ECONWPA NOTE

THIS PAPER WAS WRITTEN
IN WORDPREFECT PRESTIGE
FONT BUT WE DO NOT HAVE
THAT FONT SO IT IS
PRODUCED IN COURIER AND
THE EQUATIONS ARE WELL
NOT PERFECT
IF YOU HAVE WP AND
PRESTIGE FONT, BETTER TO
GET THE ORGINAL
DOCUMENT THAN THE .PS
AND .PDF PRODUCED HERE!

BOOTSTRAP METHODS FOR MEDIAN REGRESSION MODELS

by

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

August 1996

ABSTRACT

The least-absolute-deviations (LAD) estimator for a median-regression model does not satisfy the standard conditions for obtaining asymptotic refinements through use of the bootstrap because the LAD objective function is not smooth. This paper overcomes this problem by smoothing the objective function so that it becomes differentiable. The smoothed estimator is asymptotically equivalent to the standard LAD estimator. With bootstrap critical values, the levels of symmetrical t and χ^2 tests based on the smoothed estimator are correct through $O(n^{-\gamma})$, where $\gamma < 1$ but can be arbitrarily close to 1. In contrast, first-order asymptotic approximations make an error of size $O(n^{-\gamma})$. The bootstrap accounts for terms of size $O(n^{-\gamma})$ in the asymptotic expansions of the test statistics, whereas first-order approximations ignore these terms. These results also hold for symmetrical t and χ^2 tests for censored median regression models.

KEY WORDS: Asymptotic expansion, smoothing, L^1 regression, least absolute deviations

---

Research supported in part by NSF grant SBR-9307677. I thank Peter Bickel, Moshe Buchinsky, Oliver Linton, Paul Ruud, and Gene Savin for helpful comments and discussions.

BOOTSTRAP METHODS FOR MEDIAN REGRESSION MODELS

1. INTRODUCTION

A linear median regression model has the form

$$Y = X\beta + U, \tag{1.1}$$

where Y is an observed scalar dependent variable, X is a $1\times q$ vector of observed explanatory variables, β is a $q\times 1$ vector of constant parameters, and U is an unobserved random variable that satisfies median$(U|X=x) = 0$ almost surely. The parameters β may be estimated by the method of least absolute deviations (LAD). Bassett and Koenker (1978) and Koenker and Bassett (1982) give conditions under which the LAD estimator is $n^{1/2}$-consistent and asymptotically normal. Koenker and Bassett (1978) treat quantile regressions, which generalize (1.1) by specifying that a quantile of the conditional distribution of U (not necessarily the median) is zero. Bloomfield and Steiger (1983), Koenker (1982), and Koenker and Bassett (1978), among others, discuss the robustness properties of the LAD estimator.

The asymptotic normality of the LAD estimator makes it possible to form asymptotic t and χ^2 statistics for testing hypotheses about β in (1.1). However, first-order asymptotic approximations can be inaccurate with samples of the sizes encountered in applications. As a result, the true and nominal levels of t and χ^2 tests and the true and nominal coverage probabilities of confidence intervals for components of β can be very different when critical values based on first-order asymptotic approximations are used. Buchinsky (1995), de Angelis, *et al*. (1993), Dielman and Pfaffenberger (1984, 1986, 1988), and Monte Carlo results that are presented later in this paper provide numerical evidence on the accuracy of first-order approximations.

This paper shows that the bootstrap provides asymptotic refinements to the levels of t and χ^2 tests of hypotheses about β in (1.1). That is, as the sample size, n, increases, the differences between the true and nominal levels of the tests converge to zero more rapidly with critical values obtained from the bootstrap than with critical values obtained from first-order asymptotic theory. It is well known that under suitable conditions the bootstrap provides asymptotic refinements to the levels of tests and coverage probabilities of confidence intervals (see, e.g., Beran 1988; Hall 1986, 1992; Horowitz 1996). However, the standard theory of the bootstrap does not apply to t and χ^2 statistics based on the LAD estimator. This theory is based on an Edgeworth expansion of the distribution of the statistic of interest. The validity of the expansion is usually established by using a Taylor series to approximate the statistic by a smooth function of sample moments that satisfies conditions given, for example, by Bhattacharya and Ghosh (1978) for the existence of an Edgeworth expansion. The LAD objective function is not smooth, however, and Taylor series methods cannot be used to approximate the LAD estimator by a smooth function of sample moments. Indeed, de Angelis, *et al*. (1993) have shown that the distribution of the LAD estimator has a non-standard and very complicated asymptotic expansion.

This paper solves these problems by smoothing the LAD objective function to make it differentiable. The resulting estimator will be called the smoothed LAD (SLAD) estimator. It is first-order asymptotically equivalent to the standard LAD estimator but has much simpler higher-order asymptotics. Use of the SLAD estimator greatly eases the task of obtaining asymptotic refinements to levels of tests and, thereby, makes it possible to obtain results that go well beyond those obtained in previous research.

Previous research by de Angelis, *et al*. (1993) has shown that when U is independent of X and certain other conditions are satisfied, the error in the bootstrap approximation to the cumulative distribution function (CDF) of the LAD estimator is $o(n^{-2/5})$. Hahn (1995) showed consistency of a bootstrap approximation to the CDF without assuming independence of U and X, but he did not investigate the size of the approximation error. Neither de Angelis, *et al*. nor Hahn investigated the bootstrap's ability to correct the levels of t and χ^2 tests based on the LAD estimator.

Janas (1993) investigated the related but simpler problem of testing a hypothesis about a population median (no covariates). He showed that when a suitable version of the bootstrap is used to obtain the critical value, the difference between the true and nominal levels of a symmetrical t test of a hypothesis about a population median is $o(n^{-\gamma})$, where $\gamma < 1$ but can be arbitrarily close to 1 if the underlying population density is sufficiently smooth. By contrast, first-order approximations make an error of size $O(n^{-\gamma})$. The bootstrap accounts for a term of size $O(n^{-\gamma})$ in the asymptotic expansion of the distribution of the test statistic, whereas first-order approximations ignore this term.[1]

This paper extends the results of previous research in three ways. First, it gives conditions under which the bootstrap provides asymptotic refinements to the levels of t and χ^2 tests of hypotheses about β in (1.1). Second, in contrast to de Angelis, *et al*. (1993), it is not assumed that U and X are independent. Any form of dependence is permitted as long as $\text{median}(U|X=x) = 0$ almost surely and mild regularity conditions are satisfied. Third, it is shown that the bootstrap also provides asymptotic refinements for t and χ^2 tests of hypotheses about β in the censored median regression model of Powell (1984). Under the conditions that are given here, the differences between the true and nominal levels of symmetrical t and χ^2 tests with bootstrap critical values are $o(n^{-\gamma})$ for a suitable γ satisfying $7/9 < \gamma < 1$. By contrast, the differences between the true and nominal levels are $O(n^{-\gamma})$ with critical values based on first-order approximations. As in Janas (1993), the bootstrap accounts for a term of size $O(n^{-\gamma})$ in the asymptotic expansion of the t or χ^2 statistic, whereas first-order approximations ignore this term. The value of γ depends on the smoothness of the conditional density of U at zero and can be arbitrarily close to 1 if the density is sufficiently smooth.

Although this paper treats explicitly only the levels of symmetrical t and χ^2 tests, it will be clear that the results also apply to coverage probabilities of symmetrical confidence intervals and, with suitable modifications, to equal-tailed and one-sided t tests and confidence intervals. In addition, the methods used here can easily be extended to show that the bootstrap provides asymptotic refinements for tests and confidence intervals based on smoothed versions of the quantile-regression estimator of Koenker and Bassett (1978) and the censored quantile-regression estimator of Powell (1986).

The remainder of the paper is organized as follows. Section 2 describes the smoothed LAD estimator and gives its first-order asymptotic distribution. Section 3 describes the test statistics and procedures that are used to obtain bootstrap critical values. Section 4 presents theorems giving conditions under which the bootstrap provides asymptotic refinements to the levels of symmetrical t and χ^2 tests. Section 4 also describes the extension to censored median regressions. Section 5 presents the results of a small Monte Carlo investigation of the numerical performance of the bootstrap, and Section 6 gives concluding comments. The proofs of theorems are in the Appendix.

2. THE SMOOTHED LAD ESTIMATOR

This section describes the smoothed LAD estimator and establishes its asymptotic equivalence to the standard LAD estimator.

Let $\{Y_i, X_i: \; i = 1,...,n\}$ be a random sample of (Y,X) in (1.1). The standard LAD estimator solves

$$\underset{b \in B}{\text{minimize:}} \quad {}_n(b) \equiv n^{-1}\sum_{i=1}^{n}|Y_i - X_i b|$$

(2.1)
$$= n^{-1}\sum_{i=1}^{n}(Y_i - X_i b)[2I(Y_i - X_i b > 0) - 1],$$

where B is the parameter set and I(•) is the indicator function. $_n(b)$ has cusps and, therefore, is not differentiable at points b such that $Y_i - X_i b = 0$ for some i. The SLAD estimator smooths these cusps by replacing the indicator function in $_n$ with a smooth function.

To do this, let K be a bounded, differentiable function satisfying K(v) = 0 if v $\leq$ -1 and K(v) = 1 if v $\geq$ 1. Additional requirements that K must satisfy are given in Section 4a. Let $\{h_n\}$ be a sequence of positive real numbers (bandwidths) that converges to zero as $n \to \infty$. The SLAD estimator solves

$$\text{(2.2)} \qquad \text{minimize:} \quad H_n(b) \equiv n^{-1}\sum_{i=1}^{n}(Y_i - X_i b)\left[2K\left(\frac{Y_i - X_i b}{h_n}\right) - 1\right].$$

K is analogous to the integral of a kernel function for nonparametric estimation. K is not a kernel function itself.

It may appear that the presence of a smoothing parameter h_n in (2.2) is a disadvantage of SLAD relative to LAD, but this appearance is misleading. With median regression models, smoothing and the introduction of smoothing parameters are unavoidable for obtaining satisfactory performance of the bootstrap. Under assumptions stronger than those made here, de Angelis, *et al*. (1993) found that the error in the bootstrap approximation to the distribution of the LAD estimator converges to zero more slowly than the error made by first-order asymptotic theory unless the bootstrap samples a smoothed version of the data. Janas (1993) smooths the data to obtain bootstrap refinements for a test of a hypothesis about a population median. The smoothing methods of de Angelis, *et al.* and Janas do not extend easily to models with heteroskedasticity or censoring. In this paper, smoothing the objective function replaces smoothing the data. The resulting SLAD estimator is useful because it enables asymptotic refinements to coverage probabilities of confidence intervals and levels of tests to be obtained easily. The SLAD estimator is not needed if the only objective is to obtain a point estimate of

β.

Let $_n$ be a LAD estimator (a solution to (2.1)) and b_n be a SLAD estimator (a solution to (2.2)). Intuition suggests that $_n$ and b_n are asymptotically equivalent if h_n converges to zero sufficiently rapidly. Theorem 2.1 below shows that this intuition is correct. Regularity conditions for the theorem are given in Section 4a. They are stated in the form that is used to obtain this paper's main objective, which is to show that the bootstrap provides asymptotic refinements for tests based on the SLAD estimator. The regularity conditions are stronger than would be needed if the only objective were to prove that $_n$ and b_n are asymptotically equivalent.

<u>Theorem 2.1</u>: Under Assumptions 1-6 of Section 4a, $n^{1/2}(b_n - {}_n) = o_p(1)$.■

To state the asymptotic distribution of $n^{1/2}(b_n - \beta)$, let $f(\bullet|x)$ denote the density of U in (1.1) conditional on X = x. Assume that $f(0|x)$ exists at U = 0 for almost all x. Define $D = 2E[X'Xf(0|X)]$, and assume that D is nonsingular. It follows from Theorem (2.1) and asymptotic normality of the LAD estimator (see, e.g., Buchinsky 1995) that $n^{1/2}(b_n - \beta) \rightarrow^d N(0,V)$, where $V = D^{-1}E(X'X)D^{-1}$. To obtain a consistent estimator of V, let $K^{(1)}(v) = dK(v)/dv$. Define

$$(2.3) \qquad D_n(b) = 2(nh_n)^{-1}\sum_{i=1}^{n} X_i'X_i K^{(1)}\left(\frac{Y_i - X_i b}{h_n}\right).$$

It is not difficult to show that $D_n(b_n) \rightarrow^p D$ under the conditions given in Section 4a. E(X'X) can be estimated consistently by the sample average of X'X. However, for purposes of obtaining asymptotic refinements, it is more convenient to use an estimator of the exact finite-sample variance of the first derivative of $H_n(b)$ at $b = \beta$. This estimator is $T_n(b_n)$, where

$$T_n(b) = n^{-1}\sum_{i=1}^{n} X_i'X_i\left\{\left[2K\left(\frac{Y_i - X_i b}{h_n}\right) - 1\right]^2 + 2\left(\frac{Y_i - X_i b}{h_n}\right)K^{(1)}\left(\frac{Y_i - X_i b}{h_n}\right)\right\}.$$

Under the conditions given in Section 4a, $T_n(b_n) \rightarrow^p E(X'X)$. It follows that V is estimated consistently by $V_n \equiv D_n(b_n)^{-1}T_n(b_n)D_n(b_n)^{-1}$.

3. TESTING A HYPOTHESIS ABOUT β

a. The Symmetrical t and Chi-Square Tests

Let b_{ni} and β_i, respectively, be the i'th components of b_n and β (i = 1,...,q). Let V_{ni} be the (i,i) component of V_n. The t statistic for testing the hypothesis H_0: $\beta_i = \beta_{0i}$ is $t \equiv n^{1/2}(b_{ni} - \beta_{0i})/V_{ni}^{1/2}$. If H_0 is true, then $t \rightarrow^d N(0,1)$. The symmetrical t test rejects H_0 at the asymptotic α level if $|t| > z_{\alpha/2}$, where $z_{\alpha/2}$, the asymptotic critical value, is the 1 - $\alpha/2$ quantile of the standard normal distribution.

Now let R be an $\ell \times q$ matrix with $\ell \leq q$, and let c be an $\ell \times 1$ vector of constants. Consider a test of the hypothesis H_0: $R\beta = c$. Assume that the matrix $RD^{-1}E(X'X)D^{-1}R'$ is nonsingular. Then under H_0, the statistic

$$\chi^2 \equiv n(Rb_n - c)'(RV_nR')^{-1}(Rb_n - c)$$

is asymptotically chi-square distributed with ℓ degrees of freedom. H_0 is rejected at the asymptotic α level if χ^2 exceeds the asymptotic critical value consisting of the 1 - α quantile of the chi-square distribution.

Section 4 gives conditions under which the bootstrap provides asymptotic refinements to critical values and levels of the symmetrical t and χ^2 tests.

b. The Bootstrap Procedure

The bootstrap estimates the distribution of a test statistic by treating the estimation data as if they were the population. Thus, the bootstrap distribution of a statistic is the distribution induced by sampling the estimation data randomly with replacement. The α-level bootstrap critical value of the symmetrical t test is the 1 - α quantile of the bootstrap distribution of $|t|$. The α-level bootstrap critical value of a test based on χ^2 is the 1 - α quantile of the bootstrap distribution of χ^2.

The bootstrap distributions of $|t|$ and χ^2 can be estimated with arbitrary accuracy by Monte Carlo simulation. To specify the Monte Carlo procedure, let the bootstrap sample be denoted by $\{Y_i^*,X_i^*: i = 1,...,n\}$. Define the following bootstrap analogs of $H_n(b)$, $D_n(b)$ and $T_n(b)$:

$$H_n^*(b) \equiv n^{-1}\sum_{i=1}^{n}(Y_i^* - X_i^*b)\left[2K\left(\frac{Y_i^* - X_i^*b}{h_n}\right) - 1\right],$$

$$D_n^*(b) = (nh_n)^{-1}\sum_{i=1}^{n} X_i^{*\prime}X_i^*K^{(1)}\left(\frac{Y_i^* - X_i^*b}{h_n}\right)$$

and

$$T_n^*(b) =$$

$$n^{-1}\sum_{i=1}^{n} X_i^{*\prime}X_i^*\left\{\left[2K\left(\frac{Y_i^* - X_i^*b}{h_n}\right) - 1\right] + 2\left(\frac{Y_i^* - X_i^*b}{h_n}\right)K^{(1)}\left(\frac{Y_i^* - X_i^*b}{h_n}\right)\right\}^2.$$

Let b_n^* be a solution to (2.2) with H_n replaced by H_n^*. Let V_{ni}^* be the (i,i) component of the matrix $D_n^*(b_n^*)^{-1}T_n^*(b_n^*)D_n^*(b_n^*)^{-1}$.

The Monte Carlo procedure for estimating the bootstrap critical value of the symmetrical t test is as follows. The procedure for estimating the bootstrap critical value of χ^2 is similar.

1. Generate a bootstrap sample $\{Y_i^*,X_i^*: i = 1,...,n\}$ by sampling the estimation data randomly with replacement.[2]

2. Using the bootstrap sample, compute the bootstrap t statistic for testing the hypothesis H_0^*: $\beta_i = b_{ni}$, where b_n solves (2.2). The bootstrap t statistic is $t^* \equiv n^{1/2}(b_{ni}^* - b_{ni})/(V_{ni}^*)^{1/2}$, where b_{ni}^* is the i'th component of b_n^*.

3. Estimate the bootstrap distribution of $|t^*|$ by the empirical distribution that is obtained by repeating steps 1 and 2 many times. The bootstrap critical value of the symmetrical t test is estimated by the $1 - \alpha$ quantile of this empirical distribution.

Because the bootstrap critical value can be estimated with arbitrary accuracy by repeating steps 1 and 2 sufficiently many times, the results presented in Section 4 pertain to the true bootstrap critical value, not its Monte Carlo estimator.

4. MAIN RESULTS

This section presents theorems giving conditions under which the bootstrap provides asymptotic refinements to the levels of symmetrical t and χ^2 tests based on the SLAD estimator. As in other applications (see, e.g., Beran 1988, Hall 1992), the proof that the bootstrap provides asymptotic refinements is based on showing that the distributions of the test statistics and their bootstrap analogs have asymptotic expansions that are identical to sufficiently high order. The main technical problem that must be solved is establishing conditions under which these expansions exist. This is done in Theorems 4.1 and 4.2. Once the existence of the expansions is established, it is a relatively easy matter to show that the use of bootstrap critical values provides asymptotic refinements to the levels of symmetrical t and χ^2 tests. This is done in Theorem 4.3.

a. Assumptions

This subsection presents the assumptions under which it is proved that the bootstrap provides asymptotic refinements for symmetrical t and χ^2 tests based on the SLAD estimator. Let $r \geq 4$ be an even integer. Let $K^{(i)}(v) = d^iK(v)/dv^i$. The assumptions are:

1. $\{Y_i, X_i: i = 1,...,n\}$ is a random sample of (Y,X), where $Y = X\beta + U$, X is a $1\times q$ vector of observed random variables, U is an unobserved random

scalar, and β is a $q\times 1$ constant vector.

2. β is an interior point of B, which is a compact subset of $\Re^q$.

3. The support of the distribution of X is bounded, and E(X'X) is positive definite.

4. Let $F(\bullet|x)$ and $f(\bullet|x)$, respectively, denote the CDF and density of U conditional on X = x. (a) $F(0|x) = 0.50$ for almost every x. (b) For all u in a neighborhood of 0 and almost every x, $f(u|x)$ exists, is bounded away from zero, and is r - 1 times continuously differentiable with respect to u.

5. (a) $K(\bullet)$ is bounded, $K(v) = 0$ if $v \leq -1$, and $K(v) = 1$ if $v \geq 1$. (b) K is 4-times differentiable everywhere, $K^{(1)}(v)$ is symmetrical about v = 0, and $K^{(i)}$ $(i = 1,...,4)$ is bounded and Lipschitz continuous on $(-\infty,\infty)$. (c) Let (v) be a vector whose components are $[2K(v) - 1]$ and its derivatives through order 3, $vK^{(1)}(v)$ and its derivatives through order 3, and $[2K(v) - 1 + 2vK^{(1)}(v)]^2$ and its first derivative. For any $\theta \in \Re^{10}$ satisfying $\|\theta\| = 1$, there is a partition of [-1,1], $-1 = a_1 < a_2 <... < a_{L(\theta)} = 1$ such that $\theta'(v)$ is either strictly increasing or strictly decreasing on $(a_{\ell-1}, a_\ell)$ $(\ell = 2,...,L(\theta))$. (d) For each integer i $(1 \leq i \leq r)$,

$$\int_{-1}^{1} v^i K^{(1)}(v)dv = \begin{cases} 0 \text{ if } i < r \\ C_K \text{ (nonzero) if } i = r \end{cases}$$

6. $h_n \propto n^{-\kappa}$, where $2/(2r + 1) < \kappa < 1/3$.

Assumptions 1-5b define the model and insure that β is identified, $n^{1/2}(b_n - \beta)$ is asymptotically normal, and the Taylor series expansions used to obtain higher-order asymptotic approximations to t and χ^2 exist. The assumption that X has bounded support is not essential and can be dropped at the expense of more complex proofs. Assumption 5c is used to establish a modified form of the Cramer condition of Edgeworth analysis (lemma 9 of the Appendix). Assumption 5d, which requires $K^{(1)}$ to be a "higher-order" kernel, and Assumption 6 insure that the (first-order) asymptotic distribution of $n^{1/2}(b_n - \beta)$ has mean zero and that Taylor series remainder terms are negligibly small. Functions K satisfying assumption 5 can be constructed by integrating kernels given by Müller (1984).

b. Theorems

This section gives theorems that establish conditions under which the bootstrap provides asymptotic refinements for symmetrical t and χ^2 tests based on the SLAD estimator. Theorems 4.1 and 4.2 give conditions under which the sample and bootstrap versions of $|t|$ and χ^2 have Edgeworth-type asymptotic expansions. Theorem 4.3 shows that the bootstrap provides asymptotic refinements under the same conditions.

The following additional notation is used. Let Φ and ϕ, respectively, denote the standard normal distribution and density functions. Let P_n^* denote the bootstrap probability measure. This measure places mass 1/n at each data point (Y_i,X_i). The cumulants of t through order 4 can be approximated with an accuracy of $O[(nh_n)^{-1}]$ by using Taylor-series expansions that are described in the Appendix. Denote the approximate cumulants by the vector ν_n. The first four cumulants of t* conditional on the estimation sample can also be approximated with an accuracy of $O[(nh_n)^{-1}]$ almost surely. Let ν_n^* be the vector containing the approximate bootstrap cumulants. Define $d = \dim(\nu_n) = \dim(\nu_n^*)$.

The following theorem establishes the existence of Edgeworth-type expansions of the distributions of $|t|$ and $|t^*|$.

<u>Theorem 4.1</u>: Let assumptions 1-6 hold. Let ν be an arbitrary vector with dimension d. There is a function $q(\tau,\nu)$ such that: (a) $q(\bullet,\nu)$ is a polynomial; (b) $q(\tau,\nu_n)$ and $q(\tau,\nu_n^*)$ consist of terms whose sizes are $O[(nh_n)^{-1}]$ (almost surely in the case of $q(\tau,\nu_n^*)$);

(c)

$$P(|t| \le \tau) = 2\Phi(\tau) - 1 + q(\tau,\nu_n)\phi(\tau) + o[(nh_n)^{-1}] \tag{4.1}$$

uniformly over τ, and

(d)

$$P_n^*(|t^*| \le \tau) = 2\Phi(\tau) - 1 + q(\tau,\nu_n^*)\phi(\tau) + o[(nh_n)^{-1}]$$

uniformly over τ almost surely. ■

The coefficents of τ in q are functions of the approximate cumulants of t and t*. These, in turn, are functions of asymptotic forms of moments of products of derivatives of $H_n(\beta)$, $D_n(\beta)$, and $T_n(\beta)$ with respect to the components of β. Because the number of such moments is very large, obtaining an analytic expression for q is not feasible. It is possible, however, to calculate the rates at which the moments converge to zero, and this is sufficient to prove the theorem.

The proof of Theorem 4.1 takes place in two main steps. The first step consists of showing that t and t* can be approximated up to asymptotically negligible remainder terms by functionals of derivatives of $H_n(\beta)$, $D_n(\beta)$, and $T_n(\beta)$ (or their bootstrap analogs in the case of t*). This is done in Propositions 1 and 2 of the Appendix. The second step is to show that the distributions of the approximations to t and t* have asymptotic expansions through order $(nh_n)^{-1}$. This step is carried out using methods similar to those used to prove Theorems 5.5 and 5.6 of Hall (1992).

Now consider the χ^2 test. Let χ^{2*} be the bootstrap version of the χ^2 statistic. The first two moments of χ^2 and χ^{2*} can be approximated through $O[(nh_n)^{-1}]$. Let $\nu_{n\chi}$ and $\nu_{n\chi}^*$ denote the vectors of approximate moments. Let $F_{\chi,\ell}$ denote the chi-square distribution function with ℓ degrees of freedom. The following theorem, which is a modified version of Theorem 1b of Chandra and Ghosh (1979), gives conditions under which the distributions of χ^2 and χ^{2*} have Edgeworth expansions through $O[(nh_n)^{-1}]$.

Theorem 4.2: Let assumptions 1-6 hold. Let ν be an arbitrary 2×1 vector. There is a function $q_\chi(\tau,\nu)$ such that $q(\tau,\nu_{n\chi})$ and $q(\tau,\nu_{n\chi}^*)$ consist of terms whose sizes are $O[(nh_n)^{-1}]$ (almost surely in the case of $q(\tau,\nu_{n\chi}^*)$),

$$(4.2)\qquad P(\chi^2 < z) = \int_{-\infty}^{z} d\{[1 + q_\chi(\xi,\nu_{n\chi})]F_{\chi,\ell}(\xi)\} + o[(nh_n)^{-1}]$$

uniformly over z, and

$$P_n^*(\chi^{2*} < z) = \int_{-\infty}^{z} d\{[1 + q_\chi(\xi,\nu_{n\chi}^*)]F_{\chi,\ell}(\xi)\} + o[(nh_n)^{-1}]$$

uniformly over z almost surely. ■

The final theorem shows that the use of bootstrap critical values yields asymptotic refinements to the levels of symmetrical t and χ^2 tests. Let t_α^* denote the α-level critical value of the bootstrap symmetrical t test. That is, t_α^* is the $1 - \alpha$ quantile of the bootstrap distribution of $|t^*|$. Let c_α^* denote α-level critical value of the bootstrap χ^2 test. That is, c_α^* is the $1 - \alpha$ quantile of the bootstrap distribution of χ^{2*}.

Theorem 4.3: Let assumptions 1-6 hold. Under H_0: $\beta_i = \beta_{0i}$,

a. $P(|t| > t_\alpha^*) = \alpha + o[(nh_n)^{-1}]$.

If $RD^{-1}E(X'X)D^{-1}R'$ is nonsingular, then under H_0: $R\beta = c$,

b. $P(\chi^2 > c_\alpha^*) = \alpha + o[(nh_n)^{-1}]$. ■

First-order asymptotic approximations drop the terms $q\phi$ and $q_\chi F_{\chi,r}$ in (4.1) and (4.2). The resulting approximation errors are $O[(nh_n)^{-1}]$.

c. Censored Median Regressions

This section describes the extension of the foregoing results to the censored median regression model of Powell (1984). The model is

$$(4.3) \qquad Y = \max(0, X\beta + U),$$

where X, β, and U are as defined in (1.1). The censored LAD (CLAD) estimator of β, $_{cn}$, solves

$$\underset{b \in B}{\text{minimize:}} \quad n^{-1}\sum_{i=1}^{n} |Y_i - \max(0, X_i b)|,$$

where B is the parameter set. Equivalently, $_{cn}$ solves

$$\underset{b \in B}{\text{minimize:}} \quad {}_{cn}(b) \equiv$$

$$n^{-1}\sum_{i=1}^{n} \{(Y_i - X_i b)[2I(Y_i - X_i b > 0) - 1] - Y_i\}I(X_i b > 0),$$

Under regularity conditions, $n^{1/2}(_{cn} - \beta) \to^d N(0,V_c)$, where $V_c = D_c^{-1}T_cD_c^{-1}$, $D_c = 2E[X'Xf(0|X)I(X\beta > 0)]$, and $T_c = E[X'XI(X\beta > 0)]$ (Powell 1984).

Like the objective function of the LAD estimator, $_{cn}$ has cusps. The smoothed CLAD estimator (SCLAD) removes them by replacing the indicator functions in $_{cn}$ with smooth functions. The SCLAD estimator, b_{cn}, solves

$$\underset{b \in B}{\text{minimize:}} \quad H_{cn}(b) \equiv n^{-1}\sum_{i=1}^{n} g_c(Y_i, X_i, h_n, b),$$

where

$$g_c(y,x,h,b) = \left\{(y - xb)\left[2K\left(\frac{y - xb}{h}\right) - 1\right] - y\right\}K\left(\frac{xb}{h} - 2\right),$$

and K and h_n are as in (2.2). The smoothed version of I(xb > 0) is K(xb/h -2) instead of K(xb/h) for technical reasons relating to prevention of asymptotic bias. Under conditions stated below, $n^{1/2}(b_{cn} - {}_{cn}) = o_p(1)$ as $n \to \infty$.

To form t and χ^2 statistics based on b_{cn} it is necessary to have consistent estimators of D_c and T_c. Define

$$D_{cn}(b) = (nh_n)^{-1}\sum_{i=1}^{n} X_i{}'X_i K^{(1)}\left(\frac{Y_i - X_i b}{h_n}\right) I(Y_i > 0).$$

It is not difficult to show that $D_{cn}(b_{cn}) \to^p D_c$. T_c can be estimated consistently by the sample average of $X'XI(Xb_{cn} > 0)$ (Powell 1984). As in SLAD estimation, however, for purposes of obtaining asymptotic refinements it is more convenient to use an estimator of the exact finite-sample variance of the first derivative of $H_{cn}(b)$ at $b = \beta$. This estimator is $T_{cn}(b_{cn})$, where

$$T_{cn}(b) = n^{-1}\sum_{i=1}^{n} [\partial g_c(Y_i, X_i, h_n, b)/\partial b][\partial g_c(Y_i, X_i, h_n, b)/\partial b]'$$

V_c is estimated consistently by $V_{cn} \equiv D_{cn}(b_{cn})^{-1}T_{cn}(b_n)D_{cn}(b_{cn})^{-1}$.

The formulae for t and χ^2 statistics for testing hypotheses about β in (4.3) are the same as in Section 3a but with V_n replaced by V_{cn}. The procedure for obtaining bootstrap critical values for these statistics is the same as in Section 3b but with D_n, T_n, D_n*, and T_n* replaced with D_{cn}, T_{cn}, and their bootstrap analogs.

To establish the ability of the bootstrap to provide asymptotic refinements for t and χ^2 tests based on the SCLAD estimator, it is necessary to modify Assumptions 1 and 3 as follows:

1'. $\{Y_i, X_i: i = 1,...,n\}$ is a random sample of (Y,X), where Y = max(0, $X\beta$ + U), X is a $1\times q$ vector of observed random variables, U is an unobserved random scalar, and β is a $q\times 1$ constant vector.

3'. The support of the distribution of X is bounded, $P(X\beta = 0) = 0$, and $E[(X'X)I(Xb > \epsilon)]$ is positive definite for some $\epsilon > 0$ and all b in a neighborhood of β.

The following theorem shows that the SCLAD and CLAD estimators are asymptotically equivalent and that the bootstrap provides asymptotic refinements to the levels of symmetrical t and χ^2 tests based on the SCLAD estimator.

Theorem 4.4: Let assumptions 1', 2, 3', and 4-6 hold. Then

a. $n^{1/2}(b_{cn} - {}_{cn}) = o_p(1)$ as $n \to \infty$.

Let t_α^* and c_α^*, respectively, denote the α-level bootstrap critical values of the SCLAD symmetrical t and χ^2 tests. Under H_0: $\beta_i = \beta_{0i}$,

b. $P(|t| > t_\alpha^*) = \alpha + o[(nh_n)^{-1}]$.

If $RD_c^{-1}T_cD_c^{-1}R'$ is nonsingular, then under H_0: $R\beta = c$,

c. $P(\chi^2 > c_\alpha^*) = \alpha + o[(nh_n)^{-1}]$. ■

5. MONTE CARLO EXPERIMENTS

This section describes the results of a small Monte Carlo investigation of the finite-sample level of the SLAD t test with bootstrap critical values. The numbers of experiments and replications per experiment are small because of the very long computing times they entail, even on a fast computer.

Each experiment evaluates the level of a symmetrical t test using asymptotic or bootstrap critical values. The hypothesis being tested is H_0: $\beta_1 = 1$ in the model $Y = \beta_0 + \beta_1 X + U$, where β_0 and β_1 are scalar parameters whose true values are $(\beta_0,\beta_1) = (1,1)$ (so H_0 is true), and $X \sim U[1,5]$. There are 3 different distributions of U. In the first experiment, $U \sim N(0,2)$. In the second, $U \sim$ Student t with 3 degrees of freedom scaled to have a variance of 2. In the third experiment, $U = 0.25(1 + X)V$, where $V \sim N(0,1)$. Thus, U is heteroskedastic. The smoothing function K is

$$K(v) = \begin{cases} 0 & \text{if } v < -1 \\ 0.5 + (105/64)[v - (5/3)v^3 + (7/5)v^5 - (3/7)v^7] & \text{if } |v| \le 1 \\ 1 & \text{if } v > 1 \end{cases}$$

K is the integral of a 4th-order kernel for nonparametric density estimation (Müller 1984).[3]

The experiments with the SLAD estimator consisted of computing the empirical level of the nominal 0.05-level symmetrical t test of H_0 with bootstrap critical values. To provide a basis for evaluating the performance of the bootstrap, experiments were also carried out with the unsmoothed LAD estimator. These consisted of computing the empirical level of the nominal 0.05-level symmetrical t test of H_0 with the asymptotic critical value. The LAD estimator was studentized by using the consistent variance estimator $D_n(\,_n)^{-1}E_n[(1,X)'(1,X)]D_n(\,_n)^{-1}$, where $_n$ is the LAD estimator of (β_0,β_1), D_n is as in (2.3), and $E_n(\bullet)$ is the sample average. D_n for the LAD estimator was computed using the 2nd-order kernel $K_2(v) = (15/16)(1 - v^2)^2 I(|v| \le 1)$.

Computation of the SLAD and LAD t statistics require choosing the value of a bandwidth parameter for each. Existing theory provides little guidance on how this should be done in finite samples, so experiments were carried out using a range of bandwidth values.[4]

The experiments used a sample size of n = 50 and were carried out with a program written in GAUSS with GAUSS pseudo-random number generators. There were 500 Monte Carlo replications per experiment with the SLAD estimator and 1000 with the LAD estimator. There were fewer replications in the SLAD experiments because of the long computing times required for Monte Carlo simulations with bootstrapping. Each experiment consisted of repeating the following steps 500 or 1000 times:

A. Generate an estimation data set of size n = 50 by randomly sampling (Y,X) from the model under consideration. Obtain the SLAD or LAD estimate of

(β_0,β_1), and compute the t statistic for testing H_0: $\beta_1 = 1$. Call its value t_S if it is based on the SLAD estimator and t_L if it is based on the LAD estimator.

B. In experiments with t_S, compute the bootstrap critical value by following steps 1-3 in Section 3b. Bootstrap samples were obtained by sampling the estimation data generated in step A randomly with replacement. Denote the 0.05-level bootstrap critical value of the SLAD symmetrical t test by $t_{0.05}$*. $t_{0.05}$* was computed from 100 bootstrap samples.

C. Reject H_0 at the nominal 0.05 level based on t_S if $|t_S| > t_{0.05}$*. Reject H_0 at the nominal 0.05 level based on t_L if $|t_L| > 1.96$, the asymptotic critical value.

The results of the experiments are summarized in Figures 1-3, which show the empirical levels of the SLAD t test with bootstrap critical values and the LAD t test with the asymptotic critical value as functions of the bandwidth. In the experiments, the empirical and nominal levels of the LAD test can be made equal by choosing the bandwidth appropriately. The empirical level is very sensitive to the bandwidth, however, and it is an open question whether the "optimal" bandwidth can be estimated precisely in applications. In contrast, the empirical level of the SLAD test with bootstrap critical values is close to the nominal level over a wide range of bandwidths. Thus, use of the SLAD test with bootstrap critical values greatly decreases the importance of precisely estimating an "optimal" bandwidth. Obtaining precise bandwidth estimates is difficult even in relatively simple settings such as nonparametric density estimation, so the SLAD test's relative insensitivity to the bandwidth is an important practical advantage of this test.

6. CONCLUSIONS

This paper has shown how the bootstrap can be used to obtain asymptotic refinements for tests of hypotheses about the parameters of uncensored and censored linear median regression models with or without heteroskedasticity of unknown form. The method is based on smoothing the objective function of the relevant estimator. This approach contrasts with previous research on bootstrap methods for median regressions, which has achieved less general results under more restrictive assumptions by smoothing the data instead of the estimator. This paper has not addressed the problem of how to choose the bandwidth parameter required for smoothing. It is likely that this can also be done with the bootstrap, but the technical details are sufficiently complex and lengthy to require treatment in a separate paper.

APPENDIX

This Appendix provides proofs of the theorems stated in the text. It is assumed unless otherwise stated that assumptions 1-6 hold. Define $U_i = Y_i - X_i\beta$ and

$$G_n(b) \equiv n^{-1}\sum_{i=1}^{n}\left\{(Y_i - X_i b)\left[2K\left(\frac{Y_i - X_i b}{h_n}\right) - 1\right] - |U_i|\right\}$$

The SLAD estimator minimizes both $H_n(b)$ and $G_n(b)$ over $b \in B$. G_n is used for the proofs because it is a sum of bounded terms.

Let $\|\bullet\|$ denote the Euclidean norm. Let $X^{(j)}$ denote the j'th component of X. For $b \in B$, define $G(b) = E[|Y - Xb| - |U|]$ and

$$_n(b) = n^{-1}\sum_{i=1}^{n}(|Y_i - X_i b| - |U_i|).$$

a. Step 1: Approximating t and t*

<u>Lemma 1</u>:

$$\sup_{b \in B} |G_n(b) - G(b)| \le o(n^{-1/2}\log n) + 2h_n$$

almost surely.

<u>Proof</u>: It follows from Lemma 22 of Nolan and Pollard (1987) and Theorem 2.37 of Pollard (1984) that $|_n(b) - G(b)| = o(n^{-1/2}\log n)$ almost surely uniformly over $b \in B$. Also,

$$G_n(b) - {}_n(b) = 2n^{-1}\sum_{i=1}^{n}(Y_i - X_i b)\left[K\left(\frac{Y_i - X_i b}{h_n}\right) - I(Y_i - X_i b > 0)\right].$$

The summand differs from zero only if $|Y_i - X_i b| \le h_n$. Therefore,

$$|G_n(b) - {}_n(b)| \leq 2n^{-1}\sum_{i=1}^{n}|Y_i - X_i b| I(|Y_i - X_i b| \leq h_n) \leq 2h_n .$$

The lemma now follows from the triangle inequality. Q.E.D.

<u>Lemma 2</u>: Given any $r > 0$, $\|b_n - \beta\| \leq r$ almost surely for all sufficiently large n.

<u>Proof</u>: Let $N_r = \{b \in B: \|b - \beta\| > r\}$. By assumptions 3 and 4, β uniquely minimizes $G(b)$ over B. Therefore, $G(b) > G(\beta) + \delta$ for all $b \in N_r$ and some $\delta > 0$. By Lemma 1 and $h_n \to 0$, there is a finite n_0 such that $G_n(b) > G_n(\beta) + \delta/2 > G_n(\beta)$ almost surely for all $b \in N_r$ if $n > n_0$. But $G_n(b_n) \leq G_n(\beta)$. Therefore, b_n N_r almost surely if $n > n_0$. Q.E.D.

For $i,j,k,\ell,m = 1,...,q$, define $G_{ni}(b) = \partial G_n(b)/\partial b_i$, $G_{nij}(b) = \partial^2 G_n(b)/\partial b_i \partial b_j$, $G_{nijk}(b) = \partial^3 G_n(b)/\partial b_i \partial b_j \partial b_k$, and $G_{nijk\ell}(b) = \partial^4 G_n(b)/\partial b_i \partial b_j \partial b_k \partial b_\ell$. Also, define $D_{ni}(b) = \partial D_n(b)/\partial b_i$, $D_{nij}(b) = \partial^2 D_n(b)/\partial b_i \partial b_j$, and $T_{ni}(b) = \partial T_n(b)/\partial b_i$.

<u>Lemma 3:</u> For all $i,j,k,\ell = 1,...,q$, the following relations hold almost surely as $n \to \infty$:

(a) $\sup_{b \in B} |G_{ni}(b) - EG_{ni}(b)| = o[(\log n)/n^{1/2}]$

(b) $\sup_{b \in B} |G_{nij}(b) - EG_{nij}(b)| = o[(\log n)/(nh_n)^{1/2}]$

(c) $\sup_{b \in B} |G_{nijk}(b) - EG_{nijk}(b)| = o[(\log n)/(nh_n^3)^{1/2}]$

(d) $\sup_{b \in B} |G_{nijk\ell}(b) - EG_{nijk\ell}(b)| = o[(\log n)/(nh_n^5)^{1/2}]$

(e) $\sup_{b \in B} |D_n(b) - ED_n(b)| = o[(\log n)/(nh_n)^{1/2}]$

(f) $\sup_{b \in B} |D_{ni}(b) - ED_{ni}(b)| = o[(\log n)/(nh_n^3)^{1/2}]$

(g) $\sup_{b \in B} |D_{nij}(b) - ED_{nij}(b)| = o[(\log n)/(nh_n^5)^{1/2}]$

(h) $\sup_{b \in B} |T_n(b) - ET_n(b)| = o[(\log n)/n^{1/2}]$

(i) $\sup_{b \in B} |T_{ni}(b) - ET_{ni}(b)| = o[(\log n)/(nh_n)^{1/2}]$,

where (e)-(i) apply to the individual components of the matrices D_n, D_{ni}, D_{nij}, T_n, and T_{ni}. In addition, for all $i,j,k,\ell = 1,..., q$

(j) $EG_{ni}(\beta) = 2[(1 - r)/r!]C_K h_n^r E[X^{(i)} f^{(r-1)}(0|X)] + o(h_n^r)$

(k)

$$n^{1/2} G_{nj}(\beta) = -n^{-1/2} \sum_{i=1}^{n} X_i^{(j)} [2I(U_i > 0) - 1] + O_p(n^{1/2} h_n^r + h_n^{1/2})$$

(l) $EG_{nij}(\beta) = 2E[X^{(i)} X^{(j)} f(0|X)] + O(h_n^r)$

(m) $EG_{nijk}(b)$, $EG_{nijk\ell}(b)$, $ED_n(b)$, $ED_{ni}(b)$, $ED_{nij}(b)$, $ET_n(b)$, and $ET_{ni}(b)$, and are $O(1)$ as $n \to \infty$ for all b in a neighborhood of β.

<u>Proof</u>: Parts (a)-(i) are proved by using Lemmas 2.14 of Pakes and Pollard (1989) and 22 of Nolan and Pollard (1987) to show that the summands of the relevant G, D and T functions form Euclidean classes and then applying Theorem 2.37 of Pollard (1984). To prove (j), write $G_{nj}(\beta) = G_{nj}^{(1)} + G_{nj}^{(2)}$, where

$$G_{nj}^{(1)} = -n^{-1} \sum_{i=1}^{n} X_i^{(j)} [2K(U_i/h_n) - 1],$$

$$G_{nj}^{(2)} = -2n^{-1} \sum_{i=1}^{n} (U_i/h_n) X_i^{(j)} K^{(1)}(U_i/h_n),$$

and $U_i = Y_i - X_i\beta$. Then

$$EG_{nj}^{(1)} = -\int_{-\infty}^{\infty} x^{(j)} [2K(u/h_n) - 1] f(u|x) du dP(x).$$

Since $2K(u/h_n) - 1 = \pm 1$ unless $|u/h_n| < 1$, a change of variables gives

$$\text{(A1)} \qquad EG_{n1}^{(1)} = -\int x^{(j)} [1 - F(h_n|x) - F(-h_n|x)] dP(x)$$
$$- h_n \int_{-1}^{1} x^{(j)} [2K(\zeta) - 1] f(h_n\zeta|x) d\zeta dP(x).$$

Integration by parts yields

$$\int_{-1}^{1} \zeta^k [2K(\zeta) - 1]d\zeta = [2/(k + 1)]\left[1 - \left|\int_{-1}^{1} \zeta^{k+1} K^{(1)}(\zeta)d\zeta\right|\right]\delta_k$$

$$\equiv [2/(k + 1)](1 - c_k)\delta_k,$$

for each $k = 0, ..., r - 1$, where $\delta_k = 0$ if k is even and 1 if k is odd, and $c_k = 0$ unless $k = r - 1$. Therefore, Taylor series expansions of the integrands in (A1) about $h_n = 0$ yield

$$EG_{nj}^{(1)} = h_n \sum_{k=0}^{r-1} [(h_n)^k - (-h_n)^k]/(k + 1)!]E[X^{(j)}f^{(k)}(0|X)]$$

$$- 2h_n \sum_{k=0}^{r-1} [(1 - c_k)\delta_k/(k + 1)!]h_n^k E[X^{(j)}f^{(k)}(0|X)] + o(h_n^r)$$

(A2) $$= 2(r!)^{-1}C_K h_n^r E[X^{(j)}f^{(r - 1)}(0|X)] + o(h_n^r).$$

In addition,

$$EG_{nj}^{(2)} = -2\int_{-\infty}^{\infty} X^{(j)}(u/h_n)K^{(1)}(u/h_n)f(u|x)dudP(x)$$

$$= -2h_n \int_{-1}^{1} x^{(j)}\zeta K^{(1)}(\zeta)f(h_n\zeta|x)d\zeta dP(x).$$

A Taylor series expansion of the integrand about $h_n = 0$ yields

$$EG_{nj}^{(2)} = -2h_n \sum_{k=0}^{r-1} \int_{-1}^{1} \zeta^{k-1} K^{(1)}(\zeta)d\zeta(h_n^k/k!)E[X^{(j)}f^{(k)}(0|X)] + o(h_n^r)$$

(A3) $$= -2C_K[(r - 1)!]^{-1}h_n^r E[X^{(j)}f^{(r - 1)}(0|X)] + o(h_n^r).$$

Part (j) follows by combining (A2) and (A3).

To prove (k), observe that

$$n^{1/2}G_{nj}^{(1)} = -n^{-1/2}\sum_{i=1}^{n} X_i^{(j)}[2I(U_i > 0) - 1]$$

$$- 2n^{-1/2}\sum_{i=1}^{n} X_i^{(j)}[2K(U_i/h_n) - I(U_i > 0)].$$

The variance of the second term is $O(h_n)$, and methods similar to those used to prove (k) show that its mean is $O(n^{1/2}h_n^r)$. Similarly, $En^{1/2}G_{nj}^{(2)} = O(n^{1/2}h_n^r)$, and $\mathrm{Var}(n^{1/2}G_{nj}^{(2)}) = O(h_n)$. Part (k) now follows from Chebyshev's inequality.

To prove (l), write $G_{njk}(\beta) = G_{njk}^{(1)} + G_{njk}^{(2)}$, where

$$G_{njk}^{(1)} = 4(nh_n)^{-1}\sum_{i=1}^{n} X_i^{(j)}X_i^{(k)}K^{(1)}(U_i/h_n)$$

and

$$G_{njk}^{(2)} = 2(nh_n)^{-1}\sum_{i=1}^{n} X_i^{(j)}X_i^{(k)}(U_i/h_n)K^{(2)}(U_i/h_n)$$

Arguments similar to those applied to $EG_{nj}^{(1)}$ yield

(A4) $$EG_{njk}^{(1)} = 4E[X^{(j)}X^{(k)}f(0|X)] + O(h_n^r).$$

Similarly,

$$EG_{nj}^{(2)} = 2h_n^{-1}\int_{-\infty}^{\infty} x^{(j)}x^{(k)}(u/h_n)K^{(2)}(u/h_n)f(u|x)dudP(x)$$

$$= 2\sum_{i=0}^{r}\int\int_{-1}^{1}\zeta^{i+1}K^{(2)}(\zeta)(h_n^i/i!)f^{(i)}(0|x)d\zeta dP(x) + o(h_n^r)$$

by a change of variables and a Taylor series expansion. Integration by parts shows that

(A5) $$\int_{-1}^{1}\zeta^{i+1}K^{(2)}(\zeta)d\zeta = \begin{cases} -1 \text{ if } i = 0 \\ 0 \text{ if } 1 \le i < r \\ -(r + 1)C_K \text{ if } i = r \end{cases}$$

Therefore

$$\text{(A6)} \qquad EG_{njk}^{(2)} = -2E[X^{(j)}X^{(k)}f(0|X)] + O(h_n^r).$$

Part (l) follows by combining (A4) and (A6).

To prove (m), consider $EG_{njk\ell}(b)$. Let $\Delta b = b - \beta$. Write $G_{njk\ell}(b) = G_{njk\ell}^{(1)}(b) + G_{njk\ell}^{(2)}(b)$, where

$$G_{njk\ell}^{(1)}(b) = -6(nh_n^2)^{-1}\sum_{i=1}^{n} X_i^{(j)}X_i^{(k)}X_i^{(\ell)}K^{(2)}\left[\frac{U_i - X_i\Delta b}{h_n}\right]$$

and

$$G_{njk\ell}^{(2)}(b) = -2(nh_n^2)^{-1}\sum_{i=1}^{n} X_i^{(j)}X_i^{(k)}X_i^{(\ell)}\frac{U_i - X_i\Delta b}{h_n}K^{(2)}\left[\frac{U_i - X_i\Delta b}{h_n}\right].$$

Now

$$EG_{njk\ell}^{(1)}(b) = -6h_n^{-2}E\left\{X^{(j)}X^{(k)}X^{(\ell)}\int_{-\infty}^{\infty} K^{(2)}\left[\frac{u - X\Delta b}{h_n}\right]f(u|X)du\right\}$$

A change of variables, a Taylor series expansion, and (A5) yield

$$EG_{njk\ell}^{(1)}(b) = -6E\left\{X^{(j)}X^{(k)}X^{(\ell)}\int_{-1}^{1} \zeta K^{(2)}(\zeta)f^{(1)}(\zeta + X\Delta b|X)d\zeta\right\},$$

for between 0 and h_n, which is bounded uniformly over Δb in a neighborhood of 0 by assumption 4. Similar arguments apply to $EG_{njk\ell}^{(2)}(b)$ and the remaining G, D, and T functions. Q.E.D.

Define S_{nG} to be a vector containing the unique components of $G_{ni}(\beta)$, $G_{nij}(\beta)$, $G_{nijk}(\beta)$, and $G_{nijk\ell}(\beta)$ $(i,j,k,\ell = 1,...,q)$. Order the components of S_{nG} so that the first q are the $G_{ni}(\beta)$.

Lemma 4: Let $S_G = \text{plim}_{n\to\infty} S_{nG}$. There is a function $\Lambda_\beta(S_{nG})$ taking values in $\Re^q$ such that $\Lambda_\beta(S_G) = 0$ and

$$(b_n - \beta) = \Lambda_\beta(S_{nG}) + o[1/(n^{3/2}h_n)]$$

almost surely as $n \to \infty$.

Proof: Define $\delta_n = b_n - \beta$ and $\delta_{ni} = b_{ni} - \beta_i$ ($i = 1,...,q$). Let $G_{n\bullet}(\beta)$ be the vector whose components are the unique components of $G_{ni}(\beta)$ ($i = 1,...,q$). For fixed j, k, and ℓ, define $G_{n\bullet j}(\beta)$, $G_{n\bullet jk}(\beta)$, and $G_{n\bullet jk\ell}(\beta)$, respectively, to be the q-dimensional vectors whose components are $G_{nij}(\beta)$, $G_{nijk}(\beta)$, and $G_{nijk\ell}(\beta)$ ($i = 1,...,q$). Let Q_n be the matrix whose (i,j) element is $G_{nij}(\beta)$. By Lemma 2, b_n satisfies the first-order condition $G_{n\bullet}(b_n) = 0$ almost surely for all sufficiently large n. By assumptions 3-4 and Lemma 3, $Q_n(\beta)$ has an inverse almost surely for all sufficiently large n. Therefore, a Taylor series expansion of $G_{n\bullet}(b_n) = 0$ about $b_n = \beta$ yields

$$\text{(A7)} \qquad (b_n - \beta) = -Q_n^{-1}[G_{n\bullet}(\beta) + (1/2)G_{n\bullet jk}(\beta)\delta_{nj}\delta_{nk}$$

$$+ (1/6)G_{n\bullet jk\ell}(\beta)\delta_{nj}\delta_{nk}\delta_{n\ell} + R_n],$$

almost surely for all sufficiently large n, where the summation convention is used,

$$R_n = (1/6)[G_{n\bullet jk\ell}(\ _n) - G_{n\bullet jk\ell}(\beta)]\delta_{nj}\delta_{nk}\delta_{n\ell},$$

and $_n$ is between b_n and β. By using arguments similar to those used to prove Lemma 3(m), it may be shown that $E[G_{n\bullet jk\ell}(b) - G_{n\bullet jk\ell}(\beta)] = O(b - \beta)$ for b in a neighborhood of β. This result and Lemma 3(d) imply that

$$\|R_n\| \leq \{o[(\log n)/(nh_n^5)^{1/2}] + O(\|b_n - \beta\|\}\|b_n - \beta\|^3$$

almost surely. Given any $\nu > 0$ and $c > 0$, suppose that $\|\delta_n\| < cn^{-1/2 + \nu}$. Then it follows from Lemma 3 that the right-hand side of (A7) is less than $cn^{-1/2 + \nu}$

almost surely for all sufficiently large n. In addition, Lemma 3b and assumptions 3-4 imply that the consistent solution to $G_{n\bullet}(b) = 0$ is almost surely unique for all sufficiently large n. Therefore, application of the Brouwer fixed point theorem to the right-hand side of (A7) shows that for any $c > 0$, $\nu > 0$,

$$\|b_n - \beta\| \le cn^{-1/2+\nu} \tag{A8}$$

almost surely for all sufficiently large n. Application of the implicit function theorem to (A7) shows that there is almost surely a differentiable function Λ_β such that $\Lambda_\beta(S_G) = 0$ and

$$(b_n - \beta) = \Lambda_\beta(S_{nG} + {}_n), \tag{A9}$$

where $_n$ is a vector such that $\dim({}_n) = \dim(S_{nG})$, R_n forms the first q components of $_n$, and the remaining components of $_n$ are 0. Application of the mean value theorem to (A9) combined with (A8) shows that

$$(b_n - \beta) = \Lambda_\beta(S_{nG}) + O[(\log n)(n^4 h_n^5)^{-1/2} n^{3\nu}] \tag{A10}$$

almost surely for any $\nu > 0$. The lemma now follows from assumption 6 by making ν sufficiently small. Q.E.D.

Proof of Theorem 2.1: It follows from Lemma 3 that $Q_n \to D$ almost surely. Therefore, by (A7), (A8) and a further application of Lemma 3,

$$n^{1/2}(b_n - \beta) = D^{-1}G_{n\bullet}(\beta) + o_p(1)$$

$$= D^{-1}n^{-1/2}\sum_{i=1}^{n} X_i^{(j)}[2I(U_i > 0) - 1] + o_p(1). \tag{A11}$$

The theorem follows by observing that (A11) is the Bahadur representation of the LAD estimator. Q.E.D.

Let S_n denote the vector consisting of the unique components of S_{nG}, $D_n(\beta)$, $D_{ni}(\beta)$, $D_{nij}(\beta)$, $T_n(\beta)$, and $T_{ni}(\beta)$.

<u>Lemma 5</u>: For each i = 1,...,q, there is a real-valued function $\Lambda_{Vi}(S_n)$ such that

$$V_{ni}^{1/2} = \Lambda_{Vi}(S_n) + \zeta_n,$$

where $\zeta_n = o[(nh_n)^{-1}]$ almost surely.

<u>Proof</u>: Expand $D_n(b_n)$ and $T_n(b_n)$ in Taylor series about $b_n = \beta$ through orders $\|b_n - \beta\|^2$ and $\|b_n - \beta\|$, respectively, and use (A10) to obtain

$$\text{(A12)} \qquad V_{ni}^{1/2} = {}_{Vi}(S_n, S_{nG} + \omega_n) + o[(nh_n)^{-1}]$$

almost surely for a suitable differentiable function $_{Vi}$, where $\omega_n = o[(nh_n)^{-1}]$. The lemma follows by applying the mean value theorem to (A12). Q.E.D.

<u>Proposition 1</u>: Define $\Lambda(S_n) = \Lambda_\beta(S_{nG})/\Lambda_{Vi}(S_n)$. Then

$$\lim_{n\to\infty} \sup_z (nh_n)\{P(t \le z) - P[n^{1/2}\Lambda(S_n) \le z]\} = 0.$$

<u>Proof</u>: By Lemmas 4 and 5

$$\text{(A13)} \qquad t = \frac{n^{1/2}\Lambda_\beta(S_{nG}) + \epsilon_n}{\Lambda_{Vi}(S_n) + \nu_n},$$

where ϵ_n and ν_n are $o[(nh_n)^{-1}]$ almost surely. Define $\Delta_n = t - n^{1/2}\Lambda(S_n)$. A Taylor series approximation applied to (A13) yields $\Delta_n = o[(nh_n)^{-1}]$ almost surely. Choose the sequence $\{\omega_n\}$ such that $\omega_n = o[(nh_n)^{-1}]$ and $\Delta_n/\omega_n = o(1)$ almost surely. Then

$$P[n^{1/2}\Lambda(S_n) \le z - \omega_n] - P[n^{1/2}\Lambda(S_n) \le z] - P(\|\Delta_n\| > \omega_n)$$

$$\le P(t \le z] - P[n^{1/2}\Lambda(S_n) \le z]$$

$$\leq \quad P[n^{1/2}\Lambda(S_n) \leq z + \omega_n] - P[n^{1/2}\Lambda(S_n) \leq z] + P(\|\Delta_n\| > \omega_n)$$

for every z. Therefore, since $\Delta_n = o[(nh_n)^{-1}]$ and $\Delta_n/\omega_n = o(1)$ almost surely,

$$\text{(A14)} \qquad P(t \leq z) - P[n^{1/2}\Lambda(S_n) \leq z] \;=\; o[(nh_n)^{-1}].$$

uniformly over z. The proposition follows by multiplying both sides of (A14) by nh_n and taking the limit as $n \to \infty$. Q.E.D.

Let E_n denote the expectation with respect to P_n^*. Define $G_n^*(b)$ by replacing (Y_i,X_i) with (Y_i^*,X_i^*) in the definition of $G_n(b)$.

Lemma 6: For any $b \in B$, define $U_b = Y - Xb$ and

$$W_n(b) \;=\; n^{-1}\sum_{i=1}^{n}[(U_{bi}^*/h_n)^d g(X_i^*)f(U_{bi}^*/h_n) - E_n(U_b/h_n)^d g(X)f(U_b/h_n)],$$

where g is bounded for bounded values of its argument, d = 0 or 1, and f is a bounded, Lipschitz continuous function of bounded variation with support [-1,1]. (a) Define $\xi_n = [(h_n/n)\log n]^{1/2}$. There is a finite $C_0 > 0$ such that for all $C > C_0$ and any $\gamma \geq 0$

$$\lim_{n \to \infty} (nh_n)^{\gamma} P_n^*(\sup_{b \in B} |W_n(b)| > C\xi_n) \;=\; 0$$

almost surely (P).

(b) Define $\xi_n = [(\log n)/n]^{1/2}$. There is a finite $C_0 > 0$ such that for all $C > C_0$ and any $\gamma \geq 0$

$$\lim_{n \to \infty} (nh_n)^{\gamma} P_n^*(\sup_{b \in B} |G_{ni}^*(b) - E_n G_{ni}^*(b)| > C\xi_n) \;=\; 0$$

and

$$\lim_{n \to \infty} (nh_n)^{\gamma} P_n^*(\sup_{b \in B} |T_n^*(b) - E_n T_n^*(b)| > C\xi_n) \;=\; 0$$

almost surely (P).

(c) For any $\gamma \geq 0$ and $\eta > 0$,

$$\lim_{n \to \infty} (nh_n)^{\gamma} P_n^*(\sup_{b \in B} |G_n^*(b) - G_n(b)| > \eta) = 0$$

Proof: Only part (a) is proved. The proofs of parts (b) and (c) are similar. Partition B into subsets $\{B_j: j = 1,...,J\}$ such that $\|b_1 - b_2\| < \xi_n^2$ whenever b_1 and b_2 are in the same subset. For each $j = 1,...,J$, let b_j be a point in B_j. Observe that $J = O(\xi_n^{-2q})$. Then

$$P_n^*(\sup_{b \in B} |W_n(b)| > C\xi_n) = P_n^*(\bigcup_{j=1}^{J} \sup_{b \in B_j} |W_n(b)| > C\xi_n)$$

$$\text{(A15)} \qquad \leq \sum_{j=1}^{J} P_n^*(\sup_{b \in B_j} |W_n(b)| > C\xi_n).$$

Because g is bounded, X has bounded support, and f is bounded and Lipschitz continuous, there is an $M < \infty$ such that

$$\sup_{b \in B_j} |W_n(b)| \leq 2M(\log n)/n + |W_n(b_j)|$$

Therefore, for all sufficiently large n

$$\text{(A16)} \qquad P_n^*(\sup_{b \in B_j} |W_n(b)| > C\xi_n) \leq P_n^*(|W_n(b_j)| > C\xi_n/2)$$

By using Lemma 22 of Nolan and Pollard (1987) and Theorem 2.37 of Pollard (1984), it can be shown that $E_n[nW_n(b_j)^2] \leq c_1 h_n$ almost surely (P) for some $c_1 < \infty$ and all sufficiently large n. Therefore, by Bernstein's inequality

$$\text{(A17)} \qquad P_n^*(|W_n(b_j)| > C\xi_n/2) \leq 2\exp(-Cd\log n) = 2n^{-Cd}$$

for some finite $d > 0$ and all sufficiently large n. Combining (A15)-(A17) yields

$$(nh_n)^{\gamma}P_n^*(\sup_{b \in B}|W_n(b)| > C\xi_n) \;\le\; 2(nh_n)^{\gamma} n^{-Cd} O(\xi_n^{-2q}) \;=\; o(1)$$

as $n \to \infty$ for all sufficiently large C. Q.E.D.

The following lemma gives the bootstrap version of Lemma 2.

<u>Lemma 7</u>: For any $\gamma > 0$ and $\epsilon > 0$

$$\lim_{n \to \infty} (nh_n)^{\gamma}P_n^*(\|b_n^* - b_n\| > \epsilon) \;=\; 0.$$

almost surely (P).

<u>Proof</u>: Given any $\eta > 0$, suppose that $|G_n^*(b) - G_n(b)| \le \eta$ and $|G_n(b) - G(b)| \le \eta$ for all $b \in B$. Then since b_n^* minimizes G_n^*, $G_n(b_n) + \eta \ge G_n^*(b_n) \ge G_n^*(b_n^*)$. Also, $G_n^*(b_n^*) \ge G_n(b_n^*) - \eta$, so $G_n(b_n) + \eta \ge G_n^*(b_n^*) \ge G_n(b_n^*) - \eta$, and $G_n(b_n) - G_n(b_n^*) \ge -2\eta$. By a similar argument, $G(\beta) - G(b_n) \ge -2\eta$. Therefore, $G(\beta) - G(b_n^*) = [G(\beta) - G(b_n)] + [G(b_n) - G_n(b_n)] + [G_n(b_n) - G_n(b_n^*)] + [G_n(b_n^*) - G(b_n^*)] \ge -6\eta$. Because $G(b)$ is continuous on B with a unique minimum at β, it is possible to choose η such that $G(\beta) - G(b_n^*) \ge -6\eta$ implies $\|b_n^* - \beta\| \le \epsilon/2$. By Lemma 2 and the triangle inequality, $\|b_n^* - \beta\| \le \epsilon/2$ implies that $\|b_n^* - b_n\| \le \epsilon$ for all sufficiently large n almost surely. Therefore, $|G_n^*(b) - G_n(b)| \le \eta$ and $|G_n(b) - G(b)| \le \eta$ for all $b \in B$ imply that $\|b_n^* - b_n\| \le \epsilon$ for all sufficiently large n almost surely. The lemma follows by combining this result with Lemmas 1 and 6(c). Q.E.D.

For $i,j,k,\ell = 1,...,q$, define $G_{ni}^*(b) = \partial G_n^*(b)/\partial b_i$, $G_{nij}^*(b) = \partial^2 G_n^*(b)/\partial b_i \partial b_j$, $G_{nijk}^*(b) = \partial^3 G_n^*(b)/\partial b_i \partial b_j \partial b_k$, $G_{nijk\ell}^*(b) = \partial^4 G_n^*(b)/\partial b_i \partial b_j \partial b_k \partial b_\ell$, $D_{ni}^*(b) = \partial D_n^*(b)/\partial b_i$, $D_{nij}^*(b) = \partial^2 D_n^*(b)/\partial b_i \partial b_j$, and $T_{ni}^*(b) = \partial T_n^*(b)/\partial b_i$. The bootstrap version of Lemma 3 is:

<u>Lemma 8</u>: For all $i,j,k,\ell = 1,...,q$, any $\gamma > 0$, and all sufficiently large $C > 0$, $\lim_{n \to \infty}(nh_n)^{\gamma}P_n^*(A_n) = 0$ almost surely (P), where A_n is any of:

(a) $\sup_{b \in B} |G_{ni}*(b) - E_n G_{ni}*(b)| > C[(\log n)/n^{1/2}]$

(b) $\sup_{b \in B} |G_{nij}*(b) - E_n G_{nij}*(b)| > C[(\log n)/(nh_n)^{1/2}]$

(c) $\sup_{b \in B} |G_{nijk}*(b) - E_n G_{nijk}*(b)| > C[(\log n)/(nh_n^3)^{1/2}]$

(d) $\sup_{b \in B} |G_{nijk\ell}*(b) - E_n G_{nijk\ell}*(b)| > C[(\log n)/(nh_n^5)^{1/2}]$

(e) $\sup_{b \in B} |D_n*(b) - E_n D_n*(b)| > C[(\log n)/(nh_n)^{1/2}]$

(f) $\sup_{b \in B} |D_{ni}*(b) - E_n D_{ni}*(b)| > C[(\log n)/(nh_n^3)^{1/2}]$

(g) $\sup_{b \in B} |D_{nij}*(b) - E_n D_{nij}*(b)| > C[(\log n)/(nh_n^5)^{1/2}]$

(h) $\sup_{b \in B} |T_n*(b) - E_n T_n*(b)| > C[(\log n)/n^{1/2}]$

(i) $\sup_{b \in B} |T_{ni}(b) - ET_{ni}(b)| = o[(\log n)/(nh_n)^{1/2}]$,

and (e)-(i) apply to the individual components of the matrices D_n*, $D_{ni}*$, $D_{nij}*$, T_n*, and $T_{ni}*$. In addition, for all $i,j,k,\ell = 1,..., q$

(j) $E_n G_{ni}*(b_n) = 0$ with probability $1 - o[(nh_n)^{-\gamma}]$.

(k) $E_n G_{nij}*(b)$, $E_n G_{nijk}*(b)$, $E_n G_{nijk\ell}*(b)$, $E_n D_n*(b)$, $E_n D_{ni}*(b)$, $E_n D_{nij}*(b)$, $E_n T_n*(b)$, and $E_n T_{ni}*(b)$ are $O(1)$ almost surely (P) as $n \to \infty$ for all b in a neighborhood of β.

<u>Proof</u>: Parts (a)-(i) are immediate consequences of Lemma 6. Part (j) is the first-order condition for the bootstrap estimation problem. Part (k) follows from Lemma 3. Q.E.D.

Define $S_{nG}*$ and S_n* as S_{nG} and S_n except with (Y_i,X_i) replaced by (Y_i*,X_i*) and β replaced by b_n.

<u>Proposition 2</u>: Let Λ be the function defined in Proposition 1.

$$\lim_{n \to \infty} \sup_z (nh_n)\{P_n*(t* \le z) - P_n*[n^{1/2}\Lambda(S_n*) \le z]\} = 0$$

almost surely (P).

<u>Proof</u>: This is the bootstrap version of Proposition 1. It is proved using the same arguments that are used to prove Lemmas 4-5 and Proposition 1 but with S_{nG}, S_n, b_n, and β, respectively, replaced by $S_{nG}*$, S_n*, b_n*, and b_n.

Q.E.D.

b. Step 2: Asymptotic Expansions

For h > 0, let W(u,x,h) be a vector whose components are terms of the form $g(x)_j(u/h)$, where g(x) is the product of (not necessarily distinct) components of x that may be different in each use of g, and $_j$ is the j'th component of the vector defined in assumption 5. The following lemma gives a modified version of the Cramer condition of Edgeworth analysis.

<u>Lemma 9</u>: Let τ be a vector with the same dimension as W. Define $\psi_W(\tau,h) = E\{\exp[\iota\tau' W(X,U,h)]\}$ where $\iota = (-1)^{1/2}$. For any $\epsilon > 0$, some $C > 0$, all τ satisfying $\|\tau\| > \epsilon$, and all sufficiently small h

$$|\psi_W(\tau,h)| < 1 - Ch.$$

<u>Proof</u>: Let r index components of and W. Each component of satisfies $|\,_r(v)| = 0$ or 1 if $|v| \geq 1$. Let $\delta_r^- = \,_r(v)$ if $v \leq -1$ and $\delta_r^+ = \,_r(v)$ if $v \geq 1$. Then using the summation convention

$$\psi_W(\tau,h) = \int \exp[\iota\tau_r g_r(x)\,_r(u/h)]f(u|x)dudP(x)$$

$$= \int\left\{\int_{-\infty}^{-h} \exp[\iota\tau_r g_r(x)\delta_r^-]f(u|x)du + \int_h^{\infty} \exp[\iota\tau_r g_r(x)\delta_r^+]f(u|x)du\right.$$

$$\left. + \int_{-h}^{h} \exp[\iota\tau_r g_r(x)\,_r(u/h)]f(u|x)du\right\}dP(x)$$

$$= A_1(h) + A_2(h),$$

where

$$A_1(h) = E\{F(-h|X)\exp[\iota\tau_r g_r(X)\delta_r^-] + [1 - F(h|X)]\exp[\iota\tau_r g_r(X)\delta_r^+]\},$$

and

$$A_2(h) = \int\left\{\int_{-h}^{h} \exp[\iota\tau\, g_r(x)\, {}_r(u/h)]f(u|x)du\right\}dP(x).$$

Consider $A_1(h)$. $|A_1(h)| \leq E|A_1(h,X)|$, where

$$A_1(h,x) = F(-h|X)\exp[\iota\tau\, g_r(X)\delta_r^-] + [1 - F(h|X)]\exp[\iota\tau\, g_r(X)\delta_r^+].$$

Let ${}_r = \delta_r^+$ if $\delta_r^+ = -\delta_r^- = 1$. Note that $\delta_r^+ = \delta_r^-$ otherwise. Therefore,

$$|A_1(h,x)| = |F(-h|X)\exp[\iota\tau\, g_r(X)\, {}_r] + [1 - F(h|X)]\exp[-\iota\tau\, g_r(X)\, {}_r]|.$$

$$= \{[1 - F(h|x) + F(-h|x)]^2$$
$$- 4[1 - F(h|x)]F(-h|x)\sin^2[\tau\, g_r(x)\, {}_r]\}^{1/2}$$

$$\leq 1 - F(h|x) + F(-h|x)$$

$$= 1 - 2hf(0|x) - (1/2)h^2[f^{(1)}(h_1|x) - f^{(1)}(h_2|x)],$$

where h_1 and h_2 are between 0 and h, and the last line is obtained by a Taylor series expansion. Let $Ef(0|X) = C_1$. By assumption 4(b), $C_1 > 0$ and $E|f^{(1)}(h_1|X) - f^{(1)}(h_2|X)| < M$ for some finite M and all sufficiently small h. Therefore,

$$|A_1(h)| \leq E|A_1(h,X)| \leq 1 - C_1 h$$

for all sufficiently small h. Now consider $A_2(h)$. By a change of variables

$$A_2(h) = h\int\left\{\int_{-1}^{1} \exp[\iota\tau\, g_r(x)\, {}_r(\zeta)]f(h\zeta|x)d\zeta\right\}dP(x).$$

Given $\epsilon > 0$, choose h sufficiently small that

$$\int_{-1}^{1}|f(h\zeta|x) - f(0|x)|d\zeta dP(x) \le \epsilon\left|\int_{-1}^{1} f(0|x)d\zeta dP(x) = 2\epsilon C_1 .\right.$$

Then

$$\text{(A18)} \qquad |\psi_W(\tau,h)| \le 1 - hC_1(1 - 2\epsilon) + |A_3(\tau,h)|$$

for all τ, $\epsilon > 0$, and sufficiently small $h > 0$, where

$$A_3(\tau,h) = h\left\{\int_{-1}^{1} \exp[\iota\tau_r g_r(x)_r(\zeta)]f(0|x)d\zeta\right\}dP(x).$$

Since $g_r(x) = 0$ for every r only if $x = 0$ and $P(X = 0) < 1$, there are $\eta > 0$ and $\gamma_1 < 1$ such that

$$2\int_{\|x\| < \eta} f(0|x)dP(x) = \gamma_1 C_1 .$$

Suppose, as will be proved presently, that for some $C_2 < 1$,

$$\text{(A19)} \qquad \sup_{\|\tau\| \ge \epsilon} \int_{-1}^{1} \left| \exp[\iota\tau_r g_r(x)_r(\zeta)]d\zeta\right| = C_2$$

uniformly over x such that $\|x\| \ge \eta$. Then for $\|\tau\| \ge \epsilon$

$$\text{(A20)} \qquad |A_3(\tau,h)| \le h[\gamma_1 C_1 + (1 - \gamma_1)C_1 C_2] = h\gamma_2 C_1$$

where $\gamma_2 = [\gamma_1 + (1 - \gamma_1)C_2] < 1$. Combining (A18) with (A20) yields

$$\sup_{\|\tau\| > \epsilon} |\psi_W(\tau,h)| \le 1 - hC_1(1 - 2\epsilon - \gamma_2) \equiv 1 - Ch$$

for all sufficiently small $h > 0$ and $\epsilon > 0$, thereby establishing the lemma.

It remains to prove (A19). To do this, define $t = \|\tau\|$. Fix $\tau/\|\tau\|$ and x with $\|x\| \neq 0$. For the specified $\tau/\|\tau\|$ and x, and using the summation convention, define $f(\zeta) = \tau_r g_r(x)_r(\zeta)/\|\tau\|$. Let $-1 = a_\ell < ... < a_L = 1$ be a partition of [-1,1] that satisfies assumption 5c when $\theta_r = g_r(x)$. Then

$$\psi^*(\tau) \equiv \int_{-1}^{1} \exp[\iota t f(\zeta)]d\zeta = \sum_{\ell=2}^{L} \int_{a_{\ell-1}}^{a_L} \exp[\iota t f(\zeta)]d\zeta$$

It suffices to prove that for any $\epsilon > 0$ and some $C_3 < 1$ that does not depend on x or $\tau/\|\tau\|$

$$\text{(A21)} \qquad \sup_{|t| > \epsilon} (a_\ell - a_{\ell-1})^{-1} \left| \int_{a_{\ell-1}}^{a_\ell} \exp[\iota t f(\zeta)]d\zeta \right| \leq C_3.$$

To do this, make the change of variables $\xi = f(\zeta)$ in (A21) and set $v(\xi) = 1/\{df[\zeta(\xi)]/d\zeta\}$. Then

$$\psi^{**}(t) \equiv \int_{a_{\ell-1}}^{a_\ell} \exp[\iota t f(\zeta)]d\zeta = \int_{f(a_{\ell-1})}^{f(a_\ell)} e^{\iota t\xi} v(\xi)d\xi.$$

Observe that $|\psi^{**}(t)| \leq a_\ell - a_{\ell-1}$, so the right-hand integral is bounded. The right-hand integral can be approximated arbitrarily accurately by replacing $v(\bullet)$ with a step function. Therefore, it is enough to prove that

$$\sup_{|t| > \epsilon} \left| \int_{\alpha_1}^{\alpha_2} e^{\iota t\xi} d\xi \right| \leq (\alpha_2 - \alpha_1)C_3$$

for all $\alpha_1 < \alpha_2$ and some $C_3 < 1$ that does not depend on α_1 or α_2. But

$$\left| \int_{\alpha_1}^{\alpha_2} e^{\iota t\xi} d\xi \right| \leq (\alpha_2 - \alpha_1)\frac{\sin^2[0.5t(\alpha_2 - \alpha_1)]}{[0.5t(\alpha_2 - \alpha_1)]^2}.$$

The proof is completed by setting $C_3 = \inf_{|t| > \epsilon} [(\sin^2 t)/t]$. Q.E.D.

Define W^* as in Lemma 9 except with β replaced by b_n. Define $\psi_W^*(\tau,h_n) = E_n\{\exp[\iota\tau'W^*(U,X,h_n)]\}$. The bootstrap version of Lemma 9 is:

Lemma 10: For any $\epsilon > 0$ and $c > 0$, some $C^* > 0$, all τ satisfying $\epsilon < \|\tau\| \leq n^c$, and all sufficiently large n

$$|\psi_W^*(\tau,h_n)| < 1 - C^*h_n$$

almost surely (P).

Proof: Let $B_{n\tau} = \{\tau\colon \epsilon < \|\tau\| \leq n^c\}$. Then

$$\sup_{\|\tau\| \in B_{n\tau}} |\psi_W^*(\tau,h_n)| \leq \sup_{\|\tau\| \in B_{n\tau}} |\psi_W(\tau,h_n)| + \sup_{\|\tau\| \in B_{n\tau}} |\psi_W^*(\tau,h_n) - \psi_W(\tau,h_n)|.$$

By arguments similar to those used to prove Lemma 6 together with the Borel-Cantelli lemma, $|\psi_W^*(\tau,h_n) - \psi_W(\tau,h_n)| = o(h_n)$ almost surely uniformly over $\tau \in B_{n\tau}$. Let C be as in Lemma 9. Then Lemma 10 follows by letting C^* be any number such that $C < C^* < 1$. Q.E.D.

Let W_{n1} be a column-vector consisting of the unique components of $n^{1/2}[G_{ni}(\beta) - EG_{ni}(\beta)]$ $(i = 1,\ldots, q)$ and $n^{1/2}[T_n(\beta) - ET_n(\beta)]$. Let W_{n2} be a column-vector consisting of the unique components of $(nh_n)^{1/2}[G_{nij}(\beta) - EG_{nij}(\beta)]$, $(nh_n^3)^{1/2}[G_{nijk}(\beta) - EG_{nijk}(\beta]$, $(nh_n^5)^{1/2}[G_{nijk\ell}(\beta) - EG_{nijk\ell}(\beta)]$, $(nh_n)^{1/2}[D_n(\beta) - ED_n(\beta)]$, $(nh_n^3)^{1/2}[D_{ni}(\beta) - ED_{ni}(\beta)]$, $(nh_n^5)^{1/2}[D_{nij}(\beta) - ED_{nij}(\beta)]$, and $(nh_n)^{1/2}[T_{ni}(\beta) - ET_{ni}(\beta)]$ $(i,j,k,\ell = 1,\ldots, q)$. Set $W_n = [W_{n1}', W_{n2}']'$. Define W_n^*, W_{n1}^*, and W_{n2}^* similarly except with (Y_i,X_i) replaced by (Y_i^*,X_i^*) and β replaced by b_n. Order the components of S_n and S_n^* conformably with those of W_n and W_n^*. Let V_n be the covariance matrix of $[W_{n1}', W_{n2}'/h_n]'$ and V_n^* be the covariance matrix of $[W_{n1}^{*\prime}, W_{n2}^{*\prime}/h_n]'$ relative to P_n^*. Let w_{n1}, w_{n2}, w_{n1}^*, and w_{n2}^*, respectively, be the summands of the components of W_{n1}, W_{n2}, W_{n1}^*, and W_{n2}^*. These have the forms

$g_j(X)_j(U/h)$ and $g_j(X)_j(U_n/h)$, where $U_n = Y - Xb_n$. For any $\tau = (\tau_1', \tau_2')'$ conformable with $(w_{n1}', w_{n2}')'$, define

(A22) $$\rho_1(\tau) = -[\iota/(6h_n)]E[(\tau_2'w_{n2})^3],$$

(A23) $$\rho_2(\tau) = -(\iota/6))\{E[(\tau_1'w_{n1})^3] + (3/h_n)E[(\tau_1'w_{n1})(\tau_2'w_{n2})^2]\},$$

(A24) $$\rho_3(\tau) = -[\iota/(2h_n)]E[(\tau_2'w_{n2})^3],$$

and

(A25) $$\rho_4(\tau) = [1/(24h_n)]E[(\tau_2'w_{n2})^4] + (1/72)\{h_n^{-1}E[(\tau_2'w_{n2})^3]\}^2.$$

Define $\rho_i^*(\tau)$ ($i = 1, \ldots, 4$) by replacing w_n with w_n^* and E with E_n in (A22)-(A25). Let π_i ($i = 1, \ldots, 4$) be the signed measures whose Fourier-Stieljes transforms are

(A26) $$\int \exp(i\tau'\xi)d\pi_i(\xi) = \exp(-0.5\tau'V_n\tau)\rho_i(\tau).$$

Define π_i^* ($i = 1, \ldots, 4$) analogously by using V_n^* and ρ_i^* in place of V_n and ρ_i. Let $d_W = \dim(W_n)$. For any set α in d_W-dimensional Euclidean space, let $\partial\alpha$ denote the boundary of α and $(\partial\alpha)^\epsilon$ denote the set of all points whose distance from $\partial\alpha$ does not exceed ϵ. Let Φ_{V_n} denote probability measure according to the normal distribution with mean 0 and covariance matrix V_n. Define $\Phi_{V_n^*}$ analogously.

Lemma 11: Let A denote a class of Borel sets in d_W-dimensional Euclidean space that satisfy

$$\sup_{\alpha \in A} \int_{(\partial\alpha)^\epsilon} \exp(-0.5\|\xi\|^2)d\xi = O(\epsilon)$$

as $\epsilon \to 0^+$. Then

$$\sup_{\alpha \in A} |P(W_n \in \alpha) - \Phi_{V_n}(\alpha) - (nh_n)^{-1/2}\pi_1(\alpha) - n^{-1/2}\pi_2(\alpha)$$

$$- (h_n/n)^{1/2}\pi_3(\alpha) - (nh_n)^{-1}\pi_4(\alpha)| = o[(nh_n)^{-1}],$$

and almost surely (P)

$$\sup_{\alpha \in A} |P_n^*(W_n^* \in \alpha) - \Phi_{V_n^*}(\alpha) - (nh_n)^{-1/2}\pi_1^*(\alpha) - n^{-1/2}\pi_2^*(\alpha)$$

$$- (h_n/n)^{1/2}\pi_3^*(\alpha) - (nh_n)^{-1}\pi_4^*(\alpha)| = o[(nh_n)^{-1}],$$

Proof: This is a slightly modified version of Theorem 5.8 of Hall (1992) and is proved using the same arguments as in Hall's proof after replacing Hall's Lemma 5.6 with Lemmas 9 and 10 above. Q.E.D.

Proof of Theorem 4.1: Only parts (a), (c) and the part of (b) pertaining to $q(\tau,\nu_n)$ are proved here. The proofs of the remaining parts are similar. To begin, invert (A26) to obtain

(A27) $$\pi_i(\xi) = {}_{ni}(\xi)\phi_{V_n}(\xi),$$

where for each n and i, ${}_{ni}(\bullet)$ is a multivariate polynomial, and ϕ_{V_n} is the multivariate normal density with mean 0 and covariance matrix V_n. Let $S_n(W_n)$ be the mapping from W_n to S_n. Define $(W_n) = (nh_n)^{1/2}\Lambda[S_n(W_n)]$. By Proposition 1 it suffices to consider $P(\le \tau)$. Define $a_1 = (nh_n)^{-1/2}$, $a_2 = n^{-1/2}$, $a_3 = (h_n/n)^{1/2}$, and $a_4 = (nh_n)^{-1}$. By Lemma 11 and (A27)

$$P(\le \tau) = |\int_{\{\xi: (\xi) \le \tau\}} d[\Phi_{V_n}(\xi) + \sum_{i=1}^{4} a_i\, {}_{ni}(\xi)\phi_{V_n}(\xi)]$$

(A28) $$+ o[(nh_n)^{-1}].$$

uniformly over τ. Order the components of ξ and W_n so that the first components correspond with $(nh_n)^{1/2}[G_{ni}(\beta) - EG_{ni}(\beta)]$, where i is the component of β for which t is the t statistic. Let denote the vector consisting of all components of ξ except the first, ξ_1. Change variables in the integral of (A28) so that the variable of integration is (,')', thereby obtaining

$$P(\leq \tau) = \int\limits_{\leq \tau} d \, d J[\xi_1(,)]\{\Phi[\xi_1(,)/\sqrt{n}]$$

$$\text{(A29)} \qquad + \sum_{i=1}^{4} a_{ni}[\xi_1(,)]\phi[\xi_1(,)/\sqrt{n}]\} + o[(nh_n)^{-1}]$$

uniformly over τ, where $J(\bullet)$ is the inverse Jacobian term associated with the change of variables. Taylor series expansions in powers of n^{-1} of the terms involving $\xi_1(,)$ in (A29) yield

$$P(\leq \tau) = \Phi(\tau) + \sum_{i=1}^{5} c_{ni} {}_{i}(\tau)\phi(\tau) + o[(nh_n)^{-1}]$$

$$\text{(A30)} \qquad \equiv G_n(\tau) + o[(nh_n)^{-1}]$$

uniformly over τ, where Φ and ϕ, respectively, are the univariate standard normal distribution and density functions, the ${}_i$'s are polynomial functions of one variable, $c_{n1} = n^{-1/2}$, $c_{n2} = (nh_n)^{-1/2}$, $c_{n3} = h_n n^{-1/2}$, $c_{n4} = (nh_n^{3/2})^{-1}$, and $c_{n5} = (nh_n)^{-1}$. Let ψ and ψ_G, respectively, denote the characteristic functions of the distributions of and G_n. Then $|\psi(\tau) - \psi_G(\tau)| = o[(nh_n)^{-1}]$. A Taylor series expansion shows that in (A30) can be replaced by a multivariate polynomial in components of $S_n - E(S_n)$. The cumulants through order 4 of this polynomial may be approximated through $O[(nh_n)^{-1}]$ using standard Taylor series methods of kernel estimation. Let k_{nj} denote the approximate j'th cumulant. Expressing ψ in terms of the approximate cumulants yields $\psi(\tau) = (\tau) + o[(nh_n)^{-1}]$ uniformly over τ, where

$$(\tau) \;=\; [\exp(-\tau^2/2)]\{1 + i\tau k_{n1} + (1/2)(i\tau)^2(k_{n2} - 1) + (1/6)(i\tau)^3 k_{n3}$$

$$\text{(A31)} \qquad + (1/24)(i\tau)^4 k_{n4} + (1/2)[(i\tau)k_{n1} + (1/6)(i\tau)^3 k_{n3}]^2\}.$$

Setting ψ_G = , taking the inverse Fourier transform of the result, and setting $P(|| \leq \tau) = P(\leq \tau) - P(\leq -\tau)$ yields (4.1) with

$$q(\nu_n, \tau) \;=\; -\tau[k_{n1}^2 + (k_{n2} - 1) + (1/12)(4k_{n1}k_{n3} + k_{n4})(\tau^2 - 3)$$

$$+ (1/36)k_{n3}^2(\tau^4 - 10\tau^2 + 15)].$$

A straightforward but lengthy calculation shows that k_{n1}^2, $k_{n1}k_{n3}$, and k_{n3}^2 are $o[(nh_n)^{-1}]$, whereas $k_{n2} - 1$ and k_{n4} are $O[(nh_n)^{-1}]$ and consist of linear combinations of the terms shown in Table I. Q.E.D.

Proof of Theorem 4.2: Under H_0, $c = R\beta$, so

$$\chi^2 \;=\; (nh_n)(b_n - \beta)'R'(RV_nR')^{-1}R(b_n - \beta)$$

By arguments similar to those used to prove Propositions 1 and 2 followed by a Taylor series expansion, there is a multivariate polynomial Λ_χ such that

$$P(\chi^2 \leq z) - P[(nh_n)\Lambda_\chi(S_n) \leq z] \;=\; o[(nh_n)^{-1}].$$

uniformly over z and

$$\lim_{n \to \infty} \sup_z (nh_n)\{P_n^*(\chi^{2*} \leq z) - P_n^*[(nh_n)\Lambda_\chi(S_n^*) \leq z]\} \;=\; 0$$

almost surely (P). Set $(W_n) = (nh_n)\Lambda_\chi[S_n(W_n)]$. By arguments similar to those used to obtain (A28),

$$P(\leq \tau) \;=\; \left|\int_{\{\xi:\ (\xi) \leq z\}} d[\Phi_{V_n}(\xi) + \sum_{i=1}^{4} a_{i\,n1}(\xi)\phi_{V_n}(\xi)]\right.$$

$$+ o[(nh_n)^{-1}].$$

Now transform to polar coordinates and proceed as in the proof of Theorem 1b of Chandra and Ghosh (1979). A similar argument applies to $P(\chi^{2*} < z)$. Q.E.D.

<u>Proof of Theorem 4.3</u>: Only part (a) is proved here. The proof of part (b) is similar. Let t_α and t_α^*, respectively, denote the exact and bootstrap α-level critical values of the symmetrical t test. Let k_{ni}^* denote the bootstrap version of k_{ni} (i = 2 or 4). This is obtained from k_{ni} by replacing β with b_n and expected values with sample averages. By Theorem (4.1),

$$|P(|t| > t_\alpha^*) - \alpha| \leq \sup_\tau |P(|t| > \tau) - P_n^*(|t^*| > \tau)|$$

$$\leq \sup_\tau |[q(\tau,\nu_n) - q(\tau,\nu_n^*)]\phi(\tau)| + o[(nh_n)^{-1}]$$

$$= O(k_{n2}^* - k_{n2}) + O(k_{n4}^* - k_{n4}).$$

The proof is completed by using methods similar to those used in proving Lemma 3 to show that the difference between each of the terms in Table I and its bootstrap analog is $o[(nh_n)^{-1}]$ almost surely. Q.E.D.

<u>Proof of Theorem 4.4</u>: The proof consists of repeating each step of the proofs of Lemmas 1-11 and Theorems 4.1-4.3 with $H_{cn}(b)$ in place of $H_n(b)$ and Assumptions 1' and 3' in place of 1 and 3.

FOOTNOTES

1. When the t statistic can be approximated by a smooth function of sample moments, the difference between the true and nominal levels of a symmetrical t test with bootstrap critical values is typically $O(n^{-2})$. With critical values based on first-order asymptotic theory, the difference is typically $O(n^{-1})$. See, e.g., Hall (1992). The larger approximation errors in the case of a t statistic for a median are due to the median estimator's non-smooth objective function.

2. De Angelis, *et al*. (1993) implement the bootstrap by sampling smoothed LAD residuals. In contrast to sampling (Y,X) pairs, this method does not easily generalize to heteroskedastic or censored models.

3. K does not satisfy assumption 5b because it has only two derivatives at $v = \pm 1$. This problem can be overcome by smoothing K in neighborhoods of $v = \pm 1$, but doing so has no effect on the results of the experiments.

4. Hall and Horowitz (1990) derived the bandwidth that minimizes the asymptotic mean-square error of the variance estimator in a homoskedastic quantile regression. They suggested a plug-in estimator for this bandwidth. However, the bandwidth that optimizes the variance estimate is not necessarily optimal for computing test statistics, and little is known about the numerical performance of the Hall-Horowitz estimator in testing.

REFERENCES

Beran, R. (1988). Prepivoting test statistics: a bootstrap view of asymptotic refinements, *Journal of the American Statistical Association*, 83, 687-697.

Bhattacharya, R.N. and J.K. Ghosh (1978). On the validity of the formal Edgeworth expansion, *Annals of Statistics*, 6, 434-451.

Bloomfield, P. and W.L. Steiger (1983). *Least Absolute Deviations: Theory, Applications, and Algorithms*, Boston: Birkhauser.

Buchinsky, M. (1995). Estimating the asymptotic covariance matrix for quantile regression models: a Monte Carlo study, *Journal of Econometrics*, 68, 303-338.

Bassett, G. and R. Koenker (1978). Asymptotic theory of least absolute error regression, *Journal of the American Statistical Association*, 73, 618-621.

De Angelis, D., P. Hall, and G.A. Young (1993). Analytical and bootstrap approximations to estimator distributions in L^1 regressions, *Journal of the American Statistical Association*, 88, 1310-1316.

Dielman, T. and R. Pfaffenberger (1984). Tests of linear hypotheses and L_1 estimation: a Monte Carlo comparison, *American Statistical Association Business and Economic Statistics Section Proceedings*, 644-647.

Dielman, T. and R. Pfaffenberger (1988a). Bootstrapping in least absolute value regression: an application to hypothesis testing, *Communications in Statistics-Simulation and Computation*, 17, 843-856.

Dielman, T. and R. Pfaffenberger (1988b). Least absolute value regression: necessary samle sizes to use norma theory inference procedures, *Decision Sciences*, 19, 734-743.

Chandra, T.K. and J.K. Ghosh (1979). Valid asymptotic expansions for the likelihood ratio statistic and other perturbed chi-square variables, *Sankhya*, 41, Series A, 22-47.

Hahn, J. (1995). Bootstrapping quantile regression estimators, *Econometric Theory*, 11, 105-121.

Hall, P. (1986). On the bootstrap and confidence intervals, *Annals of Statistics*, 1431-1452.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.

Hall, P. and J.L. Horowitz (1990). Bandwidth selection in semiparametric estimation of censored linear regression models, *Econometric Theory*, 6, 123-150.

Horowitz, J.L. (1996). Bootstrap methods in econometrics: theory and numerical performance, in *Advances in Economics and Econometrics: 7th World Congress*, D. Kreps and K.W. Wallis, eds., Cambridge: Cambridge University Press, forthcoming.

Janas (1993). A smoothed bootstrap estimator for a studentized sample quantile, *Annals of the Institute of Statistical Mathematics*, 45, 317-329.

Koenker, R. (1982). Robust methods in econometrics, *Econometric Reviews*, 1, 213-255.

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

Koenker, R. and G. Bassett (1982). Robust tests for heteroscedasticity based on regression quantiles, *Econometrica*, 50, 43-61.

Müller, H.-G. (1984). Smooth optimum kernel estimators of densities, regression curves and modes, *Annals of Statistics*, 12, 766-774.

Nolan, D. and Pollard, D. (1987). U-processes: rates of convergence, *Annals of Statistics*, 15, 780-799.

Pakes, A. and Pollard, D. (1989). Simulation and the asymptotics of optimization estimators, *Econometrica*, 57, 1027-1057.

Pollard, D. (1984). *Convergence of Stochastic Processes*, New York: Springer-Verlag.

Powell, J.L. (1984) Least absolute deviations estimation for the censored regression model, *Journal of Econometrics*, 25, 303-325.

Powell, J.L. (1986). Censored regression quantiles, *Journal of Econometrics*, 32, 143-155.

TABLE I: TERMS OF APPROXIMATE CUMULANTS

Notation: $g_j(x)$, j an integer, is a product of components of x that may be different in different occurrences. $m_{1j}(x,u) = n^{-1/2}g_j(x)\{[2K(u/h_n) - 1] - 2(u/h_n)K^{(1)}(u/h_n)\}$. For i = 2 or 3, $m_{ij}(x,u) = \partial^{i-1}m_{1j}(x,u)/\partial u^{i-1}$. In addition, $m_{4j}(x,u) = n^{-1/2}g_j(x)K^{(1)}(u/h_n)$, $m_{5j}(x,u) = \partial m_4(x,u)/\partial u$, $\mu_{ij} = Em_{ij}(X,U)$, and $\nu_{ij} = [m_{ij}(X,U) - \mu_{ij}]$.

Cumulant	Terms
$k_{n2} - 1$	$nE(\nu_{11}\nu_{12})E(\nu_{13}\nu_{34})$, $nE(\nu_{11}\nu_{12})E(\nu_{13}\nu_{54})$,
	$nE(\nu_{11}\nu_{12})E(\nu_{23}\nu_{24})$, $nE(\nu_{11}\nu_{12})E(\nu_{23}\nu_{44})$,
	$nE(\nu_{11}\nu_{12})E(\nu_{43}\nu_{44})$
k_{n4}	$n^2E(\nu_{11}\nu_{12})E(\nu_{13}\nu_{14})E(\nu_{15}\nu_{36})$, $n^2E(\nu_{11}\nu_{12})E(\nu_{13}\nu_{14})E(\nu_{25}\nu_{26})$,
	$n^2E(\nu_{11}\nu_{12})E(\nu_{13}\nu_{14})E(\nu_{25}\nu_{46})$, $n^2E(\nu_{11}\nu_{12})E(\nu_{13}\nu_{14})E(\nu_{45}\nu_{46})$,
	$n^2E(\nu_{11}\nu_{12})E(\nu_{13}\nu_{14})E(\nu_{15}\nu_{56})$