Journal of the Korean Data & Information Science Society
2009, **20**(5), 913–917

한국데이터정보과학회지

Effects of curvature on leverage in nonlinear regression

Myung Wook Kahng[1]

[1]Department of Statistics, Sookmyung Women's University

Received 28 June 2009, revised 23 August 2009, accepted 7 September 2009

Abstract

The measures of leverage in linear regression has been extended to nonlinear regression models. We consider several curvature measures of nonlinearity in an estimation situation. The relationship between measures of leverage and statistical curvature are explored in nonlinear regression models. The circumstances under which the Jacobian leverage reduces to a tangent plane leverage are discussed in connection with the effective residual curvature of the nonlinear model.

Keywords: Curvature measure, effective residual curvature, intrinsic curvature array, jacobian leverage, parameter-effect array.

1. Introduction

Leverage is one of the basic components of influence in linear regression models (Chatterjee and Hadi, 1986). By Belsley, Kuh, and Welsch (1980) and Ross (1987), leverage has been generalized via linear approximation to more complex response models. Emerson, Hoaglin, and Kempthorne (1984) considered definitions of leverage for nonlinear regression models.

Bates and Watts (1980) proposed measures of intrinsic and parameter-effects curvature for assessing the adequacy of the linear approximation. These ideas are extended and refined by Bates and Watts (1981), and Hamilton, Watts and Bates (1982). In this paper, we show how the nonlinearity of the model may affect the leverage, and we explore the relationship between leverage measures and statistical curvature. We discuss the circumstances under which the Jacobian leverage reduces to a tangent plane leverage in connection with the effective residual curvature of the nonlinear model.

2. Leverages in nonlinear regression

The standard nonlinear regression model can be expressed as

$$y_i = f(\boldsymbol{x_i}, \boldsymbol{\theta}) + \epsilon_i, \quad i = 1, \cdots, n$$

in which the i-th response y_i is related to the q-dimensional vector of known explanatory variables $\boldsymbol{x}_i$ through the known model function f with domain $\boldsymbol{\Theta}$, an open subset of $\boldsymbol{R}^p$

[1] Professor, Department of Statistics, Sookmyung Women's University, Seoul 140-742, Korea.
E-mail: mwkahng@sookmyung.ac.kr

which depends on the p-dimensional unknown parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, and ϵ_i is error. We assume that f is twice continuously differentiable in $\boldsymbol{\theta}$, and errors ϵ_i are *i.i.d.* normal random variables with mean 0 and variance σ^2. In matrix notation we may write,

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) + \boldsymbol{\epsilon}$$

where $\boldsymbol{y}$ is an n-dimensional vector with elements $y_1, \cdots, y_n$, $\boldsymbol{X}$ is an $n \times q$ matrix with rows $\boldsymbol{x}_1^T, \cdots, \boldsymbol{x}_n^T$, $\boldsymbol{\epsilon}$ is an n-dimensional vector with elements $\epsilon_1, \cdots, \epsilon_n$, and $\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) = (f(\boldsymbol{x_1}, \boldsymbol{\theta}), \cdots, f(\boldsymbol{x_n}, \boldsymbol{\theta}))^T = \boldsymbol{f}(\boldsymbol{\theta}) = \boldsymbol{\eta}(\boldsymbol{\theta})$. Given the response vector $\boldsymbol{y}$, the least squares estimate of $\boldsymbol{\theta}$ is denoted by $\widehat{\boldsymbol{\theta}}$, the predicted response vector is $\widehat{\boldsymbol{y}} = \boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}) = \boldsymbol{f}(\widehat{\boldsymbol{\theta}}) = \boldsymbol{\eta}(\widehat{\boldsymbol{\theta}}) = \widehat{\boldsymbol{\eta}}$, and the residual vector is $\boldsymbol{e} = \boldsymbol{y} - \widehat{\boldsymbol{y}}$.

A tangent plane approximation to the expectation surface $\boldsymbol{M} = \{\boldsymbol{f}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ at $\widehat{\boldsymbol{\theta}}$ is used to make inferences about $\boldsymbol{\theta}$ through the derived linear model $\boldsymbol{f}(\boldsymbol{\theta}) = \boldsymbol{f}(\widehat{\boldsymbol{\theta}}) + \widehat{\boldsymbol{V}}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})$ where $\boldsymbol{V} = \boldsymbol{V}(\boldsymbol{\theta}) = \partial \boldsymbol{f}/\partial \boldsymbol{\theta}^T$ is the $n \times p$ matrix and $\widehat{\boldsymbol{V}} = \boldsymbol{V}(\widehat{\boldsymbol{\theta}})$. The elliptical confidence region for θ based on this approximation can be written as

$$\left\{ \boldsymbol{\theta} | (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})^T \widehat{\boldsymbol{V}}^T \widehat{\boldsymbol{V}} (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}) \leq ps^2 F_\alpha(p, n-p) \right\}. \tag{2.1}$$

An alternative procedure for defining the confidence region is based on the likelihood ratio. Standard asymptotic arguments suggest that the confidence region for $\boldsymbol{\theta}$ is

$$\left\{ \boldsymbol{\theta} | S(\boldsymbol{\theta}) - S(\widehat{\boldsymbol{\theta}}) \leq ps^2 F_\alpha(p, n-p) \right\}. \tag{2.2}$$

Based on a linear approximation, the tangent plane leverage matrix, denoted by $\widehat{\boldsymbol{H}}$, can be written as $\widehat{\boldsymbol{H}} = \widehat{\boldsymbol{V}}(\widehat{\boldsymbol{V}}^T \widehat{\boldsymbol{V}})^{-1} \widehat{\boldsymbol{V}}^T$. The diagonal elements $\widehat{h}_{ii}$ of $\widehat{\boldsymbol{H}}$ are frequently used as a measure of leverage in nonlinear regression (Ross, 1987).

Emerson *et al.* (1984) and St. Laurent and Cook (1992, 1993) generalized the measure of leverages due to perturbation of responses. By Kahng (2007, 2008), leverages in nonlinear regression can be obtained as follows;

$$\widehat{\boldsymbol{J}} = \widehat{\boldsymbol{V}}(\widehat{\boldsymbol{V}}^T \widehat{\boldsymbol{V}} - [\boldsymbol{e}^T][\widehat{\boldsymbol{W}}])^{-1} \widehat{\boldsymbol{V}}^T$$

which is known as the Jacobian leverage matrix. Here $\boldsymbol{W} = \boldsymbol{W}(\boldsymbol{\theta}) = \partial^2 \boldsymbol{f}/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T$ is the $n \times p \times p$ array and $\widehat{\boldsymbol{W}} = \boldsymbol{W}(\widehat{\boldsymbol{\theta}})$. The square bracket notation indicates that the summation is over the numerator index: that is, the element in the n-th face, p-th row, and q-th column of the product $\boldsymbol{A} = [\boldsymbol{B}][\boldsymbol{C}]$ is $a_{npq} = \sum_{i=1}^{n_2} b_{ni} c_{ipq}$, where $\boldsymbol{B}$ is an $n_1 \times n_2$ matrix and $\boldsymbol{C}$ is an $n_2 \times n_3 \times n_4$ array.

3. Effects of curvature on linearized regions

We consider the QR decomposition of $n \times p$ matrix $\widehat{\boldsymbol{V}}$, namely

$$\widehat{\boldsymbol{V}} = \boldsymbol{Q}\boldsymbol{R} = (\boldsymbol{Q_1} | \boldsymbol{Q_2}) \begin{pmatrix} \boldsymbol{R}_{11} \\ -- \\ \boldsymbol{0} \end{pmatrix}$$

where columns of the $n \times p$ matrix $\boldsymbol{Q}_1$ form an orthonormal basis for the tangent plane and columns of the $n \times (n-p)$ matrix $\boldsymbol{Q}_2$ are an orthonormal basis for the space orthogonal to the tangent plane, and $\boldsymbol{R}_{11}$ is a nonsingular $p \times p$ upper triangular matrix. Since $\boldsymbol{R}_{11}^T \boldsymbol{R}_{11}$ is the Cholesky decomposition of $\widehat{\boldsymbol{V}}^T \widehat{\boldsymbol{V}}$, $\boldsymbol{R}_{11}$ is unique if its diagonal elements are all positive or all negative. If $\boldsymbol{R}_{11}$ is unique, so is $\boldsymbol{Q}_1$, as $\boldsymbol{Q}_1 = \widehat{\boldsymbol{V}} \boldsymbol{R}_{11}^{-1} = \widehat{\boldsymbol{V}} \boldsymbol{K}$, where $\boldsymbol{K} = \boldsymbol{R}_{11}^{-1}$. However $\boldsymbol{Q}_2$ is not unique.

We define $\boldsymbol{U} = \boldsymbol{K}^T \widehat{\boldsymbol{W}} \boldsymbol{K} = \boldsymbol{R}_{11}^{-T} \widehat{\boldsymbol{W}} \boldsymbol{R}_{11}^{-1}$, so that pre- and post-multiplication of the $p \times p \times n$ array $\widehat{\boldsymbol{W}}$ involves multiplication of each $p \times p$ face and hence produces a $p \times p \times n$ array. Pre-multiplying the $n \times n$ matrix $\boldsymbol{Q}^T$ by $\boldsymbol{U}$ gives

$$\boldsymbol{A}_{..} = [\boldsymbol{Q}^T][\boldsymbol{U}] = [\boldsymbol{Q}_1 | \boldsymbol{Q}_2]^T [\boldsymbol{U}] = [\boldsymbol{Q}_1^T \boldsymbol{U} | \boldsymbol{Q}_2^T \boldsymbol{U}] = \boldsymbol{A}_{..}{}^T | \boldsymbol{A}_{..}{}^N$$

where $\boldsymbol{A}_{..}{}^T$ is the parameter-effect array and $\boldsymbol{A}_{..}{}^N$ is the intrinsic curvature array (Bates and Watts, 1980).

The linear approximation inference intervals and regions for parameters in nonlinear regression models can be obtained from the first order Taylor series approximation to the expectation function evaluated at $\widehat{\boldsymbol{\theta}}$. Geometrically, the linear approximation inference region (2.1) assumes that the mapping of $\boldsymbol{\theta}$ to $\boldsymbol{\eta}(\boldsymbol{\theta})$ is

$$\widehat{\boldsymbol{\eta}} + \widehat{\boldsymbol{V}}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}).$$

Thus the mapping from the tangent plane to the parameter space is linear. As pointed out by Beale (1960), this approximation will be acceptable only if the expectation surface is sufficiently flat to be replaced by the tangent plane. Bates and Watts (1980) used second derivatives of the model function to derive curvature measures of intrinsic and parameter effects nonlinearity to assess the validity of these assumptions. If the intrinsic nonlinearity is large, so that a tangent plane does not provide an adequate approximation we can use the quadratic approximation (Hamilton, Watts and Bates, 1982).

To obtain a quadratic approximation, consider a point $\boldsymbol{\theta}$ in the parameter space. This point maps through the nonlinear transformation, $\boldsymbol{\eta}(\boldsymbol{\theta})$, to the point on the tangent plane with coordinates

$$\boldsymbol{\tau} = \boldsymbol{Q}_1^T \{ \boldsymbol{\eta}(\boldsymbol{\theta}) - \widehat{\boldsymbol{\eta}} \}.$$

It is a convenient transformation for studying intrinsic curvature in that the parameter-effects curvature is zero. Reparametrizing the underlying model in terms of τ instead of $\boldsymbol{\theta}$, we have $\boldsymbol{f}(\boldsymbol{\theta}(\boldsymbol{\tau}))$ or $\boldsymbol{\eta}(\boldsymbol{\tau})$ for short. A Taylor expansion of $\boldsymbol{\eta}(\boldsymbol{\tau})$ at $\boldsymbol{\tau} = \mathbf{0}$ gives us

$$\boldsymbol{\eta}(\boldsymbol{\tau}) - \widehat{\boldsymbol{\eta}} \approx (\partial \boldsymbol{\eta} / \partial \boldsymbol{\tau}|_{\boldsymbol{\tau}=\mathbf{0}}) \, \boldsymbol{\tau} + \boldsymbol{\tau}^T \left(\partial^2 \boldsymbol{\eta} / \partial \boldsymbol{\tau} \partial \boldsymbol{\tau}^T |_{\boldsymbol{\tau}=\mathbf{0}} \right) \boldsymbol{\tau}/2. \tag{3.1}$$

Here

$$\partial \boldsymbol{\eta} / d\boldsymbol{\tau}|_{\boldsymbol{\tau}=\mathbf{0}} = \left(d\boldsymbol{\eta} / d\boldsymbol{\theta}|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \right) (\partial \boldsymbol{\theta} / \partial \boldsymbol{\tau}|_{\boldsymbol{\tau}=\mathbf{0}}) = \widehat{\boldsymbol{V}} \left(\boldsymbol{Q}_1^T \widehat{\boldsymbol{V}} \right)^{-1} = \boldsymbol{Q}_1, \tag{3.2}$$

and from the appendix of Hamilton, Watts and Bates (1982),

$$\partial^2 \boldsymbol{\eta} / \partial \boldsymbol{\tau} d\boldsymbol{\tau}^T |_{\boldsymbol{\tau}=\mathbf{0}} = [\boldsymbol{Q}_2 \boldsymbol{Q}_2^T][\boldsymbol{K}^T \widehat{\boldsymbol{W}} \boldsymbol{K}] = [\boldsymbol{Q}_2][\boldsymbol{A}_{..}{}^N]. \tag{3.3}$$

Having established the properties of our transformation we now derive a confidence region for $\boldsymbol{\theta}$ based on approximation of $\boldsymbol{\epsilon}$. Substituting (3.2) and (3.3) into (3.1), we have

$$\boldsymbol{\eta}(\boldsymbol{\tau}) \approx \widehat{\boldsymbol{\eta}} + \boldsymbol{Q}_1\boldsymbol{\tau} + [\boldsymbol{Q}_2][\boldsymbol{\tau}^T\boldsymbol{A}_{..}{}^N\boldsymbol{\tau}]/2$$

and

$$\boldsymbol{\epsilon} \approx \boldsymbol{e} - \boldsymbol{Q}_1\boldsymbol{\tau} - [\boldsymbol{Q}_2][\boldsymbol{\tau}^T\boldsymbol{A}_{..}{}^N\boldsymbol{\tau}]/2.$$

The squared total length of $\boldsymbol{\epsilon}$ is

$$\boldsymbol{\epsilon}^T\boldsymbol{\epsilon} \approx \boldsymbol{e}^T\boldsymbol{e} + \boldsymbol{\tau}^T\boldsymbol{\tau} + (\boldsymbol{\tau}^T\boldsymbol{A}_{..}{}^N\boldsymbol{\tau})^T(\boldsymbol{\tau}^T\boldsymbol{A}_{..}{}^N\boldsymbol{\tau})/4 - [\boldsymbol{e}^T\boldsymbol{Q}_2][\boldsymbol{\tau}^T\boldsymbol{A}_{..}{}^N\boldsymbol{\tau}]/2. \tag{3.4}$$

Hamilton, Watts and Bates (1982) proposed a matrix, namely

$$\boldsymbol{B} = [\boldsymbol{e}^T\boldsymbol{Q}_2][\boldsymbol{A}_{..}{}^N] = [\boldsymbol{e}^T][\boldsymbol{U}].$$

This is the $p \times p$ matrix obtained from the inner product of the rotated residual vector $\boldsymbol{e}^T\boldsymbol{Q}_2$ and the intrinsic curvature array $\boldsymbol{A}_{..}{}^N$. The matrix $\boldsymbol{B}$ is referred to as the "effective residual curvature matrix", because it gives the effective normal curvatures relative to the residual vector $\boldsymbol{e}$.

If we ignore fourth powers of $\boldsymbol{\tau}$ in (3.4), we obtain the approximation

$$\boldsymbol{\epsilon}^T\boldsymbol{\epsilon} \approx \boldsymbol{e}^T\boldsymbol{e} + \boldsymbol{\tau}^T(\boldsymbol{I} - \boldsymbol{B})\boldsymbol{\tau}. \tag{3.5}$$

Substituting this approximation (3.5) into (2.2) the confidence region leads to

$$\boldsymbol{\tau}^T(\boldsymbol{I} - \boldsymbol{B})\boldsymbol{\tau} \le ps^2F_\alpha(p, n-p). \tag{3.6}$$

This may be compared with the corresponding region based on the linear approximation in the $\boldsymbol{\tau}-$parameters obtained by setting $\boldsymbol{A}_{..}^N$ and $\boldsymbol{B}$ equal to zero, namely the sphere

$$\boldsymbol{\tau}^T\boldsymbol{\tau} \le ps^2F_\alpha(p, n-p). \tag{3.7}$$

Since the $\boldsymbol{\tau}-$parameters have a zero parameter-effect curvature, the effect of intrinsic curvature on (3.6) can be gauged by comparing the length of the axis of (3.6) with the radius of the sphere (3.7).

4. Effects of curvature on leverage

We investigate the relation between tangent plane leverage matrix and Jacobian leverage matrix, and relationship to curvature measures of nonlinearity discussed in section 3.

Since $\widehat{\boldsymbol{V}} = \boldsymbol{Q}_1\boldsymbol{R}_{11}$ and $\widehat{\boldsymbol{V}}^T\widehat{\boldsymbol{V}} = \boldsymbol{R}_{11}^T\boldsymbol{Q}_1^T\boldsymbol{Q}_1\boldsymbol{R}_{11} = \boldsymbol{R}_{11}^T\boldsymbol{R}_{11}$, the tangent plane leverage matrix can be written as,

$$\begin{aligned}\widehat{\boldsymbol{H}} &= \widehat{\boldsymbol{V}}(\widehat{\boldsymbol{V}}^T\widehat{\boldsymbol{V}})^{-1}\widehat{\boldsymbol{V}}^T \\ &= \boldsymbol{Q}_1\boldsymbol{R}_{11}(\boldsymbol{R}_{11}^T\boldsymbol{R}_{11})^{-1}\boldsymbol{R}_{11}^T\boldsymbol{Q}_1^T \\ &= \boldsymbol{Q}_1\boldsymbol{Q}_1^T.\end{aligned}$$

We note that, $\widehat{\boldsymbol{V}}^T\widehat{\boldsymbol{V}} - [\boldsymbol{e}^T][\widehat{\boldsymbol{W}}] = \boldsymbol{R}_{11}^T\boldsymbol{R}_{11} - \boldsymbol{R}_{11}^T([\boldsymbol{e}^T][\boldsymbol{U}])\boldsymbol{R}_{11} = \boldsymbol{R}_{11}^T(\boldsymbol{I}-\boldsymbol{B})\boldsymbol{R}_{11}$. This leads us to the Jacobian leverage matrix for a nonlinear model which can be written as

$$\begin{aligned}\widehat{\boldsymbol{J}} &= \widehat{\boldsymbol{V}}(\widehat{\boldsymbol{V}}^T\widehat{\boldsymbol{V}} - [\boldsymbol{e}^T][\widehat{\boldsymbol{W}}])^{-1}\widehat{\boldsymbol{V}}^T \\ &= \boldsymbol{Q}_1\boldsymbol{R}_{11}(\boldsymbol{R}_{11}^T(\boldsymbol{I}-\boldsymbol{B})\boldsymbol{R}_{11})^{-1}\boldsymbol{R}_{11}^T\boldsymbol{Q}_1^T \\ &= \boldsymbol{Q}_1(\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{Q}_1^T.\end{aligned}$$

The matrix $\boldsymbol{B}$ is referred to as the effective residual curvature matrix because it gives the effective normal curvature relative to the residual vector $\boldsymbol{e}$. As $\boldsymbol{e}$ is orthogonal to column space of $\widehat{\boldsymbol{V}}$, $\boldsymbol{B}$ is a function of only the intrinsic curvature array $\boldsymbol{A}_{..}{}^N$. The leverage matrix $\widehat{\boldsymbol{J}}$ takes into account the normal curvatures of expectation space $\boldsymbol{M}$ at $\widehat{\boldsymbol{\theta}}$ relative to the residual vector, while $\widehat{\boldsymbol{H}}$ does not. The parameter-effect array $\boldsymbol{A}_{..}{}^T$ does not play a role here, hence both $\widehat{\boldsymbol{J}}$ and $\widehat{\boldsymbol{H}}$ are invariant under reparametrization.

The leverage matrices $\widehat{\boldsymbol{J}}$ and $\widehat{\boldsymbol{H}}$ are identical if and only if the effective residual curvature matrix $\boldsymbol{B}$ is $\mathbf{0}$. Two special cases in which this occurs are when the model provides an exact fit to the data or when the model is intrinsically linear.

References

Bates, D. M. and Watts, D. G. (1980). Relative curvature measures of nonlinearity (with discussion). *Journal of the Royal Statistical Society,* Series B, **42**, 1-25.

Bates, D. M. and Watts, D. G. (1981). Parameter transformations for improved approximate confidence regions in nonlinear least squares. *Annals of Statistics,* **9**, 1152-1167.

Bates, D. M. and Watts, D. G. (1988). *Nonlinear regression analysis and its applications*, John Wiley and Sons, New York.

Beale, E. M. L. (1960). Confidence regions in nonlinear estimation (with discussion). *Journal of the Royal Statistical Society,* Series B, **22**, 41-88.

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

Chatterjee, S. and Hadi, A. S. (1986). Influential observations, high leverage points and outliers in linear regression. *Statistical Science,* **1**, 379-416.

Emerson, J. D., Hoaglin, D. C. and Kempthorne, P. J. (1984). Leverage in least squares additive-plus-multiplicative fits for two-way tables. *Journal of the American Statistical Association,* **79**, 329-335.

Hamilton, D. C., Watts, D. G. and Bates, D. M. (1982). Accounting for intrinsic nonlinearity in nonlinear regression parameter inference regions. *Annals of Statistics,* **10**, 386-393.

Kahng, M. (2007). Leverage measures in nonlinear regression. *Journal of Korean Data & Information Science Society,* **18**, 229-235.

Kahng, M. (2008). Leverage and outlier in nonlinear regression. *Journal of Korean Data & Information Science Society,* **19**, 287-292.

Ross, W. H. (1987). The geometry of case deletion and the assessments of influence in nonlinear regression. *The Canadian Journal of Statistics,* **15**, 91-103.

St. Laurent, R. T. and Cook, R. D. (1992). Leverage and superleverage in nonlinar regression. *Journal of the American Statistical Association,* **87**, 985-990.

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