Semiparametric Regression in Likelihood-Based Models

Sally HUNSBERGER*

A weighted likelihood is used to estimate the parameters in a semiparametric model involving two covariates and allowing an association between the covariates. The development is for arbitrary but specified densities of the observations. The estimators are consistent and asymptotically normal. Hypothesis testing of the parametric component can be performed using a Wald test. Simulations and analysis of data with Bernoulli observations demonstrate the estimators' application. Speckman developed kernel estimators where the conditional density of the observations is normal with p parametric covariates. Speckman's estimators and the new estimators are asymptotically equivalent, with the bias of Speckman's estimators being smaller. As an example, we study the relationship between a binary response indicating the occurrence of an intraoperative cardiac complication (ICC) in vascular surgery patients and two risk factors: duration of the operation (OR) and ASA score, which is an evaluation of the patient's overall health prior to surgery. ASA score is modeled in the parametric portion, because it appears valid to assume that ASA is linearly related to the logit of the probability of an ICC. The functional relationship between OR duration and the logit of the probability of an ICC is unknown, so it is modeled nonparametrically.

KEY WORDS: Kernel smoothers; Partial linear model; Weighted likelihood.

1. INTRODUCTION

In a regression model the objective is to describe the relationship between a response variable and a set of covariates. As a running example in this article, we study the relationship between a binary response indicating the occurrence of an intraoperative cardiac complication (ICC) in a vascular surgery patient and two risk factors. The risk factors of interest in this study are the duration of the operation recorded in minutes, (OR) duration, and an American Society of Anesthesiologist (ASA) level, which is an evaluation of the patient's overall health prior to surgery. The data were collected by the Department of Anesthesia at the Medical College of Virginia, Virginia Commonwealth University.

In describing the relationship between a response variable and a set of covariates, parametric or nonparametric models typically are used. Here a semiparametric model is used that includes both a parametric and a nonparametric component. The parametric portion includes covariates for which information concerning the functional form of the response–covariate relationship is available and thus used. The nonparametric portion includes covariates where less information is known concerning the functional form of the response–covariate relationship and thus the functional form is not specified. Combining these aspects is important, because if only partial information concerning the functional form of the response–covariate relationship is available, then a completely nonparametric model is inefficient and a completely parametric model may be wrong.

In the ICC example, the risk factor ASA is modeled in the parametric portion. It is assumed that ASA is linearly related to the logit of the probability of an ICC. This assumption is discussed in Section 4. The functional relationship between OR duration and the logit of the probability of an ICC is unknown, so it is modeled nonparametrically.

* Sally Hunsberger is Mathematical Statistician, Biostatistics Research Branch, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, MD 20892. This research was supported in part by National Science Foundation Grants DMS 8717560 and 9121555. The results in this article formed part of the author's Ph.D. thesis, written under the supervision of Joan G. Staniswalis at Virginia Commonwealth University. The author thanks Dr. Staniswalis for her guidance in that process.

Estimation in the semiparametric model for the case of normal random error has been studied using the method of penalized least squares or "partial smoothing splines," with relevant references including the work of Engle, Granger, Rice, and Weiss (1986) and Green, Jennison, and Seheult (1985). Heckman (1986) and Chen (1985) showed asymptotic normality for the parametric estimator and showed the bias to be asymptotically negligible when the covariates are assumed to be uncorrelated.

Rice (1986) showed that the bias can dominate the variance for the parametric estimator using the "partial smoothing spline" approach when the nonparametric portion and the parametric portion are correlated, unless the nonparametric portion is undersmoothed. Rice's results led to Speckman's (1988) consideration of the partial linear model with additive normal random error and allowing for an association between the covariates. Speckman used an intuitive approach and kernel smoothing to arrive at "ordinary least squares-like" estimators for the parameters in this model by regressing residuals on residuals. He showed asymptotic normality for the parametric estimator and found the bias to be asymptotically negligible. Heckman (1988) used a minimax approach, and Chen (1988) used a projection pursuit approach.

Hastie and Tibshirani (1986b) and Tibshirani and Hastie (1987) considered a generalized additive model which is any linear model in which some or all of the linear terms $\sum_j x_j\beta_j$ are replaced by the additive terms $\sum_j s_j(x_j)$, where the s_j are unspecified smooth functions. The asymptotic properties of these estimators are not known, although Buja, Hastie, and Tibshirani (1988) attempt to fill this gap.

Severini and Staniswalis (1994) used a quasi-likelihood function to estimate the parameters in the semiparametric model. This method of estimation only requires specification of the second-moment properties of the data, rather than specification of the entire distribution.

Journal of the American Statistical Association
December 1994, Vol. 89, No. 428, Theory and Methods

A weighted likelihood (Staniswalis 1989), sometimes termed a local likelihood (Hastie 1986a), is used to show that there exists a sequence of consistent estimators for the parametric and nonparametric components of the semiparametric regression model for arbitrary but specified densities of the observations in Section 2.1. Asymptotic normality and consistency for these estimators is established in Section 2.2. Section 2.3 develops the estimators for the case of Bernoulli random error, and Section 3 presents simulation results for observations with Bernoulli random error. Section 4 uses the estimators in the analysis of a data set.

2. WEIGHTED LIKELIHOOD ESTIMATORS

2.1 General Case

Consider the model with $Y_i \mid (X = x_i, T = t_i)$ having distribution $f(Y_i; \lambda_i)$. Let the parameter $\lambda_i = x_i\beta_0 + g(t_i)$ and let $f(\cdot)$, the conditional density of $Y \mid (X, T)$, be arbitrary but known. Then $x\beta_0$ is the parametric portion, with β_0 being the unknown parameter to be estimated that relates the covariate x to the response. Here g is the nonparametric portion of the model, with the only assumption on g that it be a smooth function of t with $\nu \geq 2$ continuous derivatives. Several assumptions are made that allow an association between x and t (Rice 1986; Speckman 1988). Assume the regression model $x_i = r(t_i) + \eta_i$ where $r(t)$ is a smooth function with ν continuous derivatives and η_i are independent random error terms with $E[\eta_i] = 0$ and $E[\eta_i^2] = \sigma^2$. Now λ_i can be rewritten using the model for the x's to obtain $\lambda_i = \eta_i\beta_0 + h(t_i)$, where $h(t_i) = r(t_i)\beta_0 + g(t_i)$ is the portion that depends on t. The main result of this research is to estimate β_0 and $h_i = h(t_i)$ $(i = 1, \ldots, n)$ in the semiparametric model by maximizing the weighted likelihood function

$$\mathrm{WL}(\beta, \boldsymbol{\theta}) = \sum_{i=1} \sum_{j=1} w\left(\frac{t_i - t_j}{b}\right) \log f(Y_j; \beta, \theta_i)/n^2 b$$
$$= \sum_i \mathrm{WL}(\beta, \theta_i)/n,$$

with respect to β and $\boldsymbol{\theta}$, where $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_n]'$. Throughout this article the sum is assumed to be from 1 to n. Here θ_i is used to indicate the function of the parameters of h_i. $\mathrm{WL}(\beta, \boldsymbol{\theta})$ depends on the unobserved η_i's, which can be estimated. In $\mathrm{WL}(\beta, \theta_i)$, $w(\cdot)$ is a kernel that assigns zero weights to the observations Y_j that correspond to t_j outside a neighborhood of t_i. The neighborhood is defined by the bandwidth b. The estimates of β_0 and $h_i = h(t_i)$ $(i = 1, \ldots, n)$ are found by choosing the $\hat{\beta}$ and $\hat{h}_i$ to simultaneously maximize $\mathrm{WL}(\beta, \boldsymbol{\theta})$ with respect to β and $\boldsymbol{\theta}$.

To understand the motivation for the weighted likelihood function, first examine $\mathrm{WL}(\beta, \theta_i)$. This can be seen as the portion estimating $h_i = h(t_i)$ using the approach of Staniswalis (1989). The kernel governs which observations are used to estimate $h_i = h(t_i)$. That is, because only the observations Y_j with t_j close to t_i have information about h_i, only the Y_j close to the t_i of interest are used to estimate h_i. The summation over i uses all of the individual $\mathrm{WL}(\beta, \theta_i)$ to estimate β_0, because all of the observations contain information about β_0.

Several assumptions that are used throughout are now given. The kernel is assumed to be a symmetric function with compact support on $[-1, 1]$. The smoothness of the function h governs the selection of the kernel. If the function is assumed to have ν continuous derivatives, then a kernel of order ν is used; that is,

$$\int_{-1}^{1} x^i w(x)\, dx = 1 \qquad \text{if } i = 0$$
$$= 0 \qquad \text{if } i < \nu$$
$$= w_\nu \neq 0 \qquad \text{if } i = \nu,$$

where w_ν is a constant.

The covariate t is assumed to be a design variable controlled by the experimenter with negligible error; for example, a dose of a chemical. For simplicity, it is assumed that $t_i = (i - 1)/n$, $i = 1, \ldots, n$ are equally spaced.

2.2 Asymptotic Properties

Asymptotic properties of the estimators and the sufficient conditions are now given.

Assume that

$$0 = E\left[\frac{\partial \log f(Y; \beta_0, h_j)}{\partial \beta_0} \middle| x_j, t_j\right],$$

$$0 = E\left[\frac{\partial \log f(Y; \beta_0, h_j)}{\partial h_j} \middle| x_j, t_j\right],$$

$$I_j[\beta_0] = E\left\{\left[\frac{\partial \log f(Y; \beta_0, h_j)}{\partial \beta_0}\right]^2 \middle| x_j, t_j\right\}$$
$$= -E\left[\frac{\partial^2 \log f(Y; \beta_0, h_j)}{\partial^2 \beta_0} \middle| x_j, t_j\right],$$

$$I[h_j] = E\left\{\left[\frac{\partial \log f(Y; \beta_0, h_j)}{\partial h_j}\right]^2 \middle| x_j, t_j\right\}$$
$$= -E\left[\frac{\partial^2 \log f(Y; \beta_0, h_j)}{\partial^2 h_j} \middle| x_j, t_j\right],$$

and

$$I[\beta_0, h_j] = -E\left[\frac{\partial^2 \log f(Y; \beta_0, h_j)}{\partial \beta_0 \partial h_j} \middle| x_j, t_j\right]$$
$$= -E\left[\frac{\partial^2 \log f(Y; \beta_0, h_j)}{\partial h_j \partial \beta_0} \middle| x_j, t_j\right].$$

Set

$$s_{n\beta_0}^2 = \frac{\Sigma_j I_j[\beta_0]}{n},$$

$$s_{nh_i}^2 = \frac{\Sigma_j w\left(\frac{t_i - t_j}{b}\right) I[h_j]}{nb}$$

and

$$s_{nc_i}^2 = \frac{\sum_j w\left(\frac{t_i - t_j}{b}\right) I[\beta_0, h_j]}{nb},$$

$$\tau_n^2 = \frac{\sum_j w^2\left(\frac{t_i - t_j}{b}\right) I[h_j]}{nb}.$$

Let f satisfy the following regularity conditions:

C1. The $\log f(y; \beta, h(t))$ has all continuous partial derivatives of order ν with respect to β and $h(t)$.

C2. There exist integrable $H_k(y; \beta, h(t))$ for $k = 1, \ldots, \nu$ and

$$E|H_k(Y, \beta, h(t))|^i < \infty \quad \text{for} \quad i = 1, 2, 3.$$

For all β, $h(t)$

$$\left|\frac{\partial^\nu}{\partial \beta^\nu} \log f(y; \beta, h(t))\right| \le H_\nu(y, \beta, h(t)),$$

$$\left|\frac{\partial^\nu}{\partial h_i^\nu} \log f(y; \beta, h(t))\right| \le H_\nu(y; \beta, h(t)).$$

C3. Let

$$U_1(Y; \beta, \theta) = \frac{\partial^2 \log f(Y; \beta, \theta)}{\partial \beta^2},$$

$$U_2(Y; \beta, \theta) = \frac{\partial^2 \log f(Y; \beta, \theta)}{\partial \theta^2},$$

and

$$U_3(Y; \beta, \theta) = \frac{\partial^2 \log f(Y; \beta, \theta)}{\partial \beta \partial \theta}.$$

There exists an $M > 0$ such that $|U_z(Y; \beta_i, \theta_i) - U_z(Y; \beta_j, \theta_j)| \le M|[(\beta_i - \beta_j)^2 + (\theta_i - \theta_j)^2]|^{\gamma/2}$ for $z = 1, 2, 3$ and for some $\gamma \in (0, 1)$.

C4. For $j = 1, \ldots, \nu$, the

$$E\left[\frac{\partial^j \log f(Y; \beta, \theta)}{\partial \theta^j}\bigg|_{\substack{\beta=\beta_0\\ \theta=h}} \bigg| t\right] \quad \text{and}$$

$$E\left[\frac{\partial^j \log f(Y; \beta, \theta)}{\partial \beta^j}\bigg|_{\substack{\beta=\beta_0\\ \theta=h}} \bigg| t\right]$$

must have ν continuous derivatives in t.

C5. The matrices

$$\begin{bmatrix} I_j[\beta_0] & I[\beta_0, h_j] \\ I[\beta_0, h_j] & I[h_j] \end{bmatrix} \quad \text{and} \quad \mathbf{s}_n = \begin{bmatrix} s_{n\beta}^2 & s_{nc_j}^2 \\ s_{nc_j}^2 & s_{nh_j}^2 \end{bmatrix}$$

are positive definite, and there exists M' such that for all j max $\lambda_{jn} < M'$, where λ_{jn} are the eigenvalues of the $\mathbf{s}_n$ matrix.

C6. Let n and b satisfy the following assumptions: $nb \to \infty$, $nb^2 \to \infty$, and $nb^{2\nu} \to 0$ as $n \to \infty$ and $b \to 0$.

The following theorem states the asymptotic properties for the sequence of estimators for the parameters in the semiparametric model.

Theorem 1. Under the foregoing regularity conditions and the assumptions on n and b,

a. (1) $\|(\hat\beta - \beta_0, \hat{\mathbf{h}} - \mathbf{h})\|^2 \to_p 0$ as $n \to \infty$, where $\|(\beta, \theta)\|^2 = \max[|\beta|^2, \int \theta^2(t)\, dt]$, and (2) $|\hat\beta_h - \hat\beta_{\hat h}| \to_p 0$ as $n \to \infty$. The notation $\hat\beta_{\hat h}$ denotes the estimate of β as a function of the estimates for $\mathbf{h}$ and $\hat\beta_h$ denotes the estimate of β as a function of the true values of $\mathbf{h}$.

b. $\sqrt{n} s_{n\beta_0}(\hat\beta_{\hat h} - \beta_0) \to_d N(0, 1)$.

The proof follows the method of maximum likelihood as provided by Cramér (1946) and Staniswalis (1989). The full version was presented by Hunsberger (1993); an outline is given in the Appendix.

In the proof it is shown that $\hat\beta$ has nonparametric convergence rate of $O_p(b^\nu)$ for the bias and parametric convergence rate of $O(n^{-1})$ for the variance. The nonparametric component $\hat g$ is found to have nonparametric convergence rates of $O_p(b^\nu)$ for the bias and $O_p((nb)^{-1})$ for the variance. This becomes important when the optimal bandwidth for nonparametric regression $b^* \sim n^{-1/(2\nu+1)}$ (Rice 1984) is used, because then the square of the bias dominates the variance for the parametric component. The results here are consistent with the results of Rice (1986). Severini and Wong (1992) used the concept of a least favorable curve to construct estimators of β for which the variance does dominate squared bias using b^*.

The convergence rates for the weighted likelihood estimators (WLE) are for arbitrary but specified densities of the observations and allow for an association between x and t. For the specific case of normal additive error, it is possible to compare the WLE's to Speckman's (1988) estimators. This comparison is of interest because Speckman (1988) found the bias of his estimator for β, denoted by β^s, to be $O_p(b^{2\nu})$ and found the variance to be $O_p(n^{-1})$. In this case when b^* is used, the bias is asymptotically negligible compared to the variance. Simulations (Hunsberger 1990) that compare the estimators for finite sample sizes show that the estimators are almost identical for x and t uncorrelated. This is consistent with the results of Rice (1986) in the context of partial smoothing splines.

2.3 Estimation for Bernoulli Density

In this section the estimators are developed for the specific setting of Bernoulli densities of the observations. These estimators are then used in the ICC example.

Let $p_i = P[Y_i = 1 | X_i = x_i, T_i = t_i]$ be the probability of success for the Bernoulli random variable Y given the risk factors X and T. Because p_i is a probability, it is reasonable to use the logistic function (Agresti 1984) to model the probability of response given the risk factors as

$$p_i = \frac{\exp(x_i\beta + g(t_i))}{1 + \exp(x_i\beta + g(t_i))}.$$

The log-likelihood function of the sample using the Bernoulli density is

$$l(\beta, \boldsymbol{\theta}) = \sum_i Y_i[\eta_i\beta + \theta_i] - \sum_i \log[1 + \exp(\eta_i\beta + \theta_i)].$$

The weighted log-likelihood becomes

$$\mathrm{WL}(\beta, \theta) = \sum_i \sum_j w\left(\frac{t_i - t_j}{b}\right) Y_j(\eta_j \beta + \theta_i)/n^2 b - \sum_i \sum_j w\left(\frac{t_i - t_j}{b}\right) \log[1 + \exp(\eta_j \beta + \theta_i)]/n^2 b.$$

The WLE's for h_i and β_0 are the $\hat{h}_i$ and $\hat{\beta}$ that simultaneously solve

$$0 = \left.\frac{\partial \mathrm{WL}(\beta, \theta)}{\partial \theta_i}\right|_{\substack{\beta=\hat{\beta}\\ \theta_i=\hat{h}_i}} = \sum_j \frac{w\left(\frac{t_i - t_j}{b}\right) Y_j}{nb} - \sum_j \frac{w\left(\frac{t_i - t_j}{b}\right)}{nb} \frac{\exp(\eta_j \hat{\beta} + \hat{h}_i)}{(1 + \exp(\eta_j \hat{\beta} + \hat{h}_i))}$$

and

$$0 = \left.\frac{\partial \mathrm{WL}(\beta, \theta)}{\partial \beta}\right|_{\substack{\beta=\hat{\beta}\\ \theta_i=\hat{h}_i}} = \sum_j \sum_i \frac{w\left(\frac{t_i - t_j}{b}\right) Y_j \eta_j}{n^2 b} - \sum_j \sum_i \frac{w\left(\frac{t_i - t_j}{b}\right)}{n^2 b} \frac{\eta_j \exp(\eta_j \hat{\beta} + \hat{h}_i)}{(1 + \exp(\eta_j \hat{\beta} + \hat{h}_i))}. \quad (1)$$

These equations cannot be solved explicitly. An estimator for h_i is found by making the following observations. First, rewrite the second term of Equation (1) as

$$\frac{\Sigma_j \, w\left(\frac{t_i - t_j}{b}\right)}{nb} [1 + \exp(-\eta_j \beta) \exp(-h_i)]^{-1}.$$

Now this converges in probability to $(1 + \exp(-h_i))^{-1}$ under the assumptions on the bandwidth C6 (Cheng and Lin 1981). Using this observation in (1) allows us to find an approximate solution for $\hat{h}_i$.

A Newton–Raphson algorithm is used to approximate β (Serfling 1980). The algorithm uses the first and second partial derivatives of $\mathrm{WL}(\beta, \theta)$ with respect to β evaluated at β^g, $\hat{h}$, and $\hat{\eta}$, where β^g is an initial guess for β. Now η is unobservable but can be estimated as follows: $\hat{\eta} = x - r(\hat{t})$, with $r(\hat{t})$ being the nonparametric kernel estimate of $r(t)$. In this work the Priestley–Chao estimator (Priestley and Chao 1972) is used. This is defined as

$$\hat{r}(t, b) = \sum_i w\left(\frac{t - t_i}{b}\right) X_i (nb)^{-1}.$$

3. SIMULATION RESULTS

Two Monte Carlo simulations are performed to evaluate the performance of the estimators for the case of Bernoulli random error and relatively small sample sizes of $n = 45$. The simulations examine the effect of ignoring the theoretical assumption of equal spacing of the t's, which is the case in the data set application. The programs were written in Fortran and executed on the Convex C240 Mini-supercomputer at the National Institutes of Health.

The Monte Carlo simulations consist of five blocks of 100 realizations with $n = 45$ observations in each realization. For the first simulation, an observation consists of the triplet (x_i, t_i, Y_i). Bernoulli random variables Y_i are generated, with probability p_i that depend on the fixed design variables x_i and t_i. In this simulation the t_i's are equally spaced as $t_i = i/16$ for $i = 1, \ldots, 15$. There are three observations at a t_i, with x_i taking on each of the values 0, 1, and 2. This is a balanced design with x and t uncorrelated, because there are the same number of observations in each x_i group. The equation

$$p_i = \frac{\exp(x_i \beta_0 + g(t_i))}{1 + \exp(x_i \beta_0 + g(t_i))}, \quad \text{where}$$

$$g(t_i) = 10 \exp\left[\frac{t_i - 5}{.1} - \exp\left(\frac{t_i - .5}{.1}\right)\right] - .5 \quad (2)$$

and $\beta_0 = 1$, was used. The International Mathematical Statistical Libraries (IMSL) function RNBIN was used to generate the independent Bernoulli random variables Y_i with probability of success p_i.

The second simulation differs from the first in that the design variable t is unequally spaced. Fifteen t values are generated for each level of x from a truncated normal distribution with mean zero and variance 1 using the IMSL function DRNNAO. Each sample has a different set of t values. The values are rescaled to be between (0, 1). In this simulation there is only one x value at each t.

The asymptotic variance of $\hat{\beta}$ for the first simulation is calculated from

$$\hat{\sigma}^2 = \frac{-\Sigma_i \, \hat{\eta}_i^2 \left[\left[\frac{\exp(x_i \beta_0 + g(t_i))}{1 + \exp(x_i \beta_0 + g(t_i))}\right]^2 - \frac{\exp(x_i \beta_0 + g(t_i))}{1 + \exp(x_i \beta_0 + g(t_i))}\right]}{n},$$

where $r(t_i) = 1$, $\hat{\eta}_i = x_i - r(t_i)$, $\beta_0 = 1$, and $g(t_i)$ is as in (2).

A kernel and a bandwidth must be chosen to use in the weighted likelihood. The spline kernel of Messer and Goldstein (1989) with $\nu = 2$ was used throughout. The bandwidth was selected as follows. First, a bandwidth to estimate the η's was selected by using the bandwidth that minimized $\sum_k [x_k - \hat{r}_k]^2$ with respect to the bandwidth. Here $\hat{r}_k$ is the kernel estimate of $r(t_k)$ calculated with (x_k, t_k) left out of the data set. The estimates $\hat{r}(t_i)$, $\hat{\eta}_i = x_i - \hat{r}(t_i)$, and $\hat{\eta}_{(k)} = x_k - \hat{r}_k$ are then calculated using the optimal bandwidth for the x's.

Then a bandwidth was chosen to use in the weighted likelihood. The bandwidth was chosen to minimize

$$\mathrm{CV}(b) = \sum_{k=1}^{n} \left[Y_k - \frac{\exp(\hat{\eta}_{(k)} \beta_{(k)} + h_{(k)})}{1 + \exp(\hat{\eta}_{(k)} \beta_{(k)} + h_{(k)})}\right]^2$$

with respect to b. The $\beta_{(k)}$ and $h_{(k)}$ are calculated with $(\hat{\eta}_{(k)}, t_k, Y_k)$ left out.

The results of the two simulations are summarized in Table 1. The table gives the mean $\hat{\beta}$ and the standard deviation of $\hat{\beta}$ for each block. The variance using the estimates of β and g, $\hat{\sigma}^2(\hat{\beta}, \hat{g})$, was calculated for each sample. The mean and standard deviation of $\hat{\sigma}(\hat{\beta}, \hat{g})$ for each block is given in the table.

The simulations show that the WLE is estimating $\beta_0 = 1$ well. The bias of the WLE of β for unequally spaced t is comparable to the bias of the estimate of β_0 for the case when t is equally spaced. The standard deviation of $\hat{\beta}$ for the simulation with t unequally spaced is larger. The standard deviation of $\hat{\beta}$ for the simulation with equally spaced t is about 1.64 times larger than the asymptotic standard deviation of $\hat{\beta}$. The standard deviation of $\hat{\beta}$ calculated with unequally spaced t is about 2.46 times larger than the asymptotic standard deviation, of $\hat{\beta}$.

Figure 1 is a histogram of the 500 $\hat{\beta}$'s for each of the simulations. Figure 2 is a QQ plot of the observed quantities of the standardized $\hat{\beta}$'s for each of the simulations versus the quantiles of a normal (0, 1) distribution. It is expected from the results of Theorem 1 that the histograms of the $\hat{\beta}$'s are normally distributed with mean 1 as $n \to \infty$. This is supported in Figures 1 and 2. The distributions of $\hat{\beta}$ are similar when the t's are equally spaced and unequally spaced.

The 25th, 50th, and 75th pointwise percentiles of $\hat{p}_i = [\exp(\hat{g}_i)]/[1 + \exp(\hat{g}_i)]$ for fixed t_i are shown in Figure 3 for each simulation. The percentiles are based on the 500 realizations from the Monte Carlo study. The WLE's pick up the underlying shape of the curve as expected. The peak of the curve is underestimated, but this is not surprising, because this is a known problem with kernel estimators. The percentiles do not appear to be very different for equally spaced and unequally spaced t, although the variance is slightly larger for unequally spaced t.

Because the asymptotic distribution of the estimate for β is known, it is possible to construct a Wald test for testing various hypotheses about β. Consider the typical one-sided test H_0: $\beta = 0$ vs. H_a: $\beta > 0$. The simulations can be used to find the power of the Wald test under the true condition that $\beta_0 = 1$. The power of the test at $\beta = \beta_0 = 1$ is P_1[reject H_0] $= P_1[\hat{\beta} > 0 + \hat{\sigma}(\hat{\beta}, \hat{g})z_\alpha]$, where z_α denotes the upper α point of the standard normal distribution. In our simulations this is the proportion of times $\hat{\beta} > \hat{\sigma}(\hat{\beta}, \hat{g})z_\alpha$ from the 500 estimates of β. The variance $\hat{\sigma}^2(\hat{\beta}, \hat{g})$ is calculated for each of the realizations.

The simulations can also be used to examine the P[type I error]. For this simulation, we would test H_0: $\beta = 1$ versus H_a: $\beta > 1$. The proportion of times $\hat{\beta} > 1 + \hat{\sigma}(\hat{\beta}, \hat{g})z_\alpha$ from the 500 estimates of β gives the simulated P[type I error]. We would expect this to be α.

Table 2 gives the power of the two simulations for this test for α = .025, .05, .10. The power seems to be similar for equally spaced t and unequally spaced t. Table 3 gives the P[type I error] for each simulation. In both simulations the P[type I error] is larger than α. This is not surprising, because there is a history of the actual level of the Wald test being greater than the nominal level (Cook 1990) in various settings. Due to this problem, future research could investigate other tests for hypothesis about β, such as likelihood ratio tests or score tests.

4. ANALYSIS OF ICC DATA

We finish the analysis of the ICC data set introduced earlier. The first risk factor, ASA, is a physical status classification that represents an evaluation of the patient's overall health prior to surgery as it would influence the surgery. In this analysis patients with ASA of 1 or 2 are combined into one group, because the distinction between these groups is minor. By the definitions of Klein (1985), patients in this combined group are those on whom the operation is performed for a local pathologic process and in whom the operation has no systemic effect. An example of this type of patient is an athlete who requires a hernia repair. This category also includes patients in good health but at the age extremes. Patients with ASA 3 are those patients who suffer from a significant systemic disturbance. Examples are patients with severe organic heart disease or severe diabetes. The second risk factor is OR duration, recorded in minutes.

The data set consists of 3,787 patients with each observation being of the form (x_i, t_i, Y_i), where Y_i is the response and x_i, t_i are the risk factors. The response, Y, is a binary random variable where

$$\begin{aligned} Y &= 0, \quad \text{if patient had no ICC} \\ &= 1, \quad \text{otherwise.} \end{aligned}$$

Let x_i represent the ASA score as follows:

$$\begin{aligned} x_i &= 0, \quad \text{patient has ASA score 1 or 2} \\ &= 1, \quad \text{patient has ASA score 3.} \end{aligned}$$

Table 1. Results of Simulation 1 and 2, With Mean and Standard Deviation of $\hat{\beta}$ and $\hat{\sigma}(\hat{\beta}, \hat{g})$ by Block

	Sim 1 (equally spaced t)				Sim 2 (unequally spaced t)			
N	Mean $\hat{\beta}$	Std $\hat{\beta}$	Mean $\hat{\sigma}(\hat{\beta}, \hat{g})$	Std $\hat{\sigma}(\hat{\beta}, \hat{g})$	Mean $\hat{\beta}$	Std $\hat{\beta}$	Mean $\hat{\sigma}(\hat{\beta}, \hat{g})$	Std $\hat{\sigma}(\hat{\beta}, \hat{g})$
100	1.1029	.4949	.3139	.0292	1.0549	.8295	.2931	.0406
100	.9849	.5037	.3158	.0266	.9693	.6918	.2817	.0445
100	1.0990	.5394	.3112	.0315	1.3176	.8705	.2713	.0453
100	.9728	.5080	.3124	.0275	.9766	.7189	.2784	.0355
100	1.0947	.4847	.3119	.0287	1.0750	.5965	.2757	.0400
500	1.0509	.5079	.3130		1.0787	.7556	.2800	

NOTE: When $\beta = 1$ and the true values for g are used, $\hat{\sigma}(\beta, g)$ = .309

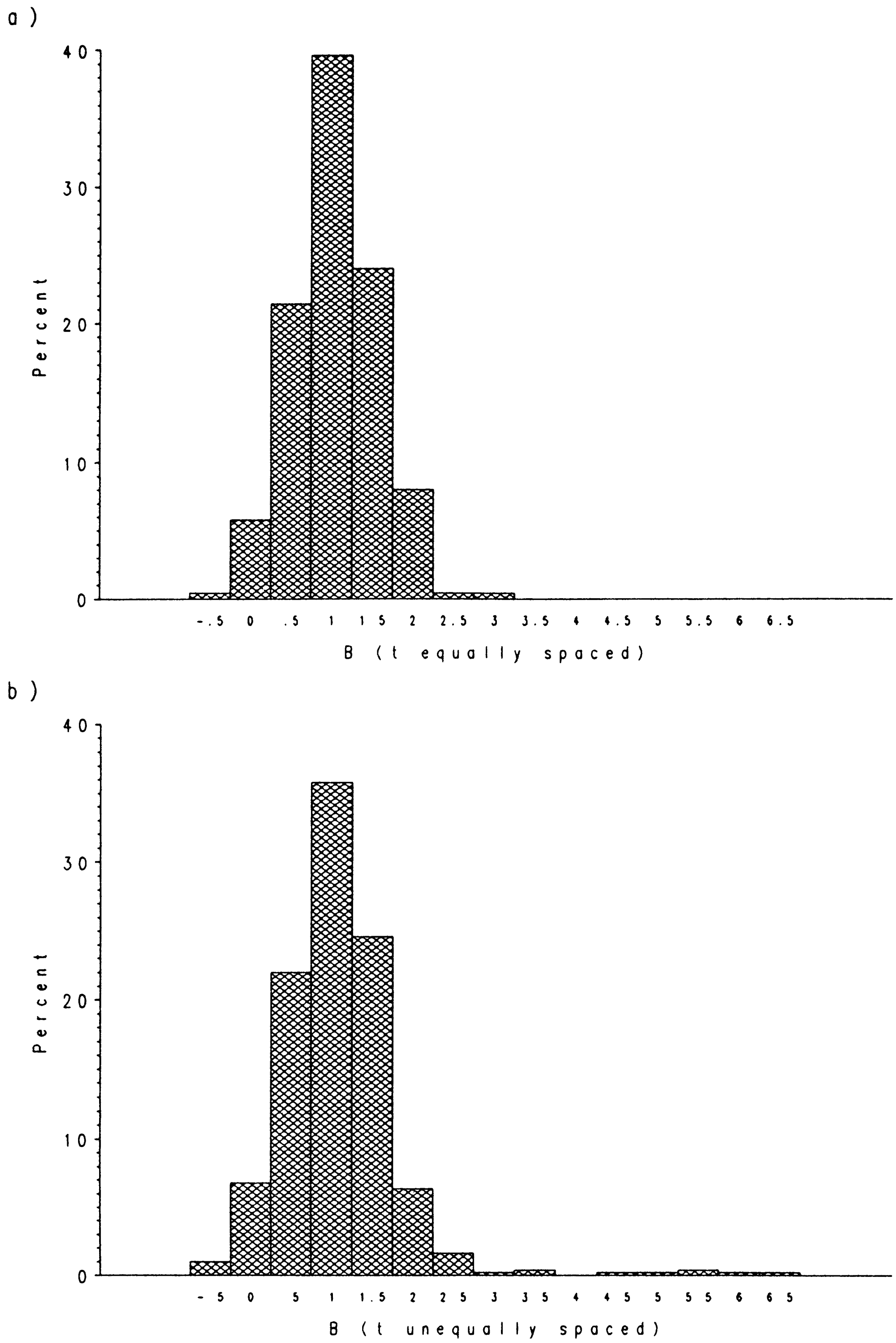

Figure 1. Histogram of 500 $\hat{\beta}$'s for Simulations 1 and 2.

The t_i are the log of the OR duration, which is rescaled to be between (0, 1). A log transformation on the OR duration values is used in an attempt to create more equally spaced values for the nonparametric component of the model.

Here as in the simulations, we assume $P[x_i = 1 \mid T_i = t_i] = r(t_i)$. When x is independent of t, the $r(t_i)$ is a constant independent of t_i. When $r(t_i)$ is a constant, the η's are homoscedastic, which is an assumption of the theorem. In the data set this would mean that the OR duration times occur in a similar pattern for each level of ASA. This assumption was examined by plotting $\hat{r}(t)$ by t. Figure 4a shows that this assumption appears to be valid, because $\hat{r}(t)$ is constant across t.

The model assumes an additive relation between the two risk factors. This assumption was checked by separating the data into the two ASA groups and finding the kernel estimate of the probability of an ICC across time for each group. Under an additive relationship, the shape of the logit curves should be similar, with the only difference being a constant shift

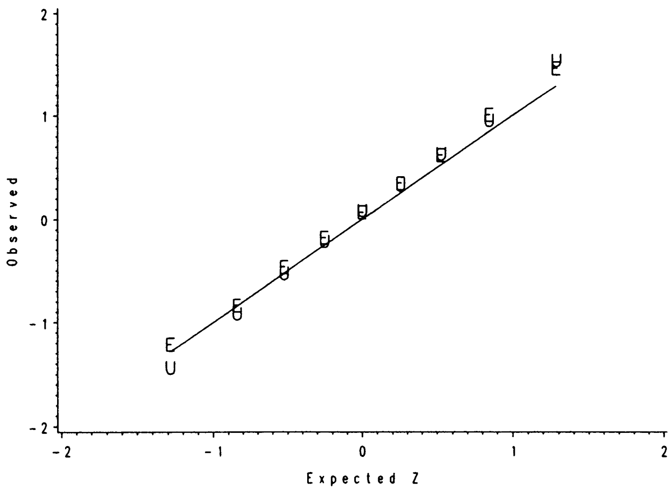

Figure 2. QQ Plots for Observed Quantiles of the Standardized $\hat{\beta}$'s. U and E are quantiles from the unequally and equally spaced simulations.

between the curves. Figure 4b supports this assumption. This figure also shows that the assumption of ASA being linearly related to the logit of the probability of an ICC is valid, because the difference in the two curves appears constant.

Because of the large sample size, the cross-validation (CV) method of choosing the bandwidth used in the simulations is not feasible. The following CV is used to select the bandwidth. A percent of the observations (details on the choice of percent are given later) for each level of x are left out, and η^*, β^*, and h^* are calculated as in Section 3. (Note that as in Section 3, a separate bandwidth is used to estimate η_i.) The observations are left out as follows. Each level of x score is sorted by the t values. Every kth observation is left out, where k is based on the desired percent of observations to be left out. The bandwidth criterion is

$$\mathrm{CV}(b) = \sum_k \left[Y_k - \frac{\exp(\eta_k^* \beta^* + h_k^*)}{1 + \exp(\eta_k^* \beta^* + h_k^*)} \right]^2,$$

where the sum is over the points left out. The bandwidth that gives the smallest CV is the bandwidth used to calculate the estimators.

The percent of observations left out in the bandwidth selection procedure was determined by trial and error. The following percents were used: 5%, 10%, and 15%. It was found that leaving out 5% of the data selected a bandwidth that was too small, and that 10% and 15% chose bandwidths that gave similar results for the estimate of β, with a good compromise between smoothing the data but not oversmoothing. Table 4 gives $\hat{\beta}$ and the bandwidth when 5%, 10%, and 15%

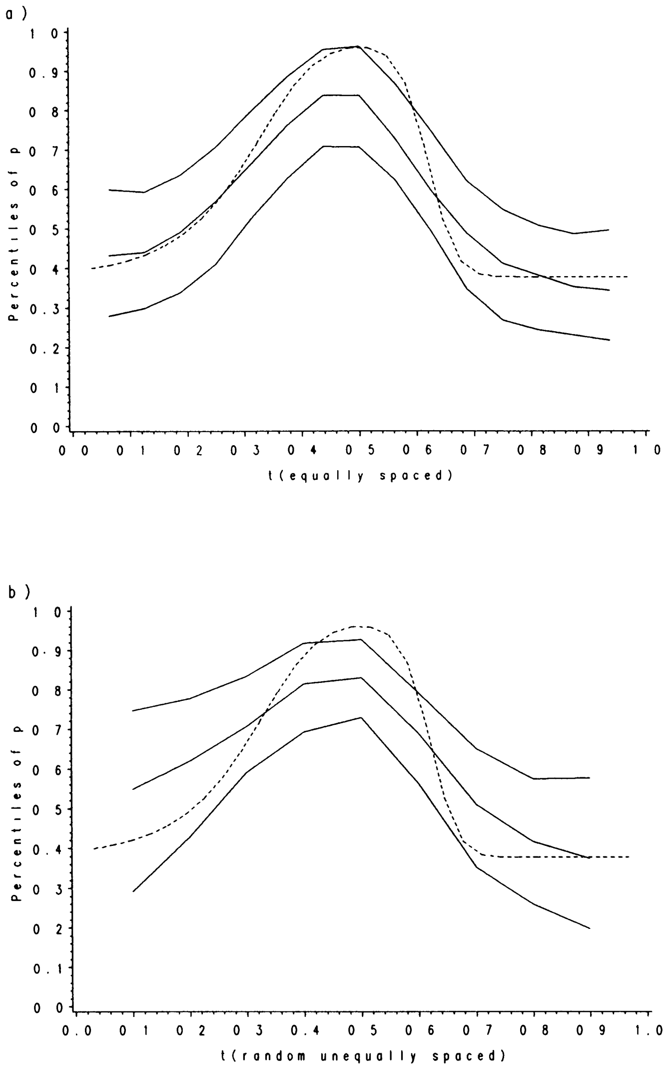

Figure 3. The 25th, 50th, 75th Percentiles of $\hat{p}_i$ Using 500 Observations for Each Logistic Simulation. The true curve is denoted by the dashed line.

of observations are left out in the bandwidth selection procedure. The var$(\hat{\beta})$ is calculated with the WLE's of β and g in the equations of Theorem 1. The last column is the p value for a Wald test of $\beta = 0$.

The interpretation of the parameter β is that the odds of an ICC increase by about $\exp(\hat{\beta}) = \exp(.724) = 2.06$ for a patient with an ASA score of 3 compared to the odds for a patient with an ASA of 1 or 2. The Wald test indicates that the increase in odds of an ICC due to having an ASA score of 3 as compared to that of 1 or 2 is significant.

Figure 5a gives the estimates of the function g by t when 5%, 10%, and 15% of the data were left out in the bandwidth

Table 2. Simulated Power of the Wald Test for Testing H_0: $\beta = 0$ versus H_a: $\beta > 0$

α level	Sim 1 (equally spaced t) Power	Sim 2 (unequally spaced t) Power
.025	78.6%	79.6%
.05	83.0%	84.6%
.10	90.0%	89.0%

Table 3. Simulated P[Type I Error] for the Test H_0: $\beta = 1$ versus H_a: $\beta > 1$

α level	Sim 1 (equally spaced t) P[type I error]	Sim 2 (unequally spaced t) P[type I error]
.025	.146	.170
.05	.196	.214
.10	.254	.276

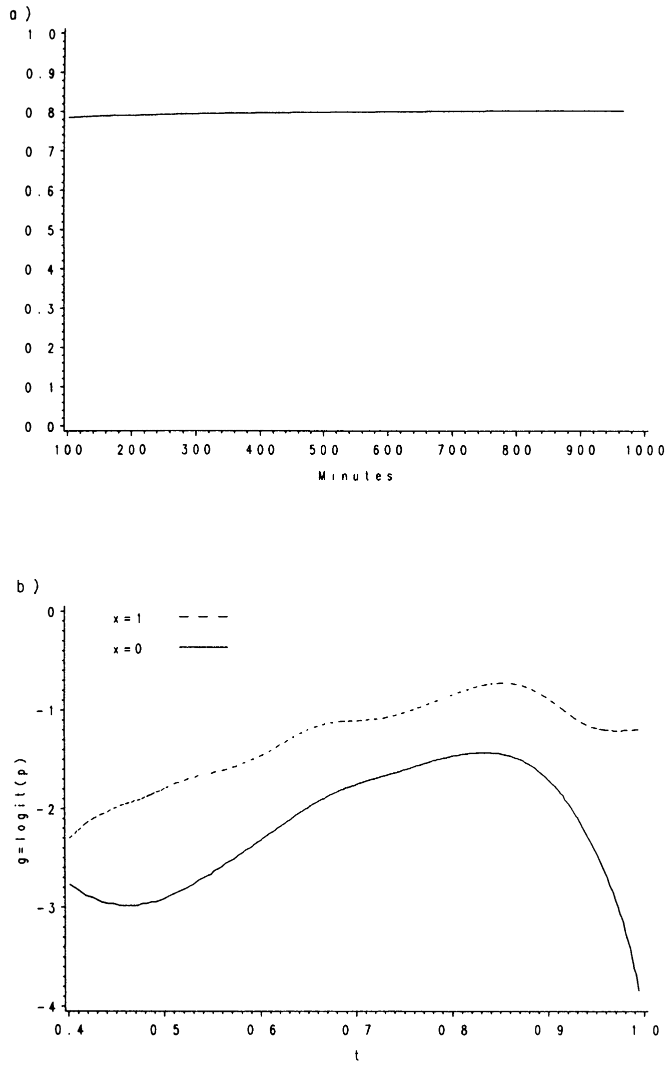

Figure 4. $\hat{f}(t)$ by t (a) and Kernel Estimate of $\hat{p}$ for Each ASA Level Estimated Separately (b).

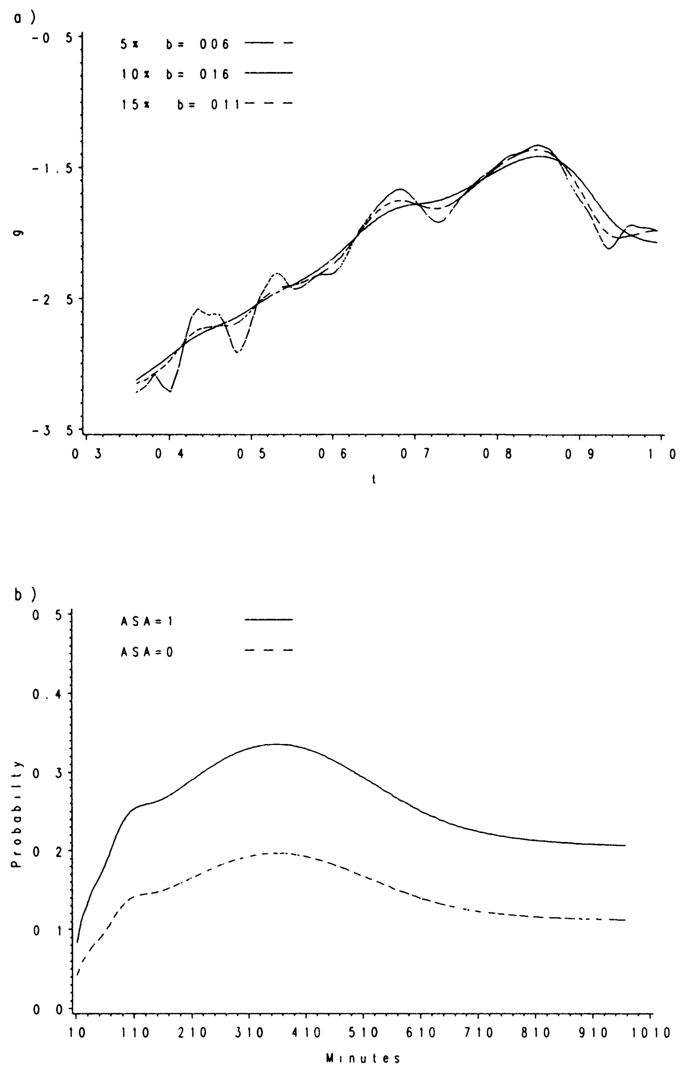

Figure 5. $\hat{g}_i$ versus t Leaving out 5%, 10%, and 15% of Data in the Bandwidth Selection Procedure (a) and $\hat{p}_i$ versus Minutes for $x = 0, 1$, Leaving Out 10% of the Data in the Bandwidth Selection Procedure (b).

selection procedure. Figure 4b gives the estimates of the probability, $p(t)$, of an ICC as a function of time for each level of x. Here the bandwidth selection procedure left out 10% of the data. The curves show that the probability of an ICC increases as the duration of the operation increases until about 400 minutes. The curve then starts to plateau or decrease. An interpretation of the plateau or decrease is that after a certain period, the probability of an ICC is a constant less than 1; in other words, all patients are not expected to have an ICC.

The usefulness of the semiparametric model is brought out in this example. Here is was possible to explore the functional relationship of OR duration and the probability of an ICC while performing tests of hypothesis and estimation about the ASA covariate. In this approach a plateau in the nonparametric curve was found that may have been missed if the wrong parametric model had been specified. The semiparametric model can be used as a tool to suggest an appropriate parametric model for the regression curve if further investigation of this covariate is desired.

Table 4. $\hat{\beta}$ From the Cardiac Complication Data Set

Percent leftout	$\hat{\beta}$	Bandwidth	var($\hat{\beta}$)	p value
5%	.731	.006	.0201	$\leqslant .05$
10%	.724	.016	.0201	$\leqslant .05$
15%	.727	.011	.0201	$\leqslant .05$

5. CONCLUSIONS

In this article a weighted likelihood is used to arrive at kernel estimators for the parametric and nonparametric components of the semiparametric regression model. The estimators are developed for the general case of arbitrary but specified densities of the observations. The WLE's are consistent and asymptotically normally distributed. The weighted likelihood method of estimation has a more general framework than other estimation methods, thereby allowing estimation for densities other the normal case.

The simulations show that the parametric component of the semiparametric model can be estimated well using the

WLE's, and the estimators extract the underlying shape of the curve for the nonparametric component. The data set analysis demonstrates that the semiparametric model has important applications. In the example the functional relationship for OR duration and the probability of an ICC was not specified, thereby allowing detection of a relationship that may otherwise have been missed. At the same time, it was possible to show that an increase in the covariate ASA significantly increased the probability of an ICC.

APPENDIX: PROOF OF THEOREM 1

Proof of Part a(1), $\|(\hat\beta - \beta_0, \hat{\mathbf{h}} - \mathbf{h})\|^2 \to_p 0$ as $n \to \infty$, where $\|(\beta, \theta)\|^2 = \max(|\beta|^2, \int_0^1 \theta^2(t)\,dt)$

In the proof of a(1) we will initially show that $\|\hat\beta - \beta_0, \hat{\mathbf{h}} - \mathbf{h}\| \to_p 0$ where $\|(\beta, \theta)\| = \max[|\beta|, (\int_0^1 \theta^p(t)\,dt)^{1/p}]$ for $p > 2$ and even. This implies the desired result $\|\hat\beta - \beta_0, \hat{\mathbf{h}} - \mathbf{h}\| \to_p 0$, where $\|(\beta, \theta)\| = \max[|\beta|, (\int_0^1 \theta^2(t)\,dt)^{1/2}]$, because, by Hölder's inequality, $(\int_0^1 \theta^2(t)\,dt)^{1/2} \le (\int_0^1 \theta^p(t)\,dt)^{1/p}$.

Before we begin the proof, we will define some notation. Let $\chi = \{(\beta, \theta): \beta \in \mathbb{R}, \theta \in \mathbb{L}_p[0, 1]\}$. The norm $\|\cdot\|$ on χ is $\|(\beta, \theta)\| = \max(|\beta|, \|\theta\|_p)$, where $\|\theta\|_p = (\int_0^1 \theta^p(t)\,dt)^{1/p}$. Let $Q_a = \{(\beta, \theta): \|(\beta, \theta)\| \le a\}$. We consider the behavior of the weighted log-likelihood on the surface of the sphere Q_a with radius a and $(\beta_0, \mathbf{h})$ the center of the sphere Q_a, where $(\beta_0, \mathbf{h})$ are the true values. We now show that the maximizer of $\mathrm{WL}(\beta, \theta)$ over $(\beta, \theta) \in Q_a$ exists.

The proof relies heavily on the work of Luenberger (1969) and will be referred to as L69. First, observe that χ is a complete normed linear space. Now it is shown that $\chi = \chi^{**}$, where $*$ is used to indicate the first dual and $**$ indicates the second dual. From L69 (p. 116), we know that $\mathbb{L}_p[0, 1] = \mathbb{L}_q[0, 1] = \mathbb{L}_p^{**}[0, 1]$, where $1/p + 1/q = 1$. Let T be in χ^*. Observe that $T_1(\beta) = T(\beta, 0)$ is a bounded linear mapping on $\mathbb{R}$ and $T_2(\theta) = T(0, \theta)$ is a bounded linear mapping on $\mathbb{L}_p[0, 1]$. Hence $T_1(\beta) \in \mathbb{R}^* = \mathbb{R}$ and $T_2(\theta) \in \mathbb{L}_p^*[0, 1] = \mathbb{L}_q[0, 1]$. Thus $T(\beta, \theta) = T(0, \theta) + T(\beta, 0) = T_2(\theta) + T_1(\beta) \in \mathbb{R} \oplus \mathbb{L}_p^*[0, 1]$. Likewise, given $T_1(\beta) \in \mathbb{R}^* = \mathbb{R}$ and $T_2(\theta) \in \mathbb{L}_p^*[0, 1] = \mathbb{L}_q[0, 1]$, we can construct $T \in \chi^*$ as $T(\beta, \theta) = T_2(\theta) + T_1(\beta)$. Hence $\chi^* = \mathbb{R} \oplus \mathbb{L}_q[0, 1]$.

Similarly, let T be in χ^{**}. Now $T_1(\beta) = T(\beta, \theta) \in \mathbb{R}^{**} = \mathbb{R}$ and $T_2(\theta) = T(0, \theta) \in \mathbb{L}_p^{**}[0, 1] = \mathbb{L}_p[0, 1]$. Thus $T(\beta, \theta) = T(0, \theta) + T(\beta, 0) = T_2(\theta) + T_1(\beta) \in \mathbb{R} \oplus \mathbb{L}_p[0, 1]$. Likewise, given $T_1(\beta) \in \mathbb{R}^{**} = \mathbb{R}$ and $T_2(\theta) \in \mathbb{L}_p^{**}[0, 1] = \mathbb{L}_p[0, 1]$, we can construct $T \in \chi^{**}$ as $T(\beta, \theta) = T_2(\theta) + T_1(\beta)$. Hence $\chi^{**} = \mathbb{R} \oplus \mathbb{L}_p[0, 1]$.

Using the triangle inequality, we show that Q_a is a closed unit ball. Suppose that $(\beta_n, \theta_n) \in Q_a$ and $\exists(\beta, \theta) \in \chi$ such that $d_n^p = \max(|\beta_n - \beta|^p, \|\theta_n - \theta\|_p^p) = \|(\beta_n - \beta, \theta_n - \theta)\|_\chi$ and $d_n^p \to 0$ as $n \to \infty$. Now we will show that $(\beta, \theta) \in Q_a$. The triangle inequality gives

$$\|(\beta, \theta)\|_\chi = \|((\beta - \beta_n), (\theta - \theta_n)) + (\beta_n, \theta_n)\|_\chi$$

$$\le \|((\beta - \beta_n), (\theta - \theta_n))\|_\chi + \|(\beta_n, \theta_n)\|_\chi,$$

$\forall\varepsilon > 0\ \exists N$, such that if $n > N$, the first term is less than ε, so $\|(\beta, \theta)\|_\chi \le \varepsilon + a$. Hence $(\beta, \theta) \in Q_a$.

Finally (L69, p. 109), $Q_a = Q_a^* = Q_a^{**}$. Therefore, from Theorem 1 of L69 (p. 128), $Q_a = Q_a^*$ is a weak* compact subset of $\chi = \chi^*$. The weighted likelihood is weak* continuous (L69, p. 128), so that the conditions of Theorem 2 of L69 (p. 128) are satisfied; therefore, the maximizer of $\mathrm{WL}(\beta, \boldsymbol\theta)$ in Q_a exists for large n.

Next we show that $\mathrm{WL}(\beta, \boldsymbol\theta) - \mathrm{WL}(\beta_0, \mathbf{h}) < 0$ with probability tending to 1 as $n \to \infty$ for sufficiently small a for $(\beta, \boldsymbol\theta)$ on the surface of Q_a. Now if $\mathrm{WL}(\beta, \boldsymbol\theta) - \mathrm{WL}(\beta_0, \mathbf{h}) < 0$ with high probability for all $(\beta, \boldsymbol\theta)$ on the surface of Q_a and $\max(\mathrm{WL}(\beta, \boldsymbol\theta) - \mathrm{WL}(\beta_0, \mathbf{h})) = 0$ at $(\beta_0, \mathbf{h})$ the center of the sphere Q_a, then it follows that $\mathrm{WL}(\beta, \boldsymbol\theta)$ has a local maximum in the interior of the sphere Q_a with high probability. Thus we will have shown $\|(\hat\beta - \beta_0, \hat{\mathbf{h}} - \mathbf{h})\|^2 \to_p 0$ as $n \to \infty$.

Now we show $\max(\mathrm{WL}(\beta, \boldsymbol\theta) - \mathrm{WL}(\beta_0, \mathbf{h})) < 0$ with high probability for $(\beta, \boldsymbol\theta)$ on the surface of Q_a, with $(\beta_0, \mathbf{h})$ the center of the sphere Q_a. We begin with a second-order Taylor expansion of $\mathrm{WL}(\beta, \boldsymbol\theta)$ about $\beta = \beta_0, \boldsymbol\theta = \mathbf{h}$. At this point we define the notation used in the Taylor expansion. The $*$ lines indicate the terms of the Taylor expansion with zero added. The line below each $*$ line is a regrouping of terms in the previous line. Let

$$* \left\{ \frac{\sum_i^n}{n} \frac{\sum_j^n w\left(\frac{t_i - t_j}{b}\right)}{nb} \frac{\partial \log f(Y_j; \beta, \theta)}{\partial\beta}\bigg|_{\substack{\beta=\beta_0\\ \theta=h_i}} - A_{n\beta} + A_{n\beta} \right\} (\beta - \beta_0)$$

$$= \{\tilde A_{n\beta} + A_{n\beta}\}(\beta - \beta_0)$$

$$* \frac{1}{n}\sum_i^n \left\{ \frac{\sum_j^n w\left(\frac{t_i - t_j}{b}\right)}{nb} \frac{\partial \log f(Y_j; \beta, \theta)}{\partial\theta}\bigg|_{\substack{\beta=\beta_0\\ \theta=h_i}} - A_{nh_i} + A_{nh_i} \right\} (\theta_i - h_i)$$

$$= \frac{1}{n}\sum_i^n \{\tilde A_{nh_i} + A_{nh_i}\}(\theta_i - h_i)$$

$$* \left\{ \frac{\sum_i^n}{n} \frac{\sum_j^n w\left(\frac{t_i - t_j}{b}\right)}{nb} \frac{\partial^2 \log f(Y_j; \beta, \theta)}{\partial\beta^2}\bigg|_{\substack{\beta=\beta_0\\ \theta=h_i}} - B_{n\beta} + B_{n\beta} \right\} (\beta - \beta_0)^2$$

$$= \{\tilde B_{n\beta} + B_{n\beta}\}(\beta - \beta_0)^2$$

$$* \frac{1}{n}\sum_{i}^{n}\left\{\frac{\sum_j^n w\left(\frac{t_i-t_j}{b}\right)}{nb}\frac{\partial^2\log f(Y_j;\beta,\theta)}{\partial\theta^2}\bigg|_{\substack{\beta=\beta_0\\ \theta=h_i}} - B_{nh_i} + B_{nh_i}\right\}(\theta_i-h_i)^2$$

$$= \frac{1}{n}\sum_{i}^{n}\{\tilde{B}_{nh_i} + B_{nh_i}\}(\theta_i-h_i)^2$$

$$* \frac{1}{n}\sum_{i}^{n}\left\{\frac{\sum_j^n w\left(\frac{t_i-t_j}{b}\right)}{nb}\frac{\partial^2\log f(Y_j;\beta,\theta)}{\partial\beta\partial\theta}\bigg|_{\substack{\beta=\beta_0\\ \theta=h_i}} - C_{nh_i} + C_{nh_i}\right\}(\beta-\beta_0)(\theta_i-h_i)$$

$$* \frac{1}{n}\sum_{j}^{n}\left\{\frac{\sum_j^n w\left(\frac{t_i-t_j}{b}\right)}{nb}\frac{\partial^2\log f(Y_j;\beta,\theta)}{\partial\theta\partial\beta}\bigg|_{\substack{\beta=\beta_0\\ \theta=h_i}} - C_{nh_i} + C_{nh_i}\right\}(\beta-\beta_0)(\theta_i-h_i)$$

$$= \frac{1}{n}\sum_{i}^{n} 2\{\tilde{C}_{nh_i} + C_{nh_i}\}(\beta-\beta_0)(\theta_i-h_i)$$

$$* \frac{\sum_i^n}{n}\frac{\sum_j^n w\left(\frac{t_i-t_j}{b}\right)}{nb}\left[\frac{\partial^2\log f(Y_j;\beta,\theta)}{\partial\beta^2}\bigg|_{\substack{\beta=\bar{\beta}_{ij}\\ \theta=\bar{h}_{ij}}} - \frac{\partial^2\log f(Y_j;\beta,\theta)}{\partial\beta^2}\bigg|_{\substack{\beta=\beta_0\\ \theta=h_i}}\right](\beta-\beta_0)^2$$

$$= R_{n\beta}(\beta-\beta_0)^2$$

$$* \frac{1}{n}\sum_{i}^{n}\left\{\frac{\sum_j^n w\left(\frac{t_i-t_j}{b}\right)}{nb}\left[\frac{\partial^2\log f(Y_j;\beta,\theta)}{\partial\theta^2}\bigg|_{\substack{\beta=\bar{\beta}_{ij}\\ \theta=\bar{h}_{ij}}} - \frac{\partial^2\log f(Y_j;\beta,\theta)}{\partial\theta^2}\bigg|_{\substack{\beta=\beta_0\\ \theta=h_i}}\right]\right\}(\theta_i-h_i)^2$$

$$= \frac{1}{n}\sum_{i}^{n} R_{nh_i}(\theta_i-h_i)^2$$

$$* \frac{1}{n}\sum_{i}^{n}\left\{\frac{\sum_j^n w\left(\frac{t_i-t_j}{b}\right)}{nb}\left[\frac{\partial^2\log f(Y_j;\beta,\theta)}{\partial\beta\partial\theta}\bigg|_{\substack{\beta=\bar{\beta}_{ij}\\ \theta=\bar{h}_{ij}}} - \frac{\partial^2\log f(Y_j;\beta,\theta)}{\partial\beta\partial\theta}\bigg|_{\substack{\beta=\beta_0\\ \theta=h_i}} + \frac{\partial^2\log f(Y_j;\beta,\theta)}{\partial\theta\partial\beta}\bigg|_{\substack{\beta=\bar{\beta}_{ij}\\ \theta=\bar{h}_{ij}}} - \frac{\partial^2\log f(Y_j;\beta,\theta)}{\partial\theta\partial\beta}\bigg|_{\substack{\beta=\beta_0\\ \theta=h_i}}\right]\right\}(\beta-\beta_0)(\theta_i-h_i)$$

$$= \frac{1}{n}\sum_{i}^{n} R_{nc_i}(\beta-\beta_0)(\theta_i-h_i),$$

where $\hat{\beta}_{ij}$ is between β and β_0 and $\bar{h}_{ij}$ is between θ_i and h_i. Now the last terms in each of the $*$ lines are defined. Let

$$A_{n\beta} = \frac{\sum_j^n}{n}\frac{\partial\log f(Y_j;\beta,\theta)}{\partial\beta}\bigg|_{\substack{\beta=\beta_0\\ \theta=h_i}},$$

$$A_{nh_i} = \frac{\sum_j^n w\left(\frac{t_i-t_j}{b}\right)}{nb}\frac{\partial\log f(Y_j;\beta,\theta)}{\partial\theta}\bigg|_{\substack{\beta=\beta_0\\ \theta=h_i}},$$

$$B_{n\beta} = \frac{\sum_j^n}{n}\frac{\partial^2\log f(Y_j;\beta,\theta)}{\partial\beta^2}\bigg|_{\substack{\beta=\beta_0\\ \theta=h_i}},$$

$$B_{nh_i} = \frac{\sum_j^n w\left(\frac{t_i-t_j}{b}\right)}{nb}\frac{\partial^2\log f(Y_j;\beta,\theta)}{\partial\theta^2}\bigg|_{\substack{\beta=\beta_0\\ \theta=h_i}},$$

$$C_{nh_i} = \frac{\sum_j^n w\left(\frac{t_i-t_j}{b}\right)}{nb}\frac{\partial^2\log f(Y_j;\beta,\theta)}{\partial\beta\partial\theta}\bigg|_{\substack{\beta=\beta_0\\ \theta=h_i}} = \frac{\sum_j^n w\left(\frac{t_i-t_j}{b}\right)}{nb}\frac{\partial^2\log f(Y_j;\beta,\theta)}{\partial\theta\partial\beta}\bigg|_{\substack{\beta=\beta_0\\ \theta=h_i}}.$$

The Taylor expansion can now be written as

$$\mathrm{WL}(\beta,\boldsymbol{\theta}) - \mathrm{WL}(\beta_0,\mathbf{h})$$
$$= A_{n\beta}(\beta-\beta_0) + \frac{\sum_i}{n}A_{nh_i}(\theta_i-h_i) + \tilde{D}_n + R_n + B_n, \text{ with} \quad \text{(A.1)}$$

$$B_n = B_{n\beta}(\beta-\beta_0)^2 + \frac{\Sigma_i}{n} B_{nh_i}(\theta_i - h_i)^2 + 2\frac{\Sigma_i}{n} C_{nh_i}(\beta-\beta_0)(\theta_i - h_i),$$

$$\tilde{D}_n = \tilde{A}_{n\beta}(\beta-\beta_0) + \frac{\Sigma_i}{n}\tilde{A}_{nh_i}(\theta_i - h_i) + \tilde{B}_{n\beta}(\beta-\beta_0)^2 + \frac{\Sigma_i}{n}\tilde{B}_{nh_i}(\theta_i - h_i)^2 + 2\frac{\Sigma_i}{n}\tilde{C}_{nh_i}(\beta-\beta_0)(\theta_i - h_i),$$

and

$$R_n = R_{n\beta}(\beta-\beta_0)^2 + \frac{\Sigma_i}{n} R_{nh_i}(\theta_i - h_i)^2 + \frac{\Sigma_i}{n} R_{nc_i}(\beta-\beta_0)(\theta_i - h_i).$$

Now five results are stated that are used later in the proof. These were proven by Hunsberger (1993).

R1. $|A_{n\beta}| < a^2$ and $\sup_i |A_{nh_i}| < a^2$, for $a > 0$, with probability tending to 1 as $n \to \infty$ so that $A_{n\beta} \to_p 0$ and $\sup_i(A_{nh_i}) \to_p 0$.

R2. $(\Sigma_i/n)(\theta_i - h_i)^p \le 2a^p$ and $(\Sigma_i/n)(\theta_i - h_i)^{p'} \le 2a^{p'}$ for $0 < p' < p$, $p > 2$ and even.

R3. $B_n = B_{n\beta}(\beta-\beta_0)^2 + (\Sigma_i/n)B_{nh_i}(\theta_i - h_i)^2 + 2(\Sigma_i/n) C_{nh_i}(\beta-\beta_0)(\theta_i - h_i) < -ca^2$, with probability tending to 1 as $n \to \infty$.

R4. $|\tilde{D}_n| < 10a^3$, for sufficiently small $a > 0$, with probability tending to 1 as $n \to \infty$.

R5. $|R_n| \le 15Ma^{2+\gamma}\|w\|_\infty$, with probability tending to 1 as $n \to \infty$. Here $\|w\|_\infty = \sup|w(x)|$.

Recall the Taylor expansion, Equation (A1):

$$\mathrm{WL}(\beta,\theta) - \mathrm{WL}(\beta_0,\mathbf{h}) = A_{n\beta}(\beta-\beta_0) + \frac{\Sigma_i}{n} A_{nh_i}(\theta_i - h_i) + \tilde{D}_n + R_n + B_n.$$

Using the five results, observe for (β,θ) on the surface of Q_a:

$$\max(\mathrm{WL}(\beta,\theta) - \mathrm{WL}(\beta_0,\mathbf{h}))$$
$$\le |A_{n\beta}|\,|\beta-\beta_0| + \frac{\Sigma_i}{n}|A_{nh_i}|\,|\theta_i - h_i| + \max_{\beta,\theta}[|\tilde{D}_n| + |R_n| + B_n]$$
$$\le a^2|\beta-\beta_0| + a^2\frac{\Sigma_i}{n}|\theta_i - h_i| + 10a^3 + 15M\|w\|_\infty a^{2+\gamma}$$
$$\le a^3 + 2a^3 + 10a^3 + 15M\|w\|_\infty a^{2+\gamma} - ca^2$$
$$\le (a^{1-\gamma} + 2a^{1-\gamma} + 10a^{1-\gamma} + 15M\|w\|_\infty)a^{2+\gamma} - ca^2,$$

with probability 1 as $n \to \infty$. So for $a^\gamma < 1$ and $13a^{1-\gamma} + 15M\|w\|_\infty < c$, $\max(\mathrm{WL}(\beta,\theta) - \mathrm{WL}(\beta_0,\mathbf{h}))$ is negative with high probability, and the proof of part a(1) is complete.

Proof of Part a(2), $|\hat{\beta}_h - \hat{\beta}_{\hat{h}}| \to_p 0$ as $n \to \infty$

This proof follows immediately from the proof of part a(1) by considering $\mathrm{WL}(\beta,\mathbf{h}) - \mathrm{WL}(\beta_0,\mathbf{h})$. $\mathrm{WL}(\beta,\mathbf{h})$ has a maximizer over β such that (β,θ) is in Q_a. Also, $\mathrm{WL}(\beta,\mathbf{h}) - \mathrm{WL}(\beta_0,\mathbf{h})$ is negative on Q_a with high probability; hence a local minimizer, call it $\hat{\beta}_h$, exists. Note that $\hat{\beta}_h$ and $\hat{\beta}_{\hat{h}}$ are both in Q_a with high probability. This proves part a(2).

Now several results are stated that will be used in the proof of part b. These results have been proven by Hunsberger (1993).

R6. $(\sqrt{n}A_{n\beta}/s_{n\beta}) \to_d N(0, 1)$.

R7. $B_{n\beta} - (-s_{n\beta}^2) \to_p 0$.

R8. $R_{n\beta}(\hat{\beta}) \to_p 0$.

R9. $\sqrt{n}3\tilde{D}_{n\beta}(\hat{\beta}) \to_p 0$.

R10. $\sqrt{n}s_{n\beta}(\hat{\beta}_h - \beta_0) \to_d N(0, 1)$, with $\mathbf{h}$ assumed known.

Proof of Part b, $\sqrt{n}s_{n\beta_0}(\hat{\beta}_{\hat{h}} - \beta_0) \to_d N(0, 1)$

We want to prove that $\lim_{n\to\infty} P[\sqrt{n}s_{n\beta}(\hat{\beta}_{\hat{h}} - \beta_0) \le t] = Z(t)$, where $Z(t)$ is the standard normal cdf. Choose and fix t such that t is a continuity point of Z. Let $\varepsilon > 0$ such that $t + \varepsilon$ and $t - \varepsilon$ are also continuity points of Z. Then

$$F_{\hat{\beta}_{\hat{h}}}(t) = P[\sqrt{n}s_{n\beta}(\hat{\beta}_{\hat{h}} - \beta_0) \le t]$$
$$\le P[\sqrt{n}s_{n\beta}(\hat{\beta}_{\hat{h}} - \beta_0) \le t, \|\hat{\mathbf{h}} - \mathbf{h}\|^2 \le \delta_\varepsilon] + P[\|\hat{\mathbf{h}} - \mathbf{h}\|^2 \ge \delta_\varepsilon]$$
$$\le P[\sqrt{n}s_{n\beta}(\hat{\beta}_{\hat{h}} - \hat{\beta}_h + \hat{\beta}_h - \beta_0) \le t, \|\hat{\mathbf{h}} - \mathbf{h}\|^2 \le \delta_\varepsilon] + P[\|\hat{\mathbf{h}} - \mathbf{h}\|^2 \ge \delta_\varepsilon]$$
$$\le P[\sqrt{n}s_{n\beta}(\hat{\beta}_h - \beta_0) \le t + \varepsilon] + P[\|\hat{\mathbf{h}} - \mathbf{h}\|^2 \ge \delta_\varepsilon].$$

Therefore, $\overline{\lim}_{n\to\infty} F_{\hat{\beta}_{\hat{h}}}(t) \le Z(t+\varepsilon)$ by part a(1) and R10. It is necessary to show $\underline{\lim}_{n\to\infty} F_{\hat{\beta}_{\hat{h}}}(t) \ge Z(t-\varepsilon)$,

$$P[\sqrt{n}s_{n\beta}(\hat{\beta}_h - \beta_0) \le t - \varepsilon]$$
$$\le P[\sqrt{n}s_{n\beta}(\hat{\beta}_h - \beta_0) + \sqrt{n}s_{n\beta}(\hat{\beta}_{\hat{h}} - \hat{\beta}_h) \le t] + P[\sqrt{n}s_{n\beta}|\hat{\beta}_{\hat{h}} - \hat{\beta}_h| \ge \varepsilon]$$
$$\le F_{\hat{\beta}_{\hat{h}}}(t) + P[\sqrt{n}s_{n\beta}|\hat{\beta}_{\hat{h}} - \hat{\beta}_h| \ge \varepsilon].$$

Now $P[\sqrt{n}s_{n\beta}|\hat{\beta}_{\hat{h}} - \hat{\beta}_h| \ge \varepsilon]$ goes to 0 as $n \to \infty$, because $\sqrt{n}s_{n\beta}|\hat{\beta}_{\hat{h}} - \hat{\beta}_h| = \sqrt{n}s_{n\beta}|\hat{\beta}_{\hat{h}} - \beta_0 + \beta_0 - \hat{\beta}_h| \le \sqrt{n}s_{n\beta}|\hat{\beta}_{\hat{h}} - \beta_0| + \sqrt{n}s_{n\beta}|\beta_0 - \hat{\beta}_h|$ and each of these terms goes to zero as $n \to \infty$ by R10 and part a(2). Thus $Z(t-\varepsilon) \le \underline{\lim}_{n\to\infty} F_{\hat{\beta}_{\hat{h}}}(t)$, and by the continuity of $Z(t)$ and because ε may be taken arbitrarily small, we have $\sqrt{n}s_{n\beta}(\hat{\beta}_{\hat{h}} - \beta_0) \to_d N(0, 1)$.

[Received April 1991. Revised October 1993.]

REFERENCES

Agresti, A. (1984), *Analysis of Ordinal Categorical Data,* New York: John Wiley.

Buja, A., Hastie, T., and Tibshirani, R. (1988), "Linear Smoothers and Additive Models," *The Annals of Statistics,* 17, 453–555.

Chen, H. (1988), "Convergence Rates for Parametric Components in a Partly Linear Model," *The Annals of Statistics,* 16, 136–146.

Cheng, K. F., and Lin, P. E. (1981), "Nonparametric Estimation of a Regression Function," *Zeitschrift fuer Wahrsheinlichkeitstheorie und Verwardte Gebiete,* 57, 223–233.

Cook, R. D., and Weisberg, S. (1990), "Confidence Curves in Nonlinear Regression," *Journal of the American Statistical Association,* 85, 544–551.

Craemer, H. (1946), *Mathematical Methods of Statistics,* Princeton, NJ: Princeton University Press.

Engle, R., Granger, C., Rice, J., and Weiss, A. (1986), "Nonparametric Estimates of the Relation Between Weather and Electricity Sales," *Journal of the American Statistical Association,* 81, 310–320.

Green, P., Jennison, C., and Seheult, A. (1985), "Analysis of Field Experiments by Least Squares Smoothing," *Journal of the Royal Statistical Society,* Ser. B, 47, 299–315.

Hastie, T., and Tibshirani, R. (1987a), "Generalized Additive Models," *Statistical Science,* 1, 297–318.

——— (1987b), "Generalized Additive Models: Some Applications," *Journal of the American Statistical Association,* 82, 371–386.

Heckman, N. (1986), "Spline Smoothing in Partly Linear Models," *Journal of the Royal Statistical Society,* Ser. B, 48, 244–248.

——— (1988), "Minimax Estimates in a Semiparametric Model," *Journal of the American Statistical Association,* 83, 1090–1096.

Hunsberger, S. (1990), "Semiparametric Regression in Likelihood Based Models," unpublished Ph.D. dissertation, Virginia Commonwealth University, Medical College of Virginia, Dept. of Biostatistics.

——— (1993), "A Proof for the Consistency and Asymptotic Normality of the Weighted Likelihood Estimators for the Parameters in a Semiparametric Model," Technical Report 9301, Virginia Commonwealth University, Medical College of Virginia, Dept. of Biostatistics.

Klein, S. L. (1985), *A Glossary of Anesthesia and Related Terminology,* New York: Springer-Verlag.

Messer, K., and Goldstein, L. (1989), "A New Class of Kernels for Nonparametric Curve Estimation," unpublished manuscript, California State University, Fullerton, CA 92634.

Priestley, M. B., and Chao, M. T. (1972), "Nonparametric Function Fitting," *Journal of the Royal Statistical Society,* Ser. B, 34, 385–392.

Rice, J. (1984), "Bandwidth Choice for Nonparametric Regression," *The Annals of Statistics,* 12, 1215–1230.

——— (1986), "Convergence Rates for Partially Splined Models," *Statistics and Probability Letters,* 4, 203–209.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics,* New York: John Wiley.

Severini, T. A., and Staniswalis, J. G. (1994), "Quasi-Likelihood Estimation in Semiparametric Models," *Journal of the American Statistical Association,* 89, 501–511.

Severini, T. A., and Wong, W. H. (1992), "Generalized Profile Likelihood and Conditionally Parametric Models," *The Annals of Statistics,* 20, 1768–1802.

Speckman, P. (1988), "Kernel, Smoothing in Partial Linear Models," *Journal of the Royal Statistical Society,* Ser. B, 50, 413–436.

Staniswalis, J. G. (1989), "The Kernel Estimate of a Regression Function in Likelihood-Based Models," *Journal of the American Statistical Association,* 84, 276–283.

Tibshirani, R., and Hastie, T. (1987), "Local Likelihood Estimation," *Journal of the American Statistical Association,* 82, 559–567.