The University of Manchester

MANCHESTER 1824

Economics Discussion Paper

EDP-0504

The Asymptotic Equivalence of Kernel-based Nonparametric Conditional Moment Test Statistics

by

Eduardo Fe-Rodriguez and Chris D. Orme

January 2005

e.ferodriguez@manchester.ac.uk
School of Social Sciences,
The University of Manchester
Oxford Road
Manchester M13 9PL
United Kingdom

The Asymptotic Equivalence of Kernel-based Nonparametric Conditional Moment Test Statistics

Eduardo Fe-Rodriguez*and Chris D. Orme
Economic Studies,
School of Social Sciences,
University of Manchester,
Manchester M13 9PL.
U.K.

January 28, 2005

Abstract

In this paper, it is shown that a consistent misspecification test statistic based on a local linear Kernel regression estimator is asymptotically equivalent to one based on the Nadaraya-Watson estimator. A variery of new, and consistent, variance estimators are also given.

1 Introduction

Testing for misspecification of functional form has been one of the most important topics for research in econometrics. Influential contributions concerning the testing of parametric models include Hausman (1978), White (1982) and Ruud (1984) and the unifying conditional moment procedures of Newey (1985) and Tauchen (1985). Subsequently, much work has focussed on improving the finite sample performance of such tests procedures, under the null. This includes examining asymptotically equivalent versions of test statistics, higher order asymptotic analysis and the development bootstrap procedures (see, for example, Orme (1991), Chesher and Spady (1991) and Horowitz (1994), respectively). For a survey of these issues, see Godfrey and Orme (2001) and references therein.

However, although methods are now available to improve the finite sample performance of parametric test procedures under the null (in the sense that one is able to obtain very good agreement between nominal and empirical significance levels, in general), these tests will not be consistent against

*Corresponding author. This paper is part of Eduardo Fe's Ph.D. dissertation. A previous version was presented in the conference "Semiparametrics in Rio", Getulio Vargas Foundation, Rio de Janeiro (2004). Email address: E.Ferodriguez@manchester.ac.uk

all deviations from the null model. One way of overcoming this is to employ nonparametric, kernel-based, methods to construct consistent misspecification tests.

During the late 1980's and the 1990's a large number of articles have been published regarding this topic. Following the work of Robinson (1988) on the estimation of semi-nonparametric models, one can find nonparametric misspecification tests in Hardle and Mammen (1993), Fan and Li (1996, 2000, 2002), Gozalo (1993), Zheng (1996), Li and Wang (1998), Lavergne and Vuong (2000), and Ellison and Ellison (2000). Virtually all the published papers consider only the framework of regression analysis. However, the methodology can, of course, be extended to test arbitrary moment conditions (see, for example, Zheng (1998) who proposes a test for conditional symmetry and Hsiao and Li (2001) who propose a consistent test for heteroskedasticity).

The paper by Ellison and Ellison (2000) proposed a nonparametric misspecification test statistic based on the Nadaraya-Watson (NW) regression estimator. This statistic is an "unweighted" version of Zheng's (1996) consistent test statistic, will was shown to be asymptotically distributed as a standard normal random variable. The NW estimator is a natural choice, since it provides a consistent estimator of a regression function and is easy to implement using a modern computer. However, work by Fan (1992,1993) has shown the NW estimator is not an optimal estimator among the class of linear smoothers. The so called Local Linear Regression (LLR) estimator (Stone, 1977) has asymptotic minimax efficiency, a "more stable" bias, the same variance as that of the NW estimator, and it does not suffer the Boundary Bias problem (see, for instance, Wand and Jones (1995)). That is, the LLR estimator has constant convergence rate through the domain of the regression function (unlike the NW estimator). Thus one might naturally consider the use of the LLR estimator in place of the NW estimator, in the construction of consistent misspecification tests. Moreover, the approach of Ellison and Ellison (2000) suggests (but does not establish) that the LLR version of the test statistic will be asymptotically distributed as a standard normal random variable, under the null. In fact this paper shows that a stronger results obtains.

This paper considers the use of the NW and LLR estimators to construct test statistics which provide consistent procedures to assess the validity of some arbitrary moment condition. It is shown that (i) the LLR version will, in fact, be asymptotically equivalent to the NW version of the test statistic; and (ii) there are a variety of different consistent variance estimators which might be employed when implementing these test procedures, which may lead to the classic conflict amongst test criteria.

The plan of this paper is as follows: Section 2 summarises the model, test statistics under consideration and derives the relevant limiting distributions. Section 3 reports the findings of a small Monte Carlo study. Section 4 concludes.

2 Model and Test Statistics

2.1 The parametric model and hypothesis under test

It is assumed that we have observations $w_i' = \{(y_i, x_i')\}_{i=1}^N$, on $W' = (Y, X')$, where X is a $(d \times 1)$ vector and Y scalar random variable. Within some specified parametric framework (e.g., maximum likelihood, non-linear least squares, generalised method of moments), these data are used to model an unknown $(k \times 1)$ parameter, θ, with $\theta \in \Theta$ and Θ being a compact convex subset of R^m. In order to test the adequacy of the parametric model, a moment condition of the form $E(\varepsilon|x) = 0$ is to be tested in which $\varepsilon = u(\theta_0; W)$ and $u(\theta; W)$ is some (scalar) random function of θ. Formally, the null and alternative hypothesis are:

$$\begin{aligned} H_0 &: \Pr\left[E\left(u\left(\theta_0; W_i\right)|X_i\right) = 0\right] = 1, \quad a.s. \quad \text{for some } \theta_0 \in \Theta \\ H_A &: \Pr\left[E\left(u\left(\theta; W_i\right)|X_i\right) = 0\right] < 1, \quad a.s. \quad \text{for all } \theta \in \Theta. \end{aligned}$$

The following basic assumptions are made, in which $|.|$ is the absolute value of a scalar, $||a|| = \sqrt{a'a}$ is the Euclidean Norm of a vector and $||A|| = \sqrt{tr\left(A'A\right)} = \sqrt{tr(AA')}$, is the Euclidean Norm of a matrix.

Assumption 1 (i) The w_i' are independently and identically distributed as $W' = (Y, X')$, with joint density $g(w)$; (ii) X is a continuous random variable and has convex compact support, $\mathcal{S} \subset R^d$, with density $f(x)$; (iii) $f(x)$ is bounded above, $\inf_x f(x) \geq \delta > 0$ and is uniformly continuous.

Assumption 2 The $u = u(\theta; W)$ have first and second order derivatives denoted $\nabla_\theta u = \dfrac{\partial u(\theta; W)}{\partial \theta}$, $(m \times 1)$ and $\nabla_{\theta\theta} u = \dfrac{\partial^2 u(\theta; W)}{\partial \theta \partial \theta'}$, $(m \times m)$, respectively, and there exists a measurable function $b(W) > 0$ satisfying $\int b(w)dw < \infty$ which dominates $|u|$, $||\nabla_\theta u||$ and $||\nabla_{\theta\theta} u||$, for all $\theta \in \Theta$. Furthermore, writing $\varepsilon = u(\theta_0; W)$, for some $\theta_0 \in \Theta$, ε has bounded fourth moment with $E\left(\varepsilon^2|X = x\right) = \sigma^2(x)$ and $E\left(\varepsilon^4|X = x\right) = \mu_4(x)$.

Assumption 3 A parametric estimator, $\hat{\theta}$, is available such that under H_0, $\hat{\theta} - \theta_0 = O_p(N^{-1/2})$

Assumption 1 requires X to be continuously distributed, although Y could be discrete. The analysis in this paper could be extended to include the case of discrete (or mixed X), along the lines of the recent paper by Racine and Li (2004). The conditions on $f(x)$ avoid problems associated with small values of the density when using Kernel-based methods (see for example, Ellison and Ellison (2000) and Fan and Li (2002)) and, thereby, simplifies the analysis. Without it, trimming out small values of the density could be employed (as, for example, in Lavergne and Vuong, 1996) but at the cost of complicating the proofs. Assumptions 2 and 3 justify the various asymptotic expansions that will be used to derive limiting distributions, and are fairly standard. The next section introduces the Kernel-based test statistics under consideration.

2.2 The Test Statistics

First, define $\zeta_{ij} = \left(\frac{X_i - X_j}{h}\right)$, with X_i having typical element X_{il}, $l = 1, ..., d$, and the product Kernel as

$$\begin{aligned} K_{ij} &= \frac{1}{h^d} \prod_{l=1}^{d} \left\{ k\left(\frac{X_{il} - X_{jl}}{h}\right) \right\} \\ &\equiv \frac{1}{h^d} K\left(\zeta_{ij}\right), \end{aligned}$$

where $K(\zeta)$ is a product kernel, and $k(s)$ is a univariate kernel, satisfying the following:

Assumption 4 Let $k(s) \geq 0$ be a bounded real-valued, symmetric, density function, $k : R \to R$, such that:
(i) $\int_{-\infty}^{\infty} k(s) ds = 1$
(ii) $\sup_s k(s) < \infty$
(iii) $|s| k(s) \to 0$ as $|s| \to \infty$.
(iv) $\int k^2(s) ds < \infty$.

The parameter h is a smoothing parameter which approaches zero as $N \to \infty$, as detailed later.

Now, let $T = \{T_{ij}\}$ be some $(N \times N)$ weighting matrix with diagonal elements equal to zero and $e = \{e_i\}$ be an $(N \times 1)$ vector. Then define a class of test statistics as

$$V_N^T(e) = \frac{N h^{d/2}}{N(N-1)} e' T e = \frac{N h^{d/2}}{N(N-1)} \sum_{i=1}^{N} \sum_{j \neq i} e_i T_{ij} e_j. \tag{1}$$

Examples include the following:

1. Zheng's (1996) test statistic is

$$\begin{aligned} V_N^K(\hat{u}) &= \frac{N h^{d/2}}{N(N-1)} \sum_i \sum_{j \neq i} \hat{u}_i K_{ij} \hat{u}_j \\ &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \hat{u}_i \hat{r}_N(X_i) \end{aligned}$$

where $\hat{u}_i = u\left(\hat{\theta}; W_i\right)$, $\hat{r}_N(X_i) = \sqrt{Nh^d}\hat{\alpha}_N(X_i)f_N(X_i)$, $f_N(X_i) = \frac{1}{N-1}\sum_{j\neq i} K_{ij}$ is the 'leave-one-out' Nadaraya-Watson (NW) density estimator and $\hat{\alpha}_N(X_i)$ is the corresponding 'leave-one-out' NW regression estimator of $E\left[\varepsilon|X_i\right]$, but with $\hat{\theta}$ replacing θ_0.

2. Ellison and Ellison (2000) proposed

$$\begin{aligned} V_N^P(\hat{u}) &= \frac{Nh^{d/2}}{N(N-1)}\sum_i\sum_{j\neq i}\hat{u}_i P_{ij}\hat{u}_j, \qquad P_{ij} = \frac{K_{ij}}{f_N(X_i)} \\ &= \frac{1}{\sqrt{N}}\sum_{i=1}^N \hat{u}_i r_N(X_i), \quad \text{say} \end{aligned}$$

where, here, $r_N(X_i) = \sqrt{Nh^d}\hat{\alpha}_N(X_i)$.

3. Replacing $\hat{\alpha}_N(X_i)$ with $\tilde{\alpha}_N(X_i)$, the 'leave-one-out' Local Linear (Kernel) Regression (LLR) estimator yielding the statistic

$$\begin{aligned} V_N^R(\hat{u}) &= \frac{1}{\sqrt{N}}\sum_{i=1}^N \hat{u}_i r_N(X_i) \\ &= \frac{Nh^{d/2}}{N(N-1)}\sum_i\sum_{j\neq i}\hat{u}_i R_{ij}\hat{u}_j \\ &= \frac{Nh^{d/2}}{N(N-1)}\hat{u}' R\hat{u} \end{aligned}$$

where, here, $r_N(X_i) = \sqrt{Nh^d}\tilde{\alpha}_N(X_i)$, and, therefore,

$$\begin{aligned} R_{ij} &= \frac{K_{ij} - \left(\frac{1}{N-1}\sum_{s\neq i} K_{is}\zeta_{is}'\right)\left(\frac{1}{N-1}\sum_{s\neq i} K_{is}\zeta_{is}\zeta_{is}'\right)^{-1} K_{ij}\zeta_{ij}}{\Delta_i} \\ &= \frac{K_{ij} - b_N(X_i)'\left(M_N(X_i)\right)^{-1} K_{ij}\zeta_{ij}}{\Delta_i}; \\ \Delta_i &\equiv \Delta_N(X_i) = \frac{1}{N-1}\sum_{s\neq i}\left\{K_{is} - b_N(X_i)'\left(M_N(X_i)\right)^{-1} K_{is}\zeta_{is}\right\} \\ &= \left\{f_N(X_i)\right\} - b_N(X_i)'\left(M_N(X_i)\right)^{-1} b_N(X_i) > 0, \end{aligned}$$

where the definitions of $f_N(X_i) \equiv f_{Ni}$, $b_N(X_i) \equiv b_{Ni}$ and $M_N(X_i) \equiv M_{Ni}$ are implicit, so that $\frac{1}{N-1}\sum_{j\neq i} R_{ij} = 1$.

This last statistic (to the best of our knowledge) has not previously been discussed or analysed. In this paper, we provide its limit distribution under the null by showing that it is asymptotically equivalent to $V_N^P(\hat{u})$, which is the Ellison and Ellison (2000) statistic. The intuition for this asymptotic equivalence runs as follows. The difference between $V_N^R(\hat{u})$ and $V_N^P(\hat{u})$ is

$$\begin{aligned} V_N^R(\hat{u}) - V_N^P(\hat{u}) &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \hat{u}_i \sqrt{Nh^d}\left(\tilde{\alpha}_N(X_i) - \hat{\alpha}_N(X_i)\right) \\ &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \hat{u}_i a_N(X_i) \end{aligned}$$

where $a_N(X_i) = \sqrt{Nh^d}\left(\tilde{\alpha}_N(X_i) - \hat{\alpha}_N(X_i)\right)$. However, under the null that $E[\varepsilon|x] = 0$, it can be shown that $a_N^0(x) = \sqrt{Nh^d}\left(\tilde{\alpha}_N^0(x) - \hat{\alpha}_N^0(x)\right)$ is $o_p(1)$, where $\hat{\alpha}_N^0(x)$ and $\tilde{\alpha}_N^0(x)$ are, respectively, the NW and LLR estimators of $E[\varepsilon|X = x]$. Consistency of $\hat{\theta}$ under the null suggests therefore that $a_N(x)$ will also be $o_p(1)$ which, in turn, indicates the possible degeneracy of $V_N^R(\hat{u}) - V_N^P(\hat{u})$. This result is formally established in this paper.

Note that, it is expected that $V_N^R(\hat{u})$, $V_N^P(\hat{u})$ and $V_N^K(\hat{u})$ will provide consistent test statistics, for a one-sided testing procedure. For example, Zheng (1996) shows that $V_N^K(\hat{u})/Nh^d$ converges (in probability) to a finite positive limit under H_A (when $u(\theta; W)$ is a regression model error).

2.3 The Limit Distributions

Following (1), define

$$V_N^Q(e) = \frac{Nh^{d/2}}{N(N-1)} \sum_i \sum_{j \neq i} e_i Q_{ij} e_j. \qquad Q_{ij} = \frac{K_{ij}}{f(X_i)}.$$

The starting point of the analysis is an application of Hall's (1984) central limit theorem for degenerate U-statistics, for which the conditions are relatively easy to verify.

Lemma 1 *As $N \to \infty$, $h \to 0$ such that $Nh^d \to \infty$,*
(i) $V_N^Q(\varepsilon) \xrightarrow{d} N(0, \Sigma)$, where $\Sigma = 2\int K^2(\zeta)\, d\zeta' \int \left\{\sigma^2(x)\right\}^2 dx'$.
(ii) Furthermore,

$$\begin{aligned} \Sigma_{1N}^Q(\varepsilon) &= \frac{2h^d}{N(N-1)} \sum_i \sum_{j \neq i} \varepsilon_i^2 Q_{ij}^2 \varepsilon_j^2 \\ \Sigma_{2N}^Q(\varepsilon) &= \frac{h^d}{N(N-1)} \sum_i \sum_{j \neq i} \varepsilon_i^2 \left\{Q_{ij}^2 + Q_{ij} Q_{ji}\right\} \varepsilon_j^2 \end{aligned}$$

are consistent for Σ.

With this result, the strategy for establishing the limit distribution of $V_N^R(\hat{u})$ is to show the following:

$$V_N^R(\hat{u}) - V_N^Q(\varepsilon) = o_p(1)$$

by noting that

$$\begin{aligned} V_N^R(\hat{u}) - V_N^Q(\varepsilon) &= V_N^R(\hat{u}) - V_N^P(\hat{u}) \\ &\quad + V_N^P(\hat{u}) - V_N^Q(\hat{u}) \\ &\quad + V_N^Q(\hat{u}) - V_N^Q(\varepsilon) \end{aligned}$$

where each of the three terms on the right hand side are $o_p(1)$. In particular, by showing that $V_N^R(\hat{u}) - V_N^P(\hat{u}) = o_p(1)$ it is established that $V_N^R(\hat{u})$ will be asymptotically equivalent to the Ellison and Ellison (2000) test statistic.

Consider first, $V_N^Q(\hat{u}) - V_N^Q(\varepsilon)$. Following very closely the arguments laid out in Zheng (1996), and the assumptions stated to date, we have the following Lemma.

Lemma 2 *Under Assumptions 1-4 with $N \to \infty$, $h \to 0$ such that $Nh^d \to \infty$,*

$$V_N^Q(\hat{u}) - V_N^Q(\varepsilon) = o_p(1)$$

so that

$$V_N^Q(\hat{u}) \xrightarrow{d} N(0, \Sigma).$$

Proof. *See Appendix* ■

To establish that $V_N^P(\hat{u}) - V_N^Q(\hat{u})$, the following rather general, result is established.

Theorem 3 *Let $V_N^K(\hat{u}) = \dfrac{Nh^{d/2}}{N(N-1)} \sum_i \sum_{j \neq i} \hat{u}_i K_{ij} \hat{u}_j$ be defined as above and consider the statistic $V_N^D(\hat{u}) = \dfrac{Nh^{d/2}}{N(N-1)} \sum_i \sum_{j \neq i} \hat{u}_i a_N(X_i) K_{ij} \hat{u}_j$ where $a_N(x)$ is a scalar random function of $X_1, ..., X_N$ such that $\sup_{x \in \mathcal{S}} |a_N(x)| = o(1)$, a.s.. Then $V_N^D(\hat{u}) = o_p(1)$.*

Proof. *See Appendix.* ■

The above Theorem allows us to show both $V_N^P(\hat{u}) - V_N^Q(\hat{u}) = o_p(1)$ and $V_N^R(\hat{u}) - V_N^P(\hat{u}) = o_p(1)$. We make the following additional assumptions which impose slightly stronger conditions on the behaviour of h (Assumption 5) and the kernel function, as follows:[1].

[1] It is possible to weaken these conditions, to obtain the same result, but at the expense of complicating the proof.

Assumption 5 As $N \to \infty$, $h \to 0$ there exists a $\delta \in (0,1)$ such that $N^{\delta} h^{2d} \to \gamma$, where γ is a positive constant.

Assumption 6 Assume that
(i) $k(s) \geq 0$ also satisfies

$$\begin{aligned} \int_{-\infty}^{\infty} |sk(s)|\, ds &< \infty & (2) \\ \int_{-\infty}^{\infty} \left|s^2 k(s)\right| ds &= \mu_2^k < \infty. & (3) \end{aligned}$$

(ii) $k(s)$ has characteristic function $\eta(t)$, with derivatives $\eta'(t)$ and $\eta''(t)$, all of which are absolutely integrable.

Note that Assumption 6(i), ensures that $\eta'(t)$ and $\eta''(t)$ exist. Furthermore, Assumption 6(ii) implies that $k(s) = (2\pi)^{-1} \int \eta(s) e^{-its} dt$, $isk(s) = (2\pi)^{-1} \int \eta'(s) e^{-its} dt$, where $i^2 = -1$, and $s^2 k(s) = -(2\pi)^{-1} \int \eta''(s) e^{-its} dt$. Then, since $K(\zeta)$ has characteristic function $\phi(\xi) = \prod_{l=1}^{d} \eta(\xi_l)$, where $\xi' = (\xi_1, ..., \xi_d)$, it follows that

$$\begin{aligned} K(\zeta) &= \prod_{l=1}^{d} k(\zeta_l) = (2\pi)^{-d} \int \phi(\xi) e^{-i\xi'\zeta} d\xi', \\ i\zeta_l K(\zeta) &= i\zeta_l k(\zeta_l) \prod_{m \neq l}^{d} k(\zeta_m) = (2\pi)^{-d} \int \nabla_l \phi(\xi) e^{-i\xi'\zeta} d\xi', \\ \zeta_l^2 K(\zeta) &= \zeta_l^2 k(\zeta_l) \prod_{r \neq l}^{d} k(\zeta_r) = -(2\pi)^{-d} \int \nabla_{ll}^2 \phi(\xi) e^{-i\xi'\zeta} d\xi' \\ \zeta_l \zeta_m K(\zeta) &= \zeta_l \zeta_m k(\zeta_l) k(\zeta_m) \prod_{\substack{r \neq l \\ r \neq m}}^{d} k(\zeta_r) = -(2\pi)^{-d} \int \nabla_{lm}^2 \phi(\xi) e^{-i\xi'\zeta} d\xi', \quad l \neq m, \end{aligned}$$

where $\nabla_l \phi(\xi) = \dfrac{\partial \phi(\xi)}{\partial \xi_l}$ and $\nabla_{lm}^2 \phi(\xi) = \dfrac{\partial^2 \phi(\xi)}{\partial \xi_l \partial \xi_m}$, and these results will be useful in what follows.

Assumption 7 The functions,

$$z_0(c) = \int |\phi(c\xi) - \phi(\xi)|\, d\xi'$$

$$z_1(c) = \int |\nabla_l \phi(c\xi) - \nabla_l \phi(\xi)|\, d\xi'$$

and

$$z_2(c) = \int \left|\nabla^2_{lm}\phi\left(c\xi\right) - \nabla^2_{lm}\phi\left(\xi\right)\right| d\xi'$$

are Locally Lipschitz of order 1 at $c = 1$; i.e., there exists a $\varepsilon > 0$ and $0 < B < \infty$ such that

$$|z_r(c)| < B\,|c-1| \quad \text{for all } c \in (1-\varepsilon, 1+\varepsilon) \ \text{ and } r = 0, 1, 2.$$

Together, Assumptions 5-7 permit the application of Van Ryzin's (1969) method in order to prove the following result.

Lemma 4 *Under Assumptions 1, 5-7*

$$\sup_{x\in\mathcal{S}} \left|f_N\left(x\right) - f(x)\right| \longrightarrow 0, \qquad a.s. \tag{4}$$

$$\sup_{x\in\mathcal{S}} \left|\left|b_N\left(x\right)\right|\right| \longrightarrow 0, \qquad a.s. \tag{5}$$

$$\sup_{x\in\mathcal{S}} \left|\left|M_N\left(x\right) - \mu_2^K f(x) I_d\right|\right| \longrightarrow 0, \qquad a.s. \tag{6}$$

Furthermore

$$\sup_{x\in\mathcal{S}} \left|f_N(x)\right|^{-1} = O(1) \quad a.s. \tag{7}$$

$$\sup_{x\in\mathcal{S}} \left|\left|M_N(x)^{-1}\right|\right| = O(1) \quad a.s. \tag{8}$$

$$\sup_{x\in\mathcal{S}} \left|\Delta_N\left(x\right) - f(x)\right| \longrightarrow 0, \qquad a.s. \tag{9}$$

$$\sup_{x\in\mathcal{S}} \left|\Delta_N(x)\right|^{-1} = O(1) \quad a.s. \tag{10}$$

Proof. *See Appendix.* ■

Exploiting Lemma 4, the following result is established, which provides an alternative asymptotic justification for the test statistic proposed by Ellison and Ellison (2000).

Corollary 5 *Under Assumptions 1-7.*

$$V_N^P(\hat{u}) - V_N^Q(\hat{u}) = o_p(1)$$

so that

$$V_N^P(\hat{u}) \xrightarrow{d} N(0, \Sigma).$$

Proof. *Note that*

$$V_N^P(\hat{u}) - V_N^Q(\hat{u}) = V_N^D(\hat{u})$$

where the $(n \times n)$ *matrix* D *has typical element*

$$\begin{aligned} D_{ij} &= \left(\frac{f(X_i) - f_N(X_i)}{f(X_i)\, f_N(X_i)} \right) K_{ij} \\ &= a_N(X_i) K_{ij}, \quad \textit{say.} \end{aligned}$$

Note that by Lemma 4. $\sup_{x \in \mathcal{S}} |a_N(x)| = o(1)$, *a.s. The result follows from Theorem 3* ■

Finally, the following Corollary completes the demonstration.

Corollary 6 *Under Assumption 1-7,*

$$V_N^R(\hat{u}) - V_N^P(\hat{u}) = o_p(1)$$

so that

$$V_N^R(\hat{u}) \xrightarrow{d} N(0, \Sigma).$$

Proof. *We have*

$$V_N^R(\hat{u}) - V_N^P(\hat{u}) = V_N^D(\hat{u})$$

where the $(n \times n)$ *matrix* D *now has typical element*

$$\begin{aligned} D_{ij} &= \frac{(f_N(X_i) - \Delta_N(X_i))K_{ij}}{\Delta_N(X_i) f_N(X_i)} - \frac{b_N(X_i)' \left(M_N(X_i)\right)^{-1} K_{ij}\zeta_{ij}}{\Delta_N(X_i)} \\ &= \frac{b_N(X_i)' \left(M_N(X_i)\right)^{-1} b_N(X_i) K_{ij}}{\Delta_N(X_i) f_N(X_i)} - \frac{b_N(X_i)' \left(M_N(X_i)\right)^{-1} K_{ij}\zeta_{ij}}{\Delta_N(X_i)} \\ &= a_{1N}(X_i)K_{ij} + a_{2N}(X_i)' K_{ij}\zeta_{ij}, \quad \textit{say,} \end{aligned}$$

so that

$$V_N^D(\hat{u}) = \frac{Nh^{d/2}}{N(N-1)} \sum_{i=1}^{N} \sum_{j \neq i} \hat{u}_i a_{1N}(X_i) K_{ij} \hat{u}_j + \frac{Nh^{d/2}}{N(N-1)} \sum_{i=1}^{N} \sum_{j \neq i} \hat{u}_i a_{2N}(X_i)' \zeta_{ij} K_{ij} \hat{u}_j.$$

Note that by Lemma 4. both $\sup_{x\in\mathcal{S}}|a_{1N}(x)| = o(1)$*, a.s, and* $sup_{x\in\mathcal{S}}||a_{2N}(x)|| = o(1)$*, a.s. Theorem 3 implies that*

$$V_{1N}^{D}(\hat{u}) = \frac{Nh^{d/2}}{N(N-1)}\sum_{i=1}^{N}\sum_{j\neq i}\hat{u}_i a_{1N}(X_i)K_{ij}\hat{u}_j = o_p(1).$$

Similarly,

$$V_{2N}^{D}(\hat{u}) = \frac{Nh^{d/2}}{N(N-1)}\sum_{i=1}^{N}\sum_{j\neq i}\hat{u}_i a_{2N}(X_i)'\zeta_{ij}K_{ij}\hat{u}_j = o_p(1).$$

The latter is slightly more complicated but the result also follows from Theorem 3 since it can be expressed as the sum of the following d terms

$$\frac{Nh^{d/2}}{N(N-1)}\sum_{i=1}^{N}\sum_{j\neq i}\hat{u}_i a_{2Nl}(X_i)\zeta_{ijl}K_{ij}\hat{u}_j$$

where $a_{2N}(X_i) = \{a_{2Nl}(X_i)\}$ *and* $\zeta_{ij} = \left\{\zeta_{ijl}\right\}$, $l = 1, ..., d$*, and noting that* $\frac{Nh^{d/2}}{N(N-1)}\sum_{i=1}^{N}\sum_{j\neq i}\hat{u}_i\zeta_{ijl}K_{ij}\hat{u}_j \equiv 0$*, for* $l = 1, ..., d$. ■

It follows, therefore, that $V_N^R(\hat{u}) - V_N^Q(\varepsilon) = o_p(1)$, as required, so that by Lemma 1, $V_N^R(\hat{u}) \xrightarrow{d} N(0,\Sigma)$. It follows that the following test statistic

$$S_N^R(\hat{u}) = \frac{V_N^R(\hat{u})}{\sqrt{\Sigma_N(\hat{u})}}$$

will have a limit standard normal distribution under the null, where $\Sigma_N(\hat{u}) > 0$ is any consistent estimator for Σ. On the other hand, it can be shown (similar to Zheng, 1996) that $S_N^R(\hat{u})/Nh^d$ converges in probability to a positive constant under H_A. Thus, an asymptotically valid test procedure is to reject H_0 for large values of $S_N^R(\hat{u})$.

A consistent estimator for Σ is given by the following Lemma.

Lemma 7 *Consistent estimators of* Σ *are given by any of the following*

$$\begin{aligned}
\Sigma_{1N}^{P}(\hat{u}) &= \frac{2h^d}{N(N-1)}\sum_i\sum_{j\neq i}\hat{u}_i^2 P_{ij}^2\hat{u}_j^2 \\
\Sigma_{2N}^{P}(\hat{u}) &= \frac{h^d}{N(N-1)}\sum_i\sum_{j\neq i}\hat{u}_i^2\left\{P_{ij}^2 + P_{ij}P_{ji}\right\}\hat{u}_j^2 \\
\Sigma_{1N}^{R}(\hat{u}) &= \frac{2h^d}{N(N-1)}\sum_i\sum_{j\neq i}\hat{u}_i^2 R_{ij}^2\hat{u}_j^2 \\
\Sigma_{2N}^{R}(\hat{u}) &= \frac{h^d}{N(N-1)}\sum_i\sum_{j\neq i}\hat{u}_i^2\left\{R_{ij}^2 + R_{ij}R_{ji}\right\}\hat{u}_j^2.
\end{aligned}$$

Proof. *See Appendix.* ■

Note that different choices of variance estimator can lead to a so-called conflict amongst test criteria. For example, writing, $t_{ij} = \hat{u}_i R_{ij} \hat{u}_j$, it follows that

$$S_N^R(\hat{u}) = \frac{\sum_i \sum_{j \neq i} t_{ij}}{\sqrt{2 \sum_i \sum_{j \neq i} t_{ij}^2}} \xrightarrow{d} N(0,1)$$

as $n \to \infty$, $h \to 0$. Equally, writing $t_{ij}^* = \dfrac{1}{2} \hat{u}_i \left\{ R_{ij} + R_{ji} \right\} \hat{u}_j$, we also have

$$S_N^{R^*}(\hat{u}) = \frac{\sum_i \sum_{j \neq i} t_{ij}^*}{\sqrt{2 \sum_i \sum_{j \neq i} t_{ij}^{*2}}} \xrightarrow{d} N(0,1)$$

but $S_N^{R^*}(\hat{u}) \geq S_N^R(\hat{u})$, in finite samples, indicating a potential conflict among test criteria depending on the choice of variance estimator.

3 Monte Carlo Simulations

This section reports the findings of a small Monte Carlo study of the finite sample performance of various test statistics of the form $S_N^T(\hat{u})$, including Zheng's (1996) statistic, the Ellison and Ellison (2000) statistic, the LLR based statistic, and modifications thereof.

The Data Generation Process (DGP) employed under the null is the following regression model

$$y_i = 1 + X_{i1} + X_{i2} + u_i \sqrt{1 + X_{i1}^2}$$

where $X_{i1} = Z_{i1} + V_i$, $X_{i2} = Z_{i2} + V_i$, with $Z_{ij} \sim U(-\pi, \pi)$, $j = 1, 2$, $V_i \sim U(-\pi, \pi)$ and $u_i \sim N(0, 1)$, with Z_{i1}, Z_{i2}, V_i and u_i all *iid*. The test statistics considered are:

1. $S_N^K(\hat{u}) = \dfrac{\sum_i \sum_{j \neq i} t_{ij}}{\sqrt{2 \sum_i \sum_{j \neq i} t_{ij}^2}}$, $t_{ij} = \hat{u}_i K_{ij} \hat{u}_j$

2. $S_N^P(\hat{u}) = \dfrac{\sum_i \sum_{j \neq i} t_{ij}}{\sqrt{2 \sum_i \sum_{j \neq i} t_{ij}^2}}$, $t_{ij} = \dfrac{1}{2} \hat{u}_i \left\{ P_{ij} + P_{ji} \right\} \hat{u}_j$

3. $S_N^R(\hat{u}) = \dfrac{\sum_i \sum_{j \neq i} t_{ij}}{\sqrt{2 \sum_i \sum_{j \neq i} t_{ij}^2}}$, $t_{ij} = \dfrac{1}{2} \hat{u}_i \left\{ R_{ij} + R_{ji} \right\} \hat{u}_j$

4. $S_N^G(\hat{u}) = \dfrac{\sum_i \sum_{j \neq i} t_{ij}}{\sqrt{2 \sum_i \sum_{j \neq i} t_{ij}^2}}$, $t_{ij} = \dfrac{1}{2} \hat{u}_i \{\Delta_N(X_i) R_{ij} + \Delta_N(X_j) R_{ji}\} \hat{u}_j$

5. $S_N^H(\hat{u}) = \dfrac{\sum_i \sum_{j \neq i} t_{ij}}{\sqrt{2 \sum_i \sum_{j \neq i} t_{ij}^2}}$, $t_{ij} = \dfrac{1}{2} \hat{u}_i \{f_N(X_i) R_{ij} + f_N(X_j) R_{ji}\} \hat{u}_j$

The Kernel employed was a standard normal product Kernel, $K(\zeta) = \frac{1}{(2\pi)^{-2}} \exp\left(-\frac{(\zeta_1 + \zeta_2)^2}{2}\right)$, with smoothing parameter $h = N^{-1/6}$, which satisfies $h \to 0$, $Nh^4 \to \infty$, and Assumptions 4,6 and 7. Both asymptotic and bootstrap critical values were employed, in a one-sided test procedure. The (wild) bootstrap DGP is given by

$$y_i^* = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \hat{u}_i \eta_i$$

where the $\hat{\beta}_j$ are the Ordinary Least Squares (OLS) estimators under the null, $\hat{u}_i$ the OLS residual and η_i an *iid* random variable which takes the value -1 or 1 with probability $\dfrac{1}{2}$. This form of the wild bootstrap was mentioned by Liu (1998) and has been advocated by Davidson and Flachaire (2000). It has been shown to perform well by Godfrey and Orme (2002, 2004). From the bootstrap sample, a typical test statistic is denoted $S_N^T(\tilde{u}^*)$, where $T = K, P, R, G$ or H, and $\tilde{u}^*$ is the bootstrap OLS residual. Repeating this for a large number of bootstrap samples one is able to construct a "bootstrap" critical value; see, for example, Li and Wang (1998). Letting $Z_N = \{y_i, X_{i1}, X_{i2}\}_{i=1}^N$ denote the observed data, then adaptations of the results in Section 2.3 show that, conditional on Z_N, $S_N^T(\tilde{u}^*) | Z_N \xrightarrow{d} N(0,1)$, which establishes the asymptotic validity of the wild bootstrap in the sense that the use of bootstrap critical values should perform at least as well as those obtained from standard $O(1)$ asymptotic theory (i.e., standard normal critical values).

For all experiments in this small Monte Carlo study, $N = 100$; 2000 replications of sample data were generated; and, 399 bootstrap samples for each replication.

Table 1 gives the empirical significance levels (at the nominal levels of 10%, 5% and 1%) for each of the statistics.

Table 1

	Empirical Significance Levels					
	Wild Bootstrap			Asymptotic		
	10%	5%	1%	10%	5%	1%
$S_N^K(\hat{u})$	9.80	5.60	1.40	6.80	3.55	1.45
$S_N^P(\hat{u})$	10.00	5.00	1.15	5.15	2.60	0.55
$S_N^R(\hat{u})$	10.35	5.15	0.95	7.00	2.60	0.30
$S_N^G(\hat{u})$	9.25	5.10	1.55	6.75	3.90	1.40
$S_N^H(\hat{u})$	10.90	5.30	0.95	5.90	2.95	0.30

As expected, the bootstrap performs well at controlling significance levels (there is very good agreement between nominal and empirical values) and this close agreement is reflected in other experiments, not reported here, which use different sample sizes, smoothing parameters, and DGPs. As previously found, the use of asymptotic critical values is not to be recommended as it leads to tests which are significantly undersized.

Table 2

	Rejection frequencies: DGP2, $\gamma = 0.25$					
	Wild Bootstrap			Asymptotic		
	10%	5%	1%	10%	5%	1%
$S_N^K(\hat{u})$	62.60	49.25	28.15	60.50	47.85	27.70
$S_N^P(\hat{u})$	67.25	55.70	30.50	59.65	45.95	21.50
$S_N^R(\hat{u})$	43.35	30.45	10.60	39.15	23.75	3.75
$S_N^G(\hat{u})$	60.20	47.20	26.25	59.85	48.30	27.50
$S_N^H(\hat{u})$	66.45	53.40	29.10	61.75	49.10	24.20

Tables 2 and 3 report rejection frequencies for the following DGP (*DGP2*):

$$y_i = y_i = 1 + X_{i1} + X_{i2} + \gamma X_{i1} X_{i2} + u_i \sqrt{1 + X_{i1}^2},$$

for $\gamma = 0.25$ (Table 2) and $\gamma = 0.5$ (Table 3). The rejection frequencies listed under "Asymptotic" are included for completeness for should be treated with caution since they are not size-adjusted. The Wild Bootstrap rejection frequencies indicate that all tests have power and, although these rejection frequencies can be sensitive to the choice of smoothing parameter, it is noteworthy that across all experiments (not all reported here to conserve space) the Ellison and Ellison (2000) statistic, $S_N^P(\hat{u})$, appeared most powerful with that based on the Local Linear Regression Estimator, $S_N^R(\hat{u})$, least powerful. This is an unexpected result, since prior intuition suggested that the Local Linear Regression technique should yield greater power.

Table 3

	Power: DGP2, $\gamma = 0.5$					
	Wild Bootstrap			Asymptotic		
	10%	5%	1%	10%	5%	1%
$S_N^K(\hat{u})$	99.50	98.90	95.55	99.45	99.05	96.55
$S_N^P(\hat{u})$	99.50	98.80	95.65	99.40	98.50	90.85
$S_N^R(\hat{u})$	92.30	82.95	53.10	91.75	79.30	33.40
$S_N^G(\hat{u})$	98.85	98.05	93.90	99.20	98.50	95.15
$S_N^H(\hat{u})$	99.45	98.85	96.15	99.45	98.60	94.95

4 Conclusion

Extending the work of Zheng (1996) and Ellison and Ellison (2000), a consistent conditional moment test has been constructed by employing the Local Linear Regression estimator and its limiting behaviour analysed. Its limit distribution is standard normal but, moreover, it is asymptotically equivalent to the Ellison and Ellison (2000) test statistics, which is based on the Nadaraya-Watson regression estimator. In addition, a variety of consistent variance estimators arise from our analysis which might be employed when implementing test procedures.

Although asymptotically equivalent, the results of a small Monte Carlo study indicates that the Ellison and Ellison (2000) procedure, which uses the Nadaraya-Watson estimator, is far more powerful than that which uses the Local Linear Regression estimator. Moreover, the Ellison and Ellison (2000) procedure is more powerful than that of Zheng (2000).

Having proposed the use of the Local Linear Regression estimator in the construction of Kernel-based nonparametric tests, Monte Carlo evidence suggests that sue of the simpler Nadaraya-Watson estimator still yields a more powerful procedure. Future research will address adaptations to these test statistics (via trimming) which avoid the requirement of compact support for the regressor density; the use of RESET versions, to avoid the curse of dimensionality; and, a comparison with other parametric and nonparametric procedures proposed in the literature; e.g., Wang (1998)

5 Appendix

Here we collect together the main of the proofs.

Proof of Lemma 1

(i) Write $V_N^Q(\varepsilon)$ as $V_N^{Q^*}(\varepsilon)$, where $Q^* = \frac{1}{2}(Q + Q')$, and Hall's conditions are easily verified.

For (ii), note that by results on variances of quadratic forms[2],

$$\begin{aligned} var\left[V_N^{Q^*}(\varepsilon)\right] &= \frac{N^2h^d}{N^2(N-1)^2}\sum_i\sum_{j\neq i}E\left[\varepsilon_i^2\varepsilon_j^2K_{ij}^2\left\{\frac{1}{f^2(X_i)}+\frac{1}{f(X_i)}\frac{1}{f(X_j)}\right\}\right] \\ &= \frac{N^2}{N(N-1)}\int K^2(\zeta)\,\sigma^2(x)\,\sigma^2(x-\zeta h) \\ &\quad \times\left\{\frac{1}{f^2(x)}+\frac{1}{f(x)}\frac{1}{f(x-\zeta h)}\right\}f(x)f(x-\zeta h)dx'd\zeta' \\ &= 2\int K^2(\zeta)\,d\zeta'\int\left\{\sigma^2(x)\right\}^2dx'+o(1). \end{aligned}$$

It is also true that

$$\begin{aligned} &\frac{N^2h^d}{N^2(N-1)^2}\sum_i\sum_{j\neq i}E\left[\varepsilon_i^2\varepsilon_j^2K_{ij}^2\frac{1}{f^2(X_i)}\right] \\ =\ &\frac{N^2}{N(N-1)}\int K^2(\zeta)\,\sigma^2(x)\,\sigma^2(x-\zeta h)\,\frac{1}{f^2(x)}f(x)f(x-\zeta h)dx'd\zeta' \\ =\ &2\int K^2(\zeta)\,d\zeta'\int\left\{\sigma^2(x)\right\}^2dx'+o(1). \end{aligned}$$

It follows, from *H-projection* arguments (see for example Zheng (1996)) that both

$$\begin{aligned} \Sigma_{1N}^Q(\varepsilon) &= \frac{2h^d}{N(N-1)}\sum_i\sum_{j\neq i}\varepsilon_i^2Q_{ij}^2\varepsilon_j^2 \\ \Sigma_{2N}^Q(\varepsilon) &= \frac{h^d}{N(N-1)}\sum_i\sum_{j\neq i}\varepsilon_i^2\left\{Q_{ij}^2+Q_{ij}Q_{ji}\right\}\varepsilon_j^2 \end{aligned}$$

are consistent for Σ.

Proof of Lemma 2.

Adapt Zheng's (1996) method of proof as follows. Take a second order Mean Value Expansion of $V_N^D(\hat{u})$ about $\hat{\theta}=\theta_0$ yielding

$$V_N^Q(\hat{u})=V_N^Q(\varepsilon)+V_{2N}'\sqrt{N}\left(\hat{\theta}-\theta_0\right)+\frac{1}{2}\sqrt{N}\left(\hat{\theta}-\theta_0\right)'V_{3N}\left(\hat{\theta}-\theta_0\right)$$

[2]See, for example, Li and Wang (1998), details under compact support conditions of Assumption 1.

where

$$
\begin{aligned}
V_N^Q(\varepsilon) &= \frac{Nh^{d/2}}{N(N-1)} \sum_i \sum_{j\neq i} \varepsilon_i Q_{ij} \varepsilon_j \\
V_{2N}' &= \frac{\sqrt{Nh^d}}{N(N-1)} \sum_i \sum_{j\neq i} \varepsilon_i Q_{ij} \left(\nabla_\theta \varepsilon_j\right)' + \frac{\sqrt{Nh^d}}{N(N-1)} \sum_i \sum_{j\neq i} \varepsilon_i Q_{ji} \left(\nabla_\theta \varepsilon_j\right)' \\
V_{3N} &= \frac{h^{d/2}}{N(N-1)} \sum_i \sum_{j\neq i} u_i^* Q_{ij} \left(\nabla_{\theta\theta} u_j^*\right) + \frac{h^{d/2}}{N(N-1)} \sum_i \sum_{j\neq i} \left(\nabla_\theta u_i^*\right) Q_{ij} \left(\nabla_\theta u_j^*\right)' \\
&\quad + \frac{h^{d/2}}{N(N-1)} \sum_i \sum_{j\neq i} u_i^* Q_{ji} \left(\nabla_{\theta\theta} u_j^*\right) + \frac{h^{d/2}}{N(N-1)} \sum_i \sum_{j\neq i} \left(\nabla_\theta u_i^*\right) Q_{ji} \left(\nabla_\theta u_j^*\right)'
\end{aligned}
$$

where $\varepsilon_i = u(\theta_0; W_i)$, $u_i^* = u(\theta^*; W_i)$, $\nabla_\theta \varepsilon_i = \dfrac{\partial u(\theta_0; W_i)}{\partial \theta}$, $\nabla_\theta u_i^* = \dfrac{\partial u(\theta^*; W_i)}{\partial \theta}$, $\nabla_{\theta\theta} u_i^* = \dfrac{\partial^2 u(\theta^*; W_i)}{\partial\theta\partial\theta'}$.
By Lemma 1, $V_N^Q(\varepsilon) \xrightarrow{d} N(0, \Sigma)$. Now we show that:(i) $V_{2N} = o_p(1)$; (ii) $V_{3N} = o_p(1)$
To prove (i), re-write the first term in V_{2N}' as

$$
A_N' = \frac{\sqrt{Nh^d}}{N(N-1)} \sum_i \sum_{j\neq i} e_i K_{ij} m_j' + \frac{\sqrt{Nh^d}}{N(N-1)} \sum_i \sum_{j\neq i} e_i K_{ij} v_j'
$$

where $e_i = \varepsilon_i / f(X_i)$, $v_j' = \nabla_\theta \varepsilon_j - m_j$ and $m_j = E\left[\nabla_\theta \varepsilon_j | X_j\right]$, and consider $A_N'\lambda$ for any $\lambda \neq 0$. The first term in $A_N'\lambda$ is $O_p\left(h^{d/2}\right)$ by Zheng (1996, Lemma 3.3b). The second term in $A_N'\lambda$ is

$$
u_N = \frac{\sqrt{Nh^d}}{N(N-1)} \sum_i \sum_{j\neq i} e_i Q_{ij} \eta_j, \quad \eta_j = v_j'\lambda,
$$

where $E\left[\eta_j | X_j\right] = 0$ and $E\left[\varepsilon_i^* | X_i\right] = 0$. Now, fairly general results on variances of quadratic forms gives

$$
\begin{aligned}
var\left[u_N\right] &\leq \frac{2Nh^d}{N^2 (N-1)^2} \sum_i \sum_{j\neq i} E\left[e_i^2 K_{ij}^2 \eta_j^2\right] \\
&= O\left(N^{-1}\right) O\left(1\right),
\end{aligned}
$$

since, by results similar to those used in Lemma 1, $\dfrac{2h^d}{N(N-1)} \sum_i \sum_{j\neq i} E\left[\varepsilon_i^{*2} K_{ij}^2 \eta_j^2\right] = O(1)$. Therefore $u_N = o_p(1)$ for all $\lambda \neq 0$. Thus, A_N' converges in probability to zero.
The same approach can be used to show that the second term in

V_{2N}, $\dfrac{\sqrt{Nh^d}}{N(N-1)} \sum_i \sum_{j \neq i} e_i K_{ji} \left(\nabla_\theta \varepsilon_j\right)'$, is $o_p(1)$

The proof of (ii), is identical to that of Zheng (1996, Lemma 3.3d).

Proof of Theorem 3.

Take a first order Mean Value Expansion of $V_N^D(\hat{u})$ about $\hat{\theta} = \theta_0$, which yields

$$V_N^D(\hat{u}) = V_N^D(\varepsilon) + V_{2N}'\sqrt{N}\left(\hat{\theta} - \theta_0\right) + \frac{1}{2}\sqrt{N}\left(\hat{\theta} - \theta_0\right)' V_{3N}\left(\hat{\theta} - \theta_0\right)$$

where

$$\begin{aligned}
V_N^D(\varepsilon) &= \frac{Nh^{d/2}}{N(N-1)} \sum_i \sum_{j \neq i} \varepsilon_i a_N(X_i) K_{ij} \varepsilon_j \\
V_{2N}' &= \frac{\sqrt{Nh^d}}{N(N-1)} \sum_i \sum_{j \neq i} \varepsilon_i a_N(X_i) K_{ij} \left(\nabla_\theta \varepsilon_j\right)' \\
&\quad + \frac{\sqrt{Nh^d}}{N(N-1)} \sum_i \sum_{j \neq i} \varepsilon_i K_{ij} a_N(X_j) \left(\nabla_\theta \varepsilon_j\right)' \\
V_{3N} &= \frac{h^{d/2}}{N(N-1)} \sum_i \sum_{j \neq i} u_i^* a_N(X_i) K_{ij} \left(\nabla_{\theta\theta} u_j^*\right) \\
&\quad + \frac{h^{d/2}}{N(N-1)} \sum_i \sum_{j \neq i} \left(\nabla_\theta u_i^*\right) a_N(X_i) K_{ij} \left(\nabla_\theta u_j^*\right)' \\
&\quad + \frac{h^{d/2}}{N(N-1)} \sum_i \sum_{j \neq i} u_i^* K_{ij} a_N(X_j) \left(\nabla_{\theta\theta} u_j^*\right) \\
&\quad + \frac{h^{d/2}}{N(N-1)} \sum_i \sum_{j \neq i} \left(\nabla_\theta u_i^*\right) K_{ij} a_N(X_j) \left(\nabla_\theta u_j^*\right)'.
\end{aligned}$$

and show that:(i) $V_N^D(\varepsilon) = o_p(1)$; (ii) $V_{2N} = o_p(1)$; (iii) $V_{3N} = o_p(1)$.

Proof of (i):

Note that $V_N^D(\varepsilon)$ has second moment

$$
\begin{aligned}
E\left|V_N^D(\varepsilon)\right|^2 &= \frac{N^2h^d}{N^2(N-1)^2}\sum_i\sum_{j\neq i} E\left[\varepsilon_i^2\varepsilon_j^2K_{ij}^2\left\{a_N^2(X_i)+a_N(X_i)a_N(X_j)\right\}\right] \\
&\leq 2\frac{N^2h^d}{N^2(N-1)^2}\sum_i\sum_{j\neq i} E\left[\varepsilon_i^2\varepsilon_j^2K_{ij}^2a_N^2(X_i)\right] \\
&\leq 2E\left[\sup_{x\in\mathcal{S}}|a_N(x)|^2\frac{N^2h^d}{N^2(N-1)^2}\sum_i\sum_{j\neq i}\varepsilon_i^2\varepsilon_j^2K_{ij}^2\right] \\
&= o(1)\frac{2N^2h^d}{N^2(N-1)^2}\sum_i\sum_{j\neq i} E\left[\varepsilon_i^2\varepsilon_j^2K_{ij}^2\right] \\
&= o(1)O(1)
\end{aligned}
$$

thus $V_N = o_p(1)$.
Proof of (ii):
Write the first term in V_{2N}' as

$$
A_N' = \frac{\sqrt{Nh^d}}{N(N-1)}\sum_i\sum_{j\neq i}\varepsilon_i a_N(X_i)K_{ij}v_j' + \frac{\sqrt{Nh^d}}{N(N-1)}\sum_i\sum_{j\neq i}\varepsilon_i a_N(X_i)K_{ij}m_j'
$$

where $m_j = E\left[\nabla_\theta\varepsilon_j|X_j\right]$, $v_j = \nabla_\theta\varepsilon_j - m_j$, and consider, first, $u_N = \frac{\sqrt{Nh^d}}{N(N-1)}\sum_i\sum_{j\neq i}\varepsilon_i a_N(X_i)K_{ij}v_j'\lambda$, for any $\lambda\neq 0$. We have, with $\eta_j = v_j'\lambda$,

$$
\begin{aligned}
var\left[u_N\right] &\leq \frac{2Nh^d}{N^2\left(N-1\right)^2}E\left[\sum_i\sum_{j\neq i}a_N^2(X_i)\varepsilon_i^2K_{ij}^2\eta_j^2\right] \\
&= o(1)\frac{2Nh^d}{N^2\left(N-1\right)^2}\sum_i\sum_{j\neq i}E\left[\varepsilon_i^2K_{ij}^2\eta_j^2\right] \\
&= o(1)O(N^{-1})
\end{aligned}
$$

Thus, in general, $u_N = o_p(1)$. Therefore $A_N' = \frac{\sqrt{Nh^d}}{N(N-1)}\sum_i\sum_{j\neq i}\varepsilon_i a_N(X_i)K_{ij}m_j' + o_p(1)$. To verify that $A_N' = o_p(1)$ it is sufficient to show that $z_N = \frac{\sqrt{Nh^d}}{N(N-1)}\sum_i\sum_{j\neq i}\varepsilon_i a_N(X_i)K_{ij}\xi_j$ is $o_p(1)$, where $\xi_j = m_j'\lambda$,for any

$\lambda \neq 0$. Consider

$$\begin{aligned}
var\left[z_N\right] &= \frac{Nh^d}{N^2\left(N-1\right)^2} E\left[\sum_i \sum_{j\neq i} a_N^2(X_i)\varepsilon_i^2 K_{ij}^2 \xi_j^2\right] \\
&\quad + \frac{Nh^d}{N^2\left(N-1\right)^2} E\left[\sum_i \sum_{j\neq i} \sum_{l\neq j, l\neq i} a_N^2(X_i)\varepsilon_i^2 K_{ij}\xi_j K_{il}\xi_l\right] \\
&\leq \sup_x \left|a_N(x)\right|^2 \left\{ \frac{Nh^d}{N^2\left(N-1\right)^2} \sum_i \sum_{j\neq i} E\left[\varepsilon_i^2 K_{ij}^2 \xi_j^2\right] \right. \\
&\quad \left. + \frac{Nh^d}{N^2\left(N-1\right)^2} \sum_i \sum_{j\neq i} \sum_{l\neq j, l\neq i} E\left[\varepsilon_i^2 K_{ij}\left|\xi_j\right| K_{il}\left|\xi_l\right|\right] \right\}.
\end{aligned}$$

As before, it can be shown that $\dfrac{h^d}{N\left(N-1\right)} \sum_i \sum_{j\neq i} E\left[\varepsilon_i^2 K_{ij}^2 \xi_j^2\right] = O(1)$, so that the first term in $var\left[z_N\right]$ is $o(N^{-1})$. Furthermore,

$$\begin{aligned}
& \frac{1}{N\left(N-1\right)\left(N-2\right)} \sum_i \sum_{j\neq i} \sum_{l\neq j, l\neq i} E\left[\varepsilon_i^{*2} K_{ij}\left|\xi_j\right| K_{il}\left|\xi_l\right|\right] \\
= & \int \left\{ \frac{1}{h^d} K\left(\frac{x_i - x_j}{h}\right) \frac{1}{h^d} K\left(\frac{x_i - x_l}{h}\right) f^{-2}(x_i)\sigma^2(x_i) \right. \\
& \left. \times \left|\xi(x_j)\right| \left|\xi(x_l)\right| f(x_i) f(x_j) f(x_l) \right\} dx_i dx_j dx_l \\
= & \int K(s)K(t)\sigma^2(x)\left|\xi(x - sh)\right| \left|\xi(x - th)\right| \frac{f(x - sh)f(x - th)}{f(x)} dx ds dt \\
\to & \int \sigma^2(x)\xi^2(x) f(x) dx = E\left[\sigma^2(X)\xi^2(X)\right] = O(1),
\end{aligned}$$

where we have made the substitutions $x \equiv x_i$, $s = \left(x_i - x_j\right)/h$, $t = \left(x_i - x_l\right)/h$, so that the second term in $var\left[z_N\right]$ is $o(h^d)$. Therefore, $z_N = o_p(1)$ Thus the first term in V_{2N}' is $o_p(1)$. By similar arguments, the second term in V_{2N}' is also $o_p(1)$.

Proof of (iii)

Consider the first term in V_{3N}. We have

$$\begin{aligned}
& \left\| \frac{h^{d/2}}{N(N-1)} \sum_i \sum_{j\neq i} u_i^* a_N(X_i) K_{ij}\left(\nabla_{\theta\theta} u_j^*\right) \right\| \\
\leq & \frac{h^{d/2}}{N(N-1)} \sum_i \sum_{j\neq i} K_{ij}\left|a_N(X_i)\right| \left|u_i^*\right| \left\|\nabla_{\theta\theta} u_j^*\right\| \\
\leq & h^{d/2} \sup_x \left|a_N(x)\right| \frac{1}{N(N-1)} \sum_i \sum_{j\neq i} K_{ij} b(W_i) b(W_j)
\end{aligned}$$

by Assumption 2.1. Again, it is easy to show using *H-projection* arguments that $\dfrac{1}{N(N-1)}\sum_i \sum_{j\neq i} K_{ij} b(W_i) b(W_j)$ is $O_p(1)$ the first term in V_{3N} is $o_p(h^{d/2})$, since $\sup_x |a_N(x)| = o_p(1)$. By similar arguments it can be shown that all such terms in V_{3N} are $o_p(1)$, so that V_{3N} is $o_p(1)$.

To prove Lemma 4, the following results are required.

Lemma 8 *(Van Ryzin, 1969) Let $\{Y_n\}_{n=1}^N$ and $\{Z_n\}_{n=1}^N$ be sequences of random variables on a probability space $(\Omega, \mathcal{F}, P)$. Let $\{\mathcal{F}_n\}$ be a sequence of Borel fields, $\mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \mathcal{F}$, where Y_n and Z_n are measurable with respect to $\mathcal{F}_n$. If*

1. $0 \leq Y_n$ *a.s.*
2. $EY_1 < \infty$
3. $E[Y_{n+1}|\mathcal{F}_n] \leq Y_n + Z_n$, *a.s.*
4. $\sum_{n=1}^{\infty} E|Z_n| < \infty$

then Y_n converges a.s. to a finite limit.

Lemma 9 *(i) The w_i' are independently and identically distributed as $W' = (Y, X')$, with joint density $g(w)$; (ii) X is a continuous $(d \times 1)$ random variable with bounded, uniformly continuous density $f(x)$. Define $\varepsilon_i = \varepsilon(W_i)$, a scalar random variable, and let $\beta(x) = E[\varepsilon(W)|X = x]$ be uniformly continuous on $\mathcal{S}$, such that $E|\varepsilon(W)|$ and $E\left|\varepsilon^2(W)\right|$ both exist and are finite. Let*

$$q_N(x) = \frac{1}{N}\sum_{j=1}^{N} K_h(x - X_j)\,\varepsilon(W_j)$$

where $K_h(s) = h^{-d}K(s/h)$ and $K(s)$ is a product kernel, with absolutely integrable characteristic function $\phi(\xi)$. Additionally, assume that (i) $h_N \longrightarrow 0$ as $N \longrightarrow \infty$, (ii) there is $\delta \in (0,1)$ and $0 < \gamma$ such that $N^{\delta} h^{2d} = \gamma$, (iii) $g(c) = \int |\phi(c\xi) - \phi(\xi)|\, d\xi'$ is Locally Lipschitz of order 1 at $c = 1$.Then,

$$\sup_x |q_N(x) - q(x)| \longrightarrow 0$$

where $q(x) = f(x)\beta(x)$.

Proof. *Let us start by noting that*

$$\sup_{x\in S}|q_N(x) - q(x)| = \sup_{x\in S}|q_N(x) - Eq_N(x)| + \sup_{x\in\mathcal{S}}|Eq_N(x) - q(x)|$$

Consider,

$$\left|Eq_N(x) - q(x)\right| = \left|\int K(s)\left\{\beta(x-sh)f(x-sh) - \beta(x)f(x)\right\}ds\right|.$$

The Kernel assumption and uniform continuity imply $\sup_{x\in\mathcal{S}}\left|Eq_N(x) - q(x)\right| \to 0$*. By Assumption 2.3,* $K(\zeta) = (2\pi)^{-d}\int\phi(\xi)e^{-i\xi'\zeta}d\xi'$*, so that*

$$q_N(x) = \frac{1}{(2\pi)^d}\int\psi_N(\zeta)\,e^{-i\zeta' x}\phi(\zeta h)\,d\zeta' \tag{11}$$

where $\psi_N(\zeta) = N^{-1}\sum_{j=1}^{N} e^{i\zeta' X_j}\varepsilon(W_j)$ *is a unbiased estimator for* $\psi(\zeta) = E\left[e^{i\zeta' X}\varepsilon(W)\right]$*. Given that* $\left|e^{-i\zeta' x}\right| = 1$*, it follows that*

$$\left|q_N(x) - Eq_N(x)\right| \leq \frac{1}{(2\pi)^d}\int\left|\psi_N(\zeta) - \psi(\zeta)\right|\left|\phi(\zeta h)\right|d\zeta'$$

which does not depend on x*. Therefore*

$$\sup_{x\in\mathcal{S}}\left|q_N(x) - Eq_N(x)\right| \leq \frac{1}{(2\pi)^d}\int\left|\psi_N(\zeta) - \psi(\zeta)\right|\left|\phi(\zeta h)\right|d\zeta'$$

or, equivalently

$$\begin{aligned}
\left\{\sup_{x\in\mathcal{S}}\left|q_N(x) - Eq_N(x)\right|\right\}^2 &\leq \left\{\frac{1}{(2\pi)^d}\int\left|\psi_N(\zeta) - \psi(\zeta)\right|\left|\phi(\zeta h)\right|d\zeta'\right\}^2 \\
&\leq \frac{1}{(2\pi)^{2d}}\int\left|\phi(\zeta h)\right|d\zeta\left\{\int\left|\psi_N(\zeta) - \psi(\zeta)\right|^2\left|\phi(\zeta h)\right|d\zeta'\right\} \\
&= \left\{\frac{1}{(2\pi)^{2d}}\int\left|\phi(\xi)\right|d\xi'\right\}Y_N
\end{aligned}$$

where

$$Y_N = \frac{1}{h^d}\int\left|\psi_N(\zeta) - \psi(\zeta)\right|^2\left|\phi(\zeta h)\right|d\zeta'. \tag{12}$$

The objective is to show that Y_N *converges to zero almost everywhere. We do so by verifying the conditions in Lemma 8. In what follows it will be convenient to recover the dependence of* h *on* N*, so that we shall now write* h_N *instead .*
Under condition (ii) *there is a* $\delta \in (0,1)$ *such that* $N^\delta h_N^{2d} = \gamma$*. When* $N = 1$ *it follows that* $h_1^{2d} = \gamma$*, so that* $h_1^d = \sqrt{\gamma}$ *and* $h_1 = \gamma^{1/2d}$*. As a result of these*

conditions,

$$\begin{aligned}
EY_1 &= \frac{1}{h_1^d}\int E\left|\psi_1(\zeta)-\psi(\zeta)\right|^2\left|\phi(\zeta h_1)\right|d\zeta' \\
&\leq \frac{1}{\gamma}\int E\left[\varepsilon^2(W)\right]\left|\phi\left(\zeta\gamma^{1/2d}\right)\right|d\zeta' \\
&= \frac{1}{\sqrt{\gamma}}E\left[\varepsilon^2(W)\right]\int\left|\phi(\xi)\right|d\xi' \\
&< \infty
\end{aligned}$$

where we have used the result in Lemma 10. Furthermore, it is easy to see that $0 \leq Y_N$*, and, by means of usual techniques,*

$$\begin{aligned}
EY_N &= \frac{1}{h_N^d}\int E\left|\psi_N(\zeta)-\psi(\zeta)\right|^2\left|\phi(\zeta h_N)\right|d\zeta' \\
&\leq \frac{1}{Nh_N^{2d}}E\left[\varepsilon^2(W)\right]\int\left|\phi(\xi)\right|d\xi' \qquad (13)
\end{aligned}$$

Therefore,

$$EY_N \to 0 \text{ as } N \to \infty$$

because by condition (ii), $Nh_N^{2d} = N^{1-\delta}\gamma \to \infty$*. We now need to show that* $E\left[Y_{N+1}|\mathcal{F}_N\right] \leq Y_N + Z_N$*, a.s. for certain quantity* Z_N *to be defined. Let us write* $\psi_{N+1}(\zeta)$ *and* $\psi(\zeta)$ *in the following way:*

$$\psi_{N+1}(\zeta) = \frac{N}{N+1}\psi_N(\zeta)+\frac{1}{N+1}e^{i\zeta' X_{N+1}} \qquad (14)$$

$$\psi(\zeta) = \frac{N}{N+1}\psi(\zeta)+\frac{1}{N+1}\psi(\zeta) \qquad (15)$$

so that

$$\begin{aligned}
Y_{N+1} &= \frac{1}{h_{N+1}^d}\int\left|\psi_{N+1}(\zeta)-\psi(\zeta)\right|^2\left|\phi(\zeta h_{N+1})\right|d\zeta' \\
&\leq \frac{1}{(N+1)^2h_{N+1}^d}\int\left|\phi(\zeta h_{N+1})\right|d\zeta' \\
&\quad +\frac{1}{h_{N+1}^d}\int\left|\frac{N}{N+1}\psi_N(\zeta)-\frac{N}{N+1}\psi(\zeta)-\frac{1}{N+1}\psi(\zeta)\right|^2\left|\phi(\zeta h_{N+1})\right|d\zeta'.
\end{aligned}$$

The second term in the right hand side can itself be expanded as follows:

$$\begin{aligned}
& \frac{1}{h_{N+1}^d}\int\left|\frac{N}{N+1}\psi_N(\zeta)-\frac{N}{N+1}\psi(\zeta)-\frac{1}{N+1}\psi(\zeta)\right|^2|\phi(\zeta h_{N+1})|\,d\zeta' \\
\leq\ & \frac{1}{h_{N+1}^d}\left(\frac{N}{N+1}\right)^2\int|\psi_N(\zeta)-\psi(\zeta)|^2|\phi(\zeta h_{N+1})|\,d\zeta' \\
& +\frac{1}{h_{N+1}^d(N+1)^2}\int|\psi(\zeta)|^2|\phi(\zeta h_{N+1})|\,d\zeta' \\
\leq\ & \frac{1}{h_{N+1}^d}\left(\frac{N}{N+1}\right)^2\int|\psi_N(\zeta)-\psi(\zeta)|^2|\phi(\zeta h_N)|\,d\zeta' \\
& +\frac{1}{h_{N+1}^d}\left(\frac{N}{N+1}\right)^2\int|\psi_N(\zeta)-\psi(\zeta)|^2|\phi(\zeta h_{N+1})-\phi(\zeta h_N)|\,d\zeta' \\
& +\frac{1}{h_{N+1}^d(N+1)^2}\int|\psi(\zeta)|^2|\phi(\zeta h_{N+1})|\,d\zeta'.
\end{aligned}$$

Thus,

$$\begin{aligned}
Y_{N+1}\ \leq\ & \frac{1}{(N+1)^2h_{N+1}^d}\int|\phi(\zeta h_{N+1})|\,d\zeta' \\
& +\frac{1}{h_{N+1}^d}\left(\frac{N}{N+1}\right)^2\int|\psi_N(\zeta)-\psi(\zeta)|^2|\phi(\zeta h_N)|\,d\zeta' \\
& +\frac{1}{h_{N+1}^d}\left(\frac{N}{N+1}\right)^2\int|\psi_N(\zeta)-\psi(\zeta)|^2|\phi(\zeta h_{N+1})-\phi(\zeta h_N)|\,d\zeta' \\
& +\frac{1}{h_{N+1}^d(N+1)^2}\int|\psi(\zeta)|^2|\phi(\zeta h_{N+1})|\,d\zeta'. \\
\leq\ & +\frac{1}{h_{N+1}^d}\left(\frac{N}{N+1}\right)^2\int|\psi_N(\zeta)-\psi(\zeta)|^2|\phi(\zeta h_N)|\,d\zeta' \\
& +\frac{1}{h_{N+1}^d}\left(\frac{N}{N+1}\right)^2\int|\psi_N(\zeta)-\psi(\zeta)|^2|\phi(\zeta h_{N+1})-\phi(\zeta h_N)|\,d\zeta' \\
& +\frac{1}{(N+1)^2h_{N+1}^d}\int|\phi(\zeta h_{N+1})|\left\{1+\{E|\varepsilon(W)|\}^2\right\}d\zeta' \\
\leq\ & \frac{h_N^d}{h_{N+1}^d}\left(\frac{N}{N+1}\right)^2Y_N+S_N+T_N
\end{aligned}$$

where we have used the inequality $|\psi(\zeta)|^2\leq\{E|\varepsilon(W)|\}^2$ *and for*

$$\begin{aligned}
S_N &= \frac{1}{h_{N+1}^d}\int|\psi_N(\zeta)-\psi(\zeta)|^2|\phi(\zeta h_{N+1})-\phi(\zeta h_N)|\,d\zeta' \\
T_N &= \frac{1}{(N+1)^2h_{N+1}^d}\int|\phi(\zeta h_{N+1})|\left\{1+\{E|Q(x)|\}^2\right\}d\zeta'
\end{aligned}$$

However, from condition (ii), $N^{\delta}h^{2d} = \gamma \Rightarrow h^d = \sqrt{\gamma}N^{-\delta/2}$ *implying that*

$$\frac{h_N^d}{h_{N+1}^d}\left(\frac{N}{N+1}\right)^2 = \left(\frac{N}{N+1}\right)^{2-\beta} \leq 1$$

where $\beta \in (0, 1/2)$*. Then, letting* $Z_N = S_N + T_N$*, it follows that*

$$Y_{N+1} \leq Y_N + Z_N$$

Furthermore, $E[Y_N|\mathcal{F}_N] = Y_N$ *and* $E[Z_N|\mathcal{F}_N] = Z_N$*, so* $E[Y_{N+1}|\mathcal{F}_N] \leq Y_N + Z_N$ *and the third condition of lemma 8 is thus verified. We have shown that* $E[Y_N] \to 0$ *as* $N \to \infty$*; Then, the theorem will be complete if we can show that* $\sum_{i=1}^{\infty} E[Z_N] < \infty$*.*
We start by noting that ,

$$\begin{aligned} ES_N &= \frac{1}{h_{N+1}^d}\int |\psi_N(\zeta) - \psi(\zeta)|^2 |\phi(\zeta h_{N+1}) - \phi(\zeta h_N)| d\zeta' \\ &\leq \frac{1}{Nh_{N+1}^d}E\left[\varepsilon^2(W)\right]\int |\phi(\zeta h_{N+1}) - \phi(\zeta h_N)| d\zeta' \end{aligned}$$

hence letting $c_N = h_{N+1}/h_N \leq 1$*,*

$$\begin{aligned} ES_N &\leq \frac{1}{Nh_{N+1}^d}E\left[\varepsilon^2(W)\right]\int |\phi(\zeta c_N h_N) - \phi(\zeta h_N)| d\zeta' \\ &= \frac{1}{Nh_{N+1}^d h_N^d}E\left[\varepsilon^2(W)\right]\int |\phi(c_N\xi) - \phi(\xi)| d\xi' \end{aligned}$$

By assumption (iii),

$$|g(c)| = \int |\phi(ct) - \phi(t)| dt \leq B|1 - c_N|$$

for some $B > 0$*. Hence*

$$\begin{aligned} ES_N &\leq \frac{1}{Nh_N^{2d}}\frac{E\left[\varepsilon^2(W)\right]}{c_N}B|1 - c_N| \\ &= \frac{1}{Nh_N^{2d}}E\left[\varepsilon^2(W)\right]B\left|\frac{h_N}{h_{N+1}} - 1\right| \end{aligned}$$

but for $\eta = \delta/2d \in (0, 1/2d)$*, we have*

$$\frac{h_N}{h_{N+1}} - 1 = \left(1 + \frac{1}{N}\right)^{\eta} - 1 \leq \frac{1}{N}$$

so that

$$ES_N \leq \frac{1}{N^2 h_N^{2d}}E\left[\varepsilon^2(W)\right]B.$$

In addition to this, $ET_N = O\left(N^2 h_N^{2d}\right)^{-1}$. *By condition (ii),* $N^2 h_N^{2d} = N^{2-\delta}\gamma \to \infty$. *Since* EZ_N *is at most* $N^2 h_N^{2d}$, *we conclude that* $EZ_N \to 0$, *so that the series* $\sum E\left[Z_N\right]$ *is also convergent. The conclusion follows that* $Y_N \to 0$ *a.s. and thus,*

$$\sup_{x\in\mathcal{S}} \left|q_N\left(x\right) - Eq_N\left(x\right)\right| \to 0$$

a.s. ■

Lemma 10 $E\left|\psi_N\left(\zeta\right) - \psi\left(\zeta\right)\right|^2 \leq \frac{1}{N}\left\{E\left[\varepsilon^2\left(W\right)\right]\right\}.$

Proof. *Firstly, note that* $\psi_N\left(\zeta\right) = \dfrac{1}{N}\sum_i \exp(i\zeta' X_i)\varepsilon(W_i)$, *so that*

$$\begin{aligned}
& E\left|\psi_N\left(\zeta\right) - \psi\left(\zeta\right)\right|^2 \\
= \; & E\left|\frac{1}{N}\sum_{i=1}^{N}\varepsilon(W_i)\cos\left(\zeta' X_i\right) - E\left[\varepsilon(W_i)\cos\left(\zeta' X_i\right)\right]\right. \\
& \left. + i\frac{1}{N}\sum_{i=1}^{N}\varepsilon(W_i)\sin\left(\zeta' X_i\right) - E\left[\varepsilon(W_i)\sin\left(\zeta' X_i\right)\right]\right|^2 \\
= \; & E\left|Z_1 + iZ_2\right|^2 \\
= \; & E\left(Z_1 + iZ_2\right)\left(Z_1 - iZ_2\right) \\
= \; & E\left(Z_1^2 + Z_2^2\right) \qquad (16)
\end{aligned}$$

then by independence

$$\begin{aligned}
& E\left(Z_1^2 + Z_2^2\right) \\
= \; & \frac{1}{N}\Big(E\left\{\varepsilon\left(W\right)\cos\left(\zeta' X\right) - E\left[\varepsilon\left(W\right)\cos\left(\zeta' X\right)\right]\right\}^2 \\
& + E\left\{\varepsilon\left(W\right)\sin\left(\zeta' X\right) - E\left[\varepsilon\left(W\right)\sin\left(\zeta' X\right)\right]\right\}^2\Big) \\
= \; & \frac{1}{N}\Big\{E\left[\varepsilon^2\left(W\right)\left\{\cos^2\left(\zeta' X\right) + \sin^2\left(\zeta' X\right)\right\}\right] \\
& - \left(\left[E\left(\varepsilon\left(W\right)\cos\left(\zeta' X\right)\right)\right]^2 + \left[E\left(\varepsilon(W)\sin\left(\zeta' X\right)\right)\right]^2\right)\Big\} \\
\leq \; & \frac{1}{N}E\left[\varepsilon^2\left(W\right)\left\{\cos^2\left(\zeta' X\right) + \sin^2\left(\zeta' X\right)\right\}\right] \\
= \; & \frac{1}{N}\left\{E\left[\varepsilon^2\left(W\right)\right]\right\}
\end{aligned}$$

where the inequality follows from the fact that $E\left(Z_1^2 + Z_2^2\right) \geq 0$ *and*

$$\left(\left[E\left(\varepsilon\left(W\right)\cos\left(\zeta' X\right)\right)\right]^2 + \left[E\left(\varepsilon(W)\sin\left(\zeta' X\right)\right)\right]^2\right) > 0.$$

The last equality follows since $\cos^2\left(\zeta' X\right) + \sin^2\left(\zeta' X\right) = 1$. ■

Proof of Lemma 4

(i)$\sup_{x\in\mathcal{S}}|f_N(x)-f(x)|\longrightarrow 0,\quad$ a.s.

This now simply follows from the previous Lemma, with $\varepsilon(W)\equiv 1$.

(ii) $\sup_{x\in\mathcal{S}}||b_N(x)||\longrightarrow 0,\quad$ a.s.

This follows very closely the previous method. By Assumptions 5 and 6, we have

$$i\zeta_l K(\zeta)=(2\pi)^{-d}\int\nabla_l\phi(\xi)\,e^{-i\xi'\zeta}d\xi'.$$

Letting $b_{Nl}(x)$ be the l^{th} element of $b_N(x)$,

$$\begin{aligned}
ib_{Nl}(x) &= i\frac{1}{Nh^d}\sum_{j=1}^{N}K\left(\frac{x-X_j}{h}\right)\left(\frac{x_l-X_{jl}}{h}\right)\\
&= \frac{1}{Nh^d}\sum_{j=1}^{N}\left\{\frac{1}{(2\pi)^d}\int_{-\infty}^{\infty}\nabla_l\phi(\xi)\exp\left(-i\xi'\left(\frac{x-X_j}{h}\right)\right)d\xi'\right\}.\\
&= \frac{1}{(2\pi)^d}\int\left\{\frac{1}{N}\sum_{j=1}^{N}\exp\left(i\zeta'X_j\right)\nabla_l\phi(\zeta h)\,e^{-i\zeta' x}d\zeta'\right\}\\
&= \frac{1}{(2\pi)^d}\int\varphi_N(\zeta)e^{-i\zeta' x}\nabla_l\phi(\zeta h)\,d\zeta',
\end{aligned}$$

which is similar to (11). It follows that, since $|b_{Nl}(x)-Eb_{Nl}(x)|\equiv|i\,(b_{Nl}(x)-Eb_{Nl}(x))|\,,$[3] we can write

$$\begin{aligned}
\left\{\sup_x|b_{Nl}(x)-Eb_{Nl}(x)|\right\}^2 &\leq \frac{1}{(2\pi)^{2d}}\left\{\int|\varphi_N(\zeta)-\varphi(\zeta)|\,|\nabla_l\phi(\zeta h)|\,d\zeta'\right\}^2\\
&\leq \frac{1}{(2\pi)^{2d}}\int|\nabla_l\phi(\zeta h)|\,d\zeta'\\
&\quad\times\int|\varphi_N(\zeta)-\varphi(\zeta)|^2\,|\nabla_l\phi(\zeta h)|\,d\zeta'\\
&= \left\{\frac{1}{(2\pi)^{2d}}\int|\nabla_l\phi(\xi)|\,d\xi'\right\}Y_N'
\end{aligned}$$

where $Y_N'=h^{-d}\int|\varphi_N(\zeta)-\varphi(\zeta)|^2\,|\nabla_l\phi(\zeta h)|\,d\zeta'\geq 0$ a.s. The method of proof then proceeds as before to show that $Y_N'\to 0$, *a.s.*. so that $\sup_x|b_{Nl}(x)-Eb_{Nl}(x)|\to 0$, *a.s.*, where we exploit the locally Lipschitz condition (Assumption 7). This implies that $\sup_x||b_N(x)-Eb_N(x)||\to 0$, *a.s.*.

[3]For any complex number $z=a+ib$, $|z|^2=z\bar{z}$, where $\bar{z}=a-ib$ is the compex conjugate of z. Thus $|z|=\left(a^2+b^2\right)^{1/2}$.

Finally,

$$\sup_x ||b_N(x)|| < \sup_x ||b_N(x) - Eb_N(x)|| + \sup_x ||Eb_N(x)||$$

where it is easy to show that $\sup_x ||Eb_N(x)|| \to 0$. This completes the proof.
(iii) $\sup_{x\in\mathcal{S}} \left|\left|M_N(x) - \mu_2^K f(x) I_d\right|\right| \longrightarrow 0, \quad$ a.s.
The same method is used, to demonstrate the result element by element. Again, by Assumptions 5 and 6, we can write

$$\zeta_l \zeta_m K(\zeta) = -(2\pi)^{-d} \int \nabla^2_{lm}\phi(\xi)\, e^{-i\xi'\zeta} d\xi'$$

so that

$$\begin{aligned}
m_{Nlm}(x) &= -\frac{1}{Nh^d}\sum_{j=1}^N K\left(\frac{x-X_j}{h}\right)\left(\frac{x_l - X_{jl}}{h}\right)\left(\frac{x_m - X_{jm}}{h}\right) \\
&= -\frac{1}{Nh^d}\sum_{j=1}^N \left\{\frac{1}{(2\pi)^d}\int_{-\infty}^{\infty} \nabla^2_{lm}\phi(\xi)\exp\left(-i\xi'\left(\frac{x-X_j}{h}\right)\right) d\xi'\right\}. \\
&= -\frac{1}{(2\pi)^d}\int\left\{\frac{1}{N}\sum_{j=1}^N \exp\left(i\zeta' X_j\right)\nabla^2_{lm}\phi(\zeta h)\, e^{-i\zeta' x} d\zeta'\right\} \\
&= -\frac{1}{(2\pi)^d}\int \varphi_N(\zeta) e^{-i\zeta' x}\nabla^2_{lm}\phi(\zeta h)\, d\zeta'.
\end{aligned}$$

Then, similar to before,

$$\begin{aligned}
\left\{\sup_x |m_{Nlm}(x) - Em_{Nlm}(x)|\right\}^2 &\leq \frac{1}{(2\pi)^{2d}}\left\{\int |\varphi_N(\zeta) - \varphi(\zeta)|\left|\nabla^2_{lm}\phi(\zeta h)\right| d\zeta'\right\}^2 \\
&\leq \frac{1}{(2\pi)^{2d}}\int\left|\nabla^2_{lm}\phi(\zeta h)\right| d\zeta' \\
&\quad \times\int |\varphi_N(\zeta) - \varphi(\zeta)|^2\left|\nabla^2_{lm}\phi(\zeta h)\right| d\zeta' \\
&= \left\{\frac{1}{(2\pi)^{2d}}\int\left|\nabla^2_{lm}\phi(\xi)\right| d\xi'\right\} Y''_N
\end{aligned}$$

where $Y''_N = h^{-d}\int |\varphi_N(\zeta) - \varphi(\zeta)|^2 \left|\nabla^2_{lm}\phi(\zeta h)\right| d\zeta' \geq 0 \quad a.s.$ The method of proof then proceeds as before to show that $Y''_N \to 0$, $a.s.$. so that $\sup_x ||m_N(x) - Em_N(x)|| \to 0$, $a.s.$
Finally,

$$\sup_x ||m_N(x)|| < \sup_x ||m_N(x) - Em_N(x)|| + \sup_x \left|\left|Em_N(x) - \mu_2^K f(x) I_d\right|\right|$$

where, as before, it is easy to show that $\sup_x \left|\left|Em_N(x) - \mu_2^K f(x) I_d\right|\right| \to 0$. This completes the proof.
(iv) $\sup_{x\in\mathcal{S}} |f_N(x)|^{-1} = O(1) \quad$ a.s.
$f(x) - \inf_{x\in\mathcal{S}} f(x) \geq 0$ and $f_N(x) - f(x) + \sup_{x\in\mathcal{S}} |f_N(x) - f(x)| \geq 0$. Therefore,

$$f_N(x) = f(x) + [f_N(x) - f(x)] \geq \inf_{x\in\mathcal{S}} f(x) - \sup_{x\in\mathcal{S}} |f_N(x) - f(x)|$$

From this it follows that

$$\sup_{x\in\mathcal{S}} \left| \frac{1}{f_N(x)} \right| \leq \frac{1}{|\inf_{x\in\mathcal{S}} f(x) - [\sup_{x\in\mathcal{S}} |f_N(x) - f(x)|]|}.$$

The term on the RHS is $O(1)$ $a.s$, since $\inf_{x\in\mathcal{S}} f(x) > \delta$ and $\sup_{x\in\mathcal{S}} |f_N(x) - f(x)| = o(1), \quad a.s.$
(v) $\sup_{x\in\mathcal{S}} \left|\left|M_N(x)^{-1}\right|\right| = O(1) \quad$ a.s.
We now that $||A||^2 = \sum_{t=1}^d \lambda_t^2$ where λ_t are the eigenvalues of the $(d \times d)$ matrix A. Then (iii) implies that

$$\sup_{x\in\mathcal{S}} \gamma'\gamma = o(1) \quad a.s.$$

where $\gamma = \gamma(x)$ is the $(d \times 1)$ vector with typical element $\gamma_t(x) = \xi_t(x) - \mu_2^K f(x)$, $t = 1, ..., d$, and $\xi_t \equiv \xi_t(x)$ is an eigenvalue of $M_N(x)$, and (of course) $\mu_2^K f(x)$ are the trivial eigenvalues of $M(x)$ (with multiplicity d). We can assume that $\xi_t > 0$ for N sufficiently large. Then since γ is finite dimensional, it must be that $\sup_x |\gamma_t(x)| = o(1)$, $a.s.$ Thus we can write

$$\begin{aligned} \xi_t(x) &= \mu_2^K f(x) + \left(\xi_t(x) - \mu_2^K f(x)\right) \\ &= \mu_2^K f(x) + \gamma_t(x) \end{aligned}$$

with $\inf_x f(x) \geq \delta > 0$ and $\sup_x |\gamma_t(x)| = o(1)$, $a.s.$ Thus

$$\xi_t(x) \geq \mu_2^K \delta - \sup_x |\gamma_t(x)|$$

so that

$$\sup_x \left| \frac{1}{\xi_t(x)} \right| \leq \frac{1}{|\mu_2^K \delta - \{\sup_x |\gamma_t(x)|\}|} = O(1), \quad a.s.$$

because $\sup_x |\gamma_t(x)| = o(1)$, $as.$ Finally $\left|\left|M_N(x)^{-1}\right|\right| = \sum_{t=1}^d \xi_t^{-2}$. Thus

$$\sup_x \left|\left|M_N(x)^{-1}\right|\right| = \sup_x \left| \sum_{t=1}^d \xi_t^{-2} \right| \leq \sup_x \sum_{t=1}^d \left| \frac{1}{\xi_t(x)} \right|^2 = O(1), \quad a.s.$$

(vi) $\sup_{x\in\mathcal{S}} |\Delta_N(x) - f(x)| \longrightarrow 0$, a.s.
This follows immediately from (i), (ii) and (v).
(vii) $\sup_{x\in\mathcal{S}} |\Delta_N(x)|^{-1} = O(1)$ a.s.
This follows in a similar vein to (iv), since $\Delta_N(x) = f(x) + [\Delta_N(x) - f(x)] \geq \inf_{x\in\mathcal{S}} f(x) - \sup_{x\in\mathcal{S}} |\Delta_N(x) - f(x)|$, and the result follows.
This completes the proof of the Lemma 4.

Proof of Lemma 7.

From Lemma 1 that $\Sigma_{1N}^{Q}(\varepsilon) = \Sigma + o_p(1)$ and $\Sigma_{2N}^{Q}(\varepsilon) = \Sigma + o_p(1)$. Furthermore, similar to Zheng (1996, Lemma 3.3e) it is easy to show that

$$\begin{aligned} \Sigma_{1N}^{Q}(\hat{u}) &= \Sigma_{1N}(\varepsilon) + o_p(1) \\ \Sigma_{2N}^{Q}(\hat{u}) &= \Sigma_{2N}(\varepsilon) + o_p(1). \end{aligned}$$

However, both these estimators are not operational since they depend upon $f(x)$ which is unknown.
We shall consider a rather general case, such as the statistic

$$\Omega_{1N}(\hat{u}) = \frac{2h^d}{N(N-1)} \sum_i \sum_{j\neq i} \hat{u}_i^2 a_N(X_i) K_{ij}^2 \hat{u}_j^2$$

where, as before, $\sup_x |a_N(x)| = o(1)$, a.s.. Then

$$\begin{aligned} |\Omega_{1N}(\hat{u})| &\leq \sup_x |a_N(x)| \frac{2h^d}{N(N-1)} \sum_i \sum_{j\neq i} \hat{u}_i^2 K_{ij}^2 \hat{u}_j^2 \\ &= \sup_x |a_N(x)| \frac{2h^d}{N(N-1)} \sum_i \sum_{j\neq i} \varepsilon_i^2 K_{ij}^2 \varepsilon_j^2 + o_p(1) \end{aligned}$$

and $\dfrac{2h^d}{N(N-1)} \sum_i \sum_{j\neq i} \varepsilon_i^2 K_{ij}^2 \varepsilon_j^2 = O_p(1)$ so that $|\Omega_{1N}(\hat{u})| = o_p(1)$. Using this result, it is straightforward to show that

$$\left| \Sigma_{1N}^{P}(\hat{u}) - \Sigma_{1N}^{Q}(\hat{u}) \right| = o_p(1),$$

using

$$\begin{aligned} |a_N(X_i)| &= \left| \frac{1}{f_N^2(X_i)} - \frac{1}{f^2(X_i)} \right| \\ &= \frac{\left| f^2(X_i) - f_N^2(X_i) \right|}{f_N^2(X_i) f^2(X_i)}. \end{aligned}$$

Similarly

$$\begin{aligned}\Sigma_{2N}^{P}(\hat{u})-\Sigma_{2N}^{Q}(\hat{u}) &= \frac{h^{d}}{N(N-1)}\sum_{i}\sum_{j\neq i}\hat{u}_{i}^{2}a_{N}(X_{i})K_{ij}^{2}\hat{u}_{j}^{2}\\ &\quad+\frac{h^{d}}{N(N-1)}\sum_{i}\sum_{j\neq i}\hat{u}_{i}^{2}K_{ij}^{2}c_{N}(X_{i},X_{j})\hat{u}_{j}^{2}\end{aligned}$$

where

$$\begin{aligned}c_{N}(X_{i},X_{j}) &= \frac{(f_{N}(X_{i})-f(X_{i}))f_{N}(X_{j})}{f_{N}(X_{i})f_{N}(X_{j})f(X_{i})f(X_{j})}-\frac{(f_{N}(X_{j})-f(X_{j}))f(X_{i})}{f_{N}(X_{i})f_{N}(X_{j})f(X_{i})f(X_{j})}\\ &= \frac{(f_{N}(X_{i})-f(X_{i}))}{f_{N}(X_{i})f(X_{i})f(X_{j})}-\frac{(f_{N}(X_{j})-f(X_{j}))}{f_{N}(X_{i})f_{N}(X_{j})f(X_{j})}\\ &= \frac{c_{1N}(X_{i})}{f_{N}(X_{i})f(X_{i})f(X_{j})}-\frac{c_{1N}(X_{j})}{f_{N}(X_{i})f_{N}(X_{j})f(X_{j})}.\end{aligned}$$

Now, $\sup_{x}|c_{1N}(x)|=o(1)$, *a.s.*, and $\sup_{x}\{f_{N}(x)\}^{-1}=O(1)$, a.s., so that $\sup_{x,y}|c_{N}(x,y)|=o(1)$, a.s. Then following previous arguments, it follows that $\left|\Sigma_{2N}^{P}(\hat{u})-\Sigma_{2N}^{Q}(\hat{u})\right|=o_{p}(1)$.
Now consider $\Sigma_{1N}^{R}(\hat{u})-\Sigma_{1N}^{P}(\hat{u})$:

$$\left|\Sigma_{1N}^{R}(\hat{u})-\Sigma_{1N}^{P}(\hat{u})\right|\leq\frac{2h^{d}}{N(N-1)}\sum_{i}\sum_{j\neq i}\hat{u}_{i}^{2}\left|R_{ij}^{2}-\frac{K_{ij}^{2}}{f_{N}^{2}(X_{i})}\right|\hat{u}_{j}^{2}$$

and $\left|R_{ij}^{2}-\frac{K_{ij}^{2}}{f_{N}^{2}(X_{i})}\right|=\left|R_{ij}-\frac{K_{ij}}{f_{N}(X_{i})}\right|\left|R_{ij}+\frac{K_{ij}}{f_{N}(X_{i})}\right|$. Now

$$\left|R_{ij}-\frac{K_{ij}}{f_{N}(X_{i})}\right|\leq|a_{1N}(X_{i})|\,K_{ij}+||a_{2N}(X_{i})||\,K_{ij}\left|\left|\zeta_{ij}\right|\right|$$

and

$$\left|R_{ij}+\frac{K_{ij}}{f_{N}(X_{i})}\right|\leq|a_{1N}(X_{i})|\,K_{ij}+||a_{2N}(X_{i})||\,K_{ij}\left|\left|\zeta_{ij}\right|\right|$$

where $a_{1N}(X_{i})=\frac{b_{N}(X_{i})'\left(M_{N}(X_{i})\right)^{-1}b_{N}(X_{i})}{\Delta_{N}(X_{i})f_{N}(X_{i})}$ and $a_{2N}(X_{i})=\frac{\left(M_{N}(X_{i})\right)^{-1}b_{N}(X_{i})}{\Delta_{N}(X_{i})}$, with $\sup_{x}|a_{1N}(x)|=o(1)$, *a.s.*, and $\sup_{x}|a_{2N}(x)|=o(1)$, *a.s.* Therefore

$$\left|\Sigma_{1N}^{R}(\hat{u})-\Sigma_{1N}^{P}(\hat{u})\right|\leq\frac{2h^{d}}{N(N-1)}\sum_{i}\sum_{j\neq i}\hat{u}_{i}^{2}D_{ij}^{2}\hat{u}_{j}^{2}$$

where

$D_{ij}^2 = |a_{1N}(X_i)|^2 K_{ij}^2 + ||a_{2N}(X_i)||^2 K_{ij}^2 \left|\left|\zeta_{ij}\right|\right|^2 + 2\,|a_{1N}(X_i)|\,||a_{2N}(X_i)||\,K_{ij}^2 \left|\left|\zeta_{ij}\right|\right|.$

Therefore,

$$
\begin{aligned}
\left|\Sigma_{1N}^R(\hat{u}) - \Sigma_{1N}^P(\hat{u})\right| \leq\ & \sup_x |a_{1N}(x)|^2 \frac{2h^d}{N(N-1)} \sum_i \sum_{j\neq i} \hat{u}_i^2 K_{ij}^2 \hat{u}_j^2 \\
& + \sup_x |a_{2N}(x)|^2 \frac{2h^d}{N(N-1)} \sum_i \sum_{j\neq i} \hat{u}_i^2 K_{ij}^2 \left|\left|\zeta_{ij}\right|\right|^2 \hat{u}_j^2 \\
& + \sup_x |a_{1N}(x)| \sup_x |a_{2N}(x)| \frac{4h^d}{N(N-1)} \sum_i \sum_{j\neq i} \hat{u}_i^2 K_{ij}^2 \left|\left|\zeta_{ij}\right|\right| \hat{u}_j^2.
\end{aligned}
$$

We know that $\frac{2h^d}{N(N-1)} \sum_i \sum_{j\neq i} \hat{u}_i^2 K_{ij}^2 \hat{u}_j^2 = O_p(1)$, and similarly it can be shown that $\frac{2h^d}{N(N-1)} \sum_i \sum_{j\neq i} \hat{u}_i^2 K_{ij}^2 \left|\left|\zeta_{ij}\right|\right|^2 \hat{u}_j^2 = O_p(1)$ and $\frac{4h^d}{N(N-1)} \sum_i \sum_{j\neq i} \hat{u}_i^2 K_{ij}^2 \left|\left|\zeta_{ij}\right|\right| \hat{u}_j^2 = O_p(1)$. Thus $\left|\Sigma_{1N}^R(\hat{u}) - \Sigma_{1N}^P(\hat{u})\right| = o_p(1)$.

Finally, slightly more tedious, but in a similar vein it is shown that $|\Sigma_{6N}(\hat{u}) - \Sigma_{4N}(\hat{u})| = o_p(1)$. Note that

$$
\begin{aligned}
\left|\Sigma_{2N}^R(\hat{u}) - \Sigma_{2N}^P(\hat{u})\right| \leq\ & |\Sigma_{5N}(\hat{u}) - \Sigma_{3N}(\hat{u})| \\
& + \frac{h^d}{N(N-1)} \sum_i \sum_{j\neq i} \hat{u}_i^2 \,|R_{ij}R_{ji} - P_{ij}P_{ji}|\, \hat{u}_j^2,
\end{aligned}
$$

in which the first term is zero. The second term is

$$
\begin{aligned}
\frac{h^d}{N(N-1)} \sum_i \sum_{j\neq i} \hat{u}_i^2 \,|R_{ij}R_{ji} - P_{ij}P_{ji}|\, \hat{u}_j^2 \leq\ & \frac{h^d}{N(N-1)} \sum_i \sum_{j\neq i} \hat{u}_i^2 \,|R_{ij}|\,|R_{ji} - P_{ji}|\, \hat{u}_j^2 \\
& + \frac{h^d}{N(N-1)} \sum_i \sum_{j\neq i} \hat{u}_i^2 \,|P_{ji}|\,|R_{ij} - P_{ij}|\, \hat{u}_j^2.
\end{aligned}
$$

Now,

$$
\begin{aligned}
|R_{ji} - P_{ji}| &\leq |a_{1N}(X_j)|\,K_{ij} + ||a_{2N}(X_j)||\,K_{ij} \left|\left|\zeta_{ij}\right|\right| \\
|R_{ij}| &\leq \frac{K_{ij}}{\Delta_N(X_i)} + ||a_{2N}(X_i)||\,K_{ij} \left|\left|\zeta_{ij}\right|\right|.
\end{aligned}
$$

Thus. since $\sup_x |a_{lN}(x)| = 0$ a.s., $l = 1, 2$, and $\sup_x \Delta_N(x) = O(1)$ a.s., it follows that

$$
\frac{h^d}{N(N-1)} \sum_i \sum_{j\neq i} \hat{u}_i^2 \,|R_{ij}|\,|R_{ji} - P_{ji}|\, \hat{u}_j^2 = o(1) \quad \text{a.s.}
$$

Similarly, $\dfrac{h^d}{N(N-1)} \sum_i \sum_{j \neq i} \hat{u}_i^2 \, |P_{ji}| \, |R_{ij} - P_{ij}| \, \hat{u}_j^2 = o(1)$, a.s.

References

[1] Chesher, A. and Spady, R. (1991) *Asymptotic Expansions of the Information Matrix Test Statistic.* Econometrica, 59, 787-815.

[2] Davidson, R. and Flachaire, E. (2000) *The Wild Bootstrap: Tamed at last.* Working Paper, GREQAM.

[3] Ellison, G and Ellison, S. (2000) *A simple framework for nonparametric specification testing.* Journal of Econometrics, 96, 1-23.

[4] Fan, Y. and Li, Q. (1996). *Consistent Model Specification Test: Omitted Variables and Semiparametric Functional Forms.* Econometrica, 64, 865-890.

[5] Fan, Y. and Li, Q. (2000). *Consistent Model Specification Tests.*Econometric Theory, 16, 1016-1041.

[6] Fan, Y. and Li, Q. (2002) . *A Consistent Model Specification Test Based on the Kernel Sum of Squares of Residuals.* Econometric Reviews, 21, 337-352.

[7] Fan, J. (1992). *Design-adaptive Nonparametric Regression.* Journal of the American Statistical Association, 87, 998-1004.

[8] Fan, J. (1993). *Local Linear Regression Smoothers and Their Minimax Efficiencies.* Annals of Statistics, 21, 196-216.

[9] Gozalo, P. (1993) *A Consistent Model Specification Test for Nonparametric Estimation of Regression Function Models.* Econometric Theory, 9, 451-477.

[10] Godfrey, L.G. and Orme, C.D. (2001) *On improving the robustness and reliability of Rao's score test.* Journal of Statistical Planning and Inference, 97, 153-176.

[11] Godfrey, L.G. and Orme, C.D. (2002) *Significance Levels of Heteroskedasticity-robust Tests for Specification and Misspecification: some results on the use of Wild Bootstraps.* Working Paper, University of York.

[12] Godfrey, L.G. and Orme, C.D. (2004) *Controlling the Finite Sample Significance Levels of Heteroskedasticity-robust Tests of Several Linear Restrictions on Regression Coefficients.* Economics Letters, 82, 281-287.

[13] Hardle, W. and Mammen, E. (1993). *Comparing Nonparametric Versus Parametric Regression Fits.* The Annals of Statistics,.21, 1926-1947

[14] Hausman, J. (1978) *Specification Tests in Econometrics.* Econometrica, 46, 1251-1272.

[15] Hsiao, C, and Li, Q. (2001) *A Consistent Test for Conditional Heteroskedasticity in Time-Series Regression Models*. Econometric Theory, 17, 1, 188-221

[16] Horowitz (1994) *Bootstrap-based Critical Values for the Information Matrix Test.* Journal of Econometrics, 61, 395-411.

[17] Lavergne, P. and Vuong, Q. (2000) *Nonparametric Significance Testing.* Econometric Theory, 16, 576-601.

[18] Li, Q, and Wang, S. (1998) *A Simple Consistent Bootstrap Test for a Parametric Regression Function.* Journal of Econometrics, 87, 145-165.

[19] Newey, W. (1985) *Maximum Likelihood Specification Testing and Conditional Moment Tests.* Econometrica, 53, 1047-1070.

[20] Orme, C.D. (1991) *The Small Sample Performance of the Information Matrix.* Journal of Econometrics, 46, 309-331.

[21] Racine, J. and Li, Q. (2004) *Nonparametric Estimation of Regression Functions with both Categorical and Continuous Data*. Journal of Econometrics, 119, 99-130.

[22] Robinson, P.M. (1988) *Root-N-Consistent Semiparametric Regression.* Econometrica, 56, 931-954.

[23] Ruud, P.A. (1984) *Tests of Specification in Econometrics.* Econometric Reviews, 3, 211-242.

[24] Stone, C.J. (1977) *Consistent Nonparametric Regression.* The Annals of Statistics, 5, 595-620.

[25] Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing.* Monographs on Statistics and Applied Probability: Chapman and Hall.

[26] Wang, Y-J. (1998) *Consistent Bootstrap Tests of Parametric Regression Functions.* Journal of Econometrics, 98, 27-46.

[27] White, H. (1982) *Maximum Likelihood Estimation of Misspecified Models.* Econometrica, 50, 1-26.

[28] Zheng, J.X. (1996). *A Consistent Test of Functional Form by Nonparametric Estimation Techniques.* Journal of Econometrics, 76, 263-289

[29] Zheng, J. X. (1998). *Consistent Specification Testing for Conditional Symmetry*. Econometric Theory, 14, 139-149..