Econometrica, Vol. 64, No.4 (July, 1996), 865–890

CONSISTENT MODEL SPECIFICATION TESTS: OMITTED VARIABLES AND SEMIPARAMETRIC FUNCTIONAL FORMS

BY YANQIN FAN AND QI LI[1]

In this paper, we develop several consistent tests in the context of a nonparametric regression model. These include tests for the significance of a subset of regressors and tests for the specification of the semiparametric functional form of the regression function, where the latter covers tests for a partially linear and a single index specification against a general nonparametric alternative. One common feature to the construction of all these tests is the use of the Central Limit Theorem for degenerate U-statistics of order higher than two. As a result, they share the same advantages over most of the corresponding existing tests in the literature: (a) They do not depend on any ad hoc modifications such as sample splitting, random weighting, etc. (b) Under the alternative hypotheses, the test statistics in this paper diverge to positive infinity at a faster rate than those based on ad hoc modifications.

KEYWORDS: Consistent tests, degenerate U-statistics, kernel estimation, omitted variables, partially linear model, single index model.

1. INTRODUCTION

RECENTLY, NONPARAMETRIC FUNCTIONAL ESTIMATION TECHNIQUES such as kernel and series methods have been used to construct consistent model specification tests.[2] These include tests for a parametric model versus a nonparametric model, tests for the significance of a subset of regressors in a nonparametric regression model, and tests for a semiparametric (partially linear or single index) model against a nonparametric alternative. For example, Fan and Li (1992a), Härdle and Mammen (1993), Hidalgo (1992), and Lee (1994) have developed consistent tests for a parametric specification by using the kernel regression estimation technique; Eubank and Spiegelman (1990), Hong and White (1995), and Wooldridge (1992) have applied the method of series estimation to consistent testing for a parametric regression model; consistent tests for omitted variables were considered by Hidalgo (1992) and Gozalo (1993), among others; Lavergne and Vuong (1996) proposed a method to select between two sets of regressors using kernel estimators; Whang and Andrews (1993) and Yatchew (1992) have developed consistent tests for a partially linear model versus a

[1] The first author would like to thank the Natural Sciences and Engineering Research Council and the Social Sciences and Humanities Research Council of Canada for their support and the second author would like to acknowledge the Social Sciences and Humanities Research Council of Canada for its support.

[2] Bierens (1982) was the first to give a consistent conditional moment model specification test; see also Bierens (1990), Bierens and Ploberger (1994), and references therein. Using nonparametric estimation technique to construct consistent model specification tests was first suggested by Ullah (1985). Robinson (1989) was the first to propose some nonparametric tests for time-series models.

nonparametric alternative; consistent tests for a single index specification have been presented in Chen (1992) and Rodriguez and Stoker (1992).

The first group of papers that make use of nonparametric estimation techniques in developing consistent tests for a parametric functional form such as Hidalgo (1992), Lee (1994), and Wooldridge (1992) among others have employed ad hoc modifications such as sample splitting or a form of weighting. Recently a number of authors have successfully used the Central Limit Theorems (CLTs) for second order degenerate U-statistics in Hall (1984) and De Jong (1987) to develop some consistent nonparametric tests for a parametric density function or a parametric regression function; see, e.g., Eubank and Hart (1992), Fan (1994), Fan and Li (1992a, b), Härdle and Mammon (1993), Hong and White (1995), Horowitz and Härdle (1994), and Li (1994), to mention only a few. However to the best of our knowledge, the existing consistent tests for a semiparametric functional form such as a partially linear model or a single index model and tests for omitted variables in a nonparametric model still employ ad hoc modifications. For example, for testing a partially linear model versus a nonparametric regression model, Whang and Andrews (1993) and Yatchew (1992) used sample splitting; for testing a single index model, Chen (1992) used sample splitting, and Rodriguez and Stoker (1992) introduced a test statistic that has a degenerate limiting distribution under the null hypothesis; for testing omitted variables, Robinson (1991) discussed the application of a form of nonstochastic weighting which is equivalent to a form of sample splitting, Hidalgo (1992) introduced random weighting, and Gozalo (1993) used a random search procedure.[3] These ad hoc modifications are introduced in the aforementioned studies in order to overcome the so called "degeneracy problem": an appropriate estimator of some measure of the distance between the models to be tested under the null hypothesis approaches zero at a rate faster than $n^{-1/2}$, where n is the sample size. As a consequence, when normalized by $n^{1/2}$, the estimator of the chosen measure does not have a well-defined limiting distribution under the null hypothesis. This degeneracy problem is caused by the fact that the estimator of the chosen measure contains in its expression some degenerate U-statistic which vanishes at a rate faster than $n^{-1/2}$. Without modifying this estimator, its asymptotic distribution under the null hypothesis would be determined by the degenerate U-statistic. In the papers we just cited, ad hoc modifications are employed such that the asymptotic distribution of the estimator of the chosen measure in each of these papers is determined by a random term that is of a larger order than the corresponding U-statistic. However, rather than introducing ad hoc methods to avoid the degeneracy problem, it seems natural, as in recent papers on consistent tests for a parametric functional form mentioned earlier, to exploit this special property by invoking the CLTs for degenerate U-statistics (of order possibly higher than two) to

[3] Recently Lavergne and Vuong (1996) constructed a nonparametric test for selecting regressors. They require that the competing sets of regressors be non-nested. Hence their test is different from the usual omitted variables test.

develop consistent tests for omitted variables and for semiparametric functional forms. This forms the core of the present paper. Tests that exploit this special property may also be more powerful than those based on ad hoc modifications,[4] because under the alternative hypothesis the corresponding test statistics diverge to $+\infty$ at a rate faster than $n^{1/2}$.

The remainder of this paper is organized as follows. We introduce the nonparametric regression model and the hypotheses to be tested in Section 2. Section 3 constructs a consistent test for omitted variables in the context of the nonparametric regression model introduced in Section 2. Section 4 presents respectively a consistent test for a partially linear specification and a consistent test for a single index specification of the regression function. The last section concludes and offers some suggestions for further research. Appendix A contains proofs of the main results in Sections 3 and 4. Appendix B contains some technical lemmas that are used in the proofs of Appendix A. In particular, it contains an extension of the CLT in Hall (1984) for degenerate U-statistics of second order to degenerate U-statistics of any finite order. This is useful because the test statistics to be constructed in this paper involve degenerate U-statistics of order higher than two.

Throughout the rest of this paper, all the limits are taken as $n \to \infty$. $\sum_i = \sum_{i=1}^n, \sum_{j \neq i} = \sum_{j \neq i, j=1}^n$, etc.

2. THE MODEL AND THE HYPOTHESES

Consider the nonparametric regression model:

$$Y_i = g(X_i) + \epsilon_i, \tag{1}$$

where $\{Y_i, X_i'\}_{i=1}^n$ is a set of n independent and identically distributed (i.i.d.) observations on $\{Y, X'\}'$ with Y the scalar dependent variable and X the $d \times 1$ regressors, $g(\cdot)$: $R^d \to R$ is the true, but unknown regression function, and ϵ_i is the error satisfying $E[\epsilon_i \mid X_i] = 0$.

Two classes of hypotheses that are often tested concerning the regression model (1) are the functional form of $g(\cdot)$ and the significance of a subset of regressors X. With respect to the functional form of g, the null hypothesis specifies either a parametric regression model or a semiparametric regression model for $g(\cdot)$. As motivated in Section 1, we will focus on the null hypothesis of a semiparametric regression model in this paper.

We first present the problem of testing for the significance of a subset of regressors. Let $X_i = (W_i', Z_i')'$, where W_i is of dimension $q_1 \times 1$ and Z_i is of dimension $q_2 \times 1$, $q_1 + q_2 = d$. Then, a subset of regressors, Z_i (say), is insignificant to the explanation of Y_i given W_i if $E[Y_i \mid X_i] = E[Y_i \mid W_i]$. Insignificant regressors should be omitted from the regression model. Thus, a test of

[4] See the last section for more discussion on this issue.

significance is also called an omitted variables test. Let $r(w) = E[Y_i \mid W_i = w]$. Then an omitted variables test is that of

$$H_0^a\colon g(x) = r(w) \quad \text{a.e.} \quad \text{against the general alternative}$$

$$H_1^a\colon g(x) \neq r(w),$$

where $x = (w', z')'$ is partitioned according to that of X_i.

Since Robinson (1988) and Powell, Stock, and Stoker (1989), partially linear and single index models have attracted much attention among econometricians,[5] because on one hand, these models are not as restrictive as parametric regression models and on the other hand, they alleviate the problem of the "curse of dimensionality" associated with nonparametric models. However, they are still not free from misspecification errors. Thus, it is important to test the validity of such semiparametric models against the general nonparametric alternative. The first attempt toward developing consistent tests of these semiparametric models has been made by Whang and Andrews (1993) and Yatchew (1992) for partially linear regression models, and by Chen (1992) and Rodriguez and Stoker (1992) for single index models. As mentioned in Section 1, these tests used some kind of ad hoc modifications (such as sample splitting) to avoid direct treatment of degenerate U-statistics. Sample splitting results in inefficient estimators. It may also cause the tests to lose power both directly and indirectly. The indirect effect of sample splitting on these tests is to slow down the divergence rate of the test statistics to $+\infty$. This motivates us to develop consistent tests for a partially linear model and for a single index model that make use of the special feature of degenerate U-statistics.

For a partially linear model, the null hypothesis is

$$H_0^b\colon g(x) = z'\gamma + \theta(w) \quad \text{a.e.}$$

$$\text{for some } \gamma \in R^{q_2} \text{ and some } \theta(\cdot)\colon R^{q_1} \to R,$$

and the alternative is

$$H_1^b\colon g(x) \neq z'\gamma + \theta(w) \quad \text{for all } \gamma \in R^{q_2} \text{ and all } \theta(\cdot)\colon R^{q_1} \to R,$$

where as before, $x = (w', z')'$.

Finally for a single index model, the null hypothesis is

$$H_0^c\colon g(x) = \varphi(\alpha' x) \quad \text{a.e.} \quad \text{for some } \alpha \in R^d \text{ and some } \varphi(\cdot)\colon R \to R,$$

and the alternative is

$$H_1^c\colon g(x) \neq \varphi(\alpha' x) \quad \text{for all } \alpha \in R^d \text{ and all } \varphi(\cdot)\colon R \to R.$$

[5] See Engle, et al. (1986), Härdle and Stoker (1989), Stock (1989), and Stoker (1992) for empirical applications of partially linear and index models.

3. A CONSISTENT TEST FOR OMITTED VARIABLES

Recall from Section 2 that under the null hypothesis H_0^a: $g(X_i) = r(W_i)$, a.e., where $X_i = (W_i', Z_i')'$. Let $u_i = Y_i - r(W_i)$. Then under the null hypothesis, the regression model becomes

$$(2) \qquad Y_i = r(W_i) + u_i,$$

where $E(u_i \mid X_i) = g(X_i) - r(W_i) = 0$ under H_0^a and $E(u_i \mid X_i) \neq 0$ under H_1^a.

Observing that $E[u_i E(u_i \mid X_i)] = E\{[E(u_i \mid X_i)]^2\} \geq 0$ and the equality holds if and only if H_0^a is true, we can construct a consistent test[6] based on $E[u_i E(u_i \mid X_i)]$. If u_i and $E(u_i \mid X_i)$ were available, then we could estimate $E[u_i E(u_i \mid X_i)]$ by its sample analogue: $n^{-1} \sum_i u_i E(u_i \mid X_i)$. To get a feasible test statistic, we need to estimate u_i by the corresponding residual from (2) and $E(u_i \mid X_i)$ by an appropriate kernel estimator. To overcome the random denominator problem in the kernel estimation, we choose to estimate a density weighted version of $n^{-1} \sum_i u_i E(u_i \mid X_i)$ given by $n^{-1} \sum_i [u_i f_{w_i}] E[u_i f_{w_i} \mid X_i] f(X_i)$, where $f_{w_i} = f_w(W_i)$, $f_w(\cdot)$ is the probability density function (p.d.f.) of W_i, and $f(\cdot)$ is the p.d.f. of X_i.

Specifically, we estimate u_i by $\tilde{u}_i = (Y_i - \hat{Y}_i)$: the nonparametric residual from (2), where $\hat{Y}_i$ is a kernel estimator of $r(W_i)$ defined as

$$(3) \qquad \hat{Y}_i = \frac{[(n-1)a^{q_1}]^{-1} \sum_{j \neq i} Y_j K_{ij}^w}{\hat{f}_{w_i}},$$

and $\hat{f}_{w_i}$ is the corresponding kernel estimator of f_{w_i} given by

$$(4) \qquad \hat{f}_{w_i} = \frac{1}{(n-1)a^{q_1}} \sum_{j \neq i} K_{ij}^w,$$

where $K_{ij}^w = K^w((W_i - W_j)/a)$ with $K^w(\cdot)$ being a product kernel formed from the univariate kernel $k^w(\cdot)$ and $a \equiv a_n$ a smoothing parameter. We estimate $E[\tilde{u}_i \hat{f}_{w_i} \mid X_i] f(X_i)$ by $[(n-1)h^d]^{-1} \sum_{j \neq i} [\tilde{u}_j \hat{f}_{w_j}] K_{ij}$, where

$$K_{ij} = K\left(\frac{X_i - X_j}{h}\right) = K\left(\frac{W_i - W_j}{h}, \frac{Z_i - Z_j}{h}\right),$$

K is a product kernel with univariate kernel function $k(\cdot)$, and $h \equiv h_n$ is a smoothing parameter.

[6] This test can be regarded as a generalized conditional moment test of omitted variables, where the weight function is a nonparametric function given by $E(u_i \mid X_i = x)$. See Newey (1985) and Tauchen (1985) for the conditional moment tests of functional forms based on a finite number of parametric weight functions.

Our test statistic will be based on

$$I_n^a \stackrel{\text{def}}{=} \frac{1}{n}\sum_i \left[\tilde{u}_i \hat{f}_{w_i}\right]\left\{\frac{1}{(n-1)h^d}\sum_{j\neq i}\left[\tilde{u}_j \hat{f}_{w_j}\right]K_{ij}\right\} \tag{5}$$

$$= \frac{1}{n(n-1)h^d}\sum_i \sum_{j\neq i}\left[\tilde{u}_i \hat{f}_{w_i}\right]\left[\tilde{u}_j \hat{f}_{w_j}\right]K_{ij},$$

where

$$\tilde{u}_i = Y_i - \hat{Y}_i = (r_i - \hat{r}_i) + u_i - \hat{u}_i, \tag{6}$$

$r_i = r(W_i)$, $\hat{r}_i$ and $\hat{u}_i$ are defined in the same way as $\hat{Y}_i$ in (3) with Y_j replaced by $r(W_j)$ and u_j respectively.

To derive the asymptotic distribution of I_n^a under H_0^a, we will use the following definitions (see Robinson (1988)) and assumptions.

DEFINITION 1: $\mathscr{K}_l$, $l \geq 1$, is the class of even functions k: $R \to R$ satisfying

$$\int_R u^i k(u)\, du = \delta_{i0} \qquad (i = 0, 1, \dots, l-1),$$

$$k(u) = O\left((1 + |u|^{l+1+\epsilon})^{-1}\right), \quad \text{some } \epsilon > 0,$$

where δ_{ij} is the Kronecker's delta.

DEFINITION 2: $\mathscr{G}_\mu^\alpha$, $\alpha > 0$, $\mu > 0$, is the class of functions g: $R^d \to R$ satisfying: g is $(m-1)$-times partially differentiable, for $m - 1 \leq \mu \leq m$; for some $\rho > 0$, $\sup_{y \in \phi_{z\rho}} |g(y) - g(z) - Q_g(y,z)|/|y-z|^\mu \leq h_g(z)$ for all z, where $\phi_{z\rho} = \{y: |y - z| < \rho\}$; $Q_g = 0$ when $m = 1$; Q_g is a $(m-1)$th degree homogeneous polynomial in $y - z$ with coefficients the partial derivatives of g at z of orders 1 through $m-1$ when $m > 1$; and $g(z)$, its partial derivatives of order $m-1$ and less, and $h_g(z)$, have finite αth moments.

Assumption A

(A1) *(a) $f_w \in \mathscr{G}_\lambda^\infty$ for some $\lambda > 0$ and $r \in \mathscr{G}_\mu^4$ for some $\mu \geq 0$; (b) $k^w \in \mathscr{K}_{l+m-1}$ for integers l and m such that $l - 1 < \lambda \leq l$ and $m - 1 < \mu \leq m$; (c) $f \in \mathscr{G}_1^\infty$, $k \in \mathscr{K}_2$; (d) The error $\epsilon = Y - g(X)$ satisfies $E|\epsilon^4| < \infty$. The conditional variance function $\sigma^2(x) \stackrel{\text{def}}{=} E[\epsilon^2 | X = x]$ and $m_4(x) \stackrel{\text{def}}{=} E[\epsilon^4 | X = x]$ are continuous. In addition, $f(x)\sigma^2(x)$ and $f(x)m_4(x)$ are bounded on R^d.*

(A2) *As $n \to \infty$, $a \to 0$, $h \to 0$, $na^{q_1} \to \infty$, $nh^d \to \infty$, $na^{2\eta}h^{d/2} \to 0$, and $h^d/a^{2q_1} \to 0$, where $\eta = \min(\lambda + 1, \mu)$.*

A few remarks on Assumption A are in order. Assumption **(A1)** (a), (c), (d) are some smoothness and moment conditions. They are quite standard and not restrictive. **(A1)** (b) requires that k be of the order of $(l+m-1)$ which may be a second order kernel $(l+m=3)$ or a higher order kernel $(l+m>3)$. The first four conditions in Assumption **(A2)** ensure that the kernel estimators involved are consistent. The last two conditions are introduced here to ensure that the limiting distribution of $nh^{d/2}I_n^a$ under H_0^a is centered correctly at zero. Heuristically, it ensures that the asymptotic mean square error of the kernel estimator: $\tilde{u}_i\hat{f}_{w_i}$ in (6) is of smaller order than $(nh^{d/2})^{-1}$, i.e., $[a^{2\eta}+(na^{q_1})^{-1}]=o((nh^{d/2})^{-1})$. With respect to the kernel estimator $\hat{Y}_i$ of the regression function $r(W_i)$ to be tested under H_0^a, **(A2)** does not impose any restrictions on the smoothing parameter a. Specifically, the data can be over-smoothed, optimally-smoothed, or under-smoothed in estimating $r(W_i)$. However, given a, the smoothing parameter h must be chosen such that the last two conditions in **(A2)** hold. Consider for example the case where $q_2\le q_1$. In this case, the last condition in **(A2)** implies that $h/a\to 0$. Hence, we must smooth the alternative regression model less than the model under the null hypothesis.

THEOREM 3.1: *Suppose Assumptions* **(A1)** *and* **(A2)** *are satisfied. Then under H_0^a, we have $nh^{d/2}I_n^a\to N(0,2\sigma_a^2)$ in distribution, where*

$$\sigma_a^2=E\big[f(W_1,Z_1)\sigma^4(W_1,Z_1)f_{w1}^4\big]\left[\int K^2(u)\,du\right].$$

In addition, the variance σ_a^2 can be consistently estimated by $\hat{\sigma}_a^2$, where

$$\hat{\sigma}_a^2=\frac{1}{n(n-1)h^d}\sum_i\sum_{j\neq i}\left[\tilde{u}_i\hat{f}_{w_i}\right]^2\left[\tilde{u}_j\hat{f}_{w_j}\right]^2K_{ij}\left[\int K^2(u)\,du\right].$$

The estimator $\hat{\sigma}_a^2$ is obtained by noting that

$$\sigma_a^2=E\left[\{u_1f_{w_1}\}^2\left\{f(W_1,Z_1)E\left[u_1^2f_{w_1}^2\mid W_1,Z_1\right]\right\}\right]\int K^2(u)\,du,$$

where the unconditional expectation is replaced with the sample average, $u_if_{w_i}$ in the sample average by $\tilde{u}_i\hat{f}_{w_i}$, and the term inside the braces by $[(n-1)h^d]^{-1}\sum_{j\neq i}[\tilde{u}_j\hat{f}_{w_j}]^2K_{ij}$.

Define

$$T^a=\frac{nh^{d/2}I_n^a}{\sqrt{2}\,\hat{\sigma}_a}. \tag{7}$$

Then Theorem 3.1 implies that $T^a\to N(0,1)$ in distribution under H_0^a. This forms the basis for the following one-sided asymptotic test for H_0^a: reject H_0^a at significance level α_0 if $T^a > Z_{\alpha_0}$, where Z_{α_0} is the upper α_0-percentile of the standard normal distribution.

The consistency of this test is the last result of this section.

THEOREM 3.2: *Assume* **(A1)** *and* **(A2)** *hold. Then the above test is consistent. In addition* $P(T^a > M_n \mid H_1^a) \to 1$, *where* M_n *is any positive, nonstochastic sequence with* $M_n = o((nh^{d/2}))$.

Theorem 3.2 follows from the facts that under H_1^a, $I_n^a \to E\{f(X)[g(X) - r(W)]^2 f_w^2(W)\}(>0)$ in probability and $\hat{\sigma}_a = O_p(1)$. The proofs of these are straightforward and are thus omitted.

4. CONSISTENT TESTS FOR SEMIPARAMETRIC MODELS

4.1. *A Consistent Test For A Partially Linear Model*

Let $v_i = Y_i - Z_i'\gamma - \theta(W_i)$. Then $E(v_i \mid X_i) = g(X_i) - [Z_i'\gamma + \theta(W_i)]$ which equals zero a.e. if and only if the null hypothesis H_0^b is true. Hence, $E[v_i E(v_i \mid X_i)] = E\{[E(v_i \mid X_i)]^2\} \geq 0$ and the equality holds if and only if H_0^b holds. Thus, as in Section 3, we will base our test for H_0^b on an estimator of $n^{-1}\sum_i [v_i f_{w_i}] E[v_i f_{w_i} \mid X_i] f(X_i)$ to overcome the random denominator problem in kernel estimation.

The estimator of $v_i f_{w_i}$ that we adopt is obtained by a two-step procedure as in Robinson (1988) and Fan, Li, and Stengos (1995). In the first step, we estimate γ by $\hat{\gamma}$ defined as (see Fan, Li, and Stengos (1995) for more details)

$$\hat{\gamma} = S_{(Z-\hat{Z})\hat{f}_w}^{-1} S_{(Z-\hat{Z})\hat{f}_w, (Y-\hat{Y})\hat{f}_w}, \tag{8}$$

where as in Robinson (1988), $S_{A\hat{f}_w, B\hat{f}_w} = n^{-1}\sum_i A_i \hat{f}_{w_i} B_i' \hat{f}_{w_i}$ and $S_{A\hat{f}_w} = S_{A\hat{f}_w, A\hat{f}_w}$ for scalar or column vectors $A_i \hat{f}_{wi}$ and $B_i \hat{f}_{wi}$. In addition, $\hat{Z}_i = [(n-1)a^{q_1}]^{-1}\sum_{j \neq i} Z_j K_{ij}^w / \hat{f}_{w_i}$ estimates $E[Z_i \mid W_i]$, $\hat{f}_{w_i}$ and $\hat{Y}_i$ are defined in (3) and (4) respectively. Let $\bar{v}_i = (Y_i - \hat{Y}_i) - (Z_i - \hat{Z}_i)'\hat{\gamma}$. Then, the density-weighted error $v_i f_{w_i}$ is estimated by

$$\begin{aligned} \bar{v}_i \hat{f}_{w_i} &= (Y_i - \hat{Y}_i)\hat{f}_{w_i} - (Z_i - \hat{Z}_i)'\gamma \hat{f}_{w_i} - (Z_i - \hat{Z}_i)'(\hat{\gamma} - \gamma)\hat{f}_{w_i} \\ &= \left[(\theta_i - \hat{\theta}_i) + v_i - \hat{v}_i\right]\hat{f}_{w_i} - (Z_i - \hat{Z}_i)'(\hat{\gamma} - \gamma)\hat{f}_{w_i} \\ &= \tilde{v}_i \hat{f}_{w_i} - (Z_i - \hat{Z}_i)'(\hat{\gamma} - \gamma)\hat{f}_{w_i}, \end{aligned} \tag{9}$$

where $\theta_i = \theta(W_i)$, $\hat{\theta}_i$ and $\hat{v}_i$ are defined as in (3) with Y_j replaced by $\theta(W_j)$ and v_j respectively, and $\tilde{v}_i \overset{\text{def}}{=} (\theta_i - \hat{\theta}_i) + v_i - \hat{v}_i$.

Hence our test statistic will be based on

$$(10)\qquad I_n^b \stackrel{\text{def}}{=} \frac{1}{n}\sum_i \left[\bar{v}_i \hat{f}_{w_i}\right]\left\{\frac{1}{(n-1)h^d}\sum_{j\neq i}\left[\bar{v}_j \hat{f}_{w_j}\right]K_{ij}\right\}$$

$$= \frac{1}{n(n-1)h^d}\sum_i\sum_{j\neq i}\left[\bar{v}_i \hat{f}_{w_i}\right]\left[\bar{v}_j \hat{f}_{w_j}\right]K_{ij}$$

$$\doteq \frac{1}{n(n-1)h^d}\sum_i\sum_{j\neq i}\left[\tilde{v}_i \hat{f}_{w_i}\tilde{v}_j \hat{f}_{w_j}\right]K_{ij} - \frac{2}{n(n-1)h^d}$$

$$\times\sum_i\sum_{j\neq i}\left[\tilde{v}_i \hat{f}_{w_i}\left(Z_j-\hat{Z}_j\right)'\hat{f}_{w_j}(\hat{\gamma}-\gamma)\right]K_{ij}$$

$$+\frac{1}{n(n-1)h^d}\sum_i\sum_{j\neq i}\left[\left(Z_i-\hat{Z}_i\right)'\hat{f}_{w_i}\left(Z_j-\hat{Z}_j\right)\right.$$

$$\left.\times\hat{f}_{w_j}(\hat{\gamma}-\gamma)'(\hat{\gamma}-\gamma)\right]K_{ij}$$

$$\stackrel{\text{def}}{=} J_n - 2J_{1n} + J_{2n},$$

where the third equality is obtained by using (9).

It is important to note that J_n is the same as I_n^a given in (5) except that $\tilde{v}_i$ in J_n replaces $\tilde{u}_i$ in I_n^a. Thus, Theorem 3.1 can be used to obtain the asymptotic distribution of J_n. We will show in Appendix A that both J_{1n} and J_{2n} are of the order $o_p((nh^{d/2})^{-1})$. Define

$$(11)\qquad T^b = \frac{nh^{d/2}I_n^b}{\sqrt{2}\,\hat{\sigma}_b},$$

where $\hat{\sigma}_b^2$ is the same as $\hat{\sigma}_a^2$ defined in Theorem 3.1 with $\tilde{u}_i$ replaced by $\tilde{v}_i$.

By the results of Theorem 3.1 and 3.2, we immediately get the main result of this subsection.

THEOREM 4.1: *Let assumptions* **(B1)** *and* **(B1)** *be the same as* **(A1)** *and* **(A2)** *except that $r(\cdot)$ in* **(A1)** *is replaced with $\theta(\cdot)$ in* **(B1)**. *Define $\xi(w) = E(Z_i \mid W_i = w)$. Then under* **(B1)**, **(B2)**, *and the assumption that $\xi \in \mathscr{G}_\mu^4$, the following results hold*:

(a) *Under H_0^b, $T^b \to N(0,1)$ in distribution.*

(b) *Under H_1^b, $P(T^b \geq M_n) \to 1$, where M_n is any positive, nonstochastic sequence with $M_n = o((nh^{d/2}))$.*

The proof of Theorem 4.1 (a) is given in Appendix A. The proof of Theorem 4.1 (b) follows from the facts that under H_1^b, $I_n^b \to E\{f(X)[g(X)-(Z'\gamma+\theta(W))]^2 f_w^2(W)\}$ (>0) and $\hat{\sigma}_b = O_p(1)$ in probability. The proofs of these are straightforward and are thus omitted.

4.2. *A Consistent Test For A Single Index Model*

The test for H_0^c versus H_1^c is constructed in a similar way to that for H_0^b versus H_1^b. Specifically, define $\nu_i = Y_i - \varphi(\alpha' X_i)$. Then a consistent test for H_0^c versus H_1^c can be based on an appropriate estimator of $E[\nu_i f_\alpha(\alpha' X_i) E\{\nu_i f_\alpha(\alpha' X_i)|X_i\} f(X_i)]$, where $f_\alpha(\cdot)$ is the density function of $\alpha' X_i$.

Let $\hat{\alpha}$ be the density-weighted average derivative estimator proposed by Powell, Stock, and Stoker (1989).[7] We assume that the conditions in Powell, Stock, and Stoker (1989) hold under H_0^c. Hence, $\hat{\alpha} - \alpha = O_p(n^{-1/2})$. Under H_0^c, the index function $\varphi(\alpha' X_i)$ can be consistently estimated by

$$\hat{E}(Y_i \mid \hat{\alpha}' X_i) = \frac{[(n-1)h_\alpha]^{-1} \sum_{j \neq i} Y_j K_{ij}^{\hat{\alpha}}}{\hat{f}_\alpha(\hat{\alpha}' X_i)}, \tag{12}$$

where $K_{ih}^{\hat{\alpha}} = k^\alpha((\hat{\alpha}' X_i - \hat{\alpha}' X_j)/h_\alpha)$ with $k^\alpha(\cdot)$ a univariate kernel function, h_α is a smoothing parameter, and

$$\hat{f}_\alpha(\hat{\alpha}' X_i) = \frac{1}{(n-1)h_\alpha} \sum_{j \neq i} K_{ij}^{\hat{\alpha}}. \tag{13}$$

Let $\bar{\nu}_i = Y_i - \hat{E}(Y_i \mid \hat{\alpha}' X_i)$. Then our test will be based on

$$\begin{aligned} I_n^c &\stackrel{\text{def}}{=} \frac{1}{n} \sum_i \left[\bar{\nu}_i \hat{f}_\alpha(\hat{\alpha}' X_i)\right] \left\{ \frac{1}{(n-1)h^d} \sum_{j \neq i} \left[\bar{\nu}_j \hat{f}_\alpha(\hat{\alpha}' X_j)\right] K_{ij} \right\} \\ &= \frac{1}{n(n-1)h^d} \sum_i \sum_{j \neq i} \left[\bar{\nu}_i \hat{f}_\alpha(\hat{\alpha}' X_i)\right]\left[\bar{\nu}_j \hat{f}_\alpha(\hat{\alpha}' X_j)\right] K_{ij}. \end{aligned} \tag{14}$$

It follows from (12) and (13) that

$$\begin{aligned} \bar{\nu}_i \hat{f}_\alpha(\hat{\alpha}' X_i) &= \left[Y_i - \hat{E}(Y_i \mid \hat{\alpha}' X_i)\right] \hat{f}_\alpha(\hat{\alpha}' X_i) \\ &= \frac{1}{(n-1)h_\alpha} \sum_{j \neq i} (Y_i - Y_j) K_{ij}^{\hat{\alpha}} \\ &= \left[Y_i - \hat{E}(Y_i \mid \alpha' X_i)\right] \hat{f}_\alpha(\alpha' X_i) \\ &\quad + \frac{1}{(n-1)h_\alpha} \sum_{j \neq i} (Y_i - Y_j)\left[K_{ij}^{\hat{\alpha}} - K_{ij}^{\alpha}\right] \\ &= \tilde{\nu}_i \hat{f}_\alpha(\alpha' X_i) + \frac{1}{(n-1)h_\alpha} \sum_{j \neq i} (Y_i - Y_j)\left[K_{ij}^{\hat{\alpha}} - K_{ij}^{\alpha}\right], \end{aligned} \tag{15}$$

where $\hat{E}(Y_i \mid \alpha' X_i)$ and $\hat{f}_\alpha(\alpha' X_i)$ are defined respectively in (12) and (13) with $\hat{\alpha}$ replaced by α, $\tilde{\nu}_i$ is the residual from kernel estimation of $Y_i = \varphi(\alpha' X_i) + \nu_i$,

[7] For the asymptotic theory for average derivatives in time series context, see Robinson (1989). Robinson (1989) and Stoker (1989) are the first papers using average derivatives in hypothesis testing.

where α is treated as known, i.e., $\tilde{\nu}_i = Y_i - \hat{E}(Y_i \mid \alpha' X_i) = (\varphi_i - \hat{\varphi}_i) + \nu_i - \hat{\nu}_i$ with $\varphi_i = \varphi(\alpha' X_i)$ and $\hat{\varphi}_i$ and $\hat{\nu}_i$ defined in the same way as $\hat{E}(Y_i \mid \alpha' X_i)$ except that Y_j is replaced by φ_j and ν_j respectively.

From (14) and (15), we get

$$
\begin{aligned}
(16)\qquad I_n^c = {}& \frac{1}{n(n-1)h^d} \sum_i \sum_{j \neq i} \left[\tilde{\nu}_i \hat{f}_\alpha(\alpha' X_i) \right] \left[\tilde{\nu}_j \hat{f}_\alpha(\alpha' X_j) \right] K_{ij} \\
& + \frac{2}{n(n-1)^2 h^d h_\alpha} \sum_i \sum_{j \neq i} \sum_{l \neq j} \left[\tilde{\upsilon}_i \hat{f}_\alpha(\alpha' X_i) \right] \\
& \times (Y_j - Y_l) \left[K_{jl}^{\hat{\alpha}} - K_{jl}^{\alpha} \right] K_{ij} \\
& + \frac{1}{n(n-1)^3 h^d h_\alpha^2} \sum_i \sum_{j \neq i} \sum_{k \neq i} \sum_{l \neq j} (Y_i - Y_k)(Y_j - Y_l) \\
& \times [K_{ik}^{\hat{\alpha}} - K_{ik}^{\alpha}] \left[K_{jl}^{\hat{\alpha}} - K_{jl}^{\alpha} \right] K_{ij}.
\end{aligned}
$$

As for I_n^c, the asymptotic distribution of the first term on the right-hand side of (16) can be derived by using Theorem 3.1 under Assumptions **(C1)** and **(C2)**, where **(C1)** and **(C2)** are the same as **(A1)** and **(A2)** except that $\varphi(\cdot)$ in **(C1)** replaces $r(\cdot)$ in **(A1)**, and in **(C2)**, $q_1 = 1$ and $a = h_\alpha$. We shall show in Appendix A that the last two terms on the right-hand side of (16) are $o_p((nh^{d/2})^{-1})$ under H_0^c and the additional Assumption **(C3)** given below:

(C3) $\quad \hat{\alpha} - \alpha = O_p(n^{-1/2})$, k^α is M-order differentiable, and $n^{(M-1)} h_\alpha^{2(M+2)} h^{-d} \to \infty$.

Define

$$
(17)\qquad T^c = \frac{nh^{d/2} I_n^c}{\sqrt{2}\, \hat{\sigma}_c},
$$

where $\hat{\sigma}_c{}^2$ is the same as $\hat{\sigma}_a{}^2$ defined in Theorem 3.1 with $\tilde{u}_i \hat{f}_{w_i}$ replaced by $\tilde{\nu}_i \hat{f}_\alpha(\hat{\alpha}' X_i)$.

By the results of Theorems 3.1 and 3.2, we immediately get the main result of this subsection.

THEOREM 4.2: *Under the assumptions* **(C1)**, **(C2)**, *and* **(C3)**, *the following results hold*:

(a) *Under* H_0^c, $T^c \to N(0,1)$ *in distribution.*

(b) *Suppose* $\hat{\alpha} - \alpha^* = o_p(1)$ *for some* α^* *under* H_1^c. *Then* $P(T^c \geq M_n \mid H_1^c) \to 1$, *where* M_n *is any positive, nonstochastic sequence with* $M_n = o((nh^{d/2}))$.

The assumption **(C3)** is introduced to ensure that the last two terms on the right-hand side of (16) are of smaller order than $(nh^{d/2})^{-1}$. The condition $n^{(M-1)} h_\alpha^{2(M+2)} h^{-d} \to \infty$ along with the differentiability of the kernel function up to order M may be relaxed by using uniform convergence results for U-processes. However, the proof would become much more involved. Since the kernel function is chosen by the researcher, one can always choose a kernel function that has as many derivatives as possible so that the condition:

$$n^{(M-1)} h_{\alpha}^{2(M+2)} h^{-d} \to \infty$$

is satisfied for a wide range of values of h_α and h.

5. CONCLUSIONS

We have proposed consistent tests for a partially linear regression model and an index model. In addition, we have developed a consistent test for omitted variables in a nonparametric regression model without specifying its functional form. These tests are constructed by invoking the CLT for degenerate U-statistics of order higher than two. As a result, they do not rely on any arbitrary modifications, and under the alternative hypotheses the test statistics diverge to $+\infty$ at a faster rate than $n^{1/2}$.

Similar to the power analysis of consistent tests for a parametric functional form provided in Fan and Li (1992a), Härdle and Mammen (1993), one could also investigate the local power properties of the tests developed in this paper. Under some additional conditions, one can show that the tests proposed in this paper can detect sequences of local alternatives that differ from the respective null hypotheses by $O((nh^{d/2})^{-1/2})$. Hence, they are more powerful than those based on arbitrary modifications, because the latter can only detect local alternatives distant apart from the respective null by $O(n^{-1/4})$. To the best of our knowledge, the tests proposed in this paper are the first consistent tests for a partially linear model, a single index model, and omitted variables that possess this desirable property.

Another issue that deserves some discussion is the support of X. In this paper, we have assumed that the support of X is the whole Euclidean space R^d. One could easily show that the tests in this paper are still valid if the support of X is a finite convex subset of R^d and the density function of X vanishes on the boundary of its support. However, if the support of X is a compact subset of R^d and the density function of X is bounded away from zero on its support, then the tests presented in this paper need to be modified. In this case, some trimming method must be introduced to overcome the boundary effect. The simplest way of doing this is to introduce a fixed weight function such that the support of the weight is a proper subset of the support of X as in Fan and Li (1992b). However, the resulting tests will be consistent only against the alternatives that differ from the null on the support of the weight function. In order for the tests to be consistent against all the alternatives, the weight function must change with respect to the sample size in such a way that its support approaches the support of X as the sample size n goes to $+\infty$.

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

and

Department of Economics, University of Guelph, Guelph, Ontario, Canada, N1G 2W1

Manuscript received December, 1992; final revision received July, 1995.

APPENDIX A: PROOFS OF THE THEOREMS

This appendix collects proofs of the main results stated in Sections 3 and 4. Throughout, the symbol C denotes a generic constant. When we evaluate the order of some terms, if there is no confusion, we will use n, $(n-1)$, and $(n-2)$ interchangeably. In addition, when an expression contains more than one summation, by slight abuse of notation, we will sometimes use the same symbol to denote this expression and different terms in the expression obtained from restricting the summation indices to different cases. For example, if we define $LA = \Sigma_i \Sigma_j A_{ij}$, then we will also use LA to denote $\Sigma_i \Sigma_{j\neq i} A_{ij}$ and $\Sigma_i A_{ii}$ corresponding to $i \neq j$ and $i=j$ respectively.

PROOF OF THEOREM 3.1: The following expression for I_n^a is immediate from (5) and (6):

$$\text{(A.1)}\qquad I_n^a = \frac{1}{n(n-1)h^d}\sum_i \sum_{j\neq i}\Big\{(r_i-\hat r_i)\hat f_{w_i}(r_j-\hat r_j)\hat f_{w_j} + u_iu_j\hat f_{w_i}\hat f_{w_j} + \hat u_i\hat f_{w_i}\hat u_j\hat f_{w_j}$$

$$+2u_i\hat f_{w_i}(r_j-\hat r_j)\hat f_{w_j} - 2\hat u_i\hat f_{w_i}(r_j-\hat r_j)\hat f_{w_j} - 2u_i\hat f_{w_i}\hat u_j\hat f_{w_j}\Big\}K_{ij}$$

$$\stackrel{\text{def}}{=} I_1 + I_2 + I_3 + 2I_4 + 2I_5 - 2I_6.$$

We shall complete the proof of Theorem 3.1 by examining $I_1, \ldots, I_6$ respectively in Propositions A.1 to A.6, and by showing that $\hat\sigma_a^2 = \sigma_a^2 + o_p(1)$ in Proposition A.7. Since the proof is similar to that of Theorem 1 of Fan and Li (1992c), we will only provide some important steps here. Throughout this appendix, $\mathscr{Z}_i = (u_i, X_i')'$ and $E_i(\cdot) = E(\cdot\,|\,X_i)$.

PROPOSITION A.1: $I_1 = o_p((nh^{d/2})^{-1})$.

PROOF: From (A.1), we get

$$\text{(A.2)}\qquad I_1 = \frac{1}{n(n-1)h^d}\sum_i\sum_{j\neq i}(r_i-\hat r_i)\hat f_{w_i}(r_j-\hat r_j)\hat f_{w_j}K_{ij}$$

$$= \frac{1}{n(n-1)^3h^d a^{2q_1}}\sum_i\sum_{j\neq i}\sum_{l\neq i}\sum_{k\neq j}(r_i-r_l)K_{il}^w(r_j-r_k)K_{jk}^wK_{ij}.$$

We will complete the proof by showing that $E[(I_1)^2] = o(n^{-2}h^{-d})$. However a direct evaluation of $E[(I_1)^2]$ would be very tedious since it contains eight summations. Below we will first show that $E(I_1) = o((nh^{d/2})^{-1})$. Then we will use this result and a symmetry argument to show that $E[(I_1)^2] = o(n^{-2}h^{-d})$.

To show that $E(I_1)=o((nh^{d/2})^{-1})$, we first consider the case where i,j,l,k in (A.2) are all different from each other and denote the resulting expression as I_{11}. Given the assumption that $na^{2\eta}h^{d/2}\to 0$, we get

$$\begin{aligned}E(I_{11}) &= h^{-d}a^{-2q_1}E[(r_1-r_2)K_{12}^w(r_3-r_4)K_{34}^w K_{13}]\\ &= h^{-d}a^{-2q_1}E\{[E_1(r_1-r_2)K_{12}^w][E_3(r_3-r_4)K_{34}^w]K_{13}\}\\ &\le Ch^{-d}a^{2\eta}E[D_r(W_1)D_r(W_3)K_{13}]\\ &= O(a^{2\eta})=o((nh^{d/2})^{-1})\end{aligned}$$

by Lemma B.1 and Lemmas 2 and 3 in Robinson (1988).

Let I_{12} denote the case where i,j,l,k take no more than three different values. It is straightforward to show that $E(I_{12})=O(n^{-1}a^{-q_1})=o((nh^{d/2})^{-1})$. Hence $E(I_1)=o((nh^{d/2})^{-1})$.

Now we show that $E[(I_1)^2]=o(n^{-2}h^{-d})$. From (A.2), it follows that

$$\text{(A.3)}\qquad \begin{aligned}E[(I_1)^2] &= \frac{1}{n^2(n-1)^6h^{2d}a^{4q_1}}\\ &\quad\times\sum_i\sum_{j\neq i}\sum_{l\neq i}\sum_{k\neq j}\sum_{i'}\sum_{j'\neq i'}\sum_{l'\neq i'}\sum_{k'\neq j'}E\{[(r_i-r_l)K_{il}^w(r_j-r_k)K_{jk}^wK_{ij}]\\ &\qquad\times[(r_{i'}-r_{l'})K_{i'l'}^w(r_{j'}-r_{k'})K_{j'k'}^wK_{i'j'}]\}\\ &\overset{\text{def}}{=} LA3.\end{aligned}$$

We first consider the case where the summation indices i,j,l,k are all different from i',j',l',k'. In this case, the two parts in two different square brackets are independent of each other. Hence by the same proof as that of $E(I_1)=o((nh^{d/2})^{-1})$ given earlier, we know that $LA3=o(n^{-2}h^{-d})$.

Next we consider the case where exactly one index from i,j,l,k equals exactly one of the indices i',j',l',k' (so there are altogether seven different summation indices). By symmetry, we only need to consider the case (i) $i=i'$, (ii) $i=l'$, and (iii) $l=l'$.

Case (i) $i=i'$: We have

$$\begin{aligned}LA3 &= \frac{1}{n^2(n-1)^6h^{2d}a^{4q_1}}\sum_i\sum_{j\neq i}\sum_{l\neq i}\sum_{k\neq j}E\Bigg\{[(r_i-r_l)K_{il}^w(r_j-r_k)K_{jk}^wK_{ij}]\\ &\qquad\times\left[\sum_{j'\neq i}\sum_{l'\neq i}\sum_{k'\neq j'}(r_i-r_{l'})K_{il'}^w(r_{j'}-r_{k'})K_{j'k'}^wK_{ij'}\right]\Bigg\}\\ &= \frac{1}{nh^{2d}a^{4q_1}}E\{(r_1-r_3)K_{13}^w(r_2-r_4)K_{24}^wK_{12}[(r_1-r_6)K_{16}^w(r_5-r_7)K_{57}^wK_{15}]\}\\ &= \frac{1}{nh^{2d}a^{4q_1}}E\{[E_1(r_1-r_3)K_{13}^w][E_2(r_2-r_4)K_{24}^w]\\ &\qquad\times K_{12}[E_1(r_1-r_6)K_{16}][E_5(r_5-r_7)K_{57}^w]K_{15}\}\\ &\le \frac{a^{4\eta}}{nh^{2d}}E\{D_r(W_1)D_r(W_2)K_{12}D_r(W_1)D_r(W_5)K_{15}\}\\ &= O(a^{4\eta}n^{-1})=o(n^{-2}h^{-d}).\end{aligned}$$

For case (ii) $i = l'$, we have

$$LA3 = \frac{1}{n^2(n-1)^6 h^{2d} a^{4q_1}} \sum_i \sum_{j \neq i} \sum_{l \neq i} \sum_{k \neq j} E\Bigg\{[(r_i - r_l) K_{il}^w (r_j - r_k) K_{jk}^w K_{ij}] \times \Bigg[\sum_{i' \neq i} \sum_{j' \neq i'} \sum_{k' \neq j'} (r_{i'} - r_i) K_{i'i}^w (r_{j'} - r_{k'}) K_{j'k'}^w K_{i'j'}\Bigg]\Bigg\}$$

$$= \frac{1}{nh^{2d} a^{4q_1}} E\{(r_1 - r_3) K_{13}^w (r_2 - r_4) K_{24}^w K_{12} [(r_5 - r_1) K_{51}^w (r_6 - r_7) K_{67}^w K_{56}]\}$$

$$= \frac{1}{nh^{2d} a^{4q_1}} E\{[E_1(r_1 - r_3) K_{13}^w][E_2(r_2 - r_4) K_{24}^w] \times K_{12} (r_5 - r_1) K_{51}^w [E_6(r_6 - r_7) K_{67}^w] K_{56}\}$$

$$\leq \frac{a^{3\eta}}{nh^{2d}} E\{|D_r(W_1) D_r(W_2)(r_5 - r_1) K_{51}^w K_{12} D_r(W_6) K_{56}|\}$$

$$= O(a^{3\eta} n^{-1}) = o(n^{-2} h^{-d}).$$

Similarly one can easily show that for case (iii), $LA3 = O(a^{3\eta} n^{-1}) = o(n^{-2} h^{-d})$.

Finally it is easy to see that when the eight summation indices $i, j, l, k, i', j', l', k'$ take no more than six different values, $LA3 = O(n^{-2} a^{-2q_1}) = o(n^{-2} h^{-d})$. Hence $E[(I_1)^2] = o(n^{-2} h^{-d})$.

PROPOSITION A.2: *$nh^{d/2} I_2 \to N(0, 2\sigma_a^2)$ in distribution, where*

$$\sigma_a^2 = E[f(X_1) \sigma^4(X_1) f_{w1}^4] \left[\int K^2(u)\, du\right].$$

PROOF: It follows from (A.1) that

$$I_2 = \frac{1}{n(n-1)h^d} \sum_i \sum_{j \neq i} u_i \hat{f}_{w_i} u_j \hat{f}_{w_j} K_{ij} = \frac{1}{n(n-1)^3 h^d a^{2q_1}} \sum_i \sum_{j \neq i} \sum_{l \neq i} \sum_{k \neq j} u_i u_j K_{il}^w K_{jk}^w K_{ij} \tag{A.4}$$

$$= \frac{1}{n(n-1)^3 h^d a^{2q_1}} \sum \sum \sum \sum_{i \neq j \neq l \neq k} u_i u_j K_{il}^w K_{jk}^w K_{ij} + I_2 R = I_2 U + I_2 R,$$

where $I_2 U$ denotes the case where all four subscripts i, j, l, k are different and $I_2 R$ denotes the sum of the remaining terms.

Rewriting $I_2 U$ in terms of a U-statistic, we get

$$I_2 U = \frac{\binom{n}{4}}{n(n-1)^3 h^d a^{2q_1}} \left[\binom{n}{4}^{-1} \sum_{1 \leq i < j < l < k \leq n} P_n(\mathscr{Z}_i, \mathscr{Z}_j, \mathscr{Z}_l, \mathscr{Z}_k)\right], \tag{A.5}$$

where $P_n(\mathscr{Z}_i, \mathscr{Z}_j, \mathscr{Z}_l, \mathscr{Z}_k) = \Sigma_{4!} u_i u_j K_{il}^w K_{jk}^w K_{ij}$ with $\Sigma_{4!}$ extending over $4! = 24$ different permutations of i, j, l, k.

Define $P_n(\mathscr{Z}_i, \mathscr{Z}_j) = E[P_n(\mathscr{Z}_i, \mathscr{Z}_j, \mathscr{Z}_l, \mathscr{Z}_k) | \mathscr{Z}_i, \mathscr{Z}_j]$. We get

$$P_n(\mathscr{Z}_i, \mathscr{Z}_j) = 2\{u_i u_j K_{ij} E[K_{il}^w K_{jk}^w | \mathscr{Z}_i, \mathscr{Z}_j] + u_i u_j K_{ij} E[K_{ik}^w K_{jl}^w | \mathscr{Z}_i, \mathscr{Z}_j]\}$$

$$= 4 u_i u_j K_{ij} E[K_{il}^w K_{jk}^w | \mathscr{Z}_i, \mathscr{Z}_j].$$

Hence,

$$\begin{aligned}
E[P_n^2(\mathscr{Z}_1,\mathscr{Z}_2)] &= 16E\{u_1^2u_2^2K_{12}^2(E[K_{13}^wK_{24}^w\,|\,\mathscr{Z}_1,\mathscr{Z}_2])^2\}\\
&= 16E\left\{\sigma^2(X_1)\sigma^2(X_2)K_{12}^2\left[\int f_w(w_3)f_w(w_4)K_{13}^wK_{24}^w\,dw_3\,dw_4\right]^2\right\}\\
&= 16E\Bigg\{\sigma^2(X_1)\sigma^2(X_2)K_{12}^2\\
&\qquad\times\left[\int f_w(W_1+au)f_w(W_2+av)K^w(u)K^w(v)a^{2q_1}\,du\,dv\right]^2\Bigg\}\\
&= 16a^{4q_1}\int f(x_1)f(x_2)\sigma^2(x_1)\sigma^2(x_2)K_{12}^2\\
&\qquad\times\left[\int f_w(w_1+au)f_w(w_2+av)K^w(u)K^w(v)\,du\,dv\right]^2dx_1\,dx_2\\
&= 16a^{4q_1}h^d\Bigg\{\int f(x_1)f(x_1+hs)\sigma^2(x_1)\sigma^2(x_1+hs)K^2(s)\\
&\qquad\times\left[\int f_w(w_1+au)f_w(w_1+hs_z+av)K^w(u)K^w(v)\,du\,dv\right]^2dx_1\,ds\Bigg\}\\
&= 16a^{4q_1}h^d\Bigg\{\int f(x_1)f(x_1)\sigma^4(x_1)K^2(s)\\
&\qquad\times\left[\int f_w^2(w_1)K^w(u)K^w(v)\,du\,dv\right]^2dx_1\,ds+o(1)\Bigg\}\\
&= 16a^{4q_1}h^d\left\{E[f(X_1)\sigma^4(X_1)f_{w1}^4]\left[\int K^2(s)\,ds\right]+o(1)\right\}\\
&\equiv 16a^{4q_1}h^d\{\sigma_a^2+o(1)\}.
\end{aligned}$$

Similar to Fan and Li (1992c), one can easily verify the conditions of Lemma B.4. Hence from (A.5) and Lemma B.4, it follows that

$$\begin{aligned}
nh^{d/2}I_2U &= \frac{\binom{n}{4}}{n(n-1)^3h^{d/2}a^{2q_1}}\left[n\binom{n}{4}^{-1}\sum_{i\le i<j<l<k\le n}P_n(\mathscr{Z}_i,\mathscr{Z}_j,\mathscr{Z}_l,\mathscr{Z}_k)\right]\\
&\to N\left(0,\frac{\binom{n}{4}^2}{n^2(n-1)^6h^da^{4q_1}}2^{-1}4^2(4-1)^2[16a^{4q_1}h^d\sigma_a^2]\right)\\
&\to N(0,2\sigma_a^2)\text{ in distribution.}
\end{aligned}$$

It is easy to see that

$$E[(I_2R)^2]=(n^2(n-1)^6h^{2d}a^{4q_1})^{-1}O(n^4h^{2d}a^{2q_1}).$$

Hence

$$I_2R=(n(n-1)^3h^da^{2q_1})^{-1}O_p(n^3h^da^{q_1})=O_p((na^{q_1})^{-1})=o_p((nh^{d/2})^{-1}).$$

Thus from (A.4) we get

$$nh^{d/2}I_2=nh^{d/2}I_2U+o_p(1)\to N(0,2\sigma_a^2)$$

in distribution.

PROPOSITION A.3: $I_3 = o_p((nh^{d/2})^{-1})$.

PROOF: From (A.1), we get

$$I_3 = \frac{1}{n(n-1)h^d} \sum_i \sum_{j \neq i} \hat{u}_i \hat{u}_j \hat{f}_{w_i} \hat{f}_{w_j} K_{ij}$$

$$= \frac{1}{n(n-1)^3 h^d a^{2q_1}} \sum_i \sum_{j\neq i} \sum_{l \neq i} \sum_{k \neq j} u_l u_k K^w_{il} K^w_{jk} K_{ij}.$$

We consider two cases: (i) $l = k$ and (ii) $l \neq k$ separately. We use I_3F and I_3S to denote these two cases.

Case (i): $l = k$. In this case, it is easy to see that

$$E(I_3F) = \frac{1}{n(n-1)^3 h^d a^{2q_1}} \sum \sum \sum_{i \neq j \neq l} E[u_l^2 K^w_{il} K^w_{jl} K_{ij}]$$

$$= n^{-1} h^{-d} a^{-2q_1} E[\sigma^2(X_3) K^w_{13} K^w_{23} K_{12}]$$

$$= O(n^{-1} a^{-q_1}) = o(n^{-1} h^{-d/2}) \quad \text{and}$$

$$E[(I_3F)^2] = \frac{1}{n^2(n-1)^6 h^{2d} a^{4q_1}} \times \sum \sum \sum_{i \neq j \neq l} \sum \sum \sum_{i' \neq j' \neq l'} E\{[u_l^2 K^w_{il} K^w_{jl} K_{ij}][u_{l'}^2 K^w_{i'l'} K^w_{j'l'} K_{i'j'}]\}.$$

By using similar arguments as in the proof of $E[(I_1)^2] = o(n^{-2}h^{-d})$ (see the proof of Proposition A.1), one can easily show that $E[(I_3F)^2] = O(n^{-2}a^{-2q_1}) = o(n^{-2}h^{-d})$. Hence $I_3F = o_p((nh^{d/2})^{-1})$.

Case (ii): $l \neq k$. Note that

$$I_3S = \frac{1}{n(n-1)^3 h^d a^{2q_1}} \sum_i \sum_{j \neq i} \sum_{l \neq i} \sum_{k \neq j,l} u_l u_k K^w_{il} K^w_{jk} K_{ij}$$

$$\stackrel{\text{def}}{=} \frac{1}{n(n-1)^3 h^d a^{2q_1}} \sum \sum \sum \sum_{i \neq j \neq l \neq k} u_l u_k K^w_{il} K^w_{jk} K_{ij} + I_3SS$$

$$\stackrel{\text{def}}{=} I_3SF + I_3SS, \quad \text{where}$$

(A.6) $$E[(I_3SF)^2]$$

$$= \frac{1}{n^8 h^{2d} a^{4q_1}} \sum \sum \sum \sum_{i \neq j \neq l \neq k} \sum_{i' \neq l,k} \sum_{j' \neq i',l,k} E[u_l^2 u_k^2 K^w_{il} K^w_{jk} K_{ij} K^w_{i'l} K^w_{j'k} K_{i'j'}]$$

$$= \frac{1}{n^6 h^{2d} a^{4q_1}} \sum \sum_{i \neq j = 3} \sum \sum_{i' \neq j' = 3} E[u_1^2 u_2^2 K^w_{i1} K^w_{j2} K_{ij} K^w_{i'1} K^w_{j'2} K_{i'j'}]$$

$$\stackrel{\text{def}}{=} LA6.$$

The leading term of $LA6$ corresponds to the case where i,j,i',j' are all different from each other. In this case,

$$LA6 = \frac{1}{n^2h^{2d}a^{4q_1}} E[u_1^2u_2^2K_{31}^wK_{42}^wK_{34}K_{51}^wK_{62}^wK_{56}]$$
$$= \frac{1}{n^2h^{2d}a^{4q_1}} E\left\{\sigma^2(X_1)\sigma^2(X_2)\left[\int f(x_3)f(x_4)K_{13}^wK_{24}^wK_{34}\,dx_3\,dx_4\right]^2\right\}$$
$$= \frac{h^{2d}}{n^2a^{4q_1}} E\left\{\sigma^2(X_1)\sigma^2(X_2)\left[\int f(w_1+hu, z_1+hs)\right.\right.$$
$$\times f(w_2+hv, z_2+ht)K^w\left(\frac{hu}{a}\right)K^w\left(\frac{hv}{a}\right)$$
$$\left.\left.\times K\left(\frac{w_1-w_2}{h}+(u-v), \frac{z_1-z_2}{h}+s-t\right)du\,ds\,dv\,dt\right]^2\right\}$$
$$= \frac{h^{2d}}{n^2a^{4q_1}} O(h^d) = O(n^{-2}h^{3d}a^{-4q_1}) = o(n^{-2}h^{-d}).$$

When some of the i,j,i',j' take the same value, $LA6$ will have at most an order of $(1/n^6h^{2d}a^{4q_1})O(n^3h^{2d}a^{3q_1}) = O(n^{-3}a^{-q_1}) = o(n^{-2}h^{-d})$. Thus we have shown that $I_3SF = o_p((nh^{d/2})^{-1})$.

Finally it is easy to see that

$$E\,|I_3SS| = \frac{1}{n^4h^da^{2q_1}} O(n^3h^da^{q_1}) = O(n^{-1}a^{-q_1}) = o((nh^{d/2})^{-1}).$$

Hence $I_3S = o_p((nh^{d/2})^{-1})$.

PROPOSITION A.4: $I_4 = o_p((nh^{d/2})^{-1})$.

PROOF: From (A.1), it follows that

$$I_4 = \frac{1}{n(n-1)h^d}\sum_i\sum_{j\neq i} u_i\hat{f}_{w_i}(r_j-\hat{r}_j)\hat{f}_{w_j}K_{ij}$$
$$= \frac{1}{n(n-1)^3h^da^{2q_1}}\sum_i\sum_{j\neq i}\sum_{l\neq i}\sum_{k\neq j} u_i(r_j-r_k)K_{il}^wK_{jk}^wK_{ij} \stackrel{\text{def}}{=} \frac{1}{\sqrt{n}}\sum_i u_iS_i.$$

Note that $E(I_4)=0$ and

$$E[(I_4)^2] = \frac{1}{n}\sum_i E(u_i^2S_i^2) = E[\sigma^2(x_1)S_1^2] \tag{A.7}$$
$$= \frac{1}{n^7h^{2d}a^{4q_1}}\sum_{j\neq 1}\sum_{l\neq 1}\sum_{k\neq j}\sum_{j'\neq 1}\sum_{l'\neq 1}\sum_{k'\neq j'} E[\sigma^2(X_1)(r_j-r_k)K_{jk}^wK_{1l}^wK_{1j}$$
$$\times(r_{j'}-r_{k'})K_{j'k'}^wK_{1l'}^wK_{1j'}]$$
$$\stackrel{\text{def}}{=} LA7.$$

If $1,j,l,k,j',l',k'$ are all different, by using Lemma B.1 and Lemmas 2 and 3 in Robinson (1988), it is easy to see that $LA7$ is of the order $O(n^{-1}a^{2\eta}) = o(n^{-2}h^{-d})$.

Next we consider the case where two of the subscript indices $l,k,j,l',k',j',1$ are the same. By symmetry, we only need to consider three cases: (i) l equals one of the other indices, (ii) k equals one of the other indices, and (iii) j equals one of the other indices. It is straightforward to show that for case (i), $LA7 = n^{-2}O(a^{\eta}a^{-q_1}) = o(n^{-2}h^{-d})$; for case (ii), $LA7 = n^{-2}O(a^{-q_1}) = o(n^{-2}h^{-d})$; and for case (iii), $LA7 = n^{-2}O(a^{-q_1}) = o(n^{-2}h^{-d})$.

Finally, when more than two indices take the same value, $LA7$ will have at most an order of $O(n^{-3}a^{-2q_1}) = o(n^{-2}h^{-d})$. Hence $E[(I_4)^2] = o(n^{-2}h^{-d})$.

PROPOSITION A.5: $I_5 = o_p((nh^{d/2})^{-1})$.

PROOF: Note that

$$I_5 = \frac{1}{n(n-1)h^d} \sum_i \sum_{j \neq i} \hat{u}_i \hat{f}_{w_i}(r_j - \hat{r}_j)\hat{f}_{w_j} K_{ij}$$
$$= \frac{1}{n(n-1)^3 h^d a^{2q_1}} \sum_i \sum_{j \neq i} \sum_{l \neq i} \sum_{k \neq j} u_l (r_j - r_k) K_{il}^w K_{jk}^w K_{ij}.$$

The rest of the proof is very similar to that of Proposition A.4 and thus is omitted.

PROPOSITION A.6: $I_6 = o_p((nh^{d/2})^{-1})$.

PROOF:

$$I_6 = \frac{1}{n(n-1)h^d} \sum_i \sum_{j \neq i} u_i \hat{f}_i \hat{u}_j \hat{f}_j K_{ij}$$
$$= \frac{1}{n(n-1)^3 h^d a^{2q_1}} \sum_i \sum_{j \neq i} \sum_{l \neq i} \sum_{k \neq j} u_i u_k K_{il}^w K_{jk}^w K_{ij}.$$

We consider two cases: (i) $i = k$ and (ii) $i \neq k$. We use I_6F and I_6S to denote these two cases.

Case (i): $i = k$.

$$E(I_6F) = \frac{1}{n(n-1)^3 h^d a^{2q_1}} \sum_i \sum_{j \neq i} \sum_{l \neq i} E[u_i^2 K_{il}^w K_{ji}^w K_{ij}]$$
$$= \frac{1}{nh^d a^{2q_1}} E[\sigma^2(X_1) K_{12}^w K_{13}^w K_{12}] + \frac{1}{n^2 h^d a^{2q_1}} E[\sigma^2(X_1)(K_{12}^w)^2 K_{12}]$$
$$= O(n^{-1}a^{-q_1}) + O(n^{-2}a^{-2q_1}) = o((nh^{d/2})^{-1}).$$
$$E[(I_6F)^2] = \frac{1}{n^2(n-1)^6 h^{2d} a^{4q_1}} \times \sum_i \sum_{j \neq i} \sum_{l \neq i} \sum_{i'} \sum_{j' \neq i'} \sum_{l' \neq i'} E\{[u_i^2 K_{il}^w K_{ji}^w K_{ij}][u_{i'}^2 K_{i'l'}^w K_{j'i'}^w K_{i'j'}]\}.$$

If all the summation indices i, j, l, i', j', l' are different from each other, by the same proof as $E(I_6F) = o((nh^{d/2})^{-1})$, we know that $E[(I_6F)^2] = o(n^{-2}h^{-d})$. If the six summation indices take at most five different values, then it is straightforward to see that

$$E[(I_6F)^2] = \frac{1}{n^2(n-1)^6 h^{2d} a^{4q_1}} O(n^5 h^{2d} a^{2q_1}) = O(n^{-3}a^{-2q_1}) = o(n^{-2}h^{-d}).$$

Case (ii): $i \neq k$. Note that

$$I_6S = \frac{1}{n(n-1)^3 h^d a^{2q_1}} \sum_i \sum_{j \neq i} \sum_{l \neq i} \sum_{k \neq j, k \neq i} u_i u_k K_{il}^w K_{jk}^w K_{ij}$$
$$\stackrel{\text{def}}{=} \frac{1}{n(n-1)^3 h^d a^{2q_1}} \sum \sum \sum \sum_{i \neq j \neq l \neq k} u_i u_k K_{il}^w K_{jk}^w K_{ij} + I_6SS$$
$$\stackrel{\text{def}}{=} I_6SF + I_6SS, \quad \text{where}$$

(A.8)
$$E[(I_6SF)^2] = \frac{1}{n^6h^{2d}a^{4q_1}} \sum \sum_{j\neq l=3} \sum \sum_{j'\neq l'=3} E[u_1^2u_2^2K_{1l}^wK_{j2}^wK_{1j}K_{1l'}^wK_{j'2}^wK_{1j'}] \stackrel{\text{def}}{=} LA8.$$

The leading term of $LA8$ is obtained when j, l, j', l' are all different. In this case, we have

$$\begin{aligned} LA8 &= \frac{1}{n^2h^{2d}a^{4q_1}} E[u_1^2u_2^2K_{14}^wK_{32}^wK_{13}K_{16}^wK_{52}^wK_{15}] \\ &= \frac{1}{n^2h^{2d}a^{4q_1}} E\{\sigma^2(X_1)\sigma^2(X_2)[E(K_{14}^wK_{32}^wK_{13}\mid X_1, X_2)]^2\} \\ &= \frac{1}{n^2h^{2d}a^{4q_1}} E\left\{\sigma^2(X_1)\sigma^2(X_2)\left[\int f(x_3)f_w(w_4)K_{14}^wK_{32}^wK_{13}\,dx_3\,dw_4\right]^2\right\} \\ &= \frac{1}{n^2a^{2q_1}} E\left\{\sigma^2(x_1)\sigma^2(X_2)\right. \\ &\quad \left.\times\left[\int f(w_1+hu, z_1+hs)f_w(w_1+av)K^w(v)K^w\left(\frac{w_1-w_2+hu}{a}\right)K(u,s)\,du\,ds\,dv\right]^2\right\} \\ &= (n^2a^{2q_1})^{-1}O(a^{q_1}) = O(n^{-2}a^{q_1}) = o(n^{-2}h^{-d}). \end{aligned}$$

Similarly one can show that when j, l, j', l' take no more than three different values,

$$LA8 = (n^6h^{2d}a^{4q_1})^{-1}O(n^3h^{2d}a^{2q_1}) = o(n^{-2}h^{-d}).$$

Finally it is easy to see that

$$I_6SS = \frac{1}{n^4h^da^{2q_1}} O_p(n^3h^da^{q_1}) = o_p((nh^{d/2})^{-1}).$$

Hence $I_6S = o_p((nh^{d/2})^{-1})$. *Q.E.D.*

PROPOSITION A.7: $\hat{\sigma}_a^2 = \sigma_a^2 + o_p(1)$.

PROOF: Since the detailed proof is similar to that of Proposition A.2, we will only sketch it here. By using (6) and (4), one can show that

$$\begin{aligned} &\frac{1}{n(n-1)h^d} \sum_i \sum_{j\neq i} [\tilde{u}_i\tilde{u}_j]^2\left[\hat{f}_{w_i}\hat{f}_{w_j}\right]^2 K_{ij} \\ &\quad = \frac{1}{n(n-1)h^d} \sum_i \sum_{j\neq i} u_i^2u_j^2\left[\hat{f}_{w_i}\hat{f}_{w_j}\right]^2 K_{ij} + o_p(1) \\ &\quad = \frac{1}{n(n-1)h^d} \sum_i \sum_{j\neq i} u_i^2u_j^2[f_{w_i}f_{w_j}]^2 K_{ij} + o_p(1) \\ &\quad = E\left[\{u_1^2f_{w1}\}^2\{f(W_1, Z_1)E[u_1^2f_{w1}^2\mid W_1, Z_1]\}\right] + o_p(1), \end{aligned}$$

which implies that $\hat{\sigma}_a^2 = \sigma_a^2 + o_p(1)$. *Q.E.D.*

PROOF OF THEOREM 4.1: By the result of Theorem 3.1, we only need to prove that both J_{1n} and J_{2n} defined in (10) are of order $o_p((nh^{d/2})^{-1})$. Note that we do not require that $\hat{\gamma} - \gamma = O_p(n^{-1/2})$. Hence the conditions in Robinson (1988) or in Fan, Li, and Stengos (1995) are not required here. Under the conditions given above, we have $\hat{\gamma} - \gamma = O_p(n^{-1/2} + (na^{q_1})^{-1} + a^{2\eta})$; see Fan, Li, and Stengos (1995) for a proof of this result. Then from (10), it follows that

$$J_{1n} = \frac{(\hat{\gamma}-\gamma)'}{n(n-1)h^d} \sum_i \sum_{j\neq i} (Z_j - \hat{Z}_j)\hat{f}_{w_j}u_i\hat{f}_{w_i} \stackrel{\text{def}}{=} (\hat{\gamma}-\gamma)'J_{11n}.$$

Define $\xi_i = E(Z_i \mid W_i)$ and $\eta_i = Z_i - \xi_i$. We have

$$J_{11n} = \frac{1}{n(n-1)h^d} \sum_i \sum_{j \neq i} \left[(\xi_j - \hat{\xi}_j)\hat{f}_{w_j} + \eta_j \hat{f}_{w_j} - \eta_j f_{w_j} \right] \left[(r_i - \hat{r}_i)\hat{f}_{w_i} + u_i \hat{f}_{w_i} - \hat{u}_i f_{w_i} \right] K_{ij}.$$

Comparing the terms in J_{11n} and the terms in I_n^a given in (A.1), it is obvious that J_{11n} has at most an order of $O_p((nh^{d/2})^{-1})$. Hence J_{1n} has the order of $O_p(n^{-1/2} + (na^{q_1})^{-1} + a^{2\eta})O_p(J_{11n}) = o_p((nh^{d/2})^{-1})$.

Similarly, define J_{22n} from $J_{2n} = (\hat{\gamma} - \gamma)' J_{22n} (\hat{\gamma} - \gamma)$. It is easy to see that $E|J_{22n}| = O(1)$. Hence J_{2n} has the order of $(\hat{\gamma} - \gamma)'(\hat{\gamma} - \gamma) = O_p(n^{-1} + (na^{q_1})^{-2} + a^{4\eta}) = o_p((nh^{d/2})^{-1})$. *Q.E.D.*

PROOF OF THEOREM 4.2: Let *II* and *III* denote the second and the third terms on the right-hand side of (16) respectively, i.e.,

$$\text{(A.9)} \qquad II = \frac{2}{n(n-1)^2 h^d h_\alpha} \sum_i \sum_{j \neq i} \sum_{l \neq j} [\tilde{\nu}_i \hat{f}_\alpha(\alpha' X_i)][Y_j - Y_l][K_{jl}^{\hat{\alpha}} - K_{jl}^{\alpha}]K_{ij},$$

$$\text{(A.10)} \qquad III = \frac{1}{n(n-1)^3 h^d h_\alpha^2} \sum_i \sum_{j \neq i} \sum_{k \neq i} \sum_{l \neq j} (Y_i - Y_k)(Y_j - Y_l)[K_{ik}^{\hat{\alpha}} - K_{ik}^{\alpha}][K_{jl}^{\hat{\alpha}} - K_{jl}^{\alpha}]K_{ij}.$$

Since $\tilde{\nu}_i \hat{f}_\alpha(\alpha' X_i) = [(n-1)h_\alpha]^{-1} \sum_{k \neq i}(Y_i - Y_k)K_{ik}^{\alpha}$, we get from (A.9):

$$\text{(A.11)} \qquad II = \frac{1}{n(n-1)^3 h^d h_\alpha^2} \sum_i \sum_{j \neq i} \sum_{k \neq i} \sum_{l \neq j} (Y_i - Y_k)(Y_j - Y_l)K_{ik}^{\alpha}[K_{jl}^{\hat{\alpha}} - K_{jl}^{\alpha}]K_{ij}.$$

By Taylor expansion, it follows that

$$\text{(A.12)} \qquad K_{jl}^{\hat{\alpha}} - K_{jl}^{\alpha} = \sum_{s=1}^{M} \frac{1}{s!} K^{\alpha,(s)}\left(\frac{\alpha'(X_j - X_l)}{h_\alpha} \right) \left[\frac{(\hat{\alpha} - \alpha)'(X_j - X_l)}{h_\alpha} \right]^s + \frac{1}{(M+1)!} k^{\alpha,(M+1)}(\psi_{jl}) \left[\frac{(\hat{\alpha} - \alpha)'(X_j - X_l)}{h_\alpha} \right]^{(M+1)},$$

where $k^{\alpha,(s)}(\cdot)$ is the sth derivative of $k^{\alpha}(\cdot)$ and ψ_{jl} is between $[\alpha'(X_j - X_l)]h_\alpha^{-1}$ and $[\hat{\alpha}'(X_j - X_l)]h_\alpha^{-1}$.

Substituting (A.12) into (A.11), we get

$$\begin{aligned} II = {} & \sum_{s=1}^{M} \frac{1}{s!} \left\{ \frac{1}{n(n-1)^3 h^d h_\alpha^2} \sum_i \sum_{j \neq i} \sum_{k \neq i} \sum_{l \neq j} (Y_i - Y_k)(Y_j - Y_l) K_{ik}^{\alpha} K_{jl}^{\alpha,(s)} \times \left[\frac{(\hat{\alpha} - \alpha)'(X_j - X_l)}{h_\alpha} \right]^s K_{ij} \right\} \\ & + \frac{1}{(M+1)!\, n(n-1)^3 h^d h_\alpha^2} \sum_i \sum_{j \neq i} \sum_{k \neq i} \sum_{l \neq j} (Y_i - Y_k)(Y_j - Y_l) K_{ik}^{\alpha} K^{\alpha,(M+1)}(\psi_{jl}) \times \left[\frac{(\hat{\alpha} - \alpha)'(X_j - X_l)}{h_\alpha} \right]^{(M+1)} K_{ij} \\ \stackrel{\text{def}}{=} {} & \sum_{s=1}^{M} \frac{1}{s!} II_s + \frac{1}{(M+1)!} II_{(M+1)}, \quad \text{where} \end{aligned}$$

$$K_{jl}^{\alpha,(s)} = k^{\alpha,(s)}\left(\frac{\alpha'(X_j - X_l)}{h_\alpha} \right).$$

We shall complete the proof for $II = o_p((nh^{d/2})^{-1})$ by showing that $(nh^{d/2})II_s = o_p(1)$ for $s = 1, \ldots, M, (M+1)$. The proofs for $s = 1, \ldots, M$ are similar. Hence, we will only provide the proofs for $s = 1$ and $s = M+1$. For $s = 1$, we have

$$\text{(A.13)} \qquad II_1 = \frac{(\hat{\alpha} - \alpha)'}{n(n-1)^3 h^d h_\alpha^2} \sum_i \sum_{j \neq i} \sum_{k \neq i} \sum_{l \neq j} (Y_i - Y_k)(Y_j - Y_l) \left[\frac{(X_j - X_l)}{h_\alpha} \right] K_{ik}^\alpha K_{jl}^{\alpha,(1)} K_{ij}$$
$$\stackrel{\text{def}}{=} (\hat{\alpha} - \alpha)' BB.$$

By comparing II_1 and I_n^α, we can see immediately that apart from $(\hat{\alpha} - \alpha)'$, the terms in II_1 are very similar to those in I_n^a. Thus, we will only consider the leading term in II_1 which is obtained from the case where $i \neq j \neq k \neq l$. In this case, if the null hypothesis holds, then

$$\text{(A.14)} \qquad E[BB] = \frac{1}{h^d h_\alpha^2} E\left\{ [\varphi(X_1 \alpha) - \varphi(X_3 \alpha)][\varphi(X_2 \alpha) - \varphi(X_4 \alpha)] \frac{(X_2 - X_4)}{h_\alpha} K_{13}^\alpha K_{24}^{\alpha,(1)} K_{12} \right\}$$
$$= \frac{1}{h^d h_\alpha^2} E\left\{ E_1[(\varphi(X_1 \alpha) - \varphi(X_3 \alpha)) K_{13}^\alpha] \right.$$
$$\left. \times E_2\left[(\varphi(X_2 \alpha) - \varphi(X_4 \alpha)) \frac{(X_2 - X_4)}{h_\alpha} K_{24}^{\alpha,(1)} \right] K_{12} \right\}$$
$$= \frac{1}{h^d h_\alpha^2} O(h_\alpha^{\eta+1}) O(h_\alpha^2) O(h^d)$$
$$= O(h_\alpha^{\eta+1}).$$

Similarly, one can show that $\text{var}[h_\alpha^{-(\eta+1)} BB] = o(1)$. Thus, $BB = O_p(h_\alpha^{\eta+1})$ which implies that $(nh^{d/2}) II_1 = O_p((nh^{d/2}) n^{-1/2} h_\alpha^{\eta+1}) = O_p([nh^d h_\alpha^{2\eta+2}]^{1/2}) = o_p(1)$.

Now, for $s = (M+1)$, we get

$$E[nh^{d/2} |II_{(M+1)}|] = O\left(\frac{n^{-(M+1)/2}}{h^d h_\alpha^2} \right) O(h^d) h_\alpha^{-(M+1)} O(h_\alpha)$$
$$= O\left(\frac{1}{n^{(M+1)/2} h_\alpha^{(M+1)}} \right)$$
$$= o((nh^{d/2})^{-1})$$

under the assumption that $n^{(M-1)} h_\alpha^{2(M+2)} h^{-d} \to \infty$. Similarly, one can show that $III = o_p((nh^{d/2})^{-1})$.

APPENDIX B: TECHNICAL LEMMAS

Lemmas B.1 and B.2 given below are slight modifications of some of the lemmas in Robinson (1988). Lemma B.3 is from the well-known H-decomposition due to Hoeffding (1961) and is used to prove Lemma B.4. Lemma B.4 generalizes the Central Limit Theorem for second order degenerate U-statistics of Hall (1984) to degenerate U-statistics of any finite order, which is used in deriving the asymptotic distributions of the test statistics proposed in this paper.

LEMMA B.1: *For λ, μ satisfying $l - 1 < \lambda \leq l$, $m - 1 < \mu \leq m$, where $l \geq 1$, $m \geq 1$ are integers, and for $\delta \geq 1$, let $f \in \mathcal{G}_\lambda^\infty$, $g \in \mathcal{G}_\mu^\delta$, $k \in \mathcal{K}_{l+m-1}$. Then*

$$|E[\{g(X_2) - g(X_1)\} K_{21} | X_1]| \leq D_g(X_1)(h^{(d+\eta)}),$$

where $D_g(\cdot)$ has finite δth moment and $\eta = \min(\lambda + 1, \mu)$.

PROOF: See the proof of Lemma 5 in Robinson (1988). *Q.E.D.*

LEMMA B.2: *For μ satisfying $m-1<\mu\leq m$, where $m\geq 1$ is an integer, and for $\delta\geq 1$, let $g\in\mathscr{G}_{\mu}^{2\delta}$ and $\zeta=\min(\mu,1)$. Suppose $\sup_u[|u|^{\delta\zeta+d}K^{\delta}(u)]<\infty$. Then*

$$|E[\{g(X_2)-g(X_1)\}^{\delta}K_{21}^{\delta}\,|\,X_1]|\leq M_g(X_1)(h^{(\delta\zeta+d)}),$$

where $M_g(\cdot)$ has finite second moment.

PROOF: Since X_1 and X_2 are independent, we have

$$\begin{aligned}
&E[\{g(X_2)-g(X_1)\}^{\delta}K_{21}^{\delta}\,|\,X_1]\\
&\quad=\int[g(x)-g(X_1)]^{\delta}K^{\delta}\left(\frac{x-X_1}{h}\right)f(x)\,dx\\
&\quad\leq C_{\delta}\int_{\phi_{X_1\rho}}[g(x)-g(X_1)-Q_g(x,X_1)]^{\delta}K^{\delta}\left(\frac{x-X_1}{h}\right)f(x)\,dx\\
&\quad\quad+C_{\delta}\int_{\phi_{X_1\rho}}Q_g^{\delta}(x,X_1)K^{\delta}\left(\frac{x-X_1}{h}\right)f(x)\,dx\\
&\quad\quad+C_{\delta}\int_{\bar{\phi}_{X_1\rho}}g^{\delta}(x)K^{\delta}\left(\frac{x-X_1}{h}\right)f(x)\,dx+C_{\delta}\int_{\bar{\phi}_{X_1\rho}}g^{\delta}(X_1)K^{\delta}\left(\frac{x-X_1}{h}\right)f(x)\,dx\\
&\quad\leq Ch_g^{\delta}(X_1)\int|x-X_1|^{\delta\mu}K^{\delta}\left(\frac{x-X_1}{h}\right)dx\\
&\quad\quad+CG^{\delta}(X_1)\int\left(\sum_{i=1}^{m-1}|x-X_1|^{\iota}\right)^{\delta}K^{\delta}\left(\frac{x-X_1}{h}\right)f(x)\,dx\\
&\quad\quad+C\sup_u\{[|u|^{\zeta+d/\delta}|K(u)|]^{\delta}\}h^{\delta\zeta+d}[E\{g^{\delta}(X)\}+g^{\delta}(X_1)]
\end{aligned}$$

where $E[|G(X_1)|^{2\delta}]<\infty$. Thus, the result holds by Lemma 1 in Robinson (1988) and the Lebesgue dominated convergence. *Q.E.D.*

The next lemma is from the well-known H-decomposition due to Hoeffding (1961). We need to introduce some notations and definitions. Let U_n be a U-statistic of order k given by

$$U_n=\binom{n}{k}^{-1}\sum_{(n,k)}\psi_n(X_{i_1},\ldots,X_{i_k}),$$

where ψ_n is a symmetric function that depends on n, $X_1,\ldots,X_n$ are independent and identically distributed random variables (or vectors), and the sum $\Sigma_{(n,k)}$ is taken over all subsets $1\leq i_1<\cdots<i_k\leq n$ of $\{1,2,\ldots,n\}$. We assume without loss of generality that U_n has been centered, so that $E[\psi_n(X_1,\ldots,X_k)]=0$ for each n. In this case, U_n is said to be degenerate if $E[\psi_n(X_1,\ldots,X_k)\,|\,X_1]=0$, almost surely.

Define, for $c=1,\ldots,k$, the conditional expectations

$$\psi_{nc}(x_1,\ldots,x_c)=E[\psi_n(X_1,\ldots,X_k)\,|\,(X_1,\ldots,X_c)=(x_1,\ldots,x_c)],$$

and their variances $\sigma_{nc}^2 = \text{var}[\psi_{nc}(X_1,\dots,X_c)]$. Obviously, $\sigma_{n1}^2 = 0$ for a degenerate U-statistic. Further, let $h_n^{(1)}(x_1) = \psi_{n1}(x_1)$ and

$$h_n^{(c)}(x_1,\dots,x_c) = \psi_{nc}(x_1,\dots,x_c) - \sum_{j=1}^{c-1} \sum_{(c,j)} h_n^{(j)}(x_{i_1},\dots,x_{i_j}), \quad \text{for} \quad c = 2,\dots,k.$$

LEMMA B.3: *For $j = 1,\dots,k$, let $H_n^{(j)}$ be the U-statistic based on the kernel $h_n^{(j)}$. Then*

$$U_n = \sum_{j=1}^{k} \binom{k}{j} H_n^{(j)},$$

where $H_n^{(j)} = \binom{n}{j}^{-1} \Sigma_{(n,j)} h_n^{(j)}(x_{i_1},\dots,x_{i_j})$ *satisfies the following properties*:

(a) *The U-statistics $H_n^{(1)},\dots,H_n^{(k)}$ are uncorrelated.*

(b) $\text{var}[H_n^{(j)}] = \binom{n}{j}^{-j} \text{var}[h_n^{(j)}(X_1,\dots,X_j)]$ *for* $j = 1,\dots,k$.

(c) $\text{var}[h_n^{(j)}(X_1,\dots,X_j)] = \sum_{c=1}^{j} (-1)^{j-c} \binom{j}{c} \sigma_{nc}^2$ *for* $j = 1,\dots,k$.

PROOF: See Lee (1990, Section 1.6).

LEMMA B.4: *Assume ψ_n is symmetric, $E[\psi_n(X_1,\dots,X_k) \mid X_1] = 0$ almost surely, and $E[\psi_n^2(X_1,\dots,X_k)] < \infty$ for each n. If $\sigma_{nc}^2/\sigma_{n2}^2 = o(n^{(c-2)})$ for $c = 3,\dots,k$ and*

$$\text{(B.1)} \qquad \frac{E[G_n^2(X_1,X_2)] + n^{-1}E[\psi_{n2}^4(X_1,X_2)]}{\{E[\psi_{n2}^2(X_1,X_2)]\}^2} \to 0$$

as $n \to \infty$, then nU_n is asymptotically normally distributed with zero mean and variance given by $2^{-1}k^2(k-1)^2\sigma_{n2}^2$, where $G_n(x,y) = E[\psi_{n2}(X_1,x)\psi_{n2}(X_1,y)]$.

PROOF: Since U_n is degenerate, $h_n^{(1)}(X_i) = 0$ almost surely. Lemma B.3 implies

$$\text{(B.2)} \qquad U_n = \frac{k(k-1)}{n(n-1)} \sum \sum_{1 \le i < j \le n} h_n^{(2)}(X_i, X_j) + R_n^{(2)},$$

where

$$R_n^{(2)} = \sum_{j=3}^{k} \binom{k}{j} H_n^{(j)}$$

is the remainder term and $h_n^{(2)}(X_i, X_j) = \psi_{n2}(X_i, X_j)$. Noting that $E[h_n^{(2)}(X_1, X_2) \mid X_1] = E[\psi_n(X_1,\dots,X_k) \mid X_1] = 0$, it follows that apart from a constant factor, the first term on the right-hand side of (B.2) is a degenerate U-statistic of second order. Thus, by Theorem 1 of Hall (1984), we know that Lemma B.4 is true if we can show that the remainder term $R_n^{(2)}$ in (B.2) is of a smaller order than the first term. The fact that $R_n^{(2)}$ is of a smaller order than the first term on the right-hand side of (B.2) follows by noting that

$$\operatorname{var}[n\sigma_{n2}^{-1}R_n^{(2)}] = \sum_{j=3}^{k}\binom{k}{j}^2 \operatorname{var}[H_n^{(j)}](n^2\sigma_{n2}^{-2})$$

$$= \sum_{j=3}^{k}\binom{k}{j}^2\binom{n}{j}^{-1} \operatorname{var}[h_n^{(j)}(X_1,\ldots,X_j)](n^2\sigma_{n2}^{-2})$$

$$= n^2\sum_{j=3}^{k}\binom{k}{j}^2\binom{n}{j}^{-1}\left[\sum_{c=2}^{j}(-1)^{j-c}\binom{j}{c}(\sigma_{nc}^2\sigma_{n2}^{-2})\right] = o(1),$$

where the first equality is obtained by Lemma B.3 (a), the second by Lemma B.3 (b), the third by Lemma B.3 (c) and the degeneracy of U_n, and the last by the assumption that $\sigma_{nc}^2/\sigma_{n2}^2 = o(n^{(c-2)})$ for $c = 3,\ldots,k$. Hence, $n\sigma_{n2}^{-1}R_n^{(2)} = o_p(1)$.

Note that if $k = 2$, Lemma B.4 reduces to Theorem 1 in Hall (1984).

REFERENCES

BIERENS, H. J. (1982): "Consistent Model Specification Tests," *Journal of Econometrics*, 20, 105–134.

——— (1990): "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58, 1443–1458.

BIERENS, H. J., AND W. PLOBERGER (1994): "Asymptotic Theory of Integrated Conditional Moment Tests," Working Paper, Department of Economics, Southern Methodist University.

CHEN, S. (1992): "A Simple Test for Single Index Model," Manuscript, Princeton University.

DE JONG, P. (1987): "A Central Limit Theorem for Generalized Quadratic Forms," *Probability Theory and Related Fields*, 75, 261–277.

ENGLE, R. F., C. W. J. GRANGER, J. RICE, AND A. WEISS (1986): "Semiparametric Estimation of the Relation Between Weather and Electricity Sales," *Journal of the American Statistical Association*, 81, 310–320.

EUBANK, R., AND J. HART (1992): "Testing Goodness-of-fit in Regression via Order Selection Criteria," *The Annals of Statistics*, 20, 1412–1425.

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. (1994): "Testing the Goodness-of-Fit of a Parametric Density Function by Kernel Method," *Econometric Theory*, 10, 316–356.

FAN, Y., AND Q. LI (1992a): "A General Nonparametric Model Specification Test," Manuscript, University of Windsor.

——— (1992b): "The Asymptotic Expansion for the Kernel Sum of Squared Residuals and Its Applications in Hypotheses Testing," Manuscript, University of Windsor.

——— (1992c): "Consistent Model Specification Tests: Omitted Variables, Parametric and Semiparametric Functional Forms," Manuscript, University of Windsor.

FAN, Y., Q. LI, AND T. STENGOS (1995): "Root-N-Consistent Semiparametric Regression with Conditional Heteroscedastic Disturbances," *Journal of Quantitative Economics*, 11, 229–240.

GOZALO, P. L. (1993): "A Consistent Model Specification Test for Nonparametric Estimation of Regression Function Models," *Econometric Theory*, 9, 451–477.

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.

HÄRDLE, W., AND T. M. STOKER (1989): "Investigating Smooth Multiple Regression by the Method of Average Derivatives," *Journal of the American Statistical Association*, 84, 986–995.
HIDALGO, J. (1992): "A General Nonparametric Misspecification Test," Manuscript, London School of Economics.
HOEFFDING, W. (1961): "The Strong Law of Large Numbers for *U*-statistics," *Institute of Statistical Mimeo Series*, 302, University of North Carolina.
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.
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 Alternative," *Journal of Quantitative Economics*, 10, 73–92.
LI, Q. (1994): "Some Simple Consistent Tests for a Parametric Regression Functional Form versus Nonparametric or Semiparametric Alternatives," Manuscript, University of Guelph.
NEWEY, W. K. (1985): "Maximum Likelihood Specification Testing and Conditional Moment Tests," *Econometrica*, 53, 1047–1070.
POWELL, J. L., J. H. STOCK, AND T. M. STOKER (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57, 1403–1430.
ROBINSON, P. M. (1988): "Root-*N*-Consistent Semiparametric Regression," *Econometrica*, 56, 931–954.
——— (1989): "Hypothesis Testing in Semiparametric and Nonparametric Models for Econometric Time Series," *Review of Economic Studies*, 56, 511–534.
——— (1991): "Consistent Nonparametric Entropy-Based Testing," *Review of Economic Studies*, 58, 437–453.
RODRIGUEZ, D., AND T. M. STOKER (1992): "Regression Test of Semiparametric Index Model Specification," Manuscript, M.I.T.
STOCK, J. H. (1989): "Nonparametric Policy Analysis," *Journal of the American Statistical Association*, 84, 567–575.
STOKER, T. M. (1989): "Tests of Additive Derivative Constraints," *Review of Economic Studies*, 56, 535–552.
——— (1992): *Lectures on Semiparametric Econometrics*. Louvain-la-Neuve: CORE Foundation.
TAUCHEN, G. (1985): "Diagnostic Testing and Evaluation of Maximum Likelihood Models," *Journal of Econometrics*, 30, 415–443.
ULLAH, A. (1985): "Specification Analysis of Econometric Models," *Journal of Quantitative Economics*, 2, 187–209.
WHANG, YOON-JAE, AND DONALD W. K. ANDREWS (1993): "Tests of Specification for Parametric and Semiparametric Models," *Journal of Econometrics*, 57, 277–318.
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.