Journal of the Korean Data & Information Science Society
2013, **24**(1), 179–187

http://dx.doi.org/10.7465/jkdi.2013.24.1.179
한국데이터정보과학회지

Accuracy of linear approximation for fitted values in nonlinear regression†

Myung-Wook Kahng[1]

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

Received 10 December 2012, revised 3 January 2013, accepted 8 January 2013

Abstract

Bates and Watts (1981) have discussed the problems of reparameterizing nonlinear models in obtaining accurate linear approximation confidence regions for the parameters. A similar problem exists with computing confidence curves for fitted values or predictions. The statistical behavior of fitted values does not depend on the parameterization. Thus, as long as the intrinsic curvature is small, standard Wald intervals for fitted values are likely to be sufficient. Accuracy of linear approximation for fitted values is investigated using confidence curves.

Keywords: Confidence curves, fitted values, intrinsic curvature, likelihood confidence region, parameter-effects curvature, subset curvature, Wald confidence region.

1. Introduction

A standard paradigm for the analysis of nonlinear models is to assume that results for the linear model hold at least approximately for large enough sample sizes. Confidence regions for parameters of a normal nonlinear regression model are commonly constructed by using linear regression methods, replacing the solution locus with the tangent plane at the maximum likelihood estimate. Such linear regions are generally easier to construct and comprehend than corresponding likelihood regions. Likelihood regions, on the other hand, are not influenced by parameter-effects nonlinearity and generally have true coverage closer to the nominal level than do linear regions. Under suitable regularity conditions and with a sufficiently large sample size, linear and likelihood regions will be in good agreement, but in any particular problem the strength of this agreement is uncertain.

Bates and Watts (1980) propose measures of intrinsic and parameter-effects curvature for assessing the adequacy of the linear approximation. These ideas are extended and refined by Cook and Goldberg (1986). They develop measures for assessing the agreement between linear and likelihood regions for an arbitrary subset of parameters from a nonlinear model. The subset curvatures developed in this paper appear to be reliable indicators of the adequacy of linear confidence regions for most nonlinear models. This ability to deal with subsets greatly extends the usefulness of the Bates-Watts methodology.

† This research was supported by the Sookmyung Women's University Research Grants 2011.
[1] Professor, Department of Statistics, Sookmyung Women's University, Seoul 140-742, Korea.
E-mail: mwkahng@sookmyung.ac.kr

To ensure good agreement between the tangent plane and likelihood regions, the maximum curvature must be considerably smaller than the Bates-Watts guide. However, this criterion can be too stringent for certain parameter subsets if whole-parameter curvatures are used. By contrast, the subset curvature describes the shape of the likelihood region in the parameter subspace of interest. Thus, the subset curvature is more directly relevant to the linearization adequacy question.

Cook and Weisberg (1990) give the graphical alternative to likelihood and Wald confidence intervals for a component of the single parameter vector. They discussed that their methodology can obtain the confidence curves for the parameter subsets but it has limitation for a function of parameters. Kahng (2003) considers obtaining graphical summaries of uncertainty in estimates of a function of parameters in nonlinear models. This method overcomes the limitation.

In this paper we will explore the accuracy of linearization inference for the fitted values through confidence curves. Section 2 considers curvature measures and confidence curves in nonlinear regression model. Accuracy of the linear approximation for the fitted values is investigated using curvature measures and confidence curves in section 3. We provide an example in section 4. Section 5 contains concluding remarks.

2. Nonlinear regression model

The standard nonlinear regression model can be expressed as

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

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, which depends on the p-dimensional unknown parameter $\boldsymbol{\theta} \in \Theta$, 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,

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) + \boldsymbol{\epsilon} \tag{2.1}$$

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 $\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) = (f(\boldsymbol{x}_1, \boldsymbol{\theta}), \cdots, f(\boldsymbol{x}_n, \boldsymbol{\theta}))^T = (f_1(\boldsymbol{\theta}), \cdots, f_n(\boldsymbol{\theta}))^T = \boldsymbol{f}(\boldsymbol{\theta})$. Given the response vector $\boldsymbol{y}$, the least squares estimate of $\boldsymbol{\theta}$ is denoted $\widehat{\boldsymbol{\theta}}$, minimizes the residual sum of squares $S(\boldsymbol{\theta})$ and maximizes the likelihood function. The predicted response vector is $\widehat{\boldsymbol{y}} = \boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}) = \boldsymbol{f}(\widehat{\boldsymbol{\theta}})$. A tangent plane approximation to the expectation surface 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 $\widehat{\boldsymbol{V}} = \boldsymbol{V}(\widehat{\boldsymbol{\theta}})$ is $\partial \boldsymbol{f} / \partial \boldsymbol{\theta}^T$ evaluated at $\widehat{\boldsymbol{\theta}}$.

2.1. Accuracy in linear approximation

A standard likelihood confidence region for $\boldsymbol{\theta}$ can be written as

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

where $s^2 = S(\widehat{\boldsymbol{\theta}})/(n-p)$ is the usual unbiased estimator of σ^2, and $F_\alpha(p, n-p)$ is the upper α point of the F-distribution with the p and $n-p$ degrees of freedom.

Suppose $\boldsymbol{\theta}$ is close to $\widehat{\boldsymbol{\theta}}$, then we have the following quadratic Taylor expansion:

$$\begin{aligned}\boldsymbol{f}(\boldsymbol{\theta}) &\approx \boldsymbol{f}(\widehat{\boldsymbol{\theta}}) + \widehat{\boldsymbol{V}}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}) + \frac{1}{2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})^T \widehat{\boldsymbol{W}}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}) \\ &= \boldsymbol{f}(\widehat{\boldsymbol{\theta}}) + \widehat{\boldsymbol{V}}\boldsymbol{\kappa} + \frac{1}{2}\boldsymbol{\kappa}^T \widehat{\boldsymbol{W}}\boldsymbol{\kappa}\end{aligned} \tag{2.3}$$

where $\boldsymbol{\kappa} = \boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{W}} = \boldsymbol{W}(\widehat{\boldsymbol{\theta}})$ is $\partial^2 \boldsymbol{f}/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T$ evaluated at $\widehat{\boldsymbol{\theta}}$. If we ignore the quadratic term, we have the linear approximation for $\boldsymbol{\theta}$ in the vicinity of $\widehat{\boldsymbol{\theta}}$

$$\boldsymbol{f}(\boldsymbol{\theta}) \approx \boldsymbol{f}(\widehat{\boldsymbol{\theta}}) + \widehat{\boldsymbol{V}}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}) = \boldsymbol{f}(\widehat{\boldsymbol{\theta}}) + \widehat{\boldsymbol{V}}\boldsymbol{\kappa} \tag{2.4}$$

This linear approximation amounts to approximating the expectation surface in the neighborhood of $\widehat{\boldsymbol{\theta}}$ by the tangent plane at $\widehat{\boldsymbol{\theta}}$. An important assumption used in this method is that the expectation surface is flat, so that the tangent plane provides an accurate approximation. Using this approximation, we have the Wald elliptical region

$$\left\{\boldsymbol{\theta} \,\middle|\, (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})^T \widehat{\boldsymbol{V}}^T \widehat{\boldsymbol{V}}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}) \le ps^2 F_\alpha(p, n-p)\right\} \tag{2.5}$$

If $\boldsymbol{f}$ is essentially linear over a sufficiently large neighborhood of $\hat{\boldsymbol{\theta}}$ or equivalently, if n is large enough, the likelihood (2.2) and linear (2.5) confidence region will be similar. Using geometric reasoning, Beale (1960) and Bates and Watts (1980) note that the confidence region based on the linear approximation will be accurate only if the expectation surface is sufficiently flat to be replaced by the tangent plane. Hamilton (1986) noted that the confidence region based on the linear approximation (2.4) has the advantage of having a simple shape which can be readily conceptualized and displayed. Unfortunately, this may not always be the case with the confidence region based on the likelihood ratio.

The validity of the linear approximation (2.4) will depend on the magnitude of the quadratic term $\boldsymbol{\kappa}^T \widehat{\boldsymbol{W}}\boldsymbol{\kappa}$ in (2.3) relative to the linear term $\widehat{\boldsymbol{V}}\boldsymbol{\kappa}$. To make this comparison Bates and Watts (1980, 1981) split the quadratic term into two orthogonal components, projections onto the tangent plane and normal to the tangent plane. These two component are given by

$$\widehat{\boldsymbol{W}} = \widehat{\boldsymbol{W}}\widehat{\boldsymbol{H}} + \widehat{\boldsymbol{W}}(\boldsymbol{I} - \widehat{\boldsymbol{H}}) = \widehat{\boldsymbol{W}}^{\boldsymbol{\tau}} + \widehat{\boldsymbol{W}}^{\boldsymbol{\eta}}$$

where $\widehat{\boldsymbol{H}} = \widehat{\boldsymbol{V}}(\widehat{\boldsymbol{V}}^T \widehat{\boldsymbol{V}})^{-1} \widehat{\boldsymbol{V}}^T$. Bates and Watts (1980) define two measures for comparing each quadratic component with the linear term, namely the maximum parameter-effects curvatures

$$\Gamma^{\boldsymbol{\tau}}(\boldsymbol{\theta}) = max\ \frac{||\boldsymbol{\kappa}\widehat{\boldsymbol{W}}^{\boldsymbol{\tau}}\boldsymbol{\kappa}||}{||\widehat{\boldsymbol{V}}\boldsymbol{\kappa}||^2} s\sqrt{p}$$

and the maximum intrinsic curvature

$$\Gamma^{\boldsymbol{\eta}}(\boldsymbol{\theta}) = max \ \frac{||\boldsymbol{\kappa}\widehat{\boldsymbol{W}}^{\boldsymbol{\eta}}\boldsymbol{\kappa}||}{||\widehat{\boldsymbol{V}}\boldsymbol{\kappa}||^2} s\sqrt{p}$$

where the maximum is taken over all $\boldsymbol{\kappa}$ in $\boldsymbol{R}^p$. Bates and Watts (1980) suggested that the statistical significant in $\Gamma^{\boldsymbol{\tau}}(\boldsymbol{\theta})$ and $\Gamma^{\boldsymbol{\eta}}(\boldsymbol{\theta})$ may be assessed by comparing these values with $1/\sqrt{F_\alpha(p, n-p)}$.

Consider the projections of confidence regions onto the tangent plane. The projections of likelihood confidence regions using quadratic approximation are found to be ellipsoids, in contrast with the spherical projections obtained with the linear approximation. Hamilton, Watts, and Bates (1982) suggested that the difference between the quadratic approximation ellipsoid and the linear approximation sphere directly indicates the effects of intrinsic nonlinearity on inference. They define the axis length ratio as the proportion of the length of axis of the ellipsoid to the radius of the linear approximation sphere. The extreme axis ratio can be used to assess the effect of intrinsic nonlinearity on likelihood confidence regions.

2.2. Confidence curves in nonlinear regression

Cook and Weisberg (1990) give the graphical alternative to likelihood and Wald confidence intervals for a component of the parameter vector $\boldsymbol{\theta}$. The reason for using confidence curves is that likelihood intervals can have a different shape for each significance level $1-\alpha$. Wald intervals are always symmetric, so if we have the 95% interval, we can always obtain the 90% and 99% intervals in a simple way. With the likelihood intervals, this is not so. Consequently, a graphical summary that does not depend on level is desirable.

The confidence curve includes two curves - a likelihood confidence curve and Wald confidence curve. In this plot, θ_2 is on the vertical axis and the plotted point corresponding to on the horizontal axis is of the form $\sqrt{[S(\widehat{\boldsymbol{\theta}}_1(\theta_2), \theta_2) - S(\widehat{\boldsymbol{\theta}})]/s^2}$ where $\widehat{\boldsymbol{\theta}}_1(\theta_2)$ denotes the $(p-1)$-dimensional vector valued function that minimizes $S(\boldsymbol{\theta})$ over $\boldsymbol{\theta}$ for θ_2 fixed.

This is a modification of the graphical summary of the standard profile log-likelihood which has θ_2 on the horizontal axis and $S(\widehat{\boldsymbol{\theta}}_1(\theta_2), \theta_2)$ on the vertical axis. The plot will be curves, with the amount of curvature giving information about the nonlinearity of the model. To this plot, two straight lines passing through $(0, \hat{\theta}_2)$ with slope $\pm se(\hat{\theta}_2)$ are added. These two straight lines represent the Wald interval. At any point on the horizontal axis of the confidence curve plot, the interval between the upper and lower curves provides a confidence interval for θ_2 or some level of $1-\alpha$. The confidence level can be determined from the calibrating distribution for either the Wald or likelihood procedure, which, in the scale of the plot, is $t(\nu)$, where ν is the corresponding degrees of freedom. If $t_\nu^{-1}(u)$ is the inverse of the t cumulative distribution function with ν degrees of freedom evaluated at u, then the confidence level at a point along the horizontal axis is $1 - 2t_\nu^{-1}(\sqrt{[S(\widehat{\boldsymbol{\theta}}_1(\theta_2), \theta_2) - S(\widehat{\boldsymbol{\theta}})]/s^2})$.

The Wald and the likelihood regions are tangent at the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$. If the likelihood is exactly quadratic, that is, if the log-likelihood is exactly normal, then the likelihood curves are the same as the Wald curves (two straight lines). Non-normality is reflected in the likelihood confidence curves failing to be straight.

Suppose that $c(\boldsymbol{\theta})$ is a continuous, twice differentiable and invertible function of $\boldsymbol{\theta}$ and is of particular interest. Kahng (2003) proposed the confidence curves for the function of parameters in nonlinear regression. The confidence curve for the function $c(\boldsymbol{\theta})$ can be obtained by using the large-sample normality of $c(\hat{\boldsymbol{\theta}})$ and the likelihood ratio statistic for testing $H_0 : c(\boldsymbol{\theta}) = \phi$.

By the large-sample normality, the maximum likelihood estimate of $c(\boldsymbol{\theta})$ is $c(\hat{\boldsymbol{\theta}})$ with standard error $se[c(\widehat{\boldsymbol{\theta}})] = c'(\widehat{\boldsymbol{\theta}})^T(\widehat{\boldsymbol{V}}^T\widehat{\boldsymbol{V}})^{-1}c'(\widehat{\boldsymbol{\theta}})$ where $c'(\widehat{\boldsymbol{\theta}}) = \partial c(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ evaluated at $\hat{\boldsymbol{\theta}}$. Using these, we can construct the Wald confidence interval for $c(\boldsymbol{\theta})$ corresponding to (2.5) having the form

$$\left\{\phi \,\middle|\, |\phi - \widehat{\phi}|/s\sqrt{c'(\widehat{\boldsymbol{\theta}})^T(\widehat{\boldsymbol{V}}^T\widehat{\boldsymbol{V}})^{-1}c'(\widehat{\boldsymbol{\theta}})} \le t_{\alpha/2}(n-p)\right\} \tag{2.6}$$

where $\widehat{\phi} = c(\widehat{\boldsymbol{\theta}})$. The likelihood confidence interval corresponding to (2.2) is

$$\left\{\phi \,\middle|\, \sqrt{[S(\boldsymbol{m}_{\boldsymbol{\theta}}(\phi)) - S(\widehat{\boldsymbol{\theta}})]/s^2} \le t_{\alpha/2}(n-p)\right\} \tag{2.7}$$

where $\boldsymbol{m}_{\boldsymbol{\theta}}(\phi)$ denotes the p-dimensional vector valued function that minimizes $S(\boldsymbol{\theta})$ over $\boldsymbol{\theta}$ for given $c(\boldsymbol{\theta}) = \phi$.

Following Cook and Weisberg (1990), we will consider constructing confidence curves for the function $c(\boldsymbol{\theta})$. In this plot, ϕ is on the vertical axis and the plotted point corresponding to on the horizontal axis is of the form $\sqrt{[S(\boldsymbol{m}_{\boldsymbol{\theta}}(\phi)) - S(\widehat{\boldsymbol{\theta}})]/s^2}$. To this plot, two straight lines passing through $(0, \widehat{\phi})$ with slope $\pm se(\widehat{\phi})$ are added. These straight lines represent the Wald interval.

2.3. Parameter transformation for improved linear approximation

Bates and Watts (1981) have discussed the problems of reparameterizing nonlinear models so as to obtain accurate linear approximation confidence regions for the parameters. A result is derived which expresses the parameter-effects under a reparameterization $c(\boldsymbol{\theta})$ in terms of the parameter-effects array for the original parameters and the first and second derivatives of the transformation.

This result can be used to determine reparameterizations for which the new parameters will have small parameter-effects curvature. This implies that in the new parameters the linear approximation confidence region should be accurate, under the assumption that higher-order derivatives are small.

A similar problem exists with computing confidence curves for fitted values or predictions. The statistical behavior of fitted values does not depend on the parameterization. Thus, as long as the intrinsic curvature is small, standard Wald intervals for $f(\boldsymbol{x};\boldsymbol{\theta})$ are likely to be sufficient (Seber and Wild; 1989).

3. Fitted values in nonlinear regression

In regression analysis, one of the major goals usually is to estimate the mean for one or more probability distributions of y. With nonlinear regression, the mean function is $E(y|\boldsymbol{x}) = f(\boldsymbol{x};\boldsymbol{\theta})$. Given estimates $\hat{\boldsymbol{\theta}}$, the estimated mean function is $\widehat{E}(y|\boldsymbol{x}) = f(\boldsymbol{x};\widehat{\boldsymbol{\theta}})$. This equation for estimated mean function can be evaluated at any value of $\boldsymbol{x}$. When it is evaluated at the observed values $\boldsymbol{x}_i$ of $\boldsymbol{x}$, we get the fitted values

$$\widehat{y}_i = \widehat{E}(y|\boldsymbol{x} = \boldsymbol{x}_i) = f(\boldsymbol{x}_i; \widehat{\boldsymbol{\theta}}) \tag{3.1}$$

for the estimate of $E(y|\boldsymbol{x}_i) = E(y|\boldsymbol{x} = \boldsymbol{x}_i) = f(\boldsymbol{x}_i; \boldsymbol{\theta})$.

The sampling distribution of $\hat{y}_i$ refers to the different values of $\hat{y}_i$ which would be obtained if repeated samples were selected, each holding the levels of the independent variable $\boldsymbol{x}$ constant, and calculating for each sample.

Using the asymptotic linearization of (2.1), we can apply existing linear methods to finding interval for the mean function at $\boldsymbol{x} = \boldsymbol{x}_i$. Since for large n, $\widehat{\boldsymbol{\theta}}$ is close to the true value $\boldsymbol{\theta}$, we have the usual Taylor expansion

$$f(\boldsymbol{x}_i; \widehat{\boldsymbol{\theta}}) \approx f(\boldsymbol{x}_i; \boldsymbol{\theta}) + \boldsymbol{v}_i^T(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \tag{3.2}$$

where $\boldsymbol{v}_i^T = (\partial f_i/\partial\theta_1, ..., \partial f_i/\partial\theta_p)\,|_{\widehat{\boldsymbol{\theta}}}$ is the i-th row of $\hat{\boldsymbol{V}}$.

From the asymptotic linearization result $\widehat{\boldsymbol{\theta}} \sim N_p(\boldsymbol{\theta}, \sigma^2(\widehat{\boldsymbol{V}}^T\widehat{\boldsymbol{V}})^{-1})$, we get $E(\widehat{y}_i) \approx E(y|\boldsymbol{x}_i)$ and $var(\widehat{y}_i) \approx \sigma^2\boldsymbol{v}_i^T(\widehat{\boldsymbol{V}}^T\widehat{\boldsymbol{V}})^{-1}\boldsymbol{v}_i$. And $\hat{y}_i$ is asymptotically $N(E(y|\boldsymbol{x}_i), \sigma^2\boldsymbol{v}_i^T(\widehat{\boldsymbol{V}}^T\widehat{\boldsymbol{V}})^{-1}\boldsymbol{v}_i)$. Hence, asymptotically, $[\widehat{y}_i - E(y|\boldsymbol{x}_i)]/se(\widehat{y}_i)$ is distributed as $t(n-p)$. A confidence interval for $E(y|\boldsymbol{x}_i)$ is constructed in the standard fashion. The limits are:

$$\widehat{y}_i \pm t_{\alpha/2}(n-p)\sqrt{s^2\boldsymbol{v}_i^T(\widehat{\boldsymbol{V}}^T\widehat{\boldsymbol{V}})^{-1}\boldsymbol{v}_i} \tag{3.3}$$

and an approximate $100(1-\alpha)\%$ confidence interval for $E(y|\boldsymbol{x}_i)$ corresponding to (2.6) is is therefore given by

$$\left\{\phi \,\middle|\, |\phi - \widehat{y}_i| / \sqrt{s^2\boldsymbol{v}_i^T(\widehat{\boldsymbol{V}}^T\widehat{\boldsymbol{V}})^{-1}\boldsymbol{v}_i} \le t_{\alpha/2}(n-p)\right\} \tag{3.4}$$

Since mean response is a function of $\boldsymbol{\theta}$, that is $c(\boldsymbol{\theta}) = E(y|\boldsymbol{x}_i) = f(\boldsymbol{x}_i; \boldsymbol{\theta})$, the likelihood confidence interval corresponding to (2.7) is

$$\left\{\phi \,\middle|\, \sqrt{[S(\boldsymbol{m}_{\boldsymbol{\theta}}(\phi)) - S(\widehat{\boldsymbol{\theta}})]/s^2} \le t_{\alpha/2}(n-p)\right\}$$

where $\boldsymbol{m}_{\boldsymbol{\theta}}(\phi)$ denotes the p-dimensional vector valued function that minimizes $S(\boldsymbol{\theta})$ over $\boldsymbol{\theta}$ for given $\phi = c(\boldsymbol{\theta}) = E(y|\boldsymbol{x}_i) = f(\boldsymbol{x}_i; \boldsymbol{\theta})$.

4. Example

We consider the data on biochemical oxygen demand (BOD) from Bates and Watts (1988), reproduced in Table 4.1. The model was derived based on exponential decay with a fixed rate constant as

$$f(x, \boldsymbol{\theta}) = \theta_1(1 - e^{-\theta_2 x})$$

where y is predicted biochemical oxygen demand and x is time.

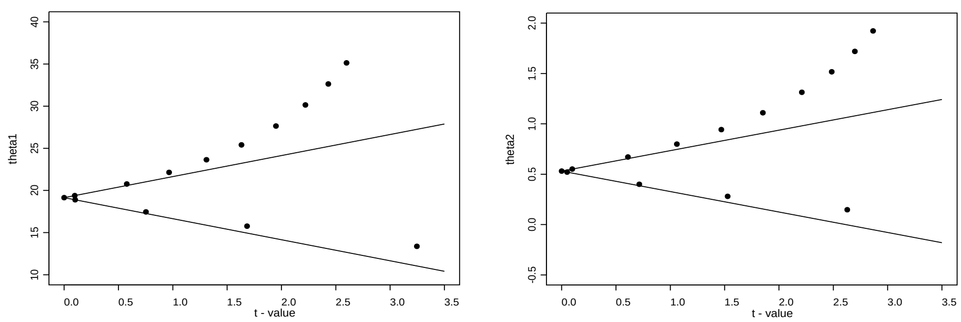
Table 4.1 BOD data

x	y
1	8.3
2	10.3
3	19.0
4	16.0
5	15.6
7	19.8

For this data. we calculated the maximum parameter-effects curvature $\Gamma^{\boldsymbol{\tau}}(\boldsymbol{\theta}) = 2.172$ and the maximum intrinsic curvature $\Gamma^{\boldsymbol{\eta}}(\boldsymbol{\theta}) = .3011$. The corresponding guide is, $1/[\sqrt{F_{.05}(2,4)}]$ =. 3794 . The maximum parameter-effects curvature is judged to be large, so that linear approximation inference will be completely unreliable. The subset curvatures proposed by Cook and Goldberg (1985) which indicate the validity of linear approximation for single parameters. The subset the maximum parameter-effects and intrinsic curvatures for θ_1 and θ_2 are $\Gamma_s^{\boldsymbol{\tau}}(\theta_1) = .3461$, $\Gamma_s^{\boldsymbol{\eta}}(\theta_1) = .4239$, $\Gamma_s^{\boldsymbol{\tau}}(\theta_2) = .3461$, $\Gamma_s^{\boldsymbol{\eta}}(\theta_2) = .4239$. The corresponding guide is $1/[2\sqrt{F.05(1,4)}] = .1801$, indicating that estimations, inferences, and diagnostics for single parameters based on the linear approximation may give misleading results.

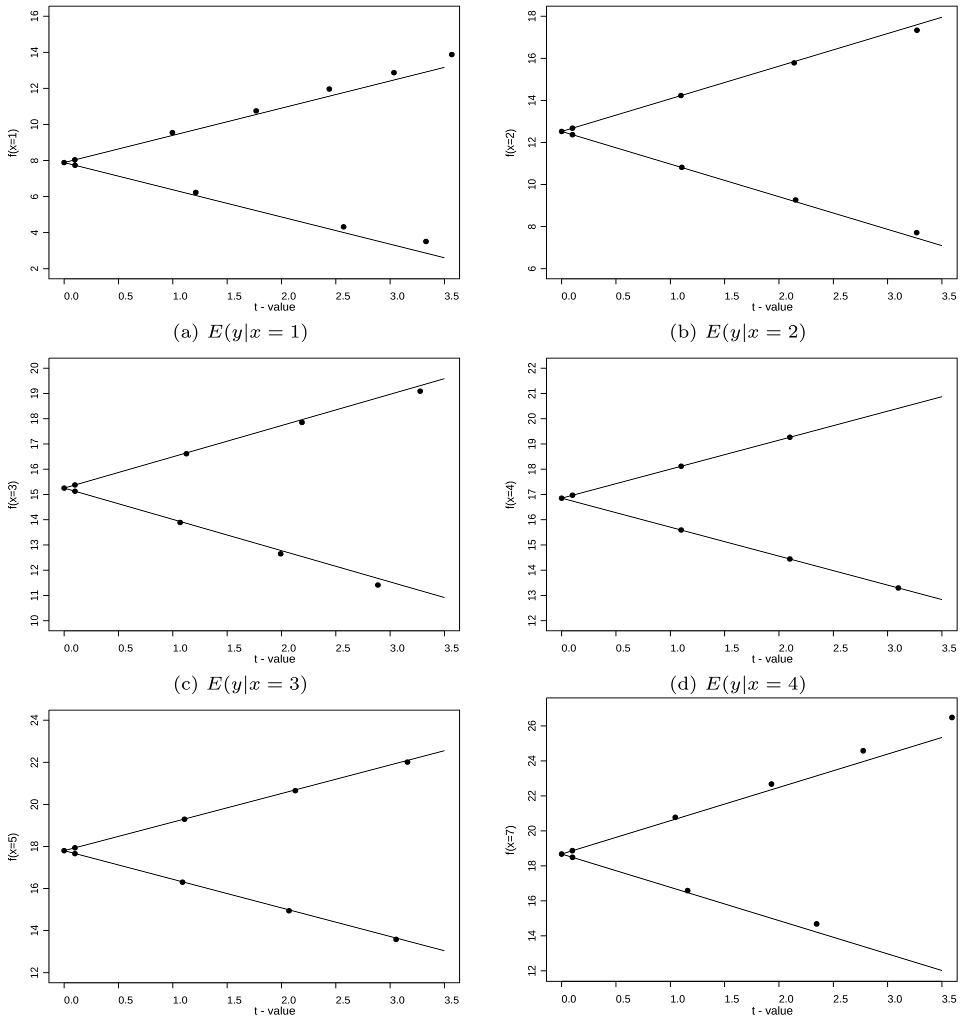
Now we consider the confidence curves discussed in section 2.2. The confidence curves θ_1 and θ_2 are given in Figure 4.1. The solid lines are Wald confidence curves, and the curves traced by the stars are likelihood confidence curves. In these figures, the likelihood confidence curves are severely curved and dissimilar to the Wald confidence curves, which demonstrates that linear approximation inference will be unsatisfactory, as was noted from the curvature measures.

Since the maximum intrinsic curvature $\Gamma^{\boldsymbol{\eta}}(\boldsymbol{\theta}) = .3011$ is small, the likelihood confidence curves for $E(y|\boldsymbol{x}_i)$ should be sufficiently close to the Wald confidence curves. The confidence curves for $E(y|\boldsymbol{x}_i) = f(\boldsymbol{x}_i;\boldsymbol{\theta})$, $i = 1, \cdots, 6$ are given in Figure 4.2. In these figures, The Wald intervals seem clearly acceptable for any confidence level. Thus the linear approximation seems to be excellent.



(a) θ_1 (b) θ_2

Figure 4.1 Confidence curves for θ_1 and θ_2



(a) $E(y|x=1)$ (b) $E(y|x=2)$

(c) $E(y|x=3)$ (d) $E(y|x=4)$

(e) $E(y|x=5)$ (f) $E(y|x=7)$

Figure 4.2 Confidence curves for $E(y|\boldsymbol{x}_i),\ i=1,\cdots,6$

5. Remarks

Accuracy of linear approximation for fitted values was investigated using various curvature measures in nonlinear regression. These curvatures appear to be reliable indicators of the adequacy of the linearization inference. Also, we considered obtaining graphical summaries of uncertainty in estimating fitted values through confidence curves. The results are applied to the problem of assessing the accuracy of linear approximation for fitted values.

References

Bates, D. M. and Watts, D. G. (1980). Relative curvature measures of nonlinearity (with discussion). *Journal of the Royal Statistical Society 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 B*, **22**, 41-88.

Cook, R. D. and Goldberg, M. L. (1986). Curvatures for parameter subsets in nonlinear regression. *The Annals of Statistics*, **14**, 1399-1418.

Cook, R. D. and Weisberg, S. (1990). Confidence curves in nonlinear regression. *Journal of the American Statistical Association*, **85**, 544-551.

Hamilton, D. (1986). Confidence regions for parameter subsets in nonlinear regression. *Biometrika*, **73**, 57-64.

Hamilton, D., 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. (2003). Confidence curves for a function of parameters in nonlinear regression. *Journal of the Korean Statistical Society*, **32**, 1-10.

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

Kahng, M. (2009). Effects of curvature on leverage in nonlinear regression. *Journal of the Korean Data & Information Science Society*, **20**, 913-917.

Seber, G. A. F. and Wild, C. J. (1989). *Nonlinear regression*, John Wiley and Sons, New York.