Statistica Sinica **20** (2010), 637-659

SIMULTANEOUS CONFIDENCE INTERVALS FOR SEMIPARAMETRIC LOGISTICS REGRESSION AND CONFIDENCE REGIONS FOR THE MULTI-DIMENSIONAL EFFECTIVE DOSE

Jialiang Li[1], Chunming Zhang[2], Kjell A. Doksum[2] and Erik V. Nordheim[2]

[1]*National University of Singapore and* [2]*University of Wisconsin, Madison*

Abstract: We construct pointwise and simultaneous confidence intervals for the link function in a semiparametric logistic regression model based on the logit link function. Simultaneous confidence intervals are especially important for semiparametirc regression models since they allow inference to be made over the whole predictor space. These intervals are used to construct confidence regions for the multi-dimensional effective dose – that is, for the set of multivariate covariate values where the probability that the binary response is one is equal to a preassigned value p. We construct methods using inequalities, asymptotics and a semiparametric bootstrap. Simulations and examples are provided to demonstrate the performance and to compare the methods. The bootstrap procedure yields accurate confidence regions for the multi-dimensional effective dose.

Key words and phrases: Bootstrap, local likelihood, multi-dimensional effective dose, semiparametric logistic regression model, simultaneous confidence interval.

1. Introduction

Semiparametric inference techniques have become an increasingly important tool for solving statistical estimation and inference problems. A semiparametric model is partly but not fully characterized by some interpretable Euclidean parameters that are usually of primary interest to researchers. The model also involves one or more infinite-dimensional components that are usually reserved for nuisance parameters. Such models can produce more accurate inference results than parametric models while maintaining the interpretability for the parametrized components. Various forms of semiparametric models have been suggested in the literature for different goals. In this paper we focus on dichotomous responses. We explore using a generalized partially linear model (Green and Yandell (1985), Hastie and Tibshirani (1990), Carroll, Fan, Gijbels and Wand (1997)) as our semiparametric modeling device to study the construction of simultaneous confidence intervals for the semiparametric logistic regression function.

We observe the sample $\{(X_{1i}, \ldots, X_{ki}, Z_i, Y_i), i = 1, \ldots, n\}$ from n independent subjects. Suppose the ith subject depends on covariates $\mathbf{X}_i = (X_{1i}, \ldots, X_{ki})^T \in \mathcal{X} \subset \mathbb{R}^k$ and $Z_i \in \mathcal{Z} \subset \mathbb{R}$, where both $\mathcal{X}$ and $\mathcal{Z}$ are bounded sets, and has a dichotomous response $Y_i \in \{0, 1\}$.

Both $\mathbf{X}$ and Z could effect the probability of the occurrence of Y and we use the following semiparametric logistic regression model to study the relationship between covariates and the response,

$$\begin{aligned} \log\{\frac{p_i}{(1-p_i)}\} &= \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \gamma(Z_i) \\ &= \mathbf{X}_i^T \boldsymbol{\beta} + \gamma(Z_i), \end{aligned} \tag{1.1}$$

where the logit transformation of the mean response $p_i = E(Y_i|\mathbf{X}_i, Z_i)$ is linearly dependent on $\mathbf{X}_i$ and is dependent on Z_i through a smooth function $\gamma(\cdot)$. Both the vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)^T$ and the function $\gamma(\cdot)$ are unknown. This model specification is particularly appealing when $\mathbf{X}$ consists of predictor variables of interest and the effects of Z are considered as a nuisance. This model allows us to make inference about the effects of $\mathbf{X}$ using a parametric form while making minimal assumptions on the effects of Z by using a fully nonparametric function.

Model (1.1) offers a flexible approach and has been used in various applications. Severini and Staniswalis (1994) employed it to study the reproduction of Torrey yucca plants under different environments. The scientifically interesting factor of growing site was modeled as a parametric component and the confounding factor of plant height was modeled as a nonparametric component. Hunsberger (1994) studied the occurrence of an intraoperative cardiac complication in vascular surgery patients subject to two risk factors using this model. The two risk factors included the level of American Society of Anesthesiologist, which needs a parametric interpretation, and was modeled accordingly, and the duration of the operation whose effects were modeled nonparametrically.

The estimation technique we consider in this paper is the local likelihood approach (Staniswalis (1989), Fan, Heckman and Wand (1995)). Asymptotic results for such estimators were given in Carroll et al. (1997). One possible alternative for fitting this model is the penalized quasi-likelihood approach (Green and Yandell (1985), Hastie and Tibshirani (1990)). Mammen and van de Geer (1997) provided asymptotic results for such estimators by using empirical process theory. These results are sufficient for most statistical inferences. However, a certain gap remains between the current theoretical knowledge and the goal of constructing a simultaneous confidence interval for the semiparametric regression function $\eta(\mathbf{x}, z) = \mathbf{x}^T\boldsymbol{\beta} + \gamma(z)$ across all $\mathbf{x} \in \mathcal{X}$ and $z \in \mathcal{Z}$. We propose a way to achieve this goal using local likelihood estimators. The method can be implemented for the penalized quasi-likelihood estimator as well, although technical justification needs to be completed in a different fashion from this paper.

Another unique feature of this paper is that we address an important but under-appreciated statistical problem. In some scientific experiments, researchers might be interested in finding what joint values of the predictor variables will produce a given response. When multiple predictors are present in a model, the question of obtaining such values for fixed response, the multi-dimensional effective dose, has not been fully addressed. Li, Nordheim, Zhang, and Lehner (2008a) studied this problem under a parametric logistic regression model. We extend this approach to the more general semiparametric logistic regression model. This allows application to a broader range of observed data. In our approach, the desired confidence region for the multi-dimensional effective dose is related to the simultaneous confidence interval for the regression function.

This paper is organized as follows. Section 2 reviews a method for fitting a semiparametric regression model. In Sections 3 and 4, we develop pointwise and simultaneous confidence intervals for the regression link function $\eta(\mathbf{x}, z) = \mathbf{x}^T\boldsymbol{\beta} + \gamma(z)$. Since this function is of a semiparametric form, special attention needs to be paid to the large sample properties of the estimators involved. In Section 5, we adapt our methods to calculate confidence regions for the multi-dimensional effective dose. Finding such regions can help scientific researchers design the "dose" levels of multiple covariates to achieve a dichotomous outcome with a specific probability. Section 6 presents Monte Carlo simulation results to assess the performance of our proposed methods. Section 7 presents an example that illustrates the use of our methods for a real-world problem. Section 8 points out future research directions.

2. Local Likelihood Estimation

We sketch a brief introduction on how to fit the semiparametric model (1.1) using the local likelihood approach (Fan and Chen (1999), Fan et al. (1995)). First, we notice that in a small neighborhood of z_0, one can approximate $\gamma(z)$ locally by a linear function

$$\gamma(z) \approx \gamma(z_0) + \gamma^{'}(z_0)(z - z_0) = a + b(z - z_0). \tag{2.1}$$

Based on independent observations $\{(\mathbf{X}_i, Z_i, Y_i)\}$, the parameters can be estimated by maximizing the local log-likelihood

$$\mathcal{L} = \sum_{i=1}^{n} [Y_i \log(p_i) + (1 - Y_i) \log(1 - p_i)] K_h(Z_i - z), \tag{2.2}$$

where

$$p_i = [1 + \exp\{-(\mathbf{X}_i^T\boldsymbol{\beta} + a + b(Z_i - z))\}]^{-1}, \tag{2.3}$$

$K_h(\cdot) = K(\cdot/h)/h$ is re-scaled from a known kernel function K, and h is a bandwidth.

We now describe the profile likelihood method for fitting the semiparametric model. For each given $\boldsymbol{\beta}$, by applying the local likelihood method with a bandwidth h, we obtain $\hat{\gamma}(z;\boldsymbol{\beta},h)$, depending on h and $\boldsymbol{\beta}$. After substituting into the original model, we obtain a pseudo parametric model:

$$\log\left\{\frac{p}{1-p}\right\} = \mathbf{X}^T\boldsymbol{\beta} + \hat{\gamma}(z;\boldsymbol{\beta},h). \tag{2.4}$$

Regarding the above as a parametric model with parameter $\boldsymbol{\beta}$, we obtain the profile likelihood estimators $\hat{\boldsymbol{\beta}}(h)$ and $\hat{\gamma}(z;\hat{\boldsymbol{\beta}}(h),h)$ by using maximum likelihood.

In this paper we do not elaborate on the computational aspect of the semiparametric model. More details about fitting the local likelihood model can be found in Zhang (2003, 2008).

3. Pointwise Confidence Interval

The asymptotic properties of the estimators obtained through maximizing the local log-likelihood (2.2) can be studied via the following theorem, extended from Theorem 1 in Carroll et al. (1997). The proof is in the Appendix.

Theorem 3.1. *Suppose for any $z \in \mathcal{Z}$, the density $f(z)$ is Lipschitz continuous and bounded away from 0, the function $\gamma(z)$ is twice continuously differentiable, $E(\mathbf{X}\mathbf{X}^T|Z=z)$ is non-singular, and both $E(\mathbf{X}\mathbf{X}^T|Z)$ and $E(\mathbf{X}\mathbf{X}^T|Z)^{-1}$ are Lipschitz continuous. Assume that the kernel $K(\cdot)$ is a symmetric density function with compact support and that the parameter space of $\boldsymbol{\beta}$ is a compact subset of $\mathbb{R}^k$. Then as $n\to\infty, h\to 0$, and $nh\to\infty$,*

$$\sqrt{nh}\begin{pmatrix} \mathbf{x}_1^T\hat{\boldsymbol{\beta}} + \hat{\gamma}(z_1) - \mathbf{x}_1^T\boldsymbol{\beta} - \gamma(z_1) - b(\mathbf{x}_1,z_1) \\ \vdots \\ \mathbf{x}_L^T\hat{\boldsymbol{\beta}} + \hat{\gamma}(z_L) - \mathbf{x}_L^T\boldsymbol{\beta} - \gamma(z_L) - b(\mathbf{x}_L,z_L) \end{pmatrix} \to_d N(0,\Sigma), \tag{3.1}$$

for every finite set of points $(\mathbf{x}_1,z_1),\ldots,(\mathbf{x}_L,z_L)$ in $\mathcal{X}\times\mathcal{Z}$.

The lth diagonal entry of the matrix Σ is

$$\sigma(\mathbf{x}_l,z_l)^2 = nh\mathrm{Var}\left(\mathbf{x}_l^T\hat{\boldsymbol{\beta}} + \hat{\gamma}(z_l)\right) = (\mathbf{x}_l^T,1)\Psi(z_l)(\mathbf{x}_l^T,1)^T, \tag{3.2}$$

where $\Psi(z_l)$ is the variance-covariance matrix of $\sqrt{nh}(\hat{\boldsymbol{\beta}}^T,\hat{\gamma}(z_l))^T$. The (l,l')th off-diagonal entry of the matrix Σ is

$$nh\,\mathrm{Cov}\left(\mathbf{x}_l^T\hat{\boldsymbol{\beta}} + \hat{\gamma}(z_l), \mathbf{x}_{l'}^T\hat{\boldsymbol{\beta}} + \hat{\gamma}(z_{l'})\right) = \begin{cases} \Upsilon(\mathbf{x}_l,\mathbf{x}_{l'}) & z_l \neq z_{l'} \\ (\mathbf{x}_l^T,1)\Psi(z_l)(\mathbf{x}_{l'}^T,1)^T & z_l = z_{l'}. \end{cases} \tag{3.3}$$

Using Theorem 3.1, a pointwise confidence interval for $\eta(\mathbf{x}, z) = \mathbf{x}^T\boldsymbol{\beta} + \gamma(z)$ is

$$\mathbf{x}^T\hat{\boldsymbol{\beta}} + \hat{\gamma}(z) - b(\mathbf{x}, z) \pm \mathsf{z}_{\alpha/2}\sigma(\mathbf{x}, z)(\sqrt{nh})^{-1}, \tag{3.4}$$

where $\mathsf{z}_{\alpha/2}$ is the upper $\alpha/2$ quantile of the standard normal distribution.

The explicit expressions of $b(\mathbf{x}, z)$, $\Psi(z)$ and $\Upsilon(\mathbf{x}_l, \mathbf{x}_{l'})$ are given by (A.16), (A.17) and (A.18), respectively, in the Appendix. It is shown that the bias $b(\mathbf{x}, z)$ is of order $O(h^2)$. Thus we can neglect its effect when h is close to 0 for the purposes of implementation. We found in numerous simulation studies that ignoring this term has little influence on the performance of the above pointwise confidence intervals in terms of coverage rates. This strategy sometimes performs even better than plugging in the estimator $\hat{b}(\mathbf{x}, z)$ which might increase the variance more than reduce the bias (Sun and Loader (1994)). In practice we replace $\sigma(\mathbf{x}, z)$ with the consistent estimator

$$\hat{\sigma}(\mathbf{x}, z) = \sqrt{(\mathbf{x}^T, 1)\hat{\Psi}(z)(\mathbf{x}^T, 1)^T} \tag{3.5}$$

based on a moment estimator of $\Psi(z)$, whose consistency follows easily from the Law of Large Numbers.

4. Simultaneous Confidence Intervals

Pointwise confidence intervals are not adequate for making inference about the regression function $\eta(\mathbf{x}, z)$ across the whole range of $\{\mathbf{x}, z\}$. We study three methods for constructing a simultaneous confidence interval for $\eta(\mathbf{x}, z)$ valid for all $x \in \mathcal{X}$ and $z \in \mathcal{Z}$.

Before we present our methods, we present a theory that describes the asymptotic behavior of the normalized process

$$W_n(\mathbf{x}, z) = \sqrt{nh}[\mathbf{x}^T\hat{\boldsymbol{\beta}} + \hat{\gamma}(z) - \mathbf{x}^T\boldsymbol{\beta} - \gamma(z)]/\hat{\sigma}(\mathbf{x}, z) - \sqrt{nh}b(\mathbf{x}, z)/\hat{\sigma}(\mathbf{x}, z) \tag{4.1}$$

$$\equiv Z_n(\mathbf{x}, z) - \delta(\mathbf{x}, z) \qquad (\mathbf{x}, z) \in \mathcal{X} \times \mathcal{Z}. \tag{4.2}$$

This defines a process indexed by $(\mathbf{x}, z) \in \mathcal{X} \times \mathcal{Z}$. We prove the following theorem in the Appendix. Let $\xrightarrow{w}$ denote process weak convergence as defined in van der Vaart (1998).

Theorem 4.1. *Assume the same conditions as in Theorem* 3.1. *Then as* $n \to \infty$, $h \to 0$, *and* $nh \to \infty$, $W_n(\mathbf{x}, z) \xrightarrow{w} W(\mathbf{x}, z)$ *which is a Gaussian random field with mean* 0 *and covariance function*

$$\rho((\mathbf{x}, z), (\mathbf{x}', z')) = \begin{cases} \langle s(\mathbf{x}, z), s(\mathbf{x}', z')\rangle & z = z' \\ \frac{\Upsilon(\mathbf{x}_l, \mathbf{x}_{l'})}{\{\sigma(\mathbf{x}, z)\sigma(\mathbf{x}', z')\}} & z \neq z', \end{cases} \tag{4.3}$$

where $s(\mathbf{x}, z) = p\lim_{n\to\infty} s_n(\mathbf{x}, z)$ *and*

$$s_n(\mathbf{x}, z) = \frac{\hat{\Psi}(z)^{1/2}(\mathbf{x}^T, 1)^T}{\sqrt{(\mathbf{x}^T, 1)\hat{\Psi}(z)(\mathbf{x}^T, 1)^T}} = \frac{\hat{\Psi}(z)^{1/2}(\mathbf{x}^T, 1)^T}{\|\hat{\Psi}(z)^{1/2}(\mathbf{x}^T, 1)^T\|}. \tag{4.4}$$

4.1. Method 1

Using an argument similar to that for deriving the Scheffé type of simultaneous confidence interval for a linear model, we apply the Cauchy Schwartz inequality to obtain

$$\begin{aligned} W_n(\mathbf{x}, z)^2 &= \frac{nh(\mathbf{x}^T\hat{\boldsymbol{\beta}} + \hat{\gamma}(z) - \mathbf{x}^T\boldsymbol{\beta} - \gamma(z) - b(\mathbf{x}, z))^2}{(\mathbf{x}^T, 1)\hat{\Psi}(z)(\mathbf{x}^T, 1)^T} \\ &\le nh\{(\hat{\boldsymbol{\beta}}^T, \hat{\gamma}(z)) - E(\hat{\boldsymbol{\beta}}^T, \hat{\gamma}(z))\}\hat{\Psi}(z)^{-1}\{(\hat{\boldsymbol{\beta}}^T, \hat{\gamma}(z)) - E(\hat{\boldsymbol{\beta}}^T, \hat{\gamma}(z))\}^T. \end{aligned} \tag{4.5}$$

The inequality follows since $(\mathbf{x}^T, 1)E(\hat{\boldsymbol{\beta}}^T, \hat{\gamma}(z))^T = \mathbf{x}^T\boldsymbol{\beta} + \gamma(z) + b(\mathbf{x}, z)$. The right-hand-side of (4.5) is asymptotically distributed as a Chi-squared variable with $k + 1$ degrees of freedom. Since this upper bound is distributed free of $\mathbf{x}$ and z, we can construct a Scheffé type simultaneous confidence interval for $\eta(\mathbf{x}, z) = \mathbf{x}^T\boldsymbol{\beta} + \gamma(z)$ at any $(\mathbf{x}, z) \in \mathcal{X} \times \mathcal{Z}$ as

$$\mathbf{x}^T\hat{\boldsymbol{\beta}} + \hat{\gamma}(z) - b(\mathbf{x}, z) \pm \sqrt{\chi^2_{k+1}(\alpha)}\hat{\sigma}(\mathbf{x}, z), \tag{4.6}$$

where $\chi^2_{k+1}(\alpha)$ is the upper α quantile of the Chi-squared distribution with $k+1$ degrees of freedom.

In finite sample simulations we find that the coverage rate of this interval is usually lower than the desired confidence level even if we use a more accurate covariance estimator $\hat{\Psi}(z)^*$ described in the next subsection. This is because the sampling distribution of the supremum of $W_n(\mathbf{x}, z)^2$ is different from the Chi-squared distribution (see Figure 1); the Chi-squred distribution tends to give a smaller upper α percentile values. The interval thus appears to be narrower than it should be and, when nh is far from ∞, the approximation made in (4.5) may not be accurate enough. The improvement is also quite slow as we enlarge the sample size. We therefore do not recommend the use of this approach in the analysis; a more practical method is introduced in the next subsection.

4.2. Method 2

We now propose a Monte Carlo (MC) approach to calculate the simultaneous confidence interval. First we use a bootstrap re-sampling method to compute the empirical covariance of the semiparametric estimators. Then a large number of simulations are performed to approximate the asymptotic distribution of

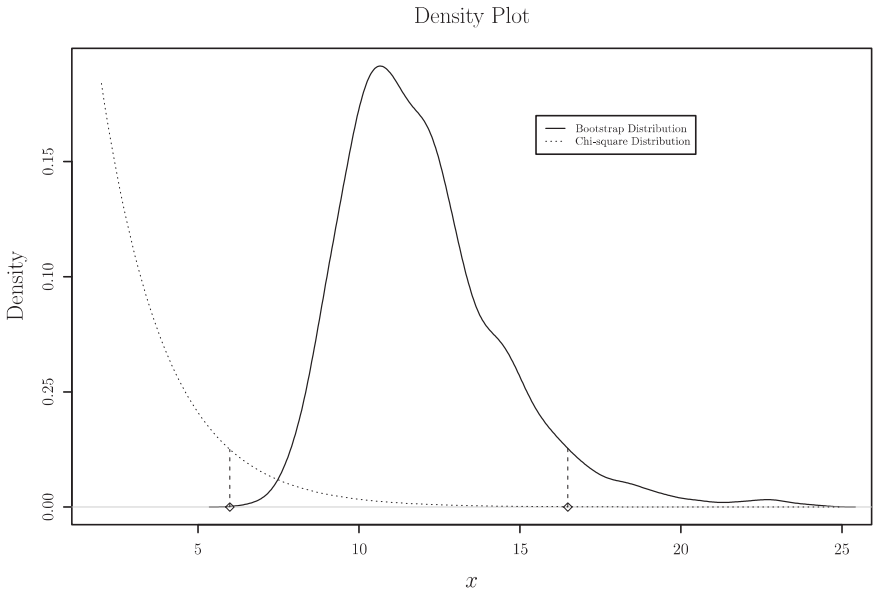

Figure 1. The Bootstrap Distribution (solid line) and Chi-square Distribution (dotted line). The upper 0.05 quantiles of these two distributions are marked at broken lines.

$\sup_{\mathbf{x},z} W_n(\mathbf{x}, z)^2$. The upper α quantile from the generated MC samples serves as the critical values c_α for the construction of simultaneous confidence bands.

We call this method a Bootstrap method in this paper since in some sense it resembles the popular double-bootstrap procedure (Lee and Young (1999), Hall, Lee and Young (2000)) in that two separate nested random simulation steps are carried out. However, we are not re-sampling from the data but from a fitted parametric distribution constructed by the first level bootstrap at the second simulation stage. The theoretical properties of bootstrap sampling are well preserved for the results obtained at the first step and the distributional properties for the results obtained at the second step are then effectively employed.

Here is an outline of our proposed algorithm for computing the bootstrap critical value.

1. Draw B bootstrap samples $\Omega_b = \{(X_{1i}^{(b)}, \ldots, X_{ki}^{(b)}, Z_i^{(b)}, Y_i^{(b)}), i = 1, \ldots, n\}$ $(b = 1, \ldots, B)$ from the original sample.
2. For each bootstrap sample, fit the semiparametric model (1.1) and obtain $\hat{\boldsymbol{\beta}}^{(b)}$ and $\hat{\gamma}(z_i)^{(b)}$ for all distinct values of z_i in such a sample.
3. Calculate the empirical bootstrap covariance matrix

$$\hat{\Psi}^*(z_i) = \begin{pmatrix} \text{Cov}^*(\hat{\boldsymbol{\beta}}^{(b)}) & \text{Cov}^*(\hat{\boldsymbol{\beta}}^{(b)}, \hat{\gamma}(z_i)^{(b)}) \\ \text{Cov}^*(\hat{\boldsymbol{\beta}}^{(b)}, \hat{\gamma}(z_i)^{(b)})^T & \text{Var}^*(\hat{\gamma}(z_i)^{(b)}) \end{pmatrix}$$

for each distinct z_i.

4. Draw M independent random $(k+1)$-vectors $\{\zeta_m(z_i), m = 1, \ldots, M\}$ from the multivariate normal distribution with mean zero and covariance $\hat{\Psi}^*(z_i)$ for each z_i.
5. Evaluate $(\mathbf{x}_i^T, 1)\zeta_m(z_i)/\hat{\sigma}^*(\mathbf{x}_i, z_i)$, where $\hat{\sigma}^*(\mathbf{x}_i, z_i) = \sqrt{(\mathbf{x}_i^T, 1)\hat{\Psi}^*(z_i)(\mathbf{x}_i^T, 1)^T}$ and find $w_m = \max_{\mathbf{x}_i, z_i} \left[\{(\mathbf{x}_i^T, 1)\zeta_m(z_i)/\hat{\sigma}^*(\mathbf{x}_i, z_i)\}\right]^2$ for $m = 1, \ldots, M$.
6. Take the upper α quantile of $\{w_m : m = 1, \ldots, M\}$ to be the critical value c_α.

The simulated sample $\{w_m\}$ displays a distribution distinct from the χ^2 distribution. Figure 1 gives an example, based on Case II as described in Section 5, with sample size $1,000$ and bandwidth 0.2. The solid line is obtained by using a kernel density estimator for $\{w_m\}$. The dotted line is the density of the χ^2 distribution with 2 degrees of freedom. For the purpose of comparison, we also mark the upper 0.05 quantiles for both distributions in the plot. The evident difference between these two distributions suggests that a non-central Chi-squared distribution with adjusted degrees of freedom might be a more appropriate approximation for the distribution of $\sup_{\mathbf{x},z} W_n(\mathbf{x}, z)^2$ in this case.

Since the empirical distribution of $\{w_m\}$ approximates the distribution of $\sup_{\mathbf{x},z} W(\mathbf{x}, z)^2$, by Theorem 4.1 we can construct a simultaneous confidence interval for $\eta(\mathbf{x}, z) = \mathbf{x}^T\boldsymbol{\beta} + \gamma(z)$ at any $(\mathbf{x}, z) \in \mathcal{X} \times \mathcal{Z}$ as

$$\mathbf{x}^T\hat{\boldsymbol{\beta}} + \hat{\gamma}(z) \pm \sqrt{\mathsf{c}_\alpha}\hat{\sigma}^*(\mathbf{x}, z). \tag{4.7}$$

5. Application to Multi-Dimensional Effective Dose

We can directly apply our results regarding the confidence intervals for the semiparametric regression function to construct confidence regions for the multi-dimensional effective dose.

Suppose in model (1.1) we are interested in the set of values of the predictor variables that yield the outcome $Y = 1$ with a given probability p. The set of values for $(\mathbf{X}, Z)$ that satisfy this condition is called a *multi-dimensional effective dose*, since such a quantity is conventionally called effective dose $100p$ in a bioassay problem where only a univariate predictor is included in the model (Finney (1978)). Both parametric (e.g., Finney (1978)) and nonparametric methods (e.g., Müller and Schmitt (1988)) have been extensively used to solve the univariate effective dose problem. Few studies have considered the setting with multiple predictors (Li et al. (2008a,b)). For a given $p \in (0, 1)$, we define the multi-dimensional effective dose under the semiparametric regression model as the set

$$\Theta_p = \Big\{(\mathbf{x}, z) : \log\big\{\frac{p}{1-p}\big\} = \mathbf{x}^T\boldsymbol{\beta} + \gamma(z)\Big\}. \tag{5.1}$$

With given values of $\boldsymbol{\beta}$ and given form of $\gamma(\cdot)$, we can determine the set Θ_p of multi-dimensional effective dose. For example, if $k = 1$, $\boldsymbol{\beta} = 5$, and $\gamma(u) = u^2$, then the set Θ_p is simply a parabolic curve in the two-dimensional (x, z) plane determined by the equation $5x + z^2 = \log p/(1-p)$ for a given p.

It is usually more convenient to deal with the $(\mathbf{x}^*_{-i}, z^*)-$conditioning effective dose where $\mathbf{x}^*_{-i} = (x^*_1, \ldots, x^*_{i-1}, x^*_{i+1}, \ldots, x^*_k)$ and z^* are fixed. The conditional dose level needed for x_i to make $Pr(Y = 1)$ equal to p is

$$\mathrm{x}_i(p)^* = \Big\{x_i : x_i = \big\{\log\frac{p}{1-p} - \sum_{j\neq i} x^*_j\beta_j - \gamma(z^*)\big\}(\beta_i)^{-1}\Big\}, \tag{5.2}$$

where $\mathbf{x}^*_{-i} = (x^*_1, \ldots, x^*_{i-1}, x^*_{i+1}, \ldots, x^*_k)$ and z^* are fixed. We note that

$$\Theta_p = \bigcup\Big\{(x^*_1, \ldots, x^*_{i-1}, x_i, x^*_{i+1}, \ldots, x^*_k, z^*) : x_i = \mathrm{x}_i(p)^*, (\mathbf{x}^*_{-i}, z^*) \in \mathcal{X}^{-i} \times \mathcal{Z}\Big\},$$

where $\mathcal{X}^{-i}$ is the set of all possible values of $\mathbf{x}_{-i}$. The union does not depend on i.

We estimate the sets in the union by fitting model (1.1) thereby obtaining $\hat{\boldsymbol{\beta}}$ and $\hat{\gamma}(\cdot)$, and then plugging in the estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\gamma}(\cdot)$ in (5.2) to obtain $\hat{\Theta}_p$ or $\hat{\mathrm{x}}_i(p)^*$. The estimators can be shown to be consistent by extending Theorem 3.1 and results in Li et al. (2008a).

The confidence region for Θ_p can be obtained by joining the confidence intervals for $\mathrm{x}_i(p)^*$ across all possible values of $(\mathbf{x}^*_{-i}, z^*)$. We note that only simultaneous confidence intervals for $\mathrm{x}_i(p)^*$ can achieve satisfactory confidence level for the entire set of Θ_p. Pointwise confidence intervals for $\mathrm{x}_i(p)^*$ may be useful when conditional inference or inference over a subregion is of interest. That is, for a particular case of interest we may know $(\mathbf{x}^*_{-i}, z^*)$ and if x_i is the dose level we can vary, we are interested in $\mathrm{x}_i(p)^*$. Hence we introduce both pointwise and simultaneous confidence intervals for $\mathrm{x}_i(p)^*$.

Without loss of generality, we consider finding a confidence interval for x_1 given $(\mathbf{x}^*_{-1}, z^*)$ for a fixed p in the following presentation. Basically the form of the resulting confidence region for Θ_p does not depend on the choice of the index i asymptotically. Even for a finite sample, only the delta method, to be introduced in Section 5.1, exhibits a difference due to the choice of i.

To facilitate notation, we partition the covariance matrix $\Psi(z)$ as

$$\begin{aligned}(nh)^{-1}\Psi(z) &= \begin{pmatrix} \mathrm{Var}\,(\hat{\beta}_1) & \mathrm{Cov}\,(\hat{\beta}_1, \hat{\boldsymbol{\beta}}_{-1})^T & \mathrm{Cov}\,(\hat{\beta}_1, \hat{\gamma}(z)) \\ \mathrm{Cov}\,(\hat{\beta}_1, \hat{\boldsymbol{\beta}}_{-1}) & \mathrm{Cov}\,(\hat{\boldsymbol{\beta}}_{-1}) & \mathrm{Cov}\,(\hat{\boldsymbol{\beta}}_{-1}, \hat{\gamma}(z)) \\ \mathrm{Cov}\,(\hat{\beta}_1, \hat{\gamma}(z)) & \mathrm{Cov}\,(\hat{\boldsymbol{\beta}}_{-1}, \hat{\gamma}(z))^T & \mathrm{Var}\,(\hat{\gamma}(z)) \end{pmatrix} \\ &= \begin{pmatrix} \psi_{11} & \boldsymbol{\psi}_{1,x} & \psi_{1,z} \\ \boldsymbol{\psi}^T_{1,x} & \boldsymbol{\psi}_{xx} & \boldsymbol{\psi}_{xz} \\ \psi_{1,z} & \boldsymbol{\psi}^T_{xz} & \psi_{zz} \end{pmatrix}. \end{aligned} \tag{5.3}$$

5.1. Pointwise confidence interval: Delta method

The pointwise confidence interval for $\mathrm{x}_1(p)^*$ can be obtained by the delta method (e.g., Bickel and Doksum (2007)). Since

$$\mathrm{x}_1(p)^* \equiv \Big\{ \log\{\frac{p}{(1-p)}\} - \sum_{j\neq 1} x_j^*\beta_j - \gamma(z^*)\Big\}(\beta_1)^{-1} \tag{5.4}$$

is a smooth function of $(\boldsymbol{\beta}, \gamma(z^*))$, it is straightforward to show that

$$\sqrt{nh}\hat{\mathrm{x}}_1(p)^* = \sqrt{nh}\Big\{ \log\Big\{\frac{p}{1-p}\Big\} - \sum_{j\neq 1} x_j^*\hat{\beta}_j - \hat{\gamma}(z^*)\Big\}(\hat{\beta}_1)^{-1} \tag{5.5}$$

is asymptotically normal with mean equal to $\sqrt{nh}\mathrm{x}_1(p)^* + \sqrt{nh}\boldsymbol{\lambda}^T\mathbf{b}^*$, where $\mathbf{b}^*$ consists of the first $k+1$ elements of $\mathbf{b}$ given in (A.16) in the Appendix, and variance equal to $\boldsymbol{\lambda}^T\Psi(z)\boldsymbol{\lambda}$, where

$$\boldsymbol{\lambda} = \Big(-\Big\{ \log\frac{p}{1-p} - \sum_{j\neq 1} x_j^*\beta_j - \gamma(z^*)\Big\}(\beta_1^2)^{-1}, -\frac{\mathbf{x}_{-1}^{*T}}{\beta_1}, -\frac{1}{\beta_1}\Big)^T. \tag{5.6}$$

The $100(1-\alpha)\%$ confidence interval for $\mathrm{x}_i(p)^*$ is

$$\hat{\mathrm{x}}_1(p)^* - \hat{\boldsymbol{\lambda}}^T\mathbf{b}^* \pm \mathsf{z}_{\alpha/2}\sqrt{\frac{\hat{\boldsymbol{\lambda}}^T\hat{\Psi}(z)\hat{\boldsymbol{\lambda}}}{nh}}, \tag{5.7}$$

where $\hat{\boldsymbol{\lambda}}$ is the estimator of $\boldsymbol{\lambda}$ obtained by substituting $\hat{\boldsymbol{\beta}}$ and $\hat{\gamma}(\cdot)$ for their estimands. The bias term $\hat{\boldsymbol{\lambda}}^T\mathbf{b}^*$ is not computed in practice since it is of a smaller order.

This yields a symmetric confidence interval around the point estimator of $\mathrm{x}_1(p)^*$, which is usually desirable in practice. However, the resulting confidence region for Θ_p is not invariant under a different choice of conditioning variables $(\mathbf{x}_{-i}^*, z^*)$ for $i \neq 1$. It can be shown that the difference due to the index choice vanishes when the sample size tends to infinity.

5.2. Pointwise confidence interval: Fieller method

A pointwise confidence interval for $\mathrm{x}_1(p)^*$ can be obtained from Fieller's theorem (Fieller (1954)). By Theorem 3.1, we have that

$$\mathsf{D} \equiv \frac{nh(x_1\hat{\beta}_1 + \sum_{j\neq 1} x_j^*\hat{\beta}_j + \hat{\gamma}(z^*) - \log(p/(1-p)) - \hat{b}(x_1, \mathbf{x}_{-1}^*, z^*))^2}{(x_1, \mathbf{x}_{-1}^{*T}, 1)\hat{\Psi}(z^*)(x_1, \mathbf{x}_{-1}^{*T}, 1)^T} \tag{5.8}$$

is asymptotically distributed as χ^2 with one degree of freedom. Re-arrange terms in the inequality $\mathsf{D} < \chi_1^2(\alpha)$ to yield

$$Ax_1^2 + Bx_1 + C < 0, \tag{5.9}$$

with

$$A=(\hat{\beta}_1+\hat{b}_1)^2-\chi_1^2(\alpha)\hat{\psi}_{11}, \tag{5.10}$$

$$B=2\big[(\hat{\beta}_1+\hat{b}_1)\{\log\frac{p}{1-p}+\hat{b}_2-\sum_{j\neq 1}x_j^*\hat{\beta}_j-\hat{\gamma}(z^*)\}-\chi_1^2(\alpha)(\mathbf{x}_{-1}^T\hat{\psi}_{1,x}+\hat{\psi}_{1,z})\big], \tag{5.11}$$

$$C=\Big\{\log\frac{p}{1-p}+\hat{b}_2-\sum_{j\neq 1}x_j^*\hat{\beta}_j-\hat{\gamma}(z^*)\Big\}^2-\chi_1^2(\alpha)(\mathbf{x}_{-1}^T\hat{\psi}_{xx}\mathbf{x}_{-1}+2\mathbf{x}_{-1}^T\hat{\psi}_{xz}+\hat{\psi}_{zz}), \tag{5.12}$$

where each $\hat{\psi}_{\cdot\cdot}$ is the consistent estimate of $\psi_{\cdot\cdot}$ in (5.3). $\hat{b}_1$ and $\hat{b}_2$ are smaller order bias terms which are generally excluded in practice. This inequality leads to a confidence interval for $\mathrm{x}_1(p)^*$ given $(\mathbf{x}_{-1}^*, z^*)$ as

$$\frac{-B\pm\sqrt{B^2-4AC}}{2A}, \tag{5.13}$$

provided that $A>0$ and $B^2-4AC>0$.

5.3. Simultaneous confidence interval: The Scheffé and bootstrap methods

We invert the simultaneous confidence region for the semi-parametric regression function obtained in Section 4 to get a simultaneous confidence interval for $\mathrm{x}_1(p)^*$ across all possible values of $(\mathbf{x}_{-1}^*, z^*)$. All three methods can be inverted as described below. In this paper, we consider only the Scheffé method (Method 1) and the Bootstrap method (Method 2).

We illustrate with the Bootstrap method (Method 2). For any $(x_1, \mathbf{x}_{-1}^{*T}, z^*)\in \mathcal{X}\times\mathcal{Z}$,

$$P\Big[\frac{nh\{x_1\hat{\beta}_1+\sum_{j\neq 1}x_j^*\hat{\beta}_j+\hat{\gamma}(z^*)-\log[p/(1-p)]-\hat{b}(x_1,\mathbf{x}_{-1}^*,z^*)\}^2}{(x_1,\mathbf{x}_{-1}^{*T},1)\hat{\Psi}(z^*)(x_1,\mathbf{x}_{-1}^{*T},1)^T}\le \mathsf{c}_\alpha\Big] = 1-\alpha, \tag{5.14}$$

where c_α is as in Section 4.2. We then arrange the inequality inside the above probability into a quadratic inequality for x_1 as (5.9), where A, B and C are as at (5.10), (5.11) and (5.12) except that we replace $\chi_1^2(\alpha)$ with c_α. The resulting simultaneous confidence interval for $\mathrm{x}_i(p)^*$ is given by (5.13) with the modified A, B and C.

We invert the simultaneous Scheffé confidence interval (Method 1) as above after replacing c_α everywhere with $\chi_{k+1}^2(\alpha)$.

Table 5.1. Coverage rates for 95% confidence intervals from 1,000 simulations for $p = 0.1$.

			Pointwise Confidence Intervals				Simultaneous Confidence Intervals			
			Delta Method		Fieller Method		Scheffé Method		Bootstrap Method	
Case	h	n	Coverage	Length	Coverage	Length	Coverage	Length	Coverage	Length
I	0.1	500	92.4	0.363	93.4	0.429	45.0	0.605	92.4	0.918
		1,000	93.4	0.252	94.0	0.274	49.3	0.359	98.8	0.552
	0.2	500	91.3	0.253	92.1	0.273	63.2	0.362	98.2	0.711
		1,000	93.8	0.178	94.2	0.187	66.5	0.239	99.5	0.417
	0.3	500	92.1	0.207	92.3	0.221	67.7	0.285	98.5	0.536
		1,000	92.4	0.146	92.6	0.149	67.1	0.190	98.9	0.374
II	0.1	500	94.2	0.425	96.8	0.506	66.5	0.717	95.8	0.729
		1,000	96.4	0.316	97.0	0.345	64.3	0.454	99.8	0.519
	0.2	500	95.8	0.319	96.2	0.349	73.6	0.458	99.5	0.594
		1,000	95.5	0.225	95.5	0.236	76.9	0.302	99.7	0.413
	0.3	500	94.3	0.261	95.5	0.277	80.0	0.358	99.5	0.535
		1,000	94.4	0.185	95.6	0.190	79.5	0.242	99.6	0.371
III	0.1	500	93.7	0.606	94.8	0.771	44.5	1.155	98.9	1.785
		1,000	95.0	0.456	95.5	0.515	48.0	0.686	98.9	0.712
	0.2	500	93.9	0.457	94.5	0.516	63.4	0.690	99.4	0.859
		1,000	95.3	0.333	95.3	0.354	72.4	0.457	99.8	0.588
	0.3	500	92.5	0.369	93.5	0.399	67.0	0.521	99.7	0.820
		1,000	93.1	0.270	93.7	0.281	69.8	0.358	99.8	0.529

6. Simulations

This section presents finite-sample simulation studies that assess the performance of the proposed methods for calculating confidence regions for multidimensional effective doses.

We conducted a numerical study with 1,000 simulations. For each simulation, we generated $n = 500$ or $1,000$ samples of X from a uniform distribution on $[0, 1]$ and we let Z be n equally spaced points on $[0, 1]$. The true regression functions were set to be the following three cases:

$$\begin{aligned} \text{Case I: } & \eta = 5X + 5Z^3 - 4, \\ \text{Case II: } & \eta = -5X + 2\sin(0.8\pi Z) + 1, \\ \text{Case III: } & \eta = 4X - 0.6\exp\{-(2Z - 2)\}. \end{aligned}$$

The binary response was then generated from the Bernoulli distribution with a success probability $p = 1/(1 + e^{-\eta})$. A semiparametric logistic regression model was fitted to the generated data. For the local likelihood (2.2), we employed the Epanechnikov kernel $K(u) = 0.75(1-u^2)_+$. We fixed the bandwidth h at $0.1, 0.2$ and 0.3. As was previously reported in Fan and Huang (2005), we noticed that semiparametric estimates are not sensitive to the choice of bandwidth.

We calculated the z^*-conditioning effective dose $\hat{\Theta}_p^*$ for $p = 0.1, 0.5$ and 0.9 and the associated 95% confidence intervals. We applied both the delta and Fieller methods to build the pointwise confidence intervals and applied both the

Table 5.2. Coverage rates for 95% confidence regions from 1,000 simulations for $p = 0.5$.

			Pointwise Confidence Intervals				Simultaneous Confidence Intervals			
			Delta Method		Fieller Method		Scheffé Method		Bootstrap Method	
Case	h	n	Coverage	Length	Coverage	Length	Coverage	Length	Coverage	Length
I	0.1	500	91.6	0.292	90.5	0.338	25.5	0.466	95.0	0.859
		1,000	91.9	0.207	91.3	0.222	35.1	0.290	98.6	0.507
	0.2	500	91.8	0.210	91.6	0.226	54.0	0.294	97.8	0.580
		1,000	92.2	0.147	92.4	0.154	58.2	0.196	99.6	0.351
	0.3	500	92.2	0.172	91.6	0.180	63.0	0.231	99.5	0.407
		1,000	92.7	0.119	92.8	0.126	68.5	0.152	99.8	0.297
II	0.1	500	90.8	0.229	89.7	0.257	19.6	0.348	98.4	0.649
		1,000	95.6	0.159	90.5	0.169	28.5	0.217	99.9	0.459
	0.2	500	85.7	0.160	84.5	0.169	30.2	0.218	99.8	0.467
		1,000	88.3	0.112	85.8	0.115	32.0	0.146	99.9	0.323
	0.3	500	85.5	0.132	85.0	0.137	39.2	0.175	99.7	0.369
		1,000	89.3	0.092	88.1	0.094	40.1	0.118	99.9	0.269
III	0.1	500	92.3	0.335	90.6	0.405	23.6	0.582	98.6	1.77
		1,000	93.1	0.232	91.8	0.254	33.4	0.335	99.8	0.579
	0.2	500	90.8	0.237	90.7	0.258	52.2	0.339	99.1	0.594
		1,000	91.9	0.170	91.8	0.178	54.6	0.227	99.9	0.416
	0.3	500	90.4	0.198	90.3	0.211	66.6	0.273	99.8	0.518
		1,000	91.5	0.138	91.2	0.142	67.9	0.181	99.9	0.344

Table 5.3. Coverage rates for 95% confidence regions from 1,000 simulations for $p = 0.9$.

			Pointwise Confidence Intervals				Simultaneous Confidence Intervals			
			Delta Method		Fieller Method		Scheffé Method		Bootstrap Method	
Case	h	n	Coverage	Length	Coverage	Length	Coverage	Length	Coverage	Length
I	0.1	500	93.8	0.504	94.6	0.608	41.9	0.865	95.4	1.021
		1,000	94.7	0.371	94.7	0.407	42.8	0.537	99.6	0.598
	0.2	500	91.3	0.360	92.5	0.394	56.2	0.518	98.4	0.792
		1,000	93.4	0.264	93.4	0.275	58.9	0.354	99.2	0.492
	0.3	500	90.5	0.304	90.7	0.322	64.1	0.416	99.8	0.723
		1,000	92.8	0.214	93.1	0.221	65.4	0.279	99.9	0.436
II	0.1	500	95.2	0.419	97.7	0.505	78.3	0.713	98.0	0.732
		1,000	97.2	0.295	98.2	0.322	88.6	0.424	99.8	0.517
	0.2	500	94.1	0.295	95.4	0.320	84.5	0.421	98.9	0.594
		1,000	96.7	0.208	94.5	0.249	79.8	0.321	99.7	0.413
	0.3	500	93.2	0.235	94.5	0.249	79.4	0.321	99.5	0.543
		1,000	94.6	0.167	94.9	0.172	85.0	0.218	99.8	0.371
III	0.1	500	94.2	0.530	95.9	0.677	62.5	1.018	98.5	1.810
		1,000	94.6	0.368	96.2	0.413	65.8	0.557	98.9	0.660
	0.2	500	94.3	0.377	95.2	0.425	70.9	0.569	99.0	0.818
		1,000	94.6	0.257	95.5	0.272	77.5	0.352	99.1	0.530
	0.3	500	93.8	0.311	92.2	0.335	72.0	0.288	96.2	0.760
		1,000	94.3	0.218	95.3	0.226	81.1	0.437	99.8	0.471

Scheffé and Bootstrap methods to build the simultaneous confidence intervals for Θ_p^*. Main results are summarized in Table 5.1, 5.2 and 5.3 for $p = 0.1, 0.5$ and 0.9, respectively.

The pointwise coverage rate for each $\mathrm{x}_i(p)^*$ at the n z^* values was recorded

for the pointwise confidence intervals. The average coverage rates over the n $\mathrm{x}_i(p)^*$ were then averaged over the 1,000 simulations. The resulting coverage rates were reported as "Coverage" under the "Pointwise Confidence Intervals" columns in the tables.

For the simultaneous confidence intervals, we checked if the n true z^*-conditioning effective doses were in the calculated intervals. Only if the n intervals cover all n points in one simulation did we report a correct simultaneous coverage. This simultaneous coverage rate, as discussed in the previous section, is the coverage rate for the multi-dimensional effective dose Θ_p. The proportion of correct simultaneous coverages in the 1,000 simulations was reported as "Coverage" under the "Simultaneous Confidence Intervals" columns in the tables.

We also studied the precision of the intervals reflected by their lengths. The median (over n points and over 1,000 simulations) lengths of the confidence intervals generated by the different methods were computed and reported as "Length".

The coverage rates for the two types of pointwise confidence intervals were very close to the nominal confidence level in general. As the sample size increased, both types of confidence interval at each z^* tended to capture the true $\mathrm{x}_i(p)^*$ more frequently. Increased sample size also decreased the length of the interval. We notice that the choice of bandwidth can affect the coverage and length of the interval even though the semiparametric estimates do not differ too much under different bandwidths. In all three tables, smaller bandwidth is associated with relatively higher coverage and wider length; on the other hand, larger bandwidth is associated with relatively lower coverage and shorter length. It seems that a reasonable balance between accuracy and precision still needs to be achieved by selecting an appropriate h. Nevertheless, the differences in all cases are rather small as only small h's are considered.

For the simultaneous confidence intervals, we notice that Scheffé method had poor coverage rates. Although the coverage rates were observed to increase as sample size increased, it does not seem to be a suitable method for calculating the simultaneous confidence intervals in a reasonable-size sample. The bootstrap method had much higher coverage rates, almost always beyond the nominal confidence level. This conservative feature is very similar to what is found for most simultaneous confidence intervals employed for linear regression problems. We thus recommend the use of the bootstrap method to construct conservative confidence intervals in practice. The differences of coverage over different choices of bandwidth were not very large for the bootstrap method when bandwidths were small. However, as Figure 2 suggests, if a large value ($h = 0.8$) bandwidth is used in the fitting procedure, the semiparametric confidence region approaches a

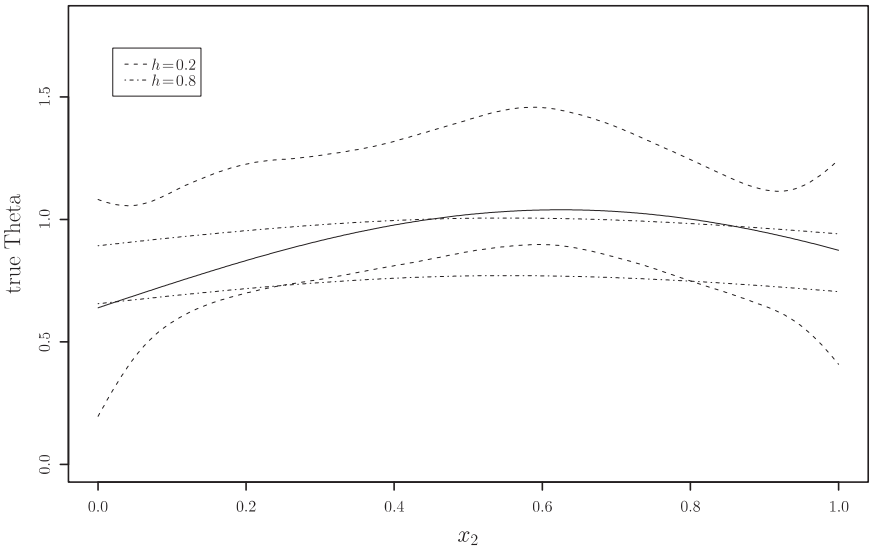

Figure 2. Confidence regions for 2-dimensional effective dose under Case II: The solid line is the true 2-dimensional effective dose; the dashed lines are the confidence limits when $h = 0.2$; the dash-and-point lines are the confidence limits when $h = 0.8$.

"parametric" type of region and thus fails to capture the true curve. From our extensive simulation results (not shown), we found that the simultaneous coverage rates of the bootstrap method can drop to only 50% when the bandwidth is set at such large values. Therefore, we recommend smaller bandwidths to obtain satisfactory confidence levels.

7. Example

In this section we re-visit the sheep example in Li et al. (2008a,b) to illustrate our methods of estimating and constructing confidence regions for the multi-dimensional effective dose. In a recent study at the University of Wisconsin, Madison, sheep were used as experimental subjects to determine whether the decompression sickness (DCS) disease might occur in humans undergoing similar dive profiles. Sheep offer an approximate animal model for DCS with susceptibility quite similar to humans since the body mass of sheep is similar to that of humans. Data from 1,108 observations were collected. One important research goal is to find joint values of risk factors that correspond to a fixed probability of developing DCS.

The major risk factors investigated in this study include exposure pressure (depth) and exposure duration. Each sheep underwent simulated dives (in a pressure chamber) with a designed pressure and duration, and its outcome for central

nervous system DCS (CNS-DCS), limb bends, respiratory DCS, and mortality was determined thereafter. The pressure was measured in absolute atmospheres and duration at depth was measured in minutes. In the following analysis, we took log base 10 transformations for both predictor variables. All observed outcomes were coded as dichotomous variables.

A nonparametric additive model (Hastie and Tibshirani (1990)) can be employed to roughly examine such dependence relationships by modeling the individual effect of each predictor variable as a smooth function without any specific parametric form. Zhang, Li and Meng (2008) recently found that under very general assumptions of independence or joint normality of the predictor variables, the estimated functional form in a nonparametric additive model can reflect the correct underlying dependence relationship between the response variable and each predictor. The fitted results for each component of the predictor variables can provide visually informative insights on the suitability of a parametric (linear) formulation for such a component. A more formal approach to assess the adequacy of a parametric model needs to be developed. A possible approach might be the adaption of the goodness-of-fit test proposed in Fan, Zhang and Zhang (2001).

We used nonparametric additive models to fit the four types of outcomes, respectively, and we examined the effects of pressure and duration. It was observed that the individual effects of pressure and duration were close to parametric linear functions for predicting respiratory DCS and mortality outcomes. Therefore, using a logistic regression model is sufficient to fit either of these outcomes. For the other two responses, CNS-DCS and limb bends, the effects of covariates differed remarkably from linear functions. Here we report on limb bends; the outcome of CNS-DCS can be analyzed similarly. The individual effects for fitting the limb bends in a nonparametric additive model are shown in Figure 3. The effect of $\log_{10}$(duration) is relatively close to a straight line over most of its support while that of $\log_{10}$(pressure) is clearly nonlinear. We thus chose $\log_{10}$(duration) as X (parametric term) and $\log_{10}$(pressure) as Z (nonparametric term) in fitting the response Y limb bends in a semiparametric logistic regression model.

The primary goal of this study was to determine the range of pressures and durations that correspond to certain risks of incurring the limb bends response. We fit the semiparametric model (1.1) and then used the methods we described in Section 4 to construct the pointwise and simultaneous confidence intervals for the two-dimensional effective dose Θ_p. The estimated $\hat{\Theta}_p$ for $p = 0.1$ and its associated confidence regions are shown in Figure 4.

The estimates of the 2-dimensional effective dose corresponding to a probability 0.1 of developing limb bends follow a curve in the ($\log_{10}$(pressure), $\log_{10}$(dur-

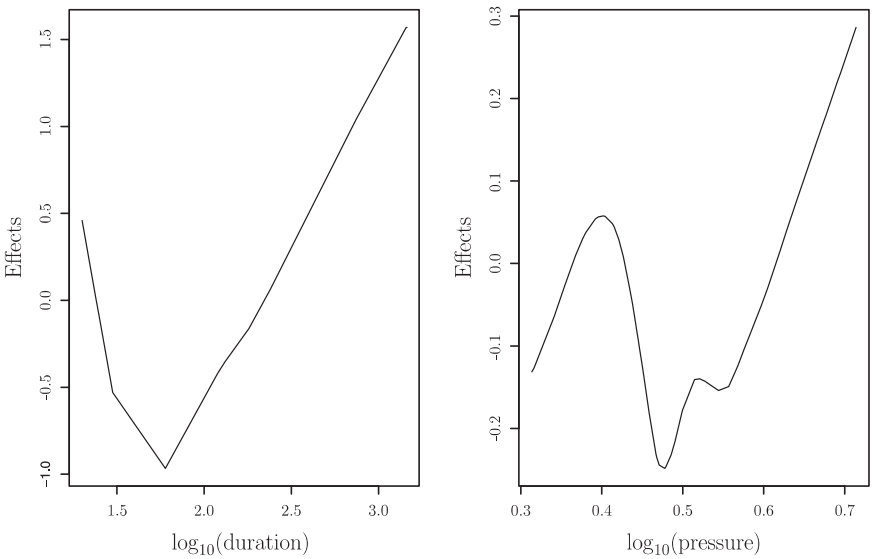

Figure 3. The Nonparametric Additive Estimates of Effects of Covariates for Limb Bends.

ation)) space. Generally a low value of pressure was associated with a high value of duration, and vice versa. However, for the moderate values of pressure, the associated duration values display a concave changing pattern: $\log_{10}$(duration) slightly increased for $\log_{10}$(pressure) when $\log_{10}$(pressure) was less than approximately 0.53, and then decreased rapidly.

The pointwise confidence interval ensures that the probability that the true value of $\log_{10}$(duration) at a given $\log_{10}$(pressure) is inside the interval is 95%. The simultaneous confidence interval ensures that the probability that all the true 2-dimensional effective dose values are in the region is 95%. The simultaneous confidence interval obtained from the bootstrap method is much wider than all pointwise confidence intervals.

The parametric logistic regression fit in this case gives a quite different answer for estimating the 2-dimensional effective dose. The misspecification of model forms can lead to an incorrect solution of the practical problem; it is more appropriate to use the semiparametric model in this case.

8. Discussion

The methodology proposed in this paper can be easily generalized to obtain pointwise and simultaneous confidence regions for any functional transformation of link functions among which the multi-dimensional effective dose is a special

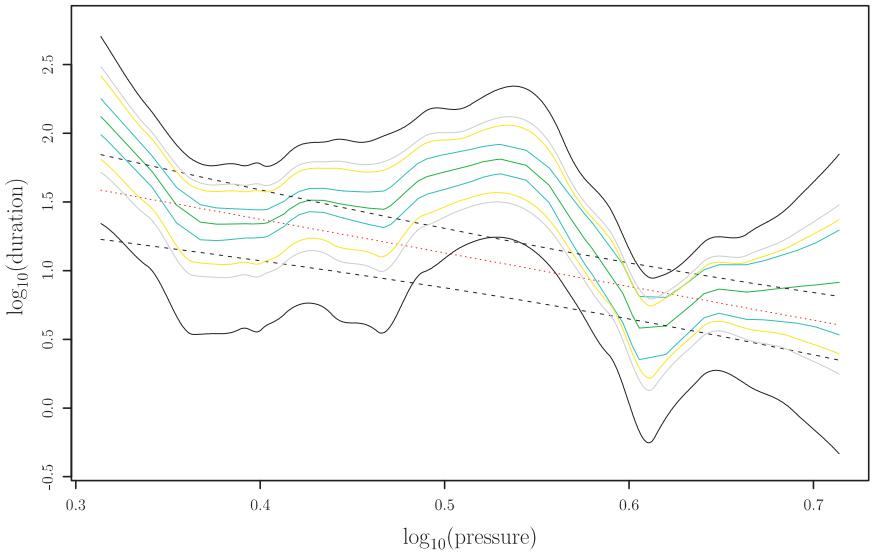

Figure 4. The Estimated 2D Effective Dose and Confidence Regions: The solid line with circles is the estimated 2D effective dose; the solid lines with triangles are the upper and lower bounds of the pointiwise confidence intervals based on the delta method; the dashed lines are the upper and lower bounds of the pointiwise confidence intervals based on the Fieller method; the gray solid lines are the upper and lower bounds of the simultaneous confidence intervals based on the Scheffé method; the black solid lines are the upper and lower bounds of the simultaneous confidence intervals based on the Bootstrap method; the dotted lines are parametric estimator and confidence regions for 2D effective dose.

case. We can also generalize our methods to responses being other members of the exponential family, such as Gaussian or Poisson.

We realize that the immediate application of multi-dimensional effective dose may not be in drug development but in some large observational studies. In fact, in drug development, dose-selection is done in Phase II clinical trials when sample sizes are relatively limited. To evaluate potential efficacy of a new agent, researchers usually choose to include small numbers of predictors. A larger study is usually followed in a Phase III trials where more sophisticated models like what we study in this paper may be developed.

The bandwidth involved in the semi-parametric fitting may be selected with the usual cross-validation methods. Such a data-driven method is usually computationally intensive. We might consider related methods such as the empirical cross-validation methods in Zhang (2003, 2008) that are asymptotically as good as the cross-validation methods. Further research which embeds such a strategy into the estimation of multi-dimensional effective dose is under development.

Acknowledgements

We are grateful to the Editor, an associate editor and a referee for their careful reading, and for some helpful comments and suggestions that have greatly improved our original submission. Li's research was supported in part by Singapore Academic Research Fundings R-155-050-066-133, R-155-050-066-101 and a National Medical Research Council grant. Zhang's research was supported in part by NSF Grants DMS-03-53941, DMS-07-05209 and Wisconsin Alumni Research Foundation.

Appendix

Proof of Theorem 3.1 Define $\mu_j = \int t^j K(t)dt$ and $\nu_j = \int t^j K^2(t)dt$.

For $l = 1, \ldots, L$, the estimators $(\hat{\boldsymbol{\beta}}, \hat{\gamma}(\cdot))$ in $\mathbf{x}_l^T\hat{\boldsymbol{\beta}} + \hat{\gamma}(z_l)$ are obtained by maximizing the local log-likelihood (2.2) in the neighborhood of z_l. With $\theta_l = (\boldsymbol{\beta}^T, a_l, b_l)^T$, the local log-likelihood of θ_l is

$$\mathcal{L}(\theta_l) = \sum_{i=1}^{n} [Y_i \log(p_{il}) + (1 - Y_i) \log(1 - p_{il})] K_h(Z_i - z_l), \tag{A.1}$$

where

$$p_{il} = \frac{1}{1 + \exp\{-(\mathbf{X}_i^T\boldsymbol{\beta} + a_l + b_l(Z_i - z_l))\}}. \tag{A.2}$$

Let $c_n = (nh)^{-1/2}$, $\mathbf{X}_{il}^* = (\mathbf{X}_i^T, 1, (Z_i - z_l))^T$, and $\hat{\theta}_l = (\hat{\boldsymbol{\beta}}^T, \hat{a}_l, \hat{b}_l)^T$. By a Taylor series expansion of $\mathcal{L}$ we obtain that

$$\mathcal{L}(\hat{\theta}_l) - \mathcal{L}(\theta_l) = \bar{\mathbf{W}}_l^T(\hat{\theta}_l - \theta_l) + \frac{1}{2}(\hat{\theta}_l - \theta)^T \bar{\mathbf{A}}_l(\hat{\theta}_l - \theta)\{1 + o_P(1)\}, \tag{A.3}$$

$$n^{-1}\bar{\mathbf{W}}_l = n^{-1}\sum_{i=1}^{n} \frac{(Y_i - p_{il})\mathbf{X}_{il}^*}{p_{il}(1 - p_{il})} K_h(Z_i - z_l), \tag{A.4}$$

$$n^{-1}\bar{\mathbf{A}}_l = n^{-1}\sum_{i=1}^{n} \frac{(Y_i(2p_{il} - 1) + p_{il}^2)\mathbf{X}_{il}^*\mathbf{X}_{il}^{*T}}{p_{il}^2(1 - p_{il})^2} K_h(Z_i - z_l). \tag{A.5}$$

It can be shown that

$$n^{-1}\bar{\mathbf{A}}_l = -f(z_l)E\left[\frac{e^{-\mathbf{X}^T\boldsymbol{\beta} - \gamma(Z)}}{(1 + e^{-\mathbf{X}^T\boldsymbol{\beta} - \gamma(Z)})^2} \begin{pmatrix} \mathbf{X}\mathbf{X}^T & \mathbf{X} & 0 \\ \mathbf{X}^T & 1 & 0 \\ 0 & 0 & \mu_2 h^2 \end{pmatrix} \middle| Z = z_l \right] + o_P(1) \tag{A.6}$$

$$= -\mathbf{A}_l + o_P(1). \tag{A.7}$$

Therefore, by (A.10) and the convexity lemma (Pollard (1991)), we obtain that $\hat{\theta}_l - \theta_l = \mathbf{A}_l^{-1}n^{-1}\bar{\mathbf{W}}_l + o_P(1)$. Now consider the concatenated vector $\boldsymbol{\theta} =$

$(\theta_1^T, \ldots, \theta_L^T)^T$ and its estimator $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1^T, \ldots, \hat{\theta}_L^T)^T$:

$$\sqrt{nh}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \sqrt{nh} \begin{pmatrix} \hat{\theta}_1 - \theta_1 \\ \vdots \\ \hat{\theta}_L - \theta_L \end{pmatrix} \tag{A.8}$$

$$= n^{-1}\sqrt{nh} \begin{pmatrix} \mathbf{A}_1^{-1} & 0 & \cdots & 0 \\ 0 & \mathbf{A}_2^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{A}_L^{-1} \end{pmatrix} \begin{pmatrix} \bar{\mathbf{W}}_1 \\ \bar{\mathbf{W}}_2 \\ \vdots \\ \bar{\mathbf{W}}_L \end{pmatrix} + \begin{pmatrix} o_P(1) \\ o_P(1) \\ \vdots \\ o_P(1) \end{pmatrix}. \tag{A.9}$$

It is easy to verify that Liapounov's condition for the Central Limit Theorem holds for the random vector $\mathbf{W} = n^{-1}\sqrt{nh}(\bar{\mathbf{W}}_1^T, \ldots, \bar{\mathbf{W}}_L^T)^T$, which therefore is asymptotically normal with mean $E(\mathbf{W}) = n^{-1}\sqrt{nh}(E(\bar{\mathbf{W}}_1)^T, E(\bar{\mathbf{W}}_2)^T, \ldots, E(\bar{\mathbf{W}}_L)^T)^T$, where

$$E(n^{-1}\sqrt{nh}\bar{\mathbf{W}}_l) = \frac{1}{2}h^2 f(z_l)\gamma''(z_l)E[\frac{e^{-\mathbf{X}^T\boldsymbol{\beta}-\gamma(Z)}}{(1+e^{-\mathbf{X}^T\boldsymbol{\beta}-\gamma(Z)})^2}(\mu_2\mathbf{X}^T, \mu_2, 0)^T | Z = z_l]$$

$$+o(h^2) \tag{A.10}$$

$$= \mathbf{b}_l + o(h^2), \tag{A.11}$$

and covariance matrix $cov(\mathbf{W})$. Here the lth diagonal block is

$$\mathrm{Var}\,(n^{-1}\sqrt{nh}\bar{\mathbf{W}}_l) = f(z_l)E\left[\frac{e^{-\mathbf{X}^T\boldsymbol{\beta}-\gamma(Z)}}{(1+e^{-\mathbf{X}^T\boldsymbol{\beta}-\gamma(Z)})^2}\begin{pmatrix} \nu_0\mathbf{X}\mathbf{X}^T & \nu_0\mathbf{X} & 0 \\ \nu_0\mathbf{X}^T & \nu_0 & 0 \\ 0 & 0 & \nu_2h^2 \end{pmatrix} \middle| Z = z_l\right]$$

$$+o(1) \tag{A.12}$$

$$= \mathbf{B}_l + o(1), \tag{A.13}$$

and the (l, l')th off-diagonal block $\mathrm{Cov}\,(\sqrt{nh}(\bar{\mathbf{W}}_l/n, \sqrt{nh}(\bar{\mathbf{W}}_{l'}/n))$ is

$$E\left[\frac{e^{-\mathbf{X}^T\boldsymbol{\beta}-\gamma(Z)}}{(1+e^{-\mathbf{X}^T\boldsymbol{\beta}-\gamma(Z)})^2}\begin{pmatrix} \mathbf{X}\mathbf{X}^T & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\right] \tag{A.14}$$

when $z_l \neq z_{l'}$ and $\mathbf{B}_l$ when $z_l = z_{l'}$.

The asymptotical normality of $\sqrt{nh}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ follows from (A.16) and its mean and covariance can be obtained through matrix multiplication.

Finally note that

$$\begin{pmatrix} \mathbf{x}_1^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})+(\hat{\gamma}(z_1)-\gamma(z_1)) \\ \mathbf{x}_2^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})+(\hat{\gamma}(z_2)-\gamma(z_2)) \\ \vdots \\ \mathbf{x}_L^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})+(\hat{\gamma}(z_L)-\gamma(z_L)) \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^T & 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{x}_2^T & 1 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & \mathbf{x}_L^T & 1 & 0 \end{pmatrix} \hat{\boldsymbol{\theta}}. \tag{A.15}$$

By multiplying matrices, we can find the lth bias term

$$\begin{aligned} b(\mathbf{x}_l, z_l) &= E\{\mathbf{x}_l^T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + (\hat{\gamma}(z_l) - \gamma(z_l))\} \\ &= (\mathbf{x}_l^T, 1, 0)\mathbf{A}_l^{-1}\mathbf{b}_l \\ &= \frac{1}{2}\mu_2\gamma''(z_l)(\mathbf{x}_l^T, 1)E\Big[\frac{e^{-\mathbf{X}^T\boldsymbol{\beta}-\gamma(Z)}}{(1+e^{-\mathbf{X}^T\boldsymbol{\beta}-\gamma(Z)})^2}\begin{pmatrix}\mathbf{X}\mathbf{X}^T & \mathbf{X}\\ \mathbf{X}^T & 1\end{pmatrix}\Big|Z = z_l\Big]^{-1} \\ &\quad E[\frac{e^{-\mathbf{X}^T\boldsymbol{\beta}-\gamma(Z)}}{(1+e^{-\mathbf{X}^T\boldsymbol{\beta}-\gamma(Z)})^2}(\mathbf{X}^T, 1)^T|Z = z_l]. \end{aligned} \tag{A.16}$$

Also note that the lth diagonal element of Σ is $(\mathbf{x}_l^T, 1)\Psi(z_l)(\mathbf{x}_l^T, 1)^T$, where

$$\begin{aligned} \Psi(z_l) &= \begin{pmatrix}\mathbf{I}_k & 0 & 0\\ 0 & 1 & 0\end{pmatrix}\mathbf{A_l^{-1}B_lA_l^{-1}}\begin{pmatrix}\mathbf{I}_k & 0\\ 0 & 1\\ 0 & 0\end{pmatrix} \\ &= \frac{\nu_0}{f(z_l)} = \Big\{E\Big[\frac{e^{-\mathbf{X}^T\boldsymbol{\beta}-\gamma(Z)}}{(1+e^{-\mathbf{X}^T\boldsymbol{\beta}-\gamma(Z)})^2}\begin{pmatrix}\mathbf{X}\mathbf{X}^T & \mathbf{X}\\ \mathbf{X}^T & 1\end{pmatrix}\Big|Z = z_l\Big]\Big\}^{-1}, \end{aligned} \tag{A.17}$$

and the (l, l')th off-diagonal elements of Σ is

$$\Upsilon(\mathbf{x}_l, \mathbf{x}_{l'}) = \mathbf{x}_l^T\Big\{E\Big[\frac{e^{-\mathbf{X}^T\boldsymbol{\beta}-\gamma(Z)}}{(1+e^{-\mathbf{X}^T\boldsymbol{\beta}-\gamma(Z)})^2}\mathbf{X}\mathbf{X}^T\Big]\Big\}^{-1}\mathbf{x}_{l'} \tag{A.18}$$

if $z_l \neq z_{l'}$ and $(\mathbf{x}_l^T, 1)\Psi(z_l)(\mathbf{x}_{l'}^T, 1)^T$ if $z_l = z_{l'}$.

Proof of Theorem 4.1. By the convergence of the finite-dimensional distributions in Theorem 3.1, we need only show asymptotic tightness of the process $W_n(\mathbf{x}, z)$, then the result follows from Theorem 18.14 in van der Vaart (1998).

For each positive ϵ and δ, we can find a partition of $\mathcal{T} = \mathcal{X} \times \mathcal{Z}$ into finitely many sets $\mathcal{T}_1, \ldots, \mathcal{T}_k$, such that for any $t_1 = (\mathbf{x}_1, z_1)$ and $t_2 = (\mathbf{x}_2, z_2) \in \mathcal{T}_i (i = 1, \ldots, k)$, $\|\mathbf{x}_1 - \mathbf{x}_2\| < \delta$ and $\|z_1 - z_2\| < \delta$, where $\|.\|$ is the Euclidean distance.

It is then straightforward to show

$$\limsup_{n\to\infty} P\big(\sup_i \sup_{t_1,t_2\in\mathcal{T}_i} |W_n(t_1) - W_n(t_2)| \geq \epsilon\big) \tag{A.19}$$

$$\begin{aligned} &\leq \limsup_{n\to\infty} E\{\frac{\sup_i \sup_{t_1,t_2\in\mathcal{T}_i} |W_n(t_1) - W_n(t_2)|}{\epsilon}\} \\ &\leq \limsup_{n\to\infty} \frac{\sup_i \sup_{t_1,t_2\in\mathcal{T}_i}\{C_{1n}\|\mathbf{x}_1 - \mathbf{x}_2\| + C_{2n}\|z_1 - z_2\|\}}{\epsilon} \\ &\leq C_3\delta. \end{aligned} \tag{A.20}$$

The first inequality follows from Tchebychef's inequality; C_{1n} and C_{2n} are finite constants which do not depend on t_1, t_2; the second inequality follows since $\boldsymbol{\beta}$ is in a compact space and $\gamma(\cdot)$ is twice continuously differentiable; C_3 is also a finite constant.

References

Bickel, P. J. and Doksum, K. A. (2007). *Mathematical Statistics: Basic Ideas and Selected Topics*. Pearson Prentice Hall; 2nd edition. Updated printing.

Carroll, R. J., Fan, J., Gijbels, I. and Wand, M. P. (1997). Generalized partially linear single-index models. *J. Amer. Statist. Assoc.* **92**, 477-489.

Fan, J. and Chen, J. (1999). One-step local quasi-likelihood estimation. *J. Roy. Statist. Soc. Ser. B* **61**, 927-943.

Fan, J., Heckman, N. E. and Wand, M. P. (1995). Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *J. Amer. Statist. Assoc.* **90**, 141-150.

Fan, J. and Huang, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli* **11**, 1031-1057.

Fan, J., Zhang, C. M. and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.* **29**, 153-193.

Fieller, E. C. (1954). Some problems in interval estimation. *J. Roy. Statist. Soc. Ser. B* **16**, 175-185.

Finney, D. J. (1978). *Statistical Method in Biological Assay*. Macmillan, New York.

Green, P. J. and Yandell, B. S. (1985). Semiparametric generalized linear models. In *Proceedings 2nd International GLIM Conference*. Lecture Notes in Statistics 32, 245-259. Springer-Verlag, New York.

Hall, P., Lee, S. M. S. and Young, G. A. (2000). Importance of interpolation when constructing double-bootstrap confidence intervals. *J. Roy. Statist. Soc. Ser. B* **62**, 479-491.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Model*. Chapman and Hall, London.

Hunsberger, S. (1994). Semiparametric regression in likelihood-based models. *J. Amer. Statist. Assoc.* **89**, 1354-1365.

Lee, S. M. S. and Young, G. A. (1999). The effect of Monte Carlo approximation on coverage error of double-bootstrap confidence intervals. *J. Roy. Statist. Soc. Ser. B* **61**, 353-366.

Li, J., Nordheim, E. V., Zhang, C. M. and Lehner, C. E. (2008a). Estimation and confidence regions for multi-dimensional effective dose. *Biometrical J.* **50**, 110-122.

Li, J., Zhang, C. M., Nordheim, E. V. and Lehner, C. E. (2008b). On the multivariate predictive distribution of multi-dimensional effective dose: a Bayesian approach. *J. Statist. Comput. Simulation* **78**, 429-442.

Mammen, E. and van de Geer, S. (1997). Penalized quasi-likelihood estimation in partial linear models. *Ann. Statist.* **25**, 1014-1035.

Müller, H. G. and Schmitt, T. (1988). Kernel and probit estimates in quantal bioassay. *J. Amer. Statist. Assoc.* **83**, 750-759.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory* **7**, 186-199.

Severini, T. A. and Staniswalis, J. G. (1994). Quasi-likelihood estimation in semiparametric models. *J. Amer. Statist. Assoc.* **89**, 501-511.

Staniswalis, J. G. (1989). On the kernel estimate of a regression function in likelihood-based models. *J. Amer. Statist. Assoc.* **84**, 276-283.

Sun, J. and Loader, C. R. (1994). Simultaneous confidence bands for linear regression and smoothing. *Ann. Statist.* **22**, 1328-1345.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

Zhang, C. M. (2003). Calibrating degrees of freedom for automatic data smoothing and effective curve checking. *J. Amer. Statist. Assoc.* **98**, 609-628.

Zhang, C. M. (2008). Prediction error estimation under Bregman divergence for non-parametric regression and classification. *Scand. J. Statist.* **35**, 496-523.

Zhang, C. M., Li, J. and Meng, J. (2008). On Stein's lemma, dependent covariates and functional monotonicity in high dimensional modeling. *Journal of Multivariate Analysis* **99**, 2285–2303.

Block 16, Level 7, 6 Science Drive 2, Singapore 117546. Department of Statistics & Applied Probability, National University of Singapore, Singapore 117546.

E-mail: stalj@nus.edu.sg

Department of Statistics, University of Wisconsin, Madison, WI 53706, U.S.A.

E-mail: cmzhang@stat.wisc.edu

Department of Statistics, University of Wisconsin, Madison, WI 53706, U.S.A.

E-mail: doksum@stat.wisc.edu

Department of Statistics, University of Wisconsin, Madison, WI 53706, U.S.A.

E-mail: nordheim@stat.wisc.edu

(Received April 2008; accepted January 2009)