RT-MAE 9724

THE MINIMUM SUM OF ABSOLUTE ERRORS REGRESSION: AN OVERVIEW

by

Subhash C. Narula, Paulo H.N. Saldiva, Carmen D.S. Andre, Silvia N.Elian, Aurea Favero Ferreira and Vera Capelozzi

Palavras-Chave: Interstitial lung disease; least squares; leverage points; multicollinearity; outliers; variables selection.

Classificação AMS: 62F35, 62J05, 92C50.
(AMS Classification)

- Setembro de 1997 -

THE MINIMUM SUM OF ABSOLUTE ERRORS REGRESSION: AN OVERVIEW

Subhash C. Narula
Virginia Commonwealth University
Richmond, Virginia, USA

Paulo H. N. Saldiva
Carmen D. S. Andre
Silvia N. Elian
Aurea Favero Ferreira
Vera Capelozzi
University of Sao Paulo
Sao Paulo, Brazil

Abstract

In this paper, our objective is to introduce the minimum sum of absolute errors regression which is a more robust alternative to the popular least squares regression whenever there are outliers in the values of the response variable, or the errors follow a long tailed distribution, or the loss function is proportional to the absolute errors rather than their squared values. We do so with a real application from the medical field. We point out some of the problems with the least squares analysis and show how these are avoided by the minimum sum of absolute errors analysis.

Key words : interstitial lung disease; least squares; leverage points; multicollinearity; outliers; variables selection.

1. INTRODUCTION

In medical studies, quantitative models have been used to diagnose and assess the response to therapy. The least squares regression is one of the most often used model; however, it is very sensitive to outliers. There are several ways to deal with this problem, for example, one may identify the outliers, reject them and fit the model to the remaining observations, or one may use a more robust procedure than least squares to estimate the parameters. Sometimes the first alternative is not acceptable or desirable because the observations are legitimate and therefore should not be discarded. The second alternative of using a more robust technique presents another difficulty of choosing a technique from among several available techniques.

Our objective, in this paper, is to introduce the minimum sum of absolute errors MSAE regression as an alternative to the least squares regression which is sensitive to outliers. The MSAE analysis is also more appropriate than least squares when the errors follow a long tailed distribution or the loss function is proportional to absolute value of the errors rather than their squared value. We introduce the technique and the process of model selection with an actual case study from medicine. The rest of the paper is organized as follows: In Section 2, we describe the medical study which started us on this problem and give the least squares analysis for the problem. In Section 3, we give an overview of the MSAE regression. In Section 4, we present the MSAE analysis for the problem and conclude the paper with a few comments in Section 5.

2. THE EXAMPLE

Interstitial Lung Disease(ILD) refers to a diffuse inflammatory process that occurs predominantly within the interstitial spaces and supporting structures of a lung. Clinical chart and x-rays (radiological pictures) of a patient with ILD usually suggest an open-chest lung biopsy to establish the diagnosis and to provide additional information on activity and stage of disease.

Pathological assessment is important to determine the prognosis and response of ILD to therapy, Carrington, Gaensler, Coutu, Fitzgerald, and Grupta (1978), Katzeinstein, and Askin (1990) and Crystal, Bitterman, Rennard, Hance, and Keogh (1984). However, in routine practice, the proper quantification of the extension and severity of pulmonary involvement is sometimes difficult and subject to frequent disagreement among different pathologists. Therefore, semi-quantitative scoring systems have been proposed, Cherniak, Colby, Flint, Thurlbeck, Waldron, Ackerson, and King (1991), Watters, King, Schwarz, Waldron, Styanford, and Cherniak (1986) and Fulmer, Robert, von Gal, and Crystal (1979), to provide the practicing pathologists a more rational basis to establish the severity of ILD.

The idea embodied in using pathological scoring systems is that the amount of alterations detected when analyzing the biopsy specimen express the severity of patient's functional and clinical impairment. However, because ILD usually affects a large part of pulmonary parenchyma, one has to be cautious when trying to establish structural-clinical correlation's based on a small tissue sample. Thus, studies trying to correlate morphological alterations of lung biopsies with data more representative of entire lung function (such as pulmonary function tests) are necessary.

In an elegant study, Watters, et. al. (1986) demonstrated that histopathological alterations of open-chest lung biopsies of patients with ILD, as determined by semi-quantitative scoring, significantly correlate with clinical, radiological and functional parameters. This finding encourages further studies focusing the role of applying quantitative histological criteria to lung biopsies to assess the severity of ILD. In this context, it is possible that the combination of conventional histopathological assessment of ILD may improve the accuracy of histopathological evaluation of lung biopsy, adding important information about the severity of disease.

This study was designed to verify the association between objective indicators of lung damage and severity of functional impairment in ILD patients. For this purpose, stereological and semi-quantitative techniques were employed on 24 open-chest lung biopsies of patients with diffuse interstitial involvement.

Patients: Twenty four biopsies of patients with ILD were selected from the file cases of open chest lung biopsies of Surgical Pathology Service of the teaching hospital of Faculdade de Medicina da Universidade de Sao Paulo. Biopsies were selected for this study on the basis of availability of patient's complete clinical and radiological data. In addition, this set of patients had the pulmonary function measurements gathered within 30 days before the biopsy.

Pulmonary Function Measurements: Forced Vital Capacity (FVC, y) was measured with a computerized modular lung analyzer as recognized by the American Thoracic Society (1991) and expressed in terms of the predicted value for each patient, according to patient's age and physical characteristics, Morris, Koski, and Johnson (1971).

Morphological Analysis: Fragments were fixed in 10% buffered formalin and embedded in paraffin for processing by routine histological procedures. Semi-thin sections (2 micrometers) were obtained from the paraffin blocks using the technique described by Junqueira, Silva, and Torloni(1989). Slides were stained with ematoxylineosin.

Pathological studies were carried out without knowledge of patient's clinical or physiological status. In the first step, morphometric studies were done at the level of alveolar interstitium to determine the areal fractions of cellular infiltration (CELL, x_8) and septal vascularization (VES, x_9) at alveoplar level. For this purpose, twelve randomly selected non-coincident 1000x power fields of lung parenchyma were studied, excluding axial components such as large bronchi and vessels. The areal fraction of each component of alveolar tissue was determined by standard-point-counting procedure, i.e., by counting 1,420 points per biopsy. In addition, differential counting of cells within the interstitial space was performed at the same moment. Cells in the pulmonary interstitium were classified into four-categories based on their appearance at light microscopy, Saldiva, Brentani, Carvalho, Auler, Calheiros, and Pacheco (1985): epithelial cells(EPIT, x_4), elongated cells(FUSI, x_5), polymorphonuclear cells(POLY, x_7) and mononucleated cells(MONO, x_6). At a lower magnification(40x), more general aspects of parenchyma remodeling were quantified by a semi-quantitative scoring system. The presence of vascular sclerosis(SCLEVASC, x_{12}), obliterate bronchiolitis(BOBLIT, x_{10}), smooth muscle hyperplasia(MUSCLE, x_{11}), honeycombing(HONEY, x_{13}), and desquamative pneumonia(DESQ, x_{14}) were individually graded from zero to four. For each of the preceding alterations, a degree zero corresponds to the absence of alteration; the degree one indicates that less than 25% of the structures of interest are altered ; the degree two indicates that 25 to 50% of the structures under analysis are affected; the degree three indicates that more than 50% but less than 75% of structures are altered; and, finally, degree four signifies that more than 75% of the structures are abnormal. Pathological scoring was carried out simultaneously by two pathologists, in double observation microscope.

In addition to the pulmonary function measurements and morphological variables, variables such as the age in years(AGE, x_2), sex(SEX, x_1) and if the patient smoked or not (SMOK, x_3) were also observed. The list of variables and the data are given in the Appendix A.

The Least Squares Analysis

We started the analysis by fitting the least squares model to all the variables. The resulting model had $R^2 = 0.764$; however, the model and all the regression coefficients were not significant at the 5% level of significance (see Appendix B for details) which may have been caused by multicollinearity.

To assess prognosis and response to treatment for ILD patients, as a next step, we decided to develop a parsimonious model that is effective, easy to understand, explain and maintain. To do so, we began the analysis of the data using a stepwise least squares regression as a procedure to select a model. The resulting model was:

$$y = 46.65 + 0.614\, x_2 - .0615\, x_4 + 107.73\, x_8 - 10.64\, x_{13},$$

with $R^2 = 0.708$. That is, this model explained 70.8 % of the variation in the response variable. An analysis of the residuals identified two outliers and a leverage point. We deleted the two outliers (observation numbers 11 and 15) and recomputed the model; the resulting model was:

$$y = 54.84 + 0.439\, x_2 - .064\, x_4 + 112.57\, x_8 - 10.48\, x_{13}.$$

Clearly, the coefficients of the model changed when the two outliers were eliminated; the major changes occurred in the values of the intercept and the coefficient of x_2. See Appendix B for more details.

On further investigation of the data, it was confirmed that these observations were correct. Therefore, it was decided not to discard them and to use a more robust procedure than least squares to estimate the parameters of the model. The minimum sum of absolute errors MSAE regression is one such alternative.

3. THE MSAE REGRESSION

3.1 Introduction

Let y be an n x 1 vector of value of the response variable corresponding X, an n x k matrix of regressor (predictors) variable values that may include a columns of ones for the intercept term. Consider the multiple linear regression model

$$y = X\beta + \varepsilon \qquad (1)$$

where β is a k x 1 vector of the unknown parameters and ε is an n x 1 vector of the unobservable random errors. The components of ε are independent and identically distributed random variables with density function f(.). Let v denote the median of εi and define the scale parameter τ as $\tau = (2f(v))^{-1}$.

The minimum sum of absolute errors MSAE estimator $\hat{\beta}$ of β minimizes $\Sigma_{i=1}^{n} |y_i - x_i\beta|$ for all values of β where y_i is the i-th element of the vector y and x_i is the i-th row of the matrix X. The MSAE criterion is a robust alternative to the least squares principle to estimate β whenever the data contains outliers or the errors follow a long tailed error distribution such as the Laplace or the Cauchy distribution, or the loss function is proportional to the absolute value of the errors rather than their squared values, Narula and Wellington (1977, 1985). Huber (1974, p. 927) stated that with regards to L_p estimators in regression, "p = 1 (*MSAE regression)* gives robustness in a technical sense (Hample, 1971), i.e., resistance against arbitrary outliers." Unlike other robust procedures, MSAE regression does not require a rejection parameter. Because it is resistant to outliers, it provides a good starting solution for one step and iteratively weighted multi-step least squares methods.

Boscovich (1757) proposed that a straight line should be fitted to three or more non-collinear points in a plane so as to satisfy two conditions (i) that the sum of the positive and the negative errors of the given points from the fitted line be equal in magnitude, and (ii) that the sum of the absolute errors be minimum. The MSAE regression reappeared in the 1880's largely due to the work of Edgeworth. His contributions included: (a) that condition (i) of Boscovich should be dropped so that the minimum in condition (ii) can obtain its smallest value (Edgeworth (1887)); (b) that the MSAE regression may not be unique; and (c) that the MSAE estimators are maximum likelihood estimators when the errors follow a Laplace distribution (Edgeworth (1888)). He also developed an algorithm to solve the simple linear MSAE problem. The interested reader may refer to Farebrother (1987) for further historical details.

Until the late 1950's, the computational and statistical inference problems associated with the MSAE regression effectively prevented its use. Karst (1958) proposed an algorithm to solve the simple linear regression problem and Wagner (1959) formulated the multiple linear regression problem as a linear programming problem. Since then, a number of very efficient algorithms have been proposed which removed the first problem. Basset and Koenker (1978) developed the asymptotic distribution of the MSAE estimators. Based on these results, statistical inference procedures have been proposed, thus removing the second difficulty in the use of the MSAE principle.

3.2 Computational algorithms

Simple Linear Regression: Let y_i denote the value of the response variable corresponding to x_i, the value of a regressor (or predictor) variable for the i-th observation, i = 1, 2, ..., n., where n is the number of observations. The simple linear regression model may be written as:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad (2)$$

where β_0 and β_1 are the unknown intercept and slope parameters of the model, and ε_i represents the unobservable random error.

Our objective is to determine the estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ of β_0 and β_1 such that $\sum_{i=1}^{n} |y_i - \hat{y}_i|$ is minimum, where $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$. Edgeworth (1888) proposed an algorithm to compute the MSAE estimates for the simple linear regression model; however, his method does not seem to have been used widely. Seventy years later, Karst (1958) developed an intuitively appealing iterative algorithm for the problem. Since then, a number of algorithms were proposed in quick succession, e.g., Barrodale and Roberts (1973), Armstrong and Kung (1978), Abdelmalek (1980), Wesolowsky (1981), Klingman and Mote (1982), and Josavanger and Sposito (1983).

Multiple Linear Regression: In an effort to determine compensation for executives (i.e., salary plus fringe benefits), Charnes, Cooper, and Ferguson (1955) pointed out that the MSAE regression problem is essentially a linear programming problem. Wagner (1959) formulated it as the following linear programming problem:

(LP)

$$\text{Minimize} \quad 1'(e^+ + e^-)$$

$$\text{Subject to}$$

$$X\hat{\beta} + e^+ - e^- = y,$$

$$e^+,\ e^- \geq 0,$$

$$\hat{\beta} \text{ unrestricted in sign}$$

where 1 is an n x 1 vector of ones, and e^+ and e^- are n x 1 vectors of residuals corresponding to over- and under-prediction of y, respectively. He stated that a dual formulation of the problem may be solved more efficiently by a simplex algorithm for bounded variable. Barrodale and Roberts (1973) proposed a very efficient special purpose algorithm for solving (LP). At present a number of very efficient and effective algorithms, for example, by Bartels, Conn, and Sinclair (1976, 1978), Armstrong, Frome, and Kung (1979), Bloomfield and Steiger (1980), Wesolowsky (1981), Coleman and Li (1992), Madsen and Nielsen (1993), Ruzinsky and Olsen (1989), and Zhang (1993) among others are available to solve the simple and multiple linear MSAE regression models. The interested reader may also consult Narula (1987).

3.3 Computer Programs

A number of computer programs are available to compute the MSAE estimates of the parameters of the simple and the multiple linear regression models. For example, computer programs for the simple and multiple linear regression appear in Barrodale and

Roberts (1974), Bartels, Conn, and Sinclair (1976), Armstrong and Kung (1978), and Armstrong, Frome, and Kung (1979).

For the simple linear regression problem, the computer program of Josavanger and Sposito (1983), which is based on the modification of the algorithm of Wesolowsky (1981), performed well in a comparative study of Gentle, Sposito, and Narula (1988). In a study to compare the relative performance of the available computer programs for the multiple linear regression problem, Gentle, Narula, and Sposito (1987) reported that within the limitations of the study the program of Armstrong, Frome, and Kung (1979) performed better than the other available programs.

The MSAE model may also be fitted using the robust regression package ROBSYS. Furthermore, a short FORTRAN program can be written to calculate the estimates using the IMSL subroutine *RLLAV*. At present, the computer programs are also available in popular statistical packages like S-Plus (*L1fit* function) and SAS (proc *IML*). Therefore, at present, it is reasonable to claim that the computational difficulties associated with the use of the MSAE regression do not exist.

3.4 Statistical Inference

It is well known that the MSAE estimators are maximum likelihood estimators and hence are asymptotically unbiased and efficient when errors follow the Laplace distribution. Basset and Koenker (1978) proved that the MSAE estimator $\hat{\beta}$ of the parameter β of the regression model is asymptotically unbiased, consistent and asymptotically follows a multinormal distribution with variance-covariance matrix $\tau^2(X'X)^{-1}$, where τ^2/n is the variance of the median of a sample of size n from the error distribution. An important implication of this result is that the MSAE estimator has a smaller confidence ellipsoid than the least squares estimator of β for any error distribution for which the sample median is a more efficient estimator than the sample mean.

Based on the asymptotic distribution results, the formulas for constructing confidence intervals and testing hypothesis on the parameters of the model have been developed, Dielman and Pfaffenberger (1982) and Narula (1987). We give a few formulae for confidence intervals and tests of hypotheses on the linear combination $r'\beta$ of the regression parameters, where r is a k x 1 vector of known constants.

- A (1 - α) 100% confidence interval on $r'\beta$ may be written as

$$r'\hat{\beta} \pm z_{\alpha/2}\hat{\tau}\{r'(X'X)^{-1}r\}^{1/2}, \qquad (3)$$

where z_p denotes the (1 - p)th percentile of the standard normal distribution and $\hat{\tau}$ is a consistent estimator of τ.

A number of estimators of τ have been proposed. One such estimator recommended by Birkes and Dodge (1993) and McKean and Schrader (1984) is

$$\hat{\tau} = \sqrt{n^*}\,(e_{(n^*-m+1)} - e_{(m)})/4, \qquad (4)$$

where $m = (n^* + 1)/2 - \sqrt{n^*}$, n^* is the number of nonzero residuals from (1), and $e_{(1)}, e_{(2)}, \ldots, e_{(n^*)}$ are the nonzero residuals arranged in an ascending order.

- For a single component of β, say β_i, the (1 - α) 100% confidence interval is

$$\hat{\beta}_i \pm z_{\alpha/2}\hat{\tau}\sqrt{(X'X)^{-1}_{ii}}, \qquad (5)$$

where $(X'X)^{-1}_{ii}$ is the i-th diagonal element of $(X'X)^{-1}$ and $\hat{\beta}_i$ is the i-th component of $\hat{\beta}$.

- To test the null hypothesis, H_0: $r'\beta = \rho$ versus the alternative hypothesis H_1: $r'\beta \neq \rho$ at the α level of significance, the decision rule is

Reject H_0 whenever $$z^* = \left| \frac{r'\hat{\beta} - \rho}{\hat{\tau}\sqrt{r'(X'X)^{-1}r}} \right| > z_{\alpha/2} \qquad (6)$$

- To test the null hypothesis H_0: $\beta_i = 0$ versus $\beta_i \neq 0$, the decision rule is

Reject H_0 whenever $$z^* = \left| \frac{\hat{\beta}_i}{\hat{\tau}\sqrt{(X'X)^{-1}_{ii}}} \right| > z_{\alpha/2} \qquad (7)$$

The statistical inference procedures for small sample size have been investigated by Dielman and Pfaffenberger (1990, 1992) and Dielman and Rose (1995), and Stangenhaus and Narula (1991) using Monte Carlo studies. Their results show that the statistical inference procedures based on normal distribution may be used for small sample sizes also. Stangenhaus, Narula, and Ferreira (1993) have proposed bootstrap procedures to draw statistical inference.

3.5 Variable Selection

It is generally tacitly assumed that the k regressors include all relevant variables and their functions and, at times, may include a few extraneous variables and their functions. Often it is possible to select a model with m (< k) variables without essentially losing any information about the response variable contained in the k regressors. A simplified model may also lead to a better understanding of the phenomenon under investigation. If prediction is the major objective of the model, it is well known that a model with fewer variables may be more desirable than the full model. Moreover, models with fewer variables are easier to understand, explain and less expensive to maintain. In fact, for economic, computational, and statistical reasons, it may be desirable to include fewer than k variables in the model.

An efficient implicit enumeration algorithm to find the best model with m (= 1, 2, ..., k-1) variables without examining all models with m variables was proposed by Narula and Wellington (1979). A computer program based on their algorithm appears in Wellington and Narula (1981). Recently, Andre, Elian, Narula, and Aubin (1996) have proposed stepwise procedures for selection of variables.

In most practical problems, as a rule, there does not exist a single "best" model but rather many "equally good" models. One possible method to select a model, from among a few good models, is to compute the sum of predictive absolute errors SPAE for each model as follows:

Leave out an observation, i say, and fit the model to the remaining n-1 observations. Predict the value of the response variable for the i-th observation using this model. Compute the difference between the observed and predicted values of the response variable for the i-th observation. Repeat this operation for each observation and compute the sum of the predictive absolute errors. The sum of predictive absolute errors is given by

$$SPAE = \sum_{i=1}^{n} |y_i - \hat{y}_{(i)}|, \tag{8}$$

where y_i is the observed value of the i-th observation and $\hat{y}_{(i)}$ is its predicted value without using this observation to estimate the parameters of the model. Then choose the model which minimizes the sum of predictive absolute errors. We hasten to add that this process of computing SPAE is computationally very intensive.

Another way to compare the models is to measure the goodness of fit of each model by its coefficient of determination R_2 proposed by McKean and Sievers (1987). Let RSAE denote the reduction in sum of absolute errors because of fitting a p-variable model, i.e.,

$$\text{RSAE} = \sum_{i=1}^{n} |y_i - median(y_i)| - \text{SAE}, \tag{9}$$

where $\sum_{i=1}^{n} |y_i - median(y_i)|$ is the sum of absolute errors for the model with no predictor variable and SAE is the sum of absolute errors associated with a p-variable model. They recommended

$$R_2 = \text{RSAE}/(\text{RSAE} + (n - p - 1)(\hat{\tau}/2)), \tag{10}$$

where $\hat{\tau}$ is given by (4). Although desirable that R_2 increases as a variable is added to a model, this is true for R_2 only if the estimate of τ is smaller for the larger model. This typically occurs, but it is not guaranteed.

In selecting the final model, however, one should always use experience, professional judgment in the subject area, and other practical and economic considerations.

3.6 Robustness

Appa and Smith (1973) have shown that (i) at least one hyperplane that minimizes the sum of absolute errors passes through k of the n observations; and (ii) under the assumption that no set of k + 1 observations lie on one hyperplane in k dimensions, the hyperplane cannot be optimal under MSAE regression unless $|n^+ - n^-| \leq k$, where n^+ and n^- denote the number of observations with positive and negative residuals, respectively.

Clearly, at least one MSAE regression hyperplane passes through k of the n observations. That is, the observed residual $e_i = y_i - \hat{y}_i$ is equal to zero for at least k observations where $\hat{y}_i$ is the predicted value of the response variable for the i-th observation. The observations with zero residuals are called basic (or defining) observations; and the others are called the nonbasic (or nondefining) observations. This also implies that the MSAE estimates are completely determined by the basic observations.

It is useful to observe that the MSAE regression is to the least squares regression what the sample median is to the sample mean. For example, both the sample mean and the least squares estimators are determined and influenced by all the observations; whereas the sample median and the MSAE estimators are determined by only a subset of observations. Just as the value of the sample median is unaffected if the magnitude of an observation is changed such that it remains on the same side of (either above or below) the sample median, a similar results is true for the MSAE regression, Narula and Wellington (1985). That is, the MSAE estimators are not altered by changes in the values of the response variable associated with the non-zero residuals as long as these observations remain on the same side of the MSAE hyperplane. This is very unlike the least squares regression, where any change in the values of an observation results in a change in the values of the least squares estimates of the parameters.

Furthermore, the fitted MSAE regression remains unchanged if the value of a predictor variable for an observation with nonzero residual remains within certain intervals, keeping the values of all other observations unchanged. The procedures to compute such intervals for simple linear regression problem have been proposed, Narula, Sposito and Wellington (1993). These intervals give the analyst useful information about the error that can be tolerated in the value of a variable in an observation without changing the fitted MSAE regression. Recently, Narula and Wellington (1997) have extended these results to find the intervals for the values of the predictor variables (changing only one value at a time) which leave the MSAE fitted model unchanged in the multiple linear regression model.

4. THE MSAE ANALYSIS FOR THE EXAMPLE

After consultations, it was decided that a model with fewer than three variables may be too small whereas a model with more than six variables may not be more useful than one with fewer variables. A way to select the best model with p predictors variables (p = 3,...,6) using the MSAE criterion is to determine the set of p variables from among $\binom{k}{p}=\binom{14}{p}$ possible subsets of size p that results in the smallest MSAE value. Using the computer program given in Wellington and Narula (1981), we found the following models with three to six variables with the minimum sum of absolute errors:

Three-variable model:

$$y = 67.76 - 0.062\, x_4 + 141.76\, x_8 - 9.53\, x_{13}, \qquad (11)$$

with R_2, the coefficient of determination for MSAE regression given in (10), equal to 0,600.

Four-variable model:

$$y = 56.55 + 0.423\, x_2 - 0.069\, x_4 + 116.67\, x_8 - 10.39\, x_{13}, \qquad (12)$$

with $R_2 = 0.632$.

Five-variable model:

$$y = 76.77 - 8.29\, x_1 + 0.249\, x_2 - 0.063\, x_4 + 109.74\, x_8 - 11.11\, x_{13}, \qquad (13)$$

with $R_2 = 0.698$.

Six-variable model:

$$y = 81.59 -8.76\, x_1 + 0.223\, x_2 - 0.061\, x_4 + 95.35\, x_8 - 1.40\, x_{10} - 9.87\, x_{13}, \qquad (14)$$

with $R_2 = 0.720$.

The minimum sum of absolute errors model found by the stepwise procedure proposed by Andre, Elian, Narula, and Aubin (1996) also selected the model with six variables given in (14).

In this problem, the SPAE for the models with three, four, five and six variables presented above are: 231.37, 195.98, 222.37 and 237.71, respectively. So the model chosen by this criterion is the four variable model. For more details for the model, see Appendix C. Observe that the variables in this model are the same as those selected by the usual stepwise procedure in the least squares regression. Furthermore, the MSAE and the least squares estimates without the outliers are also very close to each other, see

Appendix C. This shows that the MSAE coefficients are not affected by the two outliers. The standard errors of the MSAE estimates are of the same magnitude as the standard errors of the least squares estimates.

It is interesting to note that the selected model includes morphological variables relevant for pathogenesis of the Interstitial Lung Disease (ILD). For instance, epithelial cells (EPIT, x_4) and honeycombing (HONEY, x_{13}), which have a negative coefficient in the model, are widely accepted as markers of severity of ILD. Furthermore, cellular infiltration (CELL, x_8), which has a positive sign in the model, is a marker of the early phase of ILD. That is, the selected model makes sense from the medical point of view.

Because the selected model has five parameters, the fitted model goes through five observation, namely, observations 2, 3, 4, 21, and 22. For model (12), in Table 1, we give intervals which leave the fitted MSAE model unchanged. From Table 1, keeping all other values fixed, for observation 12, x_{13} can have any value between zero and four (which is the maximum value for x_{13}) without changing the fitted model; whereas for observation number 13, x_{13} has to be zero; any change in its value may change the fitted model. This is very useful as it shows that we need not concern ourselves with the value of x_{13} in observation 12 because it can take on any possible value and will have no affect on the fitted model; but for observation 13, we need to be extremely careful with the value of x_{13} as any value other than zero may change the MSAE estimates. The results for the remaining observations can be interpreted similarly.

It may be observed that some intervals are short and some are long; some original values are close to one end of the interval whereas other lie in the middle of the interval. For example, for observation 1, the intervals are narrow and the original values of the variables lie close to an end point of the interval; for x_2 the observed value (64) is close to the upper end of the interval (13.4, 64.4), for x_4 the observed value (192.405) is close to the lower end of the interval (189.771, 473.042), for x_8 the observed value (.231) is close to the upper end of the interval (.156, .233), and for x_{13} the observed value (4.0) is close to the lower end of the interval (3.9, 5.1). That is, for this observation, it seems more important that the observation has been correctly taken, recorded, and transmitted. On the other hand, for observation number 5, the observed values of the variables lie in the middle of the intervals except for the value of x_{13}.

Table 1: Intervals which leave the MSAE fit unchanged for the four variable model

OBS	y			x_2			x_4			x_8			x_{13}		
	L. L.		U.L	L. L.		U.L	L. L.		U.L	L. L.		U.L	L. L.		U.L
1.	55.8	56	∞	13.4	64	64.4	189.771	192.405	473.042	.156	0.231	.233	3.9	4	5.1
2.	Defining observation														
3.	Defining observation														
4.	Defining observation														
5.	69.3	83	∞	0*	41	63.4	110.598	310.136	590.773	.068	0.143	.260	0*	0	1.1
6.	47.8	59	∞	0*	42	64.4	24.059	187.597	468.234	.075	0.150	.246	1.9	3	4.1
7.	0*	51	68.5	9.5	32	82.6	125.199	405.836	661.071	.075	0.225	.300	0*	0	1.4
8.	0*	67	79.7	22.5	45	95.6	0*	100.237	974.580	.074	0.183	.258	0*	1	2.2
9.	0*	60	68.8	32.1	53	103.6	0*	144.290	273.054	.100	0.176	.251	0.9	2	2.8
10.	95.1	98	∞	0*	46	52.9	106.460	149.187	429.824	.176	0.251	.276	0*	0	1.1
11.	0*	48	80.9	21.5	44	94.6	0*	211.614	692.078	0*	0.174	.249	0* 3.2	0	1.4 4.2
12.	0*	82	85.9	34.6	44	94.6	0*	254.398	312.071	.033 .208	0.242	.208 .317	0*	0	5.7
13.	0*	86	92.8	27.9 40.8	57	30.9 107.6	0*	167.728	267.609	.144	0.203	.278	0*	0	0.6
14.	90.7	10 3	∞	0*	49	71.4	157.529	337.145	617.782	.238	0.313	.419	0*	0	1.1
15.	89.1	11 5	∞	14.4	65	87.4	0*	276.864	557.501	.131	0.206	.420	0*	0	1.1
16.	0*	64	72.5	5.9	26	76.6	28.569	309.206	432.840	.151	0.224	.299	0* 1.5	0	0.8 4.1
17.	56.8	57	∞	0*	46	46.5	169.739	173.373	454.010	.129	0.204	.206	2.9	3	4.1
18.	72.8	82	∞	0*	28	49.6	104.522	238.277	518.914	.103	0.178	.257	0*	0	1.1
19.	0*	50	58.8	31.1	52	102.6	0*	130.308	259.471	.099	0.175	.250	1.9	3	3.8
20.	0*	48	48.0	48.9	49	99.6	0*	165.546	166.250	.203	0.203	.278	2.9	4	4.0
21.	Defining observation														
22.	Defining observation														
23.	0*	77	79.2	66.8	72	122.6	326.631	607.268	639.082	.449	0.468	.543	0.9	2	2.2
24.	87.1	92	∞	6.3	57	68.5	333.220	404.735	685.372	.218	0.293	.335	0*	0	1.1

L.L. = lower limit U.L. = upper limit
*Since the variables can not take on negative values, the negative lower limit of the intervals have been replaced with zero.

5. CONCLUDING REMARKS

The minimum sum of absolute errors regression is more desirable than the least squares regression whenever (i) the errors follow a symmetric distribution for which the median is a more efficient estimator of the location parameter than the sample mean; or (ii) the errors follow a long tailed error distribution; or (iii) there are outliers in the data; or (iv) there is multicollinearity among the variables; or (v) the absolute error loss function is more appropriate than the quadratic loss function. It also provides a good starting solution for a number of robust regression procedures.

The example presented illustrated the desirable behavior of the MSAE regression in the presence of multicollinearity among the predictor variables and outliers in a data set. Furthermore, the intervals on the values of a predictor (response) variable which leave the fitted MSAE regression unchanged provide useful information to the scientist.

ACKNOWLEDGMENT

We thank FAPESP for the grant which partially supported the work of Professor Narula. We also thank Professor John F. Wellington of Capital University, Columbus, Ohio, for supplying the values in Table 1.

REFERENCES

Abdelmalek, N. N. (1980). L_1 solution of overdetermined system of linear equations. *ACM Transactions on Mathematical Software, 6*, 220-227.

André, C. S. D., Elian, S. N., Narula, S. C., and Aubin, E. C. Q. (1996). Stepwise procedure for selecting variables in the minimum sum of absolute errors regression. Technical Report RT-MAE 9619, Instituto de Matemática e Estatística da Universidade de São Paulo, São Paulo, Brazil.

Appa, G. and Smith, C. (1973). On L_1 and Chebychev estimation. *Mathematical Programming, 5,* 73 - 87.

Arthanari, T. S. and Dodge, Y. (1981). *Mathematical Programming in Statistics.* John Wiley and Sons, Inc., New York, N. Y.

American Thoracic Society (1991). Lung infection testing: Selection of reference values and interpretative strategies. *American Review of Respiratory Diseases, 144*, 1202-1218.

Armstrong, R. D., Frome, E., and Kung, D. S. (1979). A revised simplex algorithm for the absolute deviation curve fitting problem. *Communications in Statistics, B8,* 175-190.

Armstrong, R. D. and Kung, M. T. (1978). AS132: Least absolute value estimates for a simple linear regression problem. *Applied Statistics, 27,* 363-366.

Barrodale, I. and Roberts, F. D. K. (1973). An improved algorithm for Discrete l_1 linear approximation. *SIAM Journal on Numerical Analysis, 10,* 839-848.

Barrodale, I. and Roberts, F. D. K. (1974). Solution of an overdetermined system of equations in the L_1 norm. *Communications of the ACM, 17,* 319-320.

Bartels, R. H., Conn, A. R., and Sinclair, J. W. (1976). A FORTRAN program for solving overdetermined systems of linear equations in the L_1 sense. Technical Report No. 236, Mathematical Sciences Department, John Hopkins University, Baltimore, MD, USA.

Bartels, R. H., Conn, A. R., and Sinclair, J. W. (1978). Minimization techniques for piecewise differentiable functions: The L_1 solution to an overdetermined linear system. *SIAM Journal of Numerical Analysis, 15,* 224-241.

Boscovich, R. J. (1757). Delitteraria expeditone per pontificiam ditionem, et synopsis amplioris operis, ac habentur plura ejus ex exemplaria etiam sensorum impressa. *Bononiesi Sientiarum st Artum Institute Atque Academia Commentarii, 4,* 353 - 396.

Basset, G. and Koenker, R. (1978). Asymptotic theory of least absolute error regression. *Journal of the American Statistical Association, 73,* 618-622.

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

Bloomfield, P. and Steiger, W. (1983). *Least Absolute Deviations: Theory, Applications, and Algorithms.* Birkhauser, Boston, MA.

Carington, C. B., Gaensler, E. A., Coutu, R. E., Fitzgerald, M. X., and Grupta, R. G. (1978). Natural history and treated course of usual and desquamative interstitial pneumonia. *New England Journal of Medicine, 298,* 801-809.

Charnes, A., Cooper, W. W., and Ferguson, R. D. (1955). Optimal estimation of executive compensation by linear programming. *Management Science, 1,* 138 -151.

Cherniak, R. M., Colby, T. V., Flint, A., Thurlbbeck, W. M., Waldron, J., Ackerson, L., and King Jr., T. E. and the BAL Cooperative Group Steering Committee (1991). Quantitative assessment of lung pathology in idiophatic pulmonary fibrosis. *American Review of Respiratory Diseases, 144,* 892-900.

Coleman, T. F. and Li, Y. (1992). A globally and quadratically convergent affine scaling method for linear l_1 problems. *Mathematical Programming, 56,* 189-222.

Crystal, R. G., Bitterman, P. B., Rennard, S. I., Hance, A. J., and Keogh, B. A. (1984). Interstitial lung disease of unknown cause: Disorders characterized by chronic inflammation of the lower respiratory tract (First of two parts). *New England Journal of Medicine, 19*, 154-166.

Dielman, T. and Pfaffenberger, R. (1982). LAV (Least absolute value) estimation in the regression model: A review. *TIMS Studies in the Management Sciences, 19*, 31 - 52.

Dielman, T. and Pfaffenberger, R. (1990). Tests of linear hypotheses in LAV regression. *Communication in Statistics: Simulation and Computation, 19*, 1179-1199.

Dielman, T. and Pfaffenberger, R. (1990). A further comparison of tests of hypotheses in LAV regression. *Computational Statistics and Data Analysis, 14*, 375-384.

Dielman, T. and Rose, E. L. (1995). A bootstrap approach to hypothesis testing in least absolute value regression. *Computational Statistics and Data Analysis, 20*, 119-130.

Dodge, Y. (1987). *Statistical Data Analysis: Based on the L_1-Norm and Related Methods.* North-Holland, Amsterdam, Holland.

Edgeworth, F. Y. (1887). On observations relating to several quantities. *Hermathena, 6*, 279-285.

Edgeworth, F. Y. (1888). On a new method of reducing observations relating to several quantities. *Philosophical Magazine, 25*, 184-191.

Farebrother, R. W. (1987). The historical development of the L_1 and L_∞ estimation procedures. *Statistical Data Analysis Based on the L_1-Norm and Related Methods.* (Y. Dodge, editor). North-Holland, Amsterdam, Holland, 37-64.

Fulmer, J. D., Robert, W. C., von Gal, E. R., and Crystal, R. G. (1979). Morphologic-physiologic correlates of the severity of fibrosis and degree of cellularity in idiophatic pulmonary fibrosis. *Journal of Clinical Investigations, 63*, 665-676.

Gentle, J. E., Sposito, V. A., and Narula, S. C. (1988). Algorithms for unconstrained L_1 simple linear regression. *Computational Statistics and Data Analysis, 6*, 335-339.

Gentle, J. E., Sposito, V. A., and Narula, S. C. (1988). Algorithms for unconstrained L_1 linear regression. *Statistical Data Analysis Based on the L_1-Norm and Related Methods.* (Y. Dodge, editor). North-Holland, Amsterdam, Holland, 83-94.

Hampel, F. R. (1971). A general qualitative definition of robustness. *Annals of Mathematical Statistics, 42*, 1887 - 1896.

Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Annals of Statistics, 1*, 799 - 821.

International Mathematical and Statistical Libraries, Inc. (1980). *IMSL Library and Reference Manual*, Houston, Texas.

Josavanger, L. A. and Sposito, V. A. (1983). L_1- norm estimates for the simple regression problem. *Communications i Statistics - Simulation and Computation, B12,* 215 -221.

Junqueira, L. C. U., Silva, M. D. A., and Torloni, H. (1989). A simple procedure to obtain one-micrometer sections of routinely embedded paraffin material. *Stain Technology, 64,* 39-42.

Karst, O. J. (1958). Linear curve fitting using least deviations. *Journal of the American Statistical Association, 53,* 118 - 132.

Katzeinstein, A. L. A., and Askin, F. B. (1990). *Surgical Pathology of Non-Neoplastic Lung Disease, Second Edition.* W. B. Saunders Co.

Klingman, D. and Mote, J. (1982). Generalized network approaches for solving least absolute value and Tchebycheff regression problem. *TIMS Studies in Management Sciences, 19,* 53-66.

Madsen, K. and Nielsen, H. B. (1993). A finite smoothing algorithm for linear l_1 estimation. *SIAM Journal on Optimization, 3,* 223-235.

McKean, J. W. and Schrader, R. M. (1987). Least absolute errors analysis of variance. *Statistical Data Analysis Based on the L_1-Norm and Related Methods* (Y. Dodge, editor), Elsevier Science Publishers B. V., 297 - 305.

McKean, J. W. and Sievers, G. L. (1987). Coefficient of Determination for least absolute deviation analysis. *Statistics and Probability Letters, 5,* 49-54

Morris, J. F., Koski, A., and Johnson, L. C. (1971). Spirometric standards for healthy non-smoking adults. . *American Review of Respiratory Diseases,* 163, 57-67.

Narula, S. C. (1987). The minimum sum of absolute errors regression. *Journal of Quality Technology, 19,* 37-45.

Narula, S. C., Sposito, V. A. and Wellington, J. F. (1993). Intervals which leave the minimum sum of absolute errors regression unchanged. *Applied Statistics, 42,* 369-378.

Narula, S. C. and Wellington, J. F. (1977). Prediction, linear regression and minimum sum of absolute errors. *Technometrics, 19,* 185-190.

Narula, S. C. and Wellington, J. F. (1979). Selection of variables in linear regression using the minimum sum of weighted absolute errors criterion. *Technometrics, 21,* 299-306.

Narula, S. C. and Wellington, J. F. (1982). The minimum sum of absolute errors regression: A state of the art survey. *International Statistical Review, 50,* 317-326.

Narula, S. C. and Wellington, J. F. (1983). Selection of variables in linear regression: A pragmatic approach. Journal of Statistical Computation and Simulation, 17, 159 - 172.

Narula, S. C. and Wellington, J. F. (1985). Interior analysis for the minimum sum of absolute errors regression. Technometrics, 27, 181 -188.

Narula, S. C. and Wellington, J. F. (1997). Intervals which leave the MSAE multiple linear regression unchanged. School of Business, Virginia Commonwealth University, Richmond, Virginia, USA

Ruzinsky, S. and Olsen, E. (1989). L_1 and L_∞ Minimization via a variant of Karmarkar's algorithm. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 37,* 245-253.

Saldiva, P. H. N., Brentani, M. M., Carvalho, C. C. R., Auler Jr., J. O. C., Carheiros, D. F., and Pacheco, M. M. (1985). Changes in the pulmonary glucocorticoid receptor content in the course of interstitial disease. *Chest., 88*, 417-419.

SAS Institute, Inc. (1983). *SUGI Supplemental Library User's Guide, 1983 Edition.* SAS Institute, Inc., Cary, N. C.

Sposito, V. A., Smith, W. C., and McCormick, C. (1978). *Minimizing the Sum of Absolute Deviations.* Vandenhoeck and Ruprecht, Gottington, Germany.

Stangenhaus, G. and Narula, S. C. (1991). Inference procedures for the L_1 regression. *Computational Statistics and Data Analysis, 12*, 1, 79-85.

Stangenhaus, G., Narula, S. C. and Ferreira, P. Fo. (1993). Bootstrap confidence intervals for the minimum sum of absolute errors regression. Journal of Statistical *Computation and Simulation, 48*, 127-133.

Wagner, H. M. (1959). Linear programming techniques for regression analysis. *Journal of the American Statistical Association, 54*, 206 -212.

Watters, L. C., King, T. E., Schwarz, M. I., Waldron, J. A., Styanford, R. E., and Cherniak, R. M.(1986). A clinical, radiographic and physiologic scoring system for the longitudinal assessment of patient with idiophatic pulmonary fibrosis. . *American Review of Respiratory Diseases, 133*, 97-103.

Wellington, J. F. and Narula, S. C. (1981). Variable selection in multiple linear regression using the minimum sum of weighted absolute errors criterion. *Communications in Statistics - Simulation and Computations, B10*, 641-648.

Wesolowsky, G. O. (1981). A new descent algorithm for the least absolute value regression problem. *Communications in Statistics - Simulation and Computations, B10,* 479-491.

Zhang, Y. (1993). Primal-dual interior point approach for computing l_1 solutions and l_∞ solutions of overdetermined systems. *Journal of Optimization Theory and Applications, 77,* 323-341.

APPENDIX A

DATA FOR THE STUDY

The List of Variables

Response variable:

y = FVC (forced vital capacity):

Predictor variables:

x_1 = SEX: 1 = Male, 2 = Female

x_2 = AGE (in years)

x_3 = SMOK (smoking): 0 = Smoker, 1 = Nonsmoker

x_4 = EPIT (epithelial cells): Area fraction of epitelial cells/10 000 μm^2 of alveolar tissue;

x_5 = FUSI (elongated cells): Area fraction of fusiform cells/10 000 μm^2 of alveolar tissue;

x_6 = MONO (mononucleated cells): Area fraction of mononucleated cells/10 000 μm^2 of alveolar tissue;

x_7 = POLY (polymorphonuclear cells): Area fraction of polymorphonuclear cells /10 000 μm^2 of alveolar tissue;

x_8 = CELL (cellular infiltration): Total cellularity/10 000 μm^2 of alveolar tissue;

x_9 = VES (septal vascularization): Area fraction of capillaries/10 000 μm^2 of alveolar tissue;

x_{10} = BOBLIT (obliterate bronchiolitis): Score of bronchiolitis obliterans (zero to four);

x_{11} = MUSCLE (smooth muscle hyperplasia): Score of smooth muscle (zero to four);

x_{12} = SCLEVASC (sclerosis): Score of vascular sclerosis (zero to four);

x_{13} = HONEY (honeycombing): Score of honeycombing (zero to four);

x_{14} = DESQ (desquamative pneumonia): Score of intra alveolar cell disquamation (zero to four);

The data for the example are given in Table A.1.

Table A.1 Data for the example

	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}
1.	56	1	64	0	192.405	359.71	669.24	0.000	0.231	0.289	3	4	1	4	2
2.	75	2	39	0	398.588	441.53	163.06	20.706	0.251	0.578	0	0	3	0	0
3.	32	2	39	0	671.674	622.29	1728.57	49.308	0.043	0.203	3	0	0	0	0
4.	88	1	69	1	227.424	539.19	145.42	13.424	0.153	0.615	0	0	0	0	0
5.	83	1	41	0	310.136	419.39	88.11	3.525	0.143	0.551	0	0	0	0	0
6.	59	1	42	1	187.597	378.95	82.54	1.251	0.150	0.785	0	4	3	3	2
7.	51	1	32	1	405.836	411.85	261.54	30.062	0.225	0.240	0	0	0	0	0
8.	67	1	45	1	100.237	346.53	223.38	2.864	0.183	0.725	0	0	3	1	2
9.	60	2	53	0	144.290	397.77	129.99	0.000	0.176	0.696	3	1	4	2	0
10.	98	1	46	1	149.187	275.22	204.49	14.147	0.251	0.577	0	0	2	0	4
11.	48	2	44	0	211.614	398.81	278.35	4.883	0.174	0.703	0	1	3	0	2
12.	82	1	44	0	254.398	376.39	297.54	7.439	0.242	0.593	3	2	0	0	2
13.	86	2	57	0	167.728	384.79	4624.07	3.289	0.203	0.702	0	1	2	0	0
14.	103	2	49	0	337.145	597.76	614.08	12.410	0.313	0.554	0	0	2	0	0
15.	115	2	65	0	276.864	365.31	401.66	8.206	0.206	0.572	0	0	2	0	3
16.	64	2	26	0	309.206	512.22	99.65	27.510	0.224	0.579	0	0	3	0	1
17.	57	1	46	1	173.373	367.14	308.02	24.222	0.204	0.722	0	2	3	3	2
18.	82	1	28	1	238.277	375.29	223.85	64.037	0.178	0.685	0	0	2	0	1
19.	50	2	52	1	130.308	374.79	423.90	34.747	0.175	0.697	3	2	3	3	2
20.	48	1	49	1	165.546	318.45	284.34	37.911	0.203	0.674	4	2	2	4	2
21.	57	2	32	0	168.547	394.52	282.60	1.349	0.165	0.647	0	2	4	2	2
22.	45	1	57	0	621.861	1477.29	416.57	151.746	0.238	0.685	2	3	2	2	2
23.	77	1	72	0	607.268	171.91	2529.51	89.094	0.468	0.435	0	0	2	2	2
24.	92	1	57	1	404.735	1443.59	2022.71	93.677	0.293	0.618	0	0	3	0	0

APPENDIX B

THE LEAST SQUARES ANALYSIS

The results of the least squares for the full model are:

The full model (with all the predictor variables) is:

$$y = -7.2 + 6.6\ x_1 + 0.257\ x_2 + 3.7\ x_4 - 0.00003\ x_5 + 0.0201\ x_6 + 0.00033\ x_7 - 0.251\ x_8 + 130\ x_9 + 75.6\ x_{10} - 2.12\ x_{11} - 7.07\ x_{12} - 7.28\ x_{13} + 3.86\ x_{14}$$

Table B.1 The details for the model with all the variables

Predictor	Coefficient	St. dev	t-ratio	p-value
constant	-7.22	76.07	-0.09	0.926
x_1	6.56	15.42	0.43	0.681
x_2	0.2573	0.5144	0.50	0.629
x_3	3.70	14.24	0.26	0.801
x_4	-0.00003	0.09219	-0.00	1.000
x_5	0.02015	0.02555	0.79	0.451
x_6	0.000325	0.004433	0.07	0.943
x_7	-0.2507	0.3390	-0.74	0.478
x_8	129.82	73.06	1.78	0.109
x_9	75.60	84.49	0.89	0.394
x_{10}	-2.118	4.036	-0.52	0.612
x_{11}	-7.073	7.728	-0.92	0.384
x_{12}	-7.282	7.696	-0.95	0.369
x_{13}	-3.175	7.762	-0.41	0.692
x_{14}	3.862	4.759	0.81	0.438

Table B.2 Analysis of variance table

Source	DF	SS	MS	F	p
Regression	14	7713.0	550.9	2.09	.135
Error	9	2377.0	264.1		
Total	23	10090.0			

Notice that all the variables and the regression model are non significant at the 0.05 level. The estimated standard deviation of the error distribution is 16.25 and R^2 for the model is 0.764.

APPENDIX C

COMPARISON OF THE FOUR VARIABLE MODELS SELECTED BY THE LEAST SQUARES AND THE MSAE REGRESSION

For the four variable models selected by the least squares and the MSAE regression, the estimated values of the coefficients and their standard deviations (in parentheses) are given in Table C.1.

Table C.1 The coefficient (standard deviation) for the four variables models

Method	Remarks	Constant	x_2	x_4	x_8	x_{13}
LS	All Obs.	46.65 (11.28)	0.614 (0.2364)	-0.061 (0.0174)	107.73 (37.92)	-10.64 (1.936)
LS	Without Obs. 11, 15	54.84 (8.196)	0.439 (0.1823)	-0.064 (0.0125)	112.57 (27.30)	-10.48 (1.459)
MSAE	All Obs.	56.55 (13.231)	0.423 (0.2828)	-0.069 (0.0205)	116.67 (44.49)	-10.39 (2.272)

ÚLTIMOS RELATÓRIOS TÉCNICOS PUBLICADOS

9701 - **BOFARINE, H.; ARELLANO-VALE, R.B.** Weak nondifferential measurement error models. IME-USP, 1997. 12p. (RT-MAE-9701)

9702 - **FERRARI, S.L.P.; CORDEIRO, G. M.; CRIBARI-NETO, F.** Higher Order Asymptotic Refinements for Score Tests in Proper Dispersion Models. IME-USP, 1997. 14p. (RT-MAE-9702)

9703 - **DOREA, C.; GALVES, A.; KIRA, E.; ALENCAR, A. P.** Markovian modeling of the stress contours of Brazilian and European Portuguese. IME-USP, 1997. 11p. (RT-MAE-9703)

9704 - **FONTES, L.R.G.; ISOPI, M.; KOHAYAKAWA, Y.; PICCO, P.** The Spectral Gap of the REM under Metropolis Dynamics. IME-USP, 1997. 24p. (RT-MAE-9704)

9705 - **GIMENEZ, P.; BOLFARINE, H.; COLOSIMO, E.A.** Estimation in weibull regression model with measurement error.IME-USP, 1997. 17p. (RT-MAE-9705)

9706 - **GALVES, A.; GUIOL, H.** Relaxation time of the one-dimensional symmetric zero range process with constant rate. IME-USP, 1997. 10p. (RT-MAE-9706)

9707 - **GUIOL, H.** A note about a burton keane's theorem. IME-USP, 1997. 7p. (RT-MAE-9707)

9708 - **ARELLANO-VALLE,R.B.; FERRARI,S.L.P.; CRIBARI-NETO,F.** Bartlett and barblett-type corrections for testing linear restrictions. IME-USP, 1997. 5p. (RT-MAE-9708)

9709 - **PAULA, G.A.** One-Sided Tests in Dose-Response Models. IME-USP, 1997. 29p. (RT-MAE-9709)

9710 - **BRESSAUD, X.** Subshift on an infinite alphabet. IME-USP, 1997. 34p. (RT-MAE-9710)

9711 - **YOSHIDA, O.S.; LEITE, J.G.; BOLFARINE, H.** Stochastic Monotonicity properties of bayes estimation of the population size for capture-recapture data. IME-USP. 1997. 12p. (RT-MAE-9711)

9712 - **GUIMENEZ,P.;COLOSIMO,E.A.; BOLFARINE,H.** Asymptotic relative efficiency of wald tests in measurement error models. IME-USP. 1997. 14p. (RT-MAE-9712)

9713 - **SCHONMANN, R.H.; TANAKA, N.I.** Lack of monotonicity in ferromagnetic ising model phase diagrams. IME-UPS. 1997. 12p. (RT-MAE-9713)

9714 - **LEITE,J.G.; TORRES, V.H.S.; TIWARI, R.C.; ZALKIKAR, J.** Bayes estimation of dirichlet process parameter. IME-USP. 1997. 21p. (RT-MAE-9714)

9715 - **GIMENEZ, P.; BOLFARINE, H.; COLOSIMO, E.** Hypotheses testing based on a corrected score function for errors-in-variables models. 1997. 16p. (RT-MAE-9715)

9716 - **PAULA, G.A.; BOLFARINE,H.** Some results on the slope of the linear regression model for the analysis of pretest-posttest data. 1997. 08p. (RT-MAE-9716)

9717 - **AUBIN, E.C.Q.; CORDEIRO, G.M.** Some adjusted likelihood ratio tests for heteroscedastic regression models. 1997. 19p. (RT-MAE-9717)

9718 - **ANDRÉ, C.D.S.; NARULA, S.C.** Statistical inference for the parameters of a one-stage dose-response model using the minimum sum absolute errors estimators. 1997. 09p. (RT-MAE-9718)

9719 - **ANDRÉ, C.D.S.; NARULA, S.C.** Statistical inference for the parameters of a two-stage dose-response model using the minimum sum of absolute errors estimators. 1997. 15p. (RT-MAE-9719)

9720 - **COLLET, P.; GALVES, A.; SCHMITT, B.** Fluctuations of repetition time for gibbsian sources. 1997. 08p. (RT-MAE-9720)

9721 - **CORDEIRO, G.M.; FERRARI, S.L.P.;CYSNEIROS, A.H.M.A** A formula to improve score test statistics. 1997. 10p. (RT-MAE-9721)

9722 - **ZUAZOLA, P.I.;GASCO-CAMPOS,L.; BOLFARINE, H.** Pearson type II measurement error models. 1997. 24p. (RT-MAE-9722)

9723 - BUENO, V.C. A note on asymptotics reliability of a linearly connected system. 1997. 07p. (RT-MAE-9723)

The complete list of "Relatórios do Departamento de Estatística", IME-USP, will be sent upon request.

Departamento de Estatística
IME-USP
Caixa Postal 66.281
05315-970 - São Paulo, Brasil