A Consistent Model Specification Test Based on the Kernel Sum of Squares of Residuals *

Yanqin Fan
Department of Economics, University of Windsor
Windsor, Ontario, N9B 3P4 Canada

Qi Li
Department of Economics, Texas A&M University
College Station, TX 77843, U.S.A.

Abstract

This paper constructs a consistent model specification test based on the difference between the nonparametric kernel sum of squares of residuals and the sum of squares of residuals from a parametric null model. We establish the asymptotic normality of the proposed test statistic under the null hypothesis of correct parametric specification and show that the wild bootstrap method can be used to approximate the null distribution of the test statistic. Results from a small simulation study are reported to examine the finite sample performance of the proposed tests.

Keywords: Consistent test; Kernel method; Sum of squares of residuals; Asymptotic normality; Wild bootstrap; Simulation.

*The research of both authors is supported by grants from the Social Sciences and Humanities Research Council of Canada and the Natural Sciences and Engineering Research Council of Canada. Li's research is also supported by Ontario Premier's Research Excellence Awards, Texas A&M University Private Enterprise Research Center, and Bush program in Economics of Public Policy.

1 Introduction

Currently, there is a large literature on consistent model specification testing using nonparametric estimation methods. The idea was first proposed by Ullah (1985) in which he suggested that one may construct a consistent test for a parametric regression function based on the difference between the nonparametric kernel sum of squares of residuals and the sum of squares of residuals from the parametric null model.[1] This is quite natural given that the most popular parametric model specification test is the F-test based on restricted and unrestricted sums of squares of residuals and the kernel method is one of the most popular nonparametric estimation methods. However existing tests based on comparing nonparametric with parametric sums of squares of residuals are not completely satisfactory. For example, Lee (1994) attempted to establish the asymptotic null distribution of the test statistic proposed in Ullah (1985). However, to avoid the degeneracy of the distribution of the test statistic under the null, Lee (1994) used re-weighting, the consequence of which is the requirement that the errors must be conditionally homoskedastic; Yatchew (1992) also developed tests for a parametric regression function based on comparing sums of squares of residuals. To avoid the degeneracy of the null distribution of his test statistic, Yatchew (1992) introduced sample splitting. Despite the lack of a satisfactory theoretical justification for the test originally proposed in Ullah (1985), the test is still being used due to its intuitive appeal, see Lee and Ullah (2000).

Recently many authors have constructed consistent model specification tests that do not rely on either sample splitting or the assumption of conditionally homoskedastic errors, see Ait-Sahalia, Bickel and Stoker (1994), Delgado and Stengos (1994), Eubank and Spiegelman (1990), Fan and Li (1999), Hong and White (1995), Horowitz and Härdle (1994), Lavergne and Vuong (1996), Lewbel (1995), Robinson (1989, 1991), Wooldridge (1992), and Zheng (1996), among others. However, none of these tests is based on the kernel sum of squares of residuals originally proposed by Ullah (1985). Given that this is the first nonparametric

[1]Bierens (1982) was the first to construct a consistent model specification test. His test does not use any nonparametric estimation method, see Bierens (1990), De Jong (1996), Bierens and Ploberger (1997), Andrews (1997), and Chen and Fan (1999). For power comparisons between Bierens' test and some nonparametric kernel tests, see Fan and Li (2000). See also the recent monograph of Hart (1997) on consistent model specification tests.

test proposed in the literature and is a natural extension of the parametric F-test, we feel it is worthwhile providing a theoretical justification for it. This is the purpose of the current paper.[2]

The remainder of this paper is organized as follows. In section 2, we establish the asymptotic null distribution of the difference between the kernel sum of squares of residuals and the parametric sum of squares of residuals. Based on this, we construct a consistent test for a parametric regression model, which does not rely on either sample splitting or the assumption of conditionally homoskedastic errors. We also show the validity of a wild bootstrap procedure for approximating the asymptotic null distribution of the proposed test statistic. A small simulation experiment is carried out in section 3 to study the efficacy of the proposed test and its bootstrap version. All the technical proofs are postponed to the Appendix.

2 The Test Statistic and Its Asymptotic Distribution

We consider the following nonparametric regression model

$$y_i = m(x_i) + u_i, \qquad (i = 1, ..., n) \tag{1}$$

where $x_i \in R^d$, $m(.)$ is a smooth unknown function, and u_i is the error satisfying $E(u_i|x_i) = 0$ and $E(u_i^2|x_i) = \sigma^2(x_i)$. We are interested in testing the null hypothesis of a correct parametric specification for $m(x)$, i.e.,

$$H_0: \ \text{Prob}[m(x_i) = g(x_i, \beta_0)] = 1 \text{ for some } \beta_0 \in \mathcal{B} \ , \tag{2}$$

where $g(.,.)$ is a known function with β_0 being a $p \times 1$ unknown parameter, and $\mathcal{B}$ is a compact set in R^p. The alternative hypothesis is the negation of H_0, i.e.,

$$H_1: \ \text{Prob}[m(x_i) = g(x_i, \beta)] < 1 \text{ for all } \beta \in \mathcal{B} \ . \tag{3}$$

[2]Recently, Horowitz and Spokoiny (2000) propose an adaptive and rate-optimal test for parametric regression functional forms. To generalize the tests considered in this paper along the lines of adaptive and rate-optimal is left for future research.

2.1 A Test Based on the Kernel Sum of Squares of Residuals

Following the suggestion of Ullah (1985), we will construct a test for H_0 against H_1 on the basis of the difference between the sum of squares of residuals from the restricted parametric null model and the kernel sum of squares of residuals from the unrestricted nonparametric regression model.

Let $\hat{\beta}$ denote a $\sqrt{n}$-consistent estimator of β_0 under H_0. For instance, one can take $\hat{\beta}$ as the nonlinear least squares estimator of β_0 based on the null model. The ith residual from the parametric model is $\tilde{u}_{i,p} = y_i - g(x_i, \hat{\beta})$. Then the sum of squares of residuals from the restricted parametric model is

$$SSR_P = \frac{1}{n} \sum_i \tilde{u}_{i,p}^2. \tag{4}$$

By using standard arguments based on the Taylor series expansion, one can easily show that under condition (C1) given in the Appendix and the fact that $\hat{\beta} - \beta = O_p(n^{-1/2})$ under H_0, the following is true:

$$SSR_P = \frac{1}{n} \sum_i u_i^2 + O_p(n^{-1}). \tag{5}$$

For the unrestricted model, we estimate $E(y_i|x_i) \equiv m(x_i)$ by the following leave-one-out kernel estimator,

$$\hat{y}_i \equiv \hat{E}(y_i|x_i) = \frac{1}{nh^d} \sum_{j \neq i} y_j K(\frac{x_i - x_j}{h}) / \hat{f}(x_i), \tag{6}$$

where $K(.)$ is the kernel function, $h = h_n$ is the smoothing parameter, and

$$\hat{f}(x_i) = \frac{1}{nh^d} \sum_{j \neq i} K(\frac{x_i - x_j}{h}), \tag{7}$$

is the leave-one-out kernel estimator of $f(x_i)$, the density function of the regressor x_i. Therefore, the ith nonparametric residual is

$$\tilde{u}_i = y_i - \hat{y}_i, \tag{8}$$

and the nonparametric kernel sum of squares of residuals is given by

$$SSR_N = \frac{1}{n} \sum_i \tilde{u}_i^2 = \frac{1}{n} \sum_i (y_i - \hat{y}_i)^2. \tag{9}$$

In the Appendix we show that

$$SSR_N = \frac{1}{n}\sum_i u_i^2 - \mathcal{U}_n + c_1(n) + c_2(n) + o_p((nh^{d/2})^{-1}), \tag{10}$$

where

$$\mathcal{U}_n = \frac{1}{n^2h^d}\sum_i \sum_{j\neq i} u_i u_j [\bar{K}_{ij} - 2K_{ij}]/f(x_i), \tag{11}$$

is a second order degenerate U-statistic, $K_{ij} = K((x_i - x_j)/h)$, $\bar{K}_{ij} = \bar{K}((x_i - x_j)/h)$, in which $\bar{K}(v) = \int K(u)K(u+v)du$ is the two-fold convolution kernel obtained from $K(.)$,

$$c_1(n) = \frac{1}{n^3h^{2d}}\sum_i \sum_{j\neq i}(m(x_i) - m(x_j))^2 K_{ij}^2/\hat{f}^2(x_i), \tag{12}$$

and

$$c_2(n) = \frac{1}{n^3h^{2d}}\sum_i \sum_{j\neq i} u_i^2 K_{ij}^2/\hat{f}^2(x_i). \tag{13}$$

By lemma A.7 we know that $nh^{d/2}\mathcal{U}_n \to N(0, \sigma_0^2)$ in distribution, where $\sigma_0^2 = 2E[\sigma^4(x_i)/f(x_i)]\{\int[\bar{K}(u) - 2K(u)]^2 du\}$. Thus, (5), (10) and lemma A.7 lead to the following Theorem.

THEOREM 2.1 *Under (C1) and (C2) given in the Appendix, we have under H_0,*

$$nh^{d/2}[SSR_P - SSR_N + c_1(n) + c_2(n)] \to N(0, \sigma_0^2) \text{ in distribution.}$$

Theorem 2.1 suggests that one need to provide consistent estimators for $c_1(n)$, $c_2(n)$ and σ_0^2 in order to construct a test statistic based on $(SSR_P - SSR_N)$. The following theorem serves this purpose.

THEOREM 2.2 *Define $\hat{c}_1(n) = (n^3h^{2d})^{-1}\sum_i\sum_{j\neq i}[g(x_i,\hat{\beta}) - g(x_j,\hat{\beta})]^2 K_{ij}^2/\hat{f}_i^2$, $\hat{c}_2(n) = (n^3h^{2d})^{-1}\sum\sum_{j\neq i}\hat{u}_{i,p}^2 K_{ij}^2/\hat{f}_i^2$ and $\hat{\sigma}_0^2 = 2(n^2h^d)^{-1}\sum_i\sum_{j\neq i}\hat{u}_{i,p}^2\hat{u}_{j,p}^2(\bar{K}_{ij} - 2K_{ij})^2/\hat{f}_i^2$, where $\hat{f}_i = \hat{f}(x_i)$. Then under conditions (C1) and (C2) given in the Appendix, we have under H_0*

$$T_n \overset{def}{=} nh^{d/2}[SSR_P - SSR_N + \hat{c}_1(n) + \hat{c}_2(n)]/\hat{\sigma}_0 \to N(0,1) \text{ in distribution.}$$

Remark It is easy to show that $c_1(n) = O_p(h^2(nh^d)^{-1})$. Therefore, $nh^{d/2}c_1(n) = O_p(h^{2-d/2})$, which is $o_p(1)$ if $d \leq 3$ (d is the dimension of x_i). Thus, one can drop $c_1(n)$ and $\hat{c}_1(n)$ in Theorem 2.1 and Theorem 2.2 when $d \leq 3$.

The next Theorem establishes the consistency of the T_n test.

THEOREM 2.3 *Under (C1) and (C2), we have under* H_1,

$$Prob[T_n > c] \to 1 \text{ for any positive constant } c > 0.$$

2.2 A Bootstrap Test

It is well known that nonparametric kernel-based tests suffer from poor size performance for moderately large samples typically encountered in practice, see Härdle and Mammen (1993), Li and Wang (1998), and Lee and Ullah (2000), among others for evidence on this. In this subsection we show that the wild bootstrap can be used to approximate the null distribution of the T_n test.

Let u_i^* be the wild-bootstrap error obtained from the two point distribution based on the nonparametric residual $\tilde{u}_i$, i.e., $u_i^* = a\tilde{u}_i$ with probability $r = (\sqrt{5}+1)/(2\sqrt{5})$ and $u_i^* = b\tilde{u}_i$ with probability $1-r$, where $a = (1-\sqrt{5})/2$ and $b = (1+\sqrt{5})/2$. With $\{u_i^*\}_{i=1}^n$, the bootstrap test statistic is obtained by the following steps.

Step (i): Generate y_i^* according to the parametric *null* model: $y_i^* = g_0(x_i, \hat{\beta}) + u_i^*$, $i = 1, ..., n$, and then use the bootstrap sample $\{x_i, y_i^*\}_{i=1}^n$ to estimate β based on the parametric null model. Denote the resulting estimator as $\hat{\beta}^*$. Then obtain the parametric bootstrap residual based on the null model: $\hat{u}_{i,p}^* = y_i^* - g(x_i, \hat{\beta}^*)$, $i = 1, ..., n$.

Step (ii): Obtain the nonparametric bootstrap residual: $\tilde{u}_i^* = y_i^* - \hat{y}_i^*$, $i = 1, ..., n$, where $\hat{y}_i^* = \sum_{j \neq i} y_j^* K_{ij} / \sum_{j \neq i} K_{ij}$ is the kernel estimator of $E(y_i|x_i)$ using the bootstrap sample $\{x_i, y_i^*\}_{i=1}^n$.

Step (iii): Compute the bootstrap test statistic

$$T_n^* = nh^{d/2}[SSR_P^* - SSR_N^* + \hat{c}_1^*(n) + \hat{c}_2^*(n)]/\hat{\sigma}_0^*,$$

where SSR_N^*, SSR_P^*, $\hat{c}_j^*(n)$ ($j = 1, 2$) and $\hat{\sigma}_0^{2*}$ are obtained in the same way as SSR_N, SSR_P, $\hat{c}_j(n)$ ($j = 1, 2$) and $\hat{\sigma}_0^2$ except that $\tilde{u}_i$ and $\hat{u}_{i,p}$ are replaced by $\tilde{u}_i^*$ and $\hat{u}_{i,p}^*$, respectively.

The next Theorem shows that the wild bootstrap method works.

THEOREM 2.4 *Let $\mathcal{Z}_n = \{x_i, y_i\}_{i=1}^n$. Under conditions (C1) and (C2), we have*

$$T_n^* | \mathcal{Z}_n \to N(0,1) \text{ in distribution.}$$

The proof of Theorem 2.4 is basically the same as that of Theorem 2.2. We only provide a sketchy proof in the Appendix. Note that whether H_0 is true or not, our bootstrap sample is always generated according to the null model. Therefore, the distribution of the bootstrap statistic T_n^* will mimic the null distribution of T_n even when the null hypothesis is false.

2.3 A Density-Weighted Test

Condition (C1) requires that the density function $f(.)$ be bounded below by a positive constant, which rules out any regressor whose distribution has an unbounded support such as normal regressors. To relax this restrictive assumption, one can either introduce a trimming parameter to trim out the small values of $\hat{f}(x_i)$, or construct a density weighted test statistic. The latter is much easier to handle technically and also has the advantage of avoiding another (trimming) nuisance parameter. Below we will construct a density weighted test statistic and derive its asymptotic distribution.

Define a density-weighted kernel sum of squares of residuals by $DSSR_N = n^{-1} \sum_i \tilde{u}_i^2 \hat{f}_i^2$ and a density weighted parametric sum of squares of residuals by $DSSR_P = n^{-1} \sum_i \hat{u}_{i,p}^2 \hat{f}_i^2$.

THEOREM 2.5 *Define $\hat{c}_{1,d}(n) = (n^3 h^{2d})^{-1} \sum_i \sum_{j \neq i} [\hat{g}(x_i, \hat{\beta}) - g(x_j, \hat{\beta})]^2 K_{ij}^2$, $\hat{c}_{2,d}(n) = (n^3 h^{2d})^{-1} \sum_i \sum_{j \neq i} \tilde{u}_{i,p}^2 K_{ij}^2$, and $\hat{\sigma}_{0,d}^2 = 2(n^2 h^d)^{-1} \sum_i \sum_{j \neq i} \hat{u}_{i,p}^2 \hat{u}_{j,p}^2 \hat{f}_i \hat{f}_j [\bar{K}_{ij} - 2K_{ij}]^2$. Then under (C1) and (C2) except the condition of $inf_{x \in \mathcal{S}} f(x) \geq \delta > 0$ in (C1), we have*

(i) $T_{n,d} = nh^{d/2}[DSSR_P - DSSR_N + \hat{c}_{1,d}(n) + \hat{c}_{2,d}(n)]/\hat{\sigma}_{0,d} \to N(0,1)$ in distribution under H_0, and

(ii) $Prob[T_{n,d} > c] \to 1$ for any positive constant $c > 0$ under H_1.

The proof of Theorem 2.5 is similar to the proofs of Theorems 2.2 and 2.3 and a sketchy proof is given in the Appendix.

One can also use the wild bootstrap method to approximate the null distribution of $T_{n,d}$. The bootstrap statistic is given by

$$T^*_{n,d} = nh^{d/2}[DSSR^*_P - DSSR^*_N + \hat{c}^*_{1,d}(n) + \hat{c}^*_{2,d}(n)]/\hat{\sigma}^*_{0,d},$$

where $DSSR^*_P$, $DSSR^*_N$, $\hat{c}^*_{1,d}(n)$, $\hat{c}^*_{2,d}(n)$ and $\hat{\sigma}^*_{0,d}$ are defined in the same way as $DSSR_P$, $DSSR_N$, $\hat{c}_{1,d}(n)$, $\hat{c}_{2,d}(n)$ and $\hat{\sigma}_{0,d}$ except that $\tilde{u}_i$ and $\hat{u}_{i,p}$ are replaced by $\tilde{u}^*_i$ and $\hat{u}^*_{i,p}$, respectively, $\tilde{u}^*_i$ and $\hat{u}^*_{i,p}$ are given in section 2.2. Similar to the proof of Theorem 2.5, one can show that $T^*_{n,d} \to N(0,1)$ whether H_0 is true or not. Hence, $T^*_{n,d}$ can be used to approximate the null distribution of $T_{n,d}$.

2.4 The Dependent Data Case

Like most of the existing tests in the literature, our tests are established for independent observations. The main tool being used to obtain the asymptotic null distributions of the test statistics in this area is Hall's (1984) central limit theorem for degenerate U-statistics. Recently Fan and Li (1999) have extended Hall's (1984) result to weakly dependent data case. Using Theorem 3.1 of Fan and Li (1999), one should be able to establish the validity of the tests in section 2 for the weakly dependent data case and hence provide some theoretical justification for the finite sample results in Lee and Ullah (2000). However, a detailed proof is quite tedious and thus omitted.

3 Monte Carlo Simulations

In this section we report results from a small Monte Carlo experiment to examine the finite sample performance of our proposed T_n test and its bootstrap version. We will only study the size performance of our tests, for the power performance, see Lee and Ullah (2000). We use the same data generating process (DGP) as considered in Härdle and Mammen (1993):

$$y_i = 2x_i - x_i^2 + u_i, \tag{14}$$

where x_i is uniform [0,1] and u_i is i.i.d. $N(0, \sigma^2)$ with $\sigma = 0.1$. Li and Wang (1998) also used the same DGP to examine the size performance of an alternative nonparametric test (their J_n test, see also Zheng (1996)). We use a standard normal (second order) kernel function for

$K(.)$. Hence, the convolution kernel $\bar{K}(.)$ is a normal density with zero mean and variance 2. Condition (C2) (ii) in the Appendix requires that $h \to 0$ and $nh^{4.5} \to \infty$ as $n \to \infty$, which rules out the optimal smoothing. We choose the smoothing parameter via $h = \sigma_x n^{-1/3}$, where σ_x is the sample standard deviation of $\{x_i\}_{i=1}^n$. For comparison we also computed the test statistic suggested by Li and Wang (1998), and Zheng (1996) (denoted as J_n test). The J_n test is given by

$$J_n = nh^{d/2}[\frac{1}{n^2h^d}\sum_i\sum_{j\neq i}\hat{u}_{i,p}\hat{u}_{j,p}K_{ij}/\hat{\sigma}_1],$$

where $\hat{\sigma}_1^2 = 2(n^2h^d)^{-1}\sum_i\sum_{j\neq i}\hat{u}_{i,p}^2\hat{u}_{j,p}^2K_{ij}^2$. It is known that $J_n \to N(0,1)$ under H_0. Note that the J_n test has a simpler structure compared with the T_n test because J_n only involves two summations. Also, the conditions on h are $h \to 0$ and $nh^d \to \infty$ so that the J_n test allows optimal smoothing parameter being used. We also use a standard normal kernel for the J_n test, and the smoothing parameter is chosen via $h = \sigma_x n^{-1/5}$ for the J_n test.

Table 1 gives the estimated sizes of the T_n and the J_n tests based on the asymptotic standard normal critical values, where the number of replications is 2,000 for all cases. From Table 1 we observe that both tests are conservative. The T_n test suffers from more bias than the J_n test, while the J_n test has an standard error significantly less than 1 even for $n = 1,000$. Overall, the bias of both tests become smaller and the standard deviations get closer to 1, as sample size gets larger, as suggested by the asymptotic theories. However extremely large sample sizes may be needed in order for the asymptotic theories to work satisfactorily.

Next we turn to the performances of the bootstrap tests. Table 2 reports the estimated sizes based on wild bootstrap statistics. The number of replications for the bootstrap tests is 1,000 and within each replication, 399 bootstrap statistics are generated to yield bootstrap critical values. From Table 2 we can see that the bootstrap tests perform much better than their asymptotic counterparts. The J_n test seems to have slightly better estimated sizes than the T_n test, probably due to the fact that J_n test has a simple structure and it allows a wide range of values for h (including optimal smoothing).

Table 1: Estimated sizes based on asymptotic critical values

	T_n test					J_n test				
	Mean	Std	1%	5%	10%	Mean	Std	1%	5%	10%
n= 50	-1.587	1.373	.001	.014	.027	-0.845	.568	.001	.002	.004
n=100	-1.328	1.171	.004	.014	.025	-0.768	.630	.001	.004	.014
n=200	-1.207	1.032	.004	.012	.020	-0.726	.681	.002	.011	.018
n=500	-1.030	0.992	.003	.012	.022	-0.697	.687	.002	.008	.016
n=1000	-0.944	0.998	.004	.004	.029	-0.678	.712	.004	.130	.046

Table 2: Estimated sizes based on bootstrap critical values

	T_n test				J_n test			
	1%	5%	10%	20%	1%	5%	10%	20%
n = 50	0.012	0.057	0.118	0.237	0.012	0.053	0.109	0.216
n = 100	0.013	0.069	0.135	0.248	0.008	0.048	0.094	0.210
n = 200	0.009	0.043	0.101	0.219	0.009	0.051	0.101	0.207

APPENDIX: Proofs of the Main Results

The following definition (Robinson (1988)) will be used.

Definition 1. $\mathcal{G}_\mu^\alpha$, $\alpha > 0$, $\mu > 0$, is a class of functions g: $\mathcal{R}^d \to \mathcal{R}$ satisfying: g is μ -times partially differentiable; for some $\rho > 0$, $sup_{y \in \phi_{z\rho}} |g(y) - g(z) - Q_g(y,z)|/|y-z| \le G_g(z)$ for all z, where $\phi_{z\rho} = \{y : |y-z| < \rho\}$; $Q_g = 0$ when $\nu = 1$; Q_g is a $(\nu - 1)$ th degree homogeneous polynomial in $y - z$ with the coefficients the partial derivatives of order $\mu - 1$ and less, and $G_g(z)$ has finite αth moments.

The following conditions are used to derive the asymptotic distributions of the test statistics considered in this paper.

C1 (y_i, x_i), $i = 1, ..., n$, are independent and identically distributed as (Y, X), X has a convex compact support $\mathcal{S}$. Let $f(.)$ denote the density function of X. $f(x)$ is bounded above and $inf_{x \in \mathcal{S}} f(x) \ge \delta > 0$; under H_0, $\hat{\beta} - \beta = O_p(n^{-1/2})$, $\nabla g(X,.)$ and $\nabla^2 g(X,.)$ are continuous in X and dominated by functions with finite second moments, where $\nabla g(X,.)$ and $\nabla^2 g(X,.)$ are the $d \times 1$ vector of first order partial derivatives and $d \times d$ matrix of second order partial derivatives of m with respect to β respectively. Y has finite fourth moment; $m(x)$, $f(x)$ and $\sigma^2(x) = E[u_i^2 | x_i = x]$ and $\mu_4(x) = E[u_i^4 | x_i = x]$ all belong to $\mathcal{G}_\nu^4$ for some positive integer $\nu \geq 2$.

C2 (i) We use product kernel, in which the univariate kernel function k: $\mathcal{R} \rightarrow \mathcal{R}$ is an even and bounded νth order kernel function, i.e, $\int k(v)v^l dv = 0$ for $l = 1,...,\nu-1$ and $\int k(v)v^l dv \neq 0$; (ii) as $n \rightarrow \infty$, $h \rightarrow 0$, $nh^d \rightarrow \infty$ and $nh^{2\nu+d/2} \rightarrow \infty$.

Define $\hat{m}_i = (nh^d)^{-1}\sum_{j\neq i} m(x_j)K_{ij}$ and $\hat{u}_i = (nh^d)^{-1}\sum_{j\neq i} u_j K_{ij}$. Then from (6) we have $\hat{y}_i = \hat{m}_i + \hat{u}_i$, and

$$\begin{aligned} SSR_N &= \frac{1}{n}\sum_i [u_i + (m_i - \hat{m}_i) - \hat{u}_i]^2 = \frac{1}{n}\sum_i u_i^2 + \frac{1}{n}\sum_i (m_i - \hat{m}_i)^2 + \frac{1}{n}\sum_i \hat{u}_i^2 \\ &\quad + \frac{2}{n}\sum_i u_i(m_i - \hat{m}_i) - \frac{2}{n}\sum_i \hat{u}_i(m_i - \hat{m}_i) - \frac{2}{n}\sum_i u_i\hat{u}_i \\ &\equiv A_{1n} + A_{2n} + A_{3n} + 2A_{4n} - 2A_{5n} - 2A_{6n}, \end{aligned} \tag{A.1}$$

where the definitions of A_{jn} $(j = 1, ..., 6)$ should be apparent from (A.1).

Lemma A.0 $E[(m(x_2) - m(x_1))|x_1] \leq G_m(x_1)h^{d+\nu}$,
where $G_m(X)$ has finite fourth moments.

Proof: It is similar to the proof of Lemma 5 in Robinson (1988).

Lemma A.1 *(i) $sup_{x\in\mathcal{S}}|\hat{f}(x) - f(x)| = o(1)$ almost surely (a.s), (ii) $sup_{x\in\mathcal{S}}|1/\hat{f}(x)| = O(1)$ a.s.*

Proof: For (i), see Györfi, et al. (1989).

(ii) $\hat{f}(x) = f(x) + (\hat{f}(x) - f(x)) \leq inf_{x\in\mathcal{S}}f(x) - sup_{x\in\mathcal{S}}|\hat{f}(x) - f(x)|$. Therefore, $sup_{x\in\mathcal{S}}|1/\hat{f}(x)| \leq 1/|\,inf_{s\in\mathcal{S}}f(x) - [sup_{x\in\mathcal{S}}|\hat{f}(x) - f(x)|]\,| = O(1)$ a.s. because $inf_{x\in\mathcal{S}}f(x) \geq \delta > 0$ by (C1) and $sup_{x\in\mathcal{S}}|\hat{f}(x) - f(x)| = o(1)$ a.s. by (i) above.

Lemma A.2 $A_{2n} = c_1(n) + o_p((nh^{d/2})^{-1})$, *where $c_1(n)$ is defined in (12).*

Proof: $A_{2n} = n^{-1}\sum_i (m_i - \hat{m}_i)^2 = n^{-1}\sum_i [(m_i - \hat{m}_i)^2 \hat{f}_i^2/\hat{f}_i^2] = (n^3h^{2d})^{-1}\sum_i\sum_{j\neq i}\sum_{l\neq i}(m_i - m_j)K_{ij}(m_i - m_l)K_{il}/\hat{f}_i^2 = (n^3h^{2d})^{-1}\sum\sum_{j\neq i}(m_i - m_j)^2K_{ij}^2/\hat{f}_i^2 + (n^3h^{2d})^{-1}\sum\sum\sum_{i\neq j\neq l}(m_i - m_j)K_{ij}(m_i - m_l)K_{il}/\hat{f}_i^2 \equiv c_1(n) + A_{2n,2}$.

It is easy to show that the leading term of $A_{2n,2}(n)$ is $(n^3h^{2d})^{-1}\sum\sum\sum_{i\neq j\neq l}(m_i - m_j)K_{ij}(m_i - m_l)K_{il}/f_i^2 = (n^3h^{2d})^{-1}O_p(n^3h^{2(d+\nu)}) = O_p(h^{2\nu}) = o_p((nh^{d/2})^{-1})$ by lemma A.0.

Hence, $A_{2n} = c_1(n) + o_p((nh^{d/2})^{-1})$.

Lemma A.3 $A_{3n} = \mathcal{U}_{1n} + c_2(n) + o_p((nh^{d/2})^{-1})$,
where $\mathcal{U}_{1n} = (n^2h^d)^{-1}\sum_i\sum_{j\neq i} u_iu_j\bar{K}_{ij}/f_i$ *with* $\bar{K}_{ij} = \int K((x_i-x_j)/h+v)K(v)dv$, *and* $c_2(n) = (n^3h^{2d})^{-1}\sum_i\sum_{j\neq i} u_j^2K_{ij}^2/\hat{f}_i^2$.

Proof: $A_{3n} = (n^3h^{2d})^{-1}\sum_i\sum_{j\neq i}\sum_{l\neq i} u_ju_lK_{ij}K_{il}/\hat{f}_i^2 = (n^3h^{2d})^{-1}\sum\sum\sum_{i\neq j\neq l} u_ju_lK_{ij}K_{il}/\hat{f}_i^2 + (n^3h^{2d})^{-1}\sum\sum_{i\neq j} u_j^2K_{ij}^2/\hat{f}_i^2 = (n^3h^{2d})^{-1}\sum_i\sum_{j\neq i}\sum_{l\neq i} u_ju_lK_{ij}K_{il}/\hat{f}_i^2 + c_2(n) \equiv A_{3n,1} + c_2(n)$, where $c_2(n)$ is defined in (13). We write $A_{3n,1} = (n^3h^{2d})^{-1}\sum\sum\sum_{i\neq j\neq l} u_ju_lK_{ij}K_{il}/f_i^2 + (n^3h^{2d})^{-1}\sum\sum\sum_{i\neq j\neq l} u_ju_lK_{ij}K_{il}[(\hat{f}_i)^{-2} - (\hat{f}_i)^{-2}] \equiv B_{3n,1} + B_{3n,2}$.

By H-decomposition (e.g., Lee (1990)), the leading term of $B_{3n,1}$ is a second order degenerate U-statistic of the form of $(n^2h^{2d})^{-1}\sum\sum_{l\neq j} u_ju_lE[K_{ij}K_{ill}/f_i^2|x_j, x_l]$.

Note that $E[K_{ij}K_{il}/f_i^2|x_j, x_l] = \int[f(x_i)K((x_i-x_j)/h)K((x_i-x_l)/h)/f_i^2]dx_i = h^d\int[K(v)K((x_j-x_l)/h + v)/f(x_j + uh)]dv = (h^d/f(x_j))\bar{K}((x_j - x_l)/h) + O(h^{d+1})$, where $\bar{K}(v) = \int K(u + v)K(u)du$ is the two-fold convolution kernel obtained from $K(.)$. Therefore, we have

$B_{3n,1} = (n^2h^d)^{-1}\sum_j\sum_{l\neq j} u_ju_l\bar{K}_{jl}/f_j + (s.o) = (n^2h^d)^{-1}\sum_i\sum_{j\neq i} u_iu_j\bar{K}_{ij}/f_i + (s.o) \equiv \mathcal{U}_{1n} + (s.o.)$, where $\mathcal{U}_{1n}$ is defined in lemma A.3 above and the (s.o.) term means it has an order smaller than $\mathcal{U}_{1n}$.

Obviously, $B_{3n,1} = O_p((nh^{d/2})^{-1})$. By lemma A.1, it is easy to show that $B_{3n,2}$ has an order smaller than $B_{3n,1}$, that is $B_{3n,2} = o_p((nh^{d/2})^{-1})$. Thus we have,

$A_{3n} = \mathcal{U}_{1n} + c_2(n) + o_p((nh^{d/2})^{-1})$.

Lemma A.4 $A_{4n} = o_p((nh^{d/2})^{-1})$.

Proof: $A_{4n} = O_p(h(nh^{d/2})^{-1} + n^{-1/2}h^{\nu}) = o_p((nh^{d/2})^{-1})$ by lemma 3 of Li (1996).

Lemma A.5 $A_{5n} = o_p((nh^{d/2})^{-1})$.

Proof: $A_{5n} = n^{-1}\sum_i \hat{u}_i(m_i-\hat{m}_i) \equiv n^{-1}\sum_i \hat{u}_i\hat{f}_i(m_i-\hat{m}_i)\hat{f}_i/\hat{f}_i^2 = (n^3h^{2d})^{-1}\sum_i\sum_{j\neq i}\sum_{l\neq i} u_jK_{ij}(m_i-m_l)K_{il}/\hat{f}_i^2 = (n^3h^{2d})^{-1}\sum_i\sum_{j\neq i} u_j(m_i-m_j)K_{ij}^2/\hat{f}_i^2 + (n^3h^{2d})^{-1}\sum\sum\sum_{i\neq j\neq l} u_j(m_i-m_l)K_{ij}K_{il}/\hat{f}_i^2 \equiv A_{5n,1} + A_{5n,2}$,

By lemma A.0 it is easy to see that the second moment of $A_{5n,1}$ is $(n^6h^{4d})^{-1}O(n^3h^2h^{2d} + n^2h^d) = o(n^2h^d)$, and that the second moment of $A_{5n,2}$ is $(n^6h^{4d})O(n^6h^{4(d+\nu)} + n^4h^{3(d+\nu)} + n^2h^{2(d+\nu)}) = O(h^{2\nu} + (n^2h^d)^{-1}h^{2\nu}) = o(n^2h^d)$.

Therefore, $A_{5n} = o_p((nh^{d/2})^{-1})$.

Lemma A.6 $A_{6n} = \mathcal{U}_{2n} + o_p((nh^{d/2})^{-1})$, *where* $\mathcal{U}_{2n} = (n^2h^d)^{-1}\sum_i\sum_{j\neq i} u_iu_jK_{ij}/f_i$.

Proof: It suffices to show that $B_{6n} \stackrel{def}{=} A_{6n} - \mathcal{U}_{2n} = o_p((nh^{d/2})^{-1})$.

Now, $B_{6n} = A_{6n} - \mathcal{U}_{1n} = 2(n^2h^d)^{-1}\sum_i\sum_{j>i} u_iu_jK_{ij}[(1/\hat{f}_i) - (1/f_i)]$. The second moment of B_{6n} is $E[B_{6n}^2] = 4(n^2h^d)^{-1}\sum_i\sum_{j>i} E\{u_i^2u_j^2K_{ij}^2[(1/\hat{f}_i) - (1/f_i)]^2\} \leq E[\{sup_{x\in\mathcal{S}}|[\hat{f}(x) - f(x)]/(f(x)\hat{f}(x))|\}^2 4(n^4h^{2d})^{-1}\sum_i\sum_{j>i} u_i^2u_j^2K_{ij}^2] = o(1)(n^4h^{2d})^{-1}\sum_i\sum_{j>i} E[u_i^2u_j^2K_{ij}^2]$ $= o(1)O((n^2h^d)^{-1}) = o((n^2h^d)^{-1})$ by lemma A.1, which implies that $B_{6n} = o_p((nh^{d/2})^{-1})$.

Lemma A.7 *$nh^{d/2}\mathcal{U}_n \to N(0,\sigma_0^2)$ in distribution, where $\mathcal{U}_n = \mathcal{U}_{1n} - 2\mathcal{U}_{2n}$, $\mathcal{U}_{1n}$ and $\mathcal{U}_{2n}$ are defined in lemmas A.1 and A.3, respectively, and σ_0^2 is defined above Theorem 2.1.*

Proof: $nh^{d/2}\mathcal{U}_n = (n^2h^d)\sum_i\sum_{j\neq i} u_iu_j(\bar{K}_{ij} - 2K_{ij})/f_i$ is a degenerate U-statistic and it is easy to show that its second moment is $E\{[nh^{d/2}\mathcal{U}_n]^2\} = 4(n^2h^d)^{-1}\sum_i\sum_{j>i} E[u_i^2u_j^2(\bar{K}_{ij} - 2K_{ij})^2/f_i^2] = 2E[\sigma^4(x_i)/f(x_i)]\{\int[K(u) - \bar{K}(u)]^2du\} + o(1) = \sigma_0^2 + o(1)$. Finally it is straightforward to check that the conditions of Hall's (1984) Theorem 1 holds. Therefore by Hall's central limit theorem for degenerate U-statistic, we know that $nh^{d/2}\mathcal{U}_n \to N(0,\sigma_0^2)$ in distribution.

Proof of Theorem 2.1

Proof: Using (5), (8) and lemmas A.2 to A.6, we have

$SSR_P - SSR_N + c_1(n) + c_2(n) = \mathcal{U}_n + o_p((nh^{d/2})^{-1})$.

Therefore,

$nh^{d/2}[SSR_P - SSR_N + c_1(n) + c_2(n)] = nh^{d/2}\mathcal{U}_n + o_p(1) \to N(0,\sigma_0^2)$ in distribution by lemma A.7.

Proof of Theorem 2.2

By noting that $\hat{u}_{i,p} - u_i = O_p(n^{-1/2})$ and $c_j(n) = O_p((nh^d)^{-1})$ $(j = 1,2)$, it is easy to see that $\hat{c}_j(n) - c_j(n) = O_p(n^{-1/2})O_p((nh^d)^{-1}) = o_p((nh^{d/2})^{-1})$. Hence,

$T_n = nh^{d/2}[SSR_P - SSR_N + \hat{c}_1(n) + \hat{c}_2(n)] = nh^{d/2}\{\mathcal{U}_n + [\hat{c}_1(n) - c_1(n)] + [\hat{c}_2(n) - c_2(n)] + o_p((nh^{d/2})^{-1})\} = nh^{d/2}\mathcal{U}_n + o_p(1) \to N(0,\sigma_0^2)$ in distribution by lemma A.7.

Finally it is easy to show that $\hat{\sigma}_0^2 = \sigma_0^2 + o_p(1)$. Hence,

$nh^{d/2}[SSR_P - SSR_N + \hat{c}_1(n) + \hat{c}_2(n)]/\hat{\sigma}_0 \to (1/\sigma_0)N(0,\sigma_0^2) = N(0,1)$ in distribution under H_0.

Proof of Theorem 2.3

Let $\bar{\beta}$ denote the probability limit of $\hat{\beta}$ under H_1. We have $\hat{u}_{i,p} = y_i - g(x_i, \hat{\beta}) = u_i + m(x_i) - g(x_i, \bar{\beta}) + o_p(1) \equiv u_i + \eta_i + o_p(1)$, where $\eta_i = m(x_i) - g(x_i, \bar{\beta})$. Denote $\sigma_u^2 = E(u_i^2)$ and $\sigma_\eta^2 = E(\eta_i^2)$. Then

$$\begin{aligned} SSR_p &= n^{-1} \sum_i (u_i + \eta_i)^2 + o_p(1) \\ &= n^{-1} \sum_i u_i^2 + n^{-1} \sum_i \eta_i + o_p(1) = \sigma_u^2 + \sigma_\eta^2 + o_p(1). \end{aligned}$$

It is easy to show that, under H_1, $SSR_N = n^{-1} \sum_i u_i^2 + o_p(1) = \sigma_u^2 + o_p(1)$, $\hat{c}_j(n) = O_p((nh^d)^{-1}) = o_p(1)$ $(j = 1, 2\ )$. Hence,

$nh^{d/2}[SSR_P - SSR_N + \hat{c}_1(n) + \hat{c}_2(n)] = nh^{d/2}[\sigma_\eta^2 + o_p(1)] \to \infty$ at the rate of $nh^{d/2}$ since $\sigma_\eta^2 > 0$ under H_1. This completes the proof of Theorem 2.3.

Proof of Theorem 2.4

First using $\hat{\beta}^* - \hat{\beta} = O_p(n^{-1/2})$, it is easy to show that

$$SSR_P^* = \frac{1}{n} \sum_i u_i^{*2} + O_p(n^{-1}). \tag{A.2}$$

Next, define $\hat{m}_{i,p} = g(x_i, \hat{\beta})$, $\hat{m}_i^* = (nh^d)^{-1} \sum_{j \neq i} \hat{m}_{j,p} K_{ij}$ and $\hat{u}_i^* = (nh^d)^{-1} \sum_{j \neq i} u_j^* K_{ij}$. Then we have $\hat{y}_i^* \stackrel{def}{=} \sum_{j \neq i} y_j^* K_{ij} / \sum_{j \neq i} K_{ij} = \hat{m}_i^* + \hat{u}_i^*$, and

$$\begin{aligned} SSR_N^* &= \frac{1}{n} \sum_i [u_i^* + (\hat{m}_{i,p} - \hat{m}_i^*) - \hat{u}_i^*]^2 = \frac{1}{n} \sum_i u_i^{*2} + \frac{1}{n} \sum_i (\hat{m}_{i,p} - \hat{m}_i^*)^2 + \frac{1}{n} \sum_i \hat{u}_i^{*2} \\ &\quad + \frac{2}{n} \sum_i u_i^* (\hat{m}_{i,p} - \hat{m}_i^*) - \frac{2}{n} \sum_i \hat{u}_i^* (\hat{m}_{i,p} - \hat{m}_i^*) - \frac{2}{n} \sum_i u_i^* \hat{u}_i^* \\ &\equiv A_{1n}^* + A_{2n}^* + A_{3n}^* + 2A_{4n}^* - 2A_{5n}^* - 2A_{6n}^*, \end{aligned} \tag{A.3}$$

where the definitions of A_{jn}^* $(j = 1, ..., 6)$ should be apparent.

Following exactly the same arguments as in the proof of Theorem 2.1, one can show that

$$nh^{d/2}[SSR_P^* - SSR_N^* + c_1^*(n) + c_2^*(n)] | \mathcal{Z}_n = nh^{d/2} \mathcal{U}_n^* + o_p(1) \to N(0, \hat{\sigma}_0^2) \text{ in distribution,}$$

where $c_1^*(n)$, $c_2^*(n)$ and $\mathcal{U}_n^*$ are defined the same way as $c_1(n)$, $c_2(n)$ and $\mathcal{U}_n$ with u_i and $m(x_i)$ replaced by u_i^* and $\hat{m}_{i,p} = g(x_i, \hat{\beta})$, respectively. Hence, Theorem 2.4 follows by noting that $\hat{c}_j^*(n) - c_j^*(n) = o_p((nh^{d/2})^{-1})$ $(j = 1, 2)$, $\hat{\sigma}_0^{*2} - \hat{\sigma}_0^2 = o_p(1)$.

Proof of Theorem 2.5

$$DSSR_N = \frac{1}{n} \sum_i [u_i + (m_i - \hat{m}_i) - \hat{u}_i]^2 \hat{f}_i^2 = \frac{1}{n} \sum_i u_i^2 \hat{f}_i^2 + \frac{1}{n} \sum_i (m_i - \hat{m}_i)^2 \hat{f}_i^2 + \frac{1}{n} \sum_i \hat{u}_i^2 \hat{f}_i^2$$

$$+\frac{2}{n}\sum_i u_i(m_i-\hat{m}_i)\hat{f}_i^2-\frac{2}{n}\sum_i \hat{u}_i(m_i-\hat{m}_i)\hat{f}_i^2-\frac{2}{n}\sum_i u_i\hat{u}_i\hat{f}_i^2$$
$$\equiv \quad D_{1n}+D_{2n}+D_{3n}+2D_{4n}-2D_{5n}-2D_{6n}, \qquad \text{(A.4)}$$

where the definitions of D_{jn} $(j=1,...,6)$ should be apparent.

The proof is similar to that of Theorem 2.1. In particular, D_{4n} and D_{5n} are all of $o_p((nh^{d/2})^{-1})$ by the same arguments in the proof of $A_{jn}=o_p((nh^{d/2})^{-1})$ $(j=4,5)$.

By the same arguments as in the proof of lemma A.2, one can easily show that $D_{2n}=c_{1,d}(n)+o_p((nh^{d/2})^{-1})$, where $c_{1,d}(n)=(n^3h^{2d})^{-1}\sum_i\sum_{j\neq i}(m(x_i)-m(x_j))^2K_{ij}^2$.

Similar to the proof of lemma A.3, we have

$D_{3n}=n^{-1}\sum_i \hat{u}_i^2\hat{f}_i^2=(n^3h^{2d})^{-1}\sum_i\sum_{j\neq i}\sum_{l\neq i}u_ju_lK_{ij}K_{il}=(n^3h^{2d})^{-1}\sum_{i\neq j\neq l}u_ju_lK_{ij}K_{il}+(n^3h^{2d})^{-1}\sum_i\sum_{j\neq i}u_j^2K_{ij}^2=D_{3n,1}+c_{2,d}(n)$.

The leading term of $D_{3n,1}$ is a second order degenerate U-statistic. Using the standard H-decomposition, we have $E[u_ju_lK_{ij}K_{il}|z_j,z_l]=u_ju_lE[K_{ij}K_{il}|x_j,x_l]=u_ju_l\{h^d\bar{K}_{jl}f_j+o(h^d)\}$, where $\bar{K}_{jl}=\bar{K}((x_j-x_l)/h)$ and $\bar{K}(.)$ is the convolution kernel. Therefore, we get

$D_{3n,1}=(n^2h^d)^{-1}\sum_i\sum_{j\neq i}u_iu_jf_j\bar{K}_{ij}+(s.o.)\equiv\mathcal{U}_{1n,d}+(s.o.)$.

By the same arguments as in the proof of lemma A.6, one can show that $D_{6n}=(n^2h^d)^{-1}\sum_i\sum_{j\neq i}u_iu_jK_{ij}\hat{f}_i=(n^2h^d)^{-1}\sum_i\sum_{j\neq i}u_iu_jK_{ij}f_i+(s.o.)\equiv\mathcal{U}_{2n,d}+(s.o.)$.

D_{1n} cancels the leading term in $DSSR_P$, i.e., $DSSR_P-D_{1n}=O_p(n^{-1})$. Thus, we have

$$nh^{d/2}[DSSR_P-DSSR_N+c_{1,d}(n)+c_{2,d}(n)]=nh^{d/2}\mathcal{U}_{n,d}+o_p(1),$$

where $\mathcal{U}_{n,d}=(n^2h^d)^{-1}\sum_i\sum_{j\neq i}u_iu_jf_i(\bar{K}_{ij}-2K_{ij})$. The second moment of $\mathcal{U}_{n,d}$ is

$E[(\mathcal{U}_{n,d})^2]=4(n^2h^d)^{-1}\sum_i\sum_{j>i}E\{u_i^2u_j^2f_i^2(\bar{K}_{ij}-2K_{ij})^2\}=E[\sigma^2(x)f^3(x)][\int(\bar{K}(v)-2K(v))^2dv]+o(1)\equiv\sigma_{0,d}^2+o(1)$.

We estimate $\sigma_{0,d}^2$ by $\hat{\sigma}_{0,d}^2=2(n^2h^d)^{-1}\sum_i\sum_{j\neq i}\hat{u}_{i,p}^2\hat{u}_{j,p}^2\hat{f}_i\hat{f}_j(\bar{K}_{ij}-2K_{ij})^2$. Also we estimate $c_{1,d}(n)$ and $c_{2,d}(n)$ by $\hat{c}_{1,d}(n)=(n^3h^{2d})^{-1}\sum_i\sum_{j\neq i}[g(x_i,\hat{\beta})-g(x_j,\hat{\beta})]^2K_{ij}^2$ and $\hat{c}_{2,d}(n)=(n^3h^{2d})^{-1}\sum_i\sum_{j\neq i}\hat{u}_{j,p}^2K_{ij}^2$.

Obviously $\hat{\sigma}_{0,d}^2-\sigma_{0,d}^2=o_p(1)$ and $\hat{c}_{j,d}(n)-c_{j,d}(n)=O_p(n^{-1/2})O_p((nh^d)^{-1})=o_p((nh^{d/2})^{-1})$ $(j=1,2)$. Therefore, we have

$$\begin{aligned} T_{n,d} &\stackrel{def}{=} nh^{d/2}[DSSR_P-DSSR_N+\hat{c}_{1,d}(n)+\hat{c}_{2,d}(n)]/\hat{\sigma}_{0,d} \\ &= nh^{d/2}\mathcal{U}_{n,d}/\sigma_{0,d}+o_p(1)\to N(0,1) \text{ in distribution under } H_0. \end{aligned}$$

References

Ait-Sahalia, Y., P.J. Bickel and T.M. Stoker (1994), "Goodness-of-fit Tests for Regression Using Kernel Methods," manuscript, University of Chicago.

Andrews, D.W.K. (1997), "A Conditional Kolmogorov Test," *Econometrica* 65, 1097-1128.

Bierens, H. J. (1982), "Consistent Model Specification Tests," *Journal of Econometrics* 20, 105-134.

Bierens, H.J. (1990), "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58, 1443-1458.

Bierens, H.J. and W. Ploberger (1997), "Asymptotic Theory of Integrated Conditional Moment Tests," *Econometrica* 65, 1129-1151.

Chen, X. and Y. Fan (1999), "Consistent Hypothesis Testing in Semiparametric and Nonparametric Models for Econometric Time Series," *Journal of Econometrics* 91, 373-401.

De Jong, R.M. (1996), "The Bierens Test Under Data Dependence," *Journal of Econometrics* 72, 1-32.

Delgado, M.A. and T. Stengos (1994),"Semiparametric Testing of Non-nested Econometric Models," *Review of Economic Studies* 75, 345-367.

Eubank, R. and S. Spiegelman (1990), "Testing the Goodness of Fit of a Linear Model via Nonparametric Regression Techniques," *Journal of the American Statistical Association* 85, 387-392.

Fan, Y. and Q. Li (1999), "Central Limit Theorem for Degenerate U-statistics of Absolutely Regular Processes With Applications to Model Specification Tests," *Journal of Nonparametric Statistics* 10, 245-271.

Fan, Y. and Q. Li (2000), "Consistent Model Specification Tests: Bierens' Tests Versus Kernel Tests," *Econometric Theory* 16, 1016-1041.

Györfi, L., W. Härdle, P. Sarda and P. Vieu (1989), *Nonparametric Curve Estimation from Time Series*, Berlin: Spring-Verlag.

Hall, P. (1984), "Central Limit Theorem for Integrated Square Error of Multivariate Nonparametric Density Estimators," *Journal of Multivariate Analysis* 14, 1-16.

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

Hart, J.D. (1997), *Nonparametric Smoothing and Lack-of-fit Tests*. New York: Springer-Verlag.

Hong, Y. and H. White (1995), "Consistent Specification Testing via Nonparametric Series Regression," *Econometrica* 63, 1133-1159.

Horowitz, J.L. and W. Härdle (1994) ,"Testing a Parametric Model Against a Semiparametric Alternative," *Econometric Theory* 10, 821-848.

Horowitz, J.L. and V.G. Spokoiny (2000) ,"An adaptive, rate-optimal test of a parametric model against a nonparametric alternative," forthcoming in *Econometrica*.

Lavergne, P. and Q. Vuong (1996), "Nonparametric Selection of Regressors: The Nonnested Case," *Econometrica* 64, 207-219.

Lee, A.J. (1990), *U-Statistics: Theory and Practice*, New York and Basel: Marcel Dekker, Inc.

Lee, B.J. (1994), "Asymptotic Distribution of the Ullah-type Specification Test Against the Nonparametric Alternatives," *Journal of Quantitative Economics* 10, 73-92.

Lee, T. and A. Ullah (2000), "Nonparametric Bootstrap Tests for Neglected Nonlinearity in Time Series Regression Models," forthcoming in *Journal of Nonparametric Statistics*.

Lewbel, A. (1995), "Consistent Nonparametric Testing With an Application to Testing Slusky Symmetry," *Journal of Econometrics* 67, 379-401.

Li, Q. (1996), "On root-n-Consistent Estimation of Partially Linear Models," *Economics Letters* 51, 277-285.

Li, Q and S. Wang (1998), "A Simple Consistent Bootstrap Test for a Parametric Regression Functional Form," *Journal of Econometrics* 87, 145-165.

Robinson, P. (1988), "Root-N Consistent Semiparametric Regression," *Econometrica 56*, 931-954.

Robinson, P.M. (1989), "Hypothesis Testing in Semiparametric and Nonparametric Models for Econometric Time Series," *Review of Economic Studies* 56, 511-534.

Robinson, P.M. (1991), "Consistent Nonparametric Entropy-Based Testing," *Review of Economic Studies* 58, 437-453.

Ullah, A. (1985), "Specification Analysis of Econometric Models," *Journal of Quantitative Economics* 2, 187-209.

Wooldridge, J. (1992), "A Test for Functional Form Against Nonparametric Alternatives," *Econometric Theory*, 8, 452-475.

Yatchew, A. J. (1992), "Nonparametric Regression Tests Based on Least Squares," *Econometric Theory*, 8, 435-451.

Zheng, J.X. (1996), "A Consistent Test of Functional Form via Nonparametric Estimation Technique," *Journal of Econometrics* 75, 263-289.