Ann Inst Stat Math (2009) 61:905–918
DOI 10.1007/s10463-007-0158-9

Local influence analysis for penalized Gaussian likelihood estimation in partially linear single-index models

Qingming Zou · Zhongyi Zhu · Jinglong Wang

Received: 21 December 2006 / Revised: 23 July 2007 / Published online: 27 October 2007

Abstract Single-index model is a potentially tool for multivariate nonparametric regression, generalizes both the generalized linear models(GLM) and the missing-link function problem in GLM. In this paper, we extend Cook's local influence analysis to the penalized Gaussian likelihood estimator based on P-spline for the partially linear single-index model. Some influence measures, based on the minor perturbation of the model, are derived for the penalized least squares estimation. An illustrative example is also presented.

Keywords Local influence · P-spline · Partially linear · Single-index model · Case-weight

1 Introduction

Influence diagnostics, including detecting outliers and influential observations, and the study of the sensitivity about the departure from basic assumption, have become a part of any serious statistical analysis. Based on case deletion, an important approach for

This work was supported by Natural Science Foundation of China Grants 10671038.

Q. Zou · J. Wang
Department of Statistics, East China Normal University, Shanghai 200062,
People's Republic of China

Q. Zou
College of Economics and Management, Nanhua University, Hengyang 421001,
People's Republic of China

Z. Zhu (✉)
Department of Statistics, Fudan University, Shanghai 200433, People's Republic of China
e-mail: zhuzy@fudan.edu.cn

assessing the impact of each observation on the parameter estimate was proposed by Cook (1977). Cook's distance has a clear interpretation and implication and has been well accepted in the statistics community. However, case deletion does not directly reflect the impact of other perturbation of the model. To supplement the case deletion approach, Cook (1986) developed a local influence approach based on the sensitivity of log-likelihood against small perturbation in part of the model. The local influence analysis does not involve recomputing the parameter estimates for every case deletion, so it is often computationally simpler. Furthermore, it permits perturbation of various aspects of the model to tell us more than what the case deletion approach is designed for. For example, it can help measure leverage of a design point and evaluate the impact of a small measurement error of predictor x on our estimates.

Following the pioneering work of Cook (1986), many authors have done a lot of work for a variety of models. This approach has been extended to generalized linear models by Thomas and Cook (1989), to restricted likelihood models by Kwan and Fung (1998), and to nonlinear models by St Laurent and Cook (1993). Lesaffre and Verbeke (1998) made a thorough investigation of local influence analysis in linear mixed models. Ouwens et al. (2001) extended the work to generalized linear mixed models. Thomas (1991) constructed local influence diagnostics for the smoothing parameter in smoothing spline. Lu et al. (1997) studied a standardized influence matrix that is related to Cook's local influence. Wang and Lee (1996) studied the sensitivity analysis for structural equation models with equality functional constraints. Others related work can be found in Shi and Wang (1999), Zhu et al. (2003) and Lee and Xu (2004). For the single-index models, very little has been done for the local influence analysis and the case deletion. In this paper, we demonstrate that the local influence analysis of Cook (1986) can be extended to the penalized Gaussian likelihood estimator in the partially linear single-index model.

The single-index model (Stoker 1986; Härdle and Stoker 1989; Li 1991; Ichimura 1993) is an important tool in multivariate nonparametric regression. It generalizes linear regression by replacing the linear combination $\alpha^{\mathrm{T}}X$ with a nonparametric component, $g(\alpha^{\mathrm{T}}X)$, where X is a vector of covariates, α is a parameter vector, $g(\cdot)$ is an unknown univariate link function. By reducing the dimensionality from multivariate predictors to a univariate index $\alpha^{\mathrm{T}}X$, the single-index model avoids the so-called "curse of dimensionality" and still captures important features in high–dimensional data. Applications of the single-index model lies in a variety of fields, such as discrete choice analysis in econometric and dose-response models in biometric (Härdle et al. 1993). The estimation problem of the single-index model has been discussed by many authors including Ichimura (1993), Härdle et al. (1993), Naik et al. (2000), Delecroix et al. (2006), Carroll et al. (1997) and Xia et al. (2006) and others. Yu and Ruppert (2002) derived the penalized least square estimators of parameters in the single-index model by appling penalized spline (P-spline) approach and discussed some asymptotic properties. P-spline (Ruppert and Carroll 2000; Ruppert 2002) is a generalization of smoothing splines allowing a more flexible choice of knots and penalty. For the single-index model, since P-spline can be fitted directly by penalized nonlinear least square, which leads to straight-forward computational algorithms and statistical inference. Therefore, P-spline affords us the convenience of applying Cook's local influence analysis to the single-index model. In this paper, we focus on the local influence of observation on P-spline least square estimators for the single-index model. The rest of the paper is organized as follows. In Sect. 2, we introduce the partially linear single-index model and review the P-spline estimator of Yu and Ruppert (2002). In Sect. 3, we generalize the local influence analysis to P-spline estimate, and some local influence measures based on the minor perturbation of model are also developed. In Sect. 4, the influence diagnostics are applied to an air pollution data. Some technical details are given in the appendix.

2 Models and estimation method

In this paper, our partially linear single-index model can be written as

$$y_i = g(\alpha_0^{\mathrm{T}} X_i) + \beta_0^{\mathrm{T}} Z_i + \varepsilon_i, \tag{1}$$

where $X_i \in R^d, Z_i \in R^{d_z}, y_i \in R, \alpha_0 \in R^d$ is an unknown single-index parameter, $\beta_0 \in R^{d_z}$ is an unknown linear parameter and $g(\cdot)$ is an unknown univariate link function; $\{\varepsilon_i\}$ is a mean 0 independent error with variance σ_0^2 and independent of $\{(X_i, Z_i)\}$; and $||\alpha_0|| = 1$ and the first nonzero element of α_0 is positive for identifiability.

For model (1), based on the idea of smoothing spline, Yu and Ruppert (2002) developed the estimation for unknown parameters by a P-spline and the estimating algorithm, and showed some asymptotic properties. Assumed that

$$g(u) = \delta_0 + \delta_1 u + \cdots + \delta_p u^p + \sum_{i=1}^{k} \delta_{p+i}(u - s_i)_+^p, \tag{2}$$

where $\{s_i\}_{i=1}^k$ are spline knots. The choice of the number of knots and the knot location can be referred to Yu and Ruppert (2002). Define the spline coefficient vector $\delta = (\delta_0, \delta_1, \ldots, \delta_{p+k})^{\mathrm{T}}$ and spline basis

$$B(u) = (1, u, \ldots, u^p, (u - s_1)_+^p, \ldots, (u - s_k)_+^p)^{\mathrm{T}}. \tag{3}$$

Then we have spline model $g(u) = \delta^{\mathrm{T}} B(u)$. Let $\theta = (\delta^{\mathrm{T}}, \alpha_0^{\mathrm{T}}, \beta_0^{\mathrm{T}})^{\mathrm{T}}$, the penalized least square estimator of θ, denoting $\hat{\theta}$, minimizes the penalized Gaussian likelihood

$$L_p(\theta) = n^{-1} \sum_{i=1}^{n} [y_i - \delta^{\mathrm{T}} B(\alpha_0^{\mathrm{T}} X_i) - \beta_0^{\mathrm{T}} Z_i]^2 + \lambda \delta^{\mathrm{T}} D \delta, \tag{4}$$

where D is an appropriate positive semidefinite symmetric matrix and $\lambda \geq 0$ is a penalty parameter. For example, D can be a positive semidefinite symmetric matrix such that

$$\delta^{\mathrm{T}} \mathrm{D} \delta = \int_{\min(\alpha_0^{\mathrm{T}} X_i)}^{\max(\alpha_0^{\mathrm{T}} X_i)} [g''(u)]^2 \mathrm{d}u,$$

which yields the usual quadratic integral penalty. If D is a diagonal with its last k diagonal elements equal to 1 and the rest equal to 0, it penalizes the sum of squares of the jumps in the pth degree of g. Yu and Ruppert (2002) discussed the choice of the knots and λ. They recommended that 5-10 knots should be quite adequate and the knots should be placed at equally spaced quantiles of the estimated index value, and the penalty parameter λ was selected by minimizing the GCV score. They have shown the consistency and asymptotic normality under the mild regularity conditions.

In this paper, our interest is to consider the impact of a small perturbation on the penalized Gaussian likelihood estimation of unknown parameters in the single-index model. The choice of the knots and the selection of penalty parameter are the same as those of Yu and Ruppert (2002).

3 Local influence analysis

According to the assumptions on model (1), it is obvious that the P-spline estimator is a constrained least squares estimator, with the constrained condition $||\alpha_0|| = 1$. So, the $\hat{\theta}$ should be regarded as the solution which minimizes $L_p(\theta)$ subject to $h(\theta) = ||\alpha_0|| - 1 = 0$. Thus the assessment of local influence of some possible model perturbations on the P-spline estimator should be done under this constrained condition.

3.1 General formula

According to above interpretation, it follows from Lagrange multiplier's method that there exists a real number r such that

$$\begin{cases} \dot{L}_p(\hat{\theta}) + r\dot{h}(\hat{\theta}) = 0, \\ h(\hat{\theta}) = 0, \end{cases}$$

where $\dot{L}_p(\hat{\theta}) = (\frac{\partial L_p}{\partial \theta_i}|_{\hat{\theta}})$ and $\dot{h}(\hat{\theta}) = (\frac{\partial h}{\partial \theta_i}|_{\hat{\theta}})$ are respectively the gradient vectors of $L_p(\theta)$ and $h(\theta)$ evaluated at $\hat{\theta}$. Suppose a minor perturbation ω is introduced to the model assumptions, where ω is an $m \times 1$ vector which varies in an open set Ω. Let $L_p(\theta|\omega)$ be the penalized Gaussian likelihood function corresponding to the minor perturbation and $\hat{\theta}(\omega)$ be the perturbed estimator of θ obtained by minimizing $L_p(\theta|\omega)$ subject to $h(\theta) = 0$. In addition, we assume that there exists an $\omega_0 \in \Omega$ such that $L_p(\theta|\omega_0) = L_p(\theta)$ for all θ. This assumption obviously implies that $\hat{\theta}_{\omega_0} = \hat{\theta}$. Similar to Cook's local influence analysis, we define

$$Q(\omega) = 2\{L_p(\hat{\theta}) - L_p(\hat{\theta}(\omega))\}$$

as a measure of the influence of the perturbation ω. It is clear that the $Q(\omega)$ contains the essential information about the influence of the minor perturbation scheme on the inference of θ. If $Q(\omega)$ is large for small $\omega \in \Omega$, then ω will lead to substantial changes in the results of the inference. Therefore, $d_{\max}$, the direction which makes the $Q(\omega)$

attain the greatest change on the lifted line $\omega = \omega_0 + td$, where $||d|| = 1$, is a statistics that we mainly concern in the context of local influence analysis.

To find $d_{\max}$, we first study the first- and second-partial derivatives of $Q(\omega)$ with respect to ω evaluated at ω_0, respectively. Then it follows that

$$\dot{Q} = \left.\frac{\partial Q(\omega)}{\partial \omega}\right|_{\omega_0} = -2\left(\frac{\partial \hat{\theta}(\omega)}{\partial \omega^{\mathrm{T}}}\right)^{\mathrm{T}} \left.\dot{L}_p\right|_{\omega_0}, \tag{5}$$

$$\ddot{Q} = \left.\frac{\partial^2 Q(\omega)}{\partial \omega \partial \omega^{\mathrm{T}}}\right|_{\omega_0} = -2\left[\left.\left(\frac{\partial \hat{\theta}(\omega)}{\partial \omega^{\mathrm{T}}}\right)^{\mathrm{T}} \ddot{L}_p \left(\frac{\partial \hat{\theta}(\omega)}{\partial \omega^{\mathrm{T}}}\right)\right|_{\omega_0} - \sum_{i=1}^{q} \frac{\partial^2 \hat{\theta}(\omega)_i}{\partial \omega \partial \omega^{\mathrm{T}}} \left.\dot{L}_{pi}\right|_{\omega_0}\right], \tag{6}$$

where $\ddot{L}_p = \left(\frac{\partial^2 L_p}{\partial \theta_i \partial \theta_j}\right)$(see appendix) is the Hessian matrix of $L(\theta)$, evaluated at $\hat{\theta}$, $\dot{L}_{pi}$ and $\hat{\theta}(\omega)_i$ are the ith component of $\dot{L}_p$ and $\hat{\theta}(\omega)$ respectively, and q is the dimension of θ.

Note that $\dot{L}_p = -r\dot{h}$, so in general $\dot{L}_p = 0$ does not hold with constraints, but we still have $\dot{Q} = 0$.

Since $\hat{\theta}(\omega)$ is the solution of minimizing $L_p(\theta|\omega)$ subject to $h(\theta) = 0$, hence

$$\frac{\partial L_p(\hat{\theta}(\omega)|\omega)}{\partial \theta} + r_\omega \frac{\partial h(\hat{\theta}(\omega))}{\partial \theta} = 0, \quad h(\theta(\omega)) = 0. \tag{7}$$

Differentiating both sides of the above equations with respect to ω and evaluating at ω_0, it follows that

$$\begin{bmatrix} \ddot{L}_p + r\ddot{h} & \dot{h} \\ \dot{h}^{\mathrm{T}} & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial \theta(\omega)}{\partial \omega^{\mathrm{T}}}|_{\omega_0} \\ \frac{\partial r_\omega}{\partial \omega^{\mathrm{T}}}|_{\omega_0} \end{bmatrix} = -\begin{bmatrix} \frac{\partial^2 L_p(\theta(\omega)|\omega)}{\partial \theta \partial \omega^{\mathrm{T}}}|_{\hat{\theta},\omega_0} \\ 0 \end{bmatrix}. \tag{8}$$

According to Lee and Bentler (1980), the coefficient matrix on the left side of (8) is nonsingular. Let

$$M = \begin{bmatrix} \ddot{L} + r\ddot{h} & \dot{h} \\ \dot{h}^{\mathrm{T}} & 0 \end{bmatrix}^{-1} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}, \tag{9}$$

then

$$\left.\frac{\partial \theta(\omega)}{\partial \omega^{\mathrm{T}}}\right|_{\omega_0} = -M_{11} G, \tag{10}$$

$$\left.\frac{\partial r_\omega}{\partial \omega^{\mathrm{T}}}\right|_{\omega_0} = -M_{21} G, \tag{11}$$

where $G = \frac{\partial^2 L_p(\theta(\omega)|\omega)}{\partial \theta \partial \omega^{\mathrm{T}}}|_{\hat{\theta},\omega_0}$.

Let $A = \ddot{L}_p + r\ddot{h} + \dot{h}\dot{h}^{\mathrm{T}}$, then following Lee and Bentler (1980), it is a $q \times q$ nonsingular matrix. Let $a = \dot{h}^{\mathrm{T}} A^{-1} \dot{h}$ and $P_{\dot{h}}$ be the generalized projection matrix of $\dot{h}$ with respect to A^{-1}:

$$P_{\dot{h}} = \frac{1}{a} \dot{h}\dot{h}^{\mathrm{T}} A^{-1},$$

then it can be verified that

$$M_{11} = A^{-1}(I_q - P_{\dot{h}}).$$

Thus, it follows from (5) and (10) that

$$\dot{Q} = 2G^{\mathrm{T}} A^{-1}(I_q - P_{\dot{h}})\dot{L}_p.$$

Note that $\dot{L}_p = -r\dot{h}$, so $\dot{Q} = 0$ holds.

Being $\dot{Q} = 0$, to search the $d_{\max}$, Cook (1986) suggested considering the influence graph to display the information of the perturbation. We define influence graph as following:

$$\eta(\omega) = \begin{bmatrix} \omega \\ Q(\omega) \end{bmatrix}. \tag{12}$$

Following Cook (1986), let $\dot{\eta}$ and $\ddot{\eta}$ be the first- and second-derivatives of $\eta = \eta(\omega_0 + td)$ with respect to t and evaluated at $t = 0$ respectively, where $||d|| = 1$. Then the influence curvature along direction d is defined as

$$C_d = \frac{||(\ddot{\eta})^N||}{||\dot{\eta}||^2}, \tag{13}$$

where $(\ddot{\eta})^N$ is a normal section of $\ddot{\eta}$, a projection to the normal space at ω_0. From the definition of $\eta(\omega)$, we have

$$\dot{\eta} = \begin{bmatrix} I_m \\ 0 \end{bmatrix} d, \quad \ddot{\eta} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} d^{\mathrm{T}} \ddot{Q} d.$$

Thus, the influence curvature is given by

$$C_d = |d^{\mathrm{T}} \ddot{Q} d|. \tag{14}$$

It is obvious that the influence curvature in (14) is very similar to that without constraint. However, here is $\ddot{Q}$ not $\ddot{L}_p$. From $h(\theta(\omega)) = 0$, we have

$$\left(\frac{\partial \hat{\theta}(\omega)}{\partial \omega^{\mathrm{T}}}\right)^{\mathrm{T}} \ddot{h} \left(\frac{\partial \hat{\theta}(\omega)}{\partial \omega^{\mathrm{T}}}\right)\Bigg|_{\omega_0} + \sum_{i=1}^{q} \frac{\partial^2 \hat{\theta}(\omega)_i}{\partial \omega \partial \omega^{\mathrm{T}}} \dot{h}_i \Bigg|_{\omega_0} = 0, \tag{15}$$

where $\ddot{h} = \frac{\partial^2 h(\theta)}{\partial \theta \partial \theta^{\mathrm{T}}}$ and is evaluated at $\hat{\theta}$. $\dot{h}_i$ is the ith component of $\dot{h}$.

By $\dot{L}_p = -r\dot{h}$, it follows that

$$\sum_{i=1}^{q} \frac{\partial^2 \hat{\theta}(\omega)_i}{\partial\omega\partial\omega^{\mathrm{T}}} \dot{L}_{pi} = r_\omega \left(\frac{\partial\hat{\theta}(\omega)}{\partial\omega^{\mathrm{T}}}\right)^{\mathrm{T}} \ddot{h} \left(\frac{\partial\hat{\theta}(\omega)}{\partial\omega^{\mathrm{T}}}\right). \tag{16}$$

From (6), (10) and (16), we have

$$\begin{aligned} \ddot{Q} &= -2\left[\left(\frac{\partial\hat{\theta}(\omega)}{\partial\omega^{\mathrm{T}}}\right)^{\mathrm{T}} (\ddot{L}_p + r\ddot{h}) \left(\frac{\partial\hat{\theta}(\omega)}{\partial\omega^{\mathrm{T}}}\right)\Big|_{\omega_0}\right] \\ &= -2G^{\mathrm{T}}\left(I_q - P_{\dot{h}}\right) A^{-1} \left(\ddot{L}_p - (\dot{h}^{\mathrm{T}}\dot{h})^{-1}\dot{h}^{\mathrm{T}}\dot{L}_p\ddot{h}\right) A^{-1}\left(I_q - P_{\dot{h}}\right) G. \end{aligned} \tag{17}$$

It has been pointed out that the direction $d_{\max}$ corresponding to the maximum value $C_{\max}$ of C_d is what we look for. The direction $d_{\max}$ indicates how to perturb the model to obtain the greatest local change. Computationally, $C_{\max}$ and $d_{\max}$ can be obtained by solving the following eigenvalue equation:

$$(-\ddot{Q} - \rho I_m)d = 0, \tag{18}$$

where $C_{\max} = \rho_{\max}$, $\rho_{\max}$ is the maximum eigenvalue and $d_{\max}$ is the associated eigenvector. From the above interpretations about $\ddot{Q}$, $C_{\max}$ and $d_{\max}$ can be finally computed by solving the following eigenvalue problem

$$\left[2G^{\mathrm{T}}(I_q - P_{\dot{h}})A^{-1}(\ddot{L}_p - (\dot{h}^{\mathrm{T}}\dot{h})^{-1}\dot{h}^{\mathrm{T}}\dot{L}_p\ddot{h})A^{-1}(I_q - P_{\dot{h}})G - \rho I_m\right]d = 0. \tag{19}$$

It is obvious that, in the local influence analysis with constraints, the $\ddot{Q}$ still plays an important role as in usual setting. Although the $\ddot{Q}$ often only relates to the likelihood displacement or the residue function in the setting with no constraint, when some constraints about parameters exist, the $\ddot{Q}$ can be decomposed into two parts. The first part of $\ddot{Q}$, $\ddot{L}$, is the contribution from the likelihood displacement or the residue function. The other is from constraints impact. If there is no constrain, the $\ddot{Q}$ reduces to $\ddot{L}$. Then we obtain the same results as in usual case.

It is well known that, for a nonnegative defined matrix, there exists a similar diagonal matrix whose diagonal components are the singular magnitudes of the matrix, and the singular value is the eigenvalue of the matrix. Therefore, to detect influential case in a data set, besides computing the C_{max} and d_{max} of an influence matrix, we can also look for the largest diagonal component of an influence matrix $\ddot{Q}$. In the below example, we can find this approach work well.

Remark 1 It is important to note that the influence matrix in this paper is conditional on the fixed knots and the penalty parameter λ. Yu and Ruppert (2002) pointed out that the fixed knots is appropriate to consistency and asymptotic normality for the P-spline least squares estimator. If we include the knots and the estimate of λ(obtained by GCV) as part of the local influence analysis, no closed form would be available. Consequently, the computation of $\ddot{Q}$, would be more demanding and difficult.

Remark 2 In this paper, we study the assessment of local influence for penalized Gaussian likelihood estimation in the partially linear single-index model under some restrictions and obtain the general formula of local influence analysis. Kwan and Fung (1998), and Wang and Lee (1996) also considered the influence diagnostics with some constrained conditions in other models. In some sense, there exist some common properties between our works and their works. For example, both of us follow the idea of Cook's local influence analysis and use the Lagrange multiplier's method to obtain the general formula. However, there are also some distinct divergence. The likelihood displacements defined in our works are different. We define the likelihood displacement by the P-spline penalized Gaussian likelihood. In addition, we firstly show that, since the first-derivative of likelihood displacement is zero, so the curvature of the influence graph should be considered. Finally, our interest is to investigate the impact of minor perturbations on the parametric estimation and the nonparametric estimation in a partially linear single-index model.

3.2 Perturbation schemes

We have obtained the general result for local influence analysis for penalized least-square estimator in the partially single-index model via the general perturbation scheme. In practice, different perturbation schemes should be considered based on the investigator's special concerns. From the general result, we know that the matrix $G = \frac{\partial^2 L_p(\theta(\omega)|\omega)}{\partial\theta\partial\omega^{\mathrm{T}}}|_{\hat{\theta},\omega_0}$ is a pivotal quantity in the calculation of C_{max} and d_{max}. In the following, G corresponding to different perturbation scheme is given.

Scheme 1 Case weight perturbation. We first perturb the data by modifying the weight given to each case in the least squares criterion. This is equivalent to perturb the variance of ε_i in the model. Let $\omega = (\omega_1, \ldots, \omega_n)^{\mathrm{T}}$ be a perturbation scheme such that $\omega_0 = (1, \ldots, 1)^{\mathrm{T}}$. Assigning weights ω_i to the ith case, we have

$$L_p(\theta(\omega)|\omega) = \frac{1}{n}\sum_{i=1}^{n} \omega_i \left[y_i - \delta^{\mathrm{T}} B(\alpha_0^{\mathrm{T}} X_i) - \beta_0^{\mathrm{T}} Z_i\right]^2 + \lambda\delta^{\mathrm{T}} D\delta. \tag{20}$$

Direct calculation yields

$$G = \left.\frac{\partial^2 L_p(\theta(\omega)|\omega)}{\partial\theta\partial\omega^{\mathrm{T}}}\right|_{\hat{\theta},\omega_0} = -\frac{2}{n} U \operatorname{diag}(e_1, \ldots, e_n), \tag{21}$$

where

$$U = \begin{bmatrix} B(\hat{\alpha}_0^{\mathrm{T}} X_1) & \cdots & B(\hat{\alpha}_0^{\mathrm{T}} X_n) \\ \hat{\delta}^{\mathrm{T}} \dot{B}(\hat{\alpha}_0^{\mathrm{T}} X_1) X_1 & \cdots & \hat{\delta}^{\mathrm{T}} \dot{B}(\hat{\alpha}_0^{\mathrm{T}} X_n) X_n \\ Z_1 & \cdots & Z_n \end{bmatrix},$$

$e_i = y_i - \hat{\delta}^{\mathrm{T}} B(\hat{\alpha}_0^{\mathrm{T}} X_i) - \hat{\beta}_0^{\mathrm{T}} Z_i$, $\dot{B}(\hat{\alpha}_0^{\mathrm{T}} X_i) = \frac{dB(u)}{du}|_{u=\hat{\alpha}_0^{\mathrm{T}} X_i}$. In fact, e_i is the ith component of the residual vector. From the above formula, we know that if $\omega_i = 0$, and $\omega_j = 1$ for $j \neq i$, this perturbation scheme is reduced to the case-deletion approach.

Scheme 2 Perturbation of response. We now consider the perturbation of the response variable so that Y is replaced by $Y + \omega$, where $\omega \in R^n$. In this case,

$$L_p(\theta(\omega)|\omega) = \frac{1}{n}\sum_{i=1}^{n}[\omega_i + y_i - \delta^{\mathrm{T}} B(\alpha_0^{\mathrm{T}} X_i) - \beta_0^{\mathrm{T}} Z_i]^2 + \lambda\delta^{\mathrm{T}} D\delta. \tag{22}$$

It follows that

$$G = \frac{\partial^2 L_p(\theta(\omega)|\omega)}{\partial\theta\partial\omega^{\mathrm{T}}}|_{\hat{\theta},\omega_0} = -\frac{2}{n}U. \tag{23}$$

Scheme 3 Perturbation of covariates. Perturbation of covariates has a more complicate impact on the estimates. It is well known that measurement errors on the covariates can result in serious bias in the estimation of linear regression coefficients. Very few have been done for the bias issue in the partially linear single-index model. The local influence analysis under perturbation of covariates may provide an alternative view to measurement error models. In partially linear single-index models, there are two kinds of perturbation schemes, one is the perturbation of single-index covariates X and the other is that of covariates Z in the linear part.

Consider perturbing X_i to $X_i + l_j\omega^{\mathrm{T}} k_i$, where $\omega \in R^n$, $l_j \in R^d$, $k_i \in R^n$ and l_j, k_i are unit vectors with the jth and ith elements equal to 1 respectively. It means that only the jth covariate is being perturbed. In this case,

$$L_p(\theta(\omega)|\omega) = \frac{1}{n}\sum_{i=1}^{n}[y_i - \delta^{\mathrm{T}} B(\alpha_0^{\mathrm{T}}(X_i + l_j\omega^{\mathrm{T}} k_i)) - \beta_0^{\mathrm{T}} Z_i]^2 + \lambda\delta^{\mathrm{T}} D\delta. \tag{24}$$

Direct calculation yields

$$\begin{cases} \dfrac{\partial^2 L}{\partial\alpha_0\partial\omega^{\mathrm{T}}}\Big|_{\hat{\theta},\omega_0} = -\frac{2}{n}\sum_{i=1}^{n} e_i b_i l_j k_i^{\mathrm{T}} - \frac{2}{n}\sum_{i=1}^{n}(e_i c_i - b_i^2) X_i \hat{\alpha}_0^{\mathrm{T}} l_j k_i^{\mathrm{T}}, \\ \dfrac{\partial^2 L}{\partial\beta_0\partial\omega^{\mathrm{T}}}\Big|_{\hat{\theta},\omega_0} = -\frac{2}{n}\sum_{i=1}^{n} b_i Z_i l_j k_i^{\mathrm{T}}, \\ \dfrac{\partial^2 L}{\partial\delta\partial\omega^{\mathrm{T}}}\Big|_{\hat{\theta},\omega_0} = -\frac{2}{n}\sum_{i=1}^{n} e_i \dot{B}(\hat{\alpha}_0^{\mathrm{T}} X_i)\hat{\alpha}_0^{\mathrm{T}} l_j k_i^{\mathrm{T}} + \frac{2}{n}\sum_{i=1}^{n} b_i B(\hat{\alpha}_0^{\mathrm{T}} X_i)\hat{\alpha}_0^{\mathrm{T}} l_j k_i^{\mathrm{T}}, \end{cases} \tag{25}$$

where $b_i = \hat{\delta}^{\mathrm{T}}\dot{B}(\hat{\alpha}_0^{\mathrm{T}} X_i)$, $c_i = \hat{\delta}^{\mathrm{T}}\ddot{B}(\hat{\alpha}_0^{\mathrm{T}} X_i)$ and $\ddot{B}(\hat{\alpha}_0^{\mathrm{T}} X_i) = \frac{d^2B(u)}{du^2}|_{u=\hat{\alpha}_0^{\mathrm{T}} X_i}$.

Now consider perturbing Z_i to $Z_i + l_j\omega^{\mathrm{T}} k_i^{\mathrm{T}}$, where $\omega \in R^n$ is a perturbation scheme, $l_j \in R^{d_z}$, $k_i \in R^n$ are unit vectors with the jth and ith elements equal to 1 respectively. Then we have

$$L_p(\theta(\omega)|\omega) = \frac{1}{n}\sum_{i=1}^{n}[y_i - \delta^{\mathrm{T}} B(\alpha_0^{\mathrm{T}} X_i) - \beta_0^{\mathrm{T}}(Z_i + l_j\omega^{\mathrm{T}} k_i)]^2 + \lambda\delta^{\mathrm{T}} D\delta. \tag{26}$$

It can be shown that

$$\begin{cases} \left.\dfrac{\partial^2 L}{\partial \alpha_0 \partial \omega^{\mathrm{T}}}\right|_{\hat{\theta},\omega_0} = -\frac{2}{n}\sum_{i=1}^{n} b_i X_i \hat{\beta}_0^{\mathrm{T}} l_j k_i^{\mathrm{T}}, \\ \left.\dfrac{\partial^2 L}{\partial \beta_0 \partial \omega^{\mathrm{T}}}\right|_{\hat{\theta},\omega_0} = -\frac{2}{n}\sum_{i=1}^{n} e_i l_j k_i^{\mathrm{T}} + \frac{2}{n} Z_i \hat{\beta}_0^{\mathrm{T}} l_j k_i^{\mathrm{T}}, \\ \left.\dfrac{\partial^2 L}{\partial \delta \partial \omega^{\mathrm{T}}}\right|_{\hat{\theta},\omega_0} = -\frac{2}{n}\sum_{i=1}^{n} B(\hat{\alpha}_0^{\mathrm{T}} X_i) \hat{\beta}_0^{\mathrm{T}} l_j k_i^{\mathrm{T}}. \end{cases} \tag{27}$$

From the above conclusion about $\frac{\partial^2 L_p(\theta(\omega)|\omega)}{\partial\theta\partial\omega^{\mathrm{T}}}$, $\ddot{L}$, $\dot{h}$ and $\ddot{h}$, we can obtain our interested quantity $\ddot{Q}$. Then the influential observations for P-spline least squares estimator can be detected by solving the eigenvalue problem.

4 Illustrative example

This section gives a numerical example through which we illustrate how the local influence measures given in Sect. 3 can assist diagnostics in the partially linear single-index model. we plot the diagonal elements of the influence matrices $\ddot{Q}$ to assess the perturbation case-weight, response variable and covariate variable.

The air pollution data set was obtained from an environmental study to find how the concentration y of the air pollution ozone depends on three meteorological variables, X; wind speed, x_1; temperature, x_2; and radiation, z. The data are daily measurements of the four variables for $n = 111$ days (Härdle et al. 1993). Yu and Ruppert (2002) used this data to fit several models with reduced dimension, and concluded in their analysis that partially linear single-index model using a P-spline in which temperature and wind are the two components of the index and radiation is the linear term is appropriate. They fitted a cubic P-spline with the penalty parameter value $\hat{\lambda}$ selected by minimizing the GCV score over a grid of values of λ. They used the 30-point grid where the values of $\log_{10}(\lambda)$ are equally space between -6 and 7. According to their suggestion, we use the penalized least squares with $\lambda = 4.182$ chosen by GCV and obtain single-index coefficients and linear regression coefficient estimates: $\hat{\alpha}_{01} = 0.5450$ (SE $= 0.0069$), $\hat{\alpha}_{02} = -0.8385$ (SE $= 0.0045$) and $\hat{\beta}_0 = 0.0024$ (SE $= 0.00006$).

Basing on this estimation, the proposed diagnostic procedures can be used to identify the influential observations in this data set. We compute the local influence matrices $\ddot{Q}$ corresponding the variety of perturbations, use $\ddot{Q}_w$, $\ddot{Q}_y$, $\ddot{Q}_z$, and $\ddot{Q}_{x_1}$ to denote influence matrices of perturbation of case weight, response and covariates and plot the diagonal elements of those matrices in Fig. 1. We see from Fig. 1 that case number 80 is the most influential datum in the sample for penalized least square estimates under case-weight, response and covariates perturbations. To detect whether the case 80 is an outlier, according to the analysis of Yu and Ruppert (2002), we plot the Studentized residuals in Fig. 2 and find that case 80 has the smallest residual. Therefore, we can conclude that case 80 is an influential observation but not an outlier. A closer inspection finds that case number 80 is a rather extreme point in this data set with

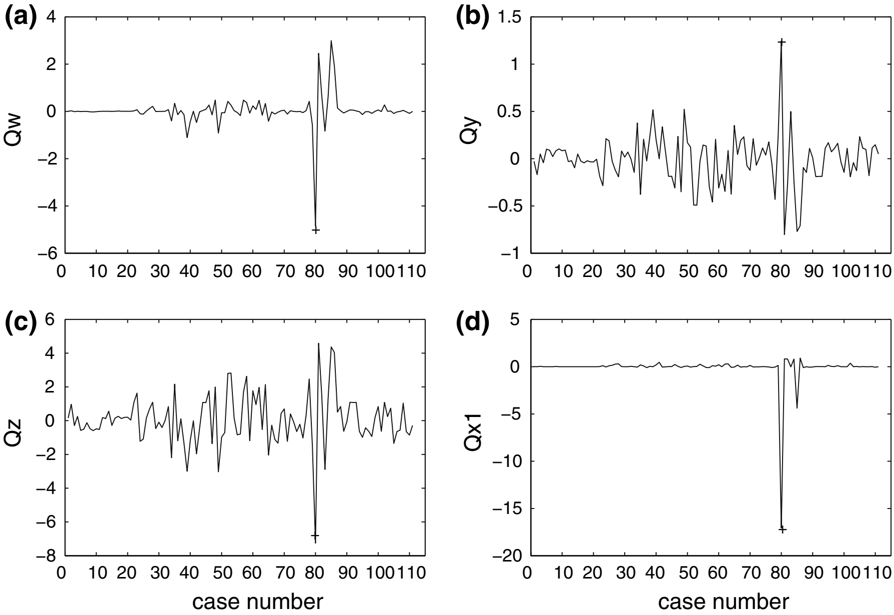

Fig. 1 Local influence analysis for the air pollution data. Diagonal elements of **a** $\ddot{Q}_w$, **b** $\ddot{Q}_y$, **c** $\ddot{Q}_z$, **d** $\ddot{Q}_{x_1}$, representing case-wise contribution to influence matrices

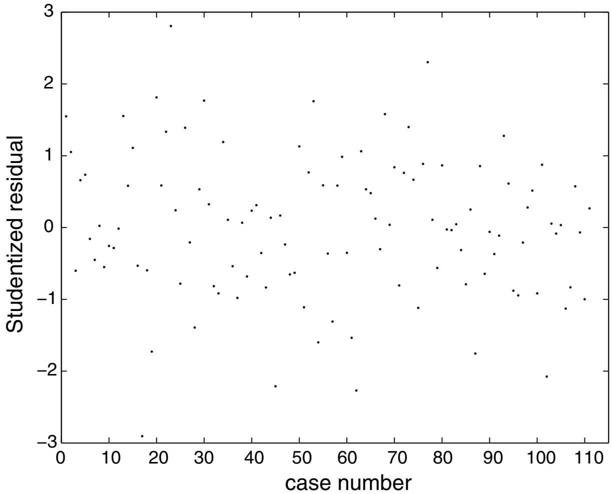

Fig. 2 Studentized residuals for the air pollution data

much larger concentration y, the temperature x_1 and the radiation z, and the smallest wind speed x_2. It means that the removal of this case would have a large impact on the parameter estimates. In addition, from Fig. 1a–c we can find that case 81 and case 85 have also large influence but they are not outliers.

Remark 3 To investigate the influence of case 80 on the estimation of unknown parameters, we delete case 80 in the air pollution data set, and then use the same algorithm program as before. We obtain single-index coefficients and linear regression coefficient estimates as $\hat{\alpha}_{01} = 0.6892$ (SE $= 0.0127$), $\hat{\alpha}_{02} = -0.7268$ (SE $= 0.0115$) and $\hat{\beta}_0 = 0.0021$ (SE $= 0.0005$) respectively. Moreover, the magnitudes of δ and λ have great changes, for example, $\lambda = 49857$. Therefore, the above diagnostic procedure of influential for the penalized Gaussian likelihood estimation in partially linear single-index model is feasible and applicable.

Remark 4 In our local influence study, the penalty parameter λ is fixed, one of nature concern is the sensitivity of the results against some perturbation of λ. We repeated the analysis using $\lambda = 2\hat{\lambda} = 8.364$ in this example. We also plot the diagonal elements of the influence $\ddot{Q}$ corresponding various perturbations. For the sake of space, those figures are omitted. We find that these figures have little difference with those in Fig. 1. The same results can be obtained using $\lambda = \hat{\lambda}/2$. The different choices of penalty parameter λ seem to have little effects on local influence analysis.

5 Discussion

In this paper, we extend the work of Cook (1986) to provide local influence measures under perturbation of case weight, response and covariate for the partially penalized Gaussian likelihood estimators in a partially linear single-index model. Due to the complexity of the partially linear single-index model, it is difficult to obtain local influence measures directly by Cook's (1986) approach. Thus we investigate the local influence analysis of a P-spline model with equality constraints for partially linear single-index model. The results in this paper have shown that the procedure is practically feasible.

As we point out before, detecting outliers is also an important issue in data analysis. After the identification of these influential observations, further steps are included to test whether they are actually outliers. Case-deletion approach is widely used to find the outliers. However, it is difficult to detect outliers for partially linear single-index model by case-deletion and further research is required.

Finally, we note that we have only used the diagonal elements of the influence matrices in our example. This is partly for convenience and partly for data without clusters of influential points. According to a referee's suggestions, we also investigate the eigenvector of the influence matrix in the illustrative example and find that there are little difference between the two approaches. Each component of the eigenvector related to case 80 is larger than the corresponding component of other eigenvectors. The eigenvectors of the influence matrix also show that there are no clusters of influential points. Like the delete-one diagnostics, this approach can suffer from masking. The eigenvectors that correspond to some of the largest eigenvalues of a influence matrix would be helpful in identifying batches of influential observations.

Acknowledgements The authors thanks the editor and two anonymous referees for their careful reading of the paper and constructive comments.

Appendix A: Expression for $\ddot{L}$

Let $Y = (y_1, \ldots, y_n)^{\mathrm{T}}$, $B = (B(\alpha_0^{\mathrm{T}} X_1), \ldots, B(\alpha_0^{\mathrm{T}} X_n))^{\mathrm{T}}$, $Z = (Z_1, \ldots, Z_n)^{\mathrm{T}}$, $X = (X_1, \ldots, X_n)^{\mathrm{T}}$, $\theta = (\delta^{\mathrm{T}}, \alpha_0^{\mathrm{T}}, \beta_0^{\mathrm{T}})^{\mathrm{T}}$, then $L_p(\theta)$ can be expressed as

$$L_p(\theta) = \frac{1}{n}||Y - B\delta - Z\beta_0||^2 + \lambda\delta^{\mathrm{T}} D\delta.$$

Therefore, the first-and second-order derivatives of L_p are given as follows:

$$\begin{aligned}
\frac{\partial L_p}{\partial \delta} &= -\frac{2}{n} B^{\mathrm{T}}(Y - B\delta - Z\beta_0) + 2\lambda D\delta, \\
\frac{\partial L_p}{\partial \alpha_0} &= -\frac{2}{n} X \mathrm{diag}(\delta^{\mathrm{T}} \dot{B}(\alpha_0^{\mathrm{T}} X_1), \ldots, \delta^{\mathrm{T}} \dot{B}(\alpha_0^{\mathrm{T}} X_n))(Y - B\delta - Z\beta_0), \\
\frac{\partial L_p}{\partial \beta_0} &= -\frac{2}{n} Z^{\mathrm{T}}(Y - B\delta - Z\beta_0), \\
\frac{\partial^2 L_p}{\partial \delta \partial \delta^{\mathrm{T}}} &= \frac{2}{n} B^{\mathrm{T}} B + 2\lambda D, \\
\frac{\partial^2 L_p}{\partial \delta \partial \alpha_0^{\mathrm{T}}} &= -\frac{2}{n} (\dot{B}(\alpha_0^{\mathrm{T}} X_1), \ldots, \dot{B}(\alpha_0^{\mathrm{T}} X_n)) \mathrm{diag}(e_1, \ldots, e_2) X^{\mathrm{T}}, \\
\frac{\partial^2 L_p}{\partial \delta \partial \beta_0^{\mathrm{T}}} &= \frac{2}{n} B^{\mathrm{T}} Z, \\
\frac{\partial^2 L_p}{\partial \alpha_0 \partial \alpha_0^{\mathrm{T}}} &= -\frac{2}{n} X[\mathrm{diag}(c_1, \ldots, c_2)\mathrm{diag}(e_1, \ldots, e_2) - \mathrm{diag}(b_1^2, \ldots, b_n^2)] X^{\mathrm{T}}, \\
\frac{\partial^2 L_p}{\partial \alpha_0 \partial \beta_0^{\mathrm{T}}} &= \frac{2}{n} X \mathrm{diag}(b_1, \ldots, b_n) Z, \\
\frac{\partial^2 L_p}{\partial \beta_0 \partial \beta_0^{\mathrm{T}}} &= \frac{2}{n} Z^{\mathrm{T}} Z,
\end{aligned}$$

where $e_i = y_i - \delta^{\mathrm{T}} B(\alpha_0^{\mathrm{T}} X_i) - \beta_0^{\mathrm{T}} Z_i$, $b_i = \delta^{\mathrm{T}} \dot{B}(\alpha_0^{\mathrm{T}} X_i)$, $c_i = \delta^{\mathrm{T}} \ddot{B}(\alpha_0^{\mathrm{T}} X_i)$, $\dot{B}(\alpha_0^{\mathrm{T}} X_i) = \frac{dB(u)}{du}|_{u=\alpha_0^{\mathrm{T}} X_i}$, $\ddot{B}(\alpha_0^{\mathrm{T}} X_i) = \frac{d^2 B(u)}{du^2}|_{u=\alpha_0^{\mathrm{T}} X_i}$.

References

Carroll, R. J., Fan, J., Gijbels, I., Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, *92*, 477–489.

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, *19*, 15–18.

Cook, R. D. (1986). Assessment of local influence (with discussion). *Journal of Royal Statistical Society B, 48*, 133–169.

Delecroix, M., Hristache, M., Patilea, V. (2006). On semiparametric M-estimation in single-index regression. *Journal of Statatical Planing and Inference*, *136*, 730–769.

Hastie, T., Tibshirani, R. (1993). Varying-coefficient models. *Journal of Royal Statistical Society B, 55*, 757–796.

Hărdle, W., Hall, P., Ichimura, H. (1993). Optimal smoothing in single-index models. *The Annals of Statistics*, *21*, 157–178.

Hărdle, W., Stoker, T. M. (1989). Investing smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association*, *84*, 986–995.

Ichimura, H. (1993). Semiparametric least square (SLS) and werighted SLS estimation of single-index models. *Journal of Econometrics*, *58*, 71–120.

Kwan, C. W., Fung, W. K. (1998). Assessing local influence for specific restricted likelihood: application to factor analysis. *Psychometrika*, *63*, 35–46.

Lee, S. Y., Bentler, P. M. (1980). Some asymptotic properties of constrained generalized least squares estimation in covariance structure models. *South African Statistics*, *14*, 121–136.

Lee, S. Y., Xu, L. (2004). Influence analysis of nonlinear mixed-effects models. *Computational Statistics & Data Analysis*, *45*, 321–341.

Lesaffre, E., Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics*, *54*, 570–582.

Li, K. C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, *86*, 316–342.

Lu, J., Ko, D., Chang, T. (1997). The standardized influence matrix and its application. *Journal of the American Statistical Association*, *92*, 1572–1580.

Naik, P., Tsai, C. L. (2000). Partial least squares estimator for single-index. *Journal of Royal Statistical Society B, 62*, 763–771.

Ouwens, M. J. N. M., Tan, F. E. S., Berger, M. P. F. (2001). Local influence to detect influential data structures for generalized linear mixed models. *Biometrics*, *54*, 1166–1172.

Ruppert, D. (2002). Selecting the number knots for penalized splines. *Journal of Computational and Graphical Statistics*, *11*, 735–757.

Shi, L., Wang, X. R. (1999). Local influence in ridge regression. *Computational Statistics & Data Analysis*, *31*, 341–353.

St Laurent, R. T., Cook, R. D. (1993). Leverage, local influence and curvature in nonlinear regression. *Biometrika*, *60*, 99–106.

Stoker, T. M. (1986). Consistent estimation of scaled coefficients. *Econometrica*, *54*, 1461–1481.

Thomas, W. (1991). Influence diagnostics for the cross-validated smoothing parameter in spline smoothing. *Journal of the American Statistical Association*, *86*, 693–698.

Thomas, W., Cook, R. D. (1989). Assessing influence on regreesion coefficient in generalized linear models. *Biometrika*, *76*, 741–749.

Wang, S. J., Lee, S. Y. (1996). Sensitivity analysis of structural equation models with equalty functional constraints. *Computational Statistics & Data Analysis*, *23*, 239–256.

Xia, Y.C., Hărdle, W. (2006). Semi-parametric estimation of partially linear single-index models. *Journal of Multivariate Analysis*, *97*, 1162–1184.

Yu, Y., Ruppert, D. (2002). Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association*, *97*, 1042–1054.

Zhu, Z. Y., He, X., Fung, W. K. (2003). Local influence analyis for penalized gaussian likelihood estimatiors in partially linear models. *Scandinavian Journal of Statatics*, *30*, 767–780.