Nonparametric Estimation of the Peak Location in a Response Surface

Matthew R. Facer
California Department of Health Services
Sacramento, CA 95814 USA

Hans-Georg Müller
Department of Statistics
University of California, Davis
Davis, CA 95616 USA

Abstract

We explore a nonparametric version of response surface analysis. Estimates for the location where maximum response occurs are proposed and their asymptotic distribution is investigated. The proposed estimates are based on kernel and local least squares methods. We construct asymptotic confidence regions for the location. The methods are illustrated for the two-dimensional case with AIDS incidence data.

Key Words: Nonparametric regression, maximal response, kernel method, local least squares, surface fitting, AIDS incidence

1. Introduction

Estimating the location and size of extrema of a nonparametric regression function often is a motivating factor in fitting such a curve. For the one-dimensional case, peak estimation in a smooth regression function has been explored in Müller (1985, 1989). Extensions to the multivariate situation for the fixed design case, which includes bioassay data, are of general interest. A classic example would be finding the amounts of two or more nutrients that optimize a growth response (Clifford et al., 1993, Müller et al., 1996). Such knowledge optimizes health benefits and maximizes cost-effectiveness. This basic premise motivated many developments within the field of response surface methodology, or RSM (Myers et al., 1989).

RSM is currently a popular method, particularly in industry, to establish conditions that provide maximal yields. Second order parametric models are the norm to find conditions that maximize product yield or quality in disciplines such as chemical engineering (Axelsson et al., 1995), biological/biochemical processes (Roberto et al., 1995; Dey et al., 2001), food science (Kitagawa et al., 1994), engineering (Shyy et al., 2001), air quality (Vogt et al., 1989), or toxicology (Carter et al., 1985). For the special problem of finding the most (cost) effective dose of a drug or drug combination, either in terms of least amount to be effective or tolerance threshold, a univariate nonparametric approach can be found in Rice and Kelly (1990). This compliments parametric approaches by Cox (1987) and Gennings et al. (1990). The field of dose-response modeling with RSM has been substantially developed, particularly with respect to (drug) interactions (see e.g., Hung, 1992, Greco et al., 1994, or Hirst et al., 1996).

The current use of RSM to find conditions that maximize the response has some limitations, which are particularly relevant for the biological sciences. First, the data are assumed to follow a normal distribution, although generalized linear models (McCullagh and Nelder, 1989) are recently being discussed (see e.g., Myers, 1999). Second, the surface and peak are determined by a parametric equation, usually of quadratic type. This implies that all interactions between the predictors are assumed to be of product type. That this assumption is too restrictive is clearly seen in our example of finding the maximum AIDS incidence in terms of age and calendar year which is discussed in section 4. More generally, parametric models have the major weakness of not being flexible in that one equation is assumed to relate the response to the predictors over the entire range of values of the predictors considered.

Most of these problems can be alleviated by using nonparametric regression to generate the response surface, but the study of extrema in this case has been limited. Multivariate nonparametric response surfaces do not rely on distributional assumptions and allow the effects of and interactions between the predictors to vary over the range of the predictors, providing flexibility for the resulting surfaces (Müller, 1988). The use of nonparametric regression within RSM has recently been recognized among its practitioners as an important part of the future direction of RSM (Myers, 1999). However, optimization, a major consideration in RSM practice, has received very little attention for nonparametric response surfaces. We note that in typical RSM applications, the levels of predictor combinations would be fixed in advance, so that a fixed design regression model is appropriate. The case for one predictor (point estimates and asymptotic properties) was studied by Müller (1985, 1989), but to our knowledge the multivariate case has received less attention. The extension from one predictor to several is nontrivial, as will be demonstrated in the following.

As an example in a biological setting, we will use AIDS surveillance data to find the maximum incidence of AIDS in California. AIDS incidence is modeled as a function of age (in years) and calendar time (year) for the two racial/ethnic groups with the most AIDS cases in California (Latinos and Whites). A confidence region for maximum AIDS incidence for each of these groups is constructed, and compared with the quadratic RSM approach.

The paper is organized as follows: In section 2 we collect preliminary results on multivariate nonparametric regression, emphasizing the fixed design case, and results from matrix theory, in particular with respect to matrix norms. Drawing on these, we obtain asymptotic results for the extrema of multivariate nonparametric regression functions (using kernel type smoothers), including the derivation of $100(1-\alpha)\%$ confidence regions for the true location of extrema of a multivariate nonparametric regression function in section 3. Using AIDS surveillance data as an example, we provide an illustration of these results and comparisons with the RSM approach in section 4. Finally, some further discussion and concluding remarks can be found in section 5. Proofs and auxilary results are relegated to the Appendix.

2. Preliminaries

We consider here the standard set up for fixed design multivariate nonparametric regression (see Müller, 1988). The model is:

$$y_i = g(\mathbf{x}_i) + \epsilon_i, \tag{2.1}$$

where $\mathbf{x}_i \in A_i \subset \mathcal{X} \subseteq \Re^m$, $m \geq 1$, for a smooth regression function g. With λ representing Lebesgue measure in $\Re^m$, we assume that the domain $\mathcal{X}$ of the data $\mathbf{x}_i$ is compact, connected, and measurable with $0 < \lambda(\mathcal{X}) < \infty$, and that $\{A_i\}_{\{i=1,\ldots,n\}}$ is a partition of $\mathcal{X}$ into n measurable, connected subsets (hence $A_i \bigcap A_j = \emptyset$, $i \neq j$). The data $(\mathbf{x}_i, y_i)$, $i = 1, \ldots, n$ are used to estimate the multivariate regression function $g : \mathcal{X} \rightarrow \Re$ in (2.1); denote this estimate as $\hat{g}$. The assumptions for the errors ϵ_i, $i = 1, \ldots, n$, are that for some $r \geq 2$,

(M1) $E(\epsilon_i) = 0$, $Var(\epsilon_i) = \sigma^2 < \infty$, $E|\epsilon_i|^r \leq c < \infty$, and the ϵ_i are iid.

Let $\nu = (\nu_1, \ldots, \nu_m)$ be a multiindex indicating that the ν_j^{th} partial derivative is to be taken in the j^{th} direction. As a special case, define $\alpha_i = (0, \ldots, 0, 1, 0, \ldots, 0)$, where 1 occurs at the i^{th} position (and the other m-1 elements are 0). Define $|\nu| = \sum_{i=1}^m \nu_i$ (hence $|\alpha_j|$=1), $\nu! = \nu_1!\nu_2! \cdots \nu_m!$, and $\nu = \tau$ as $\nu_i = \tau_i$, $i = 1, \ldots, m$. For $\mathbf{z} \in \Re^m$, define $\mathbf{z}^\tau = z_1^{\tau_1} \cdots z_m^{\tau_m}$. Let $\mathbf{b} \in \Re^m$ denote the bandwidth vector used to construct the estimate $\hat{g}$; we shall assume that $b_1 = \ldots = b_m = b$ with $b := b(n)$, a sequence of bandwidths satisfying $b \rightarrow 0$ as $n \rightarrow \infty$. We consider the following kernel type estimator (Müller, 1988) for $g^{(\nu)}(\mathbf{x})$:

$$\hat{g}^{(\nu)}(\mathbf{x}) = \frac{1}{b^{|\nu|+m}} \sum_{i=1}^n [\int_{A_i} K_\nu(\frac{\mathbf{x} - \mathbf{s}}{b}) d\mathbf{s}] y_i \tag{2.2}$$

where $K_\nu : \mathcal{X} \rightarrow \Re$ is a kernel function, with $K_\nu(\frac{\mathbf{x}-\mathbf{s}}{b})$ denoting $K_\nu(\frac{x_1-s_1}{b_1}, \ldots, \frac{x_m-s_m}{b_m})$.

Further assumptions are needed about the fixed design. Let $x_i \in A_i$, where the sets A_i, $i = 1, \ldots, n$, form a partition of $\mathcal{X}$, with

(M2) $\max_{\{1 \leq i \leq n\}} |\lambda(A_i) - \lambda(\mathcal{X}) n^{-1}| = o(n^{-1})$

(M3) $\max_{\{1 \leq i \leq n\}} \sup_{\{w, z \in A_i\}} \|w - z\|_2 = O(n^{-1/m})$;

here $\| \cdot \|_2$ is the Euclidean norm in $\Re^m$.

Let $Lip(\mathcal{X})$ denote the set of Lipschitz continuous functions on $\mathcal{X}$ and $\mathcal{C}^k(\mathcal{X})$ the set of k times continuously differentiable functions on $\mathcal{X}$, for an integer $k > |\nu|$ which defines the smoothness of g, and ultimately, if coupled with a kernel of order k, the rate of convergence $MSE \sim n^{-2(k-|\nu|)/(2k+m)}$. We require the following properties for $g^{(\nu)}$, K_ν, and b:

(M4) For the regression function g, $g^{(\nu)} \in \mathcal{C}^k(\mathcal{X})$ and $g^{(\nu)}$, $g^{(\nu+\alpha_i+\alpha_j)} \in Lip(\mathcal{X})$ for each $i, j = 1, \ldots, m$ and $k \geq 3$.

(M5) For the sequence of bandwidths $b = b(n)$, $n^{1/m}b \to \infty$ and $\gamma_n^2 = nb^{2(|\nu|+1)+m} \to \infty$ as $n \to \infty$.

(M6) For the kernel function K_ν, $K_\nu : \mathcal{T} \to \Re$ with support $\mathcal{T} \subset \Re^m$, where $\mathcal{T}$ is compact, connected, and λ-measurable with $\lambda(\mathcal{T}) = 1$ and $K_\nu \in Lip(\mathcal{T})$. Furthermore, K_ν is a kernel of the order $(|\nu|, k)$, i.e., satisfies the moment properties

(**M7**)
$$\int_{\mathcal{T}} K_\nu(\mathbf{z})\mathbf{z}^\tau d\mathbf{z} = \begin{cases} 0, & \text{if } 0 \le |\tau| \le |\nu|, \tau \neq \nu \\ (-1)^{|\nu|}\nu!, & \text{if } \tau = \nu \\ 0, & \text{if } |\nu| < |\tau| < k \end{cases}$$

We now quote some preliminary background results. It is assumed that any fixed point $\mathbf{x}$ is in the interior of $\mathcal{X}$ (that is, $\mathbf{x} \in \overset{o}{\mathcal{X}}$) to avoid boundary effects. The first two results can be found in Müller and Prewitt (1993), while the third is a straightforward extension of a theorem in Müller and Stadtmüller (1987). The more general conditions in Lemma 2.3 can be derived along the lines of the univariate case found in Theorem 11.2 in Müller (1988). The first result provides the asymptotically leading terms of the mean squared error of the multivariate kernel regression estimator.

Lemma 2.1 Assume (M1)-(M7) hold, and for $\mathbf{z} = \frac{\mathbf{x}-\mathbf{s}}{b}$ let

(2.3)
$$V = \int_{\mathcal{T}} K_\nu^2(\mathbf{z})d\mathbf{z}.$$

Then

$$\begin{aligned} \mathrm{Var}(\hat{g}^{(\nu)}(\mathbf{x})) &= \frac{\sigma^2\lambda(\mathcal{X})}{nb^{2|\nu|+m}}[V + o(1)], \\ E(\hat{g}^{(\nu)}(\mathbf{x})) &= b^{-|\nu|}\int_{\mathcal{T}} K_\nu(\mathbf{z})g(\mathbf{x} - \mathbf{z}b)d\mathbf{z}[1 + o(1)] + O(\frac{1}{n^{1/m}b^{|\nu|}}), \\ E(\hat{g}^{(\nu)}(\mathbf{x})) - g^{(\nu)}(\mathbf{x}) &= \sum_{|\rho|=k} b^{|\rho|-|\nu|}[\frac{g^{(\rho)}(\mathbf{x})(-1)^{|\rho|}}{\rho!}\int_{\mathcal{T}} K_\nu(\mathbf{z})\mathbf{z}^\rho d\mathbf{z} + o(1)] + O(\frac{1}{n^{1/m}b^{|\nu|}}). \end{aligned}$$

The second result provides the asymptotic normality of the multivariate kernel estimator.

Lemma 2.2 Assume (M1)-(M7) hold and $nb^{2k+m} \to d^2 \geq 0$ as $n \to \infty$. Then for $\mathbf{x} \in \overset{o}{\mathcal{X}}$,

$$[nb^{2|\nu|+m}]^{1/2}[\hat{g}^{(\nu)}(\mathbf{x}) - g^{(\nu)}(\mathbf{x})] \xrightarrow{\mathcal{D}} \mathcal{N}(d \sum_{|\rho|=k} \frac{g^{(\rho)}(\mathbf{x})(-1)^{|\rho|}}{\rho!} \int_{\mathcal{T}} K_\nu(\mathbf{z})\mathbf{z}^\rho d\mathbf{z} \ , \sigma^2\lambda(\mathcal{X})V).$$

The third result is on uniform convergence.

Lemma 2.3 Assume (M1)-(M7) hold, $E|\epsilon_i|^r < \infty$ for some $r > 2$, $\lim\inf_{n\to\infty} nb^{mk} > 0$, and $\lim\inf_{n\to\infty} n^{1-\frac{2}{r}} b^m [logn]^{-1} > 0$. Then

$$\sup_{\{\mathbf{x}\in\mathcal{X}\}} |\hat{g}^{(\nu)}(\mathbf{x}) - g^{(\nu)}(\mathbf{x})| = O_p(b^{k-|\nu|} + [\frac{logn}{nb^{2|\nu|+m}}]^{1/2}).$$

Note that the assumptions of Lemmas 2.2 and 2.3 constrain the value of k relative to the dimension m. In particular, if $b \sim n^{-q}$, then the assumptions

$$nb^{2k+m} \ \to \ d^2 \ \geq \ 0 \text{ as } n \ \to \ \infty \quad \text{and} \quad \lim\inf_{n\to\infty} nb^{mk} > 0$$

require that $km \ \leq \ 2k + m$. While for the case $m = 2$ there is no restriction on k, if $m = 3$ then $k \leq 3$, if $m = 4$ then $k \leq 2$, and if $m \geq 5$ then $k \leq 1$, which implies $k = 1$.

A result that extends the univariate case of nonparametric extrema studied in Müller (1985) to the multivariate case will be established next. We assume there is a unique point $\theta \in \mathcal{X}$ that maximizes $g^{(\nu)}(\cdot)$; that is, $\theta = \text{argmax}_{\{\mathbf{x}\in\mathcal{X}\}} g^{(\nu)}(\mathbf{x})$. The empirical maximum $\hat{\theta}$ is the point in $\mathcal{X}$ that maximizes the nonparametric estimate $\hat{g}^{(\nu)}(\cdot)$; that is, $\hat{\theta} = \max_{\{\mathbf{x}\in\mathcal{X}\}} \hat{g}^{(\nu)}(\mathbf{x})$. The proof of the following result is in the Appendix.

Lemma 2.4 Assume (M1)-(M7) hold and that for a null sequence β_n it holds that

$$\sup_{\{\mathbf{x}\in\mathcal{X}\}} |\hat{g}^{(\nu)}(\mathbf{x}) - g^{(\nu)}(\mathbf{x})| = O_p(\beta_n), \tag{2.4}$$

and moreover that for any m-dimensional ϵ-ball $B_\epsilon(\theta)$ surrounding θ, there exists $c > 0$ such that

$$|g^{(\nu)}(\mathbf{x}) - g^{(\nu)}(\theta)| > c\|\mathbf{x} - \theta\|_2 \text{ for } \mathbf{x} \in B_\epsilon(\theta). \tag{2.5}$$

Then

$$\|\hat{\theta}-\theta\|_2 = O_p(\beta_n) \tag{2.6}$$

and

$$|\hat{g}^{(\nu)}(\hat{\theta}) - g^{(\nu)}(\theta)| = O_p(\beta_n). \tag{2.7}$$

Combining this with Lemma 2.3, one immediately obtains a consistency result for the estimation of the peak location. The asymptotic distribution of the peak location will be investigated in the next section in more detail.

3. Limit Distribution of the Peak Location Estimate

The asymptotics of the estimate $\hat{\theta} = \max_{\{\mathbf{x}\in\mathcal{X}\}}\hat{g}^{(\nu)}(\mathbf{x})$ of $\theta = \operatorname{argmax}_{\{\mathbf{x}\in\mathcal{X}\}}g^{(\nu)}(\mathbf{x})$ follow from those of $\hat{g}^{(\nu)}$ combined with the local geometry of the curve near its extremum. By (2.1) and the Multivariate Mean Value Theorem (recall $\alpha_i = (0,\dots,0,1,0,\dots,0)$, where 1 occurs at the i^{th} position and the other m-1 elements are 0)

$$0 = \hat{g}^{(\nu+\alpha_i)}(\hat{\theta}) = \hat{g}^{(\nu+\alpha_i)}(\theta) + [\nabla\hat{g}^{(\nu+\alpha_i)}(\theta_i^*)]^T(\hat{\theta}-\theta), \quad i = 1,\dots,m \tag{3.1}$$

for some mean values $\theta_i^* \in \mathcal{X}$ such that $\theta_i^* = (1-\lambda_i)\theta + \lambda_i\hat{\theta}, \quad 0 \le \lambda_i \le 1$. Using $g^{(\nu+\alpha_i)}(\theta) = \hat{g}^{(\nu+\alpha_i)}(\hat{\theta}) = 0$ in (3.1) yields

$$g^{(\nu+\alpha_i)}(\theta) - \hat{g}^{(\nu+\alpha_i)}(\theta) = [\nabla\hat{g}^{(\nu+\alpha_i)}(\theta_i^*)]^T(\hat{\theta}-\theta). \tag{3.2}$$

Defining

$$\mathbf{d}_n = \begin{pmatrix} \hat{g}^{(\nu+\alpha_1)}(\theta) - g^{(\nu+\alpha_1)}(\theta) \\ \vdots \\ \hat{g}^{(\nu+\alpha_m)}(\theta) - g^{(\nu+\alpha_m)}(\theta) \end{pmatrix}_{m\times 1},$$

$$A = \begin{pmatrix} [\nabla g^{(\nu+\alpha_1)}(\theta)]^T \\ \vdots \\ [\nabla g^{(\nu+\alpha_m)}(\theta)]^T \end{pmatrix}_{m\times m}, \quad B_n = \begin{pmatrix} [\nabla\hat{g}^{(\nu+\alpha_1)}(\theta_1^*)]^T \\ \vdots \\ [\nabla\hat{g}^{(\nu+\alpha_m)}(\theta_m^*)]^T \end{pmatrix}_{m\times m},$$

we may rewrite (3.2) as

$$\mathbf{d}_n = B_n(\hat{\theta} - \theta) = A[(\hat{\theta} - \theta) + R_n], \tag{3.3}$$

where $R_n = [A^{-1} - B_n^{-1}]\mathbf{d}_n$. Setting $\gamma_n^2 = nb^{2(|\nu|+1)+m}$, the asymptotic behavior of $\gamma_n \mathbf{d}_n$ is of interest, and we also need to show $\gamma_n R_n \xrightarrow{p} 0$. For the latter, we need some additional assumptions. Define

$$K_\nu^{(\alpha_i)}(\mathbf{z}) = \frac{\partial}{\partial z_i} K_\nu(\mathbf{z}) \qquad \text{and} \qquad K_\nu^{(\alpha_i+\alpha_j)}(\mathbf{z}) = \frac{\partial^2}{\partial z_i \partial z_j} K_\nu(\mathbf{z}),$$

and assume $\theta, \hat{\theta} \in \overset{o}{\mathcal{X}}$ throughout. Note that if $K_\nu^{(\alpha_i)} \in Lip(\mathcal{T})$, applying an integration by parts argument to (M6) and (M7) yields

$$\int_{\mathcal{X}} K_\nu^{(\alpha_i)}(\mathbf{z})\mathbf{z}^\tau d\mathbf{z} = \begin{cases} 0, & \text{if } \ 0 \le |\tau| \le |\nu + \alpha_i| = |\nu| + 1,\ \tau \ne \nu + \alpha_i \\ (-1)^{|\nu|}\nu!(\nu_i + 1), & \text{if } \ \tau_j = \nu_j,\ j \ne i,\ \tau_i = \nu_i + 1 \\ 0, & \text{if } \ |\nu + \alpha_i| < |\tau| < k + 1 \end{cases} \tag{3.4}$$

and therefore the results of Lemmas 2.1 - 2.4 can be applied with ν and k replaced with $\nu + \alpha_i$ and $k+1$, respectively. A similar result holds if $K_\nu^{(\alpha_i+\alpha_j)} \in Lip(\mathcal{T})$ where ν and k are replaced by $\nu + \alpha_i + \alpha_j$ (with $|\nu + \alpha_i + \alpha_j| = |\nu| + 2$) and $k+2$, respectively. We add the following assumption:

(M8) $\quad K_\nu^{(\alpha_i)}\ ,\ K_\nu^{(\alpha_i+\alpha_j)} \in Lip(\mathcal{T})$ for all $i, j = 1, \ldots, m$,

and another assumption on the bandwidths will be:

(M9) $\quad$ For a constant $d \ge 0$, $\frac{logn}{nb^{2|\nu|+m+4}} \to 0$ and $nb^{2k+m+2} \to d^2 \ge 0$ as $n \to \infty$.

Define

$$\beta(\rho, \alpha_i) = \frac{1}{\rho!}\int_T K_\nu^{(\alpha_i)}(\mathbf{z})\mathbf{z}^\rho d\mathbf{z}$$

$$V(\alpha_i, \alpha_j) = \int_T K_\nu^{(\alpha_i)}(\mathbf{z})K_\nu^{(\alpha_j)}(\mathbf{z})d\mathbf{z}\ ,\ 1 \le i, j \le m.$$

We are now ready to state our main result.

Theorem 3.1 Under (M1)-(M9),

$$\gamma_n(\hat{\theta} - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}_m(dA^{-1}\mu, A^{-1}\Sigma[A^{-1}]^T)$$

where

$$\mu = [\mu_1, \ldots, \mu_m]^T, \quad \mu_j = (-1)^{k+1} \sum_{|\rho|=k+1;\rho_j \geq 1} g^{(\rho)}(\theta)\beta(\rho, \alpha_j) \ , \ j = 1, \ldots, m$$

$$A = \begin{pmatrix} [\nabla g^{(\nu+\alpha_1)}(\theta)]^T \\ \vdots \\ [\nabla g^{(\nu+\alpha_m)}(\theta)]^T \end{pmatrix}_{m\times m}, \quad \Sigma = \lambda(\mathcal{X})\sigma^2 \begin{pmatrix} V(\alpha_1, \alpha_1) \cdots V(\alpha_1, \alpha_m) \\ \vdots \quad \ddots \quad \vdots \\ V(\alpha_m, \alpha_1) \cdots V(\alpha_m, \alpha_m) \end{pmatrix}.$$

A primary application of this result is in the construction of a confidence region for θ. As $\hat{\theta}$ has an asymptotic Gaussian distribution, the $100(1-\alpha)\%$ confidence region will be based on the $(1-\alpha)^{th}$ percentile of the χ^2_m distribution, based on standard multivariate normality theory. This is stated explicitly in the following two corollaries.

Corollary 3.1 Let (M1)-(M9) hold. Then

$$\gamma_n\{A^{-1}\Sigma[A^{-1}]^T\}^{-1/2}_{m\times m}[\hat{\theta} - \theta - dA^{-1}\mu] \quad \xrightarrow{\mathcal{D}} \quad \mathcal{N}_m(0, I_m), \tag{3.5}$$

where A, Σ, and μ are as stated in Theorem 3.1 and I_m is the $m \times m$ identity matrix.

Corollary 3.2 Let (M1)-(M9) hold. As $n \to \infty$, a $100(1-\alpha)\%$ confidence region for $\theta \in \Re^m$, the maximizer of the function $g^{(\nu)}$ in (2.1) estimated by $\hat{\theta}$, the maximizer (or minimizer) of the nonparametric kernel type estimate in (2.2), is given by

$$[\hat{\theta} - \theta - dA^{-1}\mu]^T\{A^{-1}\Sigma[A^{-1}]^T\}^{-1}[\hat{\theta} - \theta - dA^{-1}\mu] \quad \leq \quad \chi^2_m(1-\alpha), \tag{3.6}$$

where $\chi^2_m(1-\alpha)$ is the $100(1-\alpha)$ percentile of the Chi square distribution with m degrees of freedom and A, Σ, and μ are as in Theorem 3.1.

We conclude this section by noting that confidence regions (3.6) can be greatly simplified when $V(\alpha_i, \alpha_j) = 0$ for each $i \neq j$, $i, j = 1, \ldots, m$. This case is of particular interest because it occurs when one uses product kernels, which are discussed briefly in the appendix.

Corollary 3.3 Assume (M1)-(M9) hold. If $V(\alpha_i, \alpha_j) = 0$ and $V(\alpha_i, \alpha_i) > 0$ for all $i \neq j$, with $i, j = 1, \ldots, m$, then the asymptotic confidence regions (3.6) may be written as

$$[\hat{\theta} - \theta - dA^{-1}\mu]^T \; H \; [\hat{\theta} - \theta - dA^{-1}\mu] \;\; \leq \;\; \chi^2_m(\alpha), \tag{3.7}$$

where

$$H \;\; = \;\; \frac{1}{\lambda(\mathcal{X})\sigma^2} \begin{pmatrix} \sum_{j=1}^m \frac{g^{(\nu+\alpha_j+\alpha_1)}(\theta) g^{(\nu+\alpha_j+\alpha_1)}(\theta)}{V(\alpha_j,\alpha_j)} & \cdots & \sum_{j=1}^m \frac{g^{(\nu+\alpha_j+\alpha_1)}(\theta) g^{(\nu+\alpha_j+\alpha_m)}(\theta)}{V(\alpha_j,\alpha_j)} \\ \vdots & \ddots & \vdots \\ \sum_{j=1}^m \frac{g^{(\nu+\alpha_j+\alpha_m)}(\theta) g^{(\nu+\alpha_j+\alpha_1)}(\theta)}{V(\alpha_j,\alpha_j)} & \cdots & \sum_{j=1}^m \frac{g^{(\nu+\alpha_j+\alpha_m)}(\theta) g^{(\nu+\alpha_j+\alpha_m)}(\theta)}{V(\alpha_j,\alpha_j)} \end{pmatrix}. \tag{3.8}$$

Applying these results in practice requires the substitution of the unknown quantities d, σ^2, and $g^{(\rho)}(\theta)$ for several different indices ρ (with $|\rho| \geq 1$) that appear in the asymptotic confidence regions. A natural estimate for $g^{(\rho)}(\theta)$ is $\hat{g}^{(\rho)}(\theta)$, which we know from Lemma 2.4 to converge in probability to $g^{(\rho)}(\theta)$, although the rate depends on $|\rho|$ (in particular, we would need to assume $\frac{logn}{nb^{2|\rho|+m}} \to 0$ as $n \to \infty$). To estimate σ^2, a residual sum of squares similar to parametric regression models can be used:

$$\hat{\sigma}^2 \;\; = \;\; \frac{1}{n} \sum_{i=1}^n \{y_i - \hat{g}(x_i)\}^2.$$

An alternative approach which was implemented for the data analysis in section 4 is to estimate $\hat{\sigma}^2$ via a binning technique. Using the framework of section 2, the j^{th} set in this partition consists of κ_j neighboring sets of the original partition $\{A_i\}_{\{i=1,\ldots,n\}}$ of the data. The new partition then contains n^* sets composed of κ_j elements for $j = 1, \ldots, n^*$. The estimator is given by

$$\hat{\sigma}^2 \;\; = \;\; \frac{1}{n^*} \sum_{j=1}^{n^*} \frac{1}{\kappa_j - 1} \sum_{l=1}^{\kappa_j} (y_{jl} - \overline{y}_l)^2 \;\;\; , \;\; \overline{y}_l \;\; = \;\; \frac{1}{\kappa_j} \sum_{l=1}^{\kappa_j} y_{jl}. \tag{3.9}$$

Under (M1)-(M3), Müller and Prewitt (1993) showed $\hat{\sigma}^2 \;\stackrel{p}{\to}\; \sigma^2$.

4. Application: Confidence Regions for Peak Location of AIDS Incidence in California

In a data application example, we compare here the confidence region for the maximizing argument of a parametric response surface as obtained by fitting a quadratic regression surface with that of the nonparametric model developed above, for the case of two predictors ($m = 2$). The data considered here are AIDS case surveillance data from the California Department of Health Services, Office of AIDS. All cases of AIDS diagnosed in California are reported, along with demographic information such as age, date, and county of residence at diagnosis, and any AIDS-defining illnesses and their date of diagnosis. Over 120,000 cases of AIDS have been diagnosed in California and reported as of January 1, 2001. We will consider examining the peak of AIDS incidence (number of AIDS cases occurring in a specified population per year per 100,000 of that population) in California among different racial/ethnic groups with respect to calendar time and age at AIDS diagnosis as predictors.

AIDS incidence was calculated each year between 1985 and 1995 for each age between 20 and 60 among Whites and Latinos in California (the two largest ethnic groups in the State) from the AIDS surveillance data base and census-based population projections available from the California Department of Finance. This provided two data sets for Whites and Latinos that include as response AIDS incidence and as predictors calendar time (in years, and recorded as 0 = 1985, 1 = 1986, ..., 10 = 1995) and age at AIDS diagnosis (in years). These data are available from the authors upon request.

4.1 Parametric Confidence Region for Two Predictors

The quadratic response surface model is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \epsilon, \tag{4.1}$$

where we assume that (M1) from section 2 holds for ϵ, and that (4.1) is to be fitted by the iid data (y_i, x_{1i}, x_{2i}), $i = 1, \ldots, n$ via least squares. The peak coordinates are found to be

$$\theta_{P1} = \frac{2\beta_{22}\beta_1 - \beta_{12}\beta_2}{\beta_{12}^2 - 4\beta_{11}\beta_{22}} \quad \text{and} \quad \theta_{P2} = \frac{2\beta_2\beta_{11} - \beta_{12}\beta_1}{\beta_{12}^2 - 4\beta_{11}\beta_{22}}. \tag{4.2}$$

We will assume that these are the coordinates of a maximum. Under regulatory conditions, we will have by the delta method, letting $\Sigma(\beta)$ be the limiting covariance matrix of parameter estimates $\hat{\beta} = [\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_{11}, \hat{\beta}_{22}, \hat{\beta}_{12}]^T$,

$$n^{1/2}\begin{pmatrix}\hat{\theta}_{P1}-\theta_{P1}\\ \\ \hat{\theta}_{P2}-\theta_{P2}\end{pmatrix}\xrightarrow{\mathcal{D}}\mathcal{N}_2(\begin{pmatrix}0\\ \\ 0\end{pmatrix},\mathcal{P}\Sigma(\beta)\mathcal{P}^T)$$

$$\text{where } \mathcal{P}^T=\begin{pmatrix}\frac{\partial\theta_{P1}}{\partial\beta_0} & \frac{\partial\theta_{P1}}{\partial\beta_1} & \frac{\partial\theta_{P1}}{\partial\beta_2} & \frac{\partial\theta_{P1}}{\partial\beta_{11}} & \frac{\partial\theta_{P1}}{\partial\beta_{22}} & \frac{\partial\theta_{P1}}{\partial\beta_{12}}\\ \\ \frac{\partial\theta_{P2}}{\partial\beta_0} & \frac{\partial\theta_{P2}}{\partial\beta_1} & \frac{\partial\theta_{P2}}{\partial\beta_2} & \frac{\partial\theta_{P2}}{\partial\beta_{11}} & \frac{\partial\theta_{P2}}{\partial\beta_{22}} & \frac{\partial\theta_{P2}}{\partial\beta_{12}}\end{pmatrix}_{\beta=\hat{\beta}}.$$

In particular, if

$$[\mathcal{P}\hat{\Sigma}(\hat{\beta})\mathcal{P}^T]^{-1}=\begin{pmatrix}\hat{\sigma}_{11} & \hat{\sigma}_{12}\\ \\ \hat{\sigma}_{21} & \hat{\sigma}_{22}\end{pmatrix}, \tag{4.3}$$

an asymptotic $100(1-\alpha)\%$ confidence region for $[\theta_{P1},\theta_{P2}]^T$ is given by

$$\begin{pmatrix}\hat{\theta}_{P1}-\theta_{P1}\\ \\ \hat{\theta}_{P2}-\theta_{P2}\end{pmatrix}^T\begin{pmatrix}\hat{\sigma}_{11} & \hat{\sigma}_{12}\\ \\ \hat{\sigma}_{21} & \hat{\sigma}_{22}\end{pmatrix}\begin{pmatrix}\hat{\theta}_{P1}-\theta_{P1}\\ \\ \hat{\theta}_{P2}-\theta_{P2}\end{pmatrix}\le\chi_2^2(1-\alpha). \tag{4.4}$$

The quadratic response surface (4.1) was fitted to the AIDS incidence data with predictor x_1 chosen as calendar time (1985 through 1995, recoded as 0 through 10), and predictor x_2 as age at AIDS diagnosis (20 through 60, in years); the response y is AIDS incidence. These parametric surfaces (one each for California's Latino and White population) are shown in Figure 1. We see that the maxima of these surfaces are located at different points. For Latinos the maximum AIDS incidence occurred at $[\hat{\theta}_{time}\ ,\ \hat{\theta}_{age}]\ =\ [1993.3\ ,\ 40.1$ years], while among Whites at $[\hat{\theta}_{time}\ ,\ \hat{\theta}_{age}]\ =\ [1991.5\ ,\ 40.4$ years]. The confidence regions for peak AIDS incidence over time and age were calculated via (4.2)-(4.4) and are presented in section 4.3.

4.2 Nonparametric Surface Estimate for AIDS Incidence Data

To generate the nonparametric response surface estimates for the AIDS incidence data, we used product kernels (see Appendix) that were implemented by linear locally weighted least squares (see Appendix), with the bandwidths calculated via the cross-validation method. The kernels used were $K(\mathbf{z})\ =\ K(z_1,z_2)\ =\ (1-z_1^2)^2(1-z_2^2)^2$ with $z_1,z_2\ \in\ [-1,1]$.

The nonparametric response surfaces for AIDS incidence among the Latino and White populations in California are shown in Figure 2. The cross-validation bandwidth for the Latino AIDS incidence data was found to be $[b_{time}\ ,\ b_{age}] \ = \ [1.65 \text{ years}\ ,\ 4.66 \text{ years}]$, and that for Whites to be $[b_{time}\ ,\ b_{age}] \ = \ [1.13 \text{ years}\ ,\ 2.56 \text{ years}]$. The larger bandwidth for the Latino data is largely attributable to the increased sparseness of the data as compared to the data for Whites, particularly for older ages early in the epidemic. According to these nonparametric surfaces, the maximum AIDS incidence for Latinos occurred at $[\hat{\theta}_{time}\ ,\ \hat{\theta}_{age}] \ = \ [1992.25\ ,\ 34.0 \text{ years}]$, while among Whites it occurred at $[\hat{\theta}_{time}\ ,\ \hat{\theta}_{age}] \ = \ [1991.75\ ,\ 36.0 \text{ years}]$.

4.3 Parametric and Nonparametric Confidence Regions for AIDS Incidence Data

We construct parametric 95% confidence regions for the age and calendar time of peak AIDS incidence in California using the results of sections 4.1 and 4.2. From (4.4), we obtain

$$\text{Latinos:} \qquad \begin{pmatrix} \theta_{time} - 1993.3 \\ \theta_{age} - 40.1 \end{pmatrix}^T \begin{pmatrix} 8.16 & 1.30 \\ 1.30 & 11.87 \end{pmatrix} \begin{pmatrix} \theta_{time} - 1993.3 \\ \theta_{age} - 40.1 \end{pmatrix} \ \leq \ 5.99$$

$$\text{Whites:} \qquad \begin{pmatrix} \theta_{time} - 1991.5 \\ \theta_{age} - 40.4 \end{pmatrix}^T \begin{pmatrix} 38.39 & 4.20 \\ 4.20 & 17.37 \end{pmatrix} \begin{pmatrix} \theta_{time} - 1991.5 \\ \theta_{age} - 40.4 \end{pmatrix} \ \leq \ 5.99$$

These regions are plotted in Figure 3 (the two highest ellipses). As these regions do not overlap, we would conclude that under the parametric modeling assumptions the age-specific peak incidence of AIDS among Latinos and Whites occurred at different times in California with 95% confidence. In particular, AIDS incidence among Whites peaked significantly earlier than among Latinos in California.

Nonparametric confidence regions for the age and calendar time of maximum AIDS incidence among Latinos and Whites in California were obtained using Lemma A.7 in the Appendix, Corollary 3.3, and estimates for unknown quantities obtained from the data. For simplicity, we assumed $d \ = \ 0$ (so that the estimates for $\theta \ = \ [\theta_{time}\ ,\ \theta_{age}]$ are assumed asymptotically unbiased). As discussed in section 4, we used $\hat{g}^{(\rho)}(\hat{\theta})$ to estimate $g^{(\rho)}(\theta)$, where in this case $\nu = (0,0)$ and $m = 2$, so that the values of ρ to be considered are $\rho = (1,0)$, $(0,1)$, $(1,1)$, $(2,0)$, $(0,2)$. The bandwidths used in these calculations were 1.25 times those of the cross-validation bandwidths estimated to generate $\hat{g}$ (for Latinos 1.25 $\times$ [1.65 years, 4.66 years], and for Whites 1.25 $\times$ [1.13 years, 2.56 years]); slightly increased bandwidths for derivative estimation were suggested in Müller (1988). More details about the estimates can be found in the Appendix. For each racial/ethnic group, $\hat{\sigma}^2$ was calculated using (3.9), for which we obtained $\hat{\sigma}^2 = 70.19$ for the Latino group and $\hat{\sigma}^2 = 147.98$ for the White group. For both groups, $\lambda(\mathcal{X}) = 10 \times 40 = 400$ (where 10 is from the 10-year period 1985-1995 and 40 from the width of the age range 20-60). As we assumed $d = 0$, the primary concern is the estimation of the variance-covariance structure of $\hat{\theta}$, for which the relevant estimates are shown in the Appendix.

Using these estimates and noting that $\chi^2_2(0.05) = 5.99$, the nonparametric 95% confidence regions of the age-specific calendar times of peak AIDS incidence were found to be defined by

$$\text{Latinos:} \qquad \begin{pmatrix} \theta_{time} - 1992.25 \\ \theta_{age} - 34.0 \end{pmatrix}^T \begin{pmatrix} 45.70 & 18.61 \\ 18.61 & 8.01 \end{pmatrix} \begin{pmatrix} \theta_{time} - 1992.25 \\ \theta_{age} - 34.0 \end{pmatrix} \leq 5.99$$

$$\text{Whites:} \qquad \begin{pmatrix} \theta_{time} - 1991.75 \\ \theta_{age} - 36.0 \end{pmatrix}^T \begin{pmatrix} 25.04 & -22.02 \\ -22.02 & 32.92 \end{pmatrix} \begin{pmatrix} \theta_{time} - 1991.75 \\ \theta_{age} - 36.0 \end{pmatrix} \leq 5.99$$

These regions are plotted in Figure 3 (the two lowest ellipses). These regions do overlap each other, and we would conclude that under the nonparametric response surface scheme, there was no significant difference between the age-specific time of peak AIDS incidence among Latinos and Whites in California. Thus we arrive at a different conclusion from that reached by fitting the second order parametric response surface model of section 4.1.

5. Concluding Remarks

A comparison of point estimates and associated confidence regions for the calendar time and age of diagnosis for the peak location in AIDS incidence between the second order parametric response surface model from section 4.1 and the nonparametric approach from section 4.2 reveals several interesting features. First, the nonparametric confidence regions are larger than their respective parametric regions, which is expected as the nonparametric fits have larger variances. Second, the nonparametric models are expected to have smaller biases, and the nonparametric ellipses in Figure 3 suggest that AIDS incidence peaked among a younger segment of both Latinos and Whites in California than suggested by the parametric fit (6 years younger for Latinos, 4 years younger for Whites). Third, the nonparametric approach suggests the peak in AIDS incidence occurred about one year earlier in the Latino group (in early 1992) than suggested by the parametric model (in early 1993), though the nonparametric confidence region extends out to late 1993. Both models agree in predicting that peak AIDS incidence occurred during the second half of 1991 for the White group.

Comparing parametric and nonparametric models, discrepancies emerge regarding the timing of peak AIDS incidence among Latinos and Whites. The parametric models suggest AIDS incidence peaked significantly earlier for Whites than Latinos (by about 1.8 years), with virtually no age difference between the two populations (Figure 3). In contrast, the nonparametric models indicate that the peaks occurred much nearer to each other (Whites about 0.5 years, or 6 months, earlier than Latinos), and that the age in the Latino population where the peak occurred was slightly younger (34) than in the White population (36). From epidemiologic considerations, it appears that the results obtained with the nonparametric approach reflect the underlying situation much better than those obtained with the parametric approach.

Acknowledgements

We thank the California Department of Health Services for permission to use the AIDS data. This research was supported in part by NSF grant DMS-9971602.

Appendix

This appendix consists of six subsections. The first contains the derivation of an auxiliary result in matrix perturbation theory that is needed in the proof of the main theorem. The second subsection contains further auxiliary results and proofs, including those needed to establish the main theorem, the proof of the main theorem itself, and proofs of the subsequent confidence region expressions. The third subsection gives a brief overview of product kernels, while the fourth provides a general outline of the asymptotic equivalence of locally weighted least squares estimators and those of kernel-type in the fixed design case. The fifth and sixth subsections detail the calculations used to generate the response surfaces and confidence regions in Figures 1-3.

A.1 Preliminary Results from Matrix Perturbation Theory

Recall $\|\cdot\|_2$ is the Euclidean norm in $\Re^m$, and let $\mathcal{M}_m(\Re)$ denote the space of all $m \times m$ real-valued matrices. Let $\|\cdot\|$ denote the matrix norm

$$\|\cdot\| : H \subseteq \Re^{m\times m} \rightarrow \Re \text{ with } \|H\| = \sup_{\{\mathbf{x}\in\Re^m:\|\mathbf{x}\|_2\leq 1\}} \|H\mathbf{x}\|_2$$

We cite two results from Lewis (1991) (Lemmas A.1 and A.2) and establish a third (Lemma A.3).

Lemma A.1 (Banach) Let $W \in M_m(\Re)$ and $\|\cdot\|$ a matrix norm on $M_m(\Re)$. If $\|W\| < 1$, then $I + W$ is invertible and

$$\frac{1}{1+\|W\|} \leq \|(I+W)^{-1}\| \leq \frac{1}{1-\|W\|}.$$

Lemma A.2 Let $V \in M_m(\Re)$ be invertible. If V is perturbed into a matrix $V + P$ where $\|P\| < \frac{1}{\|V^{-1}\|}$ for some matrix norm $\|\cdot\|$, then $V + P$ is invertible.

Lemma A.3 Let $A \in M_m(\Re)$ be invertible. For all $B \in M_m(\Re)$ such that

$$(A.1) \qquad \|A - B\| < \frac{1}{2\|A^{-1}\|},$$

B^{-1} exists and there exists $0 < c < \infty$ such that

$$(A.2) \qquad \|B^{-1} - A^{-1}\| \le c\|A^{-1}\|^2\|A - B\|.$$

Proof. Setting $V = A$ and $P = B - A$ in Lemma A.2 establishes the existence of B^{-1}. To show (A.2), note

$$\|B^{-1} - A^{-1}\| \le \|A^{-1}\|\|AB^{-1} - I\| \le \|A^{-1}\|\|A - B\|\|B^{-1}\|.$$

Rewrite $B = A[I - A^{-1}(A - B)]$. By (A.1),

$$(A.3) \qquad \|A^{-1}(A - B)\| < \frac{1}{2} \ ,$$

and Lemma A.1 shows $I - A^{-1}(A - B)$ is invertible. We can then write

$$\begin{aligned}
\|B^{-1} - A^{-1}\| &\le \|A^{-1}\|\|A - B\|\|(I - A^{-1}(A - B))^{-1}A^{-1}\| \\
&\le \|A^{-1}\|^2\|A - B\|\|(I - A^{-1}(A - B))^{-1}\| \\
&\le \frac{\|A^{-1}\|^2\|A - B\|}{1 - \|A^{-1}(A - B)\|} \\
&\le c\|A^{-1}\|^2\|A - B\|,
\end{aligned}$$

by taking $W = -A^{-1}(A - B)$ in Lemma A.1 and observing (A.3). □

A.2 Auxiliary Results and Proofs

We list several lemmas that are needed for the proof of Theorem 3.1 and proofs of results that are stated in the main sections (including Theorem 3.1 itself).

Lemma A.4 Assume (M1)-(M8) hold. Let A, B_n, and γ_n be as in Section 3, and assume A^{-1} exists. If $\frac{logn}{nb^{2|\nu|+m+4}} \to 0$, then $\|A - B_n\| = O_p(b^{k-\|\nu\|} + [\frac{logn}{nb^{2|\nu|+m+4}}]^{1/2})$.

Proof. By definition

$$\|A - B_n\| \quad = \quad \sup_{\{\mathbf{x}\in\mathcal{X}:\|\mathbf{x}\|_2\le 1\}}\| \begin{pmatrix} [\nabla g^{(\nu+\alpha_1)}(\theta) - \nabla \hat{g}^{(\nu+\alpha_1)}(\theta_1^*)]^T \\ \vdots \\ [\nabla g^{(\nu+\alpha_m)}(\theta) - \nabla \hat{g}^{(\nu+\alpha_m)}(\theta_m^*)]^T \end{pmatrix} \mathbf{x}\|_2$$

$$\leq \quad \sup_{\{\mathbf{x}\in\mathcal{X}:\|\mathbf{x}\|_2\leq 1\}}\|U\mathbf{x}\|_2$$

where $U_{m\times m} = [u_{ij}]$, $1 \leq i, j \leq m$ (here i and j correspond to the row and column, respectively) with

$$u_{ij} = \frac{\partial^2 g^{(\nu)}(\theta)}{\partial x_i \partial x_j} - \frac{\partial^2 \hat{g}^{(\nu)}(\theta_i^*)}{\partial x_i \partial x_j}.$$

Hence $\sup_{\{\mathbf{x}\in\mathcal{X}:\|\mathbf{x}\|_2\leq 1\}}\|U\mathbf{x}\|_2 \quad \leq \quad \{\sum_{i=1}^m (\max_{\{j:1\leq j\leq m\}}|[\mathbf{u}_i]_j^T|)^2\}^{1/2} \quad = \quad m \max_{\{1\leq i,j\leq m\}}|u_{ij}|$. Applying the triangle inequality, we note that

$$|u_{ij}| = |g^{(\nu+\alpha_i+\alpha_j)}(\theta) - \hat{g}^{(\nu+\alpha_i+\alpha_j)}(\theta_i^*)|$$

$$(A.4) \qquad \leq |g^{(\nu+\alpha_i+\alpha_j)}(\theta) - g^{(\nu+\alpha_i+\alpha_j)}(\theta_i^*)| + |g^{(\nu+\alpha_i+\alpha_j)}(\theta_i^*) - \hat{g}^{(\nu+\alpha_i+\alpha_j)}(\theta_i^*)|$$

Arguing as in (3.4) for $K_\nu^{(\alpha_i+\alpha_j)}$, noting $K_\nu^{(\alpha_i+\alpha_j)} \in Lip(\mathcal{X})$, and applying Lemma 2.3 to this case yields

$$|g^{(\nu+\alpha_i+\alpha_j)}(\theta_i^*) - \hat{g}^{(\nu+\alpha_i+\alpha_j)}(\theta_i^*)| = O_p(\beta_n^*) \quad \text{for all } i, j,$$

where

$$\beta_n^* \quad = \quad b^{k+2-(|\nu|+|\alpha_i|+|\alpha_j|)} + [\frac{logn}{nb^{2(|\nu|+|\alpha_i|+|\alpha_j|)+m}}]^{1/2} \quad = \quad b^{k-|\nu|} + [\frac{logn}{nb^{2|\nu|+m+4}}]^{1/2}.$$

Furthermore, we note that $\|\theta - \theta_i^*\|_2 \leq \|\theta - \hat{\theta}\|_2 = O_p(\beta_n^*)$ by Lemma 2.4. Applying the Continuous Mapping Theorem, the first term of (A.4) is bounded by $O_p(\beta_n^*)$. Thus $|u_{ij}| = O_p(\beta_n^*)$ for all i, j, and the result follows. □.

Lemma A.5 Under (M1)-(M9), $\gamma_n\|(A^{-1} - B_n^{-1})\mathbf{d}_n\| \xrightarrow{p} 0$.

Proof. By the Cauchy-Schwarz Inequality, $\gamma_n\|(A^{-1} - B_n^{-1})\mathbf{d}_n\| \leq \gamma_n\|(A^{-1} - B_n^{-1})\|\|\mathbf{d}_n\|$. By Lemma A.4, $P[\|A - B_n\| < \frac{1}{2\|A^{-1}\|}] \to 1$ as $n \to \infty$. Hence by Lemma A.3 it suffices to show $\gamma_n\|A - B_n\|\|\mathbf{d}_n\| \xrightarrow{p} 0$. Now notice that

$$\|\mathbf{d}_n\| \quad = \quad \sup_{\{\mathbf{x}\in\mathcal{X}:\|\mathbf{x}\|_2\leq 1\}}\| \begin{pmatrix} \hat{g}^{(\nu+\alpha_1)}(\theta) - g^{(\nu+\alpha_1)}(\theta) \\ \vdots \\ \hat{g}^{(\nu+\alpha_m)}(\theta) - g^{(\nu+\alpha_m)}(\theta) \end{pmatrix} \mathbf{x}\|_2 \quad = \quad O_p([\frac{1}{nb^{2[|\nu|+1]+m}}]^{1/2})$$

by Lemma 2.2, using a derivative of order $|\nu|+1$. Combining this with (2.7) and Lemma A.4 gives

$$\begin{aligned}\gamma_n\|A-B_n\|\|\mathbf{d}_n\| &\leq [nb^{2(|\nu|+1)+m}]^{1/2}\ O_p(b^{k-|\nu|}+[\frac{logn}{nb^{2|\nu|+m+4}}]^{1/2})\ O_p([\frac{1}{nb^{2[|\nu|+1]+m}}]^{1/2}) \\ &\xrightarrow{p} 0. \quad \square\end{aligned}$$

Lemma A.6 If (M1)-(M8) hold, then for all $i \neq j$, $i, j = 1, \ldots, m$, with $B(\hat{g}^{(\nu+\alpha_i)}) = E[\hat{g}^{(\nu+\alpha_i)}(\theta)] - g^{(\nu+\alpha_i)}(\theta)$:

$$(A.5) \qquad B(\hat{g}^{(\nu+\alpha_i)}) = b^{k-|\nu|}(-1)^{k+1}\sum_{|\rho|=k+1;\rho_i\geq 1}[g^{(\rho)}(\theta)\beta(\rho,\alpha_i)+o(1)]+O[\frac{1}{n^{1/m}b^{|\nu|+1}}]$$

$$(A.6) \qquad \mathrm{Var}[\hat{g}^{(\nu+\alpha_i)}(\theta)] = \frac{\sigma^2\lambda(\mathcal{X})}{nb^{2|\nu|+m+2}}[V(\alpha_i,\alpha_i)+o(1)]$$

$$(A.7) \qquad \mathrm{Cov}[\hat{g}^{(\nu+\alpha_i)}(\theta),\hat{g}^{(\nu+\alpha_j)}(\theta)] = \frac{\sigma^2\lambda(\mathcal{X})}{nb^{2|\nu|+2m+2}}[V(\alpha_i,\alpha_j)+o(1)]$$

Proof. (A.5) and (A.6) follow directly from Lemma 2.1 by substituting $|\nu+\alpha_i| = |\nu|+1$ and $k+1$ for $|\nu|$ and k, respectively, while (A.7) follows from similar arguments as in Lemma 6.3 of Müller and Prewitt (1993). $\square$

Proof of Lemma 2.4. We will consider the case of maxima. As $\mathrm{argmax}_{\{\mathbf{x}\in\mathcal{X}\}}g^{(\nu)}(\mathbf{x})$ is unique, according to (2.5) there exists $\delta > 0$ such that $g^{(\nu)}(\theta) > g^{(\nu)}(\mathbf{x}) + \delta$ for $\mathbf{x} \notin B_\epsilon(\theta)$. Rewriting,

$$(A.8) \qquad g^{(\nu)}(\theta) - \frac{\delta}{2} > g^{(\nu)}(\mathbf{x}) + \frac{\delta}{2} \text{ for } \mathbf{x} \notin B_\epsilon(\theta).$$

As $\beta_n \to 0$, (2.4) implies

$$P[\sup_{\{\mathbf{x}\in\mathcal{X}\}}|\hat{g}^{(\nu)}(\mathbf{x}) - g^{(\nu)}(\mathbf{x})| \leq \frac{\delta}{2}] \to 1, \text{ as } n\to\infty.$$

Hence,

$$P[g^{(\nu)}(\theta) - \frac{\delta}{2} \leq \hat{g}^{(\nu)}(\theta) \leq g^{(\nu)}(\theta) + \frac{\delta}{2}] \to 1, \text{ as } n\to\infty.$$

Combining this with (A.8) yields

$$P[\hat{g}^{(\nu)}(\theta) \geq g^{(\nu)}(\theta) - \frac{\delta}{2} > g^{(\nu)}(\mathbf{x}) + \frac{\delta}{2} > g^{(\nu)}(\mathbf{x})] \to 1, \text{ as } n\to\infty \text{ for } \mathbf{x}\notin B_\epsilon(\theta),$$

which implies $P[\hat{\theta} \in B_\epsilon(\theta)] \to 1,$ as $n \to \infty$. Applying (2.5) for $\hat{\theta}$, recalling that θ is the unique maximum of $g^{(\nu)}$, and using the triangle inequality then yields

$$\begin{aligned} \|\hat{\theta} - \theta\|_2 &< \frac{1}{c}[g^{(\nu)}(\theta) - g^{(\nu)}(\hat{\theta})] \\ &\leq \frac{1}{c}[(\hat{g}^{(\nu)}(\hat{\theta}) - \hat{g}^{(\nu)}(\theta)) + g^{(\nu)}(\theta) - g^{(\nu)}(\hat{\theta})] \\ &\leq \frac{1}{c}[|\hat{g}^{(\nu)}(\hat{\theta}) - g^{(\nu)}(\hat{\theta})| + |\hat{g}^{(\nu)}(\theta) - g^{(\nu)}(\theta)|] = O_p(\beta_n), \text{ by (2.4).} \end{aligned}$$

Result (2.7) follows from the triangle inequality combined with (2.4). That is,

$$|\hat{g}^{(\nu)}(\hat{\theta}) - g^{(\nu)}(\theta)| \leq \{g^{(\nu)}(\theta) - g^{(\nu)}(\hat{\theta})\} + |\hat{g}^{(\nu)}(\hat{\theta}) - g^{(\nu)}(\hat{\theta})| = O_p(\beta_n). \qquad \square$$

Proof of Theorem 3.1. Redistributing terms in (3.3) yields

$$\gamma_n A^{-1}\mathbf{d}_n - \gamma_n R_n = \gamma_n(\hat{\theta} - \theta).$$

By Lemma A.5, $\gamma_n R_n \stackrel{p}{\to} 0$, and therefore by Slutsky's Theorem and the assumption that A^{-1} exists, it suffices to show $\gamma_n \mathbf{d}_n \stackrel{\mathcal{D}}{\to} \mathcal{N}_m(\mu, \Sigma)$. Note $\gamma_n \mathbf{d}_n = \gamma_n \phi(\theta) + \gamma_n \psi(\theta)$, where

$$\phi(\theta) = [\phi_1, \ldots, \phi_m]^T, \quad \phi_i = \hat{g}^{(\nu+\alpha_i)}(\theta) - E[\hat{g}^{(\nu+\alpha_i)}(\theta)], \quad i = 1, \ldots, m$$

$$\psi(\theta) = [\psi_1, \ldots, \psi_m]^T, \quad \psi_i = E[\hat{g}^{(\nu+\alpha_i)}(\theta)] - g^{(\nu+\alpha_i)}(\theta), \quad i = 1, \ldots, m$$

Note that by the standard Taylor expansion for the bias

$$\psi_i = b^{k-|\nu|}(-1)^{k+1} \sum_{|\rho|=k+1;\rho_i \geq 1} [g^{(\rho)}(\theta)\beta(\rho, \alpha_i) + o(1)] + O[\frac{1}{n^{1/m} b^{|\rho|+1}}]$$

so that $\gamma_n \psi(\theta) \to d\mu$ uniformly in $\mathcal{X}$. For the random part $\gamma_n \phi(\theta)$, we apply the Cramer-Wold device and show that for any $\mathbf{a} \in \Re^m$

$$\mathbf{a}^T \gamma_n \phi(\theta) \stackrel{\mathcal{D}}{\to} \mathcal{N}(0, S_\theta) \tag{A.9}$$

with

$$S_\theta = \sigma^2 \lambda(\mathcal{X}) \sum_{j=1}^m a_j^2 V(\alpha_j, \alpha_j) + 2\sigma^2 \lambda(\mathcal{X}) \sum_{j=1}^m \sum_{k \neq j}^m a_j a_k V(\alpha_j, \alpha_k).$$

By definition and (2.3)

$$\mathbf{a}^T \gamma_n \phi(\theta) \quad = \quad \gamma_n \sum_{j=1}^m a_j \{\hat{g}^{(\nu+\alpha_j)}(\theta) - E[\hat{g}^{(\nu+\alpha_j)}(\theta)]\} \quad = \quad \sum_{i=1}^n W_i \epsilon_i \ ,$$

where

$$W_i \quad = \quad \frac{\gamma_n}{b^{|\nu+\alpha_j|+m}} \sum_{j=1}^m a_j [\int_{A_i} K_\nu^{(\alpha_j)}(\frac{t-s}{b}) d\mathbf{s}] \ , \quad \epsilon_i \quad = \quad y_i \ - \ E(y_i).$$

Using for example Theorem 4.2 in Müller (1988), the Lindeberg condition implies $\frac{\sum_{i=1}^n W_i \epsilon_i}{(\sigma^2 \sum_{i=1}^n W_i^2)^{1/2}} \ \stackrel{\mathcal{D}}{\rightarrow} \ \mathcal{N}(0,1)$ if

$$(A.10) \qquad G(W_i) \quad = \quad \frac{\max_{\{1 \le i \le n\}} |W_i|}{(\sum_{i=1}^n W_i^2)^{1/2}} \ \rightarrow \ 0 \ \text{ as } \ n \rightarrow \infty.$$

For the numerator of $G(W_i)$, the triangle inequality, $K_\nu^{(\alpha_i)} \in Lip(\mathcal{X})$, and mean values ξ_{ij} (with $i = 1, \ldots, n$, $j = 1, \ldots, m$) yield with (2.4) and (2.5)

$$\begin{aligned}
\max_{\{1 \le i \le n\}} |W_i| \quad &\le \quad \max_{\{1 \le i \le n\}} |\frac{[nb^{2|\nu|+2+m}]^{1/2}}{b^{|\nu|+m+1}}| \sum_{j=1}^m |a_j| | \int_{A_i} K_\nu^{(\alpha_j)}(\frac{t-s}{b}) d\mathbf{s}| \\
&= \quad [\frac{\lambda(\mathcal{X})}{[nb^m]^{1/2}} + o(\frac{1}{[nb^m]^{1/2}})] \sum_{j=1}^m |a_j| O(\frac{1}{n^{1/m} b}) \rightarrow 0 \text{ by (2.7).}
\end{aligned}$$

For the denominator of $G(W_i)$, note

$$\begin{aligned}
\sigma^2 \sum_{i=1}^n W_i^2 \quad &= \quad \text{Var}(\gamma_n \sum_{j=1}^m a_j \{\hat{g}^{(\nu+\alpha_j)}(\theta) - E[\hat{g}^{(\nu+\alpha_j)}(\theta)]\}) \\
&= \quad \frac{\gamma_n^2 \sigma^2 \lambda(\mathcal{X})}{nb^{2|\nu|+2+m}} \{\sum_{j=1}^m a_j^2 [V(\alpha_j, \alpha_j) + o(1)] \quad + \quad 2 \sum_{j=1}^m \sum_{k \ne j}^m a_j a_k [V(\alpha_j, \alpha_k) + o(1)]\} \\
&\rightarrow \quad S_\theta \text{ by Lemma A.6.}
\end{aligned}$$

Therefore, (A.10) and as a consequence, (A.9) are satisfied, whence the result follows. □

Proof of Corollary 3.3. By Corollary 3.2, the confidence region is given by (3.6). Now $V(\alpha_i, \alpha_j) = 0$ and $V(\alpha_i, \alpha_i) > 0$ for all $i \ne j$, $i, j = 1, \ldots, m$ implies that Σ^{-1} exists. As A is symmetric, $\{A^{-1} \Sigma [A^{-1}]^T\}^{-1} \ = \ A \Sigma^{-1} A$. Note that

$$A \Sigma^{-1} A = \frac{1}{\lambda(\mathcal{X}) \sigma^2} \begin{pmatrix} [\nabla g^{(\nu+\alpha_1)}(\theta)]^T \\ \vdots \\ [\nabla g^{(\nu+\alpha_m)}(\theta)]^T \end{pmatrix} \begin{pmatrix} \frac{1}{V(\alpha_1,\alpha_1)} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{V(\alpha_m,\alpha_m)} \end{pmatrix} \begin{pmatrix} [\nabla g^{(\nu+\alpha_1)}(\theta)]^T \\ \vdots \\ [\nabla g^{(\nu+\alpha_m)}(\theta)]^T \end{pmatrix} \quad = \quad H. \ \square$$

A.3 Product Kernels

We discuss here the special case where the m-dimensional kernel (with assumptions (M6) and (M7)) can be written as a product of m univariate kernels,

$$(A.11) \qquad K_\nu(\mathbf{z}) \;=\; \prod_{j=1}^{m} K_{\nu_j}(z_j).$$

It is assumed that the m univariate kernels K_{ν_j} are such that for each $j = 1, \ldots, m$ the univariate equivalents of (M6) and (M7) hold, namely

[P1] K_{ν_j} has support $[-\tau_j, \tau_j]$, $\tau_j \;>\; 0$

[P2] $K_{\nu_j} \in Lip([-\tau_j, \tau_j])$

[P3]
$$\int_{\mathcal{T}} K_{\nu_j}(z_j) z_j^\tau dz_j = \begin{cases} 0, & \text{if } 0 \le \tau < k + \nu_j - |\nu|, \, \tau \neq \nu_j \\ (-1)^{\nu_j} \nu_j!, & \text{if } \tau = \nu_j \\ \beta_j \neq 0, & \text{if } \tau = k + \nu_j - |\nu| \end{cases}$$

As discussed in Müller and Prewitt (1993), under these conditions for (A.11), it follows that $\mathcal{T} = \prod_{j=1}^{m}[-\tau_j, \tau_j]$, $K_\nu \in Lip(\mathcal{T})$, and K_ν satisfies (M6) and (M7). Extending these results further, we note that if

[P4] $K_{\nu_j} \in \mathcal{C}^k$ and $K'_{\nu_j}, K''_{\nu_j} \in Lip([-\tau_j, \tau_j])$ for each $j = 1, \ldots, m$,

then K_{ν_j} satisfies (M8). As a consequence, the results of section 3 are valid when using the product kernel set up of (A.11) with the assumptions [P1]-[P4]. This leads to the following result that is relevant in calculating confidence regions for θ as discussed in Corollaries 3.2 and 3.3.

Lemma A.7 Assume (M1)-(M5) and (M9) hold, and that the kernel K_ν in (2.2) is of the form in (A.11) such that [P1]-[P4] hold. Then

$$V(\alpha_i, \alpha_j) \;=\; \int_{\mathcal{T}} K_\nu^{(\alpha_i)}(\mathbf{z}) K_\nu^{(\alpha_j)}(\mathbf{z}) d\mathbf{z} \;=\; 0 \text{ for all } i \neq j \ , \, i, j = 1, \ldots, m.$$

Proof. Without loss of generality, we use $i = 1$, $j = 2$. Note

$$V(\alpha_1, \alpha_2) = [\int_{-\tau_1}^{\tau_1} K'_{\nu_1}(z_1)K_{\nu_1}(z_1)dz_1][\int_{-\tau_2}^{\tau_2} K'_{\nu_2}(z_2)K_{\nu_2}(z_2)dz_2][\int_{-\tau_m}^{\tau_m} \cdots \int_{-\tau_3}^{\tau_3} \prod_{j=3}^{m} K^2_{\nu_j}(z_j)dz_3 \cdots dz_m].$$

But $\int_{-\tau_1}^{\tau_1} K'_{\nu_1}(z_1)K_{\nu_1}(z_1)dz_1 = 0$ by an integration by parts argument (let $v = K_{\nu_1}(z_1)$ and $du = K'_{\nu_1}(z_1)$. This implies $dv = K'_{\nu_1}(z_1)$ and $u = K_{\nu_1}(z_1)$), and noting $K_{\nu_1}(-\tau_1) = K_{\nu_1}(\tau_1) = 0$, the result follows. □

A.4 Relation of Kernel Estimators to Locally Weighted Least Squares Estimators

Although the results to this point have been established using the kernel-type estimators of (2.2), we briefly discuss their validity when using the locally weighted least squares (LWLS) estimator. We assume the fixed design regression case. Recall the data are of the form $(\mathbf{x}_i, y_i)$, where $\mathbf{x}_i = [x_{i1}, \dots, x_{im}]^T$. Using the multiindex notation of section 2, the m-dimensional LWLS estimator fits a local polynomial of degree p at $\mathbf{u} = [u_1, \dots, u_m]^T$ as a solution to the problem

$$(A.12) \qquad \text{Minimize} \sum_{i=1}^{n} \{y_i - \sum_{|\tau|=0}^{p} \beta_{\tau_1 \cdots \tau_m}(u_1 - x_{i1})^{\tau_1} \cdots (u_m - x_{im})^{\tau_m}\}^2 \; G(\frac{\mathbf{u} - \mathbf{x}_i}{\mathbf{b}})$$

where the function $G(\frac{\mathbf{u}-\mathbf{x}_i}{\mathbf{b}}) = G(\frac{u_1-x_{i1}}{b_1}, \dots, \frac{u_m-x_{im}}{b_m})$ serves as a kernel weighting function using bandwidth $\mathbf{b} = [b_1, \dots, b_m]^T$ and where it is again assumed that $b_1 = \dots = b_m = b$.

In order to estimate $g^{(\nu)}(\mathbf{u})$, we must have $|\nu| \leq p < k$ (with k as in (M4)); for simplicity we work with the case $|\nu| = 0$ and $p = 1$, i.e., linear LWLS estimation of the function g in (2.1). This case was studied in detail by Ruppert and Wand (1994), where the estimator is shown to be $\hat{\alpha}$ in

$$(A.13) \qquad \{\hat{\alpha}, \hat{\beta}\}(\mathbf{u}) \; = \; \text{argmin}_{\{\alpha,\beta\}} \sum_{i=1}^{n} \{y_i - \alpha - \beta^T[\mathbf{u} - \mathbf{x}_i]\}^2 \; G(\frac{\mathbf{u} - \mathbf{x}_i}{b}),$$

providing

$$(A.14) \qquad \hat{g}(\mathbf{u}) \; = \; \hat{\alpha}(\mathbf{u}).$$

In the fixed design case where m=1, Müller (1987) showed that for any ν, the LWLS estimator described by (A.12) is asymptotically equivalent to that given by the kernel-type estimator of (2.2) when setting $G = K$ for nonnegative kernels $K \geq 0$ and using the unique decomposition $K = GP$

into a polynomial of degree $(k-2)$ and a nonnegative weight function $G \geq 0$ with $\int G(\mathbf{u})d\mathbf{u} = 1$. This equivalence is based on recognizing that both (2.2) and the LWLS estimator (the latter through the Gauss-Markov Theorem) can be written as

$$(A.15) \qquad \hat{g}^{(\nu)}(\mathbf{u}) \;=\; \sum_{i=1}^{n} w_{i,n,\nu}(\mathbf{u}) y_i .$$

Letting $w_{K,i}$ and $w_{G,i}$ denote the weights of (A.15) in the m=1 case for the kernel-type estimator of (2.2) and the LWLS estimator described by (A.12), respectively, Müller (1987) showed

$$(A.16) \qquad \lim_{\{n\to\infty\}} \sup_{\{1\leq i\leq n\}} \left|\frac{w_{G,i}}{w_{K,i}} - 1\right| \;=\; 0 \text{ , defining } \frac{0}{0} = 1.$$

As a consequence of this asymptotic equivalence, it was shown that given a LWLS estimate of $\hat{g}^{(\nu)}(\mathbf{u})$ obtained by (6.1) where $m = 1$, one could construct a corresponding kernel estimate of the form (2.2). The asymptotic consistency and distribution properties of both estimators would be the same, as a consequence of (6.6). These results are expected to carry over to the cases $m > 1$, $|\nu| \geq 0$.

A.5 Parametric Model Estimates for AIDS Incidence Data

The table below presents parameter values for model (4.1)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \epsilon$$

estimated by least squares for Latinos and Whites in California with

$$y = \text{AIDS incidence (cases/100,000 population)}$$

$$x_1 = \text{Time of AIDS diagnosis (year)} \text{ , } x_2 = \text{Age at AIDS diagnosis (years)}$$

Parameter Estimate	Latinos	Whites
$\hat{\beta}_0$ (intercept)	-133.9750	-200.1491
$\hat{\beta}_1$ (time)	10.2645	13.7943
$\hat{\beta}_2$ (age)	7.4376	11.6765
$\hat{\beta}_{11}$ (time^2)	-0.5916	-0.9484
$\hat{\beta}_{22}$ (age^2)	-0.0916	-0.1415
$\hat{\beta}_{12}$ (time $\times$ age)	-0.0118	-0.0367
$\hat{\theta}_{time}$ (from [4.2])	1993.2741	1991.4897
$\hat{\theta}_{age}$ (from [4.2])	40.0596	40.4065
$\hat{\sigma}^2 = s^2$	126.3482	226.0738
$\text{Var}(\hat{\beta}_0)$	40.3200	63.8896
$\text{Var}(\hat{\beta}_1)$	0.6613	1.0677
$\text{Var}(\hat{\beta}_2)$	0.0840	0.1310
$\text{Var}(\hat{\beta}_{11})$	0.0038	0.0057
$\text{Var}(\hat{\beta}_{22})$	0.000012	0.000017
$\text{Var}(\hat{\beta}_{12})$	0.0002	0.000253
$\text{Cov}(\hat{\beta}_0, \hat{\beta}_1)$	-2.3407	-3.5804
$\text{Cov}(\hat{\beta}_0, \hat{\beta}_2)$	-1.7265	-2.7208
$\text{Cov}(\hat{\beta}_0, \hat{\beta}_{11})$	0.0574	0.0950
$\text{Cov}(\hat{\beta}_0, \hat{\beta}_{22})$	0.0176	0.0275
$\text{Cov}(\hat{\beta}_0, \hat{\beta}_{12})$	0.0381	0.0563
$\text{Cov}(\hat{\beta}_1, \hat{\beta}_2)$	0.0440	0.0640
$\text{Cov}(\hat{\beta}_1, \hat{\beta}_{11})$	-0.0352	-0.0571
$\text{Cov}(\hat{\beta}_1, \hat{\beta}_{22})$	-0.0001	-0.0001
$\text{Cov}(\hat{\beta}_1, \hat{\beta}_{12})$	-0.0068	-0.0107
$\text{Cov}(\hat{\beta}_2, \hat{\beta}_{11})$	-0.0003	-0.0004
$\text{Cov}(\hat{\beta}_2, \hat{\beta}_{22})$	-0.0009	-0.0014
$\text{Cov}(\hat{\beta}_2, \hat{\beta}_{12})$	-0.0009	-0.0014
$\text{Cov}(\hat{\beta}_{11}, \hat{\beta}_{22})$	0.000012	0.000005
$\text{Cov}(\hat{\beta}_{11}, \hat{\beta}_{12})$	-0.0001	0.000001
$\text{Cov}(\hat{\beta}_{22}, \hat{\beta}_{12})$	-0.000003	0.000001

A.6 Nonparametric Model Calculations for AIDS Incidence Data

For the data in Appendix B, a linear LWLS estimate was fit with the calculated cross-validation bandwidth to produce the point estimates ($\hat{\theta}$) of the true maximizers of AIDS incidence (θ) for Latinos and Whites in California. The cross-validation bandwidths (shown in the table below) were multiplied by a factor of 1.25 to calculate $\hat{g}^{(\tau)}(\hat{\theta})$, which were used to estimate $g^{(\tau)}(\theta)$. These were done using the product polynomial kernel $(1 - z_1^2)^2(1 - z_2^2)^2$ with $z_1, z_2 \in [-1, 1]$. The degree of local polynomial fit in the j^{th} direction was $\tau_j + 1$ for $j = 1, 2$. The numbers are provided below:

Nonparametric Estimate/Calculation	Latinos	Whites
$\hat{\theta}_{time}$	1992.25	1991.75
$\hat{\theta}_{age}$	34.0	36.0
Cross-validation bandwidth, time direction $= b_{time}$	1.647	1.13
Cross-validation bandwidth, age direction $= b_{age}$	4.662	2.56
$\hat{\sigma}^2$, using estimate from Müller and Prewitt (1993)	70.19	147.98
Estimate of $g^{(1,1)}(\theta) = \hat{g}^{(1,1)}(\hat{\theta})$	-1.1925	9.8392
Estimate of $g^{(0,2)}(\theta) = \hat{g}^{(0,2)}(\hat{\theta})$	-8.0035	-17.1148
Estimate of $g^{(2,0)}(\theta) = \hat{g}^{(2,0)}(\hat{\theta})$	0.1033	1.7387

We now present calculations for the Latino AIDS incidence data and note that those for the White AIDS data are similar. Using Corollary 3.3 and Lemma A.7, the expression for a 95% confidence region for $\theta = [\theta_{time}, \theta_{age}]$ under the assumption that $d = 0$ is given by

$$(A.17) \qquad \begin{pmatrix} \theta_{time} - \hat{\theta}_{time} \\ \theta_{age} - \hat{\theta}_{age} \end{pmatrix}^T \frac{1}{\lambda(\mathcal{X})\sigma^2} H \begin{pmatrix} \theta_{time} - \hat{\theta}_{time} \\ \theta_{age} - \hat{\theta}_{age} \end{pmatrix} \leq \chi^2_2 = 5.99$$

where

$$H = \gamma_n^2 \begin{pmatrix} \frac{g^{(2\alpha_1)}(\theta)g^{(2\alpha_1)}(\theta)}{V(\alpha_1,\alpha_1)} + \frac{g^{(\alpha_1+\alpha_2)}(\theta)g^{(\alpha_1+\alpha_2)}(\theta)}{V(\alpha_2,\alpha_2)} & \frac{g^{(\alpha_1+\alpha_2)}(\theta)g^{(2\alpha_1)}(\theta)}{V(\alpha_1,\alpha_1)} + \frac{g^{(2\alpha_2)}(\theta)g^{(\alpha_1+\alpha_2)}(\theta)}{V(\alpha_2,\alpha_2)} \\ \frac{g^{(\alpha_1+\alpha_2)}(\theta)g^{(2\alpha_1)}(\theta)}{V(\alpha_1,\alpha_1)} + \frac{g^{(2\alpha_2)}(\theta)g^{(\alpha_1+\alpha_2)}(\theta)}{V(\alpha_2,\alpha_2)} & \frac{g^{(\alpha_1+\alpha_2)}(\theta)g^{(\alpha_1+\alpha_2)}(\theta)}{V(\alpha_1,\alpha_1)} + \frac{g^{(2\alpha_2)}(\theta)g^{(2\alpha_2)}(\theta)}{V(\alpha_2,\alpha_2)} \end{pmatrix}$$

From section 4.1, $\lambda(\mathcal{X}) = 400$, and as discussed in section 3, we use consistent estimates for σ^2, θ, and the elements of H, all of which are shown in the above table. By Lemma A.7, $V(\alpha_1, \alpha_2) = 0$, and so we need only to find $V(\alpha_j, \alpha_j)$, $j = 1, 2$. Now

$$\begin{aligned} K_{\mathbf{z}}^{(\alpha_1)} &= \frac{\partial}{\partial z_1}(1 - z_1^2)^2(1 - z_2^2)^2 \\ &= -4z_1(1 - z_1^2)(1 - z_2^2)^2 \\ \text{and } K_{\mathbf{z}}^{(\alpha_2)} &= -4(1 - z_1^2)^2(1 - z_2^2)z_2 \qquad \text{by symmetry.} \end{aligned}$$

This implies $V(\alpha_1, \alpha_1) = V(\alpha_2, \alpha_2)$, and we calculate

$$\begin{aligned} V(\alpha_1, \alpha_1) &= \int_{-1}^{1}\int_{-1}^{1}[K^{(\alpha_1)}(\mathbf{z})]^2 d\mathbf{z} \\ &= 16 \times [\int_{-1}^{1}(z_1^3 - z_1)^2 dz_1] \times [\int_{-1}^{1}(1 - z_2^2)^4 dz_2] \\ &= 16 \times \frac{16}{105} \times \frac{256}{315} \approx 1.98 \end{aligned}$$

Plugging the values presented above into the expression for H and noting the bandwidth values for the appropriate direction (all results were derived under the assumption that the bandwidth was the same in all directions for simplicity, which is not the case for these data) then gives

$$
\begin{aligned}
H &= \frac{\gamma_n}{1.98}\begin{pmatrix} (-8.0035)^2 + (-1.1925)^2 & -1.1925 \times (0.1033 \text{ - } 8.0035) \\ -1.1925 \times (0.1033 \text{ - } 8.0035) & (0.1033)^2 + (-1.1925)^2 \end{pmatrix} \\
&= \begin{pmatrix} 466 \times (1.647)^3 \times (4.662) \times (33.0697) & 466 \times (1.647)^2 \times (4.662)^2 \times (4.7581) \\ 466 \times (1.647)^2 \times (4.662)^2 \times (4.7581) & 466 \times (1.647) \times (4.662)^3 \times (0.7236) \end{pmatrix} \\
&= \begin{pmatrix} 320973.545 & 130722.763 \\ 130722.763 & 56272.328 \end{pmatrix}.
\end{aligned}
$$

We note that for the kernel weighting scheme, $\mathbf{z} \in [-1,1] \times [-1,1]$, and $\lambda([-1,1] \times [-1,1]) = 4$, and so we adjust $\lambda(\mathcal{X})$ from 400 to 100. Hence, $\lambda(\mathcal{X}) \times \hat{\sigma}^2 = 7019$. Plugging all of this information into (A.17) yields the 95% confidence region to be

$$
\begin{pmatrix} \theta_{time} - 1992.25 \\ \theta_{age} - 34.0 \end{pmatrix}^T \frac{1}{7019} \begin{pmatrix} 320973.545 & 130722.763 \\ 130722.763 & 56272.328 \end{pmatrix} \begin{pmatrix} \theta_{time} - 1992.25 \\ \theta_{age} - 34.0 \end{pmatrix} \leq 5.99,
$$

$$
\text{or} \qquad \begin{pmatrix} \theta_{time} - 1992.25 \\ \theta_{age} - 34.0 \end{pmatrix}^T \begin{pmatrix} 45.70 & 18.61 \\ 18.61 & 8.01 \end{pmatrix} \begin{pmatrix} \theta_{time} - 1992.25 \\ \theta_{age} - 34.0 \end{pmatrix} \leq 5.99
$$

The last equation is presented in section 4, where a similar calculation done for the White AIDS incidence data is also presented.

References

Axelsson, A.K., Nordahl, A., Carlson, R. (1995). Optimization in organic synthesis. Sequential response surface modelling for obtaining kinetic information. Part 2: An experimental study of the Williamson ether synthesis. *J. Chemometrics* **9**, 441-449.

California Department of Finance (2001). Census-based population estimates for all racial/ethnic groups are available at www.dof.ca.gov

Carter, W.H. Jr., Jones, D.E., Carchman, R.A. (1985). Application of response surface methods for evaluating the interactions of soman, atropine, and pralidioxime chloride. *Fundamental and Applied Toxicology* **5**, S232-S241.

Clifford, A.J., Bills, N.D., Peerson, J.M., Müller, H.G., Burk, G.E., Rich, K.D. (1993). A depletion-repletion folate bioassay based on growth and tissue folate concentrations of rats. *J. Nutrition* **123**, 926-932.

Cox, C. (1987). Threshold dose-response models in toxicology. *Biometrics* **43**, 511-523.

Dey, G., Mitra, A., Banerjee, R., Maiti, B.R. (2001). Enhanced production of amylase by optimization of nutritional constituents using response surface methodology. *Biochem. Engineering J.* **7**, 227-231.

Fan, J., Gijbels, I. (1996). *Local Polynomial Modeling and Its Applications.* Chapman and Hall: New York.

Gennings, C., Dawson, K.S., Carter, W.H. Jr., Myers, R.H. (1990). Interpreting plots of a multidimensional dose-response surface in a parallel coordinate system. *Biometrics* **46**, 719-735.

Greco, W.R., Bravo, G., Parsons, J.C. (1995). The search for synergy: A critical review from a response surface perspective. *Pharmacological Reviews* **47**, 331-385.

Hirst, K., Zerbe, G.O., Hay, W.W. Jr. (1996). A generalized Michaelis-Menten response surface. *Stat. Med.* **15**, 2107-2119.

Hung, H.M.J. (1992). On identifying a positive dose-response surface for combination agents. *Stat.*

Med. **11**, 703-711.

Kitagawa, R., Bruns, R.E., de Menezes, T.J.B. (1994). Optimizing the enzymatic maceration of foliole puree from hard pieces of hearts of palm (Euterpe edulis Mart.) using response surface analysis. *J. Food Science* **59**, 844-848.

Lewis, D.W. (1991). *Matrix Theory.* World Scientific Publishing: Singapore/New Jersey.

McCullagh, P., Nelder J.A. (1992). *Generalized Linear Models*, 2nd ed. Chapman and Hall: New York.

Müller, H.G. (1985). Kernel estimators of zeros and of location and size of extrema of regression functions. *Scand. J. Statist.* **12**, 221-232.

Müller, H.G. (1987). Weighted local regression and kernel methods for nonparametric curve fitting. *J. Am. Stat. Assoc.* **82**, 231-238.

Müller, H.G., Stadtmüller (1987). Variable bandwidth kernel estimators of regression curves. *Ann. Stat.* **15**, 182-201.

Müller, H.G. (1988). *Nonparametric Regression Analysis of Longitudinal Data.* Springer-Verlag: Berlin.

Müller, H.G. (1989). Adaptive nonparametric peak estimation. *Ann. Stat.* **17**, 1053-1069.

Müller, H.G., Prewitt, K.A. (1993). Multiparameter bandwidth processes and adaptive surface smoothing. *J. Multivariate Anal.* **47**, 1-21.

Müller, H.G., Facer, M.R., Bills, N.A., Clifford, A.J. (1996). Statistical interaction model for exchangeability of food folates in a rat growth bioassay. *J. Nutr.* **126**, 2585-2592.

Myers, R.H., Khuri, A.I., Carter, W.H. Jr. (1989). Response surface methodology: 1966-1988. *Technometrics* **31**, 137-157.

Myers, R.H. (1999). Response surface methodology - current status and future directions. *J. Quality Technology* **31**, 30-44.

Rice, J., Kelly, C. (1990). Monotone smoothing with application to dose-response curves and the assessment of synergism. *Biometrics* **46**, 1071-1085.

Roberto, I.C., Sato, S., de Mancilha, I.M., Taqueda, M.E.S. (1995). Influence of media composition on xylitol fermentation by Candida guiliiermondii using response surface methodology. *Biotechnol. Lett.* **17**, 1223-1228.

Ruppert, D., Wand, M.P. (1994). Multivariate weighted least squares regression. *Ann. Statist.* **22**, 1346-1370.

Shyy, W., Tucker, P.K., Vaidyanathan, R. (2001). Response surface and neural network techniques for rocket engine injector optimization. *J. Propulsion and Power* **17**, 391-401.

Vogt, N.B., Kolset, K. (1989). Composition activity relationships - CARE: Part III. Polynomial principal component regression and response surface analysis of mutagenicity in air samples. *Chemometrics and Intelligent Lab. Systems* **6**, 221-237.