Statistics & Probability Letters 22 (1995) 205–212

Consistent estimation of density-weighted average derivative by orthogonal series method

B.L.S. Prakasa Rao[1]

Indian statistical institute, 203 Barrackpore Trunk Road, Calcutta 700 035, India

Received August 1993; revised January 1994

Abstract

The problem of estimation of density-weighted average derivative is of interest in econometric problems, especially in the context of estimation of coefficients in index models. Here we propose a consistent estimator based on the orthogonal series method. Earlier work on this problem dealt with kernel method of estimation.

Keywords: Nonparametric estimation of density-weighted average derivative; Orthogonal series method; Consistency

1. Introduction

In a series of papers, Stoker (1986, 1989), Powell et al. (1989) and Hardle and Stoker (1989) proposed the problem of estimation of the density-weighted average derivative of a regression function.

Let (X_i, Y_i), $1 \leqslant i \leqslant n$ be i.i.d. bivariate random vectors distributed as (X, Y). Suppose $E(Y|X) = g(X)$ exists and X is distributed with density f. The density-weighted average derivative is defined as

$$\delta = E\left[f(X) \frac{\mathrm{d}g}{\mathrm{d}X} \right]$$

assuming that $g(\cdot)$ is differentiable.

Stoker (1986) and Powell et al. (1989) explain the motivation behind the estimation of density-weighted average derivative. For instance, weighted average derivatives are of practical interest as they are proportional to coefficients in index models. If the model indicates that $g(x) = \alpha + \beta x$, then

$$\frac{\mathrm{d}g}{\mathrm{d}x} = \beta$$

[1] Jawaharlal Nehru Centenary Chair, University of Hyderabad.

and $\delta = \beta E[f(X)]$. In general, if $g(x) = F(\alpha + \beta x)$, then

$$\frac{dg}{dx} = F'(\alpha + \beta x)\beta$$

and $\delta = E[F'(\alpha + \beta X) f(X)]\beta$.

Kernel method of estimation has been proposed and its properties are investigated in Powell et al. (1989). Here we propose an alternate method for estimation of δ by the method of orthogonal series. The method of orthogonal series for the estimation of density and the regression function has been extensively discussed in Prakasa Rao (1983).

Note that

$$\delta = E\left[f(X)\frac{dg}{dX}\right] = \int_{-\infty}^{\infty} f^2(x)\frac{dg}{dx}dx$$

$$= [g(x)f^2(x)]_{-\infty}^{\infty} - 2\int_{-\infty}^{\infty} f(x)\frac{df}{dx}g(x)\,dx$$

integrating by parts.

We assume that the density $f(x)$ and the regression function $g(x)$ satisfy the following conditions:

(A1) $\lim_{x \to \pm\infty} g(x)f^2(x) = 0$;

(A2) the density function f has an orthogonal series expansion

(i) $f(x) = \sum_{l=1}^{\infty} a_l e_l(x)$,

with respect to an orthonormal basis $\{e_l(x)\}$; the function $f(x)$ and the elements of the basis $\{e_l(x)\}$ are differentiable such that

$$\text{(ii)}\quad E\left|\sum_{l=1}^{q(N)} a_l e_l'(X) - f'(X)\right|^2 \to 0 \quad \text{as } N \to \infty$$

whenever $q(N) \to \infty$; and

(iii) $\sup_l |e_l(x)| < \infty$ and $\sup_l |e_l'(x)| < \infty$.

Assumption (A1) implies that

$$\delta \equiv E\left[f(X)\frac{dg}{dX}\right] = -2E\left[g(X)\frac{df}{dX}\right]$$

$$= -2E\left[Y\frac{df}{dX}\right], \tag{1.1}$$

since $g(X) = E[Y|X]$. Hereafter we write $f'(x)$ for df/dx and in general prime denotes differentiation.

2. Consistency of the estimator

Given a sample of independent and identically distributed observations (X_i, Y_i), $1 \leqslant i \leqslant n$, a natural estimator of δ is

$$\hat{\delta}_N = \frac{-2}{N} \sum_{i=1}^{N} Y_i \frac{\mathrm{d}\hat{f}_{Ni}}{\mathrm{d}X}\bigg|_{X=X_i} \tag{2.1}$$

from (1.1). Here $\hat{f}_{Ni}$ is an estimator of f based on the sample (X_j, Y_j), $1 \leqslant j \leqslant N$. It is convenient to choose $\hat{f}_{Ni}$ based on (X_j, Y_j), $1 \leqslant j \leqslant N$, $j \neq i$ and we will do the same in the sequel. An orthogonal series estimator of f is

$$\hat{f}_N(x) = \sum_{l=1}^{q(N)} \hat{a}_{lN}^{(i)} e_l(x)$$

where

$$\hat{a}_{lN}^{(i)} = \frac{1}{N-1} \sum_{\substack{j=1 \\ j \neq i}}^{N} e_l(X_j)$$

and $q(N) \to \infty$ as $N \to \infty$ to be chosen at a later stage. Then

$$\hat{\delta}_N = \frac{-2}{N} \sum_{i=1}^{N} Y_i \left[\sum_{l=1}^{q(N)} \hat{a}_{lN}^{(i)} e_l'(X_i) \right]. \tag{2.2}$$

Let $X_N^{(i)}$ denote the vector $(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_N)$. Hence,

$$\begin{aligned} \hat{\delta}_N &= -\frac{2}{N} \sum_{i=1}^{N} \sum_{l=1}^{q(N)} Y_i e_l'(X_i) \hat{a}_{lN}^{(i)} \\ &= -\frac{2}{N} \sum_{l=1}^{q(N)} \sum_{i=1}^{N} \psi_l(X_i, Y_i) \eta_l(X_N^{(i)}), \end{aligned} \tag{2.3}$$

where

$$\psi_l(X_i, Y_i) = Y_i e_l'(X_i) \tag{2.4}$$

and

$$\eta_l(X_N^{(i)}) = \hat{a}_{lN}^{(i)}. \tag{2.5}$$

Note that $\eta_l(X_N^{(i)})$ does not depend on the observation X_i by construction. Therefore,

$$\begin{aligned} E[\hat{\delta}_N] &= -\frac{2}{N} \sum_{l=1}^{q(N)} \sum_{i=1}^{N} E\{\psi_l(X_i, Y_i)\} E\{\eta_l(X_N^{(i)})\} \\ &= -2 \sum_{l=1}^{q(N)} E[\psi_l(X_1, Y_1)] E[e_l(X_1)] \\ &= -2 \sum_{l=1}^{q(N)} a_l E[Y e_l'(X)] \quad (\text{since } E[e_l(X_1)] = a_l) \\ &= -2E\left[Y \sum_{l=1}^{q(N)} a_l e_l'(X) \right] \end{aligned} \tag{2.6}$$

The sets $A_{1,0}, A_{2,0}, \ldots, A_{T,0}$ are determined beforehand once for all, and we store for each instant of time, the address of the buffer from which a data packet is to be sent and the link along which that packet is to be sent. Let $[w, x] \in A_{t,0}$. Then $[w + u, x + u] \in A_{t,u}$. Hence at time t, the node $(w + u)$ must send the packet originated from the node u, i.e., $P(u)$ which is stored in location $(n - w)$ of its buffer, to the node $(x + u)$. To implement this, we need to store the buffer address $(n - w)$ and the link type $\Delta(w + u, x + u)$ which is same as the link type $\Delta(w, x)$. Thus the information regarding the link $[w, x]$ in $A_{t,0}$ is sufficient to effect transmission of data packets from all the nodes in the network. If $A_{t,0} = \{[w_1, w], [x_1, x], [y_1, y], [z_1, z]\}$ we store the tth record consisting of four pairs $(b_1, l_1), (b_2, l_2), (b_3, l_3), (b_4, l_4)$ for any node $(w + u)$ as follows:

$(n - w_1)$	$\Delta(w_1, w)$	$(n - x_1)$	$\Delta(x_1, x)$	$(n - y_1)$	$\Delta(y_1, y)$	$(n - z_1)$	$\Delta(z_1, z)$
b_1	l_1	b_2	l_2	b_3	l_3	b_4	l_4

There will be $\lceil (n - 1)/4 \rceil$ such records. For $t = 0, 1, \ldots, \lceil (n - 1)/4 \rceil$, each node will fetch the tth record, and transmit the packet in location b_i along the link of type l_i.

3. SINGLE NODE SCATTER

In scattering, a node has to send $(n - 1)$ different packets to each of the other nodes in the network. Since a node can transmit at most four packets at a time, the minimum time required for single node scatter is $\lceil (n - 1)/4 \rceil$. Also, no scattering algorithm can be completed in time less than the diameter of the network. We have already shown that the diameter of $G(n; 1, s)$ is less than or equal to $\lceil (n - 1)/4 \rceil$. We will present now a time-optimal algorithm for single node scatter which requires $\lceil (n - 1)/4 \rceil$ units of time.

To describe our scattering algorithm, we assume that the node 0 is the source node. The packets will be transmitted from the node 0, along a spanning tree T rooted at node 0. T consists of four subtrees T_{+1}, T_{-1}, T_{+s}, and T_{-s} rooted at the nodes $+1, -1, +s$, and $-s$, respectively. Each of the four subtrees contains at most $\lceil (n - 1)/4 \rceil$ nodes.

With such a construction of the spanning tree, all the nodes will receive their packets within time $\lceil (n - 1)/4 \rceil$, if the following rule for transmission of packets is obeyed [3].

Node 0 sends packets to distinct nodes in the subtree (using only the links in T), giving priority to nodes farthest away from node 0 (breaking ties arbitrarily).

We also ensure that each packet travels along the shortest path to its destination by making T a shortest path tree.

3.1. Construction of the Spanning Tree

We find the sets S_k's for the graph $G(n; 1, s)$ as before. We maintain the property that if a node u of a generated pair $(u, n - u)$ is in T_{+1}, then the node $(n - u)$ will be in T_{-1} or if u is in T_{+s}, then $(n - u)$ will be in T_{-s}. We divide the total set of $(n - 1)$ nodes into two partitions of nearly equal size: partition I, consisting of the pairs which will be included in the trees T_{+1} and T_{-1}, and partition S, consisting of the pairs which will be included in the trees T_{+s} and T_{-s}.

Before going into the details of partitioning the nodes, we make the following observations on the matrix M.

Observation 1. In row k, the pair in column 1 is of the form $(k, -k)$. So we put all the pairs in column 1 in partition I.

Observation 2. All the pairs of the form $(k.s, -k.s)$ will be put in the partition S.

Observation 3. If a node u of a pair $(u, n - u)$ in S_k, is adjacent to some node u' in S_{k-1} then $(n - u)$ is adjacent to the node $(n - u')$ in S_{k-1}.

The method of grouping the nodes for partition I and partition S is almost identical for odd and even values of n. First, we describe the procedure for odd n.

3.1.1. *For odd n*

Since n is odd, there will be a total of $(n - 1)/2$ pairs in all the sets S_k's. We collect the pairs for partition I as follows. We leave out the pairs of the form $(k.s, -k.s)$. We take all the pairs in column 1. The maximum number of such pairs is $\lceil (n - 1)/4 \rceil$. If the number of pairs in column 1 is $\lceil (n - 1)/4 \rceil$ then we put all these pairs in partition I and the rest in partition S. Otherwise, from successive columns we select pairs starting at the bottom of that column and move upwards until we get $\lceil (n - 1)/4 \rceil$ pairs (see Example 3). Later, we will show that it is indeed possible to collect $\lceil (n - 1)/4 \rceil$ pairs in this way.

The pairs in partition I are connected in such a way that if one node of a pair is connected to T_{+1}, then the other node of that pair is connected to T_{-1}. Now we have the following lemmas.

LEMMA 1. *Suppose $(u, n - u)$ is a pair in partition I in some column c. Then the pair $(u, n - u)$ can always be connected to the subtrees T_{+1} and T_{-1}.*

$$= e_l(X_2)\,e_m(X_1) + e_m(X_1)(N-2)\,a_l$$
$$+ e_l(X_2)(N-2)\,a_m + (N-2)\,E[e_l(X_j)\,e_m(X_j)]$$
$$+ (N-2)(N-3)\,a_l a_m$$
$$\equiv I_2 \quad \text{(say)}. \tag{2.13}$$

Hence,

$$\begin{aligned}(N-1)^2 I_1 &= E[\psi_l(X_1, Y_1)\,\psi_m(X_2, Y_2)\,I_2]\\
&= E[\psi_l(X_1, Y_1)\,\psi_m(X_2, Y_2)\,e_l(X_2)\,e_m(X_1)]\\
&\quad + E[\psi_l(X_1, Y_1)\,\psi_m(X_2, Y_2)\,e_m(X_1)](N-2)\,a_l\\
&\quad + E[\psi_l(X_1, Y_1)\,\psi_m(X_2, Y_2)\,e_l(X_2)](N-2)\,a_m\\
&\quad + E[\psi_l(X_1, Y_1)\,\psi_m(X_2, Y_2)](N-2)\,E[e_l(X_j)\,e_m(X_j)]\\
&\quad + (N-2)(N-3)\,a_l a_m\,E[\psi_l(X_1, Y_1)\,\psi_m(X_2, Y_2)]\\
&= E[Y_1 e_l'(X_1)\,Y_2 e_m'(X_2)\,e_l(X_2)\,e_m(X_1)]\\
&\quad + (N-2)\,a_l\,E[Y_1 e_l'(X_1)\,Y_2 e_m'(X_2)\,e_m(X_1)]\\
&\quad + (N-2)\,a_m\,E[Y_1 e_l'(X_1)\,Y_2 e_m'(X_2)\,e_l(X_2)]\\
&\quad + (N-2)\,E[Y_1 e_l'(X_1)\,Y_2 e_m'(X_2)]\,E[e_l(X_1)\,e_m(X_1)]\\
&\quad + (N-2)(N-3)\,a_l a_m\,E[Y_1 e_l'(X_1)]\,E[Y_2 e_m'(X_2)].\end{aligned} \tag{2.14}$$

Let

$$b_{ml} = E[Y_1 e_l'(X_1)\,e_m(X_1)], \quad \gamma_{lm} = E[Y_1^2 e_l(X_1)\,e_m'(X_1)], \tag{2.15}$$

$$c_m = E[Y_1 e_m'(X_1)] \tag{2.16}$$

and

$$d_{lm} = E[e_l(X_1)\,e_m(X_1)]. \tag{2.17}$$

Then

$$\begin{aligned}(N-1)^2 \operatorname{cov}[\psi_l(X_i, Y_i)\,\eta_l(X_N^{(i)}), \psi_m(X_j, Y_j)\,\eta_m(X_N^{(j)})] &= b_{ml}b_{lm} + (N-2)\,a_l b_{ml} c_m\\
&\quad + (N-2)\,a_m b_{lm} c_l + (N-2)\,c_l c_m d_{lm}\\
&\quad + (N-2)(N-3)\,a_l a_m c_l c_m - a_l a_m c_l c_m.\end{aligned} \tag{2.18}$$

Case (ii): $i = j$. Then

$$\begin{aligned}
&\operatorname{cov}[\psi_l(X_1, Y_1)\eta_l(X_N^{(1)}), \psi_m(X_1, Y_1)\eta_m(X_N^{(1)})] \\
&\quad = E[\psi_l(X_1, Y_1)\psi_m(X_1, Y_1)\eta_l(X_N^{(1)})\eta_m(X_N^{(1)})] \\
&\qquad - E[\psi_l(X_1, Y_1)\eta_l(X_N^{(1)})]E[\psi_m(X_1, Y_1)\eta_m(X_N^{(1)})] \\
&\quad = E[Y_1 e_l'(X_1) Y_1 e_m'(X_1)\eta_l(X_N^{(1)})\eta_m(X_N^{(1)})] \\
&\qquad - a_l a_m c_l c_m \\
&\quad = E[Y_1^2 e_l'(X_1) e_m'(X_1)]E[\eta_l(X_N^{(1)})\eta_m(X_N^{(1)})] - a_l a_m c_l c_m \\
&\quad = \gamma_{lm} E[\eta_l(X_N^{(1)})\eta_m(X_N^{(1)})] - a_l c_l a_m c_m.
\end{aligned} \tag{2.19}$$

Let us now compute

$$\begin{aligned}
(N-1)^2 E[\eta_l(X_N^{(1)})\eta_m(X_N^{(1)})] &= E\left[\left\{\sum_{j=2}^{N} e_l(X_j)\right\}\left\{\sum_{k=2}^{N} e_m(X_k)\right\}\right] \\
&= \sum_{j=2}^{N}\sum_{k=2}^{N} E[e_l(X_j)e_m(X_k)] \\
&= (N-1)E[e_l(X_1)e_m(X_1)] + (N-1)(N-2)E[e_l(X_1)e_m(X_2)] \\
&= (N-1)d_{lm} + (N-1)(N-2)a_l a_m.
\end{aligned} \tag{2.20}$$

Hence,

$$\operatorname{cov}[\psi_l(X_1, Y_1)\eta_l(X_N^{(1)}), \psi_m(X_1, Y_1)\eta_m(X_N^{(1)})] = \gamma_{lm}\left\{\frac{d_{lm}}{N-1} + \frac{N-2}{N-1}a_l a_m\right\} - a_l c_l a_m c_m. \tag{2.21}$$

Calculations made above in the cases (i) and (ii) lead to the formula

$$\begin{aligned}
\operatorname{var}[\hat{\delta}_N] = {}& \frac{4}{N^2}\sum_{l=1}^{q(N)}\sum_{m=1}^{q(N)}\left[\gamma_{lm}\left\{\frac{d_{lm}}{N-1} + \frac{N-2}{N-1}a_l a_m\right\} - a_l c_l a_m c_m\right]N \\
& + \frac{4}{N^2}\sum_{l=1}^{q(N)}\sum_{m=1}^{q(N)}\left\{\begin{array}{c}
\dfrac{b_{ml}b_{lm}}{(N-1)^2} + \dfrac{N-2}{(N-1)^2}a_l b_{ml} c_m \\
+ \dfrac{N-2}{(N-1)^2}a_m b_{lm} c_l \\
+ \dfrac{N-2}{(N-1)^2}c_l c_m d_{lm} \\
+ \dfrac{(N-2)(N-3)}{(N-1)^2}a_l a_m c_l c_m \\
- a_l a_m c_l c_m
\end{array}\right\} N(N-1)
\end{aligned} \tag{2.22}$$

$$= \frac{4}{N(N-1)} \sum_{l=1}^{q(N)} \sum_{m=1}^{q(N)} \gamma_{lm} d_{lm} + \frac{4(N-2)}{N(N-1)} \sum_{l=1}^{q(N)} \sum_{m=1}^{q(N)} \gamma_{lm} a_l a_m$$

$$- \frac{4}{N} \left(\sum_{l=1}^{q(N)} a_l c_l \right)^2 + \frac{4N(N-1)}{N^2(N-1)^2} \sum_{l=1}^{q(N)} \sum_{m=1}^{q(N)} b_{ml} b_{lm}$$

$$+ \frac{4N(N-1)(N-2)}{N^2(N-1)^2} \sum_{l=1}^{q(N)} \sum_{m=1}^{q(N)} a_l b_{ml} c_m + \frac{4N(N-1)(N-2)}{N^2(N-1)^2} \sum_{l=1}^{q(N)} \sum_{m=1}^{q(N)} a_m b_{lm} c_{ml}$$

$$+ \frac{4N(N-1)(N-2)}{N^2(N-1)^2} \sum_{l=1}^{q(N)} \sum_{m=1}^{q(N)} c_l c_m d_{lm}$$

$$+ \frac{4N(N-1)(N-2)(N-3)}{N^2(N-1)^2} \sum_{l=1}^{q(N)} \sum_{m=1}^{q(N)} a_l a_m c_l c_m$$

$$- \frac{4N(N-1)}{N^2} \sum_{l=1}^{q(N)} \sum_{m=1}^{q(N)} a_l a_m c_l c_m. \tag{2.23}$$

Note that

$$\sup_{l,m} \gamma_{lm} < \infty, \qquad \sup_{l,m} b_{ml} < \infty, \qquad \sup_{l} a_l < \infty, \qquad \sup_{l} c_l < \infty \tag{2.24}$$

and

$$\sup_{l,m} d_{lm} < \infty \tag{2.25}$$

by assumption (A2)(iii). Observe that the coefficient of $(\sum_{l=1}^{q(N)} a_l c_l)^2$ in the expression for $\mathrm{var}(\hat{\delta}_N)$ is

$$- \frac{4}{N} + \frac{4(N-2)(N-3)}{N(N-1)} - \frac{4(N-1)}{N} = \frac{4(6-4N)}{N(N-1)}$$

$$\simeq \frac{-16}{N} + 0\left(\frac{1}{N}\right).$$

Under the assumption (A3), it follows that

$$\mathrm{var}(\hat{\delta}_N) \simeq O\left(\frac{q^2(N)}{N^2} + \frac{q^2(N)}{N} \right). \tag{2.26}$$

Theorem. *Under assumptions* (A1) *and* (A2), *if* $q(N) \to \infty$ *such that*

$$\frac{q^2(N)}{N} \to 0 \quad \text{as } N \to \infty \tag{2.27}$$

and $EY^2 < \infty$, *then*

$$\hat{\delta}_N \xrightarrow{P} \delta \quad \text{as } N \to \infty. \tag{2.28}$$

Proof. The result follows from the fact

$$\operatorname{var}(\hat{\delta}_N) \to 0 \quad \text{and} \quad E(\hat{\delta}_n) \to \delta \quad \text{as } n \to \infty.$$

3. Remarks

Let us now discuss the limiting behaviour of

$$\{\hat{\delta}_N - E(\hat{\delta}_N)\} \tag{3.1}$$

if any. Note that

$$\begin{aligned}
\{\hat{\delta}_N - E(\hat{\delta}_N)\} &= -\frac{2}{N}\sum_{i=1}^{N}\left[Y_i \frac{\partial \hat{f}_{N_i}}{\partial X}\bigg|_{X=x_i} - E\left(Y_i \frac{\partial \hat{f}_{N_i}}{\partial X}\bigg|_{X=x_i}\right)\right] \\
&= -\frac{2}{N}\sum_{l=1}^{q(N)}\sum_{i=1}^{N}\{\psi_l(X_i, Y_i)\eta_l(X_N^{(i)}) - E(\psi_l(X_i, Y_i)\eta_l(X_N^{(i)}))\} \\
&= -\frac{2}{N}\sum_{i=1}^{N}\left[\sum_{l=1}^{q(N)}\{\psi_l(X_i, Y_i)\eta_l(X_N^{(i)}) - E[\psi_l(X_i, Y_i)\eta_l(X_N^{(i)})]\}\right] \\
&= -\frac{2}{N}\sum_{i=1}^{N} Z_{Ni},
\end{aligned}$$

where

$$\begin{aligned}
Z_{Ni} = {} & [\psi_1(X_i, Y_i)\eta_1(X_N^{(i)}) + \cdots + \psi_{q(N)}(X_i, Y_i)\eta_{q(N)}(X_N^{(i)})] \\
& - E\{[\psi_1(X_i, Y_i)\eta_1(X_N^{(i)}) + \cdots + \psi_{q(N)}(X_i, Y_i)\eta_{q(N)}(X_N^{(i)})]\}.
\end{aligned}$$

Note that

$$\{Z_{Ni}, 1 \leqslant i \leqslant N\}$$

are finitely interchangeable for each N. Furthermore $E(Z_{Ni}) = 0$.

From the structure of $\{Z_{Ni}, 1 \leqslant i \leqslant N, N \geqslant 1\}$, it should be possible to study the asymptotic behaviour of the estimator $\hat{\delta}_N$. However, the limit theorems for exchangeable arrays presently available do not seem to be applicable in this context. The problem remains open.

References

Hardle, W. and T.M. Stoker (1989), Investigating smooth multiple regression by the method of average derivatives, *J. Amer. Statist. Assoc.* **84**, 986–995.

Powell, J.L., J.H. Stock, T.M. Stoker (1989), Semiparametric estimation of index coefficients, *Econometrica* **57**, 1403–1430.

Prakasa Rao, B.L.S. (1983), *Nonparametric Functional Estimation* (Academic Press, Orlando).

Stoker T.M. (1986), Consistent estimation of scaled coefficients, *Econometrica* **54**, 1461–1481.

Stoker, T.M. (1989), Tests of additive derivative constraints, *Rev. Econom. Stud.* **56**, 535–552.