AN ADAPTIVE, RATE-OPTIMAL TEST OF A PARAMETRIC MODEL AGAINST A NONPARAMETRIC ALTERNATIVE

by

Joel L. Horowitz
Department of Economics
University of Iowa
Iowa City, IA 52242
USA

and

Vladimir G. Spokoiny
Weierstrass Institute for Applied Stochastics
Mohrenstrasse, 39
Berlin 10117
GERMANY

August 1999

Abstract

We develop a new test of a parametric model of a conditional mean function against a nonparametric alternative. The test adapts to the unknown smoothness of the alternative model and is uniformly consistent against alternatives whose distance from the parametric model converges to zero at the fastest possible rate. This rate is slower than $n^{-1/2}$. Some existing tests have non-trivial power against restricted classes of alternatives whose distance from the parametric model decreases at the rate $n^{-1/2}$. There are, however, sequences of alternatives against which these tests are inconsistent and ours is consistent. As a consequence, there are alternative models for which the finite-sample power of our test greatly exceeds that of existing tests. This conclusion is illustrated by the results of some Monte Carlo experiments.

Key words: Hypothesis testing, local alternative, uniform consistency, asymptotic power

JEL listing: C12, C14, C21

We thank Gene Savin for comments on an earlier draft of this paper. The research of Joel L. Horowitz was supported in part by NSF grant SBR-9617925. Part of the research was carried out while Horowitz was visiting the Institute of Statistics and Econometrics, Humboldt University, Berlin, with support from Deutsche Forschungsgemeinschaft Sonderforschungsbereich 373, "Quantifikation und Simulation Ökonomischer Prozesse."

AN ADAPTIVE, RATE-OPTIMAL TEST OF A PARAMETRIC MODEL AGAINST A NONPARAMETRIC ALTERNATIVE

1. INTRODUCTION

This paper is concerned with testing a parametric model of a conditional mean function against a nonparametric alternative. We develop a test that is consistent against alternative models whose distance from the parametric model converges to zero as rapidly as possible as the sample size, n, increases. The test does not require *a priori* knowledge of the smoothness of the alternative model, and it has desirable power properties that are not shared by existing tests.

We consider the model

(1.1) $Y_i = f(X_i) + e_i; \quad i = 1,2,3,\ldots,$

where Y_i is a scalar random variable; $\{X_i\} \in \Re^d$ is a sequence of distinct, non-stochastic, design points; f is an unknown function; and $\{e_i\}$ is a sequence of unobserved, independent, random variables with means of zero. We test the null hypothesis, H_0, that f belongs to the parametric family $Á \equiv \{F(\cdot, q), q \in \Theta\}$, where F is a known function and Θ is a subset of a finite-dimensional space. More precisely, the null hypothesis is that there is a $q \in \Theta$ such that $f(X_i) = F(X_i, q)$ for all i. The alternative hypothesis, H_1, is that there is no $q \in \Theta$ such that $f(X_i) = F(X_i, q)$ for all i.[1]

There is a large literature on testing a parametric model of a conditional mean function against a nonparametric alternative. Many tests compare a nonparametric estimator of $f(\cdot)$ with a parametric estimator, $F(\cdot, q_n)$, where q_n is an estimator of q that is consistent under H_0 (e.g., a least-squares estimator). See, for example, Aït-Sahalia, *et al*. (1994), Eubank and Spiegelman (1990), Fan and Li (1996), Gozalo (1993), Härdle and Mammen (1993), Hart (1997), Hong and White (1995), Li and Wang (1998), Whang and Andrews (1993), Wooldridge (1992), Yatchew (1992), and Zheng (1996). Other tests do not require nonparametric estimation of f. Bierens (1982, 1990), Bierens and Ploberger (1997), and De Jong (1996) test orthogonality conditions that are implied by (1.1). Andrews (1997) develops a conditional Kolmogorov test.[2]

The asymptotic power of a test of H_0 is often investigated by deriving the asymptotic probability that the test rejects H_0 against a sequence of local alternative models. This approach is well known but, as is explained in the next paragraph, restricts attention to a class of alternative models that is too small. The form of the local alternative models is

(1.2) $f_n(x) = F(x, q_1) + r_n g(x)$

for some function g, where $q_1 \in \Theta$ and $\{r_n\}$ is a sequence of real numbers that converges to 0 as $n \to \infty$. See, for example, Andrews (1997), Bierens and Ploberger (1997), Eubank and Spiegelman (1990), Hong and White (1995), and Zheng (1996). Many tests that compare a nonparametric estimator of f with a parametric estimator have non-trivial power (that is, power exceeding the probability that a correct H_0 is rejected) only against sequences of local alternatives for which $r_n \to 0$ at a rate that is slower than $n^{-1/2}$. The tests of Aït-Sahalia, *et al*. (1994), Eubank and Spiegelman (1990), Fan and Li (1996), Gozalo (1993), Härdle and Mammen (1993), Hong and White (1995), Whang and Andrews (1993), Wooldridge (1992), Zheng (1996), and Yatchew (1992) have non-trivial power only if r_n converges more slowly than $n^{-1/2}$.

Andrews (1997), Bierens (1982), Bierens and Ploberger (1997), and Hart (1997) describe tests that have non-trivial power against local alternatives for which $r_n \propto n^{-1/2}$. Thus, at least in terms of asymptotic local power, these tests appear to dominate tests that require slower convergence of r_n. It turns out, however, that if $r_n \propto n^{-1/2}$, then no test can have non-trivial power uniformly over reasonable classes of functions g in (1.2) (e.g., functions that have derivatives of order s for some integer s). See Burnashev (1979), Ibragimov and Khasminskii (1977), and Ingster (1982). In other words, the power of any test of H_0 against the sequence of local alternatives $f_n(x) = F(x,q_1) + n^{-1/2} g_n(x)$ equals the probability that the test rejects a correct H_0 for some sequence $\{g_n\}$ of (say) twice differentiable functions. The practical consequence of this result is that any test of H_0 for which $r_n \propto n^{-1/2}$ has low finite-sample power against certain classes of smooth alternatives. Section 4.2 presents numerical examples of this phenomenon. Hong and White (1995) and Fan and Li (1999) also present examples. Because the class (1.2) excludes models of the form $f_n(x) = F(x,q_1) + r_n g_n(x)$, it cannot be used to develop tests that have good power against all smooth alternatives. This is the sense in which the class (1.2) is too small.

Another way to investigate the asymptotic power properties of tests of H_0 is the *minimax* approach of Ingster (1982, 1993a, 1993b, 1993c). This approach, which is not widely known in econometrics, permits the set of alternatives to consist of an entire smoothness class. The minimax approach forms the basis of the test that is developed here. In this approach, it is assumed that f belongs to a class of one-or-more-times-differentiable functions on $\Re^d$, such as a Hölder, Sobolev, or Besov ball, B.[3] B is separated from the null-hypothesis set *Á* by some distance r_n that converges to zero as $n \to \infty$. The aim of the minimax approach is to find the fastest rate at which r_n can approach zero while permitting consistent testing uniformly over B. This rate is called the *optimal rate of testing*. A test is consistent uniformly over B if

(1.3) $$\lim_{n\to\infty} \inf_{f \in B} \boldsymbol{P}(H_0 \text{ is rejected against } f) = 1 .$$

Thus, the optimal rate of testing is the fastest rate at which r_n can approach zero while maintaining (1.3). The optimal rate of testing for Hölder, Sobolev, or Besov classes of functions that have bounded derivatives of order $s \geq d/4$ is $n^{-2s/(4s+d)}$ (Ingster 1982, 1993a, 1993b, 1993c; Guerre and Lavergne 1999). This rate assumes that s is known *a priori*. If s is unknown, then the optimal rate of testing is $\left(n^{-1}\sqrt{\log\log n}\right)^{2s/(4s+d)}$, which differs from the rate that is achievable with known s by the very slowly increasing factor $(\log\log n)^{s/(4s+d)}$ (Spokoiny 1996). If $s < d/4$, then the optimal rate of testing is $n^{-1/4}$ (see, e.g., Guerre and Lavergne 1999).

A test that achieves the optimal rate of testing has the advantage of being sensitive to alternatives uniformly over a Hölder, Sobolev, or Besov class whose distance from the null hypothesis $\acute{A}$ converges to zero at the fastest possible rate. These classes contain sequences of alternative models against which the tests of Andrews (1997), Bierens (1982), Bierens and Ploberger (1997), and Hart (1997) are inconsistent. In practice, this means that there are smooth alternatives against which these tests have much lower finite-sample power than does a test that achieves the optimal rate of testing. Section 4.2 presents numerical illustrations.

In this paper, we construct a test of H_0 that has the optimal rate of testing uniformly over Hölder classes and does not require *a priori* knowledge of s, the order of differentiability of f. The test satisfies (1.3) with $r_n \propto \left(n^{-1}\sqrt{\log\log n}\right)^{2s/(4s+d)}$ when $s \geq d/4$. The test is called *adaptive* and *rate-optimal* because it adapts to the unknown s and has the optimal rate of testing for the case of an unknown s.[4]

A test that achieves the optimal rate of testing uniformly over a smoothness class B is necessarily oriented toward the alternatives in B that are most extreme and hardest to detect. These functions have narrow peaks or valleys whose widths decrease with increasing n. See Section 4.1 for an example. A test that is oriented toward such alternatives may have low power against functions that are less extreme. To provide some protection against this possibility, we investigate the consistency of our test against alternatives of the form (1.2). These alternatives cannot have the extreme behavior just described because g in (1.2) is a fixed function. We show that our test is consistent against alternatives of the form (1.2) whenever $r_n \geq Cn^{-1/2}\sqrt{\log\log n}$ for some finite $C > 0$. The tests of Andrews (1997), Bierens (1982), Bierens and Ploberger (1997), and Hart (1997) are consistent against alternatives of the form (1.2) whenever $r_n \to 0$ more slowly than $n^{-1/2}$. Thus, our adaptive, rate-optimal test and the other tests (which are not rate-optimal) are consistent against virtually the same alternatives of the form (1.2). In terms of consistency against alternatives of the form (1.2), there is essentially no penalty paid for the adaptiveness and rate optimality of our test.[5]

Throughout this paper, our concern is with the rate at which the distance between the null and alternative hypotheses can decrease to zero while permitting consistent testing by some procedure. We do not investigate other properties of the power functions of tests, and we do not derive the asymptotic local power function of our test. Nor do we attempt analytic comparisons of the powers of our test and others apart from noting conditions under which our test is consistent and others are not. More extensive power comparisons are left for future research. The contribution of this paper is to provide a test that (1) adapts to the unknown smoothness of the alternative model, (2) is consistent at the optimal rate uniformly over Hölder classes of alternatives, and (3) is consistent against alternatives of the form (1.2) when r_n has nearly a $n^{-1/2}$ rate of convergence. The first two properties of our test guarantee that there are alternatives against which our test has high power and tests such as those of Andrews (1997), Bierens (1982), Bierens and Ploberger (1997), and Hart (1997) have low power. The third property provides some protection against the occurrence of the opposite situation.

The test statistic is described in Section 2 of this paper. Theorems giving properties of the test under H_0 and various forms of H_1 are presented in Section 3. Section 4 presents the results of some Monte Carlo experiments that illustrate the numerical performance of the test. Concluding comments are presented in Section 5. The proofs of theorems are in the Appendix.

2. THE TEST STATISTIC

This section describes our test statistic and presents a bootstrap method for obtaining critical values of the test. The test is closely related to that of Härdle and Mammen (1993). Like Härdle and Mammen, we base the test on the distance between a kernel nonparametric estimator of f and a kernel-smoothed parametric estimator. The main difference between our test and that of Härdle and Mammen is that we compute the distance with many different values of the bandwidth parameter of the kernel smoother. We reject H_0 if the distance obtained with any of the bandwidths is too large. The rate-optimal and adaptive properties of our test arise from its use of many different bandwidths.

The remainder of this section is divided into five parts. Section 2.1 describes the parametric estimator of f. Section 2.2 describes the kernel smoothing procedure and the metric that is used to measure the distance between the nonparametric and smoothed parametric estimators of f. Section 2.3 explains how the distance between the nonparametric and smoothed parametric estimators is centered and Studentized. The test procedure is presented in Section 2.4. Section 2.5 explains how to estimate unknown population parameters that enter the test statistic.

2.1 *The Parametric Estimator*

We consider the model (1.1). The hypothesis to be tested is H_0: $f \in \Im = \{F(\cdot, q), q \in \Theta\}$, where F is a known function and Θ is an open subset of a finite-dimensional Euclidean space. We assume that there is an estimator of q, denoted by q_n, that is $n^{1/2}$-consistent under H_0. Let $q_0 \in \Theta$ denote the true value of q if H_0 is true. That is, $\boldsymbol{E}(Y_i) = F(X_i, q_0)$ for all i if H_0 is true. Then, $n^{1/2}(q_n - q_0)$ is bounded in probability under H_0.

We assume that q_n is stable if H_0 is false. By this we mean that there is a $q^* \in \Theta$ such that $n^{1/2}(q_n - q^*)$ is bounded in probability if H_0 is false. Under assumptions stated in Amemiya (1985), for example, the least-squares estimator of q has the required properties, as do many other M estimators (Millar 1982).

2.2 *The Kernel Smoother*

We now explain the kernel smoothing procedure that is used in our test. Let K denote the kernel and h denote a bandwidth. For $x \in \Re^d$, let $K_h(x) = K(x/h)$. For each $i, j = 1, 2, \ldots, n$ define

$$w_h(X_i, X_j) = \frac{K_h(X_i - X_j)}{\sum_{k=1}^{n} K_h(X_i - X_k)} .$$

The kernel nonparametric estimator of $f(X_i)$ is

$$f_h(X_i) = \sum_{j=1}^{n} w_h(X_i, X_j) Y_j .$$

The kernel-smoothed parametric estimator is

$$F_h(X_i, q_n) = \sum_{j=1}^{n} w_h(X_i, X_j) F(X_j, q_n) .$$

The distance between the nonparametric and smoothed parametric estimators of f is defined to be the sum of the squared differences $f_h(X_i) - F_h(X_i, q_n)$.[6] Accordingly, for any $q \in \Theta$, define

$$S_h(q) = \sum_{i=1}^{n} [f_h(X_i) - F_h(X_i, q)]^2 .$$

The test statistic is based on a centered, Studentized version of $S_h(q_n)$ whose asymptotic distribution has a mean of zero and variance of one.

Some vector notation will be useful in the discussion that follows. Define the $n\times 1$ vectors $Y = (Y_1, \ldots, Y_n)'$ and $F(q) = [F(X_1, q), \ldots, F(X_n, q)]'$. Let W_h be the $n\times n$ matrix whose (i, j) element is $w_h(X_i, X_j)$. Let $\|\cdot\|$ denote the ℓ_2 norm. That is, for any $z \in \Re^n$,

$$\|z\|^2 = \sum_{i=1}^{n} z_i^2 .$$

Then

$$S_h(q) = \|W_h[Y - F(q)]\|^2$$

for any $q \in \Theta$.

2.3 *Centering and Studentization*

This section explains the method that is used to center and Studentize $S_h(q_n)$. We begin by defining further notation. Suppose that H_0 is true. Then $f(X_i) = F(X_i, q_0)$ for all i. Define the $n\times 1$ vector $\boldsymbol{e} = (e_1, \ldots, e_n)'$. For $q \in \Theta$, define the $n\times 1$ vector $b_h(q) = W_h[F(q_0) - F(q)]$. Then

$$Y - F(q) = F(q_0) - F(q) + e ,$$

and

(2.1) $$S_h(q) = \|W_h e + b_h(q)\|^2 = \|W_h e\|^2 + \|b_h(q)\|^2 + 2b_h(q)'W_h e .$$

Let $a_{ij,h}$ denote the (i, j) element of the $n\times n$ matrix $A_h = W_h'W_h$. Let $s_4(X_i) = \boldsymbol{E}(e_i^4)$ and $s^2(X_i) = \boldsymbol{E}(e_i^2)$. Assume that these quantities exist.

To develop the method for centering and Studentizing $S_h(q_n)$, it is first necessary to evaluate the mean and variance of $S_h(q_0)$ under H_0. Observe that

$$S_h(q_0) = \|W_h e\|^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij,h} e_i e_j .$$

Then

(2.2) $$\boldsymbol{E}\|W_h e\|^2 \equiv N_h = \sum_{i=1}^{n} a_{ii,h} s^2(X_i) .$$

In addition, $\mathrm{Var}\|W_h e\|^2 = V_h^2 + n_h$, where

(2.3) $$V_h^2 = 2\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij,h}^2 s^2(X_i) s^2(X_j)$$

and

$$n_h = \sum_{i=1}^{n} a_{ii,h}^2 [s_4(X_i) - 3s^4(X_i)] .$$

It is not difficult to show that $n_h = o(V_h^2)$ as $n \to \infty$, so n_h is asymptotically negligible. Therefore, an asymptotically centered, normalized form of $S_h(q_0)$ is

$$T_h^0 \equiv \frac{S_h(q_0) - N_h}{V_h} = \frac{\|W_h e\|^2 - N_h}{V_h} .$$

That is, the asymptotic distribution of T_h^0 has a mean of zero and a variance of one.

To obtain the centered, Studentized form of $S_h(q_n)$, define

$$\tilde{T}_h = \frac{S_h(q_n) - N_h}{V_h} = T_h^0 + h_h ,$$

where

$$h_h = \frac{\|b_h(q_n)\|^2 + 2b_h(q_n)' W_h e}{V_h} .$$

It follows from Lemmas 4.3 and 4.5 of the Appendix that $h_h = o_p(1)$ as $n \to \infty$. Therefore, the asymptotic distribution of $\tilde{T}_h$ has mean zero and variance one. However, $\tilde{T}_h$ cannot be computed in an application because it depends on the unknown quantities $s^2(X_i)$ (i = 1, …, n). This problem can be solved by replacing each $s^2(X_i)$ in (2.2) and (2.3) with an estimator. Methods for estimating $s^2(X_i)$ are described in Section 2.5. For now, we assume that such methods exist and denote the estimator of $s^2(X_i)$ by $s_n^2(X_i)$. The centered, Studentized form of $S_h(q_n)$ is obtained from $\tilde{T}_h$ by replacing $s^2(X_i)$ with $s_n^2(X_i)$ in N_h and V_h. Specifically, define

(2.4) $$\hat{N}_h = \sum_{i=1}^{n} a_{ii,h} s_n^2(X_i) ,$$

(2.5) $$\hat{V}_h^2 = 2\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij,h}^2 s_n^2(X_i) s_n^2(X_j) ,$$

and

(2.6) $$T_h = \frac{S_h(q_n) - \hat{N}_h}{\hat{V}_h} .$$

Then T_h is a feasible statistic whose asymptotic distribution has mean zero and variance one. It is the centered, Studentized form of $S_h(q_n)$ that is used to construct our test statistic.

2.4 *The Test Procedure*

The idea of the test is to consider simultaneously a family of test statistics $\{T_h, h \in H_n\}$, where H_n is a set of bandwidth values. We assume that H_n is finite and denote the number of elements of H_n by J_n. A specific example is a geometric grid of the form

(2.7) $$H_n = \{h = h_{\max} a^k : h \geq h_{\min}, k = 0,1,2...\} ,$$

where $0 < h_{\min} < h_{\max}$, and $0 < a < 1$. In this case, $J_n \le \log_{1/a} (h_{\max}/h_{\min})$. The proposed test procedure rejects H_0 if at least one of the statistics T_h for $h \in H_n$ is sufficiently large. Thus, we define

$$(2.8) \qquad T^* = \max_{h \in H_n} T_h = \max_{h \in H_n} \frac{S_h(q_n) - \hat{N}_h}{\hat{V}_h} .$$

We use T^* as a test statistic.

We now discuss how to obtain critical values for T^*. The exact a-level critical value, t_a^*, $(0 < a < 1)$ is the $1 - a$ quantile of the exact finite-sample distribution of T^*. This critical value cannot be evaluated in applications because q_0 and the distributions of the e_i are unknown. However, it is shown in Lemmas 8-10 of the Appendix that asymptotically (as $n \to \infty$), t_a^* is determined by the variances of the e_i's, $s^2(X_i)$. The value of q_0 and other features of the distributions of the e_i's do not affect the asymptotic critical value. Therefore, an asymptotic a-level critical value, t_a, can be obtained as the $1 - a$ quantile of the distribution of T^* that is induced by the model $Y_i^* = F(X_i, q_n) + e_i^*$, where e_i^* is sampled randomly from the normal distribution $N[0, s_n^2(X_i)]$. The test proposed here rejects H_0 with asymptotic level a if $T^* > t_a$. The asymptotic critical value t_a can be estimated with any desired accuracy by using the following simulation procedure:

1. For each $i = 1, \ldots, n$, generate $Y_i^* = F(X_i, q_n) + e_i^*$, where e_i^* is sampled randomly from the normal distribution $N[0, s_n^2(X_i)]$.

2. Use the data set $\{Y_i^*, X_i: \; i = 1, \ldots, n\}$ to estimate q and $s^2(X_i)$. Denote the resulting estimates by $\hat{q}_n$ and $\hat{s}_n^2(X_i)$, respectively. Compute the statistic $\hat{T}^*$ that is obtained by replacing Y_i, q_n, and $s_n^2(X_i)$ with Y_i^*, $\hat{q}_n$, and $\hat{s}_n^2(X_i)$ on the right-hand side of (2.5).

3. Estimate t_a by the $1 - a$ quantile of the empirical distribution of $\hat{T}^*$ that is obtained by repeating steps 1-2 many times.

2.5 *Estimating* $s^2(X_i)$

This section explains how $s^2(X_i)$ can be estimated. We need an estimator that is consistent regardless of whether H_0 is true. Thus, we cannot base the estimator on the residuals of the parametric model $Y_i - F(X_i, q_n)$.[7]

Recall that the e_i's are assumed to be independently distributed. Assume for the moment that they are also identically distributed so that $s^2(X_i) = s^2$ for some constant $s^2 > 0$. If $d = 1$ (the X_i's are scalars), then we can estimate s^2 by using the method of Rice (1984), Gasser, *et al*. (1986), and Buckley, *et al*. (1988). Let $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ be the ordered sequence of design points, and let $Y_{(i)}$ and $e_{(i)}$, respectively, be the similarly ordered values of the Y_i's and e_i's. Then $Y_{(i+1)} - Y_{(i)} = e_{(i+1)} - e_{(i)} + f(X_{(i+1)}) - f(X_{(i)})$. Now, $\boldsymbol{E}(e_{(i+1)} - e_{(i)})^2 = 2s^2$. Moreover, under the assumptions of Section 3.1, $|f(X_{(i+1)}) - f(X_{(i)})| \to 0$ as $n \to \infty$. Therefore, we can estimate s^2 by

$$(2.9) \qquad s_n^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (Y_{(i+1)} - Y_{(i)})^2 .$$

This estimator is $n^{1/2}$-consistent under the assumptions of Section 3.1, regardless of whether H_0 is true (Rice 1984).

We now explain how this method can be extended to multivariate settings. We restrict the discussion to the case of $d \leq 4$. Let $j(i)$ ($i = 1, \ldots, n$) be a set of indices that is defined through the following recursion:

$$j(1) = \arg \min_{j=2,\ldots,n} \left\| X_j - X_1 \right\|$$

and

$$j(i) = \arg \min_{j \neq i,\, j(1),\ldots j(i-1)} \left\| X_j - X_i \right\|; \quad i = 2, \ldots, n.$$

The number $j(i)$ is the index of the design point that is nearest to X_i among those whose indices are not $j(1), \ldots, j(i-1)$. Then s^2 can be estimated by

$$(2.10) \qquad s_n^2 = \frac{1}{2n} \sum_{i=1}^{n} (Y_i - Y_{j(i)})^2 .$$

Under the assumptions of Section 3.1, (2.10) is a $n^{1/2}$-consistent estimator of s^2, regardless of whether H_0 is true.

The estimator s_n^2 can be extended to e_i's that have heteroskedasticity of unknown form by replacing the global sums in (2.9) and (2.10) by sums over shrinking neighborhoods of the design points X_i.[8] Let $s^2(\cdot)$ satisfy the Lipschitz condition $|s^2(X_i) - s^2(X_j)| \leq L\|X_i - X_j\|$ for some finite $L > 0$. Let b_n be a bandwidth that converges to 0 as $n \to \infty$, and let $I(\cdot)$ be the indicator function. Define $j(i)$ as before. Then under the assumptions of Section 3.1, $s^2(X_i)$ can be estimated by

$$s_n^2(X_i) = \frac{\sum_{k=1}^{n} (Y_k - Y_{j(k)})^2 I(|X_k - X_i| \le b_n)}{\sum_{k=1}^{n} I(|X_k - X_i| \le b_n)} .$$

If $b_n \to 0$ and $nh_{\min} b_n^d \to \infty$ as $n \to \infty$, then $s_n^2(X_i) - s^2(X_i) = o_p(h_{\min}^{1/2})$ as $n \to \infty$.

It is shown Lemma 8 of the Appendix that if $s_n^2(X_i) - s^2(X_i) = o_p(h_{\min}^{1/2})$, then $T^* = \max_{h \in H_n} T_{h0} + o_p(1)$, where $T_{h0} = [S_h(q^*) - N_h]/V_h$ and $q^* = q_0$ if H_0 is true. Thus, the asymptotic distribution of T^* is the same as it would be if q^* and $s^2(X_i)$ were known, regardless of whether H_0 is true.

3. THE MAIN RESULTS

This section presents theorems that give the asymptotic behavior of the proposed test. Section 3.1 states our assumptions. The behavior of the test under H_0 is given in Section 3.2. Sections 3.3-3.5, respectively, give the test's behavior under a fixed alternative hypothesis, under the sequence of local alternative hypotheses (1.2), and under smooth alternatives that are contained in a Hölder ball whose distance from the null hypothesis converges to zero at the optimal rate of testing $\left(n^{-1}\sqrt{\log\log n}\right)^{2s/(4s+d)}$. The adaptive, rate-optimal property of the test is established in Section 3.5.

3.1 *Assumptions*

Our results are obtained under the assumptions stated in this section. Define $\nabla_q F(x,q) = \partial F(x,q)/\partial q$, $\nabla_q^2 F(x,q) = \partial^2 F(x,q)/\partial q \partial q'$, $\nabla_x F(x,q) = \partial F(x,q)/\partial x$, and $\nabla_x^2 F(x,q) = \partial^2 F(x,q)/\partial x \partial x'$ whenever these derivatives exist. For any $q \times q$ matrix D, define

$$\|D\|_\infty = \sup_{v \in \Re^q} \frac{\|Dv\|}{\|v\|},$$

where $\|\cdot\|$ is the ℓ_2 norm. Let $\nabla_q F(q)$ be the $n \times q$ matrix whose (i,j) element is $\partial F(X_i, q)/\partial q_j$.

<u>Assumption 1</u> (Parametric family): *The parameter set* Θ *is an open subset of* $\Re^q$ *for some* $q \ge 1$. *The parametric family* $\Im = \{F(\cdot, q), q \in \Theta\}$ *satisfies*:

(i) Differentiability in *q: For each* $x \in [-1,1]^d$, $F(x, q)$ *is twice differentiable with respect to q. For finite constants* C_{11} *and* C_{12}, *each* $i = 1, \ldots, n$, *and each* $q \in \Theta$, $\|\nabla_q F(X_i, q)\| \le C_{11}$, *and* $\|\nabla_q^2 F(X_i, q)\|_\infty \le C_{12}$.

(ii) Differentiability in *x*: *For each* $q \in \Theta$, $F(x, q)$ *is twice differentiable with respect to* $x \in [-1,1]^d$. *Moreover,* $\|\nabla_x^2 F(x, q)\|_\infty \le C_{13}$ *for some finite constant* C_{13}.

(iii) Identifiability: *There is a finite* $C_I > 0$ *such that*

$$\sup_{q \in \Theta} \| [n^{-1} \nabla_q F(q)' \nabla_q F(q)]^{-1} \|_\infty \le C_I^{-1}$$

and for every $d > 0$

$$\inf_{q, q' \in \Theta: |q - q''| \ge d} \|F(q) - F(q')\|^2 \ge C_I d^2 n .$$

<u>Assumption 2</u> (Parametric estimator): (i) *Let* H_0 *be true. Then* $q_0 \in \Theta$ *and*

$$\lim_{n \to \infty} \boldsymbol{P}\left(n^{1/2} \|q_n - q_0\| > z\right) < d$$

for any $d > 0$ *and all sufficiently large z.* (ii) *Let* H_0 *be false. Then there is a* $q^* \in \Theta$ *such that*

$$\lim_{n \to \infty} \boldsymbol{P}\left(n^{1/2} \|q_n - q^*\| > z\right) < d$$

for any $d > 0$ *and all sufficiently large z.* (iii) *Let* $\{q_{n0}\ \ n = 1, 2, \ldots\}$ *be a sequence in* Θ *whose limit points, if any, are all in* Θ. *Let* $\{s_{ni}:\ i = 1, \ldots, n;\ n = 1, 2, \ldots\}$ *be a triangular array of real numbers that is bounded away from* 0 *and* ∞. *Define* $Y_i^* = F(X_i, q_{n0}) + s_{ni} w_i$, *where* $\{w_i : i = 1, \ldots, n\}$ *are independently distributed as* $N(0,1)$. *Let* $\hat{q}_n$ *be the estimator of* q_{n0} *that is obtained from the data set* $\{Y_i^*, X_i:\ i = 1, \ldots, n\}$. *Then*

$$\lim_{n \to \infty} \boldsymbol{P}\left(n^{1/2} \|\hat{q}_n - q_{n0}\| > z\right) < d$$

for any $d > 0$ *and all sufficiently large z.*

Assumption 2(iii) establishes a stability property of the parametric estimator that is used to justify the simulation procedure for obtaining the critical value of the test statistic.

For every $x \in \Re^d$ and every $h > 0$, define $M_h(x)$ as the number of elements in the set $\{X_i : \|X_i - x\| \le h\}$.

<u>Assumption 3</u> (Design): (i) *The design points* $X_i \in \Re^d$ $(i = 1, \ldots, n)$ *are non-stochastic and scaled so that* $\|X_i\| \le 1$ *for all i.* (ii) *There are positive constants* C_1 and C_2 *such that for all* $h \in H_n$ *and all* $i = 1, \ldots, n$, $C_1 n h^d \le M_h(X_i) \le C_2 n h^d$.

Assumption 3(i) restricts the X_i to a bounded subset of $\Re^d$. Given boundedness of the X_i, there is no loss of generality in the scaling assumption. Assumption 3(ii) is satisfied with probability approaching 1 as $n \to \infty$ if H_n satisfies Assumption 6 below and $\{X_i\}$ is sampled randomly from a distribution that is absolutely continuous with respect to Lebesgue measure, has bounded support, and whose density is bounded away from zero on its support. Therefore, our results hold conditional on $\{X_i\}$ that are generated this way. However, we do not require $\{X_i\}$ to be sampled from a distribution.

Assumption 4 (Kernel): *K is non-negative, supported on* $[-1,1]^d$, *and symmetrical about the origin. Moreover,* $K(u) \le 1$ *for all u, and* $K(u) \ge k$ *for* $\|u\| \le 1/2$ *and some* $k > 0$.

Assumption 5 (Moments of e_i): (i) *The random variables* e_i *are independent with means of zero and uniformly bounded moments of order* $4 + d$ *for some* $d > 0$. $\boldsymbol{E}|e_i|^{4+d} \le C_E$ *for some constant* $C_E < \infty$ *and all* $i = 1, \ldots, n$. (ii) $s^2(X_i) = \boldsymbol{E}(e_i^2)$ *and* $s_4(X_i) = \boldsymbol{E}(e_i^4)$ *satisfy* $|s^2(X_i) - s^2(X_j)| \le L\|X_i - X_j\|$ and $|s_4(X_i) - s_4(X_j)| \le L\|X_i - X_j\|$ *for some constant* $L < \infty$ *and all* $i, j, = 1, \ldots, n$. (iii) $s^2(X_i) \ge m_2$ *for some constant* $m_2 > 0$ *and all i.*

Assumption 6: (Bandwidths): *The set* H_n *of bandwidths has the structure* (2.7) *with* $h_{\max} > h_{\min} \ge n^{-g}$ *for some constant g such that* $0 < g < \min(1/3, 1/d)$, *and* $h_{\max} = C_H(\log\log n)^{-1}$ *for some finite constant* $C_H > 0$.

Under Assumption 6, $J_n \le O(\log n)$ as $n \to \infty$.

3.2 *Behavior of the Test Statistic under the Null Hypothesis*

Recall from Section 2.4 that t_a is the $1 - a$ quantile of the distribution of T^* that is induced by the model $Y_i^* = F(X_i, q_n) + e_i^*$, where e_i^* is sampled randomly from the normal distribution $N[0, s_n^2(X_i)]$. The main result on the behavior of the test statistic T^* under H_0 is that t_a is an asymptotically correct a-level critical value under any model in H_0. This result is established by the following theorem.

Theorem 1: *Let Assumptions* 1-6 *hold. Let* H_0 *be true. Then*

$$\lim_{n \to \infty} \boldsymbol{P}(T^* > t_a) = a .$$

3.3 *Consistency Against a Fixed Alternative*

We now show that our test is consistent against a fixed alternative model. Let (1.1) hold. Define the $n\times 1$ vector $\bar{f} = [f(X_1), \ldots, f(X_n)]'$. Measure the distance between f and the parametric family $\Im$ by the normalized ℓ_2 distance

$$r(f,\Im) = \left[\inf_{q \in \Theta} \left(n^{-1} \left\| \bar{f} - F(q) \right\|^2 \right) \right]^{1/2}. \tag{3.1}$$

If H_0 is false, then $r(f, \Im) \geq c_r$ for all sufficiently large n and some $c_r > 0$. A consistent test will reject a false H_0 with probability approaching one as $n \to \infty$. Theorem 2 establishes the consistency of our test.

Theorem 2: *Let Assumptions* 1-6 *hold. If there is an* n_0 *such that* $r(f, \Im) \geq c_r$ for all $n > n_0$ *and some* $c_r > 0$, *then*

$$\lim_{n\to\infty} \boldsymbol{P}(T^* > t_a) = 1.$$

3.4 *Consistency Against a Sequence of Local Alternatives*

This section establishes the consistency of our test under local alternatives of the form (1.2) with $r_n \geq Cn^{-1/2}\sqrt{\log\log n}$ for some constant $C > 0$.

Define the $n\times 1$ vectors $\bar{g} = [g(X_1),\ldots,g(X_n)]'$ and $\bar{f}_n = [f_n(X_1),\ldots,f_n(X_n)]'$. We assume that g is a continuous function that is normalized so that

$$\frac{1}{n}\|\bar{g}\|^2 = \frac{1}{n}\sum_{i=1}^{n}|g(X_i)|^2 \geq 1. \tag{3.2}$$

We also assume that $\bar{g}$ is not an element of the space spanned by the columns of $\nabla_q F(q_1)$. That is,

$$\|\bar{g} - \Pi_1 \bar{g}\| \geq d\|\bar{g}\| \tag{3.3}$$

for some $d > 0$, where

$$\Pi_1 = \nabla_q F(q_1)[\nabla_q F(q_1)'\nabla_q F(q_1)]^{-1}\nabla_q F(q_1)'$$

is the projection operator into the column space of $\nabla_q F(q_1)$. Conditions (3.2) and (3.3) exclude functions g for which $\left\|\bar{f}_n - F(q_{n,0})\right\| = o(r_n)$ for some non-stochastic sequence $\{q_{n,0}\} \in \Theta$. Thus, (3.2) and (3.3) insure that the rate of convergence of f_n to the parametric model $F(\cdot, q_1)$ is the same as the rate of convergence of r_n to zero. In particular, under (3.2) and (3.3),

$$\left[\inf_{q \in \Theta}\left(n^{-1}\left\| \bar{f}_n - F(q)\right\|^2\right)\right]^{1/2} \geq d r_n[1-o(1)]$$

as $n \to \infty$.

Finally, we assume that q_n is the least squares estimator of q. This assumption is made for technical convenience only and is not essential to the consistency result, which is stated in the following theorem.

Theorem 3: *Let Assumptions* 1 *and* 3-6 *hold with* $h_{\max} = C_H(\log\log n)^{-1}$ *for some finite constant* C_H. *Let* q_n *be the least-squares estimator of* q. *Let* f_n *satisfy* (1.2) *with* $r_n \geq Cn^{-1/2}\sqrt{\log\log n}$ *for some constant* $C > 0$. *Let g satisfy* (3.2) *and* (3.3). *Then*

$$\lim_{n \to \infty} \boldsymbol{P}(T^* > t_a) = 1 .$$

This result shows that the power of the adaptive, rate-optimal test approaches 1 as $n \to \infty$ for any function g and sequence $\{r_n\}$ that satisfy the assumptions of the theorem. However, the result is not uniform over all possible g's. Uniformity is addressed in the next section.

3.5 *Consistency Against a Sequence of Smooth Alternatives*

This section gives conditions under which our test is consistent uniformly over alternatives in a Hölder smoothness class whose distance from the parametric model approaches zero at the fastest possible rate. The results can be extended to Sobolev and Besov classes under some additional technical conditions on the design $\{X_i\}$.

To specify the smoothness classes that we consider, let $j = (j_1, \ldots, j_d)$, where $j_1, \ldots, j_d \geq 0$ are integers, be a multi-index. Define

$$|j| = \sum_{k=1}^{d} j_k$$

and

$$D^j f(x) = \frac{\partial^{|j|} f(x)}{\partial x_1^{j_1} \ldots \partial x_d^{j_d}}$$

whenever the derivative exists. Define the Hölder norm

$$\|f\|_{H,s} = \sup_{x \in [-1,1]^d} \sum_{|j| \leq s} |D^j f(x)| .$$

The smoothness classes that we consider consist of functions $f \in S(H,s) \equiv \{f : \|f\|_{H,s} \leq C_F\}$ for some (unknown) $s \geq \max(2, d/4)$ and $C_F < \infty$.

Theorem 4 states that our test is consistent uniformly over the sets

(3.4) $$B_{H,n} \equiv \left\{ f \in S(H,s): r(f,\Im) \geq C_a \left(n^{-1}\sqrt{\log\log n}\right)^{2s/(4s+d)} \right\}$$

for some $s \geq \max(2, d/4)$ and all sufficiently large $C_a < \infty$.

Theorem 4: *Let Assumptions* 1-6 *hold. Then for* $0 < a < 1$ *and* $B_{H,n}$ *as defined in* (3.4),

$$\lim_{n\to\infty} \inf_{f \in B_{H,n}} \boldsymbol{P}(T^* > t_a) = 1$$

for all sufficiently large $C_a < \infty$.

4. MONTE CARLO EXPERIMENTS

This section presents the results of Monte Carlo experiments that illustrate the numerical performance of the adaptive, rate-optimal test. The section has two parts. Section 4.1 presents a sequence of alternatives against which our test is consistent but the tests of Andrews (1997), Bierens (1982), Bierens and Ploberger (1997), and Härdle and Mammen (1993) are not. This sequence motivates the design of the Monte Carlo experiments. The experiments and their results are described in Section 4.2.

4.1 *An Example*

This section presents a parametric model and a sequence of alternatives against which our test is consistent but the tests of Andrews (1997), Bierens (1982), Bierens and Ploberger (1997), and Härdle and Mammen (1993) are not. All of these tests are consistent against each alternative in the sequence, however. The fact that the tests are not consistent against the sequence itself, as opposed to its individual elements, illustrates their lack of uniform consistency.

The null hypothesis model (parametric family) in the example is

(4.1) $$Y_i = b_0 + b_1 X_i + e_i,$$

where b_0 and b_1 are constants, the X_i's are scalars that are sampled from a distribution that is symmetrical about 0, and $e_i \sim N(0, s^2)$ for every i. The distribution of e_i is specified parametrically because Andrews' (1997) test requires a fully parametric model. The other tests do not require specification of the distribution of e_i. The sequence of alternative models is

(4.2) $$Y_i = X_i + t_n^4 f(X_i / t_n) + e_i,$$

where $e_i \sim N(0,1)$, f is the standard normal density function, and $t_n = C\left(n^{-1}\sqrt{\log\log n}\right)^{-1/9}$ for some finite $C > 0$. The function $f_n(x) = x + t_n^4 f(x/t_n)$ has a peak that is centered at $x = 0$ and that becomes narrower as n increases. The sequence of alternative models $\{f_n\}$ is contained in $B_{H,n}$

with $s = 2$. The distance between f_n and the parametric model (4.1) satisfies $r(f_n, \Im) \propto \left(n^{-1}\sqrt{\log\log n}\right)^{-4/9}$, so the distance converges to zero more slowly than $n^{-1/2}$.

It is not difficult to show under that the sequence (4.2), the noncentral parameters of the tests of Andrews (1997), Bierens (1982), Bierens and Ploberger (1997), and Härdle and Mammen converge to zero as $n \to \infty$. Therefore, these tests are inconsistent against (4.2). It follows from Theorem 4, however, that the adaptive, rate optimal test is consistent against this sequence if C is sufficiently large.

4.2 *Monte Carlo Experiments*

This section presents the results of Monte Carlo experiments that illustrate the numerical performance of the adaptive, rate-optimal test. In each experiment, a parametric null-hypothesis model and two alternatives are specified. Monte Carlo simulation is used to estimate the probability that the adaptive, rate-optimal test rejects the parametric model when it is correct and the test's power against the alternatives. To provide a basis for judging whether the test's power is high or low, the powers of the tests of Andrews (1997) and Härdle and Mammen (1993) are also estimated by Monte Carlo simulation. In all experiments, the nominal probability of rejecting a correct null hypothesis is 0.05. The computing time required for the experiments is lengthy because all of the tests use of Monte Carlo or bootstrap methods to obtain critical values. Accordingly, the designs of the experiments are simple so as to minimize the time required to compute the test statistics.

The null-hypothesis model in the experiments is

(4.3) $$Y_i = b_0 + b_1 X_i + e_i; \quad i = 1,2,\ldots,250$$

where each X_i is a scalar that is sampled from the $N(0,25)$ distribution truncated at its 5th and 95th percentiles. In experiments where (4.3) is correct (H_0 is true), $b_0 = b_1 = 1$. The e_i's were sampled independently from three distributions, depending on the experiment. These are $N(0,4)$, a variance mixture of normals in which e_i is sampled from $N(0,1.56)$ with probability 0.9 and from $N(0,25)$ with probability 0.1, and the Type I extreme value distribution scaled to have a variance of 4. The mixture distribution is leptokurtic with a variance of 3.9, and the Type I extreme value distribution is asymmetrical.

The alternative models have the form

(4.4) $$Y_i = 1 + X_i + (5/t)f(X_i/t) + e_i,$$

where the e_i's are sampled from one of the three distributions just described and $t = 1$ or 0.25, depending on the experiment. Figure 1 plots the function $f(x) = 1 + x + (5/t)f(x/t)$ for each value of t. The example of Section 4.1 suggests that the power of the adaptive, rate-optimal test should be high compared to the powers of the tests of Andrews (1997) and Härdle and Mammen (1993) in the case t = 0.25, where the difference between the null and alternative models consists of a narrow peak. The power advantage of the adaptive, rate-optimal test is likely to be less or even non-existent under the more moderate case t = 1. However, Theorem 3 suggests that the power of the adaptive, rate optimal test should be satisfactory in comparison to the powers of the other tests when t = 1.

The X_i's were sampled once from the specified distribution and held fixed in repeated realizations of the Y_i's. The values of b_0 and b_1 were estimated by ordinary least squares. Equation (2.9) was used to estimate s^2 in experiments with the adaptive, rate-optimal test. The Härdle-Mammen test does not require an estimator of s^2. In experiments with Andrews' test and e_i's with the normal or extreme value distribution, the distribution of the e_i's was assumed to be known up to s^2, which was estimated from (2.9). In experiments with Andrews' test and e_i's with the mixture-of-normals distribution, the mixing probabilities, 0.9 and 0.1, were assumed to be known *a priori*. The variances of the normal components of the mixture were estimated from estimates of the variance and fourth central moment of the e_i's. The variance was estimated from (2.9). The fourth central moment was estimated by

$$s_{4n} = \frac{1}{2(n-1)}\sum_{i=1}^{n-1}(Y_{(i+1)} - Y_{(i)})^4 - 6s_n^2 .$$

The kernel used for the adaptive, rate-optimal test and the test of Härdle and Mammen (1993) is $K(u) = (15/16)(1-u^2)^2 I(|u| \leq 1)$.

Implementing the test of Härdle and Mammen (1993) requires selecting a bandwidth parameter, h. Existing theory provides no guidance on how this should be done in applications. We found through preliminary simulations that in all of our experiments, the power of the test is maximized near h = 3.5 and varies little over the range $3 \leq h \leq 4$. Accordingly, we used h = 3.5 for all experiments with the test of Härdle and Mammen (1993). The set of bandwidths for the adaptive, rate optimal test was {2.5, 3, 3.5, 4, 4.5} in all of the experiments.

The experiments were carried out in GAUSS using GAUSS pseudo-random number generators. There were 1000 Monte Carlo replications in the experiments in which H_0 is true and 250 in the experiments in which H_0 is false. The larger number of replications for the experiments with a true H_0 insures that the probabilities of Type I errors are estimated reasonably precisely. The lower number of replications with a false H_0 conserves computing time while providing sufficient precision to be informative about the relative powers of the tests. Bootstrap critical values for the tests of Andrews (1997) and Härdle and Mammen (1993) were computed from 99 bootstrap resamples. There were 99 replications in the Monte Carlo procedure that was used to estimate the critical value of the adaptive, rate-optimal test.

The results of the experiments are presented in Table 1. When H_0 is true, all tests have empirical rejection probabilities that are close to the nominal probability of 0.05. None of the empirical rejection probabilities differs from the nominal rejection probability at the 0.01 level. The power of the adaptive, rate-optimal test is much higher than the powers of the other tests when H_0 is false and $t = 0.25$. All of the differences between the powers of the adaptive, rate-optimal test and the other tests are significant at the 0.01 level when $t = 0.25$. The power of the adaptive, rate-optimal test is similar to that of the Härdle-Mammen test but greater than that of Andrews' test ($p < 0.01$) when H_0 is false and $t = 1$. Thus, the simulation results are consistent with the expectation based on theory that the adaptive, rate-optimal test has higher power than the other tests in the presence of a relatively extreme alternative and has satisfactory power in comparison to the others in the presence of a more moderate alternative.

5. CONCLUSIONS

This paper has developed a new test of a parametric model of a conditional mean function against a nonparametric alternative. The test adapts to the unknown smoothness of the alternative model and is uniformly consistent against alternative models whose distance from the parametric model converges to zero at the fastest possible rate. This rate is slower than $n^{-1/2}$. Some existing tests have non-trivial power against local alternative models whose distance from the null hypothesis decreases at the rate $n^{-1/2}$. However, this rate is not achievable uniformly over reasonable classes of alternatives. As a consequence, there are situations in which the new test has much higher finite-sample power than do tests that have non-trivial power against $n^{-1/2}$ local alternatives. The new test is consistent (though not uniformly) against local alternatives whose distance from the null hypothesis decreases at a rate that is only slightly slower than $n^{-1/2}$. This property provides some protection against the occurrence of situations in which the power of the new test is much lower than that of existing tests. The predictions of theory have been illustrated numerically by the results of a small set of Monte Carlo experiments.

APPENDIX

Sections A.1-A.4 present technical lemmas that are used in the proofs of Theorems 1-4. The proofs of the theorems are in Section A.5. It is assumed throughout that Assumptions 1-6 hold. To minimize the complexity of the notation, it is assumed that $d = 1$. The generalization to the case $d > 1$ is straightforward but requires more complicated vector notation. The structure of the proofs is as follows. In Lemma 10, we show that under H_0, T^* has the same limiting distribution as the version of T^* that is obtained by sampling from the model $Y_i = F(X_i, q_0) + s(X_i)w_i$, where the w_i 's are independently distributed as $N(0,1)$. This result forms the basis of the proof of Theorem 1. Lemma 13 shows that $\boldsymbol{P}(T^* > t_a) \to 1$ as $n \to \infty$ whenever the distance between the parametric family $\Im$ and $f(\cdot)$ exceeds a specified value. This result forms the basis of the proofs of Theorems 2-4.

A.1 *Moments of $S_h(q)$*

Lemma 1: *Let A be a n×n symmetrical matrix whose* (i,j) *element is* a_{ij}. *Let* $\{e_i$: $i = 1, \ldots, n\}$ *be independent random variables with* $\boldsymbol{E}e_i^2 = 0$, $\boldsymbol{E}e_i^2 = s_i^2$, *and* $\boldsymbol{E}e_i^4 = s_i$. *Then*

$$\boldsymbol{E}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}e_i e_j = \sum_{i=1}^{n} a_{ii}s_i^2$$

and

$$\text{Var}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}e_i e_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} 2a_{ij}^2 s_i^2 s_j^2 + \sum_{i=1}^{n} a_{ii}^2(s_i - 3s_i^4)\,.$$

Proof: Obvious. Q.E.D.

Lemma 2: *There are positive constants* C_{N1}, C_{N2}, C_N, C_{V1}, *and* C_{V2} *that depend only on* C_1 *and* C_2 *in Assumption* 3, *on* C_E *in Assumption* 5, *and on K such that for all* $h \in H_n$: (i) $C_{N1}h^{-1} \le N_h \le C_{N2}h^{-1}$, (ii) $C_{V1}h^{-1} \le V_h^2 \le C_{V2}h^{-1}$, *and* (iii) $\|W_h' W_h\|_\infty \le C_N$.

Proof: Assumptions 3 and 4 imply that for all i

$$\text{(A1)}\qquad \sum_{j=1}^{n} K\left(\frac{X_i - X_j}{h}\right) \le M_h(X_i) \le C_2 nh\,,$$

$$\text{(A2)}\qquad \sum_{j=1}^{n} K\left(\frac{X_i - X_j}{h}\right)^2 \le M_h(X_i) \le C_2 nh\,,$$

$$\text{(A3)}\qquad \sum_{j=1}^{n} K\left(\frac{X_i - X_j}{h}\right) \ge kM_{h/2}(X_i) \ge kC_1 nh/2\,,$$

and

(A4) $$\sum_{j=1}^{n} K\left(\frac{X_i - X_j}{h}\right)^2 \geq k^2 M_{h/2}(X_i) \geq k^2 C_1 nh/2 .$$

Therefore,

(A5) $$\frac{K_h(X_i - X_j)}{C_2 nh} \leq w_h(X_i, X_j) \leq \frac{K_h(X_i - X_j)}{kC_1 nh/2},$$

$$\frac{k^2 C_1 nh/2}{(C_2 nh)^2} \leq \sum_{j=1}^{n} w_h(X_i, X_j)^2 \leq \frac{C_2 nh}{k^2 C_1^2 (nh/2)^2}$$

and the first assertion follows.

Next, since all elements of the matrix $A_h = W_h{}'W_h$ are non-negative,

$$\|A_h\|_\infty \leq \max_{1 \leq i \leq n} \sum_{j=1}^{n} a_{ij,h} .$$

Using (A1) and $\sum_{j=1}^{n} w_h(X_k, X_j) = 1$, we obtain for every i, $k \leq n$,

$$\sum_{j=1}^{n} a_{ij,h} = \sum_{j=1}^{n} \sum_{k=1}^{n} w_h(X_k, X_i) w_h(X_k, X_j) = \sum_{k=1}^{n} w_h(X_k, X_i) \leq \frac{C_2 nh}{kC_1 nh/2},$$

and the third assertion follows.

Now, the Cauchy-Schwarz inequality and (A2)-(A4) yield

(A6) $$a_{ij,h}^2 \leq \sum_{k=1}^{n} w_h(X_k, X_i)^2 \sum_{k=1}^{n} w_h(X_k, X_j)^2 \leq \left[\frac{C_2 nh}{(kC_1 nh/2)^2}\right]^2 \leq (C/nh)^2$$

for a suitable constant C. These inequalities give the bound

$$V_h^2 = 2\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij,h}^2 s^2(X_i) s^2(X_j) \leq 2n\left[\max_{1 \leq i \leq n} s^4(X_i)\right]\left(\max_{1 \leq i, j \leq n} a_{ij,h}\right) \max_{1 \leq i \leq n} \sum_{j=1}^{n} a_{ij,h}$$

$$\leq 2n\left[\max_{1 \leq i \leq n} s^4(X_i)\right] \frac{2CC_2}{nh}.$$

A similar argument bounds V_h^2 from below, thereby yielding (ii). Q.E.D.

A.2 *Bounding $b_h(q)$*

Lemma 3: *Let C_{11} be as in Assumption* 1 *and C_N be as in Lemma* 2. *For every d* > 0

$$\max_{h \in H_n} \sup_{q \in \Theta: \|q - q_0\| \leq d} \|b_h(q)\|^2 \leq C_{11}^2 C_N n d^2 .$$

Proof: By Assumption 1(i) and the mean value theorem, $\|F(q)-F(q_0)\|^2 \le C_{11}^2\|q-q_0\|^2$. Therefore,

$$\|b_h(q)\|^2 = \|W_h[F(q)-F(q_0)]\|^2$$

$$= [F(q)-F(q_0)]'W_h'W_h[F(q)-F(q_0)]$$

$$\le \|W_h'W_h\|_\infty \|F(q)-F(q_0)\|^2$$

$$\le C_N \sum_{i=1}^{n} C_{11}^2\|q-q_0\|^2 \le C_{11}^2 C_N n d^2.$$

Q.E.D.

Lemma 4: *As* $n \to \infty$:

$$J_n^{-1/2} \max_{h \in H_n} V_h^{-1}\|\nabla_q F(q)'W_h'W_h e\| = O_p(1)$$

and

$$J_n^{-1/2} \max_{h \in H_n} V_h^{-1}\|W_h e\| = O_p(1)\ .$$

Proof: To obtain the first result, it suffices to show that for some constant $C < \infty$

$$R_{n,1} \equiv J_n^{-1} \sum_{h \in H_n} V_h^{-2}\boldsymbol{E}\|\nabla_q F(q_0)'W_h'W_h e\|^2 \le C\ .$$

Using Assumption 1(i), we obtain

$$\boldsymbol{E}\|\nabla_q F(q_0)'W_h'W_h e\|^2 = \boldsymbol{E}tr[\nabla_q F(q_0)'W_h'W_h ee'W_h'W_h \nabla_q F(q_0)]$$

$$\le \left[\max_{1\le i\le n} s^2(X_i)\right] tr[\nabla_q F(q_0)'(W_h'W_h)^2\nabla_q F(q_0)]$$

$$\le \left[\max_{1\le i\le n} s^2(X_i)\right] C_{11}^2 tr(W_h'W_h)^2.$$

Therefore,

$$R_{n,1} \le J_n^{-1} \sum_{h \in H_n} V_h^{-2}\left[\max_{1\le i\le n} s^2(X_i)\right] C_{11}^2 tr(W_h'W_h)^2$$

$$\le \frac{\left[\max_{1\le i\le n} s^2(X_i)\right] C_{11}^2 tr(W_h'W_h)^2}{2\left[\min_{1\le i\le n} s^4(X_i)\right] tr(W_h'W_h)^2}.$$

The first result now follows from Assumption 5.

To prove the second result, it suffices to show that

$$R_{n,2} \equiv J_n^{-1} \sum_{h \in H_n} V_h^{-2} \boldsymbol{E}\|W_h e\|^2 \le C$$

for some $C < \infty$. Using Lemma 2, we get

$$R_{n,2} = J_n^{-1} \sum_{h \in H_n} V_h^{-2} N_h \le J_n^{-1} \sum_{h \in H_n} C_{N2} C_{V1}^{-1} \le C_{N2} C_{V1}^{-1},$$

which proves the second result. Q.E.D.

The following result is a corollary of Lemma 4.

Lemma 5: *Let* H_0 *hold. Then for each* $u > 0$

$$\max_{h \in H_n} \sup_{q \in \Theta: \|q - q_0\| \le n^{-1/2} u} V_h^{-1} \|b_h(q)' W_h e\| = O_p(J_n^{1/2} n^{-1/2}).$$

The following result holds when H_0 is false.

Lemma 6: *Given* $h \in H_n$, *let* $B_h = \|W_h[\bar{f} - F(q_0)]\|$. *If* $B_h \ge V_h$, *then for every* $u > 0$ *and* $d > 0$,

$$\boldsymbol{P}\left(\sup_{q \in \Theta: \|q - q_0\| \le n^{-1/2} u} \|[\bar{f} - F(q)] W_h' W_h e\| \ge d B_h^2 \right) = o(1)$$

as $n \to \infty$.

Proof: Assumption 1(i) and a Taylor series approximation to $F(q) - F(q_0)$ give

$$\sup_{q \in \Theta: \|q - q_0\| \le n^{-1/2} u} \|b_h(q) W_h e\| \le \|[\bar{f} - F(q_0)]' W_h' W_h e\|$$

$$+ n^{-1/2} u \|\nabla_q F(q_0)' W_h' W_h e\| + n^{-1/2} C_{12}^{1/2} u^2 \|W_h e\|.$$

By this result and Lemma 4, it suffices to prove that $B_h^{-4} \boldsymbol{E}\|[\bar{f} - F(q_0)]' W_h' W_h e\|^2 = o(1)$ as $n \to \infty$. Use Lemma 2 to obtain

$$B_h^{-4} \boldsymbol{E}\|[\bar{f} - F(q_0)]' W_h' W_h e\|^2$$

$$\le \left[\max_{1 \le i \le n} s^2(X_i) \right] B_h^{-4} [\bar{f} - F(q_0)]' (W_h' W_h)^2 [\bar{f} - F(q_0)]$$

$$\le \left[\max_{1 \le i \le n} s^2(X_i) \right] B_h^{-4} \|W_h' W_h\|_\infty [\bar{f} - F(q_0)]' (W_h' W_h) [\bar{f} - F(q_0)]$$

$$= \left[\max_{1 \le i \le n} s^2(X_i) \right] B_h^{-4} \| W_h' W_h \|_\infty \| W_h' [\bar{f} - F(q_0)] \|^2$$

$$= \left[\max_{1 \le i \le n} s^2(X_i) \right] B_h^{-2} C_N .$$

Since $B_h^2 \ge V_h^2$, the result follows from Lemma 2 and $h_{max} = o(1)$ as $n \to \infty$. Q.E.D.

A.3 *Sequences of Local Alternative Models*

Write the local alternative model (1.2) in the form $\bar{f}_n = F(q_1) + r_n \bar{g}$, $q_1 \in \Theta$, where $\bar{f}_n$ and $\bar{g}$ are as defined in Section 3.4. Define

$$q_{0,n} = \arg \inf_{q \in \Theta} \| \bar{f}_n - F(q) \| .$$

This quantity exists for all sufficiently large n. Let I_n denote the $n \times n$ identity matrix.

Lemma 7: *Define* $\bar{g}^{\perp} = (I_n - \Pi_1)\bar{g}$, *where* Π_1 *is as defined in Section* 3.4. *Then*

$$\| \bar{f}_n - F(q_{0,n}) - r_n \bar{g}^{\perp} \| = o(1)$$

as $n \to \infty$. *Moreover, the least-squares estimator* q_n *satisfies*

$$\| F(q_n) - F(q_{0,n}) \|^2 = O_p(1)$$

as $n \to \infty$.

Proof: See Millar (1982, Theorem 3.6). Q.E.D.

A.4 *Gaussian Approximation of Quadratic Forms*

This section presents properties of the centered, normalized quadratic forms $T_h = \hat{V}_h^{-1}[S_h(q_n) - \hat{N}_h]$ and $T_{h0} = [S_h(q^*) - N_h]/V_h$. Lemma 8 shows that $T_h = T_{h0} + o_p(1)$ for all h. Let $\tilde{e}_i = s(X_i) w_i$ (i = 1, …, n), where the w_i's are independently distributed as $N(0,1)$. Define $\tilde{T}_{h0} = [\| W_h \tilde{e} \|^2 - N_h]/V_h$. Lemmas 9-10 show that under H_0, $\max_{h \in H_n} T_h$ and $\max_{h \in H_n} \tilde{T}_{h0}$ have identical asymptotic distributions. This result is used in the proof of Theorem 1 to justify the simulation method for estimating the critical value of T^*. Lemmas 11-14 provide results that are used in the proofs of Theorems 2-3.

Define $Y_i^* = F(X_i, q_n) + \tilde{e}_i$ (i = 1, …, n). Let $\hat{q}_n$ and $\hat{s}_n^2(X_i)$ be the estimators of q_n and $s^2(X_i)$ that are obtained from the data set $\{Y_i^*, X_i\}$. Let $\hat{T}_h$ be the version of T_h that is obtained by replacing θ_n with $\hat{\theta}_n$, and $\sigma_n^2(X_i)$ with $\hat{\sigma}_n^2(X_i)$, and e_i with $\hat{\sigma}_n^2(X_i)w_i$ in (2.4)-(2.6).

Lemma 8: *Let* $\sigma_n^2(X_i) - \sigma^2(X_i) = o_p(h_{\min}^{1/2})$ *uniformly over* i = 1, ..., *n*. *Then* $T_h = T_{h0} + o_p(1)$ *and* $\hat{T}_h = \tilde{T}_{h0} + o_p(1)$ *uniformly over* $h \in H_n$.

Proof: This result follows from Lemmas 1 and 2 and an application of the delta method. Q.E.D.

Lemma 9: *As* $n . \to \infty$,

$$\max_{h \in H_n} V_h^{-1} \sum_{i=1}^{n} a_{ii,h}(e_i^2 - \sigma^2) = o_p(1)\,.$$

Proof: It suffices to show that

$$R_n \equiv \sum_{h \in H_n} V_h^{-2} \boldsymbol{E}\left[\sum_{i=1}^{n} a_{ii,h}(e_i^2 - \sigma^2)\right]^2 = o(1)$$

as $n \to \infty$. Taking the expected value gives

$$R_n = \sum_{h \in H_n} V_h^{-2} \left[\sum_{i=1}^{n} a_{ii,h}^2 (\sigma_4 - \sigma^4)\right]^2 .$$

By Assumption 5, $\sigma_4 \le \sigma^4 C_E^4$. By Lemma 2, $V_h^{-2} \le C_{V1}^{-1} h$ and $a_{ii,h} \le C_N (nh)^{-1}$. Therefore,

$$R_n \le \sum_{h \in H_n} C_{V1}^{-1} h \left[\sum_{i=1}^{n} C_N^2 (nh)^{-2} \sigma^4 C_E^4\right]$$

$$\le n^{-1} C_{V1}^{-1} C_N^2 \sigma^4 C_E^4 \sum_{h \in H_n} h$$

$$\le n^{-1} C_{V1}^{-1} C_N^2 \sigma^4 C_E^4 h_{\max} (1-a)^{-1}$$

The lemma now follows. Q.E.D.

Lemma 10: *Let* H_0 *be true. Then* $\max_{h \in H_n} T_{h0}$ *and* $\max_{h \in H_n} \tilde{T}_{h0}$ *have identical asymptotic distributions*.

Proof: By Lemmas 8 and 9, it suffices to show that the joint distributions of $V_h^{-1} \sum_{i \ne j} a_{ij,h} e_i e_j$ $(h \in H_n)$ and $V_h^{-1} \sum_{i \ne j} a_{ij,h} \tilde{e}_i \tilde{e}_j$ $(h \in H_n)$ are asymptotically the same. For $h \in H_n$, and $x_i = e_i$ or $\tilde{e}_i$ $(i = 1,...,n)$, define

$$B_{hn}(x_1,...,x_n)=V_h^{-1}\sum_{i\neq j}a_{ij,h}x_ix_j \; .$$

Let $B_n(x_1,...,x_n)$ be the vector that is obtained by stacking $B_{hn}(x_1,...,x_n)$ $(h\in H_n)$. Let g_n be a 3-times continuously differentiable function on $\Re^{J_n}$. Define

$$g_{3n}=\sup_{v\in\Re^{J_n}}\max_{i,j,k=1,...,J_n}\left|\frac{\partial^3 g_n(v)}{\partial v_i\partial v_j\partial v_k}\right| .$$

The proof takes place in two steps. The first step is to show that

(A7) $$|\boldsymbol{E}g[B_n(e_1,...,e_n)]-\boldsymbol{E}g[B_n(\tilde{e}_1,...,\tilde{e}_n)]|\leq c_H g_{3n}\left(\frac{J_n^{3/2}}{n^{1/2}h_{\min}^{3/2}}\right)$$

for any 3-times differentiable g, some finite constant c_H, and all sufficiently large n. The second step uses (A7) to prove that $V_h^{-1}\sum_{i\neq j}a_{ij,h}e_ie_j$ $(h\in H_n)$ and $V_h^{-1}\sum_{i\neq j}a_{ij,h}\tilde{e}_i\tilde{e}_j$ $(h\in H_n)$ have the same asymptotic distribution.

Step 1: Define $b_{ij,h}=a_{ij,h}/V_h$. Assume without loss of generality that $s(X_i)=1$ for all i = 1, ..., n. [If $s(X_i)\neq 1$, replace e_i with $e_i/s(X_i)$, $\tilde{e}_i$ with $\tilde{e}_i/s(X_i)$, and $b_{ij,h}$ with $b_{ij,h}s(X_i)s(X_j)$.] It is easily shown that

(A8) $$\|\boldsymbol{E}g[B_n(e_1,...,e_n)]-\boldsymbol{E}g[B_n(\tilde{e}_1,...,\tilde{e}_n)]\|$$

$$\leq\sum_{i=1}^{n}\|\boldsymbol{E}g[B_n(e_1,...,e_i,\tilde{e}_{i+1},...\tilde{e}_n)]-\boldsymbol{E}g[B_n(e_1,...e_{i-1},\tilde{e}_i,...,\tilde{e}_n)]\|,$$

where $B_n(e_1,...,e_n,\tilde{e}_{n+1})\equiv B_n(e_1,...,e_n)$ and $B_n(e_0,\tilde{e}_1,\tilde{e}_n)\equiv B_n(\tilde{e}_1,...,\tilde{e}_n)$. We now derive an upper bound on the last term of the sum on the right-hand side of (A8). Similar bounds can be derived for the other terms. Let u_{n-1},Δ_n, and $\tilde{\Delta}_n$, respectively, denote the vectors that are obtained by stacking

$$u_{h,n}=\sum_{i=1}^{n-1}\sum_{\substack{j=1\\ j\neq i}}^{n-1}b_{ij,h}e_ie_j \; ,$$

$$\Delta_{h,n}=2e_n\sum_{i=1}^{n-1}b_{in,h}e_i \; ,$$

and

$$\tilde{\Delta}_{h,n} = 2\tilde{e}_n \sum_{i=1}^{n-1} b_{in,h} e_i \ .$$

Then a Taylor-series expansion of the last term of the sum on the right-hand side of (A8) about $e_n = \tilde{e}_n = 0$ yields

$$\left\| \boldsymbol{E} g[B_n(e_1,...,e_n)] - \boldsymbol{E} g[B_n(e_1,...,e_{n-1},\tilde{e}_n)] \right\| \le |\boldsymbol{E} g'(u_{n-1})(\Delta_n - \tilde{\Delta}_n)|$$

$$+ (1/2)|\boldsymbol{E}[\Delta_n' g''(u_{n-1})\Delta_n - \tilde{\Delta}_n{}' g''(u_{n-1})\tilde{\Delta}_n]| + (g_{3n}/6)(\boldsymbol{E}\|\Delta_n\|^3 + \boldsymbol{E}\|\tilde{\Delta}_n\|^3),$$

where g' and g'', respectively, denote the gradient and matrix of second derivatives of g. Since e_n and $\tilde{e}_n$ are independent of $e_1,...,e_{n-1}$, $\boldsymbol{E} e_n = \boldsymbol{E}\tilde{e}_n = 0$, and $\boldsymbol{E} e_n^2 = \boldsymbol{E}\tilde{e}_n^2 = 1$, we have

$$\boldsymbol{E}(\Delta_n - \tilde{\Delta}_n | e_1,...,e_{n-1}) = \boldsymbol{E}[(\Delta_n\Delta_n' - \tilde{\Delta}_n\tilde{\Delta}_n') | e_1,...,e_{n-1}] = 0 \ .$$

Therefore,

(A9) $$|\boldsymbol{E} g[B_n(e_1,...,e_n)] - \boldsymbol{E} g[B_n(\tilde{e}_1,...,\tilde{e}_n)]| \le (g_{3n}/6)(\boldsymbol{E}\|\Delta_n\|^3 + \boldsymbol{E}\|\tilde{\Delta}_n\|^3).$$

To find bounds on $\boldsymbol{E}\|\Delta_n\|^3$ and $\boldsymbol{E}\|\tilde{\Delta}_n\|^3$, let b_{in} be the vector that is obtained by stacking $b_{in,h}$ (h = 1,. , J_n). Then Hölder's inequality gives

$$\boldsymbol{E}\|\Delta_n\|^3 = \boldsymbol{E}|e_n|^3 \boldsymbol{E}\left\| 2\sum_{i=1}^{n-1} b_{in} e_i \right\|^3$$

$$\le 8\boldsymbol{E}|e_n|^3 \left\{ \boldsymbol{E}\left[\sum_{h\in H_n} \left(\sum_{i=1}^{n-1} b_{in,h} e_i \right)^2 \right]^2 \right\}^{3/4}$$

$$= 8\boldsymbol{E}|e_n|^3 \left(\boldsymbol{E} \sum_{h\in H_n} \sum_{s\in H_n} \sum_{i,j,k,\ell=1}^{n-1} b_{in,h} b_{jn,h} b_{kn,s} b_{\ell n,s} e_i e_j e_k e_\ell \right)^{3/4}$$

$$= 8\boldsymbol{E}|e_n|^3 \left\{ \boldsymbol{E} \sum_{h\in H_n} \sum_{s\in H_n} \left[\sum_{\substack{i,j=1 \\ i\ne j}}^{n-1} (b_{in,h}^2 b_{jn,s}^2 + 2 b_{in,h} b_{in,s} b_{jn,h} b_{jn,s}) + \sum_{i=1}^{n-1} b_{in,h}^2 b_{in,s}^2 \boldsymbol{E}(e_i^4) \right] \right\}^{3/4}$$

$$\le c\left(\frac{J_n}{n h_{\min}} \right)^{3/2}$$

for some finite $c > 0$, where the last line follows from Lemma 2 and (A6). A similar result holds for $\boldsymbol{E}\|\tilde{\Delta}_n\|^3$. Therefore

$$\boldsymbol{E}\|\Delta_n\|^3 + \boldsymbol{E}\|\tilde{\Delta}_n\|^3 \le 2c\left(\frac{J_n}{nh_{\min}}\right)^{3/2},$$

and (A9) gives

$$\|\boldsymbol{E}g[B_n(e_1,...,e_n)] - \boldsymbol{E}g[B_n(e_1,...,e_{n-1},\tilde{e}_n)]\| \le (cg_{3n}/3)\left(\frac{J_n}{nh_{\min}}\right)^{3/2}.$$

Similar bounds hold for the other terms of the sum on the right-hand side of (A8). Summing the bounds yields (A7).

Step 2: It suffices to show that for any real z

$$\lim_{n\to\infty}\left\{\boldsymbol{P}\left[\max_{h\in H_n} B_{hn}(e_1,...,e_n) \le z\right] - \boldsymbol{P}\left[\max_{h\in H_n} B_{hn}(\tilde{e}_1,...,\tilde{e}_n) \le z\right]\right\} = 0$$

or, equivalently, that

$$\lim_{n\to\infty}\left|\boldsymbol{E}\prod_{h\in H_n} I[B_{hn}(e_1,...,e_n) \le z] - \boldsymbol{E}\prod_{h\in H_n} I[B_{hn}(\tilde{e}_1,...,\tilde{e}_n) \le z]\right| = 0 .$$

Let g be a non-decreasing function that is 3 times continuously differentiable on the real line and satisfies $g(v) = 0$ if $v \le -1$ and $g(v) = 1$ if $v \ge 0$. Let $d_n = J_n^{-2}$. Some algebra shows that

$$\text{(A10)} \quad \left|\boldsymbol{E}\prod_{h\in H_n} I[B_{hn}(e_1,...,e_n) \le z] - \boldsymbol{E}\prod_{h\in H_n} I[B_{hn}(\tilde{e}_1,...,\tilde{e}_n) \le z]\right|$$

$$\le \left|\boldsymbol{E}\prod_{h\in H_n} g\left[\frac{B_{hn}(e_1,...,e_n) - z}{d_n}\right] - \boldsymbol{E}\prod_{h\in H_n} g\left[\frac{B_{hn}(\tilde{e}_1,...,\tilde{e}_n) - z}{d_n}\right]\right|$$

$$+ \sum_{h\in H_n} \boldsymbol{E}\left|g\left[\frac{B_{hn}(e_1,...,e_n) - z}{d_n}\right] - I[B_{hn}(e_1,...,e_n) \le z]\right|$$

$$+ \sum_{h\in H_n} \boldsymbol{E}\left|g\left[\frac{B_{hn}(\tilde{e}_1,...,\tilde{e}_n) - z}{d_n}\right] - I[B_{hn}(\tilde{e}_1,...,\tilde{e}_n) \le z]\right|.$$

Each term of the summands of the second two sums on the right-hand side of (A10) is bounded from above by $J_n d_n = J_n^{-1}$. Therefore, using (A7) to bound the first term on the right-hand side of (A10) yields

(A11) $$\left| \boldsymbol{P}\left[\max_{h\in H_n} B_{hn}(e_1,...,e_n) \le z \right] - \boldsymbol{P}\left[\max_{h\in H_n} B_{hn}(\tilde{e}_1,...,\tilde{e}_n) \le z \right] \right| \le \frac{cJ_n^{15/2}}{n^{1/2} h_{\min}^{3/2}} + 2J_n^{-1}.$$

The lemma follows by taking limits as $n \to \infty$ on both sides of (A11). Q.E.D.

Lemma 11: *For any* $z \ge 1$, $h \in H_n$, *and all sufficiently large n*

$$\boldsymbol{P}(\tilde{T}_{h0} > z) \le \exp(-z^2/4).$$

Proof: Write $\tilde{e}' \mathcal{W}_h' \mathcal{W}_h \tilde{e} = w'\Sigma \mathcal{W}_h' \mathcal{W}_h \Sigma w$, where Σ is the diagonal matrix whose (i, i) element is $s(X_i)$ and w is a $n\times 1$ vector of independent $N(0,1)$ variates. Let Λ be the diagonal matrix of eigenvalues of $\Sigma \mathcal{W}_h' \mathcal{W}_h \Sigma$, $\{l_i: i = 1, ..., n\}$ be the eigenvalues, and Π be the orthogonal matrix such that $\Sigma \mathcal{W}_h' \mathcal{W}_h \Sigma = \Pi' \Lambda \Pi$. Define $Z = \Pi w$. Then the elements of Z are independent $N(0,1)$ variates,

$$\tilde{e}' \mathcal{W}_h' \mathcal{W}_h \tilde{e} = \sum_{i=1}^n l_i Z_i^2,$$

$$\boldsymbol{E}(\tilde{e}' \mathcal{W}_h' \mathcal{W}_h \tilde{e}) = \sum_{i=1}^n l_i,$$

and

$$V^2 \equiv Var(\tilde{e}' \mathcal{W}_h' \mathcal{W}_h \tilde{e}) = 2\sum_{i=1}^n l_i^2.$$

Therefore,

$$\tilde{T}_{h0} = V^{-1} \sum_{i=1}^n l_i (Z_i^2 - 1).$$

It now follows from the Chebyshev exponential inequality (see, e.g., Loève 1977, p. 160) that for every $m > 0$,

$$Q_n \equiv \boldsymbol{P}(\tilde{T}_{h0} > z) \le e^{-mz} \boldsymbol{E} \exp\left[mV^{-1} \sum_{i=1}^n l_i (Z_i^2 - 1) \right].$$

Since the Z_i are independent $N(0,1)$ variates,

$$\boldsymbol{E} \exp\left[mV^{-1} \sum_{i=1}^n l_i (Z_i^2 - 1) \right] = \prod_{i=1}^n \exp\left[-mV^{-1} l_i - (1/2)\log(1 - 2mV^{-1} l_i) \right]$$

whenever $mV^{-1} l_i < 1$. It follows from Lemma 2 and Assumption 5 that $V^{1} l_i < d$ for any $d > 0$ and all sufficiently large n. Therefore, using the inequality $-\log(1 - u) \le u + u^2$ for all sufficiently small $u > 0$, we have

$$\boldsymbol{E}\exp\left[mV^{-1}\sum_{i=1}^{n} l_i(Z_i^2-1)\right] \le \prod_{i=1}^{n}\exp\left(2m^2V^{-1}l_i^2\right) = \exp(-m^2)\,,$$

and

$$Q_n \le \exp(-mz+m^2)$$

for all sufficiently large n. The lemma follows by setting $m = z/2$. Q.E.D.

For $0 < a < 1$, define $\tilde{t}_a$ to be the $1 - a$ quantile of $\max_{h \in H_n} \tilde{T}_{h0}$.

Lemma 12: *For all sufficiently large n,* $\tilde{t}_a \le 2\sqrt{\log J_n - \log a}$.

Proof: Let $z \ge 1$. By Lemma 11,

$$\boldsymbol{P}\left(\max_{h \in H_n} \tilde{T}_{h0} > z\right) \le \sum_{h \in H_n} \boldsymbol{P}(\tilde{T}_{h0} > z)$$

$$\le \sum_{h \in H_n} \exp\left(-\frac{z^2}{4}\right)$$

$$= J_n \exp\left(-\frac{z^2}{4}\right).$$

Therefore,

$$a \le J_n \exp\left(\frac{-\tilde{t}_a^{\,2}}{4}\right).$$

The Lemma follows by taking logarithms on both sides of this inequality. Q.E.D.

Lemma 13: *Let* $\tilde{t}_a{}^* = \max\left(\tilde{t}_a, \sqrt{2\log J_n + \sqrt{2\log J_n}}\right)$. *Suppose that* $\left\|W_h[\bar{f} - \bar{F}(q^*)]\right\|^2 \ge 4V_h\tilde{t}_a{}^*$ *for some* $h \in H_n$*. Then*

$$\lim_{n\to\infty} \boldsymbol{P}(T^* > t_a) = 1\,.$$

Proof: By Lemma 8, T^* can be replaced by $\max_{h \in H_n} T_{h0}$. By Lemmas 8 and 10, t_a can be replaced by $\tilde{t}_a$. Thus, it suffices to prove that

$$\lim_{n\to\infty} \boldsymbol{P}(\max_{h \in H_n} T_{h0} > \tilde{t}_a) = 1\,,$$

which holds if

$$\lim_{n\to\infty} \boldsymbol{P}(T_{h0} > \tilde{t}_a) = 1$$

for some $h \in H_n$. For any $h \in H_n$,

$$T_{h0} = \tilde{T}_{h0} + \frac{\|b_h(q^*)\|^2 + 2b_h'(q^*)W_h e}{V_h} .$$

Therefore, by Lemma 6,

$$T_{h0} = \tilde{T}_{h0} + \frac{\|b_h(q^*)\|^2}{V_h} + o_p(1) ,$$

and

$$\lim_{n\to\infty} \boldsymbol{P}(T_{h0} > \tilde{t}_a) = \lim_{n\to\infty} \boldsymbol{P}(\tilde{T}_{h0} + \frac{\|b_h(q^*)\|^2}{V_h} > \tilde{t}_a) .$$

But $\|b_h(q^*)\|^2 = \|W_h[\bar{f} - \bar{F}(q^*)]\|^2$. Therefore, $\|W_h[\bar{f} - \bar{F}(q^*)]\|^2 \geq 4V_h \tilde{t}_a^*$,

$$\lim_{n\to\infty} \boldsymbol{P}(\tilde{T}_{h0} + \frac{\|b_h(q^*)\|^2}{V_h} > \tilde{t}_a) \geq \lim_{n\to\infty} \boldsymbol{P}(\tilde{T}_{h0} > \tilde{t}_a - 4\tilde{t}_a^*) \to 1$$

as $n \to \infty$ because $\tilde{T}_{h0}$ is bounded in probability and $\tilde{t}_a - 4\tilde{t}_a^* \to -\infty$ as $n \to \infty$. Q.E.D.

Lemma 14: *Let $h \in H_n$. Let m be the largest integer that is less than s. Let I be a subinterval of* $[0,1]$ *with length* $h_2 = (m + 1)h$. *Let x denote the center of I. Let $V_{h,\ell}$ be the* $(m + 1)\times(m + 1)$ *matrix with elements*

$$v_{k,\ell} = \sum_{i:X_i \in I} \left(\frac{X_i - x}{h}\right)^{k+\ell} .$$

There exists a number R depending only on the constants C_1 and C_2 from Assumption 3 such that

$$\|V_{h,\ell}\|_\infty \leq R$$

and

$$\|V_{h,\ell}^{-1}\|_\infty \leq R .$$

Proof: This result is proved for the case of a regular design in Ingster (1993c) and for the case of a design satisfying Assumption 3 in Härdle, *et al*. (1997, Lemma 6.6). The idea is as follows. To obtain a non-degenerate, non-singular $V_{h,\ell}$, it suffices to have $m + 1$ distinct design points inside the interval I. Under Assumption 3, I contains $O(nh)$ points, which is more than sufficient. Q.E.D.

A.5 *Proofs of Theorems*

Proof of Theorem 1: By Lemma 8, $\max_{h\in H_n} T_h = \max_{h\in H_n} T_{h0} + o_p(1)$. By Lemma 10, $\max_{h\in H_n} T_{h0} \to^d \max_{h\in H_n} \tilde{T}_{h0}$ as $n \to \infty$. A further application of Lemma 8 gives $\max_{h\in H_n} \tilde{T}_{h0} \to^d \max_{h\in H_n} \hat{T}_h + o_p(1)$. Therefore, $\max_{h\in H_n} T_h \to^d \max_{h\in H_n} \hat{T}_h + o_p(1)$. Q.E.D.

Proof of Theorem 2: By Lemma 13, it suffices to show that $\|W_h[\bar{f} - \bar{F}(q^*)]\|^2 \geq 4V_h\tilde{t}_a^*$ for some $h \in H_n$ and all sufficiently large n, where

$$q^* = \arg\inf_{q\in\Theta} \|\bar{f} - F(q)\|^2 .$$

Because $h_{\max} \to 0$ as $n \to \infty$ and $W_h[\bar{f} - \bar{F}(q)]$ is the result of smoothing the continuous function $f(\cdot) - F(\cdot, q)$ by the kernel method, $\|W_h[\bar{f} - \bar{F}(q)]\|^2 \to \|\bar{f} - \bar{F}(q)\|^2$ as $n \to \infty$. But under H_1, $\inf_{q\in\Theta} \|\bar{f} - F(q)\|^2 \geq c_r n$ for some $c_r > 0$ and all sufficiently large n. The result that $\|W_h[\bar{f} - \bar{F}(q^*)]\|^2 \geq 4V_h\tilde{t}_a^*$ now follows from Lemmas 2 and 11. Q.E.D.

Proof of Theorem 3: By Lemma 13, it suffices to show that $B_h^2 \equiv \|W_h[\bar{f} - \bar{F}(q_{0,n})]\|^2 \geq 4V_h\tilde{t}_a^*$ for some $h \in H_n$ and all sufficiently large n, where

$$q_{0,n} = \arg\inf_{q\in\Theta} \|\bar{f}_n - F(q)\|^2 .$$

To show this, use the inequality $a^2 \geq 0.5b^2 - (b-a)^2$ to write

$$B_h^2 \geq 0.5r_n^2\|W_h g^\perp\|^2 - \|W_h[\bar{f}_n - \bar{F}(q_{0,n}) - r_n g^\perp]\|^2 .$$

By Lemmas 2 and 7,

$$\|W_h[\bar{f}_n - \bar{F}(q_{0,n}) - r_n g^\perp]\|^2 \leq \|W_h'W_h\|_\infty \|\bar{f}_n - \bar{F}(q_{0,n}) - r_n g^\perp\|^2 = o(1)$$

as $n \to \infty$. Moreover, because $h_{\max} \to 0$ as $n \to \infty$ and $W_h g^\perp$ is the result of smoothing the continuous function $g^\perp$ by the kernel method, $\|W_h g^\perp\|^2 \to \|g^\perp\|^2$ as $n \to \infty$. Therefore, for sufficiently large n,

$$B_h^2 \geq 0.25r_n^2\|g^\perp\|^2 \geq 0.25r_n^2 d^2\|g\|^2 \geq 0.25nr_n^2 d^2 .$$

Set $h = h_{\max} = C_H(\log\log n)^{-2}$. Then theorem follows from the definition of r_n and Lemma 2. Q.E.D.

Proof of Theorem 4: Let $g = \bar{f} - \bar{F}(q^*)$. Then by Lemma 12, (3.4) and the definition of $S(H,s)$,

(A12) $n^{-1/2}\|g\| \geq C_a(n^{-1}\tilde{t}_a^*)^{2s/(4s+1)}$

and $\|g\|_{H,s} \leq C_g$ for some $C_g < \infty$. By Lemma 13, it suffices to show that $\|W_h g\|^2 \geq 4V_h\tilde{t}_a^*$ for some $h \in H_n$. This is done by approximating g by a piecewise polynomial function and proving that each segment of the polynomial satisfies the required condition.

Set $h_1 = (h^{-1}\tilde{t}_a^*)^{2/(4s+1)}$. Then $nh_1^{2s} = h_1^{-1/2}\tilde{t}_a^*$. Select $h \in H_n$ such that $h_1 \leq h < 2h_1$. It will now be shown that $\|W_h g\|^2 \geq 4V_h\tilde{t}_a^*$ for the selected h. First, observe that by Lemma 2(ii), $V_h \leq C_{V2}h^{-1/2}$. Moreover, since $h \geq h_1$,

$$4V_h\tilde{t}_a^* \leq 4C_2h^{-1/2}\tilde{t}_a^* \leq 4C_2h_1^{-1/2} = 4C_2nh_1^{2s} \leq 4C_2nh^{2s}.$$

Therefore, it suffices to show that

(A13) $\|W_h g\|^2 \geq 4C_{V2}nh^{2s}$.

Let m be the smallest integer less than s. Set $h_2 = (m + 1)h$. Let I be a subinterval of [0,1] with length h_2. Let x denote the center of I. The smoothness assumption $\|g\|_{H,s} \leq C_g$ implies that there exists a polynomial

$$P(u) = b_0 + b_1\frac{u-x}{h} + \ldots + b_m\left(\frac{u-x}{h}\right)^m$$

such that $|g(u) - P(u)| \leq Ch^s$ for all u with $|u - x| \leq h_2/2 + h_1$, where C depends only on C_g and m. Define

$$W_h g(X_i) = \sum_{j=1}^{n} w_h(X_i, X_j)g(X_j).$$

Define $W_h P(X_i)$ similarly. Then, since $w_h(X_i, X_j) = 0$ for all X_j with $\|X_i - X_j\| > h$, $|W_h g(X_i) - W_h P(X_i)| \leq Ch^s$. Moreover,

$$\sum_{i:X_i \in I} |g(X_i)|^2 \le 2 \sum_{i:X_i \in I} |P(X_i)|^2 + 2 \sum_{i:X_i \in I} |g(X_i) - P(X_i)|^2$$

$$\le 2 \sum_{i:X_i \in I} |P(X_i)|^2 + 2 N_I C^2 h^{2s},$$

where N_I denotes the number of design points in *I*. Similarly

$$\sum_{i:X_i \in I} |W_h g(X_i)|^2 \ge \frac{1}{2} \sum_{i:X_i \in I} |W_h P(X_i)|^2 - N_I C^2 h^{2s}.$$

Let $V_{h,\ell}$ be the $(m + 1)\times(m + 1)$ matrix with elements

$$v_{k,\ell} = \sum_{i:X_i \in I} \left(\frac{X_i - x}{h} \right)^{k+\ell}.$$

Let $b = (b_0, \ldots, b_m)'$. Then

$$\sum_{i:X_i \in I} |P(X_i)|^2 = b' V_{h,\ell} b,$$

and, by Lemma 14, $b' V_{h,\ell} b \le R \|b\|^2$. Equivalently, $\|b\|^2 \ge R^{-1} b' V_{h,\ell} b$.

Now define the numbers Z_{ik} ($i = 1, \ldots, n$; $k = 1, \ldots, m$) as the solutions to the equations

$$\left(\frac{Z_{ik} - x}{h} \right)^k = \sum_{j=1}^{n} w_h(X_i, X_j) \left(\frac{X_j - x}{h} \right)^k.$$

Define $\tilde{V}_{h,\ell}$ to be the $(m + 1)\times(m + 1)$ matrix with elements

$$\tilde{v}_{k\ell} = \sum_{i:X_i \in I} \left(\frac{Z_{ik} - x}{h} \right)^k \left(\frac{Z_{i\ell} - x}{h} \right)^\ell \qquad k, \ell = 0, 1, \ldots, m.$$

It is easy to see that $|X_i - Z_{ik}| \le h$ for all $k = 0, 1, \ldots, m$ and for all i with $X_i \in I$. Therefore,, for every k, the sequence $\{Z_{ik}: X_i \in I\}$ satisfies Assumption 3, and Lemma 14 applies to $\tilde{V}_{h,\ell}$. This yields $\|\tilde{V}_{h,\ell}\|_\infty \le R$ and $\|\tilde{V}_{h,\ell}^{-1}\|_\infty \le R$. Next, by definition of Z_{ik},

$$W_h P(X_i) = b_0 + b_1 \frac{Z_{i1} - x}{h} + \ldots + b_m \left(\frac{Z_{im} - x}{h} \right)^m,$$

so that

$$\sum_{i:X_i \in I} |W_h P(X_i)|^2 = b' \tilde{V}_{h\ell} b.$$

Similarly, $b' \tilde{V}_{h\ell} b \ge R^{-1} \|b\|^2$. Therefore,

$$\sum_{i:X_i\in I}|W_h g(X_i)|^2 \geq \frac{1}{2}\sum_{i:X_i\in I}|W_h P(X_i)|^2 - N_I C^2 h^{2s}$$

$$= (1/2) b'\tilde{V}_{h,\ell} b - N_I C^2 h^{2s}$$

$$\geq (1/2) R^{-1}\|b\|^2 - N_I C^2 h^{2s}$$

$$\geq (1/2) R^{-2} b' V_{h,\ell} b - N_I C^2 h^{2s}$$

$$= \frac{1}{2R^2}\sum_{i:X_i\in I}|P(X_i)|^2 - N_I C^2 h^{2s}$$

$$\geq \frac{1}{4R^2}\sum_{i:X_i\in I}|g(X_i)|^2 - \frac{3}{2} N_I C^2 h^{2s}.$$

Now split [0,1] into N intervals, I_1, …, I_N of length no greater than h_2. Applying the foregoing inequality to each interval yields

$$\text{(A14)} \qquad \sum_{i=1}^{n}|W_h g(X_i)|^2 = \sum_{j=1}^{N}\sum_{i:X_i\in I_j}|W_h g(X_i)|^2$$

$$\geq \frac{1}{4R^2}\sum_{j=1}^{N}\sum_{i:X_i\in I_j}|g(X_i)|^2 - \frac{3}{2}\sum_{j=1}^{N} N_{I_j} C^2 h^{2s}$$

$$= \frac{1}{4R^2}\sum_{i=1}^{n}|g(X_i)|^2 - (3/2) n C^2 h^{2s}.$$

Inequality (A14) combined with (A12) implies (A13) for sufficiently large C_a in (3.4). Q.E.D.

FOOTNOTES

1 The fixed design formulation used here includes as special cases random designs in which the distribution of X is absolutely continuous with respect to Lebesgue measure. If (Y, X) is a random variable, then the null hypothesis is that $f(X) = F(X, q)$ almost surely for some $q \in \Theta$. The alternative hypothesis is that $P[f(X) = F(X, q)] < m$ for every $q \in \Theta$ and some $m < 1$.

2 Andrews (1997) assumes that the distribution of e_i in (1.1) is known up to a finite-dimensional parameter. Thus, Andrews tests a parametric model of the conditional distribution of Y not just the conditional mean function. It is not difficult, however, to modify Andrews' test so that it becomes a test of a hypothesis about f alone. See Whang (1998).

3 Triebel (1992) provides definitions of Hölder, Sobolev, and Besov spaces.

4 The condition $s \geq d/4$ is unlikely to be restrictive in applications because the curse of dimensionality makes nonparametric estimation and testing unattractive when d is large. Hart (1997) discusses tests that have the optimal rate of testing when $s < d/4$.

5 Guerre and Lavergne (1999) describe a method for achieving the optimal rate of testing against an alternative of known smoothness. Their test is not adaptive and its behavior against alternatives of the form (1.2) is unknown.

6 Härdle and Mammen (1993) use the integrated squared difference between f_h and F_h. As they note, the properties of their test are the same with summed or integrated squared differences except, possibly, for the values of constants in the expressions for the mean and variance of the test statistic's asymptotic distribution.

7 The variance estimators described in this section are not the only possible ones. For example, Hart (1997, Section 5.3) describes an alternative estimator that is unbiased if X_i is a scalar, $F(x,q)$ is a linear function of x, and the e_i's are homoskedastic. The choice of variance estimator does not affect the asymptotic properties, adaptiveness, or rate optimality of our test. The choice may affect the small-sample performance of the test, but investigation of the small-sample performances of alternative variance estimators is beyond the scope of this paper.

8 If the form of the heteroskedasticity of the e_i's is known, then this knowledge can be used to form a variance estimator. For example, if Y_i is binary, then $s^2(X_i)$ can be estimated by $\hat{f}_n(X_i)[1 - \hat{f}_n(X_i)]$, where $\hat{f}_n(x)$ is a nonparametric estimator of $f(x)$.

TABLE 1: RESULTS OF MONTE CARLO EXPERIMENTS[1]

Distribution *e*	*t*	Andrews' Test	Härdle-Mammen Test	Rate-Optimal Test
		Probability of Rejecting Null Hypothesis		
		Hull Hypothesis Is True		
Normal		0.057	0.060	0.066
Mixture		0.053	0.053	0.054
Extreme Value		0.063	0.057	0.055
		Hull Hypothesis Is False		
Normal	1.0	0.680	0.752	0.792
Mixture	1.0	0.692	0.736	0.796
Extreme Value	1.0	0.600	0.760	0.820
Normal	0.25	0.536	0.770	0.924
Mixture	0.25	0.592	0.704	0.932
Extreme Value	0.25	0.604	0.696	0.968

[1] The differences between empirical and nominal rejection probabilities under H_0 are not significant at the 0.01 level. Under H_1, the differences between the rejection probabilities of the rate-optimal and Andrews' test are significant at the 0.01 level. Under H_1, the differences between the rejection probabilities of the rate-optimal and Härdle-Mammen tests are significant at the 0.01 level when $t = 0.25$.

REFERENCES

Aït-Sahalia, Y., P.J. Bickel, and T.M. Stoker (1994): "Goodness-of-Fit Tests for Regression Using Kernel Methods," Working Paper 3747, Alfred P. Sloan School of Management, Massachusetts Institute of Technology.

Amemiya, T. (1985): *Advanced Econometrics*. Cambridge, MA: Harvard University Press.

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

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

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

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

Buckley, M.J., G.K. Eagleson, and B.W. Silverman (1988): "The Estimation of Residual Variance in Nonparametric Regression," *Biometrika*, 75, 189-199.

Burnashev, M.V. (1979): "On the Minimax Detection of an Inaccurately Known Signal in a White Gaussian Noise Background," *Theory of Probability and Its Applications*, 24, 107-119.

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

de Jong, R.M. (1996): "On the Bierens Test under Data Dependence," *Journal of Econometrics*, 72, 1-32.

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

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

Fan, Y. and Q. Li (1999): "Consistent Model Specification Tests: Kernel-Based Tests Versus Bierens' ICM Tests," working paper, Department of Economics, Texas A&M University.

Gasser, T., L. Sroka, and C. Jennen-Steinmetz (1986): "Residual Variance and Residual Pattern in Nonlinear Regression," *Biometrika*, 73, 625-633.

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

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

Härdle, W., S. Sperlich, and V. Spokoiny (1997): "Component Analysis for Additive Models," discussion paper no. 52, Sonderforschungsbereich 373, Institute for Statistics and Econometrics, Humboldt University, Berlin, Germany.

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

Hong, Y. and H. White (1995): "Consistent Specification Testing via Nonparametric Series Regressions," *Econometrica*, 63, 1133-1160.

Ibragimov, I.A. and Khasminskii, R.Z. (1977): "One Problem of Statistical Estimation in Gaussian White Noise," *Soviet Mathematics Doklady*, 236, 1351-1354.

Ingster, Yu. I. (1982): "Minimax Nonparametric Detection of Signals in White Gaussian Noise." *Problems of Information Transmission*, 18, 130-140.

Ingster, Yu. I. (1993a): "Asymptotically Minimax Hypothesis Testing for Nonparametric Alternatives, I," *Mathematical Methods of Statistics*, 2, 85-114.

Ingster, Yu. I. (1993b): "Asymptotically Minimax Hypothesis Testing for Nonparametric Alternatives, II," *Mathematical Methods of Statistics*, 2, 171-189.

Ingster, Yu. I. (1993c): "Asymptotically Minimax Hypothesis Testing for Nonparametric Alternatives, III," *Mathematical Methods of Statistics*, 2, 249-268.

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

Loève, M. (1977): *Probability Theory* I, 4th edition. New York: Springer-Verlag.

Millar, P.W. (1982): "A General Approach to the Optimality of Minimum Distance Estimators," *Transactions of the American Mathematical Society*, 286, 377-418.

Rice, J. (1984): "Bandwidth Choice for Nonparametric Regression," *Annals of Statistics*, 12, 1215-1230.

Spokoiny, V.G. (1996): "Adaptive Hypothesis Testing Using Wavelets," *Annals of Statistics*, 24, 2477-2498.

Triebel, H. (1992): *Theory of Function Spaces II*. Basel: Birkhäuser Verlag.

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

Whang, Y.-J. and D.W.K. Andrews (1993): "Tests of Specification for Parametric and Semiparametric Models," *Journal of Econometrics*, 57, 277-318.

Wooldridge, J.M. (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 Techniques," *Journal of Econometrics*, 75, 263-289.

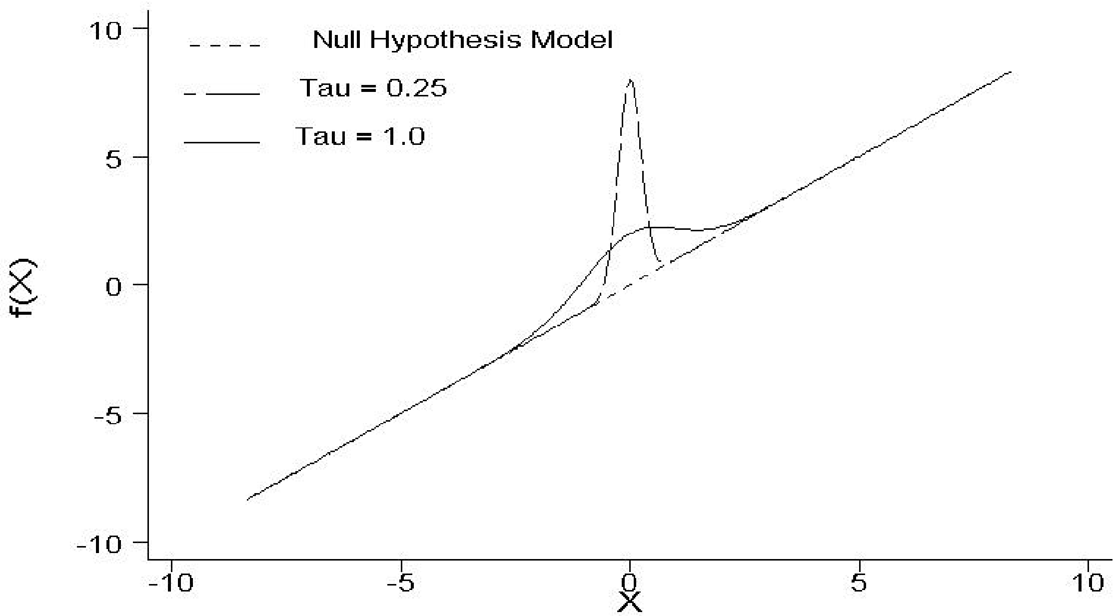

Figure 1: Null and Alternative Models