Journal of the Korean Statistical Society (2004), **33**: 4, pp 435–447

ON MARGINAL INTEGRATION METHOD IN NONPARAMETRIC REGRESSION†

YOUNG KYUNG LEE[1]

ABSTRACT

In additive nonparametric regression, Linton and Nielsen (1995) showed that the marginal integration when applied to the local linear smoother produces a rate-optimal estimator of each univariate component function for the case where the dimension of the predictor is two. In this paper we give new formulas for the bias and variance of the marginal integration regression estimators which are valid for boundary areas as well as fixed interior points, and show the local linear marginal integration estimator is in fact rate-optimal when the dimension of the predictor is less than or equal to four. We extend the results to the case of the local polynomial smoother, too.

AMS 2000 subject classifications. Primary 62G08; Secondary 62G20.
Keywords. Marginal integration, local linear smoothing, backfitting, boundary properties.

1. INTRODUCTION

Suppose that we observe $(\boldsymbol{X}_1, Y_1), \ldots, (\boldsymbol{X}_n, Y_n)$ which are *iid* copies of $(\boldsymbol{X}, Y)$, where $\boldsymbol{X}$ is a d-dimensional predictor. Let $m(\boldsymbol{x}) = E(Y|\boldsymbol{X} = \boldsymbol{x})$ is the regression function. In additive regression models, it is assumed that

$$m(\boldsymbol{x}) = m_0 + m_1(x_1) + \cdots + m_d(x_d) \tag{1.1}$$

where m_0 is a constant and m_j's are univariate functions. See Hastie and Tibshirani (1990) for the practical and theoretical aspects of additive models.

Three useful approaches to estimate the regression function are the ordinary backfitting algorithm (Friedman and Stuetzle, 1981; Buja *et al.*, 1989), the

Received June 2004; accepted September 2004.
†This research was supported by the Brain Korea 21 projects.
[1]Department of Statistics, Seoul National University, Seoul 151-747, Korea

smooth backfitting (Mammen *et al.*, 1999), and the marginal integration (Linton and Nielsen, 1995). Opsomer and Ruppert (1997), Opsomer (2000) and Mammen *et al.* (1999) provided some statistical properties of the backfitting estimators. Sperlich *et al.* (1999) compared finite sample properties of the ordinary backfitting and marginal integration methods. Severance-Lossin and Sperlich (1999) extended the approach of marginal integration to regression derivative estimation.

Linton and Nielsen (1995) showed that the marginal integration method, when applied to the local linear smoother, can achieve the one-dimensional $n^{-2/5}$ rate of convergence under the smoothness condition that m has two continuous partial derivatives. However, their proof relies on the stochastic approximation of the marginal integration estimator by the 'internal estimator' which is a multivariate Nadaraya-Watson-like estimator with the true joint density function in the denominator. The approximation error is of order $O_p(n^{-1}h^{-d})$ when a common bandwidth h is used for the initial smoother. This can dominate unless h is chosen very large, thereby excluding the optimal bandwidth $h = O(n^{-1/5})$ in the case where $d \geq 3$. In fact, Linton and Nielsen (1995) failed to show the one-dimensional $n^{-2/5}$ rate of convergence for the marginal integration method when $d \geq 3$.

In this paper, we give new formulas for the bias and variance which are valid for boundary areas as well as fixed interior points, and show that the marginal integration can afford the univariate convergence rate for $d \leq 4$. For this, we derive a relevant stochastic expansion for the estimator, which is different from the one used by Linton and Nielsen (1995). Furthermore, we extend the results to the case of the local polynomial smoother. We show that the marginal integration when applied to the local polynomial estimators of order p (odd) can have the one-dimensional $n^{-(p+1)/(2p+3)}$ rate of convergence when $d \leq 2(p+1)$.

2. Marginal Integration with Local Linear Smoother

Let q be a given weight function defined on $\mathbb{R}^d$ with $\int q(\boldsymbol{u})d\boldsymbol{u} = 1$. For $\boldsymbol{x} = (x_1, \ldots, x_d)$, we write $\boldsymbol{x}_{-j} = (x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_d)$. Define q_{-j}, $1 \leq j \leq d$, to be the marginalizations of q on $\mathbb{R}^{d-1}$, *i.e.*, $q_{-j}(\boldsymbol{u}_{-j}) = \int q(\boldsymbol{u})du_j$. Also, define $q_j(u_j) = \int q(\boldsymbol{u})d\boldsymbol{u}_{-j}$. Define

$$\alpha_j(x_j) = \int m(\boldsymbol{x})q_{-j}(\boldsymbol{x}_{-j})d\boldsymbol{x}_{-j}$$

for $j = 1, \ldots, d$. The marginal integration method is based on the fact that $\alpha_j(x_j) = c_j + m_j(x_j)$ under the additive structure (1.1), where $c_j = m_0 +$

$\sum_{k\neq j}^{d} \int m_k(x_k)q_k(x_k)dx_k$. Thus, up to constant addition, $\alpha_j(x_j)$ is the univariate component of interest, $m_j(x_j)$. Identifiability of m_j's requires an additional structural assumption. A usual practice is to assume that all the component functions are centered, *i.e.* $E\{m_j(X_j)\} = 0$, $j = 1, \ldots, d$.

Suppose that we are given an estimator $\widetilde{m}(\boldsymbol{x})$ of $m(\boldsymbol{x})$. The marginal integration estimator of $\alpha_j(x_j)$ (Linton and Nielsen, 1995) is given by

$$\widehat{\alpha}_j(x_j) = \int \widetilde{m}(\boldsymbol{x})q_{-j}(\boldsymbol{x}_{-j})d\boldsymbol{x}_{-j}.$$

With the norming conditions $E\{m_j(X_j)\} = 0$, $j = 1, \ldots, d$, the j^{th} component function $m_j(x_j)$ may be estimated by

$$\widehat{m}_j(x_j) = \widehat{\alpha}_j(x_j) - n^{-1}\sum_{i=1}^{n}\widehat{\alpha}_j(X_{ij}), \tag{2.1}$$

and the constant term m_0, which equals $E(Y)$, is then estimated by $\bar{Y}$. Thus, the estimator of the whole regression function $m(\boldsymbol{x})$ is given by

$$\widehat{m}(\boldsymbol{x}) = \bar{Y} + \sum_{j=1}^{d}\widehat{m}_j(x_j). \tag{2.2}$$

In this section, we consider the local linear smoother for $\widetilde{m}$. Thus, $\widetilde{m}(\boldsymbol{x})$ is given by $\widetilde{\beta}_0 \equiv \widetilde{\beta}_0(\boldsymbol{x})$ which, with $\widetilde{\beta}_j$ for $j = 1, \ldots, d$, minimizes

$$\sum_{i=1}^{n}\left\{Y_i - \beta_0 - \sum_{j=1}^{d}\beta_j\left(\frac{X_{ij} - x_j}{h_j}\right)\right\}^2 \prod_{j=1}^{d} K_j\left(\frac{X_{ij} - x_j}{h_j}\right),$$

where K_j's are kernel functions and h_j's are positive numbers called bandwidths. Let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ and $\boldsymbol{\mathcal{X}}^T = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$ with

$$\boldsymbol{a}_i \equiv \boldsymbol{a}_i(\boldsymbol{x}) = \left(1, \frac{X_{i1} - x_1}{h_1}, \ldots, \frac{X_{id} - x_d}{h_d}\right)^T.$$

Define $\mathbf{W} \equiv \mathbf{W}(\boldsymbol{x})$ to be the $n \times n$ diagonal matrix whose i^{th} diagonal entry equals $\prod_{j=1}^{d} K_{j,h_j}(X_{ij} - x_j)$, where $K_{j,h_j}(u_j) = K_j(u_j/h_j)/h_j$. Then, we may write

$$\widetilde{m}(\boldsymbol{x}) = \boldsymbol{e}_1{}^T\left(\boldsymbol{\mathcal{X}}^T\mathbf{W}\boldsymbol{\mathcal{X}}\right)^{-1}\boldsymbol{\mathcal{X}}^T\mathbf{W}\boldsymbol{Y}, \tag{2.3}$$

where $\boldsymbol{e}_1$ is the $(d + 1)$-dimensional unit vector with 1 appearing in the first position.

In the first two theorems below, we give a uniform stochastic expansion of $\widehat{\alpha}_j(x_j)$ and its asymptotic distribution. We emphasize here that these results do not require the regression function to have the additive structure given at (1.1). We collect the assumptions below:

(A1) The kernel functions K_j's have compact supports, say $[-1,1]$, and are Lipschitz continuous, nonnegative and symmetric with $\int K_j = 1$;

(A2) The marginal density, f, of $\boldsymbol{X}$ has a compact support, say $[0,1]^d$, is bounded away from zero and infinity on its support, and satisfies the Lipschitz condition of order 1;

(A3) The weight function q has a compact support which is contained in $[0,1]^d$, and is continuous;

(A4) The regression function $m(\boldsymbol{x})$ has continuous second partial derivatives;

(A5) The variance function $\sigma^2(\boldsymbol{x}) = \mathrm{Var}(Y|\boldsymbol{X}=\boldsymbol{x})$ is continuous;

(A6) The bandwidth h_j is asymptotic to $n^{-\gamma_j}$ with $\gamma_j > 0$ and $\gamma_1 + \cdots + \gamma_d < 1$.

To state the first theorem, let $\boldsymbol{M}(\boldsymbol{x})$ denote the $(d+1)\times(d+1)$ matrix of the (incomplete) moments of the product kernel defined by

$$\boldsymbol{M}(\boldsymbol{x}) = \int_{-x_d/h_d}^{(1-x_d)/h_d} \cdots \int_{-x_1/h_1}^{(1-x_1)/h_1} \begin{pmatrix} 1 \\ \boldsymbol{u} \end{pmatrix} (1, \boldsymbol{u}^T) \prod_{j=1}^{d} K_j(u_j) du_j,$$

where $\boldsymbol{u} = (u_1,\ldots,u_d)^T$. Define $Y_i^*(\boldsymbol{x}) = Y_i - m(\boldsymbol{x}) - \sum_{j=1}^d (X_{ij}-x_j)\{\partial m(\boldsymbol{x})/\partial x_j\}$ and

$$T_j(x_j) = n^{-1}\sum_{i=1}^{n}\int\Big\{f(\boldsymbol{x})^{-1}\boldsymbol{e}_1{}^T\boldsymbol{M}^{-1}(\boldsymbol{x})\boldsymbol{a}_i(\boldsymbol{x})\prod_{r=1}^{d}K_{r,h_r}(X_{ir}-x_r)Y_i^*(\boldsymbol{x})\Big\}\,q_{-j}(\boldsymbol{x}_{-j})d\boldsymbol{x}_{-j}.$$

THEOREM 2.1. *Assume (A1)-(A6) hold. Then, for each $j = 1,\ldots,d$, uniformly for $x_j \in [0,1]$ as $n \to \infty$, we have*

$$\widehat{\alpha}_j(x_j) - \alpha_j(x_j) = T_j(x_j) + o_p\Big(\sum_{r=1}^{d} h_r^2 + n^{-1/2}h_j^{-1/2}\Big).$$

The next theorem describes the limit distribution of $\widehat{\alpha}_j(x_j)$. To state the theorem, let $K_j^*(u, x_j)$ be the equivalent kernel (see Fan and Gijbels, 1996) defined by

$$K_j^*(u, x_j) = \left\{ \frac{\mu_{2,j}(x_j) - \mu_{1,j}(x_j)u}{\mu_{0,j}(x_j)\mu_{2,j}(x_j) - \mu_{1,j}(x_j)^2} \right\} K_j(u),$$

where

$$\mu_{k,j}(x_j) = \int_{-x_j/h_j}^{(1-x_j)/h_j} u^k K_j(u)du.$$

Define for $j = 1, \ldots, d$

$$b_j(x_j) = \int_{-x_j/h_j}^{(1-x_j)/h_j} u^2 K_j^*(u, x_j)du \int \frac{\partial^2 m(\boldsymbol{x})}{\partial x_j^2} q_{-j}(\boldsymbol{x}_{-j})d\boldsymbol{x}_{-j}, \tag{2.4}$$

$$s_j^2(x_j) = \int_{-x_j/h_j}^{(1-x_j)/h_j} \{K_j^*(u, x_j)\}^2 du \int \frac{\sigma^2(\boldsymbol{x})}{f(\boldsymbol{x})} q_{-j}^2(\boldsymbol{x}_{-j})d\boldsymbol{x}_{-j},$$

and for $r \neq j$

$$b_{r,j}(x_j) = \int_{-1}^{1} u^2 K_r(u)du \int \frac{\partial^2 m(\boldsymbol{x})}{\partial x_r^2} q_{-j}(\boldsymbol{x}_{-j})d\boldsymbol{x}_{-j}. \tag{2.5}$$

THEOREM 2.2. *Under the conditions of Theorem 2.1, we have for each $j = 1, \ldots, d$ and for every $x_j \in [0, 1]$, possibly depending on n,*

$$\widehat{\alpha}_j(x_j) - \alpha_j(x_j) = (nh_j)^{-1/2} s_j(x_j) Z_n + \frac{1}{2}\Big\{ b_j(x_j)h_j^2 + \sum_{r \neq j} b_{r,j}(x_j)h_r^2 \Big\}$$
$$+ o_p\Big(\sum_{r=1}^{d} h_r^2 + n^{-1/2} h_j^{-1/2} \Big)$$

as $n \to \infty$, where Z_n converges in distribution to the standard normal distribution.

The bias and variance formulas given in Theorem 2.2 are simplified for interior points. For $x_j \in [h_j, 1 - h_j]$, we have $K_j^*(u, x_j) \equiv K_j(u)$ by the condition (A1), and can replace the incomplete integrals in the definitions of $b_j(x_j)$ and $s_j^2(x_j)$ by the complete integrals over the entire interval $[-1, 1]$.

Now, we present some statistical properties of $\widehat{m}_j(x_j)$ defined at (2.1). First, we note that, under the additive structure (1.1), $\partial^2 m(\boldsymbol{x})/\partial x_j^2 = m_j''(x_j)$. Thus,

$b_j(x_j)$ and $b_{r,j}(x_j)$ defined at (2.4) and (2.5), respectively, may be replaced by

$$\widetilde{b}_j(x_j) = m_j''(x_j) \int_{-x_j/h_j}^{(1-x_j)/h_j} u^2 K_j^*(u, x_j) du,$$
$$\widetilde{b}_{r,j}(x_j) \equiv \widetilde{b}_r = \int m_r''(x_r) q_r(x_r) dx_r \int_{-1}^{1} u^2 K_r(u) du.$$

By Theorem 2.1, we have

$$n^{-1}\sum_{i=1}^{n} \{\widehat{\alpha}_j(X_{ij}) - \alpha_j(X_{ij})\} = n^{-1}\sum_{i=1}^{n} T_j(X_{ij}) + o_p\Big(\sum_{r=1}^{d} h_r^2 + n^{-1/2}h_j^{-1/2}\Big). \quad (2.6)$$

It may be shown that the stochastic part of $n^{-1}\sum_{i=1}^{n} T_j(X_{ij})$ makes a negligible contribution of order $o_p(n^{-1/2}h_j^{-1/2})$. Also, it is not difficult to see that

$$E\{T_j(X_{1j})\} = \frac{1}{2}\Big\{h_j^2 E\widetilde{b}_j(X_{1j}) + \sum_{r\neq j} h_r^2 \widetilde{b}_r\Big\} + o\Big(\sum_{r=1}^{d} h_r^2\Big). \quad (2.7)$$

We define the 'centered' version of $\widetilde{b}_j(x_j)$ by

$$\widetilde{b}_j^c(x_j) = \widetilde{b}_j(x_j) - E\widetilde{b}_j(X_j).$$

From (2.6), (2.7) and the fact $n^{-1}\sum_{i=1}^{n} \alpha_j(X_{ij}) - c_j = O_p(n^{-1/2})$ under the norming condition $E\{m_j(X_j)\} = 0$, we obtain the following theorem.

THEOREM 2.3. *In addition to the assumptions (A1)–(A6), we assume that the regression function has the additive structure (1.1) and* $E\{m_j(X_j)\} = 0$ *for* $1 \le j \le d$. *Then, for each* $j = 1, \ldots, d$ *and for every* $x_j \in [0,1]$, *possibly depending on* n,

$$\widehat{m}_j(x_j) - m_j(x_j) = (nh_j)^{-1/2} s_j(x_j) Z_n + \frac{1}{2}\widetilde{b}_j^c(x_j) h_j^2 + o_p\Big(\sum_{r=1}^{d} h_r^2\Big)$$

as $n \to \infty$, *where* Z_n *converges in distribution to the standard normal distribution.*

The leading bias of $\widehat{m}_j(x_j)$ does not depend on the other component functions m_r, $r \neq j$. If one takes $h_j \asymp n^{-1/5}$, then $\widehat{m}_j(x_j)$ converges to $m_j(x_j)$ at the univariate rate $n^{-2/5}$. However, this is true only when $nh_1 \times \cdots \times h_d = n^{1-(d/5)}$ converges to infinity as n goes to infinity, *i.e.*, when $d \le 4$. Thus, we conclude that the local linear marginal integration estimator achieves the univariate optimal rate $n^{-2/5}$ when $d \le 4$.

Next, we consider $\widehat{m}(\boldsymbol{x})$ defined at (2.2). Suppose that we take $h_j \asymp h = n^{-\gamma}$ for all $1 \le j \le d$ where $0 < \gamma < 1/d$. It may be shown that the covariances of $T_j(x_j)$ and $T_{j'}(x_{j'})$ for $j \ne j'$ have negligible magnitudes of order $o(n^{-1}h^{-1})$. Thus, the asymptotic moments of the estimator $\widehat{m}(\boldsymbol{x})$ are easily calculated from those of $\widehat{m}_j(x_j)$'s. Define $v_n^2(\boldsymbol{x}) = n^{-1}\sum_{j=1}^d h_j^{-1}s_j^2(x_j)$ and $\beta_n(\boldsymbol{x}) = \sum_{j=1}^d h_j^2 \widetilde{b}_j^c(x_j)/2$.

THEOREM 2.4. *In addition to the conditions of Theorem 2.3, we assume $h_j \asymp h = n^{-\gamma}$ for all $1 \le j \le d$ where $0 < \gamma < 1/d$. Then, for every $j \ne j'$ and for all $x_j, x_{j'} \in [0,1]$, possibly depending on n,*

$$\text{Cov}\left\{T_j(x_j), T_{j'}(x_{j'})\right\} = o(n^{-1}h^{-1})$$

as $n \to \infty$, and hence for all $\boldsymbol{x} \in [0,1]^d$, possibly depending on n,

$$\widehat{m}(\boldsymbol{x}) - m(\boldsymbol{x}) = v_n(\boldsymbol{x})Z_n + \beta_n(\boldsymbol{x})\left\{1 + o_p(1)\right\},$$

where Z_n converges in distribution to the standard normal distribution.

3. MARGINAL INTEGRATION WITH LOCAL POLYNOMIAL SMOOTHER

We present extensions of the theorems in the previous section to the case of the local polynomial smoothers with order p. We treat only the case where p is odd. The local p^{th} order polynomial smoother $\widetilde{m}(\boldsymbol{x})$ is given by $\widetilde{\beta}_0 \equiv \widetilde{\beta}_0(\boldsymbol{x})$ which, with $\widetilde{\beta}_j$ for $j = 1, 2, \dots$, minimizes

$$\sum_{i=1}^n \left\{Y_i - \beta_0 - \beta_1\left(\frac{X_{i1} - x_1}{h_1}\right) - \cdots - \beta_d\left(\frac{X_{id} - x_d}{h_d}\right) \right.$$
$$\left. - \cdots - \beta_J\left(\frac{X_{id} - x_d}{h_d}\right)^p\right\}^2 \prod_{j=1}^d K_j\left(\frac{X_{ij} - x_j}{h_j}\right),$$

where J is the total number of β's, which equals $J = \sum_{i=1}^p \binom{d-1+i}{d-1}$.

For a d-tuple $\boldsymbol{r} \equiv (r_1, \dots, r_d)$ and a d-dimensional vector $\boldsymbol{x}$, write

$$\begin{aligned} \boldsymbol{r}! &= r_1! \times \cdots \times r_d!, \\ |\boldsymbol{r}| &= \sum_{i=1}^d r_i, \\ \boldsymbol{x}^{\boldsymbol{r}} &= x_1^{r_1} \times \cdots \times x_d^{r_d}. \end{aligned}$$

For a function g defined on $\mathbb{R}^d$, write

$$g^{(\boldsymbol{r})}(\boldsymbol{x}) = \frac{\partial^{|\boldsymbol{r}|} g(\boldsymbol{x})}{\partial x_1^{r_1} \cdots \partial x_d^{r_d}}.$$

To state an extension of Theorem 2.1, we need to define an ordering of the d-tuples $\boldsymbol{r}$. Consider the group of $\boldsymbol{r}$'s with $|\boldsymbol{r}| = j$ for a fixed $j \in \{1, \ldots, p\}$. Recall that there are $\binom{d-1+j}{j}$ members in this group. Arrange the d-tuples in this group as follows: For two $\boldsymbol{r}^{(1)}$ and $\boldsymbol{r}^{(2)}$, $\boldsymbol{r}^{(1)}$ is positioned first if $r_1^{(1)} = r_1^{(2)}, \ldots, r_{k-1}^{(1)} = r_{k-1}^{(2)}$ and $r_k^{(1)} > r_k^{(2)}$ for some $k \in \{1, 2, \ldots, d\}$. This means $(j, 0, \ldots, 0)$ is positioned first, then followed by $(j-1, 1, 0, \ldots, 0)$, and $(0, \ldots, 0, j)$ is the last. Next, concatenate these ordered groups in the order of their sizes, *i.e.*, put the group with $|\boldsymbol{r}| = 1$ first and put the one with $|\boldsymbol{r}| = p$ last. This yields an arrangement of total J d-tuples. Call this arrangement $\mathcal{A}$. Now, define a function ι which maps a d-tuple $\boldsymbol{r}$ to its rank in the arrangement $\mathcal{A}$. For example, $\iota((1, 0, \ldots, 0)) = 1$ and $\iota((0, 0, \ldots, p)) = J$. Next, for a d-dimensional vector $\boldsymbol{u}$ define $l(\boldsymbol{u})$ to be the J-dimensional vector such that $(l(\boldsymbol{u}))_{\iota(\boldsymbol{r})} = \boldsymbol{u}^{\boldsymbol{r}}$, *i.e.*, $\boldsymbol{u}^{\boldsymbol{r}}$ is the $\iota(\boldsymbol{r})^{th}$ entry of $l(\boldsymbol{u})$.

Now, define an extension of $M(\boldsymbol{x})$ by

$$\boldsymbol{\mathcal{M}}(\boldsymbol{x}) = \int_{-x_d/h_d}^{(1-x_d)/h_d} \cdots \int_{-x_1/h_1}^{(1-x_1)/h_1} \begin{pmatrix} 1 \\ l(\boldsymbol{u}) \end{pmatrix} \left(1,\, l(\boldsymbol{u})^T\right) \prod_{j=1}^{d} K_j(u_j) du_j.$$

Let $\mathcal{Y}_i^*(\boldsymbol{x}) = Y_i - \sum_{\boldsymbol{r}:|\boldsymbol{r}|=0}^{p} \beta_{\boldsymbol{r}} (\boldsymbol{X}_i - \boldsymbol{x})^{\boldsymbol{r}} m^{(\boldsymbol{r})}(\boldsymbol{x})$ and

$$\mathcal{T}_j(x_j) = n^{-1} \sum_{i=1}^{n} \int \Big\{ f(\boldsymbol{x})^{-1} \boldsymbol{e}_1{}^T \boldsymbol{\mathcal{M}}^{-1}(\boldsymbol{x}) l\Big(\frac{\boldsymbol{X}_i - \boldsymbol{x}}{\boldsymbol{h}}\Big) \prod_{r=1}^{d} K_{r,h_r}(X_{ir} - x_r) \mathcal{Y}_i^*(\boldsymbol{x}) \Big\} \times q_{-j}(\boldsymbol{x}_{-j}) d\boldsymbol{x}_{-j},$$

where $\boldsymbol{b}/\boldsymbol{a} = (b_1/a_1, \ldots, b_d/a_d)^T$ for two d-dimensional vectors $\boldsymbol{a}$ and $\boldsymbol{b}$.

An extension of Theorem 2.1 is given below. The theorem holds under the assumptions stated in the previous section with only (A4) being replaced by

(A4′) The regression function $m(\boldsymbol{x})$ has continuous partial derivatives up to order $p+1$.

THEOREM 3.1. *Assume (A1)-(A3), (A4′), (A5) and (A6) hold. Then, for each $j = 1, \ldots, d$, uniformly for $x_j \in [0, 1]$ as $n \to \infty$, we have*

$$\widehat{\alpha}_j(x_j) - \alpha_j(x_j) = \mathcal{T}_j(x_j) + o_p\Big(\sum_{r=1}^{d} h_r^{p+1} + n^{-1/2} h_j^{-1/2}\Big).$$

Let $\mathcal{M}_j(x_j)$ denote the matrix $\mathcal{M}(\boldsymbol{x})$ with the integrals over $\big[-x_r/h_r,(1-x_r)/h_r\big]$ $(r\neq j)$ being replaced by the complete integrals over $[-1,1]$. Define

$$\mathcal{K}_j^*(u_j,x_j)=\mathbf{e}_1{}^T\mathcal{M}_j^{-1}(x_j)\begin{pmatrix}1\\ l(\boldsymbol{u})\end{pmatrix}K_j(u_j). \tag{3.1}$$

Note that the right hand side of (3.1) does not depend on u_r for $r\neq j$. Define analogues of $b_j(x_j)$, $s_j^2(x_j)$ in (2.4) by

$$\begin{aligned} b_{p,j}(x_j) &= \int_{-x_j/h_j}^{(1-x_j)/h_j} u^{p+1}\mathcal{K}_j^*(u,x_j)du\int\left\{\frac{\partial^{p+1}m(\boldsymbol{x})}{\partial x_j^{p+1}}\right\}q_{-j}(\boldsymbol{x}_{-j})d\boldsymbol{x}_{-j},\\ s_{p,j}^2(x_j) &= \int_{-x_j/h_j}^{(1-x_j)/h_j}\{\mathcal{K}_j^*(u,x_j)\}^2du\int\left\{\frac{\sigma^2(\boldsymbol{x})}{f(\boldsymbol{x})}\right\}q_{-j}^2(\boldsymbol{x}_{-j})d\boldsymbol{x}_{-j}. \end{aligned}$$

Also, for $r\neq j$ define

$$b_{p,r,j}(x_j)=\int_{-1}^{1}u^{p+1}K_r(u)du\int\left\{\frac{\partial^{p+1}m(\boldsymbol{x})}{\partial x_r^{p+1}}\right\}q_{-j}(\boldsymbol{x}_{-j})d\boldsymbol{x}_{-j}.$$

An extension of Theorem 2.2 is given as follows.

THEOREM 3.2. *Under the conditions of Theorem 3.1, we have for each $j=1,\ldots,d$ and for every $x_j\in[0,1]$, possibly depending on n,*

$$\begin{aligned} \widehat{\alpha}_j(x_j)-\alpha_j(x_j) &= (nh_j)^{-1/2}s_{p,j}(x_j)Z_n+\frac{1}{(p+1)!}\Big\{b_{p,j}(x_j)h_j^{p+1}\\ &\quad +\sum_{r\neq j}b_{p,r,j}(x_j)h_r^{p+1}\Big\}+o_p\Big(\sum_{r=1}^{d}h_r^{p+1}+n^{-1/2}h_j^{-1/2}\Big) \end{aligned}$$

as $n\to\infty$, where Z_n converges in distribution to the standard normal distribution.

Define an analogue of $\widetilde{b}_j(x_j)$ in the previous section by

$$\widetilde{b}_{p,j}^c(x_j)=\widetilde{b}_{p,j}(x_j)-E\,\widetilde{b}_{p,j}(X_j).$$

We obtain the following extension of Theorem 2.3.

THEOREM 3.3. *In addition to the conditions of Theorem 3.1, we assume that the regression function has the additive structure (1.1) and $E\{m_j(X_j)\}=0$ for $1\le j\le d$. Then, for each $j=1,\ldots,d$ and for every $x_j\in[0,1]$, possibly depending on n,*

$$\widehat{m}_j(x_j)-m_j(x_j)=(nh_j)^{-1/2}s_{p,j}(x_j)Z_n+\frac{1}{2}\widetilde{b}_{p,j}^c(x_j)h_j^{p+1}+o_p\Big(\sum_{r=1}^{d}h_r^{p+1}\Big)$$

as $n\to\infty$, where Z_n converges in distribution to the standard normal distribution.

As in the case of the local linear marginal integration, the leading bias of $\widehat{m}_j(x_j)$ does not depend on the other component functions m_r, $r \neq j$. If one takes $h_j \asymp n^{-1/(2p+3)}$, then $\widehat{m}_j(x_j)$ converges to $m_j(x_j)$ at the univariate rate $n^{-(p+1)/(2p+3)}$. Again, this is true only when $nh_1 \times \cdots \times h_d = n^{1-(d/(2p+3))}$ converges to infinity as n goes to infinity, *i.e.*, when $d \leq 2(p+1)$. This implies that the local p^{th} order polynomial marginal integration estimator achieves the univariate optimal rate $n^{-(p+1)/(2p+3)}$ when $d \leq 2(p+1)$.

An extension of Theorem 2.4 is also immediate with obviously modified definitions of $v_n^2(\boldsymbol{x})$ and $\beta_n(\boldsymbol{x})$. Thus, we do not state the extension here.

4. Proofs

We only give proofs of Theorems 2.1, 2.2 and 2.4. A proof of Theorem 2.3 is immediate from the discussion in Section 2. Theorems 3.1–3.3 can be proved in similar fashions as in the proofs of Theorems 2.1, 2.2 and 2.3, but with more notational complexity.

4.1. Proof of Theorem 2.1

Define $h_{\text{prod}} = h_1 \times \cdots \times h_d$. Write $\boldsymbol{\mathcal{X}}(\boldsymbol{x})^T = (\boldsymbol{a}_1(\boldsymbol{x}), \ldots, \boldsymbol{a}_n(\boldsymbol{x}))$. By a standard technique in the kernel density estimation, it may be shown that

$$\sup_{\boldsymbol{x}\in[0,1]^d} \left| \frac{1}{n}\boldsymbol{\mathcal{X}}(\boldsymbol{x})^T \mathbf{W}(\boldsymbol{x})\boldsymbol{\mathcal{X}}(\boldsymbol{x}) - f(\boldsymbol{x})\boldsymbol{M}(\boldsymbol{x})\right| = O_p(\rho),$$

where $\rho = \sum_{r=1}^d h_r + \sqrt{(\log n)/(nh_{\text{prod}})}$. Let $\mathbf{I}$ denote the identity matrix. It follows then from a Taylor expansion for the matrix inversion operation that

$$\begin{aligned}
&\left(\frac{1}{n}\boldsymbol{\mathcal{X}}(\boldsymbol{x})^T \mathbf{W}(\boldsymbol{x})\boldsymbol{\mathcal{X}}(\boldsymbol{x})\right)^{-1} \qquad (4.1)\\
&= \frac{1}{f(\boldsymbol{x})}\boldsymbol{M}(\boldsymbol{x})^{-1}\left[\mathbf{I} + \sum_{\ell=1}^{k}\left\{\mathbf{I} - \frac{1}{f(\boldsymbol{x})}\left(\frac{1}{n}\boldsymbol{\mathcal{X}}(\boldsymbol{x})^T \mathbf{W}(\boldsymbol{x})\boldsymbol{\mathcal{X}}(\boldsymbol{x})\right)\boldsymbol{M}(\boldsymbol{x})^{-1}\right\}^{\ell}\right]\\
&\quad + R_k(\boldsymbol{x}),
\end{aligned}$$

where $\sup_{\boldsymbol{x}\in[0,1]^d} |R_k(\boldsymbol{x})| = O_p(\rho^{k+1})$.

For the local linear estimator defined at (2.3), we can write

$$\begin{aligned}
&\widetilde{m}(\boldsymbol{x}) - m(\boldsymbol{x}) \qquad (4.2)\\
&= \boldsymbol{e}_1{}^T\left(\frac{1}{n}\boldsymbol{\mathcal{X}}(\boldsymbol{x})^T \mathbf{W}(\boldsymbol{x})\boldsymbol{\mathcal{X}}(\boldsymbol{x})\right)^{-1} \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{a}_i(\boldsymbol{x})\left\{\prod_{r=1}^{d} K_{r,h_r}(X_{ir} - x_r)\right\} Y_i^*(\boldsymbol{x}).
\end{aligned}$$

Furthermore, one may show

$$\frac{1}{n}\sum_{i=1}^{n}\int \frac{1}{f(\boldsymbol{x})}\boldsymbol{M}(\boldsymbol{x})^{-1}\left\{\mathbf{I}-\frac{1}{f(\boldsymbol{x})}\left(\frac{1}{n}\boldsymbol{\mathcal{X}}(\boldsymbol{x})^{T}\mathbf{W}(\boldsymbol{x})\boldsymbol{\mathcal{X}}(\boldsymbol{x})\right)\boldsymbol{M}(\boldsymbol{x})^{-1}\right\}^{\ell}$$

$$\times\boldsymbol{a}_i(\boldsymbol{x})\left\{\prod_{r=1}^{d}K_{r,h_r}(X_{ir}-x_r)\right\}Y_i^{*}(\boldsymbol{x})q_{-j}(\boldsymbol{x}_{-j})d\boldsymbol{x}_{-j} \quad (4.3)$$

$$=O_p\left\{\rho^{\ell}\left(\sum_{r=1}^{d}h_r^2+n^{-1/2}h_j^{-1/2}\right)\right\},$$

$$\frac{1}{n}\sum_{i=1}^{n}\int R_k(\boldsymbol{x})\boldsymbol{a}_i(\boldsymbol{x})\left\{\prod_{r=1}^{d}K_{r,h_r}(X_{ir}-x_r)\right\}Y_i^{*}(\boldsymbol{x})q_{-j}(\boldsymbol{x}_{-j})d\boldsymbol{x}_{-j} \quad (4.4)$$

$$=O_p\left(\rho^{k+1}\right).$$

Taking k large enough so that $\rho^{k+1}=o(n^{-1/2}h_j^{-1/2})$ and using (4.1)–(4.4) conclude the proof of Theorem 2.1.

4.2. Proof of Theorem 2.2

We calculate the mean and variance of $T_j(x_j)$. For $\boldsymbol{x}=(x_1,\ldots,x_d)^T$, let $I_{\boldsymbol{x}}=I_1\times\cdots\times I_d$ where $I_r=[-x_r/h_r,(1-x_r)/h_r]$. Then, we have

$$E\{T_j(x_j)\}=\frac{1}{2}\iint_{I_{\boldsymbol{x}}}\left\{\frac{f(x_1+h_1u_1,\ldots,x_d+h_du_d)}{f(\boldsymbol{x})}\right\}\boldsymbol{e}_1{}^T\boldsymbol{M}^{-1}(\boldsymbol{x})\begin{pmatrix}1\\ \boldsymbol{u}\end{pmatrix} \quad (4.5)$$

$$\times\boldsymbol{u}^T\mathbf{H}\left\{\nabla^2m(\boldsymbol{x})\right\}\mathbf{H}\boldsymbol{u}\left\{\prod_{r=1}^{d}K_r(u_r)\right\}q_{-j}(\boldsymbol{x}_{-j})d\boldsymbol{u}d\boldsymbol{x}_{-j}\{1+o(1)\},$$

where $\mathbf{H}$ is the diagonal matrix with h_r's being its diagonal entries, and $\nabla^2m(\boldsymbol{x})$ is the $d\times d$ matrix with $\partial^2m(\boldsymbol{x})/\partial x_r\partial x_s$ being its $(r,s)^{th}$ entry.

Now, because of the integration with respect to $\boldsymbol{x}_{-j}$, the integrals over I_r's $(r\neq j)$ in (4.5) and those for $\boldsymbol{M}(\boldsymbol{x})$ therein may be replaced by the complete integrals over $[-1,1]$. Let $\boldsymbol{M}_j(x_j)$ denote the matrix $\boldsymbol{M}(\boldsymbol{x})$ with the integrals over I_r's $(r\neq j)$ being replaced by the complete integrals over $[-1,1]$. Then, by symmetry of K_r's it is easy to see

$$\boldsymbol{e}_1{}^T\boldsymbol{M}_j^{-1}(x_j)\begin{pmatrix}1\\ \boldsymbol{u}\end{pmatrix}K_j(u_j)=K_j^{*}(u_j,x_j). \quad (4.6)$$

By symmetry of K_r's again, it follows from (4.5) and (4.6) that

$$E\{T_j(x_j)\} = \frac{1}{2}\Big\{b_j(x_j)h_j^2 + \sum_{r\neq j} b_{r,j}(x_j)h_r^2\Big\} + o\Big(\sum_{r=1}^{d} h_r^2\Big). \tag{4.7}$$

Next, we compute the variance. We note that

$$\text{Var}\{E\,(T_j(x_j)|\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n)\,\} = O\Big\{n^{-1}h_j^{-1}\Big(\sum_{r=1}^{d} h_r^2\Big)\Big\} = o(n^{-1}h_j^{-1}). \tag{4.8}$$

Now, define $K_{j,h_j}^*(u,x_j) = K_j^*(u/h_j,x_j)/h_j$. Then, by the equation (4.6)

$$\begin{aligned} &E\Big\{\text{Var}\Big(T_j(x_j)|\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n\Big)\Big\} \\ &= \frac{1}{n}E\Big[\Big\{\int\frac{1}{f(\boldsymbol{x})}\boldsymbol{e}_1{}^T\boldsymbol{M}^{-1}(\boldsymbol{x})\boldsymbol{a}_1(\boldsymbol{x})\prod_{r=1}^{d}K_{r,h_r}(X_{1r}-x_r)q_{-j}(\boldsymbol{x}_{-j})d\boldsymbol{x}_{-j}\Big\}^2\sigma^2(\boldsymbol{X}_1)\Big] \\ &= \frac{1}{n}E\Big[\Big\{K_{j,h_j}^*(X_{1j}-x_j,x_j)\int\frac{1}{f(\boldsymbol{x})}\prod_{r\neq j}^{d}K_{r,h_r}(X_{1r}-x_r)q_{-j}(\boldsymbol{x}_{-j})d\boldsymbol{x}_{-j}\Big\}^2\sigma^2(\boldsymbol{X}_1)\Big], \end{aligned}$$

which equals $n^{-1}h_j^{-1}s_j^2(x_j)\{1+o(1)\}$. This and (4.8) yield the variance formula for $\widehat{\alpha}_j(x_j)$. Asymptotic normality follows by a standard technique, which together with (4.7) completes the proof of the theorem.

4.3. Proof of Theorem 2.4

We only prove the first part. The second part is immediate from the first. We first note that the covariance of the conditional expectations, $E\{T_j(x_j)|\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n\}$ and $E\,\{T_{j'}(x_{j'})|\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n\}$, has an order of magnitude $o(n^{-1}h^{-1})$ by (4.8). Now, we compute the expectation of the conditional covariance. Write

$$w_{ij}(x_j) = \int\frac{1}{f(\boldsymbol{x})}\boldsymbol{e}_1{}^T\boldsymbol{M}^{-1}(\boldsymbol{x})\boldsymbol{a}_i(\boldsymbol{x})\prod_{r=1}^{d}K_{r,h_r}(X_{ir}-x_r)q_{-j}(\boldsymbol{x}_{-j})d\boldsymbol{x}_{-j}.$$

Then, we may write

$$E\,\{\text{Cov}\,(T_j(x_j),T_{j'}(x_{j'})\,|\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n)\} = \frac{1}{n}E\,\{w_{1j}(x_j)w_{1j'}(x_{j'})\sigma^2(\boldsymbol{X}_1)\}\,. \tag{4.9}$$

By (4.6), we may approximate $w_{1j}(x_j)$ by

$$\frac{q_{-j}(\boldsymbol{X}_{-j})}{f(X_1,\ldots,X_{j-1},x_j,X_{j+1},\ldots,X_d)}K_{j,h_j}^*(X_{1j}-x_j).$$

Plugging this into (4.9) and using the fact that the integral of $K_j^*(u, x_j)$ over $u \in [-x_j/h_j,\ (1-x_j)/h_j]$ equals 1, we may conclude that the right hand side of (4.9) equals

$$\frac{1}{n}\int \frac{\sigma^2(\boldsymbol{x})}{f(\boldsymbol{x})} q_{-j}(\boldsymbol{x}_{-j}) q_{-j'}(\boldsymbol{x}_{-j'}) d\boldsymbol{x}_{-(j,j')}\{1+o(1)\} = o(n^{-1}h^{-1}),$$

where $\boldsymbol{x}_{-(j,j')}$ is the vector formed from $\boldsymbol{x}$ with x_j and $x_{j'}$ being deleted.

REFERENCES

BUJA, A., HASTIE, T. AND TIBSHIRANI, R. (1989). "Linear smoothers and additive models (with discussion)", *The Annals of Statistics*, **17**, 453–555.

FAN, J. AND GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*, Chapman & Hall, London.

FRIEDMAN, J. H. AND STUETZLE, W. (1981). "Projection pursuit regression", *Journal of the American Statistical Association*, **76**, 817–823.

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

LINTON, O. B. AND NIELSEN, J. P. (1995). "A kernel method of estimating structured nonparametric regression based on marginal integration", *Biometrika*, **82**, 93–100.

MAMMEN, E., LINTON, O. AND NIELSEN, J. (1999). "The existence and asymptotic properties of a backfitting projection algorithm under weak conditions", *The Annals of Statistics*, **27**, 1442–1490.

OPSOMER, J. D. (2000). "Asymptotic properties of backfitting estimators", *Journal of Multivariate Analysis*, **73**, 166–179.

OPSOMER, J. D. AND RUPPERT, D. (1997). "Fitting a bivariate additive model by local polynomial regression", *The Annals of Statistics*, **25**, 186–211.

SEVERANCE-LOSSIN, E. AND SPERLICH, S. (1999). "Estimation of derivatives for additive separable models", *Statistics*, **33**, 241–265.

SPERLICH, S., LINTON, O. AND HÄRDLE, W. (1999). "Integration and backfitting methods in additive models-finite sample properties and comparisons", *Test*, **8**, 419–458.