Journal of Multivariate Analysis **70**, 57–65 (1999)
Article ID jmva.1999.1812, available online at http://www.idealibrary.com on IDEAL®

A Central Limit Theorem for Local Polynomial Backfitting Estimators

M. P. Wand

Harvard School of Public Health
E-mail:mwand@hsph.harvard.edu

Received March 4, 1998

Additive models based on backfitting estimators are among the most important recent contributions to modern statistical modelling. However, the statistical properties of backfitting estimators have received relatively little attention. Recently, J.-D. Opsomer and D. Ruppert (1997, *Ann. Statist.* **25**, 186-211; 1998, *J. Amer. Statist. Assoc.* **93**, 605–619) and J.-D. Opsomer (1997, preprint 96-12, Department of statistics, Iowa State University) derived their mean squared error properties in the case of local polynomial smoothers. In this paper the asymptotic distributional behaviour of backfitting estimators is investigated. © 1999 Academic Press

AMS 1991 subject classifications: primary 62G05; secondary 62E20.

Key words and phrases: additive models; kernel smoothing; limiting distribution; regression functionals.

1. INTRODUCTION

Additive models (Ezekiel, 1924) postulate that the conditional mean of the response variable is a sum of functions of each of the predictor variables. For example,

$$E(\texttt{sales}|\texttt{price}, \texttt{advertizing}) = \alpha + m_1(\texttt{price}) + m_2(\texttt{advertising}),$$

for some functions $m_1(\cdot)$ and $m_2(\cdot)$, is an additive model that may arise in a business context. Such models and generalized extension have become among the most widely used statistical tools, mainly because of the exemplary monograph of Hastie and Tibshirani (1990) and companion software as described in Chambers and Hastie (1991).

Since their introduction to modern statistics by Friedman and Stuetzle (1981) the most important practical development has been the evolution of the *backfitting* algorithm (Ezekiel, 1924; Buja, Hastie, and Tibshirani, 1989). This algorithm allows additive models to be fit through repetitive use of single-predictor smoothing operators (or "scatterplot smoothers") which greatly enhances their practical implementation. Most of the theory

0047-259X/99 $30.00
Copyright © 1999 by Academic Press
All rights of reproduction in any form reserved.

on backfitting has been confined to convergence of the algorithm (Buja, Hastie, and Tibshirani, 1989; Härdle and Hall, 1993; Ansley and Kohn, 1994; Opsomer and Ruppert, 1997) whereas the statistical properties have received considerably less attention. Stone (1985) derived optimal rates of convergence of additive models based on local polynomials, but it was not until the recent work of Opsomer and Ruppert (1997) that the asymptotic mean squared error properties of backfitting estimators were derived. In this paper we carry their theory one step further by establishing a joint central limit theorem for backfitting estimators. One of the most significant findings is that the backfitting estimates for each predictor are asymptotically independent of one another. This gives some theoretical support for the use of marginal nonparametric information in standard error calculations—an approach that is commonly used to avoid computational difficulties (Chambers and Hastie, 1991).

There has been a good deal of theoretical development for the non-backfitting approach to additive model fitting known as marginal integration (Linton and Nielson, 1995). For example, Fan, Härdle, and Mammen (1998) derive the asymptotic distribution of such estimators.

Section 2 describes the backfitting estimation framework and Section 3 presents the main result. The proof is given in an Appendix.

2. LOCAL POLYNOMIAL BACKFITTING ESTIMATORS

Let $(X_1, Y_1), ..., (X_n, Y_n)$ be a set of independent and identically distributed random pairs, where the Y_i's are real-valued and the X_i's are $\mathbb{R}^d$-valued with jth component denoted by $X_{ji}, 1 \leqslant j \leqslant d$. The additive regression model is

$$E(Y_i | X_1, ..., X_n) = \alpha + \sum_{j=1}^{d} m_j(X_{ji}), \qquad 1 \leqslant i \leqslant n,$$

where, for identification purposes, each of the $m_j(\cdot)$ functions are such that $E\{m_j(X_{ji})\} = 0$.

Consider, for the moment, the family of single-predictor models:

$$E(Y_i | X_1, ..., X_n) = m_j(X_{ji}), \qquad 1 \leqslant j \leqslant d,$$

Then any estimate of $m_j = [m_j(X_{j1}), ..., m_j(X_{jd})]^\mathsf{T}$ of the form

$$\tilde{m}_j = S_j Y,$$

where S_j is an $n \times n$ matrix depending on $X_{j1}, ..., X_{jd}$, is called a *linear smoother* with *smoother matrix* S_j.

The backfitting algorithm estimates the vectors $m_1, ..., m_d$ through the process:

(i) Set $\hat{\alpha} = \bar{Y}$ and initialise $\hat{m}_j$.

(ii) Cycle $j = 1, ..., d$: $\hat{m}_j = S_j^*(Y - \alpha \mathbf{1} - \sum_{k \neq j} \hat{m}_k)$.

(iii) Repeat (ii) until convergence.

(See, e.g., Hastie and Tibshirani, 1990, p. 91.) Here $S_j^* = (I - n^{-1}\mathbf{1}\mathbf{1}^{\mathsf{T}})\, S_j$ where $\mathbf{1}$ is the $n \times 1$ vector of ones. This adjustment is made to ensure that the $\hat{m}_j(\cdot)$ remain centred about zero during the algorithm and to enforce identifiability (e.g., Hastie and Tibshirani, 1990; Opsomer and Ruppert, 1997). Uniqueness of the estimators arising from backfitting is a delicate matter and has been addressed through the concept of *concurvity* by Buja, Hastie, and Tibshirani (1989). Assuming that concurvity is not present it can be shown that the convergent of the backfitting algorithm is given by

$$\begin{bmatrix} \hat{m}_1 \\ \hat{m}_2 \\ \vdots \\ \hat{m}_d \end{bmatrix} = \begin{bmatrix} I & S_1^* & \cdots & S_1^* \\ S_2^* & I & \cdots & S_2^* \\ \vdots & \vdots & \ddots & \vdots \\ S_d^* & S_d^* & \cdots & I \end{bmatrix} - 1 \quad \begin{bmatrix} S_1^* \\ S_2^* \\ \vdots \\ S_d^* \end{bmatrix} Y. \tag{1}$$

The right-hand side of this expression is much more expensive to compute than backfitting so is rarely used in practical implementations. However, because of its explicitness it is easier to handle theoretically. So we will take (1) to be *the* set of backfitting estimates of $\hat{m}_j$, $1 \leqslant j \leqslant d$.

Let $W_1, ..., W_d$ be the $n \times n$ matrices for which

$$\hat{m}_j = W_j Y, \qquad j = 1, ..., d.$$

From (1) it is easily seen that W_j can be expressed in terms of the S_j^*. For example, in the case $d = 2$ we have

$$W_1 = I - (I - S_1^* S_2^*)^{-1} (I - S_1^*)$$
$$W_2 = I - (I - S_2^* S_1^*)^{-1} (I - S_2^*)$$

(Hastie and Tibshirani, 1990, pp. 199–120; and Opsomer and Ruppert, 1997). In this paper we take the S_j to correspond with a pth degree local polynomial smoother with kernel K and bandwidth h_j. This means that the (i, k) entry of S_j is

$$(S_j)_{ik} = e_1^{\mathsf{T}} (X_{X_{ji}}^{\mathsf{T}} W_{X_{ji}} X_{X_{ji}})^{-1} X_{X_{ji}}^{\mathsf{T}} W_{X_{ji}} e_k,$$

where

$$X_x = \begin{bmatrix} 1 & X_{j1}-x & \cdots & (X_{j1}-x)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{jn}-x & \cdots & (X_{jn}-x)^p \end{bmatrix}, \quad W_x = \underset{1 \leqslant i \leqslant n}{\text{diag}} \; K\left(\frac{X_{ji}-x}{h_j}\right) \tag{2}$$

and e_i is a column vector with 1 in the ith position and zeroes elsewhere.

3. CENTRAL LIMIT THEOREM

We now present the main result of this paper. Let f denote the common density of the X_i and f_j denote the common density of the X_{ji}, $1 \leqslant i \leqslant n$. For each $x = (x_1, ..., x_d) \in \mathbb{R}^d$ let $v(x) = \text{Var}(Y | x_1, ..., x_d)$ and put

$$a_j(x_j) = \int (vf)(x)\, dx_1 \cdots dx_{j-1}\, dx_{j+1} \cdots dx_d.$$

Also, let N_p be the $(p+1) \times (p+1)$ matrix having (i, j) entry equal to $\int u^{i+j-2} K(u)\, du$ and $M_p(u)$ be the same as N_p, but with the first column replaced by $(1, u, ..., u^p)^T$. Then, as in Ruppert and Wand (1994), define the kernel

$$K_p(u) = \{ |M_p(u)| / |N_p| \}\, K(u).$$

Finally, let $h = [h_1, ..., h_d]^T$.

THEOREM 1. *Let $(x_1, ..., x_d)$ be a point in the support of f. Under assumptions given in the Appendix*

$$n^{1/2} \text{diag}\,(h)^{1/2} \begin{bmatrix} \hat{m}_1(x_1) - m_1(x_1) + O(h_1^{p+1}) \\ \vdots \\ \hat{m}_d(x_d) - m_d(x_1) + O(h_d^{p+1}) \end{bmatrix}$$

$$\xrightarrow[D]{} N\left(0, \underset{1 \leqslant j \leqslant d}{\text{diag}} \left\{ \left(\int K_p^2 \right) a_j(x_j)/f_j(x_j)^2 \right\} \right).$$

Remark 1. Each $\hat{m}_j(x_j)$ is independently and asymptotically normally distributed with mean $m(x_j)$ plus an $O(h_j^{p+1})$ bias and with variance $(nh_j)^{-1} (\int K_p^2)\, a_j(x_j)/f_j(x_j)^2$. In the case of homoskedastic errors, i.e.,

$v(x) = \sigma^2$ for all x, this reduces to $(nh_j)^{-1} (\int K_p^2)\, \sigma^2 / f_j(x_j)$ which is identical to the single predictor case.

Remark 2. In practical implementations of backfitting additive models it is common to use only marginal nonparametric information in the computation of standard error curves since full information is very expensive to compute. This is the case for the function `gam( )` in `S-PLUS` (Chambers and Hastie, 1991). The asymptotic independence of the $\hat{m}_j$ gives some justification for this strategy. While the asymptotic covariance matrix given in the Theorem could be used to obtain standard error estimates, it is usually preferrable to estimate them more directly through exact covariance expressions.

Remark 3. The $O(h_j^{p+1})$ terms could be replaced by leading bias expressions as given in Opsomer and Ruppert (1997) and Opsomer (1997). However, in this backfitting estimator context bias expressions are quite complicated so the interested reader is referred to those articles.

APPENDIX

Assumptions. The assumptions which we make for the proof of Theorem 1 are along the same lines as those assumed by Opsomer (1997).

(A.1) For each $1 \leqslant j \leqslant d$ we have $E(|Y_1|^{2+\delta} | X_j = x_j) < \infty$ for all x_j in the support of f_j.

(A.2) The functions m_j, $1 \leqslant j \leqslant d$, each have $p+1$ continuous and bounded derivatives.

(A.3) The densities f, f_j and $f_{jj'}$ $(1 \leqslant j, j' \leqslant d)$ are bounded and continuous, have compact support and their first derivatives have a finite number of sign changes over their supports. Also, $f_j(x_j), f_{jj'}(x_j, x_{j'}) > 0$ for all $(x_1, ..., x_d)$ in the support of f.

(A.4) The kernel K is bounded and continuous, it has compact support and its first derivative has a finite number of sign changes over its support.

(A.5) The bandwidths are sequences that, as $n \to \infty$, satisfy $h_j \to 0$, $nh_j \to \infty$ and $h_j / h_k \to C_{jk}$ where $0 < C_{jk} < \infty$ for all j and k.

Complex assumptions such as (A.3) and (A.4) are made for technical convenience, since they allow easier handling, for example, of tail behaviour of the density functions. Some of these restrictions could be removed, but at the cost of more complicated proofs in papers such as Opsomer and Ruppert (1997) upon which the proofs in this paper rely.

Proof of Theorem 1. For each $1 \leqslant j \leqslant d$ let $x_{j1}, ..., x_{jn}$ be a set of fixed distinct points in the support of f_j. To make the proof easier to follow we will redefine the smoother matrix S_j to be

$$(S_j)_{ik} = e_1^{\mathsf{T}} (X_{x_{ji}}^{\mathsf{T}} W_{x_{ij}} X_{x_{ji}})^{-1} X_{x_{ji}}^{\mathsf{T}} W_{x_{ji}} e_k$$

which corresponds to evaluation of the smooth at the x_{ji} rather than the X_{ji}. Results of Opsomer and Ruppert (1997) and Opsomer (1997) for W_j based on this modification of S_j still hold.

By Lemma 2.1 of Opsomer (1997),

$$\begin{aligned} W_j &= I - (I - S_j^* W^{(-j)})^{-1} (I - S_j^*) \\ &= (I - S_j^* W^{(-j)})^{-1} (I - n^{-1} \mathbf{1}\mathbf{1}^{\mathsf{T}}) S_j + I - (I - S_j^* W^{(-j)})^{-1}, \end{aligned}$$

where $W^{(-j)} = \sum_{k \neq j} W_k$. Therefore

$$\begin{aligned} \hat{m}_j(x_{ji}) = {} & e_i^{\mathsf{T}} (I - S_j^* W^{(-j)})^{-1} (I - n^{-1} \mathbf{1}\mathbf{1}^{\mathsf{T}}) \tilde{m}_j \\ & + e_i^{\mathsf{T}} \{ I - (I - S_j^* W^{(-j)})^{-1} \} Y, \end{aligned} \tag{3}$$

where $\tilde{m}_j = [\tilde{m}_j(x_{j1}), ..., \tilde{m}_j(x_{jn})]^{\mathsf{T}}$.

Results of Opsomer and Rupert (1997, Lemma 3.2) and Opsomer (1997, Eq. (12)) indicate that

$$(I - S_j^* W^{(-j)})^{-1} = U + o_P(n^{-1} \mathbf{1}\mathbf{1}^{\mathsf{T}}) = I + O_P(n^{-1} \mathbf{1}\mathbf{1}^{\mathsf{T}}), \tag{4}$$

where U is a deterministic $n \times n$ matrix depending on the x_{ji}. For example, in the case $d = 2$,

$$U = (I - T)^{-1}, \qquad \text{where} \quad T_{ii'} = n^{-1} \left\{ \frac{f(x_{1i}, x_{2i'})}{f_1(x_{1i}) f_2(x_{2i'})} - 1 \right\}.$$

Using arguments given in the proof of Theorem 4.1 of Ruppert and Wand (1994), for each $1 \leqslant j \leqslant d$,

$$\begin{aligned} \tilde{m}_j(x_{ji}) - E(Y_i \mid X_{ji} = x_{ji}) = {} & \sum_{i'=1}^{n} \frac{K_p((X_{ji'} - x_{ji})/h_j) \{ Y_{i'} - E(Y_i \mid X_{ji} = x_{ji}) \}}{n h_j f_j(x_{ji'})} \\ & + o_P\{(nh_j)^{-1/2}\}. \end{aligned} \tag{5}$$

Let $\alpha = [\alpha_1, ..., \alpha_d]^{\mathsf{T}}$ be an arbitrary d-vector and define

$$\hat{\theta} \equiv \alpha^{\mathsf{T}} \begin{bmatrix} \hat{m}_1(x_{11}) - m_1(x_{11}) \\ \vdots \\ \hat{m}_d(x_{d1}) - m_d(x_{d1}) \end{bmatrix}.$$

Then from (3) and (5) we have

$$\hat{\theta} = \sum_{i=1}^{n} \mathscr{X}_i + R_n + \sum_{j=1}^{d} o_P\{(nh_j)^{-1/2}\},$$

where

$$\mathscr{X}_i = \sum_{j=1}^{d} \alpha_j \sum_{i'=1}^{n} \{U(I - n^{-1}\mathbf{1}\mathbf{1}^{\mathsf{T}})\}_{1i'} \frac{K_p((X_{ji} - x_{ji'})/h_j)\{Y_i - E(Y_{i'} \mid X_{ji'} = x_{ji'})\}}{nh_j f_j(x_{ji'})}$$

and

$$R_n = \sum_{j=1}^{d} \alpha_j \Bigg[\sum_{i'=1}^{n} \{(I - S_j^* W^{(-j)})^{-1} - U\}_{1i'} \tilde{m}_j(x_{ji'}) + m_j(x_{j1}) - \sum_{i'=1}^{n} \{U(I - n^{-1}\mathbf{1}\mathbf{1}^{\mathsf{T}})\}_{1i'} E(Y_{i'} \mid X_{ji'} = x_{ji'}) \Bigg].$$

The $\mathscr{X}_i$ are independent, so by Liapounov's version of the Central Limit Theorem,

$$\frac{\sum_{i=1}^{n} \{\mathscr{X}_i - E(\mathscr{X}_i)\}}{\sqrt{\sum_{i=1}^{n} \operatorname{Var}(\mathscr{X}_i)}} \xrightarrow[D]{} N(0, 1) \tag{6}$$

provided

$$\lim_{n \to \infty} \frac{\sum_{i=1}^{n} E\,|\mathscr{X}_i - E(\mathscr{X}_i)|^{2+\delta}}{\{\sum_{i=1}^{n} \operatorname{Var}(\mathscr{X}_i)\}^{(2+\delta)/2}} = 0 \tag{7}$$

for some $\delta > 0$. Now,

$$\begin{aligned} \sum_{i=1}^{n} \operatorname{Var}(\mathscr{X}_i) = n \sum_{j=1}^{d} \alpha_j^2 \operatorname{Var} \sum_{i'=1}^{n} & \{U(I - n^{-1}\mathbf{1}\mathbf{1}^2)\}_{1i'} \\ & \times \frac{K_p((X_{ji} - x_{ji'})/h_j)\{Y_i - E(Y_{i'} \mid X_{ji'} = x_{ji'})\}}{nh_j f_j(x_{ji'})} \\ & + n \sum \sum_{j \neq j'} \alpha_j \alpha_{j'} \operatorname{Cov} \Bigg[\sum_{i'=1}^{n} \{U(I - n^{-1}\mathbf{1}\mathbf{1}^2)\}_{1i'} \\ & \times \frac{K_p((X_{ji} - x_{ji'})/h_j)\{Y_i - E(Y_{i'} \mid X_{ji'} = x_{ji'})\}}{nh_j f_j(x_{ji'})}, \\ & \times \sum_{i'=1}^{n} \{U(I - n^{-1}\mathbf{1}\mathbf{1}^2)\}_{1i'} \\ & + \frac{K_p((X_{ji'} - x_{j'i'})/h_{j'})\{Y_i - E(Y_{i'} \mid X_{j'i'} = x_{j'i'})\}}{nh_{j'} f_{j'}(x_{j'i'})} \Bigg]. \end{aligned} \tag{8}$$

Note that, from (4),

$$\{U(I-n^{-1}\,\mathbf{1}\mathbf{1}^2)\}_{1i'} = \{I+O(n^{-1}\,\mathbf{1}\mathbf{1}^{\mathsf{T}})\}_{1i'} = \begin{cases} 1+O(n^{-1}), & i'=1 \\ O(n^{-1}), & i'=2,\ldots,n, \end{cases}$$

so the dominant part of the variance term of (8) is

$$\sum_{j=1}^{d} \alpha_j^2 n^{-1} h_j^{-2} \operatorname{Var}\left[\left(\frac{X_{ji}-x_{j1}}{h_j}\right) f_j(x_{j1})^{-1}\{Y_i - E(Y_i \mid X_{ji}=x_{j1})\}\right]$$

$$= \sum_{j=1}^{d} \alpha_j^2 (nh_j)^{-1} a_j(x_{j1})\, f_j(x_{j1})^{-2}\left(\int K_p^2\right)\{1+o(1)\}.$$

Similar arguments can be used to show that the covariance term of (8) is $O(n^{-1})$ so

$$\sum_{i=1}^{n} \operatorname{Var}(\mathscr{X}_i) = \sum_{j=1}^{d} \alpha_j^2 (nh_j)^{-1} a_j(x_{j1}) f_j(x_{j1})^{-2}\left(\int K_p^2\right)\{1+o(1)\}. \tag{9}$$

By repeated use of the inequality $|a+b|^r \leqslant 2^{r-1}(|a|^r + |b|^r)$, $r \geqslant 1$,

$$\sum_{i=1}^{n} E\,|\mathscr{X}_i - E(\mathscr{X}_i)|^{2+\delta} \leqslant 2^{2+\delta+(d-1)(1+\delta)} \sum_{j=1}^{d} |\alpha_j|^{2+\delta}$$

$$\times f_j(x_j)^{-1} h_j^{-2-\delta} E\left|Kp\left(\frac{X_{j1}-x_j}{h_j}\right) Y_1\right|^{2\delta}$$

$$= 2^{2+\delta+(d-1)(1+\delta)} \sum_{j=1}^{d} |\alpha_j|^{2+\delta} f_j(x_{j1})^{-1} (nh_j)^{-1-\delta}$$

$$\times \int |K_p|^{2\delta} E(|Y_1|^{2+\delta} \mid X_{j1}=x_{j1})\{1+o(1)\}$$

so (6) follows from (7), (A.1), (A.3), and (A.4). It then follows from (9) and Slutsky's Theorem that

$$\frac{\sum_{i=1}^{n}\{\mathscr{X}_i - E(\mathscr{X}_i)\}}{\sqrt{n\alpha^{\mathsf{T}} \operatorname{diag}(h) \sum_x \alpha}} \xrightarrow[D]{} N(0,1),$$

where $\sum_x = \operatorname{diag}_{1 \leqslant j \leqslant d}\{(\int K_p^2)\, a_j(x_j)/f_j(x_j)^2\}$. Similar arguments can be used to show that

$$\sum_{i=1}^{n} E(\mathscr{X}_i) = \sum_{j=1}^{d} O(h_j^{p+1})$$

and $\{n\alpha^{\mathsf{T}} \operatorname{diag}(h)\,\alpha\}^{-1/2} R_n \to_P 0$ so further application of Slutsky's Theorem leads to

$$\frac{\hat{\theta} - \sum_{j=1}^{d} O(h_j^{p+1})}{\sqrt{n\alpha^{\mathsf{T}} \operatorname{diag}(h) \sum_x \alpha}} \xrightarrow[D]{} N(0, 1).$$

The stated result follows from this and the Cramér–Wold Device.

ACKNOWLEDGMENTS

I am grateful to the editor and two referees for helpful comments. This research was partially supported by Sonderforschungsbereich 373 at Humboldt University, Germany. Parts of this research were performed while the author was at University of New South Wales, Australia.

REFERENCES

1. C. F. Ansley and R. Kohn, Convergence of the backfitting algorithm for additive models, *J. Austral. Math. Soc. Ser. A* **57** (1994), 316–329.
2. A. Buja, T. J. Hastie, and R. J. Tibshirani, Linear smoothers and additive models, *Ann. Statist.* **17** (1989), 453–555.
3. J. M. Chambers and T. J. Hastie, "Statistical Models," Wadsworth/ Brooks Cole, Pacific Grove, CA, 1991.
4. J. Fan, W. Härdle, and E. Mammen, Direct estimation of low-dimensional components in additive models, *Ann. Satist.* (1998), 943–971.
5. J. H. Friedman and W. Stuetzle, Projection pursuit regression, *J. Amer. Statist. Assoc.* **76** (1981), 817–823.
6. M. Ezekiel, A method for handling curvilinear correlation for any number of variables, *J. Amer. Statist. Assoc.* **19** (1924), 431–453.
7. W. Härdle and P. Hall, On the backfitting algorithm for additive regression models, *Statist. Neerlandica* **47** (1993), 43–57.
8. T. J. Hastie and R. J. Tibshirani, "Generalized Additive Models," Chapman and Hall, Washington, 1990.
9. O. B. Linton and J. P. Nielsen, A kernel method of estimating structured nonparametric regression based on marginal integration, *Biometrika* **82** (1995), 93–101.
10. J.-D. Opsomer, "Optimal Bandwith Selection for Fitting an Additive Model by Local Polynomial Regression," Ph.D. dissertation, Cornell University; http://www.public.iastate.edu/~ jopsomer/research.html, 1995.
11. J.-D. Opsomer, On the existence and asymptotic properties of backfitting estimators. Preprint 96-12, Department of Statistics, Iowa State University. http://www.public.iastate.edu/~ jopsomer/research.html, 1997.
12. J.-D. Opsomer and D. Ruppert, Fitting a bivariate additive model by local polynomial regression, *Ann. Statist.* **25** (1997), 186–211.
13. J.-D. Opsomer and D. Ruppert, A fully automated bandwidth selection method for fitting additive models, *J. Amer. Statist. Assoc.* **93** (1998), 605–619.
14. D. Ruppert and M. P. Wand, Multivariate locally weighted least squares regression, *Ann. Statist.* **22** (1994), 1346–1370.
15. M. P. Wand and M. C. Jones, "Kernel Smoothing," Chapman & Hall, London, 1995.