A Bootstrap Test for Single Index Models

Wolfgang Härdle
Institut für Statistik und Ökonometrie
Wirtschaftswissenschaftliche Fakultät
Humboldt–Universität zu Berlin
D–10178 Berlin

Isabel Proença
ISEG – Universidade Técnica de Lisboa
and
Institut de Statistique
Université Catholique de Louvain
B–1348 Louvain–La–Neuve

April 12, 1995

Abstract

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, e.g., testing against a semiparametric alternative.

The bootstrap is used with the aim to find a more accurate distribution under the null than the standard normal since simulations with single index models have shown that the test has a negative bias that affects its empirical power in finite samples. We prove that the statistic and its bootstrapped version have the same asymptotic distribution. A simulation study shows that the bootstrapped critical values yield better approximations to the true critical values. Bootstrap is able to capture the negative bias of the test and consequently improves the empirical power of the test in small samples.

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 natural generalizations of the linear model $Y = X^T\theta + U$ where the index $X^T\theta$ has a linear influence on Y.

A single index model is defined by the equation

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

where F, the link function, is a real-valued function which operates possibly nonlinear on the index $X^T\theta$ and U is an error term. Assuming that $E(U|X = x) = 0$ model (1) entails

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

The single index model (2) can assume several particular forms allowing to model a variety of situations, e.g., binary responses, multinomial and ordered discrete responses, censored and truncation responses. In the context of binary choice models equation (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. Stoker (1992) describes econometric applications with a view towards semiparametric formulations. Examples are labour supply (Stoker 1992, van Soest 1992), work-trip mode choice (Horowitz 1993) and vacation expenditures (Melenberg and van Soest 1991). Survival analysis models fall into this single index models class, see Cox and Oakes (1984).

An importat element in the use of single index models is the choice of the link, that is, the specific functional form of F. There are often practical reasons for choosing a certain parametric specification for F. On the other hand, empirical evidence has shown that a parametric specification of F is not always adequate as in the study for transportation choice example presented by Horowitz (1993). It is therefore interesting to test the specification of the link F.

Horowitz and Härdle (1994) have developed a test for the specification of the function F. This socalled 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 later tests have less power because the explanatory variable X is highdimensional. The HH test avoids this "curse of dimensionality" by using the index function $X^T\theta$ to aggregate a multidimensional X.

Simulations with single index models have shown though that the power of the HH test is affected in small samples mainly due to the presence of a negative bias in the statistic, as is shown by Proença (1993) and Proença and Ritter (1993). Here, we study a bootstrap procedure with the aim to find a better aproximation in finite samples of the distribution of the HH statistic than the normal distribution. Simulations show that this bootstrap procedure works well and the critical values are better determined this way.

The plan of the paper is as follows. Section 2 describes shortly the HH test. Section 3 introduces the bootstrap and shows its asymptotic correctness. Section 4 reports the results of a Monte Carlo study with binary responses models showing the superior performance of this resampling procedure.

2 The model and the test statistic

Consider the following model for the conditional expectation of a random variable Y,

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

where $Y \in I\!R$, $x \in I\!R^k$, F and v are real-valued functions known up to a finite vector of parameters $\theta \in \Theta \subset I\!R^k$. Let $\{(X_1,Y_1),...,(X_n,Y_n)\}$ be a random sample of (X,Y). We also assume that $\hat{\theta}_n$ is a $\sqrt{n}$–consistent estimator of θ obtained with the given sample. A $\sqrt{n}$–consistent estimate is provided e.g. by the MLE for fixed F.

The problem addressed in this work is how to test the specification of the link function F assuming as given the functional form of $v(\bullet)$. This makes sense since in the stochastic utility approach, see e.g. Maddala (1983), the form of the index stems from theoretical considerations whereas the distribution of the errors is more taken as a convenience to be Normal. In this setting the functional form of the link is uncertain and hence should be tested. The test hypotheses are therefore

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

where $H(\bullet)$ is an unknown function.

The statistic presented by Horowitz and Härdle (1994) is motivated by the seminal work of Newey (1985) on conditional moment tests. The main idea relies on the fact that if the model (3) is true, a nonparametric estimation of $E\{Y|v(X,\hat{\theta}_n) = \hat{v}\}$ should be close to $F\{v(x,\hat{\theta}_n)\}$, the parametric estimate. On the other hand, if the link F is not correctly specified then the nonparametric estimate should significantly differ from the parametric The HH statistic has the form

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

Here, $\hat{F}_{ni}\{v(X_i, \hat{\theta}_n)\}$ is a leave-one-out kernel regression estimate of $E\{Y|v(X, \hat{\theta}_n) = \hat{v}\}$, h is the bandwidth used to obtain this estimate and $w\{v(X_i, \hat{\theta}_n)\}$ is a non-negative weight function. Asymptotic unbiasedness is achieved by using the method proposed by Bierens (1987) which suggests defining $\hat{F}_{ni}$ as a linear combination of two regression kernels with different bandwidths (respectively h and s) according to,

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

Here $h = cn^{-1/(2r+1)}$, $s = cn^{-\delta/(2r+1)}$ with $c > 0$, $0 < \delta < 1$ and $\hat{F}_{nti}(v)$ is a nonparametric kernel regression functional calculated without observation i and with bandwidth t, $(t = h, s)$, defined by

$$\hat{F}_{nti}(v) = \frac{\sum\limits_{\substack{j=1 \\ j \neq i}}^{n} Y_j K[\{v - v(X_j, \hat{\theta}_n)\}/t]}{\sum\limits_{\substack{j=1 \\ j \neq i}}^{n} K[\{v - v(X_j, \hat{\theta}_n)\}/t]} \qquad t = h, s \tag{7}$$

where $K(\bullet)$ is a kernel of order r $(r \geq 2)$.

Horowitz and Härdle (1994) show that T_n is asymptotically distributed as a $N(0, \sigma_T^2)$ variable with

$$\sigma_T^2 = 2C_K \int_{-\infty}^{\infty} w(v)^2 \{\sigma^2(v)\} dv \tag{8}$$

$$C_K = \int_{-\infty}^{\infty} K(u)^2 du \tag{9}$$

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

A consistent estimator for $\sigma_T^2(v)$ is given by

$$\hat{\sigma}_T^2 = \frac{2C_K}{n} \sum_{i=1}^{n} \frac{w\{v(X_i, \hat{\theta}_n)\}^2 [\hat{\sigma}^2\{v(X_i, \hat{\theta}_n)\}]^2}{\hat{p}_{nhi}\{v(X_i, \hat{\theta}_n)\}} \tag{11}$$

and,

$$\hat{p}_{nhi}(v) = (nh)^{-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} K\left\{\frac{v - v(X_j, \hat{\theta}_n)}{h}\right\}.$$

In practice $w(v)$ is defined to be equal to one in a subset of the fitted values of $v(\bullet)$ and zero otherwise. For estimating $\sigma^2(v)$ a consistent estimator under H_0 has to be used that should not become excessively large under H_1 in order to avoid a loss of power. In homoscedastic models two possible estimators (which are consistent under H_0) are

$$\hat{\sigma}_1^2 = \frac{1}{n}\sum_{i=1}^{n}[Y_i - F\{v(X_i,\hat{\theta}_n)\}]^2 \tag{12}$$

$$\hat{\sigma}_2^2 = \frac{1}{n}\sum_{i=1}^{n}[Y_i - \hat{F}_{ni}\{v(X_i,\hat{\theta}_n)\}]^2. \tag{13}$$

However, $\hat{\sigma}_1^2$ is likely to be larger than $\hat{\sigma}_2^2$ under H_1, and consequently this last estimator should be used in order to enhance the power of the test.

If the model is heteroscedastic with an unknown form of heteroscedasticity, $\sigma^2(v)$ may be estimated by a nonparametric regression of $\{Y_i - \hat{F}_{ni}(\bullet)\}^2$ on $v(X_i,\hat{\theta}_n)$. When the form of the heteroscedasticity is known it should be incorporated in the estimation of $\sigma^2(v)$ as in binary variables models where $\sigma^2(v) = P\{Y = 1|v(X,\theta) = v\}[1 - P\{Y = 1|v(X,\theta) = v\}]$ which can be estimated by $\hat{F}_{ni}\{v(X_i,\hat{\theta}_n)\}[1 - \hat{F}_{ni}\{v(X_i,\hat{\theta}_n)\}]$.

Another alternative procedure to estimate $\sigma^2(v)$ is to use the estimator of Gasser, Sroka and Jennen-Steinmetz (1986). This is a nonparametric estimator based on local linear fitting which can also be used in the presence of heteroscedasticity.

Despite the asymptotic consistency of the HH test there is no theoretical result about the power of the test in finite samples. Since the set of alternatives includes all real-valued functions of $v(X,\theta)$ the power in finite samples is likely to be small. Simulations studies must therefore be performed in order to analyse the empirical power for some particular situations, see Proença (1993) and Proença and Ritter (1993). These studies reveal that the asymptotic normal distribution of the HH statistic shows a lack of accuracy on approximating the true distribution in samples including to 1000 observations with a significant loss of power of the test.

Bootstrap is a procedure that can improve approximations in finite samples when analytical corrections are hard to deduce or to implement. The next section introduces a bootstrap procedure that aims to find better critical values for the HH test in small samples than those provided by the standard normal approximation.

3 The bootstrap approach

The main idea of the bootstrap approach is to mimic the model under the null in order to find an approximation for the distribution of the HH statistic when model (3) is true. The aim is to find automatic and possibly more accurate critical values.

Following Hall (1992) a replica of the true model is built by replacing its unknown parameters by consistent estimates obtained within the sample. In the bootstrap world (the bootstrap resamples) inference is done conditional on the original sample. The model under study described in (3) can be formulated as

$$Y = F\{v(X,\theta)\} + U$$
$$E(U|X=x) = 0.$$

Under the null only the vector of parameters θ is unknown (in some situations possibly there are other unknown parameters like the variance of U as an example). According to what has been said the following class of bootstrap procedures can be defined. Consider the bootstrap variable Y^* satisfying,

$$Y^* = F\{v(X,\hat{\theta}_n)\} + U^*, \tag{14}$$

such that,

$$E^*(U^*|X=x) = 0 \tag{15}$$

$$E^*\{U^{*2}|v(X,\hat{\theta}_n) = v\} = \sigma^{*2}(v) = \sigma^2\{v(x,\hat{\theta}_n)\} \tag{16}$$

where E^* denotes the expectation conditional on $\{(X_i,Y_i)\}_{1=1\cdots n}$. Here $F(\bullet)$, $v(\bullet)$ and $\sigma^2(\bullet)$ are the same functions as before and X is the same random variable as in model (3). Equivalent to (15) and (16) are

$$E^*(Y^*|X=x) = F\{v(x,\hat{\theta}_n)\} \tag{17}$$

$$V^*\{Y^*|v(X,\hat{\theta}_n)=v\} = \sigma^2(v). \tag{18}$$

where V^* denotes the variance conditional on $\{(X_i,Y_i)\}_{1=1\cdots n}$. Condition (18) means that the bootstrap model has the same variance structure as the true model.

Suppose now that (X_i,Y_i^*), $i=1,...,n$ is a random bootstrap sample of (X,Y^*). Denote $\hat{\theta}_n^*$ a $\sqrt{n}$–consistent for $\hat{\theta}_n$. This can be the MLE in the bootstrap for instance, it will be $\sqrt{n}$–consistent because the bootstrap resamples were generated under the parametric model (14). The bootstrap replica of the HH statistic is

$$T_n^* = \sqrt{h}\sum_{i=1}^{m} w\{v(X_i,\hat{\theta}_n^*)\}[Y_i^* - F\{v(X_i,\hat{\theta}_n^*)\}][\hat{F}_{ni}^*\{v(X_i,\hat{\theta}_n^*)\} - F\{v(X_i,\hat{\theta}_n^*)\}], \tag{19}$$

where $w(\bullet)$, h have the same meaning as before and $\hat{F}_{ni}^*(\bullet)$ is calculated as $\hat{F}_{ni}(\bullet)$ but with Y_i^* instead of Y_i and $v(\bullet,\hat{\theta}_n^*$ instead of $v(\bullet,\hat{\theta}_n^*)$.

The bootstrap is asymptotically correct if T_n^* has the same asymptotic distribution as T_n. The following theorem uses assumptions A1–A9 stated in the appendix.

Theorem 1 *Under assumptions A1–A9 from the appendix the probability that*

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

converges to 1.

The proof of the theorem follows closely that of Horowitz and Härdle (1994) with somewhat different arguments because bootstrap expectations are taken conditional on the sample. A crucial feature to put the asymptotic theory into work is the fact that $\sigma^{*2}(v)$ converges in probability to $\sigma^2(v)$. In order to achieve this conditions (17) and (18) are required to be verified. The last condition implies that the conditional variance function of Y is known. This happens for example with binomial or multinomial responses models. When the conditional variance structure is not known a priori wild bootstrap as in Härdle and Mammen (1993) can be used instead.

To prove that the bootstrap distribution of the HH statistic constitutes a better approximation than the standard normal it is necessary to go further. One way would be to apply Edgeworth expansions following Hall (1992). This is the subject of further work. In the following the investigation of the quality of the bootstrap approximation will be done by a simulation study.

4 A parametric bootstrap for binary responses

The aim of the bootstrap is to find a better approximation to the distribution of HH statistic under the null. Thus, is performed under the null hypothesis.

The parametric bootstrap for binary responses is implemented according to the following algorithm:

1. Given a (pseudo) sample (x_i, y_i), $i = 1, ..., n$ calculate the ML estimate for the parameter vector,$\hat{\theta}_n$. Here ML with bias reduction according to Firth (1992) is advised to be used. Determine the statistic T_n given in equation (5) together with the estimated $\hat{\sigma}_T$.

2. Generate B bootstrap samples with size n by drawing y_i^* from the Bernoulli distribution with parameter $\{1+\exp(-\hat{v}_i)\}^{-1}$ with $\hat{v}_i = x_i^T\hat{\theta}_n$ and calculate the HH statistic T_n^* according to (19) together with the estimate of its standard deviation, $\hat{\sigma}_T^*$ obtained from (10) replacing $\hat{\theta}_n$ by $\hat{\theta}_n^*$. The bootstrap estimate of the parameter vector is obtained by performing one iteration in the M.L. algorithm from starting point $\hat{\theta}_n$ (using y_i^* and x_i as variables). Mosbach (1992) proves that this estimate is also $\sqrt{n}$–consistent.

3. Calculate the bootstrap critical value c_α^* for a test of size α. Hall and Titterington (1989) give a formula for c_α^* in one-sided test based on the following.

First order the B values $T^*/\hat{\sigma}_T^*$ obtaining $t^*_{[1]} \cdots t^*_{[B]}$. Consider $c^*_\alpha = t^*_{[M]}$ with

$$M = (B+1)(1-\alpha) \tag{20}$$

4. Repeat all the previously procedures S times obtaining S values for the standardized statistic

$$(T/\hat{\sigma}_T)_1, ..., (T/\hat{\sigma}_T)_S$$

and S bootstrap critical values $c^{*1}_\alpha, ..., c^{*S}_\alpha$.

5. Finally the bootstrap empirical power is given by,

$$\#\{(T/\hat{\sigma}_T)_i > c^{*i}_\alpha, i = 1, ..., S\}.$$

5 Monte Carlo study

The hypothesis H_0 to be tested is,

$$\begin{aligned} H_0 \quad : \quad & E\{Y|v(X,\theta) = v\} = \{1 + \exp(-v)\}^{-1} = F(v) \\ & v(X,\theta) = X^T\theta \end{aligned} \tag{21}$$

The data generating processes considered are the following. Let the exogenous variables X_1 and X_2 be independent randomly generated from a standard normal distribution. Let the function $v(x,\theta)$ be $1 - x_1 + 2x_2$. Variable Y is generated by the following conditional expectation functions

$$E[Y|v(X,\theta) = v] \quad = \quad \{1 + \exp(-v)\}^{-1} + \frac{-v}{1.5} \times \frac{A}{1.5} \times \varphi\left(\frac{v}{1.5}\right), \tag{22}$$

$$E[Y|v(X,\theta) = v] \quad = \quad 1 - \exp\{-\exp(v)\}. \tag{23}$$

In this expression $\varphi(\bullet)$ is the standard normal density. The experiment (22) consists of a logit perturbed by a bump with height given by A. For A the values 0 and 1.25 were considered. The value 1.25 was chosen guaranteeing that the conditional

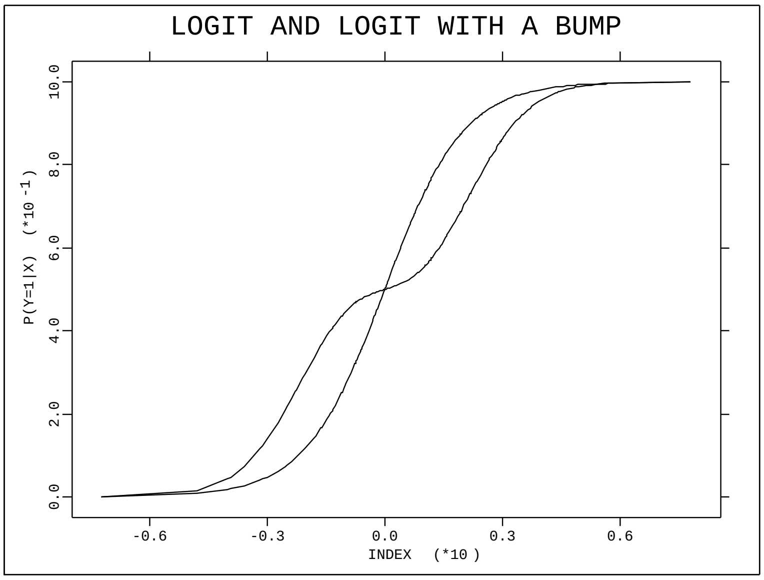

Figure 1: *The logit model versus the logit with a bump.*

expectation of Y will never decrease when v increases. Note that when $A = 0$ the function is a logit model and H_0 is true. Design (23) is a complementary log-log model. The shapes of these functions are presented in figure 1 and figure 2. For sake of comparison the data was generated using the same values for the explanatories in all designs.

In conducting the HH test, the values for $h = 1$ and $h = 0.3$ have been chosen. This choice was made after a graphical analysis of the kernel regression estimate. The value $h = 1$ oversmooths a little bit the data for the logit design and logit with a bump while $h = 0.3$ undersmooths the data. Note that the aim is not to find the best semiparametric estimate from the data but to have a tool that detects deviations from the assumed link function.

Figures 3, 4 and 5 show the true function and the respective leave–one–out kernel estimate for $h = 1$ and maximum likelihood estimate assuming a logit model for one sample. In figure 4 the data is generated under the null. The maximum likelihood

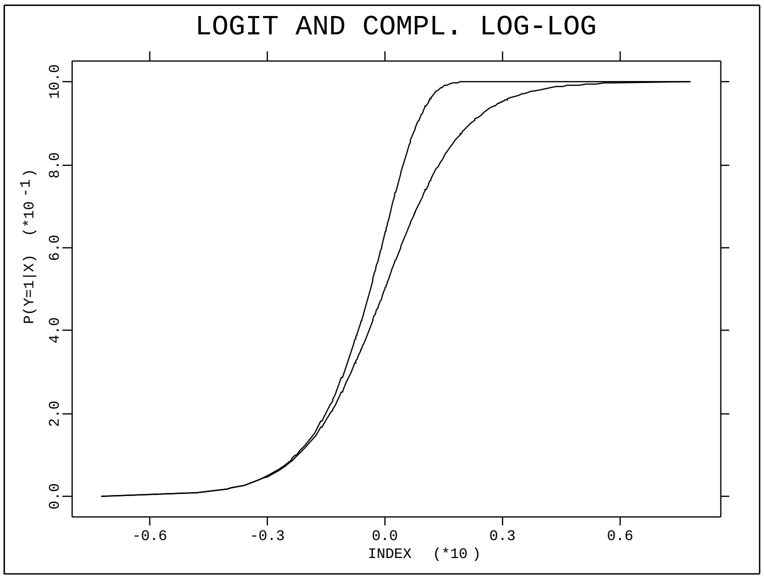

Figure 2: *The logit model versus the complementary log-log.*

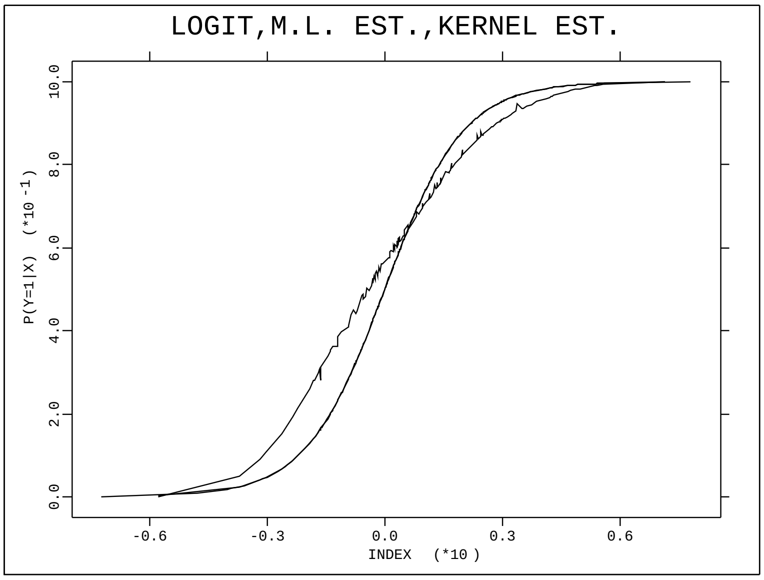

Figure 3: *The logit link together with the maximum likelihood estimate and the leave-one-out kernel regression* $(h = 1)$.

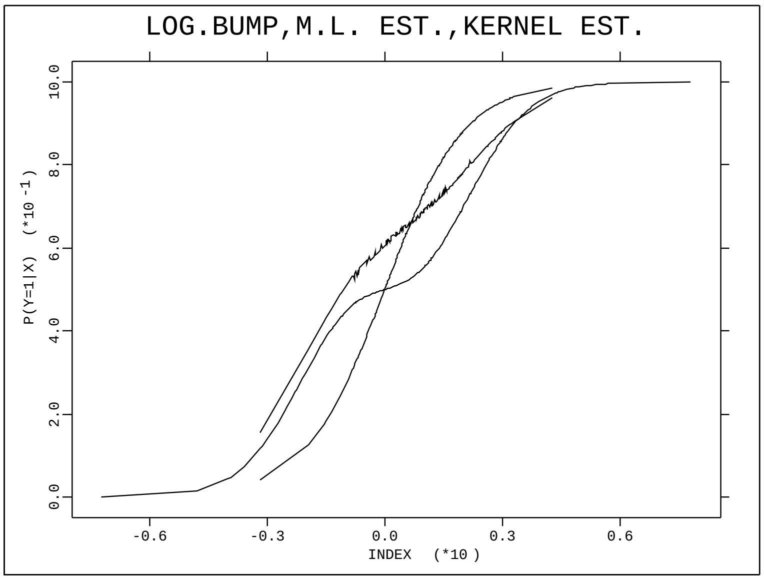

Figure 4: *The logit with a bump link together with the maximum likelihood estimate and the leave-one-out kernel regression ($h = 1$).*

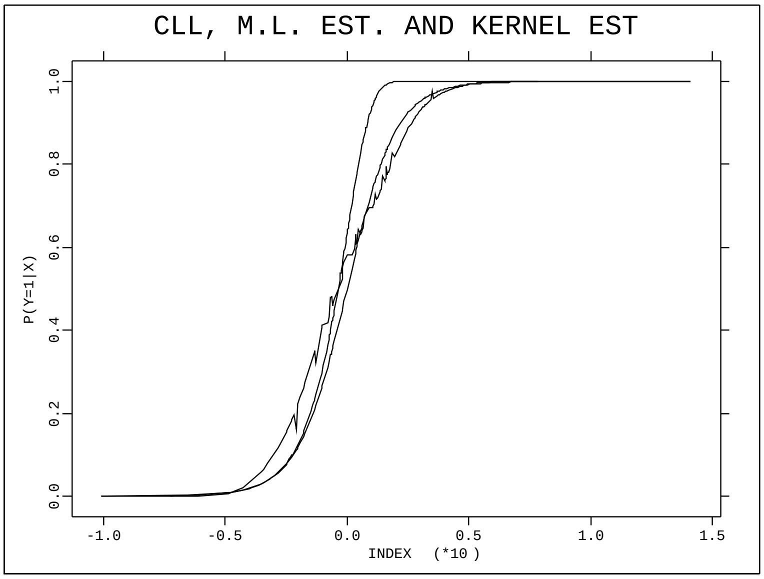

Figure 5: *The complementary log-log link together with the maximum likelihood estimate and the leave-one-out kernel regression ($h = 1$).*

estimate coincides almost exactly with the true function while the kernel estimate deviates moderately from the true due to oversmoothing. Figure 4 shows the situation when the data is generated from a logit with a bump. Here, the maximum likelihood estimate is completely unable to detect the shape of the true function. On the other side, the kernel estimate deviates considerably from the maximum likelihood estimate capturing the misspecification.

In figure 5 data was generated from a complementary log-log model. Once more the maximum likelihood estimate is not able to detect the deviation in the upper tail of the true function. However the kernel estimate follows very closely the maximum likelihood estimate inducing a poor performance of the HH test in this problem. As it was already noticed before, the kernel estimate has a tendency to follow the maximum likelihood estimate specially when the bandwidth is small.

The weight function $w(v)$ has been considered identically equal to one within an interval limited by the 95% and the 5% percentiles of the ordered fitted values of v.

All the simulations have been performed in GAUSS using $S = 500$ pseudo data sets and $B = 199$ bootstrap replications with sample size $n = 200$. B was chosen according to (20) in order to have $M = 190$ for a one-sided test at 10% level. The pictures were made in XploRe 3.1 (see XploRe (1993)).

The results obtained using standard normal critical values and bootstrap critical values are shown in Table I. Together with a two-sided test a one-sided test was also tried (e.g. using the 90% percentile to test at a 10% significance level). The one-sided test is motivated by the fact that the HH statistic has mean bigger than zero under the alternative and consequently this test can achieve higher power.

The empirical size of the test is measured by the percentage of rejections of the logit model ($A = 0$) in the 500 trials. The empirical size of the test given by the normal critical values is considerable below the nominal size of 10% which is

Table I: *Percentages of rejections using normal critical values and bootstrap critical values. Nominal size is* 10%

		normal	bootstrap
		A=0	
h = 0.3	two-sided	4.4	10.6
h = 0.3	one-sided	2.4	7.4
h = 1.0	two-sided	1.4	8.4
h = 1.0	one-sided	0.0	15.0
		A=1.25	
h = 0.3	two-sided	9.0	23.4
h = 0.3	one-sided	12.0	35.4
h = 1.0	two-sided	1.8	48.6
h = 1.0	one-sided	4.6	60.0
		CLL	
h = 0.3	two-sided	5.6	11.2
h = 0.3	one-sided	5.8	11.4
h = 1.0	two-sided	1.0	21.0
h = 1.0	one-sided	1.4	30.6

better approximated by the bootstrap test. This is clearly the consequence of the negative bias observed in the statistic. The empirical power of the classic HH test (percentage of rejections for $A = 1.25$ and for the CLL) is completely inexistant in this example which was expected given that the sample size is small. The bootstrap achieves remarkably better empirical power specially for bandwidth $h = 1$. For this bandwidth and despite the small size of the sample the rejection rate for the logit with bump can be considered very satisfactory. As expected one-sided test leads to an improvement of the power.

The results are in agreement with the studies of Proença (1993) and Proença and Ritter (1993) and show again the negative bias in the statistic. The statistic is slightly sensitive to the bandwidth under the null. For the type of alternatives used, undersmoothing induces a significant loss on the power of the test. These results show clear that the bootstrap is beneficial in this problem.

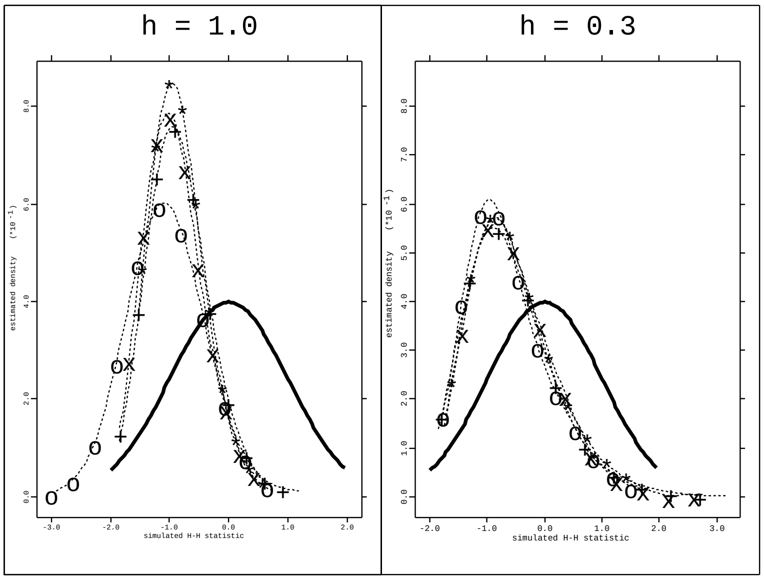

Figure 6: *This figure shows the simulated density of the HH statistic (plus), the bootstrap density under the null (cross), the bootstrap density under the alternative logit with bump (circle) and the bootstrap density under the alternative complementary log-log (star) for one sample (randomly chosen). The densities in the left window refer to the HH test conducted with bandwidth equal to 1 while for the densities in the right window the HH test was conducted with bandwidth equal to 0.3.*

The shape of the simulated density of the standardized HH statistic under H_0 together with the shapes of the bootstrap densities for each experiment are presented in figure 6. They correspond to one sample of the 500 pseudo-samples randomly chosen. The density estimates were obtained using kernel smoothing with a bandwidth of 0.8 and a quartic kernel. The standard normal density is also included as a fat solid line in the same figure. Again the negative bias and the bootstrap remedy are obvious.

The simulated density of the HH statistic is undoubtedly distinct from the standard normal having a clear mode left of zero. The bootstrap densities are very close to the simulated HH density whether under the null or the alternative (specially when the HH test is conducted with undersmoothing) supporting the believe of accurate bootstrap critical values. For $h = 1$ the bootstrap density under the logit with a bump deviates a little from the HH simulated density in the left tail and in kurtosis. However this deviation may not have consequences on defining inaccurate critical values specially for the one-sided test. The figure reveals that the distribution of the HH statistic is bandwidth dependent and is negatively biased compared to the standard normal. Bootstrap is very successful on providing a better approximation to the true distribution of the HH statisitc than the standard normal.

6 Acknowledgements.

We would like to thank Joel Horowitz, Rodney Wolff and Enno Mammen 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 nonparametric regression for model checking, *Biometrika* **76**: 1–11.

Bierens, H. J. (1987). Kernel estimators of regression functions, *in* T. F. Bewley (ed.), *Advances in Econometrics: Fifth world congress*, Vol. 1, Cambridge University Press, New York, pp. 99–144.

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

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

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

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

Hall, P. (1992). *The Bootstrap and the Edgeworth Expansion*, Springer, New York.

Hall, P. and Titterington, D. M. (1989). The effect of simulation order on level accuracy and power of monte carlo tests, *Journal of the Royal Statistic Society, Series B* **51**(3): 459–467.

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

Horowitz, J. and Härdle, W. (1994). Testing a parametric model against a semiparametric alternative, *Econometric theory.*

Horowitz, J. L. (1993). Semiparametric estimation of a work-trip mode choice model, *Journal of Econometrics* **58**(1/2): 49–70.

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*, Cambridge University Press, Cambridge.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*, 2nd edn, Chapman and Hall, London.

Melenberg, B. and van Soest, A. (1991). Parametric and semiparametric modelling of vacation expenditures, *Discussion Paper 9144*, CentER, Tilburg University, The Netherlands.

Mosbach, O. (1992). One–step bootstrapping in generalized linear models, *in* K.-H. Jöckel, G. Rothe and W. Sendler (eds), *Bootstrap and related techniques. Proceedings of an International Conference held in Trier, FRG, 1990*, Springer–Verlag, New York, pp. 143 – 147.

Newey, W. (1985). Maximum likelihood specification testing and conditional moment test, *Econometrica* **53**: 1047–1070.

Proença, I. (1993). On the performance of the h-h test, *Discussion Paper 9305*, CORE, Université Catholique de Louvain.

Proença, I. and Ritter, C. (1993). Negative bias in the h-h statistic, *Discussion Paper 9321*, Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin.

Stoker, T. M. (1992). *Lectures on Semiparametric Econometrics*, CORE Lecture Series, Louvain-La-Neuve.

van Soest, A. (1992). Discrete choice models of family labour supply, *Discussion Paper 9214*, CentER, Tilburg University, The Netherlands.

XploRe (1993). *XploRe – an Interactive Statistical Computing Environment*, XploRe Systems, Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin.

A Proofs

To begin with lets introduce the assumptions. The following set of assumptions are the same as in Horowitz and Härdle (1994).

A1 - The support of X, S_x is a compact set. At least one component of X has a probability distribution that is absolutely continuous with respect to Lebesgue measure.

A2 - For every $x \in S_x$ and $\tau \in N_\theta$ with N_θ a neigbourhood of θ that contains $\hat{\theta}_n$,

(a) $|v(x,\tau)| < M$ for some $M < \infty$ that does not depend on τ or x.

(b) The probability distribution of $v(x,\tau)$ is absolutely continuous with respect to Lebesgue measure.

(c) $v(x,\tau)$ is continuously differentiable with respect to τ and

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

for some M that does not depend on τ or x

A3 - Let p_τ denote the probability density function of $v(X,\tau)$. For each $\tau \in N_\theta$

(a) $m_p \leq p_\tau(v) \leq M_p$ for some $m_p > 0$ and $M_p < \infty$ that do not depend on τ.

(b) p_τ has r continuous derivatives that are uniformly bounded over $\tau \in N_\theta$ and $v \in \tilde{S}_v$ with $\tilde{S}_v$ an open subset of the support of v, S_v.

A4 - $w(\bullet)$ has compact support $S_w \subset int(S_v)$ and satisfies:

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

(b) $|w(v_2) - w(v_1)| \leq M_w^* |v_2 - v_1|$ for some $M_w^* < \infty$ and all v_2, v_1.

A5 - (a) $|Fv(x,\tau)|$ is uniformly bounded over $x \in S_x$ and $\tau \in N_\theta$.

(b) $F(v)$ has r continuous derivatives that are uniformly bounded over $v \in S_v$.

A6 - Let E_x denote the expectation over the distribution of X. Define

$$\Gamma(x,v,\tau) = E_X[F\{v(X,\tau)\}\{\partial v(x,\tau)/\partial\tau - \partial v(X,\tau)/\partial\tau\}|v(x,\tau) = v]$$

Let $\Gamma_j, j = 1, \cdots, k$ denote the j'th component of Γ. There is a finite number M_Γ not depending on τ or x such that for all $\tau \in N_\theta, x \in S_x, v_1, v_2 \in \tilde{S}_v$ and $j = 1, \cdots, k$

$$|\Gamma_j(x, v_2, \tau) - \Gamma_j(x, v_1, \tau)| \leq M_\Gamma |v_2 - v_1|$$

A7 - $\sigma^2(v) = V\{Y|v(x,\theta) = v\}$ is a uniformly bounded, continuous function of $v \in S_v$. $E[Y - E\{Y|v(x,\theta)\}]^4$ is uniformly bounded over $v \in S_v$.

A8 - K is an r'th order kernel with bounded support. Also, K is uniformly bounded, continuous, and symmetrical about 0. The derivative of K, K', is uniformly bounded, has an absolutely integrable Fourier transform, $\Psi(\bullet)$ and satisfies

$$\int_{-\infty}^{\infty} |uK'(u)|du < \infty$$

Next assumption is specific to the bootstrap.

A9 - $\sigma^{*2}(v) = V^*\{Y^*|v(x,\hat{\theta}_n) = v\}$ is a uniformly bounded, continuous function of $v \in S_v$. $E^*[Y^* - E^*\{Y^*|v(x,\hat{\theta}_n)\}]^4$ is uniformly bounded over $v \in S_v$.

B proof of theorem

Beside the notation already introduced the following notation is also going to be used in this proof. Consider $p_{nti}(v)$ a kernel estimate of the density of $v(x,\tau)$ with bandwidth t given by

$$p_{nti} = \frac{1}{nt}\sum_{\substack{j=1\\ j\neq i}}^{n} K\left[\frac{v - v(X_j,\tau)}{t}\right].$$

Define G_{nti}, J_{nti}, G_{ni} and J_{ni} for a $v = v(x,\tau)$ respectively as

$$\begin{aligned}
G_{nti}(v) &= \{\hat{F}_{nti}(v)p_{nti}(v) - F(v)p_{nti}(v)\}/p_\tau(v) \\
J_{nti}(v) &= \frac{\{p_{nti}(v) - p_\tau(v)\}}{p_\tau^2}\{\hat{F}_{nti}(v)p_{nti}(v) - F(v)p_{nti}(v)\} \quad t = h, s \\
G_{ni}(v) &= \left[G_{nti}(v) - \left(\frac{h}{s}\right)^r G_{nsi}(v)\right] \Big/ \left[1 - \left(\frac{h}{s}\right)^r\right] \\
J_{ni}(v) &= \left[J_{nti}(v) - \left(\frac{h}{s}\right)^r J_{nsi}(v)\right] \Big/ \left[1 - \left(\frac{h}{s}\right)^r\right]
\end{aligned} \tag{24}$$

Define the following events. For all $\epsilon > 0$ there is a $c < \infty$ such that,

$$P\left\{\sup_i \sup_v |p_{nhi}(v) - p_\tau(v)| \le c\left(\frac{\log n}{nh}\right)^{1/2}\right\} \ge 1 - \epsilon \qquad (E^\epsilon_{1n})$$

$$P[\sup_x \sqrt{nh}|p_{nhi}\{v(x,\tau_n)\} - p_\tau\{v(x,\tau)\}| \le c\sqrt{h}] \ge 1 - \epsilon \qquad (E^\epsilon_{2n})$$

$$P\left\{\sup_i \sup_v \sqrt{nh}|\hat{F}_{ni}(v) - F(v) - G_{ni}(v) + J_{ni}(v)| \le c\left(\frac{\log n}{n^{3/2}h^2}\right)\right\} \ge 1 - \epsilon \qquad (E^\epsilon_{3n})$$

$$P[\sup_i \sup_v \sqrt{nh}|\hat{F}_{ni}\{v(x,\tau_n)\} - \hat{F}_{ni}\{v(x,\tau)\}| \le c\sqrt{h}] \ge 1 - \epsilon \qquad (E^\epsilon_{4n})$$

where τ_n is a nonstochastic sequence in $I\!R^k$ that verifies $\sqrt{n}(\tau_n - \tau) = O(1)$.

We follow closely the proof in Horowitz and Härdle (1994). In a first step it will be shown that the statistic evaluated at the bootstrap estimate $\hat{\theta}^*_n$ is equal to the statistic evaluated at the true parameter in the bootstrap world, $\hat{\theta}_n$, plus something that converges in probability to zero. This is the subject of the next lemma.

Lemma 1 *Define,*

$$\tilde{T}^*_n = \sqrt{h}\sum_{i=1}^n w\{v(X_i,\hat{\theta}_n)\}[Y^*_i - F\{v(X_i,\hat{\theta}_n)\}][\hat{F}_{ni}\{v(X_i,\hat{\theta}_n)\} - F\{v(X_i,\hat{\theta}_n)\}].$$

Then as $n \to \infty$,

$$T^*_n = \tilde{T}^*_n + o_p(1).$$

Proof:

Consider θ_n an arbitrary nonstochastic sequence verifying $\sqrt{n}(\theta_n - \hat{\theta}_n) = O(1)$. Because θ_n is arbitrary $T^*_n - \tilde{T}^*_n = o_p(1)$ is verified if $\overline{T}^*_n - \tilde{T}^*_n = o_p(1)$ with

$$\overline{T}^*_n = \sqrt{h}\sum_{i=1}^n w\{v(X_i,\theta_n)\}[Y^*_i - F\{v(X_i,\theta_n)\}][\hat{F}_{ni}\{v(X_i,\theta_n)\} - F\{v(X_i,\theta_n)\}].$$

Define $V_i = v(X_i,\theta_n)$, $\hat{V}_i = v(X_i,\hat{\theta}_n)$, $\hat{U}^*_i = Y^*_i - F(\hat{V}_i)$ and $U^*_i = Y^*_i - F(V_i)$. After some algebra it follows

$$\overline{T}^*_n - \tilde{T}^*_n = \sum_{j=1}^6 R_{nj},$$

with,

$$\begin{aligned}
R_{n1} &= \sqrt{h}\sum_{i=1}^{n} w(V_i)\hat{U}_i^*\{\hat{F}_{ni}(V_i)-\hat{F}_{ni}(\hat{V}_i)\}\\
R_{n2} &= \sqrt{h}\sum_{i=1}^{n}\{w(V_i)-w(\hat{V}_i)\}\hat{U}_i^*\{\hat{F}_{ni}(\hat{V}_i)-F(\hat{V}_i)\}\\
R_{n3} &= -\sqrt{h}\sum_{i=1}^{n} w(V_i)\hat{U}_i^*\{F(V_i)-F(\hat{V}_i)\}\\
R_{n4} &= -\sqrt{h}\sum_{i=1}^{n} w(V_i)\{F(V_i)-F(\hat{V}_i)\}\{\hat{F}_{ni}(V_i)-\hat{F}_{ni}(\hat{V}_i)\}\\
R_{n5} &= -\sqrt{h}\sum_{i=1}^{n} w(V_i)\{F(V_i)-F(\hat{V}_i)\}\{\hat{F}_{ni}(\hat{V}_i)-\hat{F}(\hat{V}_i)\}\\
R_{n6} &= \sqrt{h}\sum_{i=1}^{n} w(V_i)\{F(V_i)-F(\hat{V}_i)\}^2
\end{aligned}$$

We will prove that all these terms are $o_p(1)$. We begin with $R_{n1}=o_p(1)$. Define A_{ni}^{ϵ} as the event resulting from the intersection of the events $p_{nhi}\{v(x,\theta_n)\}>\epsilon$, $p_{nsi}\{v(x,\theta_n)\}>\epsilon$ uniformly over $x\in\tilde{S}_x=\{x : v(x,\hat{\theta}_n)\in S_v\}$ and

$$\sup_{x\in\tilde{S}_x}\sqrt{h}|\hat{F}_{ni}\{v(x,\theta_n)\}-\hat{F}_{ni}\{v(x,\hat{\theta}_n)\}|\leq\epsilon$$

Let $A_n^{\epsilon}=\cup_{i=1}^{n}A_{ni}^{\epsilon}$ and define $I(\bullet)=1$ if the event in parentheses occurs and $I(\bullet)=0$ otherwise. Conditional on events $E_{1n}^{\epsilon}, E_{2n}^{\epsilon}$ and E_{4n}^{ϵ} $P[I({A_n^{\epsilon}}^C)]=o(1)$ and n is sufficiently large. Therefore,

$$R_{n1}=R_{n11}+o_p(1) \tag{25}$$

with

$$R_{n11} = \sqrt{h}\sum_{i=1}^{n} I(A_{ni}^{\epsilon})w(V_i)\hat{U}_i^*\{\hat{F}_{ni}(V_i)-\hat{F}_{ni}(\hat{V}_i)\}$$

$E^*(R_{n11})=0$ because $I(A_{ni}^{\epsilon})$ does not depend on $\hat{U}_i^*$.

$$\begin{aligned}
V^*(R_{n11}) &= n^{-1}E^*\sum_{i=1}^{n} I(A_{ni}^{\epsilon})w(V_i)^2\hat{U}_i^{*2}\{\hat{F}_{ni}(V_i)-\hat{F}_{ni}(\hat{V}_i)\}^2 + \\
&\quad n^{-1}\sum_{i}^{n}\sum_{j\neq i}^{n} I(A_{ni}^{\epsilon})I(A_{nj}^{\epsilon})w(V_i)w(V_j)\hat{U}_i^*\hat{U}_j^*\times\\
&\quad \{\hat{F}_{ni}(V_i)-\hat{F}_{ni}(\hat{V}_i)\}\{\hat{F}_{nj}(V_j)-\hat{F}_{nj}(\hat{V}_j)\}
\end{aligned}$$

Now it will be shown that

$$\sqrt{nh}E^*\{I(A^{\epsilon}_{nj})w\{v(x,\theta_n)\}\hat{U}^*_i[\hat{F}_{nj}\{v(x,\theta_n)\} - \hat{F}_{nj}\{v(x,\hat{\theta}_n)\}]\} = O_p(\{nh^{1/2}\}^{-1}). \tag{26}$$

First note that

$$\begin{aligned}
&\sqrt{nh}E^*\{I(A^{\epsilon}_{nj})w\{v(x,\theta_n)\}\hat{U}^*_i[\hat{F}_{nhj}\{v(x,\theta_n)\} - \hat{F}_{nhj}\{v(x,\hat{\theta}_n)\}]\} = \\
&(\sqrt{nh})^{-1}I(A^{\epsilon}_{nj})w\{v(x,\theta_n)\}\sigma^{*^2}(\hat{V}_i)\left|\frac{1}{p_{nhj}\{v(x,\theta_n)\}}K\left(\frac{\{v(x,\theta_n) - v(X_i,\theta_n)\}}{h}\right) - \right. \\
&\left.\frac{1}{p_{nhj}\{v(x,\hat{\theta}_n)\}}K\left(\frac{\{v(x,\hat{\theta}_n) - v(X_i,\hat{\theta}_n)\}}{h}\right)\right|
\end{aligned} \tag{27}$$

Using a Taylor expansion the expectation (27) comes

$$\begin{aligned}
&(nh)^{-1/2}\sigma^{*2}(\hat{V}_i)\left|K'\left\{\frac{v(x,\theta^0_n) - v(X_i,\theta^0_n)}{h}\right\}O(1/\{n^{1/2}h\}) + \right. \\
&\left.K\left\{\frac{v(x,\hat{\theta}_n) - v(X_i,\hat{\theta}_n)}{h}\right\}\left([p_{nhj}\{v(x,\theta_n)\}]^{-1} - [p_{nhj}\{v(x,\hat{\theta}_n)\}]^{-1}\right)\right| I(A^{\epsilon}_{nj})w\{v(x,\theta_n)\}
\end{aligned}$$

where θ^0_n lies between $\hat{\theta}_n$ and θ_n. Applying the same arguments as Horowitz and Härdle (1994)

$$E_X\left|K'\left(\frac{v(x,\theta^0_n) - v(X_i,\theta^0_n)}{h}\right)\right| = O(h)$$

and

$$E_X\left|K\left\{\frac{v(x,\tau) - v(X_i,\tau)}{h}\right\}\right| = O(h)$$

Uniformly over x. Consequently by Markov inequality,

$$K\left\{\frac{v(x,\tau) - v(X_i,\tau)}{h}\right\} = O_p(h) \tag{28}$$

and $K'(\bullet) = O_p(h)$. Given (28)

$$K\left\{\frac{v(x,\hat{\theta}_n) - v(X_i,\hat{\theta}_n)}{h}\right\} = O_p(h).$$

On the other side conditional on event E^{ϵ}_{2n}

$$|p_{nhj}\{v(x,\theta_n)\} - p_{nhj}\{v(x,\hat{\theta}_n)\}| = O_p(n^{-1/2}h^{-1}).$$

A similar argument applies when h is substituted by s. The result (26) follows from combining the results above with the fact that s converges faster to zero than h. Therefore,

$$\begin{aligned} V^*(R_{n11}) &\leq n^{-1}\sum_{i=1}^{n} I(A_{ni}^{\epsilon})w(V_i)^2\sigma^{*2}(\hat{V}_i)\epsilon^2 + o_p(1) \\ &\leq c\epsilon^2 n^{-1}\sum_{i=1}^{n} I(A_{ni}^{\epsilon}) + o_p(1) \\ &\leq c\epsilon^2 + o_p(1). \end{aligned}$$

Since ϵ is arbitrary it follows from Chebyshev inequality that R_{n11} is $o_p(1)$ and from (25) that $R_{n1} = o_p(1)$. Next it will be proved that $R_{n2} = o_p(1)$.
Remark that conditional on event E_{n3}^{ϵ}

$$\hat{F}_{ni}(\hat{V}_i) - F(\hat{V}_i) \leq G_n i(\hat{V}_i) - J_{ni}(\hat{V}_i) + O_p\left(\frac{\log n}{n^{3/2}h^2}\right)$$

Therefore,

$$R_{n2} = R_{n21} + o_p(1) \tag{29}$$

with

$$R_{n21} = \sqrt{h}\sum_{i=1}^{n}\{w(V_i) - w(\hat{V}_i)\}\hat{U}_i^*\{G_{ni}(\hat{V}_i) - J_{ni}(\hat{V}_i)\}.$$

$E^*(R_{n21}) = 0$ because G_{ni} and J_{ni} do not depend on $\hat{U}_i^*$. On the other side,

$$\begin{aligned} V^*(R_{n21}) = \; & h\sum_{i=1}^{n} E^*[\{w(V_i) - w(\hat{V}_i)\}^2\hat{U}_i^{*2}\{G_{ni}(\hat{V}_i) - J_{ni}(\hat{V}_i)\}^2] + \\ & h\sum_{i=1}^{n}\sum_{j\neq i}^{n} E^*[\{w(V_i) - w(\hat{V}_i)\}\{w(V_j) - w(\hat{V}_j)\}\hat{U}_i^*\hat{U}_j^* \\ & \{G_{ni}(\hat{V}_i) - J_{ni}(\hat{V}_i)\}\{G_{nj}(\hat{V}_j) - J_{nj}(\hat{V}_j)\}]. \end{aligned}$$

The following will analyse $E^*\{\hat{U}_i^* G_{nj}(\hat{V}_j)\}$.

$$\begin{aligned} E^*\{\hat{U}_i^* G_{nhj}(\hat{V}_j)\} &= E^*\left[U_i^*\left\{\frac{1}{nh}\sum_{i\neq j} K\left(\frac{\hat{V}_i - \hat{V}_j}{h}\right)\hat{U}_i^* - \right.\right. \\ &\qquad \left.\left. F(\hat{V}_j)\frac{1}{nh}\sum_{i\neq j} K\left(\frac{\hat{V}_i - \hat{V}_j}{h}\right)\right\}\right]\frac{1}{p(\hat{V}_j)} \\ &= \frac{1}{p(\hat{V}_j)}\sigma^{*2}(\hat{V}_i)\left\{K\left(\frac{\hat{V}_i - \hat{V}_j}{h}\right)\right\}\frac{1}{nh} \\ &= O_p(n^{-1}). \end{aligned}$$

The same is valid for $E^*\{\hat{U}_i^* G_{nsj}(\hat{V}_j)\}$ and consequently,

$$E^*\{\hat{U}_i^* G_{nj}(\hat{V}_j)\} = O_p(n^{-1}) \tag{30}$$

On the other side,

$$\begin{aligned} E^*[\hat{U}_i^* J_{nhj}(\hat{V}_j)] &= E^*\left[\hat{U}_i^*\left\{\frac{p_{nhj}(\hat{V}_j) - p(\hat{V}_j)}{p(\hat{V}_j)^2}\right\}\left\{\frac{1}{nh}\sum_{i\neq j} K\left(\frac{\hat{V}_i - \hat{V}_j}{h}\right)\hat{U}_i^* - \right.\right. \\ &\left.\left. F(\hat{V}_j)\frac{1}{nh}\sum_{i\neq j} K\left(\frac{\hat{V}_i - \hat{V}_j}{h}\right)\right\}\right] \\ &= \frac{1}{nh}\sigma^{*2}(\hat{V}_i)\frac{p_{nhj}(\hat{V}_j) - p(\hat{V}_j)}{p(\hat{V}_j)^2} K\left(\frac{\hat{V}_i - \hat{V}_j}{h}\right) \\ &= O_p\left\{\left(\frac{\log n}{nh}\right)^{1/2}\right\} O\left(\frac{1}{nh}\right) = o_p(n^{-1}), \end{aligned}$$

conditional on event E_{n1}^{ϵ} and given (28). The same is valid when h is replaced by s which tends faster to zero. Therefore

$$E^*[\hat{U}_i^* J_{nj}(\hat{V}_j)] = o_p(n^{-1}). \tag{31}$$

Consequently,

$$V^*(R_{n21}) = h\sum_{i=1}^{n}\{w(V_i) - w(\hat{V}_i)\}^2 \sigma^{*2}(\hat{V}_i)\{G_{ni}(\hat{V}_i) - J_{ni}(\hat{V}_i)\}^2 + o_p(1).$$

From assumption A4 and after a Taylor expansion of $v(X_i, \hat{\theta}_n)$, given that $||\theta_n - \hat{\theta}_n|| = O(n^{-1/2})$

$$V^*(R_{n21}) = \frac{h}{n} M_w^* \sum_{i=1}^{n} \sigma^{*2}(\hat{V}_i)\{G_{ni}(\hat{V}_i) - J_{ni}(\hat{V}_i)\}^2 + o_p(1).$$

Horowitz and Härdle (1994) show that $E\{(G_{ni} - J_{ni})^2\} = o(1)$. By Markov's inequality

$$(G_{ni} - J_{ni})^2 = o_p(1). \tag{32}$$

Therefore $V^*(R_{n21}) = o_p(1)$. By Chebyshev inequality $R_{n21} = o_p(1)$ and by (29) R_{n2} is $o_p(1)$. Next it will be proved that $R_{n3} = o_p(1)$.

Note that $E^*(R_{n3}) = 0$ and

$$V^*(R_{n3}) = h\sum_{i=1}^{n} E^*[w(V_i)^2 \hat{U}_i^{*2}\{F(V_i) - F(\hat{V}_i)\}^2]$$

By a taylor expansion argument, given that $||\theta_n - \hat{\theta}_n|| = O(n^{-1/2})$ and using assumptions A2, A4, A5 and A9

$$V^*(R_{n3}) = O(h) = o(1)$$

Concerning now the study of R_{n4} note that by assumptions A2, A4 and A5 and a Taylor expansion

$$|R_{n4}| = M\sqrt{h}||\theta_n - \hat{\theta}_n|| \sum_{i=1}^{n} w(V_i)\{\hat{F}_{ni}(V_i) - \hat{F}_{ni}(\hat{V}_i)\}.$$

Given that $||\theta_n - \hat{\theta}_n|| = O(n^{-1/2})$, conditional on E_{4n}^{ϵ}

$$|F_{ni}(V_i) - \hat{F}_{ni}(\hat{V}_i)| = O_p(n^{-1/2})$$

and consequently $R_{n4} = O_p(\sqrt{h}) = o_p(1)$. On the analysis of R_{n5} remind that $|F(V_i) - F(\hat{V}_i)| = O(n^{-1/2})$. Consequently conditional on event E_{3n}^{ϵ}

$$R_{n5} = R_{n51} + o_p(1) \tag{33}$$

with

$$R_{n51} = -\sqrt{h}\sum_{i=1}^{n} w(V_i)\{F_{ni}(V_i) - \hat{F}_{ni}(\hat{V}_i)\}G_{ni}(\hat{V}_i)$$

By a taylor series expansion

$$R_{n51} \quad = \quad -\sqrt{h}\sum_{i=1}^{n} w(V_i)F'\{v(X_i,\theta_n^0)\}v'(X_i,\theta_n^0)G_{ni}(\hat{V}_i)$$

where F' and v' are the derivatives of respectively F and v and θ_n^0 lies between θ_n and $\hat{\theta}_n$. According to Horowitz and Härdle (1994) $E\{G_{ni}(\hat{V}_i)^2\} = o(n^{-1}h^{-1})$. By Markov inequality $G_{ni}(\hat{V}_i) = o_p(n^{-1/2}h^{-1/2})$. Therefore $R_{n51} = o_p(1)$ and by (33) R_{n5} is $o_p(1)$. Finally only R_{n6} remains to be studied. By a Taylor expansion argument and assumptions A2 and A5 $\{F(V_i) - F(\hat{V}_i)\}^2$ is $O_p(n^{-1})$. Since $w(V_i)$ is bounded easily follows that R_{n6} converges in probability to zero and Lemma 1 is true. Q.E.D. The next lemma will allow to replace $\tilde{T}_n^*$ by a more manageable approximation in order to derive the limiting distribution.

Lemma 2

$$T_n^* = \sqrt{h}\sum_{i=1}^n w(\hat V_i)\hat U_i^* G_{ni}(\hat V_i) + o_p(1)$$

Conditional on event E_{3n}^{ϵ}

$$T_n^* = \sqrt{h}\sum_{i=1}^n w(\hat V_i)\hat U_i^* G_{ni}(\hat V_i) - T_{n1}^* + o_p(1)$$

with

$$T_{n1}^* = \sqrt{h}\sum_{i=1}^n w(\hat V_i)\hat U_i^* J_{ni}(\hat V_i)$$

Therefore it suffices to prove that $T_{n1}^* = o_p(1)$. $E^*(T_{n1}^*) = 0$ because J_{ni} does not depend on $\hat U_i^*$. Also

$$E^*(T_{n1}^{*2}) = h\sum_{i=1}^n w(\hat V_i)^2\sigma^{*2}(\hat V_i) J_{ni}(\hat V_i)^2 + o_p(1)$$

because by (31) $E^*\{\hat U_i^* J_{nj}(\hat V_j)\} = o_p(n^{-1})$. According to Horowitz and Härdle (1994) $E_X(J_{ni}(v)^2) = O(\{\log n/nh\}^2)$ uniformly over v. By Markov's inequality

$$J_{ni}(\hat V_i)^2 = O_p\left(\frac{\log^2 n}{n^2h^2}\right).$$

Consequently, given assumptions A5 and A9 $E^*(T_{n1}^*) = o_p(1)$. By Chebyshev's inequality T_{n1}^* is $o_p(1)$

<u>Proof of theorem 1</u>

Define $Z_i^* = (\hat U_i^*, \hat V_i)$ and

$$K_{hs}(v) = \left[K(v/h) - \left(\frac{h}{s}\right)^r K(v/s)\right] \Big/ \left[1 - \left(\frac{h}{s}\right)^r\right],$$

and

$$A_n(Z_i^*, Z_j^*) \;=\; \frac{1}{n\sqrt{h}}\{p_{\hat\theta_n}(\hat V_i)\}^{-1} w(\hat V_i)\hat U_i^*\{\hat U_j^* + F(\hat V_j) - F(\hat V_i)\}K_{hs}(\hat V_j - \hat V_i)$$

$$
\begin{aligned}
\mu(Z_i^*) &= E^*\{A_n(Z_i^*, Z_j^*) + A_n(Z_j^*, Z_i^*)|Z_i^*\} \\
H_n(Z_i^*, Z_j^*) &= A_n(Z_i^*, Z_j^*) + A_n(Z_j^*, Z_i^*) - \mu(Z_i^*) \\
\Psi_n^* &= \sum_i^n \sum_{j<i}^n H_n(Z_i^*, Z_j^*)
\end{aligned}
$$

It follows from lemma 2

$$
T_n^* = \Psi_n^* + \sum_i^n \sum_{j<i}^n \mu_n(Z_i^*) + o_p(1)
$$

On the other side,

$$
\mu(Z_i^*) = \frac{1}{nh}\{p_{\hat{\theta}_n}(\hat{V}_i)\}^{-1} w(\hat{V}_i)\hat{U}_i^*\{F(\hat{V}_j) - F(\hat{V}_i)\}K_{hs}(\hat{V}_j - \hat{V}_i)
$$

Since $E^*\{\mu(Z_i^*)\} = 0$ and

$$
\begin{aligned}
E^*\{\mu(Z_i^*)^2\} &= \frac{1}{n^2h}\{p_{\hat{\theta}_n}(\hat{V}_i)\}^{-2} w(\hat{V}_i)^2\hat{U}_i^{*2}\{F(\hat{V}_j) - F(\hat{V}_i)\}^2 K_{hs}(\hat{V}_j - \hat{V}_i)^2 \\
&= \frac{1}{n^2h}\{p_{\hat{\theta}_n}(\hat{V}_i)\}^{-2} w(\hat{V}_i)^2\sigma_i^{*2}(\hat{V}_i)\{F(\hat{V}_j) - F(\hat{V}_i)\}^2 K_{hs}(\hat{V}_j - \hat{V}_i)^2
\end{aligned}
$$

Given (28) $K_{hs}(\hat{V}_j - \hat{V}_i)^2 = O_p(h^2)$. Therefore $E^*\{\mu(Z_i^*)^2\} = O_p(hn^{-2}) = o_p(1)$. It follows from chebyshev's inequality that $\mu(Z_i^*) = o_p(1)$. Consequently $T_n^* = \Psi_n^* + o_p(1)$. It suffices to deduce the limiting distribution of Ψ_n^* to prove the theorem. Note that Ψ_n^* is a degenerate U-statistic according to Hall (1984). Following the same reference $\Psi_n^* \xrightarrow{\mathcal{L}} N(0, \sigma_T^{*2})$ if

$$
\frac{E^*\left[E^*\{H_n(Z_l^*, Z_i^*)H_n(Z_l^*, Z_j^*)|Z_i^*, Z_j^*\}^2\right]}{E^*\{H_n(Z_i^*, Z_j^*)^2\}^2} \longrightarrow 0 \tag{34}
$$

$$
\frac{E^*\left[E^*\{H_n(Z_i^*, Z_j^*)^4\right]}{E^*\{H_n(Z_i^*, Z_j^*)^2\}^2} \longrightarrow 0 \tag{35}
$$

$$
\frac{1}{2}n^2 E^*\{H_n(Z_i^*, Z_j^*)^2\} \longrightarrow \sigma_T^{*2} \tag{36}
$$

as $n \longrightarrow \infty$.

The study of the numerator of condition (34) will be done below. Observe that

$$
\begin{aligned}
H_n(Z_i^*, Z_j^*) &= \frac{1}{n\sqrt{h}}\{p_{\hat{\theta}_n}(\hat{V}_i)\}^{-1} w(\hat{V}_i)\hat{U}_i^*\hat{U}_j^* K_{hs}(\hat{V}_j - \hat{V}_i) + \\
&\quad \frac{1}{n\sqrt{h}}\{p_{\hat{\theta}_n}(\hat{V}_j)\}^{-1} w(\hat{V}_j)\hat{U}_j^*\{\hat{U}_i^* + F(\hat{V}_j) - F(\hat{V}_i)\} \\
&\quad K_{hs}(\hat{V}_j - \hat{V}_i)
\end{aligned} \tag{37}
$$

After some tedious algebra

$$E^*\{H_n(Z_l^*,Z_i^*)H_n(Z_l^*,Z_j^*)|Z_i^*,Z_j^*\} = \frac{1}{n^2h}K_{hs}(\hat V_i-\hat V_l)K_{hs}(\hat V_j-\hat V_l)$$
$$\hat U_i^*\hat U_j^*\left[\sigma^{*2}(\hat V_l)\frac{w(\hat V_l)}{p_{\hat\theta_n}(\hat V_l)}\left\{\frac{w(\hat V_l)}{p_{\hat\theta_n}(\hat V_l)}+\frac{w(\hat V_i)}{p_{\hat\theta_n}(\hat V_i)}+\frac{w(\hat V_j)}{p_{\hat\theta_n}(\hat V_j)}\right\}+\right.$$
$$\left.\frac{w(\hat V_i)}{p_{\hat\theta_n}(\hat V_i)}\frac{w(\hat V_j)}{p_{\hat\theta_n}(\hat V_j)}\left\{\sigma^{*2}(\hat V_l)+[\hat F(\hat V_l)-F(\hat V_i)][\hat F(\hat V_l)-F(\hat V_j)]\right\}\right]$$

and

$$E^*\left[E^*\{H_n(Z_l^*,Z_i^*)H_n(Z_l^*,Z_j^*)|Z_i^*,Z_j^*\}^2\right] = \frac{1}{n^4h^2}K_{hs}(\hat V_i-\hat V_l)^2K_{hs}(\hat V_j-\hat V_l)^2\hat U_i^{*2}\hat U_j^{*2}A_n$$
$$A_n = \left[\sigma^{*2}(\hat V_l)\frac{w(\hat V_l)}{p_{\hat\theta_n}(\hat V_l)}\left\{\frac{w(\hat V_l)}{p_{\hat\theta_n}(\hat V_l)}+\frac{w(\hat V_i)}{p_{\hat\theta_n}(\hat V_i)}+\frac{w(\hat V_j)}{p_{\hat\theta_n}(\hat V_j)}\right\}+\right.$$
$$\left.\frac{w(\hat V_i)}{p_{\hat\theta_n}(\hat V_i)}\frac{w(\hat V_j)}{p_{\hat\theta_n}(\hat V_j)}\left\{\sigma^{*2}(\hat V_l)+[\hat F(\hat V_l)-F(\hat V_i)][\hat F(\hat V_l)-F(\hat V_j)]\right\}\right]^2$$
$$E^*\left[E^*\{H_n(Z_l^*,Z_i^*)H_n(Z_l^*,Z_j^*)|Z_i^*,Z_j^*\}^2\right] = \frac{1}{n^4h^2}K_{hs}(\hat V_i-\hat V_l)^2K_{hs}(\hat V_j-\hat V_l)^2\times$$
$$\sigma^{*2}(\hat V_i)\sigma^{*2}(\hat V_j)A_n.$$

Given (28) and assumptions A3 – A6

$$E^*\left[E^*\{H_n(Z_l^*,Z_i^*)H_n(Z_l^*,Z_j^*)|Z_i^*,Z_j^*\}\right] = O_p(h^2n^{-4}) = o_p(n^{-4}) \tag{38}$$

Concerning the numerator of condition (35) note that by (37)

$$H_n(Z_i^*,Z_j^*)^4 = n^{-4}h^{-2}B_n$$

and the expectation of B_n conditional on X is bounded uniformly over n. Therefore

$$n^{-1}E^*\left\{H_n(Z_i^*,Z_j^*)^4\right\} = O_p\left(\frac{1}{n^5h^2}\right) \tag{39}$$

$E^*\left\{H_n(Z_i^*,Z_j^*)^2\right\}$ will be derived below. First, observe that

$$\begin{aligned} H_n(Z_i^*,Z_j^*)^2 \;&=\; \frac{1}{n^2h}\left\{\frac{w(\hat V_i)^2}{p_{\hat\theta_n}(\hat V_i)^2}\hat U_i^{*2}\hat U_j^{*2}K_{hs}(\hat V_i-\hat V_j)^2+\right.\\ &\frac{w(\hat V_i)^2}{p_{\hat\theta_n}(\hat V_i)^2}\hat U_j^{*2}[\hat U_i^{*2}+\{\hat F(\hat V_i)-F(\hat V_j)\}^2]K_{hs}(\hat V_j-\hat V_i)^2+\\ &\left.2\times\frac{w(\hat V_i)}{p_{\hat\theta_n}(\hat V_i)}\frac{w(\hat V_j)}{p_{\hat\theta_n}(\hat V_j)}\hat U_i^{*2}\hat U_j^{*2}[\hat U_i^{*2}+\hat F(\hat V_i)-F(\hat V_j)]K_{hs}(\hat V_j-\hat V_i)^2\right\}\end{aligned}$$

and

$$
\begin{aligned}
E^*\left\{H_n(Z_i^*,Z_j^*)^2\right\} &= \frac{1}{n^2h}\left\{\frac{w(\hat V_i)^2}{p_{\hat\theta_n}(\hat V_i)^2}\sigma^{*2}(\hat V_i)\sigma^{*2}(\hat V_j)K_{hs}(\hat V_i-\hat V_j)^2+\right.\\
&\quad \frac{w(\hat V_i)^2}{p_{\hat\theta_n}(\hat V_i)^2}\sigma^{*2}(\hat V_i)\sigma^{*2}(\hat V_j)K_{hs}(\hat V_j-\hat V_i)^2+\frac{w(\hat V_i)^2}{p_{\hat\theta_n}(\hat V_i)^2}\\
&\quad \sigma^{*2}(\hat V_j)\{\hat F(\hat V_i)-F(\hat V_j)\}^2]K_{hs}(\hat V_j-\hat V_i)^2+2\times\\
&\quad \left.\frac{w(\hat V_i)}{p_{\hat\theta_n}(\hat V_i)}\frac{w(\hat V_j)}{p_{\hat\theta_n}(\hat V_j)}\sigma^{*2}(\hat V_i)\sigma*2(\hat V_j)K_{hs}(\hat V_j-\hat V_i)^2\right\} \qquad (40)
\end{aligned}
$$

By the arguments of Horowitz and Härdle (1994) the expectation of (40) is

$$
E_v\left[E^*\left\{H_n(Z_i^*,Z_j^*)^2\right\}\right] = \frac{1}{n^2}4C_K\int_{-\infty}^{\infty} w(v)^2\{\sigma^{*2}(v)\}^2dv\ +o(1) \qquad (41)
$$

with $v=v(X,\hat\theta_n)$. By Markov inequality

$$
E^*\left\{H_n(Z_i^*,Z_j^*)^2\right\}=O_p(n^{-2}) \qquad (42)
$$

and given (38) condition (34) is verified. Combining results (42) and (39) condition (35) is also verified.

According to condition (36), result (41) and given that $E^*\left\{H_n(Z_i^*,Z_j^*)^2\right\}$ converges in probability to its expectation, then

$$
\sigma_T^{*2}\xrightarrow{p}2C_K\int_{-\infty}^{\infty}w(v)^2\{\sigma^{*2}(v)\}^2dv \qquad (43)
$$

with $v=v(X,\hat\theta_n)$. Thus $T_n^*\xrightarrow{\mathcal{L}}N(0,\sigma_T^{*2})$. On the other side, because $||\hat\theta_n-\theta||=O(1/\sqrt{n})$ easily follows that $v(X,\hat\theta_n)\xrightarrow{p}v(X,\theta)$. Considering $\sigma^{*2}(v)=\sigma^2\{v(x,\hat\theta_n)\}$, giving assumption A7, the convergence in probability of $v(X,\hat\theta_n)$ to $v(X,\theta)$, and applying Lebesgues' dominated convergence theorem $\sigma_T^{*2}\xrightarrow{p}\sigma_T^2$. Therefore T_n^* has the same asymptotic distribution as T_n. Q.E.D.