n-3)I_{1,6}$ is a U-statistic with kernel $H_n(Z_i, Z_j, Z_k) = u_i u_j L_{nik} \times L_{njk} K_{nij}$. Using a reasoning similar to the one followed in Proposition 3, it is not difficult to show that this U-statistic is such that $\xi_3 = O(g^{-2p_1} h^{-2p_2})$, $\xi_2 = O(g^{-2p_1} h^{-p_2}) + O(g^{-p_1} h^{-2p_2})$, $\xi_1 = O(h^{-2p_2})$, and $E[H_n] = O(\delta_n^2 g^{-p_1})$. Hence, $E(nh^{p_2/2} I_{1,6})^2 = O(\delta_n^2 n h^{p_2/2})^2 (ng^{p_1})^{-2} + O(nh^{p_2})^{-1} + O(n^2 g^{p_1} h^{p_2})^{-1} + O(ng^{p_1})^{-2} + O(ng^{p_1})^{-2}(nh^{p_2})^{-1}$. ■

PROPOSITION 5. $nh^{p_2/2} I_{2,1} = \delta_n \sqrt{n} h^{p_2/2} o_p(1) + \delta_n n h^{p_2/2} g^{(m_1+q_1)} O_p(1) + o_p(1)$.

Proof. $I_{2,1}$ is a U-statistic with kernel $H_n(Z_i, Z_j, Z_l) = u_i f_i (r_j - r_l) L_{njl} K_{nij}$.

(i) $\xi_2 = o(h^{-p_2}) + o(g^{-p_1})$. Indeed we have

$$E(H_n|Z_i, Z_j) = u_i f_i K_{nij} E((r_j - r_l) L_{njl} | Z_j),$$

$$E(H_n|Z_i, Z_l) = u_i f_i E((r_j - r_l) L_{njl} K_{nij} | Z_i, Z_l),$$

$$E(H_n|Z_j, Z_l) = \delta_n (r_j - r_l) L_{njl} E(d_i f_i K_{nij} | Z_j).$$

First, we use the fact that $E[(r_j - r_l) L_{njl} | Z_j] = O(g^{(m_1+q_1)}) = o(1)$ uniformly in Z_j by a standard Taylor expansion argument, so that

$$E[E^2(H_n|Z_i, Z_j)] = E[u_i^2 f_i^2 K_{nij}^2 E^2((r_j - r_l) L_{njl} | Z_j)]$$
$$= O(h^{-p_2}) E[u_i^2 f_i^2 \mathbf{K}_{nij} E^2((r_j - r_l) L_{njl} | Z_j)] = o(h^{-p_2}).$$

Now, by successive applications of Lemma 1,

$$E[E^2(H_n|Z_i, Z_l)] = E[u_i^2 f_i^2 E((r_j - r_l) L_{njl} K_{nij} | Z_i, Z_l) \times E((r_{j'} - r_l) L_{nj'l} K_{nij'} | Z_i, Z_l)]$$
$$= O(g^{-p_1}) E[u_i^2 f_i^2 E(|r_j - r_l| \mathbf{L}_{njl} \mathbf{K}_{nij} | Z_i, Z_l) \times E(|r_{j'} - r_l| \mathbf{K}_{nij'} | Z_i, Z_l)]$$
$$= o(g^{-p_1}),$$

$$E[E^2(H_n|Z_j, Z_l)] = \delta_n^2 E[(r_j - r_l)^2 L_{njl}^2 E^2(d_i f_i K_{nij} | Z_j)]$$
$$= O(\delta_n^2 g^{-p_1}) E[(r_j - r_l)^2 \mathbf{L}_{njl} d_j^2 f_j^2 f_{2j}^2] = o(g^{-p_1}).$$

(ii) $\xi_1 = O(g^{2(m_1+q_1)}) + o(\delta_n^2)$. Indeed, by a similar reasoning to (i),

$$E[E^2(H_n|Z_i)] = E[u_i^2 f_i^2 E^2(K_{nij} E((r_j - r_l) L_{njl} | Z_j) | Z_i)]$$
$$= O(g^{2(m_1+q_1)}) E[u_i^2 f_i^2 E^2(\mathbf{K}_{nij})] = O(g^{2(m_1+q_1)}),$$

$$E[E^2(H_n|Z_j)] = E[E^2(u_i f_i (r_j - r_l) L_{njl} K_{nij} | Z_j)] = O(g^{2(m_1+q_1)}),$$

$$E[E^2(H_n|Z_l)] = E[E^2((r_j - r_l) L_{njl} E(u_i f_i K_{nij} | Z_j) | Z_l)]$$
$$= \delta_n^2 E[E^2((r_j - r_l) L_{njl} d_j f_j f_{2j} | Z_l)] = o(\delta_n^2).$$

https://doi.org/10.1017/S0266466600164059 Published online by Cambridge University Press

(iii)

$$E(H_n) = E[u_i f_i (r_j - r_l) L_{njl} K_{nij}]$$
$$\approx \delta_n E[(r_j - r_l) L_{njl} d_j f_j f_{2j}]$$
$$\approx O(\delta_n g^{(m_1+q_1)}) E[d_j f_j f_{2j}] = O(\delta_n g^{(m_1+q_1)}).$$

(iv) $\xi_3 = o(g^{-p_1} h^{-p_2})$, as $E[H_n^2] = E[u_i^2 f_i^2 (r_j - r_l)^2 L_{njl}^2 K_{nij}^2] = o(g^{-p_1} h^{-p_2})$.

Collecting results, $E(nh^{p_2/2} I_{2,1})^2 = \delta_n^2 n^2 h^{p_2} O(g^{2(m_1+q_1)}) + nh^{p_2} O(g^{2(m_1+q_1)}) + o(\delta_n^2 n h^{p_2}) + o(1) + o(h^{p_2}/g^{p_1}) + o(ng^{p_1})^{-1}$. ■

PROPOSITION 6. $nh^{p_2/2} I_{2,3} = o_p(1)$.

Proof. $I_{2,3}$ is a U-statistic with kernel $H_n(Z_i, Z_j, Z_k, Z_l) = u_k (r_j - r_l) L_{nik} \times L_{njl} K_{nij}$. Using a reasoning similar to the one followed in Proposition 5, it is not difficult to show that this U-statistic is such that $\xi_4 = o(h^{-p_2} g^{-2p_1})$, $\xi_3 = o(g^{-p_1} h^{-p_2}) + o(g^{-2p_1})$, $\xi_2 = o(g^{-p_1})$, $\xi_1 = O(g^{2(m_1+q_1)})$, and $E[H_n] = 0$. Hence, $E(nh^{p_2/2} I_{2,3})^2 = nh^{p_2} O(g^{2(m_1+q_1)}) + o(h^{p_2}/g^{p_1})[1 + (ng^{p_1})^{-1}] + o(ng^{p_1})^{-1} + o(ng^{p_1})^{-2}$. ■

PROPOSITION 7. $nh^{p_2/2} I_3 = nh^{p_2/2} O_p(g^{2(m_1+q_1)}) + o_p(1)$.

Proof. I_3 is a U-statistic with kernel $H_n(Z_i, Z_j, Z_k, Z_l) = (r_i - r_k)(r_j - r_l) \times L_{nik} L_{njl} K_{nij}$. Similarly to the proof of Proposition 5 for $I_{2,3}$, we can show that $\xi_4 = o(h^{-p_2} g^{-2p_1})$, $\xi_3 = o(h^{-p_2} g^{-p_1}) + o(g^{-2p_1})$, $\xi_2 = o(g^{-p_1})$, $\xi_1 = o(g^{2(m_1+q_1)})$. On the other hand,

$$E[H_n] = E[(r_i - r_k)(r_j - r_l) L_{nik} L_{njl} K_{nij}] = O(g^{2(m_1+q_1)}),$$

so that $E(nh^{p_2/2} I_3)^2 = n^2 h^{p_2} O(g^{4(m_1+q_1)}) + o(1)$. ■

4.1.3. The Remaining Terms

PROPOSITION 8. $nh^{p_2/2} I_{1,1} = \delta_n^2 nh^{p_2/2} o_p(1) + \delta_n \sqrt{n} h^{p_2/2} o_p(1) + o_p(1)$.

Proof. We denote $(\hat{f}_i^j - f_i)$ by Δf_i^j. We have $I_{1,1} = (1/n^{(2)}) \sum_a u_i \Delta f_i^j u_j f_j K_{nij}$ so that

$$E(I_{1,1}^2) = \left(\frac{1}{n^{(2)}}\right)^2 \left[\sum_a u_i \Delta f_i^j u_j f_j K_{nij}\right]\left[\sum_a u_{i'} \Delta f_{i'}^{j'} u_{j'} f_{j'} K_{ni'j'}\right],$$

where the first (respectively, the second) sum is taken over all arrangements of different indices i and j (respectively, different indices i' and j').

Let $\overline{X}_1$ be the σ-algebra generated by all the X_{1i}'s and λ_n be $E[\Delta^2 f_i^j | Z_i, Z_j, Z_{i'}, Z_{j'}]$, which is $o(1)$ uniformly by Lemma 2. We consider three situations.

https://doi.org/10.1017/S0266466600164059 Published online by Cambridge University Press

(i) All indices are different: $n^{(4)}$ terms.

$$E[u_i \Delta f_i^j u_j f_j K_{nij} u_{i'} \Delta f_{i'}^{j'} u_{j'} f_{j'} K_{ni'j'}]$$

$$= \delta_n^4 E[\Delta f_i^j f_j \Delta f_{i'}^{j'} f_{j'} E(d_i d_j d_{i'} d_{j'} K_{nij} K_{ni'j'} | \bar{X}_1)]$$

$$= \delta_n^4 E[f_j f_{j'} d_i d_j d_{i'} d_{j'} K_{nij} K_{ni'j'} E(\Delta f_i^j \Delta f_{i'}^{j'} | Z_i, Z_j, Z_{i'}, Z_{j'})]$$

$$= \delta_n^4 \lambda_n E| f_j f_{j'} d_i d_j d_{i'} d_{j'} K_{nij} K_{ni'j'}| = O(\delta_n^4 \lambda_n).$$

(ii) One index is common to $\{i, j\}$ and $\{i', j'\}$: $4n^{(3)}$ terms.

$$(i' = i) \qquad E[u_i^2 \Delta f_i^j u_j f_j K_{nij} \Delta f_i^{j'} u_{j'} f_{j'} K_{nij'}]$$

$$= \delta_n^2 E[\Delta f_i^j f_j \Delta f_i^{j'} f_{j'} E(u_i^2 d_j d_{j'} K_{nij} K_{nij'} | \bar{X}_1)]$$

$$= \delta_n^2 \lambda_n E| f_j f_{j'} u_i^2 d_j d_{j'} K_{nij} K_{nij'}| = O(\delta_n^2 \lambda_n),$$

$$(j' = j) \qquad E[u_i \Delta f_i^j u_j^2 f_j^2 K_{nij} u_{i'} \Delta f_{i'}^j K_{ni'j}]$$

$$= \delta_n^2 E[\Delta f_i^j f_j^2 \Delta f_{i'}^j E(d_i u_j^2 d_{i'} K_{nij} K_{ni'j} | \bar{X}_1)]$$

$$= \delta_n^2 \lambda_n E| f_j^2 d_i u_j^2 d_{i'} K_{nij} K_{ni'j}| = O(\delta_n^2 \lambda_n),$$

$$(i' = j) \qquad E[u_i \Delta f_i^j u_j^2 f_j K_{nij} \Delta f_j^{j'} u_{j'} f_{j'} K_{njj'}]$$

$$= \delta_n^2 E[\Delta f_i^j f_j \Delta f_j^{j'} f_{j'} E(d_i u_j^2 d_{j'} K_{nij} K_{njj'} | \bar{X}_1)]$$

$$= \delta_n^2 \lambda_n E| f_j f_{j'} d_i u_j^2 d_{j'} K_{nij} K_{njj'}| = O(\delta_n^2 \lambda_n).$$

The case $j' = i$ is similar to $i' = j$.

(iii) Two indices in common to $\{i, j\}$ and $\{i', j'\}$: $2n^{(2)}$ terms. We have

$$E[u_i^2 u_j^2 (\Delta f_i^j)^2 f_j^2 K_{nij}^2] = O(\lambda_n / h^{p_2}) \quad \text{and}$$

$$E[u_i^2 u_j^2 \Delta f_i^j \Delta f_j^i f_i f_j K_{nij}^2] = O(\lambda_n / h^{p_2}).$$

Therefore,

$$E(nh^{p_2/2} I_{1,1})^2 = \delta_n^4 n^2 h^{p_2} O(\lambda_n) + \delta_n^2 n h^{p_2} O(\lambda_n) + O(\lambda_n).$$

The proposition then follows from $\lambda_n = o(1)$ uniformly (see Lemma 2). ∎

PROPOSITION 9. $nh^{p_2/2} I_{1,2} = \delta_n^2 n h^{p_2/2} o_p(1) + \delta_n \sqrt{n} h^{p_2/2} o_p(1) + o_p(1)$.

Proof. The proof is very similar to the proof of Proposition 8 for $I_{1,1}$ and is not reported.

PROPOSITION 10. $nh^{p_2/2}I_{1,4} = \delta_n^2 nh^{p_2/2}o_p(1) + \delta_n\sqrt{n}h^{p_2/2}o_p(1) + o_p(1)$.

Proof. We denote $(\hat{f}_i^{j,l} - f_i)$ by $\Delta f_i^{j,l}$. We have $I_{1,4} = (1/n^{(3)}) \times \sum_a u_i \Delta f_i^{j,l} u_l L_{njl} K_{nij}$ and

$$E(I_{1,4}^2) = \left(\frac{1}{n^{(3)}}\right)^2 \left[\sum_a u_i \Delta f_i^{j,l} u_l L_{njl} K_{nij}\right]\left[\sum_a u_{i'} \Delta f_{i'}^{j',l'} u_{l'} L_{nj'l'} K_{ni'j'}\right],$$

where the first (respectively, the second) sum is taken over all arrangements of pairwise different indices i, j, and l (respectively, pairwise different indices i', j', and l'). Let $\overline{X}_1$ be the σ-algebra generated by all the X_{1i}'s and where λ_n is $E[\Delta^2 f_i^{j,l}|Z_i, Z_j, Z_l, Z_{i'}, Z_{j'}, Z_{l'}]$, which is $o(1)$ uniformly by Lemma 2. We consider four situations and employ a similar strategy as in the proof of Proposition 8.

(i) All indices are different: $n^{(6)}$ terms.

$$E[u_i \Delta f_i^{j,l} u_l L_{njl} K_{nij} u_{i'} \Delta f_{i'}^{j',l'} u_{l'} L_{nj'l'} K_{ni'j'}]$$
$$= \delta_n^4 E[L_{njl} L_{nj'l'} d_i d_l d_{i'} d_{l'} K_{nij} K_{ni'j'} E(\Delta f_i^{j,l} \Delta f_{i'}^{j',l'} | Z_i, Z_j, Z_l, Z_{i'}, Z_{j'}, Z_{l'})]$$
$$= O(\delta_n^4 \lambda_n).$$

(ii) One index is common to $\{i,j,l\}$ and $\{i',j',l'\}$: $9n^{(5)}$ terms. For case $(i' = i)$, $E[u_i^2 \Delta f_i^{j,l} u_l L_{njl} K_{nij} \Delta f_i^{j',l'} u_{l'} L_{nj'l'} K_{nij'}] = \delta_n^2 E[\Delta f_i^{j,l} L_{njl} \Delta f_i^{j',l'} L_{nj'l'} u_i^2 d_l d_{l'} \times K_{nij} K_{nij'}] = O(\delta_n^2 \lambda_n)$. Similar computations for the other cases lead to the same result.

(iii) Two indices are common to $\{i,j,l\}$ and $\{i',j',l'\}$: $18n^{(4)}$ terms.

$(i = i' \quad \text{and} \quad j = j')$

$$E[u_i^2 \Delta f_i^{j,l} u_l L_{njl} \Delta f_i^{j,l'} u_{l'} L_{njl'} K_{nij}^2] = \delta_n^2 E[\Delta f_i^{j,l} L_{njl} \Delta f_i^{j,l'} L_{njl'} u_i^2 d_l d_{l'} K_{nij}^2]$$
$$= O(\delta_n^2 \lambda_n / h^{p_2}) = O(\lambda_n / h^{p_2}),$$

$(i = i' \quad \text{and} \quad l = l')$

$$E[u_i^2 \Delta f_i^{j,l} u_l^2 L_{njl} K_{nij} \Delta f_i^{j',l} L_{nj'l} K_{nij'}] = (\lambda_n / g^{p_1}) E[u_i^2 u_l^2 L_{njl} K_{nij} K_{nij'}]$$
$$= O(\lambda_n / g^{p_1}),$$

In other cases, similar computations lead to either $O(\lambda_n/g^{p_1})$ or $O(\lambda_n/h^{p_2})$.

(iv) Three indices are common to $\{i,j,l\}$ and $\{i',j',l'\}$: $6n^{(3)}$ terms. For instance, if $(i = i', j = j', \text{and } l = l')$, $E[u_i^2 (\Delta f_i^{j,l})^2 u_l^2 L_{njl}^2 K_{nij}^2] = O(\lambda_n/(g^{p_1}h^{p_2}))$. In the remaining cases, the corresponding expectations are all $O(\lambda_n/(g^{p_1}h^{p_2}))$.

Therefore from (4.2) we get

$$E[nh^{p_2/2}I_{1,4}]^2 = \delta_n^4 n^2 h^{p_2} O(\lambda_n) + \delta_n^2 nh^{p_2} O(\lambda_n) + h^{p_2}/g^{p_1} O(\lambda_n) + O(\lambda_n)$$
$$+ (ng^{p_1})^{-1} O(\lambda_n).$$

The proposition then follows from $\lambda_n = o(1)$ uniformly (see Lemma 2). ■

https://doi.org/10.1017/S0266466600164059 Published online by Cambridge University Press

PROPOSITION 11. $nh^{p_2/2}I_{2,2} = \delta_n^2 nh^{p_2/2}o_p(1) + \delta_n\sqrt{n}h^{p_2/2}o_p(1) + o_p(1)$.

Proof. The proof is very similar to the proof of Proposition 10 for $I_{1,4}$ and is not reported.

4.1.4. Lemmas

LEMMA 1. *For any function $l(\cdot) \in \mathcal{U}^p$ and any integrable kernel $K(\cdot)$,*

$$\sup_{x\in\mathbb{R}^p}\left|\int l(X)\,\frac{1}{h^p}\,K\left(\frac{x-X}{h}\right)dX - l(x)\int K(u)\,du\right| \to 0.$$

Proof. This result comes from the well-known Bochner lemma.

LEMMA 2. *If the density $f_1(X_1) \in \mathcal{U}^{p_1}$ and $ng^{p_1} \to \infty$, $E[\Delta^2 f_i^j|Z_i, Z_j, Z_{i'}, Z_{j'}] = o(1)$ and $E[\Delta^2 f_i^{j,l}|Z_i, Z_j, Z_l, Z_{i'}, Z_{j'}, Z_{l'}] = o(1)$ uniformly in the indices.*

Proof. From the definition of Δf_i^j,

$$E[\Delta^2 f_i^j|Z_i,Z_j,Z_{i'},Z_{j'}] = E[(\hat{f}_i^j - E(\hat{f}_i^j|Z_i,Z_j,Z_{i'},Z_{j'}))^2|Z_i,Z_j,Z_{i'},Z_{j'}] + [E(\hat{f}_i^j|Z_i,Z_j,Z_{i'},Z_{j'}) - f_i]^2.$$

Because $\hat{f}_i^j - E(\hat{f}_i^j|Z_i,Z_j,Z_{i'},Z_{j'}) = (n-2)^{-1}\sum_{k\notin\{i,j,i',j'\}}(L_{nik} - E(L_{nik}|Z_i))$, whose summands are, conditional on Z_i, independent with zero mean,

$$E[(\hat{f}_i^j - E(\hat{f}_i^j|Z_i,Z_j,Z_{i'},Z_{j'}))^2|Z_i,Z_j,Z_{i'},Z_{j'}] \le (n-2)^{-2}\sum_{k\notin\{i,j,i',j'\}} E(L_{nik}^2|Z_i) = O(n^{-1}g^{-p_1}).$$

As

$$E(\hat{f}_i^j|Z_i,Z_j,Z_{i'},Z_{j'}) = (n-2)^{-1}[L_{nii'} + L_{nij'} + (n-4)E(L_{nik}|Z_i)],$$

$$[E(\hat{f}_i^j|Z_i,Z_j,Z_{i'},Z_{j'}) - f_i]^2 = \left[\frac{1}{n-2}(L_{nii'} + L_{nij'} - f_i) + \frac{n-4}{n-2}E(L_{nik} - f_i|Z_i)\right]^2 \le [O(n^{-1}g^{-p_1}) + O(n^{-1}) + o(1)]^2 = o(1).$$

The proof for the second part is similar and is therefore not reported. ∎

https://doi.org/10.1017/S0266466600164059 Published online by Cambridge University Press

4.2. Proof of Corollary 1

Let $u_i = Y_i - c$. As $Y_i - Y_k = (u_i - u_k)$, and as $K(\cdot)$ is even, we have

$$V_n^* = \frac{1}{n^{(2)}} \sum_a u_i u_j K_{nij} - \frac{2}{n^{(3)}} \sum_a u_i u_l K_{nij} + \frac{1}{n^{(4)}} \sum_a u_k u_l K_{nij}$$

$$= V_{0n}^* - 2V_{1n}^* + V_{2n}^*.$$

Each of these terms can be studied using the same techniques as in the proof of Theorem 1, so that we have

$$nh^{p_2/2} V_{0n}^* = A_n^* + \delta_n^2 nh^{p_2/2} \mu_n^* + \delta_n \sqrt{n} h^{p_2/2} B_n^*,$$

where $A_n^* \xrightarrow{d} N(0, \omega^{*2})$, $\mu_n^* \to \mu^*$, $B_n^* \xrightarrow{d} 2N(0, \xi^* - \delta^2 \mu^{*2})$, with $\delta = \lim_{n\to\infty} \delta_n$ and $\xi^* = E[(Y-c)^2 d^2(X_2) f^2(X_2)]$, $nh^{p_2/2} V_{1n}^* = \delta_n \sqrt{n} h^{p_2/2} O_p(1) + o_p(1)$ and $nh^{p_2/2} V_{2n}^* = o_p(1)$. Collecting results, it follows that

$$nh^{p_2/2} V_n^* = A_n^* + \delta_n^2 nh^{p_2/2} \mu_n^* + \delta_n \sqrt{n} h^{p_2/2} O_p(1) + o_p(1),$$

where $A_n^* \xrightarrow{d} N(0, \omega^{*2})$ and $\mu_n^* \to \mu^*$. The end of the proof is similar to that of Theorem 1. ■

NOTES

1. Another approach uses empirical processes based on residuals of the parametric model (see Bierens, 1982, 1990; Diebolt, 1995; Stute, 1997). Extension of the empirical process approach to the comparison of two nonparametric models would inevitably also rely on smoothing techniques.
2. Unpublished related work includes Aït-Sahalia, Bickel, and Stoker (1994), Gozalo (1995), and Delgado and Gonzalez-Manteiga (1998).
3. Fan and Li (1996a) also impose that the two kernels $K(\cdot)$ and $L(\cdot)$ are product kernels with the same univariate kernel.
4. As we consider local alternatives and a finer decomposition of V_n, this prevents us from using Fan and Li's (1996) proofs.
5. One could also consider a more general form of kernel estimators as in Robinson (1983).
6. It is always possible to modify our test statistic and make it consistent against some chosen local $\sqrt{n}$-alternatives. This is done by adding to our test statistic a suitable *M*-test statistic based on the estimated residuals and by deriving the resulting limiting distribution.
7. In our setup, we have $m_1 = 2$ and $q_1 = 1$, so that our bandwidths satisfy the conditions of Theorem 1.

REFERENCES

Aït-Sahalia, Y., P. Bickel, & T.M. Stoker (1994) Goodness-of-Fit Tests for Regression Using Kernel Methods. Manuscript, University of Chicago.

Barry, D. & J.A. Hartigan (1990) An omnibus test for departures from constant mean. *Annals of Statistics* 18, 1340–1357.

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.

Buckley, M.J. (1991) Detecting a smooth signal: Optimality of cusum based procedures. *Biometrika* 78, 253–262.

https://doi.org/10.1017/S0266466600164059 Published online by Cambridge University Press

Buckley, M.J. & G.K. Eagleson (1988) An approximation to the distribution of quadratic forms in normal random variables. *Australian Journal of Statistics* 30A, 149–163.

Chen, J.C. (1994) Testing goodness-of-fit of polynomial models via spline smoothing techniques. *Statistics and Probability Letters* 19, 65–76.

Cleveland, W.S. & S.J. Devlin (1988) Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association* 83 (403), 596–610.

Delgado, M.A. & W. Gonzalez-Manteiga (1998) Significance Testing in Nonparametric Regression Based on the Bootstrap. Manuscript, Universidad Carlos III de Madrid.

Diebolt, J. (1995) A nonparametric test for the regression function: Asymptotic theory. *Journal of Statistical Planning and Inference* 44, 1–17.

Epanechnikov, V. (1969) Nonparametric estimates of a multivariate probability density. *Theory of Probability and Its Applications* 14, 153–158.

Eubank, R.L. & J.D. Hart (1992) Testing goodness-of-fit in regression via order selection criteria. *Annals of Statistics* 20 (3), 1412–1425.

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

Eubank, R.L. & V.N. Lariccia (1993) Testing for no effect in nonparametric regression. *Journal of Statistical Planning and Inference* 36, 1–14.

Eubank, R.L. & C.H. Spiegelman (1990) Testing the goodness of fit of a linear model via nonparametric regression techniques. *Journal of the American Statistical Association* 85 (410), 387–392.

Fan, Y. & Q. Li (1996a) Consistent model specification tests: Omitted variables and semiparametric functional forms. *Econometrica* 64 (4), 865–890.

Fan, Y. & Q. Li (1996b) Consistent Model Specification Tests: Nonparametric versus Bierens' Tests. Manuscript, University of Windsor.

Gozalo, P.L. (1993) A consistent model specification test for nonparametric estimation of regression functions models. *Econometric Theory* 9, 451–477.

Gozalo, P.L. (1995) Nonparametric Specification Testing with $\sqrt{n}$-Local Power and Bootstrap Critical Values. Manuscript, Brown University.

Guerre, E. & P. Lavergne (1999) Minimax Rates for Nonparametric Specification Testing in Regression Models. Manuscript, Université Paris VI.

Hall, P. (1983) Chi-squared approximations to the distribution of a sum of independent random variables. *Annals of Probability* 11, 1028–1036.

Hall, P. (1984a) Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of Multivariate Analysis* 14, 1–16.

Hall, P. (1984b) Integrated square error properties of kernel estimators of regression functions. *Annals of Statistics* 12 (1), 163–179.

Härdle, W. (1991) *Smoothing Techniques with Implementation in S.* New York: Springer-Verlag.

Härdle, W. & E. Mammen (1993) Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21 (4), 1926–1947.

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

Heffernan, P.M. (1997) Unbiased estimation of central moments by using U-statistics. *Journal of the Royal Statistical Society, Series B* 59 (4), 861–863.

Hoeffding, W. (1948) A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics* 15, 293–325.

Hong, Y. (1993) Consistent Specification Testing using Optimal Nonparametric Kernel Estimation. Manuscript, Cornell University.

Hong, Y. & H. White (1995) Consistent specification testing via nonparametric series regression. *Econometrica* 63 (5), 1133–1159.

Horowitz, J.L. & W. Härdle (1994) Testing a parametric model against a semiparametric alternative. *Econometric Theory* 10, 821–848.

Kuchibhatla, M. & J.D. Hart (1996) Smoothing-based lack-of-fit tests: Variations on a theme. *Journal of Nonparametric Statistics* 7, 1–22.

Lavergne, P. (2000) An equality test across nonparametric regions. *Journal of Econometrics*, forthcoming.

Lavergne, P. & Q.H. Vuong (1996) Nonparametric selection of regressors: The nonnested case. *Econometrica* 64 (1), 207–219.

Robinson, P.M. (1983) Nonparametric estimators for time series. *Journal of Time Series Analysis* 4, 185–207.

Samarov, A.M. (1993) Exploring regression structure using nonparametric functional estimation. *Journal of the American Statistical Association* 88 (423), 836–847.

Stute, W. (1991) Conditional *U*-statistics. *Annals of Probability* 19 (2), 812–825.

Stute, W. (1997) Nonparametric model checks for regression. *Annals of Statistics* 25, 613–641.

Yatchew, A.J. (1992) Nonparametric regression tests based on least squares. *Econometric Theory* 8, 435–451.

Zheng, J.X. (1996) A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics* 75 (2) 263–289.