BOOTSTRAP CRITICAL VALUES FOR TESTS BASED ON THE SMOOTHED MAXIMUM SCORE ESTIMATOR

by

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

March 1996

ABSTRACT

The smoothed maximum score estimator of the coefficient vector of a binary response model is consistent and asymptotically normal under weak distributional assumptions. However, the differences between the true and nominal levels of tests based on smoothed maximum score estimates can be very large in finite samples when first-order asymptotics are used to obtain critical values. This paper shows that the bootstrap provides finite-sample critical values that are more accurate than those obtained from first-order asymptotic theory. In a set of Monte Carlo experiments carried out to check numerical performance, the bootstrap essentially eliminates large finite-sample distortions of level that occur when asymptotic critical values are used.

Research supported in part by NSF grant SBR-9307677. Part of the research was carried out while I was visiting the Institute of Statistics and Econometrics, Humboldt University, Berlin, with support from Deutsche Forschungsgemeinschaft Sonderforschungsbereich 373, "Quantifikation und Simulation Ökonomischer Prozesse." I thank Oliver Linton and Gene Savin for comments on an earlier draft of this paper.

BOOTSTRAP CRITICAL VALUES FOR TESTS BASED ON THE SMOOTHED MAXIMUM SCORE ESTIMATOR

1. INTRODUCTION

The most frequently used form of the binary response model is

$$Y = I(\beta'X + U \geq 0), \tag{1.1}$$

where I is the indicator function, X is a q×1 vector of explanatory variables, U is an unobserved random variable, and β is a q×1 vector of parameters whose values must be estimated from observations of (Y,X')'. If the distribution of U conditional on X is known up to a finite set of parameters θ, β and θ can be estimated by maximum likelihood, among other ways. See, for example, McFadden (1974). If the distribution of U does not belong to a known parametric family, β can be estimated by semiparametric methods. The semiparametric estimator of Cosslett (1983) can be used if U and X are independent. If U and X are independent or if the distribution of U depends on X only through the index $\beta'X$, the semiparametric estimators of Han (1987), Horowitz and Härdle (1996), Klein and Spady (1993), Powell *et al*. (1989), and Sherman (1993) can be used. The maximum score estimator of Manski (1975, 1985) permits the distribution of U to depend on X in an unknown and very general way (heteroskedasticity of unknown form). It converges at the relatively slow rate of $n^{-1/3}$, however, and the limiting distribution of the centered, normalized estimator is very complicated (Cavanagh 1987, Kim and Pollard 1990).

Horowitz (1992) describes a smoothed version of Manski's maximum score estimator. The centered and normalized smoothed maximum score estimator is asymptotically normally distributed under assumptions that are somewhat stronger than Manski's but weak enough to accommodate heteroskedasticity of unknown form in the distribution of U. In addition to having a simpler limiting distribution, the smoothed maximum score estimator converges faster and, therefore, is asymptotically more efficient than Manski's estimator when the assumptions of the smoothed estimator are satisfied.

Because the centered, normalized smoothed maximum score estimator is asymptotically normally distributed, it can be used to form t and χ^2 statistics for testing hypotheses about one or more components of β. The asymptotic distributions of these statistics are standard normal or χ^2, depending on the statistic, and the asymptotic critical values of the tests are quantiles of the standard normal or χ^2 distributions. Monte Carlo evidence, however, has indicated that first-order asymptotic theory often provides poor approximations to the distributions of the test statistics with samples of the sizes encountered in applications. As a result, the true and nominal levels of the tests can differ greatly from one another when asymptotic critical values are used (Horowitz 1992).

One possible way to obtain improved finite-sample critical values for both t and χ^2 statistics is through the use of the bootstrap. Under certain conditions (see, e.g., Beran 1988; Hall 1986, 1992), the bootstrap provides a better finite-sample approximation to the distribution of a test statistic than does first-order asymptotic theory. Thus, the differences between the true and nominal levels of tests are often much smaller using bootstrap critical values than critical values obtained from first-order asymptotic theory (see Horowitz 1995 for examples). However, t and χ^2 statistics based on the smoothed maximum score estimator do not satisfy the standard regularity conditions under which the bootstrap provides asymptotic refinements. These conditions assume that the statistic in question can be approximated by a function of sample moments whose probability distribution has an Edgeworth expansion. However, t and χ^2 statistics based on the smoothed maximum score estimator depend on a bandwidth parameter, h_n, that decreases to zero as the sample size, n, approaches infinity. As a result, test statistics based on the smoothed maximum score estimator cannot be approximated as functions of sample moments, and the standard theory of the bootstrap does not apply to them.

This paper shows that despite this problem, the bootstrap provides improved finite-sample critical values for the smoothed maximum score t and χ^2 statistics under regularity conditions that are slightly stronger than those required for smoothed maximum score estimation. Specifically, conditions are given under which the levels of symmetrical t tests and χ^2 tests with bootstrap critical values have errors whose sizes are $o[(nh_n)^{-1}]$. In contrast, the sizes of the errors are $O[(nh_n)^{-1}]$ with critical values obtained from first-order asymptotic theory. Slightly modified versions of the arguments made in this paper can be used to show that the errors in the levels of one-tailed t tests with bootstrap critical values are $o[(nh_n)^{-1/2}]$, whereas the errors are $O[(nh_n)^{-1/2}]$ with critical values obtained from first-order asymptotic theory.

This paper also presents the results of a small-scale Monte Carlo investigation of the numerical performance of the bootstrap for the symmetrical t test. In the experiments that were carried out, the use of bootstrap critical values essentially eliminates large distortions of the finite-sample level of the symmetrical t test that occur with asymptotic critical values.

The remainder of the paper is organized as follows. Section 2 reviews the smoothed maximum score estimator. The material in this section forms the basis for the discussion that follows. Section 3 describes the test statistics and the procedure for obtaining bootstrap critical values. Section 4 presents theorems giving conditions under which the bootstrap provides asymptotic refinements to the levels of the symmetrical t and χ^2 tests. Section 5 presents the results of the Monte Carlo experiments, and Section 6 presents concluding comments. The proofs of theorems are given in a mathematical appendix, which is Section 7.

2. THE SMOOTHED MAXIMUM SCORE ESTIMATOR

This section describes the smoothed maximum score estimator and summarizes its asymptotic distributional properties.

Let $\{Y_i,X_i: i = 1,...,n\}$ be a random sample of $(Y,X')'$. The smoothed maximum score estimator of β solves

$$\underset{b \in B}{\text{maximize}} \quad H_n(b) \equiv n^{-1}\sum_{i=1}^{n}[2I(Y_i = 1) - 1]K(b'X_i/h_n), \tag{2.1}$$

where B is the parameter set, $\{h_n\}$ is a sequence of positive real numbers (bandwidths) that converges to zero as $n \to \infty$, and K is a function satisfying $|K(v)| < M$ for all v and some $M < \infty$, $\lim_{v \to -\infty} K(v) = 0$, and $\lim_{v \to \infty} K(v) = 1$. K is analogous to a cumulative distribution function or the integral of a kernel function for nonparametric estimation. K is not a kernel function itself.

The parameter β in (1.1) is identified only up to scale, so a scale normalization is needed to make estimation possible. The scale normalization used here is obtained as follows. Identification of β up to scale requires that X have at least one component whose probability distribution conditional on the remaining components is absolutely continuous with respect to Lebesgue measure (Manski 1975, 1985). Arrange the components of X so that X_1 satisfies this condition, and let β_1 denote the coefficient of X_1. The scale normalization consists of setting $|\beta_1| = 1$.

Because β_1 can have only two values, every consistent estimator b_{n1} of β_1 satisfies $\lim_{n\to\infty} P(b_{n1} = \beta_1) = 1$ and converges in probability faster than n^{-r} for any positive r. This is not true of the remaining components of β, which can take on a continuum of values. Therefore, the remaining discussion is restricted to estimating and testing hypotheses about $\tilde{\beta} \equiv (\beta_2,...,\beta_q)'$.

a. Notation

The following notation will be used to describe the smoothed maximum score estimator and test statistics based on it. Let $\tilde{X}$ be the vector consisting of components 2 - q of X. Let $Z = \beta'X$. Since $|\beta_1| = 1$, there is a one-to-one relation between $(Z,\tilde{X}')'$ and X. The assumptions made below imply that the distribution of Z conditional on $\tilde{X}$ has everywhere positive density with respect to Lebesgue measure for almost every $\tilde{X}$. Let $p(\bullet|\tilde{X}=\tilde{x}) \equiv p(\bullet|\tilde{x})$ denote this density. For each integer i > 0, define

$$p^{(i)}(z|\tilde{x}) = \partial^i p(z|\tilde{x})/\partial z^i$$

whenever the derivative exists, and define $p^{(0)}(\bullet|\tilde{x}) \equiv p(\bullet|\tilde{x})$. For each β with $|\beta_1| = 1$, let $F(\bullet|z,\tilde{x})$ denote the cumulative distribution of U conditional on Z $= z \equiv \beta'x$ and $\tilde{X} = \tilde{x}$. For each positive integer i, define

$$F^{(i)}(-z|z,\tilde{x}) = \partial^i F(-z|z,\tilde{x})/\partial z^i$$

whenever the derivative exists. For example,

$$F^{(1)}(-z|z,\tilde{x}) = [-\partial F(u|z,\tilde{x})/\partial u + \partial F(u|z,\tilde{x})/\partial z]_{u = -z}$$

Let $K^{(j)}$ denote the j'th derivative of K if it exists. For each integer $s \geq 2$, define the $(q - 1)\times 1$ vector A and the $(q-1)\times(q-1)$ matrices D and Q by

$$A = -2\alpha_A \sum_{i=1}^{s} \{[i!(s - i)!]^{-1} E[F^{(i)}(0|0,\tilde{X})p^{(s-i)}(0|\tilde{X})\tilde{X}]\}, \qquad (2.2)$$

$$D = \alpha_D E[\tilde{X}\tilde{X}'p(0|\tilde{X})],$$

and

$$Q = 2 \int \tilde{x}\tilde{x}'F^{(1)}(0|0,\tilde{x})p(0|\tilde{x})dP(\tilde{x}),$$

where

$$\alpha_A = \int_{-\infty}^{\infty} v^s K^{(1)}(v)dv$$

and

$$\alpha_D = \int_{-\infty}^{\infty} [K^{(1)}(v)]^2 dv.$$

b. Asymptotic Properties

Let b_n be a solution to (2.1), and write $b_n = (b_{n1},\tilde{b}_n')'$. Horowitz (1992) gives regularity conditions under which the following theorem holds. Section 4b of this paper gives a strengthened version of the regularity conditions that is used in the analysis of the bootstrap.

Theorem 1.1: Let $s \geq 2$ be an integer. Assume that as $n \to \infty$, $nh_n^{2s+1} \to \lambda < \infty$ and $(\log n)/(nh_n^4) \to 0$. If $F^{(i)}(-z|z,\tilde{x})$ and $p^{(i - 1)}(z|\tilde{x})$ are bounded, continuous functions of z for each integer i between 1 and s, all z in a neighborhood of 0, and almost every $\tilde{x}$, then

$$(nh_n)^{1/2}(b_n - \beta) \to^d N(-\lambda^{1/2}Q^{-1}A,\ Q^{-1}DQ^{-1}). \quad \blacksquare$$

To use this result to test hypotheses about β, it is necessary to have a consistent estimator of the covariance matrix $Q^{-1}DQ^{-1}$. To obtain such an estimator, define

$$Q_n(b) = (nh_n^2)^{-1}\sum_{i=1}^{n}[2I(Y_i = 1) - 1]X_iX_i'K^{(2)}(b'X_i/h_n). \qquad (2.3)$$

Horowitz (1992) shows that $Q_n(b_n) \to^p Q$ under the assumptions made in Section 4 of this paper. The consistent estimator of D used by Horowitz (1992) is

$$(nh_n)^{-1}\sum_{i=1}^{n}X_iX_i'[K^{(1)}(b_n'X_i/h_n)]^2.$$

However, this estimator has a bias of size $O(h_n^2)$, which is too large for the theory developed in this paper. An alternative estimator whose bias is $O(h_n^s)$ and, therefore, smaller if s > 2 can be obtained by defining

$$D_n(b) = \alpha_D(nh_n)^{-1}\sum_{i=1}^{n}X_iX_i'K^{(1)}(b'X_i/h_n).$$

It is not difficult to prove that $D_n(b_n) \to^p D$ under the assumptions made in Section 4. In the remainder of this paper, $D_n(b_n)$ is used to estimate D. The resulting consistent estimator of $Q^{-1}DQ^{-1}$ is $Q_n(b_n)^{-1}D_n(b_n)Q_n(b_n)^{-1}$.

The fastest possible rate of convergence in probability of b_n to β is $n^{-s/(2s+1)}$ (Horowitz 1992). This occurs when $h_n = (\lambda/n)^{1/(2s+1)}$ with $0 < \lambda < \infty$. However, the resulting estimator of β is asymptotically biased in the sense that the asymptotic distribution of $(nh_n)^{1/2}(b_n - \beta)$ is not centered at 0. To test hypotheses about β, it is necessary to remove the asymptotic bias. One way of doing this is by estimating the asymptotic bias and subtracting the estimate from b_n. This is the method of explicit bias removal. Another way is by setting $h_n \propto n^{-\kappa}$ where $\kappa > 1/(2s + 1)$, in which case $\lambda = 0$ and there is no asymptotic bias. This is the method of undersmoothing. This paper uses undersmoothing. Thus, h_n is assumed to converge sufficiently rapidly that

$$(nh_n)^{1/2}(b_n - \beta) \to^d N(0, Q^{-1}DQ^{-1}).$$

The method of explicit bias removal can also be considered. Horowitz (1992) describes a method for removing the asymptotic bias of b_n, but this method leaves a higher-order bias that is too large to permit $O[(nh_n)^{-1}]$ asymptotic refinements via the bootstrap. Bias removal to sufficiently high order can be achieved by replacing b_n with $b_n + \lambda^{1/2}Q_n(b_n)^{-1}A_n$, where A_n is the consistent estimator of A that is formed by replacing the derivatives and expectation on the right-hand side of (2.2) by kernel estimates and the sample average. An analysis that is notationally more complex than the one carried out here but similar in other respects can be used to show that under suitable regularity conditions, the bootstrap provides asymptotic refinements with this method of explicit bias removal.

It may seem that use of undersmoothing sacrifices asymptotic estimation efficiency relative to the method of explicit bias removal since Theorem 1.1 implies a faster rate of convergence of b_n to β in the latter case. This impression is misleading. To obtain asymptotic refinements with either method of bias removal, $F(-z|z,\tilde{x})$ and $p(z|\tilde{x})$ must have more derivatives than are required for the first-order asymptotics of Theorem 1.1. Consequently, the fastest rate of convergence permitted by Theorem 1.1 is not achievable if one wants to obtain asymptotic refinements. Moreover, F and p must have more derivatives to achieve asymptotic refinements through explicit bias removal than to achieve them through undersmoothing. If F and p have enough derivatives to obtain asymptotic refinements with explicit bias removal, the rate of convergence obtained with explicit bias removal can also be obtained with undersmoothing by choosing K so as to make $K^{(1)}$ a kernel of sufficiently high order in the sense of Assumption 6 in Section 4. Thus, there is no asymptotic efficiency advantage to using explicit bias removal instead of undersmoothing.

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 - 1). Define the matrix $V_n = Q_n(b_n)^{-1}D_n(b_n)Q_n(b_n)^{-1}$. 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

$$t = (nh_n)^{1/2}(b_{ni} - \beta_{0i})/(V_{ni}^{1/2}). \tag{3.1}$$

It follows from Theorem 1.1 together with $Q_n(b_n) \to^p Q$ and $D_n(b_n) \to^p D$ that $t \to^d N(0,1)$ if H_0 is true. Thus, if n is sufficiently large, H_0 can be tested by comparing t with the relevant quantile of the standard normal distribution. In particular, 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 $r \times (q - 1)$ matrix with $r \leq q - 1$, and let c be an $r \times 1$ vector of constants. Consider a test of the hypothesis H_0: $R\beta = c$. Assume that the matrix $RQ^{-1}DQ^{-1}R'$ is nonsingular. Then under H_0, the statistic

$$\chi^2 \equiv (nh_n)(Rb_n - c)'(RV_nR')^{-1}(Rb_n - c) \tag{3.2}$$

is asymptotically chi-square distributed with r degrees of freedom. H_0 can be tested at the asymptotic α level by comparing χ^2 with the asymptotic critical value consisting of the 1 - α quantile of the chi-square distribution.

Monte Carlo evidence (Horowitz 1992) indicates that the true and nominal levels of the asymptotic t test can be very different with samples of the sizes often found in applications. The reason is that in finite samples the true and asymptotic distributions of the t statistic can differ greatly. As a result, the true critical value for a given level of the test may be far from the one obtained from the asymptotic distribution. Section 4 of this paper gives conditions under which the bootstrap provides improved finite-sample critical values for the symmetrical t and χ^2 statistics.

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 $1 - \alpha$ quantile of the bootstrap distribution of $|t|$. The bootstrap α-level critical value of a test based on χ^2 is the $1 - \alpha$ quantile of the bootstrap distribution of χ^2.

The bootstrap distributions of $|t|$ and χ^2 cannot be calculated analytically, but they 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 $H_n(b)$, $Q_n(b)$ and $D_n(b)$:

$$H_n^*(b) = n^{-1}\sum_{i=1}^{n}[2I(Y_i^* = 1) - 1]K(b'X_i^*/h_n),$$

$$Q_n^*(b) = (nh_n^2)^{-1}\sum_{i=1}^{n}[2I(Y_i^* = 1) - 1]X_i^*X_i^{*\prime}K^{(2)}(b'X_i^*/h_n)$$

$$D_n^*(b) = \alpha_D(nh_n)^{-1}\sum_{i=1}^{n}X_i^*X_i^{*\prime}K^{(1)}(b'X_i^*/h_n).$$

Let $b_n^* = (b_{n1}^*, \tilde{b}_n^*)$ be a solution to (2.1) with H_n replaced by H_n^*. Let V_{ni}^* be the (i,i) component of the matrix $Q_n^*(b_n^*)^{-1}D_n^*(b_n^*)Q_n^*(b_n^*)^{-1}$.

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. Using the bootstrap sample, compute the bootstrap t statistic for testing the bootstrap hypothesis H_0^*: $\tilde{\beta}_i = \tilde{b}_{ni}$, where b_n solves (2.1). The bootstrap t statistic is

$$t^* = (nh_n)^{1/2}(b_{ni}^* - b_{ni})/(V_{ni}^{*1/2}),$$

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 the symmetrical t and χ^2 tests in smoothed maximum score estimation. 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 the 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 to the levels of symmetrical t and χ^2 tests based on the smoothed maximum score estimator. Let $s \geq 8$ be an even integer. The assumptions are:

<u>Assumption 1</u>: $\{Y_i, X_i: i = 1,...,n\}$ is a random sample of $(Y,X')'$, where $Y = I(\beta'X + U \geq 0)$, $X \in \Re^q$ $(q \geq 1)$, U is an unobserved random scalar, and $\beta \in \Re^q$ is a constant.

<u>Assumption 2:</u> (a) The support of the distribution of X is not contained in any proper linear subspace of $\Re^q$. (b) $0 < \Pr(Y = 1|X = x) < 1$ for almost every x. (c) $|\beta_1| = 1$, and for almost every $\tilde{x}$, the distribution of $X^{(1)}$, the first component of X, conditional on $\tilde{X} = \tilde{x}$ has everywhere positive density with respect to Lebesgue measure.

<u>Assumption 3</u>: $\text{Median}(U|X = x) = 0$ for almost every x.

<u>Assumption 4</u>: $\tilde{\beta} \equiv (\beta_2,...,\beta_q)'$ is an interior point of $\tilde{B}$, which is a compact subset of $\Re^{q-1}$.

<u>Assumption 5</u>: The components of $\tilde{X}$ are bounded with probability 1.

<u>Assumption 6</u>: (i) $K(v) = 0$ if $v \leq -1$, and $K(v) = 1$ if $v \geq 1$. (ii) K is 5-times differentiable everywhere, $K^{(1)}(v)$ is symmetrical about $v = 0$, and $K^{(i)}$ $(i = 1,...,5)$ is bounded and Lipschitz continuous on $(-\infty,\infty)$. (iii) Let $\tilde{K}(v)$ be the vector whose components are $K^{(i)}(v)$ $(i = 1,...,3)$. For $\theta \in \Re^3$ satisfying $\|\theta\| = 1$, there is a partition of $[-1,1]$, $-1 = a_1 < a_2 <... < a_{L(\theta)} = 1$ such that $\theta'\tilde{K}(v)$ is either strictly increasing or strictly decreasing on $(a_{\ell-1}, a_\ell)$ $(\ell = 2,...,L(\theta))$. (iv) For each integer i $(1 \leq i \leq s)$

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

<u>Assumption 7</u>: For each integer i such that $1 \leq i \leq s + 2$, all z in a neighborhood of 0, almost every $\tilde{x}$, and some $M < \infty$, $p^{(i)}(z|\tilde{x})$ exists and is a continuous function of z satisfying $|p^{(i)}(z|\tilde{x})| < M$. In addition, $|p(z|\tilde{x})| < M$ for all z and almost every $\tilde{x}$.

<u>Assumption 8</u>: For each integer i such that $1 \leq i \leq s + 2$, all z in a neighborhood of 0, almost every $\tilde{x}$, and some $M < \infty$, $F^{(i)}(-z|z,\tilde{x})$ exists and is a continuous function of z satisfying $|F^{(i)}(-z|z,\tilde{x})| < M$.

<u>Assumption 9</u>: The matrix Q is negative definite.

<u>Assumption 10</u>: $h_n = \lambda n^{-\kappa}$, where $0 < \lambda < \infty$ and $1/(s + 1) \leq \kappa < 1/7$.

These are strengthened versions of the assumptions of Horowitz (1992). See that paper for a discussion of their rationale and significance. The main strengthening consists of imposing more smoothness on F, p, and K than are needed for first-order asymptotics (Theorem 1.1), requiring $K^{(1)}$ to be a "higher-order" kernel, and requiring h_n to converge to zero more rapidly than the rate that minimizes the asymptotic mean-square error of b_n (undersmoothing). The strengthened smoothness assumptions are needed because the arguments showing that the bootstrap delivers asymptotic refinements to the levels of the symmetrical t and χ^2 tests are based on higher-order asymptotic expansions of the distributions of $|t|$ and χ^2. These, in turn, are based on a higher-order Taylor series expansion of the first-order condition of problem (2.1) as well as Taylor series expansions of the matrices entering V_{ni}. The Taylor series expansions include terms that are sums of derivatives of kernel functions. The strengthened smoothness assumptions insure that these sums exist, converge in probability, and are bounded with sufficiently high probability. They also insure that Taylor series remainder and asymptotic bias terms are negligibly small with sufficiently high probability.

Part (iii) of assumption 6 is used to establish a modified form of the Cramer condition of Edgeworth analysis (see lemma 9 in Section 7). Functions K satisfying assumption 6 can be constructed by integrating the kernels given by Müller (1984).

Horowitz (1992) does not assume that X has bounded support, K(v) = 0 if $v \leq -1$, or K(v) = 1 if $v \geq 1$. These assumptions are made here to simplify the analysis that follows. They are not essential, however, and can be removed at the expense of lengthier and more complex proofs.

Assumptions 6(iv) and 10 imply that $K^{(1)}$ must be a kernel of order 8 or more. The need for such a high-order kernel arises from two considerations. First, the Taylor-series expansion of the first-order condition of (2.1) includes the third derivative of a kernel estimator, whose variance is $O[(nh_n^7)^{-1}]$. Convergence of this term requires $\kappa < 1/7$. Second, the asymptotic bias of $(nh_n)^{1/2}(b_n - \beta)$ is $O[(nh_n)^{1/2} h_n^s]$. To order $(nh_n)^{-1}$, its effect on the distributions of the symmetrical t and χ^2 tests is proportional to its square. Therefore, to achieve asymptotic refinements through $O[(nh_n)^{-1}]$, the square of the asymptotic bias must be $O[(nh_n)^{-1}]$. This implies that $\kappa > 1/(s + 1)$. Since s is an even integer, $1/(s + 1) < \kappa < 1/7$ requires $s \geq 8$.

In practice, the numerical effect of the asymptotic bias is often very small even if its theoretical order of magnitude is larger than $O[(nh_n)^{-1}]$. Consequently, good numerical results may often be obtained with a $K^{(1)}$ whose order is less than 8. See Section 5 for further discussion of this point.

b. Theorems

This section gives theorems that establish conditions under which the bootstrap provides asymptotic refinements to the levels of the symmetrical t and χ^2 tests in smoothed maximum score estimation. The first two theorems give conditions under which the sample and bootstrap versions of $|t|$ and χ^2 have Edgeworth-type asymptotic expansions. The third theorem shows that the bootstrap provides asymptotic refinements under the same conditions. The proofs of the theorems are in Section 7.

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 Section 7. 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 1-10 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 $o[(nh_n)^{-1/2}]$ and no smaller than $O[(nh_n)^{-1}]$ (almost surely in the case of $q(\tau,\nu_n^*)$);

(c)

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

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 $H_n(\beta)$ and $D_n(\beta)$ 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 functions of derivatives of $H_n(\beta)$ and $D_n(\beta)$ (or their bootstrap analogs in the case of t*). This is done in Propositions 1 and 2 of Section 7. 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 $F_{\chi,r}$ denote the chi-square distribution function with r 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 $O[(nh_n)^{-1}]$.

<u>Theorem 4.2</u>: Let assumptions 1-10 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 $o[(nh_n)^{-1/2}]$ and no smaller than $O[(nh_n)^{-1}]$ (almost surely in the case of $q(\tau,\nu_{n\chi}^*)$),

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

uniformly over z, and

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

uniformly over z almost surely. ■

The final theorem shows that the use of bootstrap critical values yields asymptotic refinements to the levels of the 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*}.

<u>Theorem 4.3</u>: Let assumptions 1-10 hold. Under H_0: $\beta_i = \beta_{0i}$,

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

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

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

5. MONTE CARLO EXPERIMENTS

This section reports the results of a small set of Monte Carlo experiments that were carried out to investigate the numerical performance of the bootstrap. The number of experiments is small because of the very long computing times they require.

Each experiment evaluates levels of a symmetrical t test using asymptotic and bootstrap critical values. The hypothesis being tested is H_0: $\beta = 1$ in the model

$$Y = I(X_1 + \beta X_2 + U \geq 0). \qquad (5.1)$$

β is a scalar parameter whose true value is 1 (so H_0 is true), $X_1 \sim N(0,1)$, and $X_2 \sim N(1,1)$. There are 3 different distributions of U, depending on the experiment. These are:

Distribution L: $U \sim$ logistic with median 0 and variance 1;

Distribution T3: $U \sim$ Student's t with 3 degrees of freedom normalized to have variance 1;

Distribution H: $U \sim 0.25(1 + Z^2 + Z^4)V$, where $Z = X_1 + X_2$ and $V \sim$ logistic with median 0 and variance 1.

With distributions L and T3, U is homoskedastic and $P(Y=1|Z=z)$ has the familiar ogival shape. With distribution H, U is heteroskedastic and $P(Y=1|Z=z)$ is nonmonotonic. It has a global minimum at $z = -1/\sqrt{3}$, a global maximum at $z = 1/\sqrt{3}$, and converges to 0.5 as $z \to \pm\infty$. Experiments with multidimensional β's were not carried out owing to the very long computing times they entail.

The smoothing function is

$$K_4(v) = \begin{cases} 0 & \text{if } v < -1 \\ 0.5 + (105/64)[v - (5/3)v^3 + (7/5)v^5 - (3/7)v^7] & \text{if } |v| \leq 1 \\ 1 & \text{if } v > 1 \end{cases}$$

K_4 is the integral of a 4th-order kernel for nonparametric density estimation (Müller 1984). It does not satisfy assumptions 6(iv) and 10, which require a kernel of 8th or higher order. However, for reasons that were discussed in Section 4a, good numerical results can often be obtained with a kernel whose order is lower than 8. There are important practical advantages to using a lower-order kernel. In smoothed maximum score estimation as in other forms of kernel estimation, a lower-order kernel tends to give results that are more stable numerically and less sensitive to the choice of bandwidth than the results obtained with high-order kernels. Reduced sensitivity to the choice of bandwidth is especially important in applications, since the theory developed in this paper provides little practical guidance on bandwidth choice for achieving asymptotic refinements. K_4 provides the advantages of a relatively low-order kernel as well as the increased asymptotic efficiency that is available through use of a kernel whose order exceeds 2.[1]

K_4 does not satisfy assumption 6(ii) because it has only two derivatives at $v = \pm 1$. This problem can be overcome by smoothing K_4 in neighborhoods of $v = \pm 1$, but doing so has no effect on the results of the experiments.

Each experiment was carried out using three different bandwidths. One is the bandwidth h_{n0} that minimizes the asymptotic mean-square error of b_n as an estimator of β. The formula for this bandwidth is given in Horowitz (1992). The other bandwidths are $0.7h_{n0}$ and $0.5h_{n0}$, which represent different amounts of undersmoothing relative to the bandwidth that minimizes asymptotic mean-square error.

The experiments use a sample size of n = 250 and were carried out with a program written in GAUSS with GAUSS pseudo-random number generators. There are 1000 Monte Carlo replications per experiment. The smoothed maximum score objective function can have many local extrema, so maximizing it requires the use of a global optimization method. The results reported here are based on a search over a discrete set of b values, which is faster than a more elaborate global optimization procedure such as simulated annealing in the one-dimensional setting of the experiments.

Each experiment consisted of repeating the following steps 1000 times:

A. Generate an estimation data set of size n = 250 by randomly sampling (Y, X_1, X_2) from the model under consideration. Obtain the smoothed maximum score estimate of β in (5.1) and compute the t statistic for testing H_0: $\beta = 1$. Call its value t_0.

B. Compute the bootstrap critical value of the t statistic by following steps 1-3 in Section 3b. Bootstrap samples of size 250 are obtained by sampling the estimation data set generated in step A randomly with replacement. Denote

---

[1] I have also carried out Monte Carlo experiments using the designs described in this section and a smoothing function that is the integral of an 8th-order kernel. The results show that the bootstrap corrects distortions of the level of the symmetrical t test if the bandwidth is chosen carefully. As expected, however, the level with bootstrap critical values is more sensitive to the choice bandwidth with the 8th-order kernel than with the 4th-order one.

the α-level bootstrap critical value of the symmetrical t test by t_α^*. In experiments with $\alpha = 0.1$ and 0.05, t_α^* was computed from the empirical distribution of $|t^*|$ obtained from 100 bootstrap samples. In experiments with $\alpha = 0.01$, t_α^* was based on 500 bootstrap samples. Increasing the number of bootstrap samples beyond these values did not change the results of the experiments.

C. Reject H_0 at the nominal α level based on the bootstrap critical value if $|t_0| > t_\alpha^*$. Reject H_0 at the nominal α level based on the asymptotic critical value if $|t_0|$ exceeds the $1 - \alpha/2$ quantile of the standard normal distribution.

The results of the experiments are shown in Table 1. The empirical levels of the t test greatly exceed the nominal levels when asymptotic critical values are used. Moreover, the empirical levels with asymptotic critical values are highly sensitive to the choice of bandwidth. For example, at the nominal 0.05 and 0.01 levels, the empirical levels increase by over a factor of two when the bandwidth decreases from h_{n0} to $0.5h_{n0}$. In contrast, the differences between the nominal and empirical levels are very small when bootstrap critical values are used. None of the differences between nominal and empirical levels with bootstrap critical values is significantly different from zero at the 0.01 level. Thus, in these experiments the bootstrap essentially eliminates the distortions of the level of the t test that occur when asymptotic critical values are used. In addition, the empirical levels with bootstrap critical values display little sensitivity to variations in the bandwidth. As has already been discussed, this is a desirable property in applications.

6. CONCLUSIONS

The smoothed maximum score estimator of the coefficient vector of a binary response model is consistent and asymptotically normally distributed under weak distributional assumptions. However, first-order asymptotic theory often provides a poor approximation to the finite-sample distributions of test statistics based on this estimator. As a result, the differences between the true and nominal levels of tests based on smoothed maximum score estimates can be very large in finite samples when first-order asymptotics are used to obtain critical values. This paper has given conditions under which the bootstrap provides finite-sample critical values that are more accurate than those obtained from first-order asymptotic theory. The ability of the bootstrap to provide asymptotic refinements has long been known, but the standard theory of the bootstrap does not apply to the smoothed maximum score estimator because this estimator cannot be approximated by a function of sample moments. The theoretical results that have been presented here overcome this problem.

The numerical performance of the bootstrap was investigated through a small set of Monte Carlo experiments. In these experiments, the bootstrap essentially eliminated large finite-sample distortions of level that occurred when asymptotic critical values were used. These results illustrate the usefulness of the bootstrap for obtaining improved finite-sample critical values for tests based on the smoothed maximum score estimator. Of course, the numerical results are subject to the usual cautions about attempting to generalize from a small set of Monte Carlo experiments.

7. PROOFS OF THEOREMS

It is assumed throughout this section that assumptions 1-10 hold.

a. Step 1: Approximating t and t*

Let $\|\cdot\|$ denote the Euclidean norm. For y = 0 or 1, define J(y) = 2I(y = 1) - 1. For $b \in B$, define $H(b) = E[J(Y)I(b'X \geq 0)]$. Observe that

$$H_n(b) = n^{-1}\sum_{i=1}^{n} J(Y_i)K(b'X_i/h_n).$$

Define a Euclidean class of functions as in Nolan and Pollard (1987).

<u>Lemma 1</u>: There is a finite C > 0 such that

$$\sup_{b \in B} |H_n(b) - H(b)| \leq Ch_n$$

almost surely for all sufficiently large n.

<u>Proof</u>: It follows from lemma 22 of Nolan and Pollard (1987) and theorem 2.37 of Pollard (1984) that $|H_n(b) - EH_n(b)| = o[(\log n)/n^{1/2}]$ almost surely

uniformly over $b \in B$. Let $P_Z(\bullet|X)$ denote the conditional CDF of Z. Then, by assumption 7 and a Taylor series approximation, $|EH_n(b)| \leq P(|b'X| \leq h_n) = E\{P_Z[-(b - \beta)'X + h_n|X] - P_Z[-(b - \beta)'X - h_n|X]\} < Mh_n$ for all $b \in B$ and some $M < \infty$. 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 lemmas 2 and 3 of Manski (1985), H is continuous on B and uniquely maximized at β. Therefore, $H(b) < H(\beta) - \delta$ for all $b \in N_r$ and some $\delta > 0$. By lemma 1 and $h_n \to 0$, $H_n(b) < H_n(\beta) - \delta/2 < H_n(\beta)$ almost surely for all $b \in N_r$ if $n > n_0$ for some finite n_0. Thus, $b_n \notin N_r$ almost surely if $n > n_0$. Q.E.D.

For $i,j,k,\ell,m = 1,\ldots,q - 1$, define $G_{ni}(b) = \partial H_n(b)/\partial b_i$, $G_{nij}(b) = \partial^2 H_n(b)/\partial b_i \partial b_j$, $G_{nijk}(b) = \partial^3 H_n(b)/\partial b_i \partial b_j \partial b_k$, $G_{nijk\ell}(b) = \partial^4 H_n(b)/\partial b_i \partial b_j \partial b_k \partial b_\ell$, and $G_{nijk\ell m}(b) = \partial^5 H_n(b)/\partial b_i \partial b_j \partial b_k \partial b_\ell \partial b_m$. 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 $D_{nijk}(b) = \partial^3 D_n(b)/\partial b_i \partial b_j \partial b_k$.

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

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

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

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

(d) $\sup_{b \in B} |G_{nijk\ell}(b) - EG_{nijk\ell}(b)| = o[(\log n)/(nh_n^7)^{1/2}]$

(e) $\sup_{b \in B} |G_{nijk\ell m}(b) - EG_{nijk\ell m}(b)| = o[(\log n)/(nh_n^9)^{1/2}]$

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

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

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

(i) $\sup_{b \in B} |D_{nijk}(b) - ED_{nijk}(b)| = o[(\log n)/(nh_n^7)^{1/2}]$

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

(j) $EG_{ni}(\beta) = O(h_n^r)$

(k) $EG_{nijk}(b)$, $EG_{nijk\ell}(b)$, $EG_{nijk\ell m}(b)$, $ED_n(b)$, $ED_{ni}(b)$, $ED_{nij}(b)$, and $ED_{nijk}(b)$ are $O(1)$ as $n \to \infty$ for all $b \in B$.

<u>Proof</u>: Parts (a)-(i) are proved by using lemma 2.14 of Pakes and Pollard (1989) and lemma 22 of Nolan and Pollard (1987) to show that the summands of the relevant G or D functions form Euclidean classes and then applying theorem 2.37 of Pollard (1984). Part (j) is proved in lemma 5(a) of Horowitz (1992). To prove (k), consider $EG_{nijk}(b)$. Let $\tilde{X}_{ij}$ be the j'th component of $\tilde{X}_i$. Then

$$EG_{nijk}(b) = (nh_n^3)^{-1} \sum_{\ell=1}^{n} EJ(Y_\ell)\tilde{X}_{\ell i}\tilde{X}_{\ell j}\tilde{X}_{\ell k} K^{(3)}(b'X_\ell/h_n)$$

$$= h_n^{-3}\int[1 - 2F(-z|z,\tilde{x})]\tilde{x}_{\bullet i}\tilde{x}_{\bullet j}\tilde{x}_{\bullet k} K^{(3)}\{[z + (b - \beta)'\tilde{x}]/h_n\}p(z|\tilde{x})dzdP(\tilde{x})$$

Make the change of variables $\zeta = [z + (b - \beta)]/h_n$. Then

$$EG_{nijk}(b) =$$

$$h_n^{-2}\int_{-1}^{1}\{1 - 2F[-h_n\zeta + (b - \beta)'\tilde{x}|h_n\zeta - (b - \beta)'\tilde{x},\tilde{x}]\}\tilde{x}_{\bullet i}\tilde{x}_{\bullet j}\tilde{x}_{\bullet k} K^{(3)}(\zeta)$$

$$\bullet p[h_n\zeta - (b - \beta)'\tilde{x}|\tilde{x}]d\zeta dP(\tilde{x}). \qquad (7.1)$$

The result $E[G_{nijk}(b)] = O(1)$ is obtained by making a Taylor series expansion of the integrand of (7.1) about $h_n = 0$ through $O(h_n^2)$. Similar arguments apply to the other G and D 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 $G_{nijk\ell}(\beta)$ $(i,j,k,\ell = 1,...,q - 1)$. Order the components of S_{nG} so that the first q - 1 are the $G_{ni}(\beta)$.

<u>Lemma 4</u>: Let $S_G = \text{plim}_{n\to\infty} S_{nG}$. There is a function $\Lambda_\beta(S_{nG})$ taking values in $\Re^{q-1}$ such that $\Lambda_\beta(S_G) = 0$ and

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

with probability $1 - o[(nh_n)^{-1}]$ as $n \to \infty$.

<u>Proof</u>: Define $\delta_n = b_n - \beta$ and $\delta_{ni} = b_{ni} - \beta_i$ $(i = 1,...,q - 1)$. Let $G_{n\cdot}(\beta)$ be the vector whose components are the unique components of $G_{ni}(\beta)$ $(i = 1,...,q - 1)$. 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-1 dimensional vectors whose components are $G_{nij}(\beta)$, $G_{nijk}(\beta)$, and

$G_{nijk\ell}(\beta)$ (i = 1,...,q-1). By lemma 2, b_n satisfies the first-order condition $G_{n\bullet}(b_n) = 0$ almost surely for all sufficiently large n. It is proved in Horowitz (1992) that $\lim_{n\to\infty} EQ_n(\beta) \to Q$. It follows from this result and lemma 3 that $Q_n(\beta)$ 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

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

$$+ (1/6)G_{n\bullet jk\ell}(\beta)\delta_{nj}\delta_{nk}\delta_{n\ell} + R_n], \tag{7.2}$$

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

$$R_n = o[(\log n)/(nh_n^9)^{1/2}]\|b_n - \beta\|^4.$$

Given any $\nu > 0$ and $c > 0$, suppose that $\|\delta_n\| < c(nh_n)^{-1/2+\nu}$. Then it follows from lemma 3 that the right-hand side of (7.2) is less than $c(nh_n)^{-1/2+\nu}$ almost surely for all sufficiently large n. In addition, lemma 3b and assumptions 4 and 7-9 imply that the consistent solution to (2.1) is almost surely unique for all sufficiently large n. Therefore, application of the Brouwer fixed point theorem to the right-hand side of (7.2) shows that for any $c > 0$, $\nu > 0$,

$$\|b_n - \beta\| \leq c(nh_n)^{-1/2+\nu} \tag{7.3}$$

almost surely for all sufficiently large n. Observe that $Q_n(\beta)$ is the matrix whose elements are $G_{nij}(\beta)$. Then application of the implicit function theorem to (7.2) shows that there is almost surely a differentiable function Λ_β such that $\Lambda_\beta(S_G) = 0$ and

$$(b_n - \beta) = \Lambda_\beta(S_{nG} + \bar{R}_n), \tag{7.4}$$

where $\bar{R}_n$ is a vector such that $\dim(\bar{R}_n) = \dim(S_{nG})$, R_n forms the first q - 1 components of $\bar{R}_n$, and the remaining components of $\bar{R}_n$ are 0. Application of the mean value theorem to (7.4) combined with (7.3) shows that

$$(b_n - \beta) = \Lambda_\beta(S_{nG}) + O[(\log n)(n^5 h_n^{13})^{-1/2}(nh_n)^{4\nu}] \quad (7.5)$$

almost surely for any $\nu > 0$. The lemma now follows from assumption 10 by making ν sufficiently small. Q.E.D.

Let S_n denote denote a vector consisting of the unique components of S_{nG} and of $D_n(\beta)$, $D_{ni}(\beta)$, and $D_{nij}(\beta)$ $(i,j = 1,..., q - 1)$.

<u>Lemma 5</u>: For each $i = 1,..., q - 1$, 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 $Q_n(b_n)$ in Taylor series about $b_n = \beta$ through order $\|b_n - \beta\|^3$ and use (7.5) to obtain

$$V_{ni}^{1/2} = \Lambda_{Vi}^*(S_n, S_{nG} + \omega_n) + o[(nh_n)^{-1}] \quad (7.6)$$

almost surely for a suitable differentiable function Λ_{Vi}^*, where $\omega_n = O[(\log n)(nh_n^3)^{-2}(nh_n)^{-4\nu}]$ for any $\nu > 0$. The lemma now follows by applying the mean value theorem to (7.6). Q.E.D.

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

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

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

$$t = \frac{(nh_n)^{1/2}\Lambda_\beta(S_{nG}) + \epsilon_n}{\Lambda_{Vi}(S_n) + \nu_n}, \quad (7.7)$$

where ϵ_n and ν_n are $o[(nh_n)^{-1}]$ almost surely. Define $\Delta_n = t - (nh_n)^{1/2}\Lambda(S_n)$. A Taylor series approximation applied to (7.7) 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

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

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

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

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

$$P(t \le z) - P[(nh_n)^{1/2}\Lambda(S_n) \le z] = o[(nh_n)^{-1}]. \qquad (7.8)$$

uniformly over z. The proposition follows by multiplying both sides of (7.8) by nh_n and taking the limit as $n \to \infty$. Q.E.D.

Let P denote the population distribution of (Y,X) and $P_n{}^*$ denote the bootstrap probability measure. That is, $P_n{}^*$ places mass 1/n at each point of the estimation sample. Let E_n denote the expectation with respect to $P_n{}^*$. The following lemma gives a bootstrap version of Pollard's theorem 2.37.

<u>Lemma 6</u>: For any $b \in B$, define

$$W_n(b) = n^{-1}\sum_{i=1}^{n}[g(Y_i{}^*,X_i{}^*)f(b'X_i{}^*/h_n) - E_n g(Y,X)f(b'X/h_n)],$$

where g is a bounded function and f is a bounded, Lipschitz continuous function of bounded variation with support [-1,1]. 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 $\gamma \ge 0$

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

almost surely (P).

<u>Proof</u>: Partition B into subsets $\{B_j: j = 1,\ldots,J_n\}$ 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 $J_n = O(\xi_n{}^{-2q})$. Then

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

$$\le \sum_{j=1}^{J} P_n^*(\sup_{b \in B_j} |W_n(b)| > C\xi_n). \qquad (7.9)$$

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

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

Therefore, for all sufficiently large n

$$P_n^*(\sup_{b \in B_j} |W_n(b)| > C\xi_n) \le P_n^*(|W_n(b_j)| > C\xi_n/2) \qquad (7.10)$$

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

$$P_n^*(|W_n(b_j)| > C\xi_n/2) \le 2\exp(-Cd\log n) = 2n^{-Cd} \qquad (7.11)$$

for some finite d > 0 and all sufficiently large n. Combining (7.9)-(7.11) 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.

Define

$$H_n^*(b) = n^{-1} \sum_{i=1}^{n} J(Y_i^*)K(b'X_i^*/h_n).$$

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).

Proof: Since $E_n H_n^*(b) = H_n(b)$, lemma 6 implies that for any $\gamma > 0$ and $\eta > 0$

$$\lim_{n \to \infty} (nh_n)^\gamma P_n^*[\sup_{b \in B} |H_n^*(b) - H_n(b)| > \eta] = 0 \qquad (7.12)$$

almost surely (P). Given any $\eta > 0$, suppose that $|H_n^*(b) - H_n(b)| \le \eta$ and $|H_n(b) - H(b)| \le \eta$ for all $b \in B$. Then since b_n^* maximizes H_n^*, $H_n(b_n) - \eta \le H_n^*(b_n) \le H_n^*(b_n^*)$. Also, $H_n^*(b_n^*) \le H_n(b_n^*) + \eta$, so $H_n(b_n) - \eta \le H_n^*(b_n^*) \le H_n(b_n^*) + \eta$, and $H_n(b_n) - H_n(b_n^*) \le 2\eta$. By a similar argument, $H(\beta) - H_n(b_n) \le 2\eta$. Therefore, $H(\beta) - H(b_n^*) = [H(\beta) - H(b_n)] + [H(b_n) - H_n(b_n)] + [H_n(b_n) - H_n(b_n^*)] + [H_n(b_n^*) - H(b_n^*)] \le 6\eta$. Because $H(b)$ is continuous on B with a unique maximum at β, it is possible to choose η such that $H(\beta) - H(b_n^*) \le 6\eta$ 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, $|H_n^*(b) - H_n(b)| \le \eta$ and $|H_n(b) - H(b)| \le \eta$ 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 lemma 1 and (7.12). Q.E.D.

For $i,j,k,\ell,m = 1,\ldots,q - 1$, define $G_{ni}^*(b) = \partial H_n^*(b)/\partial b_i$, $G_{nij}^*(b) = \partial^2 H_n^*(b)/\partial b_i \partial b_j$, $G_{nijk}^*(b) = \partial^3 H_n^*(b)/\partial b_i \partial b_j \partial b_k$, $G_{nijk\ell}^*(b) = \partial^4 H_n^*(b)/\partial b_i \partial b_j \partial b_k \partial b_\ell$, and $G_{nijk\ell m}^*(b) = \partial^5 H_n(b)/\partial b_i \partial b_j \partial b_k \partial b_\ell \partial b_m$, $D_{ni}^*(b) = \partial D_n^*(b)/\partial b_i$, $D_{nij}^*(b) = \partial^2 D_n^*(b)/\partial b_i \partial b_j$, and $D_{nijk}^*(b) = \partial^3 D_n^*(b)/\partial b_i \partial b_j \partial b_k$. The bootstrap version of lemma 3 is:

Lemma 8: For all $i,j,k,\ell,m = 1,\ldots,q - 1$, 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) $\sup_{b \in B} |G_{ni}^*(b) - E_n G_{ni}^*(b)| > C[(\log n)/(nh_n)^{1/2}]$

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

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

(d) $\sup_{b \in B} |G_{nijk\ell}^*(b) - E_n G_{nijk\ell}^*(b)| > C[(\log n)/(nh_n^7)^{1/2}]$

(e) $\sup_{b \in B} |G_{nijk\ell m}^*(b) - E_n G_{nijk\ell m}^*(b)| > C[(\log n)/(nh_n^9)^{1/2}]$

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

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

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

(i) $\sup_{b \in B} |D_{nijk}^*(b) - E_n D_{nijk}^*(b)| > C[(\log n)/(nh_n^7)]^{1/2}]$

Relations (f)-(i) apply to the individual components of the matrices D_n^*, D_{ni}^*, D_{nij}^*, and D_{nijk}. In addition, for all $i,j,k,\ell,m,s = 1,\ldots, q - 1$

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

(k) $E_n G_{nij}^*(b)$, $E_n G_{nijk}^*(b)$, $E_n G_{nijk\ell}^*(b)$, $E_n G_{nijk\ell m}^*(b)$, $E_n D_n^*(b)$, $E_n D_{ni}^*(b)$, $E_n D_{nij}^*(b)$, and $E_n D_{nijk}^*(b)$ are $O(1)$ almost surely (P) as $n \to \infty$ uniformly over $b \in B$.

<u>Proof</u>: Parts (a)-(i) are immediate consequences of lemma 6. Part (j) is the first-order condition for problem (1.2). 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.

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

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 $W(y,x,h)$ be the vector whose components are terms of the form

$$J(y)g_{ri}(\tilde{x})K^{(i)}(z/h) \qquad (i = 1,\ldots, 3),$$

where $g_{ri}(\tilde{x})$ is the product of i not necessarily distinct components of $\tilde{x}$, and r indexes the product. Thus, for example, one can define $g_{11}(\tilde{x}) = x^{(2)}$, $g_{21}(\tilde{x}) = x^{(3)}$, etc. 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 $\psi_W(\tau,h) = E\{\exp[i\tau'W(Y,X,h)]\}$ where $i = (-1)^{1/2}$. For any $\epsilon > 0$, some $C > 0$, all τ satisfying $\|\tau\| > \epsilon$, and all sufficiently small h

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

Proof: Each component of W has the form $\mathcal{J}(y,\tilde{x})\mathcal{K}(z/h)$ for suitable functions $\mathcal{J}$ and $\mathcal{K}$. Therefore, letting r index components of W and using the summation convention

$$\psi_W(\tau,h) = \sum_{y=0}^{1} \iint \exp[i\tau_r \mathcal{J}_r(y,\tilde{x})\mathcal{K}_r(z/h)]P(y|z,\tilde{x})p(z|\tilde{x})dzd\mathcal{P}(\tilde{x})$$

$$= h \sum_{y=0}^{1} \iint \exp[i\tau_r \mathcal{J}_r(y,\tilde{x})\mathcal{K}_r(\zeta)]P(y|h\zeta,\tilde{x})p(h\zeta|\tilde{x})d\zeta d\mathcal{P}(\tilde{x})$$

Since the support of each $\mathcal{K}_r$ is [-1,1],

$$\psi_W(\tau,h) = h \sum_{y=0}^{1} \iint_{|\zeta| > 1} P(y|h\zeta,\tilde{x})p(h\zeta|\tilde{x})d\zeta d\mathcal{P}(\tilde{x})$$

$$+ h \sum_{y=1}^{1} \iint_{|\zeta| \leq 1} \exp[i\tau_r \mathcal{J}_r(y,\tilde{x})\mathcal{K}_r(\zeta)]P(y|h\zeta,\tilde{x})p(h\zeta|\tilde{x})dzd\mathcal{P}(\tilde{x})$$

$$= 1 - A_1(\tau,h) + A_2(\tau,h) + A_3(\tau,h),$$

where

$$A_1(\tau,h) = h \sum_{y=0}^{1} \iint_{|\zeta| \leq 1} P(y|h\zeta,\tilde{x})p(h\zeta|\tilde{x})d\zeta d\mathcal{P}(\tilde{x}),$$

$$A_2(\tau,h) = h \sum_{y=0}^{1} \iint_{|\zeta| \leq 1} \exp[i\tau_r \mathcal{J}_r(y,\tilde{x})\mathcal{K}_r(\zeta)][P(y|h\zeta,\tilde{x})p(h\zeta|\tilde{x})$$
$$- P(y|0,\tilde{x})p(0|\tilde{x})]d\zeta d\mathcal{P}(\tilde{x}),$$

and

$$A_3(\tau,h) = h \sum_{y=0}^{1} \iint_{|\zeta| \leq 1} \exp[i\tau_r \mathcal{J}(y,\tilde{x})\mathcal{K}_r(\zeta)]P(y|0,\tilde{x})p(0|\tilde{x})d\zeta d\mathcal{P}(\tilde{x}).$$

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

$$\sum_{y=0}^{1} \iint_{|\zeta| \leq 1} |P(y|h\zeta,\tilde{x})p(h\zeta|\tilde{x}) - P(y|0,\tilde{x})p(0|\tilde{x})|d\zeta d\mathcal{P}(\tilde{x})$$

$$\leq \epsilon \sum_{y=0}^{1} \int_{|\zeta| \leq 1} P(y|0,\tilde{x})p(0|\tilde{x})d\zeta dP(\tilde{x}) \equiv \epsilon C_1.$$

Then $A_1(\tau,h) \leq hC_1(1 - \epsilon)$ and $|A_2(\tau,h)| \leq \epsilon hC_1$, so

$$|\psi_W(\tau,h)| \leq 1 - hC_1(1 - 2\epsilon) + |A_3(\tau,h)| \tag{7.13}$$

for all τ, $\epsilon > 0$, and sufficiently small $h > 0$. Observe that $\|\mathfrak{J}(y,\tilde{x})\| = 0$ if and only if $\|\tilde{x}\| = 0$, which is an event whose probability is less than 1. Therefore, there are an $\eta > 0$ and a $\delta_1 < 1$ such that

$$2 \sum_{y=0}^{1} \int_{\|\mathfrak{J}(y,\tilde{x})\| < \eta} P(y|0,\tilde{x})p(0|\tilde{x})dP(\tilde{x}) = \delta_1 C_1.$$

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

$$\sup_{\|\tau\| \geq \epsilon} \left|\int_{-1}^{1} \exp[i\tau_r \mathfrak{J}_r(y,\tilde{x})K_r(\zeta)]d\zeta\right| = 2C_2 \tag{7.14}$$

uniformly over $(y,\tilde{x})$ such that $\|\mathfrak{J}(y,\tilde{x})\| \geq \eta$. Then for $\|\tau\| \geq \epsilon$

$$|A_3(\tau,h)| \leq h[\delta_1 C_1 + (1 - \delta_1)C_1 C_2] = h\delta_2 C_1 \tag{7.15}$$

where $\delta_2 = [\delta_1 + (1 - \delta_1)C_2] < 1$. Combining (7.15) with (7.13) yields

$$\sup_{\|\tau\| > \epsilon} |\psi_W(\tau,h)| \leq 1 - hC_1(1 - 2\epsilon - \delta_2) \equiv 1 - Ch$$

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

It remains to prove (7.14). To do this, define $t = \|\tau\|$. Fix $\tau/\|\tau\|$, y and $\tilde{x}$ with $\|\tilde{x}\| \neq 0$. For the specified values of $\tau/\|\tau\|$, y and $\tilde{x}$, and using the summation convention, define $f(\zeta) = \tau_r \mathfrak{J}_r(y,\tilde{x})K_r(\zeta)/\|\tau\|$. Let $-1 = a_\ell < \ldots < a_L = 1$ be a partition of $[-1,1]$ that satisfies assumption 6(iii) when $\theta_r = \mathfrak{J}_r(y,\tilde{x})$. Then

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

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

$$\sup_{|t| > \epsilon} (a_\ell - a_{\ell-1})^{-1} \left| \int_{a_{\ell-1}}^{a_\ell} \exp[itf(\zeta)]d\zeta \right| \le C_3. \qquad (7.16)$$

To do this, make the change of variables $\xi = f(\zeta)$ in (7.16) and set $v(\xi) = 1/\{df[\zeta(\xi)]/d\zeta\}$. Then

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

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

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

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

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

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 $\psi_W^*(\tau,h_n) = E_n\{\exp[i\tau'W^*(Y,X,h_n)]\}$, where $i = (-1)^{1/2}$. The bootstrap version of lemma 9 is:

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

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

almost surely (P).

<u>Proof</u>: Let $B_{n\tau} = \{\tau: \epsilon < \|\tau\| \le n^c\}$. Then

$$\sup_{\|\tau\| \in B_{n\tau}} |\psi_W^*(\tau,h_n)| \le \sup_{\|\tau\| \in B_{n\tau}} |\psi_W(\tau,h_n)|$$

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

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_n be a vector consisting of the unique components of $(nh_n)^{1/2}[G_{ni}(\beta) - EG_{ni}(\beta)]$, $(nh_n^3)^{1/2}[G_{nij}(\beta) - EG_{nij}(\beta)]$, $(nh_n^5)^{1/2}[G_{nijk}(\beta) - EG_{nijk}(\beta]$, $(nh_n^7)^{1/2}[G_{nijk\ell}(\beta) - EG_{nijk\ell}(\beta)]$, $(nh_n)^{1/2}[D_n(\beta) - ED_n(\beta)]$, $(nh_n^3)^{1/2}[D_{ni}(\beta) - ED_{ni}(\beta)]$, and $(nh_n^5)^{1/2}[D_{nij}(\beta) - ED_{nij}(\beta)]$ $(i,j,k,\ell = 1,...,q - 1)$. Define W_n^* similarly except with (Y_i,X_i) replaced by (Y_i^*,X_i^*) and β replaced by b_n. Order the components of W_n and W_n^* conformably with those of S_n and S_n^*. Let V_n be the covariance matrix of W_n and V_n^* be the covariance matrix of V_n^* relative to P_n^*. Let w_n and w_n^* be the summands of the components of W_n and W_n^*. These have the form $J(Y,X)K(Z/h)$ and $\tilde{J}(Y,\tilde{X})\tilde{K}(Z_n/h)$. For any τ conformable with w_n, define

$$\rho_{10}(\tau) = -(i/6)h_n^{-1}E[(\tau'w_n)^3]; \quad i = (-1)^{1/2} \tag{7.17}$$

$$\rho_{20}(\tau) = (1/24)h_n^{-1}E[(\tau'w_n)^4] - (1/72)\{h_n^{-1}E[(\tau'w_n)^3]\}^2, \tag{7.18}$$

and

$$\rho_{21}(\tau) = -(1/8)\{h_n^{-1}E[(\tau'w_n)^2]\}^2. \tag{7.19}$$

Define $\rho_{10}^*(\tau)$, $\rho_{20}^*(\tau)$, and $\rho_{21}^*(\tau)$ by replacing w_n with w_n^* and E with E_n in (7.17)-(7.19). Let π_{10}, π_{20}, and π_{21} be the signed measures whose Fourier-Stieljes transforms are

$$\int \exp(i\tau'\xi)d\pi_{jk}(\xi) = \exp(-0.5\tau'V_n\tau)\rho_{jk}(\tau). \tag{7.20}$$

Define π_{10}^*, π_{20}^*, and π_{21}^* analogously by using V_n^* and ρ_{jk}^* in place of V_n and ρ_{jk}. 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

$$\sup_{\alpha \in A} \left| P(W_n \in \alpha) - \Phi_{V_n}(\alpha) - \sum_{j=1}^{2} \sum_{k=0}^{j/2} (nh_n)^{-j/2} h_n^k \pi_{jk}(\alpha) \right|$$

$$= O[(nh_n)^{-3/2}],$$

and almost surely (P)

$$\sup_{\alpha \in A} \left| P_n^*(W_n^* \in \alpha) - \Phi_{V_n^*}(\alpha) - \sum_{j=1}^{2} \sum_{k=0}^{j/2} (nh_n)^{-j/2} h_n^k \pi_{jk}^*(\alpha) \right|$$

$$= O[(nh_n)^{-3/2}].$$

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 (7.20) to obtain

$$\pi_{jk}(\xi) = \tilde{\pi}_{njk}(\xi)\phi_{V_n}(\xi), \tag{7.21}$$

where for each n, j, and k, $\tilde{\pi}_{njk}(\bullet)$ 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. By Proposition 1, t in (3.1) can be replaced with $\tilde{t}(W_n) \equiv (nh_n)^{1/2}\Lambda[S_n(W_n)]$, and it suffices to consider $P(t \le \tau)$. By lemma 11 and (7.21)

$$P(t \le \tau) = \int_{\{\xi:\ t(\xi) \le \tau\}} d[\Phi_{V_n}(\xi) + \sum_{j=1}^{2} \sum_{k=0}^{j/2} (nh_n)^{-j/2} h_n^k \tilde{\pi}_{njk}(\xi)\phi_{V_n}(\xi)]$$

$$+ O[(nh_n)^{-3/2}]. \qquad (7.22)$$

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 ξ denote the vector consisting of all components of ξ except the first, ξ_1. Change variables in the integral of (7.22) so that the variable of integration is $(t,\xi')'$, thereby obtaining

$$P(t \le \tau) = \int_{t \le \tau} dt \int d\xi T[\xi_1(t,\xi),\xi]\{\Phi_{V_n}[\xi_1(t,\xi),\xi]$$

$$+ \sum_{j=1}^{2} \sum_{k=0}^{j/2} (nh_n)^{-j/2} h_n^k \tilde{\pi}_{njk}[\xi_1(t,\xi),\xi]\phi_{V_n}[\xi_1(t,\xi),\xi]\}$$

$$+ O[(nh_n)^{-3/2}], \qquad (7.23)$$

uniformly over τ, where $T(\bullet)$ is the inverse Jacobian term associated with the change of variables. Taylor series expansions of the terms involving $\xi_1(t,\xi)$ in (7.23) yield

$$P(t \le \tau) = \Phi(\tau) + \sum_{j=1}^{2} (nh_n)^{-j/2} q_{nj}(\tau)\phi(\tau) + o[(nh_n)^{-1}]$$

$$\equiv G_n(\tau) + o[(nh_n)^{-1}] \qquad (7.24)$$

uniformly over τ, where Φ and ϕ, respectively, are the univariate standard normal distribution and density functions, and the q_{nj}'s are polynomial functions of one variable. Let ψ_t and $_G\psi$, respectively, denote the characteristic functions of the distribution of t and G_n. Then $|\psi_t(\tau) - \psi_G(\tau)| = o[(nh_n)^{-1}]$. A Taylor series expansion shows that t in (7.24) can be replaced by a multivariate polynomial in components of 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_t in terms of the approximate cumulants yields $\psi_t(\tau) = \tilde{\Psi}_t(\tau) + o[(nh_n)^{-1}]$ uniformly over τ, where

$$\tilde{\Psi}_t(\tau) = [\exp(-\tau^2/2)]\{1 + i\tau k_{n1} + (1/2)(i\tau)^2(k_{n2} - 1) + (1/6)(i\tau)^3 k_{n3} + (1/24)(i\tau)^4 k_{n4} + (1/2)[(i\tau)k_{n1} + (1/6)(i\tau)^3 k_{n3}]^2\}. \quad (7.25)$$

Setting $\psi_G = \tilde{\Psi}_t$, taking the inverse Fourier transform of the result, and setting $P(|t| \le \tau) = P(t \le \tau) - P(t \le -\tau)$ 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, $k_{n2} - 1$, $k_{n1}k_{n3}$, k_{n3}^2 and k_{n4} are linear combinations of the terms shown in Table 2. The proof is completed by verifying that these terms are all $o[(nh_n)^{-1/2}]$ and no smaller than $O[(nh_n)^{-1}]$. 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

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

uniformly over z and

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

almost surely (P). Set $t(W_n) = (nh_n)\Lambda_\chi[S_n(W_n)]$. By arguments similar to those used to obtain (7.22),

$$P(\chi^2 \le z) = \int_{\{\xi:\ t(\xi) \le z\}} d[\Phi_{V_n}(\xi) + \sum_{j=1}^{2}\sum_{k=0}^{j/2}(nh_n)^{-j/2}h_n^k\tilde{\pi}_{njk}(\xi)\phi_{V_n}(\xi)] + O[(nh_n)^{-3/2}].$$

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} $(i = 1,...,4)$. This is obtained from k_{ni} by replacing β with b_n and expected values with sample averages. By Theorem (4.1),

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

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

$$= O[(k_{n1}^*)^2 - k_{n1}^2] + O(k_{n2}^* - k_{n2}) + O(k_{n1}^*k_{n3}^* - k_{n1}k_{n3})$$

$$+ O[(k_{n3}^*)^2 - k_{n3}^2] + O(k_{n4}^* - k_{n4}).$$

The proof is completed by showing that the difference between each of the terms in Table 2 and its bootstrap analog is $o[(nh_n)^{-1}]$ almost surely. To illustrate, consider $(k_{n1}^*)^2 - k_{n1}^2$. Define $\mu_n(b) = E[h_n^{-1} XX'K^{(1)}(b'X/h_n)^2]$ for any $b \in B$. Define ν_{11j} $(j = 1,...,4)$ as in Table 2, and let ν_{11j}^* be its bootstrap analog. Let $g_1(\tilde{x}) = g_3(\tilde{x}) = \tilde{x}$ and $g_2(\tilde{x}) = g_4(\tilde{x}) = \tilde{x}'$. Then $E(\nu_{111}\nu_{112}) = (nh_n)^{-1}\mu_n(\beta)$. By Theorem 2.37 of Pollard (1984)

$$E_n(\nu_{111}^*\nu_{112}^*) = (nh_n)^{-2}\sum_{i=1}^{n} X_i X_i' K^{(1)}(b_n'X_i/h_n)^2$$

$$= (nh_n)^{-1}\mu_n(b_n) + o[(nh_n)^{-1}]$$

almost surely. Since $\mu_n(b) = O(1)$ uniformly over $b \in B$, $(k_{n1}^*)^2 - k_{n1}^2 = (nh_n)^{-1}[\mu_n(b_n)^2 - \mu_n(\beta)^2] + o[(nh_n)^{-1}]$ almost surely. But $\mu_n(b_n) - \mu_n(\beta) = o(1)$ almost surely by (7.3) and continuity of $\mu_n(b)$. Similar arguments apply to the other terms in Table 2. Q.E.D.

TABLE 1: MONTE CARLO ESTIMATES OF LEVELS OF t TESTS USING ASYMPTOTIC AND BOOTSTRAP CRITICAL VALUES[a]

		Empirical Level Using					
		Asymp. Crit. Val. and Bandwidth			Boot. Crit. Val. and Bandwidth		
Model	Nominal Level	h_{n0}	$0.7h_{n0}$	$0.5h_{n0}$	h_{n0}	$0.7h_{n0}$	$0.5h_{n0}$
Logit	0.10	0.182	0.215	0.328	0.103	0.090	0.111
	0.05	0.115	0.176	0.255	0.058	0.062	0.061
	0.01	0.049	0.097	0.178	0.011	0.016	0.011
t3	0.10	0.238	0.318	0.466	0.097	0.106	0.112
	0.05	0.177	0.294	0.387	0.066	0.055	0.053
	0.01	0.076	0.171	0.279	0.018	0.013	0.012
Hetero.	0.10	0.173	0.220	0.318	0.105	0.101	0.086
	0.05	0.123	0.153	0.266	0.064	0.055	0.044
	0.01	0.040	0.101	0.200	0.011	0.012	0.007

[a] An asymptotic 0.01-level t test based on the normal approximation to the binomial distribution rejects the hypothesis that the true and nominal levels are equal for all experiments with asymptotic critical values. The same test accepts the hypothesis that the true and nominal levels are equal for all experiments with bootstrap critical values.

TABLE 2: TERMS OF APPROXIMATE CUMULANTS

Notation: $g_j(\tilde{x})$, j an integer, is a product of components of $\tilde{x}$ that may be different in different occurrences. For r = 0 or 1 and d = 1, 2, or 3, $m_{rdj}(y,z,x) = J(Y)^r h_n^{-d} K^{(d)}(z/h_n) g_j(\tilde{x})$, $\mu_{rdj} = Em_{rdj}(Y,Z,\tilde{X})$, and $\nu_{rdj} = n^{-1}[m_{rdj}(Y,Z,\tilde{X}) - \mu_{rdj}]$.

Cumulant	Terms
k_{n1}^2	$(nh_n)n^2[E(\nu_{111}\nu_{112})E(\nu_{113}\nu_{114})]$
$k_{n2} - 1$	$(nh_n)n[E(\nu_{111}\nu_{112}\nu_{013})]$, $(nh_n)n[E(\nu_{111}\nu_{112}\nu_{123})]$,
	$(nh_n)n^2[E(\nu_{111}\nu_{112})E(\nu_{113}\nu_{134})]$, $(nh_n)n^2[E(\nu_{111}\nu_{112})E(\nu_{123}\nu_{124})]$,
	$(nh_n)n^2[E(\nu_{111}\nu_{122})E(\nu_{113}\nu_{124})]$, $(nh_n)n^2[E(\nu_{111}\nu_{112})E(\nu_{113}\nu_{124})]$,
	$(nh_n)n^2[E(\nu_{111}\nu_{111})E(\nu_{123}\nu_{014})]$, $(nh_n)n^2[E(\nu_{111}\nu_{112})E(\nu_{113}\nu_{114})]$,
	$(nh_n)n^2[E(\nu_{111}\nu_{112})E(\nu_{113}\nu_{024})]$, $(nh_n)n^2[E(\nu_{111}\nu_{112})E(\nu_{013}\nu_{014})]$
$k_{n1}k_{n3}$	$n^2(nh_n)[E(\nu_{111}\nu_{112})E(\nu_{113}\nu_{114})]$,
	$n^3(nh_n)^2[E(\nu_{111}\nu_{112})E(\nu_{113}\nu_{114})E(\nu_{115}\nu_{116})]$,
	$n^3(nh_n)^2[E(\nu_{111}\nu_{112})E(\nu_{113}\nu_{114})E(\nu_{115}\nu_{126})]$,
k_{n3}^2	$n^4(nh_n)^3[E(\nu_{111}\nu_{112})E(\nu_{113}\nu_{114})E(\nu_{115}\nu_{116})E(\nu_{117}\nu_{118})]$,
	$n^4(nh_n)^3[E(\nu_{111}\nu_{112})E(\nu_{113}\nu_{114})E(\nu_{115}\nu_{116})E(\nu_{117}\nu_{128})]$,
	$n^3(nh_n)^2[E(\nu_{111}\nu_{112})E(\nu_{113}\nu_{114})E(\nu_{115}\nu_{116})]$,
	$n^3(nh_n)^2[E(\nu_{111}\nu_{112})E(\nu_{113}\nu_{114})E(\nu_{115}\nu_{126})]$,
	$n^4(nh_n)^3[E(\nu_{111}\nu_{112})E(\nu_{113}\nu_{114})E(\nu_{115}\nu_{126})E(\nu_{117}\nu_{128})]$,
	$n^2(nh_n)[E(\nu_{111}\nu_{112})E(\nu_{113}\nu_{114})]$
k_{n4}^2	$n^4(nh_n)^3[E(\nu_{111}\nu_{112})E(\nu_{113}\nu_{114})E(\nu_{115}\nu_{116})E(\nu_{117}\nu_{118})]$,
	$n^4(nh_n)^3[E(\nu_{111}\nu_{112})E(\nu_{113}\nu_{114})E(\nu_{115}\nu_{116})E(\nu_{117}\nu_{128})]$,
	$n^3(nh_n)^2[E(\nu_{111}\nu_{112})E(\nu_{113}\nu_{114})E(\nu_{115}\nu_{116})]$,
	$n^4(nh_n)^3[E(\nu_{111}\nu_{112})E(\nu_{113}\nu_{114})E(\nu_{115}\nu_{126})E(\nu_{117}\nu_{128})]$,
	$n^3(nh_n)^2[E(\nu_{111}\nu_{112})E(\nu_{113}\nu_{114})E(\nu_{115}\nu_{126})]$,
	$n^2(nh_n)[E(\nu_{111}\nu_{112})E(\nu_{113}\nu_{114})]$

REFERENCES

Beran, R. (1988). Prepivoting Test Statistics: A Bootstrap View of Asymptotic Refinements, *Journal of the American Statistical Association*, 83, 687-697.

Cavanagh, C.L. (1987). Limiting Behavior of Estimators Defined by Optimization, unpublished manuscript.

Chandra, T.K. and Ghosh, J.K. (1979). Valid Asymptotic Expansions for the Likelihood Ratio Statistic and Other Perturbed Chi-Square Variables, *Sankhya*, 41, Series A, 22-47.

Cosslett, S.R. (1983). Distribution-Free Maximum Likelihood Estimator of the Binary Choice Model, *Econometrica*, 51, 765-782.

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.

Han, A.K. (1987). Non-Parametric Analysis of a Generalized Regression Model: The Maximum Rank Correlation Estimator, *Journal of Econometrics*, 303-316.

Horowitz, J.L. (1992). A Smoothed Maximum Score Estimator for the Binary Response Model, *Econometrica*, 60, 505-531.

Horowitz, J.L. (1995). 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.

Horowitz, J.L. and Härdle, W. (1996). Direct Semiparametric Estimation of Single-Index Models with Discrete Covariates, *Journal of the American Statistical Association*, forthcoming.

Kim, J. and Pollard, D. (1990). Cube Root Asymptotics, *Annals of Statistics*, 18, 191-219.

Klein, R.L., and Spady, R.H. (1993). An Efficient Semiparametric Estimator for Discrete Choice Models, *Econometrica*, 61, 387-422.

Manski, C.F. (1975). Maximum Score Estimation of the Stochastic Utility Model of Choice, *Journal of Econometrics*, 3, 205-228.

Manski, C.F. (1985). Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator, *Journal of Econometrics*, 32, 65-108.

McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior, in *Frontiers in Econometrics*, P. Zarembka, ed., New York: Academic Press, 105-142.

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., Stock, J.H., and Stoker, T.M. (1989). Semiparametric Estimation of Index Coefficients, *Econometrica*, 51, 1403-1430.

Sherman, R.P. (1993), The Limiting Distribution of the Maximum Rank Correlation Estimator, *Econometrica*, 61, 123-138.