AN ADAPTIVE, RATE-OPTIMAL TEST OF LINEARITY FOR MEDIAN REGRESSION MODELS

by

Joel L. Horowitz
Department of Economics
Northwestern University
Evanston, IL 60208-2600
U.S.A.
joel-horowitz@northwestern.edu

and

Vladimir G. Spokoiny
Weierstrass Institute for Applied Analysis and Stochastics
Mohrenstrasse 39
10117 Berlin
GERMANY
spokoiny@wias-berlin.de

January 2002

ABSTRACT

This paper is concerned with testing the hypothesis that a conditional median function is linear against a nonparametric alternative with unknown smoothness. We develop a test that is uniformly consistent against alternatives whose distance from the linear model converges to zero at the fastest possible rate. The test does not require knowledge of the distribution of the model's random noise component, and it permits conditional heteroskedasticity of unknown form. The numerical performance and usefulness of the test are illustrated by the results of Monte Carlo experiments and an empirical example.

Key words: Hypothesis testing, local alternative, uniform consistency

---

We thank Russell Davidson and Jianqing Fan for helpful comments. The research of Joel L. Horowitz was supported in part by NSF Grant SES-9910925 and the Alexander von Humboldt Foundation.

AN ADAPTIVE, RATE-OPTIMAL TEST OF LINEARITY FOR MEDIAN REGRESSION MODELS

1. INTRODUCTION

This paper is concerned with testing a linear median-regression model against a nonparametric alternative. We develop a test that does not require knowledge of the smoothness of the alternative model, achieves the optimal rate of testing uniformly over smooth alternatives, and has other desirable power properties that are not shared by existing tests. An important feature of the test is that it does not require knowledge of the distribution of the model's random noise component, and it permits heteroskedasticity of unknown form.

We consider the model

(1.1) $Y_i = m(X_i) + \varepsilon_i; \quad i = 1,2,3,...,$

where $Y_i \in \mathbb{R}$ is a random variable; $\{X_i\} \in \mathbb{R}^d$ is a sequence of distinct, non-stochastic, design points; m is an unknown function; and $\{\varepsilon_i\}$ is a sequence of unobserved, independently but not necessarily identically distributed random variables whose medians are zero. The distributions of the ε_i's satisfy mild regularity conditions but are otherwise unknown. We test the null hypothesis, H_0, that there is a constant $\beta \in \mathbb{R}^d$ such that $m(X_i) = X_i'\beta$ for all *i*. X_i' denotes the transpose of the column vector X_i. The alternative hypothesis, H_1, is that there is no β such that $m(X_i) = X_i'\beta$ for all *i*. The test can be extended to models in which $\text{quantile}(\varepsilon_i) = 0$ for a quantile other than the median, but only the median is treated in this paper. We set the first component of each X_i equal to 1. Thus, X_i consists of $d-1$ "real" variables, and the first component of β is an intercept.

Linear quantile regression models are often used in applications. See Buchinsky (1994, 1998), Chamberlain (1994), Koenker and Geling (1999), Manning *et al*. (1995), and Poterba and Rueben (1994), among others. In contrast to mean regression models, quantile regression models do not require ε_i to have moments, are robust to outlying values of Y_i, and permit exploration of the entire conditional distribution of the dependent variable. However, there has been little research on testing the hypothesis of linearity. To our knowledge, only Zheng (1998) and Bierens and Ginther (2000) have developed tests of parametric quantile regression models against nonparametric alternatives. In contrast, there is a large literature on testing mean regression models against nonparametric alternatives. See, for example, Aït-Sahalia, *et al*. (1994), Andrews (1997), Bierens (1982, 1990), Bierens and Ploberger (1997), De Jong (1996), Eubank and Spiegelman (1990), Fan and Li (1996), Gozalo (1993), Härdle and Mammen (1993), Hart (1997), Hong and White (1995), Horowitz and Spokoiny (2001), Li and Wang (1998), Stute (1997), Whang and Andrews (1993), Wooldridge (1992), Yatchew (1992), and Zheng (1996).

The objective of this paper is to develop a test that has good theoretical and practical power properties. The power of a test is often investigated by deriving the asymptotic probability that the test rejects a false H_0 against a sequence of local alternative models. When H_0 is a linear median regression model, the form of the local alternative models is

(1.2) $$m_n(x) = x'\beta + \rho_n g(x)$$

for some $\beta \in \mathbb{R}^d$ and function g, where n is the sample size, ρ_n is a real number, and $\rho_n \to 0$ as $n \to \infty$. In (1.2), the distance between the null and alternative hypotheses converges to zero at the rate of ρ_n. See Section 4.3 for a more detailed explanation. The test of Bierens and Ginther (2000) has non-trivial power (that is, power exceeding the probability that a correct H_0 is rejected) against local alternatives for which $\rho_n \propto n^{-1/2}$. Zheng's (1998) test has non-trivial power against local alternatives for which $\rho_n \propto n^{-1/2+\nu}$ for any $\nu > 0$. However, as is explained in Horowitz and Spokoiny (2001) (hereinafter HS), the class of alternative models (1.2) is too small because it requires that $\partial^2 m_n(x)/\partial x \partial x' \to 0$ as $n \to \infty$. A less restrictive assumption is that m_n belongs to a class of smooth functions. It follows from Ingster (1993) that no test of H_0 can have non-trivial power uniformly over reasonable classes of smooth functions (e.g., Hölder classes) whose distance from the null hypothesis converges to zero at the rate $n^{-1/2}$ or $n^{-1/2+\nu}$ for any sufficiently small $\nu > 0$. For any given test, each n, and all sufficiently small $\nu > 0$, there exists a smooth alternative $m_n(x) = x'\beta + n^{-1/2+\nu} g(x)$ against which the test's power equals the probability of rejecting a correct H_0. The practical significance of this fact is that any test of H_0 for which $\rho_n \propto n^{-1/2+\nu}$ for sufficiently small $\nu > 0$ has low finite-sample power against certain classes of smooth alternatives. Section 5 presents examples.

Here, as in HS, we deal with this problem by letting the alternative hypothesis consist of a class of differentiable functions. Define the null-hypothesis set to be the set of models for which $m(X_i) = X_i'\beta$ for some $\beta \in \mathbb{R}^d$ and all $i = 1,...,n$. We assume that for each n, the alternative models belong to a Hölder class, $\mathcal{S}$, of differentiable functions on $\mathbb{R}^{d-1}$ and are separated from the null-hypothesis set by some distance. In an exact finite-sample treatment of power under this approach, the quality of a test for any given n is measured by the minimum distance, r_n, between the null hypothesis and the alternative functions from $\mathcal{S}$ that is necessary to achieve a specified level of power. Specifically, let $\rho(m)$ be the distance between a function $m \in \mathcal{S}$ and the null-hypothesis set. (See Section 4.4 for an example of ρ.) Let α_1 be the specified power. Then r_n is the smallest r such that

$$\inf_{m \in \mathcal{S},\, \rho(m) \geq r} \boldsymbol{P}(H_0 \text{ is rejected against } m) \geq \alpha_1 .$$

An optimal test minimizes r_n for the given α_1 and n while maintaining a specified probability of rejecting H_0 when it is true .

Unfortunately, an exact finite-sample treatment is difficult to implement, and tests that are optimal in finite samples have been found only for a few examples in which the ε_i's are identically distributed with a known distribution. Therefore, to make progress under assumptions that are less restrictive and more relevant to applications, we use an asymptotic approach. In this approach, we let $r_n \to 0$ as $n \to \infty$ at the fastest possible rate that permits the specified power to be achieved uniformly over sets $B_n = \{m \in \mathcal{S} : \rho(m) \geq C r_n\}$ for some finite constant C. This rate is called the optimal rate of testing. At the optimal rate of testing, for any $\alpha_1 < 1$ there exists a finite $C > 0$ such that

$$\lim_{n \to \infty} \inf_{m \in B_n} \boldsymbol{P}(H_0 \text{ is rejected against } m) \geq \alpha_1 . \tag{1.3}$$

Moreover, 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 known order $s \geq (d - 1)/4$ is $n^{-2s/(4s + d - 1)}$ (Ingster 1982, 1993a, 1993b, 1993c; Guerre and Lavergne 1999). The optimal rate of testing is $\left(n^{-1}\sqrt{\log\log n}\right)^{2s/(4s+d-1)}$ if $s \geq (d-1)/4$ is unknown (Spokoiny 1996). If $s < (d-1)/4$, then the optimal rate of testing is $n^{-1/4}$ (Guerre and Lavergne 1999).

This paper describes a test of H_0 that has the optimal rate of testing uniformly over Hölder classes and does not require knowledge of s or the (possibly non-identical) distributions of the ε_i's in (1.1). Indeed, we prove that our test satisfies (1.3) with $\alpha_1 = 1$. This property is called *uniform consistency* at the optimal rate of testing. 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. HS developed an adaptive, rate-optimal test of a parametric mean regression model against a nonparametric alternative. Fan and Huang (2000) developed an adaptive, rate-optimal test of a normal, linear mean-regression model. See, also, Ledwina (1994) and Fan (1996). This paper extends the test of HS to median regression models. Although there are similarities between the test presented here and that of HS, the properties of median and mean regression models are sufficiently different to make the extension non-trivial and to require separate treatments of median and mean regressions.

A test that achieves the optimal rate of testing has the advantage of being sensitive to alternatives uniformly over a smoothness class whose distance from the null hypothesis converges to zero at the fastest possible rate. Such a class contains sequences of alternative models against which the tests of Bierens and Ginther (2000) and Zheng (1998) 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. In addition, the optimality properties of the test that we present hold uniformly over designs $\{X_i\}$ and distributions of the ε_i's that satisfy mild regularity conditions. Similarly, the differences between the exact finite-sample and asymptotic probabilities that our test rejects H_0 converge to zero uniformly over designs and distributions of the ε_i. Thus, in large samples, the power of the test and the accuracy of the asymptotic approximations are relatively insensitive to the design and the possibly heterogeneous distributions of the ε_i's.

Since our theoretical results are asymptotic, the desirable power properties of our test do not necessarily hold when n is small. In particular, when n is small, there may be smooth alternatives that cannot be detected by our test or any other test. This happens, for example, if the null and alternative models differ only between design points. However, we require the distance between design points to decrease to zero as n increases (see Section 4.1). This enables our test to detect any smooth alternative model when n is sufficiently large. The uniform consistency property of the test insures that the same "sufficiently large" n applies to all smooth alternatives.

A test that achieves the optimal rate of testing uniformly over a smoothness class is necessarily oriented toward the alternatives within the class that are hardest to detect. Such a test may have low power against functions that are less extreme. It turns out that our test automatically protects against this possibility. Specifically, we show that our test is consistent against alternatives of the form (1.2) whenever $\rho_n \ge Cn^{-1/2}\sqrt{\log\log n}$ for some finite $C > 0$. Consistency of the tests of Bierens and Ginther (2000) and Zheng (1998) against alternatives of the form (1.2) requires $\rho_n \to 0$ more slowly than $n^{-1/2}$. Thus, in terms of consistency against such alternatives, there is essentially no penalty paid for the adaptiveness and rate optimality of our test.

Section 2 presents an empirical example that helps to motivate our test. The test is described in Section 3. Theorems giving properties of the test under H_0 and various alternative hypotheses are presented in Section 4. Section 5 presents the results of a Monte Carlo investigation of the test's finite-sample behavior. Section 6 continues the empirical example of Section 2. Section 7 presents concluding comments. The proofs of theorems are in the Appendix, which is Section 8.

2. AN EMPIRICAL EXAMPLE

Buchinsky (1998) used data from the 1993 Current Population Survey (CPS) to estimate a median regression model of the relation between the weekly wages of male workers in the U.S. and a variety of covariates. The model is

$$\log W = \beta_0 + \beta_1 X + \beta_2 X^2 + \gamma' Z + U ,$$

where W is the weekly wage, X is years of labor-force experience, and Z is a vector of covariates that includes years of education and dummy variables indicating the worker's race, the region of the country in which the worker is employed, whether the worker is employed in a metropolitan area, and whether employment is full time and for the full year. U is an unobserved random variable whose median conditional on X and Z is 0, the β's are scalar coefficients, and γ is a vector of coefficients. In this example, we investigate the relation between $\log W$ and X for white, full-time, full-year, workers with 12 years of education who were employed in a metropolitan area in the north central region of the U.S. Thus, Z is fixed in the example, and the model is

$$\log W = \beta_0 + \beta_1 X + \beta_2 X^2 + U , \tag{2.1}$$

where $\text{median}(U \mid X = x) = 0$ almost surely. The 1993 CPS contains 1656 observations of workers with the specified characteristics. The β's were estimated by LAD.

The dashed and solid lines in Figure 1 show the parametrically and nonparametrically estimated conditional median functions. The parametric estimate (dashed line) is $b_0 + b_1 X + b_2 X^2$, where b_j is the LAD estimate of β_j (j = 0, 1, 2). The nonparametric estimate (solid line) was obtained by local linear median regression (Chaudhuri 1991). There are obvious differences between the parametric and nonparametric estimates, which suggests that the parametric model is misspecified. However, the graph does not indicate whether this apparent misspecification is an artifact of random sampling error. A specification test is needed to make this distinction. As will be discussed in Section 6, our test rejects model (2.1) at the 0.05 level.

We also estimated a version of (2.1) that is augmented by adding X^3 to the specification, thereby producing the cubic model

(2.2) $\log W = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + U$.

The dotted line in Figure 1 shows the conditional median function estimated by applying LAD to (2.2). The fit of (2.2) is much better than that of (2.1). Our test does not reject the cubic model.

3. THE TEST

Section 3.1 presents an informal description of the test statistic. Section 3.2 describes a method for obtaining critical values for the test.

3.1 The Test Statistic

We assume that $d \geq 2$ and that the first component of X_i is $X_{i1} = 1$. If H_0 is true, then $Y_i = X_i'\beta + \varepsilon_i$ and $\boldsymbol{P}(\varepsilon_i \leq 0) = 0.5$ for each $i = 1, 2, \ldots$ and some $\beta \in \mathbb{R}^d$. Let b_n denote the least absolute deviations (LAD) estimator of β. Thus,

$$b_n = \arg\min_{b \in \mathbb{R}^d} \sum_{i=1}^{n} | Y_i - X_i'b | .$$

If H_0 is true, then $b_n \rightarrow^p \beta$ as $n \rightarrow \infty$ (Koenker and Bassett 1978). If H_0 is false, then β is undefined. However, it follows from Proposition 1 in the Appendix that $b_n = \beta^* + O_p(n^{-1/2})$, where β^* solves

$$\sum_{i=1}^{n} X_i \{\boldsymbol{P}[\varepsilon_i \leq X_i'b - m(X_i)] - 1/2\} = 0 . \tag{3.1}$$

Define $\beta_0 = \beta$ if H_0 is true, $\beta_0 = \beta^*$ if H_0 is false, and $\xi_i = \boldsymbol{1}(Y_i - X_i\beta_0 \leq 0) - 1/2$, where $\boldsymbol{1}$ is the indicator function. If H_0 is false, then β_0 depends on n and the design $\{X_i\}$. Moreover, $X_i'\beta_0$ can be interpreted as the best linear approximation to $m(X_i)$. If H_0 is true, then β_0 is independent of both n and the design.

Under H_0, the ξ_i's are Bernoulli random variables with $\boldsymbol{E}(\xi_i) = 0$. If H_0 is false, then $\boldsymbol{E}(\xi_i) = \boldsymbol{P}[\varepsilon_i \leq X_i'\beta_0 - m(X_i)] - 1/2 \neq 0$ for at least one *i*. Thus, a test of H_0 is equivalent to a test of H_0': $\boldsymbol{E}(\xi_i) = 0$ for all *i*. If β_0 were known, such a test could be based on the distance from 0 of a nonparametric estimator of the vector $[\boldsymbol{E}(\xi_1), \ldots, \boldsymbol{E}(\xi_n)]'$. We obtain a feasible test by replacing β_0 with b_n. Define $\hat{\xi}_i \equiv \boldsymbol{1}(Y_i - X_i b_n \leq 0) - 1/2$. Our test is based on $\{\hat{\xi}_i : i = 1, \ldots, n\}$.

To obtain the test statistic, suppose for the moment that β_0 and, therefore, the ξ_i's were known. Let K denote a kernel function (in the sense of nonparametric density estimation) of a $d-1$ dimensional argument. For $v \in \mathbb{R}^{d-1}$ and bandwidth $h > 0$, let $K_h(v) = K(v/h)$. For $i, j = 1,...,n$, define

$$w_{ij,h} = \frac{K_h(X_i - X_j)}{\sum_{k=1}^{n} K_h(X_i - X_k)}$$

and $a_{ij,h} = \sum_{k=1}^{n} w_{ki,h} w_{kj,h}$. Define

(3.2) $$S_h{*} = \sum_{i=1}^{n} \left| \sum_{j=1}^{n} w_{ij,h} \xi_j \right|^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij,h} \xi_i \xi_j .$$

Observe that $\sum_{j=1}^{n} w_{ij,h} \xi_j$ is a kernel nonparametric estimator of $\boldsymbol{E}(\xi_i)$. Therefore, $S_h{*}$ is the ℓ_2 distance from zero of the kernel estimator of $[\boldsymbol{E}(\xi_1),...,\boldsymbol{E}(\xi_n)]'$. If the ξ_i's were observable, then a test of H_0 could be based on the standardized version of $S_h{*}$. Because $\boldsymbol{E}(\xi_i) = 0$ under H_0, $\xi_i^2 = 1/4$, and ξ_i is independent of ξ_j if $i \neq j$, the standardized $S_h{*}$ is

(3.3) $$T_h{*} = \frac{S_h{*} - N_h}{V_h},$$

where

(3.4) $$N_h = (1/4) \sum_{i=1}^{n} a_{ii,h},$$

and

(3.5) $$V_h = \left[(1/8) \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} a_{ij,h}^2 \right]^{1/2} .$$

HS showed that an adaptive, rate-optimal test of H_0 can be obtained by rejecting H_0 if the maximum of $T_h{*}$ over a suitable set of bandwidths h is too large. The test proposed here uses the same idea and is obtained by replacing the unknown variable ξ_i with $\hat{\xi}_i$ in (3.2)-(3.5).

To this end, define $T_h = (S_h - N_h)/V_h$, where

(3.6) $$S_h = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij,h} \hat{\xi}_i \hat{\xi}_j .$$

We evaluate T_h at each h in a set of bandwidths and reject H_0 if T_h is too large for any bandwidth in this set. The set of bandwidths is $\mathcal{H} = \{h = h_{\min} 2^{k/[2(d-1)]} : h \le h_{\max}, k = 0,1,2,...\}$, where $h_{\max}$ and $h_{\min}$ are non-stochastic constants satisfying conditions that are stated in Section 4.1. Our test is based on the statistic

$$T = \max_{h \in \mathcal{H}} T_h .$$

The test rejects H_0 at the (asymptotic) α level if T exceeds the critical value that is described in Section 3.2.

3.2 Obtaining the Critical Value

The exact α-level critical value for T is the 1 - α quantile of the finite-sample distribution of T. This critical value cannot be evaluated in applications because the finite-sample distribution of the $\hat{\xi}_i$'s is unknown. However, it turns out that the distribution of T under H_0 depends only weakly on β and the distribution of the ε_i's in (1.1), and this dependence vanishes as the sample size increases. This is true even if the ε_i's are not identically distributed. We explain this result intuitively below and prove it in the Appendix; see Lemma 12 and the proof of Theorem 1. Therefore, a consistent estimator of the α-level critical value can be obtained as the $1-\alpha$ quantile of the distribution of T that is induced by the model $Y_i{*} = X_i' b_n + \varepsilon_i{*}$, where $\varepsilon_i{*}$ is sampled from a convenient distribution. We recommend sampling $\varepsilon_i{*}$ from the empirical distribution of the residuals (EDR) of the estimated null-hypothesis model. This sampling procedure is easy to implement, a natural choice if the ε_i's are identically distributed, and (as has just been explained) valid in large samples even if the ε_i's are non-identically distributed. The i'th residual is $Y_i - X_i b_n$. Section 6 presents Monte Carlo evidence on the finite-sample performance of the test with this distribution of $\varepsilon_i{*}$.

The EDR does not satisfy Assumption 1 of Section 4, which requires $\varepsilon_i{*}$ to have an absolutely continuous distribution with a continuous density. However, under our assumptions, the EDR approximates an absolutely continuous distribution with an error that converges to zero at the rate $n^{-1/2}$. Consequently, the difference between the results obtained with the EDR and with an absolutely continuous distribution vanishes as $n \to \infty$.

We now explain intuitively why the distribution of T under H_0 depends only weakly on β and the distribution of the ε_i's. Under H_0, the random variables $\xi_i = \boldsymbol{1}(Y_i - X_i \beta \le 0) - 1/2$

are independently and identically Bernoulli distributed with $\boldsymbol{P}(\xi_i = \pm 1/2) = 1/2$. This is true regardless of the distributions of the ε_i's or whether they are identically distributed. Therefore, under H_0 the distribution of $T_h{*}$ in (3.3) does not depend on β or the distribution of the ε_i's and can be approximated with any precision by Monte Carlo simulation. The statistic T_h is obtained from $T_h{*}$ by replacing the ξ_i's with the empirical analogs $\hat{\xi}_i = \boldsymbol{1}(Y_i - X_i b_n \le 0) - 1/2$. It is proved in Proposition 1 of the appendix that $b_n - \beta = O_p(n^{-1/2})$ under H_0. Therefore, the difference between T_h and $T_h{*}$ is small uniformly over $h \in \mathcal{H}$ if n is large. See Lemma 12 of the appendix. Consequently, the distribution of T under H_0 depends on β the distributions of the ε_i's only weakly through b_n, and this dependence vanishes as $n \to \infty$.

The recommended estimator of the critical value can be computed with any desired accuracy by using the following simulation procedure:

1. For each i = 1, …, n, generate $Y_i{*} = X_i' b_n + \varepsilon_i{*}$, where $\varepsilon_i{*}$ is sampled randomly from the residuals, $Y_i - X_i' b_n$, of the estimated null-hypothesis model.

2. Use the data set $\{Y_i{*}, X_i$: i = 1, …, $n\}$ to estimate β. Denote the resulting estimate by $\hat{b}_n$. Compute the statistic $\hat{T}$ that is obtained by replacing $\hat{\xi}_i$ $(i = 1, \ldots, n)$ with $\boldsymbol{1}(Y_i{*} - X_i \hat{b}_n \le 0) - 1/2$ in the formula for T.

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

4. THE MAIN RESULTS

This section presents theorems that give the behavior of the proposed test under the null hypothesis, local alternatives, and smooth alternatives that are contained in a Hölder class. Section 4.1 states our assumptions. The assumptions apply to any sample size, n. The results stated in Sections 4.2-4.4 hold uniformly over models, designs, and distributions of the ε_i's that satisfy the assumptions. Section 4.2 gives the behavior of the test under H_0. Sections 4.3 and 4.4, respectively, give the test's behavior under the sequence of local alternative hypotheses (1.2) and under smooth alternatives that are contained in a Hölder class whose distance from the null hypothesis converges to zero at the optimal rate of testing. The adaptive, rate-optimal property of the test is established in Section 4.4.

4.1 Assumptions

Our results are obtained under the assumptions stated in this section. Let $\|V\|$ denote the Euclidean norm of the vector V. If D is a $q \times q$ matrix, define

$$\|D\|_\infty = \sup_{v \in \mathbb{R}^q,\, v \neq 0} \frac{\|Dv\|}{\|v\|}.$$

For every $x \in \mathbb{R}^d$ and every $h > 0$, define $M_h(x)$ as the number of elements in the set $\{X_i : \|X_i - x\| \leq h\}$. Define $F_i(u) \equiv \boldsymbol{P}(\varepsilon_i \leq u)$.

<u>Assumption 1</u> (Observations): *The observations* $\{Y_i : i = 1, 2, ...\}$ *in* (1.1) *are independent. Each cumulative distribution function* F_i *is absolutely continuous with respect to Lebesgue measure with a continuously differentiable density function* f_i. *There are finite constants* $C_f > 0$, C_F *and* a *such that* $f_i(0) \geq C_f$, $f_i(u) \leq aC_F$ *and* $|f_i'(u)| \leq a^2 C_F$ *for all* $i, j = 1, ..., n$ *and u. Moreover,* $\|f_i(0) - f_j(0)\| \leq C_3 \|X_i - X_j\|$ *for some constant* $C_3 < \infty$ *and all* $i = 1, ..., n$.

<u>Assumption 2</u> (Kernel): K *is continuously differentiable, non-negative, symmetrical about the origin, and supported on* $[-1,1]^{d-1}$. *Moreover,* $K(0) = 1$ *and* $K(v)$ *is a strictly decreasing function of* $\|v\|$.

<u>Assumption 3</u> (Bandwidths): *The quantities* $h_{\min}$ *and* $h_{\max}$ *satisfy* $h_{\min} < h_{\max}$, $h_{\min} \geq C_h n^{-1/2+\gamma}$, *and* $h_{\max} = C_H (\log\log n)^{-1}$ *for finite constants* $\gamma > 0$, $C_h > 0$, *and* $C_H > 0$.

<u>Assumption 4</u> (Design): (i) *The design points* $\{X_i : i = 1, ..., n\}$ *are non-stochastic. The first component of each* X_i *is* $X_{i1} = 1$. (ii) *There are positive constants* C_{X1} *and* C_{X2} *such that for all* $h \in \mathcal{H}$ *and all* i = 1, …, *n*, $C_{X1} n h^{d-1} \leq M_h(X_i) \leq C_{X2} n h^{d-1}$. (iii) *There are finite constants* C_X *and* C_{XX} *such that* $\|X_i\| \leq C_X$ *for all i and* $\left\| \left[n^{-1} \sum_{i=1}^n f_i(0) X_i X_i' \right]^{-1} \right\|_\infty \leq C_{XX}$.

(iv) $\inf_{b: \|b - \beta_0\| > \delta} n^{-1} \sum_{i=1}^n | F_i[X_i' b - m(X_i)] - F_i[X_i' \beta_0 - m(X_i)] | > C\delta$ *for some constant* C *and each* $\delta > 0$.

Section 5.2 describes a method for choosing $h_{\min}$ and $h_{\max}$ in applications. Assumption 4(ii) is satisfied with probability approaching 1 as $n \to \infty$ if Assumption 3 holds and components $2, ..., d$ of $\{X_i\}$ are sampled from a distribution that has bounded support and a density with respect to Lebesgue measure that is bounded away from zero on its support. Therefore, our results hold conditionally on $\{X_i\}$ that are generated this way. However, we do not require $\{X_i\}$ to be sampled from a distribution. Assumption 4(iv) is an identification condition. It excludes perfectly collinear designs and the possibility that $F_i(u)$ is constant in a neighborhood of $u = X_i\beta_0 - m(X_i)$ for every $i = 1, 2, \ldots$. It is easy to verify that a local version of 4(iv) (for b satisfying $\delta \le \|b - \beta_0\| \le \delta_1$ with sufficiently small δ and $\delta_1 > \delta$) follows from Assumptions 1 and 4(iii).

4.2 Behavior of the Test Statistic under the Null Hypothesis

The null hypothesis, H_0, is that $\boldsymbol{P}(Y_i - X_i'\beta \le 0) = 1/2$ for all i and some $\beta \in \mathbb{R}^d$. Let t_α be the α-level critical value that is that is induced by the model $Y_i^* = X_i'b_n + \varepsilon_i^*$ described in Section 3.2. The main result on the behavior of T under H_0 is that t_α is an asymptotically correct α-level critical value. Moreover, t_α is an approximately correct finite-sample critical value when n is large. The approximation error converges to zero as $n \to \infty$ and depends only on the constants in Assumptions 1-4. It does not depend on β, the design, or the distributions of the ε_i's, provided that these satisfy Assumptions 1-4. This result is established by the following theorem.

Theorem 1: *Let* H_0 *be true. Then there is a sequence of positive constants* $\{\omega_{n1}\}$ *such that* $\omega_{n1} \to 0$ *monotonically as* $n \to \infty$ *and*

$$\left|\boldsymbol{P}(T > t_\alpha) - \alpha\right| \le \omega_{n1}$$

for all linear models (1.1) *that satisfy Assumptions* 1-4. *The sequence* $\{\omega_{n1}\}$ *depends only on the constants in Assumptions 1-4.*

4.3 Power against a Sequence of Local Alternatives

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

$$(4.1) \qquad \|g\|^2 \equiv \frac{1}{n}\sum_{i=1}^{n} |\, g(X_i)\,|^2 \ge 1\,.$$

This normalization depends on the design and insures that there are "enough" design points within the support of g. Let $\mathcal{X}$ be the $d \times n$ matrix whose i'th column is X_i, $\mathcal{F}$ be the $n \times n$ diagonal matrix whose (i,i) element is $f_i(0)$, and $\mathcal{G}$ be the $n \times 1$ vector whose i'th component is

$g(X_i)$. Let $\boldsymbol{I}_n$ be the $n \times n$ identity matrix. Define the $n \times n$ matrix $\Pi = \boldsymbol{I}_n - \mathcal{X}'(\mathcal{X}\mathcal{F}\mathcal{X}')^{-1}\mathcal{X}\mathcal{F}$. If the ε_i's are *iid*, then Π is the projection operator into the orthogonal complement of the space spanned by the X_i's. Assume that for all sufficiently large *n* and some $\delta > 0$,

(4.2) $\quad v^2 \equiv n^{-1}\|\Pi\mathcal{G}\|^2 \geq \delta$.

If the ε_i's are *iid*, then (4.2) states $\mathcal{G}$ has a non-zero projection into the orthogonal complement of the space spanned by the X_i's. Conditions (4.1) and (4.2) insure that the distance between the null hypothesis and the sequence of alternative models,

$$\inf_{b \in \mathbb{R}^d}\left(n^{-1}\sum_{i=1}^{n}\|X_i'b - X_i'\beta - \rho_n g(X_i)\|^2\right)^{1/2},$$

converges to 0 at the rate of ρ_n rather than a faster rate. The following theorem establishes consistency of our test under a sequence of local alternatives.

Theorem 2: *Let Assumptions* 1-4 *hold. Let* (1.2) *hold with* $\rho_n \geq Cn^{-1/2}\sqrt{\log\log n}$ *and g satisfying* (4.1)-(4.2). *There exists a constant* $C^* < \infty$, *depending on* g, δ, *and the constants in Assumptions* 1-4, *and a sequence of constants* $\{\omega_{n2}\}$ *such that* $\omega_{n2} \to 0$ *monotonically as* $n \to \infty$ *and*

$$\boldsymbol{P}(T > t_\alpha) \geq 1 - \omega_{n2}$$

whenever $C \geq C^*$. *The sequence* $\{\omega_{n2}\}$ *depends only on the constants in Assumptions 1-4.*

This result holds uniformly over β, the design $\{X_i\}$, and the possibly heterogeneous distributions of the ε_i's, but it is not uniform over "directions" g of departure from H_0. Theorem 2 is useful because, as was explained in Section 1, it insures that when n is large, our test has high power against alternatives that are less extreme than the ones that determine the optimal uniform rate of testing.

4.4 Power against a Smooth Alternative

This section gives conditions under which our test is consistent uniformly over alternatives in a Hölder smoothness class whose distance from the class of linear conditional median functions converges to zero at the fastest possible rate. The distance between the null-hypothesis set and the conditional median function $m(x)$ can be measured by

$$\rho_1(m) = \inf_{b \in \mathbb{R}^d}\left[n^{-1}\sum_{i=1}^{n}|m(X_i) - X_i'b|^2\right]^{1/2}.$$

However, it is more convenient to work with another distance function that we define by using the functions $H_i(x) = F_i[x'\beta_0 - m(x)]$, where β_0 is as defined in (3.1). Under H_0, $H_i(x) = 1/2$ for all x. Therefore, we can measure the distance between the null-hypothesis set and $m(x)$ by

$$\rho_2(H) = \left[n^{-1} \sum_{i=1}^{n} | H_i(X_i) - 1/2 |^2 \right]^{1/2},$$

where H refers to the vector of functions $(H_1, ..., H_n)$. In what follows, we will be interested in sets of functions $(H_1, ..., H_n)$ such that $\rho_2(H) \geq C_a \left(n^{-1} \sqrt{\log \log n} \right)^{2s/(4s+d-1)}$ for some constant $C_a < \infty$. One can easily check that when this condition and Assumptions 1-4 hold, then $\rho_1(m) \geq C_b \left(n^{-1} \sqrt{\log \log n} \right)^{2s/(4s+d-1)}$ for some $C_b < \infty$ and that ρ_1 and ρ_2 give the same rate of convergence of the distance between the null and alternative hypotheses.

Let $j = (j_2, \ldots, j_d)$, where $j_2, \ldots, j_d \geq 0$ are integers, be a multi-index. Define $|j| = \sum_{k=2}^{d} j_k$ and

$$D^j H_i(x) = \frac{\partial^{|j|} H_i(x)}{\partial x_2^{j_1} \ldots \partial x_d^{j_d}}$$

whenever the derivative exists. Define the Hölder norm

$$\|H_i\|_{H,s} = \sup_{x: |x_i| \leq C_X} \sum_{|j| \leq s} | D^j H_i(x) |.$$

The smoothness classes that we consider consist of functions $(H_1, ..., H_n) \in \mathcal{S}_n(s) \equiv \{H_1, ..., H_n : \|H_i\|_{H,s} \leq C_F \text{ for all } i = 1, ..., n\}$ for some $s \geq \max[2, (d - 1)/4]$ and $C_F < \infty$.

Theorem 3 states that our test is consistent uniformly over the sets

$$\mathcal{B}_n \equiv \left\{ H_1, ..., H_n \in \mathcal{S}_n(s) : \; \rho_2(H) \geq C_a \left(n^{-1} \sqrt{\log \log n} \right)^{2s/(4s+d-1)} \right\} \tag{4.3}$$

for some $s \geq \max[2, (d - 1)/4]$ and all sufficiently large $C_a < \infty$. If $(H_1, ..., H_n) \in \mathcal{B}_n$, then m belongs to a Hölder class of order s and $\rho_1(m) \geq C_b \left(n^{-1} \sqrt{\log \log n} \right)^{2s/(4s+d-1)}$ for some $C_b < \infty$.

<u>Theorem 3</u>: *Let Assumptions* 1-4 *hold. Then for* $0 < \alpha < 1$ *and* $\mathcal{B}_n$ *as defined in* (4.3), *there is a constant* C_a^* *and a sequence of constants* $\{\omega_{n3}\}$ *such that* $\omega_{n3} \to 0$ *monotonically as* $n \to \infty$ *and*

$$\inf_{m: (H_1, ..., H_n) \in \mathcal{B}_n} \boldsymbol{P}(T > t_\alpha) \geq 1 - \omega_{n3}$$

whenever $C_a \geq C_a{}^*$. C_a *and* $\{\omega_{n3}\}$ *depend only on* s *and the constants in Assumptions* 1-4.

This result shows that our test is consistent at the fastest possible rate of testing uniformly over alternatives in the smoothness classes $\mathcal{B}_n$ and over designs and distributions of the ε_i's that satisfy Assumptions 1-4. These uniformity properties insure that the test's power is not highly sensitive to the alternative model, design, or distributions of the ε_i's when n is large.

5. 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 5.1 presents an example in which our test is consistent but the tests of Bierens and Ginther (2000) and Zheng (1998) are not. This example motivates the design of the Monte Carlo experiments. The experiments and their results are described in Section 5.2.

5.1 An Example

This section presents a parametric model and a sequence of alternatives. Our test is consistent against the alternatives but the tests of Bierens and Ginther (2000) and Zheng (1998) are not. The null hypothesis model in the example is

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \,, \tag{5.1}$$

where β_0 and β_1 are constants, the X_i's are scalars that are sampled from a distribution that is symmetrical about 0, and $\varepsilon_i \sim N(0,\sigma^2)$ for every i. The sequence of alternative models is

$$Y_i = X_i + \tau_n{}^4 \phi(X_i / \tau_n) + \varepsilon_i \,, \tag{5.2}$$

where $\varepsilon_i \sim N(0,1)$, ϕ is the standard normal density function, and $\tau_n = C\left(n^{-1}\sqrt{\log\log n}\right)^{-1/9}$ for some finite $C > 0$. The function $m_n(x) = x + \tau_n{}^4\phi(x/\tau_n)$ has a peak that is centered at $x = 0$ and that becomes narrower as n increases. The sequence of alternative models $\{m_n\}$ is contained in $\mathcal{B}_n$ with $s = 2$. The distance between m_n and the parametric model (5.1) satisfies $\rho_1(m_n) \propto \left(n^{-1}\sqrt{\log\log n}\right)^{-4/9}$. It is not difficult to show that under that the sequence (5.2), the noncentral parameters of the tests of Bierens and Ginther (2000) and Zheng (1998) converge to zero as $n \to \infty$, so those tests are inconsistent against (5.2). It follows from Theorem 3, however, that the adaptive, rate optimal test is consistent against this sequence if C is sufficiently large.

5.2 Monte Carlo Experiments

This section presents the results of Monte Carlo experiments that illustrate the numerical performance of the adaptive, rate-optimal test. The designs of the experiments are motivated by the example of Section 5.1 and are taken from HS with the modification that they specify conditional medians instead of means. 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 power of Zheng's (1998) test is also estimated by Monte Carlo simulation. In all experiments, the nominal probability of rejecting a correct null hypothesis is 0.05.

The null-hypothesis model in the experiments is

(5.3) $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i; \quad i = 1, 2, ..., n$

where n = 100 or 250 and 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 (5.3) is correct (H_0 is true), $\beta_0 = \beta_1 = 1$. The ε_i's were sampled independently from three distributions, depending on the experiment. These are N(0,4), a variance mixture of normals in which ε_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 shifted and scaled to have median zero and variance of 4. The mixture distribution is leptokurtic with a variance of 3.9, and the Type I extreme value distribution is asymmetrical. Variation in X explains 77-79 percent of the variation in Y in (5.3), depending on n and the distribution of ε. Specifically, $0.77 \le 1 - Var(\varepsilon)/Var(Y) \le 0.79$.

The alternative models have the form

$$Y_i = 1 + X_i + (4/\tau)\phi(X_i/\tau) + \varepsilon_i,$$

where the ε_i's are sampled from one of the three distributions just described and τ = 1 or 0.25, depending on the experiment. Figure 2 plots the function $m(x) = 1 + x + (4/\tau)\phi(x/\tau)$ for each value of τ. 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 β_0 and β_1 were estimated by least absolute deviations (LAD). The kernel used for the adaptive, rate-optimal test and Zheng's (1998) test is $K(u) = (15/16)(1-u^2)^2 \boldsymbol{1}(|u| \le 1)$.

At the suggestion of a referee, we carried out F tests of the hypothesis $\tilde{H}_0: \beta_2 = 0$ in the augmented model $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i$. Rejection of $\tilde{H}_0$ implies that (5.3) is misspecified, but a test of $\beta_2 = 0$ is not consistent against all fixed, smooth alternatives to (5.3).

Implementation of Zheng's (1998) test requires selecting a bandwidth parameter. Zheng (1998) proposed a generalized cross validation procedure for doing this, but in our experiments it gave bandwidths that were much too large and often exceeded the range of the values of *X*. Therefore, we selected the bandwidth through Monte Carlo experimentation to maximize the test's power subject to the restriction that the empirical probability of rejecting (5.3) when it is correct be contained in a 95% confidence interval around the nominal rejection probability. Zheng's test uses a critical value that is based on the asymptotic normal distribution of his test statistic, and we used this critical value to compute his test's empirical rejection probabilities.

The adaptive, rate-optimal test requires choosing the set of bandwidths $\mathcal{H}$. We used a geometric grid consisting of the points $\omega^j h_{\min}$ $(j = 0,1,2,...,\mathcal{N} - 1)$, where $\mathcal{N}$ is the number of grid points and $\omega = (h_{\max} / h_{\min})^{1/(\mathcal{N}-1)}$. The smallest and largest bandwidths are $h_{\min} =$ $2\max(X_{i+1} - X_i)$ $(i = 1,...,n-1)$ and $h_{\max} =$ $0.4(X_n - X_1)/\log\log n$, where the X_i's are sorted in increasing order. We chose $\mathcal{N}$ according to the rule of thumb $\mathcal{N} \approx \log n$. This rule is consistent with the theory of the test, which requires $\mathcal{N} = O(\log n)$. Motivated by the rule of thumb, we did experiments with values of $\mathcal{N}$ in the range 4-10. The results varied little over this range, so we report only the results for $\mathcal{N} = 4$.

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 500 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. 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 differences between the nominal and empirical rejection probabilities is significantly different from zero at the 0.05 level. When H_0 is false, the power of the adaptive, rate-optimal test is much higher than the powers of Zheng's test and the *F*-test of $\tilde{H}_0$. All of the differences between the powers of the adaptive, rate-optimal test and either Zheng's or the *F*-test are significant at the 0.01 level.

6. CONTINUATION OF THE EMPIRICAL EXAMPLE

We now present the application of our test to models (2.1) and (2.2) using the data described in Section 2. The test was carried out using the geometric grid of bandwidths described in Section 5.2 with $\mathcal{N} = 7$, the integer closest to $\log n$. The adaptive, rate-optimal test of the quadratic model (2.1) gives $T = 2.10$ with a 0.05-level critical value of 0.88. Thus, model (2.1) is rejected at the 0.05 level. An *F*-test of the hypothesis that $\beta_3 = 0$ in (2.2) also rejects the quadratic model (p < 0.01). The adaptive, rate-optimal test of the cubic model (2.2) gives T = -0.46 with a 0.05-level critical value of 0.75. Thus, the model (2.2) is not rejected.

7. CONCLUSIONS

This paper has developed a test of the hypothesis that a conditional median function is linear against a nonparametric alternative. The test adapts to the unknown smoothness of the alternative model, does not require knowledge of the distributions of the possibly heterogeneous noise components of the model (the ε_i;s in (1.1)), and is uniformly consistent against alternative models whose distance from the class of linear functions converges to zero at the fastest possible rate. This rate is slower than $n^{-1/2}$. In addition, the new test is consistent (though not uniformly) against local alternative models whose distance from the class of linear models decreases at a rate that is only slightly slower than $n^{-1/2}$. The results of Monte Carlo simulations and an empirical application have illustrated the usefulness of the new test.

8. MATHEMATICAL APPENDIX

This appendix presents the proofs of the theorems in the text. Except as otherwise noted, it is assumed that Assumptions 1-4 hold. Throughout the proofs, $\{a_{n1}\}, \{a_{n2}\}, \ldots$, and $\{s_{n1}\}$, $\{s_{n2}\}, \ldots$ denote sequences of non-negative numbers that depend only on the constants in Assumptions 1-4 and converge monotonically to zero as $n \to \infty$. The latter property will be denoted $a_{nk} = o_n(1)$ and $s_{nk} = o_n(1)$ for $k = 1, 2...$.

8.1 Properties of the Parametric Model

The main result of this section is a proof of $n^{1/2}$ asymptotic normality of the LAD estimator b_n. Let $\tilde{F}_i$ and $\tilde{f}_i$, respectively, denote the probability distribution and density functions of Y_i. Define

$$Q_n = \frac{1}{n}\sum_{i=1}^{n} X_i X_i' \tilde{f}_i(X_i'\beta_0),$$

$$\eta_n = -Q_n^{-1} n^{-1/2} \sum_{i=1}^{n} X_i[\boldsymbol{1}(Y_i - X_i'\beta_0 \le 0) - \tilde{F}_i(X_i'\beta_0)],$$

and

$$\Sigma_n = Q_n^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n} X_i X_i' \tilde{F}_i(X_i'\beta_0)[1-\tilde{F}_i(X_i'\beta_0)]\right\} Q_n^{-1}.$$

Proposition 1: *Let Assumptions* 1-4 *hold. Let the sequence* $\{\delta_n\}$ *satisfy* $n^{-1/2}/\delta_n = o(1)$ *as* $n \to \infty$ *and* $\delta_n \le (n^{-1}\log n)^{1/2}$. *Then there exists a random set* A_{n1} *satisfying* $\boldsymbol{P}(A_{n1}) \ge 1 - a_{n1}$ *with* $a_{n1} = o_n(1)$ *and such that on* A_{n1}, $\|b_n - \beta_0\| \le \delta_n$ *and*

$$\left\| n^{1/2}(b_n - \beta_0) - \eta_n \right\| \le C_0(\delta_n \log n)^{1/2},$$

where C_0 *is a constant whose value depends only on d and the constants from Assumptions* 1-4. *Moreover,* $\Sigma_n^{-1/2}\eta_n \to^d N(0, I_d)$, *where* I_d *is the* $d \times d$ *identity matrix.*

Remark: An immediate corollary of this result is that $n^{1/2}(b_n - \beta_0)$ is asymptotically normal.

The proof relies on the following lemmas.

Lemma 1: *Define* $C_1 = dC_X/2$. *The vector* b_n *satisfies*

$$\left\| \sum_{i=1}^{n} X_i[\boldsymbol{1}(Y_i - X_i'b_n \le 0) - 1/2] \right\| \le C_1. \tag{8.1}$$

Proof: See Koenker and Bassett (1978).

Lemma 2: *Let* $\{\kappa_i : i = 1,\ldots,n\}$ *be independent Bernoulli random variables with parameters* $\{p_i\}$, *and let* $\{c_i : i = 1,\ldots,n\}$ *be constants. Let* V *be a constant such that* $\sum_{i=1}^{n} c_i^2 p_i(1-p_i) \le V^2$. *Given any real* z, *define*

$$G^2 = \max_{1\le i\le n} \exp[zc_i/(2V)].$$

If $G^2 \le 2$ *, then*

(8.2) $$\boldsymbol{P}\left[\pm\sum_{i=1}^{n} c_i(\kappa_i - p_i) > zV\right] \le \exp(-z^2/4).$$

Moreover, if $\sum_{i=1}^{n} c_i^2/4 \le V^2$ *, then for all* $z \ge 0$

(8.3) $$\boldsymbol{P}\left[\pm\sum_{i=1}^{n} c_i(\kappa_i - p_i) > zV\right] \le \exp(-z^2/2).$$

Proof: It follows from Chebyshev's exponential inequality that for every $\lambda > 0$,

$$\boldsymbol{P}\left(\sum_{i=1}^{n} c_i\kappa_i > \sum_{i=1}^{n} c_i p_i + zV\right) \le \exp\left[-\lambda zV - \lambda\sum_{i=1}^{n} c_i p_i\right]\boldsymbol{E}\left[\exp\left(\lambda\sum_{i=1}^{n} c_i\kappa_i\right)\right]$$

$$= \exp\left[-\lambda zV - \lambda\sum_{i=1}^{n} c_i p_i + \sum_{i=1}^{n}\log(1 - p_i + p_i e^{\lambda c_i})\right].$$

The function $f_p(x) = \log(1 - p - pe^x)$ satisfies $f_p(0) = 0$, $f_p'(0) = p$, and

$$f_p''(x) = \frac{p(1-p)e^x}{(1-p+pe^x)^2} \le p(1-p)e^x.$$

Therefore, $f_p(x) \le px + p(1-p)x^2e^x/2$. Set $\lambda = z/(2V)$. Then

$$-\lambda zV - \lambda\sum_{i=1}^{n} c_i p_i + \sum_{i=1}^{n}\log(1 - p_i + p_i e^{\lambda c_i}) \le -\lambda zV + \sum_{i=1}^{n} p_i(1-p_i)\lambda^2 c_i^2 e^{\lambda c_i}/2$$

$$= -\lambda zV + \lambda^2V^2G^2/2.$$

Application of this inequality with $\lambda = z/(2V)$ and $G^2 \le 2$ yields

$$-\lambda zV + \lambda^2V^2G^2/2 \le -z^2(1 - G^2/4)/2 \le -z^2/4.$$

Similarly, one can bound $\boldsymbol{P}\left[\sum_{i=1}^{n} c_i(\kappa_i - p_i) < -zV\right]$, and (8.2) follows.

Next, the inequality $ab \le (a+b)^2/4$ with $a = 1-p$ and $b = pe^x$ implies

$$f_p''(x) = \frac{p(1-p)e^x}{(1-p+pe^x)^2} \le 1/4$$

for all $x \ge 0$ and $p \in [0,1]$. Therefore,

$$-\lambda zV - \lambda\sum_{i=1}^{n} c_i p_i + \sum_{i=1}^{n}\log(1-p_i+p_i e^{\lambda c_i}) \le -\lambda zV + \frac{\lambda^2}{8}\sum_{i=1}^{n} c_i^2$$

$$= -\lambda zV + \lambda^2 V^2/2.$$

This inequality applied with $\lambda = z/V$ yields (8.3). Q.E.D.

We also present a vector version of Lemma 2. For any vector $x \in \mathbb{R}^d$, define $\|x\|_\infty = \max_{1\le j\le d} |x_j|$.

<u>Lemma 3</u>: *Let* $\{\kappa_i : i = 1,\ldots,n\}$ *be independent Bernoulli random variables with parameters* $\{p_i\}$, *and let* $\{c_i : i = 1,\ldots,n\}$ *be constant vectors in* $\mathbb{R}^d$. *Let* V *be a constant such that* $\sum_{i=1}^{n} \|c_i\|_\infty^2\, p_i(1-p_i) \le dV^2$. *Given any real* $z \ge 0$, *define*

$$G^2 = \max_{1\le i\le n} \exp[z\|c_i\|_\infty/(2V)].$$

If $G^2 \le 2$, *then*

$$\boldsymbol{P}\left[\left\|\sum_{i=1}^{n} c_i(\kappa_i - p_i)\right\|_\infty > zVd^{1/2}\right] \le 2d\exp(-z^2/4).$$

Moreover, if $\sum_{i=1}^{n}\|c_i\|_\infty^2/4 \le V^2$, *then for all* $z \ge 0$

$$\boldsymbol{P}\left[\left\|\sum_{i=1}^{n} c_i(\kappa_i - p_i)\right\|_\infty > zV\right] \le 2d\exp(-z^2/2).$$

<u>Proof</u>: Apply Lemma 4.2 to every component of $\sum_{i=1}^{n} c_i(\kappa_i - p_i)$. Q.E.D.

For any fixed $\beta \in \mathbb{R}^d$ define $\xi_i(\beta) = \boldsymbol{1}(Y_i - X_i'\beta \le 0) - \tilde{F}_i(X_i'\beta)$ and

$$\zeta(\beta) = n^{-1/2}\sum_{i=1}^{n} X_i \xi_i(\beta).$$

<u>Lemma 4</u>: *The random field* $\zeta(\beta) \in \mathbb{R}^d$ *satisfies* $\boldsymbol{E}\zeta(\beta) = 0$,

$$\boldsymbol{E}\zeta(\beta)\zeta(\beta)' = n^{-1}\sum_{i=1}^{n} X_i X_i' F_i(X_i'\beta)[1 - F_i(X_i'\beta)] \le \frac{1}{4n}\sum_{i=1}^{n} X_i X_i',$$

$$\boldsymbol{E}\|\zeta(\beta_1) - \zeta(\beta_2)\| \le C_X^2 C_{XX}\|\beta_1 - \beta_2\| + 0.5 C_X^2 C_F a^2 \|\beta_1 - \beta_2\|^2,$$

and, for every $z \ge 0$,

$$\boldsymbol{P}(\|\zeta(\beta)\| > xC_X/2) \le 2\exp(-z^2/2).$$

Proof: The first two statements obviously follow from independence of the Bernoulli random variables ξ_i. It is also straightforward to check that

$$\boldsymbol{E}\,|\xi_i(\beta_1)-\xi_i(\beta_2)|^2 \le |\tilde{F}_i(X_i'\beta_1)-\tilde{F}_i(X_i'\beta_2)|[1-|\tilde{F}_i(X_i'\beta_1)-\tilde{F}_i(X_i'\beta_2)|]$$

$$\le |\tilde{F}_i(X_i'\beta_1)-\tilde{F}_i(X_i'\beta_2)|.$$

A Taylor series expansion and Assumption 1 yield

$$|\tilde{F}_i(X_i'\beta_1)-\tilde{F}_i(X_i'\beta_2)| \le \left\|\tilde{f}_i(X_i'\beta_1)(\beta_1-\beta_2)\right\| + 0.5C_F a^2\,|X_i'(\beta_1-\beta_2)|^2$$

$$\le \tilde{f}_i(X_i'\beta_1)\|\beta_1-\beta_2\| + 0.5C_X^4 C_F a^2\|\beta_1-\beta_2\|^2.$$

Therefore,

$$\boldsymbol{E}\|\zeta(\beta_1)-\zeta(\beta_2)\|^2 = tr\boldsymbol{E}[\zeta(\beta_1)-\zeta(\beta_2)][\zeta(\beta_1)-\zeta(\beta_2)]'$$

$$= n^{-1}tr\sum_{i=1}^{n} X_i X_i'\boldsymbol{E}\,|\xi_i(\beta_1)-\xi_i(\beta_2)|^2$$

$$\le n^{-1}tr\sum_{i=1}^{n} X_i X_i'\,|\tilde{F}_i(X_i'\beta_1)-\tilde{F}_i(X_i'\beta_2)|$$

$$\le n^{-1}C_X^2\|\beta_1-\beta_2\|\sum_{i=1}^{n}\tilde{f}_i(X_i'\beta_1) + 0.5C_X^4 C_F a^2\|\beta_1-\beta_2\|^2$$

$$\le C_X^2 C_{FX}\|\beta_1-\beta_2\| + 0.5C_X^4 C_F a^2\|\beta_1-\beta_2\|^2.$$

The last statement of the lemma now follows from Lemma 3. Q.E.D.

The following lemma establishes stochastic equicontinuity of $\zeta(\beta)$.

Lemma 5: *Let $\alpha\in(1/2,1)$. There are positive constants C_{z1} and C_{z2} such that for every fixed $\beta\in\mathbb{R}^d$,*

$$\boldsymbol{P}\left[\sup_{\tilde{\beta}:\|\tilde{\beta}-\beta\|\le n^{-\alpha}}\left\|\zeta(\beta)-\zeta(\tilde{\beta})\right\|\ge C_{z1}n^{-\alpha+1/2}\right]\le 2\exp\left(-C_{z2}n^{1-\alpha}/4\right).$$

Proof: Let $\tilde{\beta}$ satisfy $\|\tilde{\beta}-\beta\|\le n^{-\alpha}$. It is easy to see that

$$\left\|\zeta(\beta)-\zeta(\tilde{\beta})\right\| = n^{-1/2}\left\|\sum_{i=1}^{n}X_i[\xi_i(\beta)-\xi_i(\tilde{\beta})]\right\|$$

$$\leq n^{-1/2}\sum_{i=1}^{n}\boldsymbol{1}(|Y_i - X_i'\beta| \leq |X'(\tilde{\beta}-\beta)|)$$

,

$$+ n^{-1/2}\left\|\sum_{i=1}^{n} X_i[\tilde{F}_i(X_i'\beta) - \tilde{F}_i(X_i'\tilde{\beta})\right\|.$$

Since $|X_i'(\tilde{\beta}-\beta)| \leq C_X n^{-\alpha}$, for some $|\theta| \leq 1$ we have

$$\left\|\sum_{i=1}^{n} X_i[\tilde{F}_i(X_i'\beta) - \tilde{F}_i(X_i'\tilde{\beta})\right\| = \left\|\sum_{i=1}^{n} X_i\,\tilde{f}_i[X_i'\beta + \theta(\tilde{\beta}-\beta)]X_i'(\tilde{\beta}-\beta)\right\|$$

$$\leq C_X^2 C_{XX}\, n\left\|\tilde{\beta}-\beta\right\|$$

$$\leq C_X^2 C_{XX}\, n^{1-\alpha}.$$

Therefore,

$$\text{(8.4)} \qquad \left\|\zeta(\beta) - \zeta(\tilde{\beta})\right\| \leq n^{-1/2} C_X \sum_{i=1}^{n}\tau_i + C_X^2 C_{XX}\, n^{1/2-\alpha},$$

where the $\tau_i \equiv \boldsymbol{1}(|Y_i - X'\beta| \leq C_X\, n^{-\alpha})$ are Bernoulli random variables with

$$p_i \equiv \boldsymbol{E}\tau_i = \boldsymbol{P}(|Y_i - X_i'\beta| \leq C_X n^{-\alpha}) = \tilde{F}_i(X_i'\beta + C_X n^{-\alpha}) - \tilde{F}_i(X_i'\beta - C_X n^{-\alpha})\,..$$

As in the proof of Lemma 4, one bounds

$$\sum_{i=1}^{n} p_i \leq 2C_X n^{-\alpha}\sum_{i=1}^{n}\tilde{f}_i(X_i'\beta) + C_X^2 C_F a^2 n^{-2\alpha+1}$$

$$\leq 2C_X C_{XX} n^{-\alpha-1} + C_X^2 C_F a^2 n^{-2\alpha+1} \leq C_{z2} n^{-\alpha+1}$$

for some constant $C_{z2} \approx 2C_X C_{XX}$. Application of Lemma 2 with $c_i = 1$, $z = V$, and $V^2 = \sum_{i=1}^{n} p_i \leq C_{z2} n^{1-\alpha}$ (so that $G^2 = e^{1/2} < 2$) yields

$$\boldsymbol{P}\left[\sum_{i=1}^{n}(\tau_i - p_i) \geq V^2\right] \leq 2\exp(-V^2/4).$$

Therefore,

$$\boldsymbol{P}\left(\sum_{i=1}^{n}\tau_i \geq 2C_{z2} n^{1-\alpha}\right) \leq 2\exp(-C_{z2} n^{1-\alpha}/4).$$

This inequality and (8.4) yield the result. Q.E.D.

The next lemma gives a uniform bound for $\zeta(\beta)-\zeta(\beta_0)$ when $\|\beta-\beta_0\|\le\delta$.

Lemma 6: *Let* $n^{-1/2}\le\delta\le 1$. *Then for some constant* C_2 *depending on* d, C_F *and* C_X *only, there exists a random set* A_{n2} *satisfying* $\boldsymbol{P}(A_{n2})\ge 1-a_{n2}$ *with* $a_{n2}=o_n(1)$ *such that on* A_{n2}

$$\sup_{\beta:\|\beta-\beta_0\|\le\delta}\|\zeta(\beta)-\zeta(\beta_0)\|\le C_2(\delta\log n)^{1/2}.$$

Proof: Let $\mathcal{D}_n$ be a ε-net in the ball $\{\beta:\|\beta-\beta_0\|\le\delta\}$ with the step $n^{-\alpha}$ for $\alpha=3/4$. This net can be constructed with cardinality $(2\delta n^{\alpha})^d\le(2n^{3/4})^d$. Fix $\beta\in\mathcal{D}_n$. By Lemma 4, $\boldsymbol{E}\|\zeta(\beta)-\zeta(\beta_0)\|^2\le dC_{z3}\|\beta-\beta_0\|$ for some constant $C_{z3}\approx C_X^2 X_{XX}/d$. Now apply Lemma 3 to $\zeta(\beta)-\zeta(\beta_0)$ with $c_i=n^{-1/2}X_i$, $V^2=C_{z3}\delta$ and $z=(4d\log n)^{1/2}$. Then

$$\log G^2\le C_X n^{-1/2}z/(2V)=C_X n^{-1/2}(d\log n)^{1/2}/(C_{z3}\delta)^{1/2}.$$

Clearly, $\delta\ge n^{-1/2}$ implies $G^2\le 2$ for n sufficiently large. By (8.3)

$$\boldsymbol{P}\left[\|\zeta(\beta)-\zeta(\beta_0)\|\ge 2d(C_{z3}\delta\log n)^{1/2}\right]\le 2de^{-d\log n}.$$

Now

$$\boldsymbol{P}\left[\sup_{\beta:\|\beta-\beta_0\|\le\delta}\|\zeta(\beta)-\zeta(\beta_0)\|\ge 2d(C_{z3}\delta\log n)^{1/2}+C_{z1}n^{-\alpha+1/2}\right]$$

$$\le\sum_{\beta\in\mathcal{D}_n}\boldsymbol{P}\left[\sup_{\tilde\beta:\|\beta-\tilde\beta\|\le n^{-\alpha}}\|\zeta(\beta)-\zeta(\tilde\beta)\|\ge C_{z1}n^{-\alpha+1/2}\right]$$

$$+\sum_{\beta\in\mathcal{D}_n}\boldsymbol{P}\left[\|\zeta(\beta)-\zeta(\beta_0)\|\ge 2d(C_{z3}\delta\log n)^{1/2}\right]$$

$$\le(2n^{3/4})^d[\exp(-C_{z2}n^{1-\alpha}/4)+2d\exp(-d\log n)]\equiv a_{n2}\to 0$$

as $n\to\infty$. The lemma follows because $\delta^{1/2}\le n^{1/4}$ and $n^{-\alpha+1/2}=n^{1/4}$. Q.E.D.

Define

$$B(\beta)=n^{-1/2}\sum_{i=1}^{n}X_i[\tilde F_i(X_i'\beta)-\tilde F_i(X_i'\beta_0)].$$

Note that $B(\beta)=\boldsymbol{E}[\zeta(\beta)-\zeta(\beta_0)]$. The next lemma states that $B(\beta)$ is nearly linear in a small neighborhood of β_0. Let $\tilde F(X'\beta)$ be the vector whose components are $\tilde F_i(X_i'\beta)$.

Lemma 7: *For all* β

$$\left\| \tilde{F}(X'\beta) - \tilde{F}(X'\beta_0) - \mathcal{F}\mathcal{X}'(\beta - \beta_0) \right\| \le 0.5 n^{1/2} C_F a^2 \left\| \beta - \beta_0 \right\|^2$$

and

$$\left\| B(\beta) - n^{1/2} Q_n (\beta - \beta_0) \right\| \le \left\| 0.5 a^2 C_F n^{-1/2} \sum_{i=1}^{n} X_i \mid X_i'(\beta - \beta_0) \mid^2 \right\| \le C_3 a^2 n^{1/2} \left\| \beta - \beta_0 \right\|^2 ,$$

where $C_3 = 0.5 C_X^2 C_F$.

Proof: This result follows from a Taylor series expansion and Assumption 1. Q.E.D.

Lemma 8: *Let the sequence* $\{\delta_n\}$ *satisfy* $n^{-1/2} / \delta_n \to 0$ *as* $n \to \infty$. *Then there exists a random set* A_{n3} *satisfying* $\boldsymbol{P}(A_{n3}) \ge 1 - a_{n3}$ *with* $a_{n3} = o_n(1)$ *such that on* A_{n3}, $\| b_n - \beta_0 \| \le \delta_n$ ..

Proof: Lemma 7 and Assumption 4 imply that

$$\inf_{\beta: \|\beta - \beta_0\| \ge \delta_n} B(\beta) \to \infty$$

as $n \to \infty$. By Lemmas 4 and 6, $\zeta(\beta)$ is bounded in probability in every neighborhod of β_0. Moreover, (8.1) implies that $\| \zeta(b_n) - B(b_n) \| \le C_1 n^{-1/2}$. The lemma follows from this inequality and monotonicity arguments. See Portnoy (1991) for details. Q.E.D.

Define $\eta_n = -Q_n^{-1} \zeta(\beta_0) = -n^{-1/2} Q_n^{-1} \sum_{i=1}^{n} X_i \xi_i(\beta_0)$.

Lemma 9: $\Sigma_n^{-1/2} \eta_n \to^d N(0, I_d)$.

Proof: By Lemma 4, $\boldsymbol{E}\eta_n = 0$ and $\boldsymbol{E}\eta_n \eta_n' = \Sigma_n$. Asymptotic normality follows from the central limit theorem for sums of uniformly bounded random variables. See Koenker and Bassett (1978) for details. Q.E.D.

Proof of Proposition 1: By definition

$$n^{-1/2} \sum_{i=1}^{n} X_i [\boldsymbol{1}(Y_i - X_i'\beta \le 0) - 1/2] = \zeta(\beta) + B(\beta) .$$

By Lemma 1,

$$\| \zeta(b_n) + B(b_n) \| \le C_1 n^{-1/2} .$$

Let δ_n satisfy $n^{1/2} \delta_n \to 0$ as $n \to \infty$. Then by Lemmas 6 and 7, on the random set A_{n2} $\| \zeta(\beta) - \zeta(\beta_0) \| \le C_2 (\delta_n \log n)^{1/2}$ and $\left\| B(b_n) - n^{1/2} Q_n (b_n - \beta_0) \right\| \le C_3 a^2 \delta_n^2 n^{1/2}$ for all β with $\| \beta - \beta_0 \| \le \delta_n$. Similarly, by Lemma 8 $\| b_n - \beta_0 \| \le \delta_n$ on A_{n3}. Define $A_{n1} = A_{n2} \cap A_{n3}$. Then $\boldsymbol{P}(A_{n1}) \ge 1 - a_{n2} - a_{n3}$, and on A_{n1},

$$\|\zeta(b_n)-\zeta(\beta_0)\| \le C_2(\delta_n \log n)^{1/2}$$

and

$$\left\|B(b_n)-n^{1/2}Q_n(b_n-\beta_0)\right\| \le C_3 a^2 \delta_n^2 n^{1/2}.$$

Thus,

$$\left\|\zeta(\beta_0)+n^{1/2}Q_n(b_n-\beta_0)\right\| \le C_1 n^{-1/2} + C_2(\delta_n \log n)^{1/2} + C_3 a^2 n^{1/2}\delta_n^2.$$

By Assumption 4, $\left\|Q_n^{-1}\right\|_\infty \le C_{FX}$, and hence

$$\left\|n^{1/2}(b_n-\beta_0)+Q_n^{-1}\zeta(\beta_0)\right\| \le C_{XX}\left[C_1 n^{-1/2} + C_2(\delta_n \log n)^{1/2} + C_3 a^2 n^{1/2}\delta_n^2\right].$$

But $\eta_n = -Q_n^{-1}\zeta(\beta_0)$. If δ_n satisfies $\delta_n \le (n^{-1}\log n)^{1/2}$, then on the set A_{n1}, $\left\|n^{1/2}(b_n-\beta_0)-\eta_n\right\| \le C_0(\delta_n \log n)^{1/2}$ with probability approaching 1, where C_0 is slightly larger than C_2. Asymptotic normality follows from Lemma 9. Q.E.D.

8.2 Properties of Nonparametric Smoothers

Let W_h be the matrix whose (i,j) element is $w_{ij,h}$. Let $\|W_h\|_\infty = \sup_{\lambda\in\mathbb{R}^n}\|W_h\lambda\|/\|\lambda\|$.

Lemma 10 (Horowitz and Spokoiny (2000)): *There exist constants* C_{w1}, C_a, C_{V1} *and* C_{V2} *depending only on the constants in Assumption 4 such that for all* $h\in\mathcal{H}$, $\sum_{i=1}^n w_{ij,h} \le C_{w1}$ $(j=1,\ldots,n)$, $\|W_h\|_\infty \le C_{w1}$, $\sum_{i=1}^n a_{ii,h} \le C_a h^{-1}$, $\sum_{i=1}^n \sum_{j\ne i} a_{ij,h}^2 \ge C_{V1}h^{-1}$, *and* $\sum_{i=1}^n \sum_{j=1}^n a_{ij,h}^2 \le C_{V2}h^{-1}$.

Proof: See HS. Q.E.D.

8.3 Asymptotic Expansion of the Statistics S_h

For every $\beta\in\mathbb{R}^d$ define

$$S_h(\beta)=\sum_{i=1}^n\left(\sum_{j=1}^n w_{ij,h}[\boldsymbol{1}(Y_j - X_j'\beta\le 0)-1/2]\right)^2$$

$$=\sum_{i=1}^n\left(\sum_{j=1}^n w_{ij,h}[\xi_j(\beta)+\tilde{F}_j(X_j'\beta)-1/2]\right)^2.$$

Also define $z_j(\beta) = \tilde{F}_j(X_j'\beta) - \tilde{F}_j(X_j'\beta_0)$. We use a matrix representation of $S_h(\beta)$. Let $\zeta(\beta)$, $\tilde{F}(X'\beta)$, and $z(\beta)$, respectively, be the vectors in $\mathbb{R}^n$ with components $\xi_j(\beta)$, $\tilde{F}_j(X_j'\beta)$, and $z_j(\beta)$. Let W_h be the $n \times n$ matrix whose (i, j) element is $w_{ij,h}$. Then

$$S_h(\beta) = \left\| W_h[\xi(\beta) + \tilde{F}(X'\beta) - 1/2] \right\|^2 = \left\| W_h[\xi(\beta) + \tilde{F}(X'\beta_0) - 1/2 + z(\beta)] \right\|^2.$$

Under the null hypothesis, $\tilde{F}_j(X_j'\beta_0) = 1/2$, so $S_h(\beta) = \left\| W_h[\xi(\beta) + z(\beta)] \right\|^2$. The test statistic is based on $S_h(b_n)$. Lemma 11 enables us to obtain an asymptotic expansion for $S_h \equiv S_h(b_n)$. Define the $n \times n$ matrices $A_h = W_h'W_h$ and $\Pi_F = \mathcal{F}\mathcal{X}'(\mathcal{X}\mathcal{F}\mathcal{X}')^{-1}\mathcal{X}$.

Lemma 11: *Let δ_n satisfy $\delta_n \le (n^{-1}\log n)^{1/2}$ and $n^{-1/2}/\delta_n \to 0$ as $n \to \infty$. There exist a constant C_9 and a random set A_{n4} satisfying $\boldsymbol{P}(A_{n4}) \ge 1 - a_{n4}$ with $a_{n4} = o_n(1)$ such that on A_4*

$$\left| S_h - \left\| W_h[(\boldsymbol{I}_n - \Pi_F)\xi(\beta_0) + \tilde{F}(X'\beta_0) - 1/2] \right\|^2 \right| \le C_9 \delta_n^{1/2} h^{-1} \log n$$

for all $h \in \mathcal{H}$.

Proof: We prove this lemma under the null hypothesis only. The general case can be considered similarly. For all β such that $\|\beta - \beta_0\| \le \delta_n$, Assumption 1 yields

$$\left| \tilde{F}_i(X_i'\beta) - \tilde{F}_i(X_i'\beta_0) \right| \le C_5 a \delta_n, \tag{8.5}$$

whre $C_5 = C_F C_X$. We now bound the differences $\left\| W_h[\xi(\beta) + z(\beta)] \right\|^2 - \left\| W_h[\xi(\beta_0) + z(\beta)] \right\|^2$ uniformly over $h \in \boldsymbol{\mathcal{H}}$ and β with $\|\beta - \beta_0\| \le \delta_n$. Define $\eta_h(\beta) = W_h \xi(\beta)$. As in Lemma 4, each element of $\eta_h(\beta)$ satisfies $\boldsymbol{E}\eta_{i,h}(\beta) = 0$, $\boldsymbol{E}\eta_{i,h}(\beta)^2 \le (1/4)\sum_{j=1}^n w_{ij,h}^2 = a_{ii,h}/4$, and

$$\boldsymbol{E}[\eta_{i,h}(\beta) - \eta_{i,h}(\beta_0)]^2 = \sum_{j=1}^n w_{ij,h}^2 \left| \tilde{F}_j(X_j'\beta) - \tilde{F}_j(X_j'\beta_0) \right|$$

$$\le C_5 a \delta_n \sum_{j=1}^n w_{ij,h}^2 = C_5 a \delta_n a_{ii,h}.$$

As in Lemma 6, there exists a random set A_{n5} satifsying $\boldsymbol{P}(A_{n5}) \ge 1 - a_{n5}$ with $A_{n5} = o_n(1)$ such that on A_{n5} $|\eta_{i,h}(\beta) - \eta_{i,h}(\beta_0)| \le C_6(a_{ii,h} a \delta_n \log n)^{1/2}$, and

$$|\eta_{i,h}(\beta_0)| \le C_7(a_{ii,h}\log n)^{1/2} \tag{8.6}$$

for some constants C_6 and C_7, all $h \in \mathcal{H}$, all β such that $\|\beta - \beta_0\| \le \delta_n$, and all $i = 1,...,n$. This and (8.5) imply that on A_n

$$\|W_h[\xi(\beta) - \xi(\beta_0)]\|^2 \le C_6^2 a\delta_n tr(A_h)\log n,$$

$$\|W_h \xi(\beta_0)\|^2 \le C_7^2 tr(A_h)\log n,$$

and

$$\|W_h[\xi(\beta_0) + z(\beta)\| \le C_7[tr(A_h)\log n]^{1/2} + C_5 a n^{1/2}\delta_n .$$

Now by the inequality $|\,\|x\|^2 - \|y\|^2\,| \le \|x - y\|(\|x - y\| + 2\|x\|)$ and Lemma 10, the following holds on A_n for all β satisfying $\|\beta - \beta_0\| \le \delta_n$:

$$\left|\|W_h[\xi(\beta) + z(\beta)\|^2 - \|W_h[\xi(\beta_0) + z(\beta)\|^2\right| \le C_6(a\delta_n)^{1/2}[C_6(a\delta_n)^{1/2} + 2C_7]tr(A_h)\log n$$

$$\le C_8(a\delta_n)^{1/2} h^{-1}\log n. \tag{8.7}$$

By Proposition 1, $\|b_n - \beta_0\| \le \delta_n$ on the set A_{n1} satisfying $\boldsymbol{P}(A_{n1}) = 1 - a_{n1}$. Therefore, on $A_{n4} = A_{n1} \cap A_{n5}$, inequality (8.7) holds when β is replaced by b_n. Also, Proposition 1 and Assumption 1 imply that on the set A_{n1},

$$\|\tilde{F}(X'b_n) - \tilde{F}(X'\beta_0) - \Pi_F \xi(\beta_0)\| \le C_0'(\delta_n \log n)^{1/2},$$

where $C_0{}' \approx C_0 C_X C_F$. Therefore, by Lemma 10,

$$\|W_h[z(b_n) - \Pi_F \xi(\beta_0)]\| = \|W_h[\tilde{F}(X'b_n) - \tilde{F}(X'\beta_0) - \Pi_F \xi(\beta_0)\|$$

$$C_9 a(\delta_n \log n)^{1/2}$$

on A_{n4}, where $C_9 = C_0' C_{w1}$. The proof is now completed similarly to (8.7). Q.E.D.

Lemma 11 implies that under the null hypothesis, S_h can be approximated by $\|W_h \xi(\beta_0) - W_h \Pi_F \xi(\beta_0)\|^2$. The second term in this expression comes from the parametric LAD fit. The next lemma shows that the effect of this term is asymptotically negligible when $h_{\max} \to 0$ as $n \to \infty$.

<u>Lemma 12</u>: *Let $h_{\max} \to 0$ as $n \to \infty$. Then under the null hypothesis there exists a random set A_{n6} satisfying $\boldsymbol{P}(A_{n6}) \ge 1 - a_{n6}$ with $a_{n6} = o_n(1)$ such that on A_{n6}*

$$\sup_{h \in \mathcal{H}} h^{1/2}\left|S_h - \|W_h \xi(\beta_0)\|^2\right| \le s_{n1} = o_n(1).$$

Proof: By Lemma 11, it suffices to show that there exists a random set A_{n7} satisfying $\boldsymbol{P}(A_{n7}) \geq 1 - a_{n7}$ *with* $a_{n7} = o_n(1)$ *such that on* A_{n7}

$$\tau_n \equiv \sup_{h \in \mathcal{H}} h^{1/2} \|W_h(\boldsymbol{I}_n - \Pi_F)\xi(\beta_0)\|^2 - \|W_h \xi(\beta_0)\|^2 \leq s_{n2} = o_n(1).$$

This would follow from

$$\sum_{h \in \mathcal{H}} h^{1/2} \boldsymbol{E}[\|W_h(\boldsymbol{I}_n - \Pi_F)\xi(\beta_0)\|^2 - \|W_h \xi(\beta_0)\|^2] = o_p(1)$$

and

$$\sum_{h \in \mathcal{H}} h^{1/2} Var[\|W_h(\boldsymbol{I}_n - \Pi_F)\xi(\beta_0)\|^2 - \|W_h \xi(\beta_0)\|^2] = o_p(1).$$

The definition of $\xi(\beta_0)$ yields $\boldsymbol{E}\xi(\beta_0)\xi(\beta_0)' = \boldsymbol{I}_n / 4$. Since Π_F is a projection operator in $\mathbb{R}^n$ onto a d-dimensional subspace, $tr(\Pi_F) = d$. This and Lemma 10 imply that

$$\begin{aligned}
\boldsymbol{E}[\|W_h \xi(\beta_0)\|^2 - \|W_h(\boldsymbol{I}_n - \Pi_F)\xi(\beta_0)\|^2] &= 2\boldsymbol{E}tr[W_h \Pi_F \xi(\beta_0)\xi(\beta_0)' W_h'] \\
&\quad - \boldsymbol{E}tr[W_h \Pi_F \xi(\beta_0)\xi(\beta_0)' \Pi_F W_h'] \\
&= (1/4) tr(W_h \Pi_F W_h') \\
&\leq \|W_h\|_\infty^2 \, tr(\Pi_F) \leq C_{w1}^2 d / 4.
\end{aligned}$$

Similarly

$$\begin{aligned}
&Var[\|W_h \xi(\beta_0)\|^2 - \|W_h(\boldsymbol{I}_n - \Pi_F)\xi(\beta_0)\|^2] \\
&\qquad = Var[\xi(\beta_0)'(\Pi_F A_h + A_h \Pi_F - \Pi_F A_h \Pi_F)\xi(\beta_0)] \\
&\qquad \leq (1/2) tr(\Pi_F A_h + A_h \Pi_F - \Pi_F A_h \Pi_F)^2 \leq C
\end{aligned}$$

where C is a constant that depends only on C_{w1} and d. Since $\mathcal{H}$ is a geometric grid,

$$\sum_{h \in \mathcal{H}} h^{1/2} \leq C_{h1} h_{\max}^{1/2} \to 0.$$

A similar result holds for $\sum_{h \in \mathcal{H}} h$. The result of the lemma follows. Q.E.D.

The results of Lemmas 10 and 12 imply that under the null hypothesis and on A_{n6},

(8.8) $$\sup_{h \in \mathcal{H}} | T_h^* - T_{h,0} | \leq s_{n3} = o_n(1),$$

where

$$T_{h,0} = \frac{\|W_h\xi(\beta_0)\|^2 - (1/4)\sum_{i=1}^{n} a_{ii,h}^2}{\left(\frac{1}{8}\sum_{i=1}^{n} a_{ii,h}^2\right)^{1/2}} .$$

8.4 Proof of Theorem 1

Relation (8.8) reduces the proof to considering $\sup_{h\in\mathcal{H}} T_{h,0}$. $T_{h,0}$ is the centered, standardized quadratic form $\|W_h\xi(\beta_0)\|^2$, and $\xi(\beta_0)$ is a vector of independently and identically distributed Bernoulli random variables with parameters 1/2 and means of zero. The distribution of $T_{h,0}$ does not depend on the unknown distributions of the ε_i's in (1.1). The distribution of $\sup_{h\in\mathcal{H}} T_{h,0}$ is investigated in HS and Spokoiny (2000). Here, we briefly review the main issues.

Let $\tilde{\xi}$ be an $n\times 1$ Gaussian random vector with zero mean and covariance matrix $\boldsymbol{I}_n/4$. Define $\tilde{T}_{h,0}$ by centering and standardizing $\|W_h\tilde{\xi}\|^2$. Then $\sup_{h\in\mathcal{H}} T_{h,0}$ is close in distribution to $\tilde{T} = \sup_{h\in\mathcal{H}} \tilde{T}_{h,0}$. Let $\tilde{t}_\alpha$ be the $1-\alpha$ quantile of the distribution of $\tilde{T}$. Then $\tilde{t}_\alpha = O(\sqrt{\log\log n})$ and $\tilde{T}$ has a bounded, continuous density at $\tilde{t}_\alpha$. This and (8.8) imply Theorem 1. See HS and Spokoiny (2000) for details.

8.5 Proofs of Theorems 2 and 3

The next proposition gives sufficient conditions for consistency of the adaptive, rate-optimal test. Define $\Delta_i = \tilde{F}_i(X_i'\beta_0) - 1/2$. Let Δ be the vector in $\mathbb{R}^n$ with elements Δ_i. Define $(V_h^*)^2 = (1/8)\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij,h}^2 = tr(A_h^2)/8$.

<u>Proposition 2</u>: *Suppose there is a sequence* $\{r_n\}$ *such that* $r_n \to \infty$ *as* $n\to\infty$ *and*

(8.9) $$\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} a_{ij,h}\Delta_i\Delta_j \ \geq\ (t_\alpha + r_n)V_h^*$$

for some $h\in\mathcal{H}$. *Then*

$$\lim_{n\to\infty} \boldsymbol{P}(T > t_\alpha) \geq 1 - s_{n4} = 1 - o_n(1) .$$

<u>Proof</u>: It suffices to show that for a given $h\in\mathcal{H}$, $\boldsymbol{P}(T_h < t_\alpha) \leq s_{n4} = o_n(1)$ as $n\to\infty$. The asymptotic expansion from Lemma 11 reduces this condition to

$$\boldsymbol{P}\left[\frac{\|W_h(\boldsymbol{I}_n-\Pi_F)\xi(\beta_0)+W_h\Delta\|^2-tr(A_h)/4}{V_h*}<t_\alpha\right]\le s_{n5}=o_n(1)\,.$$

Now

$$\|W_h(\boldsymbol{I}_n-\Pi_F)\xi(\beta_0)+W_h\Delta\|^2-tr(A_h)/4$$

$$=\|W_h\Delta\|^2+\|W_h(\boldsymbol{I}_n-\Pi_F)\xi(\beta_0)\|^2-tr(A_h)/4+2\Delta'W_h'W_h(\boldsymbol{I}_n-\Pi_F)\xi(\beta_0).$$

Note that under the alternative model, the elements $\xi_i(\beta_0)$ of the vector $\xi(\beta_0)$ are independent Bernoulli random variables with zero means and parameters $\tilde{F}(X_i'\beta_0)=\Delta_i+1/2$. By Lemma 2,

$$\boldsymbol{P}\left[\frac{\Delta'W_h'W_h(\boldsymbol{I}_n-\Pi_F)\xi(\beta_0)}{\|\Delta'W_h'W_h(\boldsymbol{I}_n-\Pi_F)\|}>r_n/2\right]\le e^{-0.5(r_n/2)^2}\to 0$$

as $n\to\infty$. Moreover, because the elements of Δ satisfy $|\Delta_i|\le 1/2$ and $\boldsymbol{I}_n-\Pi_F$ is a projection operator in $\mathbb{R}^n$, it follows that

$$\|\Delta'W_h'W_h(\boldsymbol{I}_n-\Pi_F)\|^2\le(1/4)tr(W_h'W_h)^2=2(V_h*)^2\,.$$

Therefore, there exists a random set A_{n8} satisfying $\boldsymbol{P}(A_{n8})\ge 1-a_{n8}$ with $a_{n8}=o_n(1)$ such that on A_{n8}, $2\Delta'W_h'W_h(\boldsymbol{I}_n-\Pi_F)\xi(\beta_0)\le(r_n/\sqrt{2})V_h*$. As in the proof of Lemma 12, one can show that

$$\boldsymbol{E}\left[\|W_h(\boldsymbol{I}_n-\Pi_F)\xi(\beta_0)\|^2-\boldsymbol{E}\|W_h\xi(\beta_0)\|^2\right]^2\le(V_h*+s_{n6})^2\,,$$

where $s_{n6}=o_n(1)$. This implies by the Cauchy-Schwartz inequality that

$$\boldsymbol{P}\left[\|W_h(\boldsymbol{I}_n-\Pi_F)\xi(\beta_0)\|^2-\boldsymbol{E}\|W_h\xi(\beta_0)\|^2>(r_n/4)V_h*\right]\le\frac{(V_h*+s_{n6})^2}{(r_nV_h*/4)^2}\to 0$$

as $n\to\infty$. Thus, there exists a random set A_{n9} satisfying $\boldsymbol{P}(A_{n9})\ge 1-a_{n9}$ with $a_{n9}=o_n(1)$ such that on A_{n9},

$$(V_h*)^{-1}\left|\|W_h(I-\Pi_F)\xi(\beta_0)\|^2-\boldsymbol{E}\|W_h\xi(\beta_0)\|^2\right|\le r_n/4\,.$$

Since $\boldsymbol{E}\xi_i(\beta_0)^2=1/4-\Delta_i^2$, it follows that

$$\boldsymbol{E}\|W_h\xi(\beta_0)\|^2=\boldsymbol{E}\sum_{i=1}^n\sum_{j=1}^n a_{ij,h}\xi_i(\beta_0)\xi_j(\beta_0)=\sum_{i=1}^n a_{ii,h}(1/4-\Delta_i^2)$$

so that

$$\boldsymbol{E}\|W_h\xi(\beta_0)\|^2 - (1/4)tr(A_h) = -\sum_{i=1}^{n} a_{ii,h}\Delta_i^2 \, .$$

Since, in addition, $\|W_h\Delta\|^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij,h}\Delta_i\Delta_j$, we conclude that on $A_{n8} \cap A_{n9}$

$$(V_h^*)^{-1}\left[\|W_h(\boldsymbol{I}_n - \Pi_F)\xi(\beta_0) + W_h\Delta\|^2 - tr(A_h)/4\right] - t_\alpha$$

$$\geq (V_h^*)^{-1}\sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} a_{ij,h}\Delta_i\Delta_j - t_\alpha - r_n/\sqrt{2} - r_n/4$$

$$\geq (3/4 - 1/\sqrt{2})r_n > 0$$

for sufficiently large n ,and the proposition follows. Q.E.D.

Proof of Theorem 2: Define Δ as in Proposition 2. Set $\rho_n = Cn^{-1/2}\sqrt{\log\log n}$ for some finite $C > 0$. The definition of β_0 (eq. 3.1) implies that

$$\sum_{i=1}^{n} X_i[\tilde{F}_i(X_i'\beta_0) - 1/2] = \sum_{i=1}^{n} X_i\{F_i[X_i'\beta_0 - m(X_i)] - F_i(0)] = 0 \, .$$

Under the local alternative $m(x) = x'\beta + \rho_n g(x)$, this means that

$$\sum_{i=1}^{n} X_i\{F_i[X_i'\beta_0 - X_i'\beta - \rho_n g(X_i)] - F_i(0)\} = 0 \, .$$

Assumptions 1 and 4(iv) and a Taylor series expansion yield

$$\left\|\sum_{i=1}^{n} X_i f_i(0)[X_i'\beta_0 - X_i'\beta - \rho_n g(X_i)]\right\| = \|\mathcal{X}\mathcal{F}[\mathcal{X}(\beta_0 - \beta) - \rho_n\mathcal{G}]\| < n\rho_n s_{n7} \, ,$$

where $s_{n7} = o_n(1)$. This and Assumption 4(iii) imply that

$$\|\beta_0 - \beta - \rho_n(\mathcal{X}\mathcal{F}\mathcal{X}')^{-1}\mathcal{X}\mathcal{F}\mathcal{G}\| < \rho_n s_{n8} \, ,$$

where $s_{n8} = C_{XX}s_{n7} = o_n(1)$. But $\Delta_i \equiv \tilde{F}_i(X_i'\beta_0) - 1/2 = F_i[X_i'\beta_0 - X_i'\beta\rho - \rho_n g(X_i)]$ under the local alternative. Therefore,

$$\|\Delta + \rho_n\mathcal{F}[\mathcal{X}'(\mathcal{X}\mathcal{F}\mathcal{X}')^{-1}\mathcal{X}\mathcal{F}\mathcal{G} - \mathcal{G}]\|^2 < n\rho_n^2 s_{n9}$$

with $s_{n9} = o_n(1)$ or, equivalently, $\|\Delta - \rho_n\mathcal{F}\Pi\mathcal{G}\|^2 < n\rho_n^2 s_{n9}$. As in HS, one can show that $h_{\max} \to 0$ and continuity of $f_i(0)$ and g imply that $\|W_h\mathcal{F}\Pi\mathcal{G}\| / \|\mathcal{F}\Pi\mathcal{G}\| \to 1$ as $n \to \infty$. This result and (4.2) imply that for sufficiently large n,

$$\|\Delta\|^2 \geq 0.5\|\mathcal{F}\Pi\mathcal{G}\|^2 \geq C\delta \log\log n,$$

where $\delta > 0$ is as in (4.2). By Lemma 10, $C_{V1}/h \leq (V_h^*)^2 \leq C_{V2}/h$ for finite constants C_{V1} and C_{V2}. Therefore, setting $h = h_{\max}$ and $r_n = (\log\log n)^{1/4}$ and noting that $t_\alpha = O\left(\sqrt{\log\log n}\right)$ yields $(t_\alpha + r_n)V_h^* = O(\log\log n)$. It follows that (8.9) holds for all sufficiently large C. The theorem now follows from Proposition 2. Q.E.D.

Proof of Theorem 3: It is straightforward to see that for a continuous $\Delta \in \mathcal{S}(s)$

$$\frac{\sum_{i=1}^{n} a_{ii,h}\Delta_i^2}{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij,h}\Delta_i\Delta_j} = o(1).$$

Moreover,

(8.10) $\quad \|W_h\Delta\| \geq C_{s1}\|\Delta\| - C_{2s}n^{1/2}h^s$

for constants C_{s1} and C_{s2} that depend only on the design $\{X_i : i = 1,...,n\}$. See HS (proof of Theorem 4). Now set $t_n = t_\alpha + \sqrt{2\log\log n} = O(\sqrt{\log\log n})$. Define h to be the element of $\mathcal{H}$ that is closest from below to $(n/t_n)^{-2/(4s+d-1)}$. Since $\mathcal{H}$ is a geometric grid, $h \leq (n/t_n)^{-2/(4s+d-1)}$ and $h \approx (n/t_n)^{-2/(4s+d-1)}$. By Lemma 10, $(V_h^*)^{-1} \leq C_V^{-2}h^{1/2}$ for some fixed constant C_V. Now the inequality $n^{-1/2}\|\Delta\| \geq C_{s1}^{-1}(C_{s2} + C_V)(n/t_n)^{-2s/(4s+d-1)}$ and (8.10) yield

$$(V_h^*)^{-1}\|W_h\Delta\|^2 \geq C_V^{-2}(n/t_n)^{-1/(4s+d-1)}(C_{s1}\|\Delta\| - C_{s2}n^{1/2}h^s)^2 \geq t_n.$$

Therefore, $(V_h^*)^{-1}\|W_h\Delta\|^2 - t_\alpha \to \infty$ as $n \to \infty$, as is required to prove the theorem. Q.E.D.

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.

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 D.K. Ginther (2000). Integrated Conditional Moment Testing of Quantile Regression Models, working paper, Department of Economics, Pennsylvania State University.

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

Buchinsky, M. (1994). Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression, *Econometrica*, 62, 405-458.

Buchinsky, M. (1998). Recent Advances in Quantile Regression Models, *Journal of Human Resources*,33, 88-126.

Chamberlain, G. (1994). Quantile Regression, Censoring, and the Structure of Wages, in *Advances in Econometrics, Sixth World Congress*, C.A. Sims, ed., New York: Cambridge University Press.

Chaudhuri, P. (1991). Global Nonparametric Estimation of Conditional Quantile Functions and Their Derivatives, *Journal of Multivariate Analysis*, 39, 246-269.

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, J. (1996). Test of Significance Based on Wavelet Thresholding and Neyman's Truncation, *Journal of the American Statistical Association*, 91, 674-688.

Fan, J. and L.-S. Huang (2000). Goodness of Fit Tests for Parametric Regression Models, *Journal of the American Statistical Association*, forthcoming.

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

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

Guerre, E. and P. Lavergne (1999). Minimax Rates for Nonparametric Specification Testing in Regression Models, working paper, Toulouse University of Social Sciences, Toulouse, France.

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

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

Hong, Y. and H. White (1996). Consistent Specification Testing via Nonparametric Series Regressions, *Econometrica*, 63, 1133-1160.

Horowitz, J.L. and V.G. Spokoiny (2001). An Adaptive, Rate-Optimal Test of a Parametric Mean Regression Model against a Nonparametric Alternative, *Econometrica*, 69, 599-631.

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.

Ledwina, T. (1994). Data-Driven Version of Neyman's Smooth Test of Fit, *Journal of the American Statistical Association*, 89, 1000-1005.

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

Koenker, R. and Bassett, G. (1978), "Regression Quantiles," *Econometrica*, 46, 33-50.

Koenker, R. and O. Geling (1999). Reappraising Medfly Longevity: A Quantile Regression Survival Analysis, working paper, Department of Economics, University of Illinois.

Manning, W.G., L. Blumber, and L.H. Moulton (1995). The Demand for Alcohol: The Differential Response to Price, *Journal of Health Economics*, 14, 123-148.

Portnoy, S. (1991). Asymptotic Behavior of Regression Quantiles in Non-stationary, Dependent Cases, *Journal of Multivariate Analysis*, 38, 100-113.

Poterba, J.M. and Rueben, K.S. (1994). The Distribution of Public Sector Wage Premia: New Evidence Using Quantile Regression Methods, NBER Working Paper No. 4734.

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

Spokoiny, V.G. (2000). Data Driven Testing the Fit of Linear Models, preprint 472, Weierstrass Institute for Applied Analysis and Stochastics, Berlin.

Stute, W. (1997). Nonparametric Model Checks for Regression, *Annals of Statistics*, 25, 613-641.

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.

Zheng, J.X. (1998). A Consistent Nonparametric Test of Parametric Regression Models under Conditional Quantile Restrictions, *Econometric Theory*, 14, 123-138.

TABLE 1: RESULTS OF MONTE CARLO EXPERIMENTS[1]

n	Distribution of ε	τ	Probability of Rejecting Null Hypothesis: Zheng's Test	F Test	Rate-Optimal Test
			Null Hypothesis Is True		
100	Normal		0.047	0.050	0.063
	Mixture		0.048	0.050	0.056
	Extr. Val.		0.048	0.051	0.057
250	Normal		0.048	0.052	0.055
	Mixture		0.050	0.048	0.049
	Extr. Val.		0.056	0.050	0.051
			Null Hypothesis Is False		
100	Normal	1.0	0.100	0.207	0.794
	Mixture	1.0	0.120	0.276	0.674
	Extr. Val.	1.0	0.090	0.193	0.530
	Normal	0.25	0.162	0.131	0.608
	Mixture	0.25	0.204	0.117	0.552
	Extr. Val	0.25	0.240	0.162	0.534
250	Normal	1.0	0.776	0.460	0.980
	Mixture	1.0	0.600	0.462	0.958
	Extr. Val.	1.0	0.490	0.340	0.796
	Normal	0.25	0.516	0.172	0.868
	Mixture	0.25	0.300	0.144	0.796
	Extr. Val	0.25	0.446	0.130	0.802

[1] The differences between empirical and nominal rejection probabilities under H_0 are not significant at the 0.05 level. Under H_1, the differences between the rejection probabilities of the rate-optimal and Zheng's test are significant at the 0.01 level.

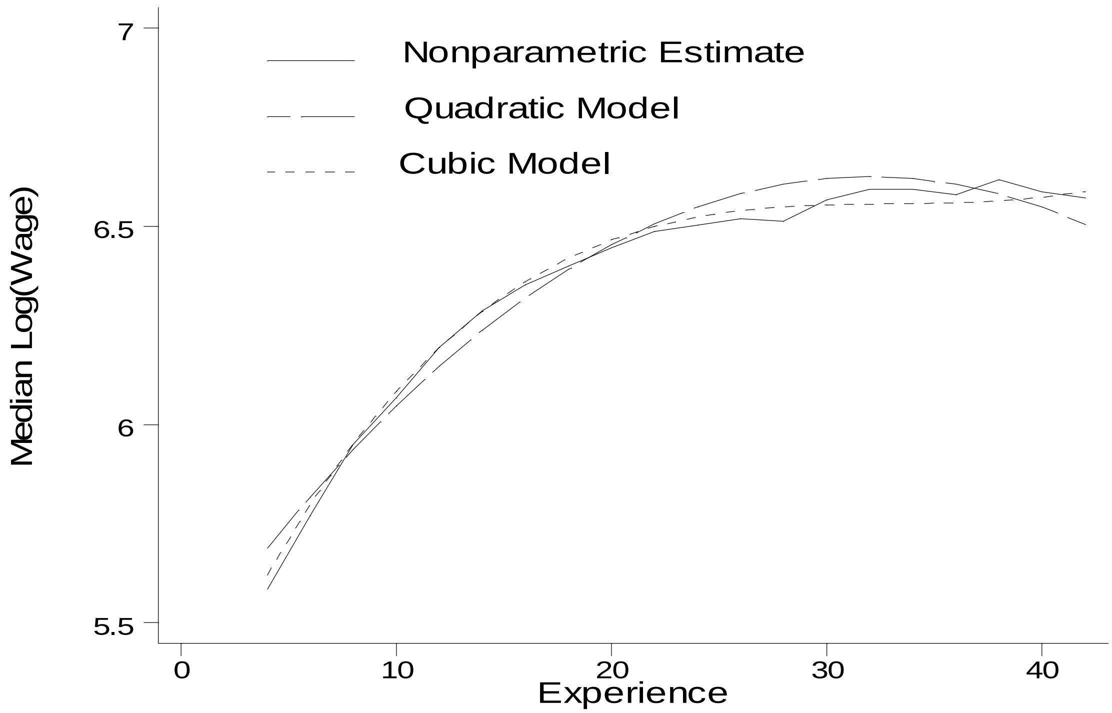

Figure 1: Estimates of Median Log(Wages)

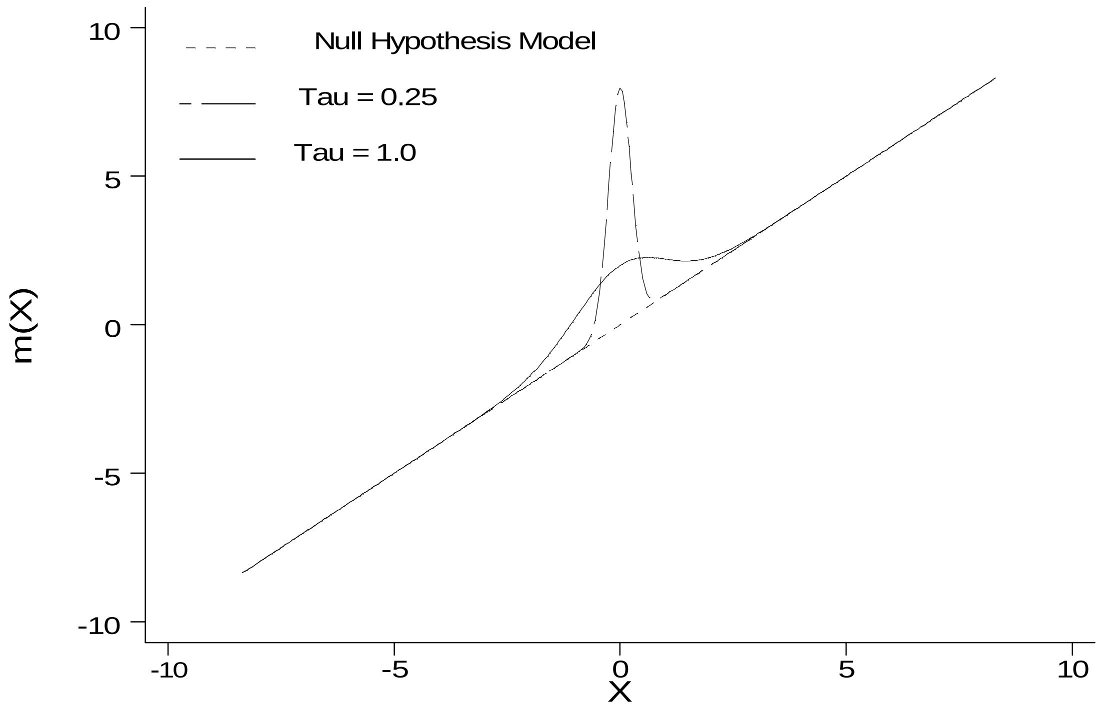

Figure 2: Null and Alternative Models