| Journal ...Journal of Statistical Computation and Simulation | Article ID ... GSCS 314236 |
|---|---|

TO: CORRESPONDING AUTHOR

AUTHOR QUERIES - TO BE ANSWERED BY THE AUTHOR

Dear Author
Please address all the numbered queries on this page which are clearly identified on the proof for your convenience.

Thank you for your cooperation

| | | |
|---|---|---|
| Q1 | Please check the affiliations given for the authors ‘E. Castillo, A.S. Hadi, and R. Mìnguez’ is correct. | |
| Q2 | Please provide expansion for ‘NIST’. | |
| Q3 | Please provide volume number for ref. [35]. | |
| Q4 | Please cite the following references [45]–[48] in text. | |

Production Editorial Department, Taylor & Francis
4 Park Square, Milton Park, Abingdon OX14 4RN

Telephone: +44 (0) 1235 828600
Facsimile: +44 (0) 1235 829000

*Journal of Statistical Computation and Simulation*
Vol. 00, No. 0, Month 2008, 1–20

Taylor & Francis
Taylor & Francis Group

# Diagnostics for non-linear regression

E. Castillo[a,b], A.S. Hadi[c,d]* and R. Mínguez[e]

[a] *Department of Applied Mathematics and Computational Sciences, University of Cantabria, Cantabria, Spain;* [b] *Department of Applied Mathematics and Computational Sciences, University of Castilla-La Mancha, Ciudad Real, Spain;* [c] *Department of Mathematics, The American University in Cairo, Cairo, Egypt;* [d] *Department of Statistical Science, Cornell University, Ithaca, New York, USA;* [e] *Department of Applied Mathematics, University of Castilla-La Mancha, Ciudad Real, Spain* **Q1**

(*Received 09 November 2007; final version received 18 April 2008*)

Sensitivity analysis in regression is concerned with assessing the sensitivity of the results of a regression model (*e.g.*, the objective function, the regression parameters, and the fitted values) to changes in the data. Sensitivity analysis in least squares linear regression has seen a great surge of research activities over the last three decades. By contrast, sensitivity analysis in non-linear regression has received very little attention. This paper deals with the problem of local sensitivity analysis in non-linear regression. Closed-form general formulas are provided for the sensitivities of three standard methods for the estimation of the parameters of a non-linear regression model based on a set of data. These methods are the least squares, the minimax, and the least absolute value methods. The effectiveness of the proposed measures is illustrated by application to several non-linear models including the ultrasonic data and the onion yield data. The proposed sensitivity measures are shown to deal effectively with the detection of influential observations in non-linear regression models.

**Keywords:** dual optimization problem; influential observations; lagrangian function; least square; least absolute value; minimax method; outlier detection; primal optimization problem

## 1. Introduction

It has long been recognized that statistical conclusions drawn from an analysis can be sensitive to changes in a model, deviations from assumptions, and other perturbations in the inputs of a statistical analysis (*e.g.*, observations in a data set). Actually, more often than not, conclusions drawn from an analysis can be completely turned around if one or few data points are changed.

It is therefore essential for data analysts to be able to assess the sensitivity of their conclusions to various perturbations in the inputs. Sensitivity analysis provides confidence in the model and the conclusions from an analysis because it allows the analyst to assess the effects of departures from the assumptions, detect outliers or wrong data values, define testing strategies, optimize resources, reduce costs, and avoid unexpected problems.

Sensitivity analysis has seen a great surge of research activities over the last three decades. Most of the work, however, focused almost exclusively on least squares (LS) linear regression.

*Corresponding author. Email: ahadi@aucegypt.edu

ISSN 0094-9655 print/ISSN 1563-5163 online

DOI: 10.1080/00949650802140703
http://www.informaworld.com

An evidence supporting this observation is the number of books and articles that have appeared in the literature. See, for example, the books [1–7], and the articles [8–23]. By contrast to the area of LS linear regression, sensitivity analysis in non-linear regression has received very little attention. This paper is an attempt to fill this void.

The intrinsically (non-linearizable) non-linear regression model can be written as

$$y_i = f(\boldsymbol{x}_i; \boldsymbol{\beta}) + \varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{1}$$

where $y_i$ is the $i$th value of the response variable, $\boldsymbol{x}_i$ is a $k \times 1$ vector of predictor variables corresponding to the $i$th observation, and $\boldsymbol{\varepsilon}_i$ is a random error. The function $f$ is known and non-linear in the parameter vector $\boldsymbol{\beta}$.

The parameters are to be estimated from the data. We consider here three estimation methods: the LS, the minimax (Min–Max), and the least absolute value (LAV) methods. All three methods can be formulated as optimization problems as follows.

The LS method is by far the most popular. The LS estimates of the parameters $\boldsymbol{\beta}$ are obtained by minimizing the sum of squared distances between observed and predicted values, that is,

$$\underset{\boldsymbol{\beta}}{\text{Minimize}}\ Z_{\text{LS}} = \sum_{i=1}^{n} (y_i - f(\boldsymbol{x}_i; \boldsymbol{\beta}))^2. \tag{2}$$

The method of LS was first named and published by Legendre in the paper 'Nouvelles méthodes pour la determination des orbites des cométes', which appeared in 1805, though the basic idea of the method of LS has occurred to Gauss in the autumn of 1794.

Two less common alternatives to the LS method are the Min–Max method and the LAV method.

The Min–Max regression estimators were studied by Euler, Lambert, and Laplace (see, for example, Sheynin [24] and Plackett [25]). Accordingly, the parameter estimates are the quantities that minimize the absolute value of the largest deviation [24]. More precisely, the Min–Max regression estimators are obtained by minimizing the maximum of the distances between observed and predicted values, *i.e.*:

$$\underset{\boldsymbol{\beta},\varepsilon}{\text{Minimize}}\ Z_{\text{Min–Max}} = \varepsilon \tag{3}$$

subject to

$$\begin{aligned} y_i - f(\boldsymbol{x}_i - \boldsymbol{\beta}) &\le \varepsilon, \quad i = 1, \ldots, n, \\ f(\boldsymbol{x}_i; \boldsymbol{\beta}) - y_i &\le \varepsilon, \quad i = 1, \ldots, n. \end{aligned} \tag{4}$$

An algorithm for finding the Min–Max residual was given in 1783 by Laplace [26] who, later in 1789, has simplified the earlier procedure in ref. [27]. The Min–Max estimates are the most non-robust estimators [28].

The LAV method minimizes the sum of the distances between observed and predicted values, *i.e.*:

$$\underset{\boldsymbol{\beta},\varepsilon_i}{\text{Minimize}}\ Z_{\text{LAV}} = \sum_{i=1}^{n} \varepsilon_i \tag{5}$$

subject to

$$\begin{aligned} y_i - f(\boldsymbol{x}_i; \boldsymbol{\beta}) &\le \varepsilon_i, \quad i = 1, \ldots, n, \\ f(\boldsymbol{x}_i; \boldsymbol{\beta}) - y_i &\le \varepsilon_i, \quad i = 1, \ldots, n. \end{aligned}$$

In the simplest linear case Boscovich was the first in proposing to estimate the parameters by quantities which satisfy the following conditions: (a) the sum of the deviations is zero; and (b) the sum of the absolute values of the deviations is a minimum, and gave a geometrical method of solution.

One question of interest is: how sensitive are the optimal values, $Z^*_{\mathrm{LS}}$, $Z^*_{\mathrm{LAV}}$ and $Z^*_{\mathrm{Min–Max}}$ and the estimated parameters $\boldsymbol{\beta}_{\mathrm{LS}}$, $\boldsymbol{\beta}_{\mathrm{Min–Max}}$, and $\boldsymbol{\beta}_{\mathrm{LAV}}$ to the data values ($x$ or $y$ values)?

One way of assessing the influence of the $i$th observation on the results of an analysis is to use the case *deletion* approach, where the influence of the $i$th observation on a given regression result is assessed by the difference between the value of the result based on all the data and the value obtained when the $i$th observation is omitted from the calculations. In LS linear regression, this approach is computationally efficient because there are closed formulas that can be obtained after fitting the model to the full data. In non-linear regression, however, this approach is computationally unrealistic because we need to solve $n + 1$ non-linear optimization problems (one for the full data and one when each of the $n$ observations in the data is omitted one at a time). For the LS model, there exist iterative formula that relate the results based on the full data with those based on the data with one observation omitted, facilitating the calculations and reducing the computational complexity.

For other models (*e.g.*, the least trimmed squares (LTS) and M-estimators), there are asymptotic representations for the difference the estimates obtained from the full data and those obtained when one or more observations are deleted, which facilitate the analysis [29–31]. In addition, by using a modern algorithm for LTS [32] one can find the most influential observations with computing times that can be reduced to 1/2 for samples sizes of $n = 50$, or to 1/100 for $n = 1000$, as compared to the times required without using these techniques.

Finally, it must be pointed out that different models can lead to very different estimates [31,33]. Thus, user needs must be taken into account when selecting the regression method for fitting the data.

Another approach to sensitivity analysis, proposed by Cook [10], is a weighted perturbation approach, where each observation is given a weight $\omega_i$, with $0 \leq \omega_i \leq 1$. The influence of an observation $x_i$ is then measured by the likelihood displacement:

$$\mathrm{LD}(\omega) = 2[L(\hat{\theta}) - L(\hat{\theta}_\omega)], \tag{6}$$

where $\omega = \{\omega_1, \ldots, \omega_n\}$, $\hat{\theta}$ is the maximum likelihood estimate of $\theta$, and $\hat{\theta}_\omega$ is the maximum likelihood estimate of $\theta$ when the $x_i$ is given weight $\omega_i$, and $L(\hat{\theta})$ is the log-likelihood function evaluated at $\hat{\theta}$. The deletion approach can be viewed as giving a weight of either 0 or 1 to each of the observations in the data. Because it is based on the likelihood function, the weighted perturbation approach normally applies to the LS normal regression, but it can also be applied to the Min–Max and LAV methods without problems using weighted residuals. The weights can be '*a priori*' selected and prescribed to the observations according to their position among the order statistics of absolute values of residuals [34]. It also has the same computational problems associated with the deletion approach in non-linear regression.

A third approach, proposed by Nyquist [16] and Hadi and Nyquist [35], is based on the sensitivity function. This, too, is computationally infeasible for non-linear regression models.

This paper uses a general, computationally feasible approach to sensitivity analysis in non-linear regression. Section 2 presents the materials that are necessary to derive the various sensitivity measures in non-linear regression. The general applicability of the proposed method for sensitivity analysis is then illustrated by its application to some real-life non-linear regression examples in Sections 3 and 4. Finally, Section 5 offers some concluding remarks.

## 2. Sensitivities in non-linear models

Many estimation methods in statistics can be expressed as non-linear programming problems, that is, to optimize an objective function subject to some constraints. This includes, for example, the LS, Min–Max, and LAV methods, as we have seen in Equations (2)–(5). Once the statistical problem is expressed as a non-linear programming problem, we show how the sensitivities of the resultant estimators can be easily obtained using some results from mathematical programming. This section gives these needed results.

Consider the following general non-linear programming problem $(P)$:

$$\underset{\boldsymbol{\theta}}{\text{Minimize}}\ Z_{\mathrm{P}} = f(\boldsymbol{\theta}; \boldsymbol{x}) \tag{7}$$

subject to

$$\boldsymbol{h}(\boldsymbol{\theta}; \boldsymbol{x}) = \mathbf{0} : \boldsymbol{\lambda}, \tag{8}$$

$$\boldsymbol{g}(\boldsymbol{\theta}; \boldsymbol{x}) \leq \mathbf{0} : \boldsymbol{\mu}, \tag{9}$$

where letters in boldface refer to vectors, $\boldsymbol{\theta} \in \mathbb{R}^n$, $\boldsymbol{x} \in \mathbb{R}^p$, $\boldsymbol{h}(\theta; \boldsymbol{x}) \in \mathbb{R}^\ell$, $\boldsymbol{g}(\boldsymbol{\theta}; \boldsymbol{x}) \in \mathbb{R}^m$, and $\lambda \in \mathbb{R}^\ell$ and $\mu \in \mathbb{R}^m$ are the *dual* variables associated with the equality and inequality constraints, respectively. The problem in Equations (7)–(9) is called the *primal* problem.

Every primal non-linear programming problem $P$ has an associated dual problem $D$, which is defined as:

$$\underset{\boldsymbol{\lambda},\boldsymbol{\mu}}{\text{Maximize}}\ Z_{\mathrm{D}} = \mathrm{Inf}_{\boldsymbol{\theta}}\{\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\mu}; \boldsymbol{x})\} \tag{10}$$

subject to

$$\boldsymbol{\mu} \geq \mathbf{0}, \tag{11}$$

where

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\mu}; \mathbf{x}) = f(\boldsymbol{\theta}; \mathbf{x}) + \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{h}(\boldsymbol{\theta}; \boldsymbol{x}) + \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{\theta}; \boldsymbol{x}), \tag{12}$$

is the *Lagrangian function* associated with the primal problem (7)–(9), and the dual variables $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ are vectors of dimensions $\ell$ and $m$, the number of equality and inequality constraints in Equations (8) and (9), respectively.

Given some regularity conditions [36–39], if the primal problem (7)–(9) has a locally optimal solution $\boldsymbol{\theta}^*$, the dual problem (10)–(11) also has a locally optimal solution $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$, and the optimal values of the objective functions of both problems coincide.

When dealing with the optimization problem (7)–(9), the following questions regarding sensitivity analysis are of interest:

(1) What is the sensitivity of $Z_{\mathrm{P}}^* = f(\boldsymbol{\theta}^*; \boldsymbol{x})$ to changes in $\boldsymbol{x}$? That is, the sensitivity of the objective function at the optimal point when the data $\boldsymbol{x}$ are modified. For our examples, this means the sensitivities of the sum of squared errors, the Min–Max value or the sum of absolute errors, to data.
(2) What is the sensitivity of $\boldsymbol{\theta}^*$, $\boldsymbol{\lambda}^*$, and $\boldsymbol{\mu}^*$ to changes in $\boldsymbol{x}$? That is, the sensitivity of the primal, $\boldsymbol{\theta}^*$, and dual variables, $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}^*$, at their optimal values when the data are modified. In estimation problems this means the sensitivities of the resultant regression coefficient estimates, or the sensitivities of the equality and inequality constraints to data, respectively.

All these sensitivities have a great practical importance because they can help in the identification of outliers and influential data points.

In this section, we give formulas that allow obtaining the local sensitivities of the objective function and the primal and dual variables with respect to data all at once. Without loss of generality, we consider that all the inequalities in Equation (9) are active. Note that after solving problem (7)–(9), it is very easy to check which of the inequalities in Equation (9) are active, and then we can ignore the inactive constraints because they do not alter the optimal solution.

Then, let $\boldsymbol{q}(\boldsymbol{\theta};\boldsymbol{x})$ be all equality constraints in Equation (8) and the active inequality constraints in Equation (9). Also, let $\boldsymbol{\eta}$ be the vector including the dual variables corresponding to $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$, that is, $\boldsymbol{\eta}=(\boldsymbol{\lambda},\boldsymbol{\mu})^{\mathrm{T}}$. In Equation [40], Castillo *et al.* show that the sensitivity of the optimal solution $(\boldsymbol{\theta}^*,\boldsymbol{\eta}^*,Z_{\mathrm{P}}^*)$ of Equations (7)–(9) to changes in the parameters can be determined by the system of equations:

$$\begin{bmatrix} \boldsymbol{F_\theta} & \boldsymbol{F_x} & 0 & -1 \\ \boldsymbol{F_{\theta\theta}} & \boldsymbol{F_{\theta x}} & \boldsymbol{Q_\theta^T} & 0 \\ \boldsymbol{Q_\theta} & \boldsymbol{Q_x} & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathrm{d}\boldsymbol{\theta} \\ \mathrm{d}\boldsymbol{x} \\ \mathrm{d}\boldsymbol{\eta} \\ \mathrm{d}Z_{\mathrm{P}} \end{bmatrix} = 0, \tag{13}$$

where $\mathrm{d}\boldsymbol{\theta}$, $\mathrm{d}\boldsymbol{x}$, $\mathrm{d}\boldsymbol{\eta}$, and $\mathrm{d}Z_{\mathrm{P}}$ are the differential perturbations, and all the matrices in Equation (13) are evaluated at the optimal solution and are defined below (with the corresponding dimensions in parenthesis):

$$\boldsymbol{F}_{\boldsymbol{\theta}_{(1\times n)}} = (\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^*,\boldsymbol{x}))^{\mathrm{T}}, \tag{14}$$

$$\boldsymbol{F}_{\boldsymbol{x}_{(1\times p)}} = (\nabla_{\boldsymbol{x}} f(\boldsymbol{\theta}^*,\boldsymbol{x}))^{\mathrm{T}}, \tag{15}$$

$$\boldsymbol{F}_{\boldsymbol{\theta\theta}_{(n\times n)}} = \nabla_{\boldsymbol{\theta\theta}} f(\boldsymbol{\theta}^*,\boldsymbol{x}) + \sum_{k=1}^{\ell+m} \lambda_k^* \nabla_{\boldsymbol{\theta\theta}} q_k(\boldsymbol{\theta}^*,\boldsymbol{x}), \tag{16}$$

$$F_{\boldsymbol{\theta x}_{(n\times p)}} = \nabla_{\boldsymbol{\theta x}} f(\boldsymbol{\theta}^*,\boldsymbol{x}) + \sum_{k=1}^{\ell+m} \lambda_k^* \nabla_{\boldsymbol{\theta x}} q_k(\boldsymbol{\theta}^*,\boldsymbol{x}), \tag{17}$$

$$\boldsymbol{Q}_{\boldsymbol{\theta}_{((\ell+m)\times n)}} = (\nabla_{\boldsymbol{\theta}} \boldsymbol{q}(\boldsymbol{\theta}^*,\boldsymbol{x}))^{\mathrm{T}}, \tag{18}$$

$$\boldsymbol{Q}_{\boldsymbol{x}_{((\ell+m)\times p)}} = (\nabla_{\boldsymbol{x}} \boldsymbol{q}(\boldsymbol{\theta}^*,\boldsymbol{x}))^{\mathrm{T}}. \tag{19}$$

Condition (13) can be written as:

$$\boldsymbol{U} \begin{bmatrix} \mathrm{d}\boldsymbol{\theta} \\ \mathrm{d}\boldsymbol{\eta} \\ \mathrm{d}Z_{\mathrm{P}} \end{bmatrix} = \boldsymbol{S}\mathrm{d}\boldsymbol{x}, \tag{20}$$

where

$$\boldsymbol{U} = \begin{bmatrix} \boldsymbol{F_\theta} & \boldsymbol{0} & -1 \\ \boldsymbol{F_{\theta\theta}} & \boldsymbol{Q_\theta^T} & 0 \\ \boldsymbol{Q_\theta} & \boldsymbol{0} & 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{S} = -\begin{bmatrix} \boldsymbol{F_x} \\ \boldsymbol{F_{\theta x}} \\ \boldsymbol{Q_x} \end{bmatrix}. \tag{21}$$

Thus, given a unit direction vector $\mathrm{d}\boldsymbol{x}$, we can solve Equation (20) to obtain the partial derivatives. If the solution exists and is unique, we obtain the corresponding directional derivatives. For a partial derivative to exist, the corresponding directional derivatives must exist and be equal in absolute value but not in sign. If the system (20) has no solution, the corresponding directional and partial derivatives do not exist.

In this paper, we consider only the regular case, that is, we assume that the matrix $\boldsymbol{U}$ is a square matrix, that is equivalent to assuming that all the $\boldsymbol{\mu}$ multipliers are non-null, and also that it is invertible. If $\boldsymbol{U}$ is invertible, we have

$$\begin{pmatrix} \dfrac{\partial\boldsymbol{\theta}}{\partial\boldsymbol{x}} \\ \dfrac{\partial\boldsymbol{\eta}}{\partial\boldsymbol{x}} \\ \dfrac{\partial Z_{\mathrm{P}}}{\partial\boldsymbol{x}} \end{pmatrix} = \boldsymbol{U}^{-1}\boldsymbol{S}. \tag{22}$$

where $\partial\boldsymbol{\theta}/\partial\boldsymbol{x}$, $\partial\boldsymbol{\eta}/\partial\boldsymbol{x}$, and $\partial Z_{\mathrm{P}}/\partial\boldsymbol{x}$ are the matrices containing all the sensitivities (partial derivatives) with respect to all data.

### 2.1. *The case without constraints*

In the particular case of an optimization problem with no constraints, the system (20) leads to

$$\begin{aligned}\boldsymbol{F}_{\boldsymbol{\theta}}\,\mathrm{d}\boldsymbol{\theta} - \mathrm{d}Z_{\mathrm{P}} &= -\boldsymbol{F}_{\boldsymbol{x}}\,\mathrm{d}\boldsymbol{x},\\ \boldsymbol{F}_{\boldsymbol{\theta\theta}}\,\mathrm{d}\boldsymbol{\theta} &= -\boldsymbol{F}_{\boldsymbol{\theta x}}\,\mathrm{d}\boldsymbol{x},\end{aligned}$$

which, if $\boldsymbol{F}_{\boldsymbol{\theta\theta}}$ is invertible, gives

$$\frac{\partial\boldsymbol{\theta}}{\partial\boldsymbol{x}} = -\boldsymbol{F}_{\boldsymbol{\theta\theta}}^{-1}\boldsymbol{F}_{\boldsymbol{\theta x}}, \tag{23}$$

$$\frac{\partial Z_{\mathrm{P}}}{\partial\boldsymbol{x}} = -\boldsymbol{F}_{\boldsymbol{\theta}}\boldsymbol{F}_{\boldsymbol{\theta\theta}}^{-1}\boldsymbol{F}_{\boldsymbol{\theta x}} + \boldsymbol{F}_{\boldsymbol{x}} = \boldsymbol{F}_{\boldsymbol{x}}, \tag{24}$$

where $\partial\boldsymbol{\theta}/\partial\boldsymbol{x}$ and $\partial Z_{\mathrm{P}}/\partial\boldsymbol{x}$ are matrices containing all the indicated partial derivatives.

### 2.2. *The case with constraints*

In the general case, if the matrix $\boldsymbol{Q}_{\boldsymbol{\theta}}$ is invertible, the solution of the system (20) becomes

$$\left.\frac{\partial\boldsymbol{\theta}}{\partial\boldsymbol{x}}\right._{(n\times p)} = -\boldsymbol{Q}_{\boldsymbol{\theta}}^{-1}\boldsymbol{Q}_{\boldsymbol{x}}, \tag{25}$$

$$\left.\frac{\partial\boldsymbol{\eta}}{\partial\boldsymbol{x}}\right._{(n\times p)} = (\boldsymbol{Q}_{\boldsymbol{\theta}}^{\mathrm{T}})^{-1}(\boldsymbol{F}_{\boldsymbol{\theta}}\boldsymbol{Q}_{\boldsymbol{\theta}}^{-1}\boldsymbol{Q}_{\boldsymbol{x}} - \boldsymbol{F}_{\boldsymbol{\theta x}}), \tag{26}$$

$$\left.\frac{\partial Z_{\mathrm{P}}}{\partial\boldsymbol{x}}\right._{(1\times p)} = -\boldsymbol{F}_{\boldsymbol{\theta}}\boldsymbol{Q}_{\boldsymbol{\theta}}^{-1}\boldsymbol{Q}_{\boldsymbol{x}} + \boldsymbol{F}_{\boldsymbol{x}} = \boldsymbol{F}_{\boldsymbol{x}} + \boldsymbol{\eta}^{\mathbf{T}}\boldsymbol{Q}_{\boldsymbol{x}}. \tag{27}$$

For the linear cases (as Min–Max and LAV) there are explicit formulas [41] in terms of the primal and dual optimal variable values.

### 2.3. *The singular case*

The singular case appears when the matrix $\boldsymbol{U}$ is not square or is singular. This occurs, for example, when the primal problem has infinitely many solutions or there are redundant constraints. As we shall see, the Min–Max and LAV regression models are prone to this singularity because of the existence of infinitely many solutions.

### 2.4. *Objective function sensitivities*

While Equation (22) gives the sensitivities of the parameters, the dual variables, and the objective functions to data values, the following theorem gives explicit and more simple formulas for the sensitivities of the objective function values with respect to data [42].

THEOREM 1 *The sensitivity of the objective function of the primal problem* (7)–(9) *with respect to* $\boldsymbol{x}$ *is given by*

$$\frac{\partial Z_{\mathrm{P}}^*}{\partial \boldsymbol{x}} = \nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{\theta}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*; \boldsymbol{x}), \tag{28}$$

*which is the partial derivative of its Lagrangian function*

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\mu}; \boldsymbol{x}) = f(\boldsymbol{\theta}; \boldsymbol{x}) + \boldsymbol{\lambda}^{\mathrm{T}} \boldsymbol{h}(\boldsymbol{\theta}; \boldsymbol{x}) + \boldsymbol{\mu}^{\mathrm{T}} \boldsymbol{g}(\boldsymbol{\theta}; \boldsymbol{x}), \tag{29}$$

*with respect to* $\boldsymbol{x}$ *evaluated at the optimal solution* $\boldsymbol{\theta}^*$, $\boldsymbol{\lambda}^*$, *and* $\boldsymbol{\mu}^*$.

The proof of this theorem is given in ref. [43], pp. 310–311.

### 2.5. *Standardized sensitivities*

Once the sensitivities in Equations (22) or (28), we need to compare them to see which data values are most influential on the obtained results. For these comparisons to be meaningful, we need to standardize the sensitivities. For example, instead of using $\partial Z_{\mathrm{P}}^*/\partial \boldsymbol{x}$ in Equation (28), we use the corresponding standardized versions

$$\frac{(\partial Z_{\mathrm{P}}^*/\partial x_{ij}) - m_j}{s_j}, \quad j = 1, \ldots, k, \tag{30}$$

where $m_j$ and $s_j$ are the mean and standard deviation of $\partial Z_{\mathrm{P}}^*/\partial x_{ij}$, $i = 1, 2, \ldots, n$, after replacing the parameters by their estimated values. When we refer to sensitivities, from now on, we mean the standardized sensitivities.

To illustrate all of the above sensitivity measures and to investigate their effectiveness in revealing influential observations in non-linear regression models, we use the following examples of real-life data:

(1) The ultrasonic data.
(2) The onion yield data.

## 3. A model for ultrasonic data

The data we use here are the result of an NIST study involving ultrasonic calibration. The **Q2** data consists of 54 observations on two variables. The response variable ($y$) is ultrasonic response and the predictor variable ($x$) is metal distance. The data in Table 1 were taken from www.itl.nist.gov/div898/strd/nls/data/LINKS/DATA/ Chwirut2.dat. In this Web site, a non-linear regression model of the form

$$y_i = f(\boldsymbol{x}_i; \boldsymbol{\beta}) + \varepsilon_i = \frac{e^{-\beta_1 x_i}}{\beta_2 + \beta_3 x_i} + \varepsilon_i, \quad i = 1, 2 \ldots, n, \tag{31}$$

is fitted to the data. here we assess the influence of observations on the results of the LS, Min–Max, and LAV methods.

Table 1. Ultrasonic calibration data.

| $i$ | $y_i$ | $x_i$ | $i$ | $y_i$ | $x_i$ | $i$ | $y_i$ | $x_i$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 92.9000 | 0.500 | 19 | 13.12 | 3.000 | 37 | 3.75 | 5.750 |
| 2 | 57.1000 | 1.000 | 20 | 59.90 | 0.750 | 38 | 11.81 | 3.000 |
| 3 | 31.0500 | 1.750 | 21 | 14.60 | 3.000 | 39 | 54.70 | 0.750 |
| 4 | 11.5875 | 3.750 | 22 | 32.90 | 1.500 | 40 | 23.70 | 2.500 |
| 5 | 8.0250 | 5.750 | 23 | 5.44 | 6.000 | 41 | 11.55 | 4.000 |
| 6 | 63.6000 | 0.875 | 24 | 12.56 | 3.000 | 42 | 61.30 | 0.750 |
| 7 | 21.4000 | 2.250 | 25 | 5.44 | 6.000 | 43 | 17.70 | 2.500 |
| 8 | 14.2500 | 3.250 | 26 | 32.00 | 1.500 | 44 | 8.74 | 4.000 |
| 9 | 8.4750 | 5.250 | 27 | 13.95 | 3.000 | 45 | 59.20 | 0.750 |
| 10 | 63.8000 | 0.750 | 28 | 75.80 | 0.500 | 46 | 16.30 | 2.500 |
| 11 | 26.8000 | 1.750 | 29 | 20.00 | 2.000 | 47 | 8.62 | 4.000 |
| 12 | 16.4625 | 2.750 | 30 | 10.42 | 4.000 | 48 | 81.00 | 0.500 |
| 13 | 7.1250 | 4.750 | 31 | 59.50 | 0.750 | 49 | 4.87 | 6.000 |
| 14 | 67.3000 | 0.625 | 32 | 21.67 | 2.000 | 50 | 14.62 | 3.000 |
| 15 | 41.0000 | 1.250 | 33 | 8.55 | 5.000 | 51 | 81.70 | 0.500 |
| 16 | 21.1500 | 2.250 | 34 | 62.00 | 0.750 | 52 | 17.17 | 2.750 |
| 17 | 8.1750 | 4.250 | 35 | 20.20 | 2.250 | 53 | 81.30 | 0.500 |
| 18 | 81.5000 | 0.500 | 36 | 7.76 | 3.750 | 54 | 28.90 | 1.750 |

### 3.1. *Sensitivity of the LS method*

The LS objective function in Equation (2) becomes:

$$Z_{\text{LS}} = \sum_{i=1}^{n} \left( y_i - \frac{e^{-\beta_1 x_i}}{\beta_2 + \beta_3 x_i} \right)^2.$$

Since there are no constraints, the sensitivities of the sum of squares to data are given by Theorem 1, that is:

$$\frac{\partial Z_{\text{LS}}}{\partial x_i} = 2\frac{\varepsilon_i(\beta_3 + \beta_1(\beta_2 + \beta_3 x_i))}{e^{\beta_1 x_i}(\beta_2 + \beta_3 x_i)^2} \quad \text{and} \quad \frac{\partial Z_{\text{LS}}}{\partial y_i} = 2\varepsilon_i.$$

The matrices we need for calculating all the sensitivities at once according to the proposed method are given in Appendix 1 and the following LS estimates for the ultrasonic data are obtained:

$$Z_{\text{LS}} = 513.05, \quad \hat{\beta}_1 = 0.16658, \quad \hat{\beta}_2 = 0.00516, \quad \hat{\beta}_3 = 0.01215.$$

We now perform a sensitivity analysis on these data. Figure 1 shows the LS fitted model, and the $Z_{\text{LS}}$, $\beta_1$, $\beta_2$, and $\beta_3$ sensitivities. From Figure 1, we see that five observations (1, 18, 28, 48, 51, and 53) exert undue influence on the LS results. Note that these are the five observations with the smallest value of $x(x = 0.5)$. Note also that observation 1 is influential only on the optimal objective function value but not on any of the estimated regression coefficients.

#### 3.1.1. *A case of a planted outlier*

To study further the effectiveness of the proposed sensitivities in the detection of outliers, we planted two outliers in this data set: Observation 55 (with $y = 11$, $x = 1.75$) and observation 56 (with $y = 25$, $x = 5.8$). These two points are shown in the scatter plot of the contaminated data in Figure 2a.

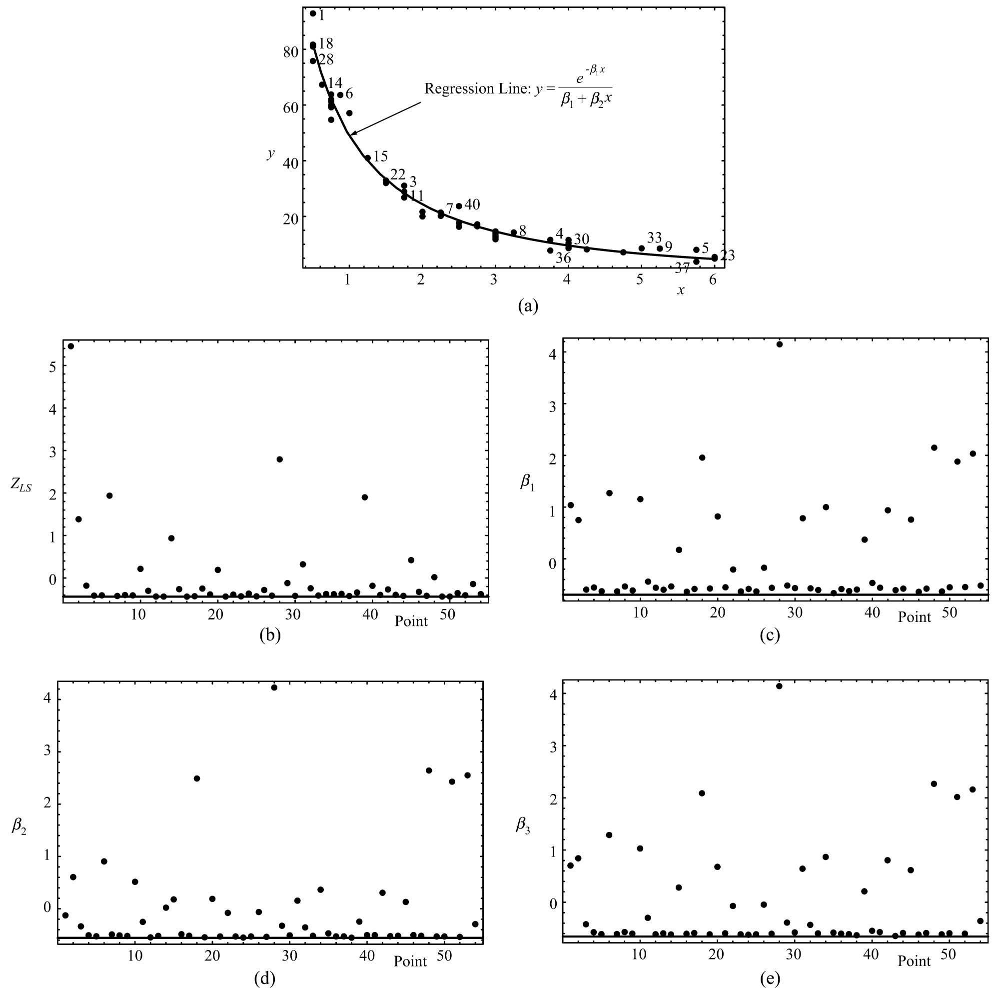


Figure 1. From left to right and top to bottom: (a) scatter plot of $y_i$ versus $x_i$, for the ultrasonic data set in Table 1 and the LS fitted model, (b)–(e) the index plot of the $Z^*_{\mathrm{LS}}$, $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ sensitivities, respectively.

The following LS estimates for the contaminated ultrasonic are obtained:

$$Z_{\mathrm{LS}} = 1180.9, \quad \hat{\beta}_1 = 0.11006, \quad \hat{\beta}_2 = 0.00408, \quad \hat{\beta}_3 = 0.01477.$$

A comparison with the previous results, shows that the addition of the two outliers has caused the LS results to change substantially (the optimal value of objective function, and the three parameter estimates have changed by 130%, 34%, 21%, and 22%, respectively).

### 3.2. *Sensitivity of the Min–Max method*

The objective function in Equation (3) is $Z_{\mathrm{Min\text{-}Max}} = \varepsilon$ and the constraints are

$$q_i^{(1)}(\beta_1, \beta_2, \beta_3, \epsilon; x_i, y_i) : y_i - \frac{e^{-\beta_1 x_i}}{\beta_2 + \beta_3 x_i} - \varepsilon \le 0; \quad i = 1, 2, \ldots, n,$$

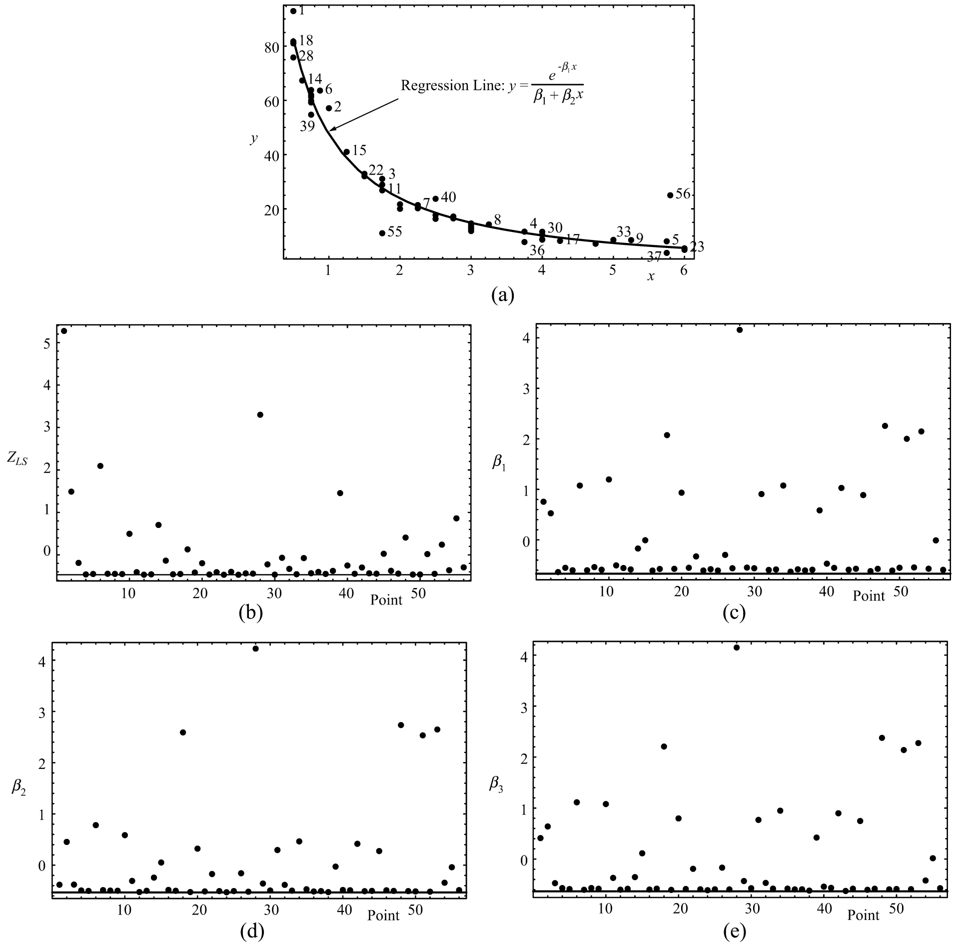


Figure 2. From left to right and top to bottom: (a) scatter plot of $y_i$ versus $x_i$, for the contaminated ultrasonic data set in Table 1 and the LS fitted model, (b)–(e) the index plot of the $Z^*_{\text{LS}}$, $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ sensitivities, respectively.

$$q_i^{(2)}(\beta_1, \beta_2, \beta_3, \varepsilon; x_i, y_i) : -y_i + \frac{e^{-\beta_1 x_i}}{\beta_2 + \beta_3 x_i} - \varepsilon \leq 0; \quad i = 1, 2, \ldots, n,$$

The matrices we need for calculating the sensitivities according to the proposed method are given in Appendix 2 and the following Min–Max estimates for the ultrasonic are obtained:

$$Z_{\text{Min-Max}} = 8.55, \quad \hat{\beta}_1 = -0.01288, \quad \hat{\beta}_2 = 0.00387, \quad \hat{\beta}_3 = 0.0161.$$

Unfortunately, this is not a regular case and the Min–Max regression problem have infinite solutions. This implies that no derivatives exist.

To avoid this non-uniqueness problem, we use the contaminated data and restrict the $\beta_1$ value to be non-negative and obtain the following estimates:

$$Z_{\text{Min-Max}} = 16.20, \quad \hat{\beta}_1 = 0, \quad \hat{\beta}_2 = 0.00355, \quad \hat{\beta}_3 = 0.01898.$$

Figure 3 shows the Min–Max model for the contaminated Ultrasonic data, and the $Z_{\text{Min-Max}}$, $\beta_1$, $\beta_2$, and $\beta_3$ sensitivities. Note that data points 1, 55, and 56 are the most influential data points. Thus, the method allows identifying the two planted outliers, together with the already existing

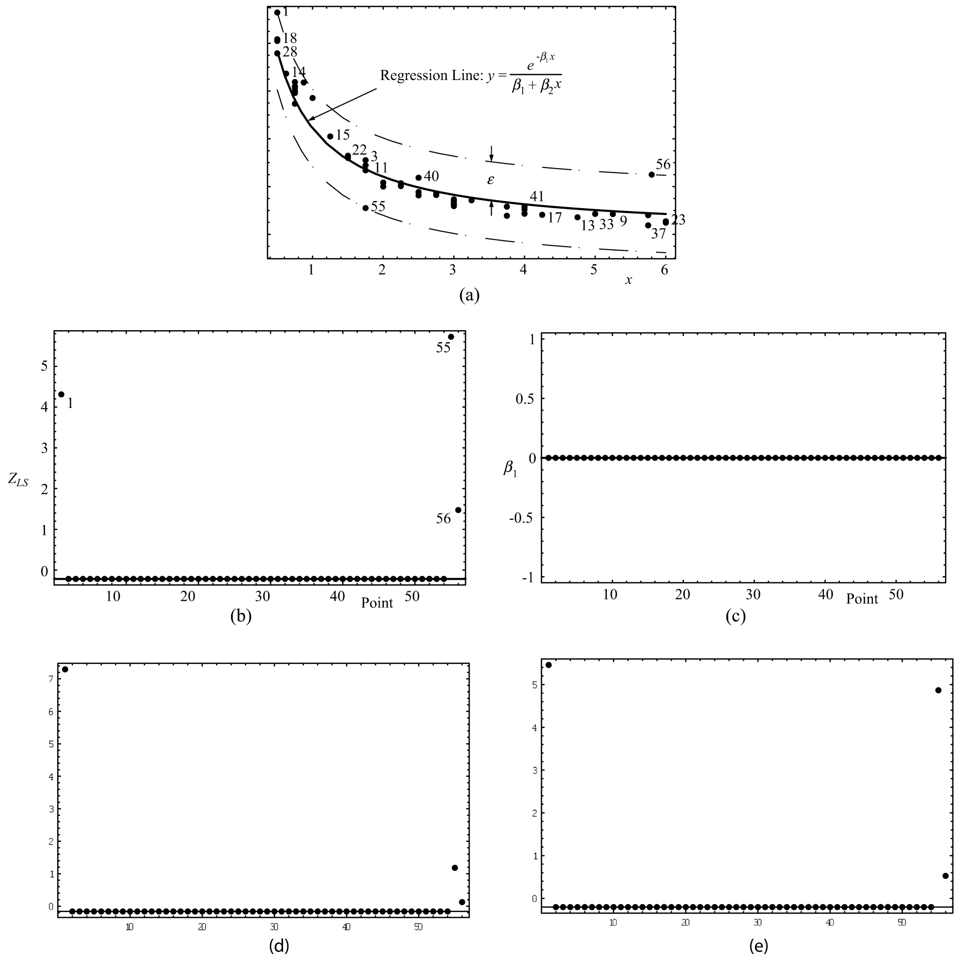


Figure 3. From left to right and top to bottom: (a) scatter plot of $y_i$ versus $x_i$, for the ultrasonic contaminated data and the Min–Max fitted model, (b)-(e) the index plot of the $Z^*_{\mathrm{LS}}$, $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ sensitivities, respectively.

outlier (observation number 1). Note that observations 1 and 56 define the upper dashed band and observation 55 defines the lower dashed line, both at a distance $\varepsilon = 16.2$ from the Min–Max-regression curve. Note also that the Min–Max results are always determined by at least $k + 1$ points, where $k$ is the number of parameters.

### 3.3. *Sensitivity of the LAV regression*

The matrices required to perform the sensitivity analysis for the LAV case are identical to those in the Min–Max case, but replacing $\varepsilon$ by $\varepsilon_i$ in all formulas.

The following LAV estimates for the ultrasonic data are obtained:

$$Z_{\mathrm{LAV}} = 105.493,\ \hat{\beta}_1 = 0.1511,\ \hat{\beta}_2 = 0.005,\ \hat{\beta}_3 = 0.0128.$$

As in the LAV case, the solution is not unique, and the derivatives do not exist. Thus, we consider the contaminated sample and obtain:

$$Z_{\mathrm{LAV}} = 142.105,\ \hat{\beta}_1 = 0.117,\ \hat{\beta}_2 = 0.0043,\ \hat{\beta}_3 = 0.0146.$$

which, unfortunately, is still not unique, therefore we conclude that this method is not adequate for estimating the regression line for this model.

## 4. A model for onion yield data

The onion yield data are taken from ref. [44], p. 58 (the Uraidla Variety). Here $y$ is the onion yield (g/plant) and $x$ is the plant density (plans/m$^2$). Ratkowsky [44] fits the model

$$\log(y) = -\log(\alpha + \beta x) \tag{32}$$

to the data, by LS.

### 4.1. *Sensitivity of the LS method*

In this case the objective function is:

$$Z_{\mathrm{LS}} = \sum_{i=1}^{n} (\log y_i + \log(\alpha + \beta x_i))^2,$$

and since there are not constraints, the sensitivities of the sum of squares to data are given by Theorem 1, that is:

$$\frac{\partial Z_{\mathrm{LS}}}{\partial x_i} = \frac{2\beta\varepsilon_i}{\alpha + \beta x_i}; \; \frac{\partial Z_{\mathrm{LS}}}{\partial y_i} = \frac{2\varepsilon_i}{y_i}.$$

The matrices we need for calculating the sensitivities according to the proposed method are:

$$\begin{aligned}
\boldsymbol{F_\theta} &: \frac{\partial Z_{\mathrm{LS}}}{\partial \alpha} = 2\sum_{i=1}^{n} \frac{\varepsilon_i}{\alpha + \beta x_i}; \; \frac{\partial Z_{\mathrm{LS}}}{\partial \beta} = 2\sum_{i=1}^{n} \frac{\varepsilon_i x_i}{\alpha + \beta x_i} \\
\boldsymbol{F_x} &: \frac{\partial Z_{\mathrm{LS}}}{\partial x_i} = 2\frac{\varepsilon_i \beta}{\alpha + \beta x_i}; \; \frac{\partial Z_{\mathrm{LS}}}{\partial y_i} = 2\frac{\varepsilon_i}{y_i}; \\
\boldsymbol{F_{\theta\theta}} &: \frac{\partial^2 Z_{\mathrm{LS}}}{\partial \alpha^2} = -2\sum_{i=1}^{n} \frac{\varepsilon_i - 1}{(\alpha + \beta x_i)^2}; \; \frac{\partial^2 Z_{\mathrm{LS}}}{\partial \beta^2} = -2\sum_{i=1}^{n} \frac{(\varepsilon_i - 1)x_i^2}{(\alpha + \beta x_i)^2} \\
&\quad \frac{\partial^2 Z_{\mathrm{LS}}}{\partial \alpha \partial \beta} = -2\sum_{i=1}^{n} \frac{(\varepsilon_i - 1)x_i}{(\alpha + \beta x_i)^2} \\
\boldsymbol{F_{\theta x}} &: \frac{\partial^2 Z_{\mathrm{LS}}}{\partial \alpha \partial x_i} = -2\frac{(\varepsilon_i - 1)\beta}{(\alpha + \beta x_i)^2}; \; \frac{\partial^2 Z}{\partial \beta \partial x_i} = 2\frac{(\varepsilon_i \alpha + x_i \beta)}{(\alpha + \beta x_i)^2} \\
&\quad \frac{\partial^2 Z_{\mathrm{LS}}}{\partial \alpha \partial y_i} = \frac{2}{y_i(\alpha + \beta x_i)^2}; \; \frac{\partial^2 Z_{\mathrm{LS}}}{\partial \beta \partial y_i} = \frac{2x_i}{y_i(\alpha + \beta x_i)^2}.
\end{aligned}$$

The LS estimates found by Ratkowsky [44], p. 59, are

$$\hat{\alpha} = 0.003462, \text{ and } \hat{\beta} = 0.000129.$$

The scatter plot of $y$ versus $x$, with the estimated LS curve, is shown in Figure 4. Observation 38 appears clearly from the graph to be an outlier. Because we have only two variables in this case,

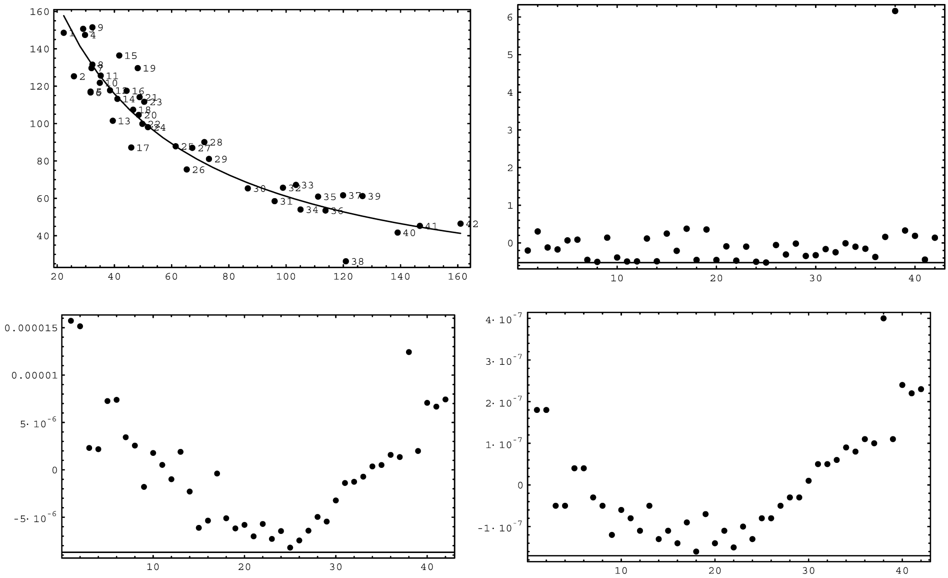

Figure 4. From left to right and top to bottom: (a) scatter plot of $y_i$ versus $x_i$, for the onion yield data set in Table 1 and the LS fitted model, (b)–(d) the index plot of the $Z^*_{\text{LS}}$, $\hat{\alpha}_1$, and $\hat{\beta}$ sensitivities, respectively.

outliers can be easily spotted upon the inspection of the scatter plot. But in higher dimensions, the detection of outliers becomes very difficult.

Ratkowsky [44] remarks that observation 38 is suspected outlier and is omitted from further analysis. For the purpose of sensitivity analysis, we include the outlier to see if any of the sensitivity measures is able to detect it. With the outlier included, we found the LS estimates to be

$$Z_{\text{LS}} = 0.912, \hat{\alpha} = 0.003462, \beta = 0.000129,$$

which agree with those found by Ratkowsky [44].

### 4.2. *Sensitivity of the Min–Max method*

In this case the objective function is $Z_{\text{Min–Max}} = \varepsilon$ and the constraints are

$$q_i^{(1)}(\alpha, \beta, \varepsilon; x_i, y_i) : \log(y_i) + \log(\alpha + \beta x_i) - \varepsilon \leq 0; \quad i = 1, 2, \ldots, n,$$
$$q_i^{(2)}(\alpha, \beta, \varepsilon; x_i, y_i) : -\log(y_i) - \log(\alpha + \beta x_i) - \varepsilon \leq 0; \quad i = 1, 2, \ldots, n.$$

The Lagrangian function is

$$\begin{aligned}\mathcal{L}(\alpha, \beta, \varepsilon; \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\eta}^{(1)}, \boldsymbol{\eta}^{(2)}) \quad &= \varepsilon + \textstyle\sum_{i=1}^{n_1} \eta_i^{(1)}(\log(y_i) + \log(\alpha + \beta x_i) - \varepsilon) + \\ &\quad + \textstyle\sum_{i=1}^{n_2} \eta_i^{(2)}(-\log(y_i) - \log(\alpha + \beta x_i) - \varepsilon)\end{aligned}$$

and, the sensitivities of the objective function to data are given by Theorem 1, that is:

$$\frac{\partial Z_{\text{Min–Max}}}{\partial x_i} = \sum_{i=1}^{n_1} \eta_i^{(1)} \frac{\beta}{\alpha + \beta x_i} - \sum_{i=1}^{n_2} \eta_i^{(2)} \frac{\beta}{\alpha + \beta x_i}; \quad \frac{\partial Z_{\text{Min–Max}}}{\partial y_i} = \sum_{i=1}^{n_1} \frac{\eta_i^{(1)}}{y_i} - \sum_{i=1}^{n_2} \frac{\eta_i^{(2)}}{y_i}.$$

The matrices we need for calculating the sensitivities according to the proposed method are given in Appendix 3. The following Min–Max estimates for the ultrasonic are obtained:

$$Z_{\text{Min–Max}} = 0.442, \hat{\alpha} = 0.00377, \hat{\beta} = 0.000171.$$

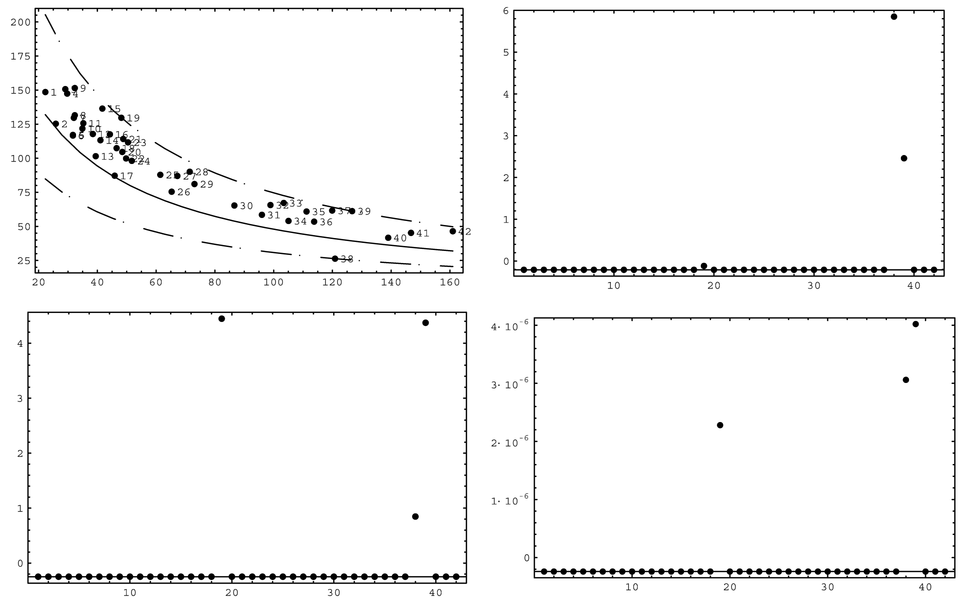

Figure 5. From left to right and top to bottom: (a) scatter plot of $y_i$ versus $x_i$, for the onion yield data set in Table 2 and the Min–Max fitted model, (b)–(d) the index plot of the $Z^*_{\mathrm{LS}}$, $\hat{\alpha}$, and $\hat{\beta}$ sensitivities, respectively.

Table 2. Onion yield data.

| $i$ | $x_i$ | $y_i$ | $i$ | $x_i$ | $y_i$ |
|---|---|---|---|---|---|
| 1 | 22.30 | 148.57 | 22 | 49.78 | 99.85 |
| 2 | 25.86 | 125.30 | 23 | 50.43 | 111.65 |
| 3 | 29.09 | 150.69 | 24 | 51.72 | 98.09 |
| 4 | 29.74 | 147.42 | 25 | 61.42 | 87.85 |
| 5 | 31.68 | 117.10 | 26 | 65.29 | 75.45 |
| 6 | 31.68 | 116.64 | 27 | 67.23 | 87.01 |
| 7 | 32.00 | 129.66 | 28 | 71.44 | 90.10 |
| 8 | 32.32 | 131.54 | 29 | 73.05 | 81.08 |
| 9 | 32.32 | 151.50 | 30 | 86.63 | 65.33 |
| 10 | 34.91 | 121.80 | 31 | 96.00 | 58.49 |
| 11 | 35.23 | 125.67 | 32 | 98.91 | 65.67 |
| 12 | 38.47 | 117.78 | 33 | 103.44 | 67.19 |
| 13 | 39.44 | 101.50 | 34 | 105.05 | 54.01 |
| 14 | 41.05 | 113.22 | 35 | 111.19 | 60.92 |
| 15 | 41.70 | 136.43 | 36 | 113.78 | 53.48 |
| 16 | 44.28 | 117.54 | 37 | 119.92 | 61.62 |
| 17 | 45.90 | 87.20 | 38 | 120.89 | 26.32 |
| 18 | 46.55 | 107.41 | 39 | 126.71 | 61.21 |
| 19 | 48.16 | 129.68 | 40 | 138.99 | 41.67 |
| 20 | 48.49 | 104.63 | 41 | 146.75 | 45.26 |
| 21 | 48.81 | 114.15 | 42 | 160.97 | 46.45 |

Figure 5 shows a scatter plot of $y_i$ versus $x_i$, for the onion yield data set in Table 2 and the LS fitted model together with the $Z_{\mathrm{Min–Max}}$, $\alpha$, and $\beta$ sensitivities. Note that observations 19 and 39 define the upper dashed band and point 38 defines the lower dashed line, both at a distance $\varepsilon = 0.442$ in the logaritmic scale from the Min–Max-regression line. They are the only influential points.

### 4.3. *Sensitivity of the LAV method*

In this case the objective function is

$$Z_{\text{LAV}} = \sum_{i=1}^{n} \varepsilon_i,$$

and the constraints are

$$q_i^{(1)}(\alpha, \beta, \varepsilon_i; x_i, y_i) : \log(y_i) + \log(\alpha + \beta x_i) - \varepsilon_i \le 0; \quad i = 1, 2, \ldots, n,$$

$$q_i^{(2)}(\alpha, \beta, \varepsilon_i; x_i, y_i) : -\log(y_i) - \log(\alpha + \beta x_i) - \varepsilon_i \le 0; \quad i = 1, 2, \ldots, n.$$

The Lagrangian function is

$$\begin{aligned}\mathcal{L}(\alpha, \beta, \varepsilon; \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\eta}^{(1)}, \boldsymbol{\eta}^{(2)}) &= \textstyle\sum_{i=1}^{n} \varepsilon_i + \sum_{i=1}^{n_1} \eta_i^{(1)}(\log(y_i) + \log(\alpha + \beta x_i) - \varepsilon_i) + \\ &\quad + \textstyle\sum_{i=1}^{n_2} \eta_i^{(2)}(-\log(y_i) - \log(\alpha + \beta x_i) - \varepsilon_i)\end{aligned}$$

and, the sensitivities of the objective function to data are given by Theorem 1:

$$\frac{\partial Z_{\text{LAV}}}{\partial x_i} = \sum_{i=1}^{n_1} \eta_i^{(1)} \frac{\beta}{\alpha + \beta x_i} - \sum_{i=1}^{n_2} \eta_i^{(2)} \frac{\beta}{\alpha + \beta x_i}; \qquad \frac{\partial Z_{\text{LAV}}}{\partial y_i} = \sum_{i=1}^{n_1} \frac{\eta_i^{(1)}}{y_i} - \sum_{i=1}^{n_2} \frac{\eta_i^{(2)}}{y_i}.$$

The matrices we need for calculating the sensitivities according to the proposed method are given in Appendix 4. The following LAV estimates for the onion yield data are obtained:

$$Z_{\text{LAV}} = 3.952, \hat{\alpha} = 0.00411, \hat{\beta} = 0.0001126.$$

Figure 6 shows a scatter plot of $y_i$ versus $x_i$, for the onion yield data set in Table 2 and the LAV fitted model together with the $Z_{\text{LAV}}$, $\alpha$, and $\beta$ sensitivities. Note that the LAV-regression line passes through data points 7 and 29. They are the only influential points. Note that the LAV curve must pass through at least as many points as the number of parameters.

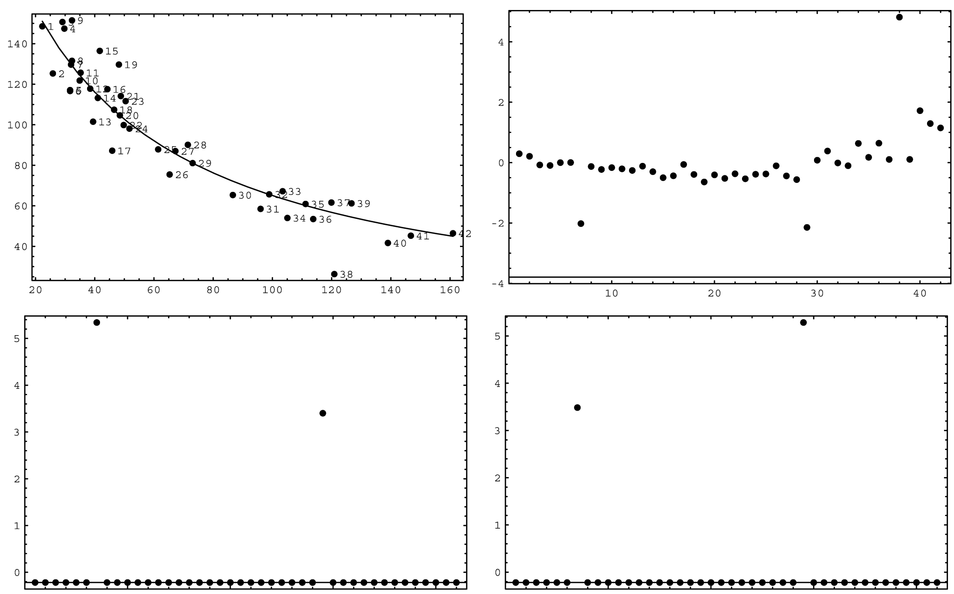

Figure 6. From left to right and top to bottom: (a) scatter plot of $y_i$ versus $x_i$, for the onion yield data set in Table 1 and the LAV fitted model, (b)-(d) the index plot of the $Z^*_{\text{LS}}$, $\hat{\alpha}_1$, and $\hat{\beta}$ sensitivities, respectively.

## 5. Conclusions

In this paper, we have performed a sensitivity analysis of the most common non-linear regression estimation methods. These include the LS, the Min–Max, and the LAV methods. The local sensitivities of the objective function, the regression coefficients, and the dual variables with respect to data are obtained using closed formulas in the regular cases. It is also shown that the Min–Max and LAV regression models are very prone to non-uniqueness, and when this occurs the corresponding partial derivatives or sensitivities do not exist.

## Appendix 1. Matrices needed for calculating estimates

### *Calculating the LS estimates in Section 3.1*

$$\boldsymbol{F_\theta}: \frac{\partial Z_{\mathrm{LS}}}{\partial \beta_1} = 2\sum_{i=1}^{n} \frac{\varepsilon_i x_i}{e^{x_i\beta_1}(\beta_2 + x_i\beta_3)}; \quad \frac{\partial Z_{\mathrm{LS}}}{\partial \beta_2} = 2\sum_{i=1}^{n} \frac{\varepsilon_i}{e^{x_i\beta_1}(\beta_2 + \beta_3 x_i)^2};$$

$$\frac{\partial Z_{\mathrm{LS}}}{\partial \beta_3} = 2\sum_{i=1}^{n} \frac{\varepsilon_i x_i}{e^{x_i\beta_1}(\beta_2 + \beta_3 x_i)^2};$$

$$\boldsymbol{F_x}: \frac{\partial Z_{\mathrm{LS}}}{\partial x_i} = 2\frac{\varepsilon_i(\beta_3 + \beta_1(\beta_2 + \beta_3 x_i))}{e^{\beta_1 x_i}(\beta_2 + \beta_3 x_i)^2}; \quad \frac{\partial Z_{\mathrm{LS}}}{\partial y_i} = 2\varepsilon_i;$$

$$\boldsymbol{F_{\theta\theta}}: \frac{\partial^2 Z_{\mathrm{LS}}}{\partial \beta_1^2} = 2\sum_{i=1}^{n} \frac{x_i^2(2 - e^{\beta_1 x_i}(\beta_2 + \beta_3 x_i)y_i)}{e^{2\beta_1 x_i}(\beta_2 + \beta_3 x_i)^2};$$

$$\frac{\partial^2 Z_{\mathrm{LS}}}{\partial \beta_2^2} = \sum_{i=1}^{n} \frac{6 - 4e^{\beta_1 x_i}(\beta_2 + \beta_3 x_i)y_i}{e^{2\beta_1 x_i}(\beta_2 + \beta_3 x_i)^4};$$

$$\frac{\partial^2 Z_{\mathrm{LS}}}{\partial \beta_3^2} = \sum_{i=1}^{n} \frac{2x_i^2(3 - 2e^{\beta_1 x_i}(\beta_2 + \beta_3 x_i)y_i)}{e^{2\beta_1 x_i}(\beta_2 + \beta_3 x_i)^4};$$

$$\frac{\partial^2 Z_{\mathrm{LS}}}{\partial \beta_1 \partial \beta_2} = \sum_{i=1}^{n} \frac{4x_i - 2e^{\beta_1 x_i}x_i(\beta_2 + \beta_3 x_i)y_i}{e^{2\beta_1 x_i}(\beta_2 + \beta_3 x_i)^3};$$

$$\frac{\partial^2 Z_{\mathrm{LS}}}{\partial \beta_1 \partial \beta_3} = \sum_{i=1}^{n} \frac{2x_i^2(2 - e^{\beta_1 x_i}(\beta_2 + \beta_3 x_i)y_i)}{e^{2\beta_1 x_i}(\beta_2 + \beta_3 x_i)^3};$$

$$\frac{\partial^2 Z_{\mathrm{LS}}}{\partial \beta_2 \partial \beta_3} = \sum_{i=1}^{n} \frac{6x_i - 4e^{\beta_1 x_i}x_i(\beta_2 + \beta_3 x_i)y_i}{e^{2\beta_1 x_i}(\beta_2 + \beta_3 x_i)^4};$$

$$\boldsymbol{F_{\theta x}}: \frac{\partial^2 Z_{\mathrm{LS}}}{\partial \beta_1 \partial x_i} = \sum_{i=1}^{n} \frac{2((\beta_2 + \beta_3 x_i(2\beta_1 x_i - e^{\beta_1 x_i}(\beta_2 + \beta_3 x_i)(\beta_1\beta_3 x_i^2 + \beta_2(\beta_1 x_i - 1))y_i))}{e^{2\beta_1 x_i}(\beta_2 + \beta_3 x_i)^3} + \sum_{i=1}^{n} \frac{2(-\beta_2 + \beta_3 x_i)}{e^{2\beta_1 x_i}(\beta_2 + \beta_3 x_i)^3};$$

$$\frac{\partial^2 Z_{\mathrm{LS}}}{\partial \beta_2 \partial x_i} = \sum_{i=1}^{n} \frac{2(3\beta_3 + (\beta_2 + \beta_3 x_i)(2\beta_1 - e^{\beta_1 x_i}(2\beta_3 + \beta_1(\beta_2 + \beta_3 x_i))y_i))}{e^{2\beta_1 x_i}(\beta_2 + \beta_3 x_i)^4};$$

$$\frac{\partial^2 Z_{\mathrm{LS}}}{\partial \beta_3 \partial x_i} = \sum_{i=1}^{n} \frac{2((\beta_2 + \beta_3 x_i)(2\beta_1 x_i - e^{\beta_1 x_i})(\beta_2(\beta_1 x_i - 1) + \beta_3 x_i(1 + \beta_1 x_i))y_i)}{e^{2\beta_1 x_i}(\beta_2 + \beta_3 x_i)^4} + \sum_{i=1}^{n} \frac{2(-\beta_2 + 2\beta_3 x_i)}{e^{2\beta_1 x_i}(\beta_2 + \beta_3 x_i)^4};$$

$$\frac{\partial^2 Z_{\mathrm{LS}}}{\partial\beta_1\partial y_i} = \sum_{i=1}^{n} \frac{2x_i}{e^{\beta_1 x_i}(\beta_2+\beta_3 x_i)}; \quad \frac{\partial^2 Z_{\mathrm{LS}}}{\partial\beta_2\partial y_i} = \sum_{i=1}^{n} \frac{2}{e^{\beta_1 x_i}(\beta_2+\beta_3 x_i)^2};$$

$$\frac{\partial^2 Z_{\mathrm{LS}}}{\partial\beta_3\partial y_i} = \sum_{i=1}^{n} \frac{2x_i}{e^{\beta_1 x_i}(\beta_2+\beta_3 x_i)^2};$$

## Appendix 2

### *Calculating the Min–Max estimates in Section 3.2*

$$\boldsymbol{F_\theta}: \frac{\partial Z_{\text{Min–Max}}}{\partial\beta_1} = \frac{\partial Z_{\text{Min–Max}}}{\partial\beta_2} = \frac{\partial Z_{\text{Min–Max}}}{\partial\beta_3} = 0; \quad \frac{\partial Z_{\text{Min–Max}}}{\partial\varepsilon} = 1;$$

$$\boldsymbol{F_x}: \frac{\partial Z_{\text{Min–Max}}}{\partial x_i} = \frac{\partial Z_{\text{Min–Max}}}{\partial y_i} = 0;$$

$$\boldsymbol{F_{\theta\theta}}: \frac{\partial^2 Z_{\text{Min–Max}}}{\partial\beta_i\partial\beta_j} = \frac{\partial^2 Z_{\text{Min–Max}}}{\partial\beta_j\partial\varepsilon} = \frac{\partial^2 Z_{\text{Min–Max}}}{\partial\varepsilon^2} = 0; \forall i, j;$$

$$\boldsymbol{F_{\theta x}}: \frac{\partial^2 Z_{\text{Min–Max}}}{\partial\beta_j\partial x_i} = \frac{\partial^2 Z_{\text{Min–Max}}}{\partial\beta_j\partial y_i} = \frac{\partial^2 Z_{\text{Min–Max}}}{\partial\varepsilon\partial x_i} = \frac{\partial^2 Z_{\text{Min–Max}}}{\partial\varepsilon\partial y_i} = 0; \forall i, j;$$

$$\boldsymbol{Q_\theta}: \frac{\partial q_i^{(1)}}{\partial\beta_1} = -\frac{\partial q_i^{(2)}}{\partial\beta_1} = \frac{x_i}{e^{\beta_1 x_i}(\beta_2+\beta_3 x_i)};$$

$$\frac{\partial q_i^{(1)}}{\partial\beta_2} = -\frac{\partial q_i^{(2)}}{\partial\beta_2} = \frac{1}{e^{\beta_1 x_i}(\beta_2+\beta_3 x_i)^2};$$

$$\frac{\partial q_i^{(1)}}{\partial\beta_3} = -\frac{\partial q_i^{(2)}}{\partial\beta_3} = \frac{x_i}{e^{\beta_1 x_i}(\beta_2+\beta_3 x_i)^2}; \quad \frac{\partial q_1^{(2)}}{\partial\varepsilon} = \frac{\partial q_i^{(2)}}{\partial\varepsilon} = -1;$$

$$\boldsymbol{Q_x}: \frac{\partial q_i^{(1)}}{\partial x_i} = -\frac{\partial q_i^{(2)}}{\partial x_i} = \frac{\beta_3+\beta_1(\beta_2+\beta_3 x_i)}{e^{\beta_1 x_i}(\beta_2+\beta_3 x_i)^2}; \quad \frac{\partial q_i^{(1)}}{\partial y_i} = -\frac{\partial q_i^{(2)}}{\partial y_i} = 1;$$

$$\boldsymbol{Q_{\theta\theta}}: \frac{\partial^2 q_i^{(1)}}{\partial\beta_1^2} = -\frac{\partial^2 q_i^{(2)}}{\partial\beta_1^2} = -\left(\frac{x_i^2}{e^{\beta_1 x_i}(\beta_2+\beta_3 x_i)}\right);$$

$$\frac{\partial^2 q_i^{(1)}}{\partial\beta_1\partial\beta_2} = -\frac{\partial^2 q_i^{(2)}}{\partial\beta_1\partial\beta_2} = -\left(\frac{x_i}{e^{\beta_1 x_i}(\beta_2+\beta_3 x_i)^2}\right);$$

$$\frac{\partial^2 q_i^{(1)}}{\partial\beta_1\partial\beta_3} = -\frac{\partial^2 q_i^{(2)}}{\partial\beta_1\partial\beta_3} = -\left(\frac{x_i^2}{e^{\beta_1 x_i}(\beta_2+\beta_3 x_i)^2}\right);$$

$$\frac{\partial^2 q_i^{(1)}}{\partial\beta_2^2} = -\frac{\partial^2 q_i^{(2)}}{\partial\beta_2^2} = \frac{-2}{e^{\beta_1 x_i}(\beta_2+\beta_3 x_i)^3};$$

$$\frac{\partial^2 q_i^{(1)}}{\partial\beta_2\partial\beta_3} = -\frac{\partial^2 q_i^{(2)}}{\partial\beta_2\partial\beta_3} = \frac{-2x_i}{e^{\beta_1 x_i}(\beta_2+\beta_3 x_i)^3};$$

$$\frac{\partial^2 q_i^{(1)}}{\partial\beta_3^2} = -\frac{\partial^2 q_i^{(2)}}{\partial\beta_3^2} = \frac{-2x_i^2}{e^{\beta_1 x_i}(\beta_2+\beta_3 x_i)^3};$$

$$\frac{\partial^2 q_i^{(j)}}{\partial\varepsilon^2} = \frac{\partial^2 q_i^{(j)}}{\partial\varepsilon\partial\beta_1} = \frac{\partial^2 q_i^{(j)}}{\partial\varepsilon\partial\beta_2} = \frac{\partial q_i^{(j)}}{\partial\varepsilon\partial\beta_3} = 0; \forall j;$$

$$\boldsymbol{Q_{\theta x}}: \frac{\partial^2 q_i^{(1)}}{\partial\beta_1\partial x_i} = -\frac{\partial^2 q_i^{(2)}}{\partial\beta_1\partial x_i} = \frac{-(\beta_1\beta_3 x_i^2)+\beta_2(1-\beta_1 x_i)}{e^{\beta_1 x_i}(\beta_2+\beta_3 x_i)^2};$$

$$\frac{\partial^2 q_i^{(1)}}{\partial\beta_2\partial x_i} = -\frac{\partial^2 q_i^{(2)}}{\partial\beta_2\partial x_i} = \frac{-2\beta_3-\beta_1(\beta_2+\beta_3 x_i)}{e^{\beta_1 x_i}(\beta_2+\beta_3 x_i)^3};$$

$$\frac{\partial^2 q_i^{(1)}}{\partial \beta_3 \partial x_i} = -\frac{\partial^2 q_i^{(2)}}{\partial \beta_3 \partial x_i} = \frac{\beta_2(1-\beta_1 x_i) - \beta_3 x_i(1+\beta_1 x_i)}{e^{\beta_1 x_i}(\beta_2+\beta_3 x_i)^3};$$

$$\frac{\partial^2 q_i^{(1)}}{\partial \epsilon \partial x_i} = \frac{\partial^2 q_i^{(1)}}{\partial \beta_1 \partial y_i} = \frac{\partial^2 q_i^{(1)}}{\partial \beta_2 \partial y_i} = \frac{\partial^2 q_i^{(1)}}{\partial \beta_3 \partial y_i} = \frac{\partial^2 q_i^{(1)}}{\partial \varepsilon \partial y_i} = 0.$$

## Appendix 3

### *Calculating the Min–Max estimates in Section 4.2*

$$\boldsymbol{F_\theta} : \frac{\partial Z_{\text{Min–Max}}}{\partial \alpha} = \frac{\partial Z_{\text{Min–Max}}}{\partial \beta} = 0; \quad \frac{\partial Z_{\text{Min–Max}}}{\partial \varepsilon} = 1;$$

$$\boldsymbol{F_x} : \frac{\partial Z_{\text{Min–Max}}}{\partial x_i} = \frac{\partial Z_{\text{Min–Max}}}{\partial y_i} = 0;$$

$$\boldsymbol{F_{\theta\theta}} : \frac{\partial^2 Z_{\text{Min–Max}}}{\partial \alpha^2} = \frac{\partial^2 Z_{\text{Min–Max}}}{\partial \beta^2} = \frac{\partial^2 Z_{\text{Min–Max}}}{\partial \alpha \partial \beta} = \frac{\partial^2 Z_{\text{Min–Max}}}{\partial \varepsilon^2}$$
$$= \frac{\partial^2 Z_{\text{Min–Max}}}{\partial \alpha \partial \varepsilon} = \frac{\partial^2 Z_{\text{Min–Max}}}{\partial \beta \partial \varepsilon} = 0;$$

$$\boldsymbol{F_{\theta x}} : \frac{\partial^2 Z_{\text{Min–Max}}}{\partial \alpha \partial x_i} = \frac{\partial^2 Z_{\text{Min–Max}}}{\partial \beta \partial x_i} = \frac{\partial^2 Z_{\text{Min–Max}}}{\partial \varepsilon \partial x_i} = \frac{\partial^2 Z_{\text{Min–Max}}}{\partial \alpha \partial y_i}$$
$$= \frac{\partial^2 Z_{\text{Min–Max}}}{\partial \beta \partial y_i} = \frac{\partial^2 Z_{\text{Min–Max}}}{\partial \varepsilon \partial y_i} = 0;$$

$$\boldsymbol{Q_\theta} : \frac{\partial q_i^{(1)}}{\partial \alpha} = -\frac{\partial q_i^{(2)}}{\partial \alpha} = \frac{1}{\alpha + \beta x_i}; \quad \frac{\partial q_i^{(1)}}{\partial \beta} = -\frac{\partial q_i^{(2)}}{\partial \beta} = \frac{x_i}{\alpha+\beta x_i}; \frac{\partial q_i^{(2)}}{\partial \varepsilon} = \frac{\partial q_i^{(2)}}{\partial \varepsilon} = -1;$$

$$\boldsymbol{Q_x} : \frac{\partial q_i^{(1)}}{\partial x_i} = -\frac{\partial q_i^{(1)}}{\partial x_i} = \frac{\beta}{\alpha+\beta x_i}; \quad \frac{\partial q_i^{(1)}}{\partial y_i} = -\frac{\partial q_i^{(2)}}{\partial y_i} = \frac{1}{y_i};$$

$$\boldsymbol{Q_{\theta\theta}} : \frac{\partial^2 q_i^{(1)}}{\partial \alpha^2} = -\frac{\partial^2 q_i^{(2)}}{\partial \alpha^2} = -\frac{1}{(\alpha+\beta x_i)^2}; \quad \frac{\partial^2 q_i^{(1)}}{\partial \beta^2} = -\frac{\partial^2 q_i^{(2)}}{\partial \beta^2} = -\frac{x_i^2}{(\alpha+\beta x_i)^2};$$
$$\frac{\partial^2 q_i^{(1)}}{\partial \alpha \partial \beta} = -\frac{\partial^2 q_i^{(1)}}{\partial \alpha \partial \beta} = -\frac{x_i}{(\alpha+\beta x_i)^2};$$
$$\frac{\partial^2 q_i^{(1)}}{\partial \varepsilon^2} = \frac{\partial^2 q_i^{(1)}}{\partial \alpha \partial \varepsilon} = \frac{\partial^2 q_i^{(1)}}{\partial \beta \partial \varepsilon} = \frac{\partial^2 q_i^{(2)}}{\partial \varepsilon^2} = \frac{\partial^2 q_i^{(2)}}{\partial \alpha \partial \varepsilon} = \frac{\partial^2 q_i^{(2)}}{\partial \beta \partial \varepsilon} = 0;$$

$$\boldsymbol{Q_{\theta x}} : \frac{\partial^2 q_i^{(1)}}{\partial \alpha \partial x_i} = -\frac{\partial^2 q_i^{(2)}}{\partial \alpha \partial x_i} = -\frac{\beta}{(\alpha+\beta x_i)^2}; \quad \frac{\partial^2 q_i^{(1)}}{\partial \beta \partial x_i} = -\frac{\partial^2 q_i^{(2)}}{\partial \beta \partial x_i} = \frac{\alpha}{(\alpha+\beta x_i)^2};$$
$$\frac{\partial^2 q_i^{(1)}}{\partial \varepsilon \partial x_i} = \frac{\partial^2 q_i^{(1)}}{\partial \alpha \partial y_i} = \frac{\partial^2 q_i^{(1)}}{\partial \beta \partial y_i} = \frac{\partial^2 q_i^{(1)}}{\partial \varepsilon \partial y_i} = \frac{\partial^2 q_i^{(2)}}{\partial \varepsilon \partial x_i} = \frac{\partial^2 q_i^{(2)}}{\partial \alpha \partial y_i} = \frac{\partial^2 q_i^{(2)}}{\partial \beta \partial y_i} = \frac{\partial^2 q_i^{(2)}}{\partial \varepsilon \partial y_i} = 0.$$

## Appendix 4

### *Calculating the LAV estimates in Section 4.3*

$$\boldsymbol{F_\theta} : \frac{\partial Z_{\text{LAV}}}{\partial \alpha} = \frac{\partial Z_{\text{LAV}}}{\partial \beta} = 0; \quad \frac{\partial Z_{\text{LAV}}}{\partial \varepsilon_i} = 1;$$

$$\boldsymbol{F_x} : \frac{\partial Z_{\text{LAV}}}{\partial x_i} = \frac{\partial Z_{\text{LAV}}}{\partial y_i} = 0;$$

$$\boldsymbol{F_{\theta\theta}} : \frac{\partial^2 Z_{\mathrm{LAV}}}{\partial \alpha^2} = \frac{\partial^2 Z_{\mathrm{LAV}}}{\partial \beta^2} = \frac{\partial^2 Z_{\mathrm{LAV}}}{\partial \alpha \partial \beta} = 0;$$

$$\frac{\partial^2 Z_{\mathrm{LAV}}}{\partial \varepsilon_i^2} = \frac{\partial^2 Z_{\mathrm{LAV}}}{\partial \alpha \partial \varepsilon_i} = \frac{\partial^2 Z_{\mathrm{LAV}}}{\partial \beta \partial \varepsilon_i} = 0;$$

$$\boldsymbol{F_{\theta x}} : \frac{\partial^2 Z_{\mathrm{LAV}}}{\partial \alpha \partial x_i} = \frac{\partial^2 Z_{\mathrm{LAV}}}{\partial \beta \partial x_i} = \frac{\partial^2 Z_{\mathrm{LAV}}}{\partial \varepsilon_i \partial x_i} = 0;$$

$$\frac{\partial^2 Z_{\mathrm{LAV}}}{\partial \alpha \partial y_i} = \frac{\partial^2 Z_{\mathrm{LAV}}}{\partial \beta \partial y_i} = \frac{\partial^2 Z_{\mathrm{LAV}}}{\partial \varepsilon_i \partial y_i} = 0;$$

$$\boldsymbol{Q_{\theta}} : \frac{\partial q_i^{(1)}}{\partial \alpha} = -\frac{\partial q_i^{(2)}}{\partial \alpha} = \frac{1}{\alpha + \beta x_i}; \quad \frac{\partial q_i^{(1)}}{\partial \beta} = -\frac{\partial q_i^{(2)}}{\partial \beta} = \frac{x_i}{\alpha + \beta x_i};$$

$$\frac{\partial q_i^{(2)}}{\partial \varepsilon_i} = \frac{\partial q_i^{(2)}}{\partial \varepsilon_i} = -1;$$

$$\boldsymbol{Q_{x}} : \frac{\partial q_i^{(1)}}{\partial x_i} = -\frac{\partial q_i^{(1)}}{\partial x_i} = \frac{\beta}{\alpha + \beta x_i}; \quad \frac{\partial q_i^{(1)}}{\partial y_i} = -\frac{\partial q_i^{(2)}}{\partial y_i} = \frac{1}{y_i};$$

$$\boldsymbol{Q_{\theta\theta}} : \frac{\partial^2 q_i^{(1)}}{\partial \alpha^2} = -\frac{\partial^2 q_i^{(2)}}{\partial \alpha^2} = -\frac{1}{(\alpha + \beta x_i)^2};$$

$$\frac{\partial^2 q_i^{(1)}}{\partial \beta^2} = -\frac{\partial^2 q_i^{(2)}}{\partial \beta^2} = -\frac{x_i^2}{(\alpha + \beta x_i)^2};$$

$$\frac{\partial^2 q_i^{(1)}}{\partial \alpha \partial \beta} = -\frac{\partial^2 q_i^{(1)}}{\partial \alpha \partial \beta} = -\frac{x_i}{(\alpha + \beta x_i)^2};$$

$$\frac{\partial^2 q_i^{(1)}}{\partial \varepsilon_i^2} = \frac{\partial^2 q_i^{(1)}}{\partial \alpha \partial \varepsilon_i} = \frac{\partial^2 q_i^{(1)}}{\partial \beta \partial \varepsilon_i} = \frac{\partial^2 q_i^{(2)}}{\partial \varepsilon_i^2} = \frac{\partial^2 q_i^{(2)}}{\partial \alpha \partial \varepsilon_i} = \frac{\partial^2 q_i^{(2)}}{\partial \beta \partial \varepsilon_i} = 0;$$

$$\boldsymbol{Q_{\theta x}} : \frac{\partial^2 q_i^{(1)}}{\partial \alpha \partial x_i} = -\frac{\partial^2 q_i^{(2)}}{\partial \alpha \partial x_i} = -\frac{\beta}{(\alpha + \beta x_i)^2};$$

$$\frac{\partial^2 q_i^{(1)}}{\partial \beta \partial x_i} = -\frac{\partial^2 q_i^{(2)}}{\partial \beta \partial x_i} = \frac{\alpha}{(\alpha + \beta x_i)^2};$$

$$\frac{\partial^2 q_i^{(1)}}{\partial \varepsilon_i \partial x_i} = \frac{\partial^2 q_i^{(1)}}{\partial \alpha \partial y_i} = \frac{\partial^2 q_i^{(1)}}{\partial \beta \partial y_i} = \frac{\partial^2 q_i^{(1)}}{\partial \varepsilon_i \partial y_i} = 0;$$

$$\frac{\partial^2 q_i^{(2)}}{\partial \varepsilon_i \partial x_i} = \frac{\partial^2 q_i^{(2)}}{\partial \alpha \partial y_i} = \frac{\partial^2 q_i^{(2)}}{\partial \beta \partial y_i} = \frac{\partial^2 q_i^{(2)}}{\partial \varepsilon_i \partial y_i} = 0.$$

## References


[1] A.C. Atkinson, *Plots, Transformations, and Regression: an Introduction to Graphical Methods of Diagnostic Regression Analysis*, Clarendon Press, Oxford, 1985.
[2] V. Barnett and T. Lewis, *Outliers in Statistical Data*, 3rd ed., John Wiley & Sons, New York, 1994.
[3] D.A. Belsley, E. Kuh, and R.E. Welsch, *Regression Diagnostics: Identifying Influential Data and Sources of Multicollinearity*, John Wiley & Sons, New York, 1980.
[4] N. Billor, S. Chatterjee, and A.S. Hadi, *A re-weighted least squares method for robust regression estimation*, Am. J. Math. Manage. Sci. 26 (2006), pp. 229–252.
[5] S. Chatterjee and A.S. Hadi, *Sensitivity Analysis in Linear Regression*, John Wiley & Sons, New York, 1988.
[6] R.D. Cook and S. Weisberg, *Residuals and Influence in Regression*, Chapman and Hall, London, 1982.
[7] D.M. Hawkins, *Identification of Outliers*, Chapman and Hall, London, 1980.
[8] B.E. Barrett and J.B. Gray, *On the use of robust diagnostics in least squares regression analysis*, Proceedings of the Statistical Computing Section, The American Statistical Association, 1997, pp. 130–135.
[9] R.D. Cook, *Detection of influential observations in linear regression*. Technometrics 19 (1977), pp. 15–18.
[10] ———, *Assessment of local influence (with discussion)*, J. R. Stat. Soc. B 48 (1986), pp. 133–169.
[11] J.B. Gray and R.F. Ling, *K-clustering as a detection tool for influential subsets in regression (with discussion)*, Technometrics 26 (1984), pp. 305–330.

[12] A.S. Hadi, *A new measure of overall potential influence in linear regression*, Comput. Stat. Data Anal. 14 (1992), pp. 1–27.

[13] A.S. Hadi and J.S. Simonoff, *Procedures for the identification of multiple outliers in linear models*, J. Am. Stat. Assoc. 88 (1993), pp. 1264–1272.

[14] W.D. Jones and R.F. Ling, *A new unifying class of influence measures for regression diagnostics*, Proceedings of the Statistical Computing Section, The American Statistical Association, 1988, pp. 305–310.

[15] M.S. Mayo and J.B. Gray, *Elemental subsets: the building blocks of regression*, J. Am. Stat. Assoc. 51 (1997), pp. 122–129.

[16] H. Nyquist, *Sensitivity analysis in empirical studies*, J. Off. Stat. 8 (1992), pp. 167–182.

[17] S.R. Paul and K.Y. Fung, *A generalized extreme studentized residual multiple-outlier-detection procedure in linear regression*, Technometrics 33 (1991), pp. 339–348.

[18] D. Peña and V. Yohai, *The detection of influential subsets in linear regression by using an influence matrix*, J. R. Stat. Soc. B, 57 (1995), pp. 145–156.

[19] B. Schwarzmann, *A connection between local-influence analysis and residual diagnostics*, Technometrics 33 (1991), pp. 103–104.

[20] J.S. Simonoff, *General approaches to stepwise identification of unusual values in data analysis*, in *Directions in Robust Statistics and Diagnostics: Part II*, W. Stahel and S. Weisberg, eds., Springer-Verlag, New York, 1991, pp. 223–242.

[21] I. Weissfeld and H. Schneider, *Influence diagnostics for the normal linear model with censored data*, Aust. J. Stat. 32 (1990a), pp. 11–20.

[22] ———, *Influence diagnostics for the Weibull model fit to censored data*, Stat. Probab. Lett. 9 (1990b), pp. 67–73.

[23] W.J. Winsnowski, D.C. Montgomery, and R.S. James, *A comparative analysis of multiple outlier detection procedures in the linear regression model*, Comput. Stat. Data Anal. 36 (2001), pp. 351–382.

[24] O.B. Sheynin, *Origin of the theory of errors*, Nature 211 (1966), pp. 1003–1004.

[25] R.L. Plackett, *Studies in the history of probability and statistics. XXIX: the method of least squares*, Biometrika 59 (1972), pp. 239–251.

[26] P.S. Laplace, *Mémoire sur la figure de la terre*, Histoire de l'Académie Royale des Sciences (1783); *Avec les Mé moires de Mathématique & de Physique, pour la même Année*, Mémoires (1786), pp. 17–46.

[27] ———, *Sur quelques points du systéme du monde*, Historie de l'Académie des Sciences, Année 1789; *Avec les Mé moires de Mathématique & de Physique, pour la même Année*, Mémoires (1793), pp. 1–87.

[28] F.R. Hampel *et al.*, *Robust Statistics*, John Wiley & Sons, New York, 1986.

[29] J.A. Visek, *On high breakdown point estimation*, Comput. Stat. 11 (1996), pp. 137–146.

[30] ———, *Sensitivity analysis of M-estimates of nonlinear regression model: influence of data subsets*, Ann. Inst. Stat. Math. 54 (2002), pp. 261–290.

[31] ———, *The least trimmed squares. sensitivity study*, Proceedings of Prague Stochastics, MatFyzpress, 2006, pp. 728–738.

[32] D.M. Hawkins and D.J. Olive, *Improved feasible solution algorithms for high breakdown estimation*, Comput. Stat. Data Analy. 30 (1999), pp. 1–11.

[33] J.A. Visek, *On the diversity of estimates*, Comput. Stat. Data Anal. 34 (2000), pp. 67–89.

[34] ———, *Instrumental weighted variables*, Austrian J. Stat. 35 (2006), pp. 379–387.

Q3 [35] A.S. Hadi and H. Nyquist, *Sensitivity analysis in statistics*. J. Stat. Stud. A Special Volume in Honor of Professor Mir Masoom. Ali's 65th Birthday. (2002), pp. 125–138.

[36] M.S. Bazaraa, H.D. Sherali, and C.M. Shetty, *Nonlinear Programming, Theory and Algorithms*, 2nd ed., John Wiley & Sons, New York, 1993.

[37] E. Castillo *et al.*, *Building and Solving Mathematical Programming Models in Engineering and Science*, John Wiley & Sons, New York, 2001.

[38] E. Castillo *et al.*, *A closed formula for local sensitivity analysis in mathematical programming*, Eng. Optim. 38 (2006), pp. 93–112.

[39] D.G. Luenberger, *Linear and Nonlinear Programming*, 2nd ed., Addison-Wesley, Reading, MA, 1989.

[40] E. Castillo *et al.*, *A perturbation approach to sensitivity analysis in mathematical programming*, J. Optim. Theory Appl. 128 (2006), pp. 49–74.

[41] E. Castillo *et al.*, *Closed formulas in local sensitivity analysis for some classes of linear and non-linear problems*, TOP 15 (2007), pp. 355–371.

[42] E. Castillo *et al.*, *A general method for local sensitivity analysis with application to regression models and other optimization problems*, Technometrics 46 (2004), pp. 430–444.

[43] A. Conejo *et al.*, *Decomposition Techniques in Mathematical Programming. Engineering and Science Applications*, Springer, Berlin, Heildelberg, 2006.

[44] D.A. Ratkowsky, *Nonlinear Regression Modeling: A United Practical Approach*, Marcel Dekker, New York, 1983.

[45] M. Barinaga, *Manic depression gene put in limbo*, Science 246 (1989), pp. 886–887.

Q4 [46] G.A.F. Seber and C.J. Wild, *Nonlinear Regression*, John Wiley & Sons, New York, 1989.

Q4 [47] J.J. Tiede and M. Pagano, *The application of robust calibration to radioimmunoassay*, Biometrics, 35 (1979), pp. 567–574.

Q4 [48] J.A. Visek, *Sensitivity Analysis of M-Estimates*, Ann. Inst. Stat. Math. 48 (1996), pp. 469–495.