Discussion Paper 8304 April 1983

REGRESSION BY MINIMUM SUM OF ABSOLUTE ERRORS: SOME RESULTS

by

Rolando A. Danao

ABSTRACT

In the multiple regression model $y = x\beta + \varepsilon$, the coefficient vector β may be estimated by minimizing the sum of absolute errors (MSAE). This paper shows the following results under MSAE estimation: (1) If k coefficients $\hat{\beta}_j$ are nonzeros, then the estimated regression equation accurately predicts at least k observations; (2) As in the case of least squares regression, β has an infinite number of estimates in the presence of perfect multicollinearity.

Regression by Minimum Sum of Absolute Errors:
Some Results

Rolando A. Danao

1. Introduction

Consider the multiple linear regression model

$$y = x\beta + \varepsilon \tag{1}$$

where y is the regressand, $x = [x_1, x_2, \ldots, x_k]$ the vector of regressors, $\beta = [\beta_1, \beta_2, \ldots, \beta_k]'$ the vector of unknown coefficients, and ε the stochastic disturbance term. The most widely used method of estimating β is by least squares, i.e., by minimizing the sum of squared errors (MSSE). Another method is that of minimizing the sum of absolute errors (MSAE), i.e., the MSAE estimate of β is obtained by minimizing $\sum_i |\varepsilon_i|$. Although MSAE estimation was suggested as far back as 1888 by Edgeworth (Bowley, 1928), its use has been limited because of computational difficulties. It was only in the 1950's that articles appeared (Charnes et al., 1955; Wagner, 1959) showing that the MSAE estimator can be obtained as a solution to a linear programming problem.

The MSAE regression problem is stated as follows:

MSAE: Minimize $\sum_{i=1}^{n} |\varepsilon_i|$

$$\text{s.t.} \quad \sum_{j=1}^{k} x_{ij}\beta_j + \varepsilon_i = y_i, \quad i = 1, 2, \ldots, n.$$

where x_{ij} is the ith observation on the jth regressor. To transform this into the standard form of the linear programming problem, we introduce the positive and negative parts of a variable v of arbitrary sign by letting

$$v^+ = \max \{0, v\}$$
$$v^- = \max \{0, -v\}$$

Then

$$v = v^+ - v^-$$
$$|v| = v^+ + v^-,$$

and $v^+ \geq 0, \; v^- \geq 0.$

Thus, the MSAE problem can be restated as follows:

MSAE-LP1: Minimize $\sum_{i=1}^{n} (\varepsilon_i^+ + \varepsilon_i^-)$

$$\text{s.t.} \quad \sum_{j=1}^{k} x_{ij} (\beta_j^+ - \beta_j^-) + \varepsilon_i^+ - \varepsilon_i^- = y_i, \quad i = 1, 2, \ldots, n.$$

$$\beta_j^+, \beta_j^-, \varepsilon_i^+, \varepsilon_i^- \geq 0.$$

If we set

$$u = [1, 1, \dots, 1]' \quad \text{(an n-vector)}$$

$$\beta^+ = [\beta_1^+, \beta_2^+, \dots, \beta_k^+]'$$

$$\beta^- = [\beta_1^-, \beta_2^-, \dots, \beta_k^-]'$$

$$\varepsilon^+ = [\varepsilon_1^+, \varepsilon_2^+, \dots, \varepsilon_n^+]'$$

$$\varepsilon^- = [\varepsilon_1^-, \varepsilon_2^-, \dots, \varepsilon_n^-]'$$

$$y = [y_1, y_2, \dots, y_n]'$$

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}$$

then the MSAE problem can be written in matrix form:

<u>MSAE-LP2</u>: Minimize $u'\varepsilon^+ + u'\varepsilon^-$

$$\text{s.t.} \quad X\beta^+ - X\beta^- + I\varepsilon^+ - I\varepsilon^- = y$$

$$\beta^+, \beta^-, \varepsilon^+, \varepsilon^- \geqq 0.$$

<u>Remark</u>: It is clear from the constraint of MSAE-LP2 that the coefficient matrix has rank n. Consequently, a basic optimal solution has n basic variables. If k_1 of the β_j's are basic variables in an optimal solution, then there are k_1 of the ε_i's that are nonbasic; hence, at least k_1 of the ε_i's equal zero, i.e., the MSAE regression hyperplane passes through at least k_1 sample points. The implication for

prediction is that the MSAE regression equation can accurately predict historical data in at least k_1 out of n cases.

Example 1

Consider the data set:

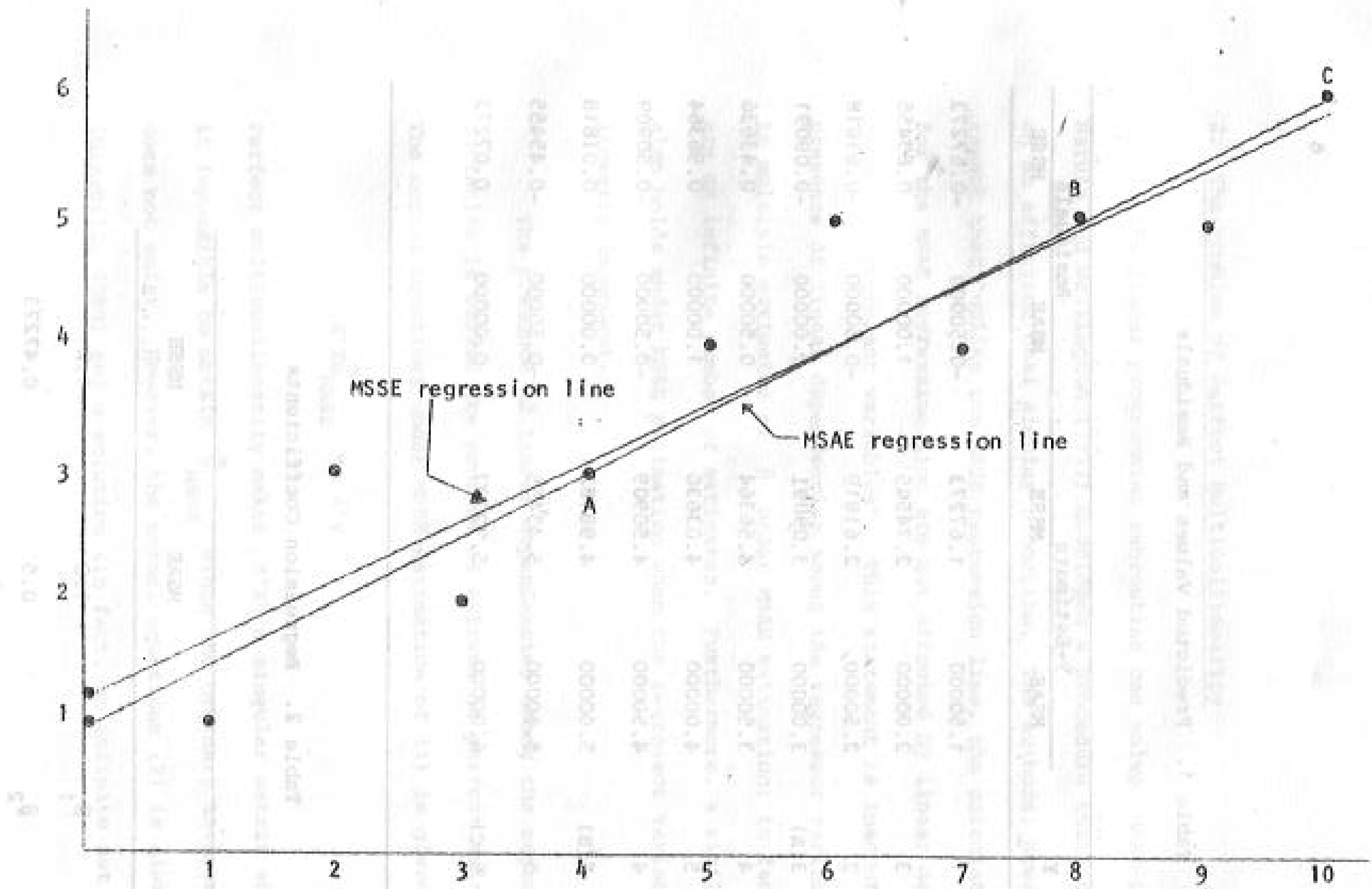
y	1	3	2	3	4	5	4	5	5	6
x	1	2	3	4	5	6	7	8	9	10

whose scatter diagram is shown in Figure 1.

Visual inspection of the scatter diagram indicates that the intercept and slope of the regression line

$$y = \beta_1 + \beta_2 x + \varepsilon$$

are nonzero, i.e., $\beta_1 \neq 0$, $\beta_2 \neq 0$. The preceding remark says that the MSAE regression line must pass through at least two points. The results of the MSAE regression done by linear programming confirms this as shown in Table 1 and Figure 1. In fact, the MSAE regression line passes through three points labeled A, B, C. The MSSE regression results are also shown for comparison.



6
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
A
B
C
MSSE regression line
MSAE regression line

Figure 1

Table 1. Predicted Values and Residuals

X	Y	Y-Estimate		Residuals	
		MSAE	MSSE	MSAE	MSSE
1	1	1.50000	1.67273	-0.50000	-0.67273
2	3	2.00000	2.14545	1.00000	0.85455
3	2	2.50000	2.61818	-0.50000	-0.61818
4	3 (A)	3.00000	3.09091	0.00000	-0.09091
5	4	3.50000	3.56364	0.50000	0.43636
6	5	4.00000	4.03636	1.00000	0.96364
7	4	4.50000	4.50909	-0.50000	-0.50909
8	5 (B)	5.00000	4.98182	0.00000	0.01818
9	5	5.50000	5.45455	-0.50000	-0.45455
10	6 (C)	6.00000	5.92727	0.00000	0.07273

Table 2. Regression Coefficients

	MSAE	MSSE
β_1	1.0	1.2
β_2	0.5	0.47273

2. The Problem of Perfect Multicollinearity

Any linear programming subroutine can solve MSAE-LP2. Narula and Wellington (1977) developed a procedure that is based on an efficient dual simplex algorithm. The authors, however, remark that "unlike the MSSE regression line, the calculations for the MSAE regression line are not affected by linear dependencies among the regressor variables". This statement is inaccurate. The presence of linear dependencies among the regressor variables results in multiple estimates of β under MSAE estimation; in fact, there are an infinite number of estimates. Furthermore, a similar situation also holds under MSSE estimation when the regressor variables are linearly dependent.

The presence of linear dependencies among the regressor variables is called the problem of extreme or perfect multicollinearity. The normal equations under MSSE estimation of (1) is given by

$$X'X\hat{\beta}_{MSSE} = X'y \tag{2}$$

Perfect multicollinearity makes $X'X$ a singular matrix which makes it impossible to obtain $\hat{\beta}_{MSSE}$ since the ordinary inverse of $X'X$ does not exist. However, the normal equations (2) is always consistent (Graybill, 1969) and a solution (in fact, an infinite set of solutions) for $\hat{\beta}_{MSSE}$ can be obtained by using the generalized inverse of $X'X$. The general solution is given by

$$\hat{\beta}_{MSSE} = (X'X)^g X'y + [I - (X'X)^g (X'X)] z \quad (3)$$

where $(X'X)^g$ is the generalized inverse of $X'X$, I is the identity matrix, and z is an arbitrary vector (Graybill, 1969).

We now show that in the presence of perfect multicollinearity, the estimate of β under MSAE estimation is not unique, i.e., MSAE-LP2 has an infinite number of optimal solutions. It is clear that MSAE-LP2 has a feasible solution given by

$$\beta_j^+ = \beta_j^- = 0, \qquad j = 1, 2, \ldots, k;$$

$$\left.\begin{aligned} \varepsilon_i^+ &= y_i \\ \varepsilon_i^- &= 0 \end{aligned}\right\} \text{ if } y_i \geq 0;$$

$$\left.\begin{aligned} \varepsilon_i^+ &= 0 \\ \varepsilon_i^- &= -y_i \end{aligned}\right\} \text{ if } y_i < 0.$$

Moreover, the objective function is bounded below by zero; hence, MSAE-LP2 has an optimal solution.

For expositional convenience, consider the case where a column of X is scalar multiple of another column. Let β_1 and β_2 be the coefficients of the linearly dependent regressors x_1 and x_2, respectively. Then, in MSAE-LP2, the variables β_1^+, β_1^-, β_2^+, β_2^- have zero coefficients in the objective function while the columns associated with them are pairwise linearly dependent. Suppose that β_1^+ is in the basis of the optimal solution obtained by the simplex algorithm. Then β_1^-, β_2^+, β_2^- cannot be in the basis since this would violate linear independence of the basis vectors. The portion of the optimal tableau (in canonical form) corresponding to these variables would look like the following:

Basic Variables	...	β_1^+	β_1^-	β_2^+	β_2^-	...	Right Hand Side
Objective Function Row →	...	0	0	0	0	...	b_o
.	...	0	0	0	0	...	.
.		:	:	:	:		.
.							.
β_1^+	...	1	-1	α	$-\alpha$	...	β_1^+
.		0	0	0	0		.
.		:	:	:	:		.
.							.

where $-b_o$ is the optimal value of the objective function. The column vector associated with β_1^+ is the unit vector since β_1^+ is in the basis. The other column vectors follow from the fact that they are scalar multiples of the vector associated with β_1^+. Note that β_2^+ is a nonbasic variable whose objective function coefficient is zero in the optimal tableau. This implies that the optimal solution is not unique since a necessary and sufficient condition for the uniqueness of an optimal solution is that the objective function coefficients of the nonbasic variables in the optimal tableau are positive (Simmonard, 1966). Another optimal solution can be obtained by pivoting on α (if $\alpha > 0$) or on $-\alpha$ (if $\alpha < 0$). This would put β_2^+ or β_2^- in the basis, replacing β_1^+ which now becomes zero. The new optimal solution has $\hat{\beta}_1 = 0$, $\hat{\beta}_2 = \frac{\beta_1^+}{\alpha}$ (if $\alpha > 0$) or $\hat{\beta}_2 = \frac{\beta_1^+}{-\alpha}$ (if $\alpha < 0$). Since the set of optimal solutions is convex, it follows that there are infinite number of optimal solutions.

Remarks: (1) The general case is proved in a similar manner. If rank $X < k$, then in the optimal solution, not all of the β's will appear as basic variables since this would violate the linear independence of the basis vectors. Suppose that $\beta_1^*, \beta_2^*, \ldots \beta_{k_1}^*$ (where $\beta_j^* = \beta_j^+$ or β_j^-) are in the optimal basis. This implies that the associated regressors $x_1, x_2, \ldots x_{k_1}$ are linearly independent. Moreover, there is a nonbasic variable β_ℓ^* whose associated regressor x_ℓ is a linear combination of $x_1, x_2, \ldots, x_{k_1}$. One can show that the column associated with β_ℓ^* in the optimal tableau is a linear combination of the columns associated with $\beta_1^*, \ldots, \beta_k^*$ with a zero objective function coefficient. This shows the existence of a nonbasic variable with a zero objective function coefficient in the optimal tableau which implies nonuniqueness of the optimal solution.

(2) In effect, MSAE estimation in the presence of perfect multicollinearity will choose a maximal set of linearly independent regressors (whose number equals the rank of X) and drops the other regressors from the equation by setting their coefficients equal to zero. This is also one of the remedies resorted to be researchers when confronted with perfect multicollinearity under MSSE estimation.

Example 2. Consider the following data set:

y	x_1	x_2
1	1	2
3	2	4
2	3	6
3	4	8
4	5	10

and the regression model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon .$$

Note that $x_2 = 2x_1$. Using the standard simplex algorithm on the MSAE-LP2 of this model, we obtain the optimal tableau show in Figure 2.

The optimal solution corresponding to this optimal tableau is given by

$$\hat{\beta}_0 = \frac{1}{3} \qquad \hat{\varepsilon}_1 = 0$$
$$\hat{\beta}_1 = 0 \qquad \hat{\varepsilon}_2 = 4/3$$
$$\hat{\beta}_2 = \frac{1}{3} \qquad \hat{\varepsilon}_3 = -1/3$$
$$\hat{\varepsilon}_4 = 0$$
$$\hat{\varepsilon}_5 = \frac{1}{3}$$

where the $\hat{\beta}_j$'s are the MSAE regression coefficients and the $\hat{\varepsilon}_i$'s are the residuals. Another optimal solution can be obtained by pivoting on the element $\frac{1}{2}$ (enclosed in a square) thus putting β_1^+ into the basis and romoving β_2^+ from the basis. This optimal solution is given by

$$\hat{\hat{\beta}}_0 = \frac{1}{3} \qquad \hat{\hat{\varepsilon}}_1 = 0$$
$$\hat{\hat{\beta}}_1 = 2/3 \qquad \hat{\hat{\varepsilon}}_2 = 4/3$$
$$\hat{\hat{\beta}}_2 = 0 \qquad \hat{\hat{\varepsilon}}_3 = -1/3$$
$$\hat{\hat{\varepsilon}}_4 = 0$$
$$\hat{\hat{\varepsilon}}_5 = 1/3$$

Basic Variables	β_0^+	β_0^-	β_1^+	β_1^-	β_2^+	β_2^-	ε_1^+	ε_1^-	ε_2^+	ε_2^-	ε_3^+	ε_3^-	ε_4^+	ε_4^-	ε_5^+	ε_5^-	Right Hand Side
Objective Function Row →	0	0	0	0	0	0	1	1	0	2	2	0	2	0	0	2	-2
β_0^+	1	-1	0	0	0	0	$\frac{4}{3}$	$-\frac{4}{3}$	0	0	0	0	$-\frac{1}{3}$	$\frac{1}{3}$	0	0	$\frac{1}{3}$
ε_2^+	0	0	0	0	0	0	$-\frac{2}{3}$	$\frac{2}{3}$	1	-1	0	0	$-\frac{1}{3}$	$\frac{1}{3}$	0	0	$\frac{4}{3}$
β_2^+	0	0	$\boxed{\frac{1}{2}}$	$-\frac{1}{2}$	1	-1	$-\frac{1}{6}$	$\frac{1}{6}$	0	0	0	0	$\frac{1}{6}$	$\frac{1}{6}$	0	0	$\frac{1}{3}$
ε_3^-	0	0	0	0	0	0	$\frac{1}{3}$	$-\frac{1}{3}$	0	0	-1	1	$\frac{2}{3}$	$-\frac{2}{3}$	0	0	$\frac{1}{3}$
ε_5^+	0	0	0	0	0	0	$\frac{1}{3}$	$-\frac{1}{3}$	0	0	0	0	$-\frac{4}{3}$	$\frac{4}{3}$	1	-1	$\frac{1}{3}$

Figure 2

This correponds to another estimate of β. Any convex combination of the optimal solutions $(\hat{\beta}; \hat{\varepsilon})$ and $(\hat{\hat{\beta}}; \hat{\hat{\varepsilon}})$ is also an optimal solution resulting in another MSAE estimate, i.e., $\theta\hat{\beta} + (1 - \theta)\hat{\hat{\beta}}$ $(0 \leqq \theta \leqq 1)$ is also an estimate of β.

References

Bowley, A.L. (1928). F.Y. Edgeworth's Contributions to Mathematical Statistics. London: Royal Statistical Society.

Charnes, A., Cooper, W.W. and Ferguson, R.O. (1955). Optimal Estimation of Executive Compensation by Linear Programming. Management Science, 1, 138-151.

Graybill, F.A. (1969). Introduction to Matrices with Applications in Statistics. Belmont, California: Wadsworth Publishing Company.

Narula, S.C., and Wellington, J.F. (1977). Multiple Linear Regression with Minimum Sum of Absolute Errors. Applied Statistics, v. 26, no. 1, pp. 106-111.

Simmonard, M. (1966). Linear Programming. Prentice-Hall, N.J.

Wagner, H.M. (1959). "Linear Programming Techniques for Regression," Journal of the American Statistical Association.