Department of
STATISTICS
Phone 409 - 845-3141

DTIC FILE COPY

AD-A182 141

KERNEL ESTIMATION OF THE DERIVATIVE
OF THE REGRESSION FUNCTION
USING REPEATED-MEASUREMENTS DATA

D. B. Holiday and Jeffrey D. Hart

Department of Statistics
Texas A&M University

Technical Report No. 4

June 1987

DTIC
ELECTE
JUL 0 9 1987
E

87 7 7 134

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 4	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Kernel Estimation of the Derivative of the Regression Function Using Repeated-Measurements Data		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) D. B. Holiday and Jeffrey D. Hart		8. CONTRACT OR GRANT NUMBER(s) N00014-85-K-0723
9. PERFORMING ORGANIZATION NAME AND ADDRESS Texas A&M University Department of Statistics College Station, TX 77843		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE June 1987
		13. NUMBER OF PAGES 42
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, If different from Report)

NA

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
Nonparametric regression; growth curves; correlated data; optimum bandwidth; mean integrated squared error; Gasser-Müller estimator.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Consider the following repeated-measurements model: $Y_s(x_t) = g(x_t) + \varepsilon(x_t)$, $s=1,\ldots,m$ (e.g., subjects), $t=1,\ldots,n$ (e.g., time points), $E(\varepsilon_s(x_t))=0$ for all s and t, $cov(\varepsilon_s(x_t), \varepsilon_u(x_v))=0$ for $s \neq u$, and $cov(\varepsilon_s(x_t), \varepsilon_s(x_v))=\sigma^2 \gamma(x_t-x_v)$, where γ is a correlation function. Asymptotic expansions of the mean squared error of the Gasser-Müller kernel estimator of an arbitrary pth derivative of g are obtained for two general classes of correlation functions. Consistency and other results based on such expansions are discussed for orders $p=1$ and $p=2$.

Kernel Estimation of the Derivative of the Regression Function

Using Repeated-Measurements Data

D. B. Holiday and J. D. Hart

TEXAS A&M UNIVERSITY

Abstract

In fixed-design kernel nonparametric regression, there has been a paucity of results for models which allow for correlated errors. Consider the following repeated-measurements model, applicable in growth curve analysis: $Y_s(x_t) = g(x_t) + e_s(x_t)$, $s=1,\ldots,m$ (e.g.,subjects), $t=1,\ldots,n$ (e.g.,time points) with errors of zero mean and within-subject covariance matrix Σ. More specifically, we assume that $\mathrm{cov}[e_s(x_t), e_u(x_v)] = \delta_{su}\sigma(x_t,x_v)$ where δ_{su} is the Kronecker-delta and $\sigma(x_t,x_v)$ is the $(t,v)^{th}$ element of Σ. Furthermore, it is assumed that the matrix elements $\sigma(x_t,x_v)$ may be represented as the product of a scalar variance term and a suitably restricted correlation function $\gamma(x_t-x_v)$. Asymptotic expansions of the mean squared error of the Gasser-Müller kernel estimator of an arbitrary p^{th} derivative of g are obtained for two general classes of correlation functions. Consistency and other results based on such expansions are discussed for orders $p=1$ and $p=2$.

1. Introduction

If one desires to estimate the pth derivative of $g(x)$, denoted $g^{(p)}(x)$, it is natural to question whether $g_n^{(p)}(x)$, the pth derivative of a kernel estimator $g_n(x)$, is a worthwhile candidate. Such estimators have been termed "indirect" estimators of $g^{(p)}$ by Georgiev (1984c) and others. Extending the notation of the general linear estimator, let

$$g_n^{(p)}(x) = \sum_{i=1}^{n} w_{ni}^{(p)}(x) \, \bar{Y}(x_{ni}), \quad p=0,1,2,\ldots \qquad (1.1)$$

denote an estimator of $g^{(p)}(x)$ in the correlated errors model

$$Y_s(x_t) = g(x_t) + e_s(x_t);$$

$$s = 1,2,\ldots,m \qquad t = 1,2,\ldots,n;$$

$$E[e_s(x_t)] = 0, \text{ all } s,t;$$

$$\mathrm{cov}[e_s(x_t),e_u(x_v)] = \delta_{su}\sigma(x_t,x_v). \qquad (1.2)$$

For an estimator of the Gasser-Müller (GM) type, which will be the main focus of this paper, we have

$$w_{ni}(x) = w_{ni}^{(0)}(x) = (1/h) \int_{A_i} K[(x-u)/h]du, \qquad (1.3)$$

where

$$A_i = [s_{i-1},s_i]; \quad x_i \le s_i \le x_{i+1}; \; s_0 = a, \; s_n = b. \qquad (1.4)$$

Without loss of generality, the range of interest $[a,b]$ in our analytic discussions will be the interval $[0,1]$. In Gasser and

Müller (1979) it is assumed that the kernel function K has finite support $[-r,r]$, which we will always assume to be $[-1,1]$ in our treatment. Unlike density estimation, where it is desirable that the estimate be a density, there is no compelling reason in regression to require that K be nonnegative. We will continue to assume that K is symmetric, has support $[-1,1]$, and integrates to one, but will not require nonnegativity. In the case of forming an indirect estimator of $g^{(p)}$, we see that the differentiated kernel terms certainly will produce weight functions which are negative over a large part of the range.

Observe that the weight function coefficient corresponding to the indirect GM estimator is obtained by differentiating (1.3), yielding

$$w_{ni}^{(p)}(x) = [d^p/dx^p]\, w_{ni}^{(0)}(x) = (1/h^{p+1}) \int_{A_i} K_p[(x-u)/h]du, \quad (1.5)$$

where

$$K_p(u) = K^{(p)}(u)\, I_{(-1,1)}(u).$$

For the Priestley-Chao estimator (PC), we observe that

$$w_{ni}^{(p)}(x) = (1/h^{p+1})\, (x_{i+1}-x_i)\, K_p[(x-x_i)/h], \quad (1.6)$$

while for the Nadaraya-Watson (NW) we have

$$w_{ni}^{(p)}(x) = [d^p/dx^p]\, \{K[(x-x_i)/h] \,/\, \textstyle\sum_t K[(x-x_t)/h]\}.$$

This last expression will produce a fairly complicated indirect estimator of $g^{(p)}$, due to the presence of x in both the numerator and denominator. Observe that this is true for any cut-and-normalized estimator. Due to these complications, we will only develop results for the non-cut-and-normalized GM estimator as specified by (1.5),

which is asymptotically less biased than the FC estimator (see Gasser and Müller (1979)).

The primary goal of this paper is to develop asymptotic expressions for the variance, bias, and mean squared error of estimators (1.1) with weights of form (1.5) for an arbitrary order p of the derivative to be estimated. The low order derivatives are of predominant interest, especially in growth curve applications. Cases for $p = 1$ and $p = 2$ will be considered in more detail.

We will consider estimation in the model given by (1.2) with a general correlation function γ, which will be a function whose argument is the spacing $z_{ni} - z_{nj}$ of design points. Throughout this paper, it is assumed that

    (i) γ is even,

    (ii) $|\gamma(\cdot)| \leq 1$,

    (iii) γ is real and continuous at 0,   (1.7)

    (iv) $\gamma(0) = 1$, and

    (v) γ is a positive definite function.

From properties (iii) - (v) it follows that γ is a characteristic function and hence is uniformly continuous over the real line (see Chung (1974,p.179)). The continuity assumption will be necessary when Taylor expansions of γ are considered.

Situations for which γ is smooth near the origin will be considered. In addition we will study situations where the first derivative of γ is not continuous at the origin, as in the Ornstein-Uhlenbeck model,

$$\gamma(u) = \exp\{-\alpha|u|\}, \quad \alpha > 0.$$

We will investigate the ramifications of our asymptotic expressions
with regard to mean square consistency, optimal local bandwidth
selection, and the corresponding rate of convergence of the MSE.
First we will review some recent developments in the estimation of
derivatives.

2. Recent Developments

Unless otherwise noted, all references in this section deal with the
traditional fixed-design model with uncorrelated errors,

$$Y(x_t) = g(x_t) + \epsilon(x_t); \quad t = 1,\ldots,n;$$

$x_1,\ldots,x_n$ are selected prior to data collection; and

$$E[\epsilon(x_t)] = 0, \quad \operatorname{cov}[\epsilon(x_t),\epsilon(x_u)] = \delta_{tu}\,\sigma^2.$$

Observe that this model is the special case of the repeated-
measurements model (1.2) with m = 1 and uncorrelated errors. It
should be noted, however, that the scalar variance term in the above
model would probably be attributed to measurement error and not to
sampling variability in the subject-to-subject response which is
associated with a population of subjects.

The estimators of $g^{(p)}(x)$ will be of linear form:

$$g_n^{(p)}(x) = \sum_{i=1}^{n} w_{ni}^{(p)}(x)\, Y(x_i), \quad p = 0,1,\ldots$$

where $w_{ni}^{(p)}(x)$ depends on the design points $x_{n1},\ldots,x_{nn}$, a bandwidth

parameter h, and a kernel function K_p. As in our development, many authors assume that the kernel function has finite support. It will be seen that the kernel must satisfy other restrictions in order for the estimator to possess desirable statistical properties.

Schuster and Yakowitz (1979) prove results involving the uniform convergence of indirect estimators of $g^{(p)}$ for the FC estimator as specified by (1.6). We state their main result.

Theorem (2.1) -- Schuster and Yakowitz (1979)

Let K be a probability density function such that

    (i) $|u^{j+1} K^{(j)}(u)| \rightarrow 0$, as $u \rightarrow \infty$, for $j = 0,1,...,p$;

    (ii) the first p+1 derivatives of K are continuous and
        bounded on $(-\infty,\infty)$; and

    (iii) $\int |u^p \Phi(u)| du < \infty$, where $\Phi(u)$ is the
        characteristic function of K.

Additional suppositions are

    (iv) $g^{(p+1)}$ exists and is continuous on $[0,1]$; and

    (v) $E[Y^2;x] \leq V$ (bounded variance for all x).

Then for every $\epsilon > 0$, there is a constant $C > 0$ such that, for n sufficiently large,

$$P[\sup_{x \in [0,1]} |g_n^{(p)}(x) - g^{(p)}(x)| > \epsilon] < C/\epsilon^2 nh^{2p+2}.$$

For results restricting K to be a symmetric density for the case p=0, consult Cheng and Lin (1981a,1981b). Their estimator is similar to that of Gasser and Müller (1979), who allow negativity of K. Gasser and Müller (1984) and Gasser, Müller, and Mammitzsch (1985) extend their original work to the estimation of derivatives. In the first paper it is true that $K_p = K^{(p)}$, as in indirect estimation, but this requirement is dropped in the latter paper since K_p may be chosen freely for each p, provided it satisfies a certain moment condition. It is not always optimal to simply estimate $g^{(p)}$ by taking the pth derivative of an optimal estimator of g. The latter work explores the choice of kernel in a class of so-called higher order kernels that depend on p. This is what Georgiev (1984c) terms direct estimation. The distinction is not particularly important since one may relate the conditions which K_p should satisfy to a set of conditions $K(=K_0)$ should satisfy when differentiated p times. This is done in Gasser and Müller (1984). We should note, however, that for p of sufficient magnitude, the associated class of higher-order kernels may not admit $K^{(p)}$, where K is a probability density. This is because the moment conditions placed on $K^{(p)}$ may imply that K must be negative over part of its support.

Gasser and Müller (1984) require that $K = K_0$ satisfy the following conditions:

(a) K is a p-times differentiable function so that

$K_p = K^{(p)}$ is defined,

(b) K_p has support $[-1,1]$,

(c) $\int K(u)du = 1$, and

(d) $K_j(1) = K_j(-1) = 0$ for $j = 0,1,\ldots,p-1$.

Notice that the Epanechnikov kernel $K(u) = (3/4)(1-u^2)$ would satisfy condition (d) for estimating the first derivative (p=1), but fail for estimating the second derivative (p=2), in which case the smoother quartic kernel $K(u) = (15/16)(1-u^2)^2$ is a candidate. Gasser, et al. (1985) have shown that, subject to certain conditions,

$$K_1(u) = (15/4)(u^3-u)$$

is superior to the derivative of the Epanechnikov,

$$K_1(u) = K^{(1)}(u) = -(3/2)u, \qquad (2.1)$$

insofar as minimum MSE is concerned. However, (2.1) is better if one is more concerned with minimization of the asymptotic variance than with minimization of the asymptotic squared bias. Under these conditions and Lipschitz continuity of g, an asymptotic expansion of $E[g_n^{(p)}(x)]$ is obtained. This result will be used in later developments in this paper. One of the main results is the following:

Theorem (2.2) — Gasser and Müller (1984)

Let g be a p-times differentiable function defined on the interval $(0,1)$, (when p=0, assume continuity). Suppose that $K_p = K^{(p)}$ is bounded. Then $g_n^{(p)}(x)$ as specified by (1.5) is MSE consistent for $g^{(p)}(x)$ if:

    (i) $g^{(p)}$ is continuous at x in $(0,1)$, and

    (ii) $h \rightarrow 0$, $nh^{2p+1} \rightarrow \infty$, as $n \rightarrow \infty$.

The authors proceed to establish results on almost-sure convergence and asymptotic normality under conditions similar to those in Cheng and Lin (1981a,1981b). An asymptotic expression for the mean squared error at a point is obtained. The rate of MSE convergence depends upon an appropriate choice of a higher-order kernel, which relates to the original differentiated kernel via the following lemma.

Lemma (2.1) — Gasser and Müller (1984)

Let β be a non-zero constant and r be an integer such that $r \geq p+2$. Suppose K is a kernel function such that

    (i) $K^{(j)}(1) = K^{(j)}(-1) = 0$ for $j=0,1,\ldots,p-1$;

$$\text{(ii)} \int_{-1}^{1} u^j K(u)du = \begin{cases} 1 & j = 0 \\ 0 & j = 1,\ldots,r-p-1 \\ (-1)^p \beta (r-p)!/r! & j = r-p. \end{cases} \quad (2.2)$$

Then the pth derivative of K, denoted $K_p = K^{(p)}$, satisfies

$$\int_{-1}^{1} u^j K_p(u)du = \begin{cases} 0 & j = 0,\ldots,p-1,p+1,\ldots,r-1 \\ (-1)^p p! & j = p \\ \beta & j = r. \end{cases} \qquad (2.3)$$

Furthermore, if K_p is a function satisfying (2.3), then there exists a p-times differentiable function K satisfying (2.2) with $K^{(p)} = K_p$.

A kernel satisfying (2.3) with a finite $\beta \neq 0$ is termed a kernel of order (p,r). The result for the expansion of MSE is now stated.

Theorem (2.3) -- Gasser and Müller (1984)

Let g be an r-times differentiable function on [0,1] and $r \geq p+2$. Let $g^{(r)}$ be continuous at a point x in (0,1). Let the Lipschitz continuous kernel K_p be a kernel of order (p,r) with support $[-L,L]$ and suppose

(i) $h \rightarrow 0$, $nh \rightarrow \infty$, as $n \rightarrow \infty$, and

(ii) $\max_j |s_j - s_{j-1} - (1/n)| = O(n^{-\delta})$, $\delta > 1$.

Then for x in $[h,(1-h)]$ and

$$g_n^{(p)}(x) = (1/h^{p+1}) \sum_{i=1}^{n} Y(x_i) \int_{A_i} K_p[(x-u)/h]du,$$

where A_i is defined in (1.4), we have

$$MSE[g_n^{(p)}(x)] = E[g_n^{(p)}(x) - g^{(p)}(x)]^2$$

$$= (\sigma^2/nh^{2p+1})\int_{-1}^{1} K_p(u)^2 du + (h^{2(r-p)}/(r!)^2)[\int_{-1}^{1} u^r K_p(u)du]^2 g^{(r)}(x)^2$$

$$+ O[(1/n^2h^{2p+1}) + (1/n^6h^{2p+1}) + (1/nh^{2p-2})] + o(h^{2(r-p)}). \qquad (2.4)$$

An asymptotically optimal local bandwidth may be obtained in the usual manner by solving for the critical value of the leading terms of (2.4), regarded as a function of h. The authors obtain

$$h_{opt}(x) = C \, n^{-1/(2r+1)}, \tag{2.5}$$

where C is a constant depending on p, r, σ^2, K_p, and $g^{(r)}(x)$. Upon insertion of this bandwidth into (2.4), the rate of convergence of the MSE is obtained:

$$MSE[g_n^{(p)}(x)] \sim C^* \, n^{-2(r-p)/(2r+1)} + o(n^{-2(r-p)/(2r+1)}). \tag{2.6}$$

The quantity C^* depends on the same quantities as C in (2.5). In (2.6) observe that the rate of convergence gets rapidly worse if one desires to estimate $g^{(p)}$ for increasing p. The bias rate may be kept constant provided r-p stays constant, which forces a larger value of r and amounts to an assumption of greater smoothness of g. Even so, the rate of the variance still gets worse for increasing p. The higher the value of p, the greater the value of the local bandwidth, at a cost of increased bias and more pronounced boundary effects.

If one desires to choose a bandwidth which minimizes a global measure of error, then one may choose h to minimize the mean integrated squared error (MISE), or an approximation thereof. Unfortunately the MISE may be dominated by the boundary effect, that is, bias which arises whenever one estimates $g^{(p)}$ near the endpoints of the range of interest. Bias near the boundary usually contributes heavily to the MSE, unless the function g happens to be smooth at the boundary. Since our weights are of kernel form and depend on a single bandwidth parameter selected to serve well over the entire interval,

some modification of h and/or K may be necessary to downweight the
contamination which arises near the boundary. It is well-known that
the rate of convergence of the MSE at a point near a boundary is
often slower than at an interior point of estimation. To circumvent
this difficulty, Gasser and Müller (1984) choose to modify the kernel
K_p, obtaining "boundary kernels" which are used whenever one
estimates within a bandwidth of either boundary. These modifications
require new kernels satisfying certain moment conditions. Contrary to
some appoaches taken by other authors, these methods do not alter the
original kernel by truncation or bandwidth shrinkage. The boundary
kernels, not necessarily nonnegative, arise from the solution of the
variational problem of MISE minimization, considered as a functional
of the kernel in the MISE expansion. The resulting kernels turn out
to be relatively simple low-order polynomials.

Härdle and Gasser (1984) consider robust estimation of $g^{(1)}$ with
an approach based on M-estimation (Huber, 1981, Ch. 3.2). When using
an estimator which acts as a linear operation on the data, such as a
kernel estimator, single outliers might mimic peaks and troughs,
corresponding to unexpected zeros in the estimated derivative. The
authors note that estimation of derivatives is likely to be more
sensitive to outliers, and hence robust methods are often called for.

Georgiev (1984c) considers direct estimators of $g^{(p)}(x)$, which,
in his notation, replace the quantity K_p in (1.5) with $K_{p,r}$, a kernel
which satisfies similar moment conditions. These results are similar
to those of Gasser and Müller (1984), but use kernels whose support

is [0,1] rather than [-1,1]. Another minor difference in the estimator is the usage of $A_i = [x_{i-1}, x_i]$ rather than $A_i = [s_{i-1}, s_i]$. Also a reversal of the order of the integrand arguments of (1.5) does not lead to a factor of $(-1)^p$ when the kernel weight is differentiated.

The estimator considered is

$$g_n^{(p)}(x) = (1/h^{p+1}) \sum_{i=1}^{n} \{ Y(x_i) \int_{A_i} K_{p,r}[(u-x)/h] du \}. \qquad (2.7)$$

We state the following theorem to afford a comparison between the approaches.

Theorem (2.4) -- Georgiev (1984a)

Let $r > p$ be a fixed integer. Assume $K_{p,r}$ is a real-valued, bounded, and continuous function on $(0,1)$ which vanishes outside $(0,1)$. Also assume

(i) $\int u^j K_{p,r}(u) du = \begin{cases} j! & \text{for } j = p \\ 0 & \text{for } j \neq p, \quad j=0,1,\ldots,r-1; \end{cases}$

(ii) $0 = x_0 \leq x_1 < \ldots < x_n \leq x_{n+1} = 1,$

(iii) $h \to 0$ as $n \to \infty,$

(iv) $\sup_{x \in [0,1]} |g^{(r)}(x)| < \infty,$ and

(v) $\delta_n = \max(x_i - x_{i-1}) = O(1/n).$

Then

$$\sup_{z \in (0,1)} \left| E \, g_n^{(p)}(z) - g^{(p)}(z) \right| = O(h^{r-p} + 1/nh^p).$$

Furthermore, if

(vi) $K_{p,r}$ is Lipschitz continuous of order $a > 0$,

(vii) $E[e^2(x_1)] \le \sigma^2 < \infty$, and

(viii) $h \sim n^{-1/(2r+1)}$,　　　　　　　　　　　　　　　　　(2.8)

we obtain

$$\sup_{z \in (0,1)} E[g_n^{(p)}(z) - g^{(p)}(z)]^2 = O(h^{-2(r-p)/(2r+1)}).$$

Observe that the order of the sequence of bandwidths in (2.8) is independent of p and depends only on r, which is termed the orthogonality parameter of $K_{p,r}$ (due to the moment conditions). Georgiev (1984a) proves other asymptotic results on the speed of convergence and also lists polynomial kernels $K_{p,r}$ satisfying the required conditions for various p and r. In Georgiev (1984b), the author examines estimator (2.7) when h has been replaced with the Euclidean distance between x and a user-specified k_n^{th} nearest neighbor of x, which is a variable bandwidth method. The latter reference contains results closely related to those of Gasser and Müller (1984).

3. An MSE Expression for Arbitrary Order p

In this section we will build up an expression for

$$MSE[g_n^{(p)}(x)] = E[g_n^{(p)}(x) - g^{(p)}(x)]^2,$$

where $g_n^{(p)}(x)$ is defined as the estimator of $g^{(p)}(x)$ with

$$g_n^{(p)}(x) = (1/h^{p+1}) \sum_{i=1}^{n} \{ \bar{Y}(x_i) \int_{A_i} K_p[(x-u)/h] du \}, \qquad (3.1)$$

where

$$\bar{Y}(x_i) = (1/m) \sum_{s=1}^{m} Y_s(x_i); \quad A_i = [s_{i-1}, s_i],$$

and

$$s_0 = 0, \quad s_n = 1, \quad x_i \leq s_i \leq x_{i+1}.$$

This estimator will be considered in the by now familiar correlated-errors model

$$Y_i(x_j) = g(x_j) + e_i(x_j); \qquad (3.2)$$

$$i = 1, \ldots, m; \quad j = 1, \ldots, n;$$

$$E[e_i(x_j)] = 0, \quad x_1, \ldots, x_n \text{ controlled};$$

$$cov[e_i(x_j), e_k(x_\ell)] = \sigma^2 \gamma(x_j - x_\ell) \delta_{ik}.$$

As before, the correlation function γ is assumed to be even with $\gamma(0) = 1$. In obtaining expressions for the asymptotic variance and bias, it will be necessary to impose mild conditions on K_p, γ, the spacings of the sequence of designs, and $g^{(j)}$, $j = 0, \ldots, p$. With further assumptions on γ and g, we will in later sections derive

expansions for the asymptotic variance and bias, respectively.

Many of the subsequent methods will use arguments similar to those found for the correlated-errors model in Hart and Wehrly (1986) for the case $p = 0$. Prior to stating the main result, we will develop the necessary tools to facilitate the proof. The following two lemmas will be used in the result for asymptotic variance.

Lemma (3.1) Let γ be Lipschitz continuous and suppose for each n, $x_{n1}, \ldots, x_{nn}$ is an ordered partition of fixed design points on $[0,1]$. Let $s_{n0}, \ldots, s_{nn}$ satisfy

$$s_{n,i-1} \leq x_{ni} \leq s_{ni}, \quad i = 1, \ldots, n,$$

where $s_{n0} = 0$ and $s_{nn} = 1$. Then there exists some $B > 0$ for which

$$|\gamma(x_{ni} - x_{nj}) - \gamma(v - u)| \leq B \sup_k [s_{nk} - s_{n,k-1}],$$

whenever

$$s_{n,j-1} \leq u \leq s_{nj} \text{ and } s_{n,i-1} \leq v \leq s_{ni}.$$

Proof: By the Lipschitz condition, there exists some $B^* > 0$ for which

$$|\gamma(x_{ni} - x_{nj}) - \gamma(v - u)| \leq B^* |(x_{ni} - x_{nj}) - (v - u)|$$

$$\leq B^* [|x_{ni} - v| + |x_{nj} - u|]$$

$$\leq B^* [(s_{ni} - s_{n,i-1}) + (s_{nj} - s_{n,j-1})]$$

$$\leq B^* [2 \cdot \max\{(s_{ni} - s_{n,i-1}), (s_{nj} - s_{n,j-1})\}]$$

$$\leq B \sup_k [s_{nk} - s_{n,k-1}],$$

where $B = 2B^* > 0$. $\blacksquare$

17

Lemma (3.2) Let h > 0 and x be a point in (0,1). Let p be a positive integer and suppose K_p is an absolutely integrable function vanishing outside [-1,1]. Let γ be a correlation function. Define

$$J_1 = h^{-2p-2} \int_0^1\int_0^1 \gamma(u-v)\, K_p[(x-u)/h]K_p[(x-v)/h]\,du\,dv. \qquad (3.3)$$

Then for h = h(x) sufficiently small, $J_1 = J_2$, where

$$J_2 = h^{-2p} \int_{-1}^1\int_{-1}^1 \gamma[(s-t)h]\, K_p(s)K_p(t)\,ds\,dt. \qquad (3.4)$$

Proof: Now (3.3) exists since γ is bounded and K_p is absolutely integrable. Taking s = (x-u)/h and t = (x-v)/h, the Jacobian is h^2. Then (3.3) may be written

$$h^{-2p-2}\{h^2 \int_S\int_S \gamma[(s-t)h]\, K_p(s)K_p(t)\,ds\,dt\},$$

where S = [(x-1)/h, x/h] contains [-1,1] for h small enough. The result follows.

∎

Next we state a lemma concerning the bias, which does not depend on n or the correlation function γ. Hence the bias is the same as in the uncorrelated case with n = 1. We now quote the previously discussed result from Gasser and Müller (1984) concerning an integral approximation to $E[g_n^{(p)}(x)]$.

Lemma (3.3) — Gasser and Müller (1984)

Let g be a p-times differentiable function with $K_p = K^{(p)}$ also existing on [-1,1], where

(i) $\int_{-1}^1 K(u)\,du = 1$, and

(ii) $K^{(r)}(-1) = K^{(r)}(1) = 0$, r = 0,1,...,p-1.

Then for $g_n^{(p)}(x)$ defined by (3.1),

$$E[g_n^{(p)}(x)] = (1/h^p) \int_{-1}^{1} g(x-hu) \, K_p(u)du + O(1/nh^p)$$

$$= \int_{-1}^{1} g^{(p)}(x-hu) \, K(u)du + O(1/nh^p).$$

We are now ready to prove the main result of this section. For reference we will restate all previous assumptions necessary to the proof.

Theorem (3.1) Let p be a positive integer. Consider estimation of $g^{(p)}$ using estimator (3.1) in model (3.2). Let $K = K_0$ be a kernel function with

(i) $K(u) = K(-u)$ with support $[-1,1]$,

(ii) $\int K(u)du = 1$ (K may be negative),

(iii) $K_r = K^{(r)}$ exist, are bounded, and vanish outside $[-1,1]$ for $r = 0,1,\ldots,p$; and

(iv) $K_r(1) = K_r(-1)$ for $r = 0,1,\ldots,p-1$.

Assume further that γ is a Lipschitz continuous correlation function, that is,

(v) γ satisfies the conditions in (1.7), and

(vi) $|\gamma(s) - \gamma(t)| \leq B|s - t|$, some $B > 0$.

Assume the design sequence satisfies

(vii) $\max_i |x_{ni} - x_{n,i-1}| = O(1/n)$,

and for the regression function g,

(viii) $g^{(r)}$ exists for $r = 1,\ldots,p$ and $g^{(p)}$

is continuous at x.

Then for $h = h_{n,m} \to 0$ as $n,m \to \infty$,

(a) $\text{Var}[g_n^{(p)}(x)] = (\sigma^2/nh^{2p}) I(h;p) + O(1/nmh^{2p})$, where

$I(h;p) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \gamma[(s-t)h] K_p(s)K_p(t)\,ds\,dt;$

(b) $\text{Bias}^2[g_n^{(p)}(x)] = B^2(h;p) + o(1/nh^p) + O(1/n^2h^{2p})$,

where $B(h;p) = \int_{-1}^{1} g^{(p)}(x-hu) K(u)\,du - g^{(p)}(x);$ and

(c) $\text{MSE}[g_n^{(p)}(x)] = \text{Var}[g_n^{(p)}(x)] + \text{Bias}^2[g_n^{(p)}(x)]$

$= (\sigma^2/nh^{2p}) I(h;p) + B^2(h;p)$

$+ o(1/nh^p) + O(1/nmh^{2p}) + O(1/n^2h^{2p}).$

Proof: Now (c) is trivial once (a) and (b) are established. We see that (b) follows immediately from Lemma (3.3) and the fact that $g^{(p)}$ is continuous at x. The proof of (a) will require (i) - (vii), except for (iv). Writing

$$w_{ni}^{(p)}(x) = (1/h^{p+1}) \int_{\Delta_i} K_p[(x-u)/h]\,du,$$

it is clear that

$$\text{Var}[g_n^{(p)}(x)] = (\sigma^2/n) \sum_i \sum_j w_{ni}^{(p)}(x) w_{nj}^{(p)}(x) \gamma(x_{ni}-x_{nj}).$$

Consider (3.4) in Lemma (3.2) and let

$$C = (\sigma^2/n)h^{-2p-1} \text{ and } D_n = \sup_k(x_{nk}-x_{n,k-1}).$$

By Lemma (3.2), with h sufficiently small (depending on x),

$$|\text{var}[g_n^{(p)}(x)] - (\sigma^2/nh^{2p})I(h;p)|$$

$$= |\text{var}[g_n^{(p)}(x)] - (\sigma^2/n)J_2|$$

$$= |\text{var}[g_n^{(p)}(x)] - (\sigma^2/n)J_1|$$

$$= C \left| \Sigma_i \Sigma_j \int_{A_i} \int_{A_j} \gamma(x_{ni}-x_{nj})\, K_p[(x-u)/h]K_p[(x-v)/h]\,du\,dv \right.$$
$$\left. - \int_0^1 \int_0^1 \gamma(u-v)\, K_p[(x-u)/h]K_p[(x-v)/h]\,du\,dv \right|$$

$$= C \left| \Sigma_i \Sigma_j \int_{A_i} \int_{A_j} [\gamma(x_{ni}-x_{nj})-\gamma(u-v)]\, K_p[(x-u)/h]K_p[(x-v)/h]\,du\,dv \right|$$

$$\leq BCD_n \int_0^1 \int_0^1 | K_p[(x-u)/h]\, K_p[(x-v)/h] | \,du\,dv$$

$$= BCD_n \left\{ \int_0^1 |K_p[(x-u)/h]|\,du \right\}^2, \tag{3.5}$$

the last inequality by Lemma (3.1). By a change of variable, the integral in (3.5) is

$$h \int_A |K_p(v)|\,dv.$$

Now $A = [(x-1)/h, x/h]$ contains $[-1,1]$ for h small enough. We then obtain

$$|\text{var}[g_n^{(p)}(x)] - (\sigma^2/nh^{2p})\, I(h;p)|$$

$$\leq BCD_n\, h^2 \left\{ \int_{-1}^1 |K_p(v)|\,dv \right\}^2$$

$$= \sigma^2 B \left\{ \int_{-1}^1 |K_p(v)|\,dv \right\}^2 (D_n/nh^{2p})$$

$$= O(1/nnh^{2p}), \text{ as } n,\ n \to \infty,\ h \to 0,$$

since (vii) implies that $D_n = O(1/n)$. ∎

4. Expansion of MSE for Arbitrary Order p

In this section the goal is to obtain an asymptotic expansion of $MSE[g_n^{(p)}(x)]$, where $g_n^{(p)}(x)$ is given by (3.1). We will need to expand $I(h;p)$ and $B(h;p)$ under smoothness assumptions on g and γ, enjoined by appropriate Taylor series requirements. Properties of the kernel function, K, are important when evaluating these expansions. We will usually assume that K satisfies the orthogonality properties of Gasser and Müller (1984).

First we set about defining notation useful for the evaluation and representation of commonly recurring terms. We define the moment notation:

$$\mu_j(K_r) = \int_{-1}^{1} u^j K_r(u)du, \tag{4.1}$$

which is the jth moment about the mean (zero) if K_r happens to be a symmetric probability density on $[-1,1]$. For $K = K_0$ we will require symmetry and $\mu_0(K_0) = 1$, but allow negativity.

A notation will be needed to handle bulky double integral expressions which arise in certain variance expansions. Define for a, b, t in $[-1,1]$, $r = 0,1,\ldots$ the notation (4.2), with special cases (4.3) and (4.4):

$$K(r,p;a,b) = \int_{-1}^{1}\int_{a}^{b} (s-t)^r K_p(s)K_p(t)ds\,dt; \tag{4.2}$$

$$C(r,p) = K(r,p;-1,1) = K(r,p;t,1) + K(r,p;-1,t); \tag{4.3}$$

$$D(r,p) = K(r,p;t,1) - K(r,p;-1,t). \tag{4.4}$$

Using the binomial expansion, we relate (4.3) to (4.1):

$$C(r,p) = \sum_{j=0}^{r} \binom{r}{j}(-1)^j \mu_j(K_p) \cdot \mu_{r-j}(K_p). \qquad (4.5)$$

Further simplification of (4.5) is possible by noting recursiveness and conditions for which terms (4.1) vanish.

Lemma (4.1) Suppose $\mu_j(K_r)$ exists for $r = 0,1,\ldots,p$ and $K_r = K^{(r)}$ is the rth derivative of an even, p-times differentiable function $K = K_0 = K^{(0)}$. Then for a given $r = 0,1,\ldots,p$,

$$\mu_j(K_r) = 0, \text{ if } j+r \text{ is odd.} \qquad (4.6)$$

Furthermore, if $K_{p-1}(1) = K_{p-1}(-1) = 0$,

$$\mu_j(K_p) = (-j)\cdot\mu_{j-1}(K_{p-1}), \quad j > 0, \text{ and} \qquad (4.7)$$

$$\mu_j(K_r) = 0 \text{ if } r > j \geq 0.$$

Proof: Equation (4.6) follows since (4.1) is an integral of an odd function over $[-1,1]$ when K_0 is even. To obtain (4.7), integrate by parts. ∎

This result causes tremendous simplification of (4.5) since at least half of the terms will vanish (every other term drops out). Our Taylor series tools will be summarized in the next lemma. This lemma has been tailored to meet our needs, since expansion will always be about zero and the argument will approach zero by assumption. Even functions and one-sided expansions will be taken into account.

<u>Lemma (4.2)</u> Let f be a real, continuous function defined in a neighborhood of the origin $N:(-t,t)$, $t > 0$. Let

      (i) $f^{(j)}$, $j = 1,\ldots,n$ exist and be continuous

          in N and let $f^{(n+1)}(0)$ exist.

Then, as $u \to 0$,

  (a) $f(u) = \sum_{k=0}^{n+1} f^{(k)}(0) u^k/k! + o(u^{n+1})$.

Furthermore, if $r = [$ greatest integer $\leq (n+1)/2]$ and

      (ii) $f(u) = f(-u)$,

then, as $u \to 0$,

  (b) $f(u) = \sum_{k=0}^{r} f^{(2k)}(0) u^{2k}/(2k)! + o(u^{2r})$.   (4.8)

Finally, if we replace (i) with

      (i)$'$ $f^{(j)}$, $j = 1,\ldots,n$ exist and are continuous in

          $N-\{0\}$ and $f^{(n+1)}(0+)$ and $f^{(n+1)}(0-)$ exist,

then assuming only (i)$'$ with u in $[0,t)$ and $u \to 0^+$,

  (c) $f(u) = \sum_{k=0}^{n+1} f^{(k)}(0+) u^k/k! + o(u^{n+1})$.

For u in $(-t,0]$ and $u \to 0^-$,

  (d) $f(u) = \sum_{k=0}^{n+1} f^{(k)}(0-) u^k/k! + o(u^{n+1})$.   (4.9)

Under (i)$'$ and (ii) with k even,

  (e) $f^{(k)}(0-) = f^{(k)}(0+)$,   (4.10)

which we will, with an abuse of notation, denote $f^{(k)}(0)$.

(Technically, this derivative may fail to exist by the definition, but we will adopt this notation convention whenever the right-hand and left-hand derivatives at a point are the same.) For k odd,

  (f) $f^{(k)}(0-) = - f^{(k)}(0+)$.

Proof: Assertions (a) and (b) are standard and follow from Fulks (1969, p.160). Assertions (c) and (d) are lesser known and are adapted from Fulks (1969, p.160).

To obtain (e) and (f), note that (ii) oscillates between an even and odd function on each successive differentiation. Using the assumed continuity of the derivatives in (i)′, we have, for k even,

$$f^{(k)}(t) = f^{(k)}(-t).$$

Letting $t \to 0^+$ produces result (e). Similarly we obtain (f). ∎

Recall that, under the conditions of Theorem (3.1),

$$\text{Var}[g_n^{(p)}(x)] = (\sigma^2/mh^{2p})\, I(h;p) + O(1/mnh^{2p}), \qquad (4.11)$$

where

$$I(h;p) = \int_{-1}^{1}\int_{-1}^{1} \gamma[(s-t)h]\, K_p(s)K_p(t)\,ds\,dt.$$

Under two situations for the smoothness properties of γ, we will now derive respective expansions of (4.11).

Theorem (4.1) Suppose γ is a correlation function satisfying (1.7) and (i)′ of Lemma (4.2) (with $n = q \geq 1$ as the order of smoothness). Then, under the conditions of Theorem (3.1),

$$\text{Var}[g_n^{(p)}(x)] = (\sigma^2/m)[S_0 + S_1 + o(h^{q+1-2p})] + O(1/mnh^{2p}), \qquad (4.12)$$

where

$$S_0 = \sum_{v=0}^{a} \gamma^{(2v)}(0)\, C(2v,p)\, h^{2(v-p)}/(2v)!,$$

$$S_1 = \sum_{v=0}^{b} \gamma^{(2v+1)}(0+)\, D(2v+1,p)\, h^{2(v-p)+1}/(2v+1)!,$$

$$a = b = q/2 \text{ for } q \text{ even, and}$$

$$a = (q+1)/2,\ b = (q-1)/2 \text{ for } q \text{ odd.}$$

The constants $C(2v,p)$ and $D(2v+1,p)$ are defined by (4.3) and (4.4), respectively. Furthermore, if γ has more smoothness at the origin, that is, γ satisfies (i) of Lemma (4.2) rather than (i)$'$, then (4.12) still holds with $S_1 = 0$.

Proof: Let $h \to 0$. Let γ satisfy (1.7) and (i)$'$ of Lemma (4.2). Observe that (ii) of Lemma (4.2) is automatically satisfied for a correlation function. We have for $s > t$,

$$\gamma[(s-t)h] = \Sigma_{j=0}^{q+1} \gamma^{(j)}(0+) (s-t)^j h^j/j! + o(h^{q+1}),$$

and for $s < t$,

$$\gamma[(s-t)h] = \Sigma_{j=0}^{q+1} \gamma^{(j)}(0-) (s-t)^j h^j/j! + o(h^{q+1}).$$

Bring in $I(h;p) = I_1 + I_2$, where

$$I_1 = \int_{-1}^{1}\int_{t}^{1} \gamma[(s-t)h] K_p(s)K_p(t)dsdt, \text{ and}$$

$$I_2 = \int_{-1}^{1}\int_{-1}^{t} \gamma[(s-t)h] K_p(s)K_p(t)dsdt,$$

so that by notation (4.2),

$$I_1 = \Sigma_{j=0}^{q+1} \gamma^{(j)}(0+) K(j,p;t,1) h^j/j! + o(h^{q+1}); \qquad (4.13)$$

$$I_2 = \Sigma_{j=0}^{q+1} \gamma^{(j)}(0-) K(j,p;-1,t) h^j/j! + o(h^{q+1}). \qquad (4.14)$$

Combining (4.13) and (4.14), we see that the coefficient of $h^j/j!$ is

$$\gamma^{(j)}(0+) K(j,p;t,1) + \gamma^{(j)}(0-) K(j,p;-1,t). \qquad (4.15)$$

When considering (4.9), (4.10), (4.3), and (4.4), expression (4.15) becomes $\gamma^{(j)}(0) C(j,p)$ when j is even and $\gamma^{(j)}(0) D(j,p)$ when j is odd. We now have

$$I(h;p) = S_0^* + S_1^* + o(h^{q+1}),$$

where

$$S_0^* = \Sigma_{v=0}^{a} \; \gamma^{(2v)}(0) \; C(2v,p) \; h^{2v}/(2v)!$$

$$S_1^* = \Sigma_{v=0}^{b} \; \gamma^{(2v+1)}(0+) \; D(2v+1,p) \; h^{2v+1}/(2v+1)!,$$

with a and b depending on whether q is even or odd. The values for a
and b follow from Lemma (4.2). The conclusion (4.12) is obtained by
making use of (4.11).

For smoother γ satisfying (i) of Lemma (4.2), we use (4.8) to
conclude that, for any s, t,

$$\gamma[(s-t)h] = \Sigma_{v=0}^{a} \; \gamma^{(2v)}(0) \; (s-t)^{2v} \; h^{2v}/(2v)! + o(h^{q+1}).$$

There is now no need to break I(h;p) into two pieces. We are led to
the expression

$$I(h;p) = S_0^* + o(h^{q+1}),$$

from which the second conclusion is clear.　　　　　　　　　　　　■

To obtain an asymptotic MSE expression, we will need to consider the
bias, which under the conditions of Lemma (3.3) may be written

$$\text{Bias}[g_n^{(p)}(x)] = B(h;p) + O(1/nh^p), \qquad (4.16)$$

where

$$B(h;p) = \int_{-1}^{1} g^{(p)}(x-hu) \; K(u)du \; - \; g^{(p)}(x).$$

Theorem (4.2) Let p, r be nonnegative integers such that $r \geq p+2$. Suppose that g is an (r-1) times continuously differentiable function and that $g^{(r)}$ exists. Also suppose $K(u) = K(-u)$ and K satisfies the assumptions of Lemma (3.3). Then as $h \to 0$,

$$\text{Bias}[g_n^{(p)}(x)] = \Sigma_{j=1}^c \, g^{(p+2j)}(x) \, \mu_{2j}(K_0) \, h^{2j}/(2j)!$$
$$+ \, o(h^{r-p}) + O(1/nh^p),$$

where c = [greatest integer $\leq (r-p)/2$] and $\mu_{2j}(K_0)$ is given by (4.1).

Proof: Now for $j = 0,1,\ldots,r-p$,

$$[d^j/dh^j] \, g^{(p)}(x-hu) = (-1)^j \, u^j \, g^{(j+p)}(x-hu),$$

which evaluated at h = 0 becomes

$$(-1)^j \, u^j \, g^{(j+p)}(x).$$

Using Lemma (4.2), an expansion of

$$\Phi(h) = \int_{-1}^1 g^{(p)}(x-hu) \, K(u)du$$

becomes (with $h \to 0$),

$$\Phi(h) = \Sigma_{k=0}^{r-p} \, \Phi^{(k)}(0) \, h^k/k! + o(h^{r-p}),$$

where

$$\Phi^{(0)}(0) = g^{(p)}(x) \, \mu_0(K_0) = g^{(p)}(x),$$

$$\Phi^{(2j)}(0) = g^{(p+2j)}(x) \, \mu_{2j}(K_0), \quad j = 1,\ldots,c, \text{ and}$$

$$\Phi^{(2j+1)}(0) = - \, g^{(p+2j+1)}(x) \, \mu_{2j+1}(K_0) = 0.$$

The last statement follows from Lemma (4.1). Therefore

$$B(h;p) = \Phi(h) - g^{(p)}(x)$$

$$= \Sigma_{j=1}^c \, g^{(p+2j)}(x) \, \mu_{2j}(K_0) \, h^{2j}/(2j)! + o(h^{r-p}),$$

from which the conclusion is immediate by using (4.16). ∎

We are now in a position to combine the results into our main result for the expansion of the MSE. So as to have a single point of reference, we will explicitly state all assumptions used in the various lemmas.

Theorem (6.3) Suppose for a given $p = 0,1,\ldots$ it is desired to estimate $g^{(p)}(x)$ at a point x in $(0,1)$ using $g_n^{(p)}(x)$ of (3.1) in the correlated-errors model (3.2). Suppose γ is a Lipschitz continuous correlation function, that is,

(A1) γ satisfies the conditions in (1.7), and

(A2) $|\gamma(s) - \gamma(t)| \leq B|s - t|$, some $B > 0$.

γ also satisfies either (A3) or (A3)':

(A3) $\gamma^{(j)}, j = 1,\ldots,q$ exist and are continuous
in $(-t,t)-\{0\}$, for some $t > 0$, and $\gamma^{(j)}(0+)$,
$\gamma^{(j)}(0-)$ exist for $j = 1,\ldots,q+1$.

(A3)' $\gamma^{(j)}, j = 1,\ldots,q$ exist and are continuous in
$(-t,t)$ for some $t > 0$ and $\gamma^{(q+1)}$ exists for
some $q \geq 1$.

Assume $g = g^{(0)}$ is a function where

(B1) $g^{(j)}$ exists and is continuous for $j = 1,\ldots,r-1$,
and, $g^{(r)}$ exists;

(B2) $r \geq p+2$.

Let $K = K_0 = K^{(0)}$ be a kernel function with K_j defined to be $K_j = K^{(j)}$, $j = 0,1,\ldots$ such that

(C1) K_j exists and is bounded with support $[-1,1]$, $j = 0,1,2,\ldots$

(C2) $K_0(u) = K_0(-u)$;

(C3) $\int_{-1}^{1} K_0(u)du = 1$;

(C4) $K_j(1) = K_j(-1) = 0$, $j = 0,\ldots,p-1$.

Assume the design sequence satisfies

(D1) $\max_i |z_{ni} - z_{n,i-1}| = O(1/n)$.

Then, under (A3) with $h \to 0$,

$$MSE[g_n^{(p)}(x)] = (\sigma^2/n)\,[S_0 + S_1 + o(h^{q+1-2p})] + O(1/nh^{2p})$$
$$+ [B + o(h^{r-p}) + O(1/nh^p)]^2, \tag{4.17}$$

where

$$S_0 = \sum_{v=0}^{a} \gamma^{(2v)}(0)\, C(2v,p)\, h^{2(v-p)}/(2v)!, \tag{4.18}$$
$$S_1 = \sum_{v=0}^{b} \gamma^{(2v+1)}(0+)\, D(2v+1,p)\, h^{2(v-p)+1}/(2v+1)!,$$

with

$$a = [\text{greatest integer} \le (q+1)/2],$$
$$b = a \text{ if } q \text{ is even}, \quad b = a-1 \text{ if } q \text{ is odd};$$

and

$$C(2v,p) = \int_{-1}^{1}\int_{-1}^{1} (s-t)^{2v}\, K_p(s)K_p(t)\,ds\,dt,$$
$$\mu_j(K_p) = \int_{-1}^{1} u^j\, K_p(u)\,du,$$
$$D(2v+1,p) = \int_{-1}^{1}\int_{t}^{1} (s-t)^{2v+1}\, K_p(s)K_p(t)\,ds\,dt,$$
$$- \int_{-1}^{1}\int_{-1}^{t} (s-t)^{2v+1}\, K_p(s)K_p(t)\,ds\,dt,$$
$$B = \sum_{j=1}^{c} g^{(p+2j)}(x)\, \mu_{2j}(K_0)\, h^{2j}/(2j)!, \text{ and}$$
$$c = [\text{greatest integer} \le (r-p)/2].$$

Furthermore, if we assume (A3)' instead of (A3), then the same result holds except that $\delta_1 = 0$.

Proof: This result simply combines the conclusions for the variance and squared bias of the theorems and lemmas of Sections 3 and 4. ∎

At first glance this appears to be a rather complicated result, but once a value of p is chosen there is great simplification when one takes advantage of (4.5) and Lemma (4.1). However, there are a myriad of situations to consider since terms drop in and drop out, depending on whether p, q, and r (or simple functions of them) are even or odd. We will consider specific situations for this theorem in the next section where we will focus on the cases $p = 1,2$. Mean square consistency, choice of local bandwidth, and the speed of convergence will be discussed.

5. Explicit Results for Orders $p = 1$ and $p = 2$

Recall from the previous section that the quantities p, q, and r are basic parameters that determine the rate of convergence of the mean squared error. Of course, the bandwidth h, the number of design points n, and the number of subjects n are also crucial quantities. We have

p = the desired order of estimation in $g^{(p)}(x)$;

q = the number of continuous derivatives (or one-sided derivatives) of γ in a neighborhood (or deleted neighborhood) of the origin -- we assume existence for q+1 such quantities, however;

r-1 = the number of continuous derivatives that g possesses--we assume that $g^{(r)}$ merely exists.

In a crude sense, the magnitude of q and r represent the degree of smoothness in the correlation function and regression function, respectively. Recall that the asymptotic results require that $r \geq p+2$.

In this section we focus on p = 1,2 and write out more explicit expansions for the MSE using Theorem (4.3) under the two cases for γ. The first situation considers correlation functions which do not have a continuous derivative at the origin, that is, satisfy (A3) of Theorem (4.3). Recall that the Ornstein-Uhlenbeck function satisfies this condition. There is not a problem with q in the Ornstein-Uhlenbeck model since we observe that it is infinitely continuously differentiable in $(-\infty,\infty)-\{0\}$. Furthermore, the use of Gasser and Müller's bias results requires that g possess at least two continuous derivatives and that $g^{(3)}$ exists. For p = 2, g must possess three continuous derivatives ($r = 4$ exist). These requirements guarantee that the MSE term $o(h^{r-p})$ actually converges to zero as $h \to 0$.

In the following special cases it will be necessary to evaluate the terms $C(2\nu,p)$ in formula (6.15) for $\hat{g}_\nu$. Table 1 has been provided

Table 1. Values of C(v,p) in the MSE Expansion

Smoothness Parameter v (of g)	Value of Order of Estimation, p		
	0	1	2
0	1	0	0
1	0	0	0
2	$2s_2(R_0)$	-2	0
3	0	0	0
4	$2s_4(R_0)+6[s_2(R_0)]^2$	$-24s_2(R_0)$	24

Note: $s_j(R_t) = \int_{-1}^{1} u^j R_t(u)du$

to facilitate the calculation of these terms.

We proceed with our explicit expansions and note that assumptions (A3), (A3)' correspond to nondifferentiable and differentiable γ at the origin, respectively. For convenience, these situations will be referred to as "peaked" γ and "smooth" γ, respectively.

Theorem (5.1) Assume that the requirements of Theorem (4.3) are satisfied for a selected order p = 1,2. Take r = p+2 as the least restrictive condition placed on the number of derivatives of g. Then for p = 1, q = 1, r = 3,

(a) $MSE[c_n^{(1)}(z)] = v_1/mh + v_2/m + o(1/m) + O(1/mh^2)$

$+ b_1 h^4 + o(h^4) + o(1/mh) + O(1/m^2h^2)$, $\qquad$ (5.1)

where, using notation (4.1) and (4.4),

$v_1 = \gamma^{(1)}(0+) \sigma^2 D(1,1)$, $\qquad$ for peaked γ

$\qquad 0$, $\qquad\qquad$ for smooth γ;

$v_2 = -\gamma^{(2)}(0) \sigma^2$, and $b_1 = [g^{(3)}(z) \mu_2(K_0)/2]^2$.

For $p = 2$, $q = 3$, $r = 4$, we have

(b) $MSE[c_n^{(2)}(z)] = v_1/mh^3 + v_2/m + v_3/m + o(1/m) + O(1/mh^4)$

$+ b_1 h^4 + o(h^4) + o(1/mh^2) + O(1/m^2h^4) + O(1/m)$, $\quad$ (5.2)

where, using notation (4.1) and (4.4),

$v_1 = \gamma^{(1)}(0+) \sigma^2 D(1,2)$, $\quad$ for peaked γ

$\qquad 0$, $\qquad\qquad$ for smooth γ;

$v_2 = \gamma^{(3)}(0+) \sigma^2 D(3,2)/6$, $\quad$ for peaked γ

$\qquad 0$, $\qquad\qquad$ for smooth γ;

$v_3 = \gamma^{(4)}(0) \sigma^2$, and $b_1 = [g^{(4)}(z) \mu(K_0)/2]^2$.

Proof: For (a), assume γ is under (A3) of Theorem (4.3). Then we have $a = [(q+1)/2] = 1$, so that

$$S_0 = \sum_{v=0}^{1} \gamma^{(2v)}(0) C(2v,1) h^{2(v-1)}/(2v)!. \qquad (5.3)$$

Now from Table 1, $C(0,1) = 0$, $C(2,1) = -2$, in which case (5.3) simplifies to

$$S_0 = -\gamma^{(2)}(0).$$

Since q is odd, $b = a-1 = 0$, and we obtain

$$S_1 = \sum_{v=0}^{0} \gamma^{(2v+1)}(0+) D(2v+1,1) h^{2(v-1)+1}/(2v+1)!$$

$$= \gamma^{(1)}(0+) D(1,1)/h.$$

Next, observe that $c = [(r-p)/2] = 1$, from which we deduce

$$B = \Sigma_{j=1}^{1} \, g^{(1+2j)}(x) \, \mu_{2j}(K_0) \, h^{2j}/(2j)!$$

$$= g^{(3)}(x) \, \mu_2(K_0) \, h^2/2.$$

Using (4.17), we have for $p = 1$, $q = 1$, $r = 3$,

$$MSE[g_n^{(1)}(x)] = (\sigma^2/n)[S_0 + S_1 + o(1)] + O(1/nmh^2)$$

$$+ [B + o(h^2) + O(1/nh)]^2,$$

which yields the result after expansion and substitution. For γ under
(A3)$'$, it is clear that $S_1 = 0$. Result (b) follows similarly and the
details are omitted.   ∎

The quantities $\gamma^{(1)}(0+)$ and $D(1,1)$ are equal to $-a$ and $-6/5$ for the
Ornstein-Uhlenbeck process and Epanechnikov kernel, respectively. The
dominant variance term v_1/nh in (5.1) is therefore equal to
$(6/5) \, a \, \sigma^2/nh$ in this case.

We also observe that a smooth γ at the origin will cause certain
leading terms in (5.1) and (5.2) to vanish, resulting in a smaller
asymptotic variance (and MSE) than if γ were peaked at the origin. A
smooth correlation function will produce smoother sample paths than
will a peaked correlation function. More precisely, when γ is
sufficiently smooth, sample paths will be more nearly parallel to the
regression function over a larger neighborhood of the point of
estimation than the corresponding path for a peaked correlation
function. Such behavior would imply that the first differences of all
realizations would tend to be practically the same. This, of course,
is why we would expect a smaller variance in the estimator $g_n^{(1)}(x)$

whenever the correlation function is smooth. Observe that if it is possible to choose the the kernel such that $D(1,1) = 0$, the same reduction in variance is achieved and the estimator has the same asymptotic MSE for either type of correlation function.

An example of a smooth correlation function (i.e., one differentiable at the origin) is

$$\gamma(u) = \exp\{-x^2/\beta\}. \tag{5.4}$$

Let $f(x)$ be a symmetric probability density. The characteristic function $\phi_\zeta(u) = E_\zeta[e^{iux}]$ is therefore real, positive definite, and has the property $\phi_\zeta(0) = 1$. This observation leads to a convenient manner for generating a class of correlation functions which have the desired smoothness properties.

Another question of interest is the comparison of $Var[g_n(x)]$ and $Var[g_n^{(1)}(x)]$ in the case of a smooth correlation function. From Hart and Wehrly (1986), we have σ^2/m as the leading term of $Var[g_n(x)]$. From (5.1) the leading term of $Var[g_n^{(1)}(x)]$ is $-\gamma^{(2)}(0)\sigma^2/m$. Hence for large n, the variance of $g_n^{(1)}(x)$ will be smaller than that of $g_n(x)$ if $\gamma^{(2)}(0) > -1$. This result differs from that in the uncorrelated case where $Var[g_n(x)]/Var[g_n^{(1)}(x)] \rightarrow 0$ as $n \rightarrow \infty$ and $h \rightarrow 0$. In the particular case (5.4), we observe that $\gamma^{(2)}(0) > -1$ if $\beta > 2$.

The most important and immediate consequence of Theorem (5.1) is mean square consistency. As reported by Hart and Wehrly (1986) for the same model with $p = 0$, we find that there is a lack of consistency unless $m \rightarrow \infty$. We next observe the conditions which are

sufficient to cause the terms in (5.1) and (5.2) to vanish asymptotically.

Theorem (5.2) Assume that the requirements of Theorem (4.3) are satisfied for a selected order $p = 1$ or 2. In addition, suppose x is in $(0,1)$.

(a) For case (a) of Theorem (5.1), assume $D(1,1) \neq 0$ for a peaked γ. Then $g_n^{(1)}(x)$ is mean square consistent for $g^{(1)}(x)$ if

(i) $h \to 0$, (ii) $nh \to \infty$, and (iii) $mh \to \infty$.

If it happens that $D(1,1) = 0$ or γ is smooth at 0, we have instead:

(i) $m \to \infty$, $h \to 0$, (ii) $nh \to \infty$, and $\qquad$ (5.5)

(iii) $mnh^2 \to \infty$. $\qquad$ (5.6)

(b) For case (b) of Theorem (5.1), assume $D(1,2) \neq 0$ and peaked γ. We have consistency of $g_n^{(2)}(x)$ if

(i) $m \to \infty$, $h \to 0$, (ii) $nh^2 \to \infty$, and (iii) $mh^3 \to \infty$.

In the case that $D(1,2) = 0$ or γ is smooth at 0, we have

(i) $m \to \infty$, $h \to 0$, (ii) $nh^2 \to \infty$, and $\qquad$ (5.7)

(iii) $mnh^4 \to \infty$. $\qquad$ (5.8)

Proof: The result follows by direct consideration of the MSE expansion in Theorem (5.1). ∎

In the above theorem, conditions (5.6) and (5.8) are simplified by the slightly stronger assumption

$$\liminf_{n\to\infty,\, m\to\infty} \quad m/n \; > \; 0.$$

which, when used with (5.5) and (5.7) imply the respective condition (iii). Of course it is still possible that $m/n \to 0$, so long as the convergence is not too fast. We are now led to consider how one should ideally choose the local bandwidth in an asymptotic sense. The usual approach is to differentiate the leading terms of expressions such as (5.1) and (5.2), regarded as a function of h. One should take care to ensure that the leading terms are actually the dominant terms asymptotically. These are the terms which decay to zero at the slowest rate (under the assumptions made) and hence ultimately make up the largest portion of the MSE as $n, m \to \infty$. Of course, for any finite sample, a solution as described above may be suboptimal due to the relative sizes of the constants in the MSE expression. Moreover, the situation is further complicated by the dependence of h on m as well as n. We observe that both m and $n \to \infty$, but they may proceed at different rates. To be complete, we must therefore specify conditions for the behavior of m/n as $n, m \to \infty$. To illustrate the arguments, consider the leading terms of (5.1):

$$v_1/mh + v_2/m + b_1 h^4. \tag{5.9}$$

Solving for the minimizing h^* we obtain

$$h^* = c\, m^{-1/5}. \tag{5.10}$$

For this choice of h, the evaluated MSE in (5.1) yields

$$O(m^{-4/5}) + O(m^{-1}) + O(n^{-1}m^{-3/5}) + o(n^{-1}m^{1/5}) + O(n^{-2}m^{2/5}). \tag{5.11}$$

To guarantee that (5.9) is the dominant term in (5.1), we need to ensure that the corresponding rate of convergence $m^{-4/5}$ is the slowest among the last four terms in (5.11). This will be accomplished if

$$\max \{m^{-1},\ n^{-1}m^{-3/5},\ n^{-1}m^{1/5},\ n^{-2}m^{2/5}\} = o(m^{-4/5}).$$

We observe that the condition

$$m/n = O(1)$$

is sufficient for this purpose. Under this condition, bandwidth (5.10) is optimum among all bandwidths satisfying $nh \to \infty$, $mh \to \infty$ and $h \to 0$, as in Theorem (5.2). The constant c in (5.10) turns out to be

$$c = (v_1/4b_1)^{1/5},$$

where

$$v_1 = \gamma^{(1)}(0+)\ \sigma^2\ D(1,1),$$

$$b_1 = [g^{(3)}(x)\ \mu_2(K_0)/2]^2,$$

and it is assumed that $\gamma^{(1)}(0+) \neq 0$. Therefore if $b_1 \neq 0$, $D(1,1) \neq 0$, and γ is peaked, we have the above result. If $D(1,1) = 0$ or γ is smooth near 0, different methods are required to obtain the asymptotically optimal bandwidth.

REFERENCES

Cheng, K. and Lin, P. (1981a), "Nonparametric Estimation of a Regression Function," Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete, 57, 223-233.

_____ (1981b), "Nonparametric Estimation of a Regression Function: Limiting Distribution," Australian Journal of Statistics, 23, 186-195.

Chung, K. L. (1974), A Course in Probability Theory, Second Edition, New York: Academic Press.

Fulks, W. (1969), Advanced Calculus, Second Edition, New York: John Wiley and Sons.

Gasser, T., and Müller, H. (1979), "Kernel Estimation of Regression Functions," In: Smoothing Techniques for Curve Estimation, eds. T. Gasser and M. Rosenblatt, LNM 757, 23-68.

_____ (1984), "Estimating Regression Functions and Their Derivatives by the Kernel Method," Scandinavian Journal of Statistics, 11, 171-185.

Gasser, T., Müller, H., and Mammitzsch, V. (1985), "Kernels for Nonparametric Curve Estimation," Journal of the Royal Statistical Society, Series B, 47, 238-252.

Georgiev, A. A. (1984a), "Speed of Convergence Nonparametric Kernel Estimation of a Regression Function and Its Derivatives," Annals of the Institute of Statistical Mathematics, 36, A, 455-462.

_____ (1984b), "On the Recovery of Functions and Their Derivatives from Imperfect Measurements," IEEE Transactions on Systems Management Cybernetics, 14, 900-903.

_____ (1984c), "Kernel Estimates of Functions and Their Derivatives with Applications," Statistics & Probability Letters, 2, 45-50.

Hardle, W. and Gasser, T. (1984), "Robust Non-parametric Function Fitting," Journal of the Royal Statistical Society, Series B, 46, 1, 42-51.

Hart, J. D. and Wehrly, T. E. (1986), "Kernel Regression Estimation Using Repeated Measurements Data," _Journal of the American Statistical Association_, 81, 1080-1088.

Huber, P. J. (1981), _Robust Statistics_, New York: John Wiley and Sons.

Schuster, E. F. and Yakowitz, S. (1979), "Contributions to the Theory of Non-parametric Regression, with Application to System Identification," _Annals of Statistics_, 7, 1, 139-149.