Testing a Parametric Model Against a Semiparametric Alternative

Horowitz, J.; Härdle, W.K.

Publication date:
1992

entER for
omic Research

Discussion paper

TILBURG UNIVERSITY

No. 9219

TESTING A PARAMETRIC MODEL AGAINST A SEMIPARAMETRIC ALTERNATIVE

by Joel L. Horowitz
and Wolfgang Härdle

June 1992

ISSN 0924-7815

TESTING A PARAMETRIC MODEL AGAINST A SEMIPARAMETRIC ALTERNATIVE

by

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

Wolfgang Härdle
CORE and Institute de Statistique
Universite Catholique de Louvain
Louvain-la-Neuve, Belgium

March 1992

ABSTRACT

This paper describes a method for testing a parametric model of the mean of a random variable Y conditional on a vector of explanatory variables X against a semiparametric alternative. The test is motivated by a conditional moment test against a parametric alternative and amounts to replacing the parametric alternative model with a semiparametric estimator. The resulting semiparametric test is consistent against a larger set of alternatives than are parametric conditional moments tests based on finitely many moment conditions. The results of Monte Carlo experiments and an application illustrate the usefulness of the new test.

We thank Whitney Newey, Theo Nijman and Tom Stoker for helpful comments. Parts of this research were carried out while both authors were visitors at the Center for Economic Research, Catholic University of Brabant, Tilburg, the Netherlands. The research of Joel L. Horowitz was supported in part by NSF grant No. SES-8922460.

TESTING A PARAMETRIC MODEL AGAINST A SEMIPARAMETRIC ALTERNATIVE

1. INTRODUCTION

Consider a parametric model for the mean of a scalar random variable Y conditional on a random variable $X \in \mathbb{R}^L$ ($L \geq 1$):

$$E(Y|X=x) = f(x,\theta), \tag{1}$$

where f is a known function and $\theta \in \mathbb{R}^K$ ($K \geq 1$) is a parameter whose value must be estimated from data. For example, f might be the mean function in a linear or nonlinear regression model, or it might be the probability that $Y = 1$ conditional on $X = x$ in a parametric binary response model. The problem addressed in this paper is to test the hypothesis that (1) is true for the specified function f and some θ.

One way of testing (1) is to specify a parametric alternative to it and test $f(x,\theta)$ against the alternative. Most familiar methods for testing (1) against a parametric alternative belong to a large class called conditional moments tests (Newey 1985). These tests can have high power against specific alternatives, but a parametric conditional moments test based on finitely many moment conditions is not consistent against all alternatives. In particular, a test of $f(x,\theta)$ against a parametric alternative model may be inconsistent if the alternative is misspecified.

A second possibility is to compare the parametric model with a nonparametric estimate of $E(Y|X=x)$. Let $\hat{\theta}_n$ denote a $n^{1/2}$-consistent estimator of θ in (1) based on a random sample of the distribution of (Y,X). If (1) is true, the nonparametric estimate and $f(x,\hat{\theta}_n)$ are equal up to random sampling error. See Härdle and Mammen (1990), le Cessie and van Houwelingen (1991), and Whang and Andrews (1991) for specification tests based on this idea. Bierens

(1990) gives a conditional moments test of a parametric model against a nonparametric alternative. These tests are consistent in all directions but have characteristics that can cause them to have low power or other kinds of poor behavior in finite samples. For example, the tests of Härdle and Mammen (1990) and le Cessie and van Houwelingen (1991) lose power through the so-called curse of dimensionality (Huber 1985) if $L > 1$. The test of Whang and Andrews (1991) requires splitting the sample into two equal parts, which reduces power and can result in poor small-sample behavior.

This paper describes a test that aims at avoiding these problems while achieving consistency against a larger set of alternatives than is the case with parametric conditional moments tests based on finitely many moment conditions. The intuition behind the test is simple. If $E(Y|X=x) = f(x,\theta)$, then

$$E[Y|f(x,\theta)=f] = f. \tag{2}$$

Therefore, a nonparametric estimate of $E[Y|f(X,\hat{\theta}_n)=f]$, considered as a function of f, differs from a 45° line only by random sampling error. One can test (1) by determining whether the difference between the nonparametric estimate and the 45° line is larger than can be explained by random sampling error.

More generally, consider the model

$$E(Y|X=x) = F[v(x,\theta)], \tag{3}$$

where F and v are known functions. If (3) is correct, nonparametric estimation of $E[Y|v(X,\hat{\theta}_n)=v]$ gives an estimate $F(v)$. Thus, (3) can be tested by comparing the nonparametric estimate of $E[Y|v(X,\hat{\theta}_n)=v]$ with $F(v)$. One way of obtaining (3) is to specify $v(x,\theta) = f(x,\theta)$ and $F(v) = v$, but other specifications may be useful in applications. For example, suppose the parametric model to be tested has the form (3) with F a nonmonotonic function. If the model is misspecified, it is possible that

$$E\{Y|F[v(X,\theta)] = f\} = f \tag{4a}$$

whereas

$$E[Y|v(X,\theta)=v] \neq F(v). \tag{4b}$$

In this case, comparison of a nonparametric estimate of $E\{Y|F[v(X,\theta)]=f\}$ with f yields an inconsistent test, whereas comparison of a nonparametric estimate of $E[Y|v(X,\theta)=v]$ with $F(v)$ yields a consistent test.

A test of (1) obtained by comparing a nonparametric estimate of $E[Y|v(X,\hat{\theta}_n)=v]$ with $F(v)$ avoids the curse of dimensionality by using the index function $v(x,\theta)$ to aggregate a multidimensional x. Because one can always set $v(x,\theta) = f(x,\theta)$ and $F(v) = v$, any model of the form (1) can be placed into the single-index form of (3). Thus, the test is not restricted to models that can be estimated in single-index form.

The remainder of this paper describes a formal test of (1) that consists of comparing a nonparametric estimate of $E[Y|v(X,\hat{\theta}_n)=v]$ with $F(v)$. We call this a test of the parametric model (1) against a semiparametric alternative because the alternatives against which the parametric model is tested and against which the test is consistent have the form $E[Y|v(X,\theta)=v] = H(v)$, where H is an unknown function but $v(x,\theta)$ is known up to the finite-dimensional parameter θ. Because the semiparametric alternative may not include the true mean of Y conditional on X, there are directions in which the semiparametric test is inconsistent. However, in a sense that is defined in Section 2, the test is consistent against a larger set of alternatives than are parametric conditional moments tests based on finitely many moments. The results of Monte

Carlo experiments and an application based on real data illustrate the usefulness of the semiparametric test.

The paper is organized as follows. The test statistic is presented in Section 2, and its asymptotic distributions under the null hypothesis and local alternatives are derived. Section 3 presents the results of the Monte Carlo experiments and the application. Concluding comments are presented in Section 4. The proofs of theorems are in the appendix.

2. THE TEST STATISTIC AND ITS ASYMPTOTIC DISTRIBUTION

a. The Null and Alternative Hypotheses

Formally the null hypothesis that we test is

$$H_0: \quad E[Y|v(X,\theta)=v] \; = \; F(v), \tag{5}$$

where Y is a scalar random variable, $X \in \mathbb{R}^L$, F and $v(\cdot,\cdot)$ are known, real functions, and $\theta \in \mathbb{R}^K$ is a parameter whose value is unknown and estimated from data. For example, if Y follows a linear-index binary probit model under H_0, F and $v(x,\theta)$, may be specified as the cumulative normal distribution function and $\theta'x$, respectively. As was discussed in Section 1, (1) can always be put into the form (5). The alternative hypothesis is

$$H_1: \quad E[Y|v(X,\theta)=v] \; = \; H(v), \tag{6}$$

where H is an unknown function.

$E[Y|v(X,\theta)=v] = F(v)$ is a necessary but not sufficient condition for $E(Y|X=x) = F[v(x,\theta)]$. It is possible that $E[Y|v(x,\theta)=v] = F(v)$ but $E(Y|X=x) \neq F[v(x,\theta)]$, in which case the test of (1) presented here is inconsistent. This possibility is discussed further in Section 2d.

b. Motivation

Suppose for the moment that H and θ were known. Consider a conditional moment test of H_0 against H_1 based on the following moment condition, which is assumed to hold under H_0:

$$E\rho(X,\theta)\{Y - F[v(X,\theta)]\} = 0,$$

where ρ is a scalar function. Let $\{Y_i,X_i: i = 1,\ldots,n\}$ be a random sample of (Y,X). Following Newey (1985), the conditional moment test statistic is proportional to

$$S_n = n^{-1/2}\sum_{i=1}^{n}\rho(X_i,\theta)\{Y_i - F[v(X_i,\theta)]\}.$$

Under H_0, $E(S_n) = 0$. Under H_1, $E(S_n) = n^{1/2}E\rho(X,\theta)\{H[v(X,\theta)] - F[v(X,\theta)]\} = \mu$. The test can be expected to have power against H_1 only if $\mu \neq 0$. This happens if

$$\rho(x,\theta) = w[v(x,\theta)]\{H[v(x,\theta)] - F[v(x,\theta)]\},$$

where $w(\cdot)$ is a non-negative weight function that is chosen so that

$$Ew[v(X,\theta)]\{H[v(X,\theta)] - F[v(X,\theta)]\}^2 > 0. \tag{7}$$

Thus, the conditional moment test in this simple case can be based on the statistic

$$S_n^* = n^{-1/2}\sum_{i=1}^{n} w[v(X_i,\theta)]\{Y_i - F[v(X_i,\theta)]\}\{H[v(X_i,\theta)] - F[v(X_i,\theta)]\}. \tag{8}$$

Since H and θ are unknown, one might consider forming a test of H_0 against H_1 by replacing H and θ in (8) with consistent estimators. This is the approach taken here. We replace θ with the $n^{1/2}$-consistent estimator $\hat{\theta}_n$ and $H[v(X_i,\theta)]$

with a kernel nonparametric regression estimator of $E[Y|v(X,\theta_n)=v(X_i,\theta_n)]$. Denote this estimator by $\hat{F}_{ni}[v(X_i,\theta_n)]$. The test statistic is

$$T_n = h^{1/2}\sum_{i=1}^{n} w[v(X_i,\theta_n)](Y_i - F[v(X_i,\theta_n)])(\hat{F}_{ni}[v(X_i,\theta_n)] - F[v(X_i,\theta_n)]),$$

where h is the bandwidth used in the kernel nonparametric regression. The normalization factor is $h^{1/2}$ rather than $n^{-1/2}$ for technical reasons associated with the rate of convergence in probability of the nonparametric regression estimator. It is shown below that, like the test based on (8), T_n is consistent against H_1 if (7) holds. In contrast to (8), however, T_n does not require *a priori* knowledge of H and θ.

c. The Kernel Nonparametric Regression Estimator

Our methods for proving the theorems in Section 2d require $\hat{F}_{ni}(\cdot)$ to be independent of Y_i and asymptotically unbiased. Independence is achieved by omitting the observation (Y_i,X_i) from the computation of $\hat{F}_{ni}$. Asymptotic unbiasedness is achieved through the use of the jackknife-like method proposed by Bierens (1987). The resulting estimator is as follows.

Let $K(\cdot)$ be the kernel function used in the nonparametric regression. Assume that K is an order r kernel. That is, for each integer i between 0 and $r \geq 2$

$$\int_{-\infty}^{\infty} u^i K(u)du = \begin{cases} 1 \text{ if } i = 0 \\ 0 \text{ if } 1 \leq i \leq r-1 \\ d_K \neq 0 \text{ if } i = r \end{cases}$$

Let $h = cn^{-1/(2r+1)}$, where $c > 0$. Let $s = cn^{-\delta/(2r+1)}$, where $0 < \delta < 1$. Define

$$\hat{F}_{nhi}(v) = \sum_{\substack{j=1 \\ j\neq i}}^{n} Y_j K\left[\frac{v - v(X_j,\theta_n)}{h}\right] \Big/ \sum_{\substack{j=1 \\ j\neq i}}^{n} K\left[\frac{v - v(X_j,\theta_n)}{h}\right], \tag{9}$$

and

$$\hat{F}_{nsi}(v) = \sum_{\substack{j=1 \\ j\neq i}}^{n} Y_j K\left[\frac{v - v(X_j,\theta_n)}{s}\right] \Big/ \sum_{\substack{j=1 \\ j\neq i}}^{n} K\left[\frac{v - v(X_j,\theta_n)}{s}\right]. \tag{10}$$

The kernel nonparametric regression estimator used in T_n is

$$\hat{F}_{ni}(v) = [\hat{F}_{nhi}(v) - (h/s)^r \hat{F}_{nsi}(v)]/[1 - (h/s)^r] \tag{11}$$

Bierens (1987) derives the properties of this estimator and proves that it is asymptotically unbiased and has the optimal rate of convergence.

d. The Asymptotic Distribution of T_n

Define $\sigma^2(v) = \mathrm{var}[Y|v(X,\theta)=v]$. The following theorem gives the asymptotic distribution of T_n under H_0:

Theorem 1: *Under* H_0 *and assumptions 1-8 of the appendix,* T_n *is asymptotically distributed as* $N(0,\sigma_T^2)$, *where*

$$\sigma_T^2 = 2C_K \int_{-\infty}^{\infty} w(v)^2 [\sigma^2(v)]^2 dv$$

and

$$C_K = \int_{-\infty}^{\infty} K(u)^2 du.$$

The proof of Theorem 1 is lengthy, but the concepts on which it is based are easily described. First, the rate of convergence in probability of the

$n^{1/2}$-consistent estimator $\hat{\theta}_n$ is faster than the rate of convergence in probability of $\hat{F}_{ni}$, which is $(nh)^{-1/2}$. As a result, the asymptotic distribution of T_n is unaffected by replacing $\hat{\theta}_n$ with θ. Thus, $T_n = T_n^* + o_p(1)$, where

$$T_n^* = h^{1/2} \sum_{i=1}^{n} w[v(X_i,\theta)]\{Y_i - F[v(X_i,\theta)]\}(F_{ni}[v(X_i,\theta)] - F[v(X_i,\theta)])$$

and $F_{ni}[v(X_i,\theta)]$ is the nonparametric regression estimator obtained by replacing $v(X_i,\hat{\theta}_n)$ with $v(X_i,\theta)$ in (9)-(11). See Lemmas 1-6 of the appendix for the proof of this result. Second, it can be shown that T_n^* is asymptotically equivalent to a certain degenerate U statistic. See Lemma 7 and the proof of Theorem 1 in the appendix. Although degenerate U statistics ordinarily are asymptotically distributed as linear combinations of χ^2 variates (see, for example, Serfling 1980), the one corresponding to T_n^* has a special form that causes it to be asymptotically normally distributed by a central limit theorem of Hall (1984). Theorem 1 is a consequence of the asymptotic normality of this U statistic.

Let $\hat{\sigma}_T{}^2$ be a consistent estimator of $\sigma_T{}^2$, and let $\hat{\sigma}_T = (\hat{\sigma}_T{}^2)^{1/2}$. It follows from Theorem 1 that H_0 can be accepted or rejected at the ς level according to whether $|T_n/\hat{\sigma}_T|$ exceeds the $1 - \varsigma/2$ quantile of the standard normal distribution. Let $\hat{\sigma}^2(v)$ be a consistent estimator of $\sigma^2(v)$. Then, under assumptions 1-8 of the appendix $\sigma_T{}^2$ is estimated consistently by

$$\hat{\sigma}_T{}^2 = (2C_K/n) \sum_{i=1}^{n} w[v(X_i,\hat{\theta}_n)]^2\{\hat{\sigma}^2[v(X_i,\hat{\theta}_n]\}^2/\hat{p}_{nhi}[v(X_i,\hat{\theta}_n],$$

where $\hat{p}_{nhi}$ is the following nonparametric estimator of the probability density function of $v(X_i,\theta)$:

$$\hat{p}_{nhi}(v) = (nh)^{-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} K\left[\frac{v - v(X_j, \theta_n)}{h}\right].$$

Methods for estimating $\sigma^2(v)$ are discussed in Section 2e.

The next theorem establishes consistency of T_n under H_1.

<u>Theorem 2</u>: *Let assumptions 1-8 of the appendix hold. Suppose that* H_1 *is true and that* $Ew(V)\{[H(V) - F(V)]^2\} > 0$, *where* $V = v(X,\theta)$ *and* θ *is the probability limit of* θ_n. *Then* $\text{plim}_{n\to\infty} T_n = \infty$.

Suppose that H_0 is false and that $E(Y|X=x) = H^*(x)$ for some function H^*. Let $E_{X|v}$ denote expectation relative to the distribution of X conditional on $v(X,\theta) = v$. It follows from Theorem 2 that the test based on T_n is consistent if $E_{X|v} H^*(X) \neq F(v)$ on a subset of the support of $w[v(X,\theta)]$ that has positive probability. The test is inconsistent if $P\{H^*(x) = F[v(x,\theta)]\} < 1$ but $E_{X|v} H^*(X) = F(v)$ almost everywhere on the support of $w(\cdot)$.

Although T_n is not consistent against all alternatives, there is a sense in which it is consistent against a larger set of alternatives than are parametric conditional moments tests based on finitely many moment conditions. Specifically, T_n is consistent against all alternatives $H[v(x,\theta)]$ such that $Ew[v(X,\theta)]\{H[v(X,\theta)] - F[v(X,\theta)]\}^2 > 0$, whereas a parametric conditional moments test is not. To see this, observe that if (3) is true, then

$$E\rho(X,\theta)\{Y - F[v(X,\theta)]\} = 0 \tag{12}$$

for any function $\rho \in \mathbb{R}^q$ for some finite $q > 0$. Accordingly, consider using the moment conditions (12) to test (1). Suppose that $E[Y|v(X,\theta)=v] = H(v)$, where H satisfies $Ew[v(X,\theta)]\{H[v(X,\theta)] - F[v(X,\theta)]\}^2 > 0$ and $E\rho(X,\theta)\{H[v(X,\theta)] - F[v(X,\theta)]\} \neq 0$. Then T_n and the conditional moments test based on (12) are both consistent against the alternative H. Now let $\Delta(v)$ be a scalar-valued function such that $E\rho(X,\theta)\Delta[v(X,\theta)] = 0$. Assume that $Ew[v(X,\theta)]\Delta[v(X,\theta)]^2 > 0$. Since ρ is finite-dimensional, there are infinitely many such functions Δ. Set $H^*(v) = F(v) + \Delta(v)$. Then T_n is consistent against H^* but the conditional moments test based on (12) is not.

We now consider the distribution of T_n under local alternative hypotheses. Define the sequence of local alternatives $H_n[v(x,\theta)]$, by

$$H_n[v(x,\theta)] = F[v(x,\theta)] + n^{-1/2}h^{-1/4}\Delta_n[v(x,\theta)],$$

where $\{\Delta_n: n = 1,2,\ldots\}$ is a sequence of uniformly bounded functions that converges uniformly to a limit function $\Delta(v)$. Note in this sequence $|H_n(v) - F(v)| = O(n^{-1/2}h^{-1/4})$ uniformly over v, whereas in tests of parametric models against local parametric alternatives the "distance" between the null and local alternative hypotheses is $O(n^{-1/2})$. Let $\hat{\theta}_n$ be an estimator of θ that is $n^{1/2}$-consistent under the sequence $\{H_n\}$.

<u>Theorem 3</u>: *Let assumptions 1-5 and 8-12 of the appendix hold. Under the sequence of local alternative models H_n, T_n is asymptotically distributed as $N(\mu,\sigma_T^2)$, where $\mu = E[w(V)\Delta(V)^2]$.*

Theorem 3 implies that T_n has power against alternatives whose distance from H_0 is $O(n^{-1/2}h^{-1/4})$. If K is a second order kernel, this distance is $O(n^{-9/20})$, which is close to the distance $O(n^{-1/2})$ that holds in tests against parametric alternative hypotheses. Subject to the regularity conditions given in the appendix, the distance $O(n^{-1/2}h^{-1/4})$ can be made arbitrarily close to $O(n^{-1/2})$ by using a kernel K of sufficiently high order.

e. Choosing $w(\cdot)$ and $\hat{\sigma}^2(v)$

The regularity conditions in the appendix require $w(\cdot)$ to be continuous and independent of the sample $\{Y_i, X_i\}$. They also require the support of $w(\cdot)$ to be contained within that of $v(X,\theta)$. The continuity requirement is not important in applications; with a finite sample there is no difference between the values of T_n obtained with a $w(\cdot)$ that has jump discontinuities and a $w(\cdot)$ in which the discontinuities have been "slightly" smoothed. The restriction on the support of $w(\cdot)$ can be important. Depending on how $\sigma_T{}^2$ is estimated, use of a $w(\cdot)$ whose support exceeds that of $v(X,\theta)$ may cause substantial overestimation of $\sigma_T{}^2$ and a corresponding loss of power. In practice it can be difficult to choose a $w(\cdot)$ that satisfies the condition on support without looking at the data. We suggest using the observed values of $v(X_i, \hat{\theta}_n)$ to choose the support of $w(\cdot)$ but not otherwise adjusting $w(\cdot)$ to the data. In the Monte Carlo experiments and application described in Section 3, we found that the T_n test works well if w is chosen to be 1 over an interval that contains 95-99% of the observed values of $v(X,\hat{\theta}_n)$ and 0 elsewhere.

Another possibility is to choose w to maximize power against a specified sequence of local alternatives. There seems to be little advantage in doing this, however. If high power against a specific alternative is desired, one should use a parametric conditional moments test that has high power against this alternative.

The main consideration involved in estimating $\sigma^2(v)$ is that the estimator must be consistent under H_0 and, to avoid loss of power, should not become excessively large under H_1. For example, suppose that Y is homoskedastic so that $\mathrm{var}[Y|v(X,\theta) = v] = \sigma^2$, where σ^2 is a constant. Two possible estimators of σ^2 are:

$$\hat{\sigma}_1^{\,2} = n^{-1}\sum_{i=1}^{n}(Y_i - F[v(X_i,\hat{\theta}_n)])^2 \tag{13}$$

and

$$\hat{\sigma}_2^{\,2} = n^{-1}\sum_{i=1}^{n}(Y_i - \hat{F}_{ni}[v(X_i,\hat{\theta}_n)])^2 \tag{14}$$

Both of these estimators are consistent under H_0, but $\hat{\sigma}_1^2$ may be very large under H_1. Accordingly, the test based on T_n is likely to have higher power if $\hat{\sigma}_2^2$ is used.

If Y has heteroskedasticity of unknown form, $\sigma^2(v)$ can be estimated by the nonparametric regression of $(Y_i - \hat{F}_{ni}[v(X_i,\hat{\theta}_n)])^2$ on $v(X_i,\hat{\theta}_n)$. In some cases the form of heteroskedasticity of Y may be known, and this information can be used to estimate $\sigma^2(v)$. For example, if Y is a binary variable, $\mathrm{var}[Y|v(X,\theta)=v] = P[Y=1|v(X,\theta)=v](1 - P[Y=1|v(X,\theta)=v])$. Therefore, $\sigma^2(v)$ can be estimated by $\hat{F}[v(X_i,\hat{\theta}_n)](1 - \hat{F}_{ni}[v(X_i,\hat{\theta}_n)])$.

3. MONTE CARLO EXPERIMENTS AND AN APPLICATION

a. Monte Carlo Experiments

The purpose of the Monte Carlo experiments was to investigate the small-sample size and power of the test based on T_n. To provide a basis for judging whether the performance of the test is good or bad, we also computed the size and power of Bierens' (1990) test against a nonparametric alternative and of the most powerful test against the correct parametric alternative model.[1]

The hypothesis H_0 tested in the Monte Carlo experiments is

$$E(Y|X=x) = \theta_0 + \theta_1 x, \tag{15}$$

where X is a scalar random variable and the θ's are constant parameters. The data were generated by random sampling from the model

$$Y = \theta_0 + \theta_1 X + 10b\phi(10X) + u \tag{16}$$

where ϕ is the standard normal density function, $\theta_0 = \theta_1 = 1$, b is a parameter whose value varies according to the experiment, $X \sim N(0,1)$ and $u \sim N(0,0.25)$. If $b = 0$, H_0 is true. Otherwise, H_0 is false and $E(Y|X=x)$ has the shape of a straight line with a bump centered at $x = 0$. The height of the bump is governed by the value of b. Figure 1 illustrates the shape of $E(Y|X=x)$ for $b = 1$ and $b = 2$. The mean function $E(Y|X=x)$ in (16) is poorly approximated by the parametric models typically used in applications (e.g., low-order polynomials in x), so it is unlikely that a most powerful or nearly most powerful parametric test of (15) would be carried out in an application if (16) were the true data-generation process. Härdle (1990) gives several applications in which the shape of $E(Y|X=x)$ is similar to Figure 1.

In the computation of T_n in the experiments, K is the standard normal density, $v(x,\theta) = \theta' x$, $F(v) = v$, $h = 0.1$, and $s = 0.8$. θ_0 and θ_1 were computed from (15) by ordinary least squares (OLS), $w(\cdot) = 1$ on an interval containing 98% of observed values of $\hat{\theta}_0 + \hat{\theta}_1 X$ and 0 elsewhere, and $\sigma^2(v)$ is given by (14).[2]

Implementation of the test of Bierens (1990) requires choosing several parameters of the test statistic and a function. We made choices similar to those used in Bierens' (1990) Monte Carlo experiments. In his notation, we set $\gamma = \rho = 0.5$, $T = [1,5]$, $K_n = 10$, and $\phi(x) = \tan^{-1}(x/2)$. Note that Bierens' ϕ is different from ϕ in (16).

The most powerful parametric test of the null hypothesis (15) against the alternative (16) is the t test of $b = 0$ based on OLS estimation of θ_0, θ_1, and b in (16).

The experiments were carried out at the nominal 0.05 level using a sample size of $n = 50$. There were 500 replications in each experiment. Random numbers were generated with the pseudo-random number generators of GAUSS.

Table 1 shows the results of the experiments. The empirical sizes of the tests are not statistically significantly different from the nominal size of 0.05 ($p > 0.10$). The test based on T_n is considerably more powerful than Bierens' test. Not surprisingly, T_n has less power than the most powerful parametric test. Of course, the power of the parametric test would be available in an application only in the unlikely event that (16) were known to be the correct alternative model, whereas T_n does not require *a priori* knowledge of the alternative.

b. An Application

Horowitz (1991) estimated a binary probit model of the choice between automobile and transit for the trip to work. The estimation data set consisted of 842 trip records drawn from the Washington, D.C., area transportation study. The specification of the probit model is

$$P(\text{Auto}|X=x) = \Phi(\theta'x), \tag{17}$$

where Φ is the cumulative normal distribution function, X is a vector of explanatory variables, and θ is a conformable vector of estimated parameters. The components of X are an intercept, the number of automobiles owned by the traveler's household, the difference between automobile and transit out-of-vehicle travel time, the difference between automobile and transit in-vehicle travel time, and the difference between automobile and transit travel cost. Horowitz (1991) carried out parametric likelihood ratio, Wald and Lagrangian multiplier tests of (17) against a random-coefficients probit model. This model is obtained from (17) by replacing $\Phi(\beta'x)$ with $\Phi[\beta'x/(x'\Sigma x)^{1/2}]$, where Σ is a positive-definite matrix. All of the tests rejected (17) ($p < 0.01$).

To investigate the performance of T_n in an application, we tested (17) using both T_n and Bierens' (1990) test. Bierens' test was carried out using the parameter and function choices described in Section 3a. The value of the test statistic was 0.43. Under the hypothesis that (17) is correctly specified, Bierens' test statistic is asymptotically distributed as χ^2 with 1 degree of freedom. Therefore, Bierens' test does not reject (17) and, thus, does not detect the misspecification of (17) found by the tests against the random-coefficients probit model.

In computing the T_n test statistic, θ_n was estimated by maximum likelihood using (17), $v(x,\theta) = \theta'x$, $w(\cdot) = 1$ on an interval containing 98% of the observed values of $\theta_n'X$ and 0 elsewhere, and $\hat{\sigma}^2(v) = \hat{F}_{ni}(v)[1 - \hat{F}_{ni}(v)]$. As is explained in footnote 2, there is no known systematic method for selecting bandwidth values for $\hat{F}_{ni}$. We used several bandwidths that were found through graphical examination of $\hat{F}_{ni}$ to span the range of reasonable choices. Values outside of this range caused the graph of $\hat{F}_{ni}$ to be either excessively wiggly or excessively flat. The value of the T_n test statistic was in the range 2.45-3.26, depending on the bandwidth. Thus T_n rejects (17) ($p = 0.001$ to 0.015, depending on the bandwidth). This is consistent with the results of the tests against the random-coefficients probit model.

4. CONCLUSIONS

This paper has described a method for testing a parametric model of the conditional mean against a semiparametric alternative. The test is motivated by a parametric conditional moments test and amounts to replacing the parametric alternative model in the conditional moments test with a semiparametric estimator. The resulting semiparametric test is not consistent against all alternatives, but in a sense that has been explained it is consistent against a larger set of alternatives than are parametric conditional moments tests based on finitely many moment conditions. The results of Monte Carlo experiments and an application using real data illustrate the usefulness of the semiparametric test.

FOOTNOTES

1. We originally intended to include the test of Whang and Andrews (1991) in the comparison. This test is based on comparing the mean square residual from parametric and nonparametric estimates of the conditional mean of Y. We dropped the test from consideration after finding that, in our Monte Carlo experiments, its empirical size at the nominal 0.05 level was between 0.24 and 0.50 for a wide range of bandwidths in the nonparametric regression.

2. A systematic procedure for choosing h and s for $\hat{F}_{ni}$ with finite samples has not been developed. Because the estimator is asymptotically unbiased, the tradeoff between asymptotic bias and variance that underlies bandwidth selection methods such as cross validation does not exist. We selected h and s graphically. With the values we used, the graph of $\hat{F}_{ni}$ is neither excessively wiggly, as happens when h and s are too small, nor excessively flat, as happens when they are too large. The regularity conditions in the appendix require K to have bounded support, but this is not essential, as is noted there.

REFERENCES

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

Bierens, H.J. (1987). Kernel estimators of regression functions. In T.F. Bewley, ed., *Advances in Econometrics: Fifth World Congress*, vol. 1, Cambridge University Press, New York, 99-144.

Bierens, J.J. (1990). A consistent conditional moment test of functional form. *Econometrica*, 58, 1443-1458.

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of Multivariate Analysis*, 14, 1-16.

Härdle, W. (1990). *Applied Nonparametric Regression*, Cambridge University Press, New York.

Härdle, W. and Mammen, E. (1990). Comparing nonparametric versus parametric regression fits. CORE discussion paper no. 9065, Universite Catholique de Louvain, Louvain-la-Neuve, Belgium.

Horowitz, J.L. (1991). Semiparametric estimation of a work-trip mode choice model. *Journal of Econometrics*, forthcoming.

Huber, P.J. (1985). Projection pursuit. *The Annals of Statistics*, 13, 435-475.

le Cessie, S. and van Houwelingen, J.C. (1991). A goodness-of-fit test for binary regression models, based on smoothing methods. *Biometrics*, 47, 1267-1282.

Newey, W.K. (1985). Maximum likelihood specification testing and conditional moment tests. *Econometrica*, 53, 1047-1070.

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, New York.

Silverman, B.W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *The Annals of Statistics*, 6, 177-184.

Whang, Y.-J. and Andrews, D.W.K. (1991). Tests of specification for parametric and semiparametric models. Cowles Foundation discussion paper no. 968, Yale University, New Haven, CT.

TABLE 1: RESULTS OF THE MONTE CARLO EXPERIMENTS

	Pr(Reject H_0 at Nominal 0.05 Level)		
b	Most Powerful Parametric Test	T_n[a]	Bierens' Test
0	0.05	0.04	0.05
0.25	0.90	0.37	0.18
0.50	0.99	0.90	0.42
0.75	1.00	0.96	0.56
1.00	1.00	0.98	0.73
1.25	1.00	1.00	0.78
1.50	1.00	0.99	0.78
1.75	1.00	1.00	0.81
2.00	1.00	1.00	0.87
2.25	1.00	0.99	0.89
2.50	1.00	1.00	0.89

[a] The fluctuations in the rejection probability when $b \geq 1.25$ are not statistically significant at the 0.10 level.

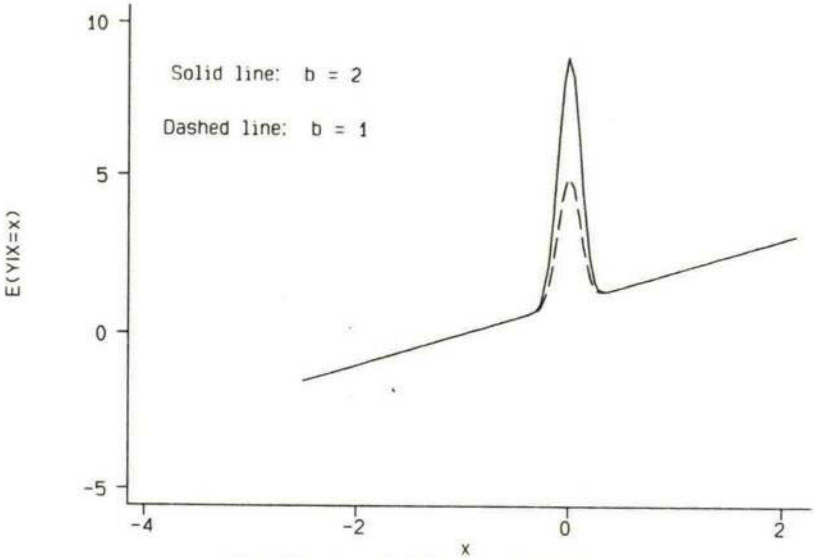

FIGURE 1: E(YIX=x) IN (16)

MATHEMATICAL APPENDIX

A1. NOTATION

In addition to the notation defined in the text:

N_θ - A neigborhood of θ.

S_X - The support of X.

$\bar{S}_v$ - An open subset of the support of $v(X,\theta)$.

S_v - A compact subset of $\bar{S}_v$

$\bar{S}_X$ - $\{x: v(x,\theta) \in S_v\}$.

$h = cn^{-1/(2r+1)}$, where $c > 0$ and $r \geq 2$ is an integer.

$s = cn^{-\delta/(2r+1)}$, where $0 < \delta < 1$.

$K(\cdot)$ - an r'th order kernel function.

For ξ = h or s:

$$p_{n\xi i}(v) = (n\xi)^{-1} \sum_{\substack{j=1 \\ j\neq i}}^{n} K\left[\frac{v - v(X_j,\theta)}{\xi}\right]$$

$$p_{n\xi}(v) = (n\xi)^{-1} \sum_{j=1}^{n} K\left[\frac{v - v(X_j,\theta)}{\xi}\right]$$

$$F_{n\xi i}(v) = \sum_{\substack{j=1 \\ j\neq i}}^{n} Y_j K\left[\frac{v - v(X_j,\theta)}{\xi}\right] \Big/ \sum_{\substack{j=1 \\ j\neq i}}^{n} K\left[\frac{v - v(X_j,\theta)}{\xi}\right]$$

$$F_{ni}(v) = [F_{nhi}(v) - (h/s)^r F_{nsi}(v)]/[1 - (h/s)^r]$$

$$g_{n\xi i}(v) = p_{n\xi i}(v) F_{n\xi i}(v)$$

A2. ASSUMPTIONS

1. S_X is compact. At least one component of X has a probability distribution that is absolutely continuous with respect to Lebesgue measure.

2. For every $\tau \in N_\theta$ and $x \in S_X$, v satisfies:

   a. $|v(x,\tau)| < M$ for some $M < \infty$ that does not depend on τ or x.

   b. The probability distribution of $v(X,\tau)$ is absolutely continuous with respect to Lebesgue measure.

   c. $v(x,\tau)$ is continuously differentiable with respect to τ, and $|\partial v(x,\tau)/\partial \tau_k| < M$ $(k = 1,\ldots,K)$ for some $M < \infty$ that does not depend on τ or x.

3. Let p_τ denote the probability density function of $v(X,\tau)$. For each $\tau \in N_\theta$

   a. $m_p \le p_\tau(v) \le M_p$ for some $m_p > 0$ and $M_p < \infty$ that do not depend on τ.

   b. p_τ has r continuous, derivatives that are uniformly bounded over $\tau \in N_\theta$ and $v \in \tilde{S}_v$.

4. $w(\cdot)$ has compact support $S_w \subset \text{int}(S_v)$ and satisfies:

   a. $0 \le w(v) < M_w$ for some $M_w < \infty$ and all $v \in S_w$.

   b. $|w(v_2) - w(v_1)| \le M_w^*|v_2 - v_1|$ for some $M_w^* < \infty$ and all v_2, v_1.

5. a. $|F[v(x,\tau)]|$ and $|H[v(x,\tau)]|$ are uniformly bounded over $x \in S_X$ and $\tau \in N_\theta$.

   b. F(v) and H(v) have r continuous derivatives that are uniformly bounded over $v \in S_v$.

6. Let E_X denote the expectation over the distribution of X. Define

$$\Gamma(x,v,\tau) = E_X\{F[v(X,\theta)][\partial v(x,\tau)/\partial \tau - \partial v(X,\tau)/\partial \tau]\,|\,v(X,\tau) = v\}.$$

   Let Γ_k $(k = 1,\ldots,K)$ denote the k'th component of Γ. There is a finite number M_Γ, not depending on τ or x, such that for all $\tau \in N_\theta$, $x \in S_X$, v_1, $v_2 \in \tilde{S}_v$, and $k = 1,\ldots,K$

$$|\Gamma_k(x,v_2,\tau) - \Gamma_k(x,v_1,\tau)| \le M_\Gamma|v_2 - v_1|$$

7. $\sigma^2(v) = \text{Var}[Y|v(X,\theta) = v]$ is a uniformly bounded, continuous function of $v \in S_v$. $E\{Y - E[Y|v(x,\theta)]\}^4$ is uniformly bounded over $v \in S_v$.

8. K is an r'th order kernel with bounded support. Also, K is uniformly bounded, continuous, and symmetrical about 0. The derivative of K, K',

is uniformly bounded, has an absolutely integrable Fourier transform, $\psi(\cdot)$, and satisfies

$$\int_{-\infty}^{\infty} |uK'(u)|du < \infty.$$

Comments:

Although theses assumptions are notationally complex, they are mainly boundedness and smoothness conditions. The requirement that K has bounded support can be removed at the cost of additional technical complexity in the proofs.

Several factors make it necessary for the assumptions to involve the nested sets $S_v \subset S_v \subset \bar{S}_v$ in addition to S_X and $\bar{S}_X$. First, the model being tested is formulated in terms of $v(X,\theta)$, not X directly. Second, it is necessary to estimate θ. Third, the derivation of the asymptotic distribution of T_n involves showing that this distribution remains unchanged if θ replaces $\hat{\theta}_n$. To do this, it is necessary to have a way of insuring that for all τ sufficiently close to θ, $v(X,\tau)$ stays away from points on the boundary of its support at which its probability density is discontinuous. The nested sets enable this to be done.

A3. THE ASYMPTOTIC DISTRIBUTION OF T_n UNDER H_0

Preliminary Lemmas:

Lemmas 1-6 show that asymptotically $\hat{\theta}_n$ can be replaced by θ in T_n. Lemma 7 gives a result that is used in deriving the U-statistic form of T_n.

Lemma 1: Define

$$G_{nhi}(v) = [g_{nhi}(v) - F(v)p_{nhi}(v)]/p_\theta(v) \qquad \text{(A1)}$$

and

$$J_{nhi}(v) = \qquad \text{(A2)}$$

$$[p_{nhi}(v) - p_\theta(v)]\{[g_{nhi}(v) - F(v)p_\theta(v)] - F(v)[p_{nhi}(v) - p_\theta(v)]\}/[p_\theta(v)]^2.$$

As $n \to \infty$,

$$\sup_{1 \le i \le n} \sup_{v \in S_v} |p_{nhi}(v) - p_\theta(v)| = O([(\log n)/(nh)]^{1/2}), \quad \text{(A3)}$$

almost surely,

$$\sup_{1 \le i \le n} \sup_{v \in S_v} |g_{nhi}(v) - p_\theta(v)F(v)| = O_p[(nh^2)^{-1/2}], \quad \text{(A4)}$$

and

$$\sup_{1 \le i \le n} \sup_{v \in S_v} |F_{nhi}(v) - F(v) - G_{nhi}(v) + J_{nhi}(v)|.$$

$$= O_p[(\log n)/(n^{3/2}h^2)]. \quad \text{(A5)}$$

These relations also hold if h is replaced by s.

Proof: Only (A3)-(A5) are proved. The proofs for the relations with h replaced by s are identical.

(A3) follows from

$$\sup_{1 \le i \le n} \sup_{v \in S_v} |p_{nhi}(v) - p_{nh}(v)| \le M/nh.$$

for some $M < \infty$ (by assumption 8) and

$$\sup_{v \in S_v} |p_{nh}(v) - p_\theta(v)| = O([(\log n)/(nh)]^{1/2})$$

almost surely (Silverman 1978).

To prove (A4), define

$$g_{nh}(v) = (nh)^{-1} \sum_{j=1}^{n} Y_j K\left[\frac{v - v(X_j, \theta)}{h}\right].$$

Observe that

$$\sup_{1 \le i \le n} \sup_{v \in S_v} |g_{nhi}(v) - p_\theta(v)F(v)| \le$$

$$\sup_{v \in S_v} |g_{nh}(v) - p_\theta(v)F(v)| + M(nh)^{-1} \sup_{1 \le i \le n} |Y_i| \quad \text{(A6)}$$

for some $M < \infty$ because K is bounded uniformly. The first term on the right-hand side of (A6) is $O_p[(nh^2)^{-1/2}]$ (Bierens 1985). Now consider the second term. Let P_Y denote the marginal CDF of Y. Given any $\epsilon > 0$,

$$\log P\Big[\sup_{1 \le i \le n} n^{-1/2}|Y_i| < \epsilon\Big] = n \log(1 - [1 - P_Y(n^{1/2}\epsilon)]) \quad (A7)$$

A Taylor series expansion of the right-hand side of (A7) yields

$$\log P\Big[\sup_{1 \le i \le n} n^{-1/2}|Y_i| < \epsilon\Big] = -n\zeta_n^{-1}[1 - P_Y(n^{1/2}\epsilon)]$$

$$= -(n^{1/2}\epsilon)^2\zeta_n^{-1}[1 - P_Y(n^{1/2}\epsilon)]/\epsilon^2, \quad (A8)$$

where ζ_n is between $P_Y(n^{1/2}\epsilon)$ and 1. By assumption 7, $E(Y^2) < \infty$, so $\lim_{u \to \infty} u^2[1 - P_Y(u)] = 0$. Therefore, the right-hand side of (A8) converges to 0 as $n \to \infty$ and

$$\sup_{1 \le i \le n} |Y_i| = o_p(n^{1/2}). \quad (A9)$$

Equation (A4) now follows from (A6) and (A9).

To prove (A5), expand $F_{nhi} = g_{nhi}/p_{nhi}$ in a Taylor series about $g_{nhi} = Fp_\theta$ and $p_{nhi} = p_\theta$, thereby obtaining

$$F_{nhi}(v) - F(v) = G_{nhi}(v) - J_{nhi}(v)$$

$$+ [g_{nhi}(v) - F(v)p_\theta(v)][p_{nhi}(v) - p_\theta(v)]^2/p_\theta(v)^3$$

$$+ g_{nhi}(v)O([p_{nhi}(v) - p_\theta(v)]^3).$$

The result now follows from (A3) and (A4). Q.E.D.

<u>Lemma 2</u>: Let $\{\theta_n: n = 1,2,\ldots\}$ be a nonstochastic sequence in $\Re^K$ such that $n^{1/2}(\theta_n - \theta) = O(1)$. For each $\ell = 1,\ldots,n$

$$\sup_{x \in \bar{S}_X} (nh)^{1/2}|\hat{F}_{nh\ell}[v(x,\theta_n)] - F_{nh\ell}[v(x,\theta)]| = o_p(h^{1/2}) \quad (A10)$$

as $n \to \infty$. The same relation holds when h is replaced by s.

Proof: Only (A10) is proved. The proof with s in place of h is identical. Define $\tilde{g}_{nh\ell}(\cdot)$ and $\tilde{p}_{nh\ell}(\cdot)$, respectively, by replacing θ with θ_n in the definitions of $g_{nh\ell}(\cdot)$ and $p_{nh\ell}(\cdot)$. It suffices to prove that

$$\sup_{x \in \bar{S}_X} (nh)^{1/2} |\tilde{g}_{nh\ell}[v(x,\theta_n)] - g_{nh\ell}[v(x,\theta)]| = o_p(h^{1/2}) \qquad \text{(A11)}$$

and

$$\sup_{x \in \bar{S}_X} (nh)^{1/2} |\tilde{p}_{nh\ell}[v(x,\theta_n) - p_{nh\ell}[v(x,\theta)]| = o_p(h^{1/2}) \qquad \text{(A12)}$$

We prove only (A11). The proof of (A12) is similar.

Define

$$D_n(x) = (nh)^{-1/2} \sum_{\substack{j=1 \\ j \neq \ell}}^{n} Y_j \left\{ K\left[\frac{v(x,\theta_n) - v(X_j,\theta_n)}{h}\right] - K\left[\frac{v(x,\theta) - v(X_j,\theta)}{h}\right] \right\}, \qquad \text{(A13)}$$

$$\Delta_{n1}(x) = D_n(x) - ED_n(x),$$

and

$$\Delta_{n2}(x) = ED_n(x).$$

Then

$$(nh)^{1/2} [\tilde{g}_{nh\ell}(x) - g_{nh\ell}(x)] = \Delta_{n1}(x) + \Delta_{n2}(x).$$

Consider Δ_{n1}. By a Taylor series expansion

$$K\left[\frac{v(x,\theta_n) - v(X_j,\theta_n)}{h}\right] - K\left[\frac{v(x,\theta) - v(X_j,\theta)}{h}\right] =$$

$$h^{-1}(\theta_n - \theta)' Z_{nj}(v) K'\left[\frac{v(x,\theta_n^*) - v(X_j,\theta_n^*)}{h}\right], \qquad \text{(A14)}$$

where θ_n^* is between θ and θ_n, and

$$Z_{nj} = Z_{nj}(v) = [\partial v(x,\tau)/\partial\tau - \partial v(X_j,\tau)/\partial\tau]_{\tau = \theta_n^*}.$$

Define

$$B_{n\ell}(x) = n^{-1/2}\sum_{\substack{j=1 \\ j\neq\ell}}^{n} Y_j Z_{nj} K'\left[\frac{v(x,\theta_n^*) - v(X_j,\theta_n^*)}{h}\right].$$

Then

$$D_n(x) = n^{1/2}(\theta_n - \theta)'(nh^3)^{-1/2}B_{n\ell}(x)$$

By assumption 8,

$$K'(u/h) = (h/2\pi)^{-1}\int_{-\infty}^{\infty} e^{-itu}\psi(ht)dt$$

for any u. Therefore

$$B_{n\ell}(x) = (h/2\pi)\int_{-\infty}^{\infty}\exp[-itv(x,\theta_n^*)]\psi(ht)n^{-1/2}\sum_{\substack{j=1 \\ j\neq\ell}}^{n} Y_j Z_{nj}\{\exp[itv(X_j,\theta_n^*)]\}dt$$

Let $B_{n\ell k}(x)$ and Z_{njk}, respectively, denote the k'th components of $B_{n\ell}(x)$ and Z_{nj}. Then for each k,

$$|B_{n\ell k}(x) - EB_{n\ell k}(x)| \leq (h/2\pi)\int_{-\infty}^{\infty}|\psi(ht)|n^{-1/2}|\sum_{\substack{j=1 \\ j\neq\ell}}^{n}\{Y_j Z_{njk}\exp[itv(X_j,\theta_n^*)]$$

$$- EY_j Z_{njk}\exp[itv(X_j,\theta_n^*)]\}|dt.$$

By assumptions 2 and 7

$$E|\sum_{\substack{j=1 \\ j\neq\ell}}^{n}\{Y_j Z_{njk}\exp[itv(X_j,\theta_n^*)] - EY_j Z_{njk}\exp[itv(X_j,\theta_n^*)]\}|$$

$$\le \ (n-1)^{1/2}(\mathrm{Var}\{Y_jZ_{njk}\cos[tv(X_j,\theta_n^*)]\}$$

$$+ \ \mathrm{Var}\{Y_jZ_{njk}\sin[tv(X_j,\theta_n^*)]\})^{1/2}$$

$$\le \ M^*(n-1)^{1/2}(EY^2)^{1/2}$$

$$\le \ Mn^{1/2}$$

for some finite M* and M. Therefore,

$$E|B_{n\ell k}(x) - EB_{n\ell k}(x)| \ \le \ (Mh/2\pi)\int_{-\infty}^{\infty}|\psi(ht)|dt$$

$$= \ (M/2\pi)\int_{-\infty}^{\infty}|\psi(t)|dt$$

and, because $|\psi|$ is integrable and $n^{1/2}(\theta_n - \theta) = O(1)$,

$$E|\Delta_{n1}(x)| \ \le \ O[(nh^3)^{-1/2}]$$

uniformly over $x \in \bar{S}_X$. It follows from equation (3.2.5) of Amemiya (1985) that

$$\Delta_{n1}(v) \ = \ O_p[(nh^3)^{-1/2}] \tag{A15}$$

uniformly over $x \in \bar{S}_X$.

Now consider Δ_{n2}. By (A13) and (A14)

$$\Delta_{n2}(x) \ = \ (n-1)/[(nh^3)^{1/2}](\theta_n - \theta)'E\left\{YZ_{n\cdot}K'\left[\frac{v(x,\theta_n^*) - v(X,\theta_n^*)}{h}\right]\right\}$$

$$= \ (n-1)/[(nh^3)]^{1/2}(\theta_n - \theta)'E\left\{F[v(X,\theta)]Z_{n\cdot}K'\left[\frac{v(x,\theta_n^*) - v(X,\theta_n^*)}{h}\right]\right\}$$

$$= \ (n-1)/[(nh^3)]^{1/2}(\theta_n - \theta)'\int_{-\infty}^{\infty}\Gamma(x,u,\theta_n^*)K'\left[\frac{v(x,\theta_n^*) - u}{h}\right]p_{\theta_n^*}(u)du$$

Let $\xi = (u - v)/h$. Then since K is symmetrical about 0

$$\Delta_{n2}(x) = -(n-1)/[(nh)^{1/2}](\theta_n - \theta)'$$

$$\cdot\int_{-\infty}^{\infty}\Gamma[x,v(x,\theta_n^*) + h\xi,\theta_n^*]K'(\xi)p_{\theta_n^*}[v(x,\theta_n^*) + h\xi]d\xi \quad (A16)$$

By assumption 2, $v(x,\theta_n^*) \in \bar{S}_v$ and $v(x,\theta_n^*) + h\xi \in \bar{S}_v$ for any ξ, any $x \in \bar{S}_X$ and all sufficiently large n. Therefore, by assumptions 2, 3, and 6

$$\|\Gamma[x,v(x,\theta_n^*) + h\xi,\theta_n^*]p_{\theta_n^*}[v(x,\theta_n^*) + h\xi]$$

$$- \Gamma[x,v(x,\theta_n^*),\theta_n^*]p_{\theta_n^*}[v(x,\theta_n^*)]\| \le Mh|\xi| \quad (A17)$$

for each ξ, all sufficiently large n, and some $M < \infty$, where $\|\cdot\|$ denotes the Euclidean norm. Moreover,

$$\int_{-\infty}^{\infty}K'(\xi)d\xi = 0$$

by symmetry of K. Therefore, it follows from (A16), (A17), and Lebesgue's dominated convergence theorem that

$$\Delta_{n2}(v) = o(h^{1/2}) \quad (A18)$$

as $n \to \infty$ uniformly over $x \in S_X$. Equation (A11) follows by combining (A15) and (A18). Q.E.D.

<u>Lemma 3</u>: Let $\{\theta_n: n = 1,2,\ldots\}$ be a nonstochastic sequence in $\mathbb{R}^K$ such that $n^{1/2}(\theta_n - \theta) = O(1)$. As $n \to \infty$,

$$\sup_{1 \le i \le n} \sup_{x \in \bar{S}_X} (nh)^{1/2}|\hat{F}_{nhi}[v(x,\theta_n)] - F_{nhi}[v(x,\theta)]| = o_p(h^{1/2})$$

The same relation holds with h replaced by s.

<u>Proof</u>: It suffices to prove that

$$\sup_{1 \le i \le n} \sup_{x \in \bar{S}_X} (nh)^{1/2} |\tilde{g}_{nhi}[v(x,\theta_n)] - g_{nhi}[v(x,\theta)]| = O_p(h^{1/2}), \quad (A19)$$

$$\sup_{1 \le i \le n} \sup_{x \in \bar{S}_X} (nh)^{1/2} |\tilde{p}_{nhi}[v(x,\theta_n)] - p_{nhi}[v(x,\theta)]| = O_p(h^{1/2}), \quad (A20)$$

and that (A19) and (A20) hold with h replaced by s, where $\tilde{g}$ and $\tilde{p}$ are defined as in Lemma 2. The proof is given only for (A19). The proofs of (A20) and the relations for s are identical.

Define

$$d_{nhi}(x) = (nh)^{-1/2} Y_i \left\{ K\left[\frac{v(x,\theta_n) - v(X_i,\theta_n)}{h}\right] - K\left[\frac{v(x,\theta) - v(X_i,\theta)}{h}\right]\right\}$$

Because of (A11) and (A12), to prove (A19) it suffices to show that

$$\sup_{1 \le i \le n} \sup_{x \in \bar{S}_X} |d_{nhi}(x)| = o_p(h^{1/2}) \quad (A21)$$

By a Taylor series expansion

$$d_{nhi}(x) = (nh^3)^{-1/2} (\theta_n - \theta)' Y_i Z_{ni} K'\left[\frac{v(x,\theta_n^*) - v(X_i,\theta_n^*)}{h}\right],$$

where Z_{ni} is defined as in Lemma 2 and θ_n^* is between θ_n and θ. Therefore, by assumptions 2 and 8

$$\sup_{1 \le i \le n} \sup_{x \in \bar{S}_X} |h^{-1/2} d_{nhi}(x)| \le M(nh^2)^{-1} \sup_{1 \le i \le n} |Y_i| \quad (A22)$$

for some $M < \infty$. But $\sup_{1 \le i \le n} |Y_i| = o_p(n^{1/2})$ by (A9). Therefore the right-hand side of (A22) is $o_p(1)$, and (A21) holds. Q.E.D.

<u>Lemma 4</u>: Let $\{\theta_n: n = 1,2,\ldots\}$ be a nonstochastic sequence in $\mathbb{R}^K$ such that $n^{1/2}(\theta_n - \theta) = O(1)$. Define G_{nsi} and J_{nsi} by replacing h with s in the definitions of G_{nhi} and J_{nsi}. Define

$$G_{ni}(v) = [G_{nhi}(v) - (h/s)^r G_{nsi}(v)]/[1 - (h/s)^r]$$

and

$$J_{ni}(v) = [J_{nhi}(v) - (h/s)^r J_{nsi}(v)]/[1 - (h/s)^r].$$

As $n \to \infty$,

$$\sup_{1 \le i \le n} \sup_{v \in S_v} (nh)^{1/2} |F_{ni}(v) - F(v) - G_{ni}(v) + J_{ni}(v)|$$

$$= O_p[(\log n)/(n^{3/2}h^2)].$$

and

$$\sup_{1 \le i \le n} \sup_{x \in \bar{S}_X} (nh)^{1/2} |\hat{F}_{ni}[v(x,\theta_n)] - F_{ni}[v(x,\theta)]| = o_p(h^{1/2}).$$

<u>Proof</u>: These results follow by combining the definitions of $\hat{F}_{ni}$ and F_{ni} with the results of Lemmas 1 and 3. Q.E.D.

<u>Lemma 5</u>: Let $\{\theta_n: n = 1,2,...\}$ be a sequence in $\mathbb{R}^K$ that converges to θ. For all sufficiently large n and $x \in S_X$, $v(x,\theta_n) \in S_v$ implies that $v(x,\theta) \in S_v$.

<u>Proof</u>: By assumption 2, $|v(x,\theta_n) - v(x,\theta)| \le M\|\theta_n - \theta\|$ for some $M < \infty$ that does not depend on x. The result now follows from the fact that $S_v \subset \text{int}(S_v)$. Q.E.D.

<u>Lemma 6</u>: Define

$$T_n^* = h^{1/2} \sum_{i=1}^{n} w[v(X_i,\theta)](Y_i - F[v(X_i,\theta)])(F_{ni}[v(X_i,\theta)] - F[v(X_i,\theta)])$$

Then as $n \to \infty$,

$$T_n = T_n^* + o_p(1).$$

<u>Proof</u>: Some algebra shows that

$$T_n = T_n^* + \sum_{j=1}^{6} R_{nj},$$

where

$$R_{n1} = h^{1/2} \sum_{i=1}^{n} w[v(X_i,\hat{\theta}_n)](Y_i - F[v(X_i,\theta)])(\hat{F}_{ni}[v(X_i,\hat{\theta}_n)] - F_{ni}[v(X_i,\theta)])$$

R_{n2} =

$$h^{1/2}\sum_{i=1}^{n}(w[v(X_i,\hat{\theta}_n)] - w[v(X_i,\theta)])(Y_i - F[v(X_i,\theta)])(F_{ni}[v(X_i,\theta)] - F[v(X ,\theta)])$$

$$R_{n3} = -h^{1/2}\sum_{i=1}^{n} w[v(X_i,\hat{\theta}_n)](Y_i - F[v(X_i,\theta)])(F[v(X_i,\hat{\theta}_n)] - F[v(X_i,\theta)])$$

R_{n4} =

$$-h^{1/2}\sum_{i=1}^{n} w[v(X_i,\hat{\theta}_n)](F[v(X_i,\hat{\theta}_n)] - F[v(X_i,\theta)])(\hat{F}_{ni}[v(X_i,\hat{\theta}_n)] - F_{ni}[v(X_i,\theta)])$$

R_{n5} =

$$-h^{1/2}\sum_{i=1}^{n} w[v(X_i,\hat{\theta}_n)](F[v(X_i,\hat{\theta}_n)] - F[v(X_i,\theta)])(F_{ni}[v(X_i,\theta)] - F[v(X_i,\theta)])$$

$$R_{n6} = h^{1/2}\sum_{i=1}^{n} w[v(X_i,\hat{\theta}_n)](F[v(X_i,\hat{\theta}_n)] - F[v(X_i,\theta)])^2.$$

We now show that $R_{nj} = o_p(1)$ for each j. In what follows, $\{\theta_n: n = 1,2,\ldots\}$ denotes an arbitrary nonstochastic sequence in $\mathbb{R}^K$ satisfying $n^{1/2}(\theta_n - \theta) = O(1)$ as $n \to \infty$.

a. Define

$$\tilde{R}_{n1} = h^{1/2}\sum_{i=1}^{n} w[v(X_i,\theta_n)](Y_i - F[v(X_i,\theta)])(\hat{F}_{ni}[v(X_i,\theta_n)] - F_{ni}[v(X_i,\theta)])$$

Since $\{\theta_n\}$ is arbitrary, it suffices to show that $\tilde{R}_{n1} = o_p(1)$. Given any $\epsilon >$ 0, let $A_{ni\epsilon}$ denote the intersection of the events $\tilde{p}_{nhi}[v(x,\theta_n)] > \epsilon$ uniformly over $x \in \bar{S}_X$, $\tilde{p}_{nsi}[v(x,\theta_n)] > \epsilon$ uniformly over $x \in \bar{S}_X$, and

$$\sup_{x \in \bar{S}_X} (nh)^{1/2}|\hat{F}_{ni}[v(x,\theta_n)] - F_{ni}[v(x,\theta)]| \le \epsilon,$$

where $\bar{p}_{nhi}$ is as defined in Lemma 2. Define

$$A_{n\epsilon} = \bigcup_{i=1}^{n} A_{ni\epsilon}.$$

Let $1(\cdot) = 1$ if the event in parentheses occurs and 0 otherwise. By (A3), (A12), and Lemma 4, $P(A_{n\epsilon}) = o(1)$. By Lemma 5, $x \in \bar{S}_X$ if $w[v(x,\theta_n)] > 0$ and n is sufficiently large. Therefore,

$$\bar{R}_{n1} = R_{n1}^* + o_p(1), \tag{A23}$$

where

$$R_{n1}^* =$$

$$h^{1/2}\sum_{i=1}^{n} 1(A_{ni\epsilon})w[v(X_i,\theta_n)](Y_i - F[v(X_i,\theta)])(\hat{F}_{ni}[v(X_i,\theta_n)] - F_{ni}[v(X_i,\theta)])$$

$E(R_{n1}^*) = 0$ because the event $A_{ni\epsilon}$ does not depend on Y_i or X_i. Define $U = Y - F[v(X,\theta)]$. Then

$$\mathrm{Var}(R_{n1}^*) =$$

$$n^{-1}E\sum_{i=1}^{n} 1(A_{ni\epsilon})w[v(X_i,\theta_n)]^2\sigma^2[v(X_i,\theta_n](nh)(\hat{F}_{ni}[v(X_i,\theta_n)] - F_{ni}[v(X_i,\theta)])^2$$

$$+ n^{-1}E\Bigg(\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} 1(A_{ni\epsilon})1(A_{nj\epsilon})w[v(X_i,\theta_n)]w[v(X_j,\theta_n)]U_iU_j$$

$$\cdot(nh)(\hat{F}_{ni}[v(X_i,\theta_n)] - F_{ni}[v(X_i,\theta)])(\hat{F}_{nj}[v(X_j,\theta_n)] - F_{nj}[v(X_j,\theta)])\Bigg)$$

It is shown below that for any x

$$(nh)^{1/2}E1(A_{nj\epsilon})w[v(x,\theta_n)]U_i(\hat{F}_{nj}[v(x,\theta_n)] - F_{nj}[v(x,\theta)]) = O[1/(nh^{1/2})].$$

uniformly over x. Therefore, since U_i is independent of U_j, $\hat{F}_{ni}$, and F_{ni},

$$\mathrm{Var}(R_{n1}^*) =$$

$$n^{-1}E\sum_{i=1}^{n} 1(A_{ni\epsilon})w[v(X_i,\theta_n)]^2\sigma^2[v(X_i,\theta_n](nh)(\hat{F}_{ni}[v(X_i,\theta_n)] - F_{ni}[v(X_i,\theta)])^2$$

$$+ o(1)$$

$$\leq n^{-1}M\epsilon^2\sum_{i=1}^{n} E1(A_{ni\epsilon}) + o(1)$$

$$\leq M\epsilon^2$$

for some $M < \infty$. Since ϵ is arbitrary, it follows from Chebyshev's inequality that $R_{n1}^* = o_p(1)$ and from (A23) that $R_{n1} = o_p(1)$.

To show that $(nh^{1/2})E1(A_{nj\epsilon})w[v(x,\theta_n)]U_i(\hat{F}_{nj}[v(x,\theta_n) - F_{nj}[v(x,\theta)]) = O[1/(nh^{1/2})]$, observe that conditional on X,

$$|EU_i(nh)^{1/2}(\hat{F}_{nhj}[v(x,\theta_n)] - F_{nhj}[v(x,\theta)])|1(A_{nj\epsilon})w[v(x,\theta_n)]$$

$$= (nh)^{-1/2}\sigma^2[v(X_i,\theta]|K([v(x,\theta_n) - v(X_i,\theta_n)]/h)\bar{p}_{nhj}[v(x,\theta_n)]^{-1}$$

$$- K([v(x,\theta) - v(X_i,\theta)]/h)p_{nhj}[v(x,\theta)]^{-1}|1(A_{nj\epsilon})w[v(x,\theta_n)].$$

By a Taylor series expansion

$$|EU_i(nh)^{1/2}(\hat{F}_{nhj}[v(x,\theta_n)] - F_{nhj}[v(x,\theta)])|1(A_{nj\epsilon})w[v(x,\theta_n)]$$

$$= (nh)^{-1/2}\sigma^2[v(X_i,\theta)]|K'([v(x,\theta_n^*) - v(X_i,\theta_n^*)]/h)O[1/(n^{1/2}h)]$$

$$+ K([v(x,\theta) - v(X_i,\theta)]/h)(\bar{p}_{nhj}[v(x,\theta_n)]^{-1} - p_{nhj}[v(x,\theta)]^{-1})|$$

$$\cdot 1(A_{nj\epsilon})w[v(x,\theta_n)]$$

conditional on X, where θ_n^* is between θ and θ_n. Integration over the distribution of X yields $E|K'([v(x,\theta_n^*) - v(X_i,\theta_n^*)]/h)| = O(h)$ uniformly over

x. Similarly, $E|K([v(x,\theta) - v(X_i,\theta)]/h)| = O(h)$. In addition, a Taylor series expansion of $\tilde{p}_{nhj}[v(X,\theta_n)] - p_{nhj}[v(x,\theta)]$ followed by taking the expectation with respect to X yields $E|\tilde{p}_{nhj}[v(X,\theta_n)] - p_{nhj}[v(x,\theta)]| = O(\|\theta_n - \theta\|/h) = O[1/(n^{1/2}h)$ uniformly over x. Therefore, it follows from the Cauchy-Schwartz inequality that

$$E|K([v(x,\theta) - v(X_i,\theta)]/h)(\tilde{p}_{nhj}[v(x,\theta_n)]^{-1} - p_{nhj}[v(x,\theta)]^{-1})|$$

$$\cdot 1(A_{nj\epsilon})w[v(x,\theta_n)] = O(n^{-1/2}).$$

uniformly over x. Combining results yields

$$(nh)^{1/2}E1(A_{nj\epsilon})w[v(x,\theta_n)]U_i(\hat{F}_{nhj}[v(x,\theta_n)] - F_{nhj}[v(x,\theta)] = O[1/(nh^{1/2})]$$

uniformly over x. A similar argument applies when h is replaced by s, so

$$(nh)^{1/2}E1(A_{nj\epsilon})w[v(x,\theta_n)]U_i(\hat{F}_{nj}[v(x,\theta_n)] - F_{nj}[v(x,\theta)] = O[1/(nh^{1/2})]$$

uniformly over x.

b. R_{n2}: Define

$$\tilde{R}_{n2} =$$

$$h^{1/2}\sum_{i=1}^{n}(w[v(X_i,\theta_n)] - w[v(X_i,\theta)])(Y_i - F[v(X_i,\theta)])(F_{ni}[v(X_i,\theta)] - F[v(X_i,\theta)])$$

It suffices to show that $\tilde{R}_{n2} = o_p(1)$. It follows from Lemma 4 that $F_{ni}[v(x,\theta) - F[v(x,\theta)] = G_{ni}[v(x,\theta)] - J_{ni}[v(x,\theta) + O_p[(\log n)/(n^{3/2}h^2)]$ uniformly over i and $x \in \tilde{S}_X$. By Lemma 5, $x \in \tilde{S}_X$ if $w[v(x,\theta_n)] > 0$ and n is sufficiently large. Therefore,

$$\tilde{R}_{n2} = R_{n2}^* + o_p(1), \tag{A24}$$

where

$$R_{n2}^* = h^{1/2}\sum_{i=1}^{n}(w[v(X_i,\theta_n)] - w[v(X_i,\theta)])$$

$$\cdot\{Y_i - F[v(X_i,\theta)]\}\{G_{ni}[v(X_i,\theta)] - J_{ni}[v(X_i,\theta)]\}.$$

$E(R_{n2}*) = 0$ because G_{ni} and J_{ni} do not depend on Y_i. It is not difficult to show that $EU_iG_{nj}[v(X_j,\theta)] = O(n^{-1})$ and $EU_iJ_{nj}[v(x,\theta)] = o(n^{-1})$ uniformly over x. Therefore, since U_i is independent of U_j, G_{ni} and J_{ni},

$$\mathrm{Var}(R_{n2}*) = hE\sum_{i=1}^{n}\{w[v(X_i,\theta_n)] - w[v(X_i,\theta)]\}^2$$

$$\cdot\sigma^2[v(X_i,\theta)]\{G_{ni}[v(X_i,\theta)] - J_{ni}[v(X_i,\theta)]\}^2 + o(1)$$

$$\leq (h/n)M\sum_{i=1}^{n}E\{G_{ni}[v(X_i,\theta)] - J_{ni}[v(X_i,\theta)]\}^2 + o(1) \qquad (A25)$$

for some $M < \infty$, where (A25) follows from assumptions 2, 4 and 7. Arguments similar to those of Bierens (1987) yield the result that the expectation in (A25) is o(1), so $\mathrm{Var}(R_{n2}*) = o(1)$. $R_{n2}* = o_p(1)$ now follows from Chebyshev's inequality. This result and (A24) imply that $\bar{R}_{n2} = o_p(1)$. Therefore, $R_{n2} = o_p(1)$.

c. R_{n3}: Define

$$R_{n3}* = -h^{1/2}\sum_{i=1}^{n}w[v(X_i,\theta_n)]\{Y_i - F[v(X_i,\theta)]\{F[v(X_i,\theta_n)] - F[v(X_i,\theta)]\}$$

It suffices to show that $R_{n3}* = o_p(1)$. To do this, observe that $E(R_{n3}*) = 0$. In addition,

$$\mathrm{Var}(R_{n3}*) = hE\sum_{i=1}^{n}w[v(X_i,\theta_n]^2\sigma^2[v(X_i,\theta_n]\{F[v(X_i,\theta_n)] - F[v(X_i,\theta)]\}^2$$

$$\leq hM\sum_{i=1}^{n}Ew[v(X_i,\theta_n)]^2\{F[v(X_i,\theta_n)] - F[v(X_i,\theta)]\}^2$$

for some $M < \infty$ by assumptions 4 and 7. By assumptions 2 and 5,

$$\{F[v(X,\theta_n)] - F[v(X,\theta)]\}^2 = O(\|\theta_n - \theta\|^2)$$

$$= O(n^{-1}).$$

uniformly over X. Therefore, $Var(R_{n2}*) = o(1)$, and $R_{n3}* = o_p(1)$ follows from Chebyshev's inequality.

d. R_{n4}: Define

$$R_{n4}* =$$

$$-h^{1/2}\sum_{i=1}^{n} w[v(X_i,\theta_n)](F[v(X_i,\theta_n)] - F[v(X_i,\theta)])(\hat{F}_{ni}[v(X_i,\theta_n)] - F_{ni}[v(X_i,\theta)])$$

It suffices to show that $R_{n4}* = o_p(1)$. To do this, observe that by assumptions 2, 4, and 5 and a Taylor series expansion

$$|R_{n4}*| \le Mh^{1/2}\|\theta_n - \theta\| M\sum_{i=1}^{n} w[v(X_i,\theta_n)]|\hat{F}_{ni}[v(X_i,\theta_n)] - F_{ni}[v(X_i,\theta)]| \quad \text{(A26)}$$

for some $M < \infty$. By Lemmas 4 and 5, the summand in (A26) is $O_p(n^{-1/2})$ uniformly over i and X_i for which $w[v(X_i,\theta_n)] > 0$. Therefore, since $\|\theta_n - \theta\| = O(n^{-1/2})$, $R_{n4} = O_p(h^{1/2})$.

e. R_{n5}: Define

$$\bar{R}_{n5} =$$

$$-h^{1/2}\sum_{i=1}^{n} w[v(X_i,\theta_n)](F[v(X_i,\theta_n)] - F[v(X_i,\theta)])(F_{ni}[v(X_i,\theta)] - F[v(X_i,\theta)])$$

It suffices to show that $\bar{R}_{n5} = o_p(1)$. Since $F[v(x,\theta_n)] - F[v(x,\theta)] = O(n^{-1/2})$ uniformly over $x \in \bar{S}_X$, it follows from Lemmas 1 and 4 that

$$\bar{R}_{n5} = R_{n5}* + o_p(1), \quad \text{(A27)}$$

where

$$R_{n5}* = -h^{1/2}\sum_{i=1}^{n} w[v(X_i,\theta_n)](F[v(X_i,\theta_n)] - F[v(X_i,\theta)])G_{ni}[v(X_i,\theta)].$$

Let F′ denote the derivative of F. By a Taylor series expansion

$$R_{n5}* = -h^{1/2}(\theta_n - \theta)' \sum_{i=1}^{n} w[v(X_i,\theta_n)]F'[v(X_i,\theta_n*)]v_\theta(X_i,\theta_n*)G_{ni}[v(X_i,\theta)]$$

where θ_n* is between θ and θ_n. By arguments identical to those of Bierens (1987), $EG_{ni}[v(x,\theta)] = o[(nh)^{-1/2}]$ uniformly over x. Therefore, $E(R_{n5}*) = o(1)$ by assumptions 2, 4, and 5. In addition

$$(R_{n5}*)^2 \leq Mh\|\theta_n - \theta\|^2 E \sum_{i=1}^{n} \sum_{j=1}^{n} \{G_{ni}[v(X_i,\theta)]G_{nj}[v(X_j,\theta)]\}.$$

By the arguments of Bierens (1987), the expectation is $O(h^{-1})$ uniformly over X. Therefore $E(R_{n5}*)^2 = o(1)$. It follows from Chebyshev's inequality that $R_{n5}* = o_p(1)$ and from this result and (A27) that $R_{n5} = o_p(1)$.

f. R_{n6}: By assumptions 2 and 5, $\{F[v(X,\hat{\theta}_n] - F[v(X,\theta)]\}^2 = O_p(n^{-1})$ uniformly over $\{X: v \in S_v\}$. Since, in addition, w is bounded uniformly, $R_{n6} = O_p(h^{1/2})$. Q.E.D.

Lemma 7: Define $V_i = v(X_i,\theta)$. Then

$$T_n = h^{1/2} \sum_{i=1}^{n} w(V_i)[Y_i - F(V_i)]G_{ni}(V_i) + o_p(1). \qquad (A28)$$

Proof: By Lemmas 4 and 6

$$T_n = h^{1/2} \sum_{i=1}^{n} w(V_i)[Y_i - F(V_i)]G_{ni}(V_i) - T_{n1} + o_p(1),$$

where

$$T_{n1} = h^{1/2} \sum_{i=1}^{n} w(V_i)[Y_i - F(V_i)]J_{ni}(V_i).$$

To prove the lemma it suffices to show that $T_{n1} = o_p(1)$. $E(T_{n1}) = 0$ because $J_{ni}(V_i)$ does not depend on Y_i. In addition, since $EU_iJ_{nj}(v) = o(n^{-1})$ uniformly over v,

$$E(T_{ni}^2) = h \sum_{i=1}^{n} E\{w(V_i)^2\sigma^2(V_i)J_{ni}(V_i)^2\} + o(1) \qquad (A28)$$

But for any $v \in S_v$

$$[1 - (h/s)^r]J_{ni}(v) \le |g_{nhi}(v) - p_\theta(v)F(v)||p_{nhi}(v) - p_\theta(v)|/p_\theta(v)^2$$

$$+ F(v)|p_{nhi}(v) - p_\theta(v)|^2/p_\theta(v)^2$$

$$+ (h/s)^r[|g_{nsi}(v) - p_\theta(v)F(v)||p_{nsi}(v) - p_\theta(v)|/p_\theta(v)^2$$

$$+ F(v)|p_{nsi}(v) - p_\theta(v)|^2/p_\theta(v)^2].$$

By (A3) and the fact that $h/s < 1$

$$[1 - (h/s)^r]J_{ni}(v) \le [|g_{nhi}(v) - p_\theta(v)F(v)|/p_\theta(v)^2$$

$$+ (h/s)^{r+1/2}|g_{nsi}(v) - p_\theta(v)F(v)|/p_\theta(v)^2]O([(\log n)/(nh)]^{1/2})$$

$$+ O[(\log n)/(nh)] \tag{A30}$$

almost surely. By arguments similar to those of Bierens (1987), $E[g_{nhi}(v) - p_\theta(v)F(v)]^2 = O[1/(nh)]$ uniformly over $v \in S_v$. Therefore, by the Cauchy-Schwartz inequality, $E|[g_{nhi}(v) - p_\theta(v)F(v)| = O[1/(nh)^{1/2}]$ uniformly over $v \in S_v$. Therefore, squaring (A30) and taking expected values on both sides of the result yields

$$[1 - (h/s)^r]^2E[J_{ni}(v)^2] = O([(\log n)/(nh)]^2) \tag{A31}$$

uniformly over $v \in S_v$. Substituting (A31) into (A28) and making use of assumption 4 yields $E[T_{n1}^2] = o(1)$. $T_{n1} = o_p(1)$ now follows from Chebyshev's inequality. Q.E.D.

<u>Proof of Theorem 1</u>

For $i = 1,\ldots,n$, define $U_i = Y_i - F(V_i)$ and $Z_i = (U_i,V_i)$. Also, for $v \in S_v$ and $i,j = 1,\ldots,n$ define

$$K_{hs}(v) = [1 - (h/s)^r]^{-1}[K(v/h) - (h/s)^{r+1}K(v/s)],$$

$$A_n(Z_i,Z_j) = [1/(nh^{1/2})]w(V_i)U_i[p_\theta(V_i)]^{-1}[U_j + F(V_j) - F(V_i)]K_{hs}(V_j - V_i)],$$

$$\mu(Z_i) = E[A_n(Z_i,Z_j) + A_n(Z_j,Z_i)|Z_i],$$

$$H_n(Z_i,Z_j) = A_n(Z_i,Z_j) + A_n(Z_j,Z_i) - \mu(Z_i),$$

and

$$\Psi_n = \sum_{1 \le i < j \le n}\sum H_n(Z_i,Z_j).$$

It follows from (A28) that

$$T_n = \Psi_n + \sum_{1 \le i < j \le n}\sum \mu(Z_i) + o_p(1). \tag{A32}$$

Observe that

$$\mu(Z_i) = [1/(nh^{1/2})]w(V_i)U_i[p_\theta(V_i]^{-1}E([F(V_j) - F(V_i)]K_{hs}(V_j - V_i)|V_i).$$

Since $E(U_i) = 0$ for all $i = 1,\ldots,n$ and the U_i are independent, $E[\mu(Z_i)] = 0$, and $E[\mu(Z_i)\mu(Z_j)] = 0$ if $i \ne j$. Moreover, arguments similar to those of Bierens (1987) yield

$$\mu(Z_i) = [1/(nh^{1/2})]w(V_i)U_i[p_\theta(V_i)]^{-1}o(h^{r + 1}). \tag{A33}$$

uniformly over Z_i. Therefore, $E[\mu(Z_i)^2] = o(h^{2r+1}/n^2)$, so the second term on the right-hand side of (A32) has mean 0 and variance $o(h^{2r+1})$. It follows from Chebyshev's inequality that the second term on the right-hand side of (A32) is $o_p(1)$ so that $T_n = \Psi_n + o_p(1)$. Therefore, to prove the theorem it suffices to show that $\Psi_n \to^d N(0,\sigma_T^2)$. Define

$$Q_n(Z_i,Z_j) = E[H_n(Z_\ell,Z_i)H_n(Z_\ell,Z_j)|Z_i,Z_j].$$

By Theorem 1 of Hall (1984), $\Psi_n \to^d N(0,\sigma_T^2)$ if

$$E[Q_n(Z_i,Z_j)^2]/(E[H_n(Z_i,Z_j)^2])^2 \to 0, \tag{A34}$$

$$n^{-1}E[H_n(Z_i,Z_j)^4]/(E[H_n(Z_i,Z_j)^2])^2 \to 0, \tag{A35}$$

and

$$(1/2)n^2E[H_n(Z_i,Z_j)^2] \rightarrow \sigma_T^{\ 2} \tag{A36}$$

as $n \rightarrow \infty$.

Analysis of $E[Q_n(Z_i,Z_j)^2]$: By (A33) and the definition of H_n

$$H_n(Z_i,Z_j) = [1/(nh^{1/2})]\{U_iU_j[w(V_i)/p_\theta(V_i) + w(V_j)/p_\theta(V_j)]$$

$$+ [F(V_j) - F(V_i)][w(V_i)U_i/p_\theta(V_i) - w(V_j)U_j/p_\theta(V_j)]$$

$$\cdot K_{hs}(V_j - V_i)\} - w(V_i)U_i[p(V_i)]^{-1}o(h^{r+1/2}/n) \tag{A37}$$

uniformly over (Z_i,Z_j). Some algebra yields the result that

$$Q_n(Z_i,Z_j) = Q_{n1}(Z_i,Z_j) + U_io(h^{r+1}/n^2) + U_jo(h^{r+1}/n^2)$$

$$+ o(h^{2r+1}/n^2) \tag{A38}$$

uniformly over (Z_i,Z_j), where

$$Q_{n1}(Z_i,Z_j) = E_{\ell|i,j}\{\sigma^2(v_\ell)U_iU_j[w(V_\ell)/p_\theta(V_\ell) + w(V_i)/p_\theta(V_i)]$$

$$\cdot[w(V_\ell)/p_\theta(V_\ell) + w(V_j)/p_\theta(V_j)] + [F(V_j) - F(V_\ell)]\sigma^2(V_\ell)U_i$$

$$\cdot[w(V_\ell)/p_\theta(V_\ell][w(V_\ell)/p_\theta(V_\ell) + w(V_i)/p_\theta(V_i)] + [F(V_i) - F(V_\ell)]$$

$$\cdot\sigma^2(V_\ell)U_j[w(V_\ell)/p_\theta(V_\ell)][w(V_\ell)/p_\theta(V_\ell) + w(V_j)/p_\theta(V_j)]$$

$$+ [F(V_\ell) - F(V_i)][F(V_\ell) - F(V_j][w(V_\ell)/p_\theta(V_\ell)]^2\sigma^2(V_\ell)\}$$

$$\cdot K_{hs}(V_i - V_\ell)K_{hs}(V_j - V_\ell), \qquad (A39)$$

and $E_{\ell|i,j}$ denotes the expectation over V_ℓ conditional on V_i and V_j. Equation (A39) has the form

$$Q_{n1}(Z_i, Z_j) = (n^2h)^{-1}\int [U_iU_jR_1(V_\ell, V_i, V_j) + U_iR_2(V_\ell, V_i, V_j)$$

$$+ U_jR_2(V_\ell, V_j, V_i) + R_3(V_\ell, V_i, V_j)]\{K[(V_i - V_\ell)/h]$$

$$- (h/s)^{r+1}K[(V_i - V_\ell)/s]\}\{K[(V_j - V_\ell)/h]$$

$$- (h/s)^{r+1}K[(V_j - V_\ell)/s]\}p_\theta(V_\ell)dV_\ell,$$

where R_1, R_2, and R_3 are bounded, continuous functions. Let $\varsigma = (V_\ell - V_i)/h$. Then

$$Q_{n1}(Z_i, Z_j) = n^{-2}\int [U_iU_jR_1(h\varsigma + V_i, V_i, V_j) + U_iR_2(h\eta + V_i, V_i, V_j)$$

$$+ U_jR_2(h\varsigma + V_i, V_j, V_i) + R_3(h\varsigma + V_i, V_i, V_j)]\{K(\varsigma)$$

$$- (h/s)^{r+1}K[(h/s)\varsigma]\}\{K[\varsigma + (V_i - V_j)/h]$$

$$- (h/s)^{r+1}K[(h/s)\varsigma + (V_i - V_j)/s]\}p_\theta(h\varsigma + V_i)d\theta$$

$$= n^{-2}\int [U_iU_jR_1(h\varsigma + V_i, V_i, V_j) + U_iR_2(h\varsigma + V_i, V_i, V_j)$$

$$+ U_jR_2(h\varsigma + V_i, V_j, V_i) + R_3(h\varsigma + V_i, V_i, V_j)]K(\varsigma)$$

$$\cdot K[\varsigma + (V_i - V_j)/h]p_\theta(h\varsigma + V_i)d\varsigma + U_iU_jo(n^{-2})$$

$$+ U_io(n^{-2}) + U_jo(n^{-2}) + o(n^{-2}). \qquad (A40)$$

It follows from (A38) that $Q_n(Z_i,Z_j)$ has the form (A40). Therefore, $E[Q_n(Z_i,Z_j)^2]$ has the form

$$E[Q_n(Z_i,Z_j)^2] =$$

$$n^{-2}\int R(h\varsigma_1 + V_i, h\varsigma_2 + V_i, V_i, V_j)K(\theta_1)K(\varsigma_2)K[\varsigma_1 + (V_i - V_j)/h]$$

$$\cdot K[\varsigma_2 + (V_i - V_j)/h]p_\theta(h\varsigma_1 + V_i)p_\theta(h\varsigma_2 + V_i)p_\theta(V_i)p_\theta(V_j)d\varsigma_1 d\varsigma_2 dV_i dV_j$$

$$+ o(n^{-4})$$

$$= o(n^{-4}).$$

Therefore

$$E[Q_n(Z_i,Z_j)^2] = o(1/n^2). \tag{A41}$$

Analysis of $n^{-1}E[H_n(Z_i,Z_j)^4]$: By (A37) $H_n(Z_i,Z_j)^4$ has the form

$$H_n(Z_i,Z_j)^4 = n^{-4}h^{-2}R_{n4}(Z_i,Z_j),$$

where R_{n4} has the property that $E(R_{n4})$ is bounded uniformly over n. Therefore,

$$n^{-1}E[H_n(Z_i,Z_j)^4] = O[1/(n^5h^2)]. \tag{A42}$$

Analysis of $E[H_n(Z_i,Z_j)^2]$: By the definition of H_n

$$H_n(Z_i,Z_j)^2 = A_n(Z_i,Z_j)^2 + A_n(Z_j,Z_i)^2 + \mu(Z_i)^2 + 2[A_n(Z_i,Z_j)A_n(Z_j,Z_i)$$

$$+ \mu(Z_i)A_n(Z_i,Z_j) + \mu(Z_i)A_n(Z_j,Z_i)]. \tag{A43}$$

But

$$A_n(Z_i,Z_j)^2 = [1/(n^2h)]w(V_i)U_i^2[p_\theta(V_i)^{-2}]\{U_j^2 + 2U_j[F(V_j) - F(V_i)]$$

$$+ [F(V_j) - F(V_i)]^2\}K_{hs}(V_j - V_i)^2.$$

Therefore,

$$n^2E[A_n(Z_i,Z_j)^2] = [1 - (h/s)^r]h^{-1}\int w(V_i)^2[p_\theta(V_i)^{-2}]\sigma^2(V_i)$$

$$\cdot\{\sigma^2(V_j) + [F(V_j) - F(V_i)]^2\}\{K[(V_j - V_i)/h]$$

$$- (h/s)^r K[(V_j - V_i)/s]\}^2 p_\theta(V_i)p_\theta(V_j)dV_j dV_i.$$

Define $\zeta = (V_j - V_i)/h$. Then

$$n^2E[A_n(Z_i,Z_j)^2] = [1 - (h/s)^r]\int w(V_i)^2[p_\theta(V_i)^{-2}]\sigma^2(V_i)$$

$$\cdot\{\sigma^2(h\zeta + V_i) + [F(h\zeta + V_i) - F(V_i)]^2\}\{K(\zeta)$$

$$- (h/s)^r K[(h/s)\zeta]\}^2 p_\theta(V_i)p_\theta(h\zeta + V_i)d\zeta dV_i$$

$$= C_K\int w(v)^2[\sigma^2(v)]^2dv + o(1). \tag{A44}$$

By the Cauchy-Schwartz inequality

$$E[A_n(Z_i,Z_j)\mu(Z_i)] \le \{E[A_n(Z_i,Z_j)^2E[\mu(Z_i)^2\}^{1/2}.$$

It now follows from (A43) and $E[\mu(Z_i)^2] = o(1/n^2)$ that

$$E[A_n(Z_i,Z_j)\mu(Z_i)] = o(1/n^2). \tag{A45}$$

In addition, it is easily seen that

$$E[A_n(Z_j,Z_i)\mu(Z_i)] = 0. \tag{A46}$$

Finally,

$$n^2E[A_n(Z_i,Z_j)A_n(Z_j,Z_i)] = [1 - (h/s)^r]h^{-1}\int w(V_i)w(V_j)\sigma^2(V_i)\sigma^2(V_j)$$

$$\cdot K_{hs}(V_j - V_i)^2 dV_j dV_i.$$

By arguments similar to those used to obtain (A43)

$$n^2 E[A_n(Z_i,Z_j)A_n(Z_j,Z_i)] = C_K \int w(v)^2[\sigma^2(v)]^2 dv + o(1). \tag{A47}$$

Combining (A43)-(A47) yields

$$n^2 E[H_n(Z_i,Z_j)^2] = 4C_K \int w(v)^2[\sigma^2(v)]^2 dv + o(1) \tag{A48}$$

Conditions (A34)-(A36) now follow by combining (A41), (A42), and (A48). Q.E.D.

A4. PROPERTIES OF T_n UNDER H_1

Proof of Theorem 2

Let $\{\theta_n\}$ be a nonstochastic sequence such that $n^{1/2}(\theta_n - \theta) = O(1)$. Let $\tilde{T}_n$ be defined as T_n with $\hat{\theta}_n$ replaced by θ_n. It suffices to show that $\text{plim}_{n\to\infty} \tilde{T}_n/(nh^{1/2}) > 0$. To do this, let $U = Y - H[v(X,\theta)]$ and $v_\theta = \partial v/\partial\theta$. Let $\theta_n{}^*$ denote a point between θ_n and θ (not necessarily the same point in each usage). Some algebra and Taylor series expansions yield

$$\tilde{T}_n/(nh^{1/2}) = \sum_{\ell=1}^{11} R_{n\ell},$$

where

$$R_{n1} = n^{-1}\sum_{i=1}^{n} w[v(X_i,\theta_n)]U_i\{\hat{F}_{ni}[v(X_i,\theta_n)] - F_{ni}[v(X_i,\theta)]\},$$

$$R_{n2} = n^{-1}\sum_{i=1}^{n} w[v(X_i,\theta_n)]U_i\{F_{ni}[v(X_i,\theta)] - H[v(X_i,\theta)]\},$$

$$R_{n3} = n^{-1}\sum_{i=1}^{n} w[v(X_i,\theta_n)]U_i\{H[v(X_i,\theta)] - F[v(X_i,\theta)]\},$$

$$R_{n4} = -[(\theta_n - \theta)'/n]\sum_{i=1}^{n} w[v(X_i,\theta_n)]U_i F'[v(X_i,\theta_n{}^*)]v_\theta(X_i,\theta_n{}^*),$$

$$R_{n5} =$$

$$n^{-1}\sum_{i=1}^{n} w[v(X_i,\theta_n)]\{H[v(X_i,\theta) - F[v(X_i,\theta)]\}\{\hat{F}_{ni}[v(X_i,\theta_n)] - F_{ni}[v(X_i,\theta)]\},$$

$$R_{n6} = n^{-1}\sum_{i=1}^{n} w[v(X_i,\theta_n)]\{H[v(X_i,\theta)] - F[v(X_i,\theta)]\}\{F_{ni}[v(X_i,\theta)] - H[v(X_i,\theta)]\},$$

$$R_{n7} = n^{-1}\sum_{i=1}^{n} w[v(X_i,\theta_n)]\{F[v(X_i,\theta)] - H[v(X_i,\theta)]\}^2,$$

$$R_{n8} = -2[(\theta_n - \theta)'/n]\sum_{i=1}^{n} w[v(X_i,\theta_n)]F'[v(X_i,\theta_n{}^*)]v_\theta(X_i,\theta_n{}^*)$$

$$\cdot\{H[v(X_i,\theta)] - F[v(X_i,\theta)]\},$$

$$R_{n9} = -[(\theta_n - \theta)'/n]\sum_{i=1}^{n} w[v(X_i,\theta_n)]F'[v(X_i,\theta_n{}^*)]v_\theta(X_i,\theta_n{}^*)$$

$$\cdot\{\hat{F}_{ni}[v(X_i,\theta_n)] - F_{ni}[v(X_i,\theta)]\},$$

$$R_{n,10} = -[(\theta_n - \theta)'/n]\sum_{i=1}^{n} w[v(X_i,\theta_n)]F'[v(X_i,\theta_n{}^*)]v_\theta(X_i,\theta_n{}^*)$$

$$\cdot\{F_{ni}[v(X_i,\theta)] - H[v(X_i,\theta)]\},$$

and

$$R_{n,11} =$$

$$[(\theta_n - \theta)'/n]\sum_{i=1}^{n} w[v(X_i,\theta_n)]F'[v(X_i,\theta_n{}^*]^2 v_\theta(X_i,\theta_n{}^*)v_{\theta'}(X_i,\theta_n{}^*)(\theta_n - \theta).$$

R_{n1} is n^{-1} times the analogous quantity in lemma 6. Therefore, $R_{n1} = o_p(1)$.

By lemmas 4 and 6

$$R_{n2} = n^{-1}\sum_{i=1}^{n} w[v(X_i,\theta_n)]U_i(G_{ni}[v(X_i,\theta)] - J_{ni}[v(X_i,\theta)]) + o_p(1)$$

$R_{n2} = o_p(1)$ now follows from a proof similar to that of lemma 7.

Because $w[v(x,\theta_n)] - w[v(x,\theta)] = O(n^{-1/2})$ uniformly over x,

$$R_{n3} = n^{-1}\sum_{i=1}^{n} w[v(X_i,\theta)]U_i(H[v(X_i,\theta)] - F[v(X_i,\theta)]) + o_p(1).$$

$R_{n3} = o_p(1)$ now follows from the strong law of large numbers.

By assumptions 2, 4, and 5 as well as $\theta_n - \theta = O(n^{-1/2})$

$$R_{n4} = (M/n)O(n^{-1/2})\sum_{i=1}^{n} |Y_i - H[v(X_i,\theta)]|.$$

The summand has a finite mean by assumptions 5 and 7, so $R_{n4} = o_p(1)$ follows from the strong law of large numbers.

By lemma 5,

$$|R_{n5}| \le n^{-1}o_p(n^{-1/2})\sum_{i=1}^{n} |H[v(X_i,\theta) - F[v(X_i,\theta)]|,$$

so $R_{n5} = o_p(1)$ follows from the strong law of large numbers.

By lemma 1 and assumption 5

$$|R_{n6}| \le o_p(1)n^{-1}\sum_{i=1}^{n} |H[v(X_i,\theta)] - F[v(X_i,\theta)]|,$$

so $R_{n6} = o_p(1)$ follows from the strong law of large numbers.

$R_{n7} \to^{P} E\{w(V)[H(V) - F(V)]^2\}$ by the strong law of large numbers and the fact that $w[v(x,\theta_n)] - w[v(x,\theta)] = O(n^{-1/2})$ uniformly over x.

By assumptions 2, 4, and 5 and $\theta_n - \theta = O(n^{-1/2})$,

$$R_{n8} \le O(n^{-1/2})n^{-1}\sum_{i=1}^{n} |H[v(X_i,\theta)] - F[v(X_i,\theta)]|,$$

so $R_{n8} = o_p(1)$ follows from the strong law of large numbers.

$R_{n9} = o_p(1)$ by lemma 4 , assumptions 2,4, and 5, and $\theta_n - \theta = O(n^{-1/2})$.

$R_{n,10}$ and $R_{n,11}$ are $o_p(1)$ by lemma 1, assumptions 2, 4,, and 5, and $\theta_n - \theta = O(n^{-1/2})$.

Collecting results yields $\bar{T}_n/(nh^{1/2}) \to^p Ew(V)\{[H(V) - F(V)]^2\} > 0$. Q.E.D.

<u>Additional Assumptions and a Lemma Used in Proving Theorem 3</u>

9. H_n has r continuous derivatives that are uniformly bounded over $v \in S_v$.

10. Define

$$\Gamma_n(x,v,\tau) = E_X\{H_n[v(X,\theta)][\partial v(x,\tau)\partial\tau - \partial v(X,\tau)/\partial\tau]|v(X,\tau) = v\}.$$

Let Γ_{nk} $(k = 1,\ldots,K)$ denote the k'th component of Γ_n. There is a finite number M_Γ, not depending on τ or x, such that for all $\tau \in N_\theta$, $x \in S_X$, v_1, $v_2 \in S_v$, and $k = 1,\ldots,K$

$$|\Gamma_{nk}(x,v_2,\tau) - \Gamma_{nk}(x,v_1,\tau)| \le M_\Gamma|v_1 - v_2|.$$

11. $\sigma^2(v)$ is a continuous function of $v \in S_v$ and is bounded uniformly over $v \in S_v$ and all sufficiently large n. $E\{Y - E[Y|v(x,\theta)]\}^4$ is bounded uniformly over $v \in S_v$ and all sufficiently large n.

12. Define

$$Q_{nr}(v,\xi) = (d^r/d\xi^r)\{[H_n(v + h\xi) - H_n(v)]p_\theta(v)\}$$

a. For some $\alpha > 1/(4\delta)$ and finite constant $C > 0$

$$|Q_{nr}(v,h\xi) - Q_{nr}(v,s\xi)| \le C|h\xi - s\xi|^\alpha.$$

b. The kernel function K satisfies

$$\int_{-\infty}^{\infty}|u^{r+\alpha}K(u)|du < \infty.$$

Assumptions 9-11 extend assumptions 5-7 to the local alternative mean functions H_n. Assumption 12 insures that the bias of $(nh)^{1/2}[F_{ni}(v) - H_n(v)]$ relative to its asymptotic distribution is $o(h^{1/4})$. This bias must be $o(h^{1/4})$ to make the result given in Theorem 3 hold,

<u>Lemma 8</u>: Let assumptions 1-5 and 8-12 hold. Under the sequence of models H_n, the conclusions of Lemmas 1-3 and 7 hold when F is replaced by H_n.

Proof: It may be verified that the proofs of Lemmas 1-3 and 7 hold line by line. Q.E.D.

Proof of Theorem 3

Let $\{\theta_n\}$ be a nonstochastic sequence such that $n^{1/2}(\theta_n - \theta) = O(1)$. Let $\bar{T}_n$ be defined as T_n with $\hat{\theta}_n$ replaced by θ_n. It suffices to show that the conclusion of this theorem holds for $\bar{T}_n$. To do this, let $U = Y - H_n[v(x,\theta)]$, $V = v(X,\theta)$, $v_\theta = \partial v/\partial\theta$, and $w_{ni} = w[v(X_i,\theta_n)]$. Let θ_n* denote a point btween θ_n and θ (not necessarily the same point in each usage). Some algebra and Taylor series expansions yield

$$\bar{T}_n = (nh^{1/2}) \sum_{i=1}^{11} R_{n\ell},$$

where $R_{n\ell}$ ($\ell = 1,\ldots,11$) is obtained by replacing H with H_n in the corresponding terms in the proof of Theorem 2.

1. $(nh^{1/2})R_{n1}$: $(nh^{1/2})R_{n1} = o_p(1)$ follows by a proof identical to that given for R_{n1} in Theorem 1.

2. $(nh^{1/2})R_{n2}$: By Lemma 8,

$$(nh^{1/2})R_{n2} = h^{1/2} \sum_{i=1}^{n} w_{ni}U_iG_{ni}(V_i) + o_p(1),$$

where H_n is used in G_{ni} instead of F. Convergence in distribution of $(nh^{1/2})R_{n2}$ now follows by arguments identical to those used in proving Theorem 1.

3. $(nh^{1/2})R_{n3}$: $E(nh^{1/2}R_{n3}) = 0$ because $E(U) = 0$.

$$\text{Var}(nh^{1/2}R_{n3}) = hE\sum_{i=1}^{n} w_{ni}{}^2\sigma^2(V_i)n^{-1/2}h^{-1/4}\Delta_n(V_i)^2 = O(h^{1/2}).$$

4. $(nh^{1/2})R_{n4}$: $E(nh^{1/2}R_{n4}) = 0$ because $E(U) = 0$. In addition,

$$\text{Var}(nh^{1/2}R_{n4}) = (\theta_n - \theta)'hE\sum_{i=1}^{n} w_{ni}{}^2\sigma^2(V_i)F'[v(X_i,\theta_n*)]^2v_\theta(X_i,\theta_n*)$$

$$\cdot\, v_{\theta'}[v(X_i,\theta_n*)](\theta_n - \theta) = O(h).$$

5. $(nh^{1/2})R_{n5}$: $(nh^{1/2})R_{n5} = O_p(h^{1/4})$ because $H_n - F = O(n^{-1/2}h^{-1/4})$ and, by Lemma 8, $\hat{F}_{ni}(v) - F_{ni}(v) = O_p(n^{-1/2})$ uniformly over i and v.

6. $(nh^{1/2})R_{n6}$: By Lemma 8

$$nh^{1/2}R_{n6} = h^{1/2}\sum_{i=1}^{n} w_{ni}n^{-1/2}h^{-1/4}\Delta_n(V_i)G_{ni}(V_i) + o_p(1).$$

where H_n replaces F in the definition of G_{ni}. Under assumption 12, $E[G_{ni}(v)] = o[h^{1/4}/(nh^{1/2})]$, $Var[G_{ni}(v)] = O[(nh)^{-1}]$, and $Cov[G_{ni}(v), G_{nj}(v')] = o(h/n)$, uniformly over $v \in S_v$, by arguments similar to those of Bierens (1987). Therefore, $nh^{1/2}R_{n6} = o_p(1)$.

7. $(nh^{1/2})R_{n7}$:

$$nh^{1/2}R_{n7} = n^{-1}\sum_{i=1}^{n} w_{ni}\Delta_n(V_i)^2 = \mu + o_p(1)$$

by uniform convergence of Δ_n to Δ and the strong law of large numbers.

8. $(nh^{1/2})R_{n8}$:

$$(nh^{1/2})R_{n8} =$$

$$-2(\theta_n - \theta)'h^{1/2}\sum_{i=1}^{n} w_{ni}F'[v(X_i,\theta_n^*)]v_\theta(X_i,\theta_n^*)n^{-1/2}h^{-1/4}\Delta_n(V_i)$$

$$= O_p(h^{1/4}).$$

9. $(nh^{1/2})R_{n9}$:

$$(nh^{1/2})R_{n9} = -(\theta_n - \theta)'h^{1/2}\sum_{i=1}^{n} w_{ni}F'[v(X_i,\theta_n^*)v_\theta(X_i,\theta_n^*)$$

$$\cdot(\hat{F}_{ni}[v(X_i,\theta_n)] - F_{ni}[v(X_i,\theta)]) = o_p(h^{1/2})$$

by Lemma 8.

10. $(nh^{1/2})R_{n,10}$

$$(nh^{1/2})R_{n,10} =$$

$$(\theta_n - \theta)'h^{1/2}\sum_{i=1}^{n} w_{ni}F'[v(X_i,\theta_n^*)]v_\theta(X_i,\theta_n^*)[F_{ni}(V_i) - H_n(V_i)].$$

$= R_{n,10}^* + o_p(1)$,

by Lemma 8, where

$R_{n,10}^* =$

$$(\theta_n - \theta)' h^{1/2} \sum_{i=1}^{n} w_{ni} F'[v(X_i,\theta_n^*)] v_\theta(X_i,\theta_n^*)[G_{ni}(V_i) - J_{ni}(V_i)]$$

Also by Lemma 8, $E(R_{n,10}^*) = o(1)$. Arguments similar to those made for $Var(n^{1/2}h^{1/4}R_{n6})$ yield the result that $Var(R_{n,10}^*) = o(1)$, so $R_{n,10} = o_p(1)$ by Chebyshev's inequality.

11. $(nh^{1/2})R_{n,11}$:

$(nh^{1/2})R_{n,11} =$

$$(\theta_n - \theta)' h \sum_{i=1}^{n} w_{ni} F''[v(X_i,\theta_n^*)]^2 v_\theta(X_i,\theta_n^*) v_{\theta'}(X_i,\theta_n^*)(\theta_n - \theta) = O_p(h).$$

Q.E.D.

Discussion Paper Series, CentER, Tilburg University, The Netherlands:

(For previous papers please consult previous discussion papers.)

No.	Author(s)	Title
9071	J. Eichberger, H. Haller and F. Milne	Naive Bayesian Learning in 2 x 2 Matrix Games
9072	G. Alogoskoufis and F. van der Ploeg	Endogenous Growth and Overlapping Generations
9073	K.C. Fung	Strategic Industrial Policy for Cournot and Bertrand Oligopoly: Management-Labor Cooperation as a Possible Solution to the Market Structure Dilemma
9101	A. van Soest	Minimum Wages, Earnings and Employment
9102	A. Barten and M. McAleer	Comparing the Empirical Performance of Alternative Demand Systems
9103	A. Weber	EMS Credibility
9104	G. Alogoskoufis and F. van der Ploeg	Debts, Deficits and Growth in Interdependent Economies
9105	R.M.W.J. Beetsma	Bands and Statistical Properties of EMS Exchange Rates
9106	C.N. Teulings	The Diverging Effects of the Business Cycle on the Expected Duration of Job Search
9107	E. van Damme	Refinements of Nash Equilibrium
9108	E. van Damme	Equilibrium Selection in 2 x 2 Games
9109	G. Alogoskoufis and F. van der Ploeg	Money and Growth Revisited
9110	L. Samuelson	Dominated Strategies and Commom Knowledge
9111	F. van der Ploeg and Th. van de Klundert	Political Trade-off between Growth and Government Consumption
9112	Th. Nijman, F. Palm and C. Wolff	Premia in Forward Foreign Exchange as Unobserved Components
9113	H. Bester	Bargaining vs. Price Competition in a Market with Quality Uncertainty
9114	R.P. Gilles, G. Owen and R. van den Brink	Games with Permission Structures: The Conjunctive Approach
9115	F. van der Ploeg	Unanticipated Inflation and Government Finance: The Case for an Independent Common Central Bank

No.	Author(s)	Title
9116	N. Rankin	Exchange Rate Risk and Imperfect Capital Mobility in an Optimising Model
9117	E. Bomhoff	Currency Convertibility: When and How? A Contribution to the Bulgarian Debate!
9118	E. Bomhoff	Stability of Velocity in the G-7 Countries: A Kalman Filter Approach
9119	J. Osiewalski and M. Steel	Bayesian Marginal Equivalence of Elliptical Regression Models
9120	S. Bhattacharya, J. Glazer and D. Sappington	Licensing and the Sharing of Knowledge in Joint Ventures
9121	J.W. Friedman and L. Samuelson	An Extension of the "Folk Theorem" with Continuous Reaction Functions
9122	S. Chib, J. Osiewalski and M. Steel	A Bayesian Note on Competing Correlation Structures in the Dynamic Linear Regression Model
9123	Th. van de Klundert and L. Meijdam	Endogenous Growth and Income Distribution
9124	S. Bhattacharya	Banking Theory: The Main Ideas
9125	J. Thomas	Non-Computable Rational Expectations Equilibria
9126	J. Thomas and T. Worrall	Foreign Direct Investment and the Risk of Expropriation
9127	T. Gao, A.J.J. Talman and Z. Wang	Modification of the Kojima-Nishino-Arima Algorithm and its Computational Complexity
9128	S. Altug and R.A. Miller	Human Capital, Aggregate Shocks and Panel Data Estimation
9129	H. Keuzenkamp and A.P. Barten	Rejection without Falsification - On the History of Testing the Homogeneity Condition in the Theory of Consumer Demand
9130	G. Mailath, L. Samuelson and J. Swinkels	Extensive Form Reasoning in Normal Form Games
9131	K. Binmore and L. Samuelson	Evolutionary Stability in Repeated Games Played by Finite Automata
9132	L. Samuelson and J. Zhang	Evolutionary Stability in Asymmetric Games
9133	J. Greenberg and S. Weber	Stable Coalition Structures with Uni-dimensional Set of Alternatives

No.	Author(s)	Title
9134	F. de Jong and F. van der Ploeg	Seigniorage, Taxes, Government Debt and the EMS
9135	E. Bomhoff	Between Price Reform and Privatization - Eastern Europe in Transition
9136	H. Bester and E. Petrakis	The Incentives for Cost Reduction in a Differentiated Industry
9137	L. Mirman, L. Samuelson and E. Schlee	Strategic Information Manipulation in Duopolies
9138	C. Dang	The D'_2-Triangulation for Continuous Deformation Algorithms to Compute Solutions of Nonlinear Equations
9139	A. de Zeeuw	Comment on "Nash and Stackelberg Solutions in a Differential Game Model of Capitalism"
9140	B. Lockwood	Border Controls and Tax Competition in a Customs Union
9141	C. Fershtman and A. de Zeeuw	Capital Accumulation and Entry Deterrence: A Clarifying Note
9142	J.D. Angrist and G.W. Imbens	Sources of Identifying Information in Evaluation Models
9143	A.K. Bera and A. Ullah	Rao's Score Test in Econometrics
9144	B. Melenberg and A. van Soest	Parametric and Semi-Parametric Modelling of Vacation Expenditures
9145	G. Imbens and T. Lancaster	Efficient Estimation and Stratified Sampling
9146	Th. van de Klundert and S. Smulders	Reconstructing Growth Theory: A Survey
9147	J. Greenberg	On the Sensitivity of Von Neuman and Morgenstern Abstract Stable Sets: The Stable and the Individual Stable Bargaining Set
9148	S. van Wijnbergen	Trade Reform, Policy Uncertainty and the Current Account: A Non-Expected Utility Approach
9149	S. van Wijnbergen	Intertemporal Speculation, Shortages and the Political Economy of Price Reform
9150	G. Koop and M.F.J. Steel	A Decision Theoretic Analysis of the Unit Root Hypothesis Using Mixtures of Elliptical Models
9151	A.P. Barten	Consumer Allocation Models: Choice of Functional Form

No.	Author(s)	Title
9152	R.T. Baillie, T. Bollerslev and M.R. Redfearn	Bear Squeezes, Volatility Spillovers and Speculative Attacks in the Hyperinflation 1920s Foreign Exchange
9153	M.F.J. Steel	Bayesian Inference in Time Series
9154	A.K. Bera and S. Lee	Information Matrix Test, Parameter Heterogeneity and ARCH: A Synthesis
9155	F. de Jong	A Univariate Analysis of EMS Exchange Rates Using a Target Zone Model
9156	B. le Blanc	Economies in Transition
9157	A.J.J. Talman	Intersection Theorems on the Unit Simplex and the Simplotope
9158	H. Bester	A Model of Price Advertising and Sales
9159	A. Özcam, G. Judge, A. Bera and T. Yancey	The Risk Properties of a Pre-Test Estimator for Zellner's Seemingly Unrelated Regression Model
9160	R.M.W.J. Beetsma	Bands and Statistical Properties of EMS Exchange Rates: A Monte Carlo Investigation of Three Target Zone Models
9161	A.M. Lejour and H.A.A. Verbon	Centralized and Decentralized Decision Making on Social Insurance in an Integrated Market Multilateral Institutions
9162	S. Bhattacharya	Sovereign Debt, Creditor-Country Governments, and
9163	H. Bester, A. de Palma, W. Leininger, E.-L. von Thadden and J. Thomas	The Missing Equilibria in Hotelling's Location Game
9164	J. Greenberg	The Stable Value
9165	Q.H. Vuong and W. Wang	Selecting Estimated Models Using Chi-Square Statistics
9166	D.O. Stahl II	Evolution of Smart$_n$ Players
9167	D.O. Stahl II	Strategic Advertising and Pricing with Sequential Buyer Search
9168	T.E. Nijman and F.C. Palm	Recent Developments in Modeling Volatility in Financial Data
9169	G. Asheim	Individual and Collective Time Consistency
9170	H. Carlsson and E. van Damme	Equilibrium Selection in Stag Hunt Games
9201	M. Verbeek and Th. Nijman	Minimum MSE Estimation of a Regression Model with Fixed Effects from a Series of Cross Sections

No.	Author(s)	Title
9202	E. Bomhoff	Monetary Policy and Inflation
9203	J. Quiggin and P. Wakker	The Axiomatic Basis of Anticipated Utility; A Clarification
9204	Th. van de Klundert and S. Smulders	Strategies for Growth in a Macroeconomic Setting
9205	E. Siandra	Money and Specialization in Production
9206	W. Härdle	Applied Nonparametric Models
9207	M. Verbeek and Th. Nijman	Incomplete Panels and Selection Bias: A Survey
9208	W. Härdle and A.B. Tsybakov	How Sensitive Are Average Derivatives?
9209	S. Albæk and P.B. Overgaard	Upstream Pricing and Advertising Signal Downstream Demand
9210	M. Cripps and J. Thomas	Reputation and Commitment in Two-Person Repeated Games
9211	S. Albæk	Endogenous Timing in a Game with Incomplete Information
9212	T.J.A. Storcken and P.H.M. Ruys	Extensions of Choice Behaviour
9213	R.M.W.J. Beetsma and F. van der Ploeg	Exchange Rate Bands and Optimal Monetary Accommodation under a Dirty Float
9214	A. van Soest	Discrete Choice Models of Family Labour Supply
9215	W. Güth and K. Ritzberger	On Durable Goods Monopolies and the (Anti-) Coase-Conjecture
9216	A. Simonovits	Indexation of Pensions in Hungary: A Simple Cohort Model
9217	J.-L. Ferreira, I. Gilboa and M. Maschler	Credible Equilibria in Games with Utilities Changing during the Play
9218	P. Borm, H. Keiding, R. Mclean, S. Oortwijn and S. Tijs	The Compromise Value for NTU-Games
9219	J.L. Horowitz and W. Härdle	Testing a Parametric Model against a Semiparametric Alternative