SIAM J. OPTIM.
Vol. 14, No. 4, pp. 1028–1042

THE FINITE SAMPLE BREAKDOWN POINT OF ℓ_1-REGRESSION*

AVI GILONI† AND MANFRED PADBERG‡

Abstract. Through a new (parametric) linear programming approach, we derive a formula for the finite sample breakdown point of ℓ_1-regression with a given design matrix $\mathbf{X}$ and contamination restricted to the dependent variable. This is done using the notion of the q-strength and the s-stability of a design matrix $\mathbf{X}$, which are introduced here. We discuss the relationship between our result and existing results in the literature. Finally, we demonstrate the usefulness of our result by calculating (via the solution of mixed-integer programs) the finite sample breakdown point of ℓ_1-regression with contamination restricted to the dependent variable for nine well-known data sets from the robust regression literature.

Key words. ℓ_1-regression, breakdown point, robust designs, robustness, linear programming, mixed-integer programming

AMS subject classifications. 62J99, 62F35, 90C05, 90C11

DOI. 10.1137/S1052623403424156

1. Introduction.

In this paper we discuss the finite sample breakdown point of the ℓ_1-regression estimator, with a fixed design matrix $\mathbf{X}$ and contamination restricted to the dependent variable $\mathbf{y}$, which we denote by $\alpha(\ell_1, \mathbf{y}|\mathbf{X})$ to indicate that the design matrix $\mathbf{X}$ is given. The finite sample breakdown point, or conditional breakdown point, $\alpha(\ell_1, \mathbf{y}|\mathbf{X})$, was introduced by Donoho and Huber [2]. It is especially important in *planned experiments*, where the design matrix is under the control of the experimenter. Its study has been addressed by, among others, He et al. [7], Ellis and Morgenthaler [6], and Mizera and Müller [11, 12]. We introduce the notions of the q-strength and the s-stability of $\mathbf{X}$ based on a parametric linear programming (LP) approach to the problem, which permits us to derive a formula for $\alpha(\ell_1, \mathbf{y}|\mathbf{X})$. We show that our result is consistent with earlier results. The advantage of our framework is that it permits us to compute the breakdown value via the solution of a mixed-integer program (MIP). We present computational results for nine data sets from the robust regression literature.

1.1. ℓ_1-regression.

In linear regression we have n observations on some "dependent" variable y and some number $p \geq 1$ of "independent" variables $x_1, \ldots, x_p$, for each of which we know n values as well. We denote

$$(1)\qquad \mathbf{y} = \begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{pmatrix}, \ \mathbf{X} = \begin{pmatrix} x_1^1 & \cdot & \cdot & \cdot & x_p^1 \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ x_1^n & \cdot & \cdot & \cdot & x_p^n \end{pmatrix} = \begin{pmatrix} \mathbf{x}^1 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{x}^n \end{pmatrix} = (\mathbf{x}_1, \ldots, \mathbf{x}_p),$$

where $\mathbf{y} \in \mathbb{R}^n$ is a vector of n observations and $\mathbf{X}$ is an $n \times p$ matrix of reals referred to as the design matrix. $\mathbf{x}_1, \ldots, \mathbf{x}_p$ are column vectors with n components, and

*Received by the editors March 12, 2003; accepted for publication (in revised form) October 16, 2003; published electronically May 25, 2004. This work was supported in part by a grant from the Office of Naval Research (N00014-96-0327).

http://www.siam.org/journals/siopt/14-4/42415.html

†Sy Syms School of Business, Yeshiva University, 500 West 185th Street, BH-428, New York, NY, 10033 (agiloni@ymail.yu.edu).

‡17, Rue Vendome, 13007 Marseille, France (manfred@padberg.com).

$\mathbf{x}^1, \ldots, \mathbf{x}^n$ are row vectors with p components corresponding to the columns and rows of $\mathbf{X}$, respectively. To rule out pathologies we assume throughout that the rank $r(\mathbf{X})$ of $\mathbf{X}$ is full, i.e., that $r(\mathbf{X}) = p$.

The statistical linear regression model is $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\beta}^T = (\beta_1, \ldots, \beta_p)$ is the vector of *parameters* of the linear model and $\boldsymbol{\varepsilon}^T = (\varepsilon_1, \ldots, \varepsilon_n)$ a vector of n random variables corresponding to the error terms in the asserted relationship. An upper index T denotes "transposition" of a vector or matrix throughout this work. In the statistical model, the dependent variable y is a random variable for which we obtain measurements or observations that contain some "noise" or measurement errors that are captured in the error terms $\boldsymbol{\varepsilon}$. For the numerical problem that we are facing we write

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{r}, \tag{2}$$

where, given some parameter vector $\boldsymbol{\beta}$, the components r_i of the vector $\mathbf{r}^T = (r_1, \ldots, r_n)$ are the *residuals* that result, given the observations $\mathbf{y}$, a fixed design matrix $\mathbf{X}$, and the chosen vector $\boldsymbol{\beta} \in \mathbb{R}^p$. In the case of ℓ_1-regression, the (optimal) parameters $\boldsymbol{\beta} \in \mathbb{R}^p$ are those that minimize the ℓ_1-norm $\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_1 = \sum_{i=1}^n |y_i - \mathbf{x}^i\boldsymbol{\beta}|$ of the residuals.

The ℓ_1-regression problem can be formulated as the LP problem

$$\begin{aligned} &\min\ \mathbf{e}_n^T\mathbf{r}^+ + \mathbf{e}_n^T\mathbf{r}^- \\ \text{s.t.}\quad &\mathbf{X}\boldsymbol{\beta} + \mathbf{r}^+ - \mathbf{r}^- = \mathbf{y}, \\ &\boldsymbol{\beta} \text{ free, } \mathbf{r}^+ \geq \mathbf{0},\ \mathbf{r}^- \geq \mathbf{0}, \end{aligned} \tag{3}$$

where $\mathbf{e}_n$ is the vector with all n components equal to 1. In (3) the residuals $\mathbf{r}$ of the general form (2) are simply replaced with a difference $\mathbf{r}^+ - \mathbf{r}^-$ of nonnegative variables; i.e., we require that $\mathbf{r}^+ \geq \mathbf{0}$ and $\mathbf{r}^- \geq \mathbf{0}$, whereas the parameters $\boldsymbol{\beta} \in \mathbb{R}^p$ are "free" to assume positive, zero, or negative values. From the properties of LP solution procedures, it follows that in any solution inspected by, e.g., the simplex algorithm, either $r_i^+ > 0$ or $r_i^- > 0$, but not both, thus giving $|r_i|$ in the objective function depending on whether $r_i > 0$ or $r_i < 0$ for any $i \in N$, where $N = \{1, \ldots, n\}$.

To characterize the optimality of coefficients $\boldsymbol{\beta} \in \mathbb{R}^p$ for ℓ_1-regression let

$$Z_{\boldsymbol{\beta}} = \left\{i \in N : r_i^{\boldsymbol{\beta}} = 0\right\},\ U_{\boldsymbol{\beta}} = \left\{i \in N : r_i^{\boldsymbol{\beta}} > 0\right\},\ L_{\boldsymbol{\beta}} = \left\{i \in N : r_i^{\boldsymbol{\beta}} < 0\right\}, \tag{4}$$

where $r_i^{\boldsymbol{\beta}} = y_i - \mathbf{x}^i\boldsymbol{\beta}$ for all $i \in N$. Let $\mathbf{X}_Z = (\mathbf{x}^i)_{i\in Z}$, $\mathbf{e}_Z = (1, \ldots, 1)^T$ with $|Z|$ components equal to 1 (i.e., $|Z|$ is the cardinality of the set Z). $\mathbf{X}_U$, $\mathbf{e}_U$, $\mathbf{r}_U$, $\mathbf{X}_L$, $\mathbf{e}_L$, and $\mathbf{r}_L$ are defined likewise.

THEOREM 1. *Let $\boldsymbol{\beta} \in \mathbb{R}^p$ and $Z_{\boldsymbol{\beta}}, U_{\boldsymbol{\beta}}, L_{\boldsymbol{\beta}}$ be as defined in* (4). *$\boldsymbol{\beta}$ is an optimal solution to $\min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_1$ if and only if there exists $\mathbf{v} \in \mathbb{R}^{|Z_{\boldsymbol{\beta}}|}$ such that*

$$\mathbf{v}\mathbf{X}_{Z_{\boldsymbol{\beta}}} = -\mathbf{e}_{U_{\boldsymbol{\beta}}}^T\mathbf{X}_{U_{\boldsymbol{\beta}}} + \mathbf{e}_{L_{\boldsymbol{\beta}}}^T\mathbf{X}_{L_{\boldsymbol{\beta}}}, \qquad -\mathbf{e}_{Z_{\boldsymbol{\beta}}}^T \leq \mathbf{v} \leq \mathbf{e}_{Z_{\boldsymbol{\beta}}}^T, \tag{5}$$

i.e., if and only if (5) *is solvable.*

Proof. The dual linear program to (3) is given by

$$\max\left\{\mathbf{u}\mathbf{y} : \mathbf{u}\mathbf{X} = \mathbf{0},\ -\mathbf{e}_n^T \leq \mathbf{u} \leq \mathbf{e}_n^T\right\} = \max\left\{\mathbf{u}\mathbf{r} : \mathbf{u}\mathbf{X} = \mathbf{0},\ -\mathbf{e}_n^T \leq \mathbf{u} \leq \mathbf{e}_n^T\right\},$$

where the equality follows because $\mathbf{u}\mathbf{y} = \mathbf{u}(\mathbf{X}\boldsymbol{\beta} + \mathbf{r}) = \mathbf{u}\mathbf{r}$ for all $\mathbf{u} \in \mathbb{R}^n$ satisfying $\mathbf{u}\mathbf{X} = \mathbf{0}$. Suppose condition (5) is satisfied. Define $u_i = 1$ for $i \in U_{\boldsymbol{\beta}}$, $u_i = -1$ for

$i \in L_{\boldsymbol\beta}$, and $\mathbf{u}_{Z_{\boldsymbol\beta}} = \mathbf{v}$. Then $\mathbf{u}$ is a feasible solution to the dual, $\mathbf{ur} = \mathbf{e}_U^T\mathbf{r}_U - \mathbf{e}_L^T\mathbf{r}_L = \|\mathbf{r}\|_1$, and thus by the weak theorem of duality of LP, $\boldsymbol\beta$ is an optimal solution. Suppose $\boldsymbol\beta$ is an optimal solution to the ℓ_1-regression problem, but that $\mathbf{v} \in \mathbb{R}^{|Z_{\boldsymbol\beta}|}$ satisfying (5) does not exist. By Farkas's lemma (see, e.g., [1, Exercise 6.5] or [13]), there exist $\boldsymbol\xi \in \mathbb{R}^p$, $\boldsymbol\eta^+, \boldsymbol\eta^- \in \mathbb{R}^{|Z_{\boldsymbol\beta}|}$ such that

$$\mathbf{X}_{Z_{\boldsymbol\beta}}\boldsymbol\xi + \boldsymbol\eta^+ - \boldsymbol\eta^- = \mathbf{0},\ \left(-\mathbf{e}_{U_{\boldsymbol\beta}}^T\mathbf{X}_{U_{\boldsymbol\beta}} + \mathbf{e}_{L_{\boldsymbol\beta}}^T\mathbf{X}_{L_{\boldsymbol\beta}}\right)\boldsymbol\xi + \mathbf{e}_{Z_{\boldsymbol\beta}}^T\boldsymbol\eta^+ + \mathbf{e}_{Z_{\boldsymbol\beta}}^T\boldsymbol\eta^- < 0, \quad \boldsymbol\eta^+ \ge \mathbf{0},\ \boldsymbol\eta^- \ge \mathbf{0}. \tag{6}$$

If $Z_{\boldsymbol\beta} = \emptyset$, then $-\mathbf{e}_{U_{\boldsymbol\beta}}^T\mathbf{X}_{U_{\boldsymbol\beta}} + \mathbf{e}_{L_{\boldsymbol\beta}}^T\mathbf{X}_{L_{\boldsymbol\beta}} \neq \mathbf{0}$ since otherwise the dual linear program, and thus (5), has a solution. In this case we choose any $\boldsymbol\xi \in \mathbb{R}^p$ such that $(-\mathbf{e}_{U_{\boldsymbol\beta}}^T\mathbf{X}_{U_{\boldsymbol\beta}} + \mathbf{e}_{L_{\boldsymbol\beta}}^T\mathbf{X}_{L_{\boldsymbol\beta}})\boldsymbol\xi < 0$. Since $\mathbf{r}_{U_{\boldsymbol\beta}} > \mathbf{0}$ and $\mathbf{r}_{L_{\boldsymbol\beta}} < \mathbf{0}$, there exists $\lambda > 0$ such that $\mathbf{r}^+_{U_{\boldsymbol\beta}}(\lambda) = \mathbf{r}_{U_{\boldsymbol\beta}} - \lambda\mathbf{X}_{U_{\boldsymbol\beta}}\boldsymbol\xi \ge \mathbf{0}$ and $\mathbf{r}^-_{L_{\boldsymbol\beta}}(\lambda) = -\mathbf{r}_{L_{\boldsymbol\beta}} + \lambda\mathbf{X}_{L_{\boldsymbol\beta}}\boldsymbol\xi \ge \mathbf{0}$. Consequently, $\boldsymbol\beta(\lambda) = \boldsymbol\beta + \lambda\boldsymbol\xi$, $\mathbf{r}^\pm_{Z_{\boldsymbol\beta}}(\lambda) = \lambda\boldsymbol\eta^\pm$, $\mathbf{r}^+_{U_{\boldsymbol\beta}}(\lambda)$, $\mathbf{r}^-_{U_{\boldsymbol\beta}}(\lambda) = \mathbf{0}$, $\mathbf{r}^+_{L_{\boldsymbol\beta}}(\lambda) = \mathbf{0}$, and $\mathbf{r}^-_{L_{\boldsymbol\beta}}(\lambda)$ define a feasible solution to the linear program (3). Calculating its objective function we get

$$\begin{aligned}\mathbf{e}_N^T\mathbf{r}^+(\lambda) + \mathbf{e}_N^T\mathbf{r}^-(\lambda) &= \mathbf{e}_{U_{\boldsymbol\beta}}^T\mathbf{r}^+_{U_{\boldsymbol\beta}}(\lambda) + \mathbf{e}_{L_{\boldsymbol\beta}}^T\mathbf{r}^-_{L_{\boldsymbol\beta}}(\lambda) + \lambda\left(\mathbf{e}_{Z_{\boldsymbol\beta}}^T\boldsymbol\eta^+ + \mathbf{e}_{Z_{\boldsymbol\beta}}^T\boldsymbol\eta^-\right)\\ &= \|\mathbf{r}\|_1 + \lambda\left(-\mathbf{e}_{U_{\boldsymbol\beta}}^T\mathbf{X}_{U_{\boldsymbol\beta}}\boldsymbol\xi + \mathbf{e}_{L_{\boldsymbol\beta}}^T\mathbf{X}_{L_{\boldsymbol\beta}}\boldsymbol\xi + \mathbf{e}_{Z_{\boldsymbol\beta}}^T\boldsymbol\eta^+ + \mathbf{e}_{Z_{\boldsymbol\beta}}^T\boldsymbol\eta^-\right) < \|\mathbf{r}\|_1,\end{aligned}$$

and consequently $\boldsymbol\beta$ is not optimal. □

As one of the referees pointed out, a different proof of Theorem 1 can be obtained by applying Theorem 2.2.1 of [8, p. 253]. We leave the details to the interested reader.

2. The breakdown point of ℓ_1-regression.

2.1. Breakdown point. The notion of the *breakdown point* of a regression estimator due to Hampel [5] can be found, e.g., in [14], and reads as follows. Suppose we estimate the regression parameters $\boldsymbol\beta$ by some technique τ from some data $(\mathbf{X}, \mathbf{y})$ to be $\boldsymbol\beta^\tau$. If we replace any number $1 \le m < n$ of the data with some arbitrary data $(\widetilde{\mathbf{x}}^i, \widetilde{y}_i)$, we obtain new data $(\widetilde{\mathbf{X}}, \widetilde{\mathbf{y}})$. The same technique τ applied to $(\widetilde{\mathbf{X}}, \widetilde{\mathbf{y}})$ yields estimates $\boldsymbol\beta^\tau(\widetilde{\mathbf{X}}, \widetilde{\mathbf{y}})$ that are different from the original ones. We can use any norm $\|\cdot\|$ on $\mathbb{R}^p$ to measure the distance $\|\boldsymbol\beta^\tau(\widetilde{\mathbf{X}}, \widetilde{\mathbf{y}}) - \boldsymbol\beta^\tau\|$ of the respective estimates. If we vary over *all* possible choices, then this distance remains either bounded or not bounded. Let

$$b(m, \tau, \mathbf{X}, \mathbf{y}) = \sup_{\widetilde{\mathbf{X}}, \widetilde{\mathbf{y}}} \left\|\boldsymbol\beta^\tau\left(\widetilde{\mathbf{X}}, \widetilde{\mathbf{y}}\right) - \boldsymbol\beta^\tau\right\| \tag{7}$$

be the *maximum bias* that results when we replace at most m of the original data $(\mathbf{x}^i, y_i)$ with arbitrary new data. Let

$$b(m, \tau, \mathbf{y}|\mathbf{X}) = \sup_{\widetilde{\mathbf{y}}} \left\|\boldsymbol\beta^\tau\left(\mathbf{X}, \widetilde{\mathbf{y}}\right) - \boldsymbol\beta^\tau\right\| \tag{8}$$

be the *maximum bias* that results when we replace at most m of the original values of the dependent variable y_i with arbitrary new data. The breakdown point of τ is

$$\alpha(\tau, \mathbf{X}, \mathbf{y}) = \min_{1 \le m < n}\left\{\frac{m}{n} : b(m, \tau, \mathbf{X}, \mathbf{y}) \text{ is infinite}\right\};$$

i.e., it is the *minimum* number of rows of $(\mathbf{X}, \mathbf{y})$ that, if replaced with arbitrary new data, make the regression technique τ break down. The conditional breakdown point of τ is

$$\alpha(\tau, \mathbf{y}|\mathbf{X}) = \min_{1 \le m < n} \left\{ \frac{m}{n} : b(m, \tau, \mathbf{y}|\mathbf{X}) \text{ is infinite} \right\};$$

i.e., it is the *minimum* number of values of $\mathbf{y}$ that, if replaced with arbitrary new data, make the regression technique τ break down. We divide by n to get $\frac{1}{n} \le \alpha(\tau, \mathbf{X}, \mathbf{y}) \le 1$.

The breakdown point of ℓ_1-regression is $\frac{1}{n}$ or, asymptotically, 0; see, e.g., [14]. However, the determination of the conditional breakdown point $\alpha(\ell_1, \mathbf{y}|\mathbf{X})$ of ℓ_1-regression is not straightforward. He et al. [7] disprove a claim of Donoho and Huber [2, p. 166] that $\alpha(\ell_1, \mathbf{y}|\mathbf{X})$ is $\frac{1}{2}$ or 50%. This was observed in [3] independently of He et al.'s work.

Example 1. Let $n = 3$, $p = 2$ with data

$$(\mathbf{X}, \mathbf{y}) = \begin{pmatrix} 1 & 0 & 2 \\ 1 & 2 & 3 \\ 1 & -1 & 2 \end{pmatrix}, \tag{9}$$

and suppose that y_2 is contaminated. We replace $y_2 = 3$ with $y_2 = 3 + \vartheta$, where $\vartheta \ge 0$ is arbitrary. We calculate that the optimal ℓ_1-regression coefficients $\boldsymbol{\beta}(\vartheta)$ are $\beta_1(\vartheta) = \frac{7}{3} + \frac{\vartheta}{3}$, $\beta_2(\vartheta) = \frac{1}{3} + \frac{\vartheta}{3}$ for all $\vartheta \ge 0$. The optimal residuals are $r_1^-(\vartheta) = \frac{1}{3} + \frac{\vartheta}{3}$, and $r_i^+(\vartheta) = r_i^-(\vartheta) = 0$ otherwise. Thus $\|\boldsymbol{\beta}(\vartheta) - \boldsymbol{\beta}(0)\|_1 = \frac{2}{3}\vartheta \to +\infty$ for $\vartheta \to +\infty$, and thus a single contaminated observation in $\mathbf{y}$ may cause ℓ_1-regression to break down. $\square$

This example can be generalized to higher dimensions. The idea that underlies the counterexample to Donoho and Huber's statement is the following: For any $\vartheta \ge 0$, the optimal basis of (3) corresponding to the data contains row $\mathbf{x}^2$ of the design matrix, but neither r_2^+ nor r_2^-. Hence the optimal $\boldsymbol{\beta}(\vartheta)$ depend on the amount ϑ of the contamination of y_2, and thus the maximum bias that results from the contamination grows beyond all bounds.

2.2. *q*-strength and *s*-stability of design matrices. To study $\alpha(\ell_1, \mathbf{y}|\mathbf{X})$, we consider the parametric linear program corresponding to (3),

$$z(\vartheta) = \min \left\{ \mathbf{e}_n^T \mathbf{r}^+ + \mathbf{e}_n^T \mathbf{r}^- : \mathbf{X}\boldsymbol{\beta} + \mathbf{r}^+ - \mathbf{r}^- = \mathbf{y} + \vartheta \mathbf{g},\, \mathbf{r}^+ \ge \mathbf{0},\, \mathbf{r}^- \ge \mathbf{0} \right\}, \tag{10}$$

where $\mathbf{g} \in \mathbb{R}^n$ is arbitrary and $\vartheta \ge 0$ is some parameter. Since $\mathbf{g} \in \mathbb{R}^n$ is arbitrary, the sign restriction on ϑ does not matter. By varying ϑ and $\mathbf{g}$, every possible contamination of the components of the observation vector $\mathbf{y} \in \mathbb{R}^n$ is obtained. It is well known (see, e.g., [13, p. 102]) that $z(\vartheta)$ is a convex, piecewise linear function of ϑ.

LEMMA 1. *Let* $\mathbf{P}^* = \{\mathbf{u} \in \mathbb{R}^n : \mathbf{u}\mathbf{X} = \mathbf{0},\, -\mathbf{e}_n^T \le \mathbf{u} \le \mathbf{e}_n^T\}$. *Then*

$$0 \le z(\vartheta) = z(\vartheta_0) + (\vartheta - \vartheta_0)\mathbf{u}^0\mathbf{g} \le z(\vartheta_0) + (\vartheta - \vartheta_0)\|\mathbf{g}\|_1$$

for all $\vartheta \ge \vartheta_0$ *with some finite* $\vartheta_0 \ge 0$ *and* $\mathbf{u}^0$ *some extreme point of* $\mathbf{P}^*$.

Proof. From (10), $z(\vartheta) \ge 0$. Since $(\boldsymbol{\beta}, \mathbf{r}^+, \mathbf{r}^-) = (\mathbf{0}, \max\{\mathbf{0}, \mathbf{y} + \vartheta\mathbf{g}\}, -\min\{\mathbf{0}, \mathbf{y} + \vartheta\mathbf{g}\})$ is feasible for (10), it follows from the triangle inequality that $z(\vartheta) \le \|\mathbf{y} + \vartheta\mathbf{g}\|_1 \le \|\mathbf{y}\|_1 + \vartheta\|\mathbf{g}\|_1$ for all $\vartheta \ge 0$. Consequently, by the duality theorem of LP,

$$z(\vartheta) = \max\{\mathbf{u}(\mathbf{y} + \vartheta\mathbf{g}) : \mathbf{u} \in \mathbf{P}^*\}. \tag{11}$$

Since $\mathbf{0} \in \mathbf{P}^*$ and $\mathbf{P}^* \subseteq \{\mathbf{u} \in \mathbb{R}^n : -\mathbf{e}_n^T \leq \mathbf{u} \leq \mathbf{e}_n^T\}$, $\mathbf{P}^*$ is a nonempty polytope. Hence, $\mathbf{P}^* = \operatorname{conv}\{\mathbf{u}^1, \ldots, \mathbf{u}^r\}$, where $\mathbf{u}^i$ is an extreme point of $\mathbf{P}^*$ and $r > 0$ a finite number. Thus

$$z(\vartheta) = \max\left\{\mathbf{u}^i(\mathbf{y} + \vartheta\mathbf{g}) : 1 \leq i \leq r\right\} = z(\vartheta_0) + (\vartheta - \vartheta_0)\mathbf{u}^0\mathbf{g} \tag{12}$$

for all $\vartheta \geq \vartheta_0$, where $\vartheta_0 \geq 0$ is some finite value of the parameter and $\mathbf{u}^0$ a corresponding optimal extreme point of $\mathbf{P}^*$ for $\vartheta = \vartheta_0$. Finally, $\mathbf{u}^0\mathbf{g} \leq \|\mathbf{g}\|_1$, since $-\mathbf{e}_n^T \leq \mathbf{u}^0 \leq \mathbf{e}_n^T$. ◻

To analyze $\alpha(\ell_1, \mathbf{y}|\mathbf{X})$, let

$$\mathcal{F}_q = \{(U, L, Z) : U \subseteq N,\ L \subseteq N - U,\ Z = N - (U \cup L),\ |U \cup L| = q\}.$$

We call a design matrix $\mathbf{X}$ *q-strong* if q is the largest integer such that

$$\mathbf{v}\mathbf{X}_Z = -\mathbf{e}_U^T\mathbf{X}_U + \mathbf{e}_L^T\mathbf{X}_L, \qquad -\mathbf{e}_Z^T \leq \mathbf{v} \leq \mathbf{e}_Z^T \tag{13}$$

is solvable for all $(U, L, Z) \in \mathcal{F}_q$. Note the similarity between (5) and (13). Geometrically, we require that q be the largest integer such that the *faces*

$$F(U, L) = \mathbf{P}^* \cap \{\mathbf{u} \in \mathbb{R}^n : u_j = 1 \text{ for } j \in U,\ u_j = -1 \text{ for } j \in L\}$$

of the dual polytope $\mathbf{P}^*$ are nonempty for all $(U, L, Z) \in \mathcal{F}_q$. Since $\mathbf{P}^* \neq \emptyset$, every design matrix is q-strong for some $q \geq 0$. The condition $|L \cup U| = q$ can be replaced with $|L \cup U| \leq q$ in the definition of q-strength. Thus $0 \leq q \leq n$ is well defined for every design matrix $\mathbf{X}$. In the numerical example of section 2.1, the dual polytope $\mathbf{P}^*$ has precisely two extreme points $\mathbf{u}^1 = (1, -\frac{1}{3}, -\frac{2}{3})$ and $\mathbf{u}^2 = -\mathbf{u}^1$. Hence the design matrix $\mathbf{X}$ of (9) is 0-strong, which explains its breakdown point of $\frac{1}{3}$.

PROPOSITION 1. *If $\mathbf{X}$ is q-strong, then $q \leq n - p$ and $\alpha(\ell_1, \mathbf{y}|\mathbf{X}) \leq \frac{q+1}{n}$.*

Proof. Suppose $q > n - p$ and let $(U, L, Z) \in \mathcal{F}_q$. Thus $|Z| = n - q < p$. Consequently, $\mathbf{X}_Z\boldsymbol{\xi} = \mathbf{0}$ has a solution $\boldsymbol{\xi} \neq \mathbf{0}$. Let $\boldsymbol{\eta}^+ = \boldsymbol{\eta}^- = \mathbf{0}$. If $(-\mathbf{e}_U^T\mathbf{X}_U + \mathbf{e}_L^T\mathbf{X}_L)\boldsymbol{\xi} \neq 0$, then $(\boldsymbol{\xi}, \boldsymbol{\eta}^+, \boldsymbol{\eta}^-)$ or $(-\boldsymbol{\xi}, \boldsymbol{\eta}^+, \boldsymbol{\eta}^-)$ solve (6). Thus by Farkas's lemma, (13) is not solvable, which is a contradiction. Suppose $(-\mathbf{e}_U^T\mathbf{X}_U + \mathbf{e}_L^T\mathbf{X}_L)\boldsymbol{\xi} = 0$. Since $r(\mathbf{X}) = p$, there exist $i \in U \cup L$ such that $\mathbf{x}^i\boldsymbol{\xi} \neq 0$. If $i \in U$, let $S = U - i$ and $T = L + i$. It follows that $(-\mathbf{e}_S^T\mathbf{X}_S + \mathbf{e}_T^T\mathbf{X}_T)\boldsymbol{\xi} = 2\mathbf{x}^i\boldsymbol{\xi} \neq 0$, and thus we contradict with $(S, T, Z) \in \mathcal{F}_q$ as in the previous case. If $i \in L$, we use $S = U + i$ and $T = L - i$. Thus $q \leq n - p$. To prove the second part we proceed as follows. Since $\mathbf{X}$ is q-strong, there exist $(U, L, Z) \in \mathcal{F}_{q+1}$ such that (13) is not solvable. Let

$$g_j = +1 \text{ for all } j \in U, \quad g_j = -1 \text{ for all } j \in L, \quad g_j = 0 \text{ otherwise,}$$

and $\mathbf{g} = (g_j)_{j \in N}$. From Lemma 1, $z(\vartheta) = z(\vartheta_0) + (\vartheta - \vartheta_0)\mathbf{u}^0\mathbf{g}$ for all $\vartheta \geq \vartheta_0$, where $\mathbf{u}^0$ is some optimal extreme point of $\mathbf{P}^*$ for $\vartheta = \vartheta_0$. Since (13) is not solvable for U and L, it follows that $-1 < u_j^0 < 1$ for some $j \in U \cup L$. Let $(\boldsymbol{\beta}(\vartheta_0), \mathbf{r}^+(\vartheta_0), \mathbf{r}^-(\vartheta_0))$ be any optimal solution to (10) for $\vartheta = \vartheta_0$. By complementary slackness, $r_j^+(\vartheta_0) = r_j^-(\vartheta_0) = 0$, and thus $\mathbf{x}^j\boldsymbol{\beta}(\vartheta_0) = y_j + \vartheta_0 g_j$. Let $\vartheta > \vartheta_0$ be arbitrary. Since $\mathbf{u}^0$ is unchanged, it follows as before that $r_j^+(\vartheta) = r_j^-(\vartheta) = 0$ in any optimal solution to (10), and thus $\mathbf{x}^j\boldsymbol{\beta}(\vartheta) = y_j + \vartheta g_j$. Consequently, $|\mathbf{x}^j\boldsymbol{\beta}(\vartheta) - \mathbf{x}^j\boldsymbol{\beta}(\vartheta_0)| = \vartheta - \vartheta_0$. By the Cauchy–Schwarz inequality, $\vartheta - \vartheta_0 = |\mathbf{x}^j(\boldsymbol{\beta}(\vartheta) - \boldsymbol{\beta}(\vartheta_0))| \leq \|\mathbf{x}^j\|\|\boldsymbol{\beta}(\vartheta) - \boldsymbol{\beta}(\vartheta_0)\|$, and hence $\|\boldsymbol{\beta}(\vartheta) - \boldsymbol{\beta}(\vartheta_0)\| \to +\infty$ for $\vartheta \to +\infty$; i.e., the maximum bias (8) grows beyond all bounds for $\mathbf{g}$. Thus $\alpha(\ell_1, \mathbf{y}|\mathbf{X}) \leq \frac{q+1}{n}$. ◻

PROPOSITION 2.

(i) *If* $\mathbf{X}$ *is* q*-strong and* $\mathbf{g} \in \mathbb{R}^n$ *in* (10) *has* q *nonzero components, then* (10) *has an optimal solution* $(\boldsymbol{\beta}(\vartheta), \mathbf{r}^+(\vartheta), \mathbf{r}^-(\vartheta))$ *with* $\lim_{\vartheta\to\infty} \|\boldsymbol{\beta}(\vartheta)\|_1 < \infty$.

(ii) *If* $\mathbf{X}$ *is* q*-strong and the solution to* (10) *is unique for* $\vartheta \geq \vartheta_0$*, then* $\alpha(\ell_1, \mathbf{y}|\mathbf{X}) = \frac{q+1}{n}$.

Proof. (i) Let $S = \{i \in N : g_i > 0\}$, $T = \{i \in N : g_i < 0\}$. Since $\mathbf{X}$ is q-strong and $|S \cup T| = q$, it follows that there exists an extreme point $\mathbf{u}^k \in \mathbf{P}^*$ such that $u^k_\ell = 1$ for all $\ell \in S$ and $u^k_\ell = -1$ for all $\ell \in T$. Since $\mathbf{u}^i\mathbf{g} \leq \|\mathbf{g}\|_1$ for all $i = 1, \ldots, r$, and (12) holds for arbitrarily large $\vartheta \geq \vartheta_0$, it follows that any optimal dual extreme point $\mathbf{u}^0$ of $\mathbf{P}^*$ satisfies $u^0_\ell = 1$ for all $\ell \in S$ and $u^0_\ell = -1$ for all $\ell \in T$. Consequently,

$$z(\vartheta) = z(\vartheta_0) + (\vartheta - \vartheta_0)\|\mathbf{g}\|_1 \text{ for all } \vartheta \geq \vartheta_0. \tag{14}$$

Let $(\boldsymbol{\beta}(\vartheta_0), \mathbf{r}^+(\vartheta_0), \mathbf{r}^-(\vartheta_0))$ be an optimal solution to (10) for $\vartheta = \vartheta_0$ and define

$$\left(\boldsymbol{\beta}(\vartheta), \mathbf{r}^+(\vartheta), \mathbf{r}^-(\vartheta)\right) = \left(\boldsymbol{\beta}(\vartheta_0), \mathbf{r}^+(\vartheta_0) + (\vartheta - \vartheta_0)\mathbf{g}^+, \mathbf{r}^-(\vartheta_0) + (\vartheta - \vartheta_0)\mathbf{g}^-\right),$$

where $\mathbf{g}^+ = \max\{\mathbf{0}, \mathbf{g}\}$ and $\mathbf{g}^- = -\min\{\mathbf{0}, \mathbf{g}\}$. $(\boldsymbol{\beta}(\vartheta), \mathbf{r}^+(\vartheta), \mathbf{r}^-(\vartheta))$ is a feasible solution to (10) for all $\vartheta \geq \vartheta_0$. Since $\mathbf{e}_n^T\mathbf{r}^+(\vartheta) + \mathbf{e}_n^T\mathbf{r}^-(\vartheta) = z(\vartheta_0) + (\vartheta - \vartheta_0)\|\mathbf{g}\|_1$, by the duality theorem, it is an optimal solution to (10). Hence $\|\boldsymbol{\beta}(\vartheta)\|_1 = \|\boldsymbol{\beta}(\vartheta_0)\|_1 < \infty$ for $\vartheta \to +\infty$ and the assertion follows.

(ii) If the solution to (10) is unique for all $\vartheta \geq \vartheta_0$, then by part (i) the maximum bias remains bounded if q components of $\mathbf{y}$ are contaminated. Since the contamination vector $\mathbf{g}$ is perfectly arbitrary, by Proposition 1, the assertion follows. □

A necessary and sufficient condition for ℓ_1-regression to have a unique solution can be found, e.g., in [4, Proposition 3].

We next give a different condition for $\mathbf{X}$ to be q-strong. It yields the basis for an algorithmic approach to find the breakdown point of ℓ_1-regression. For $(U, L, Z) \in \mathcal{F}_q$ let

$$z(U,L) = \min\{(-\mathbf{e}_U^T\mathbf{X}_U + \mathbf{e}_L^T\mathbf{X}_L)\boldsymbol{\xi} + \mathbf{e}_Z^T(\boldsymbol{\eta}^+ + \boldsymbol{\eta}^-) : \mathbf{X}_Z\boldsymbol{\xi} + \boldsymbol{\eta}^+ - \boldsymbol{\eta}^- = \mathbf{0},\ \boldsymbol{\eta}^+ \geq \mathbf{0},\ \boldsymbol{\eta}^- \geq \mathbf{0}\}$$

and note that $z(U, L) = 0$ for $U = L = \emptyset$.

THEOREM 2. *A design matrix* $\mathbf{X}$ *is* q*-strong if and only if* q *is the largest integer such that* $z(U, L) \geq 0$ *for all* $(U, L, Z) \in \mathcal{F}_q$.

Proof. We establish sufficiency first. By assumption, for every $(U, L, Z) \in \mathcal{F}_q$

$$\min\{(-\mathbf{e}_U^T\mathbf{X}_U + \mathbf{e}_L^T\mathbf{X}_L)\boldsymbol{\xi} + \mathbf{e}_Z^T\boldsymbol{\eta}^+ + \mathbf{e}_Z^T\boldsymbol{\eta}^- : \mathbf{X}_Z\boldsymbol{\xi} + \boldsymbol{\eta}^+ - \boldsymbol{\eta}^- = \mathbf{0},\ \boldsymbol{\eta}^+ \geq \mathbf{0},\ \boldsymbol{\eta}^- \geq \mathbf{0}\} = 0.$$

By the strong duality theorem of LP, the dual of this linear program,

$$\max\{\mathbf{u}\mathbf{0} : \mathbf{u}\mathbf{X}_Z = -\mathbf{e}_U^T\mathbf{X}_U + \mathbf{e}_L^T\mathbf{X}_L,\ -\mathbf{e}_Z^T \leq \mathbf{u} \leq \mathbf{e}_Z^T\},$$

has a finite optimum. Thus (13) is solvable for all choices of $(U, L, Z) \in \mathcal{F}_q$. Suppose that $\mathbf{X}$ is m-strong. It follows that $m \geq q$. Suppose that $m > q$. Then the dual program has a finite optimum for all $(U, L, Z) \in \mathcal{F}_m$ and, by strong duality, so does the primal, i.e., $z(U, L) = 0$ for all $(U, L, Z) \in \mathcal{F}_m$. This contradicts the assumption that q is the largest such integer. Consequently, $\mathbf{X}$ is q-strong. On the other hand, let $\mathbf{X}$ be q-strong. It follows that

$$P^*(U, L) = \{\mathbf{u} \in \mathbb{R}^Z : \mathbf{u}\mathbf{X}_Z = -\mathbf{e}_U^T\mathbf{X}_U + \mathbf{e}_L^T\mathbf{X}_L,\ -\mathbf{e}_Z^T \leq \mathbf{u} \leq \mathbf{e}_Z^T\} \neq \emptyset$$

for all $(U, L, Z) \in \mathcal{F}_q$. $P^*(U, L)$ is a nonempty polytope, and thus by strong duality

$$0 = \max\{\mathbf{u0} : \mathbf{u} \in P^*(U, L)\} = z(U, L).$$

Suppose that $z(U, L) \geq 0$ for all $(U, L, Z) \in \mathcal{F}_m$ with $m > q$. Then, as in the first part of the proof, we have a contradiction by the fact that q is the largest integer such that (13) is solvable. □

It follows from Proposition 2 that for any q-strong design matrix $\mathbf{X}$ and finite $\mathbf{y}$, there exist ℓ_1-regression coefficients whose bias is bounded for any set of q contaminated elements of $\mathbf{y}$. While a q-strong $\mathbf{X}$ may permit ℓ_1-regression coefficients, for which the bias grows beyond all bounds, ℓ_1-regression with q-strong $\mathbf{X}$ still is robust since one needs only to ensure in such a case that the ℓ_1-regression coefficients used are bounded. We show by a brief example that this anomaly of ℓ_1-regression due to multiple optima to (10) can indeed occur.

Example 2. Consider the data

$$(\mathbf{X}, \mathbf{y}) = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 3 & 0 \\ 1 & 4 & 0 \end{pmatrix}.$$

It is easily verified that $\mathbf{X}$ is 1-strong. If the fourth observation is contaminated, we find that the parametric linear program (10)

$$\min \sum_{i=1}^{4} (r_i^+ + r_i^-) \quad \text{s.t.} \quad \begin{array}{llllllllll} \beta_1 + \ \beta_2 & + r_1^+ & & & & - r_1^- & & & & = 0, \\ \beta_1 + 2\beta_2 & & + r_2^+ & & & & - r_2^- & & & = 0, \\ \beta_1 + 3\beta_2 & & & + r_3^+ & & & & - r_3^- & & = 0, \\ \beta_1 + 4\beta_2 & & & & + r_4^+ & & & & - r_4^- & = \vartheta, \end{array}$$

where β_1 and β_2 are free, $r_i^+ \geq 0$ and $r_i^- \geq 0$ for $i = 1, \ldots, 4$, has two basic optimal solutions given by $\beta_1^0 = \beta_2^0 = 0$, $r_4^+ = \vartheta$ and by $\beta_1^1 = -\vartheta$, $\beta_2^1 = \frac{1}{2}\vartheta$, $r_1^+ = \frac{1}{2}\vartheta$, $r_3^- = \frac{1}{2}\vartheta$, respectively, where in both cases $r_i^+ = 0$ and $r_i^- = 0$ otherwise. Thus in this example we have, in agreement with Proposition 2, the existence of optimal ℓ_1-regression coefficients $\boldsymbol{\beta}^0$ for which the bias is bounded, whereas for $\boldsymbol{\beta}^{\mathbf{1}}$ the bias grows without bounds. □

We call $\mathbf{X}$ *s-stable* if $s \geq 0$ is the largest integer such that

$$\begin{aligned} \text{(15)} \qquad \mathbf{X}_Z\boldsymbol{\xi} + \boldsymbol{\eta}^+ - \boldsymbol{\eta}^- = \mathbf{0},\ \left(-\mathbf{e}_U^T\mathbf{X}_U + \mathbf{e}_L^T\mathbf{X}_L\right)\boldsymbol{\xi} + \mathbf{e}_Z^T\left(\boldsymbol{\eta}^+ + \boldsymbol{\eta}^-\right) \leq 0, \\ \boldsymbol{\xi} \neq \mathbf{0}, \qquad \boldsymbol{\eta}^+ \geq \mathbf{0}, \qquad \boldsymbol{\eta}^- \geq \mathbf{0}, \end{aligned}$$

is *not solvable* for any $(U, L, Z) \in \mathcal{F}_s$. It follows that $s \geq 0$ is well defined for any $\mathbf{X}$ with $r(\mathbf{X}) = p$. If $|Z| < p$, then (15) is solvable: In this case there exists $\boldsymbol{\xi} \neq \mathbf{0}$ such that $\mathbf{X}_Z\boldsymbol{\xi} = \mathbf{0}$. If $(-\mathbf{e}_U^T\mathbf{X}_U + \mathbf{e}_L^T\mathbf{X}_L)\boldsymbol{\xi} \leq 0$, then $\boldsymbol{\xi}$ and $\boldsymbol{\eta}^+ = \boldsymbol{\eta}^- = \mathbf{0}$ solve (15); otherwise, we change the sign of $\boldsymbol{\xi}$. It follows that $s \leq n - p$. Note the subtle difference between (6) and (15). More precisely, every solution to (6) is feasible for (15), but not vice versa.

Example 2 (continued). Consider the 1-strong design matrix $\mathbf{X}$ of Example 2. $r(\mathbf{X}) = p$ and thus $s \geq 0$. Let $(U, L, Z) \in \mathcal{F}_1$, where $U = \{4\}$, $L = \emptyset$, and $Z = \{1, 2, 3\}$. Then $\xi_1 = -\theta$, $\xi_2 = \frac{\theta}{2}$, $\eta_1^+ = \eta_3^- = \frac{\theta}{2}$ for arbitrary real θ and $\eta_i^+ = \eta_i^- = 0$ otherwise solves (15), but not (6). Thus the 1-strong design matrix $\mathbf{X}$ is 0-stable. □

From the definition of s it follows that $z(U, L) \geq 0$ for all $(U, L, Z) \in \mathcal{F}_s$.

PROPOSITION 3. *If* $\mathbf{X}$ *is s-stable, then* $s \le n-p$ *and* $\mathbf{X}$ *is q-strong with* $q \ge s$.

Proof. For the proof, see the above discussion and Theorem 2. $\square$

As the last example shows, the inequality $q \ge s$ in Proposition 3 can be sharp.

PROPOSITION 4. *If* $\mathbf{X}$ *is s-stable, then* $\alpha(\ell_1, \mathbf{y}|\mathbf{X}) \le \frac{s+1}{n}$.

Proof. Suppose that $\mathbf{X}$ is q-strong for some $q \ge 0$. Hence $q \ge s$. Thus if $s = q$, the assertion follows from Proposition 1. Suppose $q > s$. Then there exists $(U, L, Z) \in \mathcal{F}_{s+1}$ such that (15) is solvable. Let $\boldsymbol{\xi}, \boldsymbol{\eta}^+, \boldsymbol{\eta}^-$ be any solution to (15) with $\boldsymbol{\xi} \ne \mathbf{0}$. Define $\mathbf{g} \in \mathbb{R}^n$ by $g_j = \mathbf{x}^j\boldsymbol{\xi}$ for all $j \in U \cup L$, $g_j = 0$ otherwise, and consider the linear program (10). By Lemma 1 there exists some finite $\vartheta_0 \ge 0$ such that (12) holds for all $\vartheta \ge \vartheta_0$. Since $\mathbf{X}$ is q-strong with $q \ge s+1$, it follows, as in the proof of Proposition 2, that (14) holds. Define $\boldsymbol{\beta}(\vartheta) = \boldsymbol{\beta}(\vartheta_0) + (\vartheta - \vartheta_0)\boldsymbol{\xi}$ and

$$\mathbf{r}_Z^+(\vartheta) = \mathbf{r}_Z^+(\vartheta_0) + (\vartheta - \vartheta_0)\boldsymbol{\eta}^+, \quad \mathbf{r}_Z^-(\vartheta) = \mathbf{r}_Z^-(\vartheta_0) + (\vartheta - \vartheta_0)\boldsymbol{\eta}^-, \quad \mathbf{r}_{U\cup L}^\pm(\vartheta) = \mathbf{r}_{U\cup L}^\pm(\vartheta_0).$$

It follows that $(\boldsymbol{\beta}(\vartheta), \mathbf{r}^+(\vartheta), \mathbf{r}^-(\vartheta))$ is feasible for (10) and, since by duality

$$z(\vartheta) \le \mathbf{e}^T(\mathbf{r}^+(\vartheta) + \mathbf{r}^-(\vartheta)) = z(\vartheta_0) + (\vartheta - \vartheta_0)\mathbf{e}_Z^T(\boldsymbol{\eta}^+ + \boldsymbol{\eta}^-) \le z(\vartheta_0) + (\vartheta - \vartheta_0)\|\mathbf{g}\|_1 = z(\vartheta),$$

it is optimal for (10). But $\lim_{\vartheta \to +\infty} \|\boldsymbol{\beta}(\vartheta)\| \to +\infty$, since $\boldsymbol{\xi} \ne \mathbf{0}$, and thus $\alpha(\ell_1, \mathbf{y}|\mathbf{X}) \le \frac{s+1}{n}$. $\square$

THEOREM 3. *A design matrix* $\mathbf{X}$ *is s-stable if and only if* $\alpha(\ell_1, \mathbf{y}|\mathbf{X}) = \frac{s+1}{n}$.

Proof. Let $\mathbf{X}$ be s-stable and suppose that $\alpha(\ell_1, \mathbf{y}|\mathbf{X}) \le \frac{s}{n}$. Then there exists $\mathbf{g} \in \mathbb{R}^n$ with s nonzero components such that (10) has an optimal solution $(\boldsymbol{\beta}(\vartheta), \boldsymbol{r}^+(\vartheta), \boldsymbol{r}^-(\vartheta))$ with $\|\boldsymbol{\beta}(\vartheta)\|_1 \to +\infty$ for $\vartheta \to +\infty$. By Lemma 1 there exists a finite $\vartheta_0 \ge 0$ such that $z(\vartheta) = z(\vartheta_0) + (\vartheta - \vartheta_0)\mathbf{u}^0\mathbf{g}$ for all $\vartheta \ge \vartheta_0$ and some extreme point $\mathbf{u}^0$ of $\mathbf{P}^*$. Since $\mathbf{X}$ is q-strong with $q \ge s$ it follows as before that (14) holds. By [4, Proposition 2] there exists $B \subseteq N$ with $|B| = p$ such that $\mathbf{X}_B\boldsymbol{\beta}(\vartheta) = \mathbf{y}_B + \vartheta\mathbf{g}_B$ with $\mathbf{X}_B$ nonsingular, and thus

$$\boldsymbol{\beta}(\vartheta) = \mathbf{X}_B^{-1}\mathbf{y}_B + \vartheta\mathbf{X}_B^{-1}\mathbf{g}_B = \boldsymbol{\beta}(\vartheta_0) + (\vartheta - \vartheta_0)\boldsymbol{\xi}$$

for all $\vartheta \ge \vartheta_0$, where $\boldsymbol{\xi} = \mathbf{X}_B^{-1}\mathbf{g}_B \ne \mathbf{0}$ because $\|\beta(\vartheta)\|_1 \to +\infty$. It follows that $\mathbf{r}^\pm(\vartheta) = \mathbf{r}^\pm(\vartheta_0) + (\vartheta - \vartheta_0)\boldsymbol{\eta}^\pm$ for all $\vartheta \ge \vartheta_0$, where $\boldsymbol{\eta}^+ = \max\{\mathbf{0}, \mathbf{g} - \mathbf{X}\boldsymbol{\xi}\}$ and $\boldsymbol{\eta}^- = -\min\{\mathbf{0}, \mathbf{g} - \mathbf{X}\boldsymbol{\xi}\}$, respectively. Since $\mathbf{X}\boldsymbol{\beta}(\vartheta) + \mathbf{r}^+(\vartheta) - \mathbf{r}^-(\vartheta) = \mathbf{y} + \vartheta\mathbf{g}$ for all $\vartheta \ge 0$, we calculate

$$\mathbf{X}\boldsymbol{\beta}(\vartheta_0) + \mathbf{r}^+(\vartheta_0) - \mathbf{r}^-(\vartheta_0) + (\vartheta - \vartheta_0)(\mathbf{X}\boldsymbol{\xi} + \boldsymbol{\eta}^+ - \boldsymbol{\eta}^-) = \mathbf{y} + \vartheta_0\mathbf{g} + (\vartheta - \vartheta_0)\mathbf{g}.$$

Consequently, $\mathbf{X}\boldsymbol{\xi} + \boldsymbol{\eta}^+ - \boldsymbol{\eta}^- = \mathbf{g}$. Define $(U, L, Z) \in \mathcal{F}_s$ by

$$Z = \{i \in N : g_i = 0\}, \quad U = \{i \in N - Z : g_i > 0\}, \quad L = \{i \in N - Z : g_i < 0\}.$$

From $\mathbf{X}_U\boldsymbol{\xi} + \boldsymbol{\eta}_U^+ - \boldsymbol{\eta}_U^- = \mathbf{g}_U$ and $\mathbf{X}_L\boldsymbol{\xi} + \boldsymbol{\eta}_L^+ - \boldsymbol{\eta}_L^- = \mathbf{g}_L$, we calculate

$$\left(\mathbf{e}_U^T\mathbf{X}_U - \mathbf{e}_L^T\mathbf{X}_L\right)\boldsymbol{\xi} + \mathbf{e}_U^T\boldsymbol{\eta}_U^+ + \mathbf{e}_L^T\boldsymbol{\eta}_L^- - \left(\mathbf{e}_U^T\boldsymbol{\eta}_U^- + \mathbf{e}_L^T\boldsymbol{\eta}_L^+\right) = \|\mathbf{g}\|_1.$$

Calculating the optimal objective function value of (10) from the primal solution, we find

$$z(\vartheta) = \mathbf{e}_N^T(\mathbf{r}^+(\vartheta) + \mathbf{r}^-(\vartheta)) = z(\vartheta_0) + (\vartheta - \vartheta_0)\mathbf{e}_N^T(\boldsymbol{\eta}^+ + \boldsymbol{\eta}^-) = z(\vartheta_0) + (\vartheta - \vartheta_0)\|\mathbf{g}\|_1.$$

Thus $\mathbf{e}_N^T(\boldsymbol{\eta}^+ + \boldsymbol{\eta}^-) = \|\mathbf{g}\|_1$ and, from the previous relation for $\|\mathbf{g}\|_1$, we get

$$\left(-\mathbf{e}_U^T\mathbf{X}_U + \mathbf{e}_L^T\mathbf{X}_L\right)\boldsymbol{\xi} + \mathbf{e}_Z^T\left(\boldsymbol{\eta}_Z^+ + \boldsymbol{\eta}_Z^-\right) = -2\left(\mathbf{e}_U^T\boldsymbol{\eta}_U^- + \mathbf{e}_L^T\boldsymbol{\eta}_L^+\right) \le 0.$$

Hence (15) is solvable for $(U, L, Z) \in \mathcal{F}_s$, which is a contradiction. Thus $\alpha(\ell_1, \mathbf{y}|\mathbf{X}) > \frac{s}{n}$, and by Proposition 4 we have equality. Suppose $\alpha(\ell_1, \mathbf{y}|\mathbf{X}) = \frac{s+1}{n}$. By definition, s is the smallest integer with this property. Suppose $\mathbf{X}$ is t-stable for some integer $t \geq 0$. By the first part of this theorem, $\alpha(\ell_1, \mathbf{y}|\mathbf{X}) = \frac{t+1}{n}$, and thus $s = t$. ∎

Proposition 2 shows that for q-strong $\mathbf{X}$ there exist optimal ℓ_1-regression coefficients such that the bias remains bounded when at most q data of $\mathbf{y}$ are contaminated. It is thus debatable whether or not we want to talk of a "breakdown" of ℓ_1-regression in this case. By Proposition 3 we know $q \geq s$, and from Example 2 we know that $q > s$ is possible. The question becomes whether or not the difference $q - s$ is "reasonably" small. Under the assumption that the data $\mathbf{X}$ are *in general position* we can answer the question affirmatively. ($\mathbf{X}$ is in general position if every $p \times p$ submatrix of $\mathbf{X}$ is nonsingular.)

PROPOSITION 5. *If* $\mathbf{X}$ *is in general position, q-strong, and s-stable, then* $0 \leq q - s \leq 1$.

Proof. Suppose that $q > s$. Then there exist $(U, L, Z) \in \mathcal{F}_{s+1}$ such that (15) has a solution $(\boldsymbol{\xi}, \boldsymbol{\eta}_Z^+, \boldsymbol{\eta}_Z^+)$. We claim that $\eta_i^+ + \eta_i^- > 0$ for some $i \in Z$. Otherwise, $\mathbf{X}_Z\boldsymbol{\xi} = \mathbf{0}$ with $\boldsymbol{\xi} \neq \mathbf{0}$ implies that $r(\mathbf{X_Z}) < p$. But, by Proposition 1, $|Z| = n-(s+1) \geq n-q \geq p$, and thus $r(\mathbf{X_Z}) = p$ since $\mathbf{X}$ is in general position. The claim follows. Assume that $\eta_i^+ > 0$ for some $i \in Z$. Let $Z^* = Z - i$, $U^* = U$, and $L^* = L + i$. We calculate

$$\begin{aligned}
&\left(-\mathbf{e}_{U^*}^T\mathbf{X}_{U^*} + \mathbf{e}_{L^*}^T\mathbf{X}_{L^*}\right)\boldsymbol{\xi} + \sum_{i\in Z^*}\left(\eta_i^+ + \eta_i^-\right) \\
&\quad = \left(-\mathbf{e}_U^T\mathbf{X}_U + \mathbf{e}_L^T\mathbf{X}_L\right)\boldsymbol{\xi} + \sum_{i\in Z}\left(\eta_i^+ + \eta_i^-\right) + \mathbf{x}^i\boldsymbol{\xi} - \eta_i^+ - \eta_i^- \leq -2\eta_i^+ < 0
\end{aligned}$$

because $\mathbf{x}^i\boldsymbol{\xi} + \eta_i^+ - \eta_i^- = 0$. Hence (6) has a solution for $(U^*, L^*, Z^*) \in \mathbf{F}_{s+2}$, and thus by Farkas's lemma the corresponding (13) is not solvable, i.e., $q < s+2$. If $\eta_i^- > 0$ for some $i \in Z$, we let $Z^* = Z - i$, $U^* = U + i$, and $L^* = L$ and calculate likewise. ∎

2.3. Related work on $\boldsymbol{\alpha(\ell_1, \mathbf{y}|\mathbf{X})}$. To summarize previous results on the determination of the finite sample breakdown point of ℓ_1-regression, let m_* be the largest integer such that for all $S \subseteq N$ with $|S| = m_*$

$$\inf_{\|\boldsymbol{\xi}\|=1} \frac{\sum_{i\in N-S}|\mathbf{x}^i\boldsymbol{\xi}|}{\sum_{i\in N}|\mathbf{x}^i\boldsymbol{\xi}|} > \frac{1}{2}. \tag{16}$$

He et al. [7, Theorem 5.2] show $(m_* + 1)/n \leq \alpha(\ell_1, \mathbf{y}|\mathbf{X}) \leq (m_* + 2)/n$. Mizera and Müller [11] prove that

$$\alpha(\ell_1, \mathbf{y}|\mathbf{X}) = (m_* + 1)/n. \tag{17}$$

We show that Theorem 3 is consistent with (17). From (16) it follows that

$$-\sum_{i\in S}|\mathbf{x}^i\boldsymbol{\xi}| + \sum_{i\in N-S}|\mathbf{x}^i\boldsymbol{\xi}| > 0 \quad \text{for all} \quad \boldsymbol{\xi} \in \mathbb{R}^p \quad \text{with} \quad \|\boldsymbol{\xi}\| > \mathbf{0}.$$

Hence $\sum_{i\in S}|\mathbf{x}^i\boldsymbol{\xi}| \geq \sum_{i\in U}\mathbf{x}^i\boldsymbol{\xi} - \sum_{i\in L}\mathbf{x}^i\boldsymbol{\xi}$, where $U, L \subseteq S$ with $L \cap U = \emptyset$ and $L \cup U = S$ are arbitrary. Letting $Z = N - S$ we get, from the previous inequality,

$$-\sum_{i\in U}\mathbf{x}^i\boldsymbol{\xi} + \sum_{i\in L}\mathbf{x}^i\boldsymbol{\xi} + \sum_{i\in Z}|\mathbf{x}^i\boldsymbol{\xi}| > 0 \tag{18}$$

for all $\boldsymbol{\xi} \in \mathbb{R}^p$ with $\|\boldsymbol{\xi}\| > 0$ and U and L as specified. Let $\boldsymbol{\eta}^+, \boldsymbol{\eta}^- \in \mathbb{R}^{|Z|}$ satisfy $\boldsymbol{\eta}^+, \boldsymbol{\eta}^- \geq \mathbf{0}$ and $\mathbf{X}_Z\boldsymbol{\xi} + \boldsymbol{\eta}^+ - \boldsymbol{\eta}^- = \mathbf{0}$ for some $\boldsymbol{\xi} \in \mathbb{R}^p$. It follows that

$$\eta_i^+ + \eta_i^- \geq |\mathbf{x}^i\boldsymbol{\xi}| \qquad \text{for } i \in Z \tag{19}$$

because $\mathbf{x}^i\boldsymbol{\xi} + \eta_i^+ - \eta_i^- = 0$ and $\eta_i^+ \geq 0$, $\eta_i^- \geq 0$ imply $\eta_i^+ \geq -\mathbf{x}^i\boldsymbol{\xi}$ and $\eta_i^- \geq \mathbf{x}^i\boldsymbol{\xi}$. Thus if $\mathbf{x}^i\boldsymbol{\xi} \leq 0$, then $\eta_i^+ \geq |\mathbf{x}^i\boldsymbol{\xi}|$, and if $\mathbf{x}^i\boldsymbol{\xi} > 0$, then $\eta_i^- \geq |\mathbf{x}^i\boldsymbol{\xi}|$. Equation (19) follows. From (18) and (19),

$$\left(-\mathbf{e}_U^T\mathbf{X}_U + \mathbf{e}_L^T\mathbf{X}_L\right)\boldsymbol{\xi} + \mathbf{e}_Z^T\left(\boldsymbol{\eta}^+ + \boldsymbol{\eta}^-\right) > 0$$

for all $\boldsymbol{\xi} \in \mathbb{R}^p$ with $\|\boldsymbol{\xi}\| > 0$ and U and L as specified. Consequently, since S is any subset of N with $|S| = m_*$, (15) has no solution for any $(U, L, Z) \in \mathcal{F}_{m_*}$. Hence $\mathbf{X}$ is s-stable with $s \geq m_*$. By the definition of m_*, there exists $S \subseteq N$ with $|S| = m_* + 1$ such that

$$\inf_{\|\boldsymbol{\xi}\|=1} \frac{\sum_{i\in N-S} |\mathbf{x}^i\boldsymbol{\xi}|}{\sum_{i\in N} |\mathbf{x}^i\boldsymbol{\xi}|} \leq \frac{1}{2}.$$

Hence $-\sum_{i\in S} |\mathbf{x}^i\boldsymbol{\xi}^0| + \sum_{i\in N-S} |\mathbf{x}^i\boldsymbol{\xi}^0| \leq 0$ for some $\boldsymbol{\xi}^0 \neq \mathbf{0}$. Define $U = \{i \in S : \mathbf{x}^i\boldsymbol{\xi}^0 \geq 0\}$ and $L = \{i \in S : \mathbf{x}^i\boldsymbol{\xi}^0 < 0\}$. Then $\sum_{i\in S} |\mathbf{x}^i\boldsymbol{\xi}^0| = (\mathbf{e}_U^T\mathbf{X}_U - \mathbf{e}_L^T\mathbf{X}_L)\boldsymbol{\xi}^0$. Let $\eta_i^+ = \max\{0, \mathbf{x}^i\boldsymbol{\xi}^0\}$ and $\eta_i^- = -\min\{0, \mathbf{x}^i\boldsymbol{\xi}^0\}$ for all $i \in Z = N - S$. It follows that $(\boldsymbol{\xi}^0, \boldsymbol{\eta}_Z^+, \boldsymbol{\eta}_Z^-)$ solves (15), and thus $s < m_* + 1$. Consequently, $s = m_*$ and the results agree even though the proof methodologies employed are quite different.

Mizera and Müller [12] provide an enumerative algorithm for the computation of the conditional breakdown point as follows. $m_* = |S|$ is the largest integer such that (16) holds for all $S \subseteq N$ with $|S| = m_*$ if and only if $|E| = m_* + 1$, where $m_* + 1$ is the smallest integer such that there exists an $E \subseteq N$ such that

$$\max_{\|\boldsymbol{\xi}\|=1} \frac{\sum_{i\in E} |\mathbf{x}^i\boldsymbol{\xi}|}{\sum_{i\in N} |\mathbf{x}^i\boldsymbol{\xi}|} \geq \frac{1}{2}. \tag{20}$$

Note that the restriction $\|\boldsymbol{\xi}\| = 1$ can be dropped. They then show that in order to calculate $m_* + 1$ it is sufficient to compare at most $\binom{n}{p-1}$ candidate solutions for $\boldsymbol{\xi}$ in (20).

3. Calculating the q-strength and s-stability of a design matrix.

To calculate the q-strength and s-stability of a design matrix $\mathbf{X}$, there are two roads to take. The first is to find the q-strength or the s-stability of $\mathbf{X}$ by enumeration. The second is to formulate an MIP and solve the program. Mizera and Müller [12] provide a "special purpose" enumerative algorithm. Here we use the results of section 2 to formulate the problem as an MIP. Thus, in order to calculate, a user need only generate the corresponding constraint set for the data of his/her design matrix $\mathbf{X}$ (a useful exercise for a graduate student) and solve the problem using some commercially available MIP solver, such as CPLEX.

By Theorem 2, $\mathbf{X}$ is not q-strong if for some $(U, L, Z) \in \mathcal{F}_q$ there exist $\boldsymbol{\xi} \in \mathbb{R}^p$, $\boldsymbol{\eta}^+, \boldsymbol{\eta}^- \in \mathbb{R}^{|Z|}$ satisfying (6). Thus the problem of determining the q-strength of $\mathbf{X}$ consists of finding the smallest integer such that (6) is solvable for some $(U, L, Z) \in \mathcal{F}_q$.

We claim that the following MIP, called MIP1, does just that:

$$\min \sum_{i=1}^{n} u_i + \ell_i$$

$$\text{s.t.}\quad \mathbf{x}^i\boldsymbol{\xi} + \eta_i^+ - \eta_i^- + s_i - t_i = 0 \quad \text{for } i = 1, \dots, n, \tag{21}$$

$$s_i - Mu_i \le 0, \quad s_i + Mu_i \ge 0 \quad \text{for } i = 1, \dots, n, \tag{22}$$

$$t_i - M\ell_i \le 0, \quad t_i + M\ell_i \ge 0 \quad \text{for } i = 1, \dots, n, \tag{23}$$

$$\eta_i^+ + \eta_i^- + Mu_i + M\ell_i \le M \quad \text{for } i = 1, \dots, n, \tag{24}$$

$$u_i + \ell_i \le 1 \quad \text{for } i = 1, \dots, n, \tag{25}$$

$$\sum_{i=1}^{n} s_i + t_i + \eta_i^+ + \eta_i^- \le -\varepsilon, \tag{26}$$

$$\boldsymbol{\xi}, \mathbf{s}, \mathbf{t} \text{ free}, \ \boldsymbol{\eta}^+ \ge \mathbf{0}, \ \boldsymbol{\eta}^- \ge \mathbf{0}, \quad u_i, \ell_i \in \{0, 1\} \quad \text{for } i = 1, \dots, n. \tag{27}$$

We assume that $M > 0$ is a suitably chosen large number and $\varepsilon > 0$ a small number so that the constraints (22)–(24) corresponding to the ith observation are *nonbinding* for the solution that results if we set u_i or ℓ_i equal to 1 or $u_i = \ell_i = 0$. It can be shown by standard arguments of mixed-integer programming that such M and ε exist. Specifically, let U be the set of indices in which $u_i = 1$ in any solution to MIP1, let L be given by $\ell_i = 1$, and $Z = N - U - L$. By (25), $L \cap U = \emptyset$ and thus U, L, Z is a three-way partition of N. Moreover, to every such three-way partition of N there corresponds some setting of u_i and ℓ_i equal to 0 or 1 with $u_i + \ell_i \le 1$. Equations (22)–(24) constrain s_i, t_i, η_i^+, and η_i^- such that if $u_i = 0$, then $s_i = 0$; if $\ell_i = 0$, then $t_i = 0$; and if $u_i + \ell_i = 1$, then $\eta_i^+ = \eta_i^- = 0$. It follows that the constraints of MIP1 produce the following system of constraints:

$$\mathbf{X}_Z\boldsymbol{\xi} + \boldsymbol{\eta}_Z^+ - \boldsymbol{\eta}_Z^- = \mathbf{0}, \tag{28}$$

$$\mathbf{X}_U\boldsymbol{\xi} + \mathbf{s}_U = \mathbf{0}, \qquad \mathbf{X}_L\boldsymbol{\xi} - \mathbf{t}_L = \mathbf{0}, \tag{29}$$

$$\sum_{i \in U} s_i + \sum_{i \in L} t_i + \sum_{i \in Z} \eta_i^+ + \eta_i^- \le -\varepsilon \tag{30}$$

since (22)–(24) are redundant. By (29) and (30),

$$0 > -\varepsilon \ge \sum_{i \in U} s_i + \sum_{i \in L} t_i + \sum_{i \in Z} \eta_i^+ + \eta_i^- = \left(-\mathbf{e}_U^T\mathbf{X}_U + \mathbf{e}_L^T\mathbf{X}_L\right)\boldsymbol{\xi} + \mathbf{e}_Z^T\boldsymbol{\eta}_Z^+ + \mathbf{e}_Z^T\boldsymbol{\eta}_Z^-.$$

Thus, (6) is satisfied. Since MIP1 minimizes $\sum_{i=1}^{n} u_i + \ell_i = |L \cup U|$, it determines the smallest integer k such that $\mathbf{X}$ is not k-strong. Therefore, $\mathbf{X}$ is q-strong with $q = k - 1$.

MIP1 calculates the q-strength of a design matrix $\mathbf{X}$. However, if the size of $\mathbf{X}$ is large, e.g., if $n \ge 100$, then even with today's powerful MIP solvers the calculation may take too much time. Thus, we now provide guidelines for a heuristic to determine a good upper bound of the q-strength and the s-stability of a large design matrix.

Although solving a large MIP *exactly* may take a long time, finding a *feasible* solution is often much easier and can be accomplished quite quickly. Determining a feasible solution to MIP1 provides an upper bound Q on the q-strength of a design matrix. This upper bound can be improved upon if some subset $S \subset N$ exists where $|S| < Q$, $L, U \subset S$ with $L \cup U = S$, $L \cap U = \emptyset$, and $Z = N - S$ such that no solution exists to (13). In such a case, the design matrix in question is at most ($|S|-1$)-strong. Furthermore, from preliminary computational results, when a matrix is not $|S|$-strong, it is often easy to find a subset $S \subset N$ such that no solution exists to (13).

HEURISTIC UPPER BOUND Q.

Step 1. Input MIP1 to a commercially available package such as CPLEX. Find some feasible (not necessarily optimal) solution to MIP1 with objective function value Q.

Step 2. Randomly select r subsets, $S \subset N$, of size $Q-1$. Set $i=1$.

Step 3. Randomly select p partitions of S_i, where $L_i \cup U_i = S_i$. Set $j=1$.

Step 4. For the jth partition of S_i with $Z = N - S_i$, solve a linear program corresponding to (13). If a solution exists, replace j with $j+1$. Otherwise, replace Q with $Q-1$ and goto Step 2.

Step 5. If $j > p$, replace i with $i+1$. Otherwise, goto Step 4.

Step 6. If $i \le r$, goto Step 3. Otherwise, Q is an upper bound for the q-strength of the design matrix in question and stop.

By Theorem 3, the finite sample breakdown point of ℓ_1-regression equals $\frac{s+1}{n}$, where the design matrix with n rows is s-stable. Although q-strength provides a robustness measure by itself, calculating the q-strength does not guarantee the exact value of the breakdown point. We now provide another mixed-integer linear program that calculates the s-stability and thus the breakdown point of ℓ_1-regression. This MIP, called MIP2, is similar to MIP1 except that here we calculate the smallest value of $|U \cup L|$ such that (15) is solvable.

$$\min \sum_{i=1}^{n} u_i + \ell_i$$

$$\text{s.t.}\quad \mathbf{x}^i \boldsymbol{\xi} + \eta_i^+ - \eta_i^- + s_i - t_i = 0 \quad \text{for } i = 1, \ldots, n, \tag{31}$$

$$s_i - M u_i \le 0, \quad t_i - M\ell_i \le 0 \quad \text{for } i = 1, \ldots, n, \tag{32}$$

$$\eta_i^+ + \eta_i^- + M u_i + M \ell_i \le M \quad \text{for } i = 1, \ldots, n, \tag{33}$$

$$u_i + \ell_i \le 1 \quad \text{for } i = 1, \ldots, n, \tag{34}$$

$$\sum_{i=1}^{n} \eta_i^+ + \eta_i^- - s_i - t_i \le 0, \qquad \sum_{i=1}^{n} s_i + t_i \ge \varepsilon, \tag{35}$$

$$\boldsymbol{\xi} \text{ free}, \boldsymbol{\eta}^+ \ge \mathbf{0}, \boldsymbol{\eta}^- \ge \mathbf{0}, \mathbf{s} \ge \mathbf{0}, \mathbf{t} \ge \mathbf{0}, \quad u_i, \ell_i \in \{0,1\} \quad \text{for } i = 1, \ldots, n. \tag{36}$$

As in MIP1, we assume that M is a suitably chosen large number and ε a small number so that (32) and (33) are nonbinding for the solution that results if we set u_i or ℓ_i equal to 1 or $u_i = \ell_i = 0$. The existence of such M and ε can be shown as in the case of MIP1. As in the case of MIP1, we argue that to every three-way partition of N there corresponds a feasible solution to MIP2 and vice versa. Equations (32)–(33) constrain s_i, t_i, η_i^+, and η_i^- such that if $u_i = 0$, then $s_i = 0$; if $\ell_i = 0$, then $t_i = 0$; and if $u_i + \ell_i = 1$, then $\eta_i^+ = \eta_i^- = 0$. It follows that every feasible assignment of the zero-one variables of MIP2 reduces MIP2 to the following constraints:

$$\mathbf{X}_Z \boldsymbol{\xi} + \boldsymbol{\eta}_Z^+ - \boldsymbol{\eta}_Z^- = \mathbf{0}, \tag{37}$$

$$\mathbf{X}_U \boldsymbol{\xi} + \mathbf{s}_U = \mathbf{0}, \qquad \mathbf{X}_L \boldsymbol{\xi} - \mathbf{t}_L = \mathbf{0}, \tag{38}$$

$$-\sum_{i \in U} s_i - \sum_{i \in L} t_i + \sum_{i \in Z} \eta_i^+ + \eta_i^- \le 0, \tag{39}$$

TABLE 1
Breakdown points $\alpha(\ell_1, \mathbf{y}|\mathbf{X})$ for ℓ_1-regression.

Data set	n	p	q-strength	s-stability	Breakdown
Stackloss	21	4	3	3	$\frac{4}{21}$
Aircraft	23	5	1	1	$\frac{2}{23}$
Delivery	25	3	2	2	$\frac{3}{25}$
Engine	16	5	1	1	$\frac{2}{16}$
Gessel	21	2	2	2	$\frac{3}{21}$
Salinity	28	4	3	3	$\frac{4}{28}$
Telephone	24	2	6	5	$\frac{6}{24}$
Wood	20	6	2	2	$\frac{3}{20}$
Star	47	2	4	4	$\frac{5}{47}$

where $U = \{i \in N : u_i = 1\}$ and $L = \{i \in N : \ell_i = 1\}$. By (38) and (39),

$$0 \geq -\sum_{i \in U} s_i - \sum_{i \in L} t_i + \sum_{i \in Z} \eta_i^+ + \eta_i^- = \left(-\mathbf{e}_U^T \mathbf{X}_U + \mathbf{e}_L^T \mathbf{X}_L\right) \boldsymbol{\xi} + \mathbf{e}_Z^T \boldsymbol{\eta}_Z^+ + \mathbf{e}_Z^T \boldsymbol{\eta}_Z^-.$$

Since $\sum_{i=1}^n s_i + t_i \geq \varepsilon > 0$, it follows that either $s_i > 0$, $t_i = 0$, and $\eta^+ = \eta_i^- = 0$ or $s_i = 0$, $t_i > 0$, and $\eta^+ = \eta_i^- = 0$ for some $i \in N$. Consequently, from (31), $\mathbf{x}^i \boldsymbol{\xi} = -s_i + t_i \neq 0$, and thus $\boldsymbol{\xi} \neq \mathbf{0}$. Hence, (15) is satisfied. On the other hand, if (15) is solvable, then a feasible solution to MIP2 is readily constructed. So the formulation MIP2 does the job. Since MIP2 minimizes $\sum_{i=1}^n u_i + \ell_i = |L \cup U|$, it determines the smallest integer k such that $\mathbf{X}$ is not k-stable. Therefore, $\mathbf{X}$ is s-stable with $s = k-1$.

To demonstrate the usefulness of our approach, we have calculated the q-strength and the s-stability of the design matrices for nine data sets from the robust regression literature. All of these data sets can be found in [14], except for the engine data set which can be found in [10, p. 529]. These results are listed in Table 1. In all but one data set, $s = q$ and the breakdown point equals $\frac{q+1}{n}$. The results were obtained by solving the corresponding MIPs by the commercially available CPLEX package on a Pentium 4 (2.26 GHz) processor. The median solution time and the median number of nodes in the branch and bound tree of the nine MIPs for the s-stability were 25.09 seconds and 22164, respectively. The median time spent and number of nodes traversed in order to find the optimal solution were 6.19 seconds and 10779, respectively. On the other hand, the median solution time and the median number of nodes in the branch and bound tree of the nine MIPs for the q-strength were 1.47 seconds and 1169, respectively. The median time spent and number of nodes traversed in order to find the optimal solution were .26 seconds and 405, respectively.

3.1. The telephone data. The telephone data set is data on the number of international phone calls from Belgium per year versus the years 1950 through 1973. The data has outliers in the values of the dependent variable (calls), in particular the data for the calls in the years 1964 through 1969. These outliers were due to the recording of the number of international phone *call minutes* from Belgium as opposed to the number of calls. Figure 1 contains four graphs of versions of this data set, each with a fitted ℓ_1-regression line. The graph in the upper left-hand corner is the original data set with the outliers as described above. Notice that the ℓ_1-regression line is hardly influenced by the outliers. To further demonstrate this, the graph in the upper right-hand corner contains an ℓ_1-fit to data that is most likely quite similar to

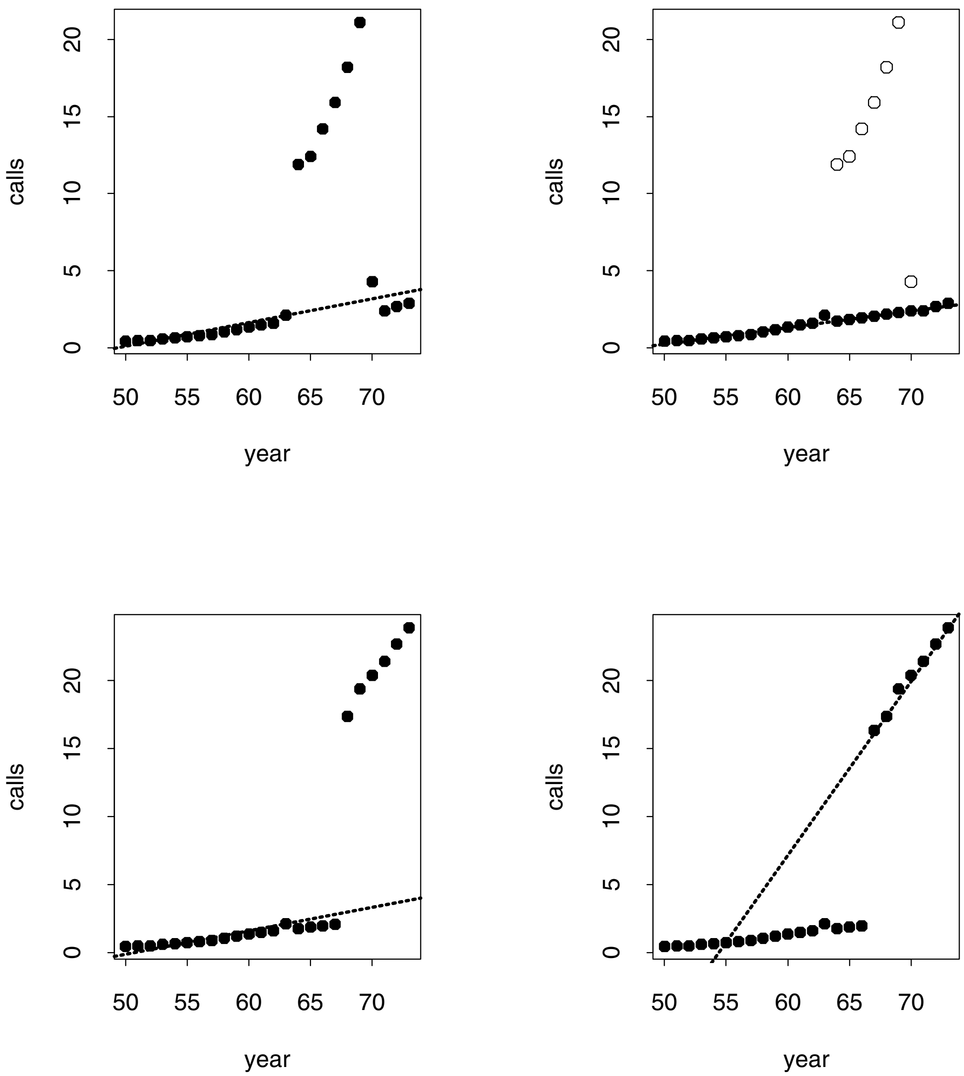

FIG. 1. *ℓ_1-regression lines for telephone data.*

the "uncontaminated" data set. The estimated observations used for the years 1964 through 1970 are determined from a least squares fit to the data excluding the years 1963 through 1970. The original outliers are plotted by empty circles but are not taken into consideration when solving for the ℓ_1-regression line. From Table 1, the q-strength for this problem is 6, its s-stability is 5, and thus $\alpha(\ell_1, \mathbf{X}|\mathbf{y}) = \frac{6}{24}$; i.e., some set of six contaminated observations may cause the ℓ_1-regression estimator to break down in this example. However, by Proposition 2 and the fact that the q-strength is 6, there exist bounded ℓ_1-regression coefficients for every set of six contaminated observations, but not for every set of seven contaminated observations. The graph in the bottom left-hand corner shows the uncontaminated data set with six new outliers for the years 1968 through 1973. The ℓ_1-regression estimator performs quite well with these six outliers in the sense that ℓ_1-regression coefficients exist that are hardly influenced by the outliers. However, for just one more contaminated observation, there are no longer ℓ_1-regression coefficients that are not influenced by the outliers; see the graph in the bottom right corner. In our minds this points to the suitability of the q-strength of a design matrix as a measure of the breakdown of ℓ_1-regression. In any case, by Proposition 5, q-strength and s-stability are, in general, reasonably close to each other.

4. Conclusion. We have provided an exact formula for the breakdown point $\alpha(\ell_1, \mathbf{y}|\mathbf{X})$ of ℓ_1-regression with contamination restricted to the dependent variable. This is done using the notion of the s-stability of the design matrix $\mathbf{X}$, which is introduced here. We have shown that our results agree with results known in the literature. We have also introduced the notion of q-strength, a new robustness measure for ℓ_1-regression. Most important, we have shown that one can indeed calculate the conditional breakdown point of ℓ_1-regression by solving an appropriate MIP. We give computational experiments using the proposed approach for nine data sets from the robust regression literature. For large data sets, we provide a heuristic that provides an upper bound on the q-strength of a design matrix. This is important in the design of robustly planned experiments, as it provides a computable assessment of the vulnerability of the experiment's design to errors in the measurement on the dependent variable. Finally, we provide an illustrative example to demonstrate the difference between q-strength and s-stability of design matrices.

REFERENCES

[1] D. Alevras and M. Padberg, *Linear Optimization and Extensions: Problems and Solutions*, Springer-Verlag, Berlin, 2001.

[2] D. L. Donoho and P. J. Huber, *The notion of breakdown point*, in A Festschrift for Erich Lehmann, P. Bickel, K. Doksum, and J. L. Hodges, eds., Wadsworth, Belmont, CA, 1983, pp. 157–184.

[3] A. Giloni, *Essays on Optimization in Data Analysis and Operations Management*, Ph.D. Thesis, Stern School of Business, New York University, New York, NY, 2000.

[4] A. Giloni and M. Padberg, *Alternative methods of linear regression*, Math. Comput. Modelling, 35 (2002), pp. 361–374.

[5] F. R. Hampel, *Contributions to the Theory of Robust Estimation*, Ph.D. Thesis, University of California, Berkeley, CA, 1968.

[6] S. P. Ellis and S. Morgenthaler, *Leverage and breakdown in l_1-regression*, J. Amer. Statist. Assoc., 87 (1993), pp. 143–148.

[7] X. He, J. Jureckova, R. Koenker, and S. Portnoy, *Tail behavior of regression estimators and their breakdown points*, Econometrica, 58 (1990), pp. 1195–1214.

[8] J. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms* I, Springer-Verlag, Berlin, 1993.

[9] R. Koenker and G. Bassett, *Regression quantiles*, Econometrica, 46 (1978), pp. 33–50.

[10] R. L. Mason, R. F. Gunst, and J. L. Hess, *Statistical Design and Analysis of Experiments*, John Wiley, New York, 1989.

[11] I. Mizera and C. H. Müller, *Breakdown points and variation exponents of robust M-estimators in linear models*, Ann. Statist., 27 (1999), pp. 1164–1177.

[12] I. Mizera and C. H. Müller, *The influence of the design on the breakdown point of ℓ_1-type M-estimators*, in MODA6—Advances in Model-Oriented Design and Analysis, A. Atkinson, P. Hackl, and W. Müller, eds., Physica-Verlag, Heidelberg, 2001, pp. 193–200.

[13] M. Padberg, *Linear Optimization and Extensions*, Springer-Verlag, Berlin, 1995.

[14] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*, Wiley, New York, 1987.