Taylor & Francis
Taylor & Francis Group

Statistics: A Journal of Theoretical and Applied Statistics

ISSN: 0233-1888 (Print) 1029-4910 (Online) Journal homepage: https://www.tandfonline.com/loi/gsta20

A bootstrap test for single index models

Wolfgang Härdle , Enno Mammen & Isabel Proenca

To cite this article: Wolfgang Härdle , Enno Mammen & Isabel Proenca (2001) A bootstrap test for single index models, Statistics: A Journal of Theoretical and Applied Statistics, 35:4, 427-451, DOI: 10.1080/02331880108802746

To link to this article: https://doi.org/10.1080/02331880108802746

Statistics **35** (2001) pp. 427–451
Reprints available directly from the publisher
Photocopying permitted by license only

A BOOTSTRAP TEST FOR SINGLE INDEX MODELS

WOLFGANG HÄRDLE[a,*], ENNO MAMMEN[b] and ISABEL PROENÇA[c]

[a]*Humboldt- Universität zu Berlin, Germany;* [b]*Ruprecht-Karls-Universität Heidelberg, Germany;* [c]*Universidade Técnica de Lisboa, Lisbon*

(Received 8 August 1999; In final form 19 April 2000)

Single index models are frequently used in econometrics and biometrics. Logit and Probit models are special cases with fixed link functions. In this paper we consider a bootstrap specification test that detects nonparametric deviations of the link function.

The bootstrap is used with the aim to find a more accurate distribution under the null than the normal approximation. We prove that the statistic and its bootstrapped version have the same asymptotic distribution. In a simulation study we show that the bootstrap is able to capture the negative bias and the skewness of the test statistic. It yields better approximations to the true critical values and consequently it has a more accurate level than the normal approximation.

Keywords: Bootstrap; Kernel estimate; Single index model; Specification test

AMS 1991 Subject Classifications: Primary: 62G09; Secondary: 62G10

1. INTRODUCTION

Single index models are frequently used in econometrics and biometrics. They are defined by the equation

$$Y = F(X^T\theta) + \varepsilon, \tag{1}$$

where the "link function" F is a known function that operates possibly nonlinearly on the "linear index" $X^T\theta$. For the error term ε we assume

*Corresponding author.

that $E(\varepsilon|X=x)=0$. Model (1) can be rewritten as

$$E(Y|X=x)=F(X^T\theta). \tag{2}$$

Single index models are natural generalizations of the linear model $Y=X^T\theta+\varepsilon$. They allow modelling of a variety of situations, *e.g.*, binary responses, multinomial and ordered discrete responses, censored and truncation responses. In the context of binary choice models Eq. (2) can have the probit formulation

$$E(Y|X=x)=P(Y=1|X=x)=\Phi(x^T\theta)$$

(with Φ the standard normal distribution), or the logit formulation

$$E(Y|X=x)=P(Y=1|X=x)=\{1+\exp(-x^T\theta)\}^{-1}.$$

McCullagh and Nelder (1989) give a survey with several applications in biometrics. Fahrmeier and Tutz (1994) apply this model in credit scoring. Stoker (1992) and Horowitz (1998) describe econometric applications with a view towards semiparametric approaches. Examples are labour supply (Stoker, 1992), work-trip mode choice (Horowitz, 1993), migration on the labour market (Burda, 1993) and innovative behaviour of firms (Bertschek and Entorf, 1996). Furthermore, some models from survival analysis are single index models, see (Cox and Oakes, 1984).

In this paper we consider the following model that is slightly more general than (2)

$$E(Y|X=x)=F\{v(x,\theta)\} \tag{3}$$

where Y is a real valued response variable and where the covariable X takes values in $\mathbb{R}^k$. The link function $F:\mathbb{R}\rightarrow\mathbb{R}$ and the index function $v:\mathbb{R}^{k+d}\rightarrow\mathbb{R}$ are known. The parameter θ is assumed to lie in a subset Θ of $\mathbb{R}^d$. Note that model (3) differs from (2) by allowing nonlinear indices $v(X,\theta)$.

An important step in the use of single index models is the choice of the link function F. There are often purely practical reasons for choosing a certain functional form for F. Consider *e.g.*, the case that model (3) is motivated by the stochastic utility approach, see Maddala (1983). In this approach the form of the index stems from theoretical considerations whereas the link function is the distribution function of some error variables. Typically only for convenience the error variables are assumed to have normal or logit distribution.

On the other hand, empirical evidence has shown that such a parametric specification of F is not always adequate, see *e.g.*, the transportation choice example in Horowitz (1993). It is therefore important to have a test for the specification of the link F. Horowitz and Härdle (1994) have developed such a test. Their HH test is different from procedures proposed by Azzalini, Bowman and Härdle (1989), Härdle and Mammen (1993), or le Cessie and van Houwelingen (1991). These latter tests are constructed for arbitrary nonparametric alternatives. If X is highdimensional such overall tests can have a poor power. The HH test avoids this "curse of dimensionality" by assuming that also on the alternative the conditional expectation of Y depends on the covariable X only *via* the index function $\nu(X, \theta)$.

Horowitz and Härdle (1994) propose to use normal approximations for the calculation of critical values. Simulations have shown that the accuracy of the normal approximation is affected by a negative bias of the HH statistic, see Proença (1993) and Proença and Ritter (1993). Here, we study bootstrap procedures. Our aim is to get approximations that work better in finite samples than the normal distribution. We will discuss several bootstrap approaches. We will study the performance of these bootstrap tests by asymptotics (see Section 3) and by simulations (see Section 4). In the next section we will give a short description of the HH test.

2. THE MODEL AND THE TEST STATISTIC

Let $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ be a random i.i.d. sample of (X, Y) that follows model (3). We assume that there exists a $\sqrt{n}$–consistent estimator $\hat{\theta}_n$ of θ. Typically, such a $\sqrt{n}$–consistent estimate is given *e.g.*, by the method of (weighted) least squares or by the (quasi) maximum likelihood approach, see McCullagh and Nelder (1989) and Severini and Staniswalis (1994).

We come now to the description of the HH test. The test is constructed for the test hypotheses

$$\begin{aligned} H_0 &: E\{Y|X\} = F(v(X,\theta)) \\ H_1 &: E\{Y|X\} = H(v(X,\theta)) \end{aligned} \tag{4}$$

where $H(\cdot)$ is an unknown function. Note that under H_0 and H_1 we have that $E\{Y|X\} = E\{Y|v(X,\theta)\}$. The HH statistic is motivated by conditional moment tests, see Newey (1985). The main idea relies on the intuition that on the hypothesis a nonparametric estimate of F should be close to F. On the other hand, if the link F is not correctly specified then the nonparametric estimate should significantly differ from F. The HH statistic has the form

$$T_n = \sqrt{h}\sum_{i=1}^{n} w(\hat{v}_i)[Y_i - F(\hat{v}_i)][\hat{F}_{ni}(\hat{v}_i) - F(\hat{v}_i)], \tag{5}$$

where $\hat{v}_i = v(X_i, \hat{\theta}_n)$ and where w is a non-negative weight function. Often the weight function w is defined as the indicator function of an appropriately chosen set. $\hat{F}_{ni}(\hat{v}_i)$ is a leave-one-out kernel regression estimate of $E\{Y|v(X,\theta)\}$. For asymptotic unbiasedness it is defined according to the proposal of Bierens (1987). It is a linear combination of two regression kernel smoothers with different bandwidths (h and s, respectively)

$$\hat{F}_{ni}(v) = \left\{\hat{F}_{nhi}(v) - \left(\frac{h}{s}\right)^r \hat{F}_{nsi}(v)\right\} \Big/ \left\{1 - \left(\frac{h}{s}\right)^r\right\} \tag{6}$$

$$\hat{F}_{nti}(v) = \frac{\sum_{j=1, j\neq i}^{n} Y_j K[\{v - v(X_j, \hat{\theta}_n)\}/t]}{\sum_{j=1, j\neq i}^{n} K[\{v - v(X_j, \hat{\theta}_n)\}/t]}, \quad \text{for } t = h \quad \text{and} \quad t = s \tag{7}$$

where $h = cn^{-1/(2r+1)}$, $s = c'n^{-\delta/(2r+1)}$ with $c, c' > 0, 0 < \delta < 1$ and a kernel $K(\cdot)$ of order $r \geq 2$.

Horowitz and Härdle (1994) show that T_n has an asymptotic normal distribution $N(0, \sigma_T^2)$ where

$$\sigma_T^2 = 2C_k \int_{-\infty}^{\infty} w(v)^2 \sigma^2(v) \mathrm{d}v, \tag{8}$$

$$C_k = \int_{-\infty}^{\infty} K(u)^2 \mathrm{d}u, \tag{9}$$

$$\sigma^2(v) = V\{Y|v(X,\theta) = v\}. \tag{10}$$

With an estimator for σ_T^2 the normal limit of the test statistic can be used for the calculation of approximate critical values, see Horowitz and Härdle (1994). However this approach can lead to very poor level accuracy, see Proença (1993) and Proença and Ritter (1993). Bootstrap is an alternative approach for getting approximate critical values. In the next section we introduce several bootstrap procedures and discuss their asymptotic properties.

3. THE BOOTSTRAP APPROACH

The main idea of the bootstrap approach is to mimic the model under the hypothesis H_0. Bootstrap gives an approximation for the distribution of the HH statistic when model (3) is true. We will see that it produces critical values that are more reliable than the critical values of the normal approximation.

We will consider a class of resampling schemes that work according to the following steps.

Step 1 Calculate variables $\hat{\sigma}_i^2$ that approximate the conditional variances $\sigma^2\ [v(X_i, \theta)] = var[Y_i | v(X_i, \theta)]$ of the responses Y_i. [Which choices of $\hat{\sigma}_i^2$ are possible will be discussed below.]

Step 2 Generate n conditionally [given the original sample] independent random variables $\varepsilon_1^*, \ldots, \varepsilon_n^*$ with conditional mean 0 and conditional variances $\hat{\sigma}_1^2, \ldots, \hat{\sigma}_n^2$. [Choices of the conditional distributions of ε_i^* will be discussed below.]

Step 3 Put

$$Y_i^* = F\{v(X_i, \hat{\theta}_n)\} + \varepsilon_i^* \tag{11}$$

for $i = 1, \ldots, n$. We will use $(X_i, Y_i^*),\ i = 1, \ldots, n$ as bootstrap sample. Note that in these resampling schemes the covariables X_i are not resampled.

Step 4 Calculate an estimate $\hat{\theta}_n^*$ that is based on the bootstrap sample (and the original sample).

Step 5 The bootstrap version of the HH statistic is now

$$T_n^* = \sqrt{h} \sum_{i=1}^{m} w(\hat{v}_i^*)[Y_i^* - F(\hat{v}_i^*)][\hat{F}_{ni}^*(\hat{v}_i^*) - F(\hat{v}_i^*)], \quad (12)$$

where $\hat{v}_i^* = v(X_i, \hat{\theta}_n^*)$, where $w(\cdot)$ is defined as above and where $\hat{F}_{ni}^*(\cdot)$ is calculated as $\hat{F}_{ni}(\cdot)$ but with Y_i replaced by Y_i^*. (The covariables X_i are not replaced in the definition of $\hat{F}_{ni}^*(\cdot)$.)

For the choice of the variables $\hat{\sigma}_i^2 (i = 1, \ldots, n)$ and $\hat{\theta}_n^*$ and of the distributions of $\varepsilon_i^* (i = 1, \ldots, n)$ we make the following assumptions:

B 1. It holds that

$$\frac{1}{hn} \sum_{j \neq i} \frac{w(v_i)^2 K^2([v_i - v_j]/h)}{p_\theta^2(v_i)} [\sigma^2(v_i)\sigma^2(v_j) - \hat{\sigma}_i^2 \hat{\sigma}_j^2] = o_P(1)$$

where $v_i = v(X_i, \theta)$ and where p_θ is the density of $v(X, \theta)$.

B 2. For the bootstrap error variables we have that there exists a $\rho > 1$ with $n^{1-\rho} h^{-2\rho} \log(n)^{2\rho} = O(1)$ and

$$\max_{1 \leq i \leq n} E\left[|\varepsilon_i^*|^{4\rho} | X_1, \ldots, X_n\right] = O_P(1).$$

B 3. The estimate $\hat{\theta}_n^*$ fulfills

$$\hat{\theta}_n^* = \hat{\theta}_n + O_P(n^{-1/2}),$$

i.e., $\hat{\theta}_n^*$ is $\sqrt{n}$ consistent in the bootstrap world.

Condition B1 holds if $\hat{\sigma}_i^2$ are consistent estimates of $\sigma^2(v_i)$, *i.e.*,

$$\max_{1 \leq i \leq n} |\hat{\sigma}_i^2 - \sigma^2(v_i)| = o_P(1). \quad (13)$$

But condition (13) is not necessary. Note that B1 holds if certain local and global averages of $\hat{\sigma}_i^2$ are asymptotically equivalent to the corresponding averages of $\sigma^2(v_i)$. In particular, we will apply our results to a resampling scheme (wild bootstrap, see below), where $\hat{\sigma}_i^2$ is equal to a squared single residual. Then under appropriate conditions B1 holds whereas (13) is not fullfilled.

In B2 we consider the conditional expectation given only the covariables (and not the responses). It can be easily verified that this condition implies that

$$E^*\left[|\varepsilon_i^*|^{4p}\right] = O_P(1), \tag{14}$$

where E^* denotes the conditional expectation given the whole sample $X_1, Y_1, \ldots, X_n, Y_n$. We will consider a resampling scheme (wild bootstrap) that fulfills B2 but not (14).

We now come to a discussion of the choice of the parametric estimate $\hat{\theta}_n^*$. An obvious choice that fulfills our condition B3 is $\hat{\theta}_n^* = \hat{\theta}_n$. Then the parametric estimation step is not mimicked in the resampling. This choice of $\hat{\theta}_n^*$ works asymptotically because the first order asymptotic distribution of our test statistic is not affected by the choice of the estimate $\hat{\theta}_n$. However we expect for finite samples and asymptotically in second order that bootstrap performs more accurately if an estimate $\hat{\theta}_n^*$ is used that mimics $\hat{\theta}_n$. This estimate has to be calculated in each cycle of the resampling algorithm. For most estimates $\hat{\theta}_n$, calculation is lengthy because it requires an iterative algorithm. In the bootstrap world we know the true underlying parameter, namely $\hat{\theta}_n$. So there we can use the same iterative algorithm with $\hat{\theta}_n$ as starting value. Theoretical considerations for many estimates (quasi-likelihood, weighted least squares) suggest that one iteration in the algorithm suffices, see *e.g.*, Mosbach (1992).

We will discuss the validity of these assumptions for several resampling schemes after the statement of the following theorem. This theorem states that on the hypothesis T_n^* has the same asymptotic distribution as T_n. This implies that bootstrap gives a consistent estimate of the critical values of T_n. So the bootstrap test has asymptotically correct level. The theorem uses assumptions A1 – A9 that are stated in the appendix.

THEOREM 1 *Suppose that the HH test is used with bandwidths* $h = cn^{-1/(2r+1)}$ *and* $s = c'n^{-\delta/(2r+1)}$ *(for constants* $c, c' > 0$ *and* $0 < \delta < 1$*) and with a* $\sqrt{n}$ *consistent estimate* $\hat{\theta}_n = \theta + O_P(n^{-1/2})$. *Under assumptions* $B1 - B3$ *stated above and* $A1 - A9$ *from the appendix we get that*

$$d_\infty\left(\mathcal{L}^*(T_n^*),\ N(0, \sigma_T^2)\right) \to 0$$

in probability. Here $\mathcal{L}^*$ *denotes the conditional distribution given the sample* $X_1, Y_1, \ldots, X_n, Y_n$. *Furthermore,* d_∞ *denotes the Kolmogorov distance (i.e., the supremum norm between the corresponding distribution functions).*

We now discuss Assumptions B1 – B3 for several bootstrap resampling schemes. We will discuss wild bootstrap and parametric bootstrap. For a discussion of bootstrap in a related model, see also Mammen and van de Geer (1997) and Härdle, Mammen and Müller (1998).

Wild bootstrap is related to proposals of Wu (1986) and Beran (1986) and was first proposed by Härdle and Mammen (1993) and Mammen (1992) in nonparametric setups. In the wild bootstrap one generates an i.i.d. sample $\eta_1, \ldots, \eta_n$ with mean 0 and variance 1 and one puts $\varepsilon_i^* = \hat{\sigma}_i \eta_i$. The choice of $\hat{\sigma}_i$ depends on our assumptions on $\sigma^2(v) = Var\,[Y|v(X,\theta) = v]$. We discuss here two types of choices for $\hat{\sigma}_i$.

- If we make no smoothness assumptions on $\sigma^2(v)$ then an appropriate choice of $\hat{\sigma}_i^2$ is $[Y_i - F(\hat{v}_i)]^2$ or $[Y_i - \hat{F}_{ni}(\hat{v}_i)]^2$, respectively (where, as above, $\hat{v}_i = v(X_i, \hat{\theta}_n)$). It can be checked that Assumptions B1 and B2 are fulfilled if the distribution of $\eta_1, \ldots, \eta_n$ have compact support. In particular then B2 follows from A7 and boundedness of F, see A5.
- Under smoothness assumptions on $\sigma^2(v)$ estimates $\hat{\sigma}_i$ can be used that are based on smoothing of squared residuals (see *e.g.*, Gasser, Sroka and Jennen-Steinmetz (1986)). Suppose that estimates $\hat{\sigma}_i$ are available such that

$$\max_{1 \le i \le n} \left|\hat{\sigma}_i^2 - \sigma^2(v_i)\right| = o_P(1)$$

and such that $\hat{\sigma}_i$ is uniformly bounded (a.s.). Then wild bootstrap fullfills Conditions B1, B2 as long as $\eta_1, \ldots, \eta_n$ have a bounded 4ρ-th moment. This is easy to check.

Another modification of wild bootstrap is the moment oriented bootstrap of Bunke (1997). In this resampling also higher order moments of the data are mimicked by the resampling. Again it is easy to see that, under appropriate conditions, this approach fits into our framework B1 – B2.

In the case of binary responses the (conditional) distribution of the response Y (given the covariable X) is determined by the value of $F(X^T\theta)$. Then it makes sense to generate Y_i^* that have (conditional) distribution according to the parameter $F(X^T\hat{\theta}_n)$. This is an example of the parametric bootstrap. It can be applied for the whole class of generalized linear models where the conditional distribution of the response belongs to an exponential family. The conditions A1 and A2 are fullfilled for parametric bootstrap if the 4ρ-th moments of the responses are uniformly bounded for parameters in a neighborhood of θ. For the case of a generalized model this holds if $F(X^T\theta)$ lies almost shurely in the interior of the natural parameter space of the exponential family.

The statement of Theorem 1 can be extended to the case of local alternatives. Suppose that

$$E[Y_i|v(X_i, \theta_n) = v] = F_n(v)$$

a.s., for a sequence of functions F_n and parameters θ_n. Furthermore, denote the conditional variance $Var\ [Y_i|v(X_i, \theta_n) = v]$ by $\sigma_n^2(v)$. The parameter θ_n now depends on n and is chosen such that $\hat{\theta}_n = \theta_n + O_P(n^{-1/2})$. (Note that $\hat{\theta}_n$ is the estimate of θ on the hypothesis H_0 where $F_n = F$. On the alternative where F_n differs from F the center θ_n of the distribution of $\hat{\theta}_n$ may change. It is too restrictive to assume that $\theta_n \equiv \theta$.) If one assumes now that F_n,θ_n and σ_n^2 converge to F, θ or σ^2, respectively, and that our smoothness assumptions on F, θ and σ^2 hold for F_n,θ_n and σ_n^2 then one can show that also on the alternative T_n^* has the same asymptotic normal limit as stated in Theorem 1. Then also on the altenative bootstrap gives a consistent estimate of the critical values of T_n. So on the alternative the bootstrap test has asymptotically the same power function as the test that uses correct critical values.

The HH test requires the choice of two bandwidth h and s. It is needed that the bandwidth h is of order $n^{-1/(2r+1)}$ and that s is of smaller order, see Theorem 1. This choice of h is motivated by estimation problems where this rate of convergence is optimal. However for testing also other choices of h make sense. A test with large h looks for more global deviations of the link function from F, whereas small choices of h make the test more powerful for local deviations. So the assumption that h is of order $n^{-1/(2r+1)}$ is too restrictive. For the HH test this assumption has been made to guarantee that the estimate $\hat{F}_{ni}$ is asymptotically unbiased. We propose now a modification of the HH test that works for all choices h with $h = o(1)$. We put

$$T_n^{mod} = \sqrt{h}\sum_{i=1}^{m} w(\hat{v}_i)[Y_i - F(\hat{v}_i)]\frac{\sum_{j=1,j\neq i}^{n}[Y_j - F\{\hat{v}_j\}]K([\hat{v}_j - \hat{v}_i]/h)}{\sum_{j=1,j\neq i}^{n} K([\hat{v}_j - \hat{v}_i]/h)}.$$

In this test statistic the smoother

$$\hat{F}_{nhi}(\hat{v}_i) = \frac{\sum_{j=1,j\neq i}^{n} Y_j K([\hat{v}_j - \hat{v}_i]/h)}{\sum_{j=1,j\neq i}^{n} K([\hat{v}_j - \hat{v}_i]/h)}.$$

is compared with an estimate of its (conditional) expectation on the hypothesis:

$$\frac{\sum_{j=1,j\neq i}^{n} F\{\hat{v}_j\} K([\hat{v}_j - \hat{v}_i]/h)}{\sum_{j=1,j\neq i}^{n} K([\hat{v}_j - \hat{v}_i]/h)}.$$

This test statistic has the same asymptotic limit as the HH test. This asymptotic limit holds under weaker conditions. In particular, it is not required that h is of order $n^{-1/(2r+1)}$ and that F and the density of $v(X_i,\theta)$ have higher order derivatives. If h is of order $n^{-1/(2r+1)}$ then the modified test is asymptotically equivalent to the HH test. This is the content of the following theorem.

THEOREM 2 *Suppose that h fulfills $h = o(1)$ and $n^{-1}h^4 \log n = o(1)$, that $\hat{\theta}_n = \theta + O_P(n^{-1/2})$ and that conditions $A1 - A8$ from the appendix apply. Then it holds that*

$$T_n^{mod} \xrightarrow{\mathcal{L}} N(0, \sigma_T^2).$$

Under the additional assumption of $A9$ and that $h = cn^{-1/(2r+1)}$ and $s = c'n^{-\delta/(2r+1)}$ (for constants c, $c' > 0$ and $0 < \delta < 1$) we get that

$$T_n^{mod} = T_n + o_P(1).$$

The next theorem states that also for this test bootstrap can be used for the approximate determination of critical values.

THEOREM 3 *Suppose that h fulfills $h = o(1)$ and $n^{-1}h^4 \log n = o(1)$, that $\hat{\theta}_n = \theta + O_P(n^{-1/2})$ and that assumptions $B1 - B3$ stated above and conditions $A1 - A8$ from the appendix apply. Then we get that*

$$d_\infty\left(\mathcal{L}^*(T_n^{mod,*}),\ N(0, \sigma_T^2)\right) \to 0$$

in probability, where $T_n^{mod,*}$ is the bootstrap version of T_n^{mod}.

We have carried out a small simulation study for binary responses. The response variables took values from $\{0, 1\}$ with

$$P\{Y = 1|v(X, \theta) = v\} = E\{Y|v(X, \theta) = v\} = F(v).$$

We considered the HH test T_n for the null hypothesis H_0:

$$H_0 : F(v) = \{1 + \exp(-v)\}^{-1}$$

with linear index $v(x, 0) = 1 + x^T\theta$ where $\theta = (-1, 2)^T$.

The covariables X had a two dimensional standard normal distribution and the generated samples had size $n = 200$. We have checked the test for the alternative *Logit with a bump*:

$$E[Y|v(X, \theta) = v] = \{1 + \exp(-v)\}^{-1} - \frac{a}{1.5}\varphi\left(\frac{a}{1.5}\right).$$

with $\varphi(\cdot)$ the standard normal density and a = 0.5, 0.75, 1, 1.25. In our simulation the HH test was performed with bandwidths $h = 0.5$, $h = 1$ and $h = 1.5$. These choices were made after a graphical inspection of the kernel regression estimates. These bandwidths corresponds to undersmoothing, nearly optimal smoothing and oversmoothing, respectively. The bandwidth s was determined according to $s = hn^{(1-\delta)/5}$ with $\delta = 0.1$. The weight function $w(v)$ was defined as the indicator function between the 5% and the 95% percentiles of the fitted index v. The critical values were calculated by parametric bootstrap. The bootstrap used 199 replications. In Table I the rejection probabilities

TABLE I Percentages of rejections using critical values from the normal approximation and bootstrap critical values. Nominal size is 5% or 10%, respectively

	Nominal size 5%		*Nominal size* 10%	
	Normal	*Bootstrap*	*Normal*	*Bootstrap*
logit link				
h = 0.5	1.6	4.8	2.2	9.8
h = 1.0	0.2	5.0	0.6	10.8
h = 1.5	0.0	7.4	0.0	14.6
logit link with bump a = 1				
h = 1.0	0.4	35.6	2.6	66.6
h = 1.5	0.8	34.4	3.0	48.8
logit link with bump a = 1.25				
h = 1.0	1.2	49.2	2.6	66.6
h = 1.5	2.2	42.4	6.4	60.8

are compared for the bootstrap test and for the HH test with critical values from the normal approximation. The normal approximation does not work. The levels are too small. The test is too conservative. Because of the inaccurate critical values this test achieves no power for the different alternatives. On the other hand, the bootstrap test performs reasonable well. Its level does not differ too much from the nominal level and it achieves remarkably better power.

Acknowledgements

We would like to thank Joel Horowitz and Rodney Wolff for valuable suggestions. We are also grateful to L. Fahrmeir for suggesting the one-iteration ML estimate for the bootstrap.

References

Azzalini, A., Bowman, A. and Härdle, W. (1989) On the use of non-parametric regression for model checking. *Biometrika*, **76**, 1 – 11.

Beran, R. (1986) Comment on "Jackknife, bootstrap, and other resampling methods in regression analysis" by Wu, C. F. J., *Annals of Statistics*, **14**, 1295 – 1298.

Bertschek, I. and Entorf, H. (1996) On nonparametric estimation of the Schumpeterian link between innovation and firm size: evidence from Belgium, France, and Germany. *Empirical Economics*, **21**, 401 – 426.

Bierens, H. (1987) Kernel estimators of regression functions. In: Bewley, T. F. (Ed.) *Advances in Econometrics*: *Fifth World Congress*, **1**, Cambridge University Press, New York, pp. 99 – 144.

Bunke, O. (1997) Bootstrapping in heteroschedastic regression situations. *Discussion Paper*, Sonderforschungsbereich 373, Humboldt university Berlin.

Burda, M. (1993) The determinants of East – West German migration. *European Economic Review*, **37**, 452 – 461.

Cox, D. R. and Oakes, D. (1984) *Analysis of Survival Data*. Chapman and Hall, New York.

de Jong, P. (1987) A central limit theorem for generalized quadratic forms. *Probab. Theory and Related Fields*, **75**, 261 – 277.

Fahrmeir, L. and Tutz, G. (1994) *Multivariate Statistical Modelling Based on Generalized Linear Models*, Springer, New York.

Firth, D. (1992) Bias reduction, the jeffreys prior and GLIM. In: Fahrmeir, L., Francis, B., Gilchrist, R. and Tutz, G. (Eds.) *Advances in CLIM and Statistical Modelling. Proceedings of the GLIM92 Conference and 7th International Workshop on Statistical Modelling*. Springer, New York, pp. 91 – 100.

Gasser, T., Sroka, L. and Jennen-Steinmetz, C. (1986) Residual variance and residual patern in nonlinear regression. *Biometrika*, **73**, 625 – 633.

Härdle, W. (1990) *Applied Nonparametric Regression*, Cambridge University Press, Cambridge.

Härdle, W. and Mammen, E. (1993) Testing parametric *versus* nonparametric regression. *Ann. Statist.*, **21**, 1926 – 1947.

Härdle, W., Mammen, E. and Müller (1998) Testing parametric *versus* semiparametric modelling in generalized linear models. *J. Amer. Statist. Assoc.*, **93**, 1461 – 1474.

Hastie, T. J. and Tibshirani, R. J. (1990) *Generalized Additive Models*. Chapman and Hall, London.
Horowitz, J. (1993) Semiparametric estimation of a work-trip mode choice model. *Journal of Econometrics*, **58**, 49–70.
Horowitz, J. (1998) *Semiparametric Methods in Econometrics*. Lecture Notes in Statistics, **131**, Springer, New York.
Horowitz, J. and Härdle, W. (1994) Testing a parametric model against a semiparametric alternative. *Econometric Theory*, **10**, 821–848.
le Cessie, S. and van Houwelingen, J. C. (1991) A goodness-of-fit test for binary regression models based on smoothing methods. *Biometrics*, **47**, 1267–1282.
Maddala, G. S. (1983) Limited-dependent and qualitative variables in econometrics. *Econometric Society Monographs* No. 4, Cambridge University Press.
Mammen, E. (1989) Asymptotics with increasing dimension for robust regression with applications to the bootstrap. *Ann. Statist.*, **17**, 382–400.
Mammen, E. (1992) When does bootstrap work? asymptotic results and simulations. *Lectures Notes in Statist.*, **77**, Springer, Heidelberg, Berlin, New York.
Mammen, E. and van de Geer, S. (1997) Penalized quasi-likelihood estimation in partial linear models. *Ann. Statist.*, **25**, 1014–1035.
McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*. Chapman and Hall, London.
Mosbach, O. (1992) One-step bootstrapping in generalized linear models. In: Jöckel, K.-H., Rothe, G. and Sendler, W. (Eds.) Bootstrap and related techniques. *Proceedings of an International Conference held in Trier*, Germany, 1990, Springer, New York.
Newey, W. (1985) Maximum Likelihood specification testing and conditional moment test. *Econometrica*, **53**, 1047–1070.
Proenca, I. and Werwatz, A. (1995) Comparing Parametric and Semi-parametric Binary Response Models. In: Härdle, W., Klinke, S. and Turlach, B. A. (Eds.) *Xplore: An Interactive Statistical Computing Environment*. pp. 251–274, Springer, New York.
Severini, T. A. and Staniswalis, J. G. (1994) Quasi-Likelihood estimation in semiparametric models. *J. Amer. Statist. Assoc.*, **89**, 501–511.
Stoker, T. M. (1992) *Lectures on Semiparametric Econometrics*. Core Lecture Series, CORE, Louvain-La-Neuve.
Whittle, P. (1960) Bounds for the moments of linear and quadratic forms in independent variables. *Theory Probab. Appl.*, **5**, 302–305.
Wu, C. F. G. (1986) Jacknife, bootstrap and other resampling methods in regression analysis. (with discussion) *Ann. Statist.*, **14**, 1291–1380.

A. CONDITIONS

For a neigbourhood N_θ of θ, for an open subset $\tilde{S}_v$ of the support of $v(x, \theta)$, and for a compact subset S_v of $\tilde{S}_v$, we make use of the following assumptions.

A1 – The covariable X has compact support S_x.
A2 – There exists a constant M such that for every $x \in S_x$ and $\tau \in N_\theta$

(a) $|v(x, \tau)| < M$,
(b) $v(x, \tau)$ is continuously differentiable with respect to τ and

$$|\partial v(x, \tau)/\partial \tau_j| < M, \ j = 1, \ldots, k.$$

A3 – The variable $v(x, \theta)$ has a density p_θ w.r.t. Lebesgue measure. For $v \in S_v$ the density is bounded from below and from above.

A4 – The weight function $w(\cdot)$ has compact support $S_w \subset int(S_v)$ and satisfies for some constants M_w and M_w^*:

(a) $0 \leq w(v) < M_w$ for all $v \in S_w$.

(b) $|w(v_2) - w(v_1)| \leq M_w^* |v_2 - v_1|$ for all $v_2, v_1 \in S_w$.

A5 –

(a) $|F\{v(x,\tau)\}|$ is uniformly bounded over $x \in S_x$ and $\tau \in N_\theta$.

(b) $|F(v_1) - F(v_2)| \leq M_F |v_2 - v_1|$ for a constant M_F and for all $v_2, v_1 \in S_v$.

A6 – $\sigma^2(v) = V\{Y | v(x,\theta) = v\}$ is a uniformly bounded, continuous function of $v \in S_v$.

A7 – There exists a $\rho > 1$ with $n^{1-\rho} h^{-2\rho} \log(n)^{2\rho} = O(1)$ such that

$$\sup_{v(x,\theta) \in S_v} E[|Y - E\{Y|X = x\}|^{4\rho}] < \infty \quad \text{a.s.}$$

A8 – K has bounded support, it is symmetrical about 0 and it has a bounded derivative K'.

A9 –

(a) K is an r'th order kernel.

(b) $F(v)$ has r continuous derivatives for $v \in S_v$.

(c) $p_\theta(v)$ has r continuous derivatives for $v \in S_v$ that are uniformly bounded for $v \in S_v$.

B. PROOFS

We start by giving a proof that T_n has an asymptotic normal distribution $N(0, \sigma_T^2)$. We show that this result holds under Assumptions A1–A9. These assumptions are slightly weaker than the assumptions used in Horowitz and Härdle (1994). The only exception is our additional Assumption A7 that we use to show that the quadratic statistic $\Delta_{n,1}(\tau)$ converges uniformly for τ with $\|\tau - \theta\| \leq Cn^{-1/2}$ to 0 (in probability), see (17). This implies that $\Delta_{n,1}(\hat{\theta}_n)$ converges to 0 (in probability). In Horowitz and Härdle (1994) it is only shown that

$\Delta_{n,1}\,(\theta_n)$ converges to 0 (in probability) for deterministic sequences θ_n with $\|\theta_n - \theta\| \leq Cn^{-1/2}$, see the proof of Lemma 6 in Horowitz and Härdle (1994).

Our proof of the asymptotic normality of T_n differs slightly from the proof given in Horowitz and Härdle (1994). It has an appropriate form that we can easily explain how the proof has to be modified to get the statements of Theorems 1−3.

Write

$$T_n(\tau) = \sqrt{h}\sum_{i=1}^{n} w\{v(X_i,\tau)\}[Y_i - F\{v(X_i,\tau)\}] \\ [\hat{F}_{ni}\{v(X_i,\tau)\} - F\{v(X_i,\tau)\}].$$

Then it holds that $T_n = T_n(\hat{\theta}_n)$, In a first step of the proof one shows that

$$T_n = T_n(\theta) + o_P(1). \tag{15}$$

This follows from

$$\sup_{\|\tau-\theta\| \leq Cn^{-1/2}} |T_n(\tau) - T_n(\theta_n)| = o_P(1). \tag{16}$$

The proof of (16) requires several steps. Most steps use uniform stochastic convergence of terms that are linear in $\varepsilon_1, \ldots, \varepsilon_n$, where $\varepsilon_i = Y_i - E[Y_i|X_i]$. At one point one has to check a term that is quadratic in $\varepsilon_1, \ldots, \varepsilon_n$,. This term has the form

$$\frac{\sqrt{h}}{n}\sum_{i=1}^{n} w\{v_i\}\varepsilon_i \sum_{j=1,j\neq 1}^{n} \varepsilon_j \Bigg\{\Bigg[\frac{K_h\{v(X_i,\tau) - v(X_j,\tau)\}}{\tilde{p}_{n,h,\tau}\{v(X_i,\tau)\}} \\ - \frac{K_h\{v_i - v_j\}}{\tilde{p}_{n,h,\theta}\{v_i\}}\Bigg] - \left(\frac{h}{s}\right)^r \\ \Bigg[\frac{K_s\{v(X_i,\tau) - v(X_j,\tau)\}}{\tilde{p}_{n,s,\tau}\{v(X_i,\tau)\}} \\ - \frac{K_s\{v_i - v_j\}}{\tilde{p}_{n,s,\theta}\{v_i\}}\Bigg]\Bigg\}\left[1 - \left(\frac{h}{s}\right)^r\right]^{-1},$$

where

$$\tilde{p}_{n,h,\tau}\{v(X_i,\tau)\} = \frac{1}{n}\sum_{j=1,j\neq 1}^{n} K_h\{v(X_i,\tau) - v(X_i,\tau)\}$$
$$= \frac{1}{n}\sum_{j=1}^{n} K_h\{v(X_i,\tau) - v(X_j,\tau)\} - \frac{1}{n}K_h(0).$$

We will discuss the quadratic term

$$\Delta_{n,1}(\tau) = \frac{\sqrt{h}}{n}\sum_{i=1}^{n} w\{v_i\}\varepsilon_i \sum_{j=1,j\neq i}^{n} \varepsilon_j[K_h\{v(X_i,\tau) - v(X_j,\tau)\} - K_h\{v_i - v_j\}]\tilde{p}_{n,h,\tau}\{v(X_i,\tau)\}^{-1},$$

and the linear terms

$$\Delta_{n,2}(\tau) = \sqrt{h}\sum_{i=1}^{n} w\{v_i\}\varepsilon_i[F\{v(X_i,\tau)\} - F\{v_i\}]\tilde{p}_{n,h,\theta}(v_i)^{-1},$$
$$\Delta_{n,3}(\tau) = \frac{\sqrt{h}}{n}\sum_{i=1}^{n} w\{v_i\}\varepsilon_i \sum_{j=1,j\neq i}^{n} [F\{v_i\} - F\{v_j\}] [K_h\{v(X_i,\tau) - v(X_j,\tau)\} - K_h\{v_i - v_j\}]\tilde{p}_{n,h,\tau}\{v(X_i,\tau)\}^{-1},$$

We will show that

$$\sup_{\|\tau-\theta\|\,\leq\, Cn^{-1/2}} |\Delta_{n,l}(\tau)| = o_P(1), \tag{17}$$

for $l=1$, $l=2$ and $l=3$. The other linear and quadratic terms can be treated similarly.

Proof of (17) *for* $l=1$ Choose $\delta_n \to 0$ such that

$$\gamma_n = \delta_n^{-4\rho} n^{1-\rho} h^{-2\rho} \log(n)^{2\rho} E[\varepsilon_i^{4\rho}\mathbf{I}\{|\varepsilon_i| > \delta_n n^{1/4} h^{1/2} \log(n)^{-1/2}\}] \to 0.$$

Such a δ_n exists because of Condition A7. Note that

$$E[\varepsilon_i^{4\rho}\mathbf{I}\{|\varepsilon_i| > \delta_n n^{1/4} h^{1/2} \log(n)^{-1/2}\}] \to 0$$

for $n \to \infty$. Put

$$\begin{aligned}
\varepsilon_i' &= \varepsilon_i \mathbf{I}\{|\varepsilon_i| \le \delta_n n^{1/4} h^{1/2} \log(n)^{-1/2}\} - \mu_i, \\
\mu_i &= E\varepsilon_i \mathbf{I}\{|\varepsilon_i| \le \delta_n n^{1/4} h^{1/2} \log(n)^{-1/2}\}, \\
\Delta_{n,1,1}(\tau) &= \sum_{j \ne 1} a_{n,i,j}(\tau)\varepsilon_i'\varepsilon_j', \\
\Delta_{n,1,2}(\tau) &= 2\sum_{j \ne i} a_{n,i,j}(\tau)\mu_i\varepsilon_j', \\
\Delta_{n,1,3}(\tau) &= \sum_{j \ne i} a_{n,i,j}(\tau)\mu_i\mu_j, \\
a_{n,i,j}(\tau) &= \frac{\sqrt{h}}{2n}\{w\{v_i\}[K_h\{v(X_i,\tau) - v(X_j,\tau)\} \\
&\qquad - K_h\{v_i - v_j\}]\tilde{p}_{n,h,\tau}\{v(X_i,\tau)\}^{-1}, \\
&\quad + w\{v_j\}[K_h\{v(X_j,\tau) - v(X_j,\tau)\} \\
&\qquad - K_h\{v_j - v_i\}]\tilde{p}_{n,h,\tau}\{v(X_j,\tau)\}^{-1}.
\end{aligned}$$

Note now that for all constants $C > 0$

$$\begin{aligned}
&P\{\Delta_{n,1}(\tau) = \Delta_{n,1,1}(\tau) + \Delta_{n,1,2}(\tau) \\
&\quad + \Delta_{n,1,3}(\tau) \text{ for all } \|\tau - \theta\| \le Cn^{-1/2}\} \\
&\ge P\{\varepsilon_i = \varepsilon_i' + \mu_i \text{ for all } 1 \le i \le n\} \\
&\ge P\{|\varepsilon_i| \le \delta_n n^{1/4} h^{1/2} \log(n)^{-1/2} \\
&\qquad \text{for all } 1 \le i \le n\} \\
&\ge 1 - nP\{|\varepsilon_1| > \delta_n n^{1/4} h^{1/2} \log(n)^{-1/2}\} \\
&\ge 1 - \gamma_n \to 1.
\end{aligned}$$

So for our claim (17) it suffices to show for all constants $C > 0$

$$\sup_{\|\tau - \theta\| \le Cn^{-1/2}} |\Delta_{n,1,j}(\tau)| = o_P(1), \tag{18}$$

for $j = 1, 2$ and 3. For the proof of (18) for $j = 1$ note first that there exists a constant C_1 such that for all $C > 0$, for all $k \ge 2$, and for all τ

with $\|\tau-\theta\| \le Cn^{-1/2}$

$$E\big[|\Delta_{n,1,1}(\tau)|^k|X_1,\ldots,X_n\big] \le C_1^k k!E \big[(\varepsilon_1')^{2k}|X_1,\ldots,X_n\big]\Big\{\sum_{i\neq j} a_{n,i,j}(\tau)^2\Big\}^{k/2}. \tag{19}$$

This bound follows by a slight modification of a bound of Whittle (1960), see also Lemma 4 and Eq. (3.23) in Mammen (1989). We use now

$$E\big[(\varepsilon_1')^{2k}|X_1,\ldots,X_n\big] \le 2^{2k}\delta_n^{2k}n^{k/2}h^k \log(n)^{-k} \quad \text{a.s.} \tag{20}$$

$$\sup_{\|\tau-\theta\|\le Cn^{-1/2}} \sum_{i\neq j} a_{n,i,j}(\tau)^2 = O_P(h^{-2}n^{-1}). \tag{21}$$

This bound (20) follows by definition of the variables ε_i'. For the proof of (21) one uses standard kernel smoothing theory. With the help of (19)–(21) we now get the following bound for $t>0$ and for all τ with $\|\tau-\theta\| \le Cn^{-1/2}$

$$\begin{aligned}&\sup_{\|\tau-\theta\|\le Cn^{-1/2}} E\big[\exp\{t\log(n)\Delta_{n,1,1}(\tau)\}|X_1,\ldots,X_n\big]\\ &\le 1+\sum_{k=2}^{\infty} C_1^k t^k \log(n)^k E\big[(\varepsilon_1')^{2k}|X_1,\ldots,X_n\big]\\ &\Big\{\sup_{\|\tau-\theta\|\le Cn^{-1/2}} \sum_{i\neq j} a_{n,i,j}(\tau)^2\Big\}^{k/2} = O_P(1).\end{aligned}$$

With a similar bound on $E[\exp\{-t\log(n)\Delta_{n,1,1}(\tau)\}\,|\,X_1,\ldots,X_n]$ we get that

$$\sup_{\|\tau-\theta\|\le Cn^{-1/2}} E\big[\exp\{t\log(n)|\Delta_{n,1,1}(\tau)|\}|X_1,\ldots,X_n\big] = O_P(1). \tag{22}$$

From inequality (22) we get the following bound for all t, C, $C'>0$

$$\begin{aligned}&\sup_{\|\tau-\theta\|\le Cn^{-1/2}} P\big[|\Delta_{n,1,1}(\tau)|>C'|X_1,\ldots,X_n\big]\\ &\le \sup_{\|\tau-\theta\|\le Cn^{-1/2}} \exp\big[-C't\log(n)\big]\\ &E\big[\exp\{t\log(n)|\Delta_{n,1,1}(\tau)|\}|X_1,\ldots,X_n\big] = O_P(n^{-C't}).\end{aligned} \tag{23}$$

Consider now finite subsets I_n of $\{\tau: \|\tau-\theta\| \le Cn^{-1/2}\}$ with number of elements bounded by $n^{C''}$ (for a constant C''). Then we get from (23)

$$\sup_{\tau \in I_n} |\Delta_{n,1,1}(\tau)| = o_P(1). \tag{24}$$

We use now the following crude bound

$$\begin{aligned} &\sup_{\|\tau_1-\theta\| \le Cn^{-1/2}, \|\tau_2-\theta\| \le Cn^{-1/2}, \|\tau_1-\tau_2\| \le n^{-3/2}h^{1/2}} |\Delta_{n,1,1}(\tau_1) - \Delta_{n,1,1}(\tau_2)| \\ &= \sup_{\|\tau_1-\theta\| \le Cn^{-1/2}, \|\tau_2-\theta\| \le Cn^{-1/2}, \|\tau_1-\tau_2\| \le n^{-3/2}h^{1/2}} \|\tau_1 - \tau_2\| O_P(\delta_n^2 n^{3/2} h^{-1/2}) \\ &= o_P(1). \end{aligned} \tag{25}$$

Claim (18) with $j=1$ follows now from (24) and (25) with an appropriate choice of I_n.

For the proof of (18) with $j=2$ and $j=3$ we use the following estimate of μ_i:

$$\begin{aligned} |\mu_i| &= |E\varepsilon_i \mathbf{I}\{|\varepsilon_i| \le \delta_n n^{1/4} h^{1/2} \log(n)^{-1/2}\}| \\ &= |-E\varepsilon_i \mathbf{I}\{|\varepsilon_i| > \delta_n n^{1/4} h^{1/2} \log(n)^{-1/2}\}| \\ &\le \delta_n^{-(4\rho-1)} n^{-(4\rho-1)/4} h^{-(4\rho-1)/2} \log(n)^{-(4\rho-1)/2} \\ &\qquad E|\varepsilon_i|^{4\rho} \mathbf{I}\{|\varepsilon_i| > \delta_n n^{1/4} h^{1/2} \log(n)^{-1/2}\} \\ &= \gamma_n n^{-1} \delta_n n^{1/4} h^{1/2} \log(n)^{1/2} E|\varepsilon_i|^{4\rho} \\ &\qquad \mathbf{I}\{|\varepsilon_i| > \delta_n n^{1/4} h^{1/2} \log(n)^{-1/2}\} \\ &= o(n^{-3/4} h^{1/2} \log(n)^{1/2}). \end{aligned} \tag{26}$$

Claim (18) with $j=2$ now follows from the following bound

$$\begin{aligned} \sup_{\|\tau-\theta\| \le Cn^{-1/2}} \sum_{j \ne i} |a_{n,i,j}(\tau)| &= O_P(h^{-1/2} n^{1/2}), \\ E|\varepsilon_j'| &= O(1). \end{aligned} \tag{27}$$

For the proof of (18) with $j=3$ one applies (26) and (27).

Proof of (17) *for* $l=2$ Note that

$$\Delta_{n,2}(\tau) = \sum_{i=1}^{n} a_{n,i}(\tau) \varepsilon_i$$

with

$$a_{n,i}(\tau) = \sqrt{h}w\{v_i\}[F\{v(X_i,\tau) - F\{v_i\}]\tilde{p}_{n,h,\theta}(v_i)^{-1}.$$

We define now

$$\Delta_{n,2,1}(\tau) = \sum_{i=1}^{n} a_{n,i}(\tau)\varepsilon_i',$$

$$\Delta_{n,2,2}(\tau) = \sum_{i=1}^{n} a_{n,i}(\tau)\mu_i,$$

where ε_i' and μ_i are defined as in the proof of (17) for $l=1$. We get now that for all $C>0$ with probability tending to 1 for all τ with $\|\tau-\theta\| \leq Cn^{-1/2}$ it holds that

$$\Delta_{n,2}(\tau) = \Delta_{n,2,1}(\tau) + \Delta_{n,2,2}(\tau).$$

So it remains to show for all $C>0$

$$\sup_{\|\tau-\theta\| \leq Cn^{-1/2}} |\Delta_{n,2,j}(\tau)| = o_P(1), \tag{28}$$

for $j=1$ and $j=2$. Claim (28) for $j=2$ follows easily from (26) and

$$\sum_{i=1}^{n} |a_{n,i}(\tau)| = O(hn^{-1/4}(\log n)^{1/2}) = o(1). \tag{29}$$

For the treatment of $\Delta_{n,2,1}(\tau)$ note that under our conditions it holds for all $C>0$ that

$$\sup_{\|\tau-\theta\| \leq Cn^{-1/2}, 1 \leq i \leq n} |a_{n,i}(\tau)| = O_P(h^{1/2}n^{-1/2}).$$

Put $a'_{n,i}(\tau) = a_{n,i}(\tau)\mathbf{I}\{|a_{n,i}(\tau)| \leq h^{1/4}n^{-1/2}\}$

$$\text{and} \quad \Delta_{n,2,3}(\tau) = \sum_{i=1}^{n} a'_{n,i}(\tau)\varepsilon_i'.$$

Then $\Delta_{n,2,1}(\tau) = \Delta_{n,2,3}(\tau)$ with probability tending to 1. It remains to show for all $C>0$

$$\sup_{\|\tau-\theta\| \leq Cn^{-1/2}} |\Delta_{n,2,3}(\tau)| = o_P(1). \tag{30}$$

We argue now that there exists a constant C_0 such that for $t > 0$, $1 \le i \le n$ and n large enough

$$\begin{aligned} &\sup_{\|\tau-\theta\| \le Cn^{-1/2}} E\big[\exp\{t \log(n) a'_{n,i}(\tau)\varepsilon'_i\}|X_1, \dots, X_n\big] \\ &\le 1 + C_0 t^2 \log(n)^2 a'_{n,i}(\tau)^2 E\big[(\varepsilon'_i)^2|X_1, \dots, X_n\big] \\ &\le \exp\big\{C_0 t^2 \log(n)^2 a'_{n,i}(\tau)^2 E\big[(\varepsilon'_i)^2|X_1, \dots, X_n\big]\big\}. \qquad (31) \end{aligned}$$

This follows by using the fact that for a constant C_1

$$\sup_{\|\tau-\theta\| \le Cn^{-1/2}, 1 \le i \le n} |\log(n) a'_{n,i}(\tau)\varepsilon'_i| \le C_1 n^{-1/4}.$$

With the help of (31) we now get the following bound for $t > 0$ and for all τ with $\|\tau - \theta\| \le Cn^{-1/2}$

$$\begin{aligned} &\sup_{\|\tau-\theta\| \le Cn^{-1/2}} E[\exp\{t \log(n) \Delta_{n,2,3}(\tau)\}|X_1, \dots, X_n] \\ &\le \sup_{\|\tau-\theta\| \le Cn^{-1/2}} \exp\left\{C_0 t^2 \log(n)^2 \sum_{i=1}^n a'_{n,i}(\tau)^2 E[(\varepsilon'_i)^2|X_1, \dots, X_n]\right\} \\ &\le \exp\{C_0 t^2 \log(n)^2 h^{1/2} O_P(1)\} = O_P(1). \end{aligned}$$

Claim (17) for $l = 2$ follows now with the same arguments as for $l = 1$.

Proof of (17) *for* $l = 3$ Note that

$$\Delta_{n,3}(\tau) = \sum_{i=1}^n b_{n,i}(\tau)\varepsilon_i$$

with

$$\begin{aligned} b_{n,i}(\tau) = \frac{\sqrt{h}}{n} w\{v_i\} \sum_{j=1, j \ne i}^n [F\{v_i\} - F\{v_j\}][K_h\{v(X_i, \tau) - v(X_j, \tau)\} \\ - K_h\{v_i - v_j\}]\tilde{p}_{n,h,\tau}\{v(X_i, \tau)\}^{-1}. \end{aligned}$$

By standard kernel smoothing theory one shows first that under our conditions for all $C > 0$ it holds that

$$\sup_{\|\tau-\theta\| \le Cn^{-1/2}, 1 \le i \le n} |b_{n,i}(\tau)| = O_P(h^{1/2} n^{1/2}).$$

One considers $\sum_{i=1}^{n} b'_{n,i}(\tau)\varepsilon'_i$ where $b'_{n,i}(\tau) = b_{n,i}(\tau)\mathbf{I}\{|b_{n,i}(\tau)| \le h^{1/4}n^{-1/2}\}$. Claim (17) for $l=3$ follows now with the same arguments as for $l=2$.

We come now to the second step of the proof of the asymptotic normality of T_n. Note that

$$T_n(\theta) = T_{n,1} + \cdots + T_{n,4},$$

where

$$T_{n,1} = \sum_{i \neq j}^{n} c_{n,i,j}\varepsilon_i\varepsilon_j,$$

$$c_{n,i,j} = \frac{\sqrt{h}}{n} w\{v_i\}K_h\{v_i - v_j\}\tilde{p}_{n,h,\theta}\{v_i\}^{-1}$$

$$T_{n,2} = T_{n,1}\left[\frac{1}{1-(h/s)^r} - 1\right],$$

$$T_{n,3} = \frac{\sqrt{h}}{n}\sum_{i \neq j}^{n} w\{v_j\}\varepsilon_i \frac{-(h/s)^r}{1-(h/s)^r} K_s\{v_i - v_j\}\varepsilon_j\tilde{p}_{n,s,\theta}\{v_i\}^{-1},$$

$$\begin{aligned} T_{n,4} = [1-(h/s)^r]^{-1}\sqrt{h}\sum_{i \neq j}^{n} & w\{v_i\}\varepsilon_i\varepsilon_j[F\{v_j\} - F\{v_i\}] \\ & [K_h\{v_i - v_j\}\tilde{p}_{n,h,\theta}\{v_i\}^{-1} \\ & - (h/s)^r K_s\{v_i - v_j\}\tilde{p}_{n,s,\theta}\{v_i\}^{-1}]. \end{aligned}$$

We now argue that

$$T_{n,j} = o_P(1) \tag{32}$$

for $j=2,3,4$. This can be shown by calculation of the second moments of $T_{n,2}$, $T_{n,3}$ and $T_{n,4}$. We would like to mention that the proof of (32) for $j=4$ is the only point of the proof where we need Condition A9. With (15) and (32) we get that

$$T_n = T_{n,1} + o_P(1). \tag{33}$$

It remains to show that $T_{n,1}$ has an asymptotic $N(0,\sigma_T^2)$ distribution. For this claim we show that

$$E^+[T_{n,1}^2 = var^+[T_{n,1}] = \sigma_T^2 + o_P(1), \tag{34}$$

$$d_\infty(\mathcal{L}^+(T_{n,1}),\ N(0, E^+[T_{n,1}^2])) = o_P(1). \tag{35}$$

Here $\mathcal{L}^+$ denotes the conditional distribution given the covariables $X_1, \ldots, X_n$ and var^+ denotes the (conditional) variance w.r.t. $\mathcal{L}^+$.

Claim (34) follows by standard kernel smoothing arguments. So it remains to show (35). According to Theorem 2.1 in de Jong (1987) for this claim it suffices to show

$$\max_{1 \le i \le n} \sum_{j=1, j \neq i}^{n} var^+[c_{n,i,j}\varepsilon_i\varepsilon_j]/var^+[T_{n,1}] = o_P(1), \tag{36}$$

$$E^+[T_{n,1}^4]/\{var^+[T_{n,1}]\}^2 = 3 + o_P(1), \tag{37}$$

where E^+ denotes the (conditional) expectation w.r.t. $\mathcal{L}^+$. Claim (36) follows from

$$\max_{1 \le i \le n} \sum_{j=1, j \neq i}^{n} c_{n,i,j}^2 = O_P(1/n)$$

and

$$\sup_{v(x,\theta) \in S_v} E[|Y - E\{Y|X = x\}|^2] < \infty \quad \text{a.s.},$$

see A7. For the proof of (37) note that

$$E^+[T_{n,1}^4] - 3\{var^+[T_{n,1}]\}^2 = 8A_{n,1} + 48A_{n,2} + 192A_{n,3},$$

where

$$A_{n,1} = \sum_{i \neq j} d_{n,i,j}^4 E^+\varepsilon_i^4 E^+\varepsilon_j^4,$$

$$A_{n,2} = \sum_{i,j,k,l \text{ pairwise different}} d_{n,i,j}d_{n,j,k}d_{n,k,l}d_{n,l,i}E^+\varepsilon_i^2E^+\varepsilon_j^2E^+\varepsilon_k^2E^+\varepsilon_l^2,$$

$$A_{n,3} = \sum_{i,j,k \text{ pairwise different}} d_{n,i,j}d_{n,i,k}^2 d_{n,j,k}E^+\varepsilon_i^3E^+\varepsilon_j^2E^+\varepsilon_k^3,$$

$$d_{n,i,j} = (c_{n,i,j} + c_{n,j,i})/2.$$

Claim (37) follows from

$$A_{n,1} \ge 0, \tag{38}$$

$$A_{n,2} \ge 0, \tag{39}$$

$$E\ A_{n,1} = o(1), \tag{40}$$

$$E\ A_{n,2} = o(1), \tag{41}$$

$$\begin{aligned} E|A_{n,3}| &\leq \sum_{i,j,k \text{ pairwise different}} E[d_{n,i,j} d_{n,i,k}^2 d_{n,j,k} E^+|\varepsilon_i|^3 E^+ \varepsilon_j^2 E^+|\varepsilon_k|^3] \\ &= o(1). \end{aligned} \tag{42}$$

Here (38) and (39) follow from $d_{n,i,j} \geq 0$ Claims (40)–(42) can be shown by using A7 and simple bounds, see the proof of Theorem 1 in Härdle und Mammen (1993) for similar calculations.

Proof of Theorem 1 Define $T_{n,1}^*$ as $T_{n,1}$ but with ε_i replaced by ε_i^*. The statement of the theorem follows from

$$T_n^* = T_{n,1}^* + o_P(1), \tag{43}$$

$$d_\infty(\mathcal{L}^*(T_{n,1}),\ N(0, \sigma_T^2)) = o_P(1), \tag{44}$$

Note that (43) implies that for all $\delta > 0$

$$P^*[|T_n^* - T_{n-1}^*| > \delta] = o_P(1),$$

because of $EP^*[|T_n^* - T_{n,1}^*| > \delta] = P[|T_n^* - T_{n,1}^*| > \delta] = o(1)$. Here P^* denotes the conditional distribution given the sample $X_1, Y_1, \ldots, X_n, Y_n$.

Claims (43) and (44) can be shown with essentially the same arguments as above. For the proof of (44) one has to show *e.g.*, that

$$A_{n,1}^* = \sum_{i \neq j} d_{n,i,j}^4 E^*(\varepsilon_i^*)^4 E^*(\varepsilon_j^*)^4 = o_P(1).$$

This follows from

$$E^+ A_{n,1}^* = \sum_{i \neq j} d_{n,i,j}^4 E^+(\varepsilon_i^*)^4 E^+(\varepsilon_j^*)^4 = o_P(1) \sum_{i \neq j} d_{n,i,j}^4,$$

see B2, and

$$\sum_{i \neq j} d_{n,i,j}^4 = o_P(1)$$

because of

$$E \sum_{i \neq j} d_{n,i,j}^4 = o(1).$$

Proof of Theorem 2 Define

$$T_n^{mod}(\tau) = \sqrt{h}\sum_{i=1}^{m} w\{v(X_i,\tau)\}[Y_i - F\{v(X_i,\tau)\}] \frac{\sum_{j=1,j\neq i}^{n}[Y_j - F\{v(X_j,\tau)\}]K([v(X_j,\tau) - v(X_i,\tau)]/h)}{\sum_{j=1,j\neq i}^{n} K([v(X_j,\tau) - v(X_i,\tau)]/h)}.$$

Then it holds that $T_n^{mod} = T_n^{mod}(\hat{\theta}_n)$. Again in a first step one shows that

$$T_n^{mod} = T_n^{mod}(\theta) + o_P(1). \tag{45}$$

This can be done with the same arguments as above by proving

$$\sup_{\|\tau-\theta\| \leq Cn^{-1/2}} |T_n^{mod}(\tau) - T_n^{mod}(\theta)| = o_P(1).$$

All steps of the proof work for bandwidth h with $h = o(1)$ and $n^{-1}h^4 (\log n)^2 = o(1)$. The latter condition is used in (29). Now we have that $T_n^{mod}(\theta) = T_{n,1}$. Furthermore, because $T_{n,1}$ converges in distribution to $N(0, \sigma_T^2)$ we get that T_n^{mod} has the asymptotic limit $N(0, \sigma_T^2)$ (in distribution). Note that we do not need Assumption A9 for the asymptotic treatment of T_n^{mod}. In particular the term $T_{n,4}$ does not appear in the asymptotic expansion of T_n^{mod}. Under the additional assumption of A9 and that $h = cn^{-1/(2r+1)}$ and $s = c'n^{-\delta/(2r+1)}$ we have $T_n = T_{n,1} + o_P(1)$, see (33). This immediately shows the claim $T_n^{mod} = T_n + o_P(1)$.

Proof of Theorem 3 See the proof of Theorem 1.

Wolfgang Härdle
Institut für Statistik
und Ökonometrie
Wirtschaftswissenschaftliche
Fakultät
Humboldt- Universität zu Berlin
Germany

Enno Mammen
Institut für Angewandte Mathematik
Ruprecht-Karls-Universität Heidelberg
Im Neuenheimer Feld 288
69120 Heidelberg
Germany

Isabel Proença
ISEG
Universidade Técnica
de Lisboa
P-1200 Lisbon
Portugal