L_2 Rate of Convergence for Interaction Spline Regression

By

Charles J. Stone

Technical Report No. 268
August 1990

This research was supported in part by NSF Grant DMS-8902016.

Department of Statistics
University of California
Berkeley, California

L_2 RATE OF CONVERGENCE FOR INTERACTION SPLINE REGRESSION[1]

BY CHARLES J. STONE

University of California, Berkeley

Let $X_1, \cdots, X_N, Y$ be random variables with $E(Y^2) < \infty$ and let $\mu(\cdot)$ denote the regression function of Y on $\mathbf{X} = (X_1, \cdots, X_N)$. Let $\mathscr{H}$ be the space of functions containing the constant and main-effect components and, possibly, some some interaction components and let d be the maximum number of variables involved in any such component. Let $\mu^*(\cdot)$ be the best approximation in $\mathscr{H}$ to $\mu(\cdot)$: $E[(\mu(\mathbf{X})-\mu^*(\mathbf{X}))^2] \le E[(\mu(\mathbf{X})-h(\mathbf{X}))^2]$ for $h \in \mathscr{H}$. Let p be a suitably defined lower bound to the smoothness of $\mu^*(\cdot)$. Consider a random sample of size n from the joint distribution of $\mathbf{X}$ and Y. Let $\mathscr{G}$ be a suitably defined finite-dimensional subspace of $\mathscr{H}$ consisting of splines. The dimension of $\mathscr{G}$ is allowed to tend to infinity along with n. Let $\hat{\mu}(\cdot)$ denote the least-squares estimate in $\mathscr{G}$ of $\mu^*(\cdot)$ based on the random sample. Under suitable conditions, the L_2 rate of convergence of $\hat{\mu}(\cdot)$ to $\mu^*(\cdot)$ is $n^{-p/(2p+d)}$ and optimal.

---

[1]This research was supported in part by National Science Foundation Grant DMS-8902016.

AMS 1980 *subject classifications*. Primary 62G20; secondary 62G05.

Key words and phrases. Functional inference, optimal rate of convergence, least-squares estimation, *B*-splines.

1. Introduction. Consider random variables $X_1, \cdots, X_N, Y$ and let $\mu(\cdot)$ denote the regression function of Y on $\mathbf{X} = (X_1, \cdots, X_N)$, so that $\mu(\mathbf{x}) = E(Y|\mathbf{X} = \mathbf{x})$. Let p denote the number of derivatives of the regression function (precise definitions will be given in Section 2). Then, under suitable conditions, the optimal L_2 rate of convergence on compacts for estimating the regression function based on a random sample of size n from the joint distribution of $\mathbf{X}$ and Y is $n^{-p/(2p+N)}$ (see Stone, 1982). One formulation of the "curse of dimensionality" is that, for fixed p, $p/(2p+N)$ is close to zero when N is large.

To get a faster rate of convergence, we could assume that $\mu(x_1, \cdots, x_N)$ is an additive function of $x_1, \cdots, x_N$ or, more realistically, replace the goal of estimating the regression function by that of estimating the best additive approximation $\mu^*(\cdot)$ to this function. The optimal rate of convergence is now $n^{-p/(2p+1)}$ (see Stone, 1985). In effect, additivity reduces the dimensionality of the estimation problem from N to 1.

More generally, we could allow some low-order interaction components into the model. It is natural to conjecture that, under suitable conditions, the optimal rate of convergence should now be $n^{-p/(2p+d)}$ with d being the maximum number of variables in any component. Suppose, for example, that $N = 3$ and let $\mu^*(\cdot)$ be the best approximation to the regression function of the form

$$\mu^*(x_1, x_2, x_3) = \mu_0^* + \mu_1^*(x_1) + \mu_2^*(x_2) + \mu_3^*(x_3) + \mu_{12}^*(x_1, x_2) + \mu_{13}^*(x_1, x_3),$$

which is a hierarchical model that includes the constant effect, all three main effects and two of the three two-factor interactions. The optimal rate of convergence should now be $n^{-p/(2p+2)}$. One main purpose of the present paper is to verify the stated conjecture.

In order to achieve the rate of convergence $n^{-p/(2p+d)}$, we will construct finite-dimensional spline spaces that can be used both to approximate $\mu^*(\cdot)$ and, in conjunction with the method of least squares, to estimate this function from the sample data. Thereby, we will accomplish the second main purpose of this paper, which is to lend further theoretical support to the use of finite-dimensional spline spaces in functional inference.

2. Statement of results. Consider random variables $X_1, \cdots, X_N, Y$, where $X_1, \cdots, X_N$ are [0, 1]-valued and Y has finite mean. Then $\mathbf{X} = (X_1, \cdots, X_N)$ ranges over $C = [0, 1]^N$. It is supposed that the following condition is satisfied.

CONDITION 1. $\mathbf{X}$ has a density f that is bounded away from zero and infinity on C.

Let M_1 and M_2 be positive numbers such that $M_1^{-1} \le f \le M_2$ on C. Then $M_1, M_2 \ge 1$.

Set

$$\langle h_1, h_2 \rangle = E[h_1(\mathbf{X})h_2(\mathbf{X})] = \int_C h_1(\mathbf{x})h_2(\mathbf{x})f(\mathbf{x})d\mathbf{x}$$

and

$$\|h\|^2 = \langle h, h \rangle = E[h^2(\mathbf{X})] = \int_C h^2(\mathbf{x})f(\mathbf{x})d\mathbf{x}$$

for square-integrable functions h_1, h_2, h on C. Two such functions are regarded as being equal if they differ only on a set of Lebesgue measure zero. Let $\mu(\cdot)$ denote the regression function of Y on $\mathbf{X}$, which is defined by $\mu(\mathbf{x}) = E(Y \mid \mathbf{X} = \mathbf{x})$ for $\mathbf{x} \in C$.

Given a subset v of $\{1, \cdots, N\}$, let $\mathcal{H}_v$ denote the collection of all square-integrable functions h on C that depend only on the coordinates x_l, $l \in v$, of $\mathbf{x} = (x_1, \cdots, x_n)$. We refer to $\#(v) - 1$ as the *interaction order* of $\mathcal{H}_v$. (If v is the empty set $\emptyset$ or, equivalently, if $\#(v) = 0$, then $\mathcal{H}_v$ is the space $\mathcal{C}$ of constant functions on C.

Let $\mathcal{N}$ be a collection of subsets of $\{1, \cdots, N\}$ and set

$$\mathcal{H} = \sum_v \mathcal{H}_v = \sum_{v \in \mathcal{N}} \mathcal{H}_v = \{\sum_v h_v : h_v \in \mathcal{H}_v \text{ for } v \in \mathcal{N}\}$$

and $d = \max_{v \in \mathcal{N}} \#(v)$. Then d–1 is the maximum interaction order of the components of $\mathcal{H}$. Observe that $d = 1$ if and only if every function in $\mathcal{H}$ is additive. It is assumed that $\mathcal{H}$ is *hierarchical*: if v is in $\mathcal{N}$ and η is a subset of v, then η is in $\mathcal{N}$. Set

$$\mathcal{H}_v^0 = \{h \in \mathcal{H}_v : h \perp \mathcal{H}_\eta \text{ for every proper subset } \eta \text{ of } v\}, \quad v \in \mathcal{N}.$$

(Here $h \perp \mathcal{H}_\eta$ means that $\langle h, k \rangle = 0$ for $k \in \mathcal{H}_\eta$.) Then (under Condition 1)

$$\mathcal{H} = \bigoplus_v \mathcal{H}_v^0;$$

each $h \in \mathcal{H}$ can be written uniquely in the form $h = \sum_v h_v$, where $h_v \in \mathcal{H}_v^0$ for $v \in \mathcal{N}$; clearly, $h_\emptyset = E[h(\mathbf{X})]$. (See Section 3 for the proof). We refer to $\mathcal{H}_v^0$, $v \in \mathcal{N}$, as the components of $\mathcal{H}$, to $\mathcal{H}_\emptyset = \mathcal{C}$ as the constant component, to $\mathcal{H}_v$ with $\#(v) = 1$ as a main-

effect component, and to $\mathcal{H}_\nu$ with $\#(\nu) \geq 2$ as an interaction component. There is a unique best approximation $\mu^*(\cdot)$ in $\mathcal{H}$ to $\mu(\cdot)$:

$$E[(\mu(\mathbf{X})-\mu^*(\mathbf{X}))^2] = \min_{h\in\mathcal{H}} E[(\mu(\mathbf{X})-h(\mathbf{X}))^2].$$

(This follows from Lemma 1 in Section 3 by a standard completeness argument in the context of Hilbert space.) We can write $\mu^*(\cdot) = \sum_\nu \mu^*_\nu(\cdot)$ for uniquely determined $\mu^*_\nu(\cdot) \in \mathcal{H}_\nu$, $\nu \in \mathcal{N}$; clearly $\mu^*_\emptyset(\cdot) = E\mu^*(\mathbf{X}) = E\mu(\mathbf{X}) = EY$. Observe that $\mu^*(\cdot) = \mu(\cdot)$ if and only if $\mu(\cdot) \in \mathcal{H}$.

Let $(\mathbf{X}_1, Y_1), \cdots, (\mathbf{X}_n, Y_n)$ be a random sample of size n from the joint distribution of $\mathbf{X}$ and Y and set $\bar{Y} = (Y_1+\cdots+Y_n)/n$. It follows from Condition 1 that $\mathbf{X}_1, \cdots, \mathbf{X}_n$ are distinct (with probability one). Set $C' = \{\mathbf{X}_1, \cdots, \mathbf{X}_n\}$. Consider the space of all real-valued functions whose domain includes C'. Let $\langle \cdot, \cdot \rangle_n$ denote the semi-inner product on this space defined by

$$\langle h_1, h_2 \rangle_n = \frac{1}{n} \sum_i h_1(\mathbf{X}_i) h_2(\mathbf{X}_i)$$

and let $\| \cdot \|_n$ denote the corresponding seminorm: $\|h\|_n^2 = \langle h, h \rangle_n$. Then $\|1\|_n^2 = 1$.

Let $K = K_n$ denote a positive integer and let I_k, $1 \leq k \leq K$, denote the subintervals of $[0, 1]$ defined by $I_k = [(k-1)/K, k/K)$ for $1 \leq k < K$ and $I_k = [1-1/K, 1]$ for $k = K$. Let m and q be fixed integers such that $m \geq 0$ and $m > q$. Let $\mathcal{S} = \mathcal{S}_n$ denote the collection of functions s on $[0, 1]$ such that

(i) the restriction of s to I_k is a polynomial of degree m (or less) for $1 \leq k \leq K$;

and, if $q \geq 0$,

(ii) s is q-times continuously differentiable on $[0, 1]$.

A function satisfying (i) is called a piecewise polynomial; if $m = 0$, it is piecewise constant. A function satisfying (i) and (ii) is called a spline. Typically, splines are considered with $q = m-1$ and then called linear, quadratic or cubic splines according as $m = 1, 2,$ or 3. Let B_j, $1 \leq j \leq J$, denote the usual basis of $\mathcal{S}$ consisting of B-splines (see de Boor, 1978). Then, in particular, $B_j \geq 0$ on $[0, 1]$ for $1 \leq j \leq J$ and $\sum B_j = 1$ on $[0, 1]$. Observe that $K \leq J \leq (m+1)K$.

Given a subset ν of $\{1, \cdots, N\}$, let $\mathcal{G}_\nu$ denote the corresponding *interaction spline*

space, defined as the span of all functions g on C of the form

$$g(\mathbf{x}) = \prod_{l \in \nu} g_l(x_l), \quad \text{where } \mathbf{x} = (x_1, \cdots, x_N) \text{ and } g_l \in \mathscr{G} \text{ for } l \in \nu.$$

Then $\mathscr{G}_\nu$ has dimension $J^{\#(\nu)}$, where $\#(\nu)$ is the number of integers in ν. (In particular, $\mathscr{G}_\emptyset = \mathscr{C}$ has dimension $J^0 = 1$.) Set $\mathscr{G} = \Sigma_\nu \mathscr{G}_\nu = \{\Sigma_\nu g_\nu : g_\nu \in \mathscr{G}_\nu \text{ for } \nu \in \mathscr{N}\}$. Also, set

$$\mathscr{G}_\nu^0 = \{g \in \mathscr{G}_\nu : g \perp_n \mathscr{G}_\eta \text{ for every proper subset } \eta \text{ of } \nu\}, \quad \nu \in \mathscr{N}.$$

(Here $g \perp_n \mathscr{G}_\eta$ means that $\langle g, h \rangle_n = 0$ for $h \in \mathscr{G}_\eta$.) Then $\mathscr{G} = \Sigma_\nu \mathscr{G}_\nu^0$.

The space $\mathscr{G}$ is said to be *singular* (with respect to C') if there is a nonzero function $g \in \mathscr{G}$ such that $g = 0$ on C'; otherwise, $\mathscr{G}$ is said to be *nonsingular*. Suppose $\mathscr{G}$ is nonsingular. Then $\langle \cdot, \cdot \rangle_n$ is an inner product on $\mathscr{G}$ and $\| \ \|$ is a norm on $\mathscr{G}$; that is, $\|g\|_{2n} > 0$ for every nonzero function $g \in \mathscr{G}$. Moreover (see Lemma 2 in Section 3), $\mathscr{G} = \bigoplus_\nu \mathscr{G}_\nu^0$: each $g \in \mathscr{G}$ can be written uniquely in the form $g = \Sigma_\nu g_\nu$, where $g_\nu \in \mathscr{G}_\nu^0$ for $\nu \in \mathscr{N}$; clearly, $g_\emptyset = n^{-1}\langle 1, g \rangle$.

Set $d_1 = \max\{\#(\eta \cup \nu) : \eta, \nu \in \mathscr{N}\}$. Then $d \le d_1 \le 2d$.

CONDITION 2. $J^{d_1} = o(n^{1-\delta})$ for some $\delta > 0$.

The next result follows from Lemmas 4 and 5 in Section 3.

THEOREM 1. *Suppose Conditions* 1 *and* 2 *hold. Then* $P(\mathscr{G}$ *is singular*$) = o(1)$.

Let $Y(\cdot)$ be defined by $Y(\mathbf{X}_i) = Y_i$ for $1 \le i \le n$. Let $\hat{\mu}(\cdot) = \Sigma_{\nu \in \mathscr{N}} \hat{\mu}_\nu(\cdot)$, where $\hat{\mu}_\nu(\cdot) \in \mathscr{G}_\nu^0$ for $\nu \in \mathscr{N}$, minimize $\|Y(\cdot) - g\|_n^2 = n^{-1}\Sigma_1^n [Y_i - g(\mathbf{X}_i)]^2$, $g \in \mathscr{G}$. Then $\hat{\mu}(\cdot)$ is the least-squares fit in $\mathscr{G}$ to the sample data and $\hat{\mu}_\emptyset(\cdot) = \bar{Y}$. We think of $\hat{\mu}(\cdot)$ as an estimate of $\mu^*(\cdot)$ and of $\hat{\mu}_\nu(\cdot)$ as an estimate of $\mu_\nu^*(\cdot)$ for $\nu \in \mathscr{N}$. If $\mathscr{G}$ is nonsingular, then $\hat{\mu}(\cdot)$ and $\hat{g}_\nu(\cdot)$, $\nu \in \mathscr{N}$, are uniquely determined.

CONDITION 3. The function $E(Y^2 \mid \mathbf{X} = \mathbf{x})$, $\mathbf{x} \in C$, is bounded.

Given the positive number b_n and the random variable Z_n for $n \ge n_0$, $Z_n = O_P(b_n)$ means that $\lim_{c \to \infty} \limsup_n P(|Z_n| > cb_n) = 0$.

THEOREM 2. *Suppose Conditions* 1–3 *hold. Then*

$$\sup_{\mathbf{x} \in C} \operatorname{var}(\hat{\mu}_\nu(\mathbf{x}) \mid \mathbf{X}_1, \cdots, \mathbf{X}_n) = O_P(J^d/n), \quad \nu \in \mathscr{N},$$

so

$$\sup_{\mathbf{x} \in C} \operatorname{var}(\hat{\mu}(\mathbf{x}) \mid \mathbf{X}_1, \cdots, \mathbf{X}_n) = O_P(J^d/n).$$

Let $0 < \beta \le 1$. A function h on C is said to satisfy a Hölder condition with exponent β if there is a positive number M such that $|h(\mathbf{x})-h(\mathbf{x}_0)| \le M|\mathbf{x}-\mathbf{x}_0|^{\beta}$ for $\mathbf{x}_0, \mathbf{x} \in C$; here $|\mathbf{x}|$ is the Euclidean norm $(x_1^2+\cdots+x_N^2)^{1/2}$ of $\mathbf{x} = (x_1, \cdots, x_N)$. Given an N-tuple $\alpha = (\alpha_1, \cdots, \alpha_N)$ of nonnegative integers, set $[\alpha] = \alpha_1+\cdots+\alpha_N$ and let D^{α} denote the differentiable operator defined by

$$D^{\alpha} = \frac{\partial^{[\alpha]}}{\partial x_1^{\alpha_1} \cdots \partial x_N^{\alpha_N}}.$$

Set $p = m + \beta$. When the following condition is satisfied, p can be thought of as a lower bound to the smoothness of g^*.

CONDITION 4. For $\nu \in \mathcal{N}$ and $[\alpha] = m$, the function $\mu_{\nu}^*(\cdot)$ on C is m-times continuously differentiable and $D^{\alpha}\mu_{\nu}^*(\cdot)$ satisfies a Hölder condition with exponent β.

THEOREM 3. *Suppose Conditions* 1–4 *hold. Then*

$$\|E(\hat{\mu}_{\nu}(\cdot)\,|\,\mathbf{X}_1, \cdots, \mathbf{X}_n) - \mu_{\nu}^*(\cdot)\| = O_P\left[J^{-p} + \sqrt{J^d/n}\right], \quad \nu \in \mathcal{N},$$

so

$$\|E(\hat{\mu}(\cdot)\,|\,\mathbf{X}_1, \cdots, \mathbf{X}_n) - \mu^*(\cdot)\| = O_P\left[J^{-p} + \sqrt{J^d/n}\right],$$

Theorems 2 and 3, which will be proven in Section 3, have the following consequence.

COROLLARY 1. *Suppose Conditions* 1–4 *hold. Then*

$$\|\hat{\mu}_{\nu}(\cdot) - \mu_{\nu}^*(\cdot)\| = O_P\left[J^{-p} + \sqrt{J^d/n}\right], \quad \nu \in \mathcal{N},$$

so

$$\|\hat{\mu}(\cdot) - \mu^*(\cdot)\| = O_P\left[J^{-p} + \sqrt{J^d/n}\right].$$

Given positive numbers a_n and b_n for $n \ge n_0$, $a_n \sim b_n$ means that a_n/b_n is bounded away from zero and infinity. Set $\gamma = 1/(2p+d)$ and $r = p/(2p+d)$. Observe that if Condition 2 holds with $J \sim n^{\gamma}$, then $p > (d_1-d)/2$. The next result follows from Corollary 1.

COROLLARY 2. *Suppose Conditions* 1–4 *hold and that* $J \sim n^{\gamma}$. *Then*

$$\|\hat{\mu}_{\nu}(\cdot)-\mu_{\nu}^*(\cdot)\| = O_P(n^{-r}), \quad \nu \in \mathcal{N},$$

so

$$\|\hat{\mu}(\cdot)-\mu^*(\cdot)\| = O_P(n^{-r}).$$

The L_2 rate of convergence in Corollary 2 does not depend on N. It is clear from Stone (1982) with $d = N$ that this rate is optimal. When $d = N$, it is possible to use the tensor-product extension of de Boor (1976) referred to in the proof of Lemma 9 below to obtain the pointwise and L_∞ rates of convergence of $\hat{\mu}(\cdot)$ to $\mu^*(\cdot)$ (see Koo, 1988).

Chen (1989) has obtained results along the lines of those of the present paper with penalized least-squares estimation with $\mathbf{X}_1, \cdots, \mathbf{X}_n$ replaced by deterministic design points $\mathbf{x}_1, \cdots, \mathbf{x}_n$. For mathematical convenience, however, he imposes the severe restriction on the design points that they form a (suitably regular) balanced complete factorial design. Under this restriction, his results for $d < N$ are new and imply those for $d < N$. He also assumes that $\mu(\cdot) \in \mathscr{H}$ and (essentially) requires that $p \geq dm$ for some positive integer m with $2m > N$, which is much more restrictive than the requirement $p > (d_1 - d)/2$ for Corollary 2. (In a private communication, Chen stated that the condition $2m > N$ in his paper can be replaced by the condition $4m > d$.)

When $d = 1$, the results in this section were obtained by Stone (1985). When $d = 1$ and $N = 1$, similar results were obtained by Agarwal and Studden (1980) in the context of suitably regular deterministic designs. The results for additive regression ($d = 1$) have been extended to robust additive regression by Mo (1990a, 1990b).

There is a growing literature on the methodological aspects of finite-dimensional additive and interactive spline modelling. See Stone and Koo (1985), Friedman and Silverman (1989), Breiman (1989) and, especially, Friedman (1991). For the corresponding methodology based on penalized least-squares estimation, see Barry (1983, 1986) and Wahba (1986). For other related additive methodologies, see Buja, Hastie and Tibshirani (1989) and Hastie and Tibshirani (1990).

3. Proofs. The arguments in this section were partly suggested by those in de Boor (1976) and Stone (1985).

LEMMA 1. *Suppose Condition* 1 *holds, set* $\delta_1 = 1 - \sqrt{1 - M_1^{-2}M_2^{-2}} \in (0, 1]$, and *let* $h_\nu \in \mathscr{H}_\nu^0$ *for* $\nu \in \mathscr{N}$. *Then*

$$E\left[\left[\sum_\nu h_\nu(\mathbf{X})\right]^2\right] \geq \delta_1^{\#(\mathscr{N})-1} \sum_\nu E[h_\nu^2(\mathbf{X})]. \tag{3.1}$$

PROOF. Recall that $M_1, M_2 \geq 1$. We will verify (3.1) by induction on $\#(\mathscr{N})$. Observe first that it is trivially true when $\#(\mathscr{N}) = 1$. Suppose $\#(\mathscr{N}) \geq 2$ and that (3.1) holds whenever $\mathscr{N}$ is replaced by $\mathscr{N}'$ with $\#(\mathscr{N}') < \#(\mathscr{N})$. Choose a "maximal" $\eta \in \mathscr{N}$ (that is, such that η is not a proper subset of any set ν in $\mathscr{N}$). We first verify that

$$E\left[\left[\sum_\nu h_\nu(\mathbf{X})\right]^2\right] \geq M_1^{-2}M_2^{-2}E[(h_\eta(\mathbf{X}))^2]. \tag{3.2}$$

If $\#(\eta) = N$, then (3.2) follows immediately from the definition of $\mathscr{H}_\eta^0$. Suppose, instead, that $1 \leq \#(\eta) \leq N-1$. We can write $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$, where $\mathbf{X}_1$ consists of X_l, $l \notin \eta$, in some order and $\mathbf{X}_2$ consists of X_l, $l \in \eta$, in some order. Then $\mathbf{X}_1$ is C_1-valued and $\mathbf{X}_2$ is C_2-valued, where $C_1 = [0, 1]^{N-\#(\eta)}$ and $C_2 = [0, 1]^{\#(\eta)}$. Let $f_{\mathbf{X}_1}$ denote the density of $\mathbf{X}_1$, $f_{\mathbf{X}_2}$ the density of $\mathbf{X}_2$ and $f_{\mathbf{X}_2|\mathbf{X}_1}$ the conditional density of $\mathbf{X}_2$ given $\mathbf{X}_1$. Then $f_{\mathbf{X}_1}$ and $f_{\mathbf{X}_2}$ are bounded above by M_2, so

$$f_{\mathbf{X}_2|\mathbf{X}_1}(\mathbf{x}_2|\mathbf{x}_1) \geq M_1^{-1}M_2^{-2}f_{\mathbf{X}_2}(\mathbf{x}_2), \quad \mathbf{x}_1 \in C_1 \text{ and } \mathbf{x}_2 \in C_2.$$

We can write $h_\eta(\mathbf{x})$ as $h_\eta(\mathbf{x}_2)$ for $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$. Since $f_{\mathbf{X}_1}$ is bounded below by M_1^{-1}, we conclude from the definition of $\mathscr{H}_\eta^0$ that

$$\begin{aligned}
E\left[\left[\sum_\nu h_\nu(\mathbf{X})\right]^2\right] &= \int_{C_1} f_{\mathbf{X}_1}(\mathbf{x}_1)d\mathbf{x}_1 \int_{C_2} \left[h_\eta(\mathbf{x}_2) + \sum_{\nu \neq \eta} h_\nu(\mathbf{x}_1, \mathbf{x}_2)\right]^2 f_{\mathbf{X}_2|\mathbf{X}_1}(\mathbf{x}_2|\mathbf{x}_1)d\mathbf{x}_2 \\
&\geq M_1^{-2}M_2^{-2} \inf_{\mathbf{x}_1 \in C_1} \int_{C_2} \left[h_\eta(\mathbf{x}_2) + \sum_{\nu \neq \eta} h_\nu(\mathbf{x}_1, \mathbf{x}_2)\right]^2 f_{\mathbf{X}_2}(\mathbf{x}_2)d\mathbf{x}_2 \\
&= M_1^{-2}M_2^{-2} \inf_{\mathbf{x}_1 \in C_1} E\left[\left[h_\eta(\mathbf{X}_2) + \sum_{\nu \neq \eta} h_\nu(\mathbf{x}_1, \mathbf{X}_2)\right]^2\right] \\
&\geq M_1^{-2}M_2^{-2}E[h_\eta^2(\mathbf{X})]
\end{aligned}$$

and hence that (3.2) again holds.

It follows from (3.2) that

$$E\left[\left[h_\eta(\mathbf{X}) - \beta \sum_{\nu \neq \eta} h_\nu(\mathbf{X})\right]^2\right] \geq M_1^{-2} M_2^{-2} E[h_\eta^2(\mathbf{X})], \quad \beta \in \mathbb{R}. \tag{3.3}$$

We conclude from (3.3) that

$$\left[E\left[h_\eta(\mathbf{X}) \sum_{\nu \neq \eta} h_\nu(\mathbf{X})\right]\right]^2 \leq (1 - M_1^{-2} M_2^{-2}) E[h_\eta^2(\mathbf{X})] E\left[\left[\sum_{\nu \neq \eta} h_\nu(\mathbf{X})\right]^2\right]$$

and hence, by the induction hypothesis, that

$$\begin{aligned}
&E\left[\left[\sum_\nu h_\nu(\mathbf{X})\right]^2\right] \\
&\quad \geq \left[1 - \sqrt{1 - M_1^{-2} M_2^{-2}}\right]\left\{E[h_\eta^2(\mathbf{X})) + E\left[\left[\sum_{\nu \neq \eta} h_\nu(\mathbf{X})\right]^2\right]\right\} \\
&\quad \geq \left[1 - \sqrt{1 - M_1^{-2} M_2^{-2}}\right]\left\{E[h_\eta^2(\mathbf{X})) + \left[1 - \sqrt{1 - M_1^{-2} M_2^{-2}}\right]^{\#(\mathcal{N})-2} \sum_{\nu \neq \eta} E[h_\nu^2(\mathbf{X})]\right\} \\
&\quad \geq \left[1 - \sqrt{1 - M_1^{-2} M_2^{-2}}\right]^{\#(\mathcal{N})-1} \sum_\nu E[h_\nu^2(\mathbf{X})].
\end{aligned}$$

Therefore (3.1) holds for $\mathcal{N}$. □

LEMMA 2. *Suppose $\mathcal{G}$ is nonsingular, $g_\nu \in \mathcal{G}_\nu^0$ for $\nu \in \mathcal{N}$ and $\Sigma_\nu g_\nu = 0$. Then $g_\nu = 0$ for $\nu \in \mathcal{N}$.*

PROOF. It suffices to show that if ν is maximal, then $g_\nu = 0$. To this end, as an application of Lemma 1, we can write $g_\nu = \Sigma^{(\nu)} \tilde{g}_\eta$, where $\tilde{g}_\eta \in \mathcal{G}_\eta$ for η a proper subset of ν and $\Sigma^{(\nu)}$ denotes summation over proper subsets of ν. Then

$$\|g_\nu\|_n^2 = \langle g_\nu, \Sigma^{(\nu)} \tilde{g}_\eta \rangle_n = 0$$

and hence $g_\nu = 0$. □

Write $\mathbf{X}_i = (X_{i1}, \cdots, X_{iN})$ for $1 \leq i \leq n$.

LEMMA 3. *Suppose Condition 1 holds and let $t > 0$. Then, except on an event having probability at most $2(m+1)^N \exp(-2nt^2)$, the inequalities*

$$\left|\frac{1}{n} \Sigma_i \Pi_l p_{1l}(X_{il}) p_{2l}(X_{il}) - E[\Pi_l p_{1l}(X_l) p_{2l}(X_l)]\right| \leq t c_m^{2N} M_1 \sqrt{E[\Pi_l p_{1l}^2(X_{1l})] E[\Pi_l p_{1l}^2(X_{1l})]}$$

hold simultaneously for all polynomials $p_{11}, \cdots, p_{1N}, p_{21}, \cdots, p_{2N}$ of degree m. Here c_m is a positive number that depends only on m.

PROOF. By an elementary compactness argument, there is a positive number c_m such that if p is a polynomial of degree m, then

$$\left[\sum_0^m \frac{|p^{(k)}(0)|}{k!}\right]^2 \le c_m^2 \int_0^1 p^2(x)dx. \tag{3.4}$$

It follows from Hoeffding's inequality (Theorem 1 of Hoeffding, 1963) that, except on an event having probability at most $2(m+1)^{2N}\exp(-2nt^2)$, the inequalities

$$\left|\frac{1}{n}\Sigma_i X_{i1}^{k_{11}} \cdots X_{iN}^{k_{1N}} X_{i1}^{k_{21}} \cdots X_{iN}^{k_{2N}} - E[X_1^{k_{11}} \cdots X_N^{k_{1N}} X_1^{k_{21}} \cdots X_N^{k_{2N}}]\right| < t \tag{3.5}$$

hold simultaneously for all choices in $\{0, \cdots, m\}$ of $k_{11}, \cdots, k_{1N}, k_{21}, \cdots, k_{2N}$. It follows from (3.4) and (3.5) that

$$|\frac{1}{n}\Sigma_i \Pi_l p_{1l}(X_{il}) p_{2l}(X_{il}) - E[\Pi_l p_{1l}(X_l) p_{2l}(X_l)]| \le t c_m^{2N} \sqrt{\Pi_l \int_0^1 p_{1l}^2(x)dx \Pi_l \int_0^1 p_{2l}^2(x)dx}.$$

Since

$$E\left[\prod_l p_{1l}^2(X_l)\right] = \int_C \prod_l p_{1l}^2(x_l) f(\mathbf{x})d\mathbf{x} \ge M_1^{-1} \int_C \prod_l p_{1l}^2(x_l)d\mathbf{x} \ge M_1^{-1} \prod_l \int_0^1 p_{1l}^2(x)dx$$

and, similarly,

$$E\left[\prod_l p_{1l}^2(X_l)\right] = M_1^{-1} \prod_l \int_0^1 p_{2l}^2(x)dx,$$

the desired result holds. □

LEMMA 4. *Suppose Conditions 1 and 2 hold and let $\varepsilon > 0$. Then, except on an event whose probability tends to zero with n,*

$$|\langle g_1, g_2 \rangle_n - E[g_1(\mathbf{X}) g_2(\mathbf{X})]| \le \varepsilon \sqrt{E[g_1^2(\mathbf{X})] E[g_2^2(\mathbf{X})]}, \tag{3.6}$$

$$g_1, g_2 \in \mathcal{G}_{\eta \cup \nu} \text{ for some } \eta, \nu \in \mathcal{N}.$$

PROOF. It suffices to verify the desired result when $q = -1$ and $d = N$. Then $d_1 = N$, $\mathcal{G}$ is the span of all functions g on C of the form

$$g(\mathbf{x}) = \prod_{l=1}^N g_l(\mathbf{x}_l), \quad \text{where } \mathbf{x} = (x_1, \cdots, x_N) \text{ and } g_l \in \mathcal{S} \text{ for } 1 \le l \le N,$$

and (3.6) simplifies to

$$|\langle g_1, g_2 \rangle_n - E[g_1(\mathbf{X}) g_2(\mathbf{X})]| \le \varepsilon \sqrt{E[g_1^2(\mathbf{X})] E[g_2^2(\mathbf{X})]}, \quad g_1, g_2 \in \mathcal{G}.$$

Given $k_1, \cdots, k_N \in \{1, \cdots, K\}$, set $\mathbf{k} = (k_1, \cdots, k_N)$ and

$$I_{\mathbf{k}} = \{\mathbf{x} = (x_1, \cdots, x_N) : x_l \in I_{k_l} \text{ for } 1 \le l \le N\}.$$

Let $\text{ind}(\cdot \in I_{\mathbf{k}})$ denote the indicator function of $I_{\mathbf{k}}$, which is defined by $\text{ind}(\mathbf{x} \in I_{\mathbf{k}}) = 1$ for $\mathbf{x} \in I_{\mathbf{k}}$ and $\text{ind}(\mathbf{x} \in I_{\mathbf{k}}) = 0$ for $\mathbf{x} \notin I_{\mathbf{k}}$. Given polynomials $p_{k_1}, \cdots, p_{k_N}$ of degree m, define the multivariate polynomial $p_{\mathbf{k}}$ by

$$p_{\mathbf{k}}(\mathbf{x}) = p_{k_1}(x_1) \cdots p_{k_N}(x_N), \quad \mathbf{x} = (x_1, \cdots, x_N).$$

Then every function g in $\mathcal{G}$ can be written in the form

$$g(\mathbf{x}) = \sum_{\mathbf{k}} p_{\mathbf{k}}(\mathbf{x}) \text{ind}(\mathbf{x} \in I_{\mathbf{k}}), \quad \mathbf{x} \in C.$$

Thus we can write

$$g_1(\mathbf{x}) = \sum_{\mathbf{k}} p_{1\mathbf{k}}(\mathbf{x}) \text{ind}(\mathbf{x} \in I_{\mathbf{k}}) \quad \text{and} \quad g_2(\mathbf{x}) = \sum_{\mathbf{k}} p_{2\mathbf{k}}(\mathbf{x}) \text{ind}(\mathbf{x} \in I_{\mathbf{k}}), \quad \mathbf{x} \in C.$$

Observe that

$$g_1(\mathbf{x}) g_2(\mathbf{x}) = \sum_{\mathbf{k}} p_{1\mathbf{k}}(\mathbf{x}) p_{2\mathbf{k}}(\mathbf{x}) \text{ind}(\mathbf{x} \in I_{\mathbf{k}}), \quad \mathbf{x} \in C.$$

Hence

$$E[g_1(\mathbf{X}) g_2(\mathbf{X})] = \sum_{\mathbf{k}} P(\mathbf{X} \in I_{\mathbf{k}}) E(p_{1\mathbf{k}}(\mathbf{X}) p_{2\mathbf{k}}(\mathbf{X}) \mid \mathbf{X} \in I_{\mathbf{k}}),$$

$$E[g_1^2(\mathbf{X})] = \sum_{\mathbf{k}} P(\mathbf{X} \in I_{\mathbf{k}}) E(p_{1\mathbf{k}}^2(\mathbf{X}) \mid \mathbf{X} \in I_{\mathbf{k}}),$$

and

$$E[g_2^2(\mathbf{X})] = \sum_{\mathbf{k}} P(\mathbf{X} \in I_{\mathbf{k}}) E(p_{2\mathbf{k}}^2(\mathbf{X}) \mid \mathbf{X} \in I_{\mathbf{k}}).$$

Set $\mathcal{J}_{\mathbf{k}} = \{i: 1 \le i \le n \text{ and } \mathbf{X}_i \in I_{\mathbf{k}}\}$. Then

$$E_n[g_1(\mathbf{X}) g_2(\mathbf{X})] = \sum_{\mathbf{k}} P_n(\mathbf{X} \in I_{\mathbf{k}}) E_n(p_{1\mathbf{k}}(\mathbf{X}) p_{2\mathbf{k}}(\mathbf{X}) \mid \mathbf{X} \in I_{\mathbf{k}}),$$

where

$$E_n[g_1(\mathbf{X}) g_2(\mathbf{X})] = \langle g_1, g_2 \rangle_n = \frac{1}{n} \sum_i g_1(\mathbf{X}_i) g_2(\mathbf{X}_i),$$

$$P_n(\mathbf{X} \in I_{\mathbf{k}}) = \frac{1}{n} \#(\mathcal{J}_{\mathbf{k}}),$$

and

$$E_n(p_{1\mathbf{k}}(\mathbf{X}) p_{2\mathbf{k}}(\mathbf{X}) \mid \mathbf{X} \in I_{\mathbf{k}}) = \frac{1}{\#(\mathcal{J}_{\mathbf{k}})} \sum_{i \in \mathcal{J}_{\mathbf{k}}} p_{1\mathbf{k}}(\mathbf{X}_i) p_{2\mathbf{k}}(\mathbf{X}_i).$$

Choose $\varepsilon_1 \in (0, 1)$ such that $\varepsilon_1^2 + 2\varepsilon_1 \le \varepsilon$. It follows from Conditions 1 and 2 and Berstein's inequality (see Theorem 3 of Hoeffding, 1963) that, except on an event whose

probability tends to zero with n, $|P_n(\mathbf{X}\in I_{\mathbf{k}}) - P(\mathbf{X}\in I_{\mathbf{k}})| \le \varepsilon_1 P(\mathbf{X}\in I_{\mathbf{k}})$ for all $\mathbf{k}$ and hence

$$\frac{(1-\varepsilon_1)}{M_1 K^N} \le P_n(\mathbf{X}\in I_{\mathbf{k}}) \le \frac{(1+\varepsilon_1)M_2}{K^N} \quad \text{for all } \mathbf{k}.$$

By Condition 2, $K^N = o(n^{1-\delta})$ for some $\delta > 0$. Thus there are positive numbers M_3 and δ such that, except on an event whose probability tends to zero with n, $\#(\mathcal{I}_{\mathbf{k}}) \ge M_3^{-1} n^{\delta}$ for all $\mathbf{k}$. We conclude from Lemma 3 that, except on an event whose probability tends to zero with n,

$$|E_n(p_{1\mathbf{k}}(\mathbf{X})p_{2\mathbf{k}}(\mathbf{X}) \mid \mathbf{X}\in I_{\mathbf{k}}) - E(p_{1\mathbf{k}}(\mathbf{X})p_{2\mathbf{k}}(\mathbf{X}) \mid \mathbf{X}\in I_{\mathbf{k}})|$$
$$\le \varepsilon_1 \sqrt{E(p_{1\mathbf{k}}^2(\mathbf{X}) \mid \mathbf{X}\in I_{\mathbf{k}}) E(p_{2\mathbf{k}}^2(\mathbf{X}) \mid \mathbf{X}\in I_{\mathbf{k}})}$$

for all $\mathbf{k}$ and all choices of $p_{1\mathbf{k}}$ and $p_{2\mathbf{k}}$. Consequently, except on an event whose probability tends to zero with n,

$$\begin{aligned}
|\langle g_1, g_2\rangle_n - E[g_1(\mathbf{X})g_2(\mathbf{X})]| &\le \varepsilon_1 \sum_{\mathbf{k}} P(\mathbf{X}\in I_{\mathbf{k}}) |E(g_1(\mathbf{X})g_2(\mathbf{X}) \mid \mathbf{X}\in I_{\mathbf{k}})| \\
&\quad + \varepsilon_1(1+\varepsilon_1) \sum_{\mathbf{k}} P(\mathbf{X}\in I_{\mathbf{k}}) \sqrt{E(g_1^2(\mathbf{X}) \mid \mathbf{X}\in I_{\mathbf{k}}) E(g_2^2(\mathbf{X}) \mid \mathbf{X}\in I_{\mathbf{k}})} \\
&\le (\varepsilon_1^2 + 2\varepsilon_1) \sum_{\mathbf{k}} P(\mathbf{X}\in I_{\mathbf{k}}) \sqrt{E(g_1^2(\mathbf{X}) \mid \mathbf{X}\in I_{\mathbf{k}}) E(g_2^2(\mathbf{X}) \mid \mathbf{X}\in I_{\mathbf{k}})} \\
&= (\varepsilon_1^2 + 2\varepsilon_1) \sum_{\mathbf{k}} \sqrt{E[g_1^2(\mathbf{X})\mathrm{ind}(\mathbf{X}\in I_{\mathbf{k}})] E[g_2^2(\mathbf{X})\mathrm{ind}(\mathbf{X}\in I_{\mathbf{k}})]} \\
&\le \varepsilon \sqrt{E[g_1^2(\mathbf{X})] E[g_2^2(\mathbf{X})]}, \quad g_1, g_2 \in \mathcal{G}. \ \square
\end{aligned}$$

LEMMA 5. *Suppose Conditions* 1 *and* 2 *hold and let* $0 < \delta_2 < \delta_1$. *Then, except on an event whose probability tends to zero with* n,

$$\|\Sigma_\nu g_\nu\|_n^2 \ge \delta_2^{\#(\mathcal{N})-1} \Sigma_\nu \|g_\nu\|_n^2, \quad g_\nu \in \mathcal{G}_\nu^0 \text{ for } \nu \in \mathcal{N}.$$

PROOF. Applying Lemma 4 with $0 < \#(\mathcal{N})\varepsilon < \delta_1^{\#(\mathcal{N})-1}$, we see that, except on an event whose probability tends to zero with n, (3.6) holds and hence

$$\|g_\nu\|_n^2 \le (1+\varepsilon)E[g_\nu^2(\mathbf{X})], \quad \nu \in \mathcal{N} \text{ and } g_\nu \in \mathcal{G}_\nu^0,$$

so

$$\sum_\nu \|g_\nu\|_n^2 \le (1+\varepsilon) \sum_\nu E[g_\nu^2(\mathbf{X})].$$

Therefore, by Lemmas 1 and 4, except on an event whose probability tends to zero with n,

$$\begin{aligned}\|\Sigma_\nu g_\nu\|_n^2 &\geq E[(\Sigma_\nu g_\nu(\mathbf{X}))^2] - \#(\mathcal{N})\varepsilon\Sigma_\nu E[g_\nu^2(\mathbf{X})] \\ &\geq (\delta_1^{\#(\mathcal{N})-1} - \#(\mathcal{N})\varepsilon)\,\Sigma_\nu E[g_\nu^2(\mathbf{X})] \\ &\geq \frac{\delta_1^{\#(\mathcal{N})-1} - \#(\mathcal{N})\varepsilon}{1+\varepsilon} - \Sigma_\nu \|g_\nu\|_n^2.\end{aligned}$$

Since ε can be made arbitrarily small, the desired result holds. □

Set $\mathcal{J}_\emptyset = \{0\}$ and $B_{\emptyset 0} = 1$. For $\nu \in \mathcal{N}$ with $\nu \neq \emptyset$, let $\mathcal{J}_\nu$ denote the collection of ordered $\#(\nu)$-tuples j_l, $l \in \nu$, with $j_l \in \{1, \cdots, J\}$ for $l \in \nu$. Then $\#(\mathcal{J}_\nu) = J^{\#(\nu)}$. For $\mathbf{j} \in \mathcal{J}_\nu$, let $B_{\nu\mathbf{j}}$ denote the function on C given by

$$B_{\nu\mathbf{j}}(\mathbf{x}) = \prod_{l\in\nu} B_{j_l}(x_l), \quad \mathbf{x} = (x_1, \cdots, x_N).$$

Then, for $\nu \in \mathcal{N}$, the functions $B_{\nu\mathbf{j}}$, $\mathbf{j} \in \mathcal{J}_\nu$, which are nonnegative and have sum one, form a basis of $\mathcal{G}_\nu$.

Suppose $\mathcal{G}$ is nonsingular and let $g \in \mathcal{G}$. Then $g = \Sigma_\nu g_\nu$, where $g_\nu \in \mathcal{G}_\nu^0$, $\nu \in \mathcal{N}$, are uniquely determined. Moreover, $g_\nu = \Sigma_{\mathbf{j}} b_{\nu\mathbf{j}} B_{\nu\mathbf{j}}$ for $\nu \in \mathcal{N}$, where the $b_{\nu\mathbf{j}}$'s are uniquely determined. Let ν and $\mathbf{j}$ be fixed. Let $G_{\nu\mathbf{j}} \in \mathcal{G}$ denote the representor of the linear functional $g \mapsto b_{\nu\mathbf{j}}$ on $\mathcal{G}$ relative to the inner product $\langle\ ,\ \rangle_n$, so that $b_{\nu\mathbf{j}} = \langle G_{\nu\mathbf{j}}, g\rangle_n$. Now $G_{\nu\mathbf{j}} = \Sigma_{\nu'} G_{\nu\mathbf{j}\nu'}$, where $G_{\nu\mathbf{j}\nu'} \in \mathcal{G}_\nu^0$ for $\nu \in \mathcal{N}$. Thus $G_{\nu\mathbf{j}\nu'} = \Sigma_{\mathbf{j}'} \gamma_{\nu\mathbf{j}\nu'\mathbf{j}'} B_{\nu'\mathbf{j}'}$ for $\nu' \in \mathcal{N}$, where the $\gamma_{\nu\mathbf{j}\nu'\mathbf{j}'}$'s are uniquely determined. Observe that

$$\langle G_{\nu\mathbf{j}}, G_{\nu'\mathbf{j}'}\rangle_n = \gamma_{\nu\mathbf{j}\nu'\mathbf{j}'}, \quad \nu, \nu' \in \mathcal{N}, \mathbf{j} \in \mathcal{J}_\nu \text{ and } \mathbf{j}' \in \mathcal{J}_{\nu'}. \tag{3.7}$$

Observe also that, for $\nu \in \mathcal{N}$ and $j \in \mathcal{J}_\nu$, $\langle G_{\nu\mathbf{j}}, B_{\nu\mathbf{j}}\rangle_n = 1$ and hence $0 < \|G_{\nu\mathbf{j}}\|_n^2 = \gamma_{\nu\mathbf{j}\nu\mathbf{j}}$.

LEMMA 6. *Suppose Conditions 1 and 2 hold. Then there is a positive number M_3, which does not depend on J, such that, except on an event whose probability tends to zero with n,*

$$\|\Sigma_\nu\Sigma_{\mathbf{j}} b_{\nu\mathbf{j}} B_{\nu\mathbf{j}}\|_n^2 \geq M_3^{-1} J^{-d}\Sigma_\nu\Sigma_{\mathbf{j}} b_{\nu\mathbf{j}}^2 \quad \text{if } \Sigma_{\mathbf{j}} b_{\nu\mathbf{j}} B_{\nu\mathbf{j}} \in \mathcal{G}_\nu^0 \text{ for } \nu \in \mathcal{N}. \tag{3.8}$$

PROOF. It follows from the basic properties of B-splines and repeated use of (viii) on page 155 of de Boor (1978) that, for some positive number M_4,

$$\int_C [\Sigma_{\mathbf{j}} b_{\nu\mathbf{j}} B_{\nu\mathbf{j}}(\mathbf{x})]^2 d\mathbf{x} \geq 2M_4^{-1} J^{-\#(\nu)}\Sigma_{\mathbf{j}} b_{\nu\mathbf{j}}^2$$

for all choices of $\nu \in \mathcal{N}$ and $b_{\nu j} \in \mathbb{R}$ and $j \in \mathcal{J}_\nu$. Thus, by Condition 1 and Lemma 4, except on an event whose probability tends to zero with n,

$$\|\Sigma_j b_{\nu j} B_{\nu j}\|_n^2 \geq M_4^{-1} J^{-\#(\nu)} \Sigma_j b_{\nu j}^2$$

for all such choices. The desired result now follows from Lemma 5. □

LEMMA 7. *Suppose Conditions 1 and 2 hold. Then, except on an event whose probability tends to zero with n,*

$$\sum_{\nu'} \sum_{j'} \gamma_{\nu j \nu' j'}^2 \leq M_3^2 J^{2d}, \quad \nu \in \mathcal{N} \text{ and } j \in \mathcal{J}_\nu \tag{3.9}$$

PROOF. Suppose that (3.8) holds and let $\nu \in \mathcal{N}$ and $j \in \mathcal{J}_\nu$. Then

$$M_3^{-1} J^{-d} \gamma_{\nu j \nu j}^2 \leq M_3^{-1} J^{-d} \sum_{\nu'} \sum_{j'} \gamma_{\nu j \nu' j'}^2 \leq \|G_{\nu j}\|_n^2 = \gamma_{\nu j \nu j},$$

so $\gamma_{\nu j \nu j} \leq M_3 J^d$ and therefore (3.9) is valid. We now obtain the desired result from Lemma 6. □

LEMMA 8. *Suppose Conditions* 1–3 *hold. Then, except on an event whose probability tends to zero with n,*

$$\max_\nu \max_{j \in \mathcal{J}_\nu} \operatorname{var}(\hat{\beta}_{\nu j} | \mathbf{X}_1, \cdots, \mathbf{X}_n) = O_P(J^d/n).$$

PROOF. Set $\sigma^2(\mathbf{x}) = \operatorname{var}(Y | \mathbf{X} = \mathbf{x})$, $\mathbf{x} \in C$. It follows from Condition 3 that $\sigma^2(\cdot)$ has a finite upper bound M_4 on C.

Suppose that $\mathcal{G}$ is nonsingular. Let Q denote orthogonal projection onto $\mathcal{G}$ relative to the inner product $\langle\ ,\ \rangle_n$. Then $\langle g, Qh \rangle_n = \langle g, h \rangle_n$ for all real-valued functions h whose domain includes C' and all $g \in \mathcal{G}$. Given such a function h, write Qh in the form

$$Qh = \sum_\nu \sum_j b_{\nu j} B_{\nu j}, \quad \text{where } \sum_j b_{\nu j} B_{\nu j} \in \mathcal{G}_\nu^0 \text{ for } \nu \in \mathcal{N}.$$

Then $b_{\nu j} = \langle G_{\nu j}, Qh \rangle_n = \langle G_{\nu j}, h \rangle_n$ and hence

$$b_{\nu j} = \sum_{\nu'} \sum_{j'} \gamma_{\nu j \nu' j'} \langle B_{\nu j}, h \rangle_n, \quad \nu \in \mathcal{N} \text{ and } j \in \mathcal{J}_\nu.$$

The least-squares estimate $\hat{\mu}(\cdot)$ can be written as

$$\hat{\mu}(\cdot) = QY(\cdot) = \sum_\nu \sum_j \hat{\beta}_{\nu j} B_{\nu j}, \quad \text{where } \sum_j \hat{\beta}_{\nu j} B_{\nu j} \in \mathcal{G}_\nu^0 \text{ for } \nu \in \mathcal{N}.$$

Thus

$$\hat{\beta}_{\nu j} = \langle G_{\nu j}, Y(\cdot) \rangle_n = \sum_{\nu'} \sum_{j'} \gamma_{\nu j \nu' j'} \langle B_{\nu j}, Y(\cdot) \rangle_n, \quad \nu \in \mathcal{N} \text{ and } j \in \mathcal{J}_\nu.$$

Consequently,

$$\begin{aligned}\operatorname{var}(\hat{\beta}_{\nu j}|\mathbf{X}_1,\cdots,\mathbf{X}_n) &= \frac{1}{n^2}\sum_i \sigma^2(\mathbf{X}_i)\Big[\sum_{\nu'}\sum_{\mathbf{j}'}\gamma_{\nu\mathbf{j}\nu'\mathbf{j}'}B_{\nu'\mathbf{j}'}(\mathbf{X}_i)\Big]^2 \\ &\le M_4 n^{-1}\|G_{\nu\mathbf{j}}\|_n^2 \\ &= M_4 n^{-1}\gamma_{\nu\mathbf{j}\nu\mathbf{j}}.\end{aligned}$$

The desired result now follows from Theorem 1 and Lemma 7. □

Theorem 2 follows from Lemma 8.

LEMMA 9. *Suppose Conditions* 1–3 *hold and that* $\mu^*(\cdot) = 0$. *Then*

$$\|E(\hat{\mu}_\nu(\cdot)|\mathbf{X}_1,\cdots,\mathbf{X}_n)\|_n = O_P\Big[\sqrt{J^d/n}\Big], \quad \nu \in \mathcal{N}.$$

PROOF. Choose $\nu \in \mathcal{N}$ and recall that $B_{\nu\mathbf{j}}$, $\mathbf{j} \in \mathcal{J}_\nu$, form a basis of $\mathcal{G}_\nu$. Let $g \in \mathcal{G}_\nu$. Then $g = \sum_{\mathbf{j}} b_{\mathbf{j}}^{(\nu)} B_{\nu\mathbf{j}}$, where the $b_{\mathbf{j}}^{(\nu)}$'s are uniquely determined. Suppose $\mathcal{G}$ is nonsingular. Let $G_{\mathbf{j}}^{(\nu)}$ denote the representor of the linear functional $g \mapsto b_{\mathbf{j}}^{(\nu)}$ on $\mathcal{G}_\nu$ relative to the inner product $\langle\ ,\ \rangle_n$, so that $b_{\mathbf{j}}^{(\nu)} = \langle G_{\mathbf{j}}^{(\nu)}, g\rangle$. Then $G_{\mathbf{j}}^{(\nu)} = \sum_{\mathbf{j}'} \gamma_{\mathbf{j}\mathbf{j}'}^{(\nu)} B_{\nu\mathbf{j}'}$, where the $\gamma_{\mathbf{j}\mathbf{j}'}^{(\nu)}$'s are uniquely determined. (Alternatively, $(\gamma_{\mathbf{j}\mathbf{j}'}^{(\nu)})$ is the inverse of the Gram matrix $(\langle B_{\nu\mathbf{j}}, B_{\nu\mathbf{j}'}\rangle)$.) Let $\mu^{(\nu)}(\cdot)$ denote the orthogonal projection of $\mu(\cdot)$ onto $\mathcal{G}_\nu$ (relative to the inner product $\langle\ ,\ \rangle_n$). Then $\mu^{(\nu)}(\cdot) = \sum_{\mathbf{j}} \beta_{\mathbf{j}}^{(\nu)} B_{\nu\mathbf{j}}$, where

$$\beta_{\mathbf{j}}^{(\nu)} = \sum_{\mathbf{j}'} \gamma_{\mathbf{j}\mathbf{j}'}^{(\nu)} \langle B_{\nu\mathbf{j}'}, \mu(\cdot)\rangle_n, \quad \mathbf{j} \in \mathcal{J}_\nu.$$

Now

$$\|\mu^{(\nu)}(\cdot)\|_n^2 = \Big\|\sum_{\mathbf{j}} \beta_{\mathbf{j}}^{(\nu)} B_{\nu\mathbf{j}}\Big\|_n^2 = \sum_{\mathbf{j}}\sum_{\mathbf{j}'} \beta_{\mathbf{j}}^{(\nu)}\beta_{\mathbf{j}'}^{(\nu)} \langle B_{\nu\mathbf{j}}, B_{\nu\mathbf{j}'}\rangle_n.$$

By Conditions 1 and 2, Bernstein's inequality applied to the binomial distribution, and the basic properties of B-splines,

$$\|\mu^{(\nu)}(\cdot)\|_n^2 = O_P(J^{-\#(\nu)}) \sum_{\mathbf{j}} (\beta_{\mathbf{j}}^{(\nu)})^2.$$

It follows from Conditions 1 and 2 by an extension of arguments in de Boor (1976) and Stone (1989) that there are numbers $M_4 \in (0, \infty)$ and $c \in (0, 1)$ (both independent of J) such that, except on an event whose probability tends to zero with n,

$$|\gamma_{\mathbf{j}\mathbf{j}'}^{(\nu)}| \le M_4 J^{\#(\nu)} c^{|\mathbf{j}'-\mathbf{j}|}, \quad \mathbf{j},\mathbf{j}' \in \mathcal{J}_\nu$$

Consequently,

$$\sum_{\mathbf{j}} (\beta_{\mathbf{j}}^{(\nu)})^2 = O_P\Big[J^{2\#(\nu)} \sum_{\mathbf{j}} \Big[\sum_{\mathbf{j}'} c^{|\mathbf{j}'-\mathbf{j}|}\, |\langle B_{\nu\mathbf{j}'}, \mu(\cdot)\rangle_n|\Big]^2\Big] = O_P(J^{2\#(\nu)} \sum_{\mathbf{j}} (\langle B_{\nu\mathbf{j}}, \mu(\cdot)\rangle_n)^2).$$

Since $\mu^*(\cdot) = 0$, we see that $E(\langle B_{\nu\mathbf{j}}, \mu(\cdot)\rangle_n) = E(B_{\nu\mathbf{j}}(\mathbf{X})\mu(\mathbf{X})) = 0$ for $\mathbf{j} \in \mathcal{J}_\nu$. Moreover,

$$\max_{\mathbf{j}} \operatorname{var}(\langle B_{\nu\mathbf{j}}, \mu(\cdot)\rangle_n) = n^{-1} \max_{\mathbf{j}} \operatorname{var}(B_{\nu\mathbf{j}}(\mathbf{X})\mu(\mathbf{X})) = O(n^{-1}J^{-\#(\nu)}).$$

Thus $E[\Sigma_{\mathbf{j}} (\langle B_{\nu\mathbf{j}}, \mu(\cdot)\rangle_n)^2] = O(1/n)$ and hence $\Sigma_{\mathbf{j}} (\langle B_{\nu\mathbf{j}}, \mu(\cdot)\rangle_n)^2 = O_P(1/n)$. Consequently, $\Sigma_{\mathbf{j}} (\beta^{(\nu)})^2 = O_P(J^{2\#(\nu)}/n)$ and therefore

$$\|\mu^{(\nu)}(\cdot)\|_n^2 = O_P(J^{\#(\nu)}/n) = O_P(J^d/n), \quad \nu \in \mathcal{N}.$$

Let $\mu_\nu^0(\cdot)$ denote the orthogonal projection of $\mu(\cdot)$ onto $\mathcal{G}_\nu^0$, which equals the orthogonal projection of $\mu^{(\nu)}(\cdot)$ onto $\mathcal{G}_\nu^0$. Then $\|\mu_\nu^0(\cdot)\|_n^2 \le \|\mu^{(\nu)}(\cdot)\|_n^2$ and hence

$$\|\mu_\nu^0(\cdot)\|_n^2 = O_P(J^d/n), \quad \nu \in \mathcal{N}.$$

Observe that $E(\hat{\mu}(\cdot)\,|\,\mathbf{X}_1, \cdots, \mathbf{X}_n)$ is the orthogonal projection of $\mu(\cdot)$ onto $\mathcal{G}$. We can write this orthogonal projection as $\Sigma_\nu \mu_\nu(\cdot)$, where $\mu_\nu(\cdot) \in \mathcal{G}_\nu^0$ for $\nu \in \mathcal{N}$. Now $\mu_\nu^0(\cdot)$ is the orthogonal projection of $\Sigma_\nu \mu_\nu(\cdot)$ onto $\mathcal{G}_\nu^0$ for $\nu \in \mathcal{N}$, so

$$\begin{aligned}
\|\Sigma_\nu \mu_\nu(\cdot)\|_n^2 &= \Sigma_\nu \langle \mu_\nu(\cdot), \Sigma_\nu \mu_\nu(\cdot)\rangle_n \\
&= \Sigma_\nu \langle \mu_\nu(\cdot), \mu_\nu^0(\cdot)\rangle_n \\
&\le \Sigma_\nu \|\mu_\nu(\cdot)\|_n \|\mu_\nu^0(\cdot)\|_n \\
&\le (\max_\nu \|\mu_\nu(\cdot)\|_n) \Sigma_\nu \|\mu_\nu^0(\cdot)\|_n.
\end{aligned}$$

We conclude from Lemma 5 that

$$\|E(\hat{\mu}(\cdot)\,|\,\mathbf{X}_1, \cdots, \mathbf{X}_n)\|_n^2 = \|\Sigma_\nu \mu_\nu(\cdot)\|_n^2 = O_P(\Sigma_\nu \|\mu_\nu^0(\cdot)\|_n^2) = O_P(J^d/n).$$

The desired result now follows by another application of Lemma 5. □

LEMMA 10. *Suppose Conditions* 1–4 *hold and that* $\mu^*(\cdot) = \mu(\cdot)$. *Then*

$$\|E(\hat{\mu}_\nu(\cdot)\,|\,\mathbf{X}_1, \cdots, \mathbf{X}_n) - \mu_\nu^*(\cdot)\|_n^2 = O_P(J^{-2p} + J^{d-1}/n), \quad \nu \in \mathcal{N}.$$

PROOF. It follows from Condition 4 (see Theorem 12.8 of Schumaker, 1981) that there is a positive number M_4 not depending on n or J such that, for $\nu \in \mathcal{N}$, there is a function $g_\nu \in \mathcal{G}_\nu$ with $\|g_\nu - \mu_\nu^*(\cdot)\|_\infty \le M_4 J^{-p}$; here $\|h\|_\infty = \sup_{\mathbf{x}\in C} |h(\mathbf{x})|$ is the L_∞ norm of a function h on C. Choose $\nu \in \mathcal{N}$ and let η be a proper subset of ν. Then $E[B_{\eta\mathbf{j}}(\mathbf{X})\mu_\nu^*(\mathbf{X})] = 0$ for $\mathbf{j} \in \mathcal{J}_\eta$ and hence

$$\max_{\mathbf{j}} |E[B_{\eta\mathbf{j}}(\mathbf{X})g_\nu(\mathbf{X})]| = O(J^{-\#(\eta)-p}).$$

Moreover,

$$\max_{\mathbf{j}} \operatorname{var}(B_{\eta\mathbf{j}}(\mathbf{X})g_\nu(\mathbf{X})) = O(J^{-\#(\eta)}).$$

Suppose $\mathcal{G}$ is nonsingular. Let $g^0_{\nu\eta}$ denote the orthogonal projection of g_ν onto $\mathcal{G}^0_\eta$ (relative to the inner product $\langle\ ,\ \rangle_n$). Arguing as in the proof of Lemma 9, we get that

$$\|g^0_{\nu\eta}\|^2_n = O_P(J^{-2p} + J^{\#(\eta)}/n) = O_P(J^{-2p} + J^{d-1}/n).$$

Write $g_\nu = \Sigma_{\eta\subset\nu} g_{\nu\eta}$, where $g_{\nu\eta} \in \mathcal{G}^0_\eta$ for $\eta \subset \nu$. Let $\Sigma^{(\nu)}$ denote summation over all proper subsets of ν. Then $g^0_{\nu\eta}$ is the orthogonal projection of $\Sigma^{(\nu)}g_{\nu\eta}$ onto $\mathcal{G}^0_\eta$. We conclude, again by arguing as in the proof of Lemma 9, that

$$\|g_\nu - g_{\nu\nu}\|^2_n = \|\Sigma^{(\nu)}g_{\nu\eta}\|^2_n = O_P(J^{-p} + J^{d-1}/n).$$

Replacing g_ν by $g_{\nu\nu}$ if necessary, we see that, for $\nu \in \mathcal{N}$, there is a function $g_\nu \in \mathcal{G}^0_\nu$ such that $\|g_\nu - \mu^*_\nu(\cdot)\|^2_n = O_P(J^{-2p} + J^{d-1}/n)$ and hence $\|\Sigma_\nu g_\nu - \mu^*(\cdot)\|^2_n = O_P(J^{-2p} + J^{d-1}/n)$.

Write the orthogonal projection $E(\hat{\mu}(\cdot)\,|\,\mathbf{X}_1, \cdots, \mathbf{X}_n)$ of $\mu(\cdot) = \mu^*(\cdot)$ onto $\mathcal{G}$ as $\Sigma_\nu \mu_\nu(\cdot)$, where $\mu_\nu(\cdot) = E(\hat{\mu}_\nu(\cdot)\,|\,\mathbf{X}_1, \cdots, \mathbf{X}_n) \in \mathcal{G}^0_\nu$ for $\nu \in \mathcal{N}$. Observe that

$$\|\Sigma_\nu \mu_\nu(\cdot) - \mu^*(\cdot)\|^2_n \le \|\Sigma_\nu g_\nu - \mu^*(\cdot)\|^2_n.$$

Thus

$$\|\Sigma_\nu \mu_\nu(\cdot) - \mu^*(\cdot)\|^2_n = O_P(J^{-2p} + J^{d-1}/n)$$

and hence

$$\|\Sigma_\nu \mu_\nu(\cdot) - \Sigma_\nu g_\nu\|^2_n = O_P(J^{-2p} + J^{d-1}/n).$$

We conclude from Lemma 5 that

$$\|\mu_\nu(\cdot) - g_\nu\|^2_n = O_P(J^{-2p} + J^{d-1}/n) \quad \nu \in \mathcal{N},$$

and therefore that

$$\|\mu_\nu(\cdot) - \mu^*_\nu(\cdot)\|^2_n = O_P(J^{-2p} + J^{d-1}/n) \quad \nu \in \mathcal{N}. \square$$

LEMMA 11. *Suppose Conditions* 1–4 *hold. There is a positive number* M_4 *not depending on n or J such that, except on an event whose probability tends to zero with n,*

$$\|g - \mu^*_\nu(\cdot)\|^2 \le M_4(\|g - \mu^*_\nu(\cdot)\|^2_n + J^{-2p}), \quad \nu \in \mathcal{N} \text{ and } g \in \mathcal{G}_\nu.$$

PROOF. Given $\nu \in \mathcal{N}$, set $h = \mu^*_\nu(\cdot)$ and let $g \in \mathcal{G}_\nu$. Then (see the proof of Lemma 4) g can be written in the form

$$g(\mathbf{x}) = \sum_{\mathbf{k}} p_{\mathbf{k}}(\mathbf{x})\operatorname{ind}(\mathbf{x} \in I_{\mathbf{k}}), \quad \mathbf{x} \in C.$$

By Condition 4, there is a function g_1 of the same form such that $\|g_1-h\|_\infty \le M_5 J^{-p}$, M_5 being a positive number that does not depend on n of J. Then $\|g_1-h\| \le M_5 J^{-p}$ and $\|g_1-h\|_n \le M_5 J^{-p}$, so

$$\|g-h\|^2 \le 2\|g-g_1\|^2 + 2M_5^2 J^{-2p}) \quad \text{and} \quad \|g-g_1\|_n^2 \le 2\|g-h\|_n^2 + 2M_5^2 J^{-2p}).$$

It follows from Lemma 4 that, except on an event whose probability tends to zero with n, $\|g-g_1\|^2 \le 2\|g-g_1\|_n^2$ and hence

$$\|g-h\|^2 \le 4\|g-g_1\|_n^2 + 2M_5 J^{-2p} \le 8\|g-h\|_n^2 + 10M_5 J^{-2p}. \square$$

PROOF OF THEOREM 3. It follows from Lemma 9 applied to the regression function $\mu(\cdot)-\mu^*(\cdot)$ and Lemma 10 applied to the regression function $\mu^*(\cdot)$ that

$$\|E(\hat{\mu}_\nu(\cdot)\,|\,\mathbf{X}_1,\cdots,\mathbf{X}_n) - \mu_\nu^*(\cdot)\|_n^2 = O_P(J^{-2p} + J^d/n), \quad \nu \in \mathcal{N}.$$

We conclude from Lemma 11 that

$$\|E(\hat{\mu}_\nu(\cdot)\,|\,\mathbf{X}_1,\cdots,\mathbf{X}_n) - \mu_\nu^*(\cdot)\|^2 = O_P(J^{-2p} + J^d/n), \quad \nu \in \mathcal{N}. \square$$

REFERENCES

AGARWAL, G. G. and STUDDEN, W. J. (1980). Asymptotic integrated mean square error using least squares and bias minimizing spline. *Ann. Statist.* **8** 1307–1325.

BARRY, D. (1983). *Nonparametric Bayesian Regression.* Ph. D. Dissertation, Department of Statistics, Yale University.

BARRY, D. (1986). Nonparametric Bayesian regression. *Ann. Statist.* **14** 934–953.

BREIMAN, L. (1989). Fitting additive models to regression data. Technical Report No. 209, Department of Statistics, University of California at Berkeley.

BUJA, A., HASTIE, T. and TIBSHIRANI, R. (1989). Linear smoothers and additive models (with discussion). *Ann. Statist.* **17** 453–555.

DE BOOR, C. (1976). A bound on the L_∞-norm of L_2-approximation by splines in terms of a global mesh ratio. *Math. Comp.* **30** 765–771.

DE BOOR, C. (1978). *A Practical Guide to Splines.* Springer–Verlag, New York.

CHEN, Z. (1989). Interaction spline models and their convergence rates. Manuscript.

FRIEDMAN, J. H. (1991). Multivariate Adaptive Regression Splines. *Ann. Statist.*, to appear.

FRIEDMAN, J. H. and SILVERMAN, B. W. (1989). Flexible parsimonious smoothing and additive modeling (with discussion). *Technometrics* 31 3–39.

HASTIE, T. J. and TIBSHIRANI, R. J. (1990). *Generalized Additive Models.* Chapman and Hall, New York.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

KOO, J.-Y. (1988) *Tensor product splines in the estimation of regression, esponential response functions and multivariate densities.* Ph. D. dissertation, Department of Statistics, University of California at Berkeley.

SCHUMAKER, L. L. (1981). *Spline Functions: Basic Theory.* Wiley, New York.

STONE, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360.

STONE, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* 10 1040–1053.

STONE, C. J. (1985). Additive regression and other nonparametric models. *Ann. Statist.* 13 689–705.

STONE, C. J. (1986). The dimensionality reduction principle for generalized additive models. *Ann. Statist.* 14 590–606.

STONE, C. J. and KOO, C.-Y. (1986) Additive splines in statistics. In *1985 Statistical Computing Section Proceedings of the American Statistical Association* 45–48. American Statistical Association, Washington.

STONE, C. J. (1989). Uniform error bounds involving logspline models. *In Probability, Statistics and Mathematics*: *Papers in Honor of Samuel Karlin* (T. W. Anderson, K. B. Athreya, and D. L. Iglehart, eds.) 335–355. Academic Press, Boston.

WAHBA, G. (1986). Partial and interaction splines for the semiparametric estimation of functions of several variables. In *Computer Science and Statistics*: *Proceedings of the 18th Symposium on the Interface*. (Boardman, T. J., ed.) 75–80. Amer. Statist. Assoc, Washington, D. C.

REFERENCES

AGARWAL, G. G. and STUDDEN, W. J. (1980). Asymptotic integrated mean square error using least squares and bias minimizing spline. *Ann. Statist.* **8** 1307–1325.

BARRY, D. (1983). *Nonparametric Bayesian Regression.* Ph. D. Dissertation, Department of Statistics, Yale University.

BARRY, D. (1986). Nonparametric Bayesian regression. *Ann. Statist.* **14** 934–953.

BREIMAN, L. (1989). Fitting additive models to regression data. Technical Report No. 209, Department of Statistics, University of California at Berkeley.

BUJA, A., HASTIE, T. and TIBSHIRANI, R. (1989). Linear smoothers and additive models (with discussion). *Ann. Statist.* **17** 453–555.

DE BOOR, C. (1976). A bound on the L_∞-norm of L_2-approximation by splines in terms of a global mesh ratio. *Math. Comp.* **30** 765–771.

DE BOOR, C. (1978). *A Practical Guide to Splines.* Springer–Verlag, New York.

CHEN, Z. (1989). Interaction spline models and their convergence rates. Manuscript.

FRIEDMAN, J. H. (1991). Multivariate Adaptive Regression Splines. *Ann. Statist.*, to appear.

FRIEDMAN, J. H. and SILVERMAN, B. W. (1989). Flexible parsimonious smoothing and additive modeling (with discussion). *Technometrics* 31 3–39.

HASTIE, T. J. and TIBSHIRANI, R. J. (1990). *Generalized Additive Models.* Chapman and Hall, New York.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

KOO, J.-Y. (1988). *Tensor product splines in the estimation of regression, esponential response functions and multivariate densities.* Ph. D. dissertation, Department of Statistics, University of California at Berkeley.

MO, M. (1990a). Robust additive regression I: population aspect. Manuscript.

MO, M. (1990b). Robust additive regression II: finite sample approximations. Manuscript.

SCHUMAKER, L. L. (1981). *Spline Functions: Basic Theory.* Wiley, New York.

STONE, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360.

STONE, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* 10 1040–1053.

STONE, C. J. (1985). Additive regression and other nonparametric models. *Ann. Statist.* 13 689–705.

STONE, C. J. (1986). The dimensionality reduction principle for generalized additive models. *Ann. Statist.* 14 590–606.

STONE, C. J. and KOO, C.-Y. (1986) Additive splines in statistics. In *1985 Statistical Computing Section Proceedings of the American Statistical Association* 45–48. American Statistical Association, Washington.

STONE, C. J. (1989). Uniform error bounds involving logspline models. *In Probability, Statistics and Mathematics: Papers in Honor of Samuel Karlin* (T. W. Anderson, K. B. Athreya, and D. L. Iglehart, eds.) 335–355. Academic Press, Boston.

WAHBA, G. (1986). Partial and interaction splines for the semiparametric estimation of functions of several variables. In *Computer Science and Statistics: Proceedings of the 18th Symposium on the Interface*. (Boardman, T. J., ed.) 75–80. Amer. Statist. Assoc, Washington, D. C.