Tatra Mt. Math. Publ. **17** (1999), 65–70

TATRA MOUNTAINS
Mathematical Publications

ON ADDITIVE PERTURBATIONS IN NONLINEAR REGRESSION MODELS

RASTISLAV POTOCKÝ — TO VAN BAN

ABSTRACT. A method is proposed of how to measure the influence of additive perturbations on parameter estimators in nonlinear regression models. The relationship between the new measure and the intrinsic curvature is explored.

1. Introduction

Consider the usual nonlinear regression model

$$y_a = \mathbf{f}(\mathbf{x}_a, \boldsymbol{\theta}) + e_a \qquad a = 1, \dots, n. \tag{1}$$

Here, $\mathbf{Y} = (y_1, \dots, y_n)'$ denotes the vector of observations, the function $\mathbf{f}(\boldsymbol{\theta}) = \big(f(\mathbf{x}_1, \boldsymbol{\theta}), \dots, f(\mathbf{x}_n, \boldsymbol{\theta})\big)$ has a known form dependent on p unknown parameters $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)'$, $\mathbf{x}_a$ are known vector-valued variables, and e_a are independent and normally distributed random errors with zero mean and variance σ^2.

The problem of finding acceptable confidence regions for $\boldsymbol{\theta}$ has been discussed by Potocký and To Van Ban [4], [5]. An interesting approach can be found in Kubáček [2]. The distribution of $\boldsymbol{\theta}^*$, the least squares estimator of $\boldsymbol{\theta}$, has been obtained by Pázman [3].

Suppose we modify $\mathbf{Y}$ by the addition of a vector $\mathbf{u}$. Let $L(\boldsymbol{\theta})$ denote the log-likelihood corresponding to the postulated model and $L(\boldsymbol{\theta}/\mathbf{u})$ the log-likelihood corresponding to the perturbed model, i.e.,

$$L(\boldsymbol{\theta}/\mathbf{u}) = -n/2 \ln(2\pi\sigma^2) - 1/2\sigma^2 \big\| \mathbf{Y} + \mathbf{u} - \mathbf{f}(\boldsymbol{\theta}) \big\|^2 .$$

Laurent and Cook [6], [7] have shown that the so-called generalized leverage vector and Jacobian leverage vector can be used to assess the influence of perturbations on fitted values in nonlinear regression. Our aim is to give a

2000 Mathematics Subject Classification: Primary 62J02; Secondary 62F99.
Keywords: parameter leverage, intrinsic curvature.

This research was supported by a grant VEGA 1/4196/97.

tool for the assessment of the influence of additive perturbations on the values of parameters themselves. Then the relationship between the new measure, the leverages introduced by Laurent and Cook and the intrinsic curvature is explored.

2. Parameter leverage

Let $\boldsymbol{\theta}^*_{\mathbf{u}}$ denote the least squares estimator of $\boldsymbol{\theta}$ under $L(\boldsymbol{\theta}/\mathbf{u})$. We have $\boldsymbol{\theta}^* = \boldsymbol{\theta}^*_0$ and $L(\boldsymbol{\theta}) = L(\boldsymbol{\theta}/\mathbf{0})$. Let $\mathbf{V}$ and $\mathbf{W}$ be the first and second derivatives of $f(\boldsymbol{\theta})$ with elements

$$V^a_j = (\partial \mathbf{f}_a/\partial\theta_j)_{\theta^*}\,, \qquad \text{and} \qquad W^a_{jk} = (\partial^2 \mathbf{f}_a/\partial\theta_j\partial\theta_k)_{\theta^*}\,,$$

respectively. For each $a = 1,\dots,n$ denote by $\mathbf{T}_a$ the $p\times p\times p$-array with element

$$(\partial^3 \mathbf{f}_a/\partial\theta_j\partial\theta_k\,\partial\theta_1)_{\theta^*}\,.$$

Put $\mathbf{e}^* = \mathbf{Y} - \mathbf{f}(\boldsymbol{\theta}^*)$.

Consider a perturbation of $\mathbf{Y}$ in the form $\mathbf{u} = c\mathbf{b}$, where $c \in \mathbb{R}$ and $\|\mathbf{b}\| = 1$. We define the parameter leverage vector due to a perturbation of the data by c in the direction $\mathbf{b}$ as

$$P(c,\mathbf{b}) = (\boldsymbol{\theta}^*_{c\mathbf{b}} - \boldsymbol{\theta}^*)/c\,,$$

and the Jacobian parameter leverage vector as

$$P(\mathbf{b}) = \lim_{c\to 0} P(c,\mathbf{b})\,.$$

Finally, we define the Jacobian parameter leverage matrix as

$$\mathbf{P} = \big(\mathbf{V}'\mathbf{V} - [\mathbf{e}^*][\mathbf{W}]\big)^{-1}\mathbf{V}'$$

where the square bracket multiplication has been defined by Bates and Watts as follows. If $\mathbf{E}$ is an $m\times n$ matrix and $\mathbf{T}$ is an $n\times p\times q$ array, then the elements of the ith face $\mathbf{M}_i$, $i = 1,\dots,m$ of the $m\times p\times q$ array $\mathbf{M} = [\mathbf{E}][\mathbf{T}]$ are E_iT_{jk}, $j = 1,\dots,p$, $k = 1,\dots,q$, where E_i is the ith row of $\mathbf{E}$ and T_{jk} is the jkth column of $\mathbf{T}$.

THEOREM 1. *The parameter leverage vector resulting from perturbation of the data by c in the direction $\mathbf{b}$ is, up to terms of order c,*

$$P(c,\mathbf{b}) = \mathbf{A}^{-1}\mathbf{V}^{-1}\mathbf{b} + 1/2\mathbf{b}'(\mathbf{P}_1 + \mathbf{P}_2 + \mathbf{P}_3 + \mathbf{P}_4)\mathbf{b}c\,, \tag{2}$$

where

$$\begin{aligned}
\mathbf{A} &= \mathbf{V}'\mathbf{V} - [\mathbf{e}^{*'}][\mathbf{W}]\,,\\
\mathbf{P}_1 &= -[\mathbf{A}^{-1}][\mathbf{V}\mathbf{A}^{-1}\mathbf{W}\mathbf{A}^{-1}\mathbf{V}']\,,\\
(\mathbf{P}_2)^s_{tu} &= \big([\mathbf{I} - \mathbf{V}\mathbf{A}^{-1}\mathbf{V}'][\mathbf{A}^{-1}\mathbf{W}\mathbf{A}^{-1}\mathbf{V}']\big)^t_{su}\,,\\
(\mathbf{P}_3)^s_{tu} &= \big([\mathbf{I} - \mathbf{V}\mathbf{A}^{-1}\mathbf{V}'][\mathbf{A}^{-1}\mathbf{W}\mathbf{A}^{-1}\mathbf{V}']\big)^u_{su}\,,\\
\mathbf{P}_4 &= [\mathbf{A}^{-1}]\big[\mathbf{V}\mathbf{A}^{-1}(\Sigma e_i\mathbf{T}_i)\mathbf{A}^{-1}\mathbf{V}'\big]\,.
\end{aligned}$$

Proof. We have

$$\left(\partial L(\boldsymbol{\theta}/\mathbf{u})/\partial\theta_i\right) = 0 \tag{3}$$

when evaluated at $\boldsymbol{\theta}^*_{\mathbf{u}}$ for $i = 1, \dots, p$ and all $\mathbf{u}$.

Differentiating both sides of (3) with respect to $\mathbf{u}$ and evaluating at $\mathbf{u} = 0$ we obtain

$$\left(\partial^2 L(\boldsymbol{\theta}/\mathbf{u})/\partial\theta_i\,\partial u_j\right)_{\theta^*} = 1/\sigma^2 V_i^j\,, \qquad j = 1, \dots, n\,.$$

Similarly we have

$$\left(\partial^2 L(\boldsymbol{\theta}/\mathbf{u})/\partial\theta_i\,\partial\theta_j\right)_{\theta^*} = -1/\sigma^2 A_{ij}\,,$$

where $A_{i,j}$ is the (i,j)-element of $\mathbf{A}$.

If we denote by $\mathbf{L}_{\theta\theta}$ the $p\times p$ matrix with the elements $\left(\partial^2 L(\boldsymbol{\theta}/\mathbf{u})/\partial\theta_i\,\partial\theta_j\right)_{\theta^*}$ it follows that

$$\mathbf{L}_{\theta\theta} = -1/\sigma^2\mathbf{A}\,. \tag{4}$$

We recall that $-\mathbf{L}_{\theta\theta}$ is the observed information for the postulated model $(\mathbf{u} = 0)$.

Let $\mathbf{J}$ be the $p\times n$-matrix with elements

$$\partial\theta^*_{i\mathbf{u}}/\partial u_j\,, \qquad i = 1, \dots p\,, \quad j = 1, \dots, n\,,$$

where $\theta^*_{i\mathbf{u}}$ is the ith component of $\boldsymbol{\theta}^*_{\mathbf{u}}$. From (3) we have

$$\mathbf{J} = \mathbf{A}^{-1}\mathbf{V}'. \tag{5}$$

Similarly

$$\begin{aligned}\left(\partial^3 L(\boldsymbol{\theta}/\mathbf{u})/\partial\theta_i\,\partial\theta_j\,\partial\theta_k\right)^*_{\theta,\mathbf{u}=0} = {}& -1/\sigma^2\Big\{\sum_t\left(W^t_{jk}V_{ti} + V_{tj}W^t_{ik}\right)\\ &+ \sum_t\left(V_{tk}W^t_{ij} - e^*_t T^t_{ijk}\right)\Big\}.\end{aligned}$$

We have

$$\begin{aligned}\sum_t W^t_{jk}V_{ti} &= \left([\mathbf{V}'][\mathbf{W}]\right)^i_{jk}\,,\\ \sum_t V_{tj}W^t_{ik} &= \left([\mathbf{V}'][\mathbf{W}]\right)^j_{ik}\,,\\ \sum_t V_{tk}W^t_{ij} &= \left([\mathbf{V}'][\mathbf{W}]\right)^k_{ij}\,,\end{aligned}$$

where an element of type S^i_{jk} means the (j,k)-element in ith face of an array $\mathbf{S}$.

If we denote by $\mathbf{L}_{\theta\theta\theta}$ the $p \times p \times p$ array with the elements $\big(\partial^3 L(\boldsymbol{\theta}/\mathbf{u})/\partial\theta_i \partial\theta_j \partial\theta_k\big)^*_{\theta,\mathbf{u}=0}$ it follows that

$$\mathbf{L}_{\theta\theta\theta} = -1/\sigma^2 \big\{[\mathbf{V}'][\mathbf{W}]^{\sim} - \Sigma e_t^* \mathbf{T}^t\big\}, \tag{6}$$

where the following convention has been used: given an array $\mathbf{M}$, $\mathbf{M}^{\sim}$ is defined to be the array with elements

$$(\mathbf{M}^{\sim})^i_{jk} = M^i_{jk} + M^j_{ki} + M^k_{ij}.$$

Differentiating both sides of (3) twice with respect to $\mathbf{u}$ and evaluating at $\mathbf{u} = \mathbf{0}$, we obtain

$$\mathbf{J}'\mathbf{L}_{\theta\theta\theta}\mathbf{J} + [\mathbf{L}_{\theta\theta}][\mathbf{J}^0] + \mathbf{J}'\mathbf{L}_{\theta\theta\mathbf{u}} + \mathbf{L}_{\theta\mathbf{u}\theta} + \mathbf{J} + \mathbf{L}_{\theta\mathbf{u}\mathbf{u}} = \mathbf{0}, \tag{7}$$

where $\mathbf{J}^0$ is the $p \times n \times n$ array with elements $\partial^2\theta^*_{i\mathbf{u}}/\partial u_j \partial\mathbf{u}_k$, $i = 1,\dots,p, j,k = 1,\dots,n$,

$\mathbf{L}_{\theta\theta\mathbf{u}}$ is the $p \times p \times n$ array with elements

$$\big(\partial^3 L(\boldsymbol{\theta}/\mathbf{u})/\partial\theta_i \partial\theta_j \partial u_k\big)^*_{\theta,\mathbf{u}=\mathbf{o}} = W^k_{ij}, \tag{8}$$

and so on. Of course $\mathbf{L}_{\theta\mathbf{u}\mathbf{u}} = \mathbf{0}$.

Solving (7) for $\mathbf{J}^0$ we obtain

$$\mathbf{J}^0 = -[\mathbf{L}^{-1}_{\theta\theta}][\mathbf{J}'\mathbf{L}_{\theta\theta\theta}\mathbf{J} + \mathbf{J}'\mathbf{L}_{\theta\theta\mathbf{u}} + \mathbf{L}_{\theta\mathbf{u}\theta}\mathbf{J}]. \tag{9}$$

It can be shown that

$$-[\mathbf{L}^{-1}_{\theta\theta}][\mathbf{J}'\mathbf{L}_{\theta\theta\theta}\mathbf{J}] = -[\mathbf{A}^{-1}]\big[\mathbf{V}\mathbf{A}^{-1}\big(([\mathbf{V}'][\mathbf{W}])^{\sim} - \Sigma e_i \mathbf{T}^i\big)\mathbf{A}^{-1}\mathbf{V}'\big], \tag{10}$$

$$-\big([\mathbf{L}^{-1}_{\theta\theta}][\mathbf{L}_{\theta\mathbf{u}\theta}]\big)^s_{tu} = (\mathbf{A}^{-1}\mathbf{W}\mathbf{A}^{-1}\mathbf{V}')^t_{su}, \tag{11}$$

$$-\big([\mathbf{L}^{-1}_{\theta\theta}][\mathbf{J}'\mathbf{L}_{\theta\theta\mathbf{u}}]\big)^s_{tu} = (\mathbf{A}^{-1}\mathbf{W}\mathbf{A}^{-1}\mathbf{V}')^u_{st}. \tag{12}$$

The result follows by inserting (10), (11), (12) into (9). □

Corollary 1. *The Jacobian parameter leverage vector due to a perturbation of the data by c in the direction $\mathbf{b}$ is $P(\mathbf{b}) = \mathbf{P}\mathbf{b}$.*

In order to get a deeper insight into the matrix $\mathbf{P}$, we express it in a standard form. Let the $\mathbf{QR}$ decomposition of $\mathbf{V}$ be given by $\mathbf{V} = \mathbf{QR}$, where $\mathbf{Q}$ is an $n \times p$ orthonormal matrix and $\mathbf{R}$ is a $p \times p$ upper triangular matrix. Consider now the reparameterization

$$\boldsymbol{\tau} = \mathbf{Q}'\big(\mathbf{f}(\boldsymbol{\theta}) - \mathbf{f}(\boldsymbol{\theta}^*)\big),$$

where $\boldsymbol{\tau}$ is called the normal parameter. It can be shown that the Jacobian parameter leverage matrix with respect to the normal parameter is

$$\mathbf{P} = (\mathbf{I} - \mathbf{B})^{-1}\mathbf{Q}', \tag{13}$$

where

$$\mathbf{B} = [\mathbf{e}^*][\mathbf{L}'\mathbf{W}\mathbf{L}], \qquad \mathbf{L} = \mathbf{R}^{-1}.$$

The relationship between the Jacobian parameter leverage matrix and the intrinsic curvature of Bates and Watts follows from the fact that

$$\mathbf{B} = [\mathbf{e}^{*'}\mathbf{N}][\mathbf{C}],$$

where $\mathbf{N}'\mathbf{e}^*$ is the rotated residual vector and $\mathbf{C}$ is the intrinsic curvature array. The norm $\|\mathbf{P}\,\mathbf{b}\|$ measures the influence of a perturbation of data in the direction $\mathbf{b}$ on values of parameters. In the case of the derived linear model $\mathbf{f}(\boldsymbol{\theta}^*) + \mathbf{V}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)$, the Jacobian parameter leverage matrix for the normal parameter is $\mathbf{H} = \mathbf{Q}'$. The matrix $\mathbf{P}$ takes into account the normal curvatures of the expectation surface at $\mathbf{f}(\boldsymbol{\theta}^*)$, while $\mathbf{H}$ does not.

For the normal parameter

$$(\mathbf{P})_{\max} = \max_{\|\mathbf{b}\|=1} \left\|(\mathbf{I}-\mathbf{B})^{-1}\mathbf{Q}'\mathbf{b}\right\| = (1-\lambda_p)^{-1}, \tag{14}$$

where λ_p is the largest eigenvalue of $\mathbf{B}$. Denoting by $\mathbf{b_i}$ the ith standard basis vector in $\mathbb{R}^n$, we have

$$P_i = \|\mathbf{P}\mathbf{b}_i\| = \left\|(\mathbf{I}-\mathbf{B})^{-1}(\mathbf{Q}')_i\right\|, \tag{15}$$

where $(\mathbf{Q}')_i$ means the ith column of $\mathbf{Q}'$. Evidently

$$P_i^2 = \mathbf{b}_i'\mathbf{Q}(\mathbf{I}-\mathbf{B})^{-2}\mathbf{Q}'\mathbf{b_i}. \tag{16}$$

We recall that the Jacobian leverage matrix introduced by Laurent and Cook is $\mathbf{J} = \mathbf{Q}(\mathbf{I}-\mathbf{B})^{-1}\mathbf{Q}'$. Its ith diagonal element j_{ii} measures the rate of change in the fitted value y_i^* with respect to the rate of change in y_i, It follows from (16) that if $(\mathbf{I}-\mathbf{B})^{-1}$ differs from $(\mathbf{I}-\mathbf{B})^{-2}$, the case maximizing j_{ii} may be different from that with the largest Jacobian parameter leverage. If $(\mathbf{I}-\mathbf{B})^{-1} = (\mathbf{I}-\mathbf{B})^{-2}$, then $\mathbf{B} = \mathbf{0}$. Two special cases in which this occurs are when the model is intrinsically linear (the elements of $\mathbf{L}'\mathbf{W}\mathbf{L}$ are zeros) and when the model provides an exact fit to the data (the residual vector $\mathbf{e}^*$ is zero).

REFERENCES

[1] COOK, R. D.: *Assessment of local influence*, J. Roy. Statistic. Soc. B **48** (1986), 133–169.

[2] KUBÁČEK, L.: *Models with a low nonlinearity*, Probastat 94, Tatra Mt. Math. Publ. **7** (1996), 149–155.

[3] PÁZMAN, A.: *Nonlinear statistical Models*, Kluwer Acad. Publ., Dordrecht, 1993.

[4] POTOCKÝ, R.—Van Ban T.: *Confidence regions in nonlinear regression models*, Appl. Math. **37** (1992), 29–39.

[5] POTOCKÝ, R.—Van Ban T.: *Confidence regions and the problem of reparametrization*, Probastat 94, Tatra Mt. Math. Publ. **7** (1996), 223–227.

[6] St. LAURENT, R. J.—COOK, R. D.: *Leverage and superleverage in nonlinear regression*, J. Amer. Statist. Assoc. **87** (1992), 985–990.

[7] St. LAURENT, R. J.—COOK, R. D.: *Leverage, local influence and curvature in nonlinear regression*, Biometrika **80** (1993), 99–106.

Received April 20, 1998

Department of Probability and Statistics
Faculty of Mathematics and Physics
Comenius University
SK–842 15 Bratislava
SLOVAKIA

E-mail: potocky@fmph.uniba.sk