A CONSISTENT SPECIFICATION TEST FOR SEMIPARAMETRIC MODELS

Juan M. Rodríguez-Póo
Universidad de Cantabria
Departamento de Economía

Stefan Sperlich
Universidad Carlos III de Madrid
Departamento de Estadística y Econometría

Philippe Vieu
Université Paul Sabatier
Laboratoire de Statistique et Probabilités

January 29, 2002

Abstract

This paper introduces a consistent specification test for semiparametric conditional densities. This test is motivated by the fact that many important econometric models need to be estimated through maximum likelihood type procedures, e.g. semiparametric limited dependent variable models. This specification is also important for prediction purposes. Our statistic combines the methodology of goodness of fit tests and nonparametric methods. It is shown to be asymptotically distributed standard normal under the null hypothesis if the semiparametric model is correctly specified. Further, the test is shown to have power against $n^{-1/2}h^{-d/2}$ local alternatives to the null hypothesis. We discuss practical issues for the application statistics and illustrate in an intensive monte carlo study both the feasibility and the performance of the procedure. [1]

Keywords: Consistent specification test, semiparametric models.

[1]This research was financially supported by the Dirección General de Enseñanza Superior del Ministerio de Educación y Ciencia under research grant PB98-0025, the Dirección General de Investigación del Ministerio de Ciencia y Tecnología under research grant BEC2001-1121 and the Danish Social Science Research fund.

1 Introduction

In the specification and estimation of structural econometric models, it is very difficult to find a situation where Economic Analysis and other information from outside data enables us to fully specify a model without taking the risk of a serious misspecification. On the other hand, some parametric modeling is certainly wanted by the empirical researcher. A reasonable way to overcome this trade-off is to develop a suitable semiparametric model. One defers the most restrictive assumptions about the structural model, i.e. where information is available, to the parametric model, whereas for the other part, i.e. where no information is available, weaker assumptions are imposed and a nonparametric structure is considered. Note that the available information not necessarily refers to the functional form as e.g. linearity, it can also be information about separability of input groups.

Although under this approach the risk of misspecification is considerably reduced, there are still some contexts where this problem is still not negligible. Consider the following relationship between some latent variable y^* and a set of explanatory variables x,

$$\begin{aligned} y^* &= m(x;\theta_0) + u, \\ y &= \psi(y^*) \end{aligned}$$

the observable outcome y is related to the unobservable endogenous variable y^* through the mapping $\psi(\cdot)$, and θ_0 are the parameters of interest. This type of representation is well known in microeconomic analysis and depending on the form of $\psi(\cdot)$, it nests very different models such as binary choice, limited dependent variable, duration models and many others. If $m(\cdot;\theta_0)$ is a partial linear model and $\psi(\cdot)$ is the identity, then Robinson (1988) proposes a root-n consistent estimator of the parameters of interest. Furthermore, if both the observability rule and the conditional distribution function of the latent variable are known, but $m(\cdot;\theta_0)$ is partially unknown, Severini and Wong (1992) propose an estimation method. This is extended in Severini and Staniswalis (1994) and Fan, Heckman and Wand (1995) under unknown cdf of the latent variable. Finally, if the cdf of the data satisfies an index restriction, then Ai (1997) proposes the so called semiparametric maximum likelihood estimator and Ai and Chen (1999) a Generalized Method of Moments.

In all the previous papers, it is of great interest to test for the set of assumptions that is needed in order to obtain root-n consistent estimator of the parameter. For example, in Severini and Staniswalis (1994) it would be of interest to test for a partially linear single index restriction, or in Severini and Wong (1992) to test for a pre-specified conditional distribution of a particular statistical model. Furthermore, the interest to test for a specific conditional distribution might come from prediction purposes.

If we want to test a fully parametric model against a broad set of alternatives, then a pleiad of consistent specification tests are available. Following Hart (1997) mainly two fundamental approaches are considered: those statistics that appear as a weighted average of the residuals, and those statistics that compare parametric and nonparametric fits. Among the second group see for example Gonzalez-Manteiga and Cao (1993), Härdle and Mammen (1993), Whang and Andrews (1993), Zheng (1996) and Fan and Li (1996). The first group does only require the estimation of the econometric model under the null (parametric) hypothesis, see for example Bierens and Ploberger (1997), Stute (1997) and Delgado and Gonzalez-Manteiga (2001). The relative performance of both types of approaches is investigated in Guerre and Lavergne (2000) but also discussed in Horowitz and Spokoiny (2001), Guerre and Lavergne (2001) or Fan and Li (2000). Apart from testing a particular parametric form for a conditional moment, it is also possible to test for some qualitative features. Eubank, Hart, Simpson and Stefanski (1995), Gozalo and Linton (2000) and Sperlich, Tjøstheim and Yang (2000), propose statistical procedures to test for additivity. Finally, some semiparametric models have been tested against nonparametric alternatives. Yatchew (1992) and Whang and Andrews (1993) have developed consistent specification tests for a partially linear model; consistent tests for a single index specification have been presented in Chen (1992) and Rodriguez and Stoker (1992).

However, most of the previous tests in parametric (semiparametric) models only focus on particular features of the statistical model (conditional moment restrictions or qualitative features). Unfortunately, in many situations, restrictions on semiparametric models can not be written in these terms. In this case, more general specification tests are needed. Only Andrews (1997), using empirical processes, and Zheng (2000), using smoothers, propose a general test for the full parametric model check.

To our knowledge this paper is the first attempt to introduce a full model test check for semiparametric models. Although the test in Andrews (1997) and Zheng (2000) is designed for a full model check, the main difference with respect to ours is that, in our case, the null hypothesis is a semiparametric model. This extension is quite important because of several different potential applications (for example single index models and multiple index models), and the aim of our paper is to show how we can deal with the technical difficulties linked to our new approach. The semiparametric model considered in the null hypothesis is one that partially specifies the conditional distribution of a vector of endogenous variables given a vector of explanatory ones. The asymptotic properties of the test are obtained without requiring a specific form for the nonparametric estimator, and only very general conditions on its asymptotic behavior are requested.

The remainder of the paper is organized as follows. Our test and an appropriate semiparametric estimator are introduced in Section 2. The asymptotic behavior and consistency is treated in Section 3. In Section 4 we discuss the choice of smoothing parameters and an intensive simulation study shows the behavior of our test in finite samples. Finally, in the Appendix we prove the main results of the paper.

2 Definition of the Test

As a statistical framework for our test, we consider a rather general family of structural semiparametric models. This can include a broad class of microeconometric applications such as qualitative response models, Tobit type specifications or duration models (see Amemiya, 1985 for a fully parametric version).

Suppose we have a sample of n independent replicates $\{(Y_i, X_i)\}_{i=1,\cdots,n}$ from the pair of random variables $Y \in I\!R$, $X \in \mathcal{X}$ being $\mathcal{X}$ a compact set $\mathcal{X} \subset I\!R^d$, such that the conditional distribution of Y given X is $\ell_{Y|X}(y, x)$. The statistical model considered here consists of a family of conditional distributions

$$(1) \qquad \left\{\ell_{Y/X}(y, x; \theta, \eta_1, \cdots, \eta_p) \quad \theta \in \Theta, \quad \eta_1 \in H_1, \cdots, \eta_p \in H_p\right\},$$

where Θ is a compact subset of $I\!R^k$ and $H_1, \cdots, H_p$ are respectively compact subsets in $I\!R$. The parameters η_1, η_2, $\cdots$, η_p are functions of x, i.e. $\eta_1 = \eta_1(x_1)$, $\eta_2 = \eta_2(x_2)$, $\cdots$, $\eta_p = \eta_p(x_p)$, and the vectors $x_i \in \mathcal{X}_{d_i}$, $i = 1, \cdots, p$, are mutually excluded subsets of x such that $\mathcal{X} = \mathcal{X}_{d_1} \times \cdots \mathcal{X}_{d_p}$. Finally, η's are assumed to be unknown smooth functions $\eta_j : \mathcal{X}_{d_j} \to H_j$ that take values in a set Γ_j

$$\Gamma_j = \left\{ \phi \in C^2 \left(\mathcal{X}_{d_j} \right) : \phi(x_j) \subset H_j \quad \text{for} \quad \text{all} \quad x_j \in \mathcal{X}_{d_j} \right\}.$$

We remark that the statistical model in (1) falls in the class of the so called conditionally parametric models (Severini and Wong, 1992). However, this model easily nests other possible situations. For example, if the x variables appear directly in the form of the conditional density and not through the η's as in (1), i.e. writing $\ell_{Y/X}(y, x; \theta)$, then we have a fully parametric model, and therefore our test can be compared with those proposed in Andrews (1997) and Zheng (2000). In our work, the null hypothesis of interest will be

$$H_0 : \ell_{Y/X}(y, x; \theta, \eta_1, \cdots, \eta_p) = g(y, \eta_1, \eta_2, \cdots, \eta_p, \theta) \quad \text{for} \quad \text{some} \quad \eta_1 \in H_1, \cdots, \eta_p \in H_p, \quad \theta \in \Theta$$

where $g(\cdot)$ is a known function.

In order to propose a test statistic several alternative approaches are available. Among the various methodologies mentioned in the introduction, a very successful one has been to directly compare the parametric versus the non parametric fit. Following this idea, we propose a test that compares our (but convoluted) semiparametric fit against a nonparametric one. We propose the following test statistic

$$I_n = \int \left[\frac{1}{nh^d} \sum_{j=1}^{n} K\left(\frac{x - X_j}{h} \right) \left(Y_j - \widehat{m_S}(X_j) \right) \right]^2 \omega(x)\, dx, \tag{2}$$

where $K(\bullet)$ is a kernel function of dimension d , h is the bandwidth and $\omega(x)$ is a weight function. Here, the regression estimator $\widehat{m_S}(x)$ is defined as follows

$$\widehat{m_S}(x) = \int y g\left(y, \hat{\eta}_1, \cdots, \hat{\eta}_p, \hat{\theta} \right) dy, \tag{3}$$

$\hat{\theta}$ is some root-n consistent estimator of the parameter vector θ and $\hat{\eta}_1, \cdots, \hat{\eta}_p$ are nonparametric estimators at observation points $x = (x_1, x_2, \ldots, x_p)$, i.e. $\hat{\eta}_1 = \widehat{\eta_1}(x_1)$,..., $\hat{\eta}_p = \widehat{\eta_p}(x_p)$. For the ease of notation but without loss of generality we set $p = 2$ for the rest of the paper. We further define as h_1, h_2 the bandwidths related to the previous estimators.

We remark that the properties of our test statistic are going to depend on the statistical properties of $\hat{\theta}$ and $\hat{\eta}_1, \cdots, \hat{\eta}_p$. Therefore, it is important to discuss how these estimators are chosen. For the parametric part, θ, some root-n consistent semiparametric estimators are available (Severini and Wong, 1992; Ai, 1997; Ai and Chen, 1999 and others). As far as their rate of convergence is square root of n, these parametric estimators will not affect the asymptotic distribution of the test since other estimators, with a slower rate of convergence are present. For the nonparametric estimators, a minimal requirement is

(C.1) The function $g(\cdot)$ is Lipschitz continuous in all its arguments. Assume that under H_0 we have

$$\sup_{x_j \in \mathcal{X}_j} \left| \widehat{\eta_j}(x_j) - \eta_j(x_j) \right| = O\left(\sqrt{\frac{\log n}{n h_j^{d_j}}} \right) + o\left(h_j^{\alpha}\right)$$

for $j = 1, 2$, and

$$\hat{\theta} = \theta + O_p\left(n^{-1/2}\right),$$

as n tends to infinity.

This is simply a uniform convergence requirement and what it is important is that various estimators in the literature have this statistical property. Now, we derive an estimator that fulfills the previous requirements to highlight feasibility and the practical use. In order to do so, let us consider the weighted local likelihood approach developed in Staniswalis (1989). This method consists basically in approximating the likelihood function locally through some smooth weighting function. Under certain hypothesis on the likelihood function it is possible to obtain estimators for η_1 and η_2 by maximizing a local quasi-likelihood function at a given value (x_1, x_2) (see e.g. Rodriguez-Póo, Sperlich and Vieu, 2001). For this value we estimate η_1 as the solution to the following optimization problem

$$\widehat{\eta}_1^{*}(x_1) = \sup_{\eta_1 \in H_1} W(\eta_1),$$

where the weighted likelihood is

$$W_1(\eta_1) = \sum_{i=1}^{n} K_1\left(\frac{x_1 - X_{1.i}}{h_1}\right) \log g\left(Y_i, \eta_1, \widetilde{\eta}_2(X_{2,i}), \hat{\theta}\right),$$

$K_1(\bullet)$ is a kernel function of dimension d_1 , and h_1 the bandwidth. Furthermore, $\widetilde{\eta}_2(X_{2,i})$ is a pilot estimator of the function η_2 at point $X_{2,i}$. The function η_2 is estimated at point x_2 in the opposite way.

This estimator presents a nice asymptotic behavior that makes it a very helpful device in the analysis of the statistical properties of the test. Previous to introduce some further assumptions that are needed to obtain the asymptotic properties of the estimator let us introduce some definitions and notation.

Denote by $p(X)$ the marginal density of $X = (X_1, X_2)$ and by $p_j(X_j)$ the marginal density of X_j. Further, set $\sigma^2(x) = E\left[(Y - m(X))^2 | X = x\right]$ and $\sigma_j^2(x_j) = E\left[(Y - m(X))^2 | X_j = x_j\right]$. Let us define also

$$\varphi(y, x_1, x_2, \theta) = \ln g(y, \eta_1(x_1), \eta_2(x_2), \theta) \quad (4)$$

$$\varphi_i^{(j)}(y, x_1, x_2, \theta) = \frac{\partial}{\partial \eta_i^j} \ln g(y, \eta_1(x_1), \eta_2(x_2), \theta) \quad i, j = 1, 2. \quad (5)$$

Now we assume the following:

(L.1) The function $\varphi(y, x_1, x_2, \theta)$ is α-times continuously differentiable in all its arguments.

(L.2) Set

$$I_j(x_j, \theta) = E\left[\varphi_j^{(1)}(Y, \eta_1(X_1), \eta_2(X_2), \theta)^2 \middle| X_j = x_j\right]$$

for $j = 1, 2$. The functions $I_j(x_j, \theta)$ are α-times continuously differentiable at point $x_j \in \mathcal{X}_j \subset I\!R^{d_j}$ and $\theta \in \Theta$.

(L.3) $E\left[\varphi_1^{(1)}(Y, \eta_1(X_1), \eta_2(X_2), \theta) \middle| X_1 = x_1\right] = 0$ and $E\left[\varphi_2^{(1)}(Y, \eta_1(X_1), \eta_2(X_2), \theta) \middle| X_2 = x_2\right] = 0$ for all $\theta \in \Theta$.

(L.4) $I_j(x_j, \theta) > 0$ for all $x_j \in \mathcal{X}_j \subset I\!R^{d_j}$ and $\theta \in \Theta$.

(L.5) The kernel function K_j is compactly supported and Lipschitz continuous of order $\beta > 0$. Furthermore,

$$\int z_1^{\alpha_1} z_2^{\alpha_2} \cdots z_{d_j}^{\alpha_j} K_j\left(z_1, z_2, \cdots, z_{d_j}\right) = 0$$

for $\alpha_1 + \alpha_2 + \cdots + \alpha_{d_j} < \alpha$ and

$$0 \neq \left| \int z_1^{\alpha_1} z_2^{\alpha_2} \cdots z_{d_j}^{\alpha_j} K_j\left(z_1, z_2, \cdots, z_{d_j}\right) \right| < \infty$$

for $\alpha_1 + \alpha_2 + \cdots + \alpha_{d_j} = \alpha$,

(L.6) h_1 and h_2 fulfill the conditions $nh_j^{d_j} \longrightarrow \infty$, $nh_j^{d_j + 2\alpha} \longrightarrow 0$, for $j = 1, 2$.

(L.7) There exists two pilot estimators of $\eta_1(x_1)$ and $\eta_2(x_2)$, $\widetilde{\eta}_1(x_1)$, $\widetilde{\eta}_2(x_2)$, such that

$$\widetilde{\eta}_j(x_j) = \eta_j(x_j) + o_p(1), \quad \text{for } j = 1, 2,$$

as n tends to infinity, for $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$.

(L.1) and (L.2) are standard smoothness conditions. (L.3) is a condition that is needed in order to remove the bias that is introduced by the smoothness of η_2 when estimating η_1, and the opposite. (L.4) is an identification condition. (L.5) is a so called higher order kernel assumption. It is introduced because otherwise the bias of the estimator is too large and thus will achieve very slow rates. The order or the kernel matches the degree of differentiability that is assumed in (L.1) and (L.2). Assumption (L.6) determines the behavior of the bandwidths. Finally (L.7) is not very retrictive since it allows to use any consistent pilot estimate, even if this estimate suffers from the curse of dimensionality. Examples of such pilot estimates involve all the usual classical multivariate smoothers (kernel, wavelets, local polynomial, splines, ...). Note that the rate at which both bandwidths must tend to zero is going to depend on the degree of the kernel, the smoothness and the dimensionality.

Then, under the conditions established above, we claim the following result

Lemma 1 *Assume conditions (L.1)-(L.7) hold. Under H_0 we have*

$$\sup_{x_j \in \mathcal{X}_j} \left|\widehat{\eta}_j^*(x_j) - \eta_j(x_j)\right| = O\left(\sqrt{\frac{\log n}{nh_j^{d_j}}}\right) + o\left(h_j^{\alpha}\right)$$

for $j = 1, 2$ as n tends to infinity.

The uniform bounds are similar to those obtained in more general nonparametric regression estimators under similar conditions. In fact, as we will see in the next section, the statistical properties of the test do not depend on a explicit expression for the estimators of η_1 and η_2. Of course, because of Lemma 1, our estimators $\hat{\eta}_1^*$ and $\hat{\eta}_2^*$ are satisfying condition (C.1) and so they are concerned with our study.

3 Asymptotic Behavior

As we have mentioned in the previous section, the level and the power of our test are going to depend crucially on the theoretical properties of $\hat{\eta}_1$ and $\hat{\eta}_2$. These estimators appear in the test through the regression function estimate $\widehat{m_S}(x)$ that is defined in equation (3). Then, we will first analyze the properties of this regression function estimator and then the test. In order to do so, we need the following additional assumptions.

(B.1) The random variable Y is uniformly bounded by some quantity $B > 0$. $|Y| < B$.

This condition is introduced in order to make the presentation simpler, but it can be weakened to the existence of absolute moments of order greater than two, by using some truncation arguments (Mack and Silverman, 1982). The first result below will be a crucial tool for the test problem we are investigating in this paper. However it is interesting to note that, as a by-product our study, we have here a new regression estimate with good reduction dimension asymptotic properties. The proof of the following result can be found in the Appendix.

Theorem 1 *Let us denote $m(x) = E[Y|X_1 = x_1, X_2 = x_2]$. Then under conditions (C.1), (B.1) and if the null hypothesis H_0 is true, we have that*

$$\sup_{x\in\mathcal{X}} \left|\widehat{m_S}(x) - m(x)\right| = O_p\left(\sqrt{\frac{\log n}{nh_1^{d_1}}} + \sqrt{\frac{\log n}{nh_2^{d_2}}}\right) + o\left(h_1^{\alpha} + h_2^{\alpha}\right)$$

as n tends to infinity.

To establish the properties of the test statistic I_n we include now the following conditions:

(S.1) The function $m(x)$ is α-times continuously differentiable at point $x_j \in \mathcal{X}_j \subset I\!R^{d_j}$.

(S.2) The marginal densities $p(X)$ and $p_j(X_j)$, for $j = 1, 2$ are α-times continuously differentiable at point $x_j \in \mathcal{X}_j \subset I\!R^{d_j}$.

(S.3) The functions $\sigma^2(x)$, $\sigma_1^2(x_1)$ and $\sigma_2^2(x_2)$ are bounded from above and below.

(K.1) The kernel function is symmetric, α-times continuously differentiable with compact support. Furthermore, $\int K(u)du = 1$.

(H.1) h must tend to zero in such a way that $nh^d \longrightarrow \infty$.

(H.2) h, h_1 and h_2 have the following relationships: $h_j^{d_j} \Big/ h^{d/2} \longrightarrow \infty$, for $j = 1, 2$.

All these conditions are standard in this type of test statistics (see e.g. Härdle and Mammen, 1993 or Zheng, 1996). However, it is important to remark that a great difference with respect to the previous tests is that a semiparametric estimator is tested instead of a parametric one. This makes it much more difficult to handle the so called bias problem. In the previous tests this problem was often solved by smoothing the parametric estimator, too. In our case, we smooth a function estimator that shows nonparametric rates of convergence, and therefore the bias of the test is large, see also the discussion in Section 4. In order to overcome this problem, we have introduced the assumption (H.2) that somehow relates the rates of decay of the bandwidths for the different function estimates. Furthermore, conditions (S.1), (S.2) and (K.1) also care for the reduction of this high order term.

The asymptotic behavior of the test statistic under the null hypothesis is given in the following result that, again, is shown in the Appendix.

Theorem 2 *Under conditions (C.1), (S.1)-(S.3), (B.1), (K.1) and (H.1)-(H.2) and if the null hypothesis H_0 is true then*

$$nh^{d/2} I_n - Bh^{-d/2} \longrightarrow_d N(0,V),$$

where

$$B = \int K^2(t)\,dt \int \sigma^2(u)\,p(u)\,\omega(u)\,du$$

and

$$V = \left[\int\int K(0)\,K(u-v)\,dudv\right]^2 \int \sigma^4(u)\,p^2(u)\,\omega^2(u)\,du$$

as n tends to infinity.

Obviously, both terms, the bias B and the variance V can be estimated straight forwardly by

$$\hat{B} = \int K^2(t)\,dt \int \hat{\sigma}^2(u)\,\hat{p}(u)\,\omega(u)\,du$$

and

$$\hat{V} = \left[\int\int K(0)\,K(u-v)\,dudv\right]^2 \int \hat{\sigma}^4(u)\,\hat{p}^2(u)\,\omega^2(u)\,du,$$

where $\hat{p}$ can be any consistent nonparametric density estimator whereas $\hat{\sigma}$ can be estimated from the residuals. However, it is well known that to obtain reliable results based on the asymptotic results given in Theorem 2 a large sample size (depending on d) is needed. Furthermore, note that the bias term is of factor $h^{-d/2}$. This makes necessary the use of a bootstrap procedure to approximate the asymptotic distribution of the test statistic. A proposal of bootstrap procedure designed to overcome this problem will be shown at the end of this section.

Next, we determine the power of our test. In order to do so, let us define the sequence of local alternatives to the null hypothesis

(6) $$H_a : \ell_{Y/X_1,X_2}(y,x_1,x_2) = g(y,\eta_1(x_1),\eta_2(x_2),\theta) + \gamma_n \psi(y,x_1,x_2),$$

where γ_n is a sequence that tends to zero such that

(A.1) $\gamma_n h^{-d/2} \longrightarrow \infty$ and $\gamma_n = o\left(\frac{1}{nh^d}\right)$

(A.2) $\gamma_n^2 \frac{nh_j^{d_j}}{\log n} \longrightarrow \infty$ for $j = 1,2$.

(A.3) The function $\psi(\cdot)$ is continuous and bounded in all its arguments.

Note that under conditions (A.1) and (A.2), $h^{d/2} \preccurlyeq \gamma_n \preccurlyeq \sqrt{\frac{1}{nh^d}}$.

Assumptions (A.1), (A.2) and (A.3) are also related to the bias problem that has been mentioned above. Now we can state the following result, proved in the Appendix.

Theorem 3 *Under conditions (S.1)-(S.3), (K.1) and (H.1)-(H.2) and if H_a is true, such that it verifies (A.1)-(A.3) then for all sequences of random variables $\{c_n : n \geq 1\}$ with $c_n = O_p(1)$ we have*

$$P\left(I_n > c_n \middle| X_1, X_2\right) \to 1$$

as n tends to infinity.

Note that Theorem 3 indicates that our test has nontrivial power only against sequences of local alternatives for which γ_n tends to zero at a rate that is smaller than $\sqrt{n}$. As seen in Andrews (1997), tests based on weighted parametric residuals have nontrivial power against local alternatives for which the rate is exactly $\sqrt{n}$. Thus, at least in terms of the asymptotic local power these tests appear to dominate tests that require slower rates. However, as shown in Guerre and Lavergne (2000) or Horowitz and Spokoiny (2001), at a exact rate of $\sqrt{n}$ no test can have nontrivial power uniformly over reasonable classes of functions $\phi(\cdot)$ in (6). Furthermore, we conjecture that a test based on smoothing residuals for our problem is going to depend explicitly on the estimator for η_1 and η_2. So a general result as ours would hardly be available.

To get the critical values of the test we suggest to use the bootstrap methodology following the same steps as in Härdle, Huet, Mammen and Sperlich (2000). Under the null hypothesis, it is easy to generate B bootstrap samples $Y_1^*, \ldots, Y_n^*$ since the conditional density is already specified. We generate the Y_i^* from $g\{\hat{\eta}_1(X_{1,i}), \hat{\eta}_2(X_{2,i}), \hat{\theta}\}$. In the case of a non-standard distribution it can be quite hard to generate these values, fortunately there are several methods available like the so called *rejection method*, see Ripley (1987). The test statistic I_N will then be compared to the B bootstrap analogs

$$I_n^* = \int \left[\frac{1}{nh^d} \sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right) \left\{Y_j^* - \widehat{m}_S^*(X_j)\right\} \right]^2 w(x)dx.$$

Under the given conditions for the $\hat{\eta}_k$, $k = 1, 2$ we conjecture the consistency of the bootstrap following the same lines as in the proof in Härdle, Huet, Mammen and Sperlich (2000). Under the assumptions of Theorem 2, it holds that

$$d_K\{\mathcal{L}^*(I_n^*), \mathcal{L}(I_n)\} \xrightarrow{P} 0,$$

where d_K denotes the Kolmogorov distance, which is defined for two probability measures μ and ν (on the real line) as

$$d_K(\mu, \nu) = \sup_{t \in I\!R} \left| \mu(X \leq t) - \nu(X \leq t) \right|.$$

4 Finite Sample Problems and Simulations

In the literature of parametric versus non- or semiparametric regression tests the choice of the smoothing parameter and the bias handling in the nonparametric part are the most out standing problems.

With respect to the bandwidth choice (h) there is a growing literature in the last three years about so called adaptive testing methods. The main and common idea is to look for a smoothing parameter that finds the maximal deviation between the null hypothesis and the alternative. Afterwards one determines the appropriate critical value by Monte Carlo Methods since the extreme value theory turned out to be of little help here. For kernel methods, the first article in this direction has been published recently by Horowitz and Spokoiny (2001), and an alternative direction is proposed in Guerre and Lavergne (2001). So far they treat simple testing problems and the solution is straightforward only for the one dimensional case. Although we believe that this principle can be extended to our problem we refer this to future research.

Let us turn to the bias problem. In the context of kernel based methods, a successful approach to take the bias into account has been always to convolute the parametric estimator with the same kernel and bandwidth as the nonparametric one. This should produce the same bias effect for both parts and thus cancel it out. Notice that we have followed this idea in our test statistic. However, a substantial additional complication in our testing problem is that we do not simply compare a parametric versus a non- or semiparametric function but a semiparametric versus a nonparametric one. This requires the additional choice of smoothing parameters for the model of the null hypothesis, here called h_j, $j = 1, 2$. Since we wanted to keep our results as general as possible, i.e. for any kind of estimator fulfilling the conditions discussed in Section 2, it is not possible to give a general rule for the bandwidth choice. Let us highlight the underlying problem concentrating on the case when H_0 is true.

Recall that we want to compare the difference $\{Y - \widehat{m}_S(h_1, h_2)\}$ with the bootstrap analog $\{Y^* - \widehat{m}^*_S(h^*_1, h^*_2)\}$. If we choose $(h^*_1, h^*_2) = (h_1, h_2)$, then since Y^* is generated from $\widehat{m}_S(h_1, h_2)$ and consequently it contains the smoothing bias, the difference in the bootstrap world will tend to underestimate the real one. This can also be interpreted in the following way. Since Y^* is generated from $\widehat{m}_S(h_1, h_2)$, the bandwidths h_1, h_2 become a parameter of the specification of our null hypothesis. E.g. if we choose h_1 and h_2 close to ∞, the hypothesis fixes the nonparametric components η_1, η_2 equal to a constant. Then, even though the data generating density is correctly specified, our test will reject the null hypothesis.

This is a quite new phenomena in non- and semiparametric testing that to our knowledge so far occurred similarly only in the problem of additivity testing, see Gozalo and Linton (2001) or Sperlich, Tjøstheim and Yang (2000). As we clearly speak of a bias problem here, a simple remedy in practice would be undersmoothing. Then we definitely test only the density specification but, as always, pay with an increase in the variance. In other words, h_1, h_2 *too small* leads to a conservative test, h_1, h_2 *too big* leads to rejections caused by bad η_1, η_2 estimates.

In theory, the solution is rather simple. Typically one recommends to choose $(h^*_1, h^*_2) > (h_1, h_2)$ and thus replicates the bias in the bootstrap. Other papers, see Härdle, Huet, Mammen and Sperlich (2000) include a bias estimate in their test statistic. We implemented the second approach in our context but surprisingly it turned out not to work at all here. Now have additionally in mind that we need to smooth the nonparametric part sufficiently to reach convergence for the Newton Raphson algorithm. So, at least if oversmoothing is necessary for numerical reasons one has to choose the first approach. Unfortunately this one does not say anything about the bandwidth size in practice. For these reasons we will present simulation studies for different cases, choosing $(h^*_1, h^*_2) = (h_1, h_2)$ but small, and $(h^*_1, h^*_2) > (h_1, h_2)$ when we had to oversmooth due to data sparseness. Further note that in our simulations we did no additional weighting or trimming, i.e. we set $w(x) \equiv 1$ throughout.

To demonstrate the performance we present various simulation results for a model with truncated variables (Tobit I). We simulated the data with

$$y_i = \begin{cases} h(x_i) + u_i & \text{if } h(x_i) + u_i > 0 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where $x_1 \sim U[0,2]^2$, $x_2 \sim U[-1,1]$ and u_i i.i.d. errors. Our H_0-density for Y was the normal one with conditional expectation

$$h(x,t) = x_1^T \gamma + \eta(x_2), \qquad x_1 = (x_{11}, x_{12})^T \tag{8}$$

and constant variance σ, i.e. the u_i were standard normal with variance σ. Restricting on finite samples, the boundedness condition (B.1) is always fulfilled.

We simulated the data from the following models:

$$h(x,t) = x_1^T \gamma + 2\sin(2x_2), \qquad \gamma = (-1.5, 2)^T \tag{9}$$

$$h(x,t) = x_1^T \gamma + (1.5x_{11}^2 - 0.75) + 2\sin(2x_2), \qquad \gamma = (-1.5, 2)^T \tag{10}$$

$$h(x,t) = x_1^T \gamma + (2x_{11}x_{12} - 2) + 2\sin(2x_2), \qquad \gamma = (-1.5, 2)^T \tag{11}$$

all with an additional $N(0,1)$ error term u_i. A further model was equal to (9) but with error distribution of variance equal to 1 and following a distribution

$$(\chi_1^2 - 1)/\sqrt{2} \tag{12}$$

Note that model (9) is the only one that holds the null hypothesis H_0. In models (10), (11), and (12) the "additional terms" are centered to zero so they do not affect the truncation threshold. We simulated also models without such centering, i.e. changing the threshold. In particular, take e.g. model (11) without the term "-2" in the brackets. In those cases the power of the test was overwhelming. Therefor we concentrate here on the presentation of the cases when detection of an alternative is difficult as e.g. in the cases enumerated above. For the simulation study we did only $B = 250$ bootstrap replications. In empirical research one should take maybe at least $B = 1000$. The estimation is done according to Severini and Wong (1992) or Rodriguez-Póo, Sperlich and Vieu (2001). For a clearer presentation we separated our study in four case studies.

CASE 1: Investigate behavior of statistic I_n with $h_1^* > h_1$ when data are sparse.

We first considered the case when the choice of a "sufficiently" small bandwidth h_1 is not possible. Then, as mentioned above, oversmoothing in the bootstrap ($h_1^* > h_1$) is necessary to yield valuable results. For bandwidths h we tried various sizes. Fortunately, the choice of h turned out to be less crucial for the outcome. To summarize, for the estimation of γ, σ and η_1 under H_0, we used bandwidth $h_1 = 3n^{-1/5}$, and for the bootstrap we increased h_1 by 30% to replicate the bias in the estimation of the H_0-model, i.e. $h_1^* = 1.\bar{3}\, h_1$. Calculating the test statistic I_n we used $h = 3n^{-1/7} s_X$. Here, s_X is the vector of empirical standard deviations for each covariate. We draw about 320 to 390 observations to end up always with only $n = 200$ for which $Y > 0$.

For only $n = 200$ and manipulating h_1^* the test could not detect the misspecification when neglecting the interaction in model (11). In contrary, neglecting this interaction rose the estimate of σ so that the bootstrap samples have a much bigger variance. Consequently the simulated $p-$value was mostly close to 0.5 without much variation. This changed when X_{11}, X_{12} were no longer independent; but it is hard to make general statements about that.

The results for the remaining models are summarized in Table 1. The test holds the quantiles under H_0 and we see that one could get rid of the bias problem in the bootstrap replications knowing the appropriate h_1^*. Further, the distribution is fit quite well. This can be seen by renormalizing the first line (H_0-case) dividing it by 1.364. Then we get for the quantiles $1.0, 5.24, 9.75$, and 14.15. We will judge the power against the different alternatives when comparing this later with CASE 2. So far, we see clearly that the deviations can be detected by the test. Recall further that the power even increases remarkable if the misspecification affects the threshold of truncation in (7).

CASE 2: We repeated the simulations from CASE 1 but increased n to see the corresponding increase in power.

For a "fair" comparison with CASE 1 we kept the same size-relation between h_1^* and h_1. Unfortunately, a negative but expected by-product was that $h_1^* = 1.\bar{3}\ h_1$ caused more bias in the bootstrap than necessary. This, additionally to the here demonstrated "power" effect, tells us a lot about the crucial problem inherent in the popular statement *the bootstrap bandwidth has to be chosen somewhat bigger*.

As already said, to see the increase in power without playing in new with the bandwidths, we chose them as above. We generated about 470 to 560 observations to have now $n = 300$ with $Y > 0$.

It turned out that with increasing N the bias problem vanished so that the increase of h_1 by 30% is too much and the test became quite conservative. However, we see very well in Table 1 how the distance between the p-values of H_0 and H_a increase with n. Again we see how well the bootstrap replicates the distribution (except the bias shift) by normalizing the first line (H_0-case) multiplying it by 3.0. This gives the wanted quantiles $1.0, 5.0, 8.05$, and 15.05.

Table 1 about here.

data generating model	simulation		rejections at significance level			
$n = 200$	runs	$p-$value	1%	5%	10%	15%
(9), H_0	500	49.95	1.364	7.150	13.30	19.30
(10)	250	12.00	27.20	52.20	63.90	72.80
$u \sim (\chi_1^2 - 1)/\sqrt{2}$	250	23.90	4.098	24.33	35.25	46.70
$n = 300$						
(9), H_0	500	63.41	0.335	1.672	2.676	5.017
(10)	250	3.908	61.36	79.28	87.65	92.03
$u \sim (\chi_1^2 - 1)/\sqrt{2}$	250	29.58	6.400	18.00	27.20	39.20

Table 1: *For* **Cases 1** *and* **2***: The bootstrap $p-$values and percentages of rejections at various significant levels for different models.*

CASE 3: The behavior of the test when $X_2 \in I\!R^2$ with $(h_1^*, h_2^*) > (h_1, h_2)$.

Now we simulated a more complicated testing problem having a two dimensional nonparametric part in the model. We still considered the data generating process (7) with H_0 as in equation (8). The data were generated from the models

$$h(x,t) = x_1^T\gamma + 2\sin(2x_{12}) + 1.5x_{22}^2, \qquad \gamma = (-1.5, 2)^T \tag{13}$$

$$h(x,t) = x_1^T\gamma + (1.5x_{11}^2 - 0.75) + 2\sin(2x_2) + 1.5x_{22}^2, \qquad \gamma = (-1.5, 2)^T \tag{14}$$

We draw about 340 to 400 observations to end up always with $n = 250$ not truncated observations. For the estimation of under H_0 we used bandwidth $h_1 = n^{-1/9}2.5$ and the 4th order optimal kernels, see Gasser, Müller and Mammitzsch (1985). For the test statistic we chose $h = n^{-1/8}2.5s_X$ with the 2nd order quartic kernel. We again had to increase h_1 by 30% due to data sparseness caused by the higher dimensionality. The results, given in Table 2, are quite appealing and show that the test works fine even for the higher dimension case with only $n = 250$ and $B = 250$ bootstrap replicates.

Table 2 about here.

test statistic	data generating model	simulation runs	$p-$value	rejections at significance level 1%	5%	10%	15%
CASE 3	(13), H_0	500	48.95	1.000	5.400	9.000	12.00
	(14)	250	10.05	33.60	56.80	70.40	74.80
	(9), H_0	500	43.77	1.600	6.200	13.60	19.20
CASE 4	(10)	250	.0000	100.0	100.0	100.0	100.0
	(11)	250	2.450	58.40	89.20	94.00	96.40
	$u \sim (\chi_1^2 - 1)/\sqrt{2}$	250	23.16	8.800	26.40	44.00	51.20

Table 2: *For* **Cases 3** *and* **4***: The bootstrap $p-$values and percentages of rejections at various significant levels for different models, $n = 250$.*

CASE 4: Finally, we investigated the behavior of statistic I_n with $h_1^* = h_1$.

Now we come to the case when we try to chose a "sufficiently" small bandwidth h_1. Recall that then oversmoothing in the bootstrap ($h_1^* > h_1$) is maybe not necessary to yield valuable results.

We draw about 400 to 470 observations to end up always with $n = 250$ data where $Y > 0$. According to the bandwidth discussion above we chose h_1 rather small, in particular $h_1 = n^{-1/5}3s_X$. Calculating the test statistic I_n we used $h = n^{-1/7}6s_X$, same for the bootstrap. The model estimation as well as the test statistic we performed with 2nd order quartic kernels. To keep things easy, the rest of the simulations were as in CASE 1.

The results are also summarized in Table 2. We see that for the chosen bandwidths the test holds almost the quantiles under H_0 but was slightly affected by the bias problem discussed above. The distribution is fit well. This can be seen by renormalizing the first line (H_0-case) dividing it by 1.25. Then we get for the quantiles $1.28, 4.96, 10.88$, and 15.36.

Note that the power against the different alternatives is pretty good. Actually, we see clearly that now all deviations from the null hypothesis got detected by our test. Recall further that the power increases even more when the misspecification affects the threshold of truncation in (7).

5 Appendix

Proof of Lemma 1

In order to simplify the proof, we make it for $\widehat{\eta}_1^*$. For the other term $\widehat{\eta}_2^*$, the proof would be analogous. Note first that if at values $x = (x_1, x_2)$, $\widehat{\eta}_1^*$ is the maximizer of $W(\eta_1)$ then, it is also the maximizer of

$$W_1^*(\eta_1) = \frac{W(\eta_1)}{\sum_{i=1}^n K_1\left(\frac{x_1 - X_{1,i}}{h_1}\right)}.$$

By substituting $W_1(\eta_1)$ by its definition we obtain

$$W_1^*(\eta_1) = \frac{\sum_{i=1}^n K_1\left(\frac{x_1 - X_{1,i}}{h_1}\right) \log g\left(Y_i, \eta_1, \widetilde{\eta}_2(X_{2,i}), \hat{\theta}\right)}{\sum_{i=1}^n K_1\left(\frac{x_1 - X_{1,i}}{h_1}\right)}.$$

Using assumption (L.1), a Taylor expansion of $\varphi_1^{(1)}$ around the point $\eta_1(X_{1,i})$ gives directly the existence of some $\overline{\eta_1}(X_{1,i})$ belonging between η_1 and $\eta_1(X_{1,i})$ such that

$$\begin{aligned}
&\varphi_1^{(1)}\left(Y_i, \eta_1, \widetilde{\eta}_2(X_{2,i}), \hat{\theta}\right) = \\
&= \varphi_1^{(1)}\left(Y_i, \eta_1(X_{1,i}), \widetilde{\eta}_2(X_{2,i}), \hat{\theta}\right) + (\eta_1 - \eta_1(X_{1,i}))\,\varphi_1^{(2)}\left(Y_i, \overline{\eta_1}(X_{1,i}), \widetilde{\eta}_2(X_{2,i}), \hat{\theta}\right) \quad .
\end{aligned}$$

This leads directly to

$$\frac{\partial W_1^*(\eta_1)}{\partial \eta_1} = A_1(\hat{\theta}) + A_2(\eta_1, \hat{\theta}) + (\eta_1 - \eta_1(x_1))\, A_3(\hat{\theta}) \ , \tag{15}$$

where

$$A_1(\hat{\theta}) = \frac{\sum_{i=1}^n K_1\left(\frac{x_1 - X_{1,i}}{h_1}\right) \varphi_1^{(1)}\left(Y_i, \eta_1(X_{1,i}), \widetilde{\eta}_2(X_{2,i}), \hat{\theta}\right)}{\sum_{i=1}^n K_1\left(\frac{x_1 - X_{1,i}}{h_1}\right)},$$

$$A_2(\eta_1, \hat{\theta}) = \frac{\sum_{i=1}^n K_1\left(\frac{x_1 - X_{1,i}}{h_1}\right) (\eta_1(x_1) - \eta_1(X_{1,i}))\,\varphi_j^{(2)}\left(Y_i, \overline{\eta_1}(X_{1,i}), \widetilde{\eta}_2(X_{2,i}), \hat{\theta}\right)}{\sum_{i=1}^n K_1\left(\frac{x_1 - X_{1,i}}{h_1}\right)}$$

and

$$A_3(\hat{\theta}) = \frac{\sum_{i=1}^n K_1\left(\frac{x_1 - X_{1,i}}{h_1}\right) \varphi_j^{(2)}\left(Y_i, \overline{\eta_1}(X_{1,i}), \widetilde{\eta}_2(X_{2,i}), \hat{\theta}\right)}{\sum_{i=1}^n K_1\left(\frac{x_1 - X_{1,i}}{h_1}\right)}.$$

Now we will study the asymptotics of the previous terms. For the term $A_3(\hat{\theta})$, assumptions (L.1) and (L.7) are sufficient to show that

$$A_3(\hat{\theta}) = \frac{\sum_{i=1}^n K_1\left(\frac{x_1 - X_{1,i}}{h_1}\right) \varphi_j^{(2)}\left(Y_i, \eta_1(X_{1,i}), \eta_2(X_{2,i}), \theta\right)}{\sum_{i=1}^n K_1\left(\frac{x_1 - X_{1,i}}{h_1}\right)} + O_p\left(\|\hat{\theta} - \theta\|^2\right) + o_p\left(A_3(\theta)\right).$$

Using condition (L.6) on the bandwidth, a strong law of large numbers and the condition $\sqrt{n}(\hat{\theta} - \theta) = O_p(1)$ then the following result holds

$$A_3(\hat{\theta}) = -I_1(x_1, \theta) + o_p(1), \tag{16}$$

with

$$I_1(x_1, \theta) = E\left[\varphi_1^{(1)}\left(Y, \eta_1(X_1), \eta_2(X_2), \theta\right)^2 \middle| X_1 = x_1\right].$$

For the term $A_1(\hat{\theta})$ note that assuming (L.1), by the mean value theorem we can write

$$A_1(\hat{\theta}) = \frac{\sum_{i=1}^n K_1\left(\frac{x_1 - X_{1,i}}{h_1}\right) \varphi_1^{(1)}\left(Y_i, \eta_1(X_{1,i}), \widetilde{\eta}_2(X_{2,i}), \theta\right)}{\sum_{i=1}^n K_1\left(\frac{x_1 - X_{1,i}}{h_1}\right)}$$

$$+\frac{\sum_{i=1}^{n}K_1\left(\frac{x_1-X_{1,i}}{h_1}\right)\frac{\partial}{\partial\theta^T}\varphi_1^{(1)}\left(Y_i,\eta_1\left(X_{1,i}\right),\widetilde{\eta}_2\left(X_{2,i}\right),\bar{\theta}\right)}{\sum_{i=1}^{n}K_1\left(\frac{x_1-X_{1,i}}{h_1}\right)}\left(\hat{\theta}-\theta\right).$$

and since $\sqrt{n}(\hat{\theta}-\theta)=O_p(1)$ then

$$A_1(\hat{\theta})=A_1(\theta)+O_p\left(n^{-1/2}\right),$$

where

$$A_1(\theta)=\frac{\sum_{i=1}^{n}K_1\left(\frac{x_1-X_{1,i}}{h_1}\right)\varphi_1^{(1)}\left(Y_i,\eta_1\left(X_{1,i}\right),\widetilde{\eta}_2\left(X_{2,i}\right),\theta\right)}{\sum_{i=1}^{n}K_1\left(\frac{x_1-X_{1,i}}{h_1}\right)}.$$

Now, taking a Taylor expansion around $\eta_2\left(X_{2,i}\right)$ we obtain

$$A_1(\theta)=A_{1,1}(\theta)+A_{1,2}(\theta),$$

where

$$A_{1,1}(\theta)=\frac{\sum_{i=1}^{n}K_1\left(\frac{x_1-X_{1,i}}{h_1}\right)\varphi_1^{(1)}\left(Y_i,\eta_1\left(X_{1,i}\right),\eta_2\left(X_{2,i}\right),\theta\right)}{\sum_{i=1}^{n}K_1\left(\frac{x_1-X_{1,i}}{h_1}\right)}$$

and

$$A_{1,2}(\theta)=\frac{\sum_{i=1}^{n}K_1\left(\frac{x_1-X_{1,i}}{h_1}\right)\left(\eta_2\left(X_{2,i}\right)-\widetilde{\eta}_2\left(X_{2,i}\right)\right)\frac{\partial^2}{\partial\eta_2\partial\eta_1}\log g\left(Y_i,\eta_1\left(X_{1,i}\right),\overline{\eta_2}\left(X_{2,i}\right),\theta\right)}{\sum_{i=1}^{n}K_1\left(\frac{x_1-X_{1,i}}{h_1}\right)}$$

where $\overline{\eta_2}\left(X_{2,i}\right)$ is between $\eta_2\left(X_{2,i}\right)$ and $\widetilde{\eta}_2\left(X_{2,i}\right)$. Since $A_{1,1}(\theta)$ is just a Nadaraya-Watson regression estimator of $E\left[\varphi_1^{(1)}\left(Y_i,\eta_1\left(X_{1,i}\right),\eta_2\left(X_{2,i}\right),\theta\right)\middle| X_{1,i}\right]$, under assumptions (L.1)-(L.6) it is straightforward to show that

$$A_{1,1}(\theta)\longrightarrow_p 0\ ,$$

$$E\left(A_{1,1}(\theta)\right)=0$$

and

$$Var(A_{1,1}(\theta))=\frac{1}{nh_1^{d_1}}\left[\int K_1^2(u)du\right]\frac{I_1(x_1,\theta)}{p_1(x_1)}.$$

For $A_{1,2}(\theta)$, proceeding as for $A_3(\theta)$ we get

$$A_{1,2}(\theta)=\frac{\sum_{i=1}^{n}K_1\left(\frac{x_1-X_{1,i}}{h_1}\right)\left(\eta_2\left(X_{2,i}\right)-\widetilde{\eta}_2\left(X_{2,i}\right)\right)\frac{\partial^2}{\partial\eta_2\partial\eta_1}\log g\left(Y_i,\eta_1\left(X_{1,i}\right),\eta_2\left(X_{2,i}\right),\theta\right)}{\sum_{i=1}^{n}K_1\left(\frac{x_1-X_{1,i}}{h_1}\right)}+o_p\left(A_{1,2}(\theta)\right).$$

But now by assumption (L.7) we obtain

$$A_{1,2}(\theta)\longrightarrow_p 0\ ,$$

$$E\left(A_{1,2}(\theta)\right)=o\left(h_1^{\alpha}\right)+o\left(n^{-1}h_1^{-d_1}\right)$$

and

$$Var(A_{1,2}(\theta))=o\left(Var(A_{1,1}(\theta))\right).$$

Hence, we finally obtain

$$A_1(\theta)\longrightarrow_p 0\ ,$$

(17) $$E\left(A_1(\theta)\right) = o\left(h_1^{\alpha}\right) + o\left(n^{-1}h_1^{-d_1}\right)$$

and

$$Var(A_1(\theta)) = \frac{1}{nh_1^{d_1}}\left[\int K_1^2(u)du\right]\frac{I_1(x_1,\theta)}{p_1(x_1)} + o\left(\frac{1}{nh_1^{d_1}}\right).$$

The term $A_2(\eta_1,\theta)$ can be dealt with the same arguments as $A_{1,2}(\theta)$. Using the same assumptions we get

$$A_2\left(\eta_1,\theta\right) \longrightarrow_p 0 \ ,$$

(18) $$E\left(A_2\left(\eta_1,\theta\right)\right) = o\left(h_1\right) + o\left(n^{-1}h_1^{-d_1}\right)$$

and

$$Var(A_2\left(\eta_1,\theta\right)) = o\left(Var\left(A_1\right)\right).$$

Combining now expressions (15), (16), (17) and (18) we have

$$\sqrt{nh_1^{d_1}}\left(\widehat{\eta}_1^*\left(x_1\right) - \eta_1\left(x_1\right)\right) = \sqrt{nh_1^{d_1}}\left(\frac{-A_1(\theta) - A_2\left(\eta_1,\theta\right)}{A_3(\theta)}\right).$$

Furthermore, since by assumption (L.6), $nh_1^{d_1+2\alpha} \longrightarrow 0$, we obtain by applying (16)

$$\sqrt{nh_1^{d_1}}\left(\widehat{\eta}_1^*\left(x_1\right) - \eta_1\left(x_1\right)\right) = \sqrt{nh_1^{d_1}}\left(\frac{A_1(\theta)}{I_1(x_1,\theta)}\right).$$

Finally, by Bernstein's inequality (see Serfling, 1980; Lemma A, p. 95) it holds

$$\sqrt{nh_1^{d_1}}\left(\widehat{\eta}_1^*\left(x_1\right) - \eta_1\left(x_1\right)\right) = o_p(1).$$

Note that the arguments of continuity are uniform because $\mathcal{X}_1$ and $\mathcal{X}_2$ are compact. Then, since by assumption (L.5) the kernels are Lipschitz continuous the proof is closed.

∎

For the proof of Theorem 1 we first need to establish the following lemma:

Lemma 2 *Assume that condition (C.1) holds. Under H_0 we have the following result*

$$\sup_{y\in\mathcal{Y}}\sup_{x_1\in\mathcal{X}_1}\sup_{x_2\in\mathcal{X}_2}\sup_{\theta\in\Theta}\left|g\left(y,\widehat{\eta}_1\left(x_1\right),\widehat{\eta}_2\left(x_2\right),\hat{\theta}\right) - g\left(y,\eta_1\left(x_1\right),\eta_2\left(x_2\right),\theta\right)\right|$$

$$= O_p\left(\sqrt{\frac{\log n}{nh_1^{d_1}}}\right) + O_p\left(\sqrt{\frac{\log n}{nh_2^{d_2}}}\right) + o\left(h_1^{\alpha} + h_2^{\alpha}\right) + O_p\left(n^{-1/2}\right)$$

for $j = 1,2$ as n tends to infinity.

Proof of Lemma 2

Since by assumption (C.1), $g(\bullet)$ is differentiable we have

$$\left|g\left(y,\widehat{\eta}_1\left(x_1\right),\widehat{\eta}_2\left(x_2\right),\hat{\theta}\right) - g\left(y,\eta_1\left(x_1\right),\eta_2\left(x_2\right),\theta\right)\right|$$

$$\leq \ C\left\{\left|\widehat{\eta}_1\left(x_1\right) - \eta_1\left(x_1\right)\right| + \left|\widehat{\eta}_2\left(x_2\right) - \eta_2\left(x_2\right)\right| + \left|\hat{\theta} - \theta\right|\right\}.$$

This is uniform because $\mathcal{X}_1$, $\mathcal{X}_2$ and Θ are compact sets. So conditions (C.1) and (L.6) apply and the proof is done.

∎

Proof of Theorem 1

The proof of this result is a direct consequence from Lemma 2 since

$$m(x) = \int y \ell_{Y/X_1,X_2}(y, x_1, x_2)\, dy.$$

Then we can write

$$|m(x) - \widehat{m_S}(x)| = \left| \int y \left(\ell_{Y/X_1,X_2}(y, x_1, x_2) - g\left(y, \widehat{\eta_1}(x_1), \widehat{\eta_2}(x_2)\right), \hat{\theta} \right) dy \right|. \tag{19}$$

Furthermore under the null hypothesis we have

$$\int y \ell_{Y/X_1,X_2}(y, x_1, x_2)\, dy = \int y g\left(y, \eta_1(x_1), \eta_2(x_2), \theta\right) dy. \tag{20}$$

Thus, the left hand side of (19) is equal to

$$\left| \int y \left(g\left(y, \eta_1(x_1), \eta_2(x_2), \theta\right) - g\left(y, \widehat{\eta_1}(x_1), \widehat{\eta_2}(x_2)\right), \hat{\theta} \right) dy \right| \tag{21}$$

and using assumption (B.1), (21) is bounded by

$$C \left| g\left(y, \eta_1(x_1), \eta_2(x_2), \theta\right) - g\left(y, \widehat{\eta_1}(x_1), \widehat{\eta_2}(x_2), \hat{\theta}\right) \right| . \tag{22}$$

Now Lemma 2 applies and the proof is done.

∎

For proving the result presented in Theorem 2 we will decompose the test statistic I_n in five terms which are separately treated in the following five lemmata.

Lemma 3 *Assume that conditions (S.1)-(S.3), (K.1) and (H.1) hold, and let*

$$S_n = \int \left(\frac{1}{nh^d}\right)^2 \sum_{j=1}^{n} K^2\left(\frac{x - X_j}{h}\right) \varepsilon_j^2 \omega(x)\, dx.$$

Then

$$\left(nh^d\right) S_n = B + o_p\left(h^{d/2}\right),$$

where

$$B = \int K^2(t)\, dt \int \sigma^2(u)\, p(u)\, \omega(u)\, du,$$

and

$$\varepsilon_j = Y_j - m(X_j)$$

as n tends to infinity.

Proof of Lemma 3

Taking iterated expectations and using the i.i.d. structure of our observations we have

$$E[S_n] = \frac{1}{nh^{2d}} \int \int \sigma^2(u) K^2\left(\frac{x-u}{h}\right) p(u)\, du \omega(x)\, dx.$$

Integrating by substitution we obtain

$$E[S_n] = \frac{1}{nh^d} \int K^2(t)\, dt \int \sigma^2(u) p(u) \omega(u)\, du + O\left(\frac{1}{n}\right).$$

For the variance we have the following expression

$$Var[S_n] = \frac{1}{n^4 h^{4d}} \sum_{j=1}^{n} Var[F_j],$$

where

$$F_j = \int \varepsilon_j^2 K^2\left(\frac{x-X_j}{h}\right) \omega(x)\, dx.$$

Again, integrating by substitution we obtain

$$Var[F_j] = O\left(h^{2d}\right)$$

and it is straightforward to show that

$$Var[S_n] = O\left(\frac{1}{n^3 h^{2d}}\right).$$

But then, using assumption (H.1) gives

$$Var[S_n] = o\left(\frac{1}{n^2 h^d}\right).$$

This closes the proof.

∎

Lemma 4 *Assume conditions (S.1)-(S.3), (K.1) and (H.1) hold, let*

$$T_n = \int \left(\frac{1}{nh^d}\right)^2 \sum_j \sum_{k \neq j} K\left(\frac{x-X_j}{h}\right) K\left(\frac{x-X_k}{h}\right) \varepsilon_j \varepsilon_k \omega(x)\, dx$$

then

$$\left(nh^{d/2}\right) \frac{T_n}{\sqrt{V}} \longrightarrow_d N(0,1),$$

where

$$V = \left[\int \int K(0) K(u-v)\, du dv\right]^2 \int \sigma^4(u) p^2(u) \omega^2(u)\, du,$$

as n tends to infinity.

Proof of Lemma 4

The proof of this result follows immediately by realizing that T_n has the structure of a U-statistic and it follows the same lines as the proof of result (7.5) in Härdle and Mammen (1993), p. 1942. Then, integrating by substitution and applying assumptions (S.1)-(S.3) we obtain

$$E\left[\int K\left(\frac{x-X_j}{h}\right)K\left(\frac{x-X_k}{h}\right)\varepsilon_j\varepsilon_k\omega\left(x\right)dx\right]^2 = h^{3d}V + o\left(h^{3d}\right).$$

Now, according to Theorem 2.1 in De Jong (1987)

$$\left(nh^{d/2}\right)\frac{T_n}{\sqrt{V}} \longrightarrow_d N\left(0,1\right).$$

Finally note that by independence $E\left(T_n\right)=0$. This closes the proof.

∎

Lemma 5 *Assume conditions (S.2), (C.1), (K.1), (H.1), (H.2) and (L.6) hold, and let*

$$U_n = \int\left[\frac{1}{nh^d}\sum_{j=1}^{n}K\left(\frac{x-X_j}{h}\right)\left(m\left(X_j\right)-\widehat{m_S}\left(X_j\right)\right)\right]^2\omega\left(x\right)dx.$$

Then, under the null hypothesis, H_0,

$$nh^{d/2}U_n \longrightarrow_p 0$$

as n tends to infinity.

Proof of Lemma 5

We rewrite

$$U_n = \frac{1}{n^2h^{2d}}\int\left[\sum_{j=1}^{n}K\left(\frac{x-X_j}{h}\right)\left(m\left(X_j\right)-\widehat{m_S}\left(X_j\right)\right)\right]^2\omega\left(x\right)dx,$$

so it can be seen that

$$\left|U_n\right| \le C\frac{1}{n^2h^{2d}}\sup_x\left|m\left(x\right)-\widehat{m_S}\left(x\right)\right|^2\int\left(\sum_{j=1}^{n}\left|K\left(\frac{x-X_j}{h}\right)\right|\right)^2\omega\left(x\right)dx.$$

Applying (S.2) and a strong law of large numbers we obtain

$$\left|U_n\right| \le C\sup_x\left|m\left(x\right)-\widehat{m_S}\left(x\right)\right|^2\int p^2(x)\omega\left(x\right)dx + o_p(1),$$

and by Theorem 1

$$\left|U_n\right| = O_p\left(\frac{\log n}{nh_1^{d_1}}+\frac{\log n}{nh_2^{d_2}}\right)+o\left(h_1^{2\alpha}+h_2^{2\alpha}\right).$$

Furthermore, applying the conditions on the bandwidths detailed in (H.2) and (L.6) we get

$$\left|U_n\right| = o_p\left(\frac{1}{nh^{d/2}}\right)$$

what closes the proof.

∎

Lemma 6 *Assume that conditions (C.1), (S.1)-(S.3), (B.1), (K.1), (L.6) and (H.1)-(H.2) hold, and let*

$$V_n = \int \left(\frac{1}{nh^d}\right)^2 \sum_{j=1}^{n} \varepsilon_j K^2\left(\frac{x - X_j}{h}\right)\left(m\left(X_j\right) - \widehat{m_S}\left(X_j\right)\right)\omega\left(x\right)dx.$$

Then under the null hypothesis H_0 *one has*

$$nh^{d/2}V_n \longrightarrow_p 0$$

as n *tends to infinity.*

Proof of Lemma 6

We have

$$\left|E\left(V_n\right)\right| \leq \sup_x \left|m\left(x\right) - \widehat{m_S}\left(x\right)\right| \frac{1}{nh^{2d}} \int E\left|\varepsilon_1\right| K^2\left(\frac{x - X_1}{h}\right)\omega\left(x\right)dx.$$

Integrating by substitution gives

$$\left|E\left(V_n\right)\right| = O\left(\sup_x \left|m\left(x\right) - \widehat{m_S}\left(x\right)\right| \frac{1}{nh^d}\right).$$

Using Theorem 1 and assumption (L.6) we obtain

$$\left|E\left(V_n\right)\right| = O\left[\frac{1}{nh^d}\left(\sqrt{\frac{\log n}{nh_1^{d_1}}} + \sqrt{\frac{\log n}{nh_2^{d_2}}}\right)\right] + o\left[\frac{1}{nh^d}\left(h_1^{d_1} + h_2^{d_2}\right)\right],$$

and under conditions (H.1) and (H.2) that relates the rates of the bandwidths h, h_1 and h_2, we have

$$\left|E\left(V_n\right)\right| = o\left(n^{-1}h^{-d/2}\right).$$

For the variance expression it holds

$$Var\left(V_n\right) = \frac{1}{n^4h^{4d}} \sum_{j=1}^{n} Var\left[\int \varepsilon_j\left(m\left(X_j\right) - \widehat{m_S}\left(X_j\right)\right)K^2\left(\frac{x - X_j}{h}\right)\omega\left(x\right)dx\right],$$

$$\left|Var\left(V_n\right)\right| \leq C\left\{\frac{1}{n^4h^{4d}} \sum_{j=1}^{n} Var\left[\int \left|\varepsilon_j\right| K^2\left(\frac{x - X_j}{h}\right)\omega\left(x\right)dx\right]\right\} \times \left(\sup_x \left|m\left(x\right) - \widehat{m_S}\left(x\right)\right|\right)^2.$$

Again, integrating by substitution and using Theorem 1 we obtain the following bound

$$\left|Var\left(V_n\right)\right| \leq C\left\{\frac{1}{n^3h^{2d}}\right\} \times \left\{\frac{\log n}{nh_1^{d_1}} + \frac{\log n}{nh_2^{d_2}}\right\} + o\left(\frac{h_1^{d_1} + h_2^{d_2}}{n^3h^{2d}}\right).$$

Now, by assumptions (H.1), (H.2) and (L.6) it can be verified that

$$\left|Var\left(V_n\right)\right| = o\left(\frac{1}{n^2h^d}\right).$$

This closes the proof.

∎

Lemma 7 *Assume conditions (C.1), (S.1)-(S.3), (K.1), (L.6) and (H.1)-(H.2) hold, and let*

$$W_n = \int \left(\frac{1}{nh^d}\right)^2 \sum_j \sum_{k\neq j} K\left(\frac{x-X_j}{h}\right) K\left(\frac{x-X_k}{h}\right) \varepsilon_k \left(m\left(X_j\right) - \widehat{m_S}\left(X_j\right)\right) \omega\left(x\right) dx.$$

Then, under the null hypothesis, H_0,

$$nh^{d/2} W_n \longrightarrow_p 0$$

as n tends to infinity.

Proof of Lemma 7

We first analyze the bias term. By the law of iterated expectations

$$\begin{aligned} E\left(W_n\right) &= E\left\{\frac{1}{n^2h^{2d}} \int \sum_j \left[\sum_{k\neq j} K\left(\frac{x-X_k}{h}\right) E\left(\varepsilon_k | X_k\right)\right]\right. \\ &\quad \left. \times\ K\left(\frac{x-X_j}{h}\right) \left(m\left(X_j\right) - \widehat{m_S}\left(X_j\right)\right) \omega\left(x\right) dx\right\}. \end{aligned}$$

But since $E\left(\varepsilon_k | X_k\right) = 0$, then $E\left(W_n\right) = 0$. For the variance term we have

$$E\left(W_n^2\right) \leq \left\{\sup_x \left(m\left(x\right) - \widehat{m_S}\left(x\right)\right)^2 \frac{1}{n^4h^{4d}} \sum_{k_1} \sum_{j_1\neq k_1} \sum_{k_2} \sum_{j_2\neq k_2} E\left|F\left(j_1, k_1, j_2, k_2\right)\right|\right\},$$

where

$$\begin{aligned} F\left(j_1, k_1, j_2, k_2\right) &= \int\int K\left(\frac{x_1-X_{j_1}}{h}\right) K\left(\frac{x_1-X_{k_1}}{h}\right) K\left(\frac{x_2-X_{j_2}}{h}\right) K\left(\frac{x_2-X_{k_2}}{h}\right) \\ &\quad \times\ \varepsilon_{j_1}\varepsilon_{j_2}\omega\left(x_1\right)\omega\left(x_2\right) dx_1 dx_2. \end{aligned}$$

By independence we obtain

$$E\left(W_n^2\right) \leq \left\{\sup_{x_1,x_2} \left(m\left(x\right) - \widehat{m_S}\left(x\right)\right)^2 \frac{1}{n^4h^{4d}} \sum_j \sum_{k_1\neq j} \sum_{k_2\neq j} E\left|F\left(j, k_1, j, k_2\right)\right|\right\}. \tag{23}$$

Now we distinguish two cases, when $k_1 = k_2$, then integrating by substitution we get

$$E\left|F\left(j, k_1, j, k_1\right)\right| = O\left(h^{3d}\right), \tag{24}$$

and when $k_1 \neq k_2$, using the same method

$$E\left|F\left(j, k_1, j, k_2\right)\right| = O\left(h^{4d}\right). \tag{25}$$

Finally, if we substitute (24) and (25) into (23) then the following inequality holds

$$E\left(W_n^2\right) \leq \left\{\sup_x \left(m\left(x\right) - \widehat{m_S}\left(x\right)\right)^2 \frac{1}{n^4h^{4d}} \left(n^2h^{3d} + n^3h^{4d}\right)\right\}, \tag{26}$$

and using assumption (H.2) on the bandwidths we get

$$E\left(W_n^2\right) = O\left(\frac{1}{n} \sup_x \left(m\left(x\right) - \widehat{m_S}\left(x\right)\right)^2\right). \tag{27}$$

Then, using assumptions (C.1), (H.1), (H.2) and (L.6) the next inequalities hold

(28)
$$E\left(W_n^2\right) = O\left(\left\{\frac{\log n}{nh_1^{d_1}} + \frac{\log n}{nh_2^{d_2}}\right\}\frac{1}{n}\right) + o\left(\frac{h_1^{d_1} + h_2^{d_2}}{n}\right) = o\left(n^{-2}h^{-d}\right).$$

This closes the proof of the lemma.

∎

Proof of Theorem 2

The test statistic I_n can be decomposed as follows

$$I_n = S_n + T_n + U_n + 2V_n + 2W_n$$

where

$$S_n = \int \left(\frac{1}{nh^d}\right)^2 \sum_{j=1}^{n} K^2\left(\frac{x - X_j}{h}\right)\varepsilon_j^2\omega(x)\,dx,$$

$$T_n = \int \left(\frac{1}{nh^d}\right)^2 \sum_j \sum_{k \neq j} K\left(\frac{x - X_j}{h}\right) K\left(\frac{x - X_k}{h}\right)\varepsilon_j\varepsilon_k\omega(x)\,dx,$$

$$U_n = \int \left[\frac{1}{nh^d}\sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right)\left(m(X_j) - \widehat{m_S}(X_j)\right)\right]^2 \omega(x)\,dx,$$

$$V_n = \int \left(\frac{1}{nh^d}\right)^2 \sum_{j=1}^{n} \varepsilon_j K^2\left(\frac{x - X_j}{h}\right)\left(m(X_j) - \widehat{m_S}(X_j)\right)\omega(x)\,dx,$$

and

$$W_n = \int \left(\frac{1}{nh^d}\right)^2 \sum_j \sum_{k \neq j} K\left(\frac{x - X_j}{h}\right) K\left(\frac{x - X_k}{h}\right)\varepsilon_j\left(m(X_k) - \widehat{m_S}(X_k)\right)\omega(x)\,dx.$$

If the null hypothesis holds then this implies that equality (20) also holds, and Lemma 3 shows that

$$\left(nh^{d/2}\right)S_n = h^{-d/2}B + o_p(1).$$

Applying Lemma 4 we obtain

$$\left(nh^{d/2}\right)T_n \longrightarrow_d N(0, V).$$

Finally, direct applications of Lemmas 5, 6 and 7 give

$$\begin{aligned}\left(nh^{d/2}\right)U_n &\longrightarrow_p 0,\\ \left(nh^{d/2}\right)V_n &\longrightarrow_p 0 \qquad \text{and}\\ \left(nh^{d/2}\right)W_n &\longrightarrow_p 0,\end{aligned}$$

as n tends to infinity. This closes the proof of this result.

∎

To prove Theorem 3 the same decomposition of I_n is used. Certainly now lemmata 5, 6 and 7 will have to be substituted taking into account the alternative H_a. This is done in the following lemmata 8 to 10.

Lemma 8 *Assume that conditions (S.1)-(S.3), (K.1), (L.6) and (H.1)-(H.2) hold, and let*

$$V_n = \int \left(\frac{1}{nh^d}\right)^2 \sum_{j=1}^{n} \varepsilon_j K^2 \left(\frac{x - X_j}{h}\right) (m(X_j) - \widehat{m_S}(X_j))\, \omega(x)\, dx.$$

Then under H_a, and if (A.1)-(A.3) hold

$$nh^{d/2} V_n \longrightarrow_p \infty$$

as n tends to infinity.

Proof of Lemma 8

The proof follows the same arguments as in Lemma 6, but now we analyze the behavior of this term V_n under H_a. Note that H_a implies

$$\int y \ell_{Y/X_1,X_2}(y, x_1, x_2)\, dy = \int yg(y, \eta_1(x_1), \eta_2(x_2), \theta)\, dy + \gamma_n \int y\psi(y, x_1, x_2)\, dy, \tag{29}$$

where

$$\vartheta(x) = \int y\psi(y, x_1, x_2)\, dy$$

is a bounded continuous function on its support because of (A.3) and then

$$|E(V_n)| \leq C\gamma_n \frac{1}{nh^{2d}} \int E|\varepsilon_1| K^2 \left(\frac{x - X_1}{h}\right) \omega(x)\, dx. \tag{30}$$

Integrating by substitution gives

$$|E(V_n)| \leq C\gamma_n \frac{1}{nh^d} + o(1).$$

For the variance expression we get

$$|Var(V_n)| \leq C\gamma_n^2 \left\{ \frac{1}{n^4 h^{4d}} \sum_{j=1}^{n} Var\left[\int |\varepsilon_j| K^2 \left(\frac{x - X_j}{h}\right) \omega(x)\, dx\right] \right\} + o(1). \tag{31}$$

Then from (30) and (31) we obtain

$$V_n = C\left(\gamma_n \frac{1}{nh^d} + \gamma_n \frac{1}{n^{3/2} h^d}\right) + o_p\left(\gamma_n \frac{1}{nh^d} + \gamma_n \frac{1}{n^{3/2} h^d}\right).$$

Apply conditions (A.1) to (A.3) and the proof is done.

∎

Lemma 9 *Assume that conditions (S.1)-(S.3), (K.1), (L.6) and (H.1)-(H.2) hold, and let*

$$U_n = \int \left[\frac{1}{nh^d} \sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right) (m(X_j) - \widehat{m_S}(X_j))\right]^2 \omega(x)\, dx,$$

then, under H_a, and if (A.1)-(A.3) hold then

$$U_n = o_p(V_n)$$

as n tends to infinity.

Proof of Lemma 9

The proof follows the lines of the one for Lemma 5, but now assuming H_a and that consequently expression (29) holds. Then we obtain

$$|U_n| \leq C\frac{1}{n^2h^{2d}}\gamma_n^2 \int \left(\sum_{j=1}^{n}\left|K\left(\frac{x-X_j}{h}\right)\right|\right)^2 \omega\left(x\right)dx.$$

Applying (S.2) and a strong law of large numbers we obtain

$$|U_n| \leq C\gamma_n^2 + o_p(1),$$

and together with condition (A.1) we get

$$U_n = o_p\left(V_n\right).$$

This closes the proof.

∎

Lemma 10 *Assume that conditions (C.1), (S.1)-(S.3), (B.1), (K.1), (L.6) and (H.1)-(H.2) hold, and let*

$$W_n = \int \left(\frac{1}{nh^d}\right)^2 \sum_j \sum_{k\neq j} K\left(\frac{x-X_j}{h}\right) K\left(\frac{x-X_k}{h}\right) \varepsilon_k \left(m\left(X_j\right) - \widehat{m_S}\left(X_j\right)\right) \omega\left(x\right)dx.$$

Then, under H_a, and if (A.1)-(A.3) hold

$$W_n = o_p\left(V_n\right)$$

as n tends to infinity.

Proof of Lemma 10

The proof of this result is based on the proof of Lemma 7. The bias term is $E\left(W_n\right) = 0$. For the variance term proceed as in the proof of Lemma 7 but using alternatively H_a, and therefore equality (29). We obtain

$$E\left(W_n^2\right) \leq \frac{C\gamma_n^2}{n}. \qquad (32)$$

This closes the proof of the lemma.

∎

Proof of Theorem 3

The proof works by following the same lines as in the proof for Theorem 2 by replacing lemmata 5, 6 and 7 now by lemmata 8, 9 and 10.

∎

References

Ai, C. (1997) A Semiparametric Maximum Likelihood Estimator. *Econometrica*, **65**, 933-963.

Ai, C. and X. Chen (1999) Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions. *Under revision in Econometrica*

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

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

Chen, S. (1992) A simple test for single index models. *Manuscript at Princeton University.*

De Jong, P. (1987) A Central Limit Theorem for Generalized Quadratic Forms. *Probability Theory and Related Fields*, **75**, 261-277.

Delgado, M. and W. Gonzalez-Manteiga (2001) Significance Testing in Nonparametric Regression Based on the Bootstrap. *Annals of Statistics.* Accepted for publication.

Eubank, R.L., J.D. Hart, D.G. Simpson and L.A. Stefanski (1995) Testing for additivity in nonparametric regression. *Annals of Statistics*, **23**, 1896-1920.

Fan, J., N.E. Heckman and M.P. Wand (1995) Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *J. Amer. Statist. Assoc.*, **90**, 141-150.

Fan, J. and Q. Li (1996) Consistent model specification tests: omitted variables and semiparametric functional forms. *Econometrica*, **64**, 865-890.

Fan, J. and Q. Li (2000) Consistent model specification tests: kernel-based tests versus Bierens' ICM tests. *Econometric Theory*, **16**, 1016-1041.

Gasser, T., H.-G. Müller and V. Mammitzsch (1985) Kernels for Nonparametric Curve Estimation. *R. Statist. Soc. B.*, **17**, 238-252.

Gonzalez-Manteiga, W. and R. Cao (1993) Testing the hypothesis of a general linear model using nonparametric regression estimation. *Test*, **2**, 161-188.

Gozalo, P.L. and O.B. Linton (2000) Testing additivity in generalized nonparametric regression models. Preprint.

Guerre, E. and P. Lavergne (2000) Minimax Rates for Nonparametric Specification Testing in Regression Models. *Manuscript at INRA, Toulouse.*

Guerre, E. and P. Lavergne (2001) Rate-optimal data-driven specification testing in regression models. *Manuscript. at INRA, Toulouse.*

Härdle, W., S. Huet, E. Mammen and S. Sperlich (2000) Bootstrap Inference in Semiparametric Generalized Additive Models. *Working Paper 00-70, Carlos III de Madrid.*

Härdle, W. and E. Mammen (1993) Comparing Nonparametric versus Parametric Regression Fits. *Annals of Statistics*, **21**, 1926-1947.

Hart, J. (1997) *Nonparametric smoothing and lack-of-fit tests.* New York: Springer Series in Statistics.

Horowitz, J. and V.G. Spokoiny (2001) An Adaptive, Rate-Optimal Test of a Parametric Mean-Regression Model Against a Nonparametric Alternative. *Econometrica*, **69**, 599-632.

Mack, Y.P. and B.W. Silverman (1982) Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **61**, 405-415.

Ripley, B.D. (1987) *Stochastic Simulations.* New York: Wiley Series in Probability and Mathematical Statistics.

Robinson, P.M. (1988) Root-N-consistent semiparametric regression. *Econometrica*, **56**, 931-954.

Rodriguez, D. and T.M. Stoker (1992) Regression Test of Semiparametric Index Model Specification. *Manuscript at M.I.T.*

Rodriguez-Póo, J.M., S. Sperlich and Ph. Vieu (2001) Normalité asymptotique d'estimateurs de maximum de vraisemblance pour modèles non-paramétriques de régression multidimensionelle. *C. R. Acad. Sci. Paris, Série I*, **333**, 61-64.

Serfling, R.J. (1980) *Approximation Theorems of Mathematical Statistics.* New York: Wiley.

Severini, T.A. and J.G. Staniswalis (1994) Quasi-likelihood estimation in semiparametric models. *Journal of the American Statistical Association*, **89**, 501-511.

Severini, T.A. and W.W. Wong (1992) Profile Likelihood and Conditionally Parametric Models. *The Annals of Statistics,* **4**, 1768-1802.

Sperlich, S., D. Tjøstheim and L. Yang (2000) Nonparametric estimation and testing in additive models. *Econometric Theory.* Accepted for publication.

Staniswalis, J.G. (1989) The kernel estimate of a regression function in likelihood-based models. *J. Amer. Statist. Assoc.* **84**, 276-283.

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

Whang, Y. and D.W.K. Andrews (1993) Tests of specification for parametric and semiparametric models. *Journal of Econometrics*, **57**, 277-318.

Yatchew, A. (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**, 263-289.

Zheng, J.X. (2000) A consistent test of conditional parametric distributions. *Econometric Theory*, **16**, 667–691.