Taylor & Francis
Taylor & Francis Group

Journal ...Journal of Statistical Computation and Simulation

Article ID ... GSCS245941

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 provide the location of the proceedings in Ref. 17.	
Q2	Please provide complete publication details for Ref. 26.	
Q3	Refs. 30–32 are not cited in the text. Pl. cite them.	
Q4	Please confirm the citation of Figure 4 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. 00, Month 2007, 1–23

Duality and local sensitivity analysis in least squares, minimax, and least absolute values regressions

ENRIQUE CASTILLO†, CARMEN CASTILLO‡, ALI S. HADI*§
and ROBERTO MÌNGUEZ¶||

†Department of Applied Mathematics and Computational Sciences, University of Cantabria, Spain
‡Civil Engineering Department, University of Castilla La Mancha, Spain
§Department of Mathematics, The American University in Cairo, Egypt
¶Department of Applied Mathematics, University of Castilla La-Mancha, Spain
||Department of Statistical Science, Cornell University, USA

(Received)

This paper deals with the problem of local sensitivity analysis in regression, i.e., how sensitive the results of a regression model (objective function, parameters, and dual variables) are to changes in the data. We use a general formula for local sensitivities in optimization problems to calculate the sensitivities in three standard regression problems (least squares, minimax, and least absolute values). Closed formulas for all sensitivities are derived. Sensitivity contours are presented to help in assessing the sensitivity of each observation in the sample. The dual problems of the minimax and least absolute values are obtained and interpreted. The proposed sensitivity measures are shown to deal more effectively with the masking problem than the existing methods. The methods are illustrated by their application to some examples and graphical illustrations are given.

Keywords: Dual problem; Dual variables; Mathematical programming; Optimization problems; Outliers; Primal problem

1. Introduction and motivation

Regression models are frequently used to analyse data and to describe the reality being observed. Various methods are used to estimate the parameters of a regression model based on data. Methods of estimation include least squares (LS), minimax (MM), and least absolute values (LAV). Though MM and LAV methods had initially a great success, they were obscured by the appearance of the LS method. Later, they somewhat recovered from this set back (see [1, 2]), when it was discovered that they correspond to maximum likelihood estimators for the uniform and double exponential residuals, respectively, but they returned to obscurity mainly due to their associated computational complexities. Recently, Portnoy and Koenter [3] have shown the interesting result that there are algorithms that make them competitive with the LS method, and even superior for some sample sizes.

*Corresponding author. Email: ahadi@aucegypt.edu

Journal of Statistical Computation and Simulation
ISSN 0094-9655 print/ISSN 1563-5163 online
http://www.tandf.co.uk/journals
DOI: 10.1080/00949650701460616

All these methods, however, can be substantially influenced by small changes in the data, hence the selected model is strongly dependent on the available data. It is, therefore, essential for data analysts to be able to assess the sensitivity of regression results to various perturbations in the data, so as to make adequate corrections when necessary. Sensitivity analysis is important because it adds quality to statistical studies.

Most of the proposed methods use the deletion approach. Let $x_1, x_2, \ldots, x_n$ be a random sample drawn from $f(x;\theta)$, which depends on a possibly vector-valued parameter θ. The deletion approach consists of taking the difference between two estimates of a parameter θ: an estimate $\hat{\theta}$ obtained from the full data and the same estimate $\hat{\theta}_{(i)}$ obtained after an observation x_i is deleted from the data. Large scaled difference indicates that the observation is influential on the parameter estimate. There is a large literature on this approach; see, for example, the books by Belsley, Kuh, and Welsch [4], Cook and Weisberg [5], Atkinson [6], Chatterjee and Hadi [7], Jones and Ling [8], Weissfeld and Schneider [9, 10], Schwarzmann [11], Paul and Fung [12], Escobar and Meeker [13], Hadi [14], Hadi and Simonoff [15], Peña and Yohai [16], Barrett and Gray [17], Mayo and Gray [18], Saltelli *et al.* [19], and Winsnowski *et al.* [20].

Another approach to sensitivity analysis, proposed by Cook [21], is a weighted perturbation approach, where each observation is given a weight ω_i, with $0 \le \omega_i \le 1$. The influence of an observation x_i is then measured by the likelihood displacement

$$LD(\omega) = 2[L(\hat{\theta}) - L(\hat{\theta}_{\omega})], \tag{1}$$

where $\omega = \{\omega_1, \ldots, \omega_n\}$, $\hat{\theta}$ is the maximum likelihood estimate of θ, and $\hat{\theta}_{\omega}$ is the maximum likelihood estimate of θ, when x_i is the given weight ω 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. The weighted perturbation approach applies to the least squares normal regression, but does not apply to the MM and LAV.

In this paper, we present methods for assessing the sensitivity of the parameter estimates in regression models to changes in the data, not the weights. Furthermore, sensitivity analysis has been almost exclusively applied to least squares regression. Castillo *et al.* [22] give the sensitivities of the objective function to data, but not the sensitivities of the regression parameters to data. In this paper, on one hand, we extend sensitivity analysis to regression parameters and, on the other hand, to alternative regression methods such as MM and LAV, and include sensitivities of dual variables to data. The approach is new and very general. In fact, it can be applied to any model, including linear and nonlinear models, and to any method of estimation that can be formulated as an optimization problem. The proposed sensitivity measures are shown to deal more effectively with the masking problem than the existing methods.

The paper is structured as follows. Section 2, reviews the important concept of duality in optimization problems and gives very important and simple formulas for local sensitivity analysis. Section 3 introduces the standard linear regression model and describes a data set to be used as an illustrative numerical example. Sections 4–6 deal with the problem of local sensitivity analysis in least squares, minimax, and least absolute value regressions, respectively, where closed-form formulas for the sensitivities of the objective function, the parameter estimates, and the primal and dual variables are obtained. Finally, a summary is given in section 7.

2. Some background on duality and sensitivity analysis

In this section, we remind the reader about duality and give some closed formulas, which allow in obtaining the sensitivities of the objective function values and the primal and dual variables of an optimization problem with respect to the data. These formulas allow in dealing with the problem of sensitivity analysis in the regression problems dealt with in this paper.

2.1 *Duality*

Consider the following general nonlinear *primal problem* (P):

$$\text{Minimize}_{\boldsymbol{\beta}} Z_P = f(\boldsymbol{\beta}; z) \tag{2}$$

subject to

$$\mathbf{h}(\boldsymbol{\beta}; z) = \mathbf{0};\ \boldsymbol{\lambda} \tag{3}$$

$$\mathbf{g}(\boldsymbol{\beta}; z) \leq \mathbf{0};\ \boldsymbol{\mu}, \tag{4}$$

where boldfaced letters refer to vectors, $\boldsymbol{\beta} \in \mathbb{R}^n, z \in \mathbb{R}^p$, $\mathbf{h}(\boldsymbol{\beta}; z) \in \mathbb{R}^\ell$ and $\mathbf{g}(\boldsymbol{\beta}; z) \in \mathbb{R}^m$, and $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ are the dual variables to be introduced below.

Every primal nonlinear programming problem P of the form (2)–(4), has an associated dual problem D, which is defined as:

$$\text{Maximize}_{\boldsymbol{\lambda},\boldsymbol{\mu}}\ Z_D = \text{In } f_{\boldsymbol{\beta}}\{\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\mu}; z)\} \tag{5}$$

subject to

$$\boldsymbol{\mu} \geq \mathbf{0}, \tag{6}$$

where

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\mu}; z) = f(\boldsymbol{\beta}; z) + \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{h}(\boldsymbol{\beta}; z) + \boldsymbol{\mu}^{\mathrm{T}}\mathbf{g}(\boldsymbol{\beta}; z), \tag{7}$$

is the Lagrangian function associated with the primal problem (2)–(4), and $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ are called dual variables and they are vectors of dimensions ℓ and m, the number of equalities and inequalities in the primal problem, respectively.

2.2 *Sensitivity in nonlinear problems without constraints*

In this section, we consider the sensitivity in unconstrained nonlinear optimization problems. Suppose now that the problem in equation (2) has no constraints. Let

$$\boldsymbol{F}_{\boldsymbol{\beta}_{(1\times n)}} = (\nabla_{\boldsymbol{\beta}} f(\boldsymbol{\beta}^*, z))^{\mathrm{T}}, \tag{8}$$

$$\boldsymbol{F}_{z_{(1\times p)}} = (\nabla_z f(\boldsymbol{\beta}^*, z))^{\mathrm{T}}, \tag{9}$$

$$\boldsymbol{F}_{\boldsymbol{\beta\beta}_{(n\times n)}} = \nabla_{\boldsymbol{\beta\beta}} f(\boldsymbol{\beta}^*, z), \tag{10}$$

$$\boldsymbol{F}_{\boldsymbol{\beta}z_{(n\times p)}} = \nabla_{\boldsymbol{\beta}z} f(\boldsymbol{\beta}^*, z), \tag{11}$$

where the asterisk refers to the optimal values. Then, provided that $\mathbf{F}_{\boldsymbol{\beta\beta}}$ is invertible, the sensitivities of the optimal solution $(\boldsymbol{\beta}^*, Z^*)$ of the problem in equation (2) to changes in the data are determined by

$$\frac{\partial \beta}{\partial z}_{(n\times p)} = -\mathbf{F}_{\beta\beta}^{-1}\mathbf{F}_{\beta z}, \tag{12}$$

$$\frac{\partial Z_P}{\partial z}_{(1\times p)} = -\mathbf{F}_{\beta}\mathbf{F}_{\beta\beta}^{-1}\mathbf{F}_{\beta z} + \mathbf{F}_z = \mathbf{F}_z. \tag{13}$$

For a more complete derivation of sensitivity results, the reader is referred to Castillo *et al.* [23, 24].

2.3 *Sensitivity in linear programming*

Consider the following LP problem

$$\text{Minimize}_{\,\beta}\, Z = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{\beta}, \tag{14}$$

subject to

$$\mathbf{A}\boldsymbol{\beta} = \mathbf{b};\, \boldsymbol{\lambda}, \tag{15}$$

where $\mathbf{c} = (c_1, c_2, \ldots, c_n)$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_n)$, $\mathbf{b} = (b_1, b_2, \ldots, b_m) \geq 0$, $\mathbf{A}$ is a matrix of dimensions $m \times n$ with elements a_{ij}; $i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$, and $\boldsymbol{\lambda}$ are the dual variables.

Then, the sensitivities of the objective function, primal and dual variables with respect to data are given by the following closed form and simple formulas (see [25]):

$$\begin{aligned} &\frac{\partial Z}{\partial c_j} = \beta_j; \quad \frac{\partial Z}{\partial a_{ij}} = -\lambda_i\beta_j; \quad \frac{\partial Z}{\partial b_i} = \lambda_i, \\ &\frac{\partial \beta_j}{\partial c_k} = 0; \quad \frac{\partial \beta_j}{\partial a_{ik}} = -a^{ji}\beta_k \quad \frac{\partial \beta_j}{\partial b_i} = a^{ji} \\ &\frac{\partial \lambda_i}{\partial c_j} = -a^{ji}; \quad \frac{\partial \lambda_i}{\partial a_{\ell j}} = -a^{ji}\lambda_\ell; \quad \frac{\partial \lambda_i}{\partial b_\ell} = 0 \end{aligned} \tag{16}$$

where a^{ji} are the elements of $\mathbf{A}^{-1}$.

3. An example

In this section, we introduce the linear regression model and an illustrative numerical example that we will use throughout this paper.

The standard linear regression model is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon, \tag{17}$$

where $\mathbf{Y} = (y_1, \ldots, y_n)^{\mathrm{T}}$ is an $n \times 1$ vector of response variables, $\mathbf{X}$ is an $n \times k$ matrix of rank k of predictor variables, $\mathbf{x}_i^{\mathrm{T}}$ is the ith row in $\mathbf{X}$, $\boldsymbol{\beta}$ is a $k \times 1$ vector of regression parameters, and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^{\mathrm{T}}$ is an $n \times 1$ vector of independent random errors.

Castillo *et al.* [22] use these results to obtain only the sensitivity of the objective function with respect to changes in the data. In this paper, we show that the above results can also be used to derive the sensitivity of the estimated parameters with respect to changes in the data. Note that these sensitivities are more interesting than those related to the objective function values. We apply them in sections 4–6 to three regression estimation problems: the least squares (LS), minimax (MM), and least absolute value (LAV).

We illustrate the methods using the star cluster data set, which is a well-known data set in the area of sensitivity analysis and outliers detection and it has been analysed by many authors. Two variables are measured for each of the 47 stars: the effective temperature at the surface of a star (x) and the light intensity of the star (y). These real data, taken from Rousseeuw and Leroy [26], p. 57, is shown in table 1. The scatter plot of Y versus X in figure 1 shows that

Table 1. The stars data (Y and X), studentized residuals (r_i), and Cook's distances (C_i).

Index (i)	y_i	x_i	r_i	C_i
1	5.23	4.37	0.43	0.002
2	5.74	4.56	1.49	0.043
3	4.93	4.26	−0.19	0.000
4	5.74	4.56	1.49	0.043
5	5.19	4.30	0.30	0.001
6	5.46	4.46	0.91	0.011
7	4.65	3.84	−1.06	0.047
8	5.27	4.57	0.66	0.009
9	5.57	4.26	0.94	0.010
10	5.12	4.37	0.23	0.001
11	5.73	3.49	0.66	0.053
12	5.45	4.43	0.86	0.010
13	5.42	4.48	0.85	0.011
14	4.05	4.01	−1.97	0.090
15	4.26	4.29	−1.35	0.020
16	4.58	4.42	−0.68	0.006
17	3.94	4.23	−1.97	0.045
18	4.18	4.42	−1.39	0.024
19	4.18	4.23	−1.54	0.028
20	5.89	3.49	0.97	0.114
21	4.38	4.29	−1.14	0.014
22	4.22	4.29	−1.42	0.022
23	4.42	4.42	−0.97	0.012
24	4.85	4.49	−0.15	0.000
25	5.02	4.38	0.06	0.000
26	4.66	4.42	−0.54	0.004
27	4.66	4.29	−0.65	0.005
28	4.90	4.38	−0.15	0.000
29	4.39	4.22	−1.18	0.017
30	6.05	3.48	1.28	0.202
31	4.42	4.38	−1.00	0.011
32	5.10	4.56	0.35	0.002
33	5.22	4.45	0.47	0.003
34	6.49	3.49	2.14	0.552
35	4.34	4.23	−1.26	0.019
36	5.62	4.62	1.33	0.043
37	5.10	4.53	0.33	0.002
38	5.22	4.45	0.47	0.003
39	5.18	4.53	0.47	0.004
40	5.57	4.43	1.08	0.015
41	4.62	4.38	−0.64	0.005
42	5.06	4.45	0.19	0.000
43	5.34	4.50	0.73	0.008
44	5.34	4.45	0.69	0.006
45	5.54	4.55	1.13	0.024
46	4.98	4.45	0.05	0.000
47	4.50	4.42	−0.82	0.008

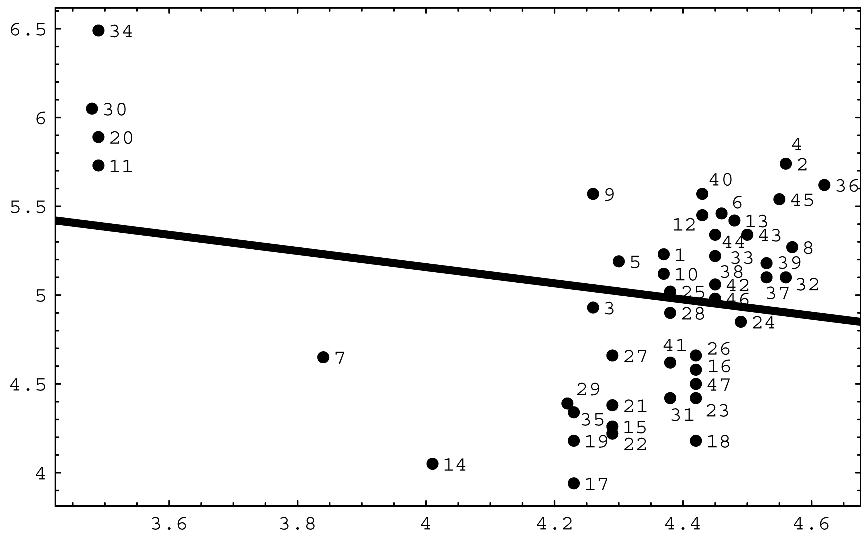

Figure 1. Scatter plot of Y versus X for the star cluster data and the resulting least squares regression line.

there is a direct relationship between the two variables for all the stars except for four unusual stars at the upper-left corner of the graph (cases 11, 20, 30, and 34). These stars, which are known as the giant stars, have low temperature with high light intensity. They can be expected to exert undue influence on the estimated regression parameters.

It should be noted here that we have selected a two-dimensional example with the purpose of illustrating concepts using some graphical displays that are only possible for two dimensions. There is no loss of generality, however, because the method works for any dimensions.

Let us now fit a linear model to the data using least squares and try to find these influential observations using existing regression diagnostic measures. The least squares line is found to be $y = 6.979 - 0.455x$. This line is drawn on the scatter plot in figure 1. Note here the effects of the four giant stars on the least squares regression line. The estimated line has a negative slope, which is contrary to the expectation of the relationship between light and temperature.

Two of the most commonly used ones are the internally studentized residuals and Cook's distances [27]. These are shown in the last two columns in table 1. Figure 2 shows the index plots of the internally studentized residual and Cook's distance. It can be seen from these graphs that the studentized residuals fail to detect any of the four giant starts. Cook's distance nominates only two observations as influential: observation 34 is clearly separated from all

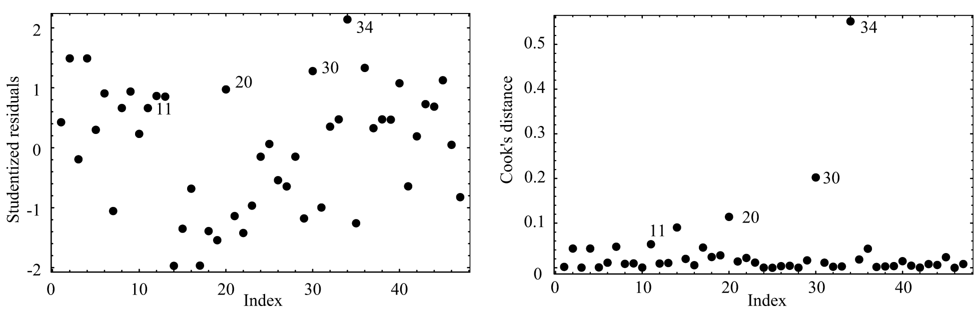

Figure 2. Index plots of the studentized and Cook's distance for the stars data.

other points followed by observation 30. Thus, the existing sensitivity methods based on least squares fails to detect the influence of the four stars on the least squares result. This is due to the well-known problem of masking (the least squares estimates are not-robust). We shall see in the following sections of the paper that the proposed sensitivity measures succeed in detecting all four points.

4. Least-squares regression

4.1 *The least-squares regression problem*

The least squares method leads to the following optimization problem

$$\text{Minimize}_{\boldsymbol{\beta}} Z_{\text{LS}} = \sum_{i=1}^{n} (y_i - \mathbf{X}_i^T \boldsymbol{\beta})^2. \tag{18}$$

Since the optimization problem in equation (18) does not have any constraints, the sensitivities can be calculated using equations (8)–(11), which in this case becomes

$$F_{\boldsymbol{\beta}(1\times k)} = (\nabla_\beta f(\bar{\beta}, z))\text{T} = -2\mathbf{e}^{\text{T}}\mathbf{X} = 0, \tag{19}$$

$$F_{\mathbf{z}_{(1\times(k+1))}} = (\nabla_z f(\bar{\beta}, z))^{\text{T}} = 2(\mathbf{e}^{\text{T}} | -\mathbf{e}^{\text{T}} \otimes \beta^{\text{T}}), \tag{20}$$

$$F_{\boldsymbol{\beta\beta}_{(k\times k)}} = \nabla_{\beta\beta} f(\bar{\beta}, z) = 2\mathbf{X}^{\text{T}}\mathbf{X}, \tag{21}$$

$$F_{\boldsymbol{\beta}\mathbf{z}_{(k\times n(k+1))}} = \nabla_{\beta z} f(\bar{\beta}, z) = 2(-\mathbf{X}^{\text{T}} | \beta^{\text{T}} \otimes \mathbf{X}^{\text{T}} - \mathbf{I}_k \otimes \mathbf{e}^{\text{T}}), \tag{22}$$

where $\mathbf{e}$ is the vector of errors, and it has been taken into account that for the LS method $\mathbf{e}^{\text{T}}\mathbf{X} = 0$, in equations (12) and (13), leading to

$$\left.\frac{\partial \boldsymbol{\beta}}{\partial \mathbf{y}}\right._{k\times n} = (\mathbf{X}^{\text{T}}\mathbf{X})^{-1}\mathbf{X}^{\text{T}}, \tag{23}$$

$$\left.\frac{\partial \boldsymbol{\beta}}{\partial \mathbf{x}}\right._{k\times nk} = -(\mathbf{X}^{\text{T}}\mathbf{X})^{-1}(\boldsymbol{\beta}^{\text{T}} \otimes \mathbf{X}^{\text{T}} - \mathbf{I}_k \otimes \mathbf{e}^{\text{T}}), \tag{24}$$

$$\left.\frac{\partial Z_{\text{LS}}^*}{\partial \mathbf{y}}\right._{1\times n} = 2\mathbf{e}^{\text{T}}, \tag{25}$$

$$\left.\frac{\partial Z_{\text{LS}}^*}{\partial \mathbf{x}}\right._{1\times nk} = -2(\mathbf{e}^{\text{T}} \otimes \boldsymbol{\beta}^{\text{T}}). \tag{26}$$

Note that the matrix with the sensitivities of the $\boldsymbol{\beta}$ with respect to $\boldsymbol{y}$ is known in the outlier detection literature as the *catcher* matrix.

From equations (23) to (26) one immediately obtains the following formulas for the sensitivities:

$$\frac{\partial \beta_j}{\partial y_i} = \sum_{r=1}^{k} c_{jr} x_{ir}, \tag{27}$$

$$\frac{\partial \beta_j}{\partial x_{st}} = -\sum_{r=1}^{k} c_{jr} [\beta_t x_{\text{sr}} - \delta_{\text{tr}} e_s], \tag{28}$$

$$\frac{\partial Z^*_{\mathrm{LS}}}{\partial y_i} = 2e_i, \tag{29}$$

$$\frac{\partial Z^*_{\mathrm{LS}}}{\partial x_{st}} = -2e_s\beta_t, \tag{30}$$

where c_{ij} are the elements of matrix $(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$, δ_{tr} is the Kronecker delta function, and e_i is the ith residual. To be meaningfully interpreted and comparable, we use the standardized versions of the above sensitivities by subtracting their means and dividing by their standard deviations. For example, the standardized version of equations (29) becomes

$$S_{\mathrm{LS}}(y_i) = \frac{e_i}{\hat{\sigma}\sqrt{1-p_{ii}}}, \tag{31}$$

where p_{ii} is the ith leverage value (the ith diagonal element of $P = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}$) and

$$\hat{\sigma}^2 = \frac{\mathbf{e}^{\mathrm{T}}\mathbf{e}}{n-k}. \tag{32}$$

Similarly, the standardized version of equation (30) becomes

$$S_{\mathrm{LS}}(x_{ij}) = \frac{e_i\hat{\beta}_j}{\hat{\sigma}\sqrt{(1-p_{ii})[\hat{\sigma}^2 c_{jj} + \hat{\beta}_j^2]}}, \quad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, k, \tag{33}$$

where c_{jj} is the jth diagonal element of $(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$.

It is interesting to note that these standardized sensitivities possess the following properties:

1. Apart from degenerate cases, all the standardized sensitivities with respect to all data points are different from zero.
2. The standardized sensitivities $\partial Z^*_{\mathrm{LS}}/\partial y_i$ and $\partial Z^*_{\mathrm{LS}}/\partial x_{ij}$ coincide in absolute value, but the sign of the second depends on the sign of the corresponding β_j.
3. The standardized sensitivities $\partial\beta_s/\partial y_i$ and $\partial\beta_r/\partial y_i$ coincide in absolute value, but their signs depend on the point positions.
4. The standardized sensitivities $\partial\beta_s/\partial x_{ij}$ and $\partial\beta_r/\partial x_{ij}$ coincide in absolute value, but their signs depend on the point positions.

4.2 *A numerical example*

Let us now compute the sensitivities for the stars data. Using equations (27)–(30) the objective function and the parameters sensitivities with respect to the data $\boldsymbol{z} = (\boldsymbol{y}, \boldsymbol{X})$ have been obtained. Table 2 shows the standardized sensitivities of Z_{LS}, β_0, β_1 with respect to the data. It can be observed that the properties of these sensitivities mentioned above hold in this table.

Figure 3 shows the scatter plot of the star cluster data, where the points are sorted by their objective function sensitivities (upper graph) and by the slope or intercept sensitivities (lower graph). Thus, the higher the number next to a point the more sensitive the results with respect to changes in the data point.

The sensitivity contours in these plots have been obtained as follows. A new data point (x_{n+1}, y_{n+1}) has been assumed to enter the sample and then the sensitivities associated with this point have been re-calculated as a function of its coordinates. In this way, these contours permit in determining the sensitivity of a new point entering the sample, or, approximately,

Table 2. Standardized sensitivities of Z, β_0, β_1 with respect to data.

	Z_{LS}			β_0			β_1		
Index	Y	X	(X, Y)	Y	X	(X, Y)	Y	X	(X, Y)
1	0.430	0.430	0.608	−0.209	−0.466	0.511	0.209	0.466	0.511
2	1.495	1.495	2.114	−0.869	−1.653	1.868	0.869	1.653	1.868
3	−0.195	−0.195	0.276	0.174	0.229	0.288	−0.174	−0.229	0.288
4	1.495	1.495	2.114	−0.869	−1.653	1.868	0.869	1.653	1.868
5	0.302	0.302	0.427	0.035	−0.286	0.288	−0.035	0.286	0.288
6	0.914	0.914	1.292	−0.521	−1.008	1.135	0.521	1.008	1.135
7	−1.036	−1.036	1.465	1.634	1.381	2.139	−1.634	−1.381	2.139
8	0.664	0.664	0.939	−0.904	−0.852	1.242	0.904	0.852	1.242
9	0.947	0.947	1.340	0.174	−0.883	0.900	−0.174	0.883	0.900
10	0.234	0.234	0.331	−0.209	−0.275	0.345	0.209	0.275	0.345
11	0.607	0.607	0.859	2.850	0.058	2.851	−2.850	−0.058	2.851
12	0.871	0.871	1.232	−0.417	−0.944	1.032	0.417	0.944	1.032
13	0.859	0.859	1.214	−0.591	−0.971	1.136	0.591	0.971	1.136
14	−1.969	−1.969	2.784	1.043	2.154	2.393	−1.043	−2.154	2.393
15	−1.366	−1.366	1.932	0.070	1.346	1.348	−0.070	−1.346	1.348
16	−0.689	−0.689	0.975	−0.382	0.584	0.698	0.382	−0.584	0.698
17	−1.986	−1.986	2.809	0.278	1.997	2.017	−0.278	−1.997	2.017
18	−1.403	−1.403	1.985	−0.382	1.279	1.335	0.382	−1.279	1.335
19	−1.558	−1.558	2.203	0.278	1.580	1.604	−0.278	−1.580	1.604
20	0.893	0.893	1.262	2.850	−0.220	2.859	−2.850	0.220	2.859
21	−1.152	−1.152	1.629	0.070	1.138	1.140	−0.070	−1.138	1.140
22	−1.438	−1.438	2.033	0.070	1.416	1.417	−0.070	−1.416	1.417
23	−0.975	−0.975	1.379	−0.382	0.862	0.943	0.382	−0.862	0.943
24	−0.151	−0.151	0.213	−0.626	0.004	0.626	0.626	−0.004	0.626
25	0.063	0.063	0.090	−0.243	−0.117	0.270	0.243	0.117	0.270
26	−0.547	−0.547	0.773	−0.382	0.445	0.587	0.382	−0.445	0.587
27	−0.652	−0.652	0.923	0.070	0.651	0.655	−0.070	−0.651	0.655
28	−0.151	−0.151	0.213	−0.243	0.091	0.260	0.243	−0.091	0.260
29	−1.191	−1.191	1.685	0.313	1.231	1.270	−0.313	−1.231	1.270
30	1.170	1.170	1.655	2.885	−0.482	2.925	−2.885	0.482	2.925
31	−1.008	−1.008	1.425	−0.243	0.926	0.957	0.243	−0.926	0.957
32	0.352	0.352	0.498	−0.869	−0.541	1.024	0.869	0.541	1.024
33	0.477	0.477	0.675	−0.487	−0.575	0.754	0.487	0.575	0.754
34	1.964	1.964	2.777	2.850	−1.263	3.117	−2.850	1.263	3.117
35	−1.272	−1.272	1.799	0.278	1.302	1.331	−0.278	−1.302	1.331
36	1.329	1.329	1.880	−1.077	−1.540	1.879	1.077	1.540	1.879
37	0.328	0.328	0.464	−0.765	−0.494	0.910	0.765	0.494	0.910
38	0.477	0.477	0.675	−0.487	−0.575	0.754	0.487	0.575	0.754
39	0.471	0.471	0.666	−0.765	−0.633	0.992	0.765	0.633	0.992
40	1.086	1.086	1.535	−0.417	−1.152	1.225	0.417	1.152	1.225
41	−0.651	−0.651	0.920	−0.243	0.578	0.627	0.243	−0.578	0.627
42	0.192	0.192	0.271	−0.487	−0.297	0.570	0.487	0.297	0.570
43	0.732	0.732	1.035	−0.660	−0.863	1.087	0.660	0.863	1.087
44	0.691	0.691	0.978	−0.487	−0.784	0.923	0.487	0.784	0.923
45	1.130	1.130	1.598	−0.834	−1.290	1.536	0.834	1.290	1.536
46	0.049	0.049	0.069	−0.487	−0.158	0.512	0.487	0.158	0.512
47	−0.832	−0.832	1.177	−0.382	0.723	0.818	0.382	−0.723	0.818

determining the sensitivity of any point already existing in the sample. Notice that the closer the point to the regression line the lower the objective function sensitivity. Note also that the closer the points to the center of gravity the smaller the sensitivity with respect to the beta parameters.

The most interesting revelation of sensitivity analysis can be seen in figure 3. The upper graph shows that only one of the four giant stars exert undue sensitivity on the objective function estimates. However, the lower graph shows that the four giant stars are the ones with the greatest sensitivities on the parameters.

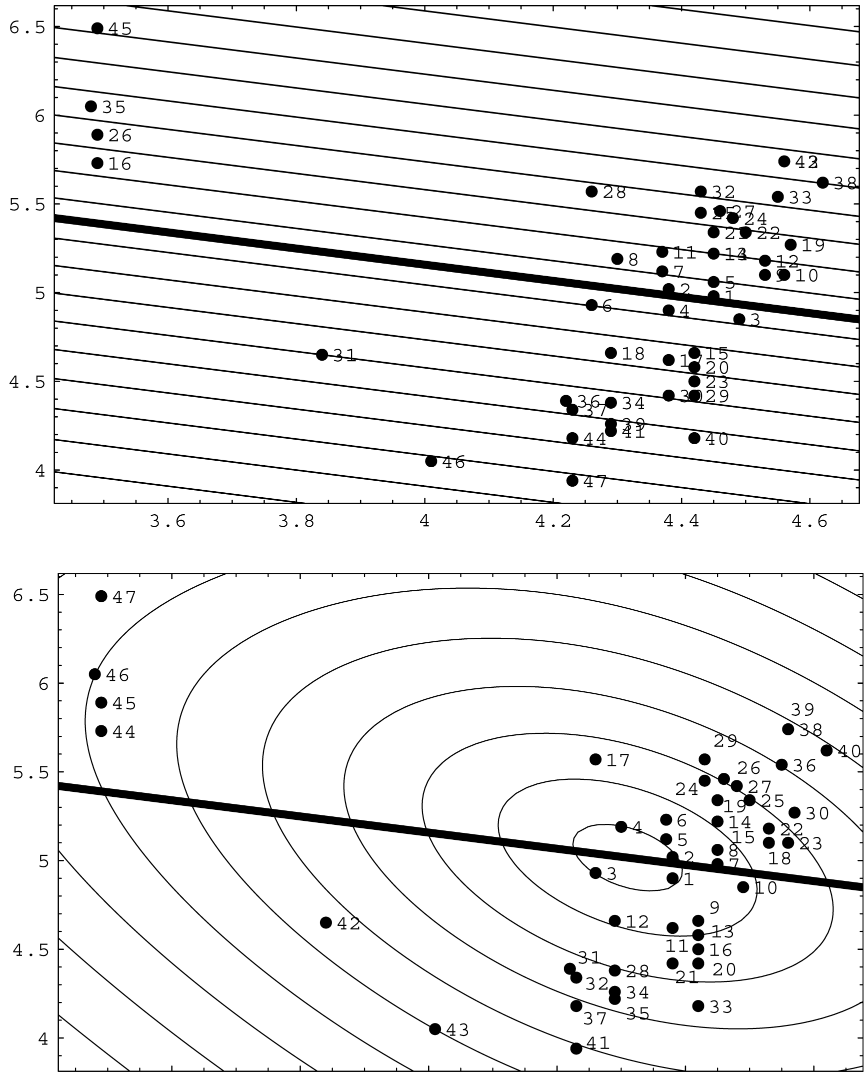

Figure 3. Scatter plot of the star cluster data with the sensitivity contours. The number next to a point refers to the rank of the point according to its sensitivity with respect to the objective function (upper graph) and the slope parameter (lower graph).

We have seen in section 3 that existing diagnostic measures (*e.g.*, the studentized residuals and Cook's distance) based on least squares fail to detect the influence of the four giant stars on the least squares regression line, which is due to the masking problem. On the contrary, the four stars are clearly separated from the rest of the points in the index plot of the Y- and XY-sensitivities. Thus, revealing the superiority of the proposed method with respect to the existing ones. One reason for the success of the proposed sensitivity measures is that they measure local sensitivities and not global sensitivities, like the existing diagnostic measures.

We should note here that the proposed sensitivity measures have been tested with other data sets and have shown similar performance, but the results are not reported here because of lack of space (figure 4).

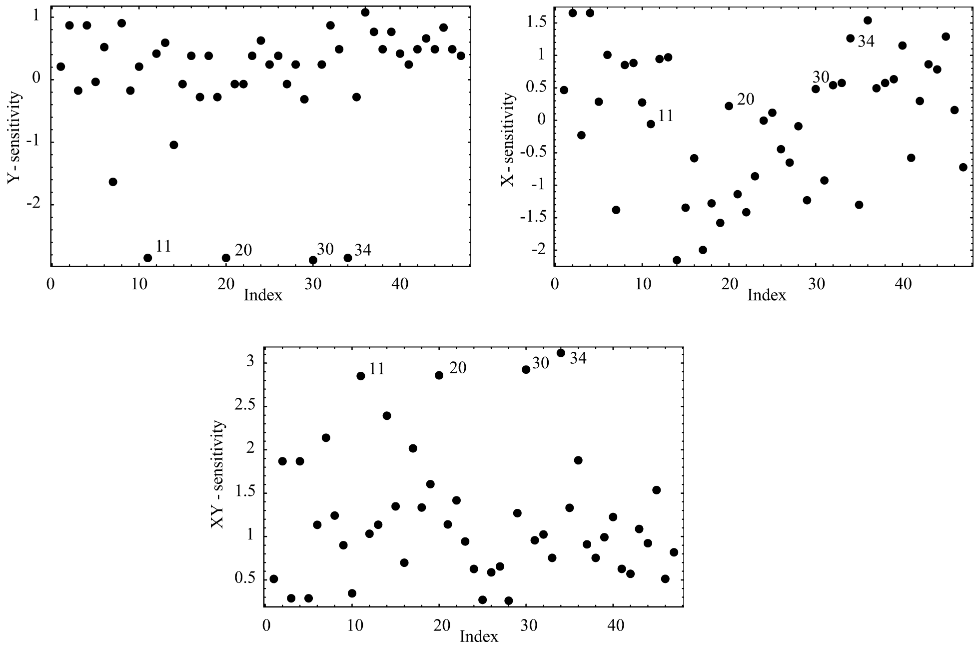

Figure 4. Index plots of the sensitivities in the last three columns in table 2.

5. Minimax regression

5.1 *The primal minimax regression problem*

The minimax method estimates the regression coefficient by minimizing the maximum error, that is,

$$\underset{\boldsymbol{\beta}}{\text{Minimize}}\;\; Z_{\text{MM}} = \max_i |y_i - \boldsymbol{x}_i^{\text{T}}\boldsymbol{\beta}|, \tag{34}$$

which is equivalent to the linear programming problem

$$\underset{\boldsymbol{\beta},\varepsilon}{\text{Minimize}}\;\; Z_{\text{MM}} = Z_{\text{MM}} = \varepsilon \tag{35}$$

subject to

$$y_i - \boldsymbol{x}_i^{\text{T}}\boldsymbol{\beta} \le \varepsilon;\, \mu_i^{(1)}, \quad i = 1, \ldots, n, \tag{36}$$

$$\boldsymbol{x}_i^{\text{T}}\beta - \dot{y}_i \le \varepsilon;\, \mu_i^{(2)}, \quad i = 1, \ldots, n, \tag{37}$$

where $\mu_i^{(1)}$ and $\mu_i^{(2)}$ are the dual variables.

We note that the constraint $\varepsilon \ge 0$, used by practically all authors, is not required because it is implied by equations (36) and (37).

To obtain the sensitivities of the $\boldsymbol{\beta}$ estimates with respect to that data, it is convenient to assume that we are not in a degenerate case, *i.e.*, we assume that a total of exactly k constraints in equations (36) and (37) are active (degenerated cases can also be dealt with in similar methods). It is also convenient to reduce the analysis to the sensitivities that are known to be non-null. Then, following the steps in section 2, we have

$$y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta} - \varepsilon = 0; \quad \mu_i^+, \; i \in I^+, \tag{38}$$

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta} - y_i - \varepsilon = 0; \quad \mu_i^-, \; i \in I^-, \tag{39}$$

i.e., the sets I^+ and I^-, with cardinals p^+ and p^-, respectively, give the data points that correspond to the active constraints. Note that $\boldsymbol{\mu}^+$ and $\boldsymbol{\mu}^-$ are the column vectors of the dual variables associated with the sets I^+ and I^-, respectively. Apart from degenerate cases, we have $p^+ + p^- = k + 1$.

Then, the problem (35)–(37) can be written as

$$\text{Minimize}_{\,\beta,\varepsilon} Z_{\mathrm{MM}} = \varepsilon \tag{40}$$

subject to

$$\mathbf{Q}\begin{pmatrix}\boldsymbol{\beta}\\ --\\ \varepsilon\end{pmatrix} = \begin{pmatrix}\mathbf{X}^+ & | & 1\\ -- & + & --\\ \mathbf{X}^- & | & 1\end{pmatrix}\begin{pmatrix}\boldsymbol{\beta}\\ --\\ \varepsilon\end{pmatrix} = \begin{pmatrix}\mathbf{Y}^+\\ --\\ \mathbf{Y}^-\end{pmatrix}, \tag{41}$$

where the meaning of matrix $\mathbf{Q}$ becomes obvious from the first equation in equation (41), and $\mathbf{X}^+, \mathbf{X}^-, \mathbf{Y}^+$, and $\mathbf{Y}^-$ refer to the $\boldsymbol{X}$ and $\boldsymbol{Y}$ matrices associated with I^+ and I^-, respectively.

Since the problem (40)–(41) is a linear programming problem, we can directly apply the formulas in equation (16) to obtain the following sensitivities:

$$\frac{\partial \beta_j}{\partial y_i} = q^{ji}, \tag{42}$$

$$\frac{\partial \beta_j}{\partial x_{st}} = -q^{js}\beta_t, \tag{43}$$

$$\frac{\partial \mu_j}{\partial y_i} = 0, \tag{44}$$

$$\frac{\partial \mu_j}{\partial x_{st}} = -q^{tj}\mu_s, \tag{45}$$

$$\frac{\partial Z^*_{\mathrm{MM}}}{\partial y_i} = -\mu_i, \tag{46}$$

$$\frac{\partial Z^*_{\mathrm{MM}}}{\partial x_{st}} = \mu_s\beta_t, \tag{47}$$

where q_{ij} are the elements of $\mathbf{Q}^{-1}$, the indices refer to the positions of the data sets in the set $I^+ \cup I^-$, and the sensitivities refer only to the data in I^+ and I^-, because the sensitivities with respect to other data items are null.

Note also that the sensitivities are proportional to the corresponding regression coefficient β_j and to the dual variable $\mu_i(s)$ value.

The standardized sensitivities of the MM objective function with respect to the response variable in equation (46) values are,

$$S_{\mathrm{MM}}(y_i) = \frac{(\partial Z^*_{\mathrm{MM}}/\partial y_i) - m}{s}, \quad i = 1, 2, \ldots, n, \tag{48}$$

where m and s are the mean and standard deviation of $\partial Z^*_{\mathrm{MM}}/\partial y_i$, $i = 1, 2, \ldots, n$.

Similarly, the standardized sensitivities with respect to the predictor variables in equation (47) are

$$S_{\mathrm{MM}}(x_{ij}) = \frac{(\partial Z^{*}_{\mathrm{MM}}/\partial x_{ij}) - m_j}{s_j}, \quad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, k, \tag{49}$$

where m_j and s_j are the mean and standard deviation of the sensitivities in equation (47), after replacing β_j by its MM estimate.

5.2 *A numerical example*

Now fitting a regression line to the star cluster data using the minimax method and solving the optimization problem (35)–(37), one gets the line

$$y = 7.89850 - 0.70093x,$$

with an optimal value $\varepsilon^* = 1.03776$. These estimates are associated with the points in the sets $I^+ = \{2, 34\}$ and $I^- = \{14\}$, *i.e.*, only data points 2, 14, and 34 have active constraints. The optimal values of the dual variables are

$$\boldsymbol{\mu}^+ = (-0.243, -0.257)^T; \quad \boldsymbol{\mu}^- = \{-0.50\}.$$

Figure 5 shows the data points, the minimax regression line, and the corresponding parallel bands at a vertical distance $\varepsilon = 1.03776$ up and down from the regression line. Note that the data points 2, 14, and 34 are at the bands.

Since point 2 and 4 are coincident, we have a degenerate case. However, we can eliminate the degeneration problem by removing point 4, because it has no influence on the final solution. In addition, the left directional derivatives of the optimal solution with respect to these two points are null, because the other points lead to the optimal solution. Then, there are no partial derivatives with respect to these two points.

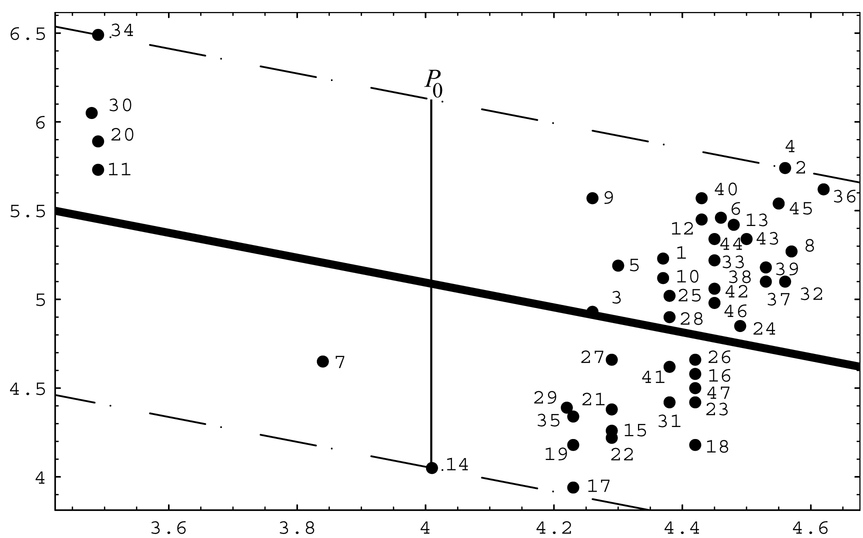

Figure 5. Minimax regression. The number next to a point refers to the rank of the point according its sensitivity with respect the objective function.

Using formulas (42) to (47), the following sensitivities are obtained:

$$\frac{\partial \boldsymbol{\beta}}{\partial z} = \begin{pmatrix} -3.505 & 4.005 & 0.5 & 27.682 & -2.457 & -31.631 & 2.807 & -3.949 & 0.35 \\ 0.935 & -0.935 & 0 & -7.382 & 0.655 & 7.382 & -0.655 & 0 & 0 \\ 0.243 & 0.257 & -0.5 & -1.919 & 0.17 & -2.03 & 0.18 & 3.949 & -0.35 \end{pmatrix},$$

$$\frac{\partial \boldsymbol{\mu}}{\partial z} = \begin{pmatrix} 0 & 0 & 0 & -0.852 & 0.227 & -0.901 & 0.24 & 1.752 & -0.467 \\ 0 & 0 & 0 & 0.973 & -0.227 & 1.029 & -0.24 & -2.002 & 0.467 \\ 0 & 0 & 0 & -0.121 & 0 & -0.128 & 0 & 0.25 & 0, \end{pmatrix}$$

$$\frac{\partial Z_{\mathrm{MM}}}{\partial z} = \begin{pmatrix} 0.243 & 0.257 & -0.500 & -1.919 & 0.170 & -2.030 & 0.180 & 3.949 & -0.350 \end{pmatrix}.$$

where we have denoted

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \varepsilon), \tag{50}$$

$$z = (y_2, y_{34}, y_{14}, x_{1,2}, x_{2,2}, x_{1,34}, x_{2,34}, x_{1,14}, x_{2,14}), \tag{51}$$

Table 3 shows the sensitivities of Z^*_{MM}, β_0, and β_1 with respect to the data for the minimax regression method. They have been extracted from the above matrices.

The results in table 3 lead to the following conclusions that are not particular to the data in this example, but of general validity:

1. One of the hyperplane bands (upper or lower, depending on the case) passes through k data points ($k = 2$, here, and the data points are 2 and 34 as can be seen in figure 5). They are all points associated with one of the active constraints (36) or (37). We call these points the band points, because they define the corresponding hyperplane band.
2. There exist one point (point 14 in figure 5) associated with the only active constraint in the other set of the pair (36) or (37). We call this point the ε point, because it gives the optimal value of ε, *i.e.*, the vertical distance from it to the hyperplane defining the band above (see figure 5).
3. With the exception of degenerate cases, no more active constraints exist. This implies a total of exactly $k + 1$ active constraints. This is due to the fact that in linear programming optimal solutions coincide with basic solutions, that are defined by $k + 1$ constraints if the space of the unknowns (the regression coefficients β and ε) is of dimension $k + 1$.
4. The sensitivities of the estimated regression parameters with respect to the ε-data point (in our example $\partial\beta_1/\partial y_{14}$ and $\partial\beta_1/\partial x_{14}$) are null because a small change in the ε-data point does not alter the estimated regression hyperplane or the bands, which is defined only by the band points.
5. The sensitivities of the objective function ε with respect to the y coordinate of the ε point has always absolute value $1/2$, because the regression line does not change, when changing only the ε point ordinate and is half way from it to the band. The sign of this sensitivity

Table 3. Sensitivities of Z^*, β_0, and β_1 with respect to data for the minimax regression method.

	Z^*_{MM}		β_0		β_1	
Index	Y	X	Y	X	Y	X
2	0.243	0.170	−3.505	−2.457	0.935	0.655
14	−0.500	−0.350	0.500	0.350	0	0
34	0.257	0.180	4.005	2.807	−0.935	−0.655

is positive or negative, depending on whether the ε point is below or above the regression hyperplane, respectively.

6. The absolute values of the estimated regression parameters sensitivities with respect to the band points ($\partial\beta_1/\partial y_2$, $\partial\beta_1/\partial y_{34}$, $\partial\beta_1/\partial x_2$, and $\partial\beta_1/\partial x_{34}$ in our example) are identical but with opposite signs, because the changes in the slope depend on those points.

5.3 *The dual minimax regression problem*

Klingman and Mote [28] give an interpretation of the dual of the minimax problem as a capacitated generalized network problem. In this section, we give an interpretation in terms of probabilities.

The problem (35)–(37) can be written in matrix form as

$$\underset{\beta,\varepsilon}{\text{Minimize}}\ \ Z_{\text{MM}} = \varepsilon, \tag{52}$$

subject to

$$\begin{pmatrix} -\mathbf{X}_{n\times k} & | & -\mathbf{1}_{n\times 1} \\ --- & + & --- \\ \mathbf{X}_{n\times k} & | & -\mathbf{1}_{n\times 1} \end{pmatrix} \begin{pmatrix} \beta \\ \varepsilon \end{pmatrix} \le \begin{pmatrix} -\boldsymbol{y}_{n\times 1} \\ --- \\ \boldsymbol{y}_{n\times 1} \end{pmatrix} : \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ -- \\ \boldsymbol{\mu}^{(2)} \end{pmatrix} \tag{53}$$

The corresponding dual problem is

$$\underset{\boldsymbol{\mu}^{(1)},\boldsymbol{\mu}^{(2)}}{\text{Maximize}}\ \sum_{i=1}^{n} y_i\left(\mu_i^{(2)} - \mu_i^{(1)}\right) \tag{54}$$

subject to

$$\begin{pmatrix} -\mathbf{X}^{\mathrm{T}}_{k\times n} & | & \mathbf{X}^{\mathrm{T}}_{k\times n} \\ --- & + & --- \\ -\mathbf{1}_{1\times n} & | & -\mathbf{1}_{1\times n} \end{pmatrix} \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ -- \\ \boldsymbol{\mu}^{(2)} \end{pmatrix} = \begin{pmatrix} \mathbf{0}_{k\times 1} \\ -- \\ 1 \end{pmatrix}, \tag{55}$$

$$\boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)} \le \mathbf{0}, \tag{56}$$

where $\boldsymbol{\mu}^{(1)}$ and $\boldsymbol{\mu}^{(2)}$ are the dual variables.

Letting $\lambda_i^{(j)} = -2\mu_i^{(j)}$; $j = 1, 2$, this dual problem can be written as

$$\underset{\boldsymbol{\lambda}^{(1)},\boldsymbol{\lambda}^{(2)}}{\text{Maximize}} \sum_{i=1}^{n} y_i\lambda_i^{(2)} - \sum_{i=1}^{n} y_i\lambda_i^{(1)} \tag{57}$$

subject to

$$-\sum_{i=1}^{n}\lambda_i^{(1)}x_{ij} + \sum_{i=1}^{n}\lambda_i^{(2)}x_{ij} = 0, \quad j = 2, 3, \ldots, k, \tag{58}$$

$$\sum_{i=1}^{n}\lambda_i^{(1)} = \sum_{i=1}^{n}\lambda_i^{(2)}, \tag{59}$$

$$\sum_{i=1}^{n}\lambda_i^{(1)} + \sum_{i=1}^{n}\lambda_i^{(2)} = 2, \tag{60}$$

$$\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(2)} \ge \mathbf{0}. \tag{61}$$

Because of equation (59) and (61) we can divide equation (57), (58), and (60) by $\sum_{i=1}^{n}\lambda_i^{(1)}$ and letting $\rho_2^{(r)} = \lambda_s^{(r)}/\sum_{i=1}^{n}\lambda_i^{(r)}$; $r = 1, 2$, then equations (57)–(61) become

$$\underset{\boldsymbol{\rho}^{(1)},\boldsymbol{\rho}^{(2)}}{\text{Minimize}} \sum_{i=1}^{n} y_i\rho_i^{(2)} - \sum_{i=1}^{n} y_i\rho_i^{(1)} \quad (62)$$

subject to

$$\sum_{i=1}^{n}\rho_i^{(1)}x_{ij} = \sum_{i=1}^{n}\rho_i^{(2)}x_{ij}, \quad j = 2, 3, \ldots, k, \quad (63)$$

$$\sum_{i=1}^{n}\rho_i^{(1)} = \sum_{i=1}^{n}\rho_i^{(2)} = 1, \quad (64)$$

$$\boldsymbol{\rho}^{(1)}, \boldsymbol{\rho}^{(2)} \geq \mathbf{0}, \quad (65)$$

showing that the dual variables $\boldsymbol{\rho}^{(1)}$ and $\boldsymbol{\rho}^{(2)}$ can be interpreted as two probability mass functions on the set

$$\mathcal{S} = \{(y_i, \mathbf{x}_i)|i = 1, 2, \ldots, n\}.$$

Hence, the objective function in equation (62) can be interpreted as the difference of marginal means $E(2)[Y] - E^{(1)}[Y]$. Similarly, the constraints in equation (63) can be interpreted as the equality of marginal means, $E^{(1)}[X_j] = E^{(2)}[X_j]j = 2, 3, \ldots, k$. Accordingly, one can think of the dual as a problem of finding two probability mass functions on the set S such that they minimize the difference of marginal means $E^{(2)}[Y] - E^{(1)}[Y]$ subject to the equality of expectations $E^{(1)}[X_j] = E^{(2)}[X_j]j = 2, 3, \ldots, k$.

The above interesting interpretations of the dual problem can be illustrated using figure 5, as an example. In this case, the probability mass function $\rho^{(1)}$ assigns probability 0.486 to point 2, and probability 0.514 to point 34, and the probability mass functions $\rho^{(2)}$ assigns probability 1 to point 14. Other points are assigned probability zero by both probability measures.

These two probabilities assigned to points 2 and 14 are inversely proportional to the distances of these points to the point P_0 in the same band whose abscissa coincides with x_{14}. Minimizing the difference of marginal expectations $E^{(2)}[Y] - E^{(1)}[Y]$ means minimizing the vertical distance between point P_0 and point 14. Note that, in fact, the set of sample points is partitioned into two sets: those above and those below the regression line. The supports of these two probabilities are inside these two sets.

6. The least-absolute-value (LAV) regression

6.1 *The primal LAV regression problem*

In the LAV regression problem (see, for example, [29], we minimize the sum of the distances between observed and predicted values instead of their squares, *i.e.*:

$$\underset{\boldsymbol{\beta}}{\text{Minimize}}\ Z_{\text{LAV}} = \sum_{i=1}^{n} |y_i - \mathbf{x}_i^T\boldsymbol{\beta}|. \quad (66)$$

This method treats all errors equally. Thus, this method must be used when the user is concerned about any level of error. In fact, what is important is the sum of all absolute errors, not a single error.

Due to the presence of the absolute-value function, it is difficult to solve equation (66) using standard regression techniques. The LAV problem in equation (66) is equivalent to the following problem

$$\underset{\boldsymbol{\beta},\varepsilon_i}{\text{Minimize}}\; Z_{\text{LAV}} = \sum_{i=1}^{n} \varepsilon_i \tag{67}$$

subject to

$$y_i - \boldsymbol{x}_i^T \boldsymbol{\beta} \le \varepsilon_i, \quad i = 1, \ldots, n, \tag{68}$$

$$\boldsymbol{x}_i^T \boldsymbol{\beta} - y_i \le \varepsilon_i, \quad i = 1, \ldots, n. \tag{69}$$

We note that the set of constraints $\varepsilon_i \ge 0;\, i = 1, \ldots, n$, used by practically all authors, is not required because it is implied by equations (68) and (69).

To obtain the sensitivities of the β estimates with respect to data it is convenient to assume that we are not in a degenerate case, *i.e.*, a total of exactly n constraints in equations (68) and (69) are active, and for exactly k points both are active. Let I^+ and I^- the sets of indices associated with the active constraints in equations (68) and (69), respectively, and $L = I^+ \cup I^-$ and $I^0 = I^+ \cap I^-$, where we keep the order of the elements in I^+ and I^-.

It is also convenient to reduce the analysis to the sensitivities that are known to be non-null. Then, following the steps in section 2, we have

$$y_i - \boldsymbol{x}_i^T \beta - \varepsilon_i = 0; \quad \mu_i^+, \;\; i \in I^+ \tag{70}$$

$$\boldsymbol{x}_i^T \beta - y_i - \varepsilon_i = 0; \quad \mu_i^-, \;\; i \in I^-, \tag{71}$$

$$\mathbf{x}_i^T \beta - y_i = 0; \quad \mu_i^0, \;\; i \in I^0, \tag{72}$$

where the sets I^+, I^- and I^0, have cardinals p^+, p^-, and p^0, respectively, where for non-degenerate cases $p^+ + p^- + p^0 = n + k$. Note that $\boldsymbol{\mu}^+$, $\boldsymbol{\mu}^-$, and $\boldsymbol{\mu}^0$ are the column vectors of dimension n with the dual variables associated with the sets I^+, I^-, and I^0, respectively, and null values, otherwise.

Then, the problem (67)–(69) can be written as

$$\underset{\beta,\varepsilon_i}{\text{Minimize}}\;\; Z_{\text{LAV}} = \sum_{i=1}^{n} \varepsilon_i \tag{73}$$

subject to

$$\mathbf{Q} = \begin{pmatrix} \beta \\ -- \\ \varepsilon^+ \\ -- \\ \varepsilon^- \end{pmatrix} = \begin{pmatrix} \mathbf{X}^+ & | & \mathbf{I} & | & \mathbf{0} \\ -- & + & -- & + & -- \\ \mathbf{X}^- & | & \mathbf{0} & | & \mathbf{I} \\ -- & + & -- & + & -- \\ \mathbf{X}^0 & | & \mathbf{0} & | & \mathbf{0} \end{pmatrix} \begin{pmatrix} \beta \\ -- \\ \varepsilon^+ \\ -- \\ \varepsilon^- \end{pmatrix} \begin{pmatrix} \mathbf{Y}^+ \\ -- \\ \mathbf{Y}^- \\ -- \\ \mathbf{Y}^0 \end{pmatrix}, \tag{74}$$

where the meaning of matrix $\mathbf{Q}$ becomes obvious from the first equation in equation (74), and $\mathbf{X}^+, \mathbf{X}^-, \mathbf{X}^0, \mathbf{Y}^+, \mathbf{Y}^-$ and Y^0 refer to the $\mathbf{X}$ and $\mathbf{Y}$ matrices associated with I^+, I^- and I^0, respectively.

Since the problem (73)–(74) is a linear programming problem, we can directly apply the formulas in equation (16) to obtain the following sensitivities:

$$\frac{\partial \beta_j}{\partial y_i} = q^{ji}, \tag{75}$$

$$\frac{\partial \beta_j}{\partial x_{st}} = -q^{js}\beta_t, \tag{76}$$

$$\frac{\partial \mu_j}{\partial y_i} = 0, \tag{77}$$

$$\frac{\partial \mu_j}{\partial x_{st}} = -q^{tj}\mu_s, \tag{78}$$

$$\frac{\partial Z^*_{\text{MM}}}{\partial y_i} = -\mu_i, \tag{79}$$

$$\frac{\partial Z^*_{\text{MM}}}{\partial x_{st}} = \mu_s\beta_t, \tag{80}$$

where q_{ij} are the elements of $\mathbf{Q}^{-1}$, the indices refer to the positions of the data sets in the set $I^+ \cup I^- \cup I^0$, and the sensitivities refer only to the data in I^+, I^-, and I^0, because the sensitivities with respect to other data items are null.

Note that the matrix $\mathbf{Q}$ can be inverted symbolically and gives

$$\mathbf{Q}^{-1} = \begin{pmatrix} \mathbf{X}^+ & | & \mathbf{I} & | & 0 \\ -- & + & -- & + & -- \\ \mathbf{X}^- & | & \mathbf{0} & | & \mathbf{I} \\ -- & + & -- & + & ----- \\ \mathbf{X}^0 & | & \mathbf{0} & | & \mathbf{0} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{0} & | & \mathbf{0} & | & (\mathbf{X}^0)^{-1} \\ -- & + & -- & + & ---- \\ \mathbf{I} & | & 0 & | & -\mathbf{X}^+(\mathbf{X}^0)^{-1} \\ -- & + & -- & + & ----- \\ \mathbf{0} & | & \mathbf{I} & | & -\mathbf{X}^-(\mathbf{X}^0)^{-1} \end{pmatrix}, \tag{81}$$

which facilitates the obtention of the above sensitivities.

Note also that the sensitivities are proportional to the corresponding regression coefficient β_j and to the dual variable $\mu_i^{(s)}$ value.

The mean and standard deviation of $\partial Z^*_{\text{LAV}}/\partial y_i$ are not known, so we use the mean, m, and standard deviation, s, of $\partial Z^*_{\text{LAV}}/\partial y_i$; $i = 1, 2, \ldots, n$, and obtain the standardized sensitivities of the LAV objective function with respect to the response values:

$$S_{\text{LAV}}(y_i) = \frac{(\partial Z^*_{\text{LAV}}/\partial y_i) - m}{s}, \quad i = 1, 2, \ldots, n. \tag{82}$$

Replacing β_j by its LAV estimate $\hat{\beta}_j$ and letting m_j and s_j be the mean and standard deviation of $\partial Z_{\text{LAV}}/\partial x_{ij}$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, k$, we obtain the standardized sensitivities of the LAV objective function with respect to x_{ij},

$$S_{\text{LAV}}(x_{ij}) = \frac{(\partial Z^*_{\text{LAV}}/\partial x_{ij}) - m_j}{s_j}, \quad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, k. \tag{83}$$

6.2 *A numerical example*

Consider again the data and the model in section 4, where now we use the LAV method. Solving the optimization problem (67)–(69) one gets the line

$$y = 8.1492 - 0.69318x.$$

Note that the LAV method is known to be robust with respect to outliers in the Y direction but not with respect to the outliers in the X space. Therefore, the LAV line has a negative slope. The LAV solution leads to an optimal value $Z_{\text{LAV}} = 22.1452$, and to the optimal values for the dual variables:

$$\boldsymbol{\mu}^+ = (1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)^{\text{T}}, \quad (84)$$

$$\boldsymbol{\mu}^- = (1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)^{\text{T}}, \quad (85)$$

$$\boldsymbol{\mu}^0 = (0.205, 0.795)^{\text{T}}, \quad (86)$$

Analysing the constraints, one gets the sets:

$$I^+ = \{1, 2, 4, 5, 6, 8, 9, 12, 13, 20, 30, 32, 33, 34, 36, 37, 38, 39, 40, 43, 44, 45\},$$

$$I^- = \{3, 7, 14, 15, 16, 17, 18, 19, 21, 22, 23, 24, 25, 26, 27, 28, 29, 31, 35, 41, 42, 46, 47\},$$

$$I^0 = \{10, 11\}.$$

Figure 6 shows the data points and the LAV regression line. Note that the data points 10 and 11 are on the LAV regression line. They are the points in I^0.

Table 3 shows the sensitivities of Z^*_{LAV}, β_0, and β_1 with respect to the data for the LAV regression method. They have been calculated using expressions (75)–(80) and the matrix in equation (81).

The results in table 4 and figure 6 lead to the following conclusions:

1. The LAV regression hyperplane passes through k data points (in our example $k = 2$ and the points are 10 and 11). We call these points regression hyperplane-points.

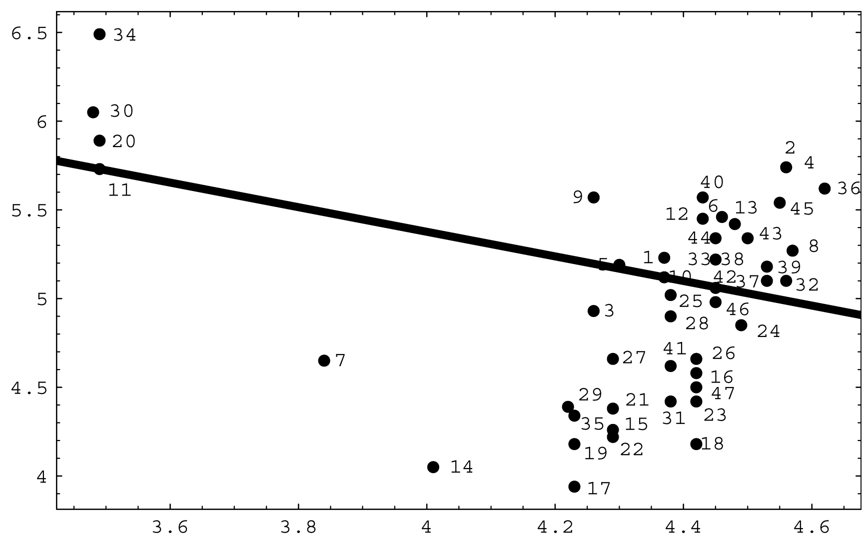

Figure 6. Data points and LAV regression line passing through points 3 and 17.

Table 4. Sensitivities of Z^*_{LAV}, β_0, β_1 with respect to data for the LAV regression method.

	Z_{LAV}		β_0		β_1	
Index	Y	X	Y	X	Y	X
1	1	0.693	0	0	0	0
2	1	0.693	0	0	0	0
3	−1	−0.693	0	0	0	0
4	1	0.693	0	0	0	0
5	1	0.693	0	0	0	0
6	1	0.693	0	0	0	0
7	−1	−0.693	0	0	0	0
8	1	0.693	0	0	0	0
9	1	0.693	0	0	0	0
10	0.205	0.142	−3.966	−2.749	1.136	0.788
11	0.795	0.551	4.966	3.442	−1.136	−0.788
12	1	0.693	0	0	0	0
13	1	0.693	0	0	0	0
14	−1	−0.693	0	0	0	0
15	−1	−0.693	0	0	0	0
16	−1	−0.693	0	0	0	0
17	−1	−0.693	0	0	0	0
18	−1	−0.693	0	0	0	0
19	−1	−0.693	0	0	0	0
20	1	0.693	0	0	0	0
21	−1	−0.693	0	0	0	0
22	−1	−0.693	0	0	0	0
23	−1	−0.693	0	0	0	0
24	−1	−0.693	0	0	0	0
25	−1	−0.693	0	0	0	0
26	−1	−0.693	0	0	0	0
27	−1	−0.693	0	0	0	0
28	−1	−0.693	0	0	0	0
29	−1	−0.693	0	0	0	0
30	1	0.693	0	0	0	0
31	−1	−0.693	0	0	0	0
32	1	0.693	0	0	0	0
33	1	0.693	0	0	0	0
34	1	0.693	0	0	0	0
35	−1	−0.693	0	0	0	0
36	1	0.693	0	0	0	0
37	1	0.693	0	0	0	0
38	1	0.693	0	0	0	0
39	1	0.693	0	0	0	0
40	1	0.693	0	0	0	0
41	−1	−0.693	0	0	0	0
42	−1	−0.693	0	0	0	0
43	1	0.693	0	0	0	0
44	1	0.693	0	0	0	0
45	1	0.693	0	0	0	0
46	−1	−0.693	0	0	0	0
47	−1	−0.693	0	0	0	0

2. Apart from degenerate cases, infinitesimal changes in the remaining points produce no change in the regression hyperplane, and so, the corresponding sensitivities are null.
3. With the exception of degenerate cases, a total of $k + n$ active constraints exist, which correspond to constraints (68) and (69). This is due to the fact that optimal solutions in linear programming coincide with basic solutions, that are defined by $k + n$ constraints, if the space of the unknowns (the k regression coefficients $\boldsymbol{\beta}$ and the n ε-variables) is of dimension $k + n$. This is the reason why in table 4 the sensitivities $\partial\beta_0/\partial y_i$, $\partial\beta_1/\partial y_i$, $\partial\beta_0/\partial x_i$, and $\partial\beta_1/\partial x_i$ for the data points not in the LAV regression hyperplane vanish.

4. Infinitesimal changes in the regression hyperplane points lead to changes in the regression hyperplane and thus in their slopes and intercept parameters. So, their corresponding sensitivities are not null.
5. The sensitivities of the objective function with respect to the regression hyperplane points are different from zero. The sensitivities with respect to y_i take value 1 or -1, depending on whether they are above or below the LAV regression hyperplane.
6. The sensitivities of the objective function with respect to x_i are different from zero and have positive or negative sign, depending on whether they are on the left or the right hand side of the LAV regression hyperplane. All of them have the same absolute value but the sign can be different, as indicated.

The above conclusions are valid in general, *i.e.*, they are not particular to this example.

6.3 *The dual LAV regression problem*

The primal problem in equations (67)–(69) can be written in matrix form as:

$$\text{Minimize}_{\boldsymbol{\beta},\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n} Z_{\text{LAV}} = \sum_{i=1}^{n} \varepsilon_i \tag{87}$$

subject to

$$\begin{pmatrix} -\mathbf{X}_{n\times k} & | & -\mathbf{I}_n \\ -- & + & -- \\ \mathbf{X}_{n\times k} & | & -\mathbf{I}_n \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\varepsilon} \end{pmatrix} \le \begin{pmatrix} -\mathbf{y}_{n\times 1} \\ --- \\ \mathbf{y}_{n\times 1} \end{pmatrix}; \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ --- \\ \boldsymbol{\mu}^{(2)} \end{pmatrix}. \tag{88}$$

The corresponding dual problem is

$$\text{Maximize}_{\boldsymbol{\mu}^{(1)},\boldsymbol{\mu}^{(2)}} \sum_{i=1}^{n} y_i \left(\mu_i^{(2)} - \mu_i^{(1)}\right) \tag{89}$$

subject to

$$\begin{pmatrix} -\mathbf{X}^{\mathrm{T}}_{k\times n} & | & -\mathbf{X}^{\mathrm{T}}_{k\times n} \\ -- & + & -- \\ -\mathbf{I}_n & | & -\mathbf{I}_n \end{pmatrix} \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ -- \\ \boldsymbol{\mu}^{(2)} \end{pmatrix} = \begin{pmatrix} \mathbf{0}_{k\times 1} \\ -- \\ \mathbf{1}_{n\times 1} \end{pmatrix}, \tag{90}$$

$$\boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)} \le \mathbf{0}, \tag{91}$$

where $\boldsymbol{\mu}^{(1)}$ and $\boldsymbol{\mu}^{(2)}$ are the dual variables.

Letting $\lambda_i^{(j)} = -\mu_i^{(j)}$; $j = 1, 2$, this dual problem can be written as

$$\text{Minimize}_{\boldsymbol{\lambda}^{(1)},\boldsymbol{\lambda}^{(2)}} \sum_{i=1}^{n} y_i \lambda_i^{(2)} - \sum_{i=1}^{n} y_i \lambda_i^{(1)} \tag{92}$$

subject to

$$\sum_{i=1}^{n} \lambda_i^{(1)} = \sum_{i=1}^{n} \lambda_i^{(2)}, \tag{93}$$

$$-\sum_{i=1}^{n} \lambda_i^{(1)} x_{ij} + \sum_{i=1}^{n} \lambda_i^{(2)} x_{ij} = 0, \quad j = 2, 3, \ldots, k, \tag{94}$$

$$\lambda_i^{(1)} + \lambda_i^{(2)} = 1, \quad i = 1, 2, \ldots, n, \tag{95}$$

$$\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(2)} \geq \mathbf{0}, \tag{96}$$

which shows that the sets of dual variables $\boldsymbol{\lambda}^{(1)}$ and $\boldsymbol{\lambda}^{(2)}$ can be interpreted as a set of n probability mass functions defined on the set

$$S = \{(-y_i, -\boldsymbol{x}_i)|i = 1, 2, \ldots, n\} \cup \{(y_i, \boldsymbol{x}_i)|i = 1, 2, \ldots, n\}\}$$

and such that

$$\Pr((-y_i, -\boldsymbol{x}_i)) = \lambda_i^{(1)}, \ \ \Pr((y_i, \boldsymbol{x}_i)) = \lambda_i^{(2)}.$$

Then, the constraints (94) can be interpreted as zero means, $E[X_{ij}]$, and the objective function as the sum of n means $\sum_{i=1}^{n} E[Y_i]$. So the dual problem consists of assigning n different probability mass functions to the set $\mathcal{S}$ such that they minimize the sum of expectations $\sum_{i=1}^{n} E[Y_i]$ subject to the equality $\sum_{i=1}^{n} \lambda_i^{(1)} = \sum_{i=1}^{n} \lambda_i^{(2)}$.

It is interesting to interpret the dual with the help of figure 6. The $\boldsymbol{\lambda}^{(1)}$ and $\boldsymbol{\lambda}^{(2)}$ assign non-zero probability to points above, below, and on the regression line, respectively (see the correspondence with sets I^+ and I^-).

7. Summary

In this paper, we presented closed formulas for assessing the sensitivity of the results of three standard regression estimation methods (LS, MM, and LAV) to changes in the data. The results include the objective function and the estimated parameters. Sensitivity contours are also presented to help in assessing the sensitivity of each observation in the sample. All sensitivities are illustrated both numerically and graphically. Additionally, interesting interpretations of the dual problems and dual variables are given. The method is new and very general because it can be applied to any model including linear and nonlinear models and to any method of estimation that can be formulated as an optimization problem. The proposed sensitivity measures are shown to deal more effectively with the masking problem than the existing methods.

References

[1] Edgeworth, F.Y., 1887, On observations relating to several quantities. *Hermathena*, **6**, 279–285.

[2] Edgeworth, F.Y., 1888, On a new method of reducing observations relating to several quantities. *Philisophical Magazine*, **25**, 184–191.

[3] Portnoy, S. and Koenker, R., 1997, The Gaussian hare and the Laplacian tortoise: computability of squared error versus absolute-error estimators. *Statistical Science*, **12**, 279–300.

[4] Belsley, D.A., Kuh, E. and Welsch, R.E., 1980, *Regression Diagnostics: Identifying Influential Data and Sources of Multicollinearity* (New York: John Wiley & Sons).

[5] Cook, R.D. and Weisberg, S., 1982, *Residuals and Influence in Regression* (London: Chapman and Hall).

[6] Atkinson, A.C., 1985, *Plots, Transformations, and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis* (Oxford: Clarendon Press).

[7] Chatterjee, S. and Hadi, A.S., 1988, *Sensitivity Analysis in Linear Regression* (New York: John Wiley & Sons).

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

[9] Weissfeld, I. and Schneider, H., 1990a, Influence diagnostics for the normal linear model with censored data. *Australian Journal of Statistics*, **32**, 11–20.

[10] Weissfeld, I. and Schneider, H., 1990b, Influence diagnostics for the weibull model fit to censored data. *Statistics and Probability Letters*, **9**, 67–73.

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

[12] Paul, S.R. and Fung, K.Y., 1991, Generalized extreme studentized residual multiple-outlier-detection procedure in linear regression. *Technometrics*, **33**, 339–348.

[13] Escobar, L.A. and Meeker, W.Q., 1992, Assessing influence in regression analysis with censored data. *Biometrics*, **48**, 507–528.

[14] Hadi, A.S., 1992, A new measure of overall potential influence in linear regression. *Computational Statistics and Data Analysis*, **14**, 1–27.

[15] Hadi, A.S., Simonoff, J.S., 1993, Procedures for the identification of multiple outliers in linear models. *Journal of the American Statistical Association*, **88**, 1264–1272.

[16] Peña, D. and Yohai, V., 1995, The detection of influential subsets in linear regression by using an influence matrix. *Journal of the Royal Statistical Society, Series B*, **57**, 145–156.

[17] Barrett, B.E. and Gray, J.B., 1997, On the use of robust diagnostics in least squares regression analysis. Proceedings of the Statistical Computing Section, The American Statistical Association, 130–135. **Q1**

[18] Mayo, M.S. and Gray, J.B., 1997, Elemental subsets: the building blocks of regression. *Journal of the American Statistical Association*, **51**, 122–129.

[19] Saltelli, A., Chan, K. and Scott, E.M., 2000, *Sensitivity Analysis* (New York: John Wiley & Sons).

[20] Winsnowski, W.J., Montgomery, D.C. and James, R.S., 2001, A comparative analysis of multiple outlier detection procedures in the linear regression model. *Computational Statistics and Data Analysis*, **36**, 351–382.

[21] Cook, R.D., 1986, Assessment of local influence (with discussion). *Journal of the Royal Statistical Society, Series B*, **48**, 133–169.

[22] Castillo, E., Hadi, A.S., Conejo, A. and Fernández-Canteli, A., 2004, A general method for local sensitivity analysis with application to regression models and other optimization problems. *Technometrics*, **46**, 430–444.

[23] Castillo, E., Conejo, A., Castillo, C., Mínguez, R. and Ortigosa, D., 2006a, A perturbation approach to sensitivity analysis in mathematical programming. *Journal of Optimization Theory and Applications*, **128**(1), 49–74.

[24] Castillo, E., Conejo, A., Mínguez, R. and Castillo, C., 2006b, A closed formula for local sensitivity analysis in mathematical programming. *Engineering Optimization*, **38**, 93–112.

[25] Conejo, A., Castillo, E., Mínguez, R. and García-Bertrand, R., 2006, *Decomposition Techniques in Mathematical Programming. Engineering and Science Applications* (Berlin, Heildelberg: Springer).

[26] Rousseeuw and Leroy (1987) p. 57. **Q2**

[27] Cook, R.D., 1977, Detection of influential observations in linear regression. *Technometrics*, **19**, 15–18.

[28] Klingman, D. and Mote, J., 1982, Generalized network approaches for solving least absolute value and Tchebycheff regression problems. In: *TIMS/Studies in the Management Sciences*, (North Holland Publishing Company), pp. 53–66.

[29] Arthanari, T.S. and Dodge, Y., 1993, *Mathematical Programming in Statistics* (New York: Wiley).

[30] Bazaraa, M.S., Sherali, H.D. and Shetty, C.M., 1993, *Nonlinear Programming, Theory and Algorithms* (2nd ed.) (New York: John Wiley & Sons). **Q3**

[31] Luenberger, D.G., 1989, *Linear and Nonlinear Programming* (2nd edn.) Reading (MA: Addison-Wesley). **Q3**

[32] Welsch, R.E. and Kuh, E., 1977, Linear regression diagnostics, Technical Report 923-77, Sloan School of Management, Massachusetts Institute of Technology. **Q3**