한국통계학회논문집 제2권 2호
1995년 12월 pp. 277-287

On the Robustness of L_1-estimator in Linear Regression Models

Bu-Yong Kim[1]

Abstract

It is well known that the L_1-estimator is robust with respect to vertical outliers in regression data, even if it is susceptible to bad leverage points. This article is concerned with the robustness of the L_1-estimator. To investigate its robustness against vertical outliers we may find intervals for the value of the response variable within which the L_1-estimates do not change. A procedure for constructing those intervals in multiple linear regression is illustrated in the sensitivity analysis context. And then vertical breakdown point of the L_1-estimator is defined on the basis of properties related to those intervals.

1. Introduction

Consider the problem of estimating the parameters of a multiple linear regression model

$$y = X\beta + \varepsilon \ ,$$

where y denotes an n-vector of response variable, X an $n \times p$ matrix of regressor variable values with rank $p<n$, β a p-vector of parameters, and ε an n-vector of random errors.

The minimum L_1-norm estimator which is also called the least absolute values estimator has long been considered as an acceptable robust alternative to least squares estimator, particularly in the presence of vertical outliers which are outlying observations in the y-direction. The L_1-estimator ($\hat{\beta}$) is defined by the solution of the following problem

$$\underset{\hat{\beta}}{\text{minimize}} \ \| e \|_1, \qquad (1)$$

where $e = y - X\hat{\beta}$, and $\| \cdot \|_1$ denotes the L_1-norm.

Statistical properties of the L_1-estimator have been studied extensively by Blattberg and Sargent (1971), Kiountouzis (1973), Rosenberg and Carlson (1977), Pfaffenberger and Dinkel (1978), Bassett and Koenker (1978), and Dielman and Pfaffenberger (1982). Bloomfield and

1) Department of Statistics, Sookmyung Women's University, Seoul 140-742, KOREA.

Steiger(1983) describe in detail the strong consistency and some robustness properties of the L_1-estimator. A necessary condition for the consistency of the L_1-estimator is proposed by Chen and Wu(1993). Birkes and Dodge(1993) present testing of hypotheses, confidence intervals, selection of variables, and so on with the L_1-estimation.

The degree of robustness of an estimator in the presence of outliers may be measured by the concept of breakdown point which is the smallest proportion of the observations that can render the estimator meaningless. Of course, it should be mentioned that the breakdown point is only one out of several measurements of robustness. It is well known that the L_1-estimator is robust with respect to vertical outliers, but it can not protect against bad leverage points, and hence the finite sample breakdown point of the L_1-estimator is equal t $1/n$. Regardless of this drawback, L_1-estimator has received considerable attention in the literature of the robust regression because the values of regressor variables are carefully chosen and are usually considered as fixed numbers in designed experiments.

Therefore, we are concerned with the robustness of the L_1-estimator with respect to vertical outliers only. One approach to investigate its robustness is to construct the intervals within which the values of the response variable can lie without changing the L_1-estimates. Recently, Narula and Wellington(1990) suggest an algorithm for finding the intervals only in case of simple linear regression. In this article, we introduce a procedure for the construction of these intervals not only in simple linear regression but also in multiple linear regression. Another approach is to measure its robustness by the breakdown point. So, we propose a definition of the breakdown point of the L_1-estimator to vertical outliers - we may call this the vertical breakdown point.

2. Algorithmic Framework of L_1-estimation

This section deals with algorithmic considerations in the L_1-estimation problem. The standard method for computing the L_1-estimates derives from the following linear programming formulation of the problem (1)

$$\underset{\hat{\beta}}{\text{minimize}} \ \{ \mathbf{1}' \boldsymbol{e}^{+} + \mathbf{1}' \boldsymbol{e}^{-} : \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{I}_n \boldsymbol{e}^{+} - \boldsymbol{I}_n \boldsymbol{e}^{-} = \boldsymbol{y},\ \boldsymbol{e}^{+} \geq 0,\ \boldsymbol{e}^{-} \geq 0 \}, \qquad (2)$$

where $\mathbf{1}$ is an n-dimensional vector of all ones, and the components of vector $\boldsymbol{e}^{+}$ and $\boldsymbol{e}^{-}$ represent positive and negative deviations, respectively.

Algorithms for computing L_1-estimates have been developed by Wagner (1959), Fisher (1961), Barrodale and Roberts (1973), Armstrong, Frome, and Kung (1979), Wesolowsky (1981), Kim (1987), Gentle, Narula, and Sposito(1987), Sherali, Skarpness, and Kim (1988), and Coleman and Li(1992). Although some of the special purpose algorithms take into account the particular features of the problem (2) in order to improve the computational efficiency and/or numerical accuracy and stability, we shall use the basic ideas of the simplex method since the objective in this article is not to deal with the computational or numerical problems.

With a slight modification of the constraint set such that $\hat{\beta} = \hat{\beta}^+ - \hat{\beta}^-$, $\hat{\beta}^+ \geq 0$, $\hat{\beta}^- \geq 0$, the linear programming problem (2) can be reformulated in the canonical form

$$\underset{r}{\text{minimize}} \ \{ c'r : Ar = y, \ r \geq 0 \}, \tag{3}$$

where $c' = [0' \ \ 0' \ \ 1' \ \ 1']$, $r' = [\hat{\beta}^{+\prime} \ \ \hat{\beta}^{-\prime} \ \ e^{+\prime} \ \ e^{-\prime}]$, and $A = [X \ -X \ I \ -I]$.

Now, the problem (3) can be solved by the simplex method. The steps of the simplex algorithm for the computation of the L_1-estimates are described below for the sake of completeness and to introduce notation. Consider the system $Ar = y$, $r \geq 0$, where A is an $n \times 2(n+p)$ matrix with rank n. After rearranging the columns of A, let $A = [B \mid N]$ where B is an $n \times n$ nonsingular matrix and N is an $n \times (n+2p)$ matrix. Here B is called the basis and N is called the nonbasic matrix. The point $r' = [r_B' \mid r_N']$ where $r_B = B^{-1}y$, $r_N = 0$, is called a basic solution of the system. If $r_B \geq 0$, then r is called a basic feasible solution of the system. The components of r_B are called basic variables, and the components of r_N are called nonbasic variables.

Initialization : Choose a starting basic solution with basis B.

Step 1 : Solve the system $Br_B = y$. Then $\begin{bmatrix} B^{-1}y \\ 0 \end{bmatrix}$ is a basic feasible solution and the objective value is given by $z_0 = c_B' r_B$.

Step 2 : Solve the system $w'B = c_B'$ (with unique solution $w' = c_B'B^{-1}$). Let a_j be the j-th column of A, and then calculate $z_j - c_j = w'a_j - c_j$ for all nonbasic variables. Let

$$z_k - c_k = \max_{j \in R} \{ z_j - c_j \},$$

where R is the current set of indices associated with the nonbasic variables. If $z_k - c_k \leq 0$, then stop with the current basic feasible solution as an optimal solution.

Step 3 : Solve the system $Bs_k = a_k$ (with unique solution $s_k = B^{-1}a_k$). If $s_k \leq 0$, then stop with the conclusion that the optimal solution is unbounded.

Step 4 : Here r_k enters the basis and the blocking variable r_{B_t} leaves the basis, where the index t is determined by the following minimum ratio test :

$$\frac{\tilde{y}_t}{s_{tk}} = \min_{1 \le i \le n} \left\{ \frac{\tilde{y}_i}{s_{ik}} : s_{ik} > 0 \right\}$$

where $\tilde{\boldsymbol{y}} = \boldsymbol{B}^{-1}\boldsymbol{y}$. Update the basis $\boldsymbol{B}$ where $\boldsymbol{a}_k$ replaces $\boldsymbol{a}_{B_t}$, and the index set R. And go to Step 1.

3. Robustness of the L_1-estimator

In this section, we address the question of how robust against vertical outliers the L_1-estimator is. In this connection, we introduce a procedure for determining intervals on the values of the response variable within which the L_1-estimates do not change. Also we suggest a definition of vertical breakdown point of the L_1-estimator.

3.1 Procedure for Intervals of $\boldsymbol{y}$

A procedure for constructing intervals of y is suggested on the basis of the sensitivity analysis. Suppose that an optimal solution is obtained by the simplex method introduced in Section 2. Let $\boldsymbol{B}_{opt}^{-1}$ be the optimal basis of $\boldsymbol{A}$. If the right-hand-side vector $\boldsymbol{y}$ has its k-th component changed, that is, $\bar{y}_k = y_k + \delta$, then to verify optimality we need only multiply the vector $\bar{\boldsymbol{y}}$ by $\boldsymbol{B}_{opt}^{-1}$.

$$\begin{aligned} \boldsymbol{B}_{opt}^{-1}\bar{\boldsymbol{y}} &= \boldsymbol{B}_{opt}^{-1}(\boldsymbol{y} + \delta \boldsymbol{h}) \\ &= \tilde{\boldsymbol{y}}^{o} + \delta(\boldsymbol{B}_{opt}^{-1})^{k} \\ &= \tilde{\boldsymbol{y}}^{o} + \delta \boldsymbol{b} \ , \end{aligned}$$

where $\boldsymbol{h}$ is a vector of zeros except for 1 at the k-th position, and $(\boldsymbol{B}_{opt}^{-1})^{k}$ $(= \boldsymbol{b})$ denotes the k-th column of $\boldsymbol{B}_{opt}^{-1}$. To determine the range on $\bar{y}_k$, we must determine when the updated right-hand-side vector remains greater than or equal to zero. That is, $\tilde{y}_i^{o} + \delta b_i \ge 0$

for $i=1,\ldots,n$.

Case 1 : If $b_i<0$, δ must be such that $\delta \le -\tilde{y}_i^o/b_i$ for all i with $b_i<0$. Thus, the maximum increase in $\bar{y}_k=y_k+\delta$ is restricted by

$$\delta_{\max}=glb\ \{ -\tilde{y}_i^o/b_i \mid b_i<0 \},$$

where *glb* means the greatest lower bound. That is, the L_1-estimates are not altered for all values of $y_k+\delta_{\max}$. If the set is empty, then the level can be increased without bound without changing the basic variable.

Case 2 : If $b_i>0$, then δ must be such that $\delta \ge -\tilde{y}_i^o/b_i$. This implies that the maximum decrease in $\bar{y}_k=y_k+\delta$ occurs when

$$\delta_{\min}=lub\ \{ -\tilde{y}_i^o/b_i \mid b_i>0 \},$$

where *lub* means the least upper bound. In other words, the L_1-estimates do not change for all values of $y_k+\delta_{\min}$.

The intervals can be readily computed by any computer programs (for instance, SAS/OR or LINDO) for the sensitivity analysis under the algorithmic framework presented in Section 2. This procedure is illustrated as follows using real and simulated data sets.

Example 1 : A data set from Montgomery and Peck(1992, p.233) is used to illustrate the procedure. Twelve observations are obtained on two regressor variables(temperature of the product, filler operating pressure) and the response variable(carbonation level of a soft drink beverage). The L_1-estimates are {-124.675, 0.830, 4.760} whereas the least squares estimates are {-147.489, 1.719, 4.557}. The lower and upper bounds on y values for all observations for which the L_1-estimates are not affected are presented in Table 1. As long as, for instance, the value of the response variable for the first observation is greater than or equal to 1.015, the L_1-regression estimates will not change, and so on. The results show that the {1, 2, 3, 5, 8}-th values of the response variable can be moved from the fitted values to positive infinity without altering the L_1-estimates, and the {4, 9, 11, 12}-th values can be moved from the fitted values to negative infinity without affecting the L_1-estimates. These nine points are called the nondefining observations. On the other hand, the {6, 7, 10}-th values are restricted

Table 1 : Intervals for the values of the response variable

x_1	x_2	y	Interval for y
31.0	21.0	2.60	[1.015, $+\infty$)
31.0	21.0	2.40	[1.015, $+\infty$)
31.5	24.0	17.32	[15.710, $+\infty$)
31.5	24.0	15.60	($-\infty$, 15.710]
31.5	24.0	16.12	[15.710, $+\infty$)
30.5	22.0	5.36	[5.333, 5.377]
31.5	22.0	6.19	[6.080, 6.600]
30.5	23.0	10.17	[10.120, $+\infty$)
31.0	21.5	2.62	($-\infty$, 3.395]
30.5	21.5	2.98	[2.955, 3.008]
31.0	22.5	6.92	($-\infty$, 8.155]
30.5	22.5	7.06	($-\infty$, 7.740]

narrowly. This implies that these three points are the defining observations which are exactly on the L_1-regression plane.

Example 2 : A data set of fifteen observations is generated by the test problem generator, L1GNR, of Hoffman and Shier(1980). There are three regressor variables, and the L_1-regression coefficients including the intercept term are specified as {1.0, 2.0, 3.0, 4.0}. The values for the regressor variables are generated from normal distribution with means {5.0, 15.0, 25.0} and variances {1.0, 2.0, 3.0}, respectively. Each row of the matrix X is generated to have unique values. The errors are generated from a normal distribution with mean 0.0 and variance 4.0. Three observations are set to be exactly on the regression hyperplane outside the basis. The intervals for the values of the response variable are summarized in Table 2. The L_1-regression hyperplane will not change if the values of the response variable stay in the intervals.

The results presented in Table 1 and 2 confirm the robustness of the L_1-estimator to vertical outliers. That is, the L_1-estimates are not affected at all, within certain range, by change in values of response variable.

Table 2 : Intervals for the values of the response variable

x_1	x_2	x_3	y	Interval for y
4.42	16.49	25.33	160.64	[160.630, $+\infty$)
5.12	14.60	22.37	144.52	[144.519, 144.530]
6.08	14.54	28.20	169.57	[168.383, 169.580]
6.31	16.50	30.39	184.77	[184.667, $+\infty$)
4.91	13.46	25.26	152.24	[152.238, $+\infty$)
4.82	17.83	23.14	156.47	($-\infty$, 156.691]
9.71	7.11	39.84	196.63	($-\infty$, 201.073]
3.71	10.93	23.20	140.04	[134.014, $+\infty$)
4.27	13.72	25.36	152.14	[152.138, 152.144]
4.92	12.12	21.59	133.56	($-\infty$, 133.561]
6.90	15.16	28.46	175.41	[174.107, $+\infty$)
4.68	15.11	23.48	147.54	($-\infty$, 149.611]
6.02	15.07	25.88	164.98	[161.763, $+\infty$)
3.74	17.61	21.46	143.02	($-\infty$, 147.157]
4.33	18.20	25.60	166.66	[166.644, 166.662]

3.2 Vertical Breakdown Point

The finite sample breakdown point of the L_1-estimator is known to be $1/n$, which tends to zero for increasing sample size. However, if we can restrict the robustness of L_1-estimator only on the vertical outliers, the finite sample breakdown point of the L_1-estimator may be much greater than that value.

It is shown in the previous section that the L_1-estimates do not change as long as the values of response variable remain within the intervals. In this section, two properties related to the intervals are verified employing some analytical arguments. One property is that at least one hyperplane giving minimum L_1-norm passes through p of the n observations. In other words, L_1-estimator has p defining observations for which the intervals are restricted narrowly. It is well known that the problem (2) has at least one solution at an extreme point of the feasible region. The constraint set of that problem has n equations and $2n+p$ variables which are p unrestricted variables and $2n$ non-negative variables. If the i-th point is not on a hyperplane, then either e_i^+ or e_i^- must be non-zero. However, $n+p$ of the $2n$

non-negative variables must be equal to zero at an extreme point. Therefore, $n-p$ points can be off the hyperplane at the optimum. This property implies that the L_1-estimates are completely determined by only a subset of p observations, and small change in the value of response variable for the defining observations may substantially influence the L_1-estimates.

Another property is that L_1-estimates are not altered by changes in the values of the response variable associated with $n-p$ nondefining observations as long as those observations remain on the same side of the L_1-regression hyperplane. In particular, the values of y corresponding to the nondefining observations can be taken from the fitted values to positive or negative infinity without changing the L_1-estimates. This property can be verified by the discrete approximation approach as follows. [This part has also been proved by Dodge(1990).] Suppose $\hat{\boldsymbol{\beta}}^*$ is the optimal solution of the L_1-estimation problem (1). Let $\boldsymbol{x}_i'$ be the i-th row of $\boldsymbol{X}$, $\widehat{y_i}^* = \boldsymbol{x}_i' \hat{\boldsymbol{\beta}}^*$, and $e_i^* = y_i - \widehat{y_i}^*$. Now we have to prove that the optimal solution $\hat{\boldsymbol{\beta}}^*$ is also the solution of the problem

$$\underset{\hat{\boldsymbol{\beta}}}{\text{minimize}} \sum_{i=1}^{n} | u_i - \boldsymbol{x}_i' \hat{\boldsymbol{\beta}} | , \qquad (4)$$

where $u_i \geq \widehat{y_i}^*$ for $i \in L$, $L = \{ i \,;\, e_i^* > 0 \}$, and $u_i \leq \widehat{y_i}^*$ for $i \in M$, $M = \{ i \,;\, e_i^* < 0 \}$.

Define $f_i(\hat{\boldsymbol{\beta}}) = | y_i - \boldsymbol{x}_i' \hat{\boldsymbol{\beta}} |$, $g_i(\hat{\boldsymbol{\beta}}) = | u_i - \boldsymbol{x}_i' \hat{\boldsymbol{\beta}} |$, and $f(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} f_i(\hat{\boldsymbol{\beta}})$, $g(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} g_i(\hat{\boldsymbol{\beta}})$. Then the problem (1) and (4) can be simply expressed, respectively, as

$$\underset{\hat{\boldsymbol{\beta}}}{\text{minimize}}\, f(\hat{\boldsymbol{\beta}}) \qquad \text{and} \qquad \underset{\hat{\boldsymbol{\beta}}}{\text{minimize}}\, g(\hat{\boldsymbol{\beta}}).$$

It is known that $\hat{\boldsymbol{\beta}}^*$ minimizes the convex function $f()$ if and only if $0 \in \partial f(\hat{\boldsymbol{\beta}}^*)$, where $\partial f(\hat{\boldsymbol{\beta}}^*)$ is a subgradient of f at $\hat{\boldsymbol{\beta}}^*$. Therefore, to show that $\hat{\boldsymbol{\beta}}^*$ is also the solution of the problem (4), we have to show that $\partial f(\hat{\boldsymbol{\beta}}^*) \subseteq \partial g(\hat{\boldsymbol{\beta}}^*)$, and hence $0 \in \partial g(\hat{\boldsymbol{\beta}}^*)$.

Since $f_i(\hat{\boldsymbol{\beta}})$ and $g_i(\hat{\boldsymbol{\beta}})$ are convex functions, it is obvious that $\partial f(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} \partial f_i(\hat{\boldsymbol{\beta}})$ and $\partial g(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} \partial g_i(\hat{\boldsymbol{\beta}})$. So it suffices to show that

$$\partial f_i(\hat{\boldsymbol{\beta}}^*) \subseteq \partial g_i(\hat{\boldsymbol{\beta}}^*). \qquad (5)$$

The subdifferential of f_i at $\hat{\beta}^*$ is calculated, from the definition of $f_i(\hat{\beta})$, as

$$\partial f_i(\hat{\beta}^*) = \begin{cases} -\eta_i x_i' & \text{if } e_i^* \neq 0 \\ \lambda_i x_i' & \text{if } e_i^* = 0, \end{cases}$$

where $\eta_i = sign(e_i^*)$, and $-1 \le \lambda_i \le 1$. Similarly, we readily obtain that

$$\partial g_i(\hat{\beta}^*) = \begin{cases} -\eta_i x_i' & \text{if } u_i - \widehat{y_i}^* \neq 0 \\ \lambda_i x_i' & \text{if } u_i - \widehat{y_i}^* = 0. \end{cases}$$

From the above results, we can prove the relationship (5) as follows : i) If $e_i^* > 0$ and $u_i - \widehat{y_i}^* > 0$, then $\partial f_i(\hat{\beta}^*) = \partial g_i(\hat{\beta}^*)$. ii) If $e_i^* > 0$ and $u_i - \widehat{y_i}^* = 0$, then $\partial f_i(\hat{\beta}^*) \subseteq \partial g_i(\hat{\beta}^*)$. iii) If $e_i^* < 0$ and $u_i - \widehat{y_i}^* < 0$, then $\partial f_i(\hat{\beta}^*) = \partial g_i(\hat{\beta}^*)$. iv) If $e_i^* < 0$ and $u_i - \widehat{y_i}^* = 0$, then $\partial f_i(\hat{\beta}^*) \subseteq \partial g_i(\hat{\beta}^*)$.

This property implies that L_1-estimation is unaffected by any change in the data when the values of regressors remain the same and the y-values of the nondefining observations change so as to maintain the same signs of the residuals.

Consequently, the finite sample vertical breakdown point of the L_1-estimator is at most $(n-p)/2n$. Taking the limit for $n \to \infty$ with p fixed, we find that the vertical breakdown point of the L_1-estimator is as high as 0.5 which is the highest possible value.

4. Concluding Remarks

An attempt has been made in this article to assess the robustness of L_1-estimator by constructing the intervals for the values of the response variable within which the L_1-estimates are not affected, and by measuring the breakdown point with respect to vertical outliers. From the viewpoint of vertical outliers only, L_1-estimator is not less robust than other high breakdown point estimators such as the least median of squares, repeated median, and least trimmed squares estimators. It is known that a great deal of computation (for instance, 4833 seconds of CPU time for the exact least median of squares estimation of Hawkins-Bradu-Kass data on 486 IBM-compatible microcomputer with MS-FORTRAN compiler) is required for those high breakdown point estimators which are widely used in the robust regression. In particular, the computer program PROGRESS provided by Rousseeuw and Leroy(1987) for the LMS-estimation adapts the approximation methods, which are the faster version and the extensive search version, in order to save the execution time for large data set. However, the estimates obtained by PROGRESS have undesirable features in terms of numerical accuracy. On the other hand, the L_1-estimation requires less computation than the least median of squares estimation, and it yields exact estimates with no computational complexity. Moreover, statistical inferences on the L_1-regression have been developed relatively well. In this context, L_1-estimator may be preferred when data sets are obtained from designed experiments or verified not having leverage points by the outlier detection procedure.

References

[1] Armstrong, R. D., Frome, E. L., and Kung, D. S. (1979). A Revised Simplex Algorithm for the Absolute Deviation Curve Fitting Problem, *Communications in Statistics - Simulation and Computation*, Vol. B8, 175-190.

[2] Barrodale, I. and Roberts, F. D. K. (1973). An Improved Algorithm for Discrete L_1 linear Approximation, *SIAM Journal of Numerical Analysis*, Vol. 10, 839-848.

[3] Bassett, G. and Koenker, R. (1978), Asymptotic Theory of Least Absolute Error Regression, *Journal of the American Statistical Association*, Vol. 73, 618-622.

[4] Birkes, D. and Dodge, Y. (1993). *Alternative Methods of Regression,* John Wiley and Sons, New York.

[5] Blattberg, R. and Sargent, T. (1971). Regression with Non-Gaussian Stable Disturbances ; Some Sampling Results, *Econometrica*, Vol. 39, 501-510.

[6] Bloomfield, P. and Steiger, W. L. (1983). *Least Absolute Deviations : Theory, Applications, and Algorithms,* Birkhauser Boston, Inc.

[7] Chen, X. R. and Wu, Y. (1993). On a Necessary Condition for the Consistency of the L_1-estimates in Linear Regression Models, *Communications in Statistics - Theory Meth,* Vol. 22(3), 631-639.

[8] Coleman, T. F. and Li, Y. (1992). A Globally and Quadratically Convergent Affine Scaling Method for Linear L_1 problem, *Mathematical Programming*, Vol. 56, 189-222.

[9] Dielman, T. and Pfaffenberger, R. (1982). Least Absolute Value Estimation in Linear Regression ; A Review, *TIMS/Studies in the Management Sciences*, Vol. 19, 31-52.

[10] Dodge, Y. (1990). L_1-norm Regression for Detection of Outliers in both Response and Explanatory Variables, *Technical Report 90-1*, Statistics Group, University of Neuchatel, Switzerland.

[11] Fisher, W. D. (1961). A Note on Curve Fitting with Minimum Deviations by Linear Programming, *Journal of the American Statistical Association*, Vol. 56, 359-362.

[12] Gentle, J. E., Narula, S. C. and Sposito, V. A. (1987). Algorithms for Unconstrained L_1 Linear Regression, *Statistical Data Analysis ; Based on the L_1-norm and Related Methods*, edited by Y. Dodge, North-Holland, 83-94.

[13] Hoffman, K. L. and Shier, D. R. (1980). A Test Problem Generator for Discrete Linear L_1

Approximation Problems, *ACM Transactions on Mathematical Software*, Vol. 6, 587-593.

[14] Kim, B. Y. (1987). A Robust Estimation Procedure for the Linear Regression Model, *Journal of the Korean Statistical Society*, Vol. 16, 80-91.

[15] Kiountouzis, E. A. (1973). Linear Programming Techniques in Regression Analysis, *Applied Statistics*, Vol. 22, 69-73.

[16] Montgomery, D. C. and Peck, E. A. (1992). *Introduction to Linear Regression Analysis,* John Wiley & Sons, New York.

[17] Narula, S. C. and Wellington, J. F. (1990). On the Robustness of the Simple Linear Minimum Sum of Absolute Errors Regression, *Robust Regression ; Analysis and Applications*, edited by K. D. Lawrence and J. L. Arthur, Marcel Dekker, Inc, 129-141.

[18] Pfaffenberger, R. C. and Dinkel, J. J. (1978). Absolute Deviations Curve Fitting ; An Alternative to Least Squares, *Contributions to Survey Sampling and Applied Statistics*, edited by H. A. David, Academic Press, NY, 279-294.

[19] Rosenberg, B. and Carlson, D. (1977). A Simple Approximation of the Sampling Distribution of Least Absolute Residuals Regression Estimates, *Communications in Statistics - Simulation and Computation*, Vol. B6, 421-437.

[20] Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection*, Wiley-Interscience, New York.

[21] Sherali, H. D., Skarpness, B. O., and Kim, B. Y. (1988). An Assumption-Free Convergence Analysis for a Perturbation of the Scaling Algorithm for Linear Programs, with Application to the L_1 Estimation Problem, *Naval Research Logistics*, Vol. 35, 473-492.

[22] Wagner, H. M. (1959). Linear Programming Techniques for Regression Analysis, *Journal of the American Statistical Association*, Vol. 54, 206-212.

[23] Wesolowsky, G. O. (1981). A New Descent Algorithm for the Least Absolute Value Regression problem, *Communications in Statistics - Simulation and Computation*, Vol. B10, 479-491.