Simultaneous Confidence Bands for a Mean Response and Its Derivatives

Benjamin Hall*, Richard Charnigo, and Cidambi Srinivasan

University of Kentucky

Abstract

Though many techniques are available to provide simultaneous confidence bands around a nonparametric estimate of a mean response function, little attention has been given to constructing simultaneous confidence bands for the mean response and one or more derivatives, where simultaneous now refers both to values of the covariate and to all derivatives under consideration. In this paper we propose a method for constructing confidence bands that are simultaneous over both the covariate space and for the mean response and one or more derivatives. Our method works for *any* nonparametric regression technique that is both self-consistent and linear in the observed responses. We also address the estimation of bias, interpolation error, and noise variance in case assumptions about these quantities are not available or defensible. Besides pro-

*Corresponding author. Postal address: 857 Patterson Office Tower, Department of Statistics, University of Kentucky, Lexington KY 40506-0027. E-mail address: Benjamin.Hall@uky.edu.

viding theoretical justifications for our methodology, we assess its performance through simulations and an application to real data.

Key words: Nonparametric regression, self-consistency, bias correction, compound estimator.

1 Introduction

Suppose that $\mu(x)$ is the mean response function for the nonparametric regression model

$$Y_i = \mu(x_i) + \epsilon_i \quad \text{for } i \in \{1, \ldots, n\}, \tag{1}$$

where the design points x_i belong to a compact interval $\mathcal{X} \subset \mathbb{R}$, $\mu(x)$ is a real-valued function defined on the covariate space $\mathcal{X}$ that has $(q+1)$ continuous derivatives for some $q \in \mathbb{N}$, and the ϵ_i are independent, normally distributed errors with mean zero and variance $\sigma^2 \in (0, \infty)$.

Many nonparametric regression methods have been devised to estimate $\mu(x)$ and its derivatives in this situation (Hardle 1990, Loader 1999, Gyorfi, Kohler, Krzyzak, and Walk 2002, and references cited therein). In addition, many techniques have been proposed for placing simultaneous confidence bands around the estimate of $\mu(x)$ (Sun and Loader 1994; Wu and Zhou 2007, and references cited therein). However, the only technique we are aware of for constructing confidence bands that are simultaneous over both the covariate space and for the mean response and one or more derivatives is that of Claeskens and Van Keilegom (2003). Their technique is specifically designed for local polynomial estimation via maximum likelihood.

Local polynomial estimation is not the only method used to recover derivatives. Charnigo and Srinivasan (2009a) discuss situations where, to be of practical use, estimates of the mean response and its derivatives need to be self-consistent, meaning that estimates of the derivatives of the mean response are derivatives of the estimate of the mean response. Local polynomial estimation is not self-consistent and would, for example, lead to conflicting conclusions about the timing of a growth spurt when $\mu(x)$ represents a height function, according to whether one searches for a maximum of the first derivative estimate or a zero of the second derivative estimate.

The present paper develops a method for constructing confidence bands that are simultaneous over both the covariate space and for the mean response and one or more derivatives. Our methodology works for any nonparametric regression technique that has the following two properties:

Self-consistency. An estimator $\widehat{\mu}(x)$ is self-consistent if $\frac{d^j}{dx^j}\widehat{\mu}(x)$ exists and equals $\widehat{\frac{d^j}{dx^j}\mu(x)}$ for every $j \in \{1, 2, \ldots, q\}$.

Linear in the observed responses. An estimator $\widehat{\mu}(x)$ is linear in the observed responses if $\widehat{\mu}(x) := \sum_{i=1}^{n} l_i(x)Y_i$ for functions $l_1, \ldots, l_n$ which do not depend on $Y_1, \ldots, Y_n$.

Kernel smoothing (Priestley and Chao 1972), spline smoothing (Schoenberg 1964, Reinsch 1967), and compound estimation (Charnigo and Srinivasan 2009a, 2009b) are examples of nonparametric regression techniques with which simultaneous confidence bands can be constructed as proposed in this paper.

Quantities like bias, interpolation error, and noise variance are generally unknown but are, for convenience, often treated as if they were known or as if upper bounds for them were available. We explicitly address the estimation of bias, interpolation error, and noise variance. While our methodology for mean responses and derivatives can work with upper bounds for such quantities, it does not rely on the availability of these upper bounds.

The rest of this paper is organized as follows. Section 2 demonstrates how to obtain simultaneous confidence bands given assumptions about bias and interpolation error. In Section 3 we remove the assumption about the bias. In Section 4, we propose estimates of the interpolation error. A modification for unknown variance is proposed in Section 5. Section 6 includes results from a simulation study and application to real data.

2 Simultaneous Confidence Bands

We say that, $L_{p_1}(x), ..., L_{p_J}(x)$ and $U_{p_1}(x), ..., U_{p_J}(x)$ form $100(1-\alpha)\%$ confidence bands for derivatives $p_1, \ldots, p_J \in \{0, 1, \ldots, q\}$ for some $J \leq q+1$ if $P(L_{p_1}(x) \leq \mu^{(p_1)}(x) \leq U_{p_1}(x), ..., L_{p_J}(x) \leq \mu^{(p_J)}(x) \leq U_{p_J}(x), \forall x \in \mathcal{X}) \geq 1-\alpha$. Note that 'simultaneous' here has a double meaning: the confidence bands are simultaneous over both $\mathcal{X}$ and $\{p_1, \ldots, p_J\}$.

Knafl, Sacks, and Ylvisaker (1985) considered an estimator linear in the observed responses and developed simultaneous confidence bands for the mean response only, i.e. $J = 1$ and $p_1 = 0$. They showed that if we take $\mathbf{G} = \{\xi_1, ..., \xi_G\}$ to be a uniform grid of points from $\mathcal{X}$, including the boundaries, and define

$$Z(x) := \frac{\sum_{i=1}^n l_i(x)\epsilon_i}{\sigma D(x)}, \text{where } D(x) = \sqrt{\sum_{i=1}^n l_i(x)^2}, \text{ then}$$

$$\begin{bmatrix} Z(\xi_1) & Z(\xi_2) & \cdots & Z(\xi_G) \end{bmatrix}^t \sim MVN(\mathbf{0}, \Sigma)$$

where Σ has diagonal entries of 1 and off-diagonal entries of

$$\Sigma_{kj} = \sum_{i=1}^{n} l_i(\xi_k) l_i(\xi_j) / [D(\xi_k) D(\xi_j)].$$

They can then compute a z_α so that $P\left(\max_{x \in \mathbf{G}} |Z(x)| > z_\alpha\right) \leq \alpha$. Then if $B(x)$ represents the absolute value of the bias of $\widehat{\mu}(x)$ and $M := \sup_{x \in \mathcal{X}} |\mu(x) - \mu_I(x)|$, they demonstrate that $\widehat{\mu}_I(x) \pm [M + B_I(x) + z_\alpha \sigma D_I(x)]$ are $100(1-\alpha)\%$ confidence bands for the mean response. Here, and in what follows, the subscript I denotes linear interpolation between the grid points so that for $f(x)$ defined on $\mathcal{X}$ and $\xi_j < x < \xi_{j+1}$,

$$f_I(x) := \left[\frac{\xi_{j+1} - x}{\xi_{j+1} - \xi_j}\right] f(\xi_j) + \left[\frac{x - \xi_j}{\xi_{j+1} - \xi_j}\right] f(\xi_{j+1}).$$

In the present paper, however, we seek to formulate simultaneous confidence bands not only for the mean response, but also over one or more derivatives. To do this, for each $p \in \{p_1, ..., p_J\}$, we define

$$Z_p(x) := \frac{\sum_{i=1}^{n} l_i^{(p)}(x) \epsilon_i}{\sigma D_p(x)}, \text{where } D_p(x) = \sqrt{\sum_{i=1}^{n} l_i^{(p)}(x)^2}. \tag{2}$$

The vector of all of the $Z_p(x)$'s evaluated at the grid points will be multivariate normally distributed. The mean will be the zero vector and the covariance matrix will have 1's on the diagonal with off-diagonal entries:

$$\operatorname{Cov}(Z_a(\xi_j), Z_b(\xi_k)) = \frac{\sum_{i=1}^{n} l_i^{(a)}(\xi_j) l_i^{(b)}(\xi_k)}{D_a(\xi_j) D_b(\xi_k)} \text{ for } a, b \in \{p_1, p_2, \cdots, p_J\}. \tag{3}$$

It then follows that

$$\begin{aligned} & P\left(\cup_{r=1}^{J} \left\{\max_{x \in \mathbf{G}} |Z_{p_r}(x)| > z_\alpha\right\}\right) \\ \leq \; & P\left(\cup_{r=1}^{J} \{|Z_{p_r}(\xi_1)| > z_\alpha\}\right) \\ & + \sum_{j=1}^{G-1} P\left(\{\cap_{r=1}^{J} \{|Z_{p_r}(\xi_j)| \leq z_\alpha\}\} \cap \{\cup_{r=1}^{J} \{|Z_{p_r}(\xi_{j+1})| > z_\alpha\}\}\right). \end{aligned} \tag{4}$$

We then choose z_α so that (4) is equal to α. The benefit of (4) is that probabilities involving multivariate normal vectors of dimension $2J$ are numerically much easier to evaluate than probabilities involving multivariate normal vectors of dimension GJ. The conservatism involved with (4) will be small as long as the correlations between $Z_{p_r}(\xi_j)$ and $Z_{p_r}(\xi_{j+1})$ are high.

If (4) is too conservative, then we could instead consider that

$$\begin{aligned} & P\left(\cup_{r=1}^{J}\{\max_{x\in\mathbf{G}}|Z_{p_r}(x)| > z_\alpha\}\right) \\ \leq\ & P\left(\cup_{r=1}^{J}\{|Z_{p_r}(\xi_1)| > z_\alpha\}\right) \\ & +P\left(\{\cap_{r=1}^{J}\{|Z_{p_r}(\xi_1)| \leq z_\alpha\}\} \cap \{\cup_{r=1}^{J}\{|Z_{p_r}(\xi_2)| > z_\alpha\}\}\right) \\ & +\sum_{j=1}^{G-2} P\left(\{\cap_{r=1}^{J}\{|Z_{p_r}(\xi_j)| \leq z_\alpha\}\}\right. \\ & \left.\cap\{\cap_{r=1}^{J}\{|Z_{p_r}(\xi_{j+1})| \leq z_\alpha\}\} \cap \{\cup_{r=1}^{J}\{|Z_{p_r}(\xi_{j+2})| > z_\alpha\}\}\right). \end{aligned} \tag{5}$$

Choosing z_α so that (5) is equal to α is less conservative than doing so for (4). Even more conservatism could be eliminated since (4) and (5) are the first two upper bounds available from a decreasing sequence of $G-1$ upper bounds whose last member is exactly equal to the probability being bounded.

We now state our first theorem:

Theorem 2.1 *Assume that model (1) holds. Let $\widehat{\mu}(x)$ be self-consistent and linear in the observed responses. Let $\boldsymbol{G}$ be a uniform grid of points from $\mathcal{X}$. Let $B_p(x)$ be the absolute value of the bias of $\widehat{\mu}^{(p)}(x)$ and $M_p = \sup_{x\in\mathcal{X}}|\mu^{(p)}(x) - \mu_I^{(p)}(x)|$ for $p \in \{p_1, ..., p_J\}$. Let $D_p(x)$ be defined as in (2). Then*

$$P(L_{p_1}(x) \leq \mu^{(p_1)}(x) \leq U_{p_1}(x), ..., L_{p_J}(x) \leq \mu^{(p_J)}(x) \leq U_{p_J}(x), \forall x \in \mathcal{X}) \geq 1-\alpha,$$

$$\text{where } U_p(x), L_p(x) := \widehat{\mu}_I^{(p)}(x) \pm (M_p + B_{pI}(x) + z_\alpha \sigma D_{pI}(x)).$$

Note that the assumptions of known $B_p(x)$ and M_p (which depend on $\mu(x)$), and of known σ, will be removed in Sections 3, 4, and 5, respectively. The proof of Theorem 2.1 and of all other theoretical results may be found in the Appendix.

3 Bias correction

In the previous section we worked from the assumption that bounds on the bias are known. One may or may not have a good rationale for making such assumptions. In this section we investigate data-based means of dealing with the bias.

Since $Bias_\mu[\widehat{\mu(x)}] = \sum_{i=1}^n l_i(x)\mu(x_i) - \mu(x)$, a proposed estimate of the bias in estimating the mean response is (Loader 1999, p. 168):

$$\widehat{Bias}(x) := \sum_{i=1}^n l_i(x)\widehat{\mu}(x_i) - \widehat{\mu}(x). \tag{6}$$

However, Loader cautions that one should not simply shift both upper and lower bands for the mean response by subtracting the estimate of the bias. Such efforts merely result in bands centered around undersmoothed estimates of the mean response.

We note that (6) can be extended to provide bias estimates for derivatives. Since $Bias_\mu[\widehat{\mu^{(p)}(x)}] = \sum_{i=1}^n l_i^{(p)}(x)\mu(x_i) - \mu^{(p)}(x)$, a proposed estimate of the bias is:

$$\widehat{Bias_p}(x) := \sum_{i=1}^n l_i^{(p)}(x)\widehat{\mu}(x_i) - \widehat{\mu}^{(p)}(x). \tag{7}$$

As when constructing confidence bands for mean responses, we should not simply use (7) to shift both the upper and lower bands for $\mu^{(p)}(x)$. However, we can take the absolute value of (7) as an estimate of $B_p(x)$. For convenience this estimate is hereafter denoted $\widehat{B}_p(x)$. Then substituting $\widehat{B}_p(x)$ for $B_p(x)$ in the $U_p(x), L_p(x)$ formulas of Theorem 2.1 addresses the bias issue without assuming that a bound for the bias is known. If there is concern about the disparity between $\widehat{B}_p(x)$ and $B_p(x)$, in that the former may underestimate the latter, then that disparity itself can be estimated and incorporated into the confidence bands. We will elaborate on that idea later in this section.

Supposing for now that the disparity between $\widehat{B}_p(x)$ and $B_p(x)$ is not worrisome, some conservatism in the confidence bands can be eliminated. This is because the use of $B_p(x)$ in Theorem 2.1 does not exploit any information about the sign of the bias. If the bias were known to be positive, then reducing the lower confidence band by $B_p(x)$ would be reasonable, but there would be no need to raise the upper confidence band by $B_p(x)$. Likewise, if the bias were known to be negative, then raising the upper confidence band by $B_p(x)$ would be reasonable, but there would be no need to reduce the lower confidence band by $B_p(x)$. Of course, if one is assuming rather than estimating an upper bound for $B_p(x)$, then typically one does not know the sign of the bias. On the other hand, if one is estimating an upper bound for $B_p(x)$ via the absolute value of (7), then one does have information about the sign of the bias. Thus, some conservatism in the confidence bands can be eliminated if, instead of replacing $B_p(x)$ by $\widehat{B}_p(x)$, we replace $B_p(x)$ in the lower band by $\widehat{B}_p(x)1_{\widehat{Bias_p(x)}>0}$ and in the upper band by $\widehat{B}_p(x)1_{\widehat{Bias_p(x)}<0}$. This is done at each grid point prior to interpolation.

We now return to the question of handling the disparity between $\widehat{B}_p(x)$ and

$B_p(x)$. Since the difference between $\widehat{B}_p(x)$ and $B_p(x)$ depends in large part on the difference between $\sum_{i=1}^n l_i^{(p)}(x)\widehat{\mu}(x_i)$ and $\sum_{i=1}^n l_i^{(p)}(x)\mu(x_i)$, we note that for $a_p(x) > 0$,

$$\begin{aligned}
& P\left(\left|\sum_{i=1}^n l_i^{(p)}(x)\widehat{\mu}(x_i) - \sum_{i=1}^n l_i^{(p)}(x)\mu(x_i)\right| \geq a_p(x)\right) \\
\leq \; & \frac{1}{a_p(x)^2}\mathbb{E}\left[\left(\sum_{i=1}^n l_i^{(p)}(x)\widehat{\mu}(x_i) - \sum_{i=1}^n l_i^{(p)}(x)\mu(x_i)\right)^2\right] \qquad (8) \\
= \; & \frac{1}{a_p(x)^2}\mathbb{E}\left[\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{m=1}^n l_i^{(p)}(x)l_j(x_i)l_k^{(p)}(x)l_m(x_k)Y_jY_m\right. \\
& \left. -2\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n l_i^{(p)}(x)l_j(x_i)l_k^{(p)}(x)Y_j\mu(x_k) + \sum_{i=1}^n\sum_{j=1}^n l_i^{(p)}(x)l_j^{(p)}(x)\mu(x_i)\mu(x_j)\right] \\
= \; & \frac{1}{a_p(x)^2}\left[\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{m=1}^n l_i^{(p)}(x)l_j(x_i)l_k^{(p)}(x)l_m(x_k)\{\mu(x_j)\mu(x_m) + \sigma^2 1_{j=m}\}\right. \\
& -2\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n l_i^{(p)}(x)l_j(x_i)l_k^{(p)}(x)\mu(x_j)\mu(x_k) \\
& \left. +\sum_{i=1}^n\sum_{j=1}^n l_i^{(p)}(x)l_j^{(p)}(x)\mu(x_i)\mu(x_j)\right], \qquad (9)
\end{aligned}$$

where line (8) follows from the Markov Inequality. By substituting $\widehat{\mu}(x)$ for $\mu(x)$, (9) can be set equal to $P_0 \in (0,1)$ and the equation can then be solved to yield an estimate, $\widehat{a}_p(x)$. The value of $\widehat{a}_p(x)$ will need to be evaluated at each of the grid points and the upper band is then shifted up by $\widehat{a}_p(x)$ while the lower band is shifted down by $\widehat{a}_p(x)$. In our experience, substitution of $\widehat{B}_p(x)1_{\widehat{Bias_p(x)}>0}$ or $\widehat{B}_p(x)1_{\widehat{Bias_p(x)}<0}$ for $B_p(x)$ sometimes leads to bands with actual coverage probability less than the nominal coverage probability of $1-\alpha$. However, the adjustment by $\widehat{a}_p(x)$ often yields an actual coverage probability greater than the nominal coverage probability, even when P_0 is as large as α. This is because inequality (8) is rather conservative.

The rate at which $a_p(x)$ goes to zero will depend on the nonparametric regression estimator being employed and the order of the derivative p. The following proposition demonstrates this for kernel regression with a compactly supported kernel.

Proposition 3.1 *Assume the conditions of Theorem 2.1 hold. Assume that nonparametric regression is performed using the kernel estimator*

$$\widehat{\mu}(x) = (nh)^{-1} \sum_{i=1}^{n} K((x - x_i)/h) Y_i$$

with the kernel function supported on $[-1, 1]$*. Further suppose that*

$$\sup_x |\widehat{\mu}^{(p)}(x) - \mu^{(p)}(x)| = O_p\left(\left(\frac{n}{\log n} \right)^{(p-J-1)/(2J+3)} \right)$$

with bandwidth $h \propto n^{-1/(2J+3)}$*. Then* $a_p(x) = O_p(n^{(p-J-1)/(2J+3)} \log n)$.

If one wishes to *guarantee* that, asymptotically, the actual coverage probability does not fall short of the nominal coverage probability, then one may proceed as follows. Let r_p be the convergence rate of the nonparametric regression technique being employed (i.e. $n^{r_p}(\widehat{\mu}^{(p)}(x) - \mu^{(p)}(x)) = O_p(1)$) and define

$$\widehat{\beta}_{pL}(x) := \max\left\{ \left| \widehat{Bias}_p(x) \right| 1_{\widehat{Bias}_p(x) > 0} + a_p(x), g(n) \right\} \text{ and}$$
$$\widehat{\beta}_{pU}(x) := \max\left\{ \left| \widehat{Bias}_p(x) \right| 1_{\widehat{Bias}_p(x) < 0} + a_p(x), g(n) \right\}, \qquad (10)$$

where $g(n) \to 0$ and $n^{r_p} g(n) \to \infty$.

We then obtain a modification of Theorem 2.1:

Theorem 3.1 *Let* $\widehat{\beta}_{aL}(x)$ *and* $\widehat{\beta}_{aU}(x)$ *be as in (10). Then, under the conditions of Theorem 2.1,*

$$\liminf_{n \to \infty} P\left(L_{p_1}(x) \le \mu^{(p_1)}(x) \le U_{p_1}(x), \right.$$

$$..., L_{p_J}(x) \leq \mu^{(p_J)}(x) \leq U_{p_J}(x), \forall x \in \mathcal{X}\Big) \geq 1 - \alpha$$

$$\textit{where } L_p(x) := \widehat{\mu}_I^{(p)}(x) - (M_p + \widehat{\beta}_{pL}(x)_I + z_\alpha \sigma D_{pI}(x))$$

$$\textit{and } U_p(x) := \widehat{\mu}_I^{(p)}(x) - (M_p + \widehat{\beta}_{pU}(x)_I + z_\alpha \sigma D_{pI}(x)).$$

4 Interpolating between the grid points

To obtain confidence bands that are simultaneous over the entire covariate space rather than simply over a finite grid of points we must interpolate between the grid points. As mentioned previously, this can be accomplished if we assume that upper bounds on $\sup_{x \in \mathcal{X}} |\mu^{(p)}(x) - \mu_I^{(p)}(x)|$, call them M_p for $p \in \{p_1, ..., p_J\}$, are available.

If such bounds are unavailable, then we propose using the following estimates:

$$\widehat{M}_p := \begin{cases} \max_{i \in \{1,\dots,n\}} |\widehat{\mu}^{(p)}(x_i) - \widehat{\mu}_I^{(p)}(x_i)| & : \quad n < N_{m_0} \\ \max_{\gamma_i \in \Gamma_m} |\widehat{\mu}^{(p)}(\gamma_i) - \widehat{\mu}_I^{(p)}(\gamma_i)| & : \quad N_m \leq n < N_{m+1} \end{cases}$$

for $m \geq m_0$, where m_0 is an arbitrary positive integer, $\Gamma_m = \{\gamma_1, \dots, \gamma_m\} \subset \mathcal{X}$ is a grid that becomes dense in $\mathcal{X}$ as $m \to \infty$ and $\{N_m : m \in \{1, 2, \dots\}\}$ is a strictly increasing sequence of positive integers.

The large-sample justification for using these estimates is shown in Theorem 4.1:

Theorem 4.1 *If $\hat{\mu}^{(p)}(x) \to^P \mu^{(p)}(x)$ as $n \to \infty, \forall x \in \mathcal{X}$ and $p \in \{p_1, \dots, p_J\}$, then the sequence $\{N_m : m \in \{1, 2, \dots\}\}$ may be chosen so that $\widehat{M}_p \to^P M_p$ as $n \to \infty$. In this case, under the conditions of Theorem 3.1,*

$$\liminf_{n \to \infty} P\Big(L_{p_1}(x) \leq \mu^{(p_1)}(x) \leq U_{p_1}(x),$$

$$..., L_{p_J}(x) \leq \mu^{(p_J)}(x) \leq U_{p_J}(x), \forall x \in \mathcal{X}\Big) \geq 1 - \alpha$$

$$\textit{where } L_p(x) := \widehat{\mu}_I^{(p)}(x) - (\widehat{M}_p + \epsilon + \widehat{\beta}_{pL}(x)_I + z_\alpha \sigma D_{pI}(x))$$

$$\textit{and } U_p(x) := \widehat{\mu}_I^{(p)}(x) - (\widehat{M}_p + \epsilon + \widehat{\beta}_{pU}(x)_I + z_\alpha \sigma D_{pI}(x)),$$

and ϵ *is an arbitrarily small positive number.*

We close this section with an observation about the interpolation error. For typical nonparametric regression estimators, $B_p(x)$ and $D_p(x)$ will tend to 0 as $n \to \infty$. Thus, their contributions to the confidence bands will become negligible asymptotically. On the other hand, the bound M_p for the interpolation error will not change with n. This has the unsettling consequence that, asymptotically, the region bounded by the confidence bands will have area at least $2 \times M_p \times length(\mathcal{X})$. In effect, there is a limit to how tight the confidence bands can become. One way to address this problem, as indicated in the following proposition, is to let the grid $\mathbf{G}$ change with n.

Proposition 4.1 *Suppose that* $\mathbf{G}$ *becomes dense in* $\mathcal{X}$ *as* $n \to \infty$. *Then, under the conditions of Theorem 3.1,* $M_p \to 0$ *as* $n \to \infty$.

5 Modifications for unknown variance

All preceding theoretical results assume that the variance σ^2 is known. We now consider the situation in which σ^2 is unknown. If $\widehat{\sigma}^2$ is a consistent estimator of σ^2 and

$$\mathbf{Z} := \begin{bmatrix} Z_{p_1}(\xi_1) & \cdots & Z_{p_1}(\xi_G) & \cdots & Z_{p_J}(\xi_1) & \cdots & Z_{p_J}(\xi_G) \end{bmatrix}^t \to^L \mathbf{Z}^* \quad (11)$$

where $\mathbf{Z}^* \sim MVN\left(\mathbf{0}, \Sigma^*\right)$ for some symmetric positive definite matrix Σ^*, then by Slutsky's Theorem

$$\widehat{\mathbf{Z}} := \left[\begin{array}{ccccccc} \widehat{Z}_{p_1}(\xi_1) & \cdots & \widehat{Z}_{p_1}(\xi_G) & \cdots & \widehat{Z}_{p_J}(\xi_1) & \cdots & \widehat{Z}_{p_J}(\xi_G) \end{array}\right]^t \rightarrow^L \mathbf{Z}^*,$$

where $\widehat{Z}_p(\xi_j) := (\sigma/\widehat{\sigma})\, Z_p(\xi_j)$ for $p \in \{p_1, \ldots, p_J\}$ and $j \in \{1, \ldots, G\}$. To demonstrate that assuming the existence of such Σ^* is not unreasonable, we give the explicit form of Σ^* for a compactly supported kernel smooth in Proposition 5.1 below.

Then for any Borel set $A \in \mathbb{R}^{G*J}$, $P(\mathbf{Z} \in A) \rightarrow P(\mathbf{Z}^* \in A)$ and $P(\widehat{\mathbf{Z}} \in A) \rightarrow P(\mathbf{Z}^* \in A)$. Thus $P(\widehat{\mathbf{Z}} \in A) - P(\mathbf{Z} \in A) \rightarrow 0$ justifying $P(\mathbf{Z} \in A)$ from expressions (4) and (5) as an approximation to $P(\widehat{\mathbf{Z}} \in A)$.

Theorem 5.1 *Let $\hat{\sigma}^2$ be a consistent estimator of σ^2, and suppose that (11) holds. Then under the conditions of Theorem 4.1,*

$$\liminf_{n\rightarrow\infty} P\Big(L_{p_1}(x) \le \mu^{(p_1)}(x) \le U_{p_1}(x),$$

$$..., L_{p_J}(x) \le \mu^{(p_J)}(x) \le U_{p_J}(x), \forall x \in \mathcal{X}\Big) \ge 1 - \alpha$$

$$\textit{where } L_p(x) := \widehat{\mu}_I^{(p)}(x) - (\widehat{M}_p + \epsilon + \widehat{\beta}_{pL}(x)_I + z_\alpha \hat{\sigma} D_{pI}(x))$$

$$\textit{and } U_p(x) := \widehat{\mu}_I^{(p)}(x) - (\widehat{M}_p + \epsilon + \widehat{\beta}_{pU}(x)_I + z_\alpha \hat{\sigma} D_{pI}(x)),$$

and ϵ is an arbitrarily small positive number.

We propose using the following estimator of σ^2 (Craven and Wahba 1979, Loader 1999):

$$\widehat{\sigma}^2 := \frac{\sum_{i=1}^n \left[Y_i - \widehat{\mu}(x_i)\right]^2}{n - \nu}, \text{ where } \nu := \sum_{m=1}^n l_m(x_m). \tag{12}$$

The quantity $\nu\widehat{\sigma}^2/\sigma^2$ has an approximate chi-square distribution with ν degrees of freedom (Satterthwaite 1946, Cleveland 1979). So for better coverage probability with small samples, we can employ a multivariate t-distribution in place of a multivariate normal distribution with Σ as the scale matrix of the multivariate t-distribution in calculating z_α in (4) and (5).

Proposition 5.1 *Assume that a kernel regression estimator with kernel function $K(u)$ compactly supported on $\mathcal{X}$ is being employed to estimate $\mu(x)$ from model (1). Assume the bandwidth parameter $h \to 0$ as $n \to \infty$ and let the design points be uniform on $\mathcal{X}$. Then $\Sigma \to \Sigma^*$, where Σ is the variance-covariance matrix of $\boldsymbol{Z}$ and Σ^* is such that the covariance of $Z_a(\xi_j)$ and $Z_b(\xi_k)$ is*

$$\frac{\int_{\mathbb{R}} K^{(a)}(u)K^{(b)}(u)du}{\sqrt{\int_{\mathbb{R}} K^{(a)}(u)^2du \int_{\mathbb{R}} K^{(b)}(u)^2du}}$$

if $j = k$ and 0 otherwise.

6 Simulations and application to ethanol data

We performed simulations to assess this methodology for constructing simultaneous confidence bands for a mean response and its derivatives. To do this we generated 1000 data sets from (1) with $\mu(x) := \sin(2\pi x)+\cos(2\pi x)+\log(4/3+x)$ with $x_1, ..., x_n$ equispaced on $\mathcal{X} = [-1, 1]$ and $\sigma = .1$. We did this for $n \in \{50, 100\}$ and $\alpha \in \{.05, .20\}$. We estimated the mean response and its derivatives using compound estimation (Charnigo and Srinivasan 2009a) with filtration and extrapolation with $J = 4$, 27 centering points, $\beta = 15$ (150 during filtration and extrapolation), nearest neighbor local regression pointwise estimators using nearest neighbor fraction .30 (".15 during filtration and extrapolation), and $\kappa = 1.1$. We then placed simultaneous confidence bands for the mean response and first derivative ($p_1 = 0, p_2 = 1$) and simultaneous confidence bands for the mean response and first two derivatives ($p_1 = 0, p_2 = 1, p_3 = 2$), in each case using $G = 25$.

The results are in Table 1. In each case, the confidence bands are conservative, achieving greater than the nominal coverage probability. The bands constructed using $\alpha = .20$ for the mean response and first two derivatives capture the true mean response and first two derivatives only 86.3 and 84.0 percent of the time for sample sizes of 50 and 100, respectively. This is reassuring because it indicates that while the bands are indeed conservative, they are not unreasonably so. Figures 1 and 2 display some simulated data sets and their accompanying simultaneous confidence bands.

We also applied our methodology to a data set from Brinkman (1981) involving exhaust emissions. This data set has been examined elsewhere using nonparametric regression techniques (Cleveland 1979, Loader 1999). Loader discusses how to estimate the mean response and first two derivatives of the concentration of certain pollutants (NOx) with respect to the equivalence ratio (E) using local regression. Loader does not, however, discuss how to place confidence bands around the estimates.

We estimate the mean response and first two derivatives using compound estimation with filtration and extrapolation using the same tuning parameters as in the simulations. In Figure 3 we obtain simultaneous 95% confidence bands for the mean response and the first derivative. In Figure 4 we obtain simultaneous 95% confidence bands for the mean response and the first two derivatives.

An initial glance at these figures may lead to the perception that the bands seem wide. This perception is due to a couple of factors. The first is that this data has a relatively low signal-to-noise ratio. The estimated mean response has a range of only 3.4 while the estimated standard deviation is .9. The wide bands reflect the uncertainty inherent in this relatively large standard deviation. The second factor is that intuition about confidence intervals around the estimate of the mean response at a given point does not translate easily to confidence bands that are simultaneous over the mean response and one or more derivatives at all points. Despite the low signal-to-noise ratio, the confidence bands do allow us to identify ranges of E over which pollution is clearly increasing (.76 to .81, per Figure 3) and decreasing (.99 to 1.06).

The price of having bands that are simultaneous over the mean response and two derivatives as opposed to bands that are simultaneous over the mean response and just one derivative is wider bands. However, in this case that price is small. The band for the mean response in Figure 4 is .8% wider on average than the band for the mean response in Figure 3. The band for the first derivative of the mean response in Figure 4 is also .8% wider on average than the band for the first derivative in Figure 3.

7 Discussion and future research

We have provided a method for constructing simultaneous confidence bands for a mean response and its derivatives. Importantly, our methodology will work for *any* nonparametric regression estimator that is both self-consistent and linear in the observed responses. We have addressed how to estimate bias, interpolation error, and noise variance when a data analyst is unwilling to assume upper bounds for these quantities a priori. In each instance, we have provided a theorem demonstrating that, for large samples, the actual coverage probability still meets or exceeds the nominal coverage probability. Small sample pragmatics have not been neglected, however, as our simulations have demonstrated that the actual coverage probability meets or exceeds the nominal coverage probability for small samples also.

In addition, our methodology can be extended to provide confidence bands for derivative j_1 on a compact interval $E_1 \subset \mathcal{X}$, derivative j_2 on $E_2 \subset \mathcal{X}$, and so forth. Such an extension is potentially useful for two reasons. First, a data analyst simply may not be interested in examining every derivative on the full covariate space but may prefer, on scientific grounds, to identify a different region of interest for each derivative. In this case, the simultaneous confidence bands constructed over $E_1, \ldots, E_J$ will be narrower than if they had been constructed over all of $\mathcal{X}$ and then restricted to $E_1, \ldots, E_J$. Second, while a data analyst may be uncomfortable assuming upper bounds for some quantities on $\mathcal{X}$, he or she may be willing to assume them on $E_1, \ldots, E_J$. If these bounds are tight over $E_1, \ldots, E_J$, then the simultaneous confidence bands constructed over $E_1, \ldots, E_J$ may be further narrowed.

The methodology we have described does require that errors be normal, independent, and have constant variance. Yet there are applications for which simultaneous confidence bands would be desirable which do not satisfy these requirements (Pagan and Ullah 1999, chap. 4). Thus, generalizations which account for non-normal, correlated, and heteroscedastic errors are avenues for future research.

Acknowledgements. This material is based upon work supported by the National Science Foundation under Grant No. DMS-0706857. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] Brinkman, B.N. (1981), "Ethanol fuel - a single-cylinder engine study of efficiency and exhaust emissions," *SAE Transactions*, **90**, 1410-1424.

[2] Charnigo, R., and Srinivasan, C. (2009a), "Self-consistent estimation of mean response functions and their derivatives," Submitted for publication and available at {www.richardcharnigo.net/TechReports}.

[3] —— (2009b), "On simultaneous estimation of a mean response and its derivatives," Submitted for publication and available at {www.richardcharnigo.net/TechReports}.

[4] Claeskens, G., and Van Keilegom, I. (2003), "Bootstrap confidence bands for regression curves and their derivatives," *Annals of Statistics*, **31**, 1852-1884.

[5] Cleveland, W.S. (1979), "Robust locally weighted regression and smoothing scatterplots," *Journal of the American Statistical Association*, **74**, 820-836.

[6] Craven, P., and Wahba, G. (1979), "Smoothing noisy data with spline functions," *Numerische Mathematik*, **31**, 377-403.

[7] Gyorfi, L., Kohler, M., Krzyzak, A., and Walk, H. (2002), *A Distribution-Free Theory of Nonparametric Regression.* Springer, New York.

[8] Hardle, W. (1990), *Applied Nonparametric Regression.* Cambridge University Press, Cambridge.

[9] Knafl, G., Sacks, J., and Ylvisaker, D. (1985), "Confidence bands for regression functions," *Journal of the American Statistical Association*, **80**, 683-691.

[10] Loader, C. (1999), *Local Regression and Likelihood.* Springer, New York.

[11] Pagan, A., and Ullah, A. (1999), *Nonparametric Econometrics.* Cambridge University Press, Cambridge.

[12] Priestley, M.B., and Chao, M.T. (1972), "Nonparametric function fitting," *Journal of the Royal Statistical Society, Series B*, **34**, 385-392.

[13] Reinsch, H. (1967), "Smoothing by spline functions," *Numerische Mathematik*, **10**, 177-183.

[14] Satterthwaite, F.E. (1946), "An approximate distribution of estimates of variance components," *Biometrics Bulletin*, **2**, 110-114.

[15] Schoenberg, I.J. (1964), "Spline functions and the problem of graduation," *Mathematics*, **52**, 947-950.

[16] Sun, J., and Loader, C. (1994), "Confidence bands for linear regression smoothing," *Annals of Statistics*, **22**, 1328-1345.

[17] Wu, W.B., and Zhou, Z. (2007), "Inference of trends in time series," *Journal of the Royal Statistical Society, Series B*, **69**, 391-410.

A Appendix: Proofs of Technical Results

Proof of Theorem 2.1. Then

$$
\begin{aligned}
& P\left(\cap_{p=p_1}^{p_J}\left\{|\widehat{\mu}_I^{(p)}(x) - \mu^{(p)}(x)| \leq M_p + B_{pI}(x) + z_\alpha \sigma D_{pI}(x)\right\}, x \in \mathcal{X}\right) \\
\geq\ & P(\cap_{p=p_1}^{p_J}\left\{|\widehat{\mu}_I^{(p)}(x) - \mu_I^{(p)}(x)| \leq B_{pI}(x) + z_\alpha \sigma D_{pI}(x)\right\}, x \in \mathcal{X}) \qquad (13) \\
=\ & P\left(\cap_{p=p_1}^{p_J}\left\{|\widehat{\mu}^{(p)}(x) - \mu^{(p)}(x)| \leq B_p(x) + z_\alpha \sigma D_p(x)\right\}, x \in \mathbf{G}\right) \\
=\ & P\left(\cap_{p=p_1}^{p_J}\left\{\left|\sum_{i=1}^{n} l_i^{(p)}(x)[\mu(x_i) + \epsilon_i] - \mu^{(p)}(x)\right|\right.\right. \\
& \leq B_p(x) + z_\alpha \sigma D_p(x)\}, x \in \mathbf{G}) \\
\geq\ & P\left(\cap_{p=p_1}^{p_J}\left\{\left|\sum_{i=1}^{n} l_i^{(p)}(x)\epsilon_i\right| \leq z_\alpha \sigma D_p(x)\right\}, x \in \mathbf{G}\right) \qquad (14) \\
\geq\ & 1 - \alpha, \qquad (15)
\end{aligned}
$$

where lines (13) and (14) follow from the triangle inequality and line (15) follows from the determination of z_α.

Proof of Theorem 3.1. Let

$$
\begin{aligned}
A \ := \ & \left\{\widehat{\mu}_I^{(p)}(x) - z_\alpha \sigma D_{pI}(x) - M_p - \left|Bias_\mu[\widehat{\mu^{(p)}(x)}]\right|_I \leq \mu^{(p)}(x)\right. \\
& \leq \widehat{\mu}_I^{(p)}(x) + z_\alpha \sigma D_{pI}(x) + M_p + \left|Bias_\mu[\widehat{\mu^{(p)}(x)}]\right|_I, \\
& \forall x \in \mathcal{X}, \forall p \in \{p_1, ..., p_J\}\}, \\
B \ := \ & \left\{\widehat{\mu}_I^{(p)}(x) - z_\alpha \sigma D_{pI}(x) - M_p - \widehat{\beta}_{pL}(x)_I \leq \mu^{(p)}(x)\right. \qquad (16) \\
& \left.\leq \widehat{\mu}_I^{(p)}(x) + z_\alpha \sigma D_{pI}(x) + M_p + \widehat{\beta}_{pU}(x)_I, \forall x \in \mathcal{X}, \forall p \in \{p_1, ..., p_J\}\right\}, \text{ and} \\
C \ := \ & \left\{\widehat{\beta}_{pU}(x)_I \geq \left|Bias_\mu[\widehat{\mu^{(p)}(x)}]\right|_I\right. \qquad (17) \\
& \left.\text{and } \widehat{\beta}_{pL}(x)_I \geq \left|Bias_\mu[\widehat{\mu^{(p)}(x)}]\right|_I, \forall x \in \mathcal{X}, \forall p \in \{p_1, ..., p_J\}\right\}.
\end{aligned}
$$

Now note that $(A \cap C) \subset B$ and so $P(B) \geq P(A \cap C) = P(A) - P(A \cap C^c)$. By Theorem 2.1, $P(A) \geq 1 - \alpha$ and since $\widehat{\mu}^{(a)}(x)$ converges to $\mu^{(a)}(x)$ at the rate r_a, $P(A \cap C^c) \leq P(C^c) \rightarrow 0$ which implies the desired result.

Lemma A.1 *If $f(x)$ is continuous on $\mathcal{X}$, where $\{\gamma_i\}_{i=1}^m$ constitute a grid of points that become dense in $\mathcal{X}$ as $m \to \infty$, then $\max_{i\in\{1,...,m\}} f(\gamma_i) \to \sup_{x\in\mathcal{X}} f(x)$ as $m \to \infty$.*

Proof of Lemma A.1. Let $f(t_0) = \sup_{x\in\mathcal{X}} f(x)$. Then $\forall\delta > 0, \exists M \in \mathbb{N}$ such that $\forall m \geq M, \exists i^* \in \{1, ..., m\}$ such that $|\gamma_{i^*} - t_0| < \delta$. Then by the uniform continuity of f on $\mathcal{X}$, $\forall\epsilon > 0, \exists M$ such that $\forall m \geq M, \exists i^* \in \{1, ..., m\}$ such that $|f(\gamma_{i^*}) - \sup_{x\in\mathcal{X}} f(x)| < \epsilon$. Since $f(\gamma_{i^*}) \leq \max_{i\in\{1,...,m\}} f(\gamma_i) \leq \sup_{x\in\mathcal{X}} f(x)$, this implies that $\forall\epsilon > 0$, $\exists M$ such that $\forall m \geq M$, $|\max_{i\in\{1,...,m\}} f(\gamma_i) - \sup_{x\in\mathcal{X}} f(x)| < \epsilon$ which yields the desired result.

Proof of Theorem 4.1. Let $\widehat{f}(x) := |\widehat{\mu}^{(p)}(x) - \widehat{\mu}_I^{(p)}(x)|$ and $f(x) := |\mu^{(p)}(x) - \mu_I^{(p)}(x)|$. Define $\widehat{M}_p(m) := \max_{i\in\{1,...,m\}} \widehat{f}(\gamma_i)$ and $M_p(m) := \max_{i\in\{1,...,m\}} f(\gamma_i)$. The fact that $\widehat{\mu}^{(p)}(x) \to^P \mu^{(p)}(x), \forall x \in \mathcal{X}$, implies $\forall m \in \mathbb{N}$, $\widehat{M}_p(m) \to^P M_p(m)$. So $\forall m \in \mathbb{N}$, $\exists N_0(m)$ such that $\forall n \geq N_0(m)$, $P\left(\left|\widehat{M}_p(m) - M_p(m)\right| \geq 1/m\right) \leq 1/m$.

Now for $m > 1$ define $N_m := \max\left(N_0(m), N_{m-1} + 1\right)$ and define

$$\widetilde{M}_p := \begin{cases} M_p(n) & : \quad n < N_{m_0} \\ M_p(m) & : \quad N_m \leq n < N_{m+1} \end{cases}$$

for $m \geq m_0$, where m_0 is an arbitrary positive integer.

Let $\epsilon > 0$. Then $\exists \widetilde{m}_1 \in \mathbb{N}$ such that $1/\widetilde{m}_1 \leq \epsilon/2$. Also by Lemma A.1, $\exists \widetilde{m}_2$ such that $\forall m \geq \widetilde{m}_2$,

$$P\left(\left|\max_{i\in\{1,...,m\}} f(\gamma_i) - \sup_{x\in\mathcal{X}} f(x)\right| \geq \epsilon/2\right) = 0 \leq \epsilon/2.$$

Let $m_\epsilon = \max\{\widetilde{m}_1, \widetilde{m}_2\}$. For $N_{m_\epsilon+1} \leq n$, $\left|\widehat{M}_p - \widetilde{M}_p\right| = \left|\widehat{M}_p(m_n) - M_p(m_n)\right|$

where $N_{m_n} \le n < N_{m_\epsilon+1}$ and so

$$\begin{aligned}
& P\left(\left|\widehat{M}_p - \sup_{x\in\mathcal{X}} f(x)\right| \ge \epsilon\right) \\
\le\ & P\left(\left|\widehat{M}_p - \widetilde{M}_p\right| + \left|\widetilde{M}_p - \sup_{x\in\mathcal{X}} f(x)\right| \ge \epsilon\right) \\
\le\ & P\left(\left|\widehat{M}_p(m_n) - M_p(m_n)\right| \ge \epsilon/2\right) + P\left(\left|M_p(m_n) - \sup_{x\in\mathcal{X}} f(x)\right| \ge \epsilon/2\right) \\
\le\ & P\left(\left|\widehat{M}_p(m_n) - M_p(m_n)\right| \ge 1/m_\epsilon\right) + \epsilon/2 \\
\le\ & P\left(\left|\widehat{M}_p(m_n) - M_p(m_n)\right| \ge 1/m_n\right) + \epsilon/2 \\
\le\ & 1/m_n + \epsilon/2 \\
\le\ & 1/m_\epsilon + \epsilon/2 \\
\le\ & \epsilon,
\end{aligned}$$

which implies the desired result.

Proof of Proposition 4.1. Let $\epsilon > 0$ and $x \in \text{int}(\mathcal{X})$. There exists $\delta > 0$ depending on ϵ but not x such that if $y \in \mathcal{X}$ and $|x-y| < \delta$, then $\left|\mu^{(p)}(x) - \mu^{(p)}(y)\right| < \epsilon/2$ by the uniform continuity of $\mu^{(p)}$ on $\mathcal{X}$. Also, since $\mathbf{G}$ becomes dense in $\mathcal{X}$ as $|\mathbf{G}| \to \infty$, there exists G (depending on δ but not on x) such that $\forall|\mathbf{G}| \ge G$, there exist $a, b \in \mathbf{G}$ such that $a \le x \le b$, $|b-a| < \delta$. So there exists G such that $\forall|\mathbf{G}| \ge G$, there exist $a, b \in \mathbf{G}$ such that $a \le x \le b$, $|\mu^{(p)}(a) - \mu^{(p)}(x)| < \epsilon/2$ and $|\mu^{(p)}(b) - \mu^{(p)}(a)| < \epsilon/2$. If $a = x = b$, then $|\mu^{(p)}(x) - \mu_I^{(p)}(x)| = 0$. Otherwise,

$$\begin{aligned}
& \left|\mu^{(p)}(x) - \mu_I^{(p)}(x)\right| \\
=\ & \left|\mu^{(p)}(x) - \frac{x-a}{b-a}\left[\mu^{(p)}(b) - \mu^{(p)}(a)\right] - \mu^{(p)}(a)\right| \\
\le\ & \left|\mu^{(p)}(x) - \mu^{(p)}(a)\right| + \left|\mu^{(p)}(b) - \mu^{(p)}(a)\right| \qquad (18) \\
\le\ & \epsilon,
\end{aligned}$$

where line (18 holds when the grid size is greater than or equal to G. Now since

$\left|\mu^{(p)}(x) - \mu_I^{(p)}(x)\right| \leq \epsilon, \forall x \in \text{int}(\mathcal{X})$, where $\text{int}(\mathcal{X})$ denotes the interior of the set $\mathcal{X}$, then $\sup_{x\in\mathcal{X}} \left|\mu^{(p)}(x) - \mu_I^{(p)}(x)\right| \leq \epsilon$, which implies the desired result.

Proof of Proposition 5.1. Let B and C be defined as in (16) and (17), respectively. Define

$$\begin{aligned} D &:= \{\widehat{M}_p + \epsilon \geq M_p, \forall p \in \{p_1, ..., p_J\}\}, \\ E &:= \{\cup_{r=1}^J \max_{x\in\mathbf{G}} |\widehat{Z}_{p_r}(x)| \leq z_\alpha\}, \\ F &:= \{\cup_{r=1}^J \max_{x\in\mathbf{G}} |Z_{p_r}(x)| \leq z_\alpha\}. \end{aligned}$$

Then $C \cap D \cap E \subset B$ and therefore

$$\begin{aligned} \limsup_{n\to\infty} P(B^c) &\leq \limsup_{n\to\infty} P(C^c) + P(D^c) + P(E^c) \\ &\leq \limsup_{n\to\infty} P(C^c) + \limsup_{n\to\infty} P(D^c) \\ &\quad + \limsup_{n\to\infty} P(F^c) + \limsup_{n\to\infty}[P(E^c) - P(F^c)] \\ &= \limsup_{n\to\infty} P(F^c) \\ &\leq \alpha. \end{aligned}$$

Proof of Proposition 5.1. Without loss of generality, take $\mathcal{X} := [-1, 1]$. First consider the elements of Σ corresponding to $j \neq k$. Consider n large enough so that $|\xi_j - \xi_k| > 2h$. Consider that $K^{(a)}((\xi_j - x_i)/h) \neq 0$ only if $|\xi_j - x_i| \leq h$. But if $|\xi_j - x_i| \leq h$ then

$$|\xi_k - x_i| = |\xi_k - \xi_j + \xi_j - x_i| \geq ||\xi_k - \xi_j| - |\xi_j - x_i|| > h,$$

which means that $K^{(b)}((\xi_k - x_i)/h) = 0$. So for n large enough that $|\xi_j - \xi_k| > 2h$, the elements of Σ corresponding to $j \neq k$ are 0.

For $j = k$ we have

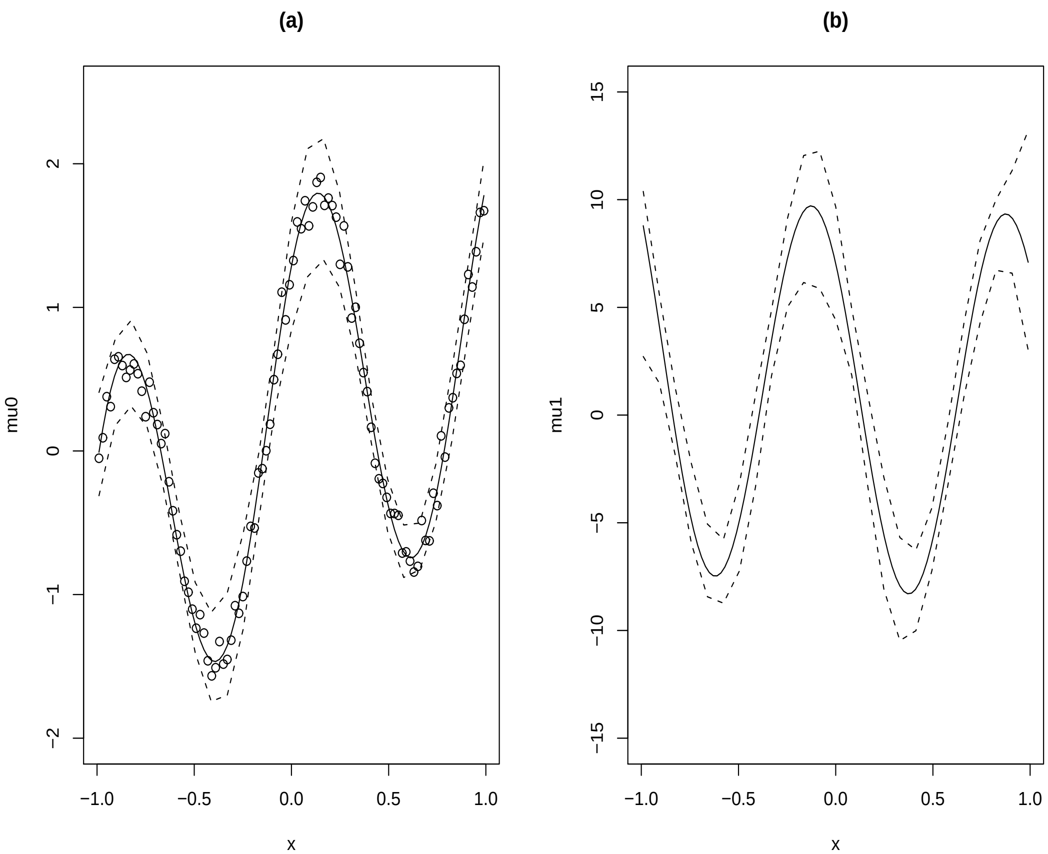
$$\begin{aligned}
& \frac{\sum_{i=1}^{n} l_i^{(a)}(\xi_j) l_i^{(b)}(\xi_k)}{\sqrt{\sum_{i=1}^{n} l_i^{(a)}(\xi_j)^2 \sum_{i=1}^{n} l_i^{(b)}(\xi_j)^2}} \\
= \quad & \frac{\sum_{i=1}^{n} K^{(a)}(\frac{\xi_j - x_i}{h}) K^{(b)}(\frac{\xi_j - x_i}{h})}{\sqrt{\sum_{i=1}^{n} K^{(a)}(\frac{\xi_j - x_i}{h})^2 \sum_{i=1}^{n} K^{(b)}(\frac{\xi_j - x_i}{h})^2}} \\
\to \quad & \frac{\int_{-1}^{1} K^{(a)}(\frac{\xi_j - x}{h}) K^{(b)}(\frac{\xi_j - x}{h}) dx}{\sqrt{\int_{-1}^{1} K^{(a)}(\frac{\xi_j - x}{h})^2 dx \int_{-1}^{1} K^{(b)}(\frac{\xi_j - x}{h})^2 dx}} \\
= \quad & \frac{\int_{(\xi_j - 1)/h}^{(\xi_j + 1)/h} K^{(a)}(u) K^{(b)}(u) du}{\sqrt{\int_{(\xi_j - 1)/h}^{(\xi_j + 1)/h} K^{(a)}(u)^2 du \int_{(\xi_j - 1)/h}^{(\xi_j + 1)/h} K^{(b)}(u)^2 du}} \\
= \quad & \frac{\int_{\mathbb{R}} K^{(a)}(u) K^{(b)}(u) du}{\sqrt{\int_{\mathbb{R}} K^{(a)}(u)^2 du \int_{\mathbb{R}} K^{(b)}(u)^2 du}}.
\end{aligned}$$

Table 1: Simulation results for simultaneous confidence bands

	α=.05		α=.20	
	0,1	0,1,2	0,1	0,1,2
n=50	100%	99.3%	99.6%	86.3%
n=100	100%	96.7%	98.0%	84.0%

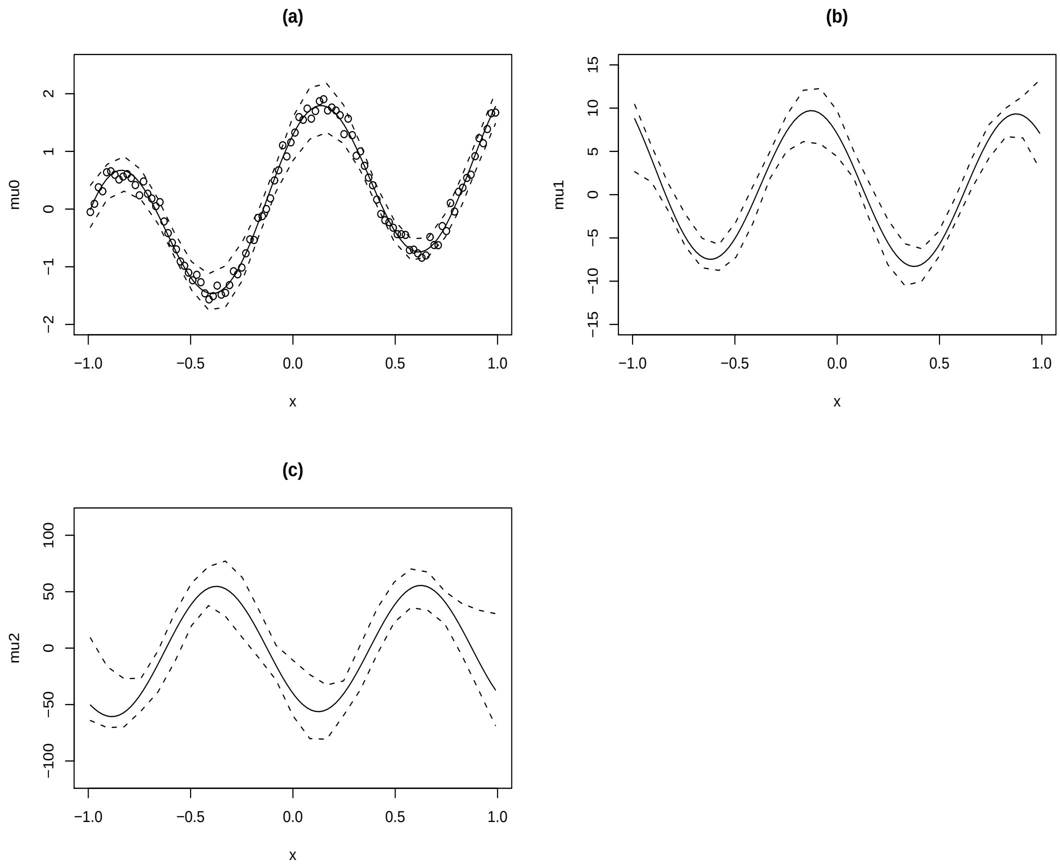
The columns labeled $0, 1$ indicate that simultaneous confidence bands were constructed for the mean response and its first derivative. The columns labeled $0, 1, 2$ indicate that simultaneous confidence bands were constructed for the mean response and its first two derivatives. The entries represent the percentages of simultaneous confidence bands that contained the true mean response and derivative(s).

Figure 1: Simultaneous Confidence Bands for the Mean Response and First Derivative



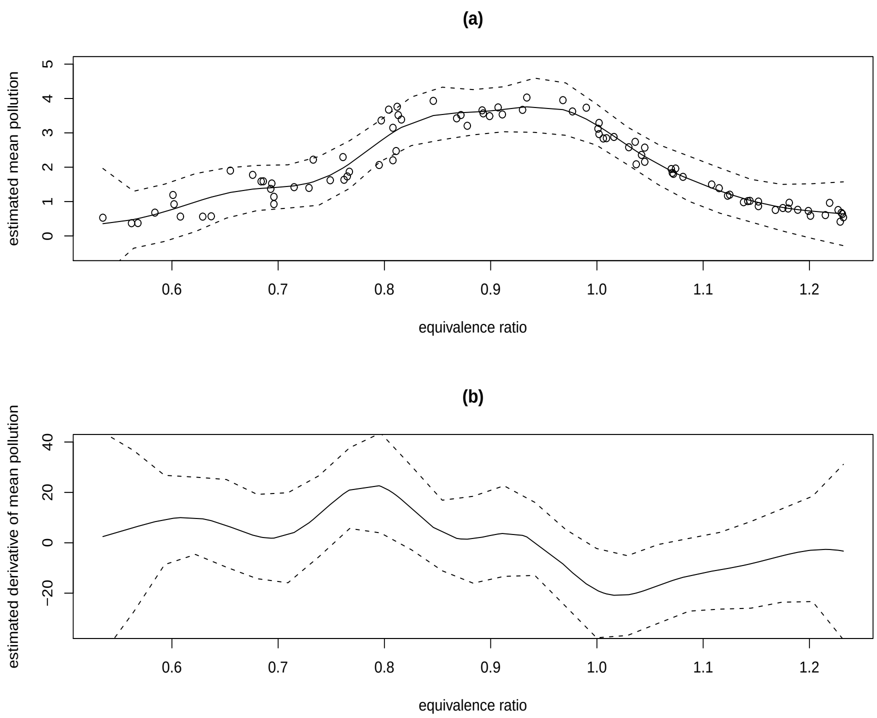
The solid curves indicate the true mean response and first derivative in panels (a) and (b), respectively. The simulated data for this sample of size 100 is also displayed in panel (a). The dashed curves represent the confidence bands with $\alpha = .05$. In this case, the confidence bands contain the true mean response and first derivative.

Figure 2: Simultaneous Confidence Bands for the Mean Response and Two Derivatives



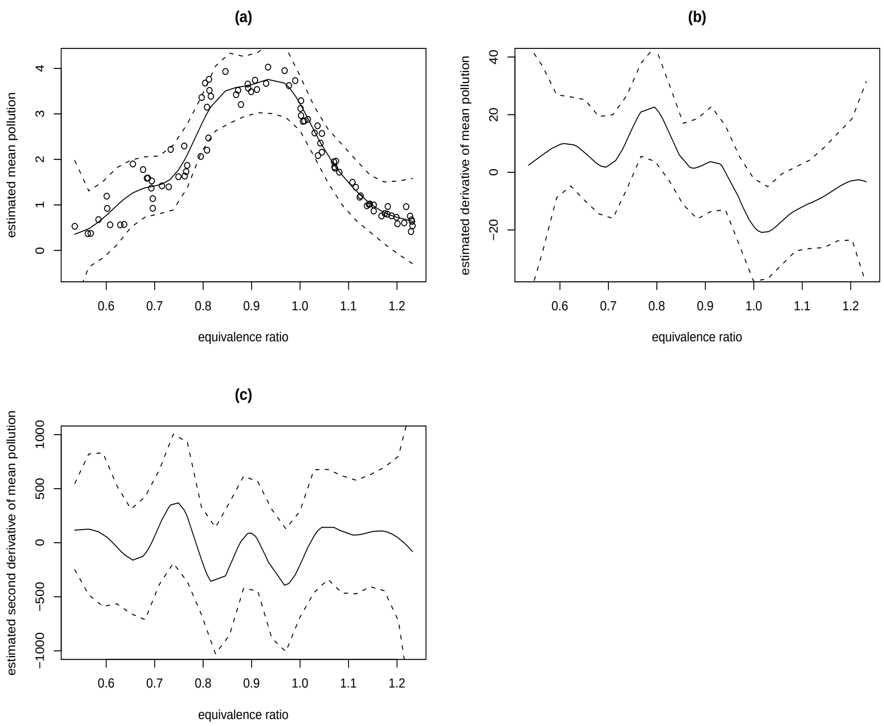
The solid curves indicate the true mean response, first derivative, and second derivative in panels (a), (b), and (c), respectively. The simulated data for this sample of size 100 is also displayed in panel (a). The dashed curves represent the confidence bands with $\alpha = .05$. In this case, the confidence bands contain the true mean response and first two derivatives.

Figure 3: Mean response and first derivative of ethanol data



The solid curves in panels (a) and (b) indicate the estimated mean response and first derivative, respectively. The circles represent the observed data. The dashed curves indicate the 95% confidence bands.

Figure 4: Mean response and first two derivatives of ethanol data



The solid curves in panels (a), (b), and (c) indicate the estimated mean response, first derivative, and second derivative, respectively. The circles represent the observed data. The dashed curves indicate the 95% confidence bands.