EQUIVALENCE OF DIRECT, INDIRECT AND
SLOPE ESTIMATORS OR AVERAGE DERIVATIVES

by Thomas M. Stoker

Professor Thomas Stoker, Sloan School of Management, M.I.T.
E52-455, Cambridge, MA 02139

Abstract

If the regression of a response variable y on a k-vector of predictors x is denoted $g(x)=E(y|x)$, then the average derivative of y on x is defined as $\delta=E(g')$, where $g'\equiv\partial g/\partial x$. This paper compares the statistical properties of four estimators of δ: a "direct" estimator $\hat{\delta}_g$, formed by averaging pointwise kernel estimators of the derivative $g'(x)$; an "indirect" estimator $\hat{\delta}_f$ proposed by Härdle and Stoker(1987), based on averaging kernel density estimates; and two "slope" estimators $\hat{d}_g$ and $\hat{d}_f$, which are instrumental variables estimators of the (linear) coefficients of y regressed on x. Each estimator is shown to be a $\sqrt{N}$-consistent, asymptotically normal estimator of δ. Moreover, all the estimators are shown to be first-order equivalent. Some relative merits of the various estimators are discussed.

EQUIVALENCE OF DIRECT. INDIRECT AND SLOPE ESTIMATORS

OF AVERAGE DERIVATIVES

by Thomas M. Stoker

1. Introduction

If the regression of a response variable y on a k-vector of predictors x is denoted $g(x)=E(y|x)$, then the average derivative of y on x is defined as $\delta=E(g')$, with $g'\equiv\partial g/\partial x$. The estimation of average derivatives provides a semiparametric approach to measuring coefficients in index models – if $g(x)$ is structured as $g(x)=F(x'\beta)$, then δ is proportional to β, and so an estimator of δ will measure β up to scale (c.f. Stoker(1986) and Powell. Stock and Stoker(1987)). Following this connection. Härdle and Stoker(1987 – hereafter HS) have argued for the usefulness of average derivatives as generalized "coefficients" of y on x for inferring smooth multivariate regression relationships. Moreover. as detailed later. HS proposed an estimator of δ based on a two-step approach: i) nonparametrically estimate the marginal density $f(x)$ of x, and then ii) form a sample average using the estimated values of the density (and its derivatives).

In this paper we study several nonparametric estimators of the average derivative δ. which differ from the HS estimator in how the two-step approach is implimented. First is the natural sample analogue estimator of $\delta=E(g')$. namely the average of estimated values of the derivative $g'(x)$. Second are procedures that estimate δ as the (linear) slope coefficients of y regressed on x. where nonparametric estimators are used to construct appropriate instrumental variables. With reference to the title. a "direct" estimator is one based on approximating the conditional expectation $g(x)$. an "indirect" estimator is one based on approximating the marginal density $f(x)$, and a "slope" estimator refers to the slope coefficients of y regressed on x,

1

estimated with certain instrumental variables.[1] Each of the estimators is based on an observed random sample (y_i, x_i), i=1,....,N, and each procedure uses nonparametric kernel estimators to approximate either $f(x)$ or $g(x)$ (and their derivatives).

After laying out the framework. we introduce each of the average derivative estimators. and then present the results of the paper (proofs and technical assumptions are collected in the Appendix). In overview, we show that each procedure gives a $\sqrt{N}$ consistent. asymptotically normal estimator of δ. so that each procedure has precision properties that are comparable to parametric estimators. In addition. we show that all of the procedures are first-order equivalent. The implications of these findings as well as some relative practical merits of the estimators are then discussed.

2. The Framework and Nonparametric Ingredients

Formally. we assume that the observed data $(y_i.x_i)$, i=1.....N is a random sample from a distribution with density $f^*(y,x)$. Denote the marginal density of x as $f(x)$. its derivative as $f'(x)=\partial f/\partial x$. and (minus) the log-density derivative as $\ell(x)=-\partial \ln f/\partial x=-f'/f$. If $G(x)$ denotes the function:

$$(2.1) \qquad G(x) = \int y \, f^*(y,x) \, dy$$

then the regression $g(x)=E(y|x)$ of y on x is given as

$$(2.2) \qquad g(x) = \frac{G(x)}{f(x)} \quad .$$

The regression derivative $g' \equiv \partial g/\partial x$ is then expressed as

$$(2.3) \qquad g'(x) = \frac{G'(x)}{f(x)} - \frac{G(x)f'(x)}{f(x)^2}$$

Our interest is in estimating the average derivative $\delta=E(g')$. where expectation is taken with respect to x. For this. we utilize kernel estimators

of the functions $f(x)$, $g(x)$, etc., introduced as follows. Begin by defining the kernel estimator $\hat{G}(x)$ of $G(x)$, and the associated derivative estimator $\hat{G}'(x)$ of $G'(x)$:

$$(2.4) \qquad \hat{G}(x) = \frac{1}{Nh^k} \sum_{j=1}^{N} K\left[\frac{x - x_j}{h}\right] y_j$$

$$(2.5) \qquad \hat{G}'(x) = \frac{\partial \hat{G}(x)}{\partial x} = \frac{1}{Nh^{k+1}} \sum_{j=1}^{N} K'\left[\frac{x - x_j}{h}\right] y_j$$

where $K(u)$ is a kernel function.[2] $K' \equiv \partial K/\partial u$, and $h = h_N$ is a bandwidth parameter such that $h \to 0$ as $N \to \infty$. Next define the (Rosenblatt-Parzen) kernel estimator $\hat{f}(x)$ of the density $f(x)$, and the associated estimator $\hat{f}'(x)$ of the density derivative $f'(x)$ as[3]:

$$(2.6) \qquad \hat{f}(x) = \frac{1}{Nh^k} \sum_{j=1}^{N} K\left[\frac{x - x_j}{h}\right]$$

$$(2.7) \qquad \hat{f}'(x) = \frac{\partial \hat{f}(x)}{\partial x} = \frac{1}{Nh^{k+1}} \sum_{j=1}^{N} K'\left[\frac{x - x_j}{h}\right]$$

and the associated estimator of the negative log-density derivative $\ell(x)$:

$$(2.8) \qquad \hat{\ell}(x) \equiv -\frac{\partial \ln \hat{f}(x)}{\partial x} = -\frac{\hat{f}'(x)}{\hat{f}(x)}$$

With these constructs, define the (Nadaraya-Watson) kernel regression estimator $\hat{g}(x)$ as[4]

$$(2.9) \qquad \hat{g}(x) = \frac{\hat{G}(x)}{\hat{f}(x)}$$

The associated kernel estimator of the derivative $g'(x)$ is

$$(2.10) \qquad \hat{g}'(x) \equiv \frac{\partial \hat{g}(x)}{\partial x} = \frac{\hat{G}'(x)}{\hat{f}(x)} - \frac{\hat{G}(x)\hat{f}'(x)}{\hat{f}(x)^2}$$

We can now introduce the average derivative estimators to be studied.

3. Various Kernel Estimators of Average Derivatives

3.1 The Direct Estimator

In many ways the most natural technique for estimating $\delta = E(g')$ is to use a sample analogue, or an average of the values of g' across the sample. In this spirit, we define the "direct" estimator $\hat{\delta}_g$ of δ as the (trimmed) sample average of the estimated values $\hat{g}'(x_i)$, or

$$(3.1) \qquad \hat{\delta}_g = \frac{1}{N} \sum_{i=1}^{N} \hat{g}'(x_i) \, \hat{I}_i$$

where $\hat{I}_i = I[\hat{f}(x_i) > b]$ is the indicator variable that drops terms with estimated density smaller than a bound $b = b_N$, where $b \to 0$ as $N \to \infty$. The use of trimming is required for the technical analysis of $\hat{\delta}_g$, but also may represent a sensible practical correction. In particular, because $\hat{g}'(x)$ of involves division by $\hat{f}(x)$, erratic behavior may be induced into $\hat{\delta}_g$ by terms involving negligible estimated density.[5]

3.2 The Indirect Estimator of HS

For estimating average derivatives, the focus can be shifted from approximating the regression $g(x)$ to approximating the density $f(x)$, by applying integration by parts to $\delta = E(g')$:

$$(3.2) \qquad \delta = \int g'(x)f(x)dx = \int g(x) \left[- \frac{f'(x)}{f(x)} \right] f(x)dx = E[\ell(x)y] .$$

where the density $f(x)$ is assumed to vanish on its boundary of its support. HS propose the estimation of δ by the trimmed sample analogue of the RHS

expectation, where $\hat{\ell}(x)$ is used in place of $\ell(x)$. We define this indirect estimator of δ as

$$(3.3) \qquad \hat{\delta}_f = \frac{1}{N} \sum_{i=1}^{N} \hat{\ell}(x_i) \ \hat{I}_i \ v_i$$

As above, trimming is necessary for our technical analysis because $\hat{\ell}(x)$ involves division by the estimated density $\hat{f}(x)$.

3.3 Two Slope Estimators

By a "slope estimator", we refer to the coefficients of the linear equation

$$(3.4) \qquad y_i = \hat{c} + x_i^T \hat{d} + \hat{v}_i \qquad\qquad i=1.....N$$

which are estimated using appropriate instrumental variables, where x^T denotes the transpose of the (column) vector x.[6]

We can define an "indirect" slope estimator, following Stoker(1986). Since $E[\ell(x)]=0$, we can write the average derivative δ as the covariance $\Sigma_{\ell y}$ between $\ell(x)$ and y:

$$(3.5) \qquad \delta = E[\ell(x)y] = \Sigma_{\ell y} \qquad .$$

The connection to linear instrumental variables estimators is seem by applying (3.2.5) to the average derivative of x^T. In particular, since $\partial x^T / \partial x = Id$, the $k \times k$ identity matrix, we have

$$(3.6) \qquad Id = E\left[\frac{\partial x^T}{\partial x}\right] = E[\ell(x)x^T] = \Sigma_{\ell x}$$

where $\Sigma_{\ell x}$ is the matrix of covariances between components of $\ell(x)$ and x. Therefore, we can write δ as

5

(3.7) $\qquad \delta = (\Sigma_{\ell x})^{-1} \Sigma_{\ell y}$

This expression motivates the use of $(1, \ell(x_i))$ as an instrumental variable for estimating the coefficients of (3.4). The indirect slope estimator is the estimated analogue of this, namely the coefficient estimates that utilize $(1, \hat{\ell}(x_i)\hat{I}_i)$ as the instrumental variable, or

(3.8) $\qquad \hat{d}_f = (S_{\ell x})^{-1} S_{\ell y}$

where $S_{\ell x}$, $S_{\ell y}$ are the sample covariance matrices between $\hat{\ell}(x_i)\hat{I}_i$ and x, y respectively.

The "direct" slope estimator is defined using similar considerations. By a realignment of terms, the direct estimator $\hat{\delta}_g$ can be written as

(3.9) $\qquad \hat{\delta}_g = \dfrac{1}{N} \displaystyle\sum_{i=1}^{N} \hat{\omega}(x_i) \, y_i$

where $\hat{\omega}(x_i)$ takes the form

(3.10) $\qquad \hat{\omega}(x_i) = \dfrac{1}{Nh^k} \displaystyle\sum_{j=1}^{N} \dfrac{\hat{I}_j}{\hat{f}(x_j)} \left[h^{-1} K' \left[\dfrac{x_j - x_i}{h} \right] - \dfrac{\hat{f}'(x_j)}{\hat{f}(x_j)} K \left[\dfrac{x_j - x_i}{h} \right] \right]$

By following the same logic as for the indirect slope estimator, we can define the direct slope estimator as the coefficient estimates of (3.4) that use $(1, \hat{\omega}(x_i))$ as the instrumental variable, namely

(3.11) $\qquad \hat{d}_g = (S_{\omega x})^{-1} S_{\omega y}$

where $S_{\omega x}$, $S_{\omega y}$ are the sample covariance matrices between $\hat{\omega}(x_i)$ and x, y respectively.

6

4. Equivalence of Direct, Indirect and Slope Estimators of Average Derivatives

The central result of the paper is Theorem 1, which characterizes the asymptotic distribution of the direct estimator $\hat{\delta}_g$:

Theorem 1: Given Assumptions 1 through 6 stated in the Appendix, as

(i) $N \to \infty$, $h \to 0$, $b \to 0$, $h^{-1}b \to 0$;

(ii) for some $\varepsilon > 0$, $b^4 N^{1-\varepsilon} h^{2k+2} \to \infty$;

(iii) $Nh^{2p-2} \to 0$;

then

$$(4.1) \qquad \sqrt{N} \, (\hat{\delta}_g - \delta) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \{r(y_i, x_i) - E[r(y,x)]\} \quad + o_p(1)$$

where

$$(4.2) \qquad r(y,x) \equiv g'(x) - [y - g(x)] \frac{f'(x)}{f(x)} \quad .$$

so that $\sqrt{N}(\hat{\delta}_g - \delta)$ has a limiting normal distribution with mean 0 and variance Σ, where Σ is the covariance matrix of $r(y,x)$.

Comparing Theorem 1 with Theorem 3.1 of HS indicates that $\sqrt{N}(\hat{\delta}_g - \delta)$ and $\sqrt{N}(\hat{\delta}_f - \delta)$ have a limiting normal distribution with the same variance. Moreover, examination of the proof of HS Theorem 3.1 shows that

$$(4.3) \qquad \sqrt{N} \, (\hat{\delta}_f - \delta) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \{r(y_i, x_i) - E[r(y,x)]\} \quad + o_p(1)$$

was established. Consequently, under the conditions of both of these theorems, we can conclude that the direct estimator $\hat{\delta}_g$ and the indirect estimator $\hat{\delta}_f$ are first-order equivalent:

Corollary 1: Given Assumptions 1 through 7 stated in the Appendix and conditions i)-iii) of Theorem 1, we have

$$(4.4) \qquad \sqrt{N} \; [\hat{\delta}_g - \hat{\delta}_f] = o_p(1).$$

This result follows from the coincidence of the conclusions of two separate arguments. In particular, the author searched without success for reasons that the equivalence could be obvious, but $\hat{\delta}_g$ and $\hat{\delta}_f$ approach the common limiting distribution in significantly different ways.

The asymptotic distribution of the indirect slope estimator $\hat{d}_f$ arises from examining its differences from $\hat{\delta}_f$ in two steps. First we establish that the difference between the underlying sample moments and the sample covariances are asymptotically inconsequential, or

Corollary 2: Given Assumptions 1 through 9 stated in the Appendix and conditions i)-iii) of Theorem 1, we have

$$(4.5) \qquad \sqrt{N} \; [\hat{\delta}_f - S_{\ell y}] = o_p(1).$$

Second, we establish that $S_{\ell x}$ converges to its limit Id at a faster rate than $\sqrt{N}$, which allows us to conclude

Corollary 3: Given Assumptions 1 through 11 stated in the Appendix and conditions i)-iii) of Theorem 1. we have

$$(4.6) \qquad \sqrt{N} \ [\hat{d}_f - \hat{\delta}_f] = o_p(1).$$

Analogous arguments apply to the relationship between the direct slope estimator $\hat{d}_g$ and the direct estimator $\hat{\delta}_g$. These are summarized as

Corollary 4: Given Assumptions 1 through 6 and 8 stated in the Appendix and conditions i)-iii) of Theorem 1. we have

$$(4.7) \qquad \sqrt{N} \ [\hat{\delta}_g - S_{\omega y}] = o_p(1).$$

Corollary 5: Given Assumptions 1 through 6, 8 and 10 stated in the Appendix and conditions i)-iii) of Theorem 1. we have

$$(4.8) \qquad \sqrt{N} \ [\hat{d}_g - \hat{\delta}_g] = o_p(1).$$

This completes the results of the paper. In sum. Theorem 1 establishes the asymptotic distribution of $\hat{\delta}_g$. and Corollaries 2-5 state that the difference between any two of $\hat{\delta}_g$. $\hat{\delta}_f$. $\hat{d}_g$. $\hat{d}_f$ (and $S_{\ell y}$. $S_{\omega y}$) is $o_p(1/\sqrt{N})$, so that all of the estimators are first-order equivalent, each estimating δ with the same statistical efficiency.

5. Remarks and Discussion

The most complicated demonstration of the paper is the proof of Theorem 1, which shows the characterization of $\hat{\delta}_g$ directly. This proof follows the format of HS and Powell, Stock and Stoker (1987): $\hat{\delta}_g$ is linearized by appealing to uniformity properties, the linearized version is approximated by the (asymptotically normal) sum of U-statistics with kernels that vary with N, and the bias is analyzed on the basis of the pointwise bias of $\hat{g}'$. The bandwidth and trimming bound conditions are interpretable as in HS: the trimming bound b must converge to 0 slowly, the rate of convergence of the bandwidth h is bounded above to insure the proper bias properties, and bounded below by the requirements of asymptotic normality and uniform pointwise convergence. As typically necessary for $\sqrt{N}$ convergence of averages of nonparametric estimators, the approximating functions must be (asymptotically) undersmoothed.

The fact that all of the estimators are asymptotically equivalent permits flexibility in the choice of estimating procedure, without loss of efficiency. The equivalence of direct and indirect procedures (of either type) gives a statistical justification for choosing either to approximate the regression function g(x) or approximate the density function f(x) for estimating the average derivative δ. Corollaries 2 and 4 state that the same asymptotic behavior arises from statistics computed from the basic data or data that is written as deviations from sample means, and indicates that the same asymptotic behavior is obtained for slope estimators whether a constant is included in the linear equation or not. Corollaries 3 and 5 permit the use of a instrumental variables coefficient, or "ratio of moments" type of average derivative estimator.

Choice between the various estimators should be based on features of the application at hand, although some indications of estimator performance can be

learned through Monte Carlo simulation and the study of second-order

efficiency (or "deficiency") properties. The estimators are of comparable

computational simplicity: while direct estimators involve slightly more

complex formulae than indirect ones, for given values of h and b, each

estimator involves only a single computation of order at most N^2.

In closing, some general observations can be made on the relative merits

of the various procedures. First, on the grounds of the required technical

assumptions, the smoothness conditions required for asymptotic normality of

direct estimators are slightly stronger than those for indirect estimators:

namely $G(x)$ and $f(x)$ are assumed to be p^{th} order differentiable for Theorem 1.

but only $f(x)$ for Theorem 3.1 of HS. Consequently, if one is studying a

problem where $g'(x)$ exists a.s., but is suspected to be discontinuous for one

or more values of x, then the indirect estimators are preferable.

The other main difference in the required assumptions is the condition

that $f(x)$ vanish on the boundary (Assumption 7), that is required for

analyzing indirect estimators but not required for direct estimators. The role

of this condition can impinge on the way each estimator measures $\delta=E(g')$ in

small samples. In particular, for a fixed value of the trimming constant b. $\hat{\delta}_f$

and $\hat{d}_f$ measure $E[y(-f'/f)I(f(x)>b)]$, and $\hat{\delta}_g$ and $\hat{d}_g$ measure $E[g'I(f(x)>b)]$.

These two expectations differ by boundary terms, that under Assumption 7 will

vanish in the limit as b approaches zero. These differences are unlikely to be

large in typical situations, as they involve only terms in the "tail" regions

of $f(x)$. But some related points can be made that favor the direct estimators

$\hat{\delta}_g$ and $\hat{d}_g$, in cases where such "tail regions" are not negligible. When the

structure of g' is relatively stable over the region of small density, $\hat{\delta}_g$ may

be subject to more limited bias. For instance, in an index model problem where

$g(x)$ takes the form $g(x)=F(x'\beta)$ for coefficients β. then $g'(x)=dG/d(x'\beta) \beta \equiv$

$\gamma(x) \beta$ is proportional to β for all values of x, where the proportionality

constant $\gamma(x)$ can vary with x. In this case $\hat{\delta}_g$ and $\hat{d}_g$ measure $E[g'I(f(x)>b)] = E[\gamma(x)I(f(x)>b)] \beta = \gamma_b \beta$. which is still proportional to β. whereas $\hat{\delta}_f$ and $\hat{d}_f$ measures an expectation that differs from $\gamma_b \beta$ by boundary terms. Thus, for estimating β up to scale, direct estimators avoid some small sample trimming biases inherent to indirect estimators.

Moreover, for certain practical situations, it may be useful to estimate the average derivative over subsets of the data. If the density $f(x)$ does not vanish on the boundary of the subsets, then boundary term differences can be introduced between "subset based" direct and indirect estimators. Formally, let $I_A(x)=I[x\in A]$ be the indicator function of a convex subset A with nonempty interior, and $\delta_A = E[g'I_A]$ the average derivative over the subset A. In this case, the direct estimator $\hat{\delta}_{gA} = N^{-1} \sum \hat{g}'(x_i) \hat{I}_i I_A(x_i)$ (or its "slope" version) can be shown to be a $\sqrt{N}$ consistent estimator of δ_A. but $\hat{\delta}_{fA} = -N^{-1} \sum y_i [\hat{f}'(x_i)/\hat{f}(x_i)] \hat{I}_i I_A(x_i)$ will be a $\sqrt{N}$-consistent estimator of $E[y(-f'/f)I_A]$. The difference between δ_A and the latter expectation will be boundary terms of the form $g(x)f(x)$ evaluated on the boundary of A, which may be significant if A is restricted to a region of high density. Therefore, direct estimators are preferable for measuring average derivatives over certain kinds of data subsets.

Finally, the potential practical advantages of the slope estimators derive from their "ratio" form, which may molify some small sample errors of approximation. For instance. $\hat{\delta}_f$ is affected by the overall level of the values $\hat{\ell}(x_i)$, whereas $\hat{d}_f$ is not. Similarly, deviations due to outliers and other small sample problems may influence slope estimators less than their "sample moment" counterparts.

Appendix: Assumptions and Proofs

Assumptions for Theorem 1.

1. The support Ω of f is a convex, possibly unbounded subset of R^k with nonempty interior. The underlying measure of (y,x) can be written as $\nu_y \times \nu_x$, where ν_x is Lebesgue measure.

2. All derivatives of $f(x)$ and $G(x)=g(x)f(x)$ of order $p=k+2$ exist.

3. The kernel function $K(u)$ has bounded support $S=\{u|\ |u|\leq 1\}$, is symmetric, $K(u)=0$ for $u\in dS=\{u|\ |u|=1\}$, and is of order p:

$$\int K(u)du = 1$$

$$\int u_1^{\ell_1}\ldots u_k^{\ell_{p'}} K(u)du = 0 \qquad \ell_1+\ldots+\ell_{p'} < p$$

$$\int u_1^{\ell_1}\ldots u_k^{\ell_{p'}} K(u)du \neq 0 \qquad \ell_1+\ldots+\ell_{p'} = p$$

4. The components of the random vectors $\partial g/\partial x$ and $[\partial \ln f/\partial x]y$ have finite second moments. Also, f, g, ℓ satisfy the following local Lipschitz conditions: For ν in a open neighborhood of 0, there exists functions ω_f, ω_g, ω_{fg}, $\omega_{f'}$, $\omega_{g'}$ and ω_ℓ such that,

$$|\ f(x+\nu) - f(x)\ | < \omega_f(x)\ |\nu|$$
$$|\ g(x+\nu) - g(x)\ | < \omega_g(x)\ |\nu|$$
$$|\ (gf)(x+\nu) - (gf)(x)\ | < \omega_{gf}(x)\ |\nu|$$

$$|\ f'(x+\nu) - f'(x)\ | < \omega_{f'}(x)\ |\nu|$$

$$|\ g'(x+\nu) - g'(x)\ | < \omega_{g'}(x)\ |\nu|$$

$$|\ \ell(x+\nu) - \ell(x)\ | < \omega_\ell(x)\ |\nu|$$

with $E[g'(x)\omega_f(x)]^2<\infty$, $E[\{f'(x)/f(x)\}\omega_g(x)]^2<\infty$, $E[\omega_{fg}(x)]^2<\infty$, $E[g(x)\omega_{f'}(x)]^2<\infty$, $E[\omega_{g'}(x)]^2<\infty$, $E[y\ \omega_\ell(x)]^2<\infty$. Finally, $M_2(x)=E(y^2|x)$ is continuous in x.

Let $A_N=\{x|f(x)>b\}$ and $B_N=\{x|f(x)\leq b\}$.

5. As $N \to \infty$.

$$\int_{B_N} g'(x)f(x)dx = o(N^{-1/2})$$

6. If $G_\iota^{(p)}$, $f_\iota^{(p)}$ denote any p^{th} order partial derivatives of G and f respectively, then $G_\iota^{(p)}$, $f_\iota^{(p)}$ are locally Hölder continuous: there exists $\gamma > 0$ and functions $c_G(x)$ and $c_f(x)$ such that for all ν in an open neighborhood of 0. $|G_\iota^{(p)}(x+\nu) - G_\iota^{(p)}(x)| \le c_G(x)|\nu|^\gamma$ and $|f_\iota^{(p)}(x+\nu) - f_\iota^{(p)}(x)| \le c_f(x)|\nu|^\gamma$. The $p+\gamma$ moments of $K(.)$ exist. Moreover

$$\int_{A_N} G_\iota^{(p)}(x)dx \le M < \infty$$

$$h^\gamma \int_{A_N} c_G(x)\, dx \le M < \infty$$

$$h\int_{A_N} [f'(x)/f(x)]G_\iota^{(p)}(x)dx \le M < \infty$$

$$h^{\gamma+1} \int_{A_N} c_G(x)\, [f'(x)/f(x)]dx \le M < \infty$$

$$\int_{A_N} g(x)f_\iota^{(p)}(x)dx \le M < \infty$$

$$h^\gamma \int_{A_N} c_f(x)\, g(x)dx \le M < \infty$$

$$h\int_{A_N} [-g'(x)+g(x)(f'(x)/f(x))]f_\iota^{(p)}(x)dx \le M < \infty$$

$$h^{\gamma+1} \int_{A_N} c_f(x)[-g(x)+g(x)(f'(x)/f(x))]dx \le M < \infty$$

Note that if there exists $B>0$ such that when $f(x)\neq0$, then $f(x)>B$, the integral conditions of Assumptions 5 and 6 are vacuous.

14

Additional Assumption for Corollary 1 (from Theorem 3.1 of HS):

7. $f(x)=0$ for all $x \in d\Omega$, where $d\Omega$ is the boundary of Ω. With reference to Assumption 4, $E[\ell(x)y\omega_f]^2$ exists. There exists a function $\omega_{\ell g}$ such that

$$| \ell(x+\nu)g(x+\nu) - \ell(x)g(x) | < \omega_{\ell g}(x) |\nu|$$

such that $E[\omega_{\ell g}]^2$ exists. Finally, as $N \to \infty$,

$$\int_{B_N} g(x)f'(x)dx = o(N^{-1/2})$$

Additional Assumptions for Corollaries 2 and 4:

8. The variance of y exists. The conditions of 4-6 are obeyed for $y=g(x)=1$.

9. The conditions of 7 are obeyed for $y=g(x)=1$.

Additional Assumption for Corollaries 3 and 5:

10. The variance of x exists. The conditions of 4-6 are obeyed for $y=g(x)=x_\varrho$, $\varrho=1 \ldots k$, where x_ϱ denotes the ϱth component of x.

11. The conditions of 7 are obeyed for $y=g(x)=x_\varrho$, $\varrho=1 \ldots,k$, where x_ϱ denotes the ϱth component of x.

<u>Proof of Theorem 1</u>:

The proof strategy follows that of Theorem 3.1 of Härdle and Stoker(1987 - hereafter HS), although the details differ substantively. As a preliminary, we note several results on the uniform convergence of kernel estimators, that will be used occasionally in the main proof. As in HS, we note that condition (iii) requires that the pointwise mean square errors of $\hat{f}(x)$, $\hat{f}'(x)$, $\hat{G}(x)$ and $\hat{G}'(x)$ are dominated by their variances. Consequently, since the set $\{x|f(x)\geq b\}$ is compact and $b^{-1}h\rightarrow 0$, by following the arguments of Collomb and Härdle(1986) or Silverman(1978), we can assert

(A.1a) $\sup |\hat{f}(x) - f(x)| \; I[f(x)>b] = 0_p[(N^{1-(\varepsilon/2)}h^k)^{-1/2}]$

(A.1b) $\sup |\hat{f}'(x) - f'(x)| \; I[f(x)>b] = 0_p[(N^{1-(\varepsilon/2)}h^{k+2})^{-1/2}]$

(A.1c) $\sup |\hat{G}(x) - G(x)| \; I[f(x)>b] = 0_p[(N^{1-(\varepsilon/2)}h^k)^{-1/2}]$

(A.1d) $\sup |\hat{G}'(x) - G'(x)| \; I[f(x)>b] = 0_p[(N^{1-(\varepsilon/2)}h^{k+2})^{-1/2}]$

for any $\varepsilon>0$. (The $N^{-\varepsilon/2}$ term is included instead of the $(\ln N)^c$ term introduced through discretization, as in Stone(1980). This is done to avoid further complication in the exposition by always carrying along the $(\ln N)$ terms).

We define two (unobservable) "estimators" to be studied and then related to $\hat{\delta}_g$. First, define the estimator $\bar{\delta}$ based on trimming with respect to the true density value:

(A.2) $\bar{\delta} = N^{-1} \sum_{i=1}^{N} \hat{g}'(x_i) \; I_i$

where $I_i \equiv I[f(x_i)>b]$, $i=1,\ldots,N$. Next, for the body of the analysis of asymptotic distribution, a Taylor expansion of $\hat{g}'$ suggests defining the "linearized" estimator $\tilde{\delta}$:

16

$$(A.3) \qquad \widetilde{\delta} = N^{-1} \sum_{i=1}^{N} \left[g'(x_i)I_i + [\hat{G}'(x_i) - G'(x_i)] \frac{I_i}{f(x_i)} \right.$$

$$- [\hat{G}(x_i) - G(x_i)] \frac{f'(x_i)I_i}{f(x_i)^2} - [\hat{f}_i'(x) - f_i'(x)] \frac{g(x_i)I_i}{f(x_i)}$$

$$\left. + [\hat{f}(x_i) - f(x_i)] \left[-\frac{g'(x_i)}{f(x_i)} + \frac{g(x_i)f'(x_i)}{f(x_i)^2} \right] I_i \right]$$

The linearized estimator $\widetilde{\delta}$ can be rewritten as the sum of "average kernel estimators" as

$$(A.4) \qquad \widetilde{\delta} = \widetilde{\delta}_0 + \widetilde{\delta}_1 - \widetilde{\delta}_2 - \widetilde{\delta}_3 + \widetilde{\delta}_4$$

where

$$(A.5) \qquad \widetilde{\delta}_0 = N^{-1} \sum_{i=1}^{N} g'(x_i)I_i \ : \qquad \widetilde{\delta}_1 = N^{-1} \sum_{i=1}^{N} \hat{G}'(x_i) \frac{I_i}{f(x_i)}$$

$$\widetilde{\delta}_2 = N^{-1} \sum_{i=1}^{N} \hat{G}(x_i) \frac{f'(x_i)I_i}{f(x_i)^2} \ : \qquad \widetilde{\delta}_3 = N^{-1} \sum_{i=1}^{N} \hat{f}'(x_i) \frac{g(x_i)I_i}{f(x_i)}$$

$$\widetilde{\delta}_4 = N^{-1} \sum_{i=1}^{N} \hat{f}(x_i) \left[-\frac{g_i'(x_i)}{f(x_i)} + \frac{g(x_i)f'(x_i)}{f(x_i)^2} \right] I_i$$

With these definitions, the proof now consists of four steps. summarized as

Step 1. <u>Linearization</u>: $\sqrt{N}(\bar{\delta} - \widetilde{\delta}) = o_p(1)$.

Step 2. <u>Asymptotic Normality</u>: $\sqrt{N}[\widetilde{\delta} - E(\widetilde{\delta})]$ has a limiting normal distribution with mean 0 and variance Σ.

Step 3. <u>Bias</u>: $[E(\widetilde{\delta}) - \delta] = o(N^{-1/2})$.

Step 4. <u>Trimming</u>: $\sqrt{N}(\hat{\delta}_g - \delta)$ has the same limiting behavior as $\sqrt{N}(\bar{\delta} - \delta)$.

The combination of Steps 1-4 yields Theorem 1

Step 1. <u>Linearization</u>: First define the notation

(A.6a) $\quad \zeta_f(x) = f(x)^{-1}[\hat{f}(x) - f(x)] \, I[f(x)>b]$

(A.6b) $\quad \zeta_{f'}(x) = f(x)^{-1}[\hat{f}'(x) - f'(x)] \, I[f(x)>b]$

(A.6c) $\quad \zeta_G(x) = f(x)^{-1}[\hat{G}(x) - G(x)] \, I[f(x)>b]$

(A.6d) $\quad \zeta_{G'}(x) = f(x)^{-1}[\hat{G}'(x) - G'(x)] \, I[f(x)>b]$

(A.6e) $\quad \zeta_{\hat{f}}(x) = \hat{f}(x)^{-1}[\hat{f}(x) - f(x)] \, I[f(x)>b]$

Some arithmetic gives the following expression, where all summations run from $i=1,\ldots,N$, and i subscripts denote evaluation at x_i (i.e. $f_i \equiv f(x_i)$, $\zeta_{fi} \equiv \zeta_f(x_i)$, etc.);

(A.7)
$$\sqrt{N}(\bar{\delta} - \tilde{\delta}) = N^{-1/2}\Sigma \, 2g_i\zeta_{f'i}\zeta_{fi} - N^{-1/2}\Sigma \, \zeta_{Gi}\zeta_{f'i} + N^{-1/2}\Sigma \, 2(f_i'/f_i)\zeta_{Gi}\zeta_{fi}$$

$$- N^{-1/2}\Sigma \, \zeta_{G'i}\zeta_{fi} + N^{-1/2}\Sigma \, (g_i' - G_i f_i'/f_i^2)\zeta_{fi}\zeta_{\hat{f}i} - N^{-1/2}\Sigma \, (G_i f_i'/f_i^2)\zeta_{\hat{f}i}^2$$

$$+ N^{-1/2}\Sigma \, 2\zeta_{Gi}\zeta_{f'i}\zeta_{fi} + N^{-1/2}\Sigma \, 2(f_i'/f_i)\zeta_{Gi}\zeta_{fi}\zeta_{\hat{f}i} + N^{-1/2}\Sigma \, 2g_i\zeta_{f'i}\zeta_{fi}\zeta_{\hat{f}i}$$

$$- N^{-1/2}\Sigma \, (f_i'/f_i)\zeta_{Gi}\zeta_{\hat{f}i}^2 - N^{-1/2}\Sigma \, g_i\zeta_{f'i}\zeta_{\hat{f}i}^2 - N^{-1/2}\Sigma \, \zeta_{Gi}\zeta_{f'i}\zeta_{\hat{f}i}^2$$

$$+ N^{-1/2}\Sigma \, 2\zeta_{Gi}\zeta_{f'i}\zeta_{fi}\zeta_{\hat{f}i}$$

Examining $\sqrt{N}(\bar{\delta}-\tilde{\delta})$ term by term gives the result. In particular, by (A.1a-d),
$\sup|\zeta_f(x)|=O_p[b^{-1}(N^{1-(\varepsilon/2)}h^k)^{-1/2}]$, $\sup|\zeta_{f'}(x)|=O_p[b^{-1}(N^{1-(\varepsilon/2)}h^{k+2})^{-1/2}]$,
$\sup|\zeta_G(x)|=O_p[b^{-1}(N^{1-(\varepsilon/2)}h^k)^{-1/2}]$, $\sup|\zeta_{G'}(x)|=O_p[b^{-1}(N^{1-(\varepsilon/2)}h^{k+2})^{-1/2}]$ and
$\sup|\zeta_{\hat{f}}(x)|=O_p[b^{-1}(N^{1-(\varepsilon/2)}h^k)^{-1/2}]$, the latter using $b^2 N^{1-(\varepsilon/2)}h^k \to \infty$, which is implied by condition (ii). For the first term of (A.7)

$$\left| N^{-1/2}\Sigma \, 2g_i\zeta_{f'i}\zeta_{fi} \right| \le \sqrt{N} \, \sup|\zeta_{f'}(x)| \, \sup|\zeta_f(x)| \, \frac{\Sigma|g_i|I_i}{N}$$

$$= O_p(b^{-2}N^{-(1/2)+(\varepsilon/2)}h^{-(2k+2)/2}) = o_p(1)$$

since $\Sigma|g_i|I_i/N$ is bounded in probability by Chebyshev's inequality, and $b^4 N^{1-\varepsilon}h^{2k+2} \to \infty$ by condition (ii). The other terms are analyzed similarly, allowing us to conclude

$$\sqrt{N}|\bar{\delta} - \tilde{\delta}| = 0_p(b^{-2}N^{-(1/2)+(\varepsilon/2)}h^{-(2k+2)/2}) = o_p(1)$$

QED 1.

Step 2: <u>Asymptotic Normality</u>: We now show that $\sqrt{N}[\tilde{\delta} - E(\tilde{\delta})]$ has a limiting normal distribution, by showing that each of the terms of (A.4) is $\sqrt{N}$ equivalent to an ordinary sample average. For $\tilde{\delta}_0$, we have that

$$(A.8) \qquad \sqrt{N}[\tilde{\delta}_0 - E(\tilde{\delta}_0)] = N^{-1/2}\left[\sum_{i=1}^{N}\{r_0(x_i) - E[r_0(x)]\}\right] + o_p(1)$$

where $r_0(x) = g'(x)$.

since Var(g') exists and $b\to 0$ and $N\to\infty$.

The analyses of the average kernel estimators $\tilde{\delta}_1$, $\tilde{\delta}_2$, $\tilde{\delta}_3$ and $\tilde{\delta}_4$ are similar, and so we present the details for $\tilde{\delta}_1$. Note that $\tilde{\delta}_1$ can be approximated by a U-statistic, namely

$$U_1 = \left[\begin{array}{c} N \\ 2 \end{array}\right]^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} p_{1N}(z_i, z_j)$$

where $z_i = (y_i, x_i)$ and

$$p_{1N}(z_i, z_j) = \frac{1}{2}\left[\frac{1}{h}\right]^{k+1} K'\left[\frac{x_i - x_j}{h}\right]\left[\frac{y_j I_i}{f(x_i)} - \frac{y_i I_j}{f(x_j)}\right]$$

where $K' \equiv \partial K/\partial u$. In particular, we have

$$\sqrt{N}[\tilde{\delta}_1 - E(\tilde{\delta}_1)] = \sqrt{N}[U_1 - E(U_1)] - N^{-1}\{\sqrt{N}[U_1 - E(U_1)]\}$$

$$+ \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\frac{1}{Nh^{k+1}}\right]K'(0)\left[\left[\frac{y_i I_i}{f(x_i)}\right] - E\left[\frac{y_i I_i}{f(x_i)}\right]\right]$$

The second term on the RHS will converge in probability to zero provided $\sqrt{N}[U_1 - E(U_1)]$ has a limiting distribution, which we show later. In general, if

19

$K'(0)=c$, then the third term converges in probability to zero, because its variance is bounded by $cc'(1/Nh^{k+2})^2(h/b)^2 E(y_i^2 I_i)=o(1)$, since $Nh^{k+2} \to \infty$ and $h/b \to 0$. However, for analyzing $\tilde{\delta}_1$, we note that the symmetry of $K(.)$ implies $K'(0)=0$, so the third term is identically zero. Consequently, we have

$$(A.9) \qquad \sqrt{N}[\tilde{\delta}_1 - E(\tilde{\delta}_1)] = \sqrt{N}[U_1 - E(U_1)] + o_p(1)$$

We now focus on U_1.

By Lemma 3.1 of Powell, Stock and Stoker(1987), we can assert

$$(A.10) \qquad \sqrt{N}[U_1 - E(U_1)] = N^{-1/2}\left[\sum_{i=1}^{N}\{r_{1N}(z_i) - E[r_{1N}(z)]\}\right] + o_p(1)$$

$$\text{where } r_{1N}(z) = 2E[p_{1N}(z,z_j)|z]$$

provided that $E[|p_{1N}(z_i,z_j)|^2]=o(N)$. To verify this condition, let $M_1(x_i)=E(y_i I_i|x_i)$, $M_2(x_i) \equiv E(y_i^2|x_i)$ and $R(x_i)=I(f(x_i)>b)$. Then

$$E[|p_{1N}(z_i,z_j)|^2]$$

$$\leq \frac{1}{4b^2 h^{2k+2}}\int\left|K'\left[\frac{x_i - x_j}{h}\right]\right|^2 [M_2(x_i)R(x_j)+M_2(x_j)R(x_i)-2M_1(x_i)M_1(x_j)]$$

$$f(x_i)f(x_j)dx_i dx_j$$

$$= \frac{1}{4b^2 h^{k+2}}\int|K'(u)|^2[M_2(x_i)R(x_i+hu) + M_2(x_i+hu)R(x_i)$$

$$- 2M_1(x_i)M_1(x_i+hu)] f(x_i)f(x_i+hu)dx_i du$$

$$= O(b^{-2}h^{-(k+2)}) = O[N(b^2 Nh^{k+2})^{-1}] = o(N)$$

since $b^2 Nh^{k+2} \to \infty$ is implied by condition (ii). Thus (A.10) is valid.

The final step in the analysis of $\tilde{\delta}_1$ is to show that the average of $r_{1N}(z_i)=E[2p_{1N}(z_i,z_j)|z_i]$ of (A.10) is equivalent to a sample average whose components do not vary with N. For this, we write out $r_{1N}(z_i)$ as

$$r_{1N}(z_i) = \frac{1}{h^{k+1}} \int K'\left[\frac{x_i - x}{h}\right] \left[\frac{g(x)I_i}{f(x_i)} - \frac{y_i I[f(x)>b]}{f(x)}\right] f(x)dx$$

$$= -\frac{I_i}{f(x_i)} \int \frac{1}{h} K'(u) \, g(x_i+hu)f(x_i+hu)du$$

$$+ y_i \int \frac{1}{h} K'(u) \, I[f(x_i+hu)>b]du$$

$$= \frac{I_i}{f(x_i)} \int K(u) \, (gf)'(x_i+hu)du$$

$$+ y_i \int \frac{1}{h} K'(u) \, I[f(x_i+hu)\leq b]du$$

since $(y_i/h)(\int K'(u)du)=0$. Now define $r_1(z_i)$ and $t_{1N}(z_i)$ as

(A.11) $\qquad r_1(z_i) = g'(x_i) + g(x_i) \dfrac{f'(x_i)}{f(x_i)}$.

(A.12) $\quad t_{1N}(z_i) \equiv r_{1N}(z_i)-r_1(z_i) = \dfrac{I_i}{f(x_i)} \int K(u) \, [(gf)'(x_i+hu)-(gf)'(x_i)]du$

$$+ (1-I_i)r_1(z_i) + y_i \int \frac{1}{h} K'(u) \, I[f(x_i+hu)\leq b]du$$

It is easy to see that the second moment $E[t_{1N}(z)^2]$ vanishes as $N\to\infty$. By the Lipschitz condition of Assumption 4 the second moment of the first RHS term is bounded by $(h/b)^2(\int|u|K(u)du)^2 E(\omega_{gf}(x)^2)=O(h/b)^2=o(1)$, and the second moment of the second RHS term vanishes because $b\to0$ and $Var(r_1)$ exists. For the final RHS term, notice that each component of the integral

$$a(x) = \int \frac{1}{h} K'(u) \, I[f(x+hu)\leq b]du$$

will be nonzero only if there exists x^* such that $|x-x^*|<h$ and $f(x^*)=b$, because if $f(x+hu)<b$ for all u, $|u|\leq1$, then $a(x)=\int K'(u)du=0$. Now, consider the ℓth component $a_\ell(x)$ of $a(x)$, and define the "marginal density" $K_{(\ell)}=\int K(u)du_\ell$ and the "conditional density" $K_\ell=K/K_{(\ell)}$. The kernel K can be chosen such that K_ℓ is bounded (say by choosing K as a product of univariate kernels

$K = \Pi K_\ell(u_\ell))$. Since K vanishes for u such that $|u|=1$, applying integration by parts absorbs h^{-1}, and shows that $a_\ell(x)$ is the expected value of $K_\ell(u)$ over u values such that $f(x+hu)=b$, where expectation is taken with respect to $K_{(\ell)}$. Consequently, each component of $a(x)$ is bounded, so that there exists a^* such that $|a(x)|<a^*$. Finally define $b^* = \sup\{f(x+hu)|f(x)=b,|u|\leq 1\}$. We then have

$$E\left[\left|y\int \frac{1}{h} K'(u)\ I[f(x+hu)\leq b]du\right|^2\right] \leq (a^*)^2\ E[y^2 I[f(x)\leq b^*]]$$

$$= o(1)$$

since $b^*\to 0$ as $h\to 0$ and $b\to 0$. Thus $E[t_{1N}(z)^2]=o(1)$.

This observation suffices to show asymptotic normality of U_1, because

$$(A.12) \qquad N^{-1/2}\sum_{i=1}^{N}\{r_{1N}(z_i)-E[r_{1N}(z_i)]\} = N^{-1/2}\sum_{i=1}^{N}\{r_1(z_i)-E[r_1(z_i)]\}$$

$$+ N^{-1/2}\sum_{i=1}^{N}\{t_{1N}(z_i)-E[t_{1N}(z_i)]\}\ ,$$

and the last term converges in probability to zero, since its variance is bounded by $E[t_{1N}(z)^2] = o(1)$. Consequently, combining (A.12), (A.10) and (A.9), we conclude that

$$(A.13) \qquad \sqrt{N}[\tilde{\delta}_1 - E(\tilde{\delta}_1)] = N^{-1/2}\left[\sum_{i=1}^{N}\{r_1(z_i) - E[r_1(z)]\}\right] + o_p(1)$$

$$\text{where } r_1(z) = g'(x) + g(x)[f'(x)/f(x)]$$

By analogous reasoning, we show the asymptotic normality of $\tilde{\delta}_2$, $\tilde{\delta}_3$ and $\tilde{\delta}_4$. Summarizing the conclusions, for $\tilde{\delta}_2$ we have that

$$(A.14) \qquad \sqrt{N}[\tilde{\delta}_2 - E(\tilde{\delta}_2)] = N^{-1/2}\left[\sum_{i=1}^{N}\{r_2(z_i) - E[r_2(z)]\}\right] + o_p(1)$$

$$\text{where } r_2(z) = [y + g(x)][f'(x)/f(x)]$$

For $\tilde{\delta}_3$ we have

$$(A.15) \qquad \sqrt{N}[\tilde{\delta}_3 - E(\tilde{\delta}_3)] = N^{-1/2}\left[\sum_{i=1}^{N}\{r_3(z_i) - E[r_3(z)]\}\right] + o_p(1)$$

where $r_3(z) = -g'(x) + g(x)[f'(x)/f(x)]$

and for $\tilde{\delta}_4$ we have

$$(A.16) \qquad \sqrt{N}[\tilde{\delta}_4 - E(\tilde{\delta}_4)] = N^{-1/2}\left[\sum_{i=1}^{N}\{r_4(z_i) - E[r_4(z)]\}\right] + o_p(1)$$

where $r_4(z) = -2g'(x) + 2g(x)[f'(x)/f(x)]$

Finally, from (A.8), (A.13), (A.14), (A.15) and (A.16), we conclude that $\tilde{\delta}$ of (A.4) has the representation

$$(A.17) \qquad \sqrt{N}[\tilde{\delta} - E(\tilde{\delta})] = N^{-1/2}\left[\sum_{i=1}^{N}\{r(z_i) - E[r(z)]\}\right] + o_p(1)$$

where $r(z_i)$ is given as

$$(A.18) \qquad r(z_i) = \sum_{\ell=0}^{4} r_\ell(z_i) = g'(x_i) - [y_i - g(x_i)]\frac{f'(x_i)}{f(x_i)} \quad .$$

where $r(z)=r(y,x)$ of (4.2). Application of the Lindberg-Levy Central Limit Theorem to (A.17) implies asymptotic normality of $\tilde{\delta}$.   QED 2.

Step 3: Bias: Using (A.3), write the bias of $\tilde{\delta}$ as

$$(A.19) \qquad E(\tilde{\delta}) - \delta = \tau_{0N} + \tau_{1N} - \tau_{2N} - \tau_{3N} + \tau_{4N}$$

where

$$\tau_{0N} = E[g'(x)I] - \delta \quad ; \quad \tau_{1N} = E\left[[\hat{G}'(x) - G'(x)]\frac{I}{f(x)}\right]$$

$$\tau_{2N} = E\left[[\hat{G}(x) - G(x)]\frac{f'(x)I}{f(x)^2}\right] \quad ; \quad \tau_{3N} = E\left[[\hat{f}'(x) - f'(x)]\frac{g(x)I}{f(x)}\right]$$

$$\tau_{4N} = E\left[[\hat{f}(x) - f(x)] \left[-\frac{g'(x)}{f(x)} + \frac{g(x)f'(x)}{f(x)^2} \right] I \right]$$

Let $A_N = \{x \mid f(x) > b\}$ and $B_N = \{x \mid f(x) \le b\}$. Then

$$\tau_{0N} = \int_{A_N} g'(x) \ f(x) dx - \int g'(x) \ f(x) dx = \int_{B_N} g'(x) \ f(x) dx = o(N^{-1/2})$$

by Assumption 5.

We show that $\tau_{1N} = o(N^{-1/2})$, with the proofs of $\tau_{2N} = o(N^{-1/2})$, $\tau_{3N} = o(N^{-1/2})$ and $\tau_{4N} = o(N^{-1/2})$ quite similar. Let ι denote an index set $(\ell_1, \ldots, \ell_k)$, where $\Sigma \ell_j = p$. For a k-vector $u = (u_1, \ldots, u_k)$, define $u^\iota = u_1^{\ell_1} u_2^{\ell_2} \ldots u_k^{\ell_k}$, and let $G_\iota^{(p)}$ denote the pth partial derivative of $G = gf$ with respect to the u components indicated by ι, namely $G_\iota^{(p)} = \partial^p G / (\partial u)^\iota$. By partial integration we have

$$\tau_{1N} = \int_{A_N} \int K(u) \ [G'(x-hu) - G'(x)] \ du \ dx$$

$$= \int_{A_N} \sum_\iota \int K(u) \ h^{p-1} G_\iota^{(p)}(\xi) u^\iota \ du \ dx$$

where the summation is over all index sets ι with $\Sigma \ell_j = p$, and where ξ lies on the line segment between x and $x - hu$. Thus

$$\tau_{1N} = h^{p-1} \int_{A_N} \sum_\iota G_\iota^{(p)}(x) \int K(u) u^\iota \ du \ dx$$

$$+ h^{p-1} \int_{A_N} \sum_\iota \int K(u) \ [G_\iota^{(p)}(\xi) - G_\iota^{(p)}(x)] \ u^\iota \ du \ dx = O(h^{p-1})$$

by Assumption 6. Therefore, by condition (iii), we have $\tau_{1N} = O[N^{-1/2}(N^{1/2} h^{p-1})] = o(N^{-1/2})$, as required. By analogous arguments, τ_{2N}, τ_{3N} and τ_{4N} are each shown to be $o(N^{-1/2})$. Consequently $E(\tilde{\delta}) - \delta = o(N^{-1/2})$.

QED 3.

Step 4: Trimming: Steps 1-3 have shown that $\sqrt{N}(\bar{\delta} - \delta) = \sqrt{N} \ R + o_p(1)$, where

24

$R = N^{-1} \sum [r(y_i, x_i) - E(r)]$. We now show the same property for $\hat{\delta}_g$.

Let $c_N = c_f (N^{1-(\varepsilon/2)} h^k)^{-1/2}$, where c_f is an upper bound consistent with (A.1a). Define two new trimming bounds as $b_u = b + c_N$ and $b_\ell = b - c_N$, and the associated trimmed kernel estimators:

$$\bar{\delta}_u = N^{-1} \sum_{i=1}^{N} \hat{g}'(x_i) \, I[f(x_i) > b_u]$$

$$\bar{\delta}_\ell = N^{-1} \sum_{i=1}^{N} \hat{g}'(x_i) \, I[f(x_i) > b_\ell]$$

Since $b^{-1} c_N \to 0$ by condition (ii), $\bar{\delta}_u$ and $\bar{\delta}_\ell$ each obey the tenets of steps 1-3, so $\sqrt{N}(\bar{\delta}_u - \delta)$ and $\sqrt{N}(\bar{\delta}_\ell - \delta)$ are each equivalent to $\sqrt{N} \, R$. Let $\eta > 0$, then by construction we have that

$$\text{Prob}\{\sqrt{N} |\bar{\delta}_u - \delta - R| < \eta, \ |\hat{f}(x_i) - f(x_i)| \le c_N, \ i=1,\ldots,N\}$$

$$\le \text{Prob}\{\sqrt{N} |\hat{\delta}_g - \delta - R| < \eta, \ |\hat{f}(x_i) - f(x_i)| \le c_N, \ i=1,\ldots,N\}$$

$$\le \text{Prob}\{\sqrt{N} |\bar{\delta}_\ell - \delta - R| < \eta, \ |\hat{f}(x_i) - f(x_i)| \le c_N, \ i=1,\ldots,N\}$$

By (A.1a), as $N \to \infty$, $\text{Prob}\{\sup |\hat{f}_{hi}(x_i) - f(x_i)| > c_N\} \to 0$. Weak consistency is then immediate: as $N \to \infty$, we have $\lim \text{Prob}\{\sqrt{N} |\hat{\delta}_g - \delta - R| \le \eta\} = 1$ for all $\eta > 0$. Strong consistency also follows by construction, because the above inequalities hold for all sample sizes greater than N.

<div align="right">QED 4.   QED Theorem 1.</div>

Corollary 1 follows as decribed in the text. For Corollaries 2 and 3, the following Lemma is used:

<u>Lemma A1</u>: Under Assumptions 1 through 11 and conditions i)-iii) of Theorem 1, we have

$$(A.20) \quad \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \hat{\ell}(x_i) \hat{I}_i = o_p(1)$$

$$(A.21) \quad \sqrt{N} \left[\frac{1}{N} \sum_{i=1}^{N} \hat{\ell}(x_i) \hat{I}_i x_i^T - Id \right] \equiv o_p(1)$$

Proof of Lemma A1: (A.20) and (A.21) follow directly from Theorem 1 and Corollary 1. For (A.20), apply (4.1) for $y_i=1$, noting that $r(1,x)=0$. For (A.21), let x_ℓ denote the ℓth component of x. Set $y=x_\ell$ and apply (4.1), noting that $r(x_\ell,x)=e_\ell$, the vector with 1 in the ℓth position and 0's elsewhere. Collect the answers for $\ell=1,....,k$.   QED Lemma A1.

Proof of Corollary 2: Note that

$$(A.22) \quad \sqrt{N} [\hat{\delta}_f - S_{\ell y}] = \left[\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \hat{\ell}(x_i) \hat{I}_i \right] \bar{y}$$

where $\bar{y} = \Sigma y_i/N$. Since $\bar{y}$ is bounded in probability (for instance by Chebyshev's inequality), the result follows from (A.20) of Lemma A1.

QED Corollary 2.

Proof of Corollary 3: By the delta method, $\sqrt{N}(\hat{\bar{d}}_f - \delta)$ can be shown to be a weighted sum of the departures $\sqrt{N}(S_{\ell y} - \delta)$ and $\sqrt{N}(S_{\ell x} - Id)$. But from (A.21), and Corollary 2 applied with y set to each component x_ℓ of x, we have that $\sqrt{N}(S_{\ell x} - Id) = o_p(1)$. Consequently, we have that

$$(A.23) \quad \sqrt{N}[\hat{\bar{d}}_f - \delta] = (Id)^{-1} \sqrt{N}(S_{\ell y} - \delta) + o_p(1)$$

so that Corollary 2 gives the result.   QED Corollary 3.

Corollaries 4 and 5 follow in the same fashion as Corollaries 2 and 3, where Lemma A2 plays the same role as Lemma A1.

Lemma A2: Under Assumptions 1 through 6, 8 and 10 and conditions i)-iii) of Theorem 1, we have

$$(A.24) \qquad \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \hat{\omega}(x_i) = o_p(1)$$

$$(A.25) \qquad \sqrt{N} \left[\frac{1}{N} \sum_{i=1}^{N} \hat{\omega}(x_i) \, x_i^T - Id \right] \equiv o_p(1)$$

Proof of Lemma A2: (A.24) and (A.25) follow directly from Theorem 1. For (A.24), apply (4.1) for $y_i = 1$, noting that $r(1,x)=0$. For (A.25), let x_ϱ denote the ϱth component of x. Set $y=x_\varrho$ and apply (4.1), noting that $r(x_\varrho,x)=e_\varrho$, the vector with 1 in the ϱth position and 0's elsewhere. Collect the answers for $\varrho=1,\ldots,k$. \hfill QED Lemma A2.

1. These terms are intended to suggest the estimation approach (as opposed to being a tight mathematical characterization), as it may be possible to write certain sample average estimators as "slope" estimators, etc.).

2. K(.) is assumed to be a kernel of order p=k+2 (see Assumption 3). involving positive and negative local weighting: see Powell, Stock and Stoker(1987) and HS among many others.

3. Nonparametric density estimators are surveyed by Prakasa-Rao(1983) and Silverman(1986).

4. Nonparametric regression and derivative estimators are surveyed by Prakasa-Rao(1983) and Härdle(1987).

5. The simulation results of HS on the indirect estimator $\hat{\delta}_f$ below indicate that while some trimming is useful. estimated values are not sensitive to trimming in the range of 1%-5% of the sample values (for k=4 predictor variables and N=100 observations).

6. We have defined "slope" coefficients by including a constant term in (3.4), but as shown later, this makes no difference to the asymptotic distribution of the estimators (c.f. Corollaries 2 and 4 below).

References

Hardle. W.(1987), <u>Applied Nonparametric Regression</u>, manuscript, Wirtschafttheorie II. Universitat Bonn.

Collomb. G. and W. Hardle(1986), "Strong Uniform Convergence Rates in Robust Nonparametric Time Series Analysis and Prediction: Kernel Regression Estimation from Dependent Observations." <u>Stochastic Processes and Their Applications</u>. 23. 77-89.

Hardle. W. and T.M. Stoker(1987), "Investigating Smooth Multiple Regression by the Method of Average Derivatives." Discussion Paper A-107. Sonderforschungsbereich 303. Universitat Bonn, revised April 1988.

Prakasa Rao. B.L.S.(1983), <u>Nonparametric Functional Estimation</u>. Academic Press. New York.

Powell. J.L., J.H. Stock and T.M. Stoker(1987), "Semiparametric Estimation of Index Coefficients." MIT Sloan School of Management Working Paper No. 1793-86. revised October 1987.

Silverman. B.W.(1978), "Weak and Strong Uniform Consistency of the Kernel Estimate of a Density Function and Its Derivatives." <u>Annals of Statistics</u>, 6. 177-184 (Addendum 1980. <u>Annals of Statistics</u>. 8, 1175-1176).

Silverman, B.W.(1986), <u>Density Estimation for Statistics and Data Analysis</u>. Chapman and Hall. London.

Stone. C.J.(1980), "Optimal Rates of Convergence for Nonparametric Estimators." <u>Annals of Statistics</u>. 8. 1348-1360.