SPECIFICATION TESTING WHEN THE NULL IS NONPARAMETRIC OR SEMIPARAMETRIC

Juan M. Rodríguez-Póo, **Stefan Sperlich**, **Philippe Vieu**[*]

December 2, 2016

Abstract

This paper discusses the problem of testing misspecifications in semiparametric regression models for a large family of econometric models under rather general conditions. We focus on mainly three issues that typically arise in econometrics. First, many econometric models are estimated through maximum likelihood or pseudo-ML methods like for example limited dependent variable or gravity models. Second, often one might not want to fully specify the null hypothesis. Instead, one rather wants to impose some important structure like separability. In order to address these points we introduce an adaptive omnibus test. Special emphasis is given on practical issues like bias reduction, adaptive bandwidth choice, general but simple requirements on the estimates, and finite sample performance, including the resampling approximations. [1]

Keywords: Specification test, semiparametric econometrics, adaptive testing, limited dependent variables, separability.

1 Introduction

When estimating structural econometric models, it is quite difficult to find a situation where economic theory or other information from outside the data enables us to fully specify a model without taking the risk of a serious misspecification. On the other hand, to a certain extent (parametric) modeling is wanted by the researcher, or required by the nature of data in order to overcome problems like identification, estimation, interpretation, numerical performance, etc. To manage this trade-off, that is where semiparametric models are made for. They enable us to include available information and necessary restrictions keeping the rest unspecified. The specification imposed is not limited to parametric functional forms; it can be either the separability of inputs, monotonicity or the conditional distribution. The latter for example might be explored to estimate the mean function of limited dependent responses where alternative estimators are either difficult to obtain or are hard to calculate (cf. Lewbel and Linton, 2002).

Although under such semiparametric approaches the risk of misspecification is considerably reduced, the problem is not negligible. For example, within the framework of conditionally parametric models the root-n consistency of the parameters of interest is not obtained for free; it is obtained under stronger conditions including the correctness of all prior information. Consider the so-called Tobit model where a linear structure in the index function might be recommended by economic theory. However, it does not guide us when choosing the conditional distribution of the latent variable. The censored Gaussian density is typically chosen just for convenience. Similarly, the pseudo-ML fails if either the moments of interest (typically the mean) are not correctly specified or the score functions do not constitute a valid set of moment conditions. For example, the Poisson pseudo-ML for gravity models is inconsistent even if there is just a small zero-inflation, i.e. a hurdle for trade.

As a conclusion of the above discussion we concentrate on problems where the main

interest is related to the conditional mean. Regression tests are mainly available for purely parametric null hypotheses. However, economic theory does not fully specify the index plus the conditional distribution. Instead, it might be that it introduces assumptions such as additivity transformations , homothetical (Lewbel and Linton, 2007) or weak separability. In those cases, the null hypothesis to be tested is semiparametric like for example the one of insignificant interactions (Sperlich, Tjøstheim, and Yang, 2002). Therefore we are testing semiparametric null hypotheses that are related to the mean function which is often estimated by (pseudo-) ML-methods. So our null hypothesis is about the mean, not the conditional distribution, though it may be affected by its particular choice.

For testing a parametric regression model against a broad set of alternatives, many specification tests are available. Following Hart (1997), mainly two approaches are considered: statistics that compare parametric vs nonparametric fits, and those that appear as a weighted average of the residuals. The literature on specification testing has been increasing with almost any new volume of an econometrics or statistics journal; see Gonzalez-Manteiga and Crujeiras (2013) for a review. Nonetheless, nonparametric testing is still less common for semiparametric latent variable models (see Pardo-Fernández, Van Keilegom, and González-Manteiga, 2007). Here we introduce a feasible omnibus test where the null hypothesis can be any parametric or semiparametric regression model. The testing target and the test statistic are both original. Moreover, the test is data adaptive giving special emphasize on the calibration (see Sperlich, 2013). In order to provide a broadly applicable procedure, we require the estimates of the null model to fulfill rather general conditions. Many of the semiparametric econometric models are based on kernel smoothing, especially when thinking of profiled likelihood based estimates. We therefore have decided for a test statistic that evaluates kernel-convoluted differences[2]. Using kernels or other smoothing methods, a bandwidth needs to be chosen which regularizes the smoothness of the potential

---

[2]Else, empirical process technologies had to be applied which are based on somewhat different assumptions and therefore test for slightly different alternatives.

alternatives.

Although the problem of bandwidth selection for a test is related to the question of calibration and power, it is often not treated very carefully. The so-called adaptive tests try to balance the need for calibration on the one, and the power maximization on the other hand; see for example Kallenberg and Ledwina (1995), Spokoiny (2001), Härdle, Sperlich, and Spokoiny (2001), or Guerre and Lavergne (2005). We adapt the ideas of the latter three articles to our problem. However, there is an important difference between these mentioned articles and ours: Our null hypothesis is semiparametric whereas theirs are fully parametric. This leads to some conditions on how the smoothing of the null model relates to the one of the alternative in order to obtain a consistent test. Like in the above mentioned papers, and since the asymptotic properties are little helpful in finite samples, to approximate the critical values the inference is done via resampling methods. This reminds us that for a nonparametric test with a semiparametric null hypothesis one has to choose three regularization parameters at least: one for estimating the null, one for the test (referring to the smoothness of the alternative), and one for estimating the critical value. The latter is either the bandwidth for generating bootstrap samples under the null or the size of subsamples when subsampling is applied instead. Note that this triple choice problem is not specific to our problem but typical for all smoothing based tests with a semiparametric null[3]. While the bandwidth for the null model should be chosen along standard criteria for regression estimation (see Köhler, Schindler and Sperlich, 2013), the testing bandwidth should ideally maximize power. This is why we propose an adaptive test which automatically provides such a testing bandwidth. We further discuss methods to select the regularization parameter of the resampling procedure to reach calibration of the test. As we give special emphasis to the practicability of our procedures, we study in detail the resulting performance of our proposals.

Summarizing, we present a practical though general approach that allows for testing

---

[3]You need even be more if for example ideas of Zheng (1996) are used to construct the test.

semiparametric hypothesis that are quite frequently used in the specification of econometric models such as qualitative response models, truncated and Tobit type specifications, duration models, etc. A main theoretical result of this contribution is that a nonparametric test of a semiparametric null hypothesis is considered. It highlights the conditions on the smoothing that have to be performed in the different steps.

2 Proposal of the Test Statistic

Suppose that we have a sample of n independent replicates $\{(Y_i, X_i)\}_{i=1,\cdots,n}$ from the pair of random variables $Y \in \mathbb{R}$, $X \in \mathcal{X}$ being $\mathcal{X}$ a compact set $\mathcal{X} \subset \mathbb{R}^d$, such that the conditional distribution of Y given X is $\ell_{Y|X}(y, x)$. As we already pointed out in the introduction, our interest is focused in testing the correct specification of the regression function

$$m(x) = \int y \, \ell_{Y|X}(y, x) \, dy. \tag{1}$$

There exist a pleiad of consistent regression specification tests that do not depend on a pre-specified choice of $\ell_{Y|X}(y, x)$. However, in some situations of interest for economists, such as multi-index models (Tobit type models, Amemiya, 1985) practitioners prefer to work under standard assumptions such as Gaussian errors of the underlying latent variable model.

If this is the case, the null hypothesis will take the form

$$m_0(x) = \int y \, \ell^0_{Y|X}(y, x; \theta, \eta_1, \cdots, \eta_p) \, dy, \tag{2}$$

with a family of conditional distributions

$$\left\{ \ell^0_{Y|X}(y, x; \theta, \eta_1, \cdots, \eta_p) \; : \; \theta \in \Theta, \; \eta_1(x_1) \in H_1, \cdots, \eta_p(x_p) \in H_p \right\}, \tag{3}$$

where Θ is a compact subset of $\mathbb{R}^k$, $H_1, \cdots, H_p$ are respectively compact subsets in

$\mathbb{R}$, and V is a compact subset in $\mathbb{R}^\nu$. The vectors $x_j \in \mathcal{X}_{d_j}$, $j = 1, \cdots, p$, are mutually exclusive subsets of x such that $\mathcal{X} = \mathcal{X}_{d_1} \times \cdots \times \mathcal{X}_{d_p}$. Further, the η are assumed to be unknown smooth functions $\eta_j : \mathcal{X}_{d_j} \to H_j$ that take values in a set

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

In order to motivate our problem of interest consider the latent regression model

$$y^* = x_1^T \gamma + \eta \left(x_2 \right) + u, \tag{4}$$

where the u_i's are assumed to be random drawings from $N \left(0, \sigma^2 \right)$. γ is an unknown vector of parameters that needs to be estimated and we introduce a nonparametric relationship $\eta \left(\cdot \right)$ that also needs to be estimated. Assume the following mechanism of censoring,

$$y = \begin{cases} y^* & \text{if} \quad y^* > 0 \\ 0 & \text{if} \quad y^* \leq 0 \end{cases}.$$

Then

$$\begin{aligned} \log \ell_{Y|X}^0 \left(y, x; \theta, \eta \right) &= \mathbf{1}_{\{y=y^*\}} \frac{1}{2} \left[-\ln \left(2\pi \right) - \ln \left(\sigma^2 \right) - \left(\frac{y - x_1^T \gamma - \eta \left(x_2 \right)}{\sigma} \right)^2 \right] \\ &\quad + \mathbf{1}_{\{y_i=0\}} \ln \left[1 - \Phi \left(\frac{x_1^T \gamma + \eta \left(x_2 \right)}{\sigma} \right) \right]. \end{aligned}$$

and the regression function takes the form

$$m_0 \left(x \right) = \Phi \left(\frac{x_1^T \gamma + \eta \left(x_2 \right)}{\sigma} \right) \left\{ x_1^T \gamma + \eta \left(x_2 \right) + \sigma \frac{\phi \left(\frac{x_1^T \gamma + \eta(x_2)}{\sigma} \right)}{1 - \Phi \left(\frac{x_1^T \gamma + \eta(x_2)}{\sigma} \right)} \right\},$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ stand respectively for the pdf and the Gaussian distribution function. Note that for the null hypothesis to be true, we do not only need a correct specification of (4). We also need correct specification of both, the censoring mechanism, the conditional distribution and the homoskedasticity assumption. Of course

there are situations such as non-existing moments, quantiles or other robust quantities where focusing on the likelihood would be more natural. In many cases the estimator for $m(x)$ under the null can be calculated directly from some consistent estimators $\hat{\theta}$ and $\hat{\eta}_1, \cdots, \hat{\eta}_p$, and can generally be written as

$$\widehat{m}_S(x) = \int y \, \ell^0_{Y|X}\left(y, x; \hat{\theta}, \hat{\eta}_1, \cdots, \hat{\eta}_p\right) dy.$$

We will specify the above mentioned consistent estimators in assumption (C.1). Let $\omega(u) \geq 0$ be a bounded weight function. Then the testing problem can be written as

$$H_0 : m(x) = m_0(x) \quad \forall x \in \mathcal{X} \text{ for which } \omega(x) > 0,$$

$$H_1 : m(x) \neq m_0(x) \quad \text{with} \quad x \in \mathcal{X} \text{ for which } \omega(x) > 0.$$

Several alternative testing approaches are thinkable. Several comparison studies (Dette, von Lieres und Wilkau and Sperlich, 200; Roca-Pardiñas and Sperlich, 2007) show that a quite successful one is to construct a statistic with convoluted differences in order to handle the bias problem. For kernel $K : \mathbb{R}^d \to \mathbb{R}$, bandwidth h, define

$$I_h\left(\hat{\eta}_1, \hat{\eta}_2, \cdots, \hat{\eta}_p, \hat{\theta}\right) = \int \left[\frac{1}{nh^d} \sum_{j=1}^n K\left(\frac{x - X_j}{h}\right)(Y_j - \widehat{m}_S(X_j))\right]^2 \omega(x) \, dx. \quad (5)$$

The statistical properties of this statistic depend both on the choice of the bandwidth h, and the asymptotic behavior of $\hat{\theta}$ and $\hat{\eta}_1, \cdots, \hat{\eta}_p$. On these grounds, it is important to discuss how the smoothing parameters and the estimators are chosen. With respect to the estimators, for the parametric part θ, we require a root-n consistent estimator. In this way, the parametric estimator does not affect the asymptotic distribution of the test statistic. This may happen with other estimators that exhibit a slower rate of convergence. For the nonparametric estimators, only one minimal requirement of uniform convergence at a certain nonparametric rate, r, is assumed.

(C.1) The function $\ell^0_{Y|X}$ is Lipschitz continuous in all its arguments. Under H_0 we

have

$$\sup_{x_j \in \mathcal{X}_j} |\widehat{\eta}_j(x_j) - \eta_j(x_j)| = O_p\left(n^{-r_j}(\log n)^{r_j}\right)$$

for $\frac{1}{4} \leq r_j < \frac{1}{2}$, $j = 1, \cdots, p$, and $\hat{\theta} = \theta + O_p\left(n^{-1/2}\right)$, as n tends to infinity.

This requirement is rather general since it is fulfilled by most part of nonparametric estimators that appear in the econometric literature. Furthermore, by imposing this condition we avoid that the statistical properties of the test neither depend on the form of the estimator for the η's nor in their bandwidth parameters. As it was already established in Stone (1982), the optimal rate of convergence, r_j, depends on the smoothness class in which η_j is assumed to be, the dimension of x_j or the type of deviation function we are choosing. More precisely, the uniform maximal deviation converges slowly to zero by a factor $\log n$ (see Mack and Silverman, 1982, section 3). Note that r_j is strictly smaller than $1/2$ as it is also expected in the nonparametric literature. For example, if the second derivative of η_j is Hölder continuous and the dimension of x_j is d_j, the optimal rate will take the value $r_j = \frac{2}{4+d_j}$. We also impose that $r_j \geq \frac{1}{4}$. That is, the rate of convergence does not have to be too slow. The same condition is also imposed in Severini and Wong (1992) for the nonparametric estimator that is included in the conditionally parametric model. For an example of estimator for η_j that fulfils these conditions see Rodriguez-Poo, Sperlich and Vieu (2013), Section 5.

The choice of bandwidth h in (5) should ideally maximize the power of the test. As the distribution of I_h will vary with h, it is natural to consider the standardized version

$$T_h = \frac{nh^{d/2}I_h\left(\hat{\eta}_1, \hat{\eta}_2, \cdots, \hat{\eta}_p, \hat{\theta}\right) - h^{-d/2}\hat{B}}{\sqrt{\hat{V}}}, \tag{6}$$

where $\hat{B}$ is an estimator of the expectation of $nh^{d/2}I_h$ under H_0,

$$B = \int K^2(u)du \int \sigma^2(u)p(u)\omega(u)du, \tag{7}$$

and $\hat{V}$ an estimator of the variance of $nh^{d/2}I_n$, i.e.

$$V = \int K^{(2)}(u)^2 du \int \sigma^4(u)p^2(u)\omega^2(u)du. \tag{8}$$

Here, we denote by $K^{(2)}$ the convolution of kernel K, p is the marginal density of X, and $\sigma^2(x)$ is the conditional variance of Y given X. The standardization (6) creates a family of test statistics $\{T_h, \ h \in H_n\}$, where the choice of h marks the difference between the null and the global alternative. Following the idea of adaptive testing (Spokoiny, 2001) we propose to use test statistic

$$T^* = \max_{h \in H_n} T_h \tag{9}$$

where the value of h is taken from a set of bandwidths H_n with cardinality J_n, namely

$$H_n = \{h = h_{\max}a^k : h \geq h_{\min}, \ k = 0, 1, 2, \cdots \}, \quad 0 < h_{\min} < h_{\max}, \quad 0 < a < 1. \tag{10}$$

Another relevant issue in the computation of the test statistic is the estimation of $\sigma^2(x)$. Note that this variance expression appears in (7) and (8). Horowitz and Spokoiny (2001) propose an estimator for $\sigma^2(x)$ that needs to be consistent under the alternative. Often such an estimator is not easily available or it is only under the costs of further restrictions. The difference estimator proposed by them works only reasonably well in one dimensional regression problems. Here, we do not need such a strong condition but instead we just ask for an estimator of $\sigma^2(x)$ fulfilling the following assumption

(C.2) $\sup_x |\hat{\sigma}^2(x)\hat{p}(x) - \sigma^2(x)p(x)| = o_p\left(h^{\frac{\mu d}{2(2+\mu)}}/\log n\right)$, for some $\mu > 0$ and for all $h \in H_n$.

Guerre and Lavergne (2005) discuss some drawbacks of the Horowitz-Spokoiny approach and propose an alternative choice of h that will be considered in Section 4 together with its practical implementation.

An analytical expression for the critical values of T^* needs extreme value theory. As this is cumbersome and it gives little helpful approximations in practice, we propose resampling strategies. No matter whether (wild) bootstrap or subsampling is applied, for the resampling estimates, say $\hat{m}_S(x)$ and $\hat{\sigma}^2(x)$, we ask for similar conditions as in Horowitz and Spokoiny (2001), that is:

(C.3) As n tends to infinity, conditionally on the sample $\{x_1, x_2, \ldots, x_n\}$

$$\sup_x \left| \hat{m}_S(x) - \hat{\hat{m}}_S(x) \right| = O_p \left(|m(x) - \hat{m}_S(x)| \right),$$

$$\sup_x \left| \hat{\sigma}^2(x) - \hat{\hat{\sigma}}^2(x) \right| = O_p \left(|\sigma^2(x) - \hat{\sigma}^2(x)| \right).$$

Note that assumption (C.3) is rather general. In order to be more precise about the consistency of the resampling estimates we have to be more specific about the model and the estimators considered, loosing thereby the pretended generality. For the consistency of related bootstrap problems see for example Giné and Zinn (1990) or Härdle, Huet, Mammen, and Sperlich (2004). The alternative subsampling (Politis, Romano, and Wolf, 1999) procedure works similarly. One draws subsamples of size $n_s < n$ from the original sample $\{X_i, Y_i\}_{i=1}^n$. Take $b = h \, (n/n_s)^r$ as the bandwidth for the subsampling estimator to calculate the subsample analogues, $\hat{T}^*$. r is the chosen bandwidth rate. Finally, compute the empirical $1 - \alpha$ quantile as above. In cases where H_0 is violated, as n tends to infinity, $nh^{d/2}I_h$ grows faster than $n_s b^{d/2} I_b^*$, and therefore $\hat{T}^*$ will become larger than T^* resulting in a rejection of H_0. General results on asymptotic theory can be found in Politis, Romano, and Wolf (2001). To our knowledge, the choice of optimal subsample size n_s for non- and semiparametric specification testing has only been treated by Delgado, Rodriguez-Poó, and Wolf (2001), and Neumeyer and Sperlich (2006).

One may also estimate $\hat{B}$, $\hat{V}$, and this way implicitly also $\sigma^2(x)$ by resampling. For the bootstrap we generate samples $\{X_i, Y_i^*\}_{i=1}^n$ with Y_i^* drawn from the estimated

null distribution, and calculate $\hat{T}^*$ like before T^*. Repeating this many times we can calculate the empirical $1 - \alpha$ quantile of $\hat{T}^*$, t_α, and use it as critical value for T^*.

3   Asymptotic Behavior

For the ease of notation but without loss of generality we set $p = 2$ for the rest of the paper. Recall that level and power of our test will depend on the theoretical properties of $\hat{\eta}_1$ and $\hat{\eta}_2$ ($p = 2$), appearing in the test through $\widehat{m_S}(x)$. Therefore, we first analyze the properties of this regression function estimator and afterward the test. We use the following assumption on Y:

(B.1) For some $s > 0$, $E|Y|^s < \infty$ and $\int |Y|^s \, dy = M_s < \infty$.

The first result is a crucial tool when studying our test statistic. Note that by having considering a separable model we avoid the curse of dimensionality under the null.

Lemma 1. *Under conditions (C.1) and (B.1), if the null hypothesis H_0 is true, for n tending to infinity,*

$$\sup_{x \in \mathcal{X}} |\widehat{m_S}(x) - m(x)| = O_p\left(b_n n^{-r_1}(\log n)^{r_1} + b_n n^{-r_2}(\log n)^{r_2}\right) + O_p\left(b_n n^{-1/2}\right).$$

Where b_n is an increasing sequence for which $\sum_n b_n^{-s}$ converges.

To study the properties of statistic T^* we need the following additional conditions:

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

(S.2) Densities p and p_j of X_j, $j = 1, 2$, are β-times continuously differentiable on $\mathcal{X}_j$.

(S.3) Let $\alpha(x)_{2+\mu} = E\left(Y^{2+\mu}| X = x\right)$. The function $\alpha(x)_{2+\mu}$ is bounded above and below, for $\mu = 0$ and for some $\mu > 0$ in any place in the support of $\mathcal{X}$.

(K.1) K is a compactly supported and continuously differentiable kernel of order β satisfying $\int K(u)du = 1$.

These are standard conditions on kernel and smoothness of densities and conditional expectations. The next set of conditions defines the set of bandwidths for the test statistic and relates it to the smoothing applied when estimating the null model $\widehat{m_S}$.

(H.1) The set H_n of bandwidths has the structure (10) with $h_{max} > h_{min} \geq n^{-\frac{1}{d}}$ and h_{max} is at most $C_H n^{\frac{2}{d}(2r-1)} b_n^2 (\log n)^{-2r}$ for some finite constant $C_H > 0$, and $r = \min\{r_1, r_2\}$. Furthermore, $J_n \sim \log n$ as n tends to infinity.

Assumption (H.1) is similar to Horowitz and Spokoiny (2001). However, there is an important difference. The fact that we have a non- or semiparametric null hypothesis increases the bias problem. Often, the bias problem has been solved by smoothing the (parametric) estimator under the null hypothesis, too. In our case, we smooth the function estimate under the null which already shows nonparametric rates of convergence, recall Section 4. To control for this problem the upper bound of the set H_n depends on the rate of convergence of $\hat{m}_S(x)$. In the case $r = \frac{1}{2}$, that is the null hypothesis is fully parametric, the upper bound coincides with that assumed by Horowitz and Spokoiny (2001). However, as it is in our case, if the estimator under the null exhibits a nonparametric rate, then h_{max} must tend to zero at a rate that adapts to the convergence rate of the estimator. As usual, conditions (S.1), (S.2), (K.1) take care for the reduction of the higher order terms of the bias. Then, the asymptotic behavior of the test under H_0 is given by

Theorem 1. *Assume that conditions (C.1)-(C.3), (S.1)-(S.3), (B.1), (K.1) and (H.1) hold. If the null hypothesis H_0 is true, then for n tending to infinity*

$$P\left(T^* > t_\alpha\right) \to \alpha.$$

In order to determine the power, we define a sequence of local alternatives.

$$H_a : m(x_1, x_2) \;=\; m_0(x_1, x_2) + \gamma_n \int \psi\,(y, x_1, x_2)\, dy \;, \tag{11}$$

where function $\psi(\cdot)$, not depending on n, shall describe the deviation from the conditional density under the null, respectively from the mean function, and the sequence γ_n is supposed to be such that:

(A.1) For all $h \in H_n$, $\gamma_n \sqrt{nh^{d/2}} \longrightarrow \infty$.

(A.2) $\psi(\cdot)$ is continuous and $\vartheta(x) := \int y\, \psi\,(y, x_1, x_2)\, dy \neq 0$.

It is clear that assumptions (A.1) and (A.2) on the one had are also related to the bias problem that has been mentioned above. On the other hand they also guarantee the detection of local alternatives by our statistic, i.e. consistency of the test. Note that conditions (A.1) and (H.1) do not entail mutual constraints.

Theorem 2. *Assume that conditions (C.2)-(C.3), (S.1)-(S.3), (K.1) and (H.1) hold. If the alternative H_a is true with (A.1) and (A.2), then we have for $n \to \infty$ that*

$$P\,(T^* > t_\alpha) \to 1.$$

Theorem 2 says 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 $\sqrt{n}$. It is known that 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 discussed in Guerre and Lavergne (2002) or Horowitz and Spokoiny (2001), at an exact rate of $\sqrt{n}$ no omnibus test can have nontrivial power uniformly over non-artificial classes of functions $\psi(\cdot)$, respectively $\vartheta(x)$ in (11). Moreover, what we can see, recall (A.1), is that for γ_n the optimal rate is $o(\frac{1}{\sqrt{nh^{d/2}}})$ as expected, cf. Guerre and Lavergne

(2002). How fast this can be, this depends obviously on h, and the fastest rate therefore on h_{max}. Concluding, we can see from (A.1) and (H.1) how the rate slows down for semi- and nonparametric (separable) null hypotheses compared to testing problems where the null is parametric.

4   Finite Sample Performance and an alternative Bandwidth Choice

To illustrate the performance of our testing procedure we present simulation results for different models with censored responses:

$$
y = \begin{cases} \kappa(x) + u & \text{if } \kappa(x) + u > 0 \\ 0 & \text{otherwise} \end{cases} , \tag{12}
$$

where $x = (x_1, x_2)^T$ with $X_1 \sim U[0,2]^2$, $X_2 \sim U[-1,1]^q$ and i.i.d. errors u. Under H_0 we suppose $u \sim N(0, \sigma^2)$ and

$$
\kappa_0(x) = x_1^T \gamma + \eta(x_2), \qquad x_1 = (x_{11}, x_{12})^T \tag{13}
$$

where γ, $\eta(\cdot)$, and error variance σ^2 are arbitrary. Then we have

$$
E[y|x] = \Phi(\kappa_0(x)/\sigma) \left\{ \kappa_0(x) + \sigma\lambda(\kappa_0(x)/\sigma) \right\}, \quad \lambda(v) = \phi(v)\Phi^{-1}(v) \tag{14}
$$

with ϕ and Φ indicating the normal density and its cumulated distribution function. The different real data generating processes (DGP) are as follows:

$$
\begin{aligned}
\kappa_1(x) &= x_1^T \gamma + \sin(2.5x_2), & \gamma = (-1, 1.5)^T , & \tag{15} \\
\kappa_2(x) &= \kappa_1(x) + (2x_{11}x_{12} - 2) , & & \tag{16} \\
\kappa_3(x) &= \kappa_1(x) + 2(x_{11} - 1)^2 - 2/3 , & & \tag{17}
\end{aligned}
$$

all with an additional $N(0, 0.5^2)$ error term u. Further, we generate data as in (15) but replacing error u by

$$u^* \sim (\chi_1^2 - 1) \cdot 8^{-1/2} , \text{ and } u^{**} \sim U[-\frac{\sqrt{3}}{2}, \frac{\sqrt{3}}{2}] \text{ respectively.} \qquad (18)$$

So the error variance is always 0.5^2. Finally, we also consider models with $x_2 = (x_{21}, x_{22}) \in I\!\!R^2$, $X_2 \sim U[-1, 1]^2$, and

$$\kappa_4(x) = x_1^T \gamma + x_{21} x_{22}, \qquad \gamma = (-1, 1.5)^T \qquad (19)$$

$$\kappa_5(x) = \kappa_4(x) + (x_{11} - 1)^2 - 1/3 . \qquad (20)$$

Note that among (15) to (20), models (15) and (19) are the only ones that belong to the null hypothesis H_0, compare (13). For the other models, the additional terms (deviating from the null model) are centered to zero such that they do not affect the censoring threshold in (12). We simulated also models without such centering (not shown here). In those cases the power of the test is always overwhelming. We concentrate here on the presentation of the cases when detection of an alternative is a hard problem. Notice that the variation due to the deviation from H_0 (i.e. $E[\kappa_0(X) - \kappa_j(X)]^2$, $j = 1, 2$) is 28/9 in (16), whereas only 16/45 in (17). Recall that we only worry about the effect of a misspecification on the regression estimation.

For the simulations we use only 250 bootstrap samples, respectively subsamples. In empirical research one should certainly take more. Having consistent (under H_0) estimates of κ_0 and σ^2 (even root-n of the latter), the bootstrap sample can be generated by $Y_i^* = max \left\{ 0, \widehat{\kappa_0}(x_i) + N(0, \widehat{\sigma^2}) \right\}$. This is actually a bootstrap from a known parametric distribution, i.e. the normal censored at zero, with semiparametric first and parametric second moment. The estimation of the model under the null hypothesis is done accordingly to Rodríguez-Póo, Sperlich and Vieu (2003). The percentages of rejections under the null are calculated out of 500 simulation runs whereas only 250 are used to approximate the rejection levels under H_1. As we use uniform distributions

for all covariates in this simulation, we set $w(x) = 1$ for the entire simulation study. An else obvious choice is either a trimming weight or the density of X such that the integral can be replaced by the sample average.

As has been discussed earlier in this paper, there exist some alternative approaches to make a nonparametric test adaptive in the sense of choosing the test bandwidth h adaptively. Among them the approach of Guerre and Lavergne (2005) seems to be a quite promising one, but they yield their excellent results applying the method to a rather different test statistic and problem. We briefly explain how we modified and implemented their idea for our statistic T_h, respectively I_h. Since they supposed to dispose of a centered (under H_0) test statistic, let us consider $I_h^c := \{I_h - E_0[I_h]\}$, where E_0 refers to the expectation under H_0. Then, select h by

$$h_{GL} = arg \max_{h \in H_n} \left(I_h^c - 2\sqrt{2 \ln J_n} Var_0 \left[I_h^c - I_{h_{max}}^c \right] \right), \tag{21}$$

where Var_0 refers to the variance under H_0. In our context h_{max} is the largest bandwidth in H_n, cf. Guerre and Lavergne (2005). The final test statistic is $I_{h_{GL}}^c / Var_0 \left[I_{h_{max}}^c \right]$. For each bandwidth h, the bootstrap can be used to approximate $E_0[I_h]$, $Var_0 \left[I_h^c - I_{h_{max}}^c \right]$, and $Var_0[I_h^c]$. This is what we have done in our simulation study. In our context this approach gives basically the same results for all simulated models and rejection levels. We conclude therefore that both approaches seem to be reasonable alternatives for the choice of the testing bandwidth h in (5) albeit Guerre and Lavergne (2005) showed some theoretical advantages of their approach. In practice this might vary for different testing problems, models and implementations.

The bandwidths h_j from (C.1) to estimate the nonparametric part $\eta(\cdot)$ of the null model (13) can be chosen by jackknife or generalized cross validation, depending on the particular estimation problem. This gives bandwidths that produce slightly under-smoothed estimates. Trying out various bandwidths for each h_j on a reasonable range we always got rather similar results. This might be due to the adaptive selection of

h, determined by a naive grid search with $J_n = 15$ and $a = 1.05$, cf. (10). Let us call h_j^* the bandwidths to (pre-)estimate the null model from which the bootstrap samples will be drawn from. The h_j^* should be chosen depending on h_j to fulfill Assumption (C.3) but in practice are often set just equal to h_j.

In Table 1 and Table 2 are given the rejection levels and bootstrap p-values for models (15) to (18) for sample size $n = 200$ applying bootstrap with $h_j = 1.75\sigma_x/n^{1/5}$, $h_j^* = h_j$ and $h_j^* = h_j n^{1/5-1/6}$ respectively. Here, σ_x denotes the vector of sample standard deviation(s) of X_2. We used second order quartic kernels, and we set $h_{max} = 3\sigma_x/n^{1/7}$ in (10). Table 1 gives the results when using the $h-$adaptive test of Horowitz and Spokoiny (2001); Table 2 the results when applying the $h-$adaptive method of Guerre and Lavergne (2003). The bootstrap p-values are the (average) percentage of bootstrap statistics which have been larger than the original test statistic. First, we see that the test almost holds the level under H_0 for $h_j^* = h_j$ but is a bit less reliable if $h_j^* = h_j n^{1/5-1/6}$. The differences between the two bandwidth adaptive methods is minor what is in accordance with the simulation results of Guerre and Lavergne (2005) for more 'regular' (or say 'smooth') alternatives. The test detects clearly when the functional form of the conditional expectation is misspecified. Even though the deviation from H_0 is much larger in (16), our test detects more easily the one in model (17). One reason could be that model (17) is smoother and thus the alternative gets estimated more easily; another reason could be that the bootstrap samples were generated with the residuals from H_0. When the functional form is specified correctly and only the error distribution deviates from H_0, the test does not reject as long as this misspecification has no effect on the regression estimation. In Table 3 we see that (18), u_i^{**}, has no effect on the estimation of the parametric part. Nevertheless, we find nontrivial power for the case of asymmetric error distribution, u^*, even though it affects the regression estimation only very mildly. Actually, when we increase n to 400, at the nominal 5% significance level we reject in about 7.5%, and for $n = 500$ in about 10% of all cases.

DGP	Horowitz-Spokoiny bootstrap $p-$value	significance levels 1%	5%	10%	15%	Guerre-Lavergne bootstrap $p-$value	significance levels 1%	5%	10%	15%
(15), H_0	.452	.012	.058	.120	.170	.469	.012	.054	.096	.162
(16)	.005	.896	.972	.992	.992	.005	.896	.972	.992	.992
(17)	.000	1.00	1.00	1.00	1.00	.000	1.00	1.00	1.00	1.00
(18), u_i^*	.413	.028	.072	.144	.224	.413	.028	.072	.144	.224
(18), u_i^{**}	.433	.012	.068	.124	.180	.433	.012	.068	.124	.180

Table 1: The bootstrap p-values and percentages of rejections at various significance levels for different models when sample size is $n = 200$ when $h_j^* = h_j$.

DGP	Horowitz-Spokoiny bootstrap p-value	significance levels 1%	5%	10%	15%	Guerre-Lavergne bootstrap p-value	significance levels 1%	5%	10%	15%
(15), H_0	.435	.008	.072	.132	.184	.459	.012	.054	.096	.162
(16)	.009	.896	.968	.980	.984	.009	.896	.968	.980	.984
(17)	.000	1.00	1.00	1.00	1.00	.000	1.00	1.00	1.00	1.00
(18), u_i^*	.421	.024	.080	.132	.188	.421	.024	.080	.132	.188
(18), u_i^{**}	.459	.008	.048	.096	.144	.459	.008	.048	.096	.144

Table 2: The bootstrap p-values and percentages of rejections at various significance levels for different models when sample size is $n = 200$ when $h_j^* = h_j n^{1/5 - 1/6}$.

When we consider models with multidimensional nonparametric parts (speaking of nonparametric parts that cannot be decomposed into lower dimensional separable components), then much larger samples are required to make the asymptotics work. For models like (19) and (20), 1000 or more observations are necessary. The problem is that for a semiparametric null hypothesis the variance decomposition does not work sufficiently well to make the bootstrap work. This easily leads to tests that are either to liberal or have almost trivial power. In most of the literature it is just proposed to choose bandwidths h_j^* *adequately*. So far there do not exist results nor clear guidelines for its practical choice. In Figure 1 we see how the bootstrap rejection levels converge for model (19) to the nominal size (α) for increasing n when using $h_j = 1.75\sigma_x/n^{1/9}$,

true coef.	estimates in model (DGP) (15), H_0	(16)	(17)	(18), u_i^*	(18), u_i^{**}
-1.0	-1.000 (.067)	1.266 (.541)	-1.033 (.119)	-1.011 (.067)	-1.001 (.060)
1.5	1.505 (.069)	3.525 (.725)	1.511 (.103)	1.525 (.068)	1.495 (.064)

Table 3: Average and standard deviation (in parentheses) of the parameter estimates for the different models when sample size is $n = 200$.

$h_j^* = h_j n^{1/9-1/13}$, and a 4th order optimal product kernel. Note that we tried many other bandwidth combinations: However, for moderate sample sizes we can not get rid of the size problem for any reasonable bandwidth. A detailed discussion of the calibration problem in nonparametric testing is given in Sperlich (2013). The power of the test seems to be rather strong now but clearly, this may be misleading due to the size problem.
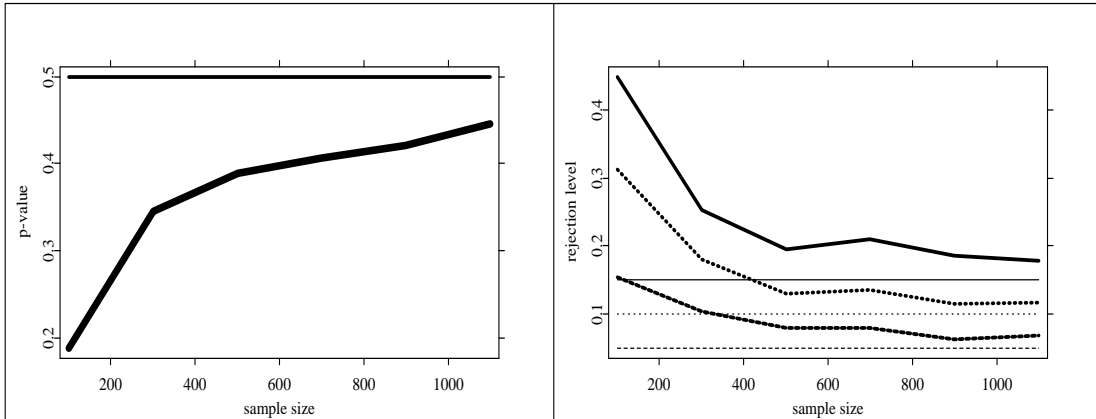

Figure 1: Bootstrap p-values (left) and rejection levels (right) in model (19) for increasing n. Straight lines indicate the 50%-line (left), and the nominal significance levels (right). Dashed lines refer to the 5%, dotted to the 10%, and solid to the 15% levels.

There exists the subsampling as an alternative resampling method. This - at least in our simulations - has turned out to be rather reliable concerning the size of the test, even for small samples and certainly with some loss of power. As the problem of finding the optimal subsample size in nonparametric testing is known (see e.g. Neumeyer and Sperlich, 2006), we concentrate here on size and power of the test but take the subsample size as given.

In Table 4 are given the number of rejections when $n = 300$, subsample size $n_s = 250$, bandwidth $h_1 = h_2 = 2\sigma_x/n^{1/9}$ with 4th order optimal product kernels, and $h_{max} = 3 \cdot stdev(X) \cdot n^{-1/8}$ in (10). Again we tried several bandwidths, and again the results do not vary much as long as h_1, h_2 are providing a reasonable smoothing, i.e. do not strongly over- or undersmooth. Taking a cross validation bandwidth is again a good choice. Table 4 shows that even for small samples and multidimensional nonparametric

parts, subsampling achieves both holding the size (being conservative for our subsample size) and obtaining nontrivial power.

DGP	$p-$value	wanted significance level			
		1%	5%	10%	15%
(19), H_0	.388	.002	.026	.052	.104
(20)	.049	.385	.645	.817	.900

Table 4: The subsampling $p-$values and percentages of rejections at various significant levels for different models when sample size is $n = 300$, and subsample size $n_s = 250$.

5 Appendix

5.1 Proof of Lemma 1

For the proof of Lemma 1 we first need to establish the following proposition:

Proposition 1. *Assume that condition (C.1) holds. Under H_0 we have*

$$\sup_{y} \sup_{x_1 \in \mathcal{X}_1} \sup_{x_2 \in \mathcal{X}_2} \sup_{\theta \in \Theta} \left| \ell^0_{Y|X} \left(y, x_1, x_2; \hat{\theta}, \widehat{\eta}_1, \widehat{\eta}_2 \right) - \ell^0_{Y|X} \left(y, x_1, x_2; \theta, \eta_1, \eta_2 \right) \right|$$

$$= O_p \left(n^{-r_1} (\log n)^{r_1} + n^{-r_2} (\log n)^{r_2} \right) + O_p \left(n^{-1/2} \right)$$

as n tends to infinity.

Proof of Proposition 1. Since by assumption (C.1), $\ell_{Y|X}$ is differentiable we have

$$\left| \ell^0_{Y|X} \left(y, x_1, x_2; \hat{\theta}, \widehat{\eta}_1, \widehat{\eta}_2 \right) - \ell^0_{Y|X} \left(y, x_1, x_2; \theta, \eta_1, \eta_2 \right) \right|$$

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

This is uniform because $\mathcal{X}_1$, $\mathcal{X}_2$ and Θ are compact sets. So condition (C.1) applies. ∎

Proof of Lemma 1. The proof of this result will be a direct consequence of Proposition 1 since

$$m(x) = \int y \ell_{Y|X}(y, x_1, x_2)\, dy.$$

Thus we can write

$$|m(x) - \widehat{m_S}(x)| = \left| \int y \left(\ell_{Y|X}(y, x_1, x_2) - \ell^0_{Y|X}\left(y, x_1, x_2; \hat{\theta}, \hat{\eta}_1(x_1), \hat{\eta}_2(x_2)\right) \right) \right|. \quad (22)$$

Furthermore, under the null hypothesis we have

$$\int y \ell_{Y|X}(y, x_1, x_2)\, dy = \int y \ell^0_{Y|X}(y, x_1, x_2; \theta, \eta_1(x_1), \eta_2(x_2))\, dy \quad (23)$$

Thus, the left hand side of (22) is equal to

$$\left| \int y \left(\ell^0_{Y|X}(y, x_1, x_2; \theta, \eta_1(x_1), \eta_2(x_2)) - \ell^0_{Y|X}\left(y, x_1, x_2; \hat{\theta}, \hat{\eta}_1(x_1), \hat{\eta}_2(x_2)\right) \right) dy \right|. \quad (24)$$

Note that (24) is bounded by the following term

$$\left| \int y\, dy \right| \sup_{y} \sup_{x_1 \in \mathcal{X}_1} \sup_{x_2 \in \mathcal{X}_2} \sup_{\theta \in \Theta} \left| \ell^0_{Y|X}\left(y, x_1, x_2; \hat{\theta}, \widehat{\eta}_1(x_1), \widehat{\eta}_2(x_2)\right) - \ell^0_{Y|X}(y, x_1, x_2; \theta, \eta_1(x_1), \eta_2(x_2)) \right|.$$

Noting that since $P\left(|y_n| > b_n\right) \le b_n^{-s} E\,|y|^s$, it follows with probability one that $|y_n| \le b_n$ for all sufficiently large n, and hence since b_n is increasing,

$$|y_j| \le b_n, \quad \text{for all} \quad j \le n,$$

therefore

$$\int_{|y| < b_n} |y|\, dy = O\left(b_n\right).$$

Now, using the fact that by a standard argument,

$$\int_{|y| \ge b_n} |y|\, dy \le M_s b_b^{-s}, \quad (25)$$

21

it follows that (25) is bounded by $O\left(b_b^{-s}\right)$ with probability one. Apply now Proposition 1 and the proof is done.

∎

5.2 Proof of Theorem 1

Before proving Theorem 1 we need first to state a few technical results. This is done in lemmas 2-7 below. In the following, as above, double hat refers to estimators from the re-samples like e.g. $\hat{\hat{\eta}}_1$.

Lemma 2. *Assume conditions (S.1)-(S.2), (B.1), (C.1), (K.1) and (H.1) 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, for some $\mu > 0$, under the null hypothesis, H_0,

$$E\left[U_n\right]^{2+\mu} = o\left(\frac{1}{n^{2+\mu}h^{\frac{d}{2}(2+\mu)}}\right)$$

as n tends to infinity.

Proof of Lemma 2. We rewrite $E\left[U_n\right]^{2+\mu} =$

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

so it can be seen that $E\left[U_n\right]^{2+\mu} \leq$

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

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

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

and by Lemma 1

$$E\left[U_n\right]^{2+\mu} = O\left(n^{-(2+\mu)r}\left(\log n\right)^{(2+\mu)r} b_n^{2+\mu}\right),$$

where recall that $r = \max\{r_1, r_2\}$. Furthermore, since $h \in H_n$, by condition (H.1)

$$E\left[U_n\right]^{2+\mu} = o\left(\frac{1}{n^{2+\mu}h^{\frac{d}{2}(2+\mu)}}\right)$$

what closes the proof. ∎

Now, for the following we define $\varepsilon_j = Y_j - m(X_j)$.

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

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

Then, for some $\mu > 0$ under the null hypothesis H_0 one has

$$E\left[V_n\right]^{2+\mu} = o\left(\frac{1}{n^{2+\mu}h^{\frac{d}{2}(2+\mu)}}\right)$$

as n tends to infinity.

Proof of Lemma 3. We have

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

Integrating by substitution and using a standard inequality for expectations gives

$$E\left[V_n\right]^{2+\mu} \leq \left(\frac{1}{nh^{d/2}}\right)^{4+2\mu} \left(\sup_x \left|m\left(x\right) - \widehat{m_S}\left(x\right)\right|\right)^{2+\mu} \left(\int K^2(u)du\right)^{2+\mu} \sum_{j=1}^{n} E\left[\varepsilon_j \omega\left(X_j\right)\right]^{2+\mu}.$$

Using Lemma 1 and assumptions (S.3) and (H.1) and recalling that $h \in H_n$ we obtain

$$E\left[V_n\right]^{2+\mu} = o\left(\frac{1}{n^{2+\mu}h^{\frac{d}{2}(2+\mu)}}\right).$$

This closes the proof. ∎

Lemma 4. *Assume conditions (C.1), (S.1)-(S.3), (K.1), (L.6) and (H.1) 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,

$$E\left[W_n\right]^{2+\mu} = o\left(\frac{1}{n^{2+\mu}h^{\frac{d}{2}(2+\mu)}}\right)$$

as n tends to infinity.

Proof of Lemma 4. If we use for any integer $\mu > 0$ the notation

$$F(j_1, k_1, \quad \ldots, \quad j_{2+\mu}, k_{2+\mu})$$

$$=$$

$$\int_{x_1} \ldots \int_{x_{2+\mu}} K(\frac{x_1 - X_{j_1}}{h})K(\frac{x_1 - X_{k_1}}{h}) \quad \ldots \quad K(\frac{x_{2+\mu} - X_{j_{2+\mu}}}{h})K(\frac{x_{2+\mu} - X_{k_{2+\mu}}}{h}) \times$$

$$\times \quad \epsilon_{k_1} \ldots \epsilon_{k_{2+\mu}} w(x_1) \quad \ldots \quad w(x_{2+\mu})dx_1 \ldots dx_{2+\mu},$$

then we have directly that

$$E(W_n^{2+\mu}) \leq \left(\sup_x \{m(x) - \hat{m}_S(x)\}\right)^{2+\mu} \times$$

$$\times \frac{1}{n^{2(2+\mu)}h^{2(2+\mu)d}} \sum_{k_1} \sum_{j_1 \neq k_1} \cdots \sum_{k_{2+\mu}} \sum_{j_{2+\mu} \neq k_{2+\mu}} E\left[F(j_1, k_1, \ldots, j_{2+\mu}, k_{2+\mu})\right].$$

If we denote by $\mathcal{I}$ the set composed of all the $2(2+\mu)$-integers $(j_1, k_1, \ldots j_{2+\mu}, k_{2+\mu})$ such that $j_i \neq k_i, \forall i$, this last inequality can be written as:

$$E(W_n^{2+\mu}) \leq \left(\sup_x \{m(x) - \hat{m}_S(x)\}\right)^{2+\mu} \frac{1}{n^{2(2+\mu)}h^{2(2+\mu)d}} \sum_{\mathcal{I}} E\left[F(j_1, k_1, \ldots, j_{2+\mu}, k_{2+\mu})\right].$$

Now we will decompose the set $\mathcal{I}$ of indices according to the number of distinct integers among the different k_i. More precisely, we can write $\mathcal{I} = \mathcal{I}_1 \cup \ldots \cup \mathcal{I}_{2+\mu}$, where $\mathcal{I}_m = \{(j_1, k_1, \ldots j_{2+\mu}, k_{2+\mu}) \in \mathcal{I}$ such that $Card\{k_1, \ldots k_{2+\mu}\} = m\}$, so that the previous inequality can be rewritten as $E(W_n^{2+\mu}) \leq$

$$\left(\sup_x \{m(x) - \hat{m}_S(x)\}\right)^{2+\mu} \frac{1}{n^{2(2+\mu)}h^{2(2+\mu)d}} \sum_{m=1,\ldots 2+\mu} \sum_{\mathcal{I}_m} E\left[F(j_1, k_1, \ldots, j_{2+\mu}, k_{2+\mu})\right].$$

By independence of the data, note that if $m > 1 + \mu/2$, i.e. if one among the k_i appears only once, we have $E\left[F(j_1, k_1, \ldots, j_{2+\mu}, k_{2+\mu}) = 0\right]$. So we have finally $E(W_n^{2+\mu}) \leq$

$$\left(\sup_x \{m(x) - \hat{m}_S(x)\}\right)^{2+\mu} \frac{1}{n^{2(2+\mu)}h^{2(2+\mu)d}} \sum_{m=1,\ldots 1+\mu/2} \sum_{\mathcal{I}_m} E\left[F(j_1, k_1, \ldots, j_{2+\mu}, k_{2+\mu})\right].$$

Next, for some $m \in \{1, \ldots, 2 + \mu\}$, let us compute $E\left[F(j_1, k_1, \ldots, j_{2+\mu}, k_{2+\mu})\right]$. Using again the same technique, we can define:

$$\mathcal{I}_{m,s} = \{(j_1, k_1, \ldots j_{2+\mu}, k_{2+\mu}) \in \mathcal{I}_m, \text{ such that } Card\{j_1, k_1, \ldots j_{2+\mu}, k_{2+\mu}\} = s\},$$

and we have $\mathcal{I}_m = \mathcal{I}_{m,1} \cup \ldots \cup \mathcal{I}_{m,2+\mu+m}$. For $(j_1, k_1, \ldots j_{2+\mu}, k_{2+\mu}) \in \mathcal{I}_{m,s}$ we can do

exactly $s + 1 + \mu/2$ integrations by substitution, so that

$$\forall (j_1, k_1, \ldots j_{2+\mu}, k_{2+\mu}) \in \mathcal{I}_{m,s}, \; E\left[F(j_1, k_1, \ldots, j_{2+\mu}, k_{2+\mu})\right] = O(h^{(1+\mu/2+s)d}).$$

Because of $Card(\mathcal{I}_{m,s}) = O(n^s)$, we have directly

$$\sum_{\mathcal{I}_{m,s}} E\left[F(j_1, k_1, \ldots, j_{2+\mu}, k_{2+\mu})\right] = O(n^s h^{(1+\mu/2+s)d}).$$

Finally, we get $\quad \sum_{\mathcal{I}_m} E\left[F(j_1, k_1, \ldots, j_{2+\mu}, k_{2+\mu})\right] =$

$$= \sum_{s=1\ldots,2+\mu+m} O(n^s h^{(1+\mu/2+s)d}) = O(n^{2+\mu+m} h^{(3(1+\mu/2)+m)d}).$$

Let us now come back to the term $E[W_n^{2+\mu}]$. By using the last equality for any $m = 1, \ldots 1 + \mu/2$, we have $\quad E[W_n^{2+\mu}] \leq$

$$\leq \left(\sup_x (m(x) - \hat{m}_S(x))\right)^{2+\mu} \frac{1}{n^{2(2+\mu)} h^{2(2+\mu)d}} \sum_{m=1,\ldots 1+\mu/2} O(n^{2+\mu+m} h^{(3(1+\mu/2)+m)d})$$

$$\leq \left(\sup_x (m(x) - \hat{m}_S(x))\right)^{2+\mu} \frac{1}{n^{2(2+\mu)} h^{2(2+\mu)d}} \sum_{m=1,\ldots 1+\mu/2} O(n^{3(1+\mu/2)} h^{2(2+\mu)d})$$

$$= O\left(n^{-(1+\mu/2)} \sup_x (m(x) - \hat{m}_S(x))\right)^{2+\mu}.$$

Then, because of Lemma 1, given that $h \in H_n$ and assumption (H.1), we have

$$\sup_x \{m(x) - \hat{m}_S(x)\} = o\left(n^{-1/2} h^{-d/2}\right),$$

and we arrive finally at $E(W_n^{2+\mu}) = o(n^{-2-\mu} h^{-d-\frac{d\mu}{2}})$. ■

Lemma 5. *Assume that conditions (C.1)-(C.2), (S.1)-(S.3), (B.1), (K.1) and (H.1) hold. Then, for all z, we have under H_0 and as n tends to infinity:*

$$P\left(\max_{h \in H_n} |T_h - T_{h0}| > z\right) = o(1),$$

where

$$T_{h0} = \frac{nh^{d/2}I_h^0(\eta_1, \eta_2, \theta) - h^{-d/2}B}{\sqrt{V}}, \tag{26}$$

$$I_h^0(\eta_1, \eta_2, \theta) = \int \left[\frac{1}{nh^d}\sum_{j=1}^{n}K\left(\frac{x - X_j}{h}\right)(Y_j - m(X_j))\right]^2 \omega(x)\,dx. \tag{27}$$

Proof of Lemma 5. By assumptions (S.2) and (S.3) and using definitions (5) and (26), a Taylor expansion around V gives

$$T_h - T_{h0} = \frac{nh^{d/2}\left\{I_h\left(\hat{\eta}_1, \hat{\eta}_2, \hat{\theta}\right) - I_h^0(\eta_1, \eta_2, \theta)\right\} - h^{-d/2}\left(\hat{B} - B\right)}{\sqrt{V}} + o_p\left(\hat{V} - V\right)$$

as n tends to infinity. Then, using the triangle inequality we obtain

$$P\left(\max_{h \in H_n}|T_h - T_{h0}| > z\right) \leq P\left(\max_{h \in H_n}\left|\frac{nh^{d/2}\left\{I_h\left(\hat{\eta}_1, \hat{\eta}_2, \hat{\theta}\right) - I_h^0(\eta_1, \eta_2, \theta)\right\}}{\sqrt{V}}\right| > z/2\right)$$

$$+ P\left(\max_{h \in H_n}\left|\frac{h^{-d/2}\left\{\hat{B} - B\right\}}{\sqrt{V}}\right| > z/2\right). \tag{28}$$

We treat each of the above terms separately. For the first term of the r. h. s. in (28) the following inequality holds

$$P\left(\max_{h \in H_n}\left|\frac{nh^{d/2}\left\{I_h\left(\hat{\eta}_1, \hat{\eta}_2, \hat{\theta}\right) - I_h^0(\eta_1, \eta_2, \theta)\right\}}{\sqrt{V}}\right| > z\right)$$

$$\leq \frac{J_n\left(nh^{d/2}\right)^{2+\mu}}{z^2}E\left[I_h\left(\hat{\eta}_1, \hat{\eta}_2, \hat{\theta}\right) - I_h^0(\eta_1, \eta_2, \theta)\right]^{2+\mu},$$

for some $\mu > 0$. Let us now decompose $I_h\left(\hat{\eta}_1, \hat{\eta}_2, \hat{\theta}\right)$ and $I_h^0(\eta_1, \eta_2, \theta)$ as

$$I_h^0(\eta_1, \eta_2, \theta) = S_h^0(\eta_1, \eta_2, \theta) + R_h^0(\eta_1, \eta_2, \theta),$$

where

$$S_h^0(\eta_1, \eta_2, \theta) = \int \left(\frac{1}{nh^d}\right)^2 \sum_{j=1}^n K^2\left(\frac{x - X_j}{h}\right) \varepsilon_j^2 \omega(x)\, dx,$$

and

$$R_h^0(\eta_1, \eta_2, \theta) = \int \left(\frac{1}{nh^d}\right)^2 \sum_j \sum_{k \neq j} K\left(\frac{x - X_j}{h}\right) K\left(\frac{x - X_k}{h}\right) \varepsilon_j \varepsilon_k \omega(x)\, dx.$$

$$I_h\left(\hat{\eta}_1, \hat{\eta}_2, \hat{\theta}\right) = S_h^0(\eta_1, \eta_2, \theta) + R_h^0(\eta_1, \eta_2, \theta) + U_n + 2V_n + 2W_n$$

where U_n, V_n and W_n are defined in previous lemmas. Then, the following inequality holds

$$E\left[I_h\left(\hat{\eta}_1, \hat{\eta}_2, \hat{\theta}\right) - I_h^0(\eta_1, \eta_2, \theta)\right]^{2+\mu} = E\left[U_n + 2V_n + 2W_n\right]^{2+\mu}$$

$$\leq C_1 E\left[U_n\right]^{2+\mu} + C_2 E\left[V_n\right]^{2+\mu} + C_3 E\left[W_n\right]^{2+\mu},$$

for some constants $C_1, C_2, C_3 > 0$.

Now, applying Lemmas 2, 3 and 4, and assumption (H.1) the first term of the r.h.s. in (28) tends to zero as n tends to infinity.

For the second term holds the following bound,

$$P\left(\max_{h \in H_n} \left|\frac{h^{-d/2}\{\hat{B} - B\}}{\sqrt{V}}\right| > z\right) \leq \frac{J_n h^{-\frac{\mu d}{2}}}{z^2} E\left[\hat{B} - B\right]^{2+\mu},$$

and substituting $\hat{B}$ and B by their definitions in (7) we obtain

$$E\left[\hat{B} - B\right]^{2+\mu} = \left\{\int K^2(t)dt\right\}^{2+\mu} \times E\left\{\int \left[\hat{\sigma}^2(u)\hat{p}(u) - \sigma^2(u)p(u)\right]\omega(u)du\right\}^{2+\mu}$$

and

$$E\left[\hat{B} - B\right]^{2+\mu} = O\left(\left\{\sup_x \left|\hat{\sigma}^2(x)\hat{p}(x) - \sigma^2(x)p(x)\right|\right\}^{2+\mu}\right).$$

Now using assumption (C.2) and (H.1) we get

$$\frac{J_n h^{-\frac{\mu d}{2}}}{z^2} E\left[\hat{B} - B\right]^{2+\mu} = o(1).$$

This closes the proof. ∎

Lemma 6. *Assume that conditions (C.1)-(C.2), (S.1)-(S.3), (B.1), (K.1) and (H.1) hold. Then, as n tends to infinity,*

$$\max_{h \in H_n} T_{h0} - \max_{h \in H_n} \tilde{T}_{h0} \to_d 0,$$

where

$$\tilde{T}_{h0} = \frac{nh^{d/2}\tilde{I}_h\left(\eta_1, \eta_2, \theta\right) - h^{-d/2}B}{\sqrt{V}}, \tag{29}$$

$$\tilde{I}_h\left(\eta_1, \eta_2, \theta\right) = \int \left[\frac{1}{nh^d}\sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right)\left(\tilde{Y}_j - m\left(X_j\right)\right)\right]^2 \omega\left(x\right) dx, \tag{30}$$

and $\tilde{Y}_j = m\left(X_j\right) + \sigma\left(X_j\right)\varepsilon_j^*$ with ε_j^* being resampled errors (having mean zero and variance one).

Proof of Lemma 6. Let $\tilde{I}_h\left(\eta_1, \eta_2, \theta\right) = \tilde{S}_h^0\left(\eta_1, \eta_2, \theta\right) + \tilde{R}_h^0\left(\eta_1, \eta_2, \theta\right)$, where

$$\tilde{S}_h^0\left(\eta_1, \eta_2, \theta\right) = \int \left(\frac{1}{nh^d}\right)^2 \sum_{j=1}^{n} K^2\left(\frac{x - X_j}{h}\right)\sigma^2\left(X_j\right)\varepsilon_j^{*2}\omega\left(x\right) dx,$$

$$\tilde{R}_h^0\left(\eta_1, \eta_2, \theta\right) = \int \left(\frac{1}{nh^d}\right)^2 \sum_{k \neq j} K\left(\frac{x - X_j}{h}\right) K\left(\frac{x - X_k}{h}\right)\sigma\left(X_j\right)\sigma\left(X_k\right)\varepsilon_j^*\varepsilon_k^*\omega\left(x\right) dx.$$

Moreover, as it is known from the proof of Lemma 5 that

$$I_h^0\left(\eta_1, \eta_2, \theta\right) = S_h^0\left(\eta_1, \eta_2, \theta\right) + R_h^0\left(\eta_1, \eta_2, \theta\right).$$

In order to prove the lemma we will first show that

$$\max_{h \in H_n} V^{-1/2}\left\{\tilde{S}_h^0\left(\eta_1, \eta_2, \theta\right) - S_h^0\left(\eta_1, \eta_2, \theta\right)\right\} = o_p(1). \tag{31}$$

If this result holds, to show the lemma it suffices to prove that the joint distributions of $V^{-1/2}\tilde{R}_h^0(\eta_1, \eta_2, \theta)$ and $V^{-1/2}R_h^0(\eta_1, \eta_2, \theta)$ are asymptotically the same.

We first show (31). Write

$$\max_{h \in H_n} \left\{ V^{-1/2}\tilde{S}_h^0(\eta_1, \eta_2, \theta) - V^{-1/2}S_h^0(\eta_1, \eta_2, \theta) \right\}$$

$$= \max_{h \in H_n} V^{-1} \int \left(\frac{1}{nh^d}\right)^2 \sum_{j=1}^n K^2\left(\frac{x - X_j}{h}\right) \left\{\varepsilon_j^2 - \sigma^2(X_j)\varepsilon_j^{*2}\right\} \omega(x)dx.$$

To prove (31), it suffices to show that for $n \to \infty$

$$\sum_{h \in H_n} V^{-1} E \left[\int \left(\frac{1}{nh^d}\right)^2 \sum_{j=1}^n K^2\left(\frac{x - X_j}{h}\right) \left\{\varepsilon_j^2 - \sigma^2(X_j)\varepsilon_j^{*2}\right\} \omega(x)dx \right]^2 = o(1). \quad (32)$$

Note that taking iterated expectations, using the i.i.d. structure of our observations and integrating by substitution, it is straightforward to show that

$$E \left[\int \left(\frac{1}{nh^d}\right)^2 \sum_{j=1}^n K^2\left(\frac{x - X_j}{h}\right) \left\{\varepsilon_j^2 - \sigma^2(X_j)\varepsilon_j^{*2}\right\} \omega(x)dx \right]^2 =$$

$$= O\left(\frac{1}{n^2 h^{2d}}\right) + O\left(\frac{1}{n^3 h^{3d}}\right) \quad .$$

Hence, (32) is of order $O\left(\frac{1}{n^2}\sum_{h \in H_n}\frac{1}{h^{2d}}\right)$. Now, using (10) and because $h \in H_n$, this is indeed of order $O\left(\frac{1}{n^2 h_{\max}^{2d}}\right)$, and (31) is proved.

Then, the proof that the joint distributions of $V^{-1/2}\tilde{R}_h^0(\eta_1, \eta_2, \theta)$ and $V^{-1/2}R_h^0(\eta_1, \eta_2, \theta)$ are asymptotically the same follows the same lines of the proof of Lemma 10 in Horowitz and Spokoiny (2001), pp 622. ∎

Lemma 7. *Assume that conditions (C.1)-(C.3), (S.1)-(S.3), (B.1), (K.1) and (H.1) hold. Then, for all z, under H_0, and for $n \to \infty$*

$$P \left(\max_{h \in H_n} \left|\hat{T}_h - \tilde{T}_{h0}\right| > z \right) = o(1),$$

where

$$\hat{T}_h = \frac{nh^{d/2} I_h\left(\hat{\hat{\eta}}_1, \hat{\hat{\eta}}_2, \hat{\hat{\theta}}\right) - h^{-d/2}\hat{\hat{B}}}{\sqrt{\hat{\hat{V}}}}, \tag{33}$$

$$I_h\left(\hat{\hat{\eta}}_1, \hat{\hat{\eta}}_2, \hat{\hat{\theta}}\right) = \int \left[\frac{1}{nh^d}\sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right)\left(Y_j - \hat{\hat{m}}_S(X_j)\right)\right]^2 \omega(x)\,dx, \tag{34}$$

$$\hat{\hat{B}} = \int K^2(u)du \int \hat{\hat{\sigma}}^2(u)\hat{p}(u)\omega(u)du, \tag{35}$$

$$\hat{\hat{V}} = \int K^{(2)}(u)^2 du \int \hat{\hat{\sigma}}^4(u)\hat{p}^2(u)\omega^2(u)du, \tag{36}$$

and $\hat{\hat{m}}_S(x)$ and $\hat{\hat{\sigma}}^2(x)$ are resampling estimators.

Proof of Lemma 7. Using (29) and (33), under conditions (S.2) and (S.3) a Taylor expansion around V gives

$$\hat{T}_h - \tilde{T}_{h0} = \frac{nh^{d/2}\left\{I_h\left(\hat{\hat{\eta}}_1, \hat{\hat{\eta}}_2, \hat{\hat{\theta}}\right) - \tilde{I}_h(\eta_1, \eta_2, \theta)\right\} - h^{-d/2}\left(\hat{\hat{B}} - B\right)}{\sqrt{V}} + o\left(\hat{\hat{V}} - V\right)$$

as n tends to infinity. Now, by the triangle inequality

$$P\left(\max_{h \in H_n}\left|T_h^* - \tilde{T}_{h0}\right| > z\right) \le P\left(\max_{h \in H_n}\left|\frac{nh^{d/2}\left\{I_h\left(\hat{\hat{\eta}}_1, \hat{\hat{\eta}}_2, \hat{\hat{\theta}}\right) - \tilde{I}_h(\eta_1, \eta_2, \theta)\right\}}{\sqrt{V}}\right| > z\right)$$

$$+ P\left(\max_{h \in H_n}\left|\frac{h^{-d/2}\left\{\hat{\hat{B}} - B\right\}}{\sqrt{V}}\right| > z\right). \tag{37}$$

We treat each of the above terms by separate. By a standard inequality,

$$P\left(\max_{h \in H_n}\left|\frac{nh^{d/2}\left\{I_h\left(\hat{\hat{\eta}}_1, \hat{\hat{\eta}}_2, \hat{\hat{\theta}}\right) - \tilde{I}_h(\eta_1, \eta_2, \theta)\right\}}{\sqrt{V}}\right| > z\right)$$

$$\leq \frac{J_n n^{2+\mu} h^{\frac{d}{2}(2+\mu)}}{z^2} E\left[I_h\left(\hat{\hat{\eta}}_1, \hat{\hat{\eta}}_2, \hat{\hat{\theta}}\right) - \tilde{I}_h\left(\eta_1, \eta_2, \theta\right)\right]^{2+\mu},$$

for some $\mu > 0$. Noting that

$$I_h\left(\hat{\hat{\eta}}_1, \hat{\hat{\eta}}_2, \hat{\hat{\theta}}\right) - \tilde{I}_h\left(\eta_1, \eta_2, \theta\right) = S_h\left(\hat{\hat{\eta}}_1, \hat{\hat{\eta}}_2, \hat{\hat{\theta}}\right) - \tilde{S}_h^0\left(\eta_1, \eta_2, \theta\right)$$

$$+R_h\left(\hat{\hat{\eta}}_1, \hat{\hat{\eta}}_2, \hat{\hat{\theta}}\right) - \tilde{R}_h^0\left(\eta_1, \eta_2, \theta\right) + U_h\left(\hat{\hat{\eta}}_1, \hat{\hat{\eta}}_2, \hat{\hat{\theta}}\right) + 2V_h\left(\hat{\hat{\eta}}_1, \hat{\hat{\eta}}_2, \hat{\hat{\theta}}\right) + 2W_h\left(\hat{\hat{\eta}}_1, \hat{\hat{\eta}}_2, \hat{\hat{\theta}}\right),$$

where

$$S_h\left(\hat{\hat{\eta}}_1, \hat{\hat{\eta}}_2, \hat{\hat{\theta}}\right) = \int \left(\frac{1}{nh^d}\right)^2 \sum_{j=1}^n K^2\left(\frac{x-X_j}{h}\right) \hat{\sigma}^2\left(X_j\right) \varepsilon_j^{*2} \omega\left(x\right) dx,$$

$$R_h\left(\hat{\hat{\eta}}_1, \hat{\hat{\eta}}_2, \hat{\hat{\theta}}\right) = \int \left(\frac{1}{nh^d}\right)^2 \sum_j \sum_{k\neq j} K\left(\frac{x-X_j}{h}\right) K\left(\frac{x-X_k}{h}\right) \hat{\sigma}\left(X_j\right) \hat{\sigma}\left(X_k\right) \varepsilon_j^* \varepsilon_k^* \omega\left(x\right) dx,$$

$$U_h\left(\hat{\hat{\eta}}_1, \hat{\hat{\eta}}_2, \hat{\hat{\theta}}\right) = \int \left[\frac{1}{nh^d} \sum_{j=1}^n K\left(\frac{x-X_j}{h}\right) \left(\widehat{m}_S\left(X_j\right) - \widehat{\widehat{m}}_S\left(X_j\right)\right)\right]^2 \omega\left(x\right) dx,$$

$$V_h\left(\hat{\hat{\eta}}_1, \hat{\hat{\eta}}_2, \hat{\hat{\theta}}\right) = \int \left(\frac{1}{nh^d}\right)^2 \sum_{j=1}^n \varepsilon_j^* K^2\left(\frac{x-X_j}{h}\right) \left(\widehat{m}_S\left(X_j\right) - \widehat{\widehat{m}}_S\left(X_j\right)\right) \omega\left(x\right) dx,$$

and

$$W_h\left(\hat{\hat{\eta}}_1, \hat{\hat{\eta}}_2, \hat{\hat{\theta}}\right) = \int \left(\frac{1}{nh^d}\right)^2 \sum_j \sum_{k\neq j} K\left(\frac{x-X_j}{h}\right) K\left(\frac{x-X_k}{h}\right) \varepsilon_j^* \left(\widehat{m}_S\left(X_j\right) - \widehat{\widehat{m}}_S\left(X_j\right)\right) \omega\left(x\right) dx,$$

the following bound holds

$$E\left[\tilde{I}_h\left(\eta_1, \eta_2, \theta\right) - I_h\left(\hat{\hat{\eta}}_1, \hat{\hat{\eta}}_2, \hat{\hat{\theta}}\right)\right]^{2+\mu} = E\left[S_h - S_h^0 + R_h - R_h^0 + U_h + 2V_h + 2W_h\right]^{2+\mu}$$

$$\leq C_1' E\left[S_h - S_h^0\right]^{2+\mu} + C_2' E\left[T_h - T_h^0\right]^{2+\mu} + C_3' E\left[U_h\right]^{2+\mu} + C_4' E\left[V_h\right]^{2+\mu} + C_5' E\left[W_h\right]^{2+\mu},$$

for some constants $C_1', C_2', C_3', C_4', C_5' > 0$.

$$E\left[S_h - S_h^0\right]^{2+\mu}$$

$$\leq \left(\frac{1}{nh^d}\right)^{2(2+\mu)} \left(\sup_x \left|\hat{\sigma}^2\left(X_j\right) - \sigma^2\left(X_j\right)\right|\right)^{2+\mu} E\left[\int \sum_{j=1}^n K^2\left(\frac{x-X_j}{h}\right)\varepsilon_j^{*2}\omega(x)dx\right]^{2+\mu}.$$

Furthermore, integrating by substitution there exists a constant $C > 0$ such that

$$E\left[\int \sum_{j=1}^n K^2\left(\frac{x-X_j}{h}\right)\varepsilon_j^{*2}\omega(x)dx\right]^{2+\mu} \leq C\left(\int K^2(u)du\right)^{2+\mu}\sum_{j=1}^n E\left[\varepsilon_j^{*2}\omega(X_j)\right]^{2+\mu}.$$

(38)

Applying (C.2) and the bound achieved in (38) we get

$$n^{2+\mu}h^{\frac{d}{2}(2+\mu)}\log n E\left[S_h - S_h^0\right]^{2+\mu} = o\left(1\right).$$

Following the same lines as above it is straightforward to show that the second term,

$$R_h - R_h^0$$

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

is also of order $o\left(\frac{1}{n^{2+\mu}h^{\frac{d}{2}(2+\mu)}\log n}\right)$, as n tends to infinity.

The terms $U_h\left(\hat{\hat{\eta}}_1, \hat{\hat{\eta}}_2, \hat{\hat{\theta}}\right)$, $V_h\left(\hat{\hat{\eta}}_1, \hat{\hat{\eta}}_2, \hat{\hat{\theta}}\right)$ and $W_h\left(\hat{\hat{\eta}}_1, \hat{\hat{\eta}}_2, \hat{\hat{\theta}}\right)$ can be treated as in Lemmas 2, 3 and 4 respectively by replacing assumption (C.1) by assumption (C.3).

■

Proof of Theorem 1. The proof of this result is done if we can show that $\hat{T}^*$ and T^* have identical asymptotic distribution. This follows directly from previous lemma, since as n tends to infinity, we have by Lemma 5 that $\max_{h\in H_n} T_h = \max_{h\in H_n} T_{h0} + o_p(1)$, while Lemma 6 insures that $\max_{h\in H_n} T_{h0}$ and $\max_{h\in H_n} \tilde{T}_{h0}$ have identical asymptotic distribution. So, because Lemma 7 insures that $\max_{h\in H_n} \tilde{T}_{h0} = \max_{h\in H_n} \hat{T}_h + o_p(1)$, we get that $\max_{h\in H_n} \hat{T}_h$ and $\max_{h\in H_n} T_h$ have the same asymptotic distribution. ■

5.3   Proof of Theorem 2

The proof works by following the lines of the proof for Theorem 1 but replacing Lemmas 2-7 by similar results stated under the alternative hypothesis. We start the proof by stating the following Lemmas, along which the notations used above.

Lemma 8. *Under the assumptions of Theorem 2 we have for any $h \in H_n$ and as n tends to infinity: $U_n = C\gamma_n^2 + o_p\left(n^{-1}h^{-d/2}\right)$.*

Proof of Lemma 8. The proof follows the lines of the one for Lemma 2, but now assuming H_1 and that consequently expression (39) holds. Then we obtain

$$U_n \equiv U_{1n} + U_{2n} + U_{12n},$$

where

$$U_{1n} = \gamma_n^2 \int \left[\frac{1}{nh^d} \sum_i K\left(\frac{x - X_i}{h}\right) \vartheta\left(X_i\right) \right]^2 \omega\left(x\right) dx,$$

$$U_{2n} = \int \left[\frac{1}{nh^d} \sum_i K\left(\frac{x - X_i}{h}\right) \{m\left(X_i\right) - \hat{m}_S\left(X_i\right)\} \right]^2 \omega\left(x\right) dx,$$

$$U_{12n} = 2\gamma_n \int \left[\frac{1}{nh^d} \sum_i K\left(\frac{x - X_i}{h}\right) \vartheta\left(X_i\right) \right]$$
$$\left[\frac{1}{nh^d} \sum_i K\left(\frac{x - X_i}{h}\right) \{m\left(X_i\right) - \hat{m}_S\left(X_i\right)\} \right] \omega\left(x\right) dx.$$

Applying (S.2) and a strong law of large numbers we obtain $U_{1n} = C\gamma_n^2 + o_p(1)$. Furthermore, $|U_{2n}| = o_p\left(n^{-1}h^{-d/2}\right)$ by Lemma 2. Finally, integrating by substitution, applying a strong law of large numbers and conditions (H.1) and (A.1), $U_{12n} = o_p\left(U_{1n}\right)$.

∎

Lemma 9. *Under the assumptions of Theorem 2 we have for any $h \in H_n$ and n tending to infinity: $V_n = o_p\left(U_n\right)$*

Proof of Lemma 9. The proof basically follows the arguments as in Lemma 3, but now we analyze the behavior under H_1. Note that H_1 implies

$$\int y\ell_{Y|X_1,X_2}(y,x_1,x_2)\,dy = \int y\ell^0_{Y|X_1,X_2}(y,x_1,x_2\theta)\,dy + \gamma_n \int y\psi(y,x_1,x_2)\,dy, \quad (39)$$

where $\vartheta(x) = \int y\psi(y,x_1,x_2)\,dy$ is a bounded continuous function on its support, see (A.3). Then

$$V_n \equiv V_{1n} + V_{2n} = \gamma_n \int \left(\frac{1}{nh^d}\right)^2 \sum_j \varepsilon_j K^2\left(\frac{x-X_j}{h}\right)\vartheta(X_j)\omega(x)dx \quad (40)$$

$$+\int \left(\frac{1}{nh^d}\right)^2 \sum_j \varepsilon_j K^2\left(\frac{x-X_j}{h}\right)\{m(X_j) - \hat{m}_S(X_j)\}\omega(x)dx. \quad (41)$$

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

Integrating by substitution and using the smoothness conditions on $\omega(x)$ then

$$V_{1n} = \frac{\gamma_n}{nh^d}\left[\int K^2(u)\,du \times \frac{1}{n}\sum_{j=1}^{n}\vartheta(X_j)\omega(X_j)\varepsilon_j\right] \quad (42)$$

$$+ \frac{\gamma_n h^2}{nh^d}\left[\int u^2 K^2(u)\,du \times \frac{1}{n}\sum_{j=1}^{n}\vartheta(X_j)\omega''(X_j)\varepsilon_j\right] \quad (43)$$

$$+ o_p\left(\frac{\gamma_n h^4}{nh^d}\right). \quad (44)$$

Under the conditions established in the lemma, standard central limit theorems apply so that $\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\vartheta(X_j)\omega(X_j)\varepsilon_j$ and $\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\vartheta(X_j)\omega''(X_j)\varepsilon_j$ have expectation zero and they are bounded in probability. Specifically,

$$V_{1n} = C\frac{\gamma_n}{n^{3/2}h^d} + o_p\left(\frac{\gamma_n}{n^{3/2}h^d}\right). \quad (45)$$

Further, the term V_{2n} can be bounded as follows,

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

Then integrating by substitution

$$|E\left(V_{2n}\right)| = O_p\left(\sup_x |m\left(x\right) - \widehat{m_S}\left(x\right)| \frac{1}{nh^d}\right) + o_p\left(\sup_x |m\left(x\right) - \widehat{m_S}\left(x\right)| \frac{1}{nh^d}\right).$$

Using assumptions (C.1) and (H.1) and recalling that $h \in H_n$,

$$|E\left(V_{2n}\right)| = o_p\left(n^{-1}h^{-d/2}\right). \tag{46}$$

For the variance expression we obtain

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

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

Integrating by substitution, using conditions (C.1) and (H.1) and recalling that $h \in H_n$,

$$|Var\left(V_{2n}\right)| = o_p\left(\frac{1}{n^2 h^d}\right). \tag{47}$$

Then, from (40), (41), (45), (46) and (47) we finally get

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

Apply conditions (A.1) and (H.1) and for all $h \in H_n$ the proof is done.

∎

Lemma 10. *Under the assumptions of Theorem 2 we have for any $h \in H_n$ and as n tends to infinity:* $W_n = o_p(U_n)$.

Proof of Lemma 10. The proof of this result is based on the proof of Lemma 4. The bias term is $E\left(W_n\right) = 0$. For the variance term proceed as in the proof of Lemma 7 but using H_1, and therefore equality (39). We obtain $E\left(W_n^2\right) \leq \frac{C\gamma_n^2}{n}$. This closes the

proof of the Lemma.

∎

Lemma 11. *Under the assumptions required either for Theorem 1 or Theorem 2, we have* $|t_\alpha| \le M < \infty$.

Proof of Lemma 11. Because of Lemmas 7 and 8, under the conditions of either Theorem 1 or 2 we have $\max_h \hat{T}_h = \max_h T_{h0} + o_p(1)$. So it suffices to see that

$$\forall\, h, T_{h0} = O_p\left(nh^{d/2}\right) \tag{48}$$

to get the proof. Note that (48) is obtained as long as we have

$$\forall\, h, I_h^0\left(\eta_1, \eta_2, \theta\right) \le C < \infty \tag{49}$$

which is easy to obtain by working the terms out. ∎

Lemma 12. *Under the assumptions of Theorem 2, and $n \to \infty$, we have for any z*

$$P\left(\max_{h \in H_n} |T_h - T_{h0}| > z\right) \to 1.$$

Proof of Lemma 12. Using the same kind of decomposition as in Lemma 5 the claimed result follows directly from Lemmas 9, 8, 10 and 11. ∎

Proof of Theorem 2. Note that for the proof of Lemma 6 we did not use the null hypothesis H_0. So, the results of this lemma are still available here. Under the alternative, H_1, and adding assumptions (A.1) to (A.3) the results in Lemma 7 also apply. So, by combining results of Lemmas 6 and 7, we get directly that

$$\max_{h \in H_n} T_{h0} = \max_{h \in H_n} \hat{T}_h + o_p(1).$$

37

So it suffices to combine this with Lemma 12 to close the proof of Theorem 2. ∎

References

Amemiya, T. (1973) Regression Analysis when the Dependent Variable is truncated Normal. *Econometrica*, **41**, 997-1016.

Delgado, M.A., J.M. Rodriguez-Poó, and M. Wolf (2001) Subsampling Cube Root Asymptotics with an Application to Manski's MSE. *economics letters*, **73**, 241-250.

Dette, H., C.v. Lieres und Wilkau & S. Sperlich (2005) A comparison of different non-parametric methods for inference on additive models. *Journal of Nonparametric Statistics*, **17**, 57-81.

Giné, E., and J. Zinn (1990) Bootstraping general empirical measures. *Annals of Probability*, **18**, 851-869.

Gonzalez-Manteiga, W. & R.M. Crujeiras (2013) An updated review of Goodness-of-Fit tests for regression models. With discussion *Test*, **22**, 361-447.

Gourieroux, C., A. Monfort & C. Trognon (1984) Pseudo-maximum likelihood Methods: theory. *Econometrica*, **52**, 681-700.

Guerre, E. & P. Lavergne (2002) Optimal Minimax Rates for Nonparametric Specification Testing in Regression Models. *Econometric Theory*, **18**, 1139-1171.

Guerre, E. & P. Lavergne (2005) Rate-optimal data-driven specification testing for regression models. *Annals of Statistics*, **33**, 840-870.

Härdle, W., S. Huet, E. Mammen & S. Sperlich (2004) Bootstrap Inference in Semi-parametric Generalized Additive Models. *Econometric Theory*, **20**, 265-300.

Härdle, W., S. Sperlich & V. Spokoiny (2001) Structural tests in additive regression. *Journal of the American Statistical Association*, **96**, 1333-1347

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

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

Kallenberg, W.C.M. & T. Ledwina (1995) Consistency and Montecarlo Simulation of a data driven version of smooth goodness-of-fit tests. *Annals of Statistics*, **23**, 1594-1608.

Köhler M., A. Schindler & S. Sperlich (2013) A review and comparison of bandwidth selection methods for kernel regression. *International Statistical Review*, in press.

Lewbel, A. & O. Linton (2002) Nonparametric censored and truncated regression. *Econometrica*, **70** 765-779.

Lewbel, A. & O. Linton (2007) Nonparametric Matching and Homothetically Separable Functions. *Econometrica*, **75**, 1209-1228.

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

Neumeyer, N. & S. Sperlich (2006) Comparison of Separable Components in Different Samples. *Scandinavian Journal of Statistics*, **33**, 477-501.

Pardo-Fernández, J.C., I. Van Keilegom & W. González-Manteiga (2007) Goodness-of-fit tests for parametric models in censored regression. *Canadian Journal of Statistics*, **35**, 249-264.

Politis, D.N., J.P. Romano & M. Wolf (1999). *Subsampling.* Springer-Verlag, New York.

Politis, D.N., J.P. Romano & M. Wolf (2001) On the Asymptotic Theory of Subsampling. *Statistica Sinica*, **11**, 1105-1124.

Roca-Pardiñas, J. & S. Sperlich (2007) Testing the link when the index is semiparamtric – A comparative study *Computational Statistics & Data Analysis*, **12**, 6565-6581.

Rodríguez-Póo, J., S. Sperlich & P. Vieu (2003) Semiparametric Estimation of Separable Models with Possibly Limited Dependent Variables. *Econometric Theory*, **19**, 1008-1040.

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

Sperlich, S. (2013) On the choice of regularization parameters in specification testing: a critical discussion. *Empirical Economics*, in press

Sperlich, S., D. Tjøstheim & L. Yang (2002) Nonparametric estimation and testing of interaction in additive models. *Econometric Theory*, **18**, 197-251.

Spokoiny, V. (2001) Data-driven testing the fit of linear models. *Mathematical methods of statistics*, **10**, 465-497.

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