Confidence intervals in a regression with both linear and non-linear terms

Robert Davies

Statistics Research Associates
Wellington, New Zealand

Christopher Withers

Applied Mathematics Group
Industrial Research Limited
Lower Hutt, New Zealand

and

Saralees Nadarajah

School of Mathematics
University of Manchester
Manchester, UK

Abstract: We present a simple way for calculating confidence intervals for a class of scalar functions of the parameters in least squares estimation when there are linear together with a small number of non-linear terms. We do not assume normality.

Contents

1. Introduction

Many inferential methods for non-linear models have been developed. For most excellent accounts, we refer the readers to Bates and Watts [1], Carroll and Ruppert [3], Seber and Wild [15], Lindsey [13], Huet *et al.* [11], Carroll *et al.* [4] and references therein.

Often one wishes to fit models that contain both linear and non-linear terms. Such models arise in many situations. We mention: stiffness and friction in biological rhythmic movements (Beek *et al.* [2]); data precision in optimal regression models (Shacham and Brauner [16]); local chlorophyll concentration from satellite images (Trentin *et al.* [19]); and models of total and presumed wildlife sources (Siewicki *et al.* [17]).

Suppose we are fitting a model involving both linear parameters and one or two non-linear parameters by least squares. A convenient approach is to use a linear least squares program to fit the model for a variety of values of the non-linear parameters and then find confidence intervals for the non-linear parameters from the graph of the residual sum of squares plotted against the trial values of the non-linear parameters. In this note, we derive an easy way of finding confidence intervals for the other parameters, considered one at a time, or certain functions of them from the graph of the residual sum of squares.

The purpose here is not to achieve high accuracy. There are of course many methods for finding highly accurate confidence intervals for non-linear models. These methods are based on complicated and computationally expensive methods like bootstrapping and Markov Chain Monte Carlo (see, for example, Hu and Kalbfleisch [10] and Tian *et al.* [18]). Besides, the methods are relative new, and software for their implementation may not be widely available in the public domain - for example, the authors could not find references for Tian *et al.* [18] in the widely used R statistical package.

Here, we derive a simple method for finding confidence intervals that can be implemented easily in almost any platform, possibly even using a hand calculator. We show that our method is reasonably accurate. We compare our method with the delta method (Casella and Berger [5]), another simple and commonly used method for constructing confidence intervals for non-linear functions. We establish that ours is far superior.

Suppose our data is represented by the n-dimensional vector

$$\mathbf{Y} = \mathbf{f}\left(\mathbf{X_\theta}, \boldsymbol{\beta}\right) + \boldsymbol{\varepsilon},$$

where $\mathbf{f}(\cdot, \cdot)$ is a real valued $n \times 1$ function; $\mathbf{X_\theta}$ is an $n \times m$ matrix function of $\boldsymbol{\theta}$; $\boldsymbol{\beta}$ is an m-dimensional vector of unknown *linear* parameters, *linear* in the sense that $\mathbf{f}(\mathbf{X_\theta}, \boldsymbol{\beta}) = \mathbf{f}(\mathbf{X_\theta'}\boldsymbol{\beta})$, where $\mathbf{A}'$ denotes the transpose of a matrix or vector $\mathbf{A}$; and $\boldsymbol{\varepsilon}$ is an n-dimensional vector of uncorrelated random variables with zero expectation and variance σ^2. We do not assume normality. The vector, $\boldsymbol{\theta}$, consists of s unknown *non-linear* parameters, *non-linear* is the sense that $\mathbf{Y}$ is a non-linear function of $\boldsymbol{\theta}$. We wish to find a confidence interval (Λ_-, Λ_+) for a function

$$\Lambda = g\left(\boldsymbol{\lambda_\theta}, \boldsymbol{\beta}\right), \tag{1}$$

where $g(\cdot, \cdot)$ is a real valued scalar function; $\boldsymbol{\lambda_\theta}$ is an m-dimensional function of $\boldsymbol{\theta}$. Three possible examples of (1) are now given.

Example 1.1. A single element of $\boldsymbol{\beta}$, say β_i, $1 \leq i \leq m$. Take $g(\boldsymbol{\lambda_\theta}, \boldsymbol{\beta}) = \boldsymbol{\lambda'_\theta}\boldsymbol{\beta}$ and $\boldsymbol{\lambda_\theta}$ to be an $m \times 1$ vector of zeros except for a one at the ith position. Then $\Lambda = \beta_i$.

Example 1.2. Predicted value, $a + c(x - \theta)^+$, of the model

$$y_i = a + c\,(x_i - \theta)^+ + \epsilon_i$$

for $i = 1, 2, \ldots, n$ and for a given x, where ϵ_i are independent errors. This is a particular case of the two phase regression model due to Hawkins [8] and Davies [6], see Section 3 for details. Here, $(x)^+ = 0$ if $x \leq 0$, $(x)^+ = x$ if $x \geq 0$ and a, c, θ are unknown scalar parameters, so $m = 2$ and $s = 1$. Take $g(\boldsymbol{\lambda_\theta}, \boldsymbol{\beta}) = \boldsymbol{\lambda'_\theta}\boldsymbol{\beta}$, $\boldsymbol{\beta} = (a, c)'$, and $\boldsymbol{\lambda_\theta} = (1, (x - \theta)^+)'$. Then $\Lambda = a + c(x - \theta)^+$.

Example 1.3. Predicted value, $a + bx + c(x - \theta)^+$, of the model

$$y_i = a + bx_i + c\,(x_i - \theta)^+ + \epsilon_i$$

for $i = 1, 2, \ldots, n$ and for a given x, where ϵ_i are independent errors. This is the two phase regression model due to Hawkins [8] and Davies [6], see Section 3 for details. Here, a, b, c, θ are unknown scalar parameters, so $m = 3$ and $s = 1$. Take $g(\boldsymbol{\lambda_\theta}, \boldsymbol{\beta}) = \boldsymbol{\lambda'_\theta}\boldsymbol{\beta}$, $\boldsymbol{\beta} = (a, b, c)'$, and $\boldsymbol{\lambda_\theta} = (1, x, (x - \theta)^+)'$. Then $\Lambda = a + bx + c(x - \theta)^+$.

Often $\mathbf{f}(\cdot, \cdot)$, $g(\cdot, \cdot)$, $\mathbf{X_\theta}$ and $\boldsymbol{\lambda_\theta}$ are highly non-linear and non-differentiable (Kaniovskaya [12]; Manski [14], Chapter 9; [9]). In these cases, we might wish to base our confidence interval directly on the residual sum of squares function $R_{\boldsymbol{\theta},\boldsymbol{\beta}} = \|\mathbf{Y} - \mathbf{f}(\mathbf{X_\theta}, \boldsymbol{\beta})\|^2$, where $\| \cdot \|$ denotes the l_2 norm of a vector. See Donaldson and Schnabel [7].

Define

$$S_m^2 = \min_{\boldsymbol{\theta},\boldsymbol{\beta}} R_{\boldsymbol{\theta},\boldsymbol{\beta}}.$$

Let

$$\widehat{\sigma}^2 = S_m^2/(n - m - s),$$

$$\Delta = S_m^2 + \widehat{\sigma}^2 F_{1,n-m-s}(\alpha),$$

where α denotes the confidence level and $F_{p,q}(\alpha)$ denotes the α probability level of an $F_{p,q}$ distribution. Then an approximate confidence set for Λ may be defined using the least squares analogue of the profile likelihood:

$$\left\{ \lambda : \min_{\boldsymbol{\theta},\boldsymbol{\beta},g(\boldsymbol{\lambda_\theta},\boldsymbol{\beta})=\lambda} (R_{\boldsymbol{\theta},\boldsymbol{\beta}}) < \Delta \right\}. \tag{2}$$

Typically, this set will be an interval (Λ_-, Λ_+), where Λ_- and Λ_+ can be found from (2) if the inequality is replaced by an equality.

In Section 2, we describe the method for finding (Λ_-, Λ_+) and illustrate it for Examples 1.1–1.3. In Section 3, some numerical studies are discussed to assess the performance of the confidence interval.

2. Main results

Theorem 2.1 derives the confidence set for (1) for general $\mathbf{f}(\mathbf{X}_{\boldsymbol{\theta}}, \boldsymbol{\beta})$ and $g(\boldsymbol{\lambda}_{\boldsymbol{\theta}}, \boldsymbol{\beta})$. Theorem 2.2 illustrates the proof of Theorem 2.1 for the particular case $\mathbf{f}(\mathbf{X}_{\boldsymbol{\theta}}, \boldsymbol{\beta}) = \mathbf{X}_{\boldsymbol{\theta}}\boldsymbol{\beta}$ and $g(\boldsymbol{\lambda}_{\boldsymbol{\theta}}, \boldsymbol{\beta}) = \boldsymbol{\lambda}'_{\boldsymbol{\theta}}\boldsymbol{\beta}$. In this case, define $\boldsymbol{\beta}_{\boldsymbol{\theta}} = \boldsymbol{A}_{\boldsymbol{\theta}}^{-1}\mathbf{X}'_{\boldsymbol{\theta}}\mathbf{Y}$, $R_{\boldsymbol{\theta}} = \mathbf{Y}'(\mathbf{I} - \mathbf{X}_{\boldsymbol{\theta}}\mathbf{A}_{\boldsymbol{\theta}}^{-1}\mathbf{X}'_{\boldsymbol{\theta}})\mathbf{Y}$ and $a_{\boldsymbol{\theta}} = \boldsymbol{\lambda}'_{\boldsymbol{\theta}}\mathbf{A}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\lambda}_{\boldsymbol{\theta}}$, where $\mathbf{A}_{\boldsymbol{\theta}} = \mathbf{X}'_{\boldsymbol{\theta}}\mathbf{X}_{\boldsymbol{\theta}}$ is assumed to be non-singular for all $\boldsymbol{\theta}$. The general proof of Theorem 2.1 is similar to that given.

Theorem 2.1. *Assume* $\mathbf{f}(\cdot, \cdot)$ *and* $g(\cdot, \cdot)$ *are differentiable. Then the confidence set for (1) defined by (2) is* $\cup_{\boldsymbol{\theta}} I_{\boldsymbol{\theta}}$, *where* $I_{\boldsymbol{\theta}}$ *is the open interval*

$$I_{\boldsymbol{\theta}} = \left\{ \lambda : \mathbf{Y}'(\mathbf{Y} - \mathbf{f}) < \Delta + \gamma \mathbf{f}' \frac{\partial g(\boldsymbol{\lambda}_{\boldsymbol{\theta}}, \boldsymbol{\beta})}{\partial \mathbf{f}} \right\},$$

where $\mathbf{f} = \mathbf{f}(\mathbf{X}_{\boldsymbol{\theta}}, \boldsymbol{\beta})$ *and* γ *are the simultaneous solutions of the equations* $\mathbf{Y} - \mathbf{f} = \gamma (\partial g(\boldsymbol{\lambda}_{\boldsymbol{\theta}}, \boldsymbol{\beta}) / \partial \mathbf{f})$, *and* $g(\boldsymbol{\lambda}_{\boldsymbol{\theta}}, \boldsymbol{\beta}) = \lambda$.

Theorem 2.2. *Let* $\mathbf{f}(\mathbf{X}_{\boldsymbol{\theta}}, \boldsymbol{\beta}) = \mathbf{X}_{\boldsymbol{\theta}}\boldsymbol{\beta}$ *and* $g(\boldsymbol{\lambda}_{\boldsymbol{\theta}}, \boldsymbol{\beta}) = \boldsymbol{\lambda}'_{\boldsymbol{\theta}}\boldsymbol{\beta}$. *Then the confidence set for (1) defined by (2) is* $\cup_{\boldsymbol{\theta}} I_{\boldsymbol{\theta}}$, *where* $I_{\boldsymbol{\theta}}$ *is the open interval*

$$I_{\boldsymbol{\theta}} = \left(\boldsymbol{\lambda}'_{\boldsymbol{\theta}}\boldsymbol{\beta}_{\boldsymbol{\theta}} - \{a_{\boldsymbol{\theta}}(\Delta - R_{\boldsymbol{\theta}})\}^{1/2}, \boldsymbol{\lambda}'_{\boldsymbol{\theta}}\boldsymbol{\beta}_{\boldsymbol{\theta}} + \{a_{\boldsymbol{\theta}}(\Delta - R_{\boldsymbol{\theta}})\}^{1/2} \right)$$

and the union is taken over all $\boldsymbol{\theta}$ *with* $R_{\boldsymbol{\theta}} < \Delta$.

If $\boldsymbol{\lambda}_{\boldsymbol{\theta}}$ *and* $\mathbf{X}_{\boldsymbol{\theta}}$ *are continuous functions of* $\boldsymbol{\theta}$ *and* $\{\boldsymbol{\theta} : R_{\boldsymbol{\theta}} < \Delta\}$ *is connected then the confidence set will be an interval* (Λ_-, Λ_+), *where*

$$\Lambda_- = \min_{\boldsymbol{\theta} : R_{\boldsymbol{\theta}} < \Delta} \left[\boldsymbol{\lambda}'_{\boldsymbol{\theta}}\boldsymbol{\beta}_{\boldsymbol{\theta}} - \{a_{\boldsymbol{\theta}}(\Delta - R_{\boldsymbol{\theta}})\}^{1/2} \right], \tag{3}$$

$$\Lambda_+ = \max_{\boldsymbol{\theta} : R_{\boldsymbol{\theta}} < \Delta} \left[\boldsymbol{\lambda}'_{\boldsymbol{\theta}}\boldsymbol{\beta}_{\boldsymbol{\theta}} + \{a_{\boldsymbol{\theta}}(\Delta - R_{\boldsymbol{\theta}})\}^{1/2} \right]. \tag{4}$$

Proof. We want to find the set

$$\left[\lambda : \min_{\boldsymbol{\theta}} \left\{ \min_{\boldsymbol{\beta}} (R_{\boldsymbol{\theta}, \boldsymbol{\beta}} : \boldsymbol{\lambda}'_{\boldsymbol{\theta}}\boldsymbol{\beta} = \lambda) \right\} < \Delta \right]. \tag{5}$$

Using Lagrange multipliers to evaluate the inner minimum: we need to solve $\mathbf{X}'_{\boldsymbol{\theta}}(\mathbf{Y} - \mathbf{X}_{\boldsymbol{\theta}}\boldsymbol{\beta}) - \gamma \boldsymbol{\lambda}_{\boldsymbol{\theta}} = \mathbf{0}$ and $\boldsymbol{\lambda}'_{\boldsymbol{\theta}}\boldsymbol{\beta} = \lambda$. So,

$$\begin{pmatrix} \boldsymbol{\beta} \\ \gamma \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{\boldsymbol{\theta}} & \boldsymbol{\lambda}_{\boldsymbol{\theta}} \\ \boldsymbol{\lambda}'_{\boldsymbol{\theta}} & 0 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}'_{\boldsymbol{\theta}}\mathbf{Y} \\ \lambda \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{A}_{\boldsymbol{\theta}}^{-1} - \mathbf{A}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\lambda}_{\boldsymbol{\theta}}\boldsymbol{\lambda}'_{\boldsymbol{\theta}}\mathbf{A}_{\boldsymbol{\theta}}^{-1}/a_{\boldsymbol{\theta}} & \mathbf{A}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\lambda}_{\boldsymbol{\theta}}/a_{\boldsymbol{\theta}} \\ \boldsymbol{\lambda}'_{\boldsymbol{\theta}}\mathbf{A}_{\boldsymbol{\theta}}^{-1}/a_{\boldsymbol{\theta}} & -1/a_{\boldsymbol{\theta}} \end{pmatrix} \begin{pmatrix} \mathbf{X}'_{\boldsymbol{\theta}}\mathbf{Y} \\ \lambda \end{pmatrix}.$$

That is,

$$\boldsymbol{\beta} = \boldsymbol{\beta_\theta} - \mathbf{A}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\lambda_\theta}\boldsymbol{\lambda}_{\boldsymbol{\theta}}'\boldsymbol{\beta_\theta}/a_{\boldsymbol{\theta}} + \mathbf{A}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\lambda_\theta}\lambda/a_{\boldsymbol{\theta}},$$

and

$$\gamma = \left(\boldsymbol{\lambda}_{\boldsymbol{\theta}}'\boldsymbol{\beta_\theta} - \lambda\right)/a_{\boldsymbol{\theta}}.$$

So, the inner minimum of (5) becomes

$$
\begin{aligned}
\left(\mathbf{Y} - \mathbf{X_\theta}\boldsymbol{\beta}\right)'\left(\mathbf{Y} - \mathbf{X_\theta}\boldsymbol{\beta}\right) &= \mathbf{Y}'\left(\mathbf{Y} - \mathbf{X_\theta}\boldsymbol{\beta}\right) - \boldsymbol{\beta}'\mathbf{X}_{\boldsymbol{\theta}}'\left(\mathbf{Y} - \mathbf{X_\theta}\boldsymbol{\beta}\right) \\
&= \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X_\theta}\boldsymbol{\beta} - \gamma\boldsymbol{\beta}'\boldsymbol{\lambda_\theta} \\
&= \mathbf{Y}'\mathbf{Y} \\
&\quad - \mathbf{Y}'\mathbf{X_\theta}\left(\boldsymbol{\beta_\theta} - \mathbf{A}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\lambda_\theta}\boldsymbol{\lambda}_{\boldsymbol{\theta}}'\boldsymbol{\beta_\theta}/a_{\boldsymbol{\theta}} + \mathbf{A}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\lambda_\theta}\lambda/a_{\boldsymbol{\theta}}\right) \\
&\quad - \gamma\lambda \\
&= \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X_\theta}\boldsymbol{\beta_\theta} + \boldsymbol{\beta}_{\boldsymbol{\theta}}'\boldsymbol{\lambda_\theta}\boldsymbol{\lambda}_{\boldsymbol{\theta}}'\boldsymbol{\beta_\theta}/a_{\boldsymbol{\theta}} - \boldsymbol{\beta}_{\boldsymbol{\theta}}'\boldsymbol{\lambda_\theta}\lambda/a_{\boldsymbol{\theta}} \\
&\quad - \left(\boldsymbol{\lambda}_{\boldsymbol{\theta}}'\boldsymbol{\beta_\theta} - \lambda\right)\lambda/a_{\boldsymbol{\theta}} \\
&= \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X_\theta}\boldsymbol{\beta_\theta} + \left(\boldsymbol{\lambda}_{\boldsymbol{\theta}}'\boldsymbol{\beta_\theta} - \lambda\right)^2/a_{\boldsymbol{\theta}} \\
&= \mathbf{Y}'\left(\mathbf{I} - \mathbf{X_\theta}\mathbf{A}_{\boldsymbol{\theta}}^{-1}\mathbf{X}_{\boldsymbol{\theta}}'\right)\mathbf{Y} + \left(\boldsymbol{\lambda}_{\boldsymbol{\theta}}'\boldsymbol{\beta_\theta} - \lambda\right)^2/a_{\boldsymbol{\theta}} \\
&= R_{\boldsymbol{\theta}} + \left(\boldsymbol{\lambda}_{\boldsymbol{\theta}}'\boldsymbol{\beta_\theta} - \lambda\right)^2/a_{\boldsymbol{\theta}},
\end{aligned}
$$

so that (5) becomes

$$
\begin{aligned}
&\left[\lambda : \min_{\boldsymbol{\theta}}\left\{R_{\boldsymbol{\theta}} + \left(\boldsymbol{\lambda}_{\boldsymbol{\theta}}'\boldsymbol{\beta_\theta} - \lambda\right)^2/a_{\boldsymbol{\theta}}\right\} < \Delta\right] \\
={}& \left[\lambda : R_{\boldsymbol{\theta}} + \left(\boldsymbol{\lambda}_{\boldsymbol{\theta}}'\boldsymbol{\beta_\theta} - \lambda\right)^2/a_{\boldsymbol{\theta}} < \Delta \text{ for at least some } \boldsymbol{\theta}\right] \\
={}& \cup_{\boldsymbol{\theta}}\left\{\lambda : R_{\boldsymbol{\theta}} + \left(\boldsymbol{\lambda}_{\boldsymbol{\theta}}'\boldsymbol{\beta_\theta} - \lambda\right)^2/a_{\boldsymbol{\theta}} < \Delta\right\} \\
={}& \cup_{\boldsymbol{\theta}}\left[\lambda : \left|\boldsymbol{\lambda}_{\boldsymbol{\theta}}'\boldsymbol{\beta_\theta} - \lambda\right| < \left\{a_{\boldsymbol{\theta}}\left(\Delta - R_{\boldsymbol{\theta}}\right)\right\}^{1/2}\right],
\end{aligned}
$$

where the union is over values of $\boldsymbol{\theta}$ with $R_{\boldsymbol{\theta}} < \Delta$. This equals $\cup_{\boldsymbol{\theta}}I_{\boldsymbol{\theta}}$ as required. If $\boldsymbol{\lambda_\theta}$ and $\mathbf{X_\theta}$ are continuous functions of $\boldsymbol{\theta}$ and $\{\boldsymbol{\theta} : R_{\boldsymbol{\theta}} < \Delta\}$ is connected; then so is $\{\boldsymbol{\lambda}_{\boldsymbol{\theta}}'\boldsymbol{\beta_\theta} : R_{\boldsymbol{\theta}} < \Delta\}$ and hence so is our confidence set. In this case, it is an interval with end points (3) and (4). This completes the proof. $\square$

We now return to the three examples discussed in Section 1. We show how the results of Theorem 2.2 apply to each of the examples. Throughout, we denote by A^{ij} the (i,j)th element of $\mathbf{A}^{-1}$.

Example 2.1. Consider Example 1.1. Then the confidence set for β_i is (Λ_-, Λ_+), where

$$\Lambda_- = \min_{\boldsymbol{\theta}:\mathbf{Y}'\left(\mathbf{I}-\mathbf{X}_{\boldsymbol{\theta}}\mathbf{A}_{\boldsymbol{\theta}}^{-1}\mathbf{X}_{\boldsymbol{\theta}}'\right)\mathbf{Y}<\Delta} \left[\beta_{\boldsymbol{\theta},i} - \left\{A_{\boldsymbol{\theta}}^{ii}\left(\Delta - \mathbf{Y}'\left(\mathbf{I} - \mathbf{X}_{\boldsymbol{\theta}}\mathbf{A}_{\boldsymbol{\theta}}^{-1}\mathbf{X}_{\boldsymbol{\theta}}'\right)\mathbf{Y}\right)\right\}^{1/2}\right],$$

and

$$\Lambda_+ = \max_{\boldsymbol{\theta}:\mathbf{Y}'\left(\mathbf{I}-\mathbf{X}_{\boldsymbol{\theta}}\mathbf{A}_{\boldsymbol{\theta}}^{-1}\mathbf{X}_{\boldsymbol{\theta}}'\right)\mathbf{Y}<\Delta} \left[\beta_{\boldsymbol{\theta},i} + \left\{A_{\boldsymbol{\theta}}^{ii}\left(\Delta - \mathbf{Y}'\left(\mathbf{I} - \mathbf{X}_{\boldsymbol{\theta}}\mathbf{A}_{\boldsymbol{\theta}}^{-1}\mathbf{X}_{\boldsymbol{\theta}}'\right)\mathbf{Y}\right)\right\}^{1/2}\right],$$

where $\mathbf{A}_{\boldsymbol{\theta}} = \mathbf{X}_{\boldsymbol{\theta}}'\mathbf{X}_{\boldsymbol{\theta}}$ and $\boldsymbol{\beta}_{\boldsymbol{\theta}} = \mathbf{A}_{\boldsymbol{\theta}}^{-1}\mathbf{X}_{\boldsymbol{\theta}}'\mathbf{Y}$. We have assumed that $\mathbf{X}_{\boldsymbol{\theta}}$ is a continuous function of $\boldsymbol{\theta}$ and that $\{\boldsymbol{\theta} : \mathbf{Y}'(\mathbf{I} - \mathbf{X}_{\boldsymbol{\theta}}\mathbf{A}_{\boldsymbol{\theta}}^{-1}\mathbf{X}_{\boldsymbol{\theta}}')\mathbf{Y} < \Delta\}$ is connected.

Example 2.2. Consider Example 1.2. Then the confidence set for $a + c(x - \theta)^+$ is (Λ_-, Λ_+), where

$$\Lambda_- = \min_{\theta:R_\theta<\Delta}\left\{A_\theta^{11}\sum_{i=1}^n y_i + A_\theta^{12}\sum_{i=1}^n (x_i - \theta)^+ y_i\right.$$
$$+(x-\theta)^+\left[A_\theta^{12}\sum_{i=1}^n y_i + A_\theta^{22}\sum_{i=1}^n (x_i - \theta)^+ y_i\right]$$
$$\left.- \left\{\left[A_\theta^{11} + (x-\theta)^2 A_\theta^{22} + 2(x-\theta)^+ A_\theta^{12}\right](\Delta - R_\theta)\right\}^{1/2}\right\},$$

and

$$\Lambda_+ = \max_{\theta:R_\theta<\Delta}\left\{A_\theta^{11}\sum_{i=1}^n y_i + A_\theta^{12}\sum_{i=1}^n (x_i - \theta)^+ y_i\right.$$
$$+(x-\theta)^+\left[A_\theta^{12}\sum_{i=1}^n y_i + A_\theta^{22}\sum_{i=1}^n (x_i - \theta)^+ y_i\right]$$
$$\left.+ \left\{\left[A_\theta^{11} + (x-\theta)^2 A_\theta^{22} + 2(x-\theta)^+ A_\theta^{12}\right](\Delta - R_\theta)\right\}^{1/2}\right\},$$

where

$$R_\theta = \sum_{i=1}^n y_i^2 - \left[A_\theta^{11}\left(\sum_{i=1}^n y_i\right)^2 + A_\theta^{22}\left(\sum_{i=1}^n (x_i - \theta)^+ y_i\right)^2\right.$$
$$\left.+ 2A_\theta^{12}\left(\sum_{i=1}^n y_i\right)\left(\sum_{i=1}^n (x_i - \theta)^+ y_i\right)\right],$$
$$A_\theta^{11} = \sum_{i=1}^n (x_i - \theta)^2 / D,$$

$$A_\theta^{12} = -\sum_{i=1}^{n} (x_i - \theta)^+ / D,$$

$$A_\theta^{22} = n/D,$$

where

$$D = n \sum_{i=1}^{n} (x_i - \theta)^2 - \left(\sum_{i=1}^{n} (x_i - \theta)^+ \right)^2.$$

We have assumed that $\{\theta : R_\theta < \Delta\}$ is connected.

Example 2.3. Consider Example 1.3. Then the confidence set for $a + bx + c(x - \theta)^+$ is (Λ_-, Λ_+), where

$$
\begin{aligned}
\Lambda_- \;=\; \min_{\theta : R_\theta < \Delta} \bigg\{ &- \Big[A_\theta^{11} + x^2 A_\theta^{22} + (x - \theta)^2 A_\theta^{33} + 2x A_\theta^{12} \\
&+ 2(x - \theta)^+ A_\theta^{13} + 2x(x - \theta)^+ A_\theta^{23} \Big]^{1/2} \left(\Delta - R_\theta \right)^{1/2} \\
&+ A_\theta^{11} \sum_{i=1}^{n} y_i + A_\theta^{12} \sum_{i=1}^{n} x_i y_i + A_\theta^{13} \sum_{i=1}^{n} (x_i - \theta)^+ y_i \\
&+ x \left[A_\theta^{12} \sum_{i=1}^{n} y_i + A_\theta^{22} \sum_{i=1}^{n} x_i y_i + A_\theta^{23} \sum_{i=1}^{n} (x_i - \theta)^+ y_i \right] \\
&+ (x - \theta)^+ \left[A_\theta^{13} \sum_{i=1}^{n} y_i + A_\theta^{23} \sum_{i=1}^{n} x_i y_i + A_\theta^{33} \sum_{i=1}^{n} (x_i - \theta)^+ y_i \right] \bigg\},
\end{aligned}
$$

and

$$
\begin{aligned}
\Lambda_+ \;=\; \max_{\theta : R_\theta < \Delta} \bigg\{ &\Big[A_\theta^{11} + x^2 A_\theta^{22} + (x - \theta)^2 A_\theta^{33} + 2x A_\theta^{12} \\
&+ 2(x - \theta)^+ A_\theta^{13} + 2x(x - \theta)^+ A_\theta^{23} \Big]^{1/2} \left(\Delta - R_\theta \right)^{1/2} \\
&+ A_\theta^{11} \sum_{i=1}^{n} y_i + A_\theta^{12} \sum_{i=1}^{n} x_i y_i + A_\theta^{13} \sum_{i=1}^{n} (x_i - \theta)^+ y_i \\
&+ x \left[A_\theta^{12} \sum_{i=1}^{n} y_i + A_\theta^{22} \sum_{i=1}^{n} x_i y_i + A_\theta^{23} \sum_{i=1}^{n} (x_i - \theta)^+ y_i \right] \\
&+ (x - \theta)^+ \left[A_\theta^{13} \sum_{i=1}^{n} y_i + A_\theta^{23} \sum_{i=1}^{n} x_i y_i + A_\theta^{33} \sum_{i=1}^{n} (x_i - \theta)^+ y_i \right] \bigg\},
\end{aligned}
$$

where

$$R_\theta = \sum_{i=1}^n y_i^2 - \left[A_\theta^{11} \left(\sum_{i=1}^n y_i \right)^2 + A_\theta^{22} \left(\sum_{i=1}^n x_i y_i \right)^2 \right.$$

$$+ A_\theta^{33} \left(\sum_{i=1}^n (x_i - \theta)^+ y_i \right)^2$$

$$+ 2A_\theta^{12} \left(\sum_{i=1}^n y_i \right) \left(\sum_{i=1}^n x_i y_i \right)$$

$$+ 2A_\theta^{13} \left(\sum_{i=1}^n y_i \right) \left(\sum_{i=1}^n (x_i - \theta)^+ y_i \right)$$

$$\left. + 2A_\theta^{23} \left(\sum_{i=1}^n x_i y_i \right) \left(\sum_{i=1}^n (x_i - \theta)^+ y_i \right) \right],$$

$$A_\theta^{11} = \left[\left(\sum_{i=1}^n (x_i - \theta)^2 \right) \left(\sum_{i=1}^n x_i^2 \right) - \left(\sum_{i=1}^n x_i (x_i - \theta)^+ \right)^2 \right] / D,$$

$$A_\theta^{12} = \left[- \left(\sum_{i=1}^n (x_i - \theta)^2 \right) \left(\sum_{i=1}^n x_i \right) \right.$$

$$\left. + \left(\sum_{i=1}^n x_i (x_i - \theta)^+ \right) \left(\sum_{i=1}^n (x_i - \theta)^+ \right) \right] / D,$$

$$A_\theta^{13} = \left[\left(\sum_{i=1}^n x_i (x_i - \theta)^+ \right) \left(\sum_{i=1}^n x_i \right) - \left(\sum_{i=1}^n x_i^2 \right) \left(\sum_{i=1}^n (x_i - \theta)^+ \right) \right] / D,$$

$$A_\theta^{22} = \left[n \sum_{i=1}^n (x_i - \theta)^2 - \left(\sum_{i=1}^n (x_i - \theta)^+ \right)^2 \right] / D,$$

$$A_\theta^{23} = \left[-n \sum_{i=1}^n x_i (x_i - \theta)^+ + \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n (x_i - \theta)^+ \right) \right] / D,$$

$$A_\theta^{33} = \left[n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right] / D,$$

where

$$D = n \left(\sum_{i=1}^n (x_i - \theta)^2 \right) \left(\sum_{i=1}^n x_i^2 \right) - n \left(\sum_{i=1}^n x_i (x_i - \theta)^+ \right)^2$$

$$- \left(\sum_{i=1}^{n} (x_i - \theta)^2 \right) \left(\sum_{i=1}^{n} x_i \right)^2$$

$$+ 2 \left(\sum_{i=1}^{n} x_i (x_i - \theta)^+ \right) \left(\sum_{i=1}^{n} (x_i - \theta)^+ \right) \left(\sum_{i=1}^{n} x_i \right)$$

$$- \left(\sum_{i=1}^{n} (x_i - \theta)^+ \right)^2 \left(\sum_{i=1}^{n} x_i^2 \right).$$

We have assumed that $\{\theta : R_\theta < \Delta\}$ is connected.

All of the terms required for (3) and (4) can be found from linear least squares programs. The minimum and maximum can be found by scanning over values of θ at suitably small intervals. In Section 3, for example, the minimum and maximum are found by taking θ from -9 to 9 at intervals of 0.01. Linear least square programs are widely available in every statistical package, even in many hand calculators. Hence, the method presented by Theorem 2.2 is simple and could have wide spread applicability.

In principle, the confidence sets, $\cup_\theta I_\theta$, in Theorems 2.1 and 2.2 could be pieces of sets. However, if $\mathbf{f}(\cdot, \cdot)$, $g(\cdot, \cdot)$, $\mathbf{X}_\theta$, and $\boldsymbol{\lambda}_\theta$ are suitably regular functions - as they will be in most applications - then $\cup_\theta I_\theta$ will be a whole interval. In particular, if $\boldsymbol{\lambda}_\theta$ and $\mathbf{X}_\theta$ are continuous functions of $\boldsymbol{\theta}$ and $\{\boldsymbol{\theta} : R_{\boldsymbol{\theta}} < \Delta\}$ is connected then $\cup_\theta I_\theta$ will be a whole interval as shown by Theorem 2.2. The conditions $\boldsymbol{\lambda}_\theta$ and $\mathbf{X}_\theta$ are continuous functions of $\boldsymbol{\theta}$ can be easily verified analytically. For example, these conditions are satisfied in the cases of Examples 1.2–1.3 and Examples 2.2–2.3. The condition $\{\boldsymbol{\theta} : R_{\boldsymbol{\theta}} < \Delta\}$ is connected can be tested either analytically or at least numerically for cases like Examples 1.2–1.3 and Examples 2.2–2.3.

3. Numerical studies

Here, we pursue the model discussed by Examples 1.2–1.3 and 2.2–2.3, i.e. the continuous two phase regression model due to Hawkins [8] and Davies [6]:

$$y_i = a + bx_i + c (x_i - \theta)^+ + \epsilon_i \tag{6}$$

for $i = 1, 2, \ldots, n$, where ϵ_i are assumed to be independent standard normal. Take $\boldsymbol{\beta} = (a, b, c)'$, an unknown parameter vector, and let $\boldsymbol{\lambda}_\theta = (1, x, (x - \theta))'$ for some given x. We wish to find a confidence interval for the predicted value

$$y = g(\boldsymbol{\lambda}_\theta, \boldsymbol{\beta}) = \boldsymbol{\lambda}_\theta' \boldsymbol{\beta} = a + bx + c(x - \theta)^+. \tag{7}$$

The results in Example 2.3 can be used to find the confidence interval.

Computer simulation shows that actual confidence levels tend to be slightly lower than nominal values. There is no evidence to suggest that the confidence
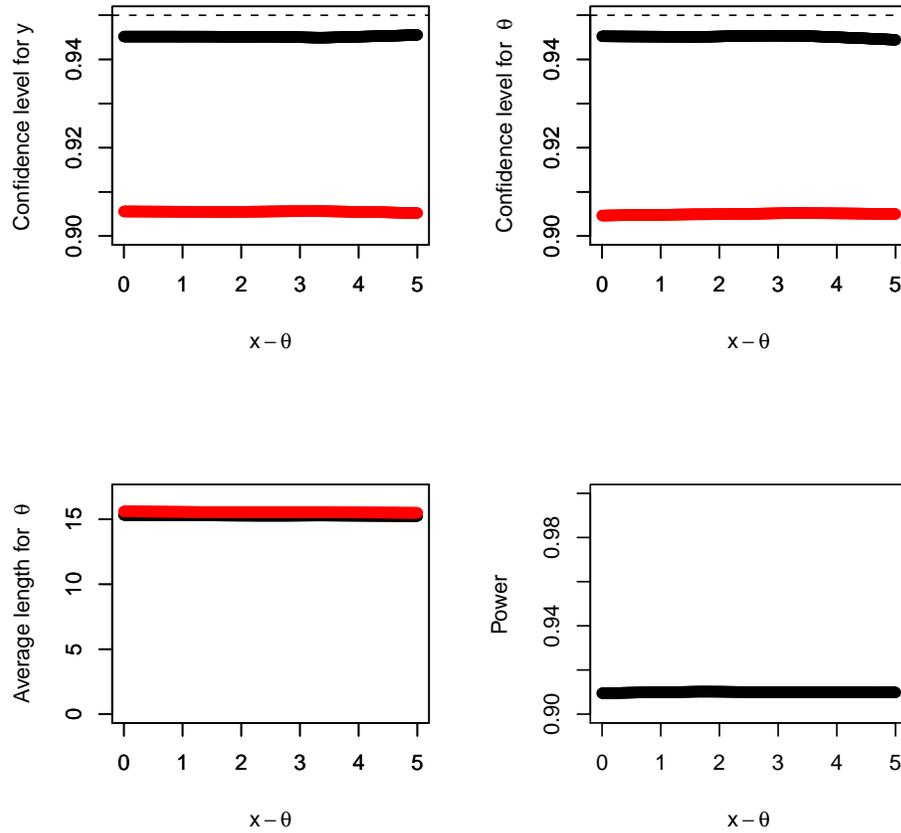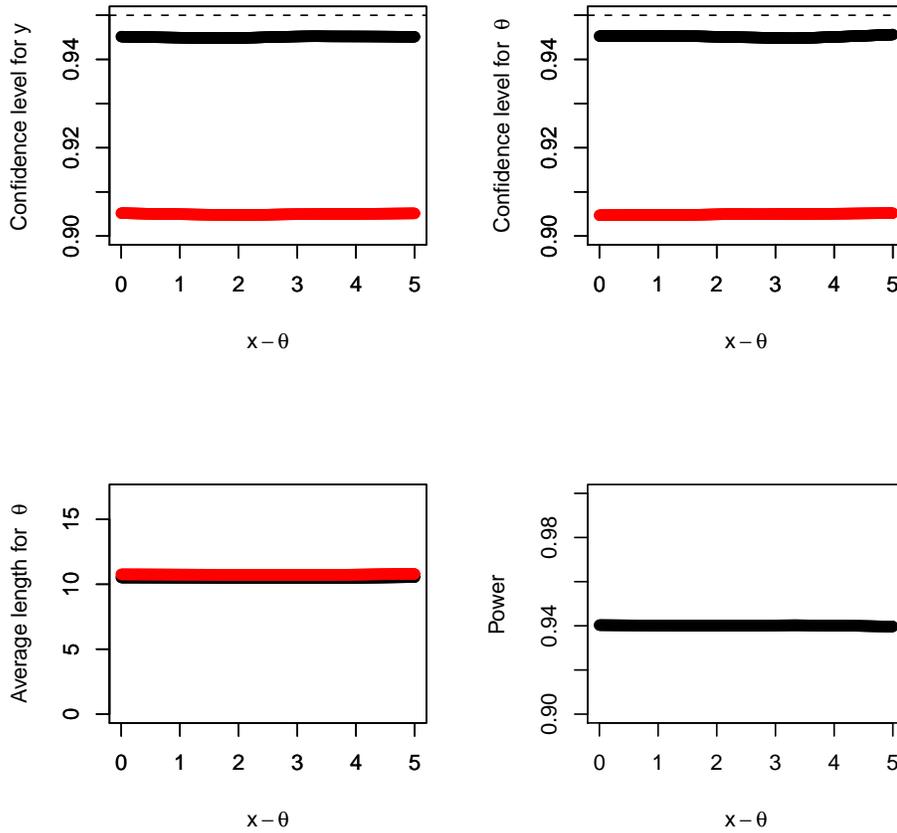
FIG 1. *Confidence level for y (top left), confidence level for θ (top right), average confidence length for θ (bottom left) and power of Davies (1987)'s test for $c \neq 0$ (bottom right) for $c = 0.2$ and $\theta = 0$. The proposed method and the delta method are in black and red, respectively.*

intervals are better for smaller/larger c or for smaller/larger $\mid x - \theta \mid$. Figures 1–5 show typical results. The four plots in each of the figures correspond to:

- the top left gives the fraction of trials that the calculated confidence interval covered y in (7).
- the top right gives the fraction of successes for the confidence interval found as those values of θ for which $R_\theta > \Delta$.
- the bottom left gives the average length of the confidence interval found as those values of θ for which $R_\theta > \Delta$.
- the bottom right gives the power of the 5% test for testing the hypothesis c is non-zero as given by Davies [6] with the variance of the series being estimated.
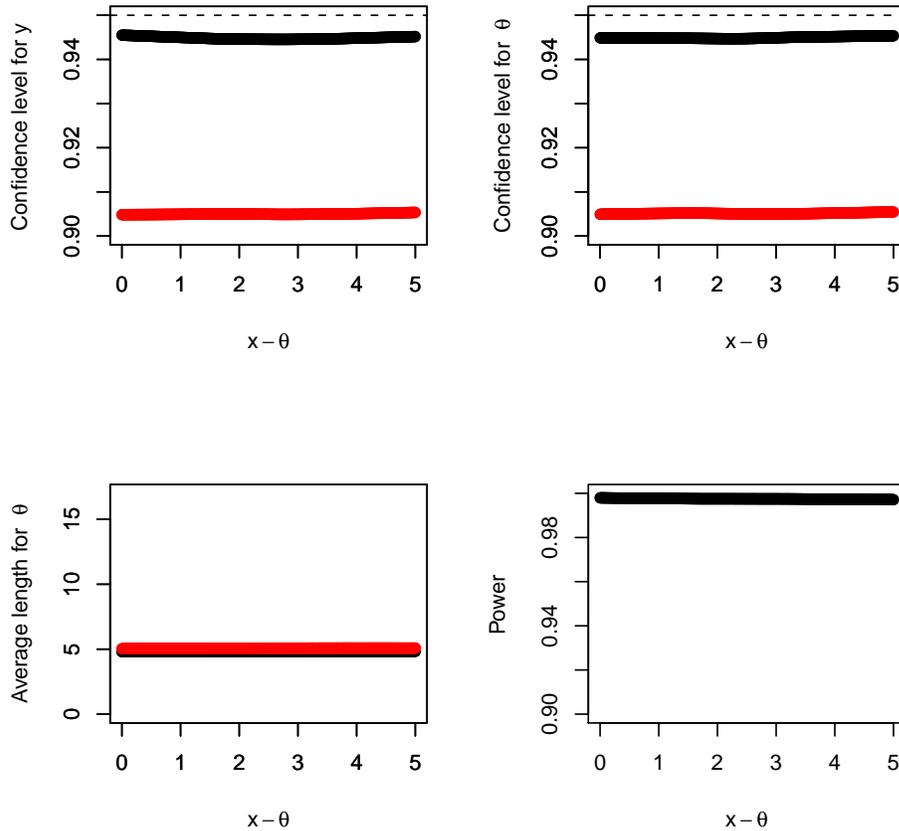
FIG 2. *Confidence level for y (top left), confidence level for θ (top right), average confidence length for θ (bottom left) and power of Davies [6]'s test for $c \neq 0$ (bottom right) for $c = 0.5$ and $\theta = 0$. The proposed method and the delta method are in black and red, respectively.*

Also shown in Figures 1–5 are the confidence levels for y, θ and the average length of the confidence interval for θ computed using the delta method.

In the simulations $n = 21$, $x_i = (-10, -9, \ldots, 10)$, the residual standard deviation equals 1.0, the formulas were calculated for values of θ from -9 to 9 at intervals of 0.01. In each case, there were 10,000 simulations and the nominal confidence level was 0.95. The choice $n = 21$ corresponds to a small sample size. Other small sample sizes yielded similar results.

Figures 1–5 suggest that the confidence interval for y is reasonably accurate whether the confidence interval for θ is short or not, whether c is well distinguished from zero or not, and whether x and θ are well separated or not.

The bottom right plots in Figures 1–5 show - as expected - that the power increases to 1 as c becomes more distinguished from zero. The bottom left plots show that the average length becomes shorter.

FIG 3. *Confidence level for y (top left), confidence level for θ (top right), average confidence length for θ (bottom left) and power of Davies [6]'s test for c ≠ 0 (bottom right) for c = 1 and θ = 0. The proposed method and the delta method are in black and red, respectively.*

The top left, the top right and the bottom left plots in Figures 1–5 suggest that the proposed method has higher confidence levels and shorter lengths than the delta method for all values of c, θ and x considered. The gain in average length does not appear substantial. The gain in confidence level appears substantial. The average confidence level for the proposed method is approximately 0.945. That for the delta method is approximately 0.905.

We now perform a simulation study to verify that the confidence level of the interval given by Theorem 2.2 converges to the nominal value with increasing sample size. We computed the confidence levels for y as in Figures 1–5 for $x_i = (-(n-1)/2, 1-(n-1)/2, \ldots, (n-1)/2-1, (n-1)/2)$, $n = 15, 17, \ldots$, and $(c, \theta, x) = (0, 0, 0)$, $(0.2, 0, 0)$, $(0.5, 0, 0)$, $(0.5, 0, 2)$, $(0.5, 0, 20)$, $(1, 0, 0)$, $(2, 0, 0)$, $(5, 0, 0)$. The plots of the confidence levels versus n are shown in Figure 6.
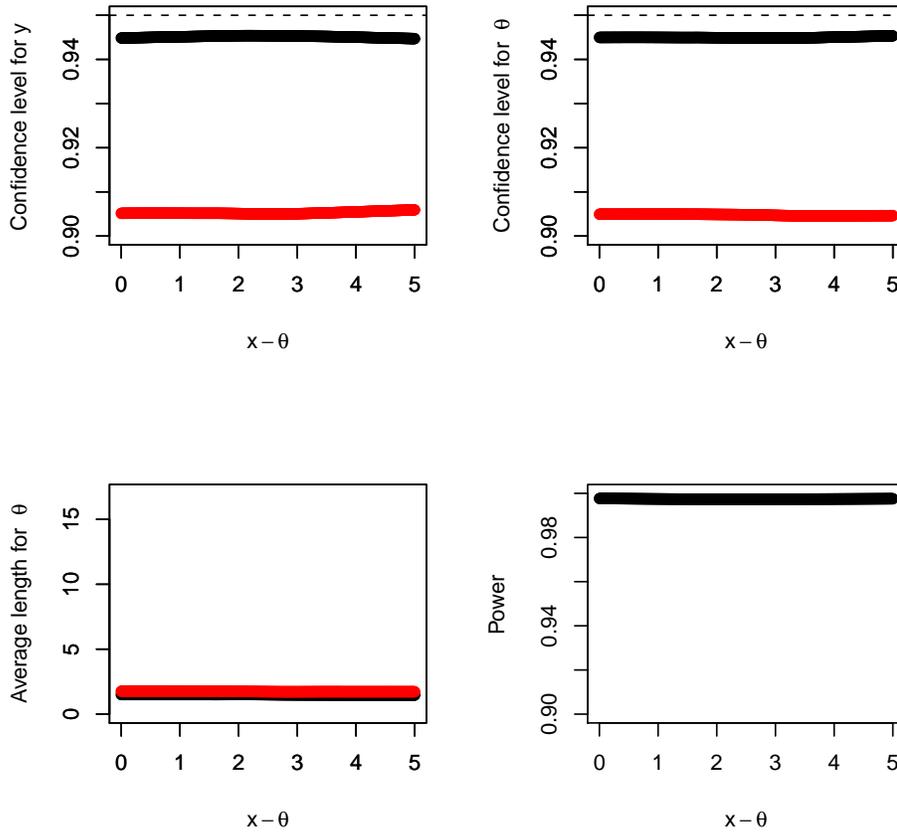
FIG 4. *Confidence level for y (top left), confidence level for θ (top right), average confidence length for θ (bottom left) and power of Davies [6]'s test for c ≠ 0 (bottom right) for c = 2 and θ = 0. The proposed method and the delta method are in black and red, respectively.*

It is evident from Figure 6 that the confidence levels for y approach the nominal level with increasing sample size for each of the selected values of (c, θ, x). For practical use, one could consider samples of size 40 or higher sufficient.

Next, we wish to verify that calculations using (2) and Theorem 2.2 lead to the same results. For this, we computed the confidence levels of y using (2) for the same ranges of parameters considered by Figure 6. These confidence levels were plotted on the same graphs. As we can see, the values computed using the two formulas appear indistinguishable. This supports correctness of the mathematical derivations for Theorem 2.2.

The above results assume that the residuals in (6) are normal. To show robustness of our methods in Section 2, these results were recalculated for a variety of non-normal and asymmetric errors, including Student's t with 1 degree of freedom, Student's t with 10 degrees of freedom, Student's t with 20 degrees
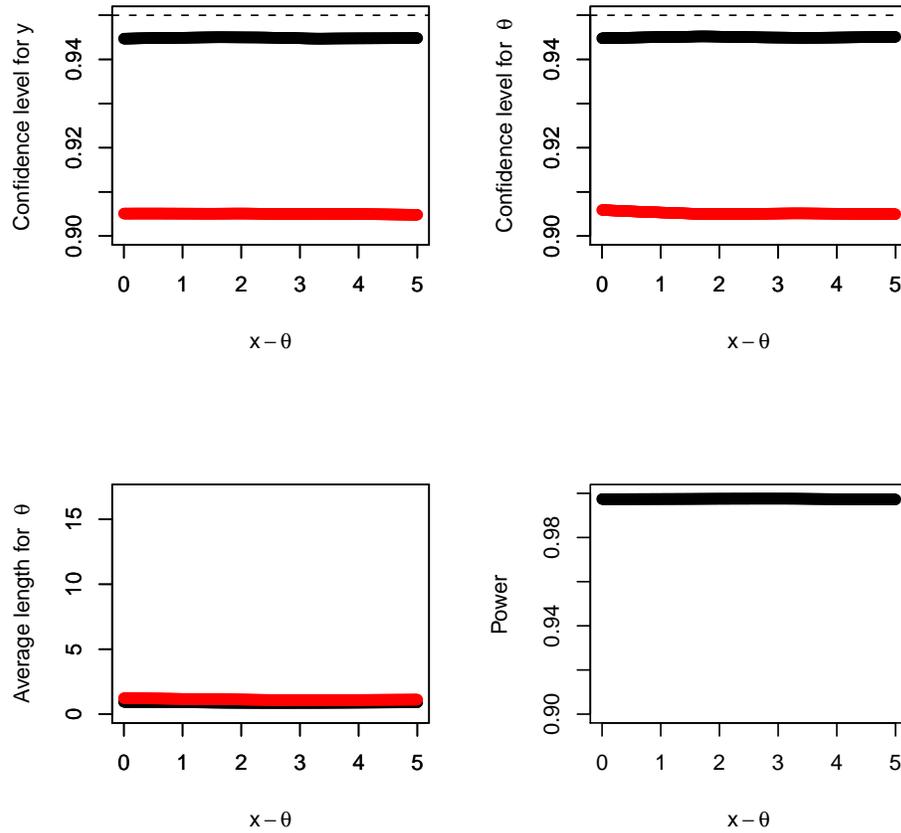
FIG 5. *Confidence level for y (top left), confidence level for θ (top right), average confidence length for θ (bottom left) and power of Davies [6]'s test for c ≠ 0 (bottom right) for c = 5 and θ = 0. The proposed method and the delta method are in black and red, respectively.*

of freedom, Laplace, logistic, uniform, Gumbel and skew normal. The corresponding Figures 1–6 were similar, suggesting robustness of our methods. These figures are not reported here because of their similarity and to avoid repetitive discussion.
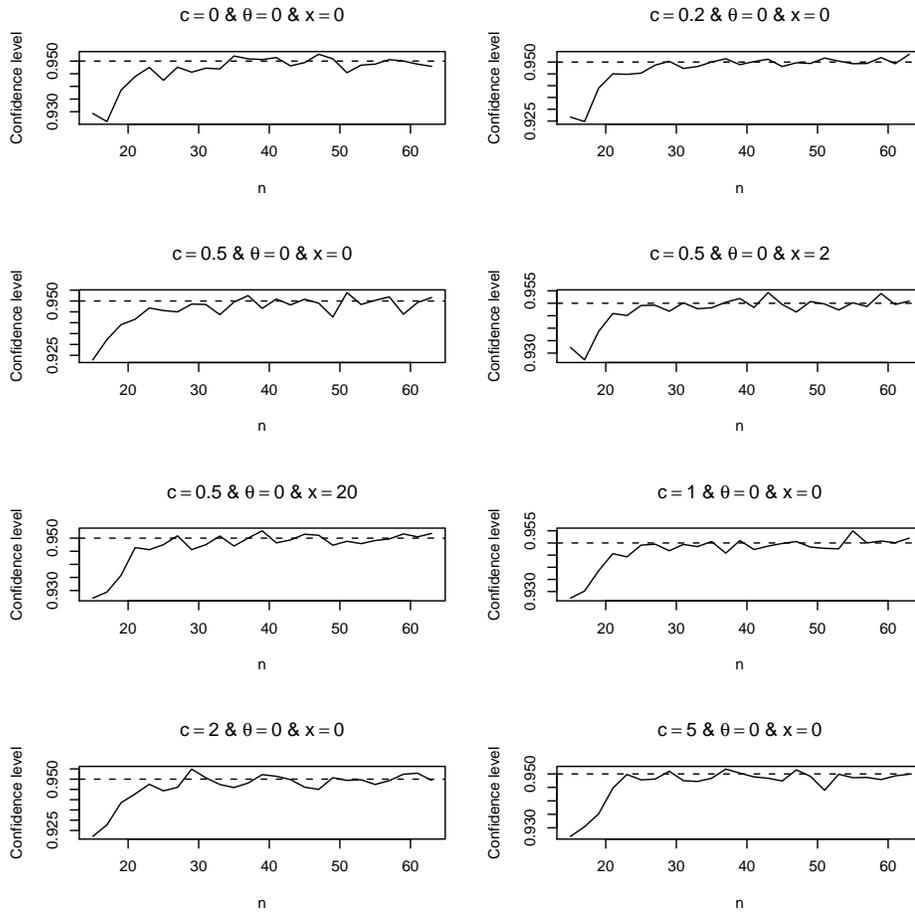
Acknowledgments

Fig 6. *Confidence level for y versus n for* $(c, \theta, x) = (0, 0, 0)$, $(0.2, 0, 0)$, $(0.5, 0, 0)$, $(0.5, 0, 2)$, $(0.5, 0, 20)$, $(1, 0, 0)$, $(2, 0, 0)$, *and* $(5, 0, 0)$.

References

[1] BATES, D. M. and WATTS, D. G. (1988). *Nonlinear Regression Analysis and Its Applications.* John Wiley and Sons, New York. MR1060528

[2] BEEK, P. J., SCHMIDT, R. C., MORRIS, A. W., SIM, M. Y. and TURVEY, M. T. (1995). Linear and nonlinear stiffness and friction in biological rhythmic movements. *Biological Cybernetics* **73** 499-507.

[3] CARROLL, R. J. and RUPPERT, D. (1988). *Transformation and Weighting in Regression.* Chapman and Hall, New York. MR1014890

[4] CARROLL, R. J., RUPPERT, D., STEFANSKI, L. A. and CRAINICEANU, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, second edition. Chapman and Hall/CRC, Boca Raton, Florida. MR2243417

[5] CASELLA, G. and BERGER, R. L. (2002). *Statistical Inference*, second edition. Duxbury Press, Belmont, California. MR2059318

[6] DAVIES, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **74** 33-43. MR0885917

[7] DONALDSON, J. R. and SCHNABEL, R. B. (1987). Computational experience with confidence regions and confidence intervals for nonlinear least squares. *Technometrics* **29** 67-82. MR0876886

[8] HAWKINS, D. M. (1980). A note on continuous and discontinuous segmented regressions. *Technometrics* **22** 443-444.

[9] HONORE, B. E. (1992). Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects. *Econometrica* **60** 533-565. MR1162998

[10] HU, F. and KALBFLEISCH, J. D. (2000). The estimating function bootstrap. *Canadian Journal of Statistics* **28** 449-499. MR1793106

[11] HUET, S., BOUVIER, A., POURSAT, M. -A. and JOLIVET, E. (2004). *Statistical Tools for Nonlinear Regression: A Practical Guide with S-PLUS and R Examples*, second edition. Springer Verlag, New York. MR2024808

[12] KANIOVSKAYA, I. YU. (1980). On the asymptotic distribution of the Robbinsmhy Monro procedure with a nonsmooth regression function. *Theory of Probability and Mathematical Statistics* **23** 69-74. MR0614578

[13] LINDSEY, J. K. (2001). *Nonlinear Models in Medical Statistics*. Oxford University Press, Oxford. MR1866908

[14] MANSKI, C. F. (1988). *Analog Estimation Methods in Econometrics*. Chapman and Hall, New York. MR0996421

[15] SEBER, G. A. F. and WILD, C. J. (1989). *Nonlinear Regression*. John Wiley and Sons, New York. MR0986070

[16] SHACHAM, M. and BRAUNER, N. (2000). A general framework for considering data precision in optimal regression models. In: *Proceedings of the Fifth International Conference on Foundations of Computer-Aided Process Design*, editors M. F. Malone and J. A. Trainham, volume 96, pp. 470-473.

[17] SIEWICKI, T. C., PULLARO, T., PAN, W., McDANIEL, S., GLENN, R. and STEWART, J. (2007). Models of total and presumed wildlife sources of fecal coliform bacteria in coastal ponds. *Journal of Environmental Management* **82** 120-132.

[18] TIAN, L., LIU, J. and WEI, L. J. (2007). Implementation of estimating function based inference procedure with MCMC samplers. *Journal of the American Statistical Association* **102** 897-900. MR2411656

[19] TRENTIN, E., MAGNONI, L. and ANDRONICO, A. (2003). Toward a modular connectionist model of local chlorophyll concentration from satellite images. In: *Proceedings of the International Joint Conference on Neural Networks*, volumes 1-4, pp. 2317-2321.