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

\[ y_i = m(x_i) + \epsilon_i; \quad i = 1, 2, 3, \ldots, \]

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 \( \{\epsilon_i\} \) is a sequence of unobserved, independently but not necessarily identically distributed random variables whose medians are zero. The distributions of the \( \epsilon_i \)'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 \( \beta \) such that \( m(x_i) = x_i'\beta \) for all \( i \). The test can be extended to models in which quantile(\( \epsilon_i \)) = 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 \( \beta \) 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 \( \epsilon_i \) 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),

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

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

for some $\beta \in \mathbb{R}^d$ and function $g$, where $n$ is the sample size, $\rho_n$ 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 $\rho_n$. 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 $i = 1, \ldots, n$. 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 $m \in S$ and the null-hypothesis set. (See Section 4.4 for an example of $\rho$.) Let $\alpha_1$ be the specified power. Then $r_n$ is the smallest $r$ such that

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

An optimal test minimizes $r_n$ for the given $\alpha_1$ 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 $\epsilon_i$’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 $B_n = \{m \in S: \rho(m) \geq C r_n\}$ 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

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

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

$$n^{-1/\sqrt{\log \log n}}$$ 

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 $\epsilon_i$’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 \( \varepsilon_i \)'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 \( \varepsilon_i \). 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 \( \varepsilon_i \)'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 \( \rho_n \geq Cn^{-1/2} \frac{\log \log n}{\sqrt{\log \log n}} \) 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 \( \rho_n \rightarrow 0 \) 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 \( \beta \)'s are scalar coefficients, and \( \gamma \) 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

\[
(2.1) \quad \log W = \beta_0 + \beta_1 X + \beta_2 X^2 + U,
\]

where \( \text{median}(U \mid X = x) = 0 \) almost surely. The 1993 CPS contains 1656 observations of workers with the specified characteristics. The \( \beta \)'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 \( \beta_j \) (\( j = 0, 1, 2 \)). 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

\[
\log W = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + U.
\]

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 $Y_i = X_i' \beta + \epsilon_i$ and $P(\epsilon_i \leq 0) = 0.5$ for each $i = 1, 2, \ldots$ and some $\beta \in \mathbb{R}^d$. Let $b_n$ denote the least absolute deviations (LAD) estimator of $\beta$. 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 $\beta$ is undefined. However, it follows from Proposition 1 in the Appendix that $b_n = \beta^* + O_p(n^{-1/2})$, where $\beta^*$ solves

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

Define $\beta_0 = \beta$ if $H_0$ is true, $\beta_0 = \beta^*$ if $H_0$ is false, and $\xi_i = I(Y_i - X_i\beta_0 \leq 0) - 1/2$, where $I$ is the indicator function. If $H_0$ is false, then $\beta_0$ 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 $\beta_0$ is independent of both $n$ and the design.

Under $H_0$, the $\xi_i$’s are Bernoulli random variables with $E(\xi_i) = 0$. If $H_0$ is false, then $E(\xi_i) = P(\epsilon_i \leq X_i'\beta_0 - m(X_i)) \leq -1/2 \neq 0$ for at least one $i$. Thus, a test of $H_0$ is equivalent to a test of $H_0'$: $E(\xi_i) = 0$ for all $i$. If $\beta_0$ were known, such a test could be based on the distance from 0 of a nonparametric estimator of the vector $[E(\xi_1), \ldots, E(\xi_n)]'$. We obtain a feasible test by replacing $\beta_0$ with $b_n$. Define $\hat{\xi}_i = I(Y_i - X_i b_n \leq 0) - 1/2$. Our test is based on $\{\hat{\xi}_i : i = 1, \ldots, n\}$.
To obtain the test statistic, suppose for the moment that $\beta_0$ and, therefore, the $\xi_i$’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 $a_{ij,h} = \sum_{k=1}^{n} w_{ki,h} w_{kji,h}$. Define

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

Observe that $\sum_{j=1}^{n} w_{ij,h} \xi_j$ is a kernel nonparametric estimator of $E(\xi_i)$. Therefore, $S_h^*$ is the $\ell_2$ distance from zero of the kernel estimator of $[E(\xi_1), ..., E(\xi_n)]’$. If the $\xi_i$’s were observable, then a test of $H_0$ could be based on the standardized version of $S_h^*$. Because $E(\xi_i) = 0$ under $H_0$, $\xi_i^2 = 1/4$, and $\xi_i$ is independent of $\xi_j$ if $i \neq j$, the standardized $S_h^*$ is

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

where

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

and

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

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 $\xi_i$ with $\hat{\xi}_i$ in (3.2)-(3.5).

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

$$S_h = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij,h} \hat{\xi}_i \hat{\xi}_j .$$ (3.6)
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 $\mathcal{H} = \{h = h_{\text{min}} 2^{k/(2(d-1))} : h \leq h_{\text{max}}, k = 0,1,2,...\}$, where $h_{\text{max}}$ and $h_{\text{min}}$ 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) $\alpha$ level if $T$ exceeds the critical value that is described in Section 3.2.

### 3.2 Obtaining the Critical Value

The exact $\alpha$-level critical value for $T$ is the $1 - \alpha$ quantile of the finite-sample distribution of $T$. This critical value cannot be evaluated in applications because the finite-sample distribution of the $\xi_i$’s is unknown. However, it turns out that the distribution of $T$ under $H_0$ depends only weakly on $\beta$ and the distribution of the $\epsilon_i$’s in (1.1), and this dependence vanishes as the sample size increases. This is true even if the $\epsilon_i$’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 $\alpha$-level critical value can be obtained as the $1 - \alpha$ quantile of the distribution of $T$ that is induced by the model $Y_i^* = X_i'b_n + \epsilon_i^*$, where $\epsilon_i^*$ is sampled from a convenient distribution. We recommend sampling $\epsilon_i^*$ 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 $\epsilon_i$’s are identically distributed, and (as has just been explained) valid in large samples even if the $\epsilon_i$’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 $\epsilon_i^*$.

The EDR does not satisfy Assumption 1 of Section 4, which requires $\epsilon_i^*$ 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 $\beta$ and the distribution of the $\epsilon_i$’s. Under $H_0$, the random variables $\xi_i = I(Y_i - X_i'\beta \leq 0) - 1/2$
are independently and identically Bernoulli distributed with \( P(\xi_i = \pm 1/2) = 1/2 \). This is true regardless of the distributions of the \( \varepsilon_i \)'s or whether they are identically distributed. Therefore, under \( H_0 \) the distribution of \( T_h^* \) in (3.3) does not depend on \( \beta \) or the distribution of the \( \varepsilon_i \)'s and can be approximated with any precision by Monte Carlo simulation. The statistic \( T_h \) is obtained from \( T_h^* \) by replacing the \( \xi_i \)'s with the empirical analogs \( \hat{\xi}_i = I(Y_i - X_ib_n \leq 0) - 1/2 \). 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 \( \beta \) the distributions of the \( \varepsilon_i \)'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 \( i = 1, \ldots, n \), generate \( Y_i^* = X_i\hat{b}_n + \varepsilon_i^* \), where \( \varepsilon_i^* \) is sampled randomly from the residuals, \( Y_i - X_i\hat{b}_n \), of the estimated null-hypothesis model.

2. Use the data set \{\( Y_i^*, X_i: \ i = 1, \ldots, n \)\} to estimate \( \beta \). Denote the resulting estimate by \( \hat{\beta}_n \). Compute the statistic \( \hat{T} \) that is obtained by replacing \( \xi_i (i = 1, \ldots, n) \) with \( I(Y_i^*-X_i\hat{b}_n \leq 0) - 1/2 \) in the formula for \( T \).

3. Estimate \( t_{\alpha} \) by the \( 1 - \alpha \) 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 H"older 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 \( \varepsilon_i \)'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 H"older 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

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

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 \( F_i(u) \equiv P(\varepsilon_i \leq u) \).

**Assumption 1** (Observations): The observations \( \{Y_i : i = 1, 2, \ldots\} \) 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 \( |f_i'(u)| \leq a^2C_F \) for all \( i, j = 1, \ldots, n \) and \( u \). Moreover, \( f_i(0) - f_j(0) \leq C_3 \|X_i - X_j\| \) for some constant \( C_3 < \infty \) and all \( i, j = 1, \ldots, n \).

**Assumption 2** (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 \( \|v\| \).

**Assumption 3** (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 \( C_H > 0 \).

**Assumption 4** (Design): (i) The design points \( \{X_i : i = 1, \ldots, n\} \) 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 \( i = 1, \ldots, n \), \( C_X1nh^{d-1} \leq M_h(X_i) \leq C_X2nh^{d-1} \). (iii) There are finite constants \( C_X \) and \( C_{XX} \) such that \( \|X_i\| \leq C_X \) for all \( i \) and \( \left\| n^{-1} \sum_{i=1}^n f_i(0)X_iX_i \right\|_{\infty} \leq C_{XX} \).

(iv) \( \inf_{b \in [-\beta_0, \beta_0]} n^{-1} \sum_{i=1}^n |F_i[X_i'b - m(X_i)] - F_i[X_i'\beta_0 - m(X_i)]| > C\delta \) 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 \( 2, \ldots, d \) 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 \( \delta \leq \|b-\beta_0\| \leq \delta_1 \) with sufficiently small \( \delta \) 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 \( P(Y_i - X_i'\beta \leq 0) = 1/2 \) for all \( i \) and some \( \beta \in \mathbb{R}^d \). Let \( t_\alpha \) be the \( \alpha \)-level critical value that is induced by the model \( Y_i^* = X_i'b_\delta + \epsilon_i^* \) described in Section 3.2. The main result on the behavior of \( T \) under \( H_0 \) is that \( t_\alpha \) is an asymptotically correct \( \alpha \)-level critical value. Moreover, \( t_\alpha \) 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 \( \beta \), the design, or the distributions of the \( \epsilon_i \)'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 \( \{\omega_{\alpha n}\} \) such that \( \omega_{\alpha 1} \to 0 \) monotonically as \( n \to \infty \) and
\[
P(T > t_\alpha - \alpha) \leq \omega_{\alpha n}
\]
for all linear models (1.1) that satisfy Assumptions 1-4. The sequence \( \{\omega_{\alpha n}\} \) 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 \( \rho_n \geq C n^{-1/2} \sqrt{\log \log n} \) for some constant \( C > 0 \). Normalize \( g \) so that
\[
\|g\|^2 = \frac{1}{n} \sum_{i=1}^{n} g(X_i)^2 \geq 1.
\]
This normalization depends on the design and insures that there are “enough” design points within the support of \( g \). Let \( X' \) be the \( d \times n \) matrix whose \( i \)'th column is \( X_i \), \( F \) be the \( n \times n \) diagonal matrix whose \( (i,i) \) element is \( f_i(0) \), and \( G \) be the \( n \times 1 \) vector whose \( i \)'th component is
\( g(X_i) \). Let \( I_n \) be the \( n \times n \) identity matrix. Define the \( n \times n \) matrix \( \Pi = I_n - X'(X'X)^{-1}X' \).

If the \( \epsilon_i \)'s are \( iid \), then \( \Pi \) 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 \( \delta > 0 \),

\[
(4.2) \quad v^2 = n^{-1} \| \Pi G \|^2 \geq \delta.
\]

If the \( \epsilon_i \)'s are \( iid \), then (4.2) states \( G \) 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,

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

converges to 0 at the rate of \( \rho_n \) 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 \( \rho_n \geq C n^{-1/2} \sqrt{\log \log n} \) and \( g \) satisfying (4.1)-(4.2). There exists a constant \( C^* < \infty \), depending on \( g \), \( \delta \), 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

\[
P(T > t_{\alpha}) \geq 1 - \omega_{n2}
\]

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

This result holds uniformly over \( \beta \), the design \( \{ X_i \} \), and the possibly heterogeneous distributions of the \( \epsilon_i \)'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

\[
\rho_t(m) = \inf_{b \in \mathbb{R}^d} \left[ n^{-1} \sum_{i=1}^{n} \left| m(X_i) - X_i'b \right|^2 \right]^{1/2}.
\]
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 \( \beta_0 \) 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 \( (H_1, \ldots, H_n) \). In what follows, we will be interested in sets of functions \( (H_1, \ldots, H_n) \) such that

\[
\rho_1(m) \geq C_a \left( n^{-1} \sqrt{\log \log n} \right)^{2s/(4s+d-1)}
\]

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 \( \rho_1 \) and \( \rho_2 \) 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

\[
D^j H_i(x) = \frac{\partial^{j_2} H_i(x)}{\partial x_2^{j_2}} \ldots \frac{\partial^{j_d} H_i(x)}{\partial x_d^{j_d}}
\]

whenever the derivative exists. Define the Hölder norm

\[
\|H_i\|_{H,s} = \sup_{x:|x| \leq C_s} |D^j H_i(x)|.
\]

The smoothness classes that we consider consist of functions \( (H_1, \ldots, H_n) \in S_n(s) \equiv \{H_1, \ldots, H_n : \|H_i\|_{H,s} \leq C_F \text{ for all } i = 1, \ldots, n\} \) 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

\( B_n = \{H_1, \ldots, H_n \in S_n(s) : \rho_2(H) \geq C_a \left( n^{-1} \sqrt{\log \log n} \right)^{2s/(4s+d-1)}\} \)

for some \( s \geq \max\{2, (d-1)/4\} \) and all sufficiently large \( C_a \). If \( (H_1, \ldots, H_n) \in B_n \), 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 \( C_b < \infty \).

**Theorem 3:** Let Assumptions 1-4 hold. Then for \( 0 < \alpha < 1 \) and \( B_n \) as defined in (4.3), there is a constant \( C_a^* \) and a sequence of constants \( \{\omega_n\} \) such that \( \omega_n \to 0 \) monotonically as \( n \to \infty \) and

\[
\inf_{m:(H_1, \ldots, H_n) \in B_n} P(T > t_{\alpha}) \geq 1 - \omega_n
\]
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 $\varepsilon_i$’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 $\varepsilon_i$’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

\[(5.1) \quad Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i,\]

where $\beta_0$ and $\beta_1$ 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

\[(5.2) \quad Y_i = X_i + \tau_n^4 \phi(X_i / \tau_n) + \varepsilon_i,\]

where $\varepsilon_i \sim N(0,1)$, $\phi$ 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

\[ Y_i = \beta_0 + \beta_1 X_i + \epsilon_i; \quad i = 1, 2, ..., n \]  

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 \( \epsilon_i \)’s were sampled independently from three distributions, depending on the experiment. These are \( N(0,4) \), a variance mixture of normals in which \( \epsilon_i \) 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 \( \epsilon \). Specifically, \( 0.77 \leq 1 - \text{Var}(\epsilon)/\text{Var}(Y) \leq 0.79 \).

The alternative models have the form

\[ Y_i = 1 + X_i + (4/\tau)\phi(X_i/\tau) + \epsilon_i, \]

where the \( \epsilon_i \)’s are sampled from one of the three distributions just described and \( \tau = 1 \) 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 \( \tau \). 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 \( \beta_0 \) and \( \beta_1 \) were estimated by least absolute deviations (LAD). The kernel used for the adaptive, rate-optimal test and Zheng’s (1998) test is

\[ K(u) = (15/16)(1-u^2)^2I(|u| \leq 1). \]

At the suggestion of a referee, we carried out \( F \) tests of the hypothesis \( H_0: \beta_2 = 0 \) in the augmented model

\[ Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i. \]

Rejection of \( H_0 \) 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 $\omega^j h_{\min}$ ($j = 0, 1, 2, ..., N - 1$), where $N$ is the number of grid points and $\omega = (h_{\max}/h_{\min})^{1/(N-1)}$. 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 $N$ according to the rule of thumb $N = \log n$. This rule is consistent with the theory of the test, which requires $N = O(\log n)$. Motivated by the rule of thumb, we did experiments with values of $N$ in the range 4-10. The results varied little over this range, so we report only the results for $N = 4$.

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 $H_0$. 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 $N = 7$, 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 $\beta_j = 0$ 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 $\epsilon_i$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 $k = 1, 2, \ldots$. 
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

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

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

and

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

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 $\delta_n \leq (n^{-1} \log n)^{1/2}$. Then there exists a random set $A_{n1}$ satisfying $P(A_{n1}) \geq 1 - a_{n1}$ with $a_{n1} = o_n(1)$ and such that on $A_{n1}$, $b_n - \beta_0 \leq \delta_n$ and

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

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 $C_1 = d C_X / 2$. The vector $b_n$ satisfies

$$\left\| \sum_{i=1}^{n} X_i [I(Y_i - X_i^\prime \beta_0 \leq 0) - 1/2] \right\| \leq C_1.$$ 

Proof: See Koenker and Bassett (1978).

Lemma 2: Let $\{X_i : i = 1, \ldots, n\}$ 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

$$\sum_{i=1}^{n} c_i^2 p_i (1 - p_i) \leq V^2.$$ 

Given any real $z$, define

$$G^2 = \max_{1 \leq i \leq n} \exp\{zc_i/(2V)\}.$$
If $G^2 \leq 2$, then

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

for all $x \geq 0$ and $p \in [0, 1]$. Therefore,
\[-\lambda zV - \lambda \sum_{i=1}^{n} c_i p_i + \sum_{i=1}^{n} \log(1 - p_i + p_i e^{\lambda c_i}) \leq -\lambda zV + \frac{\lambda^2}{8} \sum_{i=1}^{n} c_i^2 \]

\[= -\lambda zV + \frac{\lambda^2}{2} V^2.\]

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 

\[\|x\|_\infty = \max_{1 \leq j \leq d} |x_j| .\]

**Lemma 3:** Let \(\{\kappa_i : i = 1, ..., n\}\) be independent Bernoulli random variables with parameters \(\{p_i\}\), and let \(\{c_i : i = 1, ..., n\}\) be constant vectors in \(\mathbb{R}^d\). Let \(V\) be a constant such that 

\[\sum_{i=1}^{n} \|c_i\|_\infty^2 p_i (1 - p_i) \leq dV^2.\]

Given any real \(z \geq 0\), define 

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

If \(G^2 \leq 2\), then 

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

Moreover, if \(\sum_{i=1}^{n} \|c_i\|_\infty^2 / 4 \leq V^2\), then for all \(z \geq 0\)

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

**Proof:** 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 \(\xi(\beta) = I(Y_i - X_i' \beta \leq 0) - F_i(X_i' \beta)\) and 

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

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

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

\[E \|\zeta(\beta_1) - \zeta(\beta_2)\| \leq C_X^2 C_{XX} \|\beta_1 - \beta_2\| + 0.5C_X^2 C_{F} a^2 \|\beta_1 - \beta_2\|^2 ,\]

and, for every \(z \geq 0\),

\[P(\|\zeta(\beta)\| > zC_X / 2) \leq 2 \exp(-z^2 / 2).\]

20
Proof: The first two statements obviously follow from independence of the Bernoulli random variables $\xi_i$. It is also straightforward to check that

$$E \left| \xi_i(\beta_1) - \xi_i(\beta_2) \right|^2 \leq \left| \tilde{f}_i(X'_1\beta_1) - \tilde{f}_i(X'_1\beta_2) \right| [1 - \left| \tilde{f}_i(X'_1\beta_1) - \tilde{f}_i(X'_1\beta_2) \right|]$$

$$\leq \left| \tilde{f}_i(X'_1\beta_1) - \tilde{f}_i(X'_1\beta_2) \right|. $$

A Taylor series expansion and Assumption 1 yield

$$\left| \tilde{f}_i(X'_1\beta_1) - \tilde{f}_i(X'_1\beta_2) \right| \leq \left\| \tilde{f}_i(X'_1\beta_1)(\beta_1 - \beta_2) \right\| + 0.5C_F a^2 \|X'_1(\beta_1 - \beta_2)\|^2$$

$$\leq \tilde{f}_i(X'_1\beta_1) \|\beta_1 - \beta_2\| + 0.5C_X C_F a^2 \|\beta_1 - \beta_2\|^2. $$

Therefore,

$$E \left\| \xi(\beta_1) - \xi(\beta_2) \right\|^2 = trE[\xi(\beta_1) - \xi(\beta_2)][\xi(\beta_1) - \xi(\beta_2)]'$$

$$= n^{-1}tr \sum_{i=1}^n X_iX_i' E \left| \xi_i(\beta_1) - \xi_i(\beta_2) \right|^2$$

$$\leq n^{-1}tr \sum_{i=1}^n X_iX_i' \left| \tilde{f}_i(X'_1\beta_1) - \tilde{f}_i(X'_1\beta_2) \right|$$

$$\leq n^{-1}C_X^2 \|\beta_1 - \beta_2\| \sum_{i=1}^n \tilde{f}_i(X'_1\beta_1) + 0.5C_X^4 C_F a^2 \|\beta_1 - \beta_2\|^2$$

$$\leq C_X^2 C_F n \|\beta_1 - \beta_2\| + 0.5C_X^4 C_F a^2 \|\beta_1 - \beta_2\|^2. $$

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

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

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

$$P \left[ \sup_{\beta \in \mathbb{R}^d} \|\xi(\beta) - \xi(\tilde{\beta})\| \geq C_{\xi 1} n^{-\alpha + 1/2} \right] \leq 2 \exp \left( -C_{\xi 2} n^{1-\alpha} / 4 \right).$$

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

$$\|\xi(\beta) - \xi(\tilde{\beta})\| = n^{-1/2} \left\| \sum_{i=1}^n X_i [\xi_i(\beta) - \xi_i(\tilde{\beta})] \right\|$$
\[ \leq n^{-1/2} \sum_{i=1}^{n} I(\|Y_i - X'_i \beta \| \leq X'(\tilde{\beta} - \beta)) \]
\[ + n^{-1/2} \left\| \sum_{i=1}^{n} X_i [\tilde{f}_i(X'_i \beta) - \tilde{f}_i(X'_i \tilde{\beta})] \right\| . \]

Since \( |X'_i(\tilde{\beta} - \beta)| \leq C_X n^{-\alpha} \), for some \(|\theta| \leq 1\) we have
\[ \left\| \sum_{i=1}^{n} X_i [\tilde{f}_i(X'_i \beta) - \tilde{f}_i(X'_i \tilde{\beta})] \right\| = \left\| \sum_{i=1}^{n} X_i \tilde{f}_i[X'_i \beta + \theta(\tilde{\beta} - \beta)]X'_i(\tilde{\beta} - \beta) \right\| \]
\[ \leq C_X^2 C_{XX} n \|\tilde{\beta} - \beta\| \]
\[ \leq C_X^2 C_{XX} n^{1-\alpha}. \]

Therefore,
\[
(8.4) \quad \left\| \zeta(\beta) - \zeta(\tilde{\beta}) \right\| \leq n^{-1/2} C_X \sum_{i=1}^{n} \tau_i + C_X^2 C_{XX} n^{1/2-\alpha},
\]
where the \( \tau_i \equiv I(\|Y_i - X'_i \beta \| \leq C_X n^{-\alpha}) \) are Bernoulli random variables with
\[ p_i \equiv E \tau_i = P(|Y_i - X'_i \beta| \leq C_X n^{-\alpha}) = \tilde{f}_i(X'_i \beta + C_X n^{-\alpha}) - \tilde{f}_i(X'_i \beta - C_X n^{-\alpha}). \]

As in the proof of Lemma 4, one bounds
\[ \sum_{i=1}^{n} p_i \leq 2 C_X n^{-\alpha} \sum_{i=1}^{n} \tilde{f}_i(X'_i \beta) + C_X^2 C_F a^2 n^{-2\alpha+1} \]
\[ \leq 2 C_X C_{XX} n^{-\alpha-1} + C_X^2 C_F a^2 n^{-2\alpha+1} \leq C_2 n^{-\alpha+1} \]
for some constant \( C_2 = 2 C_X C_{XX} \). Application of Lemma 2 with \( c_i = 1, \ z = V \), and
\[ V^2 = \sum_{i=1}^{n} p_i \leq C_2 n^{-\alpha} \] (so that \( G^2 = e^{1/2} < 2 \)) yields
\[ P \left[ \sum_{i=1}^{n} (\tau_i - p_i) \geq V^2 \right] \leq 2 \exp(-V^2 / 4). \]

Therefore,
\[ P \left( \sum_{i=1}^{n} \tau_i \geq 2 C_2 n^{-\alpha} \right) \leq 2 \exp(-C_2 n^{1-\alpha} / 4). \]

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 \( \|\beta - \beta_0\| \leq \delta \).

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

\[
\sup_{\beta \in \mathcal{D}_n} \|\zeta(\beta) - \zeta(\beta_0)\| \leq C_2 (\delta \log n)^{1/2}.
\]

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

\[
\log G^2 \leq C_X n^{-1/2} z / (2V) = C_X n^{-1/2} (d \log n)^{1/2} /
\]

\[
(C_{\delta})^{1/2}.
\]

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

\[
P\left[ \|\zeta(\beta) - \zeta(\beta_0)\| \geq 2d(C_{\delta} \delta \log n)^{1/2} \right] \leq 2e^{-d \log n}.
\]

Now

\[
P\left[ \sup_{\beta \in \mathcal{D}_n} \|\zeta(\beta) - \zeta(\beta_0)\| \geq 2d(C_{\delta} \delta \log n)^{1/2} + C_{\delta} n^{-\alpha + 1/2} \right]
\]

\[
\leq \sum_{\beta \in \mathcal{D}_n} P\left[ \sup_{\beta \in \mathcal{D}_n} \|\zeta(\beta) - \zeta(\beta_0)\| \geq C_{\delta} n^{-\alpha + 1/2} \right]
\]

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

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

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

Define

\[
B(\beta) = n^{-1/2} \sum_{i=1}^{n} X_i [\tilde{F}(X_i \beta) - \tilde{F}_0(X_i \beta_0)].
\]

Note that \( B(\beta) = E[\zeta(\beta) - \zeta(\beta_0)] \). The next lemma states that \( B(\beta) \) is nearly linear in a small neighborhood of \( \beta_0 \). Let \( \tilde{F}(X' \beta) \) be the vector whose components are \( \tilde{F}_i(X'_i \beta) \).
Lemma 7: For all $\beta$
\[
\left\| \hat{F}(X'\beta) - \hat{F}(X'\beta_0) - F(X'(\beta - \beta_0)) \right\| \leq 0.5n^{1/2}C_F a^2 \left\| \beta - \beta_0 \right\|^2
\]
and
\[
\left\| B(\beta) - n^{1/2}Q_n(\beta - \beta_0) \right\| \leq \left\| 0.5a^2C_F n^{-1/2} \sum_{i=1}^n X_i | X_i'(\beta - \beta_0) |^2 \right\| \leq C_3 a^2 n^{1/2} \left\| \beta - \beta_0 \right\|^2.
\]
where $C_3 = 0.5C_X C_F$.

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 \(P(A_{n3}) \geq 1 - a_{n3}\) with \(a_{n3} = o_n(1)\) such that on \(A_{n3}\), \(\left\| p_n - \beta_0 \right\| \leq \delta_n\).

Proof: Lemma 7 and Assumption 4 imply that as \(n \to \infty\), \(B(\beta) \to \infty\).

As \(n \to \infty\). By Lemmas 4 and 6, \(\zeta(\beta)\) is bounded in probability in every neighborhood of \(\beta_0\).

Moreover, (8.1) implies that \(\left\| \zeta(b_n) - B(b_n) \right\| \leq C_1 n^{-1/2}\). The lemma follows from this inequality and monotonicity arguments. See Portnoy (1991) for details. Q.E.D.

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

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

Proof: By Lemma 4, \(\Sigma \eta_n = 0\) and \(\Sigma \eta_n \eta'_n = \Sigma_n\). 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
\[
n^{-1/2} \sum_{i=1}^n X_i [I(Y_i - X_i' \beta \leq 0) - 1/2] = \zeta(\beta) + B(\beta).
\]
By Lemma 1,
\[
\left\| \zeta(b_n) + B(b_n) \right\| \leq C_1 n^{-1/2}.
\]
Let \(\delta_n\) satisfy \(n^{1/2} \delta_n \to 0\) as \(n \to \infty\). Then by Lemmas 6 and 7, on the random set \(A_{n2}\)
\[
\left\| \zeta(\beta) - \zeta(\beta_0) \right\| \leq C_2 (\delta_n \log n)^{1/2} \text{ and } \left\| B(b_n) - n^{1/2}Q_n(b_n - \beta_0) \right\| \leq C_3 a^2 \delta_n^2 n^{1/2} \text{ for all } \beta \text{ with } \left\| \beta - \beta_0 \right\| \leq \delta_n\.
\]
Similarly, by Lemma 8 \(\left\| b_n - \beta_0 \right\| \leq \delta_n\) on \(A_{n3}\). Define \(A_{n1} = A_{n2} \cap A_{n3}\). Then \(P(A_{n1}) \geq 1 - a_{n2} - a_{n3}\) and on \(A_{n1}\),

24
Thus, \[
\left\| \zeta(b_n) - n^{1/2} \varphi_n(b_n - \beta_0) \right\| \leq C_1 n^{-1/2} + C_2 (\delta_n \log n)^{1/2} + C_3 a_n^2 n^{1/2} \delta_n^2.
\]

By Assumption 4, \(n^{-1} \leq C_{FX}\), and hence
\[
\left\| n^{1/2} (b_n - \beta_0) + Q_n^{-1} \xi(b_n) \right\| \leq C_{XX} \left[ C_1 n^{-1/2} + C_2 (\delta_n \log n)^{1/2} + C_3 a_n^2 n^{1/2} \delta_n^2 \right].
\]
But \(\eta_n = -Q_n^{-1} \xi(b_0)\). If \(\delta_n\) satisfies \(\delta_n \leq (n^{-1} \log n)^{1/2}\), then on the set \(A_{n1}\), \(n^{1/2} (b_n - \beta_0) - \eta_n \leq C_0 (\delta_n \log n)^{1/2}\) 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 \(\|W_h\|_\alpha = \sup_{x \in \mathbb{R}} \|W_h x\|_\alpha\).

**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 \(h \in \mathcal{H}\),
\[
\sum_{j=1}^{n} w_{ij,h} \leq C_{w1} \quad (j = 1, \ldots, n), \quad \|W_h\|_\alpha \leq C_{w1}, \quad \sum_{i=1}^{n} a_{ii,h} \leq C_a h^{-1}, \quad \sum_{i=1}^{n} \sum_{j \neq i} a_{ij,h}^2 \geq C_{V1} h^{-1}, \quad \text{and}
\]
\[
\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij,h}^2 \leq C_{V2} h^{-1}.
\]

**Proof**: See HS. Q.E.D.

8.3 Asymptotic Expansion of the Statistics \(S_h\)

For every \(\beta \in \mathbb{R}^d\) define
\[
S_h(\beta) = \sum_{i=1}^{n} \left( \sum_{j=1}^{n} w_{ij,h} [I(Y_j - X_j' \beta \leq 0) - 1/2] \right)^2,
\]
and
\[
= \sum_{i=1}^{n} \left( \sum_{j=1}^{n} w_{ij,h} [\hat{\xi}_j(\beta) - \tilde{F}_j(X_j' \beta) - 1/2] \right)^2.
\]
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

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

Under the null hypothesis, \( \tilde{F}_j(X'_j\beta_0) = 1/2 \), so \( S_h(\beta) = \left\| W_h[\zeta(\beta) + z(\beta)] \right\|^2 \). The test statistic is based on \( S_h(b_n) \). Lemma 11 enables us to obtain an asymptotic expansion for \( S_h = S_h(b_n) \).

Define the \( n \times n \) matrices \( A_h = W_h W_h \) and \( \Pi_F = \mathcal{F}\mathcal{X}'(\mathcal{F}\mathcal{X})^{-1} \mathcal{X} \).

**Lemma 11**: Let \( \delta_n \) satisfy \( \delta_n \leq (n^{-1} \log n)^{1/2} \) and \( n^{-1/2} / \delta_n \to 0 \) as \( n \to \infty \). There exist a constant \( C_0 \) and a random set \( A_{n4} \) satisfying \( P(A_{n4}) \geq 1 - a_{n4} \) with \( a_{n4} = o_n(1) \) such that on \( A_{n4} \)

\[
\left\| S_h - \left\| W_h[(I_n - \Pi_F)\xi(\beta_0) + \tilde{F}(X'\beta_0) - 1/2] \right\|^2 \right\| \leq C_0 \delta_n^{3/2} h^{-1} \log n
\]

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 \( \beta \) such that \( \| \beta - \beta_0 \| \leq \delta_n \), Assumption 1 yields

\[
(8.5) \quad \left\| \tilde{F}_i(X'_i\beta) - \tilde{F}_i(X'_i\beta_0) \right\| \leq C_5 a \delta_n,
\]

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

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

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

As in Lemma 6, there exists a random set \( A_{n5} \) satisfying \( P(A_{n5}) \geq 1 - a_{n5} \) with \( a_{n5} = o_n(1) \) such that on \( A_{n5} \)

\[
\left( \eta_{i,h}(\beta) - \eta_{i,h}(\beta_0) \right) \leq C_6 (a_{ii,h} \delta_n \log n)^{1/2},
\]

and

\[
\left( \eta_{i,h}(\beta_0) \right) \leq C_7 (a_{ii,h} \log n)^{1/2}
\]
for some constants $C_6$ and $C_7$, all $h \in \mathcal{H}$, all $\beta$ such that $\|\beta - \beta_0\| \leq \delta_n$, and all $i = 1, \ldots, n$. This and (8.5) imply that on $A_n$

$$\left\| W_h [\xi(\beta) - \xi(\beta_0)] \right\|^2 \leq C_6^2 a \delta_n \text{tr}(A_h) \log n,$$

$$\left\| W_h \xi(\beta_0) \right\|^2 \leq C_7^2 \text{tr}(A_h) \log n,$$

and

$$\left\| W_h [\xi(\beta_0) + z(\beta)] \right\| \leq C_7 \left[ \text{tr}(A_h) \log n \right]^{1/2} + C_5 an^{1/2} \delta_n.$$

Now by the inequality $\|x^2 - y^3\| \leq \|x - y\| (\|x - y\| + 2\|x\|)$ and Lemma 10, the following holds on $A_n$ for all $\beta$ satisfying $\|\beta - \beta_0\| \leq \delta_n$:

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

(8.7) \[ \leq C_8 (a \delta_n)^{1/2} h^{-1} \log n. \]

By Proposition 1, $\|b_n - \beta_0\| \leq \delta_n$ on the set $A_{n_1}$ satisfying $P(A_{n_1}) = 1 - a_{n_1}$. Therefore, on $A_{n_4} = A_{n_1} \cap A_{n_5}$, inequality (8.7) holds when $\beta$ is replaced by $b_n$. Also, Proposition 1 and Assumption 1 imply that on the set $A_{n_1}$,

$$\left\| \widetilde{F}(X b_n) - \widetilde{F}(X' \beta_0) \right\| \leq C'_0 (\delta_n \log n)^{1/2},$$

where $C'_0 = C_0 C_X C_F$. Therefore, by Lemma 10,

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

on $A_{n_4}$, where $C_9 = C'_0 C_{w_1}$. 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

$$\left\| W_h \xi(\beta_0) - W_h \Pi_F \xi(\beta_0) \right\|^2.$$

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 $n \to \infty$.

**Lemma 12:** Let $h_{\max} \to 0$ as $n \to \infty$. Then under the null hypothesis there exists a random set $A_{n_6}$ satisfying $P(A_{n_6}) \geq 1 - a_{n_6}$ with $a_{n_6} = o_n(1)$ such that on $A_{n_6}$

$$\sup_{h \in \mathcal{H}} h^{1/2} \left\| S_h - \left\| W_h \xi(\beta_0) \right\|^2 \right\| \leq s_{n_1} = o_n(1).$$
Proof: By Lemma 11, it suffices to show that there exists a random set \( A_{n,7} \) satisfying

\[
P(A_{n,7}) \geq 1 - a_{n,7} \text{ with } a_{n,7} = o_n(1) \text{ such that on } A_{n,7}
\]

\[
\tau_n = \sup_{h \in \mathcal{H}} h^{1/2} \left\| W_h (I_n - \Pi_F) \xi(\beta_0) \right\|^2 - \left\| W_h \xi(\beta_0) \right\|^2 \leq s_{n,2} = o_n(1).
\]

This would follow from

\[
\sum_{h \in \mathcal{H}} h^{1/2} E\left[ \left\| W_h (I_n - \Pi_F) \xi(\beta_0) \right\|^2 - \left\| W_h \xi(\beta_0) \right\|^2 \right] = o_p(1)
\]

and

\[
\sum_{h \in \mathcal{H}} h^{1/2} \text{Var} \left[ \left\| W_h (I_n - \Pi_F) \xi(\beta_0) \right\|^2 - \left\| W_h \xi(\beta_0) \right\|^2 \right] = o_p(1).
\]

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

\[
E\left[ \left\| W_h \xi(\beta_0) \right\|^2 - \left\| W_h (I_n - \Pi_F) \xi(\beta_0) \right\|^2 \right] = 2Etr[W_h \Pi_F \xi(\beta_0) \xi(\beta_0)' W_h'] - Etr[W_h \Pi_F \xi(\beta_0) \xi(\beta_0)' \Pi_F W_h']
\]

\[
= (1/4)tr(W_h \Pi_F W_h')
\]

\[
\leq \left\| W_h \right\|^2 tr(\Pi_F) \leq C_{w^1} d / 4.
\]

Similarly

\[
\text{Var} \left[ \left\| W_h \xi(\beta_0) \right\|^2 - \left\| W_h (I_n - \Pi_F) \xi(\beta_0) \right\|^2 \right] = \text{Var}[\xi(\beta_0)'(\Pi_F A_h + A_h \Pi_F - \Pi_F A_h \Pi_F) \xi(\beta_0)]
\]

\[
\leq (1/2)tr(\Pi_F A_h + A_h \Pi_F - \Pi_F A_h \Pi_F)^2 \leq C
\]

where \( C \) is a constant that depends only on \( C_{w^1} \) and \( d \). Since \( \mathcal{H} \) is a geometric grid,

\[
\sum_{h \in \mathcal{H}} h^{1/2} \leq C_{h_1} d^{1/2} \to 0.
\]

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 \( A_{n,6} \),

\[
\sup_{h \in \mathcal{H}} |T_h^* - T_{h,0}| \leq s_{n,3} = o_n(1),
\]

where

\[
(8.8)
\]
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 \( \| W_h \xi(\beta_0) \|^2 - (1/4) \sum_{i=1}^{n} a_{ii,h}^2 \), 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 \( \epsilon_i \)'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 \( I_n/4 \).

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 \( \Delta \) be the vector in \( \mathbb{R}^n \) with elements \( \Delta_i \). Define

\[
(V_h^*)^2 = (1/8) \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij,h}^2 = \text{tr}(A_h^2)/8.
\]

Proposition 2: Suppose there is a sequence \( \{r_n\} \) such that \( r_n \to \infty \) as \( n \to \infty \) and

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

for some \( h \in \mathcal{H} \). Then

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

Proof: It suffices to show that for a given \( h \in \mathcal{H} \), \( P(T_h < t_\alpha) \leq s_{n4} = o_n(1) \) as \( n \to \infty \). The asymptotic expansion from Lemma 11 reduces this condition to
\[
P \left[ \frac{\| W_h (I_n - \Pi_F) \hat{\xi} (\beta_0) + W_h \Delta \|^2}{V_h^*} - \frac{\text{tr}(A_h)}{4} < t_\alpha \right] \leq s_{n5} = o_n (1) .
\]

Now
\[
\| W_h (I_n - \Pi_F) \hat{\xi} (\beta_0) + W_h \Delta \|^2 - \frac{\text{tr}(A_h)}{4}
\]
\[
= \| W_h \Delta \|^2 + \| W_h (I_n - \Pi_F) \hat{\xi} (\beta_0) \|^2 - \frac{\text{tr}(A_h)}{4} + 2 \Delta^* W_h^* W_h (I_n - \Pi_F) \hat{\xi} (\beta_0) .
\]

Note that under the alternative model, the elements \( \hat{\xi} (\beta_0) \) of the vector \( \hat{\xi} (\beta_0) \) are independent Bernoulli random variables with zero means and parameters \( \bar{F}(X'_\beta) = \Delta_i + 1/2 \). By Lemma 2,
\[
P \left[ \frac{\Delta^* W_h^* W_h (I_n - \Pi_F)}{\| \Delta^* W_h^* W_h (I_n - \Pi_F) \|} > r_n / 2 \right] \leq e^{-0.5(r_n / 2)^2} \to 0
\]
as \( n \to \infty \). Moreover, because the elements of \( \Delta \) satisfy \( | \Delta_i | \leq 1/2 \) and \( I_n - \Pi_F \) is a projection operator in \( \mathbb{R}^n \), it follows that
\[
\| \Delta^* W_h^* W_h (I_n - \Pi_F) \|^2 \leq (1/4) \text{tr}(W_h^* W_h)^2 = 2(V_h^*)^2 .
\]

Therefore, there exists a random set \( A_{n8} \) satisfying \( P(A_{n8}) \geq 1 - a_{n8} \) with \( a_{n8} = o_n (1) \) such that on \( A_{n8} \),
\[
2 \Delta^* W_h^* W_h (I_n - \Pi_F) \hat{\xi} (\beta_0) \leq (r_n / \sqrt{2}) V_h^* .
\]

As in the proof of Lemma 12, one can show that
\[
E \left[ \| W_h (I_n - \Pi_F) \hat{\xi} (\beta_0) \|^2 - E \| W_h \hat{\xi} (\beta_0) \|^2 \right]^2 \leq (V_h^* + s_{n6})^2 ,
\]
where \( s_{n6} = o_n (1) \). This implies by the Cauchy-Schwartz inequality that
\[
P \left[ \| W_h (I_n - \Pi_F) \hat{\xi} (\beta_0) \|^2 - E \| W_h \hat{\xi} (\beta_0) \|^2 > (r_n / 4) V_h^* \right] \leq \frac{(V_h^* + s_{n6})^2}{(r_n V_h^*)^2} \to 0
\]
as \( n \to \infty \). Thus, there exists a random set \( A_{n9} \) satisfying \( P(A_{n9}) \geq 1 - a_{n9} \) with \( a_{n9} = o_n (1) \) such that on \( A_{n9} \),
\[
(V_h^*)^{-1} \| W_h (I - \Pi_F) \hat{\xi} (\beta_0) \|^2 - E \| W_h \hat{\xi} (\beta_0) \|^2 \leq r_n / 4 .
\]

Since \( E \hat{\xi} (\beta_0)^2 = 1/4 - \Delta_i^2 \), it follows that
\[
E \| W_h \hat{\xi} (\beta_0) \|^2 = E \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij,h} \hat{\xi}_i (\beta_0) \hat{\xi}_j (\beta_0) = \sum_{i=1}^{n} a_{ii,h} (1/4 - \Delta_i^2)
\]
so that
\[ E \|W_h \xi(\beta_0)\|^2 - (1/4) tr(A_h) = - \sum_{i=1}^{n} a_{ii,h} \Delta_i^2. \]

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 \( A_{h8} \cap A_{h9} \)

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

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

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

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

Proof of Theorem 2: Define \( \Delta \) as in Proposition 2. Set \( \rho_n = C n^{-1/2} \sqrt{\log \log n} \) for some finite \( C > 0 \). The definition of \( \beta_0 \) (eq. 3.1) implies that

\[ \sum_{i=1}^{n} X_i \left[ \tilde{f}_i(X_i \beta_0) - 1/2 \right] = \sum_{i=1}^{n} X_i \left[ f_i(X_i \beta_0 - m(X_i)) - f_i(0) \right] = 0. \]

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 \( \Delta_i \equiv \tilde{f}_i(X_i' \beta_0) - 1/2 = f_i(X_i' \beta_0 - X_i' \beta \beta - \rho_n g(X_i)) \) 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, \( \| \Delta - \rho_n \mathcal{F} \mathcal{T} \mathcal{G} \|^2 < n \rho_n^2 s_{n9} \). As in HS, one can show that \( h_{\max} \to 0 \) and continuity of \( f_j(0) \) and \( g \) imply that \( \|W_h \mathcal{F} \mathcal{T} \mathcal{G}\|/\|\mathcal{F} \mathcal{T} \mathcal{G}\| \to 1 \) as \( n \to \infty \). This result and (4.2) imply that for sufficiently large \( n \),
\[ \|\Delta\|^2 \geq 0.5 \|\mathcal{F}I\|_2^2 \geq C \delta \log \log n, \]

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 \( h = h_{\text{max}} \) 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 \( \Delta \in S(s) \)

\[
\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij,h} \Delta_i^2 \leq \sum_{i=1}^{n} a_{ij,h} \Delta_i \Delta_j
\]

Moreover,

\[
(8.10) \quad \|W_h \Delta\| \geq C_{s1} \|\Delta\| - C_{s2} n^{1/2} h^s
\]

for constants \( C_{s1} \) and \( C_{s2} \) that depend only on the design \( \{X_i : i = 1, \ldots, n\} \). 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 \( h = (n/t_n)^{-2/(4s+d-1)} \). 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, \( (V^*_h)^{-1} \|W_h \Delta\|^2 \to t_\alpha \to \infty \) as \( n \to \infty \), as is required to prove the theorem. Q.E.D.
REFERENCES


### TABLE 1: RESULTS OF MONTE CARLO EXPERIMENTS

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Zheng’s Test</th>
<th>$F$ Test</th>
<th>Rate-Optimal Test</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Null Hypothesis Is True</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 Normal</td>
<td>0.047</td>
<td>0.050</td>
<td>0.063</td>
</tr>
<tr>
<td>Mixture</td>
<td>0.048</td>
<td>0.050</td>
<td>0.056</td>
</tr>
<tr>
<td>Extr. Val.</td>
<td>0.048</td>
<td>0.051</td>
<td>0.057</td>
</tr>
<tr>
<td>250 Normal</td>
<td>0.048</td>
<td>0.052</td>
<td>0.055</td>
</tr>
<tr>
<td>Mixture</td>
<td>0.050</td>
<td>0.048</td>
<td>0.049</td>
</tr>
<tr>
<td>Extr. Val.</td>
<td>0.056</td>
<td>0.050</td>
<td>0.051</td>
</tr>
</tbody>
</table>

| **Null Hypothesis Is False** | | | |
| 100 Normal   | 1.0          | 0.100    | 0.207            |
| Mixture      | 1.0          | 0.120    | 0.276            |
| Extr. Val.   | 1.0          | 0.090    | 0.193            |
| Normal       | 0.25         | 0.162    | 0.131            |
| Mixture      | 0.25         | 0.204    | 0.117            |
| Extr. Val    | 0.25         | 0.240    | 0.162            |
| 250 Normal   | 1.0          | 0.776    | 0.460            |
| Mixture      | 1.0          | 0.600    | 0.462            |
| Extr. Val    | 1.0          | 0.490    | 0.340            |
| Normal       | 0.25         | 0.516    | 0.172            |
| Mixture      | 0.25         | 0.300    | 0.144            |
| Extr. Val    | 0.25         | 0.446    | 0.130            |

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