UNCLASSIFIED

DTIC FILE COPY

SECURITY CLASSIFICATION OF THIS PAGE

(2)

REPORT DOCUMENTATION PAGE

1a. AD-A198 228	1b. RESTRICTIVE MARKINGS
2a.	3. DISTRIBUTION/AVAILABILITY OF REPORT
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	Distribution unlimited
4. PERFORMING ORGANIZATION REPORT NUMBER(S): Technical Report # 4	5. MONITORING ORGANIZATION REPORT NUMBER(S): AFOSR-TR- 88-0842
6a. NAME OF PERFORMING ORGANIZATION: Texas A&M University; 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION: Air Force Office of Scientific Research
6c. ADDRESS (City, State and ZIP Code): College Station, TX 77853	7b. ADDRESS (City, State and ZIP Code): Same as 8C
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: AFOSR; 8b. OFFICE SYMBOL (If applicable): NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: ~~AFOSR~~ F49620-85-C-0144
8c. ADDRESS (City, State and ZIP Code): Bldg 410, Bolling Air Force Base, Washington, DC 20332	10. SOURCE OF FUNDING NOS.: PROGRAM ELEMENT NO. 61102F; PROJECT NO. 2304; TASK NO. A6; WORK UNIT NO.
11. TITLE (Include Security Classification): VARIANCE FUNCTION ESTIMATION IN REGRESSION: THE EFFECT OF ESTIMATING THE MEAN	
12. PERSONAL AUTHOR(S): Hall, Peter and Carroll, R.J.	
13a. TYPE OF REPORT: ~~technical~~ Journal; 13b. TIME COVERED: FROM 8/87 TO 8/88	14. DATE OF REPORT (Yr., Mo., Day); 15. PAGE COUNT: 15
16. SUPPLEMENTARY NOTATION	
17. COSATI CODES: FIELD / GROUP / SUB. GR.	18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number): heteroscedasticity, nonparametric regression, rates of convergence, variance functions

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

We consider estimation of a variance function g in regression problems. Such estimation requires simultaneous estimation of the mean function f. We obtain sharp results on the extent to which the smoothness of f influences best rates of convergence for estimating g. For example, in nonparametric regression with two derivatives on g, "classical" rates of convergence are possible if and only if the unknown f satisfies a Lipschitz condition of order $\frac{1}{3}$ or more. If a parametric model is known for g, then g may be estimated $n^{\frac{1}{2}}$ - consistently if and only if f is Lipschitz of order $\frac{1}{4}$ or more. Optimal rates of convergence are attained by kernel estimators. (KR)

DTIC ELECTE AUG 2 9 1988 S E D

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION: UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL: Major Brian Woodruff	22b. TELEPHONE NUMBER (Include Area Code): (202) 767-5026; 22c. OFFICE SYMBOL: NM

DD FORM 1473, 83 APR   EDITION OF 1 JAN 73 IS OBSOLETE.

88 8 25 096

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

AFOSR-TR- 88-0842

Variance Function Estimation in Regression: The Effect of Estimating the Mean

Technical Report #4

Peter Hall and R. J. Carroll

Variance Function Estimation in Regression: The Effect of Estimating the Mean

Peter Hall
Department of Statistics
Australian National University
Canberra ACT 2601
Australia

R.J. Carroll
Department of Statistics
Texas A & M University
College Station, TX 77843
USA

SUMMARY

We consider estimation of a variance function g in regression problems. Such estimation requires simultaneous estimation of the mean function f. We obtain sharp results on the extent to which the smoothness of f influences best rates of convergence for estimating g. For example, in nonparametric regression with two derivatives on g, "classical" rates of convergence are possible if and only if the unknown f satisfies a Lipschitz condition of order $\frac{1}{3}$ or more. If a parametric model is known for g, then g may be estimated $n^{\frac{1}{2}}$-consistently if and only if f is Lipschitz of order $\frac{1}{2}$ or more. Optimal rates of convergence are attained by kernel estimators.

Keywords: Heteroscedasticity; Nonparametric Regression; Rates of Convergence; Variance Functions.

1. INTRODUCTION

Consider a heteroscedastic regression problem of the form

$$Y_i = f(x_i) + g(x_i)^{\frac{1}{2}}\epsilon_i\,, \quad 1 \le i \le n\,, \tag{1.1}$$

where the design variables x_i may be either regularly or randomly spaced, and where the ϵ_i's are independent with zero mean and unit variance. Estimation of the variance function g is important in many contexts. Besides the classic need to estimate variance so as to compute weighted least squares estimates of the mean function f, variance function estimates are needed in quality control (Box & Ramirez, 1987); immunoassay (Butt, 1984); prediction, where knowledge of g is required to supply confidence intervals for f (Carroll, 1987); calibration (Watters, Spiegelman & Carroll, 1987); and the estimation of detection limits (Carroll, Davidson & Smith, 1987). These applications are discussed in detail by Carroll & Ruppert (1988). In the present paper we provide a concise description of the effect which not knowing f has on estimation of g.

The results are curious and unexpected. For example, if f is not known parametrically but has at least half a derivative (i.e. satisfies a Lipschitz condition of order $\frac{1}{2}$ or more), then g can be estimated with an accuracy which would be optimal if f were completely known. This result applies to problems where g is known parametrically, and also to problems where g must be estimated nonparametrically. However, the result fails if f is so rough that it does not have half a derivative. There, the roughness of f completely determines the convergence rate if g has known parametric form, and influences the rate if g is known nonparametrically. These remarks apply to optimal estimators of g, as well as to kernel estimators. We show that kernel estimators achieve best possible rates of convergence.

In more detail, the fastest achievable L^2 rate of convergence is

$$\max(n^{-2\nu_2/(2\nu_2+1)}, n^{-4\nu_1/(2\nu_1+1)}) \tag{1.2}$$

if f has ν_1 derivatives and g has ν_2 derivatives. If $\nu_1 \geq \frac{1}{2}$, this equals $n^{-2\nu_2/(2\nu_2+1)}$ and so does not depend on ν_1. Rates in the case where g is known parametrically may be obtained by taking $\nu_2 = \infty$ in (1.2), in which event (1.2) becomes $\max(n^{-1}, n^{-4\nu_1/(2\nu_1+1)})$. The latter equals n^{-1} if $\nu_1 \geq \frac{1}{2}$.

Section 2 presents these conclusions in detail for the case where design points x_i in (1.1) are regularly spaced. Section 3 outlines analogous results for the case of random designs.

2. REGULAR DESIGN

2.1 Introduction. In this section we take the model to be

$$Y_i = f(i/n) + g(i/n)^{\frac{1}{2}}\epsilon_i\,, \quad 1 \leq i \leq n\,, \tag{2.1}$$

where f and g are bounded functions on the interval $[0,1]$, $g \geq 0$, and $\epsilon_1, \epsilon_2, \ldots$ are independent random variables with zero mean, unit variance and uniformly bounded fourth moment. Given $\nu > 0$, write $\langle\nu\rangle$ for the largest integer strictly less than ν. We say that a function a, such as f or g, is ν-smooth if (i) derivatives $a^{(0)}, \ldots, a^{(\langle\nu\rangle)}$ exist and are bounded on [0,1]; and (ii) $a^{(\langle\nu\rangle)}$ satisfies a Lipschitz condition of order $\nu - \langle\nu\rangle$ on [0,1]:

$$|a^{(\langle\nu\rangle)}(x) - a^{(\langle\nu\rangle)}(y)| \leq C|x-y|^{\nu-\langle\nu\rangle}\,, \quad \text{all } x, y \in [0,1]\,.$$

A function with k bounded derivatives on [0,1] is k-smooth.

In subsection 2.2 we show that if f is ν_1-smooth and g is ν_2-smooth, then kernel-type estimators of g converge in mean square at rate $\max(n^{-2\nu_2/(2\nu_2+1)}, n^{-4\nu_1/(2\nu_1+1)})$. Subsection 2.3 demonstrates that if the errors ϵ_i are Gaussian then this rate is optimal, in the sense that no estimator can converge to g more rapidly in mean square. Subsection 2.4 treats the case $\nu_2 = \infty$, which amounts to postulating a parametric model for g.

2.2 Kernel-type estimators. We begin by defining an analogue of a kernel sequence for regular designs. Suppose $0 < h \leq 1$, and $m \geq 0$ is an integer. Let $c_k = c_k(h, m)$,

$-\infty < k < \infty$, be constants satisfying

$$|c_k| \le Ch, c_k = 0 \quad \text{for} \quad |k| \ge Ch^{-1}, \quad \Sigma_k c_k = 1$$
$$\text{and} \quad \Sigma_k k^i c_k = 0 \quad \text{for } 1 \le i \le m, \tag{2.2}$$

where the constant C does not depend on h. Then $\Sigma_k |k|^\alpha |c_k| \le 2C^{\alpha+2} h^{-\alpha}$ for each $\alpha \ge 0$, and $\Sigma_k c_k^2 \le 2C^3 h$. The c_k's may be constructed starting from a smooth function K, vanishing outside the interval $[-1,1]$ and satisfying $\int K(x)\,dx = 1$, $\int x^i K(x)\,dx = 0$ for $1 \le i \le m$. Minor adjustments to K, giving a new function K_1 say, ensure that at least for small h, $c_k = hK_1(hk)$ yields an appropriate sequence of constants. For example, if $m = 0$ or 1, take K to be a bounded, continuous density, symmetric about the origin and vanishing outside $[-1,1]$. Define $\kappa(h)$ by $\kappa(h)^{-1} \equiv \Sigma_k hK(hk)$, so that $\kappa(h) \to 1$ as $h \to 0$. Then $c_k \equiv \kappa(h)hK(hk)$ satisfies our conditions on c_k.

Next we define an estimator of f. Suppose the data Y_i, $1 \le i \le n$, are generated by model (2.1). If the mean function f is ν_1-smooth, choose a sequence of constants $a_k \equiv c_k(h_1, \langle \nu_1 \rangle)$ satisfying condition (2.2), and put

$$\hat{f}(i/n) \equiv \Sigma_k a_k Y_{i+k}, \quad 0 \le i \le n, \tag{2.3}$$

where Y_j is defined to be zero if $j < 1$ or $j > n$. Use linear interpolation on $\hat{f}(i/n)$ to construct $\hat{f}(x)$ for general $x \in [0,1]$. We show in Appendix (i) that if f is ν_1-smooth and g is bounded, and if $h_1 \to 0$ and $nh_1 \to \infty$ as $n \to \infty$, then for each $0 < \delta < \frac{1}{2}$,

$$\sup_{\delta \le x \le 1-\delta} |E\hat{f}(x) - f(x)| = O\{(nh_1)^{-\nu_1}\}, \tag{2.4}$$
$$\sup_{\delta \le x \le 1-\delta} \operatorname{var}\{\hat{f}(x)\} = O(h_1). \tag{2.5}$$

Therefore the mean squared error of $\hat{f}$ satisfies

$$\sup_{\delta \le x \le 1-\delta} E\{\hat{f}(x) - f(x)\}^2 = O\{h_1 + (nh_1)^{-2\nu_1}\}, \tag{2.6}$$

which is minimized at $O(n^{-2\nu_1/(2\nu_1+1)})$ by choosing h_1 to be of size $n^{-2\nu_1/(2\nu_1+1)}$.

Now we construct estimators of g. The estimated residuals are

$$\hat{r}_i \equiv Y_i - \hat{f}(i/n), \quad 1 \leq i \leq n.$$

Our hope is that $\hat{r}_i$ will be close to the "true" residual, $r_i \equiv Y_i - f(i/n) = g(i/n)^{\frac{1}{2}}\epsilon_i$. (Define $r_i = \hat{r}_i = 0$ if $i < 1$ or $i > n$.) Of course, r_i^2 admits the model type (2.1):

$$r_i^2 = g(i/n) + g(i/n)\eta_i, \quad 1 \leq i \leq n, \tag{2.7}$$

where $\eta_i^2 \equiv \epsilon_i^2 - 1$ are independent and identically distributed with zero mean. If the r_i's were observable, we could estimate g from $\{r_i^2\}$ in exactly the same way that we estimated f from $\{Y_i\}$: assuming g to be ν_2-smooth, choose a sequence of constants $b_k \equiv c_k(h_2, (\nu_2))$ satisfying (2.2), and put

$$\tilde{g}(i/n) \equiv \Sigma_k b_k r_{i+k}^2, \quad 1 \leq i \leq n.$$

Construct $\tilde{g}(x)$ by linear interpolation. We see directly from (2.6) that if $h_2 \to 0$ and $nh_2 \to \infty$ then

$$\sup_{\delta \leq x \leq 1-\delta} E\{\tilde{g}(x) - g(x)\}^2 = O\{h_2 + (nh_2)^{-2\nu_2}\}. \tag{2.8}$$

Of course, $\tilde{g}$ is not a realistic estimator, since the true residuals are not observable. If we replace true residuals by their estimates we obtain the practical estimator,

$$\hat{g}(i/n) \equiv \Sigma_k b_k \hat{r}_{i+k}^2, \quad 1 \leq i \leq n. \tag{2.9}$$

Construct $\hat{g}(x)$ by linear interpolation. We show in Appendix (ii) that for each $0 < \delta < \frac{1}{2}$,

$$\sup_{\delta \leq x \leq 1-\delta} E\{\hat{g}(x) - g(x)\}^2 = O\left[\{h_2 + (nh_2)^{-2\nu_2}\} + \{h_1 + (nh_1)^{-2\nu_1}\}^2\right]. \tag{2.10}$$

The second term on the right-hand side of (2.10) distinguishes that expression from (2.8), and is a consequence of our imperfect knowledge about f. Notice that it is the square of the right-hand side of (2.6).

To optimize the rate at which the right-hand side of (2.10) converges to zero, choose h_i of size $n^{-2\nu_i/(2\nu_i+1)}$ for $i = 1$ and 2. Then

$$\sup_{\delta \le x \le 1-\delta} E\{\hat{g}(x) - g(x)\}^2 = O\{\max(n^{-2\nu_2/(2\nu_2+1)}, n^{-4\nu_1/(2\nu_1+1)})\}. \tag{2.11}$$

A necessary and sufficient condition for the term in ν_2 here to dominate, is $4\nu_1/(2\nu_1+1) \ge 2\nu_2/(2\nu_2+1)$, or equivalently,

$$\nu_1 \ge \nu_2/\{2(\nu_2+1)\}. \tag{2.12}$$

Should this condition fail, the rate of convergence of $\hat{g}$ to g is limited by smoothness (or more correctly, lack of smoothness) of f, not by smoothness of g. On the other hand, if (2.12) holds then the rate of convergence of $\hat{g}$ to g is determined by smoothness of g. Note that $\nu_2/\{2(\nu_2+1)\} < \frac{1}{2}$ for all $\nu_2 > 0$, and so condition (2.12) is assured if $\nu_1 \ge \frac{1}{2}$ — that is, if f has at least "half a derivative".

2.3 Optimal rates of convergence. Let $C(\nu, B)$ denoted the class of ν-smooth functions $a : [0,1] \to \mathbb{R}$, such that $\sup |a^{(j)}| \le B$ for $0 \le j \le \langle \nu \rangle$ and

$$|a^{(\langle \nu \rangle)}(x) - a^{(\langle \nu \rangle)}(y)| \le B|x-y|^{\nu - \langle \nu \rangle}, \quad \text{all } x, y \in [0,1].$$

Write $C_+(\nu, B)$ for the set of $a \in C(\nu, B)$ with $a \ge 0$. We showed in Subsection 2.1 that if $f \in C(\nu_1, B)$ and $g \in C_+(\nu_2, B)$, then we may construct a nonparametric estimator $\hat{g}$ of g such that

$$\sup_{\delta \le x \le 1-\delta} E\{\hat{g}(x) - g(x)\}^2 = O\{\max(n^{-2\nu_2/(2\nu_2+1)}, n^{-4\nu_1/(2\nu_1+1)})\}$$

for each $\delta \in (0, \frac{1}{2})$. See (2.11). It is a simple matter to sharpen our proof of this result so that it applies uniformly in f and g:

$$\sup_{f \in C(\nu_1,B), g \in C_+(\nu_2,B)} \sup_{\delta \le x \le 1-\delta} E_{f,g}\{\hat{g}(x) - g(x)\}^2 = O\{\max(n^{-2\nu_2/(2\nu_2+1)}, n^{-4\nu_1/(2\nu_1+1)})\}.$$

We claim that this rate of convergence is best possible, in the following sense. If $\hat{g}$ is *any* nonparametric estimator of g, if $0 < x_0 < 1$, and if the errors ϵ_i are Gaussian, then for some $C > 0$ and all sufficiently large n,

$$M \equiv \sup_{f \in \mathcal{C}(\nu_1, B), g \in \mathcal{C}_+(\nu_2, B)} E_{f,g}\{\hat{g}(x_0) - g(x_0)\}^2 \geq C \max(n^{-2\nu_2/(2\nu_2+1)}, n^{-4\nu_1/(2\nu_1+1)}) . \tag{2.13}$$

This statement is a combination of two results, declaring that

$$M_n \geq C n^{-2\nu_2/(2\nu_1+1)} \tag{2.14}$$

and

$$M_n \geq C n^{-4\nu_1/(2\nu_1+1)} \tag{2.15}$$

respectively. The first of these inequalities has a relatively simple proof, which we now outline. Take $f \equiv 0$, so that we observe the "true" residuals $r_i \equiv g(i/n)^{\frac{1}{2}} \epsilon_i$. The sequence $r_1^2, \ldots, r_n^2$ is sufficient for g. Therefore the problem is that of estimating g under model (2.7). Techniques described by Stone (1980) are easily modified to produce the inequality

$$\sup_{g \in \mathcal{C}_+(\nu_2, B)} E_g\{\hat{g}(x_0) - g(x_0)\}^2 \geq C n^{-2\nu_2/(2\nu_2+1)} ,$$

where $\hat{g}$ is any nonparametric estimator of g based on $r_1^2, \ldots, r_n^2$, and where $f \equiv 0$. This gives (2.14). Appendix (iii) presents a proof of (2.15).

2.4 Parametric model for variance. In some circumstances it is appropriate to consider a parametric model for g, such as $g(x) \equiv \exp(cx + d)$. As far as rates of convergence go, this amounts to taking $\nu_2 = \infty$ in the preceding work, as we now relate.

Suppose g has known parametric form. If f were available we could compute the "true" residuals $r_i \equiv Y_i - f(i/n)$, and from them compute an estimator $\tilde{g}$ satisfying $E\{\tilde{g}(x) - g(x)\}^2 = O(n^{-1})$. More practically, assume f is ν_1-smooth and compute our kernel-type estimator $\hat{f}$, defined at (2.3). Calculate the estimated residuals $\hat{r}_i \equiv Y_i - \hat{f}(i/n)$. Since the constants a_k in (2.3) vanish for $|k| \geq C h_1^{-1}$ (see (2.2)), we avoid "edge effects" by using only those $\hat{r}_i$'s with $Ch_1^{-1} \leq i \leq n - Ch_1^{-1}$. Modify $\hat{g}$ by (i) including only these indices i, and (ii) replacing r_i by $\hat{r}_i$. Call the new estimator $\hat{g}$. Then for each $0 < \delta < \frac{1}{2}$,

$$\sup_{\delta \leq x \leq 1-\delta} E\{\hat{g}(x) - g(x)\}^2 = O\left[n^{-1} + \{h_1 + (nh_1)^{-2\nu_1}\}^2\right] . \qquad (2.16)$$

This is an analogue of (2.10). To optimize the rate of convergence of the right-hand side, choose h_1 to be of size $n^{-2\nu_1/(2\nu_1+1)}$, obtaining

$$\sup_{\delta \leq x \leq 1-\delta} E\{\hat{g}(x) - g(x)\}^2 = O\{\max(n^{-1}, n^{-4\nu_1/(2\nu_1+1)})\} . \qquad (2.17)$$

This is just (2.11) with $\nu_2 = \infty$.

A necessary and sufficient condition for the n^{-1} term to dominate the right-hand side of (2.17), is $\nu_1 \geq \frac{1}{2}$; this is just (2.12) with $\nu_2 = \infty$. If $\nu_1 < \frac{1}{2}$, or equivalently if f has "less than half a derivative", then estimation of even a parametric g is a nonparametric problem with nonparametric rates of convergence. When $\nu_1 = \frac{1}{2}$, $E\{\hat{g}(x) - g(x)\}^2 = O(n^{-1})$, although constants C_1 and C_2 in asymptotic formulae such as

$$E\{\tilde{g}(x) - g(x)\}^2 \sim C_1(x)n^{-1} , \quad E\{\hat{g}(x) - g(x)\}^2 \sim C_2(x)n^{-1}$$

can differ. But when $\nu_1 > \frac{1}{2}$, our imperfect knowledge about f vanishes from the asymptotics, and

$$E\{\tilde{g}(x) - g(x)\}^2 = \{1 + o(1)\}E\{\hat{g}(x) - g(x)\}^2 = O(n^{-1}) \qquad (2.18)$$

as $n \to \infty$. (This result has an analogue in the nonparametric case, when $\nu_1 > \nu_2/\{2(\nu_2 + 1)\}$.)

It is tedious to verify all these formulae in the general case, owing to the wide variety of possible parametric models and associated estimators. We treat only the case $g = g(x) \equiv \sigma^2$ (constant) on [0,1]. Here, $\tilde{g} \equiv n^{-1}\Sigma_{1\leq i\leq n} r_i^2$ and, with m denoting the smallest integer greater than Ch_1^{-1},

$$\hat{g} \equiv (n-2m)^{-1} \sum_{i=m}^{n-m+1} \hat{r}_i^2 = (n-2m)^{-1} \sum_{i=m}^{n-m+1} r_i^2 + (n-2m)^{-1} \sum_{i=m}^{n-m+1} \{\hat{f}(i/n) - f(i/n)\}^2 + 2g^{\frac{1}{2}}(n-2m)^{-1} \sum_{i=m}^{n-m+1} \epsilon_i\{f(i/n) - \hat{f}(i/n)\} .$$

Writing $B_i \equiv E\hat{f}(i/n) - f(i/n)$ for bias, and $\tilde{g}_m \equiv (n-2m)^{-1}\Sigma_{m\le i\le n-m+1} r_i^2$, we obtain

$$(\hat{g}-\tilde{g}_m)^2 \le Cn^{-2}\Bigg[\Bigg(\sum_{i=m}^{n-m+1} B_i^2\Bigg)^2 + \Bigg\{\sum_{i=m}^{n-m+1}(\Sigma_k a_k \epsilon_{i+k})^2\Bigg\}^2 + \Bigg(\sum_{i=m}^{n-m+1} B_i\epsilon_i\Bigg)^2 + \Bigg\{\sum_{i=m}^{n-m+1}\epsilon_i(\Sigma_k a_k \epsilon_{i+k})\Bigg\}^2\Bigg].$$

Now, $|B_i| = O\{(nh_1)^{-2\nu_1}\}$ uniformly in $m \le i \le n-m+1$, and so

$$E(\hat{g}-\tilde{g}_m)^2 = O\big[\{h_1 + (nh_1)^{-2\nu_1}\}^2\big] + o(n^{-1}).$$

Results (2.16)–(2.18) follow from this formula.

The lower bound (2.13), this time with $\nu_2 = \infty$, continues to hold in parametric circumstances such as the one above. In fact, our proof of (2.13) in Appendix (iii) is applicable to the parametric case.

3. RANDOM DESIGN

We now consider kernel regression estimators in the random design case. Let h be the density of the design. Typically, when h is known it is relatively easy to show that the L^2 rate of convergence satisfies (1.2). We concentrate instead on the case of an unknown design density. Under (2.12), we show that one can estimate the variance function g as accurately as though f were known.

Observe independent pairs (Y_i, x_i), $1 \le i \le n$. The x_i's have common density h, and given $\{x_i\}$, $Y_i = f(x_i) + g(x_i)^{\frac{1}{2}}\epsilon_i$. The ϵ_i's are assumed to have mean zero, variance one, and uniformly bounded fourth moments. Given $\nu > 0$, define $\langle\nu\rangle$ and "ν-smoothness" as in Subsection 2.1. Assume f is ν_1-smooth and g is ν_2-smooth, where $\nu_1 > 0$ and $\nu_2 > 0$. Suppose that, uniformly in a neighborhood of x_0, the density d of x is $\{\max(\nu_1,\nu_2)\}$-smooth and bounded away from zero and infinity. For $j = 1, 2$, let K_j be continuous functions with support $[-1,1]$, integrating to one, uniformly Lipschitz continuous of order one, and with i'th moment equal to zero for $1 \le i \le \langle\nu_j\rangle$. Let $h_j \equiv n^{-1/(2\nu_j+1)}$ for $j = 1, 2$.

Define

$$\hat{d}_j(x) \equiv (nh_j)^{-1} \sum_{k=1}^{n} K_j\{(x_k - x)/h_j\}\,, \quad \hat{d}_{1i}(x) \equiv (nh_1)^{-1} \sum_{k \neq i} K_1\{(x_k - x)/h_1\}\,.$$

A kernel regression estimator of f is

$$\hat{f}_i(x) \equiv (nh_1)^{-1} \sum_{k \neq i} Y_k K\{(x_k - x)/h_1\}/\hat{d}_{1i}(x)\,.$$

If the mean function f were known, a kernel regression estimator of g would be

$$\tilde{g}(x) \equiv (nh_2)^{-1} \sum_{i=1}^{n} \{Y_i - f(x_i)\}^2 K_2\{(x_i - x)/h_2\}/\hat{d}_2(x)\,.$$

If f is unknown, the natural analogue of $\tilde{g}$ is

$$\hat{g}(x) \equiv (nh_2)^{-1} \sum_{i=1}^{n} \{Y_i - \hat{f}_i(x_i)\}^2 K_2\{(x_i - x)/h_2\}/\hat{d}_2(x)\,.$$

Classical results on kernel regression function estimation may be used to prove that $|\tilde{g}(x_0) - g(x_0)| = O_p(n^{-\nu_2/(2\nu_2+1)})$; this is the analogue of (2.8) for an optimal choice of window size h_2. In analogy with (2.11),

$$|\hat{g}(x_0) - g(x_0)| = O_p\{\max(n^{-\nu_2/(2\nu_2+1)}, n^{-2\nu_1/(2\nu_1+1)})\}\,. \tag{3.1}$$

As in Section 2, a necessary and sufficient condition for the term in ν_2 here to dominate, is $\nu_1 \geq \nu_2/\{2(\nu_2+1)\}$. If this inequality is strict then $\hat{g}$ is asymptotically equivalent to the "ideal" estimator $\tilde{g}$, in the sense that

$$|\hat{g}(x_0) - \tilde{g}(x_0)| = o_p(n^{-\nu_2/(2\nu_2+1)})\,. \tag{3.2}$$

To prove (3.2), first observe from Stute (1984) that

$$\sup_{|x-x_0|\leq c} \{|\hat{d}_j(x) - d(x)|\} = O_p(n^{-\nu_j/(2\nu_j+1)} \log n)$$

for some $c > 0$. From this it follows that

$$\sup_{|x-x_0|\leq c} \max_{1\leq i\leq n} |\hat{d}_{1i}(x) - d(x)| = O_p(n^{-\nu_1/(2\nu_1+1)} \log n)\,. \tag{3.3}$$

Therefore to prove (3.2) it suffices to show that

$$\max(|A_n|, |B_n|) = o_p(n^{-\nu_2/(2\nu_2+1)}) , \quad (3.4)$$

where

$$A_n \equiv (nh_2)^{-1} \sum_{i=1}^{n} \{\hat{f}_i(x_i) - f(x_i)\}^2 K_2\{(x_i - x_0)/h_2\} ,$$

$$B_n \equiv (nh_2)^{-1} \sum_{i=1}^{n} g(x_i)^{\frac{1}{2}} \epsilon_i \{\hat{f}_i(x_i) - f(x_i)\} K_2\{(x_i - x_0)/h_2\} .$$

Appendix (iv) sketches a proof of (3.4).

The rate of convergence described by (3.1) is optimal. In fact, if the density d is fixed, if $\mathcal{C}(\nu_1, B)$ and $\mathcal{C}_+(\nu_2, B)$ are the function classes defined in Subsection 2.3 but with interval [0,1] replaced by $(-\infty, \infty)$, and if $\hat{g}$ is any nonparametric estimator of g, then for some $C > 0$,

$$\liminf_{n\to\infty} \sup_{f\in\mathcal{C}(\nu_1,B), g\in\mathcal{C}_+(\nu_2,B)} P_{f,g}\{|\hat{g}(x_0) - g(x_0)| > C\max(n^{-\nu_2/(2\nu_2+1)}, n^{-2\nu_1/(2\nu_1+1)})\} > 0 .$$

This is an analogue of (2.13), and has an almost identical proof.

All the results above have versions for parametric estimation of g, corresponding to $\nu_2 = \infty$. In this circumstance we usually do not require parametric knowledge about the design density d, since parametric estimation of g does not involve estimation of d. It is usually sufficient to ask that d be ν_1-smooth.

ACKNOWLEDGEMENT

The work of R.J. Carroll was supported by the U.S. Air Force Office of Scientific Research and performed while visiting the Institute of Advanced Studies at the Australian National University.

Appendix (i): Proof of (2.4) and (2.5).

Since $\hat{f}$ is defined by interpolation from $\hat{f}(i/n)$, it suffices to show that

$$\sup_{\delta n\leq i\leq n-\delta n} |E\hat{f}(i/n) - f(i/n)| = O\{(nh_1)^{-\nu_1}\} , \quad \sup_{n\leq i\leq n-\delta n} \mathrm{var}\{\hat{f}(i/n)\} = O(h_1) . \quad (A.1)$$

Observe from definition (2.3) and properties of $\{a_k\}$ that

$$E\hat{f}(i/n) - f(i/n) = \Sigma_k a_k(\langle\nu_1\rangle!)^{-1}(k/n)^{\langle\nu_1\rangle}[f^{(\langle\nu_1\rangle)}\{(i+\theta_k k)/n\} - f^{(\langle\nu_1\rangle)}(i/n)],$$

where $0 \le \theta_k \le 1$. Since f is ν_1-smooth then $|f^{(\langle\nu_1\rangle)}(x) - f^{(\langle\nu_1\rangle)}(y)| \le C_1|x-y|^{\nu_1-\langle\nu_1\rangle}$, from which it follows that

$$|E\hat{f}(i/n) - f(i/n)| \le C_1\Sigma_k|(k/n)^{\langle\nu_1\rangle}a_k|\,|k/n|^{\nu_1-\langle\nu_1\rangle}$$
$$= C_1 n^{-\nu_1}\Sigma_k|k|^{\nu_1}|a_k| \le C_2(nh_1)^{-\nu_1},$$

which gives the first part of (A.1). The second part follows from

$$\mathrm{var}\,\{\hat{f}(i/n)\} = \Sigma a_k^2 g\{(i+k)/n\} \le (\sup g)\Sigma a_k^2 = O(h_1).$$

Appendix (ii): Proof of (2.10).

Put $D_i \equiv E\hat{f}(i/n) - f(i/n)$, $\Delta_i \equiv \Sigma_k a_k g\{(i+k)/n\}^{\frac{1}{2}}\epsilon_{i+k}$. Then $\hat{r}_i = g(i/n)^{\frac{1}{2}}\epsilon_i - D_i - \Delta_i$, so that $\hat{g}(i/n) - g(i/n) = \Sigma_{1\le j\le 6}S_j$, where

$$S_1 \equiv \Sigma_l b_l g\{(i+l)/n\}(\epsilon_{i+l}^2 - 1), \quad S_2 \equiv \Sigma_l b_l D_{i+l}^2, \quad S_3 \equiv \Sigma_l b_l \Delta_{i+l}^2,$$
$$S_4 \equiv -2\Sigma_l b_l g\{(i+l)/n\}^{\frac{1}{2}} D_{i+l}\epsilon_{i+l}, \quad S_5 \equiv -2\Sigma_l b_l g\{(i+l)/n\}^{\frac{1}{2}}\epsilon_{i+l}\Delta_{i+l},$$
$$S_6 \equiv 2\Sigma_l b_l D_{i+l}\Delta_{i+l}.$$

It suffices to show that

$$\sup_{\delta n\le i\le n-\delta n,\,1\le j\le 6}\left[\{ES_j(i)\}^2 + \mathrm{var}\,S_j(i)\right] = O\{h_2 + (nh_2)^{-2\nu_2} + h_1^2 + (nh_1)^{-4\nu_1}\}. \quad \text{(A.2)}$$

Observe that $E(S_j) = 0$ for $j = 1, 4$ and 6; $|D_i| = O\{(nh_1)^{-\nu_1}\}$, by (A.1); $E(\Delta_i^2) = O(\Sigma a_k^2) = O(h_1)$; and $E(\epsilon_i\Delta_i) = a_0 g(i/n) = O(h_1)$. Therefore $E(S_2) = O\{(nh_1)^{-2\nu_1}\}$, $E(S_3) = O(h_1) = E(S_5)$. Hence, each $(ES_j)^2$ admits the bound claimed in (A.2). Trivially, $\mathrm{var}\,(S_1) = O(\Sigma b_l^2) = O(h_2)$, $\mathrm{var}\,(S_2) = 0$, $\mathrm{var}\,(S_4) = O(\Sigma b_l^2) = O(h_2)$. Furthermore,

$$E(S_3^2) = \Sigma_{l_1}\Sigma_{l_2}\Sigma_{k_1}\ldots\Sigma_{k_4} b_{l_1}b_{l_2}a_{k_1}\ldots a_{k_4}\big[g\{(i+l_1+k_1)/n\}g\{(i+l_1+k_2)/n\}$$
$$\times\, g\{(i+l_2+k_3)/n\}g\{(i+l_2+k_4)/n\}\big]^{\frac{1}{2}}E(\epsilon_{i+l_1+k_1}\epsilon_{i+l_1+k_2}\epsilon_{i+l_2+k_3}\epsilon_{i+l_2+k_4}).$$

The expectation on the right-hand side vanishes unless either $k_1 = k_2$ and $k_3 = k_4$; or $l_1 - l_2 = k_3 - k_1 = k_4 - k_2$; or $l_1 - l_2 = k_4 - k_1 = k_3 - k_2$. In the first case, all nonzero terms except those corresponding to $k_1 = k_2 = k_3 = k_4$, cancel perfectly from the difference $E(S_3^2) - (ES_3)^2$; and in the second and third cases, once l_1, l_2, k_1 and k_2 are given, k_3 and k_4 are completely determined. Therefore, since $|a_k| \le C_1 h_1$,

$$\begin{aligned}\operatorname{var}(S_3) &\le C_2\big(\Sigma_{l_1}\Sigma_{l_2}\Sigma_k |b_{l_1} b_{l_2} a_k| h_1^3 + \Sigma_{l_1}\Sigma_{l_2}\Sigma_{k_1}\Sigma_{k_2} |b_{l_1} b_{l_2} a_{k_1} a_{k_2}| h_1^2\big) \\ &= O(h_1^2)\,.\end{aligned}$$

Similar but simpler arguments show that $\operatorname{var}(S_5) = O(h_1^2 + h_2)$, $\operatorname{var}(S_6) = O\{h_1(nh_1)^{-2\nu_1}\}$. Hence, each $\operatorname{var}(S_j)$ admits the bound claimed in (A.2).

Appendix (iii): Proof of (2.15).

We may assume that $\nu_1 \le \frac{1}{2}$ and $\nu_2 \ge \nu_1$, for otherwise (2.15) follows from (2.14). For simplicity we further suppose that $B > 2$. Let ψ be a nondegenerate, twice-differentiable function on $(-\infty, \infty)$ satisfying $\psi(x) = 0$ for $x \le 0$ and $x \ge 1$, and $\sup|\psi'| \le 1$. Fix $c_1 > 0$, and write m_1, m for integers such that $m_1 \sim c_1 n^{2\nu_1/(2\nu_1+1)}$, $m_1 m \le n$ and $m_1 m \sim n$. Then $m \sim c_1^{-1} n^{1/(2\nu_1+1)}$. Put $\delta_1 \equiv m_1/n$ and $\delta \equiv \delta_1^{2\nu_1}$. Let $I_1, \ldots, I_m$ be a sequence of 0's and 1's, and define $f = f(\cdot \mid I_1, \ldots, I_m)$ by

$$\begin{aligned} f[\{(i-1)m_1 + j\}/n] &= \delta^{\frac{1}{2}} I_i \psi(j/n\delta_1) \quad \text{if } 1 \le i \le m \text{ and } 1 \le j \le m_1\,, \\ f(x) &= 0 \quad \text{if } x \le 0 \text{ or } x \ge m_1 m/n\,. \end{aligned} \tag{A.3}$$

Write $\mathcal{F}$ for the set of all such f's. Define constant functions $g_0 \equiv 1$ and $g_1 \equiv 1 + c_2\delta$, where $c_2 \ne 0$, and let $\mathcal{G} = \{g_0, g_1\}$. For large n, $\mathcal{F} \subseteq C(\nu_1, B)$ and $\mathcal{G} \subseteq C_+(\nu_2, B)$.

We claim that if $0 < x_0 < 1$ and $\hat{g}$ is a nonparametric estimator of g,

$$\sup_{f \in \mathcal{F}, g \in \mathcal{G}} E_{f,g}\{\hat{g}(x_0) - g(x_0)\}^2 \ge C n^{-4\nu_1/(2\nu_1+1)}\,, \tag{A.4}$$

where $C > 0$. It suffices to prove this result for estimators which are functions of Y_i for $i \le m_1 m$. Let $I_1, \ldots, I_m$ be independent symmetric 0–1 variables, independent also of the

ϵ_i's. For these I_i's, write f^* for the (random) function defined as f at (A.3), and let J denote the likelihood ratio rule for discriminating between the hypotheses

$$H_0 : Y_i = f^*(i/n) + g_0(i/n)^{\frac{1}{2}}\epsilon_i\,, \quad H_1 : Y_i = f^*(i/n) + g_1(i/n)^{\frac{1}{2}}\epsilon_i\,.$$

Define $\hat{J} = 0$ if $|\hat{g}(x_0) - g_0(x_0)| \le |\hat{g}(x_0) - g_1(x_0)|$, and $\hat{J} = 1$ otherwise. Write P_i and E_i for probability and expectation under H_i. Then

$$\sup_{f\in\mathcal{F},g\in\mathcal{G}} E_{f,g}\{\hat{g}(x_0) - g(x_0)\}^2 \ge \max_{i=1,2} E_i\{\hat{g}(x_0) - g_i(x_0)\}^2$$
$$\ge (\tfrac{1}{2}c_2\delta)^2 \max\{P_0(\hat{J}=1), P_1(\hat{J}=0)\} \ge \tfrac{1}{8}(c_2\delta)^2\{P_0(\hat{J}=1) + P_1(\hat{J}=0)\}$$
$$\ge \tfrac{1}{8}(c_2\delta)^2\{P_0(J=1) + P_1(J=0)\}\,,$$

by the optimality of the likelihood ratio rule. Therefore (A.4) will follow if we prove

$$\liminf_{n\to\infty} P_0(J=1) > 0\,. \tag{A.5}$$

Let (g, H) denote either (g_0, H_0) or (g_1, H_1). If $k = (i-1)m_1 + j$ where $1 \le i \le m$ and $1 \le j \le m_1$, write Y_{ij} for Y_k and ϵ_{ij} for ϵ_k. Assuming standard normal errors ϵ_{ij}, the likelihood of H given $Y_1, \ldots, Y_{m_1 m}$ is proportional to

$$L(H) \equiv g^{-m_1 m/2} \prod_{i=1}^{m}\left(\exp\left(-\tfrac{1}{2}g^{-1}\sum_{j=1}^{m_1} Y_{ij}^2\right) + \exp\left[-\tfrac{1}{2}g^{-1}\sum_{j=1}^{m_1}\{Y_{ij} - \delta^{\frac{1}{2}}\psi(j/n\delta_1)\}^2\right]\right).$$

If H_0 is true then

$$L(H) = g^{-m_1 m/2}\exp(-\tfrac{1}{2}g^{-1}\Sigma_i\Sigma_j\epsilon_{ij}^2)$$
$$\times\, \Pi_i\big[\exp\{-\tfrac{1}{2}I_i(d_1 + 2d_1^{\frac{1}{2}}N_i)g^{-1}\} + \exp\{-\tfrac{1}{2}(1-I_i)(d_1 - 2d_1^{\frac{1}{2}}N_i)g^{-1}\}\big]\,,$$

where $d_1 \equiv \delta\Sigma_j\psi^2(j/n\delta_1) \sim d \equiv c_1^{2\nu_1+1}\int\psi^2$, and $N_i \equiv d_1^{-\frac{1}{2}}\delta^{\frac{1}{2}}\Sigma_j\psi(j/n\delta_1)\epsilon_{ij}$ is standard normal. Therefore, using the symmetry of N_i,

$$R \equiv 2\log\{L(H_1)/L(H_0)\} = m_1 m(1 - g_1^{-1} + \log g_1^{-1}) - 2(g_1^{-1} - 1)mD + o_p(m_1 m\delta^2 + m\delta)\,,$$

where $D \equiv E[\{1 + \exp(\tfrac{1}{2}d + d^{\frac{1}{2}}N_1)\}^{-1}(\tfrac{1}{2}d + d^{\frac{1}{2}}N_1)]$. Note that

$$|g_1^{-1} - 1||\Sigma_i\Sigma_j(\epsilon_{ij}^2 - 1)| = O_p\{(m_1 m\delta^2)^{\frac{1}{2}}\} = o_p(m_1 m\delta^2)\,.$$

Choose c_1 so that $D \neq 0$, let $c_3 > 0$ and put $c_2 \equiv c_3 \operatorname{sgn}(D)$. Since $g_1 = 1 + c_2\delta$ then

$$R = -\tfrac{1}{2} m_1 m \delta^2 c_3^2 \{1 + o_p(1)\} + m\delta c_3 |D| \{1 + o_p(1)\} .$$

Choose c_3 so small that $c_4 \equiv c_3|D| - \frac{1}{2} c_1^{2\nu_1+1} c_3^2 > 0$. Then $R \sim c_4 m\delta \to \infty$, so that $P_0(J = 1) \to 1$, proving (A.5).

Appendix (iv): Sketch proof of (3.4).

Let $s(x) \equiv f(x)\, d(x)$ and $\hat{s}_i(x) \equiv \hat{f}_i(x) \hat{d}_{1i}(x)$. Assume $\nu_1 > \nu_2/\{2(\nu_2+1)\}$, and put $\xi_n \equiv \max(n^{-2\nu_1/(2\nu_1+1)}, n^{-2\nu_2/(2\nu_2+1)})(\log n)^2$. Equation (3.4) will follow if $|A_n| = O_p(\xi_n)$, $|B_n| = O_p(\xi_n)$. Dropping the argument x,

$$\begin{aligned} \hat{f}_i - f &= (\hat{s}_i - s)/d - (\hat{s}_i - s)(\hat{d}_{1i} - d)/(d\hat{d}_{1i}) - s(\hat{d}_{1i} - d)/(d\hat{d}_{1i}) \\ &= (\hat{s}_i - s)/d - (\hat{s}_i - s)(\hat{d}_{1i} - d)/(d\hat{d}_{1i}) - s(\hat{d}_{1i} - d)/d^2 \\ &\quad + s(\hat{d}_{1i} - d)^2/(d^2 \hat{d}_{1i}) . \end{aligned} \tag{A.6}$$

For A_n, note that

$$(\hat{f}_i - f)^2 \leq 10\{(\hat{s}_i - s)^2/d^2 + (\hat{s}_i - s)^2(\hat{d}_{1i} - d)^2/(d\hat{d}_{1i})^2 + (s/d)^2(\hat{d}_{1i} - d)^2/\hat{d}_{1i}^2\} .$$

This bounds A_n by the sum of three terms, say A_{n1}, A_{n2} and A_{n3}. By (3.3), $A_{n3} = O_p(\xi_n)$. If we show that $A_{n1} = O_p(\xi_n)$, the same easily follows for A_{n2} by (3.3). Define

$$\begin{aligned} v_1(x_i) &\equiv (nh_1)^{-1} \sum_{k \neq i} \{f(x_k) - f(x_i)\} K_1\{(x_k - x_i)/h_1\}/d(x_i) , \\ v_2(x_i) &\equiv (nh_1)^{-1} \sum_{k \neq i} g(x_k)^{\frac{1}{2}} \epsilon_i K_1\{(x_k - x_i)/h_1\}/d(x_i) , \\ v_3(x_i) &\equiv f(x_i)\{\hat{d}_{1i}(x_i) - d(x_i)\}/d(x_i) . \end{aligned}$$

Since $Y_k - f(x_i) = f(x_k) - f(x_i) + g(x_k)^{\frac{1}{2}} \epsilon_k$ then $A_{n1} \leq A_{n11} + A_{n12} + A_{n13}$, where

$$A_{n1j} = 10(n\delta_2)^{-1} \sum_{i=1}^{n} |K_2\{(x_i - x_0)/\delta_2\}| v_j^2(x_i) .$$

By (3.3) for the last and moment calculations for the first two, it is seen that each $A_{n1i} = O_p(\xi_n)$.

To study B_n, split it into four terms $B_{n1} + B_{n2} + B_{n3} + B_{n4}$ based on (A.6). Using (3.3), $B_{n4} = O_p(\xi_n)$. Since $EB_{n3} = 0$, one proves that $B_{n3} = O_p(\xi_n)$ by showing that $\mathrm{var}\,(B_{n3}) = O(\xi_n^2)$, which is an easy calculation. For B_{n2} apply Cauchy-Schwarz, (3.3) and the arguments used to bound A_{n1}, to show that $B_{n2} = O_p(\xi_n)$. This leaves us to study B_{n1}. Now $B_{n1} = B_{n11} + B_{n12} + B_{n13}$, where

$$B_{n1j} = (nh_2)^{-1} \sum_{i=1}^{n} g(x_i)^{\frac{1}{2}} \epsilon_i K_2\{(x_i - x_0)/h_2\} v_j(x_i)\,.$$

Each of these random variables has mean zero and variance $O(\xi_n^2)$, completing the proof.

REFERENCES

Box, G.E.P. & Ramirez, J. (1987). An analysis of the Taguchi method of offline quality control (with discussion). Technometrics, to appear.

Butt, W.R. (1984). Practical Immunoassay. Marcel Dekker, Inc., New York.

Carroll, R.J. (1987). The effect of variance function estimation on prediction intervals. Fourth Purdue Symposium on Statistical Decision Theory and Related Topics, Volume II. J.O. Berger and S.S. Gupta, editors. Springer-Verlag, Heidelberg.

Carroll, R.J., Davidson, M. & Smith, W.C. (1987). Variance functions and the minimum detectable concentration in assays. Preprint.

Carroll, R.J. & Ruppert, D. (1988). Transformations and Weighting in Regression. Chapman & Hall, London.

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

Stute, W. (1984). The oscillation behaviour of empirical processes: the multivariate case. Ann. Prob., 12, 361–379.

Watters, R.L., Spiegelman, C.H. & Carroll, R.J. (1987). Error modeling and confidence interval estimation for inductively coupled plasma calibration curves. Anal. Chem., in press.