UNIVERSITY OF WISCONSIN
DEPARTMENT OF BIOSTATISTICS

Technical Report #119

February 1997

RESIDUAL DIAGNOSTICS AND LOCAL INFLUENCE IN RIDGE REGRESSION AND NONPARAMETRIC REGRESSION

Wen-Hsiang Wei
and
Michael R. Kosorok, PhD
Department of Biostatistics

MADISON, WISCONSIN

Residual Diagnostics and Local Influence in Ridge Regression and Nonparametric Regression

By WEN-HSIANG WEI AND MICHAEL R. KOSOROK

Departments of Statistics, University of Wisconsin, Madison, Wisconsin 53706, U.S.A.

SUMMARY

Residual and local-influence diagnostic methods, corresponding to score residuals proposed by Therneau, Grambsch and Fleming (1990) and to Cook's (1986) local influence method, respectively, are developed for a variety of regression models. The interrelations between the two diagnostic measures are discussed. Several applications are given. One numerical study based on simulated data as well as analyses of real data is presented.

KEY WORDS: Regression Diagnostics; Influential Observations; Penalized Likelihood; Ridge Regression; Smoothing.

1. INTRODUCTION

In the linear regression model, a ridge estimator can be employed to achieve a reduction in variance as well as to solve certain collinearity problems. Ridge estimators are appropriate in two settings (Draper and Nostrand, 1979). One is when the norm of the parameters is fixed and the other is under a Bayesian formulation with a multivariate normal spherical prior. The ridge estimator can be obtained by minimizing the following negative penalized log-likelihood

$$\tilde{L}(\boldsymbol{\beta}) = -L(\boldsymbol{\beta}) + k\boldsymbol{\beta}^T\boldsymbol{\beta}, \tag{1}$$

where $L(\boldsymbol{\beta})$ is the log-likelihood function, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ is a $p \times 1$ vector of unknown parameters and k is the shrinkage parameter.

In nonparametric regression models, the penalized estimator can frequently be derived by minimizing the approximated negative penalized log-likelihood (Wahba, 1990, pp. 112-113; Hastie

and Tibshirani, 1990, pp. 149-151)

$$\tilde{L}(\boldsymbol{c}) = -L(\boldsymbol{c}) + \lambda \boldsymbol{c}^T \boldsymbol{P} \boldsymbol{c}, \tag{2}$$

where $L(\boldsymbol{c})$ is the approximated log-likelihood, $\boldsymbol{P}$ is a certain penalty matrix and $\boldsymbol{c} = (c_1, \ldots, c_p)^T$ is a $p \times 1$ unknown coefficient vector. Penalized estimators are appropriate in two settings. One is when information about the smoothness of the estimate, as reflected by $\lambda \boldsymbol{c}^T \boldsymbol{P} \boldsymbol{c}$, needs to be considered. The other is under a Bayes formulation (Wahba, 1990, pp. 19-20; Hastie and Tibshirani, 1990, pp. 129-130). When $\lambda = 0$, $\tilde{L}(\boldsymbol{c})$ reduces to the approximated negative log-likelihood in regression spline or polynomial regression.

There are several methods currently available for identifying influential observations in regression settings. Deletion diagnostics are well-developed for linear models (see Belsley, Kuh and Welsch, 1980; Cook and Weisberg, 1982; Chatterjee and Hadi, 1988). However, little work has been done for identifying influential observations in ridge regression and nonparametric regression models. Eubank and Gunst (1986) developed several standard diagnostics, including scaled residuals, leverage values and influence measures, for a class of penalized least-squares estimators. Walker and Birch (1988) showed that the influence of some observations on the ridge estimator can change as a function of the shrinkage parameter. They also proposed an approximate jackknife statistic for diagnostic purposes.

Cook (1986) developed a local influence method based on likelihood displacement to assess the effect of small perturbations of several likelihood components, including response vectors, case weights, and explanatory variables. The local influence approach has been applied to a variety of statistical models owing to the flexibility of likelihood displacement. Important examples include Thomas and Cook (1989, 1990), Pettitt and Daud (1989) and Laurent and Cook (1993). Other local influence measures based on geometrical measures or on perturbed functions different from likelihood displacement can be developed. Important examples include Lawrance (1988) and Wu and Luo (1993). Score components can also be employed to detect influential observations. Therneau, Grambsch and Fleming (1990) employ score components of the partial

likelihood function to identify influential points in the proportional hazards model. They refer to these score components as "score residuals". As described in Wei and Kosorok (1995b), we refer to individual components corresponding to the first derivative of the negative penalized log-likelihood as "estimating function residuals". In addition, we develop fairly general local influence measures for a variety of statistical models. Our approach can be readily utilized when only a subset of regression parameters is of interest. Thus the local influence approach presented here is a useful generalization of Cook's (1986) local influence graph approach. Section 2 introduces residual diagnostics, while the local influence method is presented in section 3. Several applications are given in section 4, while one extensive numerical study is presented in section 5. Finally, the interrelations among different influence measures, conclusions and a brief discussion are given in section 6.

2. RESIDUAL DIAGNOSTICS

Let $\boldsymbol{y} = (y_1, \ldots, y_N)^T$ be the response. Assume that the score function $\boldsymbol{S}(\boldsymbol{\beta})$, the first derivative of the log-likelihood, has the form

$$\boldsymbol{S}(\boldsymbol{\beta}) \;=\; \sum_{i=1}^{N} \boldsymbol{S}_{i\cdot}^{T}(y_i, \boldsymbol{\beta}),$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector and $\boldsymbol{S}_{i\cdot} = \{S_{i1}(y_i, \boldsymbol{\beta}), \ldots, S_{ip}(y_i, \boldsymbol{\beta})\}$. The penalized estimators $\hat{\boldsymbol{\beta}}$ can be obtained by solving estimating equations $\boldsymbol{G}(\boldsymbol{\beta}) = 0$, where $\boldsymbol{G}(\boldsymbol{\beta}) = \{G_1(\boldsymbol{\beta}), \ldots, G_p(\boldsymbol{\beta})\}^T$, and

$$G_j(\boldsymbol{\beta}) \;=\; \sum_{i=1}^{N} S_{ij}(y_i, \boldsymbol{\beta}) + k\beta_j. \tag{3}$$

The estimating function residual vector for the j'th parameter is

$$\boldsymbol{S}_{\cdot j} = \{S_{1j}(y_i, \hat{\boldsymbol{\beta}}), \ldots, S_{Nj}(y_N, \hat{\boldsymbol{\beta}})\}^T.$$

In many regression settings, estimating function residuals can be written as the product of residuals from the fit and the covariate value. Thus, estimating function residuals can serve as measures of influence. The most influential elements of the data can then be identified by plotting estimating function residuals versus observation number or versus covariate values.

The above estimating equations can be extended for nonparametric regression models. Assume the predictor

$$f \approx \sum_{j=1}^{p} c_j B_j,$$

where the B_j's are some basis functions and c_j's are unknown coefficients to be estimated. Assume the score function $\boldsymbol{S}(\boldsymbol{c})$, the first derivative of the log-likelihood function with respect to $\boldsymbol{c}$, has the form

$$\boldsymbol{S}(\boldsymbol{c}) = \sum_{i=1}^{N} \boldsymbol{S}_{i.}^{T}(y_i, \boldsymbol{c}),$$

where $\boldsymbol{S}_{i.} = \{S_{i1}(y_i, \boldsymbol{c}), \ldots, S_{ip}(y_i, \boldsymbol{c})\}$. Then, the estimates $\hat{\boldsymbol{c}} = (\hat{c}_1, \ldots, \hat{c}_p)^T$ of the vector of coefficients $\boldsymbol{c}$ can be obtained by solving $\boldsymbol{G}(\boldsymbol{c}) = 0$, where $\boldsymbol{G}(\boldsymbol{c}) = \{G_1(\boldsymbol{c}), \ldots, G_p(\boldsymbol{c})\}^T$ and

$$G_j(\boldsymbol{c}) \;=\; \sum_{i=1}^{N} S_{ij}(y_i, \boldsymbol{c}) + \lambda(\boldsymbol{P}\boldsymbol{c})_j. \tag{4}$$

The estimating function residual vector is $\boldsymbol{S}$, where $\boldsymbol{S} = \{\left|\left|\boldsymbol{S}_{1.}^{T}(y_i, \hat{\boldsymbol{c}})\right|\right|, \ldots, \left|\left|\boldsymbol{S}_{N.}^{T}(y_i, \hat{\boldsymbol{c}})\right|\right|\}^T$, and where $\left|\left|\boldsymbol{S}_{i.}^{T}(y_i, \hat{\boldsymbol{c}})\right|\right| = \{\sum_{j=1}^{p} S_{ij}^2(y_i, \hat{\boldsymbol{c}})\}^{1/2}$.

3. LOCAL INFLUENCE

We modify Cook's (1986) influential graph and propose an "influential curve". Before proceeding, define $F : R^N \to R^1$ to be some statistic of interest, computed from the data, where N is the sample size. We introduce a weighting scheme into the model through the $N \times 1$ vector $\boldsymbol{w}$ which is restricted to some open subset Ω, mimicking Cook (1986). Let $F(\boldsymbol{w})$ be the statistics corresponding to this weighting scheme. For example, we could choose $F(\boldsymbol{w}) = LD(\boldsymbol{w})$, where $LD(\boldsymbol{w})$ is the weighted likelihood displacement used in Cook (1986). Assume the $F(\boldsymbol{w})$ are twice continuously differentiable functions with respect to $\boldsymbol{w}$, where $\boldsymbol{w} = (w_1, ..., w_N)^T$. Let $\boldsymbol{w}_0 = (1, ..., 1)^T$ and let every component of $\boldsymbol{w}_{(k_1, k_2, \ldots, k_m)}$ equal 1 except set the $(k_1, k_2, \ldots, k_m)'$th components equal to 0. Let $\boldsymbol{w} = \boldsymbol{w}_0 - a\boldsymbol{l}$ and $\boldsymbol{l} = (l_1, ..., l_N)^T$ subject to the restrictions $||\boldsymbol{l}|| = 1$ and $0 \leq a \leq 1$. Then $F(\boldsymbol{w}_0) = F$ and $F(\boldsymbol{w}_{(k_1, \ldots, k_m)})$ is

the statistic F without the contributions from the $(k_1, \ldots, k_m)'$th observations. The influence curve, $\boldsymbol{C} : R^1 \to R^2$, is defined to be

$$\boldsymbol{C}(a) = \{a, F(\boldsymbol{w}_0 - a\boldsymbol{l})\}^T . \tag{5}$$

The direction vector $\boldsymbol{l}$ determines the scheme by which the weight vector is attached to the data points. For instance,

$$\boldsymbol{l}_{(k)} = \begin{cases} 1 & \text{for } i = k \\ 0 & \text{for } i \neq k \end{cases}$$

corresponds to the deletion of the k'th observation. The parameter a determines the scale of the weighting scheme. Therefore, the curve $\boldsymbol{C}(a)$ in R^2 reflects the behavior of the influence of the weighting scheme on the statistics F.

We employ the differential geometry measure of rate of angular change to characterize the behavior of the influence curve for some weighted statistics. We propose a diagnostic procedure based on the rate of angular change of the influence curve around the null point $\boldsymbol{w}_0$.

The rate of angular change of $\boldsymbol{C}(a)$ at $\boldsymbol{w}_0$ corresponding to some weighting scheme $\boldsymbol{l}$ can be expressed as

$$\frac{\boldsymbol{l}^t \boldsymbol{A} \boldsymbol{l}}{\boldsymbol{l}^t \boldsymbol{B} \boldsymbol{l}}, \tag{6}$$

where $\boldsymbol{A} = \ddot{\boldsymbol{F}}$, $\boldsymbol{B} = \boldsymbol{I} + \dot{\boldsymbol{F}}\dot{\boldsymbol{F}}^T$, and where $\dot{\boldsymbol{F}} = \{\partial F(\boldsymbol{w})/\partial \boldsymbol{w}\}_{\boldsymbol{w}=\boldsymbol{w}_0}$ and $\ddot{\boldsymbol{F}} = \left\{\partial^2 F(\boldsymbol{w})/\partial \boldsymbol{w}^T \partial \boldsymbol{w}\right\}_{\boldsymbol{w}=\boldsymbol{w}_0}$. Additional justifications for these local influence diagnostics are given in Wei and Kosorok (1995a).

Directions corresponding to large rates of angular change (in absolute value) can identify groups of observations with potentially large joint influence. Unit directions corresponding to large rates of angular change (in absolute value) are the eigenvectors which correspond to large eigenvalues (in absolute value) of the matrix $\boldsymbol{B}^{-1}\boldsymbol{A}$. Plots of these directions versus observation number or versus covariate values can be used to identify influential points or groups of jointly influential points.

3.1 Perturbed Statistics

Let the perturbed estimating equations be $\boldsymbol{G}(\boldsymbol{w},\boldsymbol{\beta}) = \{G_1(\boldsymbol{w},\boldsymbol{\beta}),\ldots,G_p(\boldsymbol{w},\boldsymbol{\beta})\}^T$, where

$$G_j(\boldsymbol{w},\boldsymbol{\beta}) = \sum_{i=1}^{N} w_i S_{ij}(\boldsymbol{y}_i,\boldsymbol{\beta}) + k\beta_j.$$

Then the perturbed statistic

$$F_1(\boldsymbol{w}) \quad = \quad \left|\left|\hat{\boldsymbol{\beta}}(\boldsymbol{w}) - \hat{\boldsymbol{\beta}}\right|\right|_{\boldsymbol{V}}^2,$$

can be used for diagnostic purpose in a parametric model, where $\hat{\boldsymbol{\beta}}(\boldsymbol{w}) = \{\hat{\beta}_1(\boldsymbol{w}),\ldots,\hat{\beta}_p(\boldsymbol{w})\}^T$ is the perturbed parameter estimate obtained by solving $\boldsymbol{G}(\boldsymbol{w},\boldsymbol{\beta}) = 0$, and where, for any $p \times p$ positive definite matrix $\boldsymbol{V}$, the norm $||.||_{\boldsymbol{V}}$ on R^p is defined by $||v||_{\boldsymbol{V}} = (\boldsymbol{v}^T\boldsymbol{V}\boldsymbol{v})^{1/2}$. We can choose $\boldsymbol{V} = \left\{\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}})\right\}^{-1}$, the inverse of the variance estimate, or $\boldsymbol{V} = -\left\{\partial^2 L(\boldsymbol{\beta})/\partial\boldsymbol{\beta}^T\partial\boldsymbol{\beta}\right\}_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}$.

In nonparametric models, assume design points $\boldsymbol{x} = (\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N)^T$. Define the perturbed estimating equation $\boldsymbol{G}(\boldsymbol{w},\boldsymbol{c}) \equiv \{G_1(\boldsymbol{w},\boldsymbol{c}),\ldots,G_p(\boldsymbol{w},\boldsymbol{c})\}^T$, where

$$G_j(\boldsymbol{w},\boldsymbol{c}) \equiv \sum_{i=1}^{N} w_i S_{ij}(\boldsymbol{y}_i,\boldsymbol{c}) + \lambda(\boldsymbol{P}\boldsymbol{c})_j.$$

We replace $F_1(\boldsymbol{w}) = \left|\left|\hat{\boldsymbol{\beta}}(\boldsymbol{w}) - \hat{\boldsymbol{\beta}}\right|\right|_{\boldsymbol{V}}^2$ by

$$F_2(\boldsymbol{w}) = ||\hat{\boldsymbol{y}}(\boldsymbol{w}) - \hat{\boldsymbol{y}}||_{\boldsymbol{V}}^2$$

in order to carry out our diagnostic procedure in nonparametric regression models, where $\hat{\boldsymbol{y}}(\boldsymbol{w}) = \{\hat{y}_1(\boldsymbol{w}),\ldots,\hat{y}_N(\boldsymbol{w})\}^T$ and $\hat{\boldsymbol{y}} = (\hat{y}_1,\ldots,\hat{y}_N)^T$ are the predictions of the response vector obtained from the perturbed and unperturbed estimating equations. One possible choice for $\boldsymbol{V}$ is the variance estimate. Let $\hat{\boldsymbol{c}}(\boldsymbol{w}) = \{\hat{c}_1(\boldsymbol{w}),\ldots,\hat{c}_p(\boldsymbol{w})\}^T$ be the estimates solved from the estimating equation $\boldsymbol{G}(\boldsymbol{w},\boldsymbol{c}) = 0$. Then, $\hat{y}_i(\boldsymbol{w}) = \sum_{j=1}^{p}\hat{c}_j(\boldsymbol{w})B_j(\boldsymbol{x}_i)$ and $\hat{y}_i = \sum_{j=1}^{p}\hat{c}_j B_j(\boldsymbol{x}_i)$.

In the additive and generalized additive models, let the predictor $f(X_1,\ldots,X_q) = \sum_{k=1}^{q} f_k(X_k)$, where $f_1(X_1),\ldots,f_q(X_q)$ are predictor variables corresponding to covariates $X_1,\ldots,X_p$. If covariate X_k is of interest, diagnostics can be developed based on the influence curve $\{a, \left|\left|\hat{\boldsymbol{f}}_k(\boldsymbol{w}) - \hat{\boldsymbol{f}}_k\right|\right|^2\}$, where $\hat{\boldsymbol{f}}_k(\boldsymbol{w})$ and $\hat{\boldsymbol{f}}_k$ are the predictor estimates obtained from the perturbed and unperturbed, respectively, estimating equations corresponding to the k'th covariate.

3.2 Local Influence Diagnostics

In parametric models, the first order approximation of $\hat{\boldsymbol{\beta}}(\boldsymbol{w}) - \hat{\boldsymbol{\beta}}$ evaluated at the null point $\boldsymbol{w}_0$ and by the implicit function theorem, we have $F_1(\boldsymbol{w}) \approx ||\boldsymbol{w}_0 - \boldsymbol{w}||^2_{\boldsymbol{M}_1}$, where

$$\boldsymbol{M}_1 = \boldsymbol{S}^* \left[\left\{ \frac{\partial \boldsymbol{G}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} \right\}^{-1}_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \right]^T \boldsymbol{V} \left\{ \frac{\partial \boldsymbol{G}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} \right\}^{-1}_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} (\boldsymbol{S}^*)^T, \tag{7}$$

and where $\boldsymbol{S}^*$ is an $N \times p$ matrix with elements $S^*_{ij} = \{\delta G_j(\boldsymbol{w}, \boldsymbol{\beta})/\delta w_i\}_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \boldsymbol{w}=\boldsymbol{w}_0}$, for $i = 1, \ldots, N$ and $j = 1, \ldots, p$. The directions for local influence diagnostics are the unit eigenvectors corresponding to large eigenvalues of $\boldsymbol{M}_1$.

In nonparametric models, by the first order approximation of $\hat{\boldsymbol{c}}(\boldsymbol{w}) - \hat{\boldsymbol{c}}$ evaluated at the null point $\boldsymbol{w}_0$ and by the implicit function theorem, we have $r(\boldsymbol{w}) \approx ||\boldsymbol{w}_0 - \boldsymbol{w}||^2_{\boldsymbol{M}_2}$, where

$$\boldsymbol{M}_2 = \boldsymbol{S}^* \left[\left\{ \frac{\partial \boldsymbol{G}(\boldsymbol{c})}{\partial \boldsymbol{c}^T} \right\}^{-1}_{\boldsymbol{c}=\hat{\boldsymbol{c}}} \right]^T \boldsymbol{Z}^T \boldsymbol{V} \boldsymbol{Z} \left\{ \frac{\partial \boldsymbol{G}(\boldsymbol{c})}{\partial \boldsymbol{c}^T} \right\}^{-1}_{\boldsymbol{c}=\hat{\boldsymbol{c}}} (\boldsymbol{S}^*)^T, \tag{8}$$

$\boldsymbol{Z}$ is the $N \times p$ matrix with (i, j)'th entry $B_j(x_i)$, and where $\boldsymbol{S}^*$ is an $N \times p$ matrix with elements $S^*_{ij} = \{\delta G_j(\boldsymbol{w}, \boldsymbol{c})/\delta w_i\}_{\boldsymbol{c}=\hat{\boldsymbol{c}}, \boldsymbol{w}=\boldsymbol{w}_0}$, for $i = 1, \ldots, N$ and $j = 1, \ldots, p$. The directions for local influence diagnostics are the unit eigenvectors of $\boldsymbol{M}_2$.

4. APPLICATIONS

The estimating function residuals and the local influence diagnostics can be developed for a variety of statistical models (see Wei, 1996), including ridge regression; the smoothing spline in Gaussian regression models; the smoothing spline in generalized linear model; the additive and generalized additive models proposed by Hastie and Tibshirani; and in response transformation models. The following examples demonstrate some of the applications of the proposed estimating function residuals and local influence.

4.1 Ridge Regression Model

Consider the standard linear regression model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)^T$ is an $N \times p$ matrix with column vectors $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^T$ and the elements ϵ_i of the $N \times 1$ vector $\boldsymbol{\epsilon}$ are assumed to be independent normal random variables with mean 0 and variance σ^2. The ridge estimator $\hat{\boldsymbol{\beta}}$ is the minimizer of

$$(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) + k\boldsymbol{\beta}^T\boldsymbol{\beta},$$

where k is some constant.

The estimating function residual for the j'th parameter and i'th residual is

$$S_{ij} = x_{ij}(y_i - \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}).$$

When the variance of ϵ_i is σ^2/w_i, the ridge estimator $\hat{\boldsymbol{\beta}}(\boldsymbol{w})$ is the minimizer of

$$(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T\boldsymbol{W}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) + k\boldsymbol{\beta}^T\boldsymbol{\beta},$$

where $\boldsymbol{W} = \mathrm{diag}(w_i)$. We then have the ridge estimator

$$\hat{\boldsymbol{\beta}}(\boldsymbol{w}) = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} + k\boldsymbol{I})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Y}.$$

Let $\boldsymbol{S} = (\boldsymbol{S}_{1.}^T, \ldots, \boldsymbol{S}_{N.}^T)^T$ be an $N \times p$ matrix, where $\boldsymbol{S}_{i.} = (S_{i1}, \ldots, S_{ip})$. Following the approach given in section 3.3, the local influence diagnostics are the eigenvectors of the matrices

$$\boldsymbol{M}_1 = \boldsymbol{S}(\boldsymbol{X}^T\boldsymbol{X} + k\boldsymbol{I})^{-1}\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X} + k\boldsymbol{I})^{-1}\boldsymbol{S}^T,$$

if $\boldsymbol{V} = -\left\{\partial^2 L(\boldsymbol{\beta})/\partial\boldsymbol{\beta}^T\partial\boldsymbol{\beta}\right\}_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}$ in (7), or

$$\boldsymbol{M}_1 = \boldsymbol{S}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{S}^T,$$

if $\boldsymbol{V} = \left\{\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}})\right\}^{-1}$.

4.2 Regression Spline in Gaussian Regression Models

Consider the following nonparametric regression model

$$\boldsymbol{Y} = E(\boldsymbol{Y} \mid \boldsymbol{X}) + \epsilon,$$

where $E(\boldsymbol{Y} \mid \boldsymbol{X})$ is the conditional expectation of variable $\boldsymbol{Y}$ given $\boldsymbol{X}$. Approximating $E(\boldsymbol{Y} \mid \boldsymbol{X})$ by a linear combination of some basis functions $B_1, \ldots, B_p$, we have that

$$E(\boldsymbol{Y} \mid \boldsymbol{X} = \boldsymbol{x}_i) \approx \sum_{j=1}^{p} c_j B_j(\boldsymbol{x}_i),$$

where $\boldsymbol{B}_i : R^p \rightarrow R^1$. Let $\hat{\boldsymbol{c}}$ minimize

$$(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{c})^T(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{c}),$$

where $\boldsymbol{Z}$ is the $N \times p$ matrix with the (i,j)'th entry $B_j(\boldsymbol{x}_i)$. The estimating function residuals are

$$\left(\sum_{j=1}^{p} B_j^2(\boldsymbol{x}_i)\right)^{1/2} (y_i - \boldsymbol{z}_i\hat{\boldsymbol{c}}),$$

where $\boldsymbol{z}_i$ is the i'th row of $\boldsymbol{Z}$. Let

$$S_{ij} = B_j(\boldsymbol{x}_i)(y_i - \boldsymbol{z}_i\hat{\boldsymbol{c}})$$

and let $\hat{\boldsymbol{c}}(\boldsymbol{w})$ be the minimizer of

$$(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{c})^T\boldsymbol{W}(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{c}).$$

The local influence diagnostics are the eigenvectors of the matrix

$$\boldsymbol{M}_2 = \boldsymbol{S}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{S}^T.$$

4.3 Spline Smoothing in Generalized Linear Models

Consider the standard generalized linear model in which each component of the response vector has a distribution taking the form

$$f(y_i;\theta,\phi) \quad = \quad \exp\left\{\frac{y_i\theta - b(\theta)}{a(\phi)} + c(y_i,\phi)\right\},$$

where θ and ϕ are scalar parameters, and $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$ are specific functions. The dependence of the response y_i on the associated explanatory variables can be modeled through

the link function $d(\cdot)$, where $\theta_i = d(\boldsymbol{x}_i^T\boldsymbol{\beta})$. We will assume $a(\phi) = 1$ and the natural link hereafter. However, when the information related to the dependence of the predictor θ_i on the covariates is not well understood, it might be more appropriate to assume a nonparametric form for θ_i rather than a parametric form. Thus we can assume $\theta_i = f(\boldsymbol{x}_i)$ and try to estimate f by the penalized likelihood approach. Let f be an element of some Hilbert space $\mathcal{H}$. The estimate of f is the minimizer of the penalized negative logarithm of the likelihood,

$$\sum_{i=1}^{N} \left[b\left\{f(\boldsymbol{x}_i)\right\} - y_i f(\boldsymbol{x}_i)\right] + \lambda \left|\left|P_1 f\right|\right|_{\mathcal{H}}^2,$$

where P_1 is the orthogonal projection in $\mathcal{H}$ onto a subspace $\mathcal{H}_1$ and $||P_1 f||_{\mathcal{H}}^2$ is the penalty functional employed to incorporate information about the "smoothness" of the estimate. Let $\boldsymbol{c} = (c_1, \ldots, c_p)^T$. Approximating f by a linear combination of some basis functions $B_1, \ldots, B_p$ in $\mathcal{H}$, we have that

$$f \approx \sum_{j=1}^{p} c_j B_j$$

(see Wahba, 1990, pp. 111-113). Let $\hat{\boldsymbol{c}}$ minimize

$$\sum_{i=1}^{N} b\left\{(\boldsymbol{Zc})_i\right\} - \boldsymbol{y}^T\boldsymbol{Zc} + \lambda \boldsymbol{c}^T\Sigma\boldsymbol{c},$$

where Σ is the $p \times p$ matrix with the (k, k')'th inner product $< P_1 B_k, P_1 B_{k'} >$, and where $(\boldsymbol{Zc})_i$ is the i'th entry of the vector $\boldsymbol{Zc}$. The estimating function residuals are

$$\left(\sum_{j=1}^{p} B_j^2(\boldsymbol{x}_i)\right)^{1/2} \left[b\{(\boldsymbol{Zc})_i\} - y_i\right]_{\boldsymbol{c}=\hat{\boldsymbol{c}}}.$$

Let

$$S_{ij} = B_j(\boldsymbol{x}_i)\left[b\{(\boldsymbol{Zc})_i\} - y_i\right]_{\boldsymbol{c}=\hat{\boldsymbol{c}}}$$

and let $\hat{\boldsymbol{c}}(\boldsymbol{w})$ be the minimizer of

$$\sum_{i=1}^{N} w_i b\left\{(\boldsymbol{Zc})_i\right\} - \boldsymbol{y}^T\boldsymbol{WZc} + \lambda \boldsymbol{c}^T\Sigma\boldsymbol{c}.$$

The local influence diagnostics are the eigenvectors of the matrix

$$\boldsymbol{M}_2 = \boldsymbol{S}(\boldsymbol{Z}^T\boldsymbol{D}\boldsymbol{Z} + \lambda\Sigma)^{-1}\boldsymbol{Z}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{D}\boldsymbol{Z} + \lambda\Sigma)^{-1}\boldsymbol{S}^T,$$

where $\boldsymbol{D}$ is a diagonal matrix with i'th diagonal element $\left\{\partial^2 b(\theta)/\partial\theta^2\right\}_{\theta=(\boldsymbol{Z}\boldsymbol{c})_i}$.

4.4 Additive Models in an Exponential Family

An additive model in an exponential family differs from the model in the previous example in the form of the predictor. Let $f_k \approx \sum_{j=1}^{p_k} c_{kj}B_{kj}$, where B_{kj}'s are some basis functions. Let $\boldsymbol{c}_k = (c_{k1}, \ldots, c_{kp_k})^T$, and let $\hat{\boldsymbol{c}}_k$'s be the minimizer of the penalized negative logarithm of the likelihood,

$$-L(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_q) = \sum_{i=1}^{N} b\left[k\left\{\sum_{k=1}^{q}(\boldsymbol{z}_k\boldsymbol{c}_k)_i\right\}\right] - \sum_{i=1}^{n} y_i k\left\{\sum_{k=1}^{q}(\boldsymbol{z}_k\boldsymbol{c}_k)_i\right\} + \sum_{k=1}^{q}\lambda_k\boldsymbol{c}_k^T\Sigma_k\boldsymbol{c}_k,$$

where $\boldsymbol{z}_k$ is the $N \times p_k$ matrix with the (i,j)'th entry $B_{kj}(\boldsymbol{x}_{ik})$ and Σ_k are certain penalty matrices. Hereafter, we assume the natural link. The estimating function residuals corresponding to the k'th predictor $\boldsymbol{x}_k = (x_{1k}, \ldots, x_{Nk})^T$ and the i'th observation are

$$\left(\sum_{j=1}^{p_k} B_{kj}^2(\boldsymbol{x}_{ik})\right)^{1/2}\left[b\{(\boldsymbol{z}_k\boldsymbol{c}_k)_i\} - y_i\right]_{\boldsymbol{c}_k=\hat{\boldsymbol{c}}_k}.$$

Let $\hat{\boldsymbol{c}}_k(\boldsymbol{w}) = \{c_{k1}(\boldsymbol{w}), \ldots, c_{kp_k}(\boldsymbol{w})\}^T$ be the minimizer of

$$-L(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_q, \boldsymbol{w}) = \sum_{i=1}^{N} w_i b\left\{\sum_{k=1}^{q}(\boldsymbol{z}_k\boldsymbol{c}_k)_i\right\} - \sum_{i=1}^{N} w_i y_i\left\{\sum_{k=1}^{q}(\boldsymbol{z}_k\boldsymbol{c}_k)_i\right\} + \sum_{k=1}^{q}\lambda_k\boldsymbol{c}_k^T\Sigma_k\boldsymbol{c}_k.$$

Denote $\boldsymbol{Z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_q)$ to be an $N \times p$ matrix and $\boldsymbol{c} = (\boldsymbol{c}_1^T, \ldots, \boldsymbol{c}_q^T)^T$, where $p = \sum_{k=1}^{q} p_k$. Also denote

$$S_{ijk} = B_{kj}(\boldsymbol{x}_{ik})\left[b\{(\boldsymbol{z}_k\boldsymbol{c}_k)_i\} - y_i\right]_{\boldsymbol{c}_k=\hat{\boldsymbol{c}}_k}$$

and let $\boldsymbol{S} = (\boldsymbol{S}_{1\cdot}^T, \ldots, \boldsymbol{S}_{N\cdot}^T)^T$ be an $N \times p$ matrix, where $\boldsymbol{S}_{i\cdot} = (\boldsymbol{S}_{i1}, \ldots, \boldsymbol{S}_{iq})$ and $\boldsymbol{S}_{ik} = (S_{ik1}, \ldots, S_{ikp_k})$. Let $\Sigma_\lambda = \mathrm{diag}(\lambda_k\Sigma_k)$. For ease of exposition, assume the first covariate is of interest. Then, the local influence diagnostics are the eigenvectors of the matrices

$$\boldsymbol{M}_2 = \boldsymbol{S}\boldsymbol{E}_1^T\boldsymbol{z}_1^T\boldsymbol{z}_1\boldsymbol{E}_1\boldsymbol{S}^T,$$

where $\boldsymbol{E}_1$ is a $p_1 \times p$ matrix of which the rows are the first p_1 rows of $(\boldsymbol{Z}^T\boldsymbol{D}\boldsymbol{Z} + \lambda\Sigma_\lambda)^{-1}$ and where $\boldsymbol{D}$ is a diagonal matrix with the i'th diagonal element $\{\partial^2 b(\theta)/\partial\theta^2\}_{\theta=(\boldsymbol{Z}\boldsymbol{c})_i}$.

5. NUMERICAL ILLUSTRATIONS

5.1 Simulated Data Results: Ridge Regression

300 observations were generated from the following model: observations 250, 251 and 252 were generated from $y = 1 + 0.15x_1 + 0.25x_2 + \epsilon$ for $x_1,\ x_2 = 250,\ 251$ and 252, observation 175 was simulated from the model $y = 1 + 0.11x_1 + 0.21x_2 + \epsilon$ for $x_1,\ x_2 = 175$, observations 125 and 126 were generated from $y = 1 + 0.1x_1 + 0.2x_2 + \epsilon$ for $x_1 = 130,\ 135$ and $x_2 = 125,\ 135$ and the other observations were generated from $y = 1 + 0.1x_1 + 0.2x_2 + \epsilon$ for $x_1,\ x_2$ equal to integers from 1 to 300 excluding 125, 126, 175, 250, 251, and 252, where ϵ is a normal random error with mean 0 and standard deviation 0.1 and $x_2 = x_1$.

Observations 125 with unequal covariates x_1 and x_2 results in full rank of the design matrix. A high level of collinearity is present. The value of the shrinkage parameter that minimizes a prediction-oriented criterion proposed by Myers (1986, pp. 249) is 0.013. Observations 250, 251 and 252 have larger jackknife statistics. These three observations might be potentially influential points. Although generated from a different model, observation 175 has a small jackknife statistic and a small estimating function residual. That might be because observation 175 was generated from a model not significantly different from the main model. Similarly, observations 125 and 126 also can be regarded as typical observations because their covariate values are not significantly different from the main model.

These potentially influential observations were identified by examining the estimating function residuals and by the eigenvectors of matrix $\check{\boldsymbol{P}}$ in (7). The plots of the estimating function residuals and eigenvectors obtained by the general local influence method are shown in Figures 1, 2 and 3. 100 replicates of the above numerical experiment were conducted. Observations 251, 252 and 253 were identified in each experiment by the general local influence method or by examining the estimating function residuals when the cutoff value for "identification" is specified

as 0.5 for the local influence method and 0.4 for the estimating function residuals corresponding to both covariates. No other observations were identified as influential.

5.2 Simulated Data Results: Cubic B-Spline

Let $f(x) = 4.26(e^{-x} - 4e^{-2x} + 3e^{-3x})$. 99 observations were generated from the following model: observations 7, 8 were generated from $y = 4.26(e^{-x} - 3e^{-2x} + 3e^{-3x}) + \epsilon$ for $x = 0.07$ and 0.08, observation 17 and 18 were generated from $y = 4.26(e^{-x} - 4.05e^{-2x} + 3e^{-3x}) + \epsilon$ for $x = 0.17,\ 0.18$, and the other observations were generated from $y = f(i/100) + \epsilon$ for i equal to integers from 1 to 99 excluding 7, 8, 17 and 18, where ϵ is a normal random error with mean 0 and standard deviation 0.2.

These potentially influential observations, observations 7, 8, 17 and 18 were identified by examining the estimating function residuals and by the eigenvectors of matrix $\check{\boldsymbol{P}}$ in (8). The plots of the estimating function residuals and eigenvectors obtained by the general local influence method are shown in Figures 4 and 5.

5.3 Performance of Computerized System Study

The different diagnostic measures are now applied to the performance of a computerized system study (Hill, 1977). The data relate to the performance of a computerized system for processing military personnel action forms. The data can be found in Walker and Birch (1988). Walker and Birch (1988) fit the data by ridge regression. The level of collinearity of this data set is moderate, indicated by a scaled (uncentered) condition number of 57.14 (see Walker and Birch, 1988).

The value of the shrinkage parameter which minimizes Myers's (1986) criterion is 0.03.

Observations 1, 2, 8, 12, 13 and 15 were identified by examining the estimating function residuals and by the eigenvectors of matrix $\check{\boldsymbol{P}}$ in (7). Observations 1, 2, 8, 12 and 15 were also identified by DFFITS (see table 3, Walker and Birch, 1988). The plots of the estimating function residuals and eigenvectors obtained by the general local influence method are shown in Figures 6 and 7.

5.4 New York Ozone Concentration Data

The data are taken from a study by Bruntz et al. (1974) of dependence of ozone concentration on some meteorological variables on 111 days in 1973 in New York City. The meteorological variables are solar radiation, temperature and wind speed. Cleveland, Devlin and Grosse (1988) employed the additive model,

$$(\text{ozone concentration})^{1/3} = f_1(radiation) + f_2(windspeed) + f_3(temperature) + \epsilon,$$

for this data. The plot of the eigenvector of $\check{\boldsymbol{P}}$ in (8) corresponding to the the largest eigenvalue and plots of the estimating function residuals are shown in Figures 8 and 9. From the additive model fit, high wind speed results in low ozone concentrations while high temperature results in high concentration (also see Hastie and Tibshirani, 1990, pp. 256-259). Observation 17, identified by examining local influence diagnostics, estimating function residuals and a version of Cook's distance in Hastie and Tibshirani (1990), with wind speed close to mean value has lowest ozone concentration. Observation 77 identified by examining estimating function residuals with temperature close to the mean value has the highest ozone concentration.

6. CONCLUSIONS AND DISCUSSIONS

The new local influence approach can be extended for use with multivariate statistics. If the statistics of interest are $F_1, \ldots, F_q$, the influence curve, $\boldsymbol{C} : R^1 \to R^{q+1}$,

$$\boldsymbol{C}(a) = \{a, F_1(\boldsymbol{w}_0 - a\boldsymbol{l}), \ldots, F_q(\boldsymbol{w}_0 - a\boldsymbol{l})\}^T, \tag{9}$$

can be used for diagnostic purposes. The diagnostics can also be developed based on other geometrical measures different from the rate of angular change. For example, the length of the velocity vectors of the influence curve corresponding to the weighting scheme at the null point $\boldsymbol{w}_0$ is

$$||\boldsymbol{v}(a)||_{a=0} = \sqrt{1 + \sum_{i=1}^{q} (\dot{\boldsymbol{F}}_i^T \boldsymbol{l})^2} \tag{10}$$

where $\boldsymbol{v}(a) = \{\partial \boldsymbol{C}(a)/\partial a\}$ and $\dot{\boldsymbol{F}}_i = \{\partial F_i(\boldsymbol{w})\partial \boldsymbol{w}\}_{\boldsymbol{w}=\boldsymbol{w}_0}$.

Assume the influence curve $\{a, G_j(\boldsymbol{w}_0 - a\boldsymbol{l}, \hat{\boldsymbol{\beta}})\}^T$ is employed. Approximating $G_j(\boldsymbol{w}_0 - a\boldsymbol{l}, \hat{\boldsymbol{\beta}})$ by a first order Taylor's expansion evaluated at the null point $\boldsymbol{w}_0$, we have

$$G_j(\boldsymbol{w}, \hat{\boldsymbol{\beta}}) \approx a\boldsymbol{S}_{.j}^T\boldsymbol{l}. \tag{11}$$

Based on this approximation, the first order diagnostic is the vector of estimating function residuals for the j'th parameter.

To see the interrelation between estimating function residuals and standard jackknife statistics, let $\hat{\boldsymbol{\beta}}(\boldsymbol{w}_{(i)})$ be the perturbed parameter estimate evaluated at the vector $\boldsymbol{w}_{(i)}$, where every component of $\boldsymbol{w}_{(i)}$ is equal to 1 except the i'th component is equal to 0. By the first order approximation of $\hat{\boldsymbol{\beta}}(\boldsymbol{w}_{(i)}) - \hat{\boldsymbol{\beta}}$ evaluated at the null point $\boldsymbol{w}_0$ and by the implicit function theorem, we have

$$\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)} \approx \left\{-\frac{\partial \boldsymbol{G}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T}\right\}^{-1}_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \boldsymbol{S}_{i.}^T,$$

where $\hat{\boldsymbol{\beta}}_{(i)}$ is the parameter estimate after deletion of the i'th observation and

$$\boldsymbol{S}_{i.} = \{S_{i1}(\boldsymbol{y}_i, \hat{\boldsymbol{\beta}}), \ldots, S_{ip}(\boldsymbol{y}_N, \hat{\boldsymbol{\beta}})\}.$$

These interrelations among different influence measures also exist in nonparametric regression models and can be developed along the lines described in this paper.

The local influence method and estimating function residuals proposed in this paper are potentially useful extensions of both Cook's local influence diagnostics and score residuals. General local influence measures for assessing the effect of small perturbations of response vectors or explanatory variables can be developed along lines similar to what was done in Section 3. We have also described interrelations among first order diagnostics, score residuals and jackknife statistics. In our limited experience, these three diagnostic methods tend to identify the same influential observations especially when the data set is large and the orthogonality of the parameters is significant.

It might be computer-intensive to use the local influence method in a large data set because the method involves obtaining the eigenvalues and eigenvectors of an $N \times N$ matrix. Using estimating function residuals as a diagnostic tool can be less computer-intensive.

Similar results to those given in section 5.2 can be obtained when using smoothing splines if the degree of freedom of the smoother matrix (Hastie and Tibshirani, 1990, pp. 305-306) is chosen appropriately. However, when the degree of freedom of the smoother matrix changes, the influence of the observations may also change.

In this article, diagnostic procedures are carried out after the shrinkage or smoothing parameter is chosen. The influential points are thus "influential" for a fixed model. Nevertheless, the points might also influence the estimate of the smoothing parameter. In other word, the "influential" points are also related to model selection. The smoothing spline fits for the simulated data in section 5.2 are given in Figure 10. The solid line is the fitted curve for the complete data while the broken line is the fitted curve without the contributions from observations 7, 8, 17, and 18. The estimates of the smoothing parameters by generalized cross-validation (GCV, Wahba, 1990, chapter 4) corresponding to the fits are presented in the legend. The deletion of the four observations results in rapid changes of the smoothing parameter. Therefore, the "smoothness" has been changed dramatically. It would be worthwhile to develop influence diagnostics for identifying observations which dramatically effect model selection. The other concern is about the Bayesian confidence intervals (Wahba, 1990, chapter 5). The deletion of the four observations also results in significant changes of the widths of the Bayesian confidence intervals. It would be worthwhile to develop influence diagnostics for identifying observations which dramatically effect model selection as well as the width of the Bayesian confidence intervals.

REFERENCES

Belsley, D. A., Kuh, E. and Welsch, R. E. (1980). *Regression diagnostics : identifying influential data and source of collinearity.* New York: John Wiley.

Bruntz, S. M., Cleveland, W. S., Kleiner, B. and Warner, J. L. (1974) The dependence of ambient ozone on solar radiation, temperature, and mixing heights. In *Symposium on atmospheric diffusion and air pollution.* Boston; American Meteorological Society 125-8.

Chatterjee, S. and Hadi, A. S. (1988). *Sensitivity Analysis in Linear Regression.* New York: John Wiley.

Cleveland, W. S., Devlin, S. J. and Grosse, E. H. (1988) Regression by local fitting: methods, properties and computational algorithms. *Journal of Econometrics*, **37**, 87-114.

Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression.* New York: Chapman and Hall.

Cook, R. D. (1986) Assessment of local influence (with discussion). *J. R. Statist. Soc. B*, **48**, 133-169.

Draper, N. R. and Nostrand, R. C. V. (1979) Ridge regression and Jame-Stein estimation: review and comments. *Technometrics*, **21**, 451-466.

Eubank, R. L. and Gunst, R. F. (1986). Diagnostics for penalized least-squares estimators. *Statist. Probab. Lett.*, **4**, 256-272.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models.* New York: Chapman and Hall.

Hill, R. W. (1977). Robust regression when there are outliers in the carriers. Unpublished Ph.D. dissertation, Harvard University, Dept. of Statistics.

Laurent, R. T. ST. and Cook, R. D. (1993) Leverage, local influence and curvature in nonlinear regression. *Biometrika*, **80**, 99-106.

Lawrance, A. J. (1988) Regression transformation diagnostics using local influence. *J. Am. Statist. Ass.*, **83**, 1067-1072.

Myers, R. H. (1986). *Classical and Modern Regression With Applications.* Boston: Duxbury Press.

Pettitt, A. N. and Daud, I. B. (1989) Case-weighted measures of influence for proportional hazards regression. *Appl. Statist.*, **38**, 51-67.

Therneau, T. M., Grambsch, P. M. and Fleming, T. R. (1990) Martingale-based residuals for survival models. *Biometrika*, **77**, 147-169.

Thomas, W. and Cook, R. D. (1989) Assessing influence on regression coefficients in generalized linear models. *Biometrika*, **76**, 741-749.

Thomas, W. and Cook, R. D. (1990) Assessing influence on predictions from generalized linear models. *Technometrics*, **32**, 59-65.

Wahba, G. (1990) *Spline Models for Observational Data.* Philadelphia: Society for Industrial and Applied Mathematics.

Walker, E. and Birch, J. B. (1988). Influence measures in ridge regression. *Technometrics*, **30**, 221-227.

Wei, W. H. and Kosorok, M. R. (1995a). First and second order local influence diagnostics. Technical Report # 93, Department of Biostatistics, University of Wisconsin-Madison.

Wei, W. H. and Kosorok, M. R. (1995b). Residual diagnostics and local influence. Technical Report # 94, Department of Biostatistics, University of Wisconsin-Madison.

Wei, W. H. (1996) Residual diagnostics and local influence. Ph.D. dissertation, Dept. of Statistics, University of Wisconsin-Madison.

Wu, X. and Luo, Z. (1993) Second-order approach to local influence. *J. R. Statist. Soc. B*, **55**, 929-936.

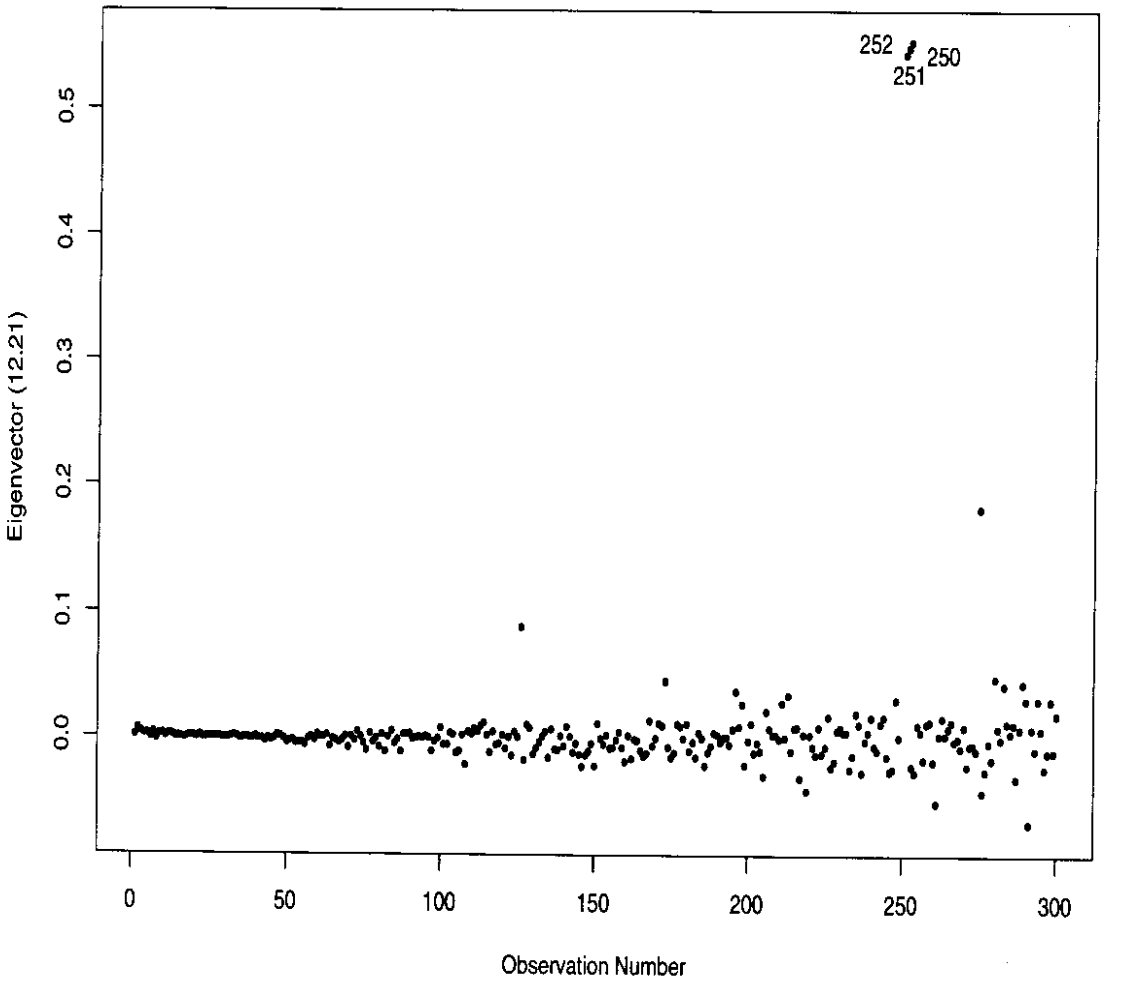

Figure 1: *Local influence method based on case weights for the ridge estimator. Observations 250, 251 and 252 stand out. The eigenvalue corresponding to the given eigenvector is presented in parentheses in the y-axis label.*

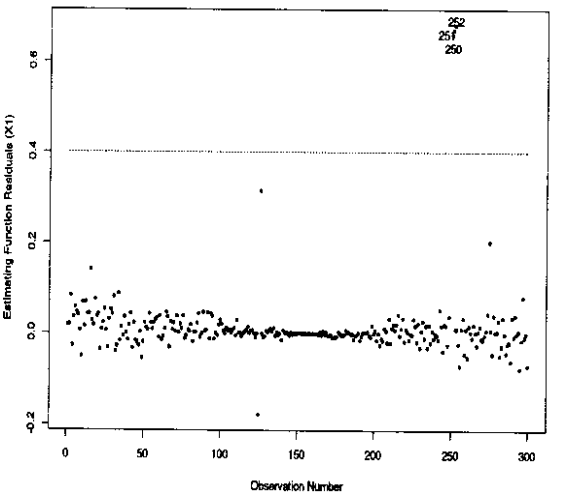

Figure 2: *Estimating function residuals corresponding to the covariate* X_1 *for the ridge estimator. Observations 250, 251 and 252 stand out.*

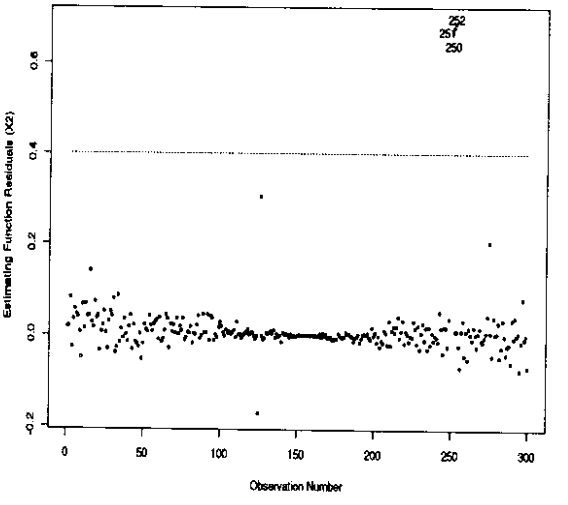

Figure 3: *Estimating function residuals corresponding to the covariate* X_2 *for the ridge estimator. Observations 250, 251 and 252 stand out.*

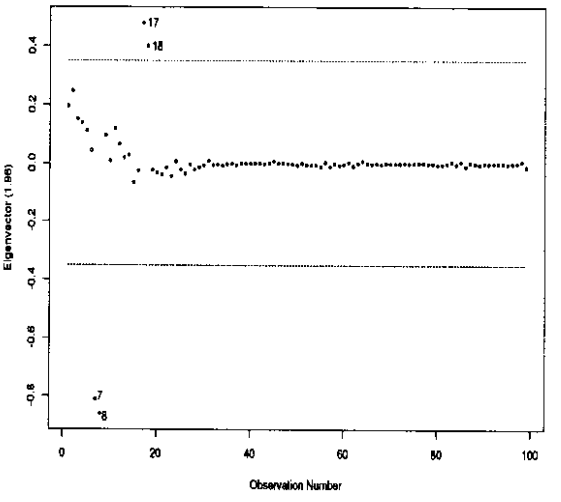

Figure 4: *Local influence method based on case weights for regression splines. Observations 7, 8, 17 and 18 stand out. The eigenvalue corresponding to the given eigenvector is presented in parentheses in the y-axis label.*

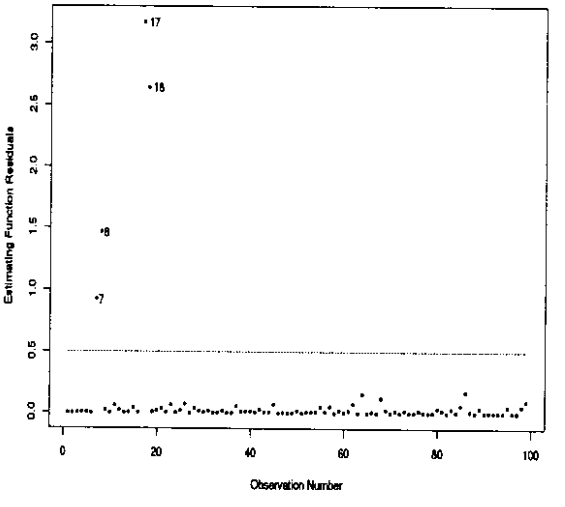

Figure 5: *Estimating function residuals for the ridge estimator. Observations 7, 8, 17 and 18 stand out.*

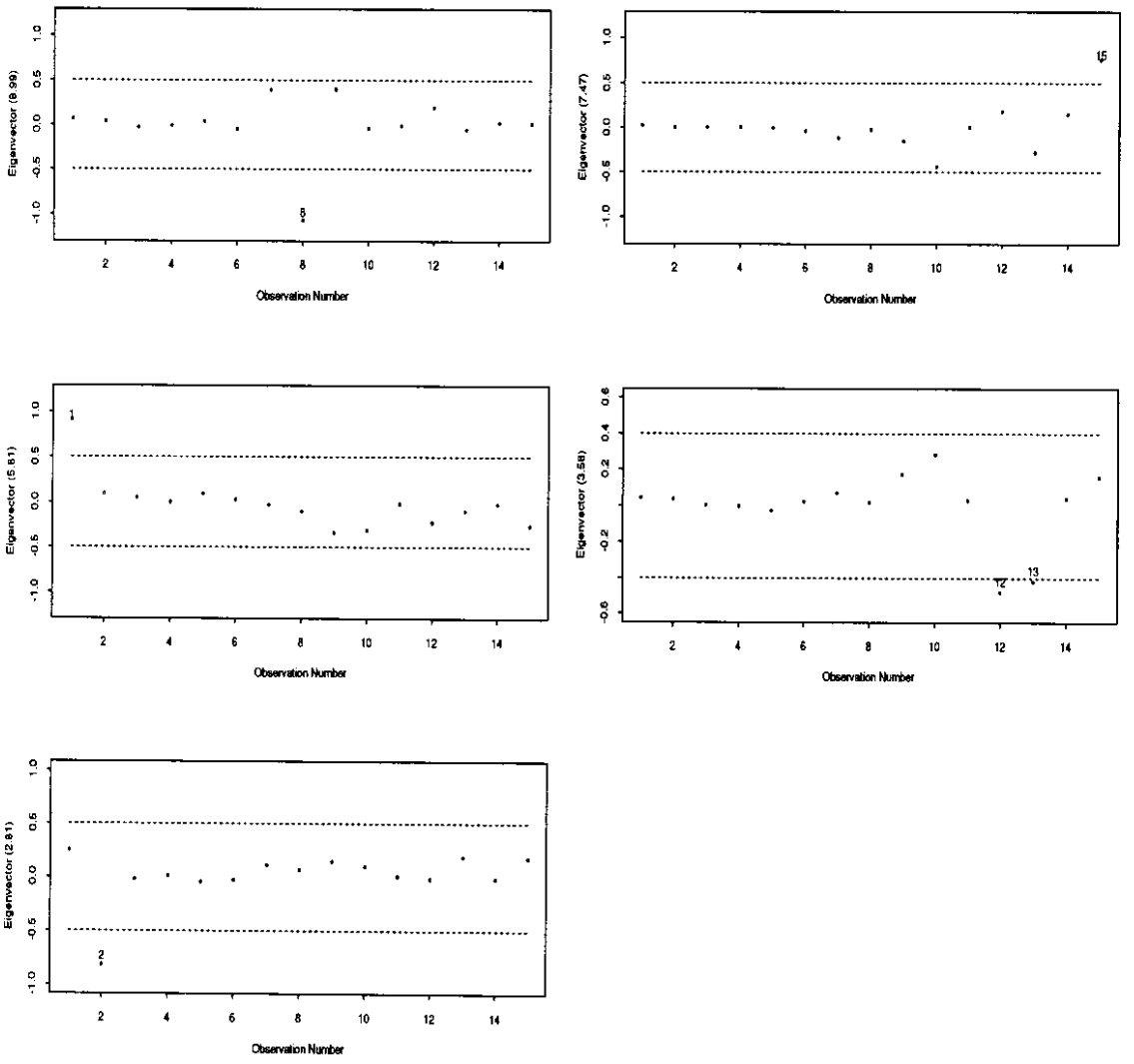

Figure 6: *Local influence method based on case weights for the ridge estimator. Observations 1, 2, 8, 12, 13 and 15 stand out. The eigenvalue corresponding to the given eigenvector is presented in parentheses in the y-axis label.*

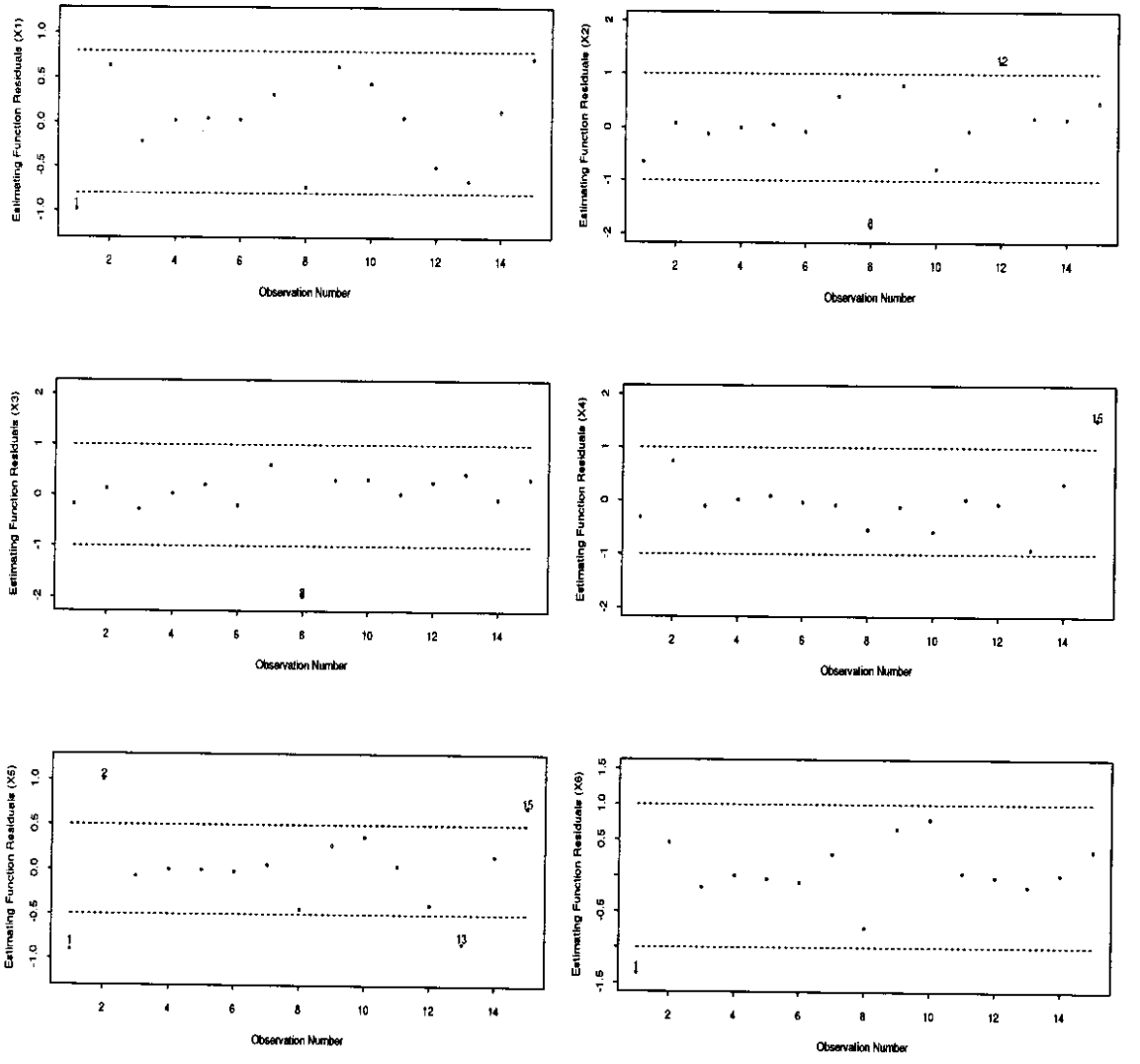

Figure 7: *Estimating function residuals for the ridge estimator. Observations 1, 2, 8, 12, 13, 15 stand out.*

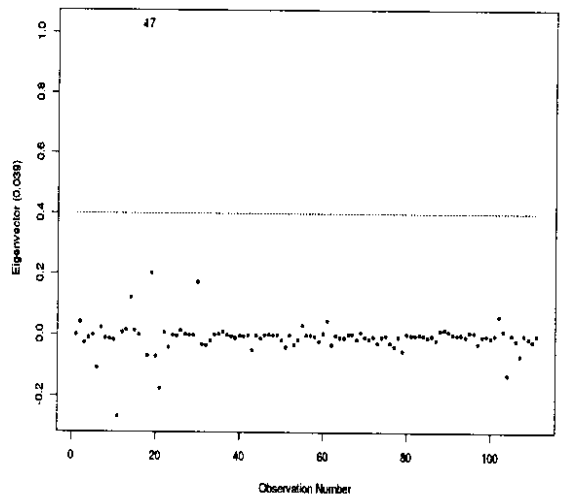

Figure 8: *Local influence method based on case weights for regression splines. Observation 17 stands out. The eigenvalue corresponding to the given eigenvector is presented in parentheses in the y-axis label.*

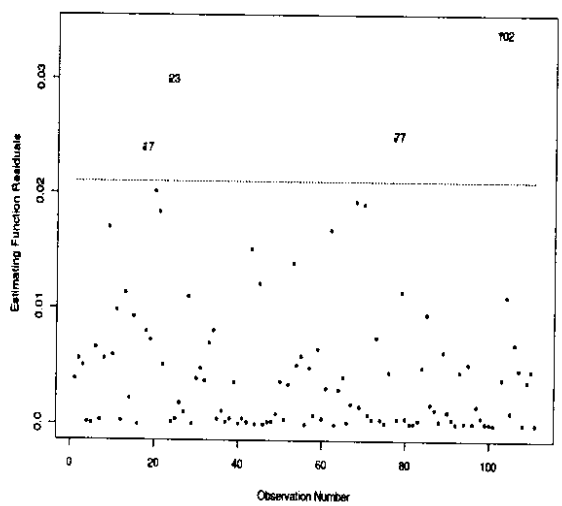

Figure 9: *Estimating function residuals for the ridge estimator. Observations 17, 23, 77 and 102 stand out.*

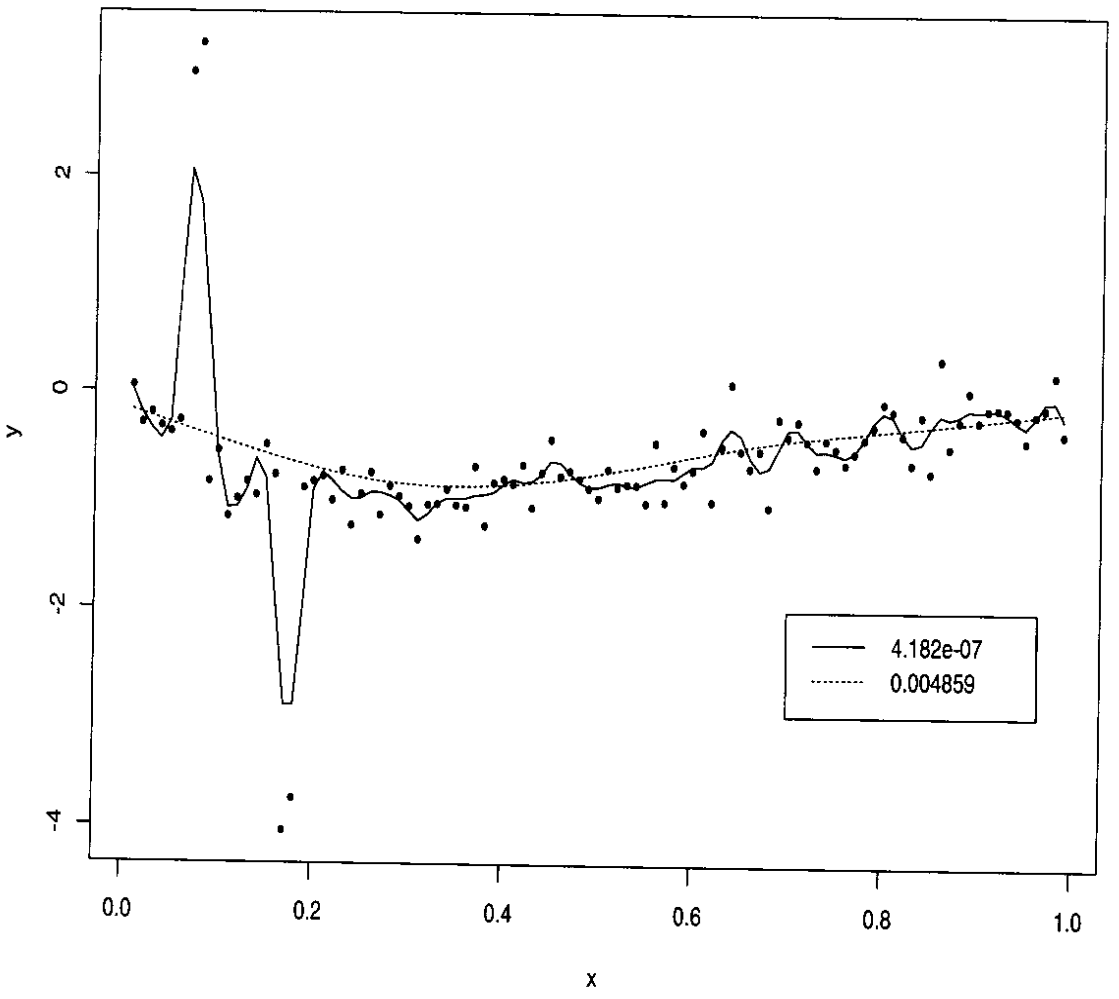

Figure 10: *The fitted curves with and without observations 7, 8, 17, and 18. The solid line is the curve with complete data while the dashed line is the curve without these observations. The estimates of the smoothing parameters are given in the legend.*