Journal of the Korean
Statistical Society
Vol. 25, No. 4, 1996

Kernel Regression Estimation for Permutation Fixed Design Additive Models†

Jangsun Baek[1] and Thomas E. Wehrly[2]

Abstract

Consider an additive regression model of Y on $X = (X_1, X_2, \cdots, X_p)$, $Y = \sum_{j=1}^{p} f_j(X_j) + \varepsilon$, where f_js are smooth functions to be estimated and ε is a random error. If X_js are fixed design points, we call it the fixed design additive model. Since the response variable Y is observed at fixed p-dimensional design points, the behavior of the nonparametric regression estimator depends on the design. We propose a fixed design called permutation fixed design, and fit the regression function by the kernel method. The estimator in the permutation fixed design achieves the univariate optimal rate of convergence in mean squared error for any $p \geq 2$.

Key Words : Nonparametric regression estimation; Fixed design additive model; Optimal rate of convergence.

†This research was supported by a grant No. KOSEF 951-0103-025-1 from the Korea Science and Engineering Foundation.

[1]Department of Statistics, Chonnam National University, Kwangju, 500-757, Korea.
[2]Department of Statistics, Texas A&M University, College Station, TX 77843, U.S.A.

1. INTRODUCTION

If we have n bivariate observations $\{(X_i, Y_i)\}_{i=1}^n$, the regression relationship can be modeled as

$$Y_i = m(X_i) + \varepsilon_i, \ i = 1, 2, \cdots, n,$$

with the unknown regression function m and uncorrelated observation errors ε_i. We can measure the response Y_i at X_i, a value of controlled variable which is fixed in advance (fixed design). In this case m is the mean function of the response variable with the nonstochastic predictor variable. The approximation of m by local average is called nonparametric estimation of the regression function. Much work has been done on nonparametric estimation of regression functions with one predictor using various nonparametric smoothers (Eubank 1988, Müller 1988, Härdle 1990, and Wand and Jones 1995).

When we are interested in the estimation of a regression surface with more than one predictor, $m(X_1, \cdots, X_p)$, the basic idea of univariate nonparametric smoothing can be extended to higher demensions by using a p-dimensional multivariate smoother (Wahba 1979, Müller 1988, and Georgiev 1989, 1990). The multivariate local averaging procedure still gives asymptotically consistent estimators to the regression surface. However, there are two major problems with this approach. First, the optimal rate of mean squared error (MSE) convergence of the multivariate estimator is $n^{-2l/(2l+p)}$, where l is an index of smoothness of the regression surface, so that it tends to be very slow in high demension. This is often called the *curse of dimensionality* since a regression surface $m(X_1, \cdots, X_p)$ will require large data sets for even moderate p. Second, estimates are difficult to interpret for $p > 3$.

One way of avoiding these problems is to impose an additive structure on the regression function. More precisely, the regression function takes the form $m(X_1, \cdots, X_p) = \alpha + \sum_{j=1}^p f_j(X_j)$ in the additive model. Friedman and Stuetzle(1981), Breiman and Friedman(1985), Buja, Hastie and Tibshirani(1989), and Hastie and Tibshirani(1990) proposed various procedures to estimate an additive regression function. Stone(1985) studied rates of convergence for additive models with the functions estimated by regression splines. He proved that the optimal rate of convergence for an estimate of the additive model is the same as that for a one-dimensional function. Thus an increase in the dimension p does not decrese the rate of convergence, as it does if one is estimating a general (nonadditive) p-dimensional function. Härdle and Tsybakov(1990) investigated a kernel estimator in the random design additive models, and established the asymptotic distrubution of the estimator.

In medical or experimental research studying the response at different points of the controlled variables, the levels of the controlled variables are fixed in advance by the experimenter resulting in a fixed design. In the fixed design additive model, the behavior of the regression estimator will depend on the experimenter's fixed design for the controlled variables. We propose a fixed design which is called the permutation fixed design, where we observe the response variable at design points containing randomly chosen levels for each controlled variables. The asymptotic behavior of kernel estimator for this fixed design additive model will be studied in this article.

We conclude the introduction with an outline of the rest of the article. In section 2, we define the permutation fixed design. It is basically an imitation of random design. We use the Gasser-Müller estimator on all the Ys to estimate the functions of the controlled variables. The estimator attains the univariate optimal rate of convergence for any $p \geq 2$. We observe that kernel estimates perform well for a 2-dimensional fixed design by simulation in section 3.

2. PERMUTATION FIXED DESIGN ADDITIVE MODELS

Suppose that we have a response variable Y, p controlled variables (or factors) $X_1, X_2, \cdots, X_p$, and X_j has n_j equally spaced design points, $j = 1, 2, \cdots, p$. The additive model

$$Y = \alpha + \sum_{j=1}^{p} f_j(X_j) + \varepsilon,$$

where $\alpha = E(Y)$ at $(X_1, \cdots, X_p) = (x_1, \cdots, x_p)$ with $f_j(x_j) = 0$, $j = 1, \cdots, p$ is assumed. Suppose that the kernel K and the marginal regression functions f_js satisfy the following conditions for the fixed design.

($A1$) The jth controlled variable X_j takes its levels $X_{i_j,j} = i_j/n_j$, $i_j = 1, 2, \cdots, n_j$, $j = 1, 2, \cdots, p$. Thus the support of X_j is $S = [0, 1]$.

($A2$) K is bounded, symmetric, nonnegative on the support $[-1, 1]$, and Hölder continuous, *i.e.*, $|K(x) - K(y)| \leq c|x - y|$ for all x and y, and some constant c.

($A3$) $\int K(u)du = 1$, $\int u^j K(u)du = 0$, $j = 1, 2, \cdots, l-1$, and $\int u^l K(u)du < \infty$.

(*A*4) The function $f_j(\cdot)$ is bounded, l times continuously differentiable, and lth derivative $f_j^{(l)}(\cdot)$ is Hölder continuous such that $|f_j^{(l)}(u) - f_j^{(l)}(v)| \leq M|u-v|^\beta$, where $0 < \beta \leq 1$, $j = 1, \cdots, p$.

(*A*5) $\int_S f_j(t)dt = 0$, where S is the support of X_j, $j = 1, 2, \cdots, p$.

Condition (*A*5) is needed for the identifiability of the regression function. Otherwise there will be free constants in each marginal regression function f_j. Since we can estimate the constant α with the average of the response observations ($\bar{Y}$) due to (*A*5), we omit the constant term in the additive model from now on.

Suppose that we have p controlled variables (or factors) and each controlled variable has n equally spaced design points (or levels) $\{X_{1j}, \cdots, X_{nj}\}$ where $X_{1j} < \cdots < X_{nj}$, $j = 1, \cdots, p$. Consider the n-tuple of numbers, $(i_1, i_2, \cdots, i_n)$ where $i_j \in N = \{1, 2, \cdots, n\}$ and $i_j \neq i_k$ for different j and k, $j = 1, 2, \cdots, n$, $k = 1, 2, \cdots, n$. Since there are $n!$ permutations of the set N, we have $n!$ different n-tuples, $(i_1, i_2, \cdots, i_n)$. We pick one permutatoin up randomly for each controlled variable, and use it as indices of the levels of that controlled variable to assign to the experimental units. Thus we can construct a design with n points such that kth design point consists of the kth elements of the p permuted n-tuples, as its coordinate indices, $k = 1, 2, \cdots, n$, and no two design points have equal X_{ij} values on any marginal j. Thanks to this permutation design, we observe the response variable at all n different values of each marginal X_j variable.

Now assume we have p randomly selected n-tuples $\{(i_{11}, i_{21}, \cdots, i_{n1}), (i_{12}, i_{22}, \cdots, i_{n2}), \cdots, (i_{1p}, i_{2p}, \cdots, i_{np})\}$. We define $\left\{(X_{i_{k1}1}, X_{i_{k2}2}, \cdots, X_{i_{kp},p});\ k = 1, 2, \cdots, n\right\}$ as the set of p-dimensional design points at which we observe a response variable. Then, the regression model in the permutation fixed design is

$$Y_{i_{k1}, i_{k2}, \cdots, i_{kp}} = \sum_{j=1}^{p} f_j(X_{i_{kj}j}) + \varepsilon_{i_{k1}, i_{k2}, \cdots, i_{kp}},$$

where the f_js are smooth functions and $\varepsilon_{i_{k1}, i_{k2}, \cdots, i_{kp}}$s are independent random variables with mean 0 and variance σ^2.

Consider estimation of the regression function $m(x) = \sum_{j=1}^{p} f_j(x_j)$ at a fixed point $x = (x_1, x_2, \cdots, x_p)$. The regression function $m(x)$ is estimated by the Gasser-Müller estimator by smoothing observations on the design points of each controlled variable. Though the local linear kernel estimator (Fan 1992) is popular, we use the Gasser-Müller estimator for its simplicity in the proof of consistency and for having the same asymptotic bias and variance as of the local linear estimator at the interior points in fixed design. Before describing the estimation procedure, we need to arrange the data so that it is possible.

For the jth controlled variable, $\{(X_{i_k j j}); k = 1, \cdots, n\}$ is rearranged in increasing order and call it $\{(X_{ij})\}_{i=1}^{n}$. We need also to arrange the observations of the response variable corresponding to $\{(X_{ij})\}_{i=1}^{n}$, and call it $\{(Y_{ij})\}_{i=1}^{n}$. Thus we have p sets of data $\{\{(Y_{ij}, X_{ij})\}_{i=1}^{n}; j = 1, 2, \cdots, p\}$ which are used to estimate the f_js. Note that $\{(Y_{ij}, X_{ij})\}_{i=1}^{n}$ has the same Y values for $j = 1, 2, \cdots, p$, but they are in a different order.

The estimate $\hat{m}(x)$ at $x = (x_1, x_2, \cdots, x_p)$ is formed as follows :

$$\hat{m}(x) = \sum_{j=1}^{p} \hat{f}_j(x_j), \ x_j \in [0,1], \ j = 1, 2, \cdots, p, \tag{2.1}$$

with

$$\hat{f}_j(x_j) = \frac{1}{\lambda_n} \sum_{i=1}^{n} Y_{ij} \int_{s_{i-1,j}}^{s_{i,j}} K\left(\frac{x_j - s}{\lambda_n}\right) ds, \ s_{i-1,j} \leq X_{ij} \leq s_{i,j},$$

where the bandwidth $\lambda_n > 0$. Here also we use the same bandwidth λ_n to estimate each function f_j. The asymptotic properties of the mean and variance of the estimator are derived under the assumptions $(A1)$ through $(A5)$. We will examine the rate of convergence of the estimator in a mean squared error sense.

The following theorem 1 insures the optimal rate of convergence of $\hat{m}(x)$ defined in (2.1) which will be shown in the Corollary 1. The proofs of the following theorem 1 and a corollary 1 are in the Appendix.

Theorem 1. Assume $X_{1j}, \cdots, X_{nj}$ are equally spaced on $[0,1]$ for $j = 1, \cdots, p$. For $x = (x_1, x_2, \cdots, x_p)$ such that $x_j \in [\lambda_n, 1 - \lambda_n]$, $j = 1, 2, \cdots, p$,

$$E(\hat{m}(x)) = \sum_{j=1}^{p} f_j(x_j) + \frac{\lambda_n^l}{l!} \int u^l K(u) du \sum_{j=1}^{p} f_j^{(l)}(x_j) + o(\lambda_n^l) + O\left(\frac{1}{n}\right), \tag{2.2}$$

$$\operatorname{Var}(\hat{m}(x)) = \frac{(p\sigma^2 + (p-1)\sum_{j=1}^{p} D_j)}{n\lambda_n} \int_{-1}^{1} K^2(u) du + O\left(\frac{1}{(n\lambda_n)^2}\right) + O\left(\frac{1}{n}\right), \tag{2.3}$$

where $D_j = \int_0^1 f_j^2(t)dt$, $j = 1, 2, \cdots, p$.

Now we explore the MSE of $\hat{m}(x)$ of (2.1) for the permutation fixed design in the following corollary.

Corollary 1. Assume (A2)-(A5), and $X_{1j}, \cdots, X_{nj}$ are equally spaced on [0,1] for $j = 1, 2, \cdots, p$. Then, if $n \to \infty$, and $\lambda_n \to 0$ in such a way that $n\lambda_n \to \infty$, then

$$\begin{aligned} \text{MSE}(\hat{m}(x)) \quad &\sim \quad \frac{(p\sigma^2 + (p-1)\sum_{j=1}^p D_j)}{n\lambda_n} \int_{-1}^1 K^2(u)du \\ &+ \quad \frac{\lambda_n^{2l}}{(l!)^2} \left(\int u^l K(u)du \right)^2 \left(\sum_{j=1}^p f_j^{(l)}(x_j) \right)^2, \end{aligned} \tag{2.4}$$

where D_j is defined in the Theorem 1. The asymptotically optimal bandwidth is

$$\lambda_{n(opt)} = \left\{ \frac{(l!)^2(p\sigma^2 + (p-1)\sum_{j=1}^p D_j)\int_{-1}^1 K^2(u)du}{2nl(\int_{-1}^1 u^l K(u)du)^2(\sum_{j=1}^p f_j^{(l)}(x_j))^2} \right\}^{1/(2l+1)}, \tag{2.5}$$

where $\sum_{j=1}^p f_j^{(l)}(x_j) \neq 0$, and

$$\begin{aligned} &\text{MSE}(\hat{m}(x); \lambda_{n(opt)}) \\ &\sim (2l+1)\left\{(l!)^2(2l)^{2l}\right\}^{-1/(2l+1)} \left\{ \left((p\sigma^2 + (p-1)\sum_{j=1}^p D_j) \int K^2(u)du \right)^{2l} \right. \\ &\left. \cdot \left(\int u^l K(u)du \right)^2 \left(\sum_{j=1}^p f_j^{(l)}(x_j) \right)^2 \right\}^{1/(2l+1)} n^{-2l/(2l+1)}. \end{aligned} \tag{2.6}$$

The asymptotic results about the univariate regression function estimate are the special case of those of the additive one with $p = 1$. The asymptotic bias and variance of the univariate Gasser-Müller estimator are the same as the results of the Theorem 1 where $p = 1$, respectively. Therefore the asymptotic pointwise MSE of the univariate regression function estimate coincides with the results of the Corollary 1 when $p = 1$.

It follows from the Corollary 1 that the rate of convergence of $\hat{m}(x)$ is free from the *curse of dimensionality* when $p \geq 2$, since

$$\mathrm{MSE}(\hat{m}(x); \lambda_{n(opt)}) \sim O(n^{-2l/(2l+1)}),$$

the same optimal rate of convergence as for a one-dimensional function. It is due to the small remainder term $O(n^{-1})$ in the bias of the permutation design estimate as shown in the Theorem 1. The bias remainder term decreases fast enough to guarantee its estimate to attain the univariate optimal rate of MSE convergence. The asymptotic variance of the fixed design estimate has the rate $O((n\lambda_n)^{-1})$. Therefore it does not hinder the asymptotic MSE of the permutation fixed design from reaching the optimal convergence rate $O(n^{2l/(2l+1)})$.

3. SIMULATIONS

We conducted a simulation study to examine the performance of the proposed kernel estimator for a 2-dimensional additive model. The kernel function used was $K(u) = (3/4)(1 - u^2)$ for $|u| \leq 1$. The Generalized Cross Validation (GCV, Eubank 1988) was used for fast implementation to select a bandwidth. The estimate with the selected bandwidth was used to compose the regression function estimate. The random errors ε_i in the model were assumed to be independent, and follow the normal distribution with mean 0 and variance 1. We considered a 2-dimensional additive model so that it is possible to visualize the regression function estimates.

The model considered in the permutation fixed design was

$$Y = f_1(X_1) + f_2(X_2) + \varepsilon,$$

where $f_1(X_1) = (1/3)(e^{2(X_1+1)} - e^2)I_{(-1<X_1<0)}(X_1) + (1/3)(-e^{-2(X_1-1)} + e^2)$ $I_{(0\leq X_1<1)}(X_1)$, $f_2(X_2) = 4X_2^3 - 2X_2$, and the levels of X_1 and X_2 are equally spaced in the interval $[-1, 1]$. We took 400 observations at the permutation design points. Figures 1(a) and 1(b) show the estimates $\hat{f}_1$ and $\hat{f}_2$ with their true functions at 50 equally spaced points, respectively. Fugure 1(c) shows that $\hat{m}$ is close to the true regression surface m in Fugure 1(d).

So far we have examined the performance of our proposed estimator with one random sample. To check if the permutation fixed design estimator behaves well in general, for given value of n, 200 random samples were drawn from the above model. For each random sample, $\hat{f}_j(x_{ij})$ was computed at $x_{ij} = -1 + 2i/100$, $i = 1, \cdots, 100$, $j = 1, 2$. At each x_{ij}, the 200 values of

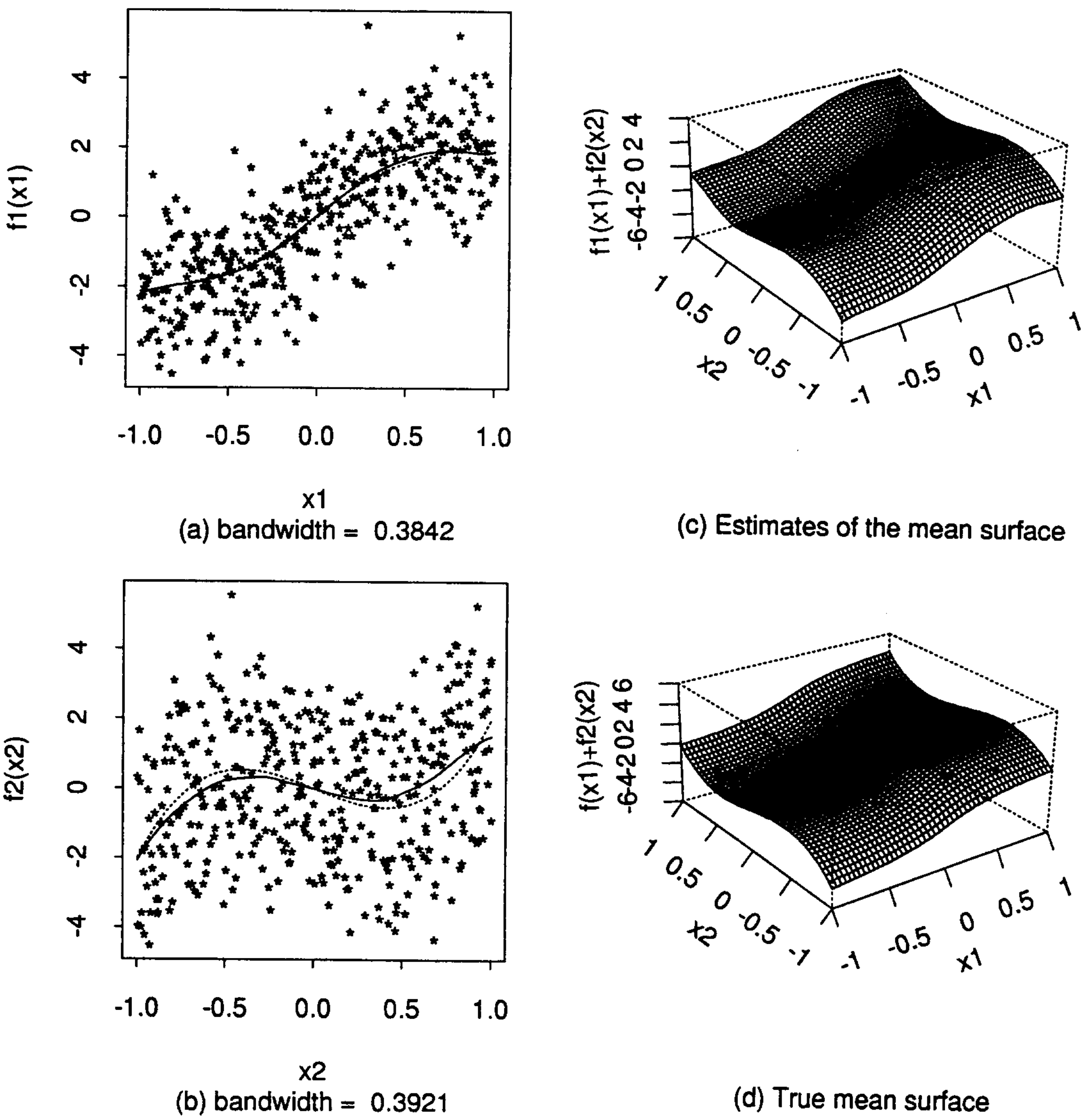

Figure 1. (a) Estimates $\hat{f}_1$ (the solid line) and the true function f_1 (the dotted line), (b) Estimates $\hat{f}_2$ (the solid line) and the true function f_2 (the dotted line), (c) Estimates $\hat{m}$, and (d) True mean surface m in the permutation fixed design. $*$ in (a), (b) are $\{Y_{i1}\}_{i=1}^{n}$ and $\{Y_{i2}\}_{i=1}^{n}$, respectively.

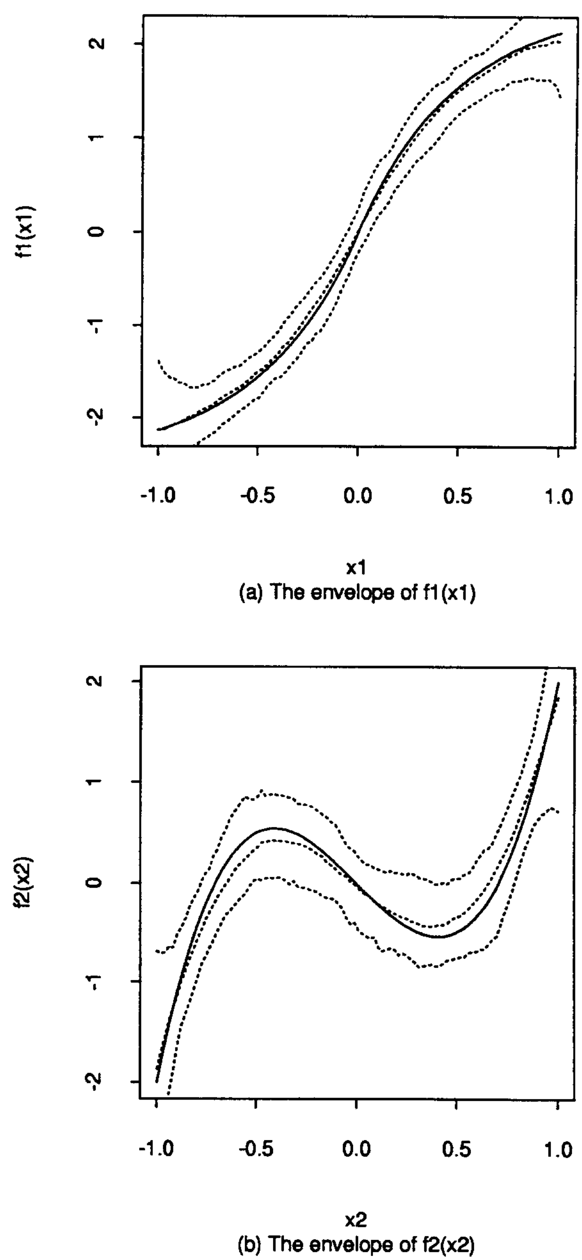

Figure 2. Envelopes of $\hat{f}_1$ and $\hat{f}_2$ in the permutation fixed design. The upper, middle, and lower dotted line is 95th, 50th, and 5th percentile, respectively. The solid line is the true function.

$\hat{f}_j(x_{ij})$ were ordered. The median, 5th percentile and 95th percentile were computed. The resulting envelopes of the estimates along with the true curves are plotted in Figure 2(a) and Figure 2(b). The medians of the estimates are very close to their true curves on the entire ranges of the controlled variables. The widths of both envelopes are constant in the interior of the intervals. Figure 2 gives evidence of the good performance of the proposed estimator in the interior.

4. CONCLUDING REMARKS

The objective of this research was to develop kernel estimaton procedures for additive regression models with fixed design. Since the major attraction of the additive model is the achievability of the univariate optimal rate of convergence, the optimal rate of convergence of the estimators is important. We proposed a fixed design called permutation fixed design where the kernel estimator of the additive mean function attains the optimal rate of convergence. We used the Gasser-Müller estimator for the regression model. The estimator of the additive mean function $\sum_{j=1}^{p} f_j(X_j)$ was defined as a sum of the estimators $\hat{f}_j$, and $\hat{f}_j$ was constructed by smoothing response observations. The kernel estimator in the permutation fixed design attains the optimal rate of convergence for any $p \geq 1$.

5. APPENDIX: PROOFS

Proof of Theorem 1.

For $k \neq j$, let l_{ki} be the index at which $X_{l_{ki}k}$ is the level of kth controlled variable when Y_{ij} is observed. Then we can express the regression model as follows :

$$Y_{ij} = f_j(X_{ij}) + \sum_{\substack{k=1 \\ k \neq j}}^{p} f_k(X_{l_{ki}k}) + \varepsilon_{ij},\ i = 1, \cdots, n,\ j = 1, \cdots, p. \tag{5.1}$$

Note that $X_{l_{ki}k}$ is a discrete uniform random variable when $X_j = X_{ij}$. For $i = 1, 2, \cdots, n$, when $X_j = X_{ij}$ the kth controlled variable level $X_{l_{ki}k}$ takes any value with the equal probability from the remaining $(n-i+1)$ fixed levels after $(i-1)$ levels of X_k were taken for $X_j = X_{1j}, \cdots, X_{i-1,j}$ by the permutation design procedure. Therefore $\{X_{l_{ki}k}\}_{i=1}^{n}$ is a random sample without replacement from a finite population $\{X_{1k}, X_{2k}, \cdots, X_{nk}\}$, and is independent of $\{X_{1j}, X_{2j}, \cdots, X_{nj}\}$. Given $(X_{1j}, X_{2j}, \cdots, X_{nj})$,

$$\begin{aligned} E(\hat{f}_j(x_j)) &= \frac{1}{\lambda_n}\sum_{i=1}^{n} f_j(X_{ij}) \int_{s_{i-1,j}}^{s_{i,j}} K\Big(\frac{x_j - s}{\lambda_n}\Big) ds \\ &+ \sum_{\substack{k=1 \\ k\neq j}}^{p} \frac{1}{\lambda_n}\sum_{i=1}^{n} E(f_k(X_{l_{ki}k})) \int_{s_{i-1,j}}^{s_{i,j}} K\Big(\frac{x_j - s}{\lambda_n}\Big) ds, \; s_{i-1,j} \le X_{ij} \le s_{i,j}. \end{aligned} \tag{5.2}$$

Using the Mean Value Theorem and the Taylor's expansion we can show the first term of (5.2),

$$\begin{aligned} \frac{1}{\lambda_n}\sum_{i=1}^{n} f_j(X_{ij}) \int_{s_{i-1,j}}^{s_{i,j}} K\Big(\frac{x_j - s}{\lambda_n}\Big) ds &= f_j(x_j) + \frac{{\lambda_n}^l}{l!} f_j^{(l)}(x_j) \int u^l K(u) du \\ &+ o({\lambda_n}^l) + O\Big(\frac{1}{n}\Big). \end{aligned} \tag{5.3}$$

Now consider the second term of (5.2). We evaluate $E(f_k(X_{l_{ki}k}))$ first. For $i = 1$, $E(f_k(X_{l_{k1}k})) = (1/n)\sum_{l=1}^{n} f_k(X_{lk})$ since $X_{l_{k1}k}$ can take any level with the same probability $1/n$ from $\{X_{1k}, X_{2k}, \cdots, X_{nk}\}$. For $1 < i \le n$, however, the index l_{ki} depends on the set of indices $L_{k(i-1)} = \{l_{k1}, l_{k2}, \cdots, l_{k(i-1)}\}$ that was chosen before by the permutation design procedure.

$$\begin{aligned} E(f_k(X_{l_{ki}k})) &= E(E(f_k(X_{l_{ki}k}) | (X_{l_{k1}k}, X_{l_{k2}k}, \cdots, X_{l_{k(i-1)}k}))) \\ &= E\Big(\frac{1}{n-i+1} \sum_{l \in N \setminus L_{k(i-1)}} f_k(X_{lk})\Big), \text{ where } N = \{1, 2, \cdots, n\}, \\ &= \frac{1}{\binom{n}{i-1}} \sum_{L_{k(i-1)}} \Big(\frac{1}{n-i+1} \sum_{l \in N \setminus L_{k(i-1)}} f_k(X_{lk})\Big) \\ &= \frac{\binom{n-1}{i-1} \sum_{l=1}^{n} f_k(X_{lk})}{\binom{n}{i-1}(n-i+1)} = \frac{1}{n}\sum_{l=1}^{n} f_k(X_{lk}). \end{aligned}$$

The second equation to the last is obtained as follows: For $m = 1, 2, \cdots, n$, $f_k(X_{mk})$ appears $\binom{n-1}{i-1}$ times in $\sum_{L_{k(i-1)}} \sum_{l \in N \setminus L_{k(i-1)}}$ since we have $f_k(X_{mk})$ in the double summation as many times as we choose $L_{k(i-1)}$ from $N \setminus \{m\}$.

Since $E(f_k(X_{l_{ki}k})) = (1/n)\sum_{l=1}^{n} f_k(X_{lk}),\ i = 1, 2, \cdots, n$, it is easy to show that the second term of (5.2) is $O(1/n)$. Therefore from (5.3) and the last argument,

$$E(\hat{f}_j(x_j)) = f_j(x_j) + \frac{\lambda_n{}^l}{l!} f_j^{(l)}(x_j) \int u^l K(u)du + o(\lambda_n{}^l) + O\Big(\frac{1}{n}\Big).$$

Since $E(\hat{m}(x)) = \sum_{j=1}^{p} E(\hat{f}_j(x_j))$, (2.2) is immediate.

Now we prove (2.3). For notational convenience we use $\sum_{j\neq k}^{p}$ to indicate $\sum_{j=1}^{p}\sum_{k=1}^{p}$ throughout the proof. $_{j\neq k}$

$$\mathrm{Var}(\hat{m}(x)) = \sum_{j=1}^{p} \mathrm{Var}(\hat{f}_j(x_j)) + \sum_{j\neq k}^{p} Cov(\hat{f}_j(x_j), \hat{f}_k(x_k)). \tag{5.4}$$

Referring to (5.1), let $X_{(-j)} = \{(X_{l_{k1}k}, X_{l_{k2}k}, \cdots, X_{l_{kn}k}); k = 1, \cdots, j-1, j+1, \cdots, p\},\ j = 1, 2, \cdots, p$. Then

$$\mathrm{Var}(\hat{f}_j(x_j)) = E\{\mathrm{Var}(\hat{f}_j(x_j)|X_{(-j)})\} + \mathrm{Var}\{E(\hat{f}_j(x_j)|X_{(-j)})\}. \tag{5.5}$$

Note

$$\begin{aligned} E\{\mathrm{Var}(\hat{f}_j(x_j)|X_{(-j)})\} &= \frac{1}{\lambda_n^2}\sum_{i=1}^{n}\Big(\int_{s_{i-1,j}}^{s_{i,j}} K\Big(\frac{x_j - s}{\lambda_n}\Big)ds\Big)^2 \mathrm{Var}(Y_{ij}|X_{(-j)}) \\ &= \frac{\sigma^2}{\lambda_n^2}\sum_{i=1}^{n}\Big(\int_{s_{i-1,j}}^{s_{i,j}} K\Big(\frac{x_j - s}{\lambda_n}\Big)ds\Big)^2, \end{aligned}$$

since Y_{ij}s are independent for given $X_{(-j)},\ i = 1, 2, \cdots, n$. Assume $s_{i,j} = (X_{ij} + X_{i+1,j})/2,\ i = 1, 2, \cdots, n$.

By the Mean Value Theorem and Hölder's continuity of $K(\cdot)$, it is easy to show that

$$\begin{aligned} \sum_{i=1}^{n}\Big(\int_{s_{i-1,j}}^{s_{i,j}} K\Big(\frac{x_j - s}{\lambda_n}\Big)ds\Big)^2 &= \frac{1}{n}\int_0^1 K\Big(\frac{x_j - s}{\lambda_n}\Big)^2 ds + O\Big(\frac{1}{n^2}\Big) \\ &= \frac{\lambda_n}{n}\int_{-1}^{1} K^2(u)du + O\Big(\frac{1}{n^2}\Big), \end{aligned}$$

for $x_j \in [\lambda_n, 1 - \lambda_n]$. Hence

$$E\{\mathrm{Var}(\hat{f}_j(x_j)|X_{(-j)})\} = \frac{\sigma^2}{n\lambda_n}\int_{-1}^{1} K^2(u)du + O\Big(\frac{1}{(n\lambda_n)^2}\Big). \tag{5.6}$$

Since

$$E(\hat{f}_j(x_j)|X_{(-j)}) = \frac{1}{\lambda_n}\sum_{i=1}^{n} f_j(X_{ij})\int_{s_{i-1,j}}^{s_{i,j}} K\Big(\frac{x_j - s}{\lambda_n}\Big)ds + \sum_{\substack{k=1 \\ k\neq j}}^{p} \frac{1}{\lambda_n}\sum_{i=1}^{n} f_k(X_{l_{ki}k})\int_{s_{i-1,j}}^{s_{i,j}} K\Big(\frac{x_j - s}{\lambda_n}\Big)ds,$$

it is easy to see that

$$\begin{aligned}\mathrm{Var}\{E(\hat{f}_j(x_j)|X_{(-j)})\} &= \sum_{\substack{k=1 \\ k\neq j}}^{p}\Big[\frac{1}{\lambda_n^2}\mathrm{Var}\Big\{\sum_{i=1}^{n} f_k(X_{l_{ki}k})\int_{s_{i-1,j}}^{s_{i,j}} K\Big(\frac{x_j - s}{\lambda_n}\Big)ds\Big\}\Big] \\ &= \sum_{\substack{k=1 \\ k\neq j}}^{p}\frac{1}{\lambda_n^2}\Big[\sum_{i=1}^{n}\Big(\int_{s_{i-1,j}}^{s_{i,j}} K\Big(\frac{x_j - s}{\lambda_n}\Big)ds\Big)^2 \mathrm{Var}(f_k(X_{l_{ki}k})) \\ &\quad + \sum_{i\neq m}^{n}\Big(\int_{s_{i-1,j}}^{s_{i,j}} K\Big(\frac{x_j - s}{\lambda_n}\Big)ds\Big)\Big(\int_{s_{m-1,j}}^{s_{m,j}} K\Big(\frac{x_j - s}{\lambda_n}\Big)ds\Big) \\ &\quad \cdot\mathrm{Cov}(f_k(X_{l_{ki}k}), f_k(X_{l_{km}k}))\Big]. \qquad (5.7)\end{aligned}$$

Following the similar steps to evaluate $E(f_k(X_{l_{ki}k}))$, we can show that for $i = 1, 2, \cdots, n$,

$$\begin{aligned}\mathrm{Var}(f_k(X_{l_{ki}k})) &= \frac{1}{n}\sum_{l=1}^{n} f_k^2(X_{lk}) - \Big(\frac{1}{n}\sum_{l=1}^{n} f_k(X_{lk})\Big)^2 \\ &= \int_0^1 f_k^2(t)dt + O\Big(\frac{1}{n}\Big) - O\Big(\frac{1}{n^2}\Big) = D_k + O\Big(\frac{1}{n}\Big),\end{aligned}$$

where $D_k = \int_0^1 f_k^2(t)dt$. The first term in the bracket of (5.7) becomes $(\lambda_n D_k/n)\int_{-1}^1 K^2(u)du + O(1/n^2)$. For any i and $m\,(i \neq m)$, we can see by applying standard technique $|\mathrm{Cov}(f_k(X_{l_{ki}k}), f_k(X_{l_{km}k}))| = O(1/n)$, and

$$\Big|\sum_{i\neq m}^{n}\Big(\int_{s_{i-1,j}}^{s_{i,j}} K\Big(\frac{x_j - s}{\lambda_n}\Big)ds\Big)\Big(\int_{s_{m-1,j}}^{s_{m,j}} K\Big(\frac{x_j - s}{\lambda_n}\Big)ds\Big)\mathrm{Cov}(f_k(X_{l_{ki}k}), f_k(X_{l_{km}k}))\Big| = O\Big(\frac{\lambda_n^2}{n}\Big).$$

Hence

$$\mathrm{Var}\{E(\hat{f}_j(x_j)|X_{(-j)})\} = \frac{\sum_{k=1, k\neq j}^{p} D_k}{n\lambda_n}\int_{-1}^{1} K^2(u)du + O\Big(\frac{1}{(n\lambda_n)^2}\Big) + O\Big(\frac{1}{n}\Big). \quad (5.8)$$

Now consider the covariance term in (5.4). For $j \neq k$,

$$\begin{aligned}
&\mathrm{Cov}(\hat{f}_j(x_j), \hat{f}_k(x_k)) \\
&= E\Big[E\Big\{\Big(\frac{1}{\lambda_n}\sum_{i=1}^{n}\int_{s_{i-1,j}}^{s_{i,j}} K\Big(\frac{x_j - s}{\lambda_n}\Big)ds\,(Y_{ij} - E(Y_{ij}))\Big) \\
&\quad \cdot\Big(\frac{1}{\lambda_n}\sum_{l=1}^{n}\int_{s_{l-1,k}}^{s_{l,k}} K\Big(\frac{x_k - s}{\lambda_n}\Big)ds\,(Y_{lk} - E(Y_{lk}))\Big)\Big|X_{(-j)}\Big\}\Big] \\
&= E\Big\{\frac{1}{\lambda_n^2}\sum_{i=1}^{n}\sum_{l=1}^{n}\Big(\int_{s_{i-1,j}}^{s_{i,j}} K\Big(\frac{x_j - s}{\lambda_n}\Big)ds\Big)\Big(\int_{s_{l-1,k}}^{s_{l,k}} K\Big(\frac{x_k - s}{\lambda_n}\Big)ds\Big) \\
&\quad \cdot E(\varepsilon_{ij}\varepsilon_{lk})\Big\}, \qquad (5.9)
\end{aligned}$$

where $\varepsilon_{ij}, \varepsilon_{lk}$ are the random errors corresponding to Y_{ij} and Y_{lk}, respectively. Since $\{(\varepsilon_{ij})\}_{i=1}^{n}$ and $\{(\varepsilon_{lk})\}_{l=1}^{n}$ have the same elements, there exists one random error which is the same as ε_{ij} in $\{(\varepsilon_{lk})\}_{l=1}^{n}$. So

$$E(\varepsilon_{ij}\varepsilon_{lk}) = \begin{cases}\sigma^2, & \text{if} \quad \varepsilon_{ij} = \varepsilon_{lk} \\ 0, & \text{otherwise.}\end{cases}$$

Denote $[s_{l_i-1}, s_{l_i}]$ be the interval where $\varepsilon_{ij} = \varepsilon_{lk}$ on the support of kth controlled variable. Thus (5.9) is equal to

$$\frac{\sigma^2}{\lambda_n^2}\sum_{i=1}^{n}\Big(\int_{s_{i-1,j}}^{s_{i,j}} K\Big(\frac{x_j - s}{\lambda_n}\Big)ds\Big)E\Big(\int_{s_{l_i-1}}^{s_{l_i}} K\Big(\frac{x_k - s}{\lambda_n}\Big)ds\Big).$$

Note for $i = 1, \cdots, n$,

$$E\Big(\int_{s_{l_i-1}}^{s_{l_i}} K\Big(\frac{x_k - s}{\lambda_n}\Big)ds\Big) = \frac{1}{n}\sum_{l=1}^{n}\int_{s_{l-1,k}}^{s_{l,k}} K\Big(\frac{x_k - s}{\lambda_n}\Big)ds = \frac{\lambda_n}{n}.$$

Therefore

$$\begin{aligned}
|\mathrm{Cov}(\hat{f}_j(x_j), \hat{f}_k(x_k))| &= \Big|\frac{\sigma^2}{n\lambda_n}\sum_{i=1}^{n} K\Big(\frac{x_j - \xi_i}{\lambda_n}\Big)(s_{i,j} - s_{i-1,j})\Big| \\
&\leq \frac{\sigma^2|K|_\infty}{n^2\lambda_n}\sum_{i=1}^{n} I\Big(\frac{x_j - \xi_i}{\lambda_n} \in [-1, 1]\Big) = O\Big(\frac{1}{n}\Big),
\end{aligned}$$

where $\xi_i \in (s_{i-1,j}, s_{i,j})$. Hence

$$\mathrm{Cov}(\hat{f}_j(x_j), \hat{f}_k(x_k)) = O\left(\frac{1}{n}\right). \tag{5.10}$$

Combining the results from (5.4), (5.5), (5.6), (5.8), and (5.10), (2.3) is immediate.

Proof of Corollary 1.

Note $\mathrm{MSE}(\hat{m}(x)) = \mathrm{Var}(\hat{m}(x)) + (E(\hat{m}(x)) - m(x))^2$. Equation (2.4) is immediate from the Theorem 1. (2.5) follows from differentiation of (2.4), and (2.6) is obtained by substituting (2.5) into (2.4).

REFERENCES

(1) Breiman, L. and Friedman, J. H. (1985), Estimating Optimal Transformations for Multiple Regression and Correlation (with discussion), *Journal of the American Statistical Association*, **80**, 580-619.

(2) Buja, A., Hastie, T. and Tibshirani, R. (1989), Linear Smoothers and Additive Models (with discussion), *The Annals of Statistics*, **17**, 453-555.

(3) Eubank, R. L. (1988), *Spline Smoothing and Nonparametric Regression*, New York, Marcel Dekker.

(4) Fan, J. (1992), Design-adaptive nonparametric regression, *Journal of the American Statistical Association*, **87**, 998-1004.

(5) Friedman, J. H. and Stuetzle, W. (1981), Projection Pursuit Regression, *Journal of the American Statistical Association*, **76**, 817-823.

(6) Gasser, T. and Müller, H. G. (1979), Kernel Estimation of Regression Functions, *In Smoothing Techniques for Curve Estimation, eds.* T. Gasser and M. Rosenblatt, Heidelberg, Springer-Verlag, 23-68.

(7) Georgiev, A. A. (1989), Asymptotic Properties of the Multivariate Nadaraya-Watson Regression Function Estimate: The Fixed Design Case, *Statistics and Probability Letters*, **7**, 35-40.

(8) Georgiev, A. A. (1990), Nonparametric Multiple Function Fitting, *Statistics and Probability Letters*, **10**, 203-211.

(9) Härdle, W. (1990), *Applied Nonparametric Regression*, Econometric Society Monographs No. 19, Cambridge University Press.

(10) Härdle, W. and Tsybakov, A. B. (1990), How Many Terms Should be Added into an Additive Model?, *CORE Discussion paper N9068*, Center for Operations Research & Econometrics, Univ. Catholique de Louvain, Belgium.

(11) Hastie, T. J. and Tibshirani, R. J. (1990), *Generalized Additive Models*, Monographs on Statistics and Applied Probability 43, London, Chapman and Hall.

(12) Jones, M. C., Davies, S. J. and Park, B. U. (1994), Versions of Kernel-Type Regression Estimators, *Journal of the American Statistical Association*, **89**, 825-832.

(13) Müller, H. G. (1988), *Nonparametric Regression Analysis of Longitudinal Data*, Springer Lecture Notes in Statistics, New York, Springer-Verlag.

(14) Stone, C. J. (1985), Additive Regression and Other Nonparametric Models, *The Annals of Statistics*, **13**, 685-705.

(15) Wahba, G. (1979), Convergence Rates of Thin Plate Smoothing Splines When the Data are Noisy, *In Smoothing Techniques for Curve Estimation, eds.* T. Gasser and M. Rosenblatt, New York, Springer-Verlag, 233-245.

(16) Wand, M. P. and Jones, M. C. (1995), *Kernel Smoothing*, Monographs on Statistics and Applied Probability 60, London, Chapman and Hall.