*Econometric Theory*, **16**, 2000, 1016–1041. Printed in the United States of America.

# CONSISTENT MODEL SPECIFICATION TESTS

## *Kernel-Based Tests Versus Bierens' ICM Tests*

YANQIN FAN
*University of Windsor*

QI LI
*Texas A&M University*
*and*
*University of Guelph*

We point out the close relationship between the integrated conditional moment tests in Bierens (1982, *Journal of Econometrics* 20, 105–134) and Bierens and Ploberger (1997, *Econometrica* 65, 1129–1152) with the complex-valued exponential weight function and the kernel-based tests in Härdle and Mammen (1993, *Annals of Statistics* 21, 1926–1947), Li and Wang (1998, *Journal of Econometrics* 87, 145–165), and Zheng (1996, *Journal of Econometrics* 75, 263–289). It is well established that the integrated conditional moment tests of Bierens (1982) and Bierens and Ploberger (1997) are more powerful than kernel-based nonparametric tests against Pitman local alternatives. In this paper we analyze the power properties of the kernel-based tests and the integrated conditional moment tests for a sequence of "singular" local alternatives, and show that the kernel-based tests can be more powerful than the integrated conditional moment tests for these "singular" local alternatives. These results suggest that integrated conditional moment tests and kernel-based tests should be viewed as complements to each other. Results from a simulation study are in agreement with the theoretical results.

## 1. INTRODUCTION

Consider the following nonparametric regression model:

$$Y_j = g(X_j) + u_j, \qquad j = 1, \dots, n, \tag{1}$$

We thank two referees and D.W.K. Andrews for helpful comments that improved the paper. The research of both authors was supported by the Natural Sciences and Engineering Research Council of Canada and the Social Sciences and Humanities Research Council of Canada. Li's participation was also supported by Ontario Premier's Research Excellence Award and by the Bush Program in the Economics of Public Policy. An earlier version of this paper was presented at the 1996 North American summer meeting of the Econometric Society at Iowa City, Iowa; at the 1996 Statistical Symposium in Dekalb, Illinois; and at the 1997 Canadian Econometric Study Group meeting in Waterloo, Ontario. Address correspondence to: Qi Li, Department of Economics, Texas A&M University, College Station, TX 77843, USA.

  0266-4666/00 $9.50

where $\{Y_j, X_j'\}_{j=1}^n$ is independently and identically distributed (i.i.d.) as $\{Y, X'\}$, $g(\cdot)$: $R^d \to R$ is the true but unknown regression function and $u_j$ is the error satisfying $E[u_j | X_j] = 0$.

In parametric regression analysis, it is assumed that the functional form of the conditional expectation $E(Y_j|X_j)(= g(X_j))$ is known apart from a finite number of unknown parameters. Given this assumption, the researcher proceeds to estimate the vector of unknown parameters from the data and bases the statistical inference on the resulting estimate. It has long been recognized that misspecifying the functional form of the regression function may result in misleading inferences. This motivated the development of model specification tests. Early tests such as the conditional moment (CM) tests of Newey (1985) and Tauchen (1985) employ only a finite number of moment conditions implied by the null hypothesis of correct model specification and thus are not consistent against the general alternative that the null hypothesis is false. Having realized the inconsistency of such parametric tests, researchers have recently devoted much effort to the development of consistent model specification tests. Bierens (1982) was the first to give consistent model specification tests.

With respect to the regression model given by (1), consistent model specification tests refer to consistent tests for

$$H_0: P(g(X) = g_0(X, \beta_0)) = 1 \quad \text{for some } \beta_0 \in \mathcal{B} \subset R^p \quad \text{against}$$
$$H_1: P(g(X) = g_0(X, \beta)) < 1 \quad \text{for all } \beta \in \mathcal{B},$$

where $g_0(x, \beta)$ is a known function apart from the unknown parameter $\beta$.

The existing consistent model specification tests in the literature can be classified into two groups: one employs some nonparametric regression estimator, and the other does not. Tests that belong to the first group include Aït-Sahalia, Bickel, and Stoker (1994), Fan and Li (1992, 1996, 1999), Gozalo (1993), Härdle and Mammen (1993), Hong (1993), Hong and White (1995), Horowitz and Härdle (1994), Lavergne and Vuong (1996, 2000), Li (1999), Li and Wang (1998), Robinson (1989), Yatchew (1992), Whang and Andrews (1993), Wooldridge (1992), and Zheng (1996), to mention only a few. Consistent model specification tests that do not employ any nonparametric estimator of the regression function are presented in Andrews (1997), Bierens (1982, 1990), Bierens and Ploberger (1997), Chen and Fan (1999), Delgado (1993), Ploberger and Bierens (1995), and Whang (1998), among others. For brevity, we will refer to these two groups of tests as nonparametric tests and Bierens' tests hereafter. It should be noted that Stinchcombe and White (1998) extended Bierens' approach to specification testing in a more general context than the standard regression framework. We refer interested readers to Stinchcombe and White (1998) for details on the extension.

Although both groups of tests achieve consistency, they have very different properties. For example, nonparametric tests typically have asymptotic normal distribution under $H_0$, whereas Bierens' tests are usually nuisance parameter dependent and have non-normal asymptotic null distributions. The nonparametric tests can detect Pitman local alternatives converging to the null at rates slower than $n^{-1/2}$, whereas Bierens' tests can detect Pitman local alternatives that are distant apart from $H_0$ by $O(n^{-1/2})$. Hence, Bierens' tests are more powerful than nonparametric tests (e.g., Ploberger and Bierens, 1995) against Pitman local alternatives. However simulation results in Fan and Li (1992) and Hong and White (1995) show that the power of Bierens' tests does not always dominate that of nonparametric tests. For some alternatives, Bierens' tests are more powerful; and for other alternatives, nonparametric tests are more powerful.

The main purpose of this paper is to attempt to provide a theoretical justification for the simulation results alluded to earlier and to point out the relationship between some kernel-based tests and the integrated conditional moment (ICM) tests of Bierens (1982) and Bierens and Ploberger (1997). Although there exist quite a number of consistent nonparametric tests, we will focus our attention on kernel-based tests, in particular, the tests in Härdle and Mammen (1993), Li and Wang (1998), and Zheng (1996), and we will make slight modifications of these tests (see Section 2 for details). Other existing kernel-based tests are either inconsistent against deviations from the null outside a proper subset of the support of the regressor $X$ as a result of fixed trimming, or are less powerful than the preceding two tests because of sample splitting or other ad hoc modifications, or are not strictly asymptotically locally unbiased against Pitman local alternatives because of the presence of the conditional bias of the kernel regression estimator.

The main results of this paper are as follows. First, we show that the kernel-based tests in Härdle and Mammen (1993), Li and Wang (1998), and Zheng (1996) with a fixed smoothing parameter can be regarded as ICM tests of Bierens (1982) and Bierens and Ploberger (1997) with specific weight functions. Second, we show via a class of "singular" local alternatives that kernel-based tests can detect such alternatives converging in probability[1] to the null model at a rate faster than $n^{-1/2}$ and this rate can be made arbitrarily close to $n^{-3/4}$, whereas the ICM tests can only detect such "singular" alternatives that approach the null at rate $n^{-1/2}$. The main feature of the "singular" local alternatives is that they have narrow spikes and change rapidly as the sample size $n$ increases. Loosely speaking, these "singular" local alternatives can be thought of as representing high frequency alternatives and the Pitman local alternatives as representing low frequency alternatives. Hence, the ICM tests have higher power than kernel-based tests for low frequency alternatives and have lower power than kernel-based tests for high frequency alternatives. These results suggest that kernel-based tests provide complements to, rather than substitutes for, Bierens' ICM tests.

Another way of investigating the asymptotic power properties of tests of the null hypothesis of a parametric regression model against general nonparametric alternatives is the minimax approach of Ingster (1982, 1993a, 1993b, 1993c). This approach was largely unknown in econometrics until the recent work of

Horowitz and Spokoiny (1999). As explained clearly in Horowitz and Spokoiny (1999), the aim of the minimax approach is to find the optimal rate of testing. The papers by Ingster (1982, 1993a, 1993b, 1993c) establish the optimal rate of testing for several classes of smooth functions with the order of smoothness known. In the case where the order of smoothness of the underlying function class is unknown, Spokoiny (1996) presents the optimal rate of testing. For more details on this approach, we refer readers to the previously mentioned papers. Horowitz and Spokoiny (1999) take the minimax approach and propose an adaptive, rate-optimal test of a parametric model against nonparametric alternatives.

The remainder of the paper is organized as follows. Section 2 first provides a brief review of the two kernel-based tests and Bierens' ICM tests and then derives the relationship between the two kernel-based tests and the ICM tests of Bierens (1982) and Bierens and Ploberger (1997). In Section 3, we analyze the local power properties of the kernel-based tests and the ICM tests for the sequence of "singular" local alternatives. Section 4 presents some simulation results. The technical proofs are given in the Appendix.

## 2. NONPARAMETRIC TESTS VERSUS BIERENS' TESTS

Nonparametric estimation techniques such as kernel, series, spline, and sieve have all been used in consistent model specification testing. In this section, we examine some of the tests based on the kernel estimation of the regression function. We are particularly concerned with their relationship with the ICM tests of Bierens (1982) and Bierens and Ploberger (1997).

Throughout the rest of this paper, we use $\sum_j$ to denote $\sum_{j=1}^{n}$ and $\sum_l \sum_{j\neq l}$ to denote $\sum_{l=1}^{n} \sum_{j\neq l, j=1}^{n}$.

### 2.1. Kernel-Based Tests

In this subsection, we briefly review two kernel-based tests: (i) the test in Härdle and Mammen (1993) and (ii) the test in Li and Wang (1998) and Zheng (1996). The test for $H_0$ versus $H_1$ established in Härdle and Mammen (1993) employs the Nadaraya–Watson kernel estimator of $g(x)$ for $x \in R^d$ given by

$$\hat{g}(x) = \frac{(1/nh^d)\sum_j Y_j K_{jx}}{\hat{f}(x)}, \tag{2}$$

where $K_{jx} = K[(X_j - x)/h]$, $K(\cdot): R^d \to R$ is a symmetric kernel function, $h = h_n \to 0$ is a smoothing parameter, and $\hat{f}(x)$ is the kernel estimator of the probability density function (p.d.f.) $f(x)$ of the regressor $X$ given by

$$\hat{f}(x) = \frac{1}{nh^d}\sum_j K_{jx}. \tag{3}$$

Let $\tilde{\beta}$ be the nonlinear least squares (NLS) estimator of $\beta_0$ under $H_0$. Define $\tilde{g}(x) = g_0(x, \tilde{\beta})$. Härdle and Mammen (1993) use a weighted integrated squared difference between the kernel estimator $\hat{g}(x)$ and the kernel-smoothed parametric estimator $\tilde{g}(x)$ to construct a consistent test for $H_0$ versus $H_1$:

$$T_n = \int [\hat{g}(x) - K_{h,n} * \tilde{g}(x)]^2 W(x)\,dx, \tag{4}$$

where the range of integration in (4) is the support of $X$, $W(\cdot): R^d \to R$ is a smooth weight function, and

$$K_{h,n} * \tilde{g}(x) = \frac{(nh^d)^{-1} \sum_j \tilde{g}(X_j) K_{jx}}{\hat{f}(x)}. \tag{5}$$

For regressor $X$ with compact support, Härdle and Mammen (1993) show that under certain conditions, $T_n$ is asymptotically normally distributed under $H_0$ and the test based on $T_n$ is a consistent test of $H_0$ against $H_1$. In addition, the test can detect Pitman local alternatives that converge to the null at rate $O((nh^{d/2})^{-1/2})$ (more slowly than $n^{-1/2}$). They also provide a "wild bootstrap" procedure to approximate the finite sample distribution of $T_n$ under $H_0$ when the sample size $n$ is small.

Li and Wang (1998) and Zheng (1996) independently propose a simple consistent test for $H_0$ versus $H_1$ based on

$$I_n = \frac{1}{n(n-1)h^d} \sum_l \sum_{j \neq l} \tilde{u}_l \tilde{u}_j K_{lj}, \tag{6}$$

where $\tilde{u}_j = Y_j - \tilde{g}(X_j)$ is the (parametric) residual and $K_{lj} = K[(X_l - X_j)/h]$.

The asymptotic properties of the test based on $I_n$ are similar to those of the test in Härdle and Mammen (1993). However, its simplicity (it is easy to compute) may appeal to applied researchers.

We note that an alternative expression for $I_n$ is given by

$$I_n = \frac{1}{n} \sum_l \tilde{u}_l [\hat{g}_{-l}(X_l) - K_{h,n,-l} * \tilde{g}(X_l)] \hat{f}_{-l}(X_l), \tag{7}$$

where $\hat{g}_{-l}(X_l)$, $\hat{f}_{-l}(X_l)$, and $K_{h,n,-l} * \tilde{g}(X_l)$ are leave-one-out versions of (2), (3), and (5), respectively. It is obvious from (4) and (7) that the conditional bias of the kernel regression estimator under $H_0$ has been removed from both $T_n$ and $I_n$, which results in the strictly asymptotically locally unbiasedness of the tests based on $T_n$ and $I_n$ (see Härdle and Mammen, 1993; Zheng, 1996). This desirable feature distinguishes $T_n$ and $I_n$ tests from other kernel-based tests such as those in Hong (1993) and Aït-Sahalia et al. (1994).

## 2.2. Bierens' Tests

In contrast to the nonparametric tests mentioned earlier, Bierens' tests do not make use of any nonparametric estimator of the regression function $g(\cdot)$. In fact, they can be viewed as extensions of the class of conditional moment tests proposed by Newey (1985) and Tauchen (1985). To achieve consistency, Bierens (1982, 1990) and Bierens and Ploberger (1997) employ an infinite number of conditional moment conditions obtained by using a class of weight functions indexed by a continuous nuisance parameter (for details, see Bierens and Ploberger, 1997, and the references therein). The asymptotic null distributions of these tests are typically non-normal and nuisance parameter dependent.

The ICM tests proposed by Bierens (1982) and Bierens and Ploberger (1997) are based on

$$\rho = \int_{\Xi} |\hat{Z}(\xi)|^2 d\mu(\xi), \tag{8}$$

where $\Xi$ is a compact subset of $R^d$, $\mu(\xi)$ is a probability measure on $\Xi$,

$$\hat{Z}(\xi) = \frac{1}{\sqrt{n}} \sum_j \tilde{u}_j w_j(\xi), \tag{9}$$

$w_j(\xi)$ is a weight function that can take the form $w_j(\xi) = w(\xi'\Phi(X_j))$ or $w_j(\xi) = w(\xi + \Phi(X_j))$, and $\Phi(\cdot)$ is a bounded Borel measurable mapping from $R^d$ to $R^d$ such that $\Phi(X_j)$ and $X_j$ generate the same euclidean Borel field. To achieve consistency, one can take $\Xi$ to be a neighborhood of the origin and take $w(\cdot)$ to satisfy certain conditions. For example, if $w_j(\xi) = w(\xi'\Phi(X_j))$, then $w(\cdot)$ must satisfy the conditions of Theorem 1 in Bierens and Ploberger (1997); if $w_j(\xi) = w(\xi + \Phi(X_j))$, then $w(\cdot)$ must be proportional to the moment generating function or the characteristic function of an absolutely continuous distribution with density having bounded $d$-dimensional support (see Bierens and Ploberger, 1997, Theorem 2). In Bierens (1982), $w_j(\xi) = \exp(i\xi'\Phi(X_j))$ and $d\mu(\xi) = d\xi$, where $i^2 \equiv -1$. The ICM test with the preceding choice of $w_j(\cdot)$ and $\mu(\cdot)$ was originally constructed from a Fourier transform characterization of conditional expectations (see Bierens, 1982). Corollary 3 of Bierens (1982) ensures consistency of this test. Bierens (1982) also derives upper bounds for the asymptotic critical values of this test based on Chebeyshev's inequality. Recently, Bierens and Ploberger (1997) show that for a general weight function and a general probability measure, the asymptotic null distribution of $\rho$ is given by that of an infinite sum of weighted independent $\chi_1^2$ random variables. Because the asymptotic critical values of the null distribution are case dependent, Bierens and Ploberger (1997) propose to approximate these critical values by the conditional Monte Carlo approach of Hansen (1996) and De Jong (1996) and derive sharper upper bounds for the asymptotic critical values of these tests than obtained from Chebeyshev's inequality.

### 2.3. The Relationship

The tests in Härdle and Mammen (1993), Li and Wang (1998), and Zheng (1996) and the ICM tests of Bierens (1982) and Bierens and Ploberger (1997) are developed from completely different ideas. Moreover, the existing expressions for the test statistics look completely different (compare (4)–(6) and (8), (9)). However, in this subsection we will show by using Fourier transforms that alternative expressions exist for the appropriately modified versions of $T_n$, $\rho_1$ (say), and $I_n$, $\rho_2$ (say), that are in essence the same as the ICM test ($\rho$) with specific weight functions and probability measures.

To motivate the modified version of $T_n$, the basis of the test in Härdle and Mammen (1993), we follow Bierens (1982) by noting that $E[\{Y_j - g_0(X_j,\beta)\}|X_j] = E[\{Y_j - g_0(X_j,\beta)\}|Z_j]$ a.s., where $Z_j = \Phi(X_j)$ and $\Phi(\cdot)$ is a bounded Borel measurable mapping from $R^d$ to $R^d$ such that $X_j$ and $Z_j$ generate the same euclidean Borel field. Hence, $H_0$ and $H_1$ are, respectively, equivalent to

$$H_0': P(E[\{Y_j - g_0(X_j,\beta_0)\}|Z_j] = 0) = 1 \quad \text{for some } \beta_0 \in \mathcal{B} \subset R^p \quad \text{and}$$

$$H_1': P(E[\{Y_j - g_0(X_j,\beta)\}|Z_j] = 0) < 1 \quad \text{for all } \beta \in \mathcal{B}.$$

Thus, we can test $H_0$ versus $H_1$ by testing $H_0'$ versus $H_1'$ via the test in Härdle and Mammen (1993). Specifically, let $f_Z(\cdot)$ and $\hat{f}_Z(\cdot)$ be, respectively, the p.d.f. of $Z = \Phi(X)$ and the kernel estimate of $f_Z(\cdot)$. In addition, let $\hat{g}_Z(\cdot)$ and $K_{h,n,Z} * \tilde{g}(\cdot)$ denote the kernel estimate of $E(Y|Z = \cdot)$ and $E(\tilde{g}(X)|Z = \cdot)$. By choosing $W(\cdot) = \hat{f}_Z^2(\cdot)$ in (4) and letting $\rho_1$ denote the resulting test statistic, we get

$$\begin{aligned}\rho_1 &= \int_{R^d} [\hat{g}_Z(z) - K_{h,n,Z} * \tilde{g}(z)]^2 \hat{f}_Z^2(z)\, dz \\ &= \int_{R^d} \left[\hat{g}_Z(z)\hat{f}_Z(z) - \frac{1}{nh^d}\sum_j \tilde{g}(X_j)K\left(\frac{Z_j - z}{h}\right)\right]^2 dz \\ &= \int_{R^d} \left[\frac{1}{nh^d}\sum_j (Y_j - \tilde{g}(X_j))K\left(\frac{Z_j - z}{h}\right)\right]^2 dz \\ &= \int_{R^d} \left[(nh^d)^{-1}\sum_j \tilde{u}_j K\left(\frac{Z_j - z}{h}\right)\right]^2 dz, \end{aligned} \tag{10}$$

where $\tilde{u}_j = Y_j - \tilde{g}(X_j)$ and $Z_j = \Phi(X_j)$. We point out here that the range of integration in (10) is $R^d$, although $Z$ is bounded. One can show that under appropriate conditions including $h \to 0$, $\rho_1$ yields an asymptotically valid test for $H_0'$ versus $H_1'$.

Let $A(z) = (nh^d)^{-1} \sum_j \tilde{u}_j K[(Z_j - z)/h]$ and $\bar{A}(\cdot)$ be the Fourier transform of $A(\cdot)$. By definition and the fact that $A(z)$ is well defined for all $z \in R^d$, we get

$$\begin{aligned}
\bar{A}(t) &= (2\pi)^{-d/2} \int_{R^d} \exp(it'z) A(z)\, dz \\
&= \frac{1}{nh^d (2\pi)^{d/2}} \sum_j \tilde{u}_j \int_{R^d} \exp(it'z) K\left(\frac{Z_j - z}{h}\right) dz \\
&= \frac{1}{n(2\pi)^{d/2}} \sum_j \tilde{u}_j \int_{R^d} \exp(it'[Z_j + hz]) K(z) dz \\
&= \frac{1}{n(2\pi)^{d/2}} \sum_j \tilde{u}_j \exp(it'Z_j) \int_{R^d} \exp(it'hz) K(z) dz \\
&= \frac{1}{n} \sum_j \tilde{u}_j \exp(it'Z_j) \bar{K}(th), \qquad \textbf{(11)}
\end{aligned}$$

where $\bar{K}(\cdot)$ is the Fourier transform of $K(\cdot)$. Thus, by Parseval's identity and symmetry of $K(\cdot)$, we obtain the following lemma.

LEMMA 2.1. *$\rho_1 = \int_{R^d} |(1/n) \sum_j \tilde{u}_j \exp(it'Z_j)|^2 \bar{K}^2(th)\, dt$, where $Z_j = \Phi(X_j)$.*

It is interesting to note that the alternative expression for $\rho_1$ given in Lemma 2.1 is in essence the basis of the ICM test in Bierens (1982) and Bierens and Ploberger (1997), where the weight function is $w_j(\xi) = \exp(i\xi'\Phi(X_j))$ and the probability measure (apart from a constant) is $d\mu(\xi) = \bar{K}^2(\xi h) d\xi$. It follows immediately from Corollary 3 in Bierens (1982) or Theorem 1 in Bierens and Ploberger (1997) that for a fixed $h$, the test based on $\rho_1$ is a consistent test for $H_0$ versus $H_1$, provided that the kernel function $K(\cdot)$ is absolutely integrable and its Fourier transform $\bar{K}(\cdot)$ satisfies the condition that there exists a compact subset $\Xi$ (say) of $R^d$ containing the origin such that the set $\{\xi \in \Xi : \bar{K}(\xi h) = 0\}$ has Lebesgue measure zero. This condition is satisfied by most of the commonly used kernel functions such as the uniform kernel, the standard normal kernel, the triangular kernel, the double exponential kernel, and the Epanechninov kernel. It is important to note that Lemma 2.1 only relates the test in Härdle and Mammen (1993) to a special class of the ICM tests. In general the weight function and the probability measure that characterize this special class may not be optimal.

A similar relationship exists between the test based on $\rho_2$, defined by adding the $j = l$ terms to $I_n$ given in (6) and the ICM tests of Bierens (1982) and Bierens and Ploberger (1997):

$$\rho_2 = \frac{1}{n^2 h^d} \sum_l \sum_j \tilde{u}_l \tilde{u}_j K\left(\frac{Z_l - Z_j}{h}\right). \qquad \textbf{(12)}$$

To see this, note that by inverse Fourier transform, we get

$$K\left(\frac{Z_l - Z_j}{h}\right) = (2\pi)^{-d/2} \int_{R^d} \exp\left[-it'\left((Z_l - Z_j)/h\right)\right] \bar{K}(t)\, dt.$$

From (12) it follows that

$$\rho_2 = \frac{1}{(2\pi)^{d/2} n^2 h^d} \sum_j \sum_l \tilde{u}_j \tilde{u}_l \int_{R^d} \exp\left[-it'\left(\frac{Z_l - Z_j}{h}\right)\right] \bar{K}(t)\, dt$$

$$= \frac{1}{(2\pi)^{d/2} n^2} \sum_j \sum_l \tilde{u}_j \tilde{u}_l \int_{R^d} \exp[-it'(Z_l - Z_j)] \bar{K}(th)\, dt$$

$$= \frac{1}{(2\pi)^{d/2}} \int_{R^d} \left| \frac{1}{n} \sum_j \tilde{u}_j \exp(it'Z_j) \right|^2 \bar{K}(th)\, dt. \tag{13}$$

LEMMA 2.2. $\rho_2 = [1/(2\pi)^{d/2}] \int_{R^d} |(1/n) \sum_j \tilde{u}_j \exp(it'Z_j)|^2 \bar{K}(th)\, dt$, *where* $Z_j = \Phi(X_j)$.

Lemma 2.2 implies that for the test based on $\rho_2$ with a fixed $h$ to be consistent, the Fourier transform of $K(\cdot)$ must be such that there exists a compact subset $\Xi$ (say) of $R^d$ containing the origin ($\Xi = R^d$ is allowed) such that $\bar{K}(th)$ vanishes outside $\Xi$ and the set $\{t \in \Xi : \bar{K}(th) \leq 0\}$ has Lebesgue measure zero. The standard normal kernel, the triangular kernel, and the double exponential kernel satisfy this condition, but the uniform kernel and the Epanechnikov kernel do not. By comparing Lemma 2.1, Lemma 2.2, (8), and (9), we arrive at the following two conclusions: (a) the tests in Härdle and Mammen (1993), Li and Wang (1998), and Zheng (1996) (after some modifications), in a certain sense, are the same tests; (b) both tests are related to the special class of the ICM tests of Bierens (1982) and Bierens and Ploberger (1997) with the complex-valued exponential weight function $w_j(\xi) = w(\xi' Z_j) = \exp(i\xi' Z_j)$ except that the "smoothing parameter" $h$ is treated differently (vanishing $h$ versus fixed $h$).

We recall that the weight function in the ICM tests can also take the form $w_j(\xi) = w(\xi + Z_j)$. As long as $w(\cdot)$ is proportional to the moment generating function or the characteristic function of an absolutely continuous distribution with density having bounded $d$-dimensional support, Theorem 2 in Bierens and Ploberger (1997) ensures that the corresponding ICM test is consistent. Interestingly, this class of ICM tests is also related to some kernel tests. For example, consider the case with $d = 1$. Let $\mu(\cdot)$ be the uniform probability measure on $[-1,1]$. Then, one can see from (8), (9), and change of variables that

$$\rho = \frac{1}{n} \int_{-1}^{1} \left| \sum_j \tilde{u}_j w(Z_j - \xi) \right|^2 d\xi. \tag{14}$$

Obviously one can interpret the preceding ICM test as a kernel test with a fixed smoothing parameter ($h = 1$) and a kernel function $w(\cdot)$ provided that $w(\cdot)$

satisfies certain conditions. In fact, it is not difficult to see that $\rho$ in (14) is a special case of $T_n$ in (4). Alternatively, if we let $\mu(\cdot)$ be the probability measure of the random variable $(-Z)$ and replace the unknown $\mu(\cdot)$ with the empirical distribution function of $\{-Z_j\}_{j=1}^n$ in $\rho$, then we get the following ICM test:

$$\rho = \frac{1}{n^2} \sum_{k=1}^{n} \left| \sum_j \tilde{u}_j w(Z_j - Z_k) \right|^2. \tag{15}$$

This can also be interpreted as a kernel test with a fixed smoothing parameter.

## 3. FIXED $h$ OR VANISHING $h$

The results in Section 2 reveal that there exists a close relationship between the tests in Härdle and Mammen (1993) (also in Li and Wang, 1998; Zheng, 1996) and a special class of the ICM tests in Bierens and Ploberger (1997). The difference lies in the treatment of the "smoothing parameter" $h$. In the former, $h$ plays the role of a smoothing parameter so that $h \to 0$ as $n \to \infty$, whereas in the latter $h$ is fixed. Fixing $h$ and choosing a vanishing $h$ lead to completely different tests with different asymptotic distributions under $H_0$ and with different power properties. The existing results in the literature suggest that fixing $h$ improves the power of the tests because the ICM tests (fixed $h$) can detect Pitman local alternatives that approach the null at the rate of $O(n^{-1/2})$ as shown in Bierens and Ploberger (1997), whereas the tests in Härdle and Mammen (1993), Li and Wang (1998), and Zheng (1996) can only detect such local alternatives converging to the null at rate $O((nh^{d/2})^{-1/2})$, slower than $n^{-1/2}$ because $h \to 0$. However intuition suggests that if the alternative changes drastically or is of high frequency, then tests with a shrinking $h$ may be more powerful than tests based on a fixed $h$ (at least when sample size $n$ is large), because the kernel estimate with a fixed $h$ may oversmooth the true regression function and thus obscure the main feature of the alternative. Simulation results in Fan and Li (1992) show that this is indeed the case. This motivates us to study power properties of these tests for high frequency alternatives.

To facilitate theoretical analysis, we consider a class of "singular" local alternatives that represent high frequency alternatives, and we investigate the power properties of the kernel-based tests and the ICM tests for the class of "singular" local alternatives. These alternatives were first proposed by Rosenblatt (1975) and later used by Ghosh and Huang (1991) and Fan (1994) in the context of testing goodness of fit of a density function. In the context of a regression model, they take the following form:

$$LH_s : Y_j = g_0(X_j, \beta_0) + \gamma_n \delta_n(X_j) + u_j, \tag{16}$$

where $\gamma_n$ is a deterministic sequence and $\alpha_n \equiv \int \delta_n^2(x)dx \to 0$, as $n \to \infty$.

In view of the close relationship between $T_n$ $(\rho_1)$ and $I_n$ $(\rho_2)$ and the simple nature of $I_n$ or $\rho_2$, we will focus on the analysis of $I_n$ or $\rho_2$ in the rest of this paper. Subsequently we provide a detailed analysis of the local power properties of the tests based on $I_n$ and $\rho_2$. To unify the analysis of $I_n$ and $\rho_2$, we do not transform $X_j$'s explicitly but rather allow for bounded $X_j$'s. We adopt the following conditions.

(C1).

(i) Random variables $(Y_j, X_j')$, $j = 1,2,\ldots,n$ are i.i.d. as $(Y, X')$, and $X$ admits a bounded density function $f(\cdot)$ on $\mathcal{S}$ ($\mathcal{S}$ is the support of $f(\cdot)$), where $\mathcal{S}$ is a convex subset of $R^d$ ($\mathcal{S} = R^d$ is included as a special case) and $f(\cdot)$ is continuous in the interior of $\mathcal{S}$;
(ii) $\nabla g_0(X,\cdot)$ and $\nabla^2 g_0(X,\cdot)$ are continuous in $X$ and dominated by a function (say, $M(X)$) with finite second moments, where $\nabla g_0(X,\cdot)$ and $\nabla^2 g_0(X,\cdot)$ are $p \times 1$ vector of first order partial derivatives and $p \times p$ matrix of second order partial derivatives of $g_0$ with respect to $\beta$, respectively;
(iii) $E[\nabla g_0(X,\beta)\nabla' g_0(X,\beta)]$ is nonsingular for $\beta$ in a neighborhood of $\operatorname{plim}\tilde{\beta}$.

(C2). Kernel $K: \mathcal{R}^d \to \mathcal{R}$ is bounded and symmetric with $\int K(z)dz = 1$ and $\int K(z)\|z\|dz = c$, where $0 < c < \infty$ is a constant, and $\|\cdot\|$ denotes euclidean norm.

(C3). As $n \to \infty$, $nh^d \to \infty$, $h^d = o(\alpha_n)$, $\int|\delta_n(x)|dx = O(\alpha_n)$, $nh^{d/2}\gamma_n^2\alpha_n \equiv C_{1n} \to C_1$, $[\int f^2(x)\delta_n^2(x)\,dx]/\alpha_n \equiv C_{2n} \to C_2$, where $C_1$ is a finite non-negative constant and $C_2$ is a finite positive constant.

(C4).

(i) $\delta_n(x)$ is uniformly bounded in both $n$ and $x$ and satisfies either $E[\delta_n(X)\nabla g_0(X,\beta_0)] = 0$ or $E[\delta_n(X)\nabla g_0(X,\beta_0)] = \alpha_n$ with $\gamma_n\alpha_n$ of magnitude $O(n^{-1/2})$ or smaller;
(ii) Let $\sigma^2(x) = E(U^2|X = x)$ and $\mu_4(x) = E[U^4|X = x]$, where $U = Y - g_0(X,\beta_0)$. Then $(f\delta_n)$, $(\sigma^2\delta_n)$, $(f\sigma^2)$, $(fM)$ and $\mu_4$ all satisfy some Lipschitz type conditions:
(a) $|(f\delta_n)(x+v) - (f\delta_n)(x)| \le \alpha_n^{-1/d}G(x)\|v\|$,
(b) $|(\sigma^2\delta_n)(x+v) - (\sigma^2\delta_n)(x)| \le \alpha_n^{-1/d}G(x)\|v\|$,
(c) $|(f\sigma^2)(x+v) - (f\sigma^2)(x)| \le G(x)\|v\|$,
(d) $|(fM)(x+v) - (fM)(x)| \le G(x)\|v\|$,
(e) $|\mu_4(x+v) - \mu_4(x)| \le G(x)\|v\|$,
where $E[G^2(X)] < \infty$ and the following condition holds for $D(x) = G(x)$, $G^2(x)$, $\sigma^2(x)G(x)$, and $\sigma^2(x)$:
(f) $\int f(x)|D(x)\delta_n(x)|dx = O(\alpha_n)$.

Some discussions on the assumptions are in order. Condition (C1) is a standard condition used in the literature on the parametric model specified under the null hypothesis. It implies that under $H_0$, $\sqrt{n}(\tilde{\beta} - \beta_0) = O_p(1)$. Condition (C2) imposes some basic assumptions on the kernel function $K(\cdot)$. The first condition in (C3) is a standard assumption on the smoothing parameter $h$. The second one, $h^d = o(\alpha_n)$, can be better understood via the example $\delta_n(x) = L((x - x_0)/\zeta_n)$, where $L(\cdot)$ is a bounded function, $x_0$ is a fixed point in $R^d$, and

$\zeta_n \to 0$ as $n \to \infty$. For this example, $\alpha_n = O(\zeta_n^d)$. Hence, this condition is equivalent to $h = o(\zeta_n)$. If we take $L(\cdot)$ to be the uniform density over $[0,1]$, then the support of $\delta_n(x)$ is $[x_0, \zeta_n + x_0]$. The second condition in (C3) just states that the smoothing parameter $h$ must be smaller (in order) than the length of the support of $\delta_n(\cdot)$. The third condition in (C3) is satisfied by the preceding example. The fourth and fifth conditions in (C3) ensure that the test based on $I_n$ has nontrivial power against alternatives in $LH_s$ ($C_1 > 0$). The third condition is satisfied by choosing $h$, $\gamma_n$, and $\alpha_n$ appropriately, and the fourth condition is satisfied by the preceding example as long as $f(x)$ is continuous at $x = x_0$ and $f(x_0) \neq 0$, because it can be shown that $\int f^2(x) L^2((x - x_0)/\zeta_n)dx/\alpha_n \to f^2(x_0)$. Condition (C4) involves some assumptions on the local alternatives. The first one along with (C3) ensures that under the sequence of "singular" local alternatives specified in (16), the NLS $\tilde{\beta}$ still converges to $\beta_0$ at rate $n^{-1/2}$. To see this, note that under condition (C1), one can easily show that

$$\sqrt{n}(\tilde{\beta} - \beta_0) = A(\beta_0)^{-1}\left[\frac{1}{\sqrt{n}}\sum_j u_j \nabla g_0(X_j, \beta_0)\right] + A(\beta_0)^{-1}\left[\frac{\gamma_n}{\sqrt{n}}\sum_j \delta_n(X_j)\nabla g_0(X_j, \beta_0)\right], \tag{17}$$

where $A(\beta_0) = p\lim_{n\to\infty} n^{-1}\sum_j \nabla g_0(X_j, \beta_0)\nabla' g_0(X_j, \beta_0)$. It is a simple exercise to show that the first term on the right hand side of (17) is $O_p(1)$. Under (C4)(i), the second term has a mean given by either zero or $A(\beta_0)^{-1}O(n^{1/2}\gamma_n\alpha_n)$ that approaches a finite constant, and it has a variance that is of the same order as $\gamma_n^2 \int \delta_n^2(x)\nabla' g_0(x, \beta_0)\nabla g_0(x, \beta_0) f(x)dx = O(\gamma_n^2 \alpha_n^{1/2})$. Condition (C3), in particular $nh^{d/2}\gamma_n^2\alpha_n = O(1)$ and $nh^d \to \infty$, implies $\gamma_n^2\alpha_n = o(h^{d/2})$. Hence, $\gamma_n^2\alpha_n^{1/2} = o(h^{d/2}\alpha_n^{-1/2}) = o(1)$ given $h^d = o(\alpha_n)$ assumed in (C3). Consequently, the second term on the right hand side of (17) is either $o_p(1)$ or $O_p(1)$. The second condition in (C4) is employed to simplify the proof. Loosely speaking, it ensures that the second term of the Taylor series expansion of the various terms in the decomposition of $I_n$ that involve $\delta_n(x)$ is of smaller order than the first. It is easy to verify that this condition is satisfied by the preceding example of "singular" local alternatives. With a more involved proof, (C4) can be weakened.

The following theorem gives the asymptotic distributions of $I_n$ and $\rho_2$ under $LH_s$.

THEOREM 3.1. *Under conditions* (C1)–(C4) *and* $LH_s$, *we have*

(i) $nh^{d/2}I_n \to N(C_1 C_2, \sigma_0^2)$ *in distribution,*

*where* $C_1$ *and* $C_2$ *are the finite constants defined in* (C3) *and* $\sigma_0^2 = 2[\int K^2(u)du]E[f(X)\sigma^4(X)]$.

(ii) *If, in addition,* $\gamma_n \alpha_n = o(h^{d/2})$, *then* $nh^{d/2}[\rho_2 - (nh^d)^{-1}K(0)\int \sigma^2(x)f(x)dx] \to N(C_1C_2, \sigma_0^2)$ *in distribution.*

(iii) *Let* $\bar{I}_n = nh^{d/2}I_n/\hat{\sigma}_0$ *and* $\bar{\rho}_2 = nh^{d/2}[\rho_2 - (n^2h^d)^{-1}K(0)\sum_j \tilde{u}_j^2]/\hat{\sigma}_0$, *where* $\hat{\sigma}_0^2 = [2/n(n-1)h^d]\sum_l \sum_{j\neq l} \tilde{u}_l^2 \tilde{u}_j^2 K_{lj}^2$. *Then* $\bar{I}_n \to N((C_1C_2/\sigma_0),1)$ *and* $\bar{\rho}_2 \to N((C_1C_2/\sigma_0),1)$ *in distribution.*

First we note that under $H_0$, $C_2 = 0$. Theorem 3.1(iii) implies that $\bar{I}_n \to N(0,1)$ in distribution, which provides the following one-sided asymptotic test for $H_0$ versus $H_1$ at significance level $\alpha$: reject $H_0$ if $\bar{I}_n > z_\alpha$, where $z_\alpha$ is the upper $\alpha$-percentile of the standard normal distribution. Second, we note from the definitions of $I_n$, $\bar{I}_n$, $\rho_2$, and $\bar{\rho}_2$ that $\bar{I}_n = \bar{\rho}_2$. Hence, the two measures $I_n$ and $\rho_2$ result in the same asymptotic test. This will be referred to as the $I_n$ test, which is the test in Li and Wang (1998) and Zheng (1996).

The most important implication of Theorem 3.1 is that under the conditions of Theorem 3.1, the $I_n$ test can detect "singular" local alternatives in $LH_s$ as long as $C_1 \neq 0$. Thus, as long as the smoothing parameter $h$ of the $I_n$ test and the parameters $\gamma_n$ and $\alpha_n$ of the "singular" local alternative are such that $C_1 = \lim_{n\to\infty}(nh^{d/2}\gamma_n^2\alpha_n) \neq 0$, the corresponding test has nontrivial power against such an alternative; otherwise, if $C_1 = 0$, then the corresponding test has only trivial power.

Subsequently we demonstrate that under the conditions of Theorem 3.1, if $C_1 \neq 0$, then the $I_n$ test can detect certain "singular" local alternatives that approach the null model in probability at a rate $(\gamma_n\alpha_n)$ arbitrarily close to $n^{-3/4}$ and the ICM tests can only detect "singular" local alternatives with rate $\gamma_n\alpha_n = O(n^{-1/2})$. Hence the kernel-based tests are more powerful than the ICM tests for such cases.[2]

Consider the class of "singular" local alternatives that was used in Rosenblatt (1975), Ghosh and Huang (1991), and Fan (1994) in the context of testing goodness of fit of a density function. It is represented by $\delta_n(x) = \sum_{j=1}^q W_j[(x - l_j)/\zeta_n]$, where $q$ is a positive integer, $l_1,\dots,l_q$ are constant vectors in $R^d$, and $W_1(\cdot),\dots,W_q(\cdot)$ are bounded functions with uniformly bounded first order derivatives and satisfy $\int|W_j(x)|dx < \infty$, and $\zeta_n \to 0$ as $n \to \infty$. Obviously $\alpha_n = O(\zeta_n^d)$ and $|\delta_n(x+v) - \delta_n(x)| \le C\alpha_n^{-1/d}\|v\|$, where $C$ is a finite constant. It is easy to see that (C4) holds if $f(x)$, $\sigma^2(x)$, $\mu_4(x)$, and $M(x)$ satisfy some Lipschitz conditions. Let $h = n^{-\eta'}$ and $\zeta_n = n^{-\eta}$, where $1/(d+\epsilon) < \eta < \eta' < 1/d$ for some small positive constant $\epsilon$. Then $\alpha_n = n^{-d\eta}$. Thus $nh^d \to \infty$ and $h^d = o(\alpha_n)$ are both satisfied. Let $\gamma_n = n^{-\lambda}$, where $\lambda$ is a constant. From $nh^{d/2}\alpha_n\gamma_n^2 = C_{1n} = 1$, say, we get $n^{1-(d\eta')/2-d\eta-2\lambda} = n^0 = 1$, which leads to

$$\lambda = \{1 - (d\eta')/2 - d\eta\}/2. \tag{18}$$

Hence the rate at which the local alternative approaches the null model is $\alpha_n\gamma_n = n^{-d\eta}n^{-[1-(d\eta')/2-d\eta]/2}$, which can be arbitrarily close to $n^{-3/4}$ because $d\eta + \{1 - (d\eta')/2 - d\eta\}/2 = (\frac{1}{2}) + (d/2)[\eta - (\eta'/2)] \ge (\frac{1}{2}) + (d/2)[1/(d+\epsilon) -$

$1/(2d)] = (\frac{1}{2}) + (1-\epsilon)/\{4[1+(\epsilon/d)]\}$ can be arbitrarily close to $(\frac{1}{2}) + (\frac{1}{4}) = (\frac{3}{4})$ if $\epsilon$ is much smaller than $d$.

It is interesting to examine further the "singular" local alternatives for which the kernel-based tests have nontrivial power. These alternatives converge to the null model at rate $\alpha_n \gamma_n$. To see the role played by either $\gamma_n$ or $\alpha_n$, note the following. (i) If $\gamma_n = O(n^{-1/2})$, then $C_1 = 0$. Hence, the tests have only trivial power. (ii) If $\alpha_n = O(n^{-1/2})$, then $C_1 \neq 0$ for $\lambda = (1 - d\eta')/4 > 0$. In this case, $\gamma_n \alpha_n = o(n^{-1/2})$. (iii) If $\alpha_n = O(n^{-1+\epsilon})$, where $\epsilon$ is a small positive number, and $\gamma_n = O(n^{1/4}) \to \infty$, then $\alpha_n \gamma_n$ can be made arbitrarily close to $n^{-3/4}$. (iv) If $\lambda \geq 0$, then $\alpha_n \gamma_n$ can be made arbitrarily close to $n^{-2/3}$. Conditions (i) and (ii) show clearly that the power advantage of the kernel-based tests stems from the highly changing feature of $\delta_n(\cdot)$ or $\alpha_n$. It follows from (iii) and (iv) that the existence of a shrinking $\gamma_n$ does not help to improve the power of these tests. As a matter of fact, if we let $\gamma_n$ approach infinity, then the power of the tests against "singular" local alternatives can be made arbitrarily close to $O(n^{-3/4})$; otherwise the power can only be made arbitrarily close to $O(n^{-2/3})$.

We now study the power properties of ICM tests for the class of "singular" local alternatives in (16). It follows from Bierens and Ploberger (1997) that under their Assumption A.5, the ICM tests have nontrivial power if and only if there exists a continuous function $\eta(\xi)$ on $\Xi$ such that

$$\int_{\Xi} \eta^2(\xi) d\mu(\xi) > 0 \tag{19}$$

and

$$\frac{\gamma_n}{\sqrt{n}} \sum_j \delta_n(X_j)\phi_j(\xi) \to \eta(\xi) \quad \text{in probability uniformly on } \Xi,$$

where

$$\phi_j(\xi) \equiv \phi(X_j,\xi) = w_j(\xi) - b'(\beta_0,\xi)A(\beta_0)^{-1}\nabla g_0(X_j,\beta_0)$$

and $b(\beta_0,\xi)$ is the uniform probability limit of $n^{-1}\sum_j \nabla' g_0(X_j,\beta_0)w_j(\xi)$ on $\mathcal{B} \times \Xi$ (see Bierens and Ploberger, 1997, Assumption A.5). For the preceding class of "singular" local alternatives, one can easily show that under general conditions, $(\gamma_n/\sqrt{n})\sum_j \delta_n(X_j)\phi_j(\xi) = O_p(\sqrt{n}\gamma_n\alpha_n) = o_p(1)$ if $\alpha_n\gamma_n = o(n^{-1/2})$. Thus, condition (19) is violated if $\gamma_n\alpha_n = o(n^{-1/2})$. Therefore, the ICM tests have only trivial power against "singular" local alternatives that approach the null model in probability at a rate of smaller order than $n^{-1/2}$.

We note that the previously demonstrated power advantage of the kernel-based test over the ICM test depends on the condition that $C_1 \neq 0$. If $C_1 = 0$, then the kernel-based test has only trivial power against "singular" local alternatives $LH_s$, whereas the ICM test can still have nontrivial power against such alternatives for which $\gamma_n\alpha_n = O(n^{-1/2})$.

To summarize the results of this section, we have shown that (against "singular" local alternatives) (i) when $C_1 \neq 0$ and $\gamma_n \alpha_n = o(n^{-1/2})$, the kernel test has nontrivial power, whereas the ICM test has only trivial power; (ii) when $C_1 = 0$ and $\gamma_n \alpha_n = O(n^{-1/2})$, the kernel test has only trivial power, whereas the ICM test can have nontrivial power. Our simulations results reported in the next section illustrate that the previous asymptotic results have finite sample implications.

## 4. MONTE CARLO RESULTS

In this section we report results from a simulation study carried out to compare the finite sample performances of the kernel-based test $I_n$ and the ICM tests. We first compare the power performances of these tests based on (size adjusted) empirical critical values. The empirical critical values are generated via 5,000 replications of the null model. Because in practice, empirical critical values are not available, we also report results obtained by using a "wild bootstrap" procedure to approximate the asymptotic null distributions of these test statistics. For the detailed steps of the wild bootstrap procedure for the $I_n$ test, and also a proof that the wild bootstrap works, see Li and Wang (1998).

We compare the $I_n$ test with four ICM tests: two with the complex-valued exponential weight function and two with the real-valued exponential weight function. Given the equivalency between ICM tests with the complex-valued exponential weight function and the $I_n$ test with a fixed value of $h$, we use the $I_n$ test given in (6) with a fixed value of $h$ to represent a special class of the ICM tests. In our simulation experiments, we choose two values for $h$ and denote the resulting tests by $J_n$ and $\tilde{J}_n$, respectively. The ICM test with the real-valued exponential weight function and probability measure $\mu(\cdot)$ is given by

$$\rho = \frac{1}{n} \sum_j \sum_l \int \tilde{u}_j \tilde{u}_l \exp(\xi'(Z_j + Z_l))\, d\mu(\xi),$$

where we use two different measures for $\mu(\cdot)$: (i) a product uniform $[-b,b]$ distribution and (ii) a product normal $N(0,a^2)$ distribution. The resulting two tests are

$$\rho_{(i)} = \frac{1}{n} \sum_j \sum_l \tilde{u}_j \tilde{u}_l \prod_{s=1}^{d} \left\{ \frac{1}{(Z_{sj} + Z_{sl})} [\exp(b(Z_{sj} + Z_{sl})) - \exp(-b(Z_{sj} + Z_{sl}))] \right\}$$

and

$$\rho_{(ii)} = \frac{1}{n} \sum_j \sum_l \tilde{u}_j \tilde{u}_l \prod_{s=1}^{d} \exp((Z_{sj} + Z_{sl})^2/2),$$

where $Z_{sj} = \tan^{-1}((X_{sj} - \bar{X}_s)/X_{s,sd})$ $(s = 1,\ldots,d)$ and $\bar{X}_s$ and $X_{s,sd}$ are the mean value and standard deviation of $\{X_{sj}\}_{j=1}^n$ with $X_{sj}$ the $s$th component of $X_j$. We note that $\rho_{(ii)}$ is invariant to different values of $a$.

We take as the null model a linear regression model with two regressors (see DGP1, which follows). For $j = 1,\dots,n$, let $V_j$, $V_{1j}$, and $V_{2j}$ be independent random drawings from the uniform $[-\pi,\pi]$ distribution and $\epsilon_j$ be independent random drawings from a standard normal distribution. The regressors are given by $X_{1j} = V_j + V_{1j}$ and $X_{2j} = V_j + V_{2j}$ for $j = 1,\dots,n$. The following data generating processes (DGP) are used in our simulation experiments:

$$DGP1: Y_j = \alpha_0 + \alpha_1 X_{1j} + \alpha_2 X_{2j} + \epsilon_j,$$

$$DGP2: Y_j = \alpha_0 + \alpha_1 X_{1j} + \alpha_2 X_{2j} + 0.15 V_{1j} V_{2j} + \epsilon_j,$$

$$DGP3: Y_j = \alpha_0 + \alpha_1 X_{1j} + \alpha_2 X_{2j} + (0.2/\sqrt{n})(X_{1j}^2 + X_{2j}^2) + \epsilon_j,$$

$$\begin{aligned} DGP4: Y_j = {} & \alpha_0 + \alpha_1 X_{1j} + \alpha_2 X_{2j} \\ & + 10 n^{-\lambda} \sin(X_{1j}/\zeta_n)\sin(X_{2j}/\zeta_n) I(|X_{1j}| \le \zeta_n) I(|X_{2j}| \le \zeta_n) + \epsilon_j, \end{aligned}$$

$$DGP5: Y_j = \alpha_0 + \alpha_1 X_{1j} + \alpha_2 X_{2j} + 2\sin(mX_{1j})\sin(mX_{2j}) + \epsilon_j,$$

where DGP1 is the null model; DGP2 corresponds to a fixed alternative; DGP3 is a Pitman local alternative that approaches the null at rate $O_p(n^{-1/2})$; DGP4 is a singular local alternative with $\zeta_n = n^{-1/10}$ and $\lambda$ satisfying (18) ($h \sim n^{-1/6}$): $\lambda = [1 - (d\eta'/2) - d\eta]/2 = [1 - (\frac{1}{6}) - (\frac{1}{5})]/2 = \frac{19}{60}$; and DGP5 is a fixed alternative that represents high (low) frequency alternative for small (large) values of $m$. Both DGP1 and DGP2 are taken from Bierens (1990).

In the simulation experiments, we chose $(\alpha_0,\alpha_1,\alpha_2) = (1,1,1)$. The sample sizes are $n = 50,100,200$ ($n = 400$ is used for DGP3). We used a product (standard) normal kernel with $h_s = c_0 X_{s,sd} n^{-1/6}$ ($s = 1,2$) for the $I_n$ test and $h_s = X_{s,sd}$ ($s = 1,2$) for the $J_n$ test. The $J_n$ test is the same as the $I_n$ test except that $h_s = X_{s,sd}$, which does not go to zero as $n \to \infty$. We also computed another fixed $h$ test $\tilde{J}_n$, where $\tilde{J}_n$ is obtained from (6) with $h = 1$ and $K((X_j - X_l)/h)$ in (6) is replaced by $K(Z_j - Z_l)$ (because $h = 1$) with $Z_{sj} = \tan^{-1}((X_{sj} - \bar{X}_s)/X_{s,sd})$. For the $\rho_{(i)}$ test, we used $b = 0.5,1,2$ and found that the results for different values of $b$ are almost identical. Hence we only report the case of $b = 1$.

**TABLE 1.** Estimated power (DGP2)

| | $n = 50$ | | | $n = 100$ | | | $n = 200$ | | |
|---|---|---|---|---|---|---|---|---|---|
| Tests | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% |
| $I_n$ | 0.089 | 0.246 | 0.368 | 0.290 | 0.566 | 0.682 | 0.753 | 0.895 | 0.952 |
| $J_n$ | 0.053 | 0.203 | 0.336 | 0.173 | 0.491 | 0.655 | 0.660 | 0.908 | 0.962 |
| $\tilde{J}_n$ | 0.061 | 0.221 | 0.347 | 0.221 | 0.501 | 0.648 | 0.613 | 0.883 | 0.946 |
| $\rho_{(i)}$ | 0.137 | 0.318 | 0.457 | 0.342 | 0.596 | 0.719 | 0.685 | 0.901 | 0.958 |
| $\rho_{(ii)}$ | 0.103 | 0.235 | 0.355 | 0.254 | 0.468 | 0.605 | 0.538 | 0.768 | 0.876 |

**TABLE 2.** Estimated power (DGP3)

| | $n = 50$ | | | $n = 100$ | | | $n = 200$ | | | $n = 400$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Tests | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% |
| $I_n$ | 0.133 | 0.284 | 0.387 | 0.119 | 0.287 | 0.380 | 0.105 | 0.263 | 0.367 | 0.098 | 0.253 | 0.369 |
| $J_n$ | 0.238 | 0.426 | 0.538 | 0.248 | 0.459 | 0.566 | 0.249 | 0.476 | 0.596 | 0.293 | 0.512 | 0.613 |
| $\tilde{J}_n$ | 0.207 | 0.391 | 0.511 | 0.220 | 0.434 | 0.536 | 0.203 | 0.450 | 0.565 | 0.275 | 0.463 | 0.596 |
| $\rho_{(i)}$ | 0.269 | 0.475 | 0.579 | 0.278 | 0.496 | 0.618 | 0.256 | 0.506 | 0.634 | 0.294 | 0.537 | 0.651 |
| $\rho_{(ii)}$ | 0.270 | 0.464 | 0.580 | 0.286 | 0.504 | 0.619 | 0.291 | 0.519 | 0.649 | 0.315 | 0.552 | 0.668 |

**TABLE 3.** Estimated power (DGP4)

| | $n = 50$ | | | $n = 100$ | | | $n = 200$ | | |
|---|---|---|---|---|---|---|---|---|---|
| Tests | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% |
| $I_n$ | 0.064 | 0.245 | 0.380 | 0.123 | 0.384 | 0.533 | 0.259 | 0.515 | 0.694 |
| $J_n$ | 0.011 | 0.073 | 0.161 | 0.009 | 0.080 | 0.170 | 0.013 | 0.081 | 0.159 |
| $\tilde{J}_n$ | 0.021 | 0.126 | 0.270 | 0.025 | 0.121 | 0.232 | 0.018 | 0.112 | 0.208 |
| $\rho_{(i)}$ | 0.043 | 0.142 | 0.230 | 0.028 | 0.100 | 0.169 | 0.018 | 0.085 | 0.150 |
| $\rho_{(ii)}$ | 0.030 | 0.091 | 0.163 | 0.017 | 0.067 | 0.126 | 0.013 | 0.064 | 0.122 |

Because $\rho_{(ii)}$ is invariant to different values of $a$, we chose $a = 1$ for the $\rho_{(ii)}$ test. The empirical 1%, 5%, and 10% critical values are obtained from 5,000 random samples of size $n$ from the null model (DGP1). Using these empirical critical values, we estimate the power of different tests for DGP2–DGP5 via 2,000 replications. The estimated powers based on empirical critical values are given in Tables 1–5.

Table 1 reports the power performances of different tests for DGP2. In general all the tests have similar power against DGP2, with $\rho_{(i)}$ being slightly more powerful than the other tests in most cases. It should be emphasized that we have only used the exponential weight function for ICM tests in Bierens and Ploberger (1997) in our simulations. The optimal choice for the weight function in the ICM tests does not seem available in the literature.

Table 2 gives the estimated power for a Pitman local alternative (DGP3) that departs from the null at a rate of $n^{-1/2}$. We observe that the rejection rate for the $I_n$ test decreases slightly as $n$ increases, which is consistent with the theoretical result because $I_n$ has only trivial power asymptotically against DGP3. In contrast to the $I_n$ test, the other four tests, $J_n$, $\tilde{J}_n$, $\rho_{(i)}$, and $\rho_{(ii)}$, have moderate power, and the power improves slightly as $n$ increases. For DGP3, the $\rho_{(ii)}$ test is the most powerful test.

**TABLE 4.** Estimated power (DGP5: $m = \frac{1}{4}$)

| | $n = 50$ | | | $n = 100$ | | | $n = 200$ | | |
|---|---|---|---|---|---|---|---|---|---|
| Tests | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% |
| $I_n$ | 0.391 | 0.604 | 0.697 | 0.757 | 0.910 | 0.943 | 0.990 | 0.998 | 1.000 |
| $J_n$ | 0.570 | 0.757 | 0.842 | 0.924 | 0.980 | 0.990 | 1.000 | 1.000 | 1.000 |
| $\tilde{J}_n$ | 0.689 | 0.851 | 0.913 | 0.967 | 0.994 | 0.998 | 1.000 | 1.000 | 1.000 |
| $\rho_{(i)}$ | 0.798 | 0.926 | 0.962 | 0.991 | 0.999 | 0.999 | 1.000 | 1.000 | 1.000 |
| $\rho_{(ii)}$ | 0.719 | 0.871 | 0.933 | 0.969 | 0.996 | 0.998 | 1.000 | 1.000 | 1.000 |

**TABLE 5.** Estimated power (DGP5: $m = 1$)

| | $n = 50$ | | | $n = 100$ | | | $n = 200$ | | |
|---|---|---|---|---|---|---|---|---|---|
| Tests | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% |
| $I_n$ | 0.311 | 0.613 | 0.752 | 0.928 | 0.993 | 0.998 | 1.000 | 1.000 | 1.000 |
| $J_n$ | 0.027 | 0.133 | 0.241 | 0.051 | 0.263 | 0.439 | 0.255 | 0.657 | 0.811 |
| $\tilde{J}_n$ | 0.035 | 0.112 | 0.210 | 0.060 | 0.234 | 0.343 | 0.183 | 0.464 | 0.607 |
| $\rho_{(i)}$ | 0.131 | 0.281 | 0.407 | 0.207 | 0.392 | 0.508 | 0.335 | 0.568 | 0.684 |
| $\rho_{(ii)}$ | 0.121 | 0.237 | 0.346 | 0.175 | 0.351 | 0.463 | 0.316 | 0.483 | 0.602 |

Table 3 summarizes the results for DGP4, a "singular" local alternative. The results are in agreement with our theoretical local power analysis: the $I_n$ test has nontrivial power against DGP4, whereas the powers of the remaining four tests decrease as the size of $n$ increases. For example, at the 5% significance level and $n = 200$, the rejection rate of the $I_n$ test is 51.5%, whereas the largest rejection rate of the remaining four tests is 11.2%.

Tables 4 and 5 present rejection rates for DGP5, a fixed low (high) frequency alternative. As expected, for the low frequency ($m = \frac{1}{4}$) alternative, $I_n$ is the least powerful test, whereas for the high frequency ($m = 1$) alternative, $I_n$ is the most powerful test.

Summarizing Tables 2–5, we conclude that our Monte Carlo results are in agreement with theoretical findings on local power properties of kernel-based tests versus ICM tests provided in previous work and also in this paper: For Pitman local alternatives, the ICM tests tend to be more powerful than the kernel tests, whereas for "singular" local alternatives, the kernel tests are more powerful. More important, these results carry over to the fixed alternatives considered in this section.

**TABLE 6.** Estimated size (DGP1)

| | $n = 50$ | | | $n = 100$ | | | $n = 200$ | | |
|---|---|---|---|---|---|---|---|---|---|
| Tests | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% |
| $I_n$ | 0.012 | 0.050 | 0.094 | 0.013 | 0.052 | 0.106 | 0.012 | 0.049 | 0.098 |
| $J_n$ | 0.007 | 0.052 | 0.097 | 0.014 | 0.049 | 0.102 | 0.011 | 0.044 | 0.099 |
| $\tilde{J}_n$ | 0.007 | 0.054 | 0.108 | 0.012 | 0.051 | 0.110 | 0.012 | 0.040 | 0.080 |
| $\rho_{(i)}$ | 0.011 | 0.054 | 0.113 | 0.019 | 0.065 | 0.117 | 0.016 | 0.052 | 0.104 |
| $\rho_{(ii)}$ | 0.006 | 0.057 | 0.102 | 0.007 | 0.051 | 0.108 | 0.008 | 0.050 | 0.108 |

**TABLE 7.** Estimated power (DGP2)

| | $n = 50$ | | | $n = 100$ | | | $n = 200$ | | |
|---|---|---|---|---|---|---|---|---|---|
| Tests | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% |
| $I_n$ | 0.077 | 0.234 | 0.366 | 0.320 | 0.556 | 0.684 | 0.770 | 0.902 | 0.936 |
| $J_n$ | 0.048 | 0.190 | 0.329 | 0.190 | 0.508 | 0.660 | 0.670 | 0.892 | 0.950 |
| $\tilde{J}_n$ | 0.077 | 0.251 | 0.378 | 0.222 | 0.516 | 0.674 | 0.608 | 0.870 | 0.935 |
| $\rho_{(i)}$ | 0.011 | 0.294 | 0.432 | 0.312 | 0.556 | 0.682 | 0.653 | 0.893 | 0.948 |
| $\rho_{(ii)}$ | 0.086 | 0.216 | 0.349 | 0.248 | 0.438 | 0.596 | 0.526 | 0.769 | 0.862 |

Finally, we repeated the preceding experiments by replacing the empirical critical values with the wild bootstrap critical values. We computed the size and power for all five tests for DGP1 and DGP2. The number of replications is 1,000, and with each replication, 1,000 wild bootstrap statistics were generated to produce bootstrap critical values. The estimated sizes based on wild bootstrap are given in Table 6, and the estimated powers for DGP2 using wild bootstrap critical values are reported in Table 7. We observe from Table 6 that the estimated sizes (for all the tests) are quite close to the corresponding nominal sizes, indicating that the wild bootstrap procedure approximates the null distribution of the test statistics quite well. Also as expected, the results in Table 7 are quite similar to those of Table 2.

## NOTES

1. Throughout this paper, the rate of convergence of a "singular" local alternative to the null model always refers to the rate of convergence in probability. Or alternatively, one can interpret this as using the $L_1$ norm to measure the difference between the "singular" local alternative and the null model. For discussion about using other norms to measure the difference, see note 2.

2. We emphasize again that the rate of convergence of a "singular" local alternative to the null model refers to the rate of convergence in probability. This is given by $\gamma_n \alpha_n$. Or alternatively one can interpret this as using the $L_1$ norm to measure the difference between the "singular" alternative and the null model. The preceding claimed rates will be different if one uses a different norm such as the $L_2$ norm to measure the difference. However, it does not matter which norm one uses; this section shows that as long as $\gamma_n \alpha_n = o(n^{-1/2})$, the ICM tests have only trivial power and the kernel test $I_n$ can have nontrivial power even for $\gamma_n \alpha_n$ arbitrarily close to $n^{-3/4}$.

## REFERENCES

Ahmad, I.A. (1982) Nonparametric estimation of the location and scale parameters based on density estimation. *Annals of the Institute of Statistical Mathematics* 34, 39–53.

Aït-Sahalia, Y., P.J. Bickel, & 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. & W. Ploberger (1997) Asymptotic theory of integrated conditional moment tests. *Econometrica* 65, 1129–1152.

Chen, X. & 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. (1993) Testing the equality of nonparametric curves. *Statistics and Probability Letters* 17, 199–204.

Fan, Y. (1994) Testing the goodness-of-fit of a parametric density function by kernel method. *Econometric Theory* 10 (2), 316–356.

Fan, Y. & Q. Li (1992) A General Nonparametric Model Specification Test. Manuscript, Department of Economics, University of Windsor.

Fan, Y. & Q. Li (1996) Consistent model specification tests: Omitted variables and semiparametric functional forms. *Econometrica* 64, 865–890.

Fan, Y. & 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.

Ghosh, B.K. & Wei-Min Huang (1991) The power and optimal kernel of the Bickel-Rosenblatt test for goodness-of-fit. *Annals of Statistics* 19 (2), 999–1009.

Gozalo, P.L. (1993) A consistent model specification test for nonparametric estimation of regression function models. *Econometric Theory* 9, 451–477.

Hansen, B.E. (1996) Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64, 413–430.

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

Hong, Y. (1993) Consistent Specification Testing Using Optimal Nonparametric Kernel Estimation. Working paper, Cornell University.

Hong, Y. & H. White (1995) Consistent specification testing via nonparametric series regression. *Econometrica* 63, 1133–1160.

Horowitz, J.T. & W. Härdle (1994) Testing a parametric model against a semiparametric alternative. *Econometric Theory* 10, 821–848.

Horowitz, J.T. & V.G. Spokoiny (1999) An adaptive, rate-optimal test of a parametric model against a nonparametric alternative. *Econometrica*, forthcoming.

Ingster, Y.I. (1982) Maximum nonparametric detection of signals in White Gaussian noise. *Problems of Information Transmission* 18, 130–134.

Ingster, Y.I. (1993a) Asymptotic minimax hypothesis testing for nonparametric alternatives, part I. *Mathematical Methods of Statistics* 2, 85–114.

Ingster, Y.I. (1993b) Asymptotic minimax hypothesis testing for nonparametric alternatives, part II. *Mathematical Methods of Statistics* 2, 171–189.

Ingster, Y.I. (1993c) Asymptotic minimax hypothesis testing for nonparametric alternatives, part III. *Mathematical Methods of Statistics* 2, 249–268.

Lavergne, P. & Q. Vuong (1996) Nonparametric selection of regressors: The nonnested case. *Econometrica* 64, 207–219.

Lavergne, P. & Q. Vuong (2000) A nonparametric significance test. *Econometric Theory* 16, 576–601.

Li, Q. (1994) A Consistent Test for Linearity in Partially Linear Regression Models. Discussion paper 1994-7, Department of Economics, University of Guelph.

Li, Q. (1999) Consistent model specification tests for time series econometric models. *Journal of Econometrics* 92, 101–147.

Li, Q. & S. Wang (1998) A simple consistent bootstrap test for a parametric regression function. *Journal of Econometrics* 87, 145–165.

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

Ploberger, W. & H.J. Bierens (1995) Asymptotic Power of the Integrated Conditional Moment Test against Large Local Alternatives. Working paper, Department of Economics, Southern Methodist University.

Robinson, P.M. (1989) Hypothesis testing in semiparametric and nonparametric models for econometric time series. *Review of Economic Studies* 56, 511–534.

Rosenblatt, M. (1975) A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Annals of Statistics* 3, 1–14.

Spokoiny, V.G. (1996) Adaptive hypothesis testing using wavelets. *Annals of Statistics* 24, 2477–2498.

Stinchcombe, M.B. & H. White (1998) Consistent specification testing with unidentified nuisance parameters using duality and Banach space limit theory. *Econometric Theory* 14, 295–324.

Tauchen, G. (1985) Diagnostic testing and evaluation of maximum likelihood models. *Journal of Econometrics* 2, 187–209.

Whang, Y.-J. (1998) A Consistent Kolmogorov Test of Parametric Regression Functions. Manuscript, Department of Economics, Ewha University, Seoul, Korea.

Whang, Y.J. & D.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.

Zheng, J.X. (1996) A consistent test of functional form via nonparametric estimation technique. *Journal of Econometrics* 75, 263–289.

# APPENDIX: PROOF OF THEOREM 3.1

We will only provide the proof for Theorem 3.1(i); (ii) follows immediately from (i), and (iii) follows from (i), (ii), and the fact that $\hat{\sigma}_0^2 = \sigma_0^2 + o_p(1)$ (e.g., Ahmad, 1982).

From (6), $\tilde{u}_j = Y_j - \tilde{g}(X_j)$, and (16), it follows that

$$\begin{aligned} I_n &= \frac{1}{n(n-1)h^d} \sum_l \sum_{j \neq l} K_{lj} \{[g_0(X_l,\beta_0) - \tilde{g}(X_l) + \gamma_n \delta_n(X_l) + u_l] \\ &\qquad \times [g_0(X_j,\beta_0) - \tilde{g}(X_j) + \gamma_n \delta_n(X_j) + u_j]\} \\ &= \frac{1}{n(n-1)h^d} \sum_l \sum_{j \neq l} K_{lj} \{[g_0(X_l,\beta_0) - \tilde{g}(X_l)][g_0(X_j,\beta_0) - \tilde{g}(X_j)] \\ &\qquad + \gamma_n^2 \delta_n(X_l)\delta_n(X_j) + u_l u_j + 2[g_0(X_l,\beta_0) - \tilde{g}(X_l)]\gamma_n \delta_n(X_j) \\ &\qquad + 2[g_0(X_l,\beta_0) - \tilde{g}(X_l)]u_j + 2\gamma_n \delta_n(X_l) u_j\} \\ &\equiv A_1 + A_2 + A_3 + 2A_4 + 2A_5 + 2A_6. \end{aligned} \tag{A.1}$$

We will prove Theorem 3.1 by evaluating the order of $A_j$ for $j = 1,\ldots,6$.

(i) Proof of $A_1 = O_p(n^{-1})$. From (A.1) and the Taylor series expansion, we get

$$A_1 = (\tilde{\beta} - \beta_0)' \left\{ \frac{1}{n(n-1)h^d} \sum_l \sum_{j \neq l} K_{lj} [\nabla g_0(X_l, \beta_*)]' [\nabla g_0(X_j, \beta_*)] \right\} (\tilde{\beta} - \beta_0)$$

$$\equiv (\tilde{\beta} - \beta_0)' B_1 (\tilde{\beta} - \beta_0),$$

where $\beta_*$ is on the line segment between $\tilde{\beta}$ and $\beta_0$. It is easy to show that given (C1), $E|B_1| = O(1)$. Hence $A_1 = O_p(n^{-1})$ because $(\tilde{\beta} - \beta_0) = O_p(n^{-1/2})$.

(ii) Proof of $nh^{d/2} A_2 = C_1 C_2 + o_p(1)$. We first evaluate $E(A_2)$: From (A.1) and (C4)(ii)(a) and (f), we have

$$E(A_2) = \frac{\gamma_n^2}{n(n-1)h^d} \sum_l \sum_{j \neq i} E[K_{lj} \delta_n(X_l) \delta_n(X_j)] = \frac{\gamma_n^2}{h^d} E[K_{12} \delta_n(X_1) \delta_n(X_2)]$$

$$= \gamma_n^2 \iint f(x) \delta_n(x) f(x + hu) \delta_n(x + hu) K(u) \, du dx$$

$$\leq \gamma_n^2 \left\{ \int f^2(x) \delta_n^2(x) dx + h \alpha_n^{-1/d} \left[ \int f(x) |\delta_n(x) G(x)| dx \right] \times \left[ \int K(u) \|u\| du \right] \right\}$$

$$= C_{2n} \gamma_n^2 \alpha_n + h \alpha_n^{-1/d} O(\gamma_n^2 \alpha_n).$$

Thus, $E(A_2) = \gamma_n^2 \alpha_n \{C_2 + o(1)\}$. Next,

$$\mathrm{Var}(nh^{d/2} A_2) = \frac{\gamma_n^4}{(n-1)^2 h^d} \sum_l \sum_{j \neq l} \sum_{l'} \sum_{j' \neq l'} \mathrm{Cov}\{K_{lj} \delta_n(X_l) \delta_n(X_j), K_{l'j'} \delta_n(X_{l'}) \delta_n(X_{j'})\}$$

$$\equiv \frac{\gamma_n^4}{(n-1)^2 h^d} \sum_l \sum_{j \neq l} \sum_{l'} \sum_{j' \neq l'} \{A_{lj,l'j'}\}$$

$$= \frac{4\gamma_n^4}{(n-1)^2 h^d} \left\{ 2 \sum_l \sum_{j<l} \sum_{l' \neq l, l' < j} A_{lj,l'j} + \sum_l \sum_{j<l} A_{lj,lj} \right\}$$

$$= \frac{4\gamma_n^4}{(n-1)^2 h^d} \{ O(n^3 h^{2d} \alpha_n) + O(n^2 h^d \alpha_n) \}$$

$$= O(\gamma_n^4 \alpha_n n h^d) = O((nh^{d/2} \gamma_n^2 \alpha_n)^2 (nh^d)^{-1} (h^d \alpha_n^{-1}))$$

$$= o(1), \quad \text{by (C3)}$$

because $A_{lj,l'j'} = 0$ if the four indices $l, j, l', j'$ are all different from each other and $A_{12,32} = E\{[K_{12} \delta_n(X_1) \delta_n(X_2) - E(K_{12} \delta_n(X_1) \delta_n(X_2))][K_{32} \delta_n(X_3) \delta_n(X_2) - E(K_{32} \delta_n(X_3) \delta_n(X_2))]\}$ has the same order as

$$E\{[K_{12}K_{32}\delta_n(X_1)\delta_n^2(X_2)\delta_n(X_3)]$$

$$= h^{2d}\iiint f(x)\delta_n^2(x)\delta_n(x+hu)\delta_n(x+hv)f(x+hu)$$

$$\times f(x+hv)K(u)K(v)\,dudvdx$$

$$\le h^{2d}\left\{\int f^3(x)\delta_n^4(x)\,dx + 2h\alpha_n^{-1/d}\left[\int f^2(x)|\delta_n^3(x)G(x)|dx\right]\right.$$

$$\times\left[\int K(u)\|u\|du\right]$$

$$\left. + h^2\alpha_n^{-2/d}\left[\int f(x)\delta_n^2(x)G^2(x)\,dx\right]\left[\int K(u)\|u\|du\right]^2\right\}$$

$$= O(h^{2d}\alpha_n), \quad \text{by (C4).}$$

Similarly one can easily show that $A_{12,12} = O(h^d\alpha_n)$. Hence $A_2 = \gamma_n^2\alpha_n\{C_2 + o_p(1)\}$ and $nh^{d/2}A_2 = nh^{d/2}\gamma_n^2\alpha_n\{C_2 + o_p(1)\} = C_1C_2 + o_p(1)$.

(iii) Proof of $nh^{d/2}A_3 \to N(0,\sigma_0^2)$ in distribution. When $\mathcal{S} = R^d$, this was proved in Corollary 3 of Li (1994) and Lemma 3.3 of Zheng (1996). When $\mathcal{S}$ is a compact, convex subset of $R^d$, this was proved by Li and Wang (1998).

(iv) Proof of $nh^{d/2}A_4 = nh^{d/2}O_p(n^{-1/2}\gamma_n\alpha_n) = o_p(1)$. From (A.1) and the Taylor series expansion, we obtain

$$A_4 = (\tilde\beta - \beta_0)'\left\{\frac{\gamma_n}{n(n-1)h^d}\sum_l\sum_{j\ne l}K_{lj}\nabla g_0(X_l,\beta_*)\delta_n(X_j)\right\}$$

$$\equiv (\tilde\beta - \beta_0)'B_4.$$

By (C1), $|\nabla g(X_j,\beta_*)| \le M(X_j)$, one can show that

$$E|B_4| \le \frac{\gamma_n}{n(n-1)h^d}\sum_l\sum_{j\ne l}E|M(X_l)\delta_n(X_j)K_{lj}| = \frac{\gamma_n}{h^d}E|M(X_1)\delta_n(X_2)K_{12}|$$

$$= \gamma_n\int f(x)|\delta_n(x)f(x+hu)M(x+hu)|K(u)dudx$$

$$\le \gamma_n\left\{\int f^2(x)|M(x)\delta_n(x)|dx + h\alpha_n^{-1/d}\int f(x)|\delta_n(x)G(x)|dx\right.$$

$$\left.\times\left[\int K(u)\|u\|du\right]\right\}$$

$$= O(\gamma_n\alpha_n).$$

Hence $A_4 = O_p(n^{-1/2}\gamma_n\alpha_n)$ because $\tilde\beta - \beta_0 = O_p(n^{-1/2})$.

(v) Proof of $A_5 = O_p(n^{-1})$. Again, applying Taylor series expansion to the fifth term on the right hand side of (A.1) yields

$$
\begin{aligned}
A_5 &= (\tilde{\beta}-\beta_0)'\left\{\frac{1}{n(n-1)h^d}\sum_l\sum_{j\neq l} K_{lj}\nabla g_0(X_l,\beta_0)u_j\right\}\\
&\quad + (\tilde{\beta}-\beta_0)'\left\{\frac{1}{n(n-1)h^d}\sum_l\sum_{j\neq l} K_{lj}\nabla^2 g_0(X_l,\beta_*)u_j\right\}(\tilde{\beta}-\beta_0)\\
&\equiv (\tilde{\beta}-\beta_0)'B_5 + (\tilde{\beta}-\beta_0)'D_5(\tilde{\beta}-\beta_0).
\end{aligned}
$$

We now show that $B_5 = O_p(n^{-1/2})$ and $D_5 = O_p(1)$. First, note that $E(B_5) = 0$ and

$$
\begin{aligned}
E[(B_5)^2] &= \frac{1}{n^2(n-1)^2h^{2d}}\\
&\quad \times \{n(n-1)(n-2)E[K_{12}K_{23}\nabla g_0(X_1,\beta_0)(\nabla g_0(X_3,\beta_0))'\sigma^2(X_3)]\\
&\quad\quad + 2n(n-1)E[K_{12}^2\nabla g_0(X_1,\beta_0)(\nabla g_0(X_1,\beta_0))'\sigma^2(X_2)]\}\\
&= \frac{1}{n^2(n-1)^2h^{2d}}\{O(n^3h^{2d}) + O(n^2h^d)\} = O(n^{-1}).
\end{aligned}
$$

Hence $B_5 = O_p(n^{-1/2})$. Obviously given (C1), $E|D_5| = O(1)$. Thus $A_5 = O_p(n^{-1})$ because $\tilde{\beta} - \beta_0 = O_p(n^{-1/2})$.

(vi) Proof of $nh^{d/2}A_6 = nh^{d/2}O_p(n^{-1/2}\alpha_n^{1/2}\gamma_n) = o_p(1)$. Note that from (A.1), it follows that $E(A_6) = 0$ and

$$
\begin{aligned}
E[(A_6)^2] &= \frac{\gamma_n^2}{n^2(n-1)^2h^{2d}}\sum_l\sum_{j\neq l}\sum_{l'\neq j} E\{K_{lj}K_{l'j}\delta_n(X_l)\delta_n(X_{l'})u_j^2\}\\
&= \frac{\gamma_n^2}{n^2(n-1)^2h^{2d}}\{n(n-1)E[K_{12}^2\delta_n^2(X_1)\sigma^2(X_2)]\\
&\quad + n(n-1)(n-2)E[K_{12}K_{32}\delta_n(X_1)\delta_n(X_3)\sigma^2(X_2)]\}\\
&= \frac{\gamma_n^2}{n^2(n-1)^2h^{2d}}\{O(n^2h^d\alpha_n) + O(n^3h^{2d}\alpha_n)\}\\
&= O(n^{-1}\gamma_n^2\alpha_n) = o((n^2h^d)^{-1}),
\end{aligned}
$$

because

$$
\begin{aligned}
&E[K_{12}^2\delta_n^2(X_1)\sigma^2(X_2)]\\
&\quad = h^d\iint f(x)\delta_n^2(x)f(x+hu)\sigma^2(x+hu)K^2(u)\,du\,dx\\
&\quad \leq h^d\left\{\int f^2(x)\delta_n^2(x)\sigma^2(x)\,dx\left[\int K^2(u)du\right]\right.\\
&\quad\quad \left. + h\alpha_n^{-1/d}\int f(x)\delta_n^2(x)|G(x)|dx\left[\int \|u\|K^2(u)\,du\right]\right\}\\
&\quad = O(h^d\alpha_n)
\end{aligned}
$$

and

$$
\begin{aligned}
&E[K_{12}K_{32}\delta_n(X_1)\delta_n(X_3)\sigma^2(X_2)] \\
&\quad = h^{2d}\iiint f(x)\delta_n(x)f(x+hu)f(x+hv)\delta_n(x+hv)\sigma^2(x+hu)K(u)K(v)\,dudvdx \\
&\quad \le h^{2d}\Bigg\{\int f^3(x)\delta_n^2(x)\sigma^2(x)\,dx + h^2\alpha_n^{-2/d}\int f(x)|\delta_n(x)|G^2(x)dx \\
&\qquad\qquad \times\left[\int K(u)\|u\|du\right]^2 \\
&\qquad\qquad + h\alpha_n^{-1/d}\left[\int f^2(x)\delta_n^2(x)|G(x)|dx + \int f^2(x)|\delta_n(x)|\sigma^2(x)|G(x)|dx\right] \\
&\qquad\qquad \times\left[\int K(u)\|u\|du\right]\Bigg\} \\
&\quad = O(h^{2d}\alpha_n). \qquad \blacksquare
\end{aligned}
$$

# LINKED CITATIONS





## References

**A Conditional Kolmogorov Test**
Donald W. K. Andrews
*Econometrica*, Vol. 65, No. 5. (Sep., 1997), pp. 1097-1128.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199709%2965%3A5%3C1097%3AACKT%3E2.0.CO%3B2-X

**A Consistent Conditional Moment Test of Functional Form**
Herman J. Bierens
*Econometrica*, Vol. 58, No. 6. (Nov., 1990), pp. 1443-1458.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199011%2958%3A6%3C1443%3AACCMTO%3E2.0.CO%3B2-E

**Asymptotic Theory of Integrated Conditional Moment Tests**
Herman J. Bierens; Werner Ploberger
*Econometrica*, Vol. 65, No. 5. (Sep., 1997), pp. 1129-1151.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199709%2965%3A5%3C1129%3AATOICM%3E2.0.CO%3B2-P

**Testing the Goodness of Fit of a Parametric Density Function by Kernel Method**
Yanqin Fan
*Econometric Theory*, Vol. 10, No. 2. (Jun., 1994), pp. 316-356.
Stable URL:
http://links.jstor.org/sici?sici=0266-4666%28199406%2910%3A2%3C316%3ATTGOFO%3E2.0.CO%3B2-Q

# LINKED CITATIONS



**Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms**
Yanqin Fan; Qi Li
*Econometrica*, Vol. 64, No. 4. (Jul., 1996), pp. 865-890.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199607%2964%3A4%3C865%3ACMSTOV%3E2.0.CO%3B2-1

**The Power and Optimal Kernel of the Bickel-Rosenblatt Test for Goodness of Fit**
B. K. Ghosh; Wei-Min Huang
*The Annals of Statistics*, Vol. 19, No. 2. (Jun., 1991), pp. 999-1009.
Stable URL:
http://links.jstor.org/sici?sici=0090-5364%28199106%2919%3A2%3C999%3ATPAOKO%3E2.0.CO%3B2-T

**A Consistent Model Specification Test for Nonparametric Estimation of Regression Function Models**
Pedro L. Gozalo
*Econometric Theory*, Vol. 9, No. 3. (Sep., 1993), pp. 451-477.
Stable URL:
http://links.jstor.org/sici?sici=0266-4666%28199309%299%3A3%3C451%3AACMSTF%3E2.0.CO%3B2-K

**Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis**
Bruce E. Hansen
*Econometrica*, Vol. 64, No. 2. (Mar., 1996), pp. 413-430.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199603%2964%3A2%3C413%3AIWANPI%3E2.0.CO%3B2-0

**Comparing Nonparametric Versus Parametric Regression Fits**
W. Hardle; E. Mammen
*The Annals of Statistics*, Vol. 21, No. 4. (Dec., 1993), pp. 1926-1947.
Stable URL:
http://links.jstor.org/sici?sici=0090-5364%28199312%2921%3A4%3C1926%3ACNVPRF%3E2.0.CO%3B2-S

**Consistent Specification Testing Via Nonparametric Series Regression**
Yongmiao Hong; Halbert White
*Econometrica*, Vol. 63, No. 5. (Sep., 1995), pp. 1133-1159.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199509%2963%3A5%3C1133%3ACSTVNS%3E2.0.CO%3B2-1

# LINKED CITATIONS




**Testing a Parametric Model against a Semiparametric Alternative**
Joel L. Horowitz; Wolfgang Härdle
*Econometric Theory*, Vol. 10, No. 5. (Dec., 1994), pp. 821-848.
Stable URL:
http://links.jstor.org/sici?sici=0266-4666%28199412%2910%3A5%3C821%3ATAPMAA%3E2.0.CO%3B2-L

**Semiparametric Estimation of a Proportional Hazard Model with Unobserved Heterogeneity**
Joel L. Horowitz
*Econometrica*, Vol. 67, No. 5. (Sep., 1999), pp. 1001-1028.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199909%2967%3A5%3C1001%3ASEOAPH%3E2.0.CO%3B2-B

**Nonparametric Selection of Regressors: The Nonnested Case**
Pascal Lavergne; Quang H. Vuong
*Econometrica*, Vol. 64, No. 1. (Jan., 1996), pp. 207-219.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199601%2964%3A1%3C207%3ANSORTN%3E2.0.CO%3B2-7

**Nonparametric Significance Testing**
Pascal Lavergne; Quang Vuong
*Econometric Theory*, Vol. 16, No. 4. (Aug., 2000), pp. 576-601.
Stable URL:
http://links.jstor.org/sici?sici=0266-4666%28200008%2916%3A4%3C576%3ANST%3E2.0.CO%3B2-H

**Maximum Likelihood Specification Testing and Conditional Moment Tests**
Whitney K. Newey
*Econometrica*, Vol. 53, No. 5. (Sep., 1985), pp. 1047-1070.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28198509%2953%3A5%3C1047%3AMLSTAC%3E2.0.CO%3B2-4

**Hypothesis Testing in Semiparametric and Nonparametric Models for Econometric Time Series**
P. M. Robinson
*The Review of Economic Studies*, Vol. 56, No. 4. (Oct., 1989), pp. 511-534.
Stable URL:
http://links.jstor.org/sici?sici=0034-6527%28198910%2956%3A4%3C511%3AHTISAN%3E2.0.CO%3B2-2

# LINKED CITATIONS



**A Quadratic Measure of Deviation of Two-Dimensional Density Estimates and A Test of Independence**
M. Rosenblatt
*The Annals of Statistics*, Vol. 3, No. 1. (Jan., 1975), pp. 1-14.
Stable URL:
http://links.jstor.org/sici?sici=0090-5364%28197501%293%3A1%3C1%3AAQMODO%3E2.0.CO%3B2-V

**Adaptive Hypothesis Testing Using Wavelets**
V. G. Spokoiny
*The Annals of Statistics*, Vol. 24, No. 6. (Dec., 1996), pp. 2477-2498.
Stable URL:
http://links.jstor.org/sici?sici=0090-5364%28199612%2924%3A6%3C2477%3AAHTUW%3E2.0.CO%3B2-5

**Consistent Specification Testing with Nuisance Parameters Present Only under the Alternative**
Maxwell B. Stinchcombe; Halbert White
*Econometric Theory*, Vol. 14, No. 3. (Jun., 1998), pp. 295-325.
Stable URL:
http://links.jstor.org/sici?sici=0266-4666%28199806%2914%3A3%3C295%3ACSTWNP%3E2.0.CO%3B2-X

**A Test for Functional Form against Nonparametric Alternatives**
Jeffrey M. Wooldridge
*Econometric Theory*, Vol. 8, No. 4. (Dec., 1992), pp. 452-475.
Stable URL:
http://links.jstor.org/sici?sici=0266-4666%28199212%298%3A4%3C452%3AATFFFA%3E2.0.CO%3B2-B

**Nonparametric Regression Tests Based on Least Squares**
Adonis John Yatchew
*Econometric Theory*, Vol. 8, No. 4. (Dec., 1992), pp. 435-451.
Stable URL:
http://links.jstor.org/sici?sici=0266-4666%28199212%298%3A4%3C435%3ANRTBOL%3E2.0.CO%3B2-0