L_1-Norm Based Data Analysis and Related Methods

(1632-1989)

Bijan Bidabad[1]

Keywords: L_1 norm, Regression, Algorithm, Computer

Abstract:

This paper gives a rather general view on the L_1-norm criterion on the area of data analysis and related topics. We tried to cover all aspects of mathematical properties, historical development, computational algorithms, simultaneous equations estimation, statistical modeling, and application of the L_1-norm in different fields of sciences.

<u>1. Introduction</u>

[1] Research professor of economics, Monetary and Banking Research Academy, Central Bank of Iran, Pasdaran, Zarabkhaneh, Tehran, 16619, Iran. bijan_bidabad@msn.com.
I should express my sincere thanks to Professor Yadolah Dodge from Neuchatel University of Switzerland who taught me many things about L_1 norm when he was my Ph.D.dissertation advisor.

Although the L_1-norm is an old topic in science, but lack of a general book or paper on this subject induced me to gather a relatively complete list of references in this paper. The methods related to L_1-norm are very broad and summarizing them is very difficult. However, it has been tried to have a glance on almost all related areas. The sections are designed as separate modules, so that the reader may skip some of the sections without loss of continuity of the subject.

While the least squares method of estimation of the regression parameters is the most commonly used procedure, some alternative techniques have received widespread attention in recent years. Conventionally, interest in other methods of estimation has been generated by the unsatisfactory performance of least squares estimators in certain situations when some model assumptions fail to hold or when large correlations exist among the regressors. However, the least squares regression is very far from optimal in many non-Gaussian situations, especially when the errors follow distributions with longer tails. In particular, when the variance of the error is infinite. While intuition may dispel consideration of errors with infinite variance, in many cases, studies have shown that, in fact, certain distributions with infinite variances may be quite appropriate models. An infinite variance means thick tail error distribution with lots of outliers. Of course, observed distributions of economic variables will never display infinite variances. However, the important issue is not that the second moment is actually infinite, but the interdecile range in relation to interquartile range is sufficiently large that one is justified in acting as though the variance is infinite. Even when the majority of the errors in the model follow a normal distribution, it often occurs that a small number of observations are from a different distribution. That is the sample is contaminated with outliers. Since least squares gives a lot of weight to outliers, it becomes extremely sample dependent and it is well known that the performance of this estimator is markedly degraded in this situation. It has been stated that even when errors follow a normal distribution, alternative to least squares may be required; especially if the form of the model is not exactly known or any other specification error exists. Further, least squares is not very satisfactory if the quadratic loss function is not a satisfactory measure of loss. Loss denotes the seriousness of the nonzero prediction error to the investigator, where prediction error is the difference between the predicted and the observed values of the response variable. It has been shown that for certain economic problems least absolute errors gives more satisfactory results than least squares, because the former is less sensitive than the latter to extreme errors, and consequently is resistant to outliers. It should be noted that the least absolute errors estimates have maximum likelihood properties and hence are asymptotically efficient when the errors follow the Laplace distribution.

Although least absolute errors estimator is very old, it has emerged in the literature again and has attracted attention in the last two decades because of unsatisfactory properties of least squares. Now, this method is discussed in the econometrics textbooks such as Kmenta (1986) and Maddala (1977). Many Master's and Ph.D dissertations have been written on this subject in different departments such as Lawson (1961), Burgoyne (1965), Gentleman (1965), Barrodale (1967), Oveson (1968), Lewis (1969), Cline (1970), Hunt (1970), Groucher (1971), Henriksson (1972), Bassett (1973), Forth (1974), Anderson (1975), Ronner (1977), Nyquist (1980), Clarke (1981), Kotiuga (1981), Gonin (1983), Busovaca (1985), Kim ( ), Bidabad (1989a,b) which are more recent (see bibliography for the corresponding departments and universities).

Robust property of this estimator is its advantage to deal with large variance error distributions. Since many economic phenomena such as distribution of personal income, security returns, speculative prices, stock and commodity prices, employment, asset size of business firms, demand equations, interest rate, treasury cash flows, insurance, price expectations, and many other economic variables fall within the category of infinite variance (see, Ganger and Orr (1972), Nyquist and Westlund (1977), Fama (1965), Goldfeld and Quandt (1981), Sharpe (1971)) it is necessary to turn the economists attention to this estimator. There are many other works which confirm the superiority of least absolute to least squares estimator such as interindustry demand analysis of Arrow and Hoffenberg (1959), investment models of Meyer and Glauber (1964), security and portfolios analysis of Sharpe (1971), Danish investment analysis of Kaergard (1987) and so forth.

Many new economic theories weaken the assumption of rationality of human behavior. This relative irrationality is a major source of large variances and outliers in economic data. Therefore the least absolute errors estimator becomes a relevant estimator in the cases that rationality is a strong assumption.

Another major application of this estimator is on data with measurement errors. This type of errors makes variances large and forces the observations to locate far from reality which obviously causes outliers. Existence of two important types of measurement errors, sampling and non sampling errors, specifically in countries with poor statistics such as developing countries make this estimator a basic tool of analysis.

Unknown specification errors in regression models because of complexity of human behavior are always occurred in mathematical formulation of human related problems. Specification error occurs whenever formulation of the regression equation or one of the underlying assumptions is incorrect. In this context when any assumption of the underlying theory or the formulation of the model does not hold, a relevant explanatory variable is omitted or an irrelevant one is included, qualitative change of the explanatory variable is disregarded, incorrect mathematical form of the regression is adopted, or incorrect specification of the way in which the disturbance enters the regression equation is used and so on; specification error exits (see also, Kmenta (1986)). Since specification errors are not always clear to researcher, least squares is a poor estimator and other alternatives as least absolute errors estimators become attractive.

Although least absolute errors estimator benefits from optimal properties in many econometric problems, it is not a commonly used tool. This is to some extent due to difficulties of calculus with absolute value functions. When the model is enlarged and equations enter simultaneously, difficulties of computation increase. Another problem with this estimator is that the properties of the solution space is not completely clear and the corresponding closed form of the solution have not been derived yet.

Thus these three important problems of algebraic closed form, computational difficulties and solution space properties are the main obstacles that prevent the regular use of L_1-norm estimator. Any attempt to remove these obstacles are worthy.

2. L_p norm and regression analysis

Given a point $\mathbf{u}=(u_1,...,u_n)$ in $\mathbf{R}^n$, Minkowski norm or L_p norm can be written as the following expression,

$$||\mathbf{u}||_p = d_p(\mathbf{u},\mathbf{0}) = [\sum_{i=1}^{n} |u_i|^p]^{1/p} \tag{1}$$

When p=2 we are confronted with Euclidian or L_2 norm. Thus, Euclidian distance is a special case of L_p distance (see, Ralston and Rabinowitz (1985)).

The following overdetermined system of equations is given,

$$\mathbf{y} = \mathbf{X}\mathbf{ß} + \mathbf{u} \tag{2}$$

where, $\mathbf{y}$ is a nx1 vector of dependent variables, $\mathbf{X}$, a nxm matrix of independent or explanatory variables with n>m, $\mathbf{ß}$, a mx1 vector of unknown parameters and $\mathbf{u}$ is a nx1 vector of random errors. The problem is to find the unknown vector $\mathbf{ß}$ such that the estimated value of $\mathbf{y}$ be close to its observed value. A class of procedures which obtains these estimated values is L_p norm minimization criterion (see, Narula (1982)). In this class $||\mathbf{u}||_p$ is minimized to find the $\mathbf{ß}$ vector,

$$\min_{\mathbf{ß}} S = \min_{\mathbf{ß}} ||\mathbf{u}||_p = \min_{\mathbf{ß}} ||\mathbf{y}-\mathbf{Xß}||_p =$$

$$\min_{\mathbf{ß}} [\sum_{i=1}^{n} |y_i-\mathbf{x}_i\mathbf{ß}|^p]^{1/p} = \min_{\mathbf{ß}} [\sum_{i=1}^{n} |y_i-\sum_{j=1}^{m} ß_j x_{ij}|^p]^{1/p} \Longrightarrow$$

$$\min_{\mathbf{ß}} \sum_{i=1}^{n} |y_i-\mathbf{x}_i\mathbf{ß}|^p = \min_{\mathbf{ß}} \sum_{i=1}^{n} |y_i-\sum_{j=1}^{m} ß_j x_{ij}|^p \tag{3}$$

where y_i is the i^{th} element of **y** and $\mathbf{x}_i$ is the i^{th} row of the matrix **X**. Any value of $p\varepsilon[1,\infty]$ may be used to find **ß** in (3) (see, Money et al (1978a), Rice (1933)), but each value of p is relevant for special types of error distributions. Many authors have investigated this problem (see, Barrodale (1968), Barr et al (1980a,b,c,81a,b), Money et al (1978b,82), Gonin and Money (1985a,b), Sposito and Hand (1980), Sposito and Hand and Skarpness (1983), Sposito (1987b)). However, justification of p comes from the following theorem (see, Kiountouzis (1971), Rice and White (1964), Hogan (1976), Taguchi (1974,78)).

Theorem: If in model (2), **X** is nonstochastic and $E(\mathbf{u})=\mathbf{0}$, $E(\mathbf{u}\mathbf{u}^T)=\sigma^2\mathbf{I}$, and **u** distributed with $f(u)=h.\exp(-k|u|^p)$, where h and k are constants and $p\varepsilon[1,\infty]$; then the "best" **ß** with maximum likelihood properties is a vector which comes from minimization of (3).

Certain values of p have particular importances (see, Box and Tiao (1962), Theil (1965), Anscombe (1967), Zeckhauser and Thompson (1970), Blattberg and Sargent (1971), Kadiyala (1972), Maddala (1977)). L_∞ norm minimization of (3) is called Tchebyshev or uniform norm minimization or minimum maximum deviations and has the maximum likelihood properties when u has a uniform probability distribution function. When p=2 we are confronted with least squares method. In this case if the errors distribution is normal it is the best unbiased estimator (see, Anderson (1962), Theil (1971)). When p=1, we have L_1 norm or Gershgorin norm minimization problem. It is also called least or minimum sum of absolute errors (MSAE, LSAE), minimum or least absolute deviations, errors, residuals, or values (MAD, MAE, MAR, MAV, LAD, LAE, LAR, LAV), L_1 norm fit, approximation, regression or estimation.

Harter (1974a,b,75a,b,c,76) monumental papers provide a chronology of works on nearly all the estimators which includes L_1 norm estimation too. A concise review of data analysis based on the L_1 norm is presented by Dodge (1987) and a brief discussion is given by Gentle (1977) too. Narula and Wellington (1982) and Narula (1987) give a brief and concise presentation of L_1 norm regression. Blattberg and Sargent (1971) show that if the errors of the regression follow the second law of Laplace (two- tailed exponential distribution) with probability density function

$$f(u)=(1/2\theta).\exp(-|u|/\theta) \qquad (4)$$

where $var(u)=2\theta^2$, then L_1 norm minimization leads to maximum likelihood estimator.

3. Properties of the L_1 norm estimation

Similar to other criteria, the L_1 norm estimation has its own properties which are essential in computational and statistical viewpoints. The more important properties are as follows.

3.1 Invariance property

An estimator $\hat{\mathbf{ß}}(\mathbf{y},\mathbf{X})$ of population parameter **ß** is invariant if,

$$\hat{\mathbf{ß}}(\theta\mathbf{y},\mathbf{X}) = \theta\hat{\mathbf{ß}}(\mathbf{y},\mathbf{X}), \qquad \theta\varepsilon[0,\infty) \qquad (5)$$

Gentle and Sposito (1976), Koenker and Bassett (1978) have proved that the L_p norm estimator of **ß** is invariant when the regression model is linear. The L_p norm estimator is not invariant for general nonlinear models. The invariance property is the homogeneity of degree one of the **ß^** solution function.

3.2 Transformation of variables

If $\theta\varepsilon\mathbf{R}^m$, by transforming **y** to $\mathbf{y}+\mathbf{X}\theta$ the optimal value of **ß^** will increase by θ, (see, Koenker and Bassett (1978));

$$\hat{\mathbf{ß}}(\mathbf{y}+\mathbf{X}\theta,\mathbf{X}) = \hat{\mathbf{ß}}(\mathbf{y},\mathbf{X}) + \theta \qquad (6)$$

If **A** is a mxm nonsingular matrix, transformation of **X** to **XA** premultiplies optimal **ß^** by inverse of **A** (see, Taylor (1974), Koenker and Bassett (1978), Bassett and Koenker (1978)).

$$\hat{\mathbf{ß}}(\mathbf{y},\mathbf{XA}) = \mathbf{A}^{-1}\hat{\mathbf{ß}}(\mathbf{y},\mathbf{X}) \qquad (7)$$

3.3 convexity of the objective function

To show the convexity of S in (3), suppose m=1; the objective function (3) reduces to

$$S = \sum_{i=1}^{n} \left| y_i - \beta_1 x_{i1} \right| = \sum_{i=1}^{n} S_i \qquad (8)$$

where $S_i=\left| y_i-\beta_1 x_{i1} \right|$. If we plot S_i as a function of β_1 then we will have a broken line in $S x \beta_1$ plane and its function value is zero at $\beta_{i1}=y_i/x_{i1}$. The slope of the half-lines to the left and right of β_{i1} are $-\left| x_{i1} \right|$ and $\left| x_{i1} \right|$ respectively. So, S_i's are all convex and hence their sum S is also convex with slope at any β_1 equal to the sum of the slopes of the S_i's at that value of β_1 (see, Karst (1958), Taylor (1974)).

Consider now (3) when m=2,

$$S = \sum_{i=1}^{n} \left| y_i - \beta_1 x_{i1} - \beta_2 x_{i2} \right| = \sum_{i=1}^{n} S_i \qquad (9)$$

Where $S_i=\left| y_i-\beta_1 x_{i1}-\beta_2 x_{i2} \right|$. We may plot S_i as a function of β_1 and β_2. Every S_i is composed of two half planes in $S x\beta_1 x \beta_2$ space that intersect in the $\beta_1 x \beta_2$ plane. Thus S_i is convex downward which its minimum locates on the intersection of the two half-planes. Since S_i's are all convex, their sum S surface is convex too. Extension to m independent variables is straightforward. In this case each S_i consists of two m dimensional half- hyperplanes in $S x\beta_1 x \ldots x\beta_m$ space intersecting in the $\beta_1 x \ldots x\beta_m$ hyperplane, and as before is convex in the opposite direction of the S axis. S, which is the sum of all these half-hyperplanes forms a polyhedronal hypersurface which is convex too.

3.4 Zero residuals in optimal solution

L_1-norm regression hyperplane always passes through r of thy n data points, where r is rank of the **X** matrix. Usually **X** is of full rank and thus r is equal to m. So, for number of parameters there exist zero residuals for the minimal solution of (3). This implies that L_1-norm regression hyperplane must pass through m observation points (see, Karst (1959), Taylor (1974), Money et al (1978), Appa and Smith (1973), Gentle and Sposito and Kennedy (1977)).

This phenomenon is because of the polyhedronal shape of the S. It is obvious that the minimum solution occurs on at least one of the corners of S, and the corners of S are the loci of changes in slopes of the polygonal hypersurface. Note that these corners and also edges of S will be above the intersections of m subset of the following n hyperplanes.

$$y_i - \sum_{j=1}^{m} \beta_j x_{ij} = 0 \quad i \varepsilon \{1,\ldots,n\} \qquad (10)$$

Since each of these hyperplanes corresponds to a particular m subset of observations, there will be m observations that lie on the regression hyperplane (see, Taylor (1974)).

3.5 Optimality condition

This condition is derived from the Kuhn-Tucker necessary condition of nonlinear programming and proved by Gonin and Monpy (1987b) and Charalambous (1979). Define $A=\{i \left| y_i-\mathbf{x}_i\boldsymbol{\beta}^*=0\right.\}$ and $I=\{i \left| y_i-\mathbf{x}_i\boldsymbol{\beta}^*\neq 0\right.\}$; in linear L_1-norm regression, a necessary and sufficient condition for $\boldsymbol{\beta}^*$ to be a global L_1-norm solution is the existence of multipliers $\alpha_i \varepsilon[-1,1]$ such that:

$$\sum_{i\varepsilon A} \alpha_i \mathbf{x}_i + \sum_{i\varepsilon I} \operatorname{sgn}(y_i-\mathbf{x}_i\boldsymbol{\beta}^*)\mathbf{x}_i = 0 \qquad (11)$$

(see also, El-Attar and Vidyasagar and Dutta (1976). Appa and Smith (1973) showed that this solution is a hyperplane such that:

$$\left| n^+ - n^- \right| \le m \qquad (12)$$

where n^+ and n^- are the number of observations above and below the regression hyperplane respectively.

3.6 Unique and non-unique qolutions

Since S is a convex polyhedronal hypersurface, it always has a minimum. This solution is often unique. Sometimes the shape of S is such that a line or a closed polygon or polyhedron or hyperpolyhedron segment of S is parallel to $\beta_1 x \ldots x \beta_m$ hyperplane. On this case the L_1 norm regression parameters are not unique and infinite points of the mentioned hyperpolyhedron are all solutions (see, Moroney (1961), Sielken and Hartley (1973), Taylor (1974), Farebrother (1985), Sposito (1982), Harter (1977)).

3.7 Interior and sensitivity analysis

Narula and Wellington (1985) Showed that the L_1 norm estimates may not be affected by certain data points. Thus deleting those points does not change the estimated values of the regression parameters. In another discussion, they called sensitivity of L_1 norm estimates, determined the amounts by which the value of response variable y_i can be changed before the parameters estimates are affected. Specifically, if value of y_i increases or decreases without changing the sign of u_i, the solution of the parameters will not change (see, Gauss (1809), Farebrother (1987b)).

For topology of L_1 norm approximation and its properties see Kripke and Rivlin (1965), Vajda (1987), Hromadka II et al (1987). Other properties of the L_1 norm regression are discussed by Gentle and Kennedy and Sposito (1976,77), Assouad (1977), Sposito and Kennedy and Gentle (1980), Bassett (1987,88a,b).

4. Chronology and historical development (1632-1928)

The origin of L_1 norm estimation may be traced back to Galilei (1632). In determining the position of a newly discovered star, he proposed the least possible correction in order to obtain a reliable result (see, Ronchetti (1987) for some direct quotations). Boscovich (1757) for the first time formulated and applied the minimum sum of absolute errors for obtaining the best fitting line given three or more pairs of observations for a simple two variable regression model. He also restricts the line to pass through the means of the observation points. That is,

$$\begin{aligned} \min_{\beta_0,\beta_1} :\ & \sum_{i=1}^{n} \left| y_i - \beta_0 - \beta_1 x_{i1} \right| \\ \text{s.to:}\ & \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1}) = 0 \end{aligned} \qquad (13)$$

Boscovich (1760) gives a simple geometrical solution to his previous suggestion. This paper has been discussed by Eisenhart (1961) and Sheynin (1973). In a manuscript Boscovich poses the problem to Simpson and Simpson gives an analytical solution to the problem (see, Stigler (1984)).

Laplace (1773) provides an algebraic formulation of an algorithm for the L_1 norm regression line which passes through the centroid of observations. In Laplace (1779), extension of L_1 norm regression to observations with different weights has also been discussed. Prony (1804) gives a geometric interpretation of Laplace's (1779) method and compares it with other methods through an example. Svanberg (1805) applies Laplace's method in determining a meridian arc and Von Lindenau (1806) uses this method in determination of the elliptic meridian.

Gauss (1809) suggests the minimization of sum of absolute errors without constraint. He concludes that this criterion necessarily sets m of the residuals equal to zero, where m is number of parameters and further, the solution obtained by this method is not changed if the value of dependent variable is increased or decreased without changing the sign of the residual. This conclusion, is recently discussed by Narula and Wellington (1985) which explained in previous section under the subject of interior and sensitivity analysis. He also noted that Boscovich or Laplace estimators which minimize the sum of absolute residuals with zero sum of residuals constraint, necessarily set m-1 of the residuals equal to zero (see, Stigler (1981), Farebrother (1987b)).

Mathieu (1816) used Laplace's method to compute the eccentricity of the earth. Van Beeck-Calkoen (1816) advocates the using of the least absolute values criterion in fitting curvilinear equation obtained by using powers of the independent variable.

Laplace (1818) adapted Boscovich's criterion again and gave an algebraic procedure (see, Farebrother (1987b)). Let x_1* and y* be the means of x_{i1} and y_i then,

$$ß_0 = y^* - ß_1 x_1^* \qquad (14)$$

Value of $ß_1$ is found by,

$$\min_{ß_1}: S = \sum_{i=1}^{n} |y_i\sim - ß_1 x_{i1}\sim| \qquad (15)$$

where, $x_{i1}\sim$ and $y_i\sim$ are deviations of x_{i1} and y_i from these means respectively. By rearranging the observations in descending order of $y_i\sim/x_{i1}\sim$ values, Laplace notes that S is infinite when $ß_1$ is infinite and decreases as $ß_1$ is reduced. $ß_1$ reaches the critical value $y_t\sim/x_{t1}\sim$ when it again begins to increase. This critical value of $ß_1$ is determined when,

$$\sum_{i=1}^{t-1} |x_{i1}\sim| < ½ \sum_{i=1}^{n} |x_{i1}\sim| \le \sum_{i=1}^{t} |x_{i1}\sim| \qquad (16)$$

This procedure to find a1 is called weighted median, and has been used in many other algorithms such as Rhodes (1930), Singleton (1940), Karst (1958), Bloomfield and Steiger (1980), Bidabad (1987a,b,88a,b) later. Bidabad (1987a,b,88a,b) derives the condition (16) via discrete differentiation method.

Fourier (1824) formulates least absolute residuals regression as what we would now call linear programming; that is minimization of a linear objective function subject to linear inequality constraints.

Edgeworth (1883) presents a philosophical discussion on differences between minimizing mean square errors and mean absolute errors. Edgeworth (1887a,b) proposed a simple method for choosing the regression parameters. By fixing m-1 of the parameters, he used Laplace's procedure to determine the optimal value of the remaining parameter. Repeating this operation for a range of values for m-1 fixed parameters he obtained a set of results for each of m possible choices of the free parameters. Edgeworth drops the restriction of passing through the centroid of data. Turner (1887) discusses the problem of non unique solutions under the least absolute error criterion as a graphical variant of Edgeworth (1887a) as a possible drawback to the method. Edgeworth (1888) replies to Turner's criticism by proposing a second method for choosing the two parameters of least absolute error regression of a simple linear model which makes no use of the median loci of his first method. Edgeworth, in this paper, followed Turner's suggestion for graphical analysis of steps to reach the minimum solution.

Before referring to double median method of Edgeworth (1923), it should be noted that Bowley (1902) completes the Edgeworth's (1902) paper by a variant of double median method which presented after him by Edgeworth (1923). This variant ignores the weights attached to errors.

Edgeworth (1923) discussed the more general problem of estimating the simple linear regression parameters by minimizing the weighted sum of the absolute residuals. He restates the rationale for the method and illustrates its usage through several examples. He also considers the non unique solution problem. His contribution is called double median method.

Estienne (1926-28) proposes replacing the classical theory of errors of data based on least squares with what he calls a rational theory based on the least absolute residual procedure. Bowley (1928) summarizes the Edgeworth's contributions to mathematical statistics which includes his work on L_1-norm regression. Dufton (1928) also gives a graphical method of fitting a regression line.

Farebrother (1987b) summarizes the important contributions to L_1-norm regression for the period of 1793-1930. For more references see also Crocker (1969), Harter (1974a,b,75a,b,c,76), Dielman (1984).

Up to 1928, all algorithms had been proposed for simple linear regression. Though some of them use algebraic propositions, are not so organized to handle multiple L_1-norm regression problem. In the next section we will discuss the more elaborated computational methods for simple and multiple L_1-norm regressions not in a chronological sense; because many digressions have been occurred. We may denote the period of after 1928 the time of modern algorithms in the subject of L_1-norm regression.

5. Computational augorithms

Although, a closed form of the solution of L_1-norm regression has not been derived yet, many algorithms have been proposed to minimize its objective function (see, Cheney (1966), Chambers (1977), Dielman and Pfaffenberger (1982,84)). Generally, we can classify all L_1-norm algorithms in three major categories as, direct descent algorithms, simplex type algorithms and other algorithms which will be discussed in the following sections sequentially.

5.1 Direct descent alborithms

The essence of the algorithms which fall within this category is finding an steep path to descend down the polyhedron of the L_1-norm regression objective function. Although the Laplace's method (explained herein before) is a special type of direct descent algorithms; origin of this procedure in the area of L_1-norm can be traced back to the algorithms of Edgeworth which were explained in the previous section.

Rhodes (1930) found Edgeworth's graphical solution laborious, therefore, he suggested an alternative method for general linear model which may be summarized as follows (see, Farebrother (1987b)). Suppose, we have n equations with m<n unknown parameters. To find L_1-norm solution of this overdetermined system of equations he tries to reduce the m parameter model to a weighted median one parameter problem by solving m-1 of n equations (see also Bidabad (1989a,b)). Rhodes (1930) explained his algorithm by an example and did not give any proof for convergence. Bruen (1938) reviews the L_1-norm regression methods presented by earlier authors. He also compares L_1, L_2 and L_∞ norms regressions.

Singleton (1940) applied Cauchy's steepest descent method (see, Panik (1976)) for the general linear L_1-norm regression. In this paper a geometrical interpretation of gradient on L_1-norm polyhedron and some theorems about existence and uniqueness of solution and convexity property all were given. This paper has not been clearly written, for discussion of the algorithm see Bidabad (1989a,b).

Bejar (1956,57) focuses on consideration of residuals rather than on the vector of parameters. He puts forth a procedure with the essence of Rhodes (1930). However, he is concerned with two and three parameter linear models.

Karst (1958) gives an expository paper for one and two parameter regression models. In his paper, Karst without referring to previous literature actually reaches to the Laplace proposition to solve the one parameter restricted linear model and for the two parameter model, he proposed an algorithm similar to that of Rhodes (1930). His viewpoint is both geometrical and algebraic and no proof of convergence for his iterative method is offered. Sadovski (1974) uses a simple "bubble sort" procedure and implements Karst algorithm in Fortran. Sposito (1976) pointed out that the Sadovski's program may not converge in general. Sposito and Smith (1976) offered another algorithm to remove this problem. Farebrother (1987c) recodes Sadovski's implementation in Pascal language with some improvement such as applying "straight insert sort".

Usow (1967b) presents an algorithm for L_1-norm approximation for discrete data and proves that it converges in a finite number of steps. A similar algorithm on L_1-norm approximation for continuous data is given by Usow (1967a). The Usow's algorithm is to descend on the convex polytope from vertex to vertex along connecting edges of the polytope in such a way that certain intermediate vertices are by-passed. This descent continues until the lowest vertex is reached. (see also, Abdelmalek (1974), Bidabad (1989a,b)).

Relation of this algorithm with simplex method has been discussed by Abdelmalek (1974). He shows that Usow's algorithm is completely equivalent to a dual simplex algorithm applied to a linear programming model with nonnegative bounded variables, and one iteration in the former is equivalent to one or more iterations in the latter. Bloomfield and Steiger (1980) devise an efficient algorithm based on the proposition of Usow explained above.

Sharpe (1971) by applying the L_1-norm regression to portfolio and its rate of return, gives an algorithm for the two parameter linear regression model it must be possible to assign half of the points above and half below the regression line (see also Bidabad (1989a,b)).

Rao and Srinivasan (1972) interpret Sharpe's procedure as the solution of parametric dual linear programming formulation of the problem. They give an alternate and about equally efficient procedure for solving the same problem. Brown (1980) gives a distinct but similar approach to those of Edgeworth (1923) and Sharpe (1971). He emphasizes on the median properties of the estimator. The similarity comes from graphical approach of the three authors. Kawara (1979) also develops a graphical method for the simple regression model.

Bartels and Conn and Sinclair (1978) apply the method of Conn (1976) to the L_1-norm solution of overdetermined linear system. Their approach is minimization technique for piecewise differentiable functions (see also Bidabad (1989a,b)). This algorithm has also been modified for the case of degeneracy (see also, Bartels and Conn and Sinclair (1976)). Bartels and Conn (1977) showed that how L_1-norm, restricted L_1-norm, L_∞ norm regressions and general linear programming can all be easily expressed as a piecewise linear minimization problem. By some simplifications this algorithm corresponds precisely to the algorithm proposed by Bartels and Conn and Sinclair (1978). The contribution of this paper is putting a wide class of problems in the mould of two algorithms mentioned above. The techniques are easily extended to the models with norm restrictions (see also Bidabad (1989a,b)).

Bloomfield and Steiger (1980) proposed a descent method for the L_1-norm multiple regression. Their algorithm is also explained in Bloomfield and Steiger (1983). In some steps this algorithm is related to that of Singleton (1940) and Usow (1967b). The basis of this method is to search for a set of m observations which locate on the optimal L_1-norm regression. This set is found iteratively by successive improvement. In each iteration one point from the current set is identified as a good prospect for deletion. This point is then replaced by the best alternative. The novel features of this method are in an efficient procedure for finding the optimal replacement and a heuristic method for identifying the point to be deleted from the pivot (see also Bidabad (1989a,b)). In this paper relationship of this algorithm to linear programming is also discussed.

Seneta and Steiger (1984) proposed an algorithm for L_1-norm solution of slightly overdetermined system of equations. Their proposition is based on the above algorithm of Bloomfield and Steiger. It is more efficient than the former if m is near n.

Seneta (1983) reviews the iterative use of weighted median to estimate the parameters vector in the classical linear model when the fitting criterion is L_1-norm and also Cauchy criterion.

Wesolowsky (1981) presents an algorithm for multiple L_1-norm regression based on the notion of edge descent along the polyhedron of the objective function (see also Bidabad (1989a,b)). This algorithm is closely related to those of Rhodes (1930) and Bartels and Conn and Sinclair (1978) which explained before. Consider the multiple linear regression as before. In this paper Wesolowsky also discusses the problem of multicolinearity and gives an appropriate solution.

Josvanger and Sposito (1983) modify Wesolowsky's algorithm for the two parameter simple linear regression model. The modification is an alternative way to order observations instead of sorting all of them to find the necessary weighted median value. Suppose the problem has been reduced to a weighted median problem. They place smaller values of factors to be sorted with corresponding weights below the current solution point and larger or equal values above it, then recheck the inequalities (16) of weighted median. If the inequalities do not satisfy then an appropriate adjustment is made. In particular, if the right hand side is overly weighted, then the weight corresponding to the smallest sorting factor is transferred to the left hand side, and the check is made again. A computer program for this algorithm is also given by the authors.

"Generalized gradient" method introduced by Clarke (see, Clarke (1983)) is a general procedure for non-smooth optimization functions and problems (see, Osborne and Pruess and Womersley (1986)). A subclass of this method is called "reduced gradient" explained by Osborne (1985) is a general algorithm which contains linear programming, piecewise linear optimization problems and polyhedral convex function optimization algorithms inside. The reduced gradient algorithm is a special case of descent method which possesses two important characteristics. Identify direction and taking a step in this direction to reduce the function value (see also, Anderson and Osborne (1975), Osborne and Watson (1985) Osborne (1985,87)). The algorithms of Bartels and Conn and Sinclair (1978), Armstrong and Frome and Kung (1979), Bloomfield and Steiger (1980) are all special cases of reduced gradient method.

Imai and Kato and Yamamoto (1987) present a linear time algorithm for computing the two parameter L_1-norm linear regression by applying the pruning technique. Since the optimal solution in the a0xa1 plane lies at the intersection of data lines, so, at each step a set of data lines which does not determine the optimum solution are discarded. In this paper algebraic explanation of the problem is also offered.

Pilibossian (1987) also gives an algorithm similar to Karst (1958) for the simple two parameter linear L_1-norm regression.

Bidabad (1987a,b,88a,b) proposed a descent methods for the simple and multiple L_1-norm regressions. These algorithms with many improvements discussed by Bidabad (1989a,b). Since the algebraic closed form of the L_1-norm estimator has not been derived yet, he tried to give some insight to this problem by applying discrete differentiation technique to differentiate the L_1-norm objective function. This differentiation on discrete domain variables accompanying with regular differentiation on variables with continuous domains increases our knowledge on the algebraic closed form of the problem. In order to improve the accuracy, speed and generally the efficiency of computation of the L_1-norm estimator, he proposed four algorithms which two of them are for simple and others two are for multiple regression models. By inspecting the properties of proposed algorithms, many characteristics of the solution space are clarified. In Bidabad (1989a,b) to find the minimum of the L_1-norm objective function of the regression, m-1 points on the polyhedron of the objective function are selected and from this set the m^{th} point is found by descending in steepest direction. Delete an appropriate point and enter the last m^{th} point for next descending step. The procedure is continued until the global minimum is reached. Although, most of the descent methods use a similar procedure, the steps are well organized and modified for the special shape of the L_1-norm objective function. In this paper the new convergence theorems related to the proposed algorithms are proved and their properties are discussed.

5.2 Simplex type algorithms

The essence of linear programming in solving L_1-norm problem may be found in the work of Edgeworth (1888). Harris (1950) suggested that the L_1-norm estimation problem is connected with linear programming. Charnes and Cooper and Ferguson (1955) formulated the problem as linear programming model. This article is the first known to use linear programming for this case. Adaptation of linear programming to L_1-norm estimation problem is shown below,

$$
\begin{aligned}
&\min_{\boldsymbol{\beta}}: \mathbf{1}_n^T(\mathbf{w}+\mathbf{v}) \\
&\text{s.to: } \mathbf{X}\boldsymbol{\beta}+\mathbf{I}_n(\mathbf{w}-\mathbf{v})=\mathbf{y} \\
&\qquad\quad \mathbf{w},\mathbf{v}\geq\mathbf{0} \\
&\qquad\quad \boldsymbol{\beta} \text{ unrestricted in sign}
\end{aligned}
\tag{17}
$$

Where $\mathbf{1}_n$ is a vector of size nx1 of 1's and $\mathbf{I}_n$ is a n^{th} order identity matrix. The vectors **v** and **w** are of size nx1 and their elements may be interpreted as vertical deviations above and below the fitted regression hyperplane respectively. This problem has n equality constraints in m+2n variables. When n is large, this formulation generally requires a large amount of storage and computation time.

Wagner (1959) shows that the formulation of the L_1-norm regression may be reduced to m equality constraints linear programming problem. Thus, this dual formulation reduces n equations of primal form to m equations of dual form and considerably reduces the storage and computation time.

Fisher (1961) reviews the formulation of the L_1-norm estimation in relation with primal form of linear programming. Barrodale and Young (1966) developed a modified simplex algorithm for determining the best fitting function to a set of discrete data under the L_1-norm criterion. The method is given as Algol codes (for critics see, McCormick and Sposito (1975)). Davies (1967) demonstrates the use of the L_1 norm regression estimates. Rabinowitz (1968) also discusses the application of linear programming in this field. Crocker (1969) cautions against using the L_1 norm criterion merely to restrain unwanted negative coefficient estimates which occur in least squares regression. Multicolinearity is one of the cases which causes this result. Robers and Ben-Israel (1969) by using interval linear programming, proposed an algorithm to solve the L_1 norm estimation problem. Rabinowitz (1970), Shanno and Weil (1970) discuss some connections between linear programming and approximation problem. Barrodale (1970) summarizes the linear and nonlinear L_1 norm curve fitting on both continuous and discrete data. Spyropoulos and Kiountouzis and Young (1973) suggest two algorithms for fitting general functions and particularly fast algorithm with minimum storage requirements for fitting polynomials based on the algebraic properties of linear programming formulation. Robers and Robers (1973) have supplied a special version of the general method of Robers and Ben-Israel (1969) which is designed specifically for the L_1 norm problem. A Fortran code is also provided.

Barrodale and Roberts (1973) present a modification of simplex method which needs smaller amount of storage and by skipping over simplex vertices is more efficient than usual simplex procedure. Define the vector **ß** as a difference of two nonnegative vectors **c** and **d**, their formulation can be stated as follows,

$$\begin{aligned}&\min_{\mathbf{c},\mathbf{d}}: \mathbf{1}_n^T(\mathbf{w}+\mathbf{a})\\ &\text{s.to: } \mathbf{X}(\mathbf{c}-\mathbf{d})+\mathbf{I}_n(\mathbf{w}-\mathbf{v})=\mathbf{y}\\ &\qquad \mathbf{w},\mathbf{v},\mathbf{c},\mathbf{d}\geq \mathbf{0}\end{aligned} \tag{18}$$

Because of the relationships among variables, computation can be performed by using only (n+2)x(m+2) amount of array storage, including labels for the basic and non-basic vectors. An initial basis is given by **w** if all y_i are nonnegative. If a y_i is negative, sign of the corresponding row is changed and the unit column from the corresponding element of **v** is taken as part of the basis. The algorithm is implemented in two stages. First stage restricts the choice of pivotal column during the first m iterations to the vectors elements c_j and d_j recording to the associated maximum nonnegative marginal costs. The vector that leaves the basis causes the maximum decrease in the objective function. Thus the pivot element is not necessarily the same as in the usual simplex. Second stage involves interchanging non basic w_i or v_i with the basic w_i or v_i. The basic vectors corresponding to c_j and d_j are not allowed to leave the basis. The algorithm terminates when all marginal costs are non-positive (see, Kennedy and Gentle (1980)). A Fortran code for this procedure is given by Barrodale and Roberts (1974). Peters and Willms (1983) give algorithms accompanying with computer codes for up-and-down dating the solution of the problem when a column or row inserted to or deleted from **X**, or **y** is changed. These algorithms are all based on Barrodale and Roberts (1973,74) procedure.

Abdelmalek (1974) describes a dual simplex algorithm for the L_1 norm problem with no use of artificial variables. For this algorithm, the Haar condition (see, Osborne (1985), Moroney (1961)) need not be satisfied anymore. This algorithm seemed to be very efficient at the time of publication. An improved dual simplex algorithm for L_1 norm approximation is proposed by Abdelmalek (1975a). In this algorithm, certain intermediate iterations are skipped and in the case of ill-conditioned problems, the basis matrix can lend itself to triangular factorization and thus ensure stable solution. Abdelmalek (1980a) improves his previous algorithm by using triangular decomposition. A Fortran translation of the algorithm is given by Abdelmalek (1980b). Sposito and McCormick and Kennedy (1978) summarizes much of the works on L_1 norm estimation including problem statement, linear programming formulation, efficient computational algorithms and properties of the estimators.

Armstrong and Kung (1978) propose an algorithm for simple two parameter L_1 norm regression. The method is a specification of linear programming of Barrodale and Roberts (1973) algorithm. A Fortran code is given too.

Armstrong and Frome and Kung (1979) use LU (Lower-Upper triangular) decomposition of Bartels and Golub (1969) in maintaining the current basis on revised simplex procedure. A Fortran translation is also enclosed. Armstrong and Godfrey (1979) show that the primal method of Barrodale and Roberts (1973) and dual method of Abdelmalek (1975) are essentially equivalent. With a given initial basis for the two methods, they show that, both algorithms will generate corresponding bases at each iteration. The only difference is the choice of initial basis and heuristic rules for breaking ties. Armstrong and Kung (1982b) presents a dual linear programming formulation for the problem. Various basis entry and initialization procedures are considered. It has been shown that the dual approach is superior to primal one if a good dual feasible solution is readily available (see also, Steiger (1980)). Banks and Taylor (1980) suggest a modification of Barrodale and Roberts (1973) algorithm. The objective function is altered to include magnitudes of the elements of the both errors and solution vectors. For a general discussion on simplex for piecewise linear programming see Fourer (1985a,b) and for a survey of the corresponding problem on the L_1 norm see Fourer (1986). Narula and Wellington (1987) propose an efficient linear programming algorithm to solve the both L_1 and L_∞ norms linear multiple regressions. The algorithm exploits the special structure and similarities between the two problems.

Brennan and Seiford (1987) develop a geometrical interpretation of linear programming in L_1 norm regression. They give a geometric insight into the solving process in the space of observations. McConnell (1987) shows how the method of vanishing Jacobians which has been used to optimize quadratic programming problems can also be used to solve the special linear programming problem associated with computing linear discrete L_1 norm approximation. For the possibility of applying other types of linear programming solutions such as Karmarkar solution to L_1 norm problem see Meketon (1986).

5.3 Other algorithms

This category consists of algorithms which were not classified in the two last sections.

Rice (1964c) applies the bisection method to L_1 norm regression. In this method at each step the domain of S is broken to two segments and the appropriate segment is selected for the next iteration. Solution is reached when the last segment is less than a predetermined small value (see, Bidabad (1989) for discussing bisection method).

Abdelmalek (1971) develops an algorithm for fitting functions to discrete data points and solving overdetermined system of linear equations. The procedure is based on determining L_1 norm solution as the limiting case of L_p norm approximation when p tends to one from right in limit. This technique thus obtains a solution to a linear problem by solving a sequence of nonlinear problems.

Schlossmacher (1973) computed the L_1 norm estimates of regression parameters by an iterative weighted least squares procedure. Instead of minimizing sum of absolute deviations he minimized sum of weighted squared errors with $1/|u_i|$ as weights. Once least squares is applied to the problem and residuals are computed. The absolute value of the inverse of the residuals are again used as corresponding weights in the next iteration for minimizing the sum of weighted squared errors (see also, Holland and Welsh (1977)). Fair (1974) observed that the estimated values of **ß** did not change after the second or third iterations. In cases where any residual is zero, continuation of procedure is impossible, because the corresponding weight to this residual is infinite. This problem is also discussed by Sposito and Kennedy and Gentle (1977), Soliman and Christensen and Rouhi (1988). Absolute convergence of this algorithm has not been proved, but non-convergent experiment has not been reported.

Soliman and Christensen and Rouhi (1988) used left pseudoinverse (see, Dhrymes (1978) for description of this inverse) to solve the general linear L_1 norm regression. According to this procedure one should calculate the least squares solution using the left pseudo-inverse or least squares approximation. Calculate the residual vector. Select the m observations with the smallest absolute values of the residuals and partition the matrices as the selected observations locate on the top and solve **ß** for the top partitions. Although this procedure is operationally simple, its solution is not the same as other exact methods and no proof is presented to show that the solution is in the neighborhood of the exact solution of the L_1 norm minimization problem.

Application of median polish (see, Tukey (1977)) and n-median polish to L_1 norm estimation are discussed and developed by Bloomfield and Steiger (1983), Kemperman (1984), Sposito (1987a), Bradu (1987a,b).

Application of Karmarkar's algorithm for linear programming and its relation to L_1 norm is given by Sherali and Skarpness and Kim (1987). For using homotopy method in L_1 norm see Garcia and Gould (1983), Schellhorn (1987). An algorithm for linear L_1 norm approximation for continuous function is given by Watson (1981), (see also, Baboolal and Watson (1981)).

5.4 Initial value problem

It is discussed by many authors how the algorithms should be started. Selection of initial value is an important factor in the execution time of various algorithms. On the other hand, a good starting point leads to the solution faster and reduces number of iterations. There are several papers which consider the problem for the L_1 norm minimization algorithms. Duris and Sreedharan (1968) briefly refers to this problem. McCormick and Sposito (1976) used the least squares estimator to construct an starting point for the algorithm of Barrodale and Roberts (1973). This initial value reduced the number of iterations in most cases. Sposito and Hand and McCormick (1977) show that the total CPU time needed to obtain optimal regression coefficients under the L_1 norm can generally be reduced if one first computes a near-best L_1 norm estimator such as least squares and then solve the modified procedure of Barrodale and Roberts (1973). A similar discussion about L_∞ norm estimation is given by Hand and Sposito (1980). Sklar and Armstrong (1982) demonstrate that utilizing the least squares residuals to provide an advanced start for the algorithm of Armstrong and Frome and Kung (1978) results in a significant reduction in computational effort.

5.5 Computer programs and packages

Although many authors have coded the computer programs for their own algorithms, which were referenced before, there are also other packages which solve the L_1 norm regression problem and compute the necessary statistics. Some of these packages are IMSL (see, Rice (1985)); BLINWDR (see, Dutter (1987)); ROBETH and ROBSYS (see, Marazzi (1987), Marazzi and Randriamiharisoa (1985)) and XploRe (see, Hardle (1987)). Since these softwares have their own special characteristics we do not go through the details of them. The interested reader may consult the references.

5.6 Comparison of the algorikhms

Generally, the comparison of algorithms is not a straightforward task. As it is indicated by Dutter (1977), factors such as quality of computer codes and computing environment should be considered. In the case of the L_1 norm algorithms, three specific factors of number of observations, number of parameters and the condition of data are more important. Kennedy and Gentle and Sposito (1977a,b), and Hoffman and Shier (1980a,b) describe methods for generating random test data with known L_1 norm solution vectors. Gilsinn et al (1977) discuss a general methodology for comparing the L_1 norm algorithms.

Table 1. Summary of the characteristics of the existing algorithms.

Ref.	**Compared with**	**m range**	**n range**	**Time performances**
BCS	BR	2-8	201	roughly equal speed
AFK	BR	5-20	100-1500	30%-50% AFK is faster
A	BR	1-11	15-203	nearly equal speed
BS	BR	2-6	100-1800	BS is faster for larger n
W	AFK, AK	2-25	100-1800	W is faster for larger n, smaller m
SS	BS	4-34	10-50	SS is faster for m near n
B4	AFK, BS, BR	3-10	20-10000	B4 is faster and more accurate; AFK and BS failed in large samples
AK	S	2	50-500	AK is faster
JS	AK	2	10-250	JS is faster
B2	JS	2	20-10000	B2 is faster

n number of observations.
m number of parameters.

BCS	Bartels,Conn,Sinclair (1978).	BS	Bloomfield,Steiger (1980).
BR	Barrodale,Roberts (1973,74).	W	Wesolowsky (1981).
AK	Armstrong,Kung (1978).	JS	Josvanger,Sposito (1983).
S	Sadovski (1974).	SS	Seneta,Steiger (1984).
AFK	Armstrong,Frome,Kung (1979).	B2	Bidabad (1989a,b), algorithm 2.
A	Abdelmalek (1980a,b).	B4	Bidabad (1989a,b), algorithm 4.

Kennedy and Gentle (1977) examine the rounding error of L_1 norm regression and present two techniques for detecting inaccuracies of the computation (see also, Larson and Sameh (1980)).

Many authors have compared their own algorithms with those already proposed. Table 1 gives a summary of the characteristics of the algorithms proposed by different authors. It is important to note that since the computing environment and condition of data with respect to distribution of the regression errors of the presented algorithms by table 1 are not the same, definitive conclusion and comparison should not be drawn from this table.

Armstrong and Frome (1976a) compare the iterative weighted least squares of Schlossmacher (1973) with Barrodale and Roberts (1973) algorithm. The result was high superiority of the latter. Anderson and Steiger (1980) compare the algorithms of Bloomfield and Steiger (1980), Bartels and Conn and Sinclair (1978) and Barrodale and Roberts (1973). It was concluded that as number of observations n increases the BR locates in a different complexity class than BCS and BS. All algorithms are linear in number of parameters m, and BS is less complex than BCS. Complexities of BS and BCS are linear in n. There is a slight tendency for all algorithms to work proportionately harder for even m than for odd m. BR and BS had the most difficulty with normal error distribution and the least difficulty with Pareto distribution with corresponding Pareto density parameter equal to 1.2.

Gentle and Narula and Sposito (1987) performs a rather complete comparison among some of the L_1 norm algorithms. They limited this comparison to the codes that are openly available for L_1 norm linear regression of unconstrained form. Table 2 shows the required array storage and stopping constants of the corresponding algorithms and the algorithms of Bidabad (1989a,b). Table 2. Array storage requirement for selected algorithms.

Table 2. Array stqrage requirement for selected algrrithms.

Program name	Ref.	Required array storage	Stopping constants
L1	BR	3n+m(n+5)+4	BIG=1.0E+75 TOLER=10**(-D+2/3) D=No.of decimal digits of accuracy
L1	A	6n+m(n+3m/2+15/2)	PREC=1.0E-6 ESP=1.0E-4
L1NORM	AFK	6n+m(n+m+5)	ACU=1.0E-6 BEG=1.0E+15
BLAD1	BS	4n+2m(n+2)	-----------------
BL1	B4	2n+m(3n+m+2)-2	-----------------
LONESL	S	4n	PREC=1.0E-6 BIG=1.0E+19
SIMLP	AK	4n	ACU=1.0E-6 BIG=1.0E+19
DESL1	JS	5n	TOL=1.0E-6
BL1S	B2	5n	-----------------

See table 1 for abbreviations.
Sources: Gentle, Narula, Sposito (1987), Bidabad (1989a,b).

They concluded that BS program performs quite well on smaller problems but in larger cases, because of accumulated round-off error it fails to produce correct answers. Increasing the precision of the coded program to avoid rounding error increases the execution time, so it is not clear what would happen to the relative efficiency of BS after modification.

The Wesolowsky program was not usable and deleted in their study. Because of superiority of AFK to BR and AK to S which had been indicated in previous studies, BR and S algorithm did not enter in their study. Gentle and Sposito and Narula (1988) also compare the algorithms for unconstrained L_1 norm simple linear regression. This investigation is essentially an extraction of Gentle and Narula and Sposito (1987). The attained results are completely similar.

Bidabad (1989a,b) compare the algorithm B2 with JS and B4 with the algorithms of AFK, BS and BR. He concludes that B2 is faster than JS and B4 is faster for smaller m and accurate for larger n. He also observed the failure of AFK and BS for larger problems.

5.7 Nonlinear form computational methods

Suppose again **y**, **X**, **u** and **ß** are defined as before. In nonlinear L_1 norm regression, the problem is to estimate **ß** vector in the nonlinear model,

$$y_i = f_i(\mathbf{x}_i,\boldsymbol{\beta}) + u_i \qquad i=1,...,n;\ n\geq m \tag{19}$$

Where f_i is the response function and $\mathbf{x}_i$ is the i^{th} row of **X**. L_1 norm regression parameters are derived by minimizing the following sum:

$$\min_{\boldsymbol{\beta}}: \sum_{i=1}^{n} \left| y_i - f_i(\mathbf{x}_i,\boldsymbol{\beta}) \right| \tag{20}$$

The function (20) can be reformulated as a nonlinear programming problem as,

$$\begin{aligned} &\min_{\boldsymbol{\beta}}: \sum_{i=1}^{n} w_i \\ &\text{s.to: } y_i - f_i(\mathbf{x}_i,\boldsymbol{\beta}) - w_i \leq 0 \\ &\qquad -y_i + f_i(\mathbf{x}_i,\boldsymbol{\beta}) - w_i \leq 0 \\ &\qquad w_i \geq 0 \\ &\qquad i=1,...,n \end{aligned} \tag{21}$$

Over the last three decades numerous algorithms have been proposed for solving the nonlinear L_1 norm regression problem. These methods can be classified into the following three main categories (see, Gonin and Money (1987b); for another categorization see Watson (1986), McLean and Watson (1980)).

The first category consists of the methods using only first order derivative. In these algorithms the original nonlinear problem is reduced to a sequence of linear L_1 norm problems, which each of them can be solved efficiently by standard linear programming procedures. These methods are of the Gauss-Newton type. The main algorithms which fall into this category have been presented by authors like Osborne and Watson (1971), Anderson and Osborne (1977a,b), Shrager and Hill (1980), McLean and Watson (1980), Jittorntrum and Osborne (1980), Osborne (1980), Watson (1980,84a), Bartels and Conn (1982), Hald and Madsen (1985).

The second category consists of methods which by using second order derivative, transform the original problem into a sequence of unconstrained minimization problems. The non differentiability of the objective function is then overcome. This procedure is known as the penalty function method of nonlinear programming. The contributors are El-attar and Vidyasagar and Dutta (1979), Fletcher (1981,84), Tishler and Zang (1982), Conn (1984), Conn and Gould (1987).

In the last category the objective function is linearized but quadratic approximations are incorporated to take curvature effects into account (see, Murray and Overton (1981), Overton (1982), Bartels and Conn (1982)).

Other characteristics of nonlinear L_1 norm problem are discussed by Rice (1964a,b), Osborne and Watson (1978), Charalambous (1979), Glashoff and Schultz (1979), Hald (1981a,b), Wagner (1982), Watson (1982,87), Powell and Yuan (1984).

5.8 L_p norm computation

Suppose our linear regression model of the form discussed before. The L_p norm estimation of **ß** may be found by minimizing sum of the p^{th} power of the absolute values of the errors. That is,

$$\min_{\beta} \sum_{i=1}^{n} \left| y_i - \sum_{j=1}^{m} \beta_j x_{ij} \right|^p \qquad (22)$$

The above problem can be reformulated as a mathematical programming problem. Rewrite the error vector as difference of two nonnegative vectors **w** and **v** which present positive and negative deviations respectively. That is **u**=**w**-**v**; **w**,**v**≥0. The L_p norm approximation problem reduces as follows (see, Kiountouzis (1972)),

$$\begin{aligned} &\min_{\beta} \sum_{i=1}^{n} (w_i^p + v_i^p) \\ &\text{s.to: } w_i - v_i + \sum_{j=1}^{m} \beta_j x_{ij} = y_i \\ &\qquad w_i, v_i \geq 0 \\ &\qquad \beta_j \text{ unrestricted in sign} \\ &\qquad i=1,\ldots,n;\ j=1,\ldots,m \end{aligned} \qquad (23)$$

It should be noted that this formulation is extremely flexible as it allows that any other constraint to be added (see, Money and Affleck-Graves and Hart (1978)). Another nice specification is that we can change the model to nonlinear form by removing the summation term in the first n constraints and inserting $f_i(\mathbf{x}_i,\boldsymbol{\beta})$ instead. That is,

$$\begin{aligned} &\min_{\beta} \sum_{i=1}^{n} (w_i^p + v_i^p) \\ &\text{s.to: } w_i - v_i + f_i(\mathbf{x}_i,\boldsymbol{\beta}) = y_i \\ &\qquad w_i, v_i \geq 0 \\ &\qquad \beta_j \text{ unrestricted in sign} \\ &\qquad i=1,\ldots,n;\ j=1,\ldots,m \end{aligned} \qquad (24)$$

The resultant is the formulation of nonlinear L_p norm estimation problem.

For general L_p norm regression, there exist various computational methods for linear as well as nonlinear models (for details of the discussion, interested readers may see, Descloux (1963), Rice (1964,69), Barrodale and Young (1966), Sreedharan (1969,71), Ekblom and Henriksson (1969), Karlovitz (1970a,b), Barrodale and Roberts (1970), Barrodale and Roberts and Hunt (1970), Fletcher and Grant and Hebden (1971,74b), Kiountouzis (1972), Forsythe (1972), Kahng (1972), Ekblom (1973a,b), Anton and Duris (1973), Watson (1973,77,78,84b,85a), Shisha (1974), Merle and Spath (1974), Oettli (1975), Rey (1975), Mond and Schechter (1976), Borowsky (1976), Shier and Witzgall (1978), Kennedy and Gentle (1978), Wolfe (1979), Porter and Winstanley (1979), Barr and Affleck-Graves and Money and Hart (1980a), Harter (1981), Madsen (1985), Gonin and du Toit (1987), Fichet (1987b)).

In the case of L_∞ norm solution of overdetermined system of equations, there are similar methods as well (for more information, interested readers may see the following selected articles and also their references, Kelley (1958), Goldstein and Cheney (1958), Cheney and Goldstein (1958), Stiefel (1960), Veidinger (1960), Valentine and Van Dine (1963), Aoki (1965), Osborne and Watson (1967), Bartels and Golub (1968a,b), Gustafson and Kortanek and Rom (1970), Barrodale and Powell and Roberts (1972), Cline (1972,76), Duris and Temple (1973), Watson (1973), Barrodale and Phillips (1974,75), Boggs (1974), Fletcher and Grant and Hebden (1974a), Madsen (1975), Abdelmalek (1975b,76,77a,b), Conn (1975), Coleman (1978), Charalambous and Conn (1978), Bartels and Conn and Charalambous (1978), Armstrong and Kung (1979), Klingman and Mote (1982), Bartels and Conn and Li (1987), Brannigan and Gustafson (1987)).

6. Simultaneous equations system

The L_1-norm estimation has been extensively studied for single equation regression model and its properties are well recognized. But despite of the wide variety of econometric applications of L_1-norm estimation to simultaneous equation systems, there have been only few investigators in this area which their works are summarized in this section. Suppose the following equation as a first equation of a structural system,

$$\mathbf{y} = \mathbf{Y}\theta + \mathbf{X}_1\mathbf{ß} + \mathbf{u} = [\mathbf{Y}|\mathbf{X}_1]\begin{bmatrix}\theta \\ --- \\ \mathbf{ß}\end{bmatrix} + \mathbf{u} \equiv \mathbf{Z}\alpha + \mathbf{u} \qquad (25)$$

Where **y** is a vector of dependent endogenous, **Y**, matrix of independent endogenous, $\mathbf{X}_1$, matrix of exogenous variables; θ and **ß** are vectors of regression parameters and **u** is random error vector. The reduced form for **Y** is given by,

$$\mathbf{Y} = \mathbf{X}\Pi + \mathbf{v} \qquad (26)$$

Direct and indirect least absolute deviations (DLAD, IDLAD) analogues of direct and indirect least squares (DLS, IDLS) may be applied to the systems (25) and (26) respectively. The L_1-norm objective function analogue of two stage least squares (2SLS) for estimation of α may be defined as,

$$\min_{\alpha}: \sum_{i=1}^{n} \left| y_i - \mathbf{P}_i^T\mathbf{Z}\alpha \right| \qquad (27)$$

Where y_i is the i^{th} element of **y**, $\mathbf{P}_i^T$ is the i^{th} row of $\mathbf{P}=(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ (see, Fair (1974)). Amemiya (1982) by comparing the problem (27) with Theil's interpretation of 2SLS,

$$\min_{\alpha}: \sum_{i=1}^{n} (y_i - \mathbf{P}_i^T\mathbf{Z}\alpha)^2 \qquad (28)$$

and interpretation of 2SLS as the instrumental variables estimator, namely, the minimization of,

$$\min_{\alpha}: \sum_{i=1}^{n} (\mathbf{P}_i^T\mathbf{y} - \mathbf{P}_i^T\mathbf{Z}\alpha)^2 \qquad (29)$$

defines two stage least absolute deviations (2SLAD) as,

$$\mathrm{miy}_{\alpha}: \sum_{i=1}^{n} \left| \mathbf{P}_i^T\mathbf{y} - \mathbf{P}_i^T\mathbf{Z}\alpha \right| \qquad (30)$$

Amemiya (1982) combines the two ideas and proposes 2SLAD as a class of estimators obtained by minimizing,

$$\min_{\alpha}: \sum_{i=1}^{n} \left| qf_i + (1-q)\mathbf{P}_i^T\mathbf{y} - \mathbf{P}_i^T\mathbf{Z}\alpha \right| \qquad (31)$$

Where, q is a parameter to be determined by researcher. When q=0, problem (31) is equivalent to (30) and yields the estimator which is asymptotically equivalent to 2SLS. When q=1 then (31) is equivalent to (27). For any value of $q\varepsilon[0,\infty)$ Amemiya (1982) proves the strong consistency of 2SLAD and gives its asymptotic variance under three different cases of normal, partially normal and non normal distribution of **u** and **v**. Powell (1983) demonstrates the asymptotic normality of Amemiya (1982) proposed estimators for more general distributions of error terms.

Amemiya (1982) also proposes another alternative LAD analogue of 2SLS. Once IDLAD is applied to each equation of reduced form and Π^ is computed. Then by minimizing the following expression,

$$\min_{\boldsymbol{\beta}} : \sum_{i=1}^{n} \left| y_i - \mathbf{X}_i^T \hat{\Pi} \theta - \mathbf{X}_{1i}^T \boldsymbol{\beta} \right| \tag{32}$$

$\boldsymbol{\theta}$^ and $\boldsymbol{\beta}$^ are derived. He calls this estimator double two stage least absolute deviations (D2SLAD). A similar discussion for different values of q has also been done. Powell (1983) shows an asymptotic equivalence proposition for the sub-class of D2SLAD estimators. This result is analogous to the finite sample equivalence of Theil's interpretation of 2SLS, and its instrumental variable interpretation.

Glahe and Hunt (1970) as pioneers of introducing L_1-norm in simultaneous system of equations, compare small sample properties of least absolutes and least squares estimators for an overidentified simultaneous system of two equations via Monte Carlo experiments. Estimators where used are DLAD, DLS, IDLAD, IDLS, 2SLAD and 2SLS. All comparisons were done for all three pairs of direct, indirect and two stage least absolute and least squares estimators for different sample sizes of ten and twenty with considering various cases of multicolinearity, heteroskedasticity and misspecification. They concluded that the L_1-norm estimators should prove equal or superior to the L_2 norm estimators for models using a structure similar to that of their study, with very small sample sizes and randomly distributed errors.

The same structure is used by Hunt and Dowling and Glahe (1974) with Laplace and normal error distributions. The estimators in their study are DLAD, DLS, 2SLAD and 2SLS. They concluded that the L_1-norm estimators provided 100% of the best results in the case of Laplace distribution, and 37.5% of the best results in the case of normal distribution of errors.

Nyquist and Westlund (1977) performs a similar study with an overidentified three equations simultaneous system with error terms obeying symmetric stable distributions. The estimators used in this study were similar to those of Glahe and Hunt (1970) mentioned above. They concluded that with normal distribution, L_2 norm estimators are favorable. In non normal case L_1-norm estimators tend to perform better as the degree of non normality increases. When sample size increases, the relative performance of 2SLAD to DLS is increased too. In the normal distribution case 2SLS is the best, and for non normal distributions 2SLAD is the leading alternative closely followed by IDLAD and for extremely non normal cases IDLAD seems to be more robust than 2SLAD.

7. Statistical aspects

Since, the L_1-norm criterion has discovered many interesting extension in statistics, this section has a glance at some of its features on the various fields of statistics.

7.1 Sampling distribution

Ashar and Wallace (1963), Rice and White (1964), Meyer and Glauber (1964), Glahe and Hunt (1970), Fama and Roll (1971), Smith and Hall (1972), Kiountouzis (1973), Brecht (1976), Ramsay (1977), Hill and Holland (1977), Rosenberg and Carlson (1977), Pfaffenberger and Dinkel (1978) have examined small sample properties of L_1-norm fitting via Monte Carlo method in different conditions. The relative efficiency of this estimator to least squares is occurred if errors distribution has big tails.

Wilson (1978) concludes that L_1-norm estimator is 80% as efficient as least squares when errors follow contaminated normal distribution. When outliers are present, L_1-norm estimator becomes more efficient. His approach is Monte Carlo too and a wide variety of experiments are examined.

Cogger (1979) performed ex-post comparisons between L_1-and L_2 norms forecasts from Box-Jenkins autoregressive time series models. The comparisons indicated that L_1-norm approaches to the estimation of ARIMA (integrated autoregression moving average) models of time series data should receive further attention in practice.

For multivariate regression with a symmetric disturbance term distribution, Rosenberg and Carlson (1973) showed that the error in the L_1-norm estimation is approximately, normally distributed with mean zero and variance covariance matrix $\delta^2(\mathbf{X}^T\mathbf{X})^{-1}$, where, δ^2/n is the variance of the median of errors (see also, Sposito and Tvejte (1984), Ronner (1984)). They concluded that, the L_1-norm estimates have smaller variance than least squares in regression with high kurtosis error distribution (see also, Bloomfield and Steiger (1983)).

Sielken and Hartley (1973), Farebrother (1985) have shown that when the errors follow a symmetric distribution, and the L_1 norm estimates may not be unique, the problem may be formulated in such a way as to yield unbiased estimators. A similar discussion for general L_p norm may be found in Sposito (1982).

Bassett and Koenker (1978) showed that the L_1 norm estimates of regression parameters in general linear model are consistent and asymptotically Gaussian with covariance matrix $\delta^2(\mathbf{X}^T\mathbf{X})^{-1}$, where δ^2/n is the asymptotic variance of the sample median from random samples of size n taken from the error distribution (see, Bassett and Koenker (1982), Koenker and Bassett (1984), Bloomfield and Steiger (1983), Oberhofer (1982), Wu (1988). A simple approximation method for computing the bias and skewness of the L_1 norm estimates is given by Withers (1987) which shows that bias and skewness of $\hat{\beta}$ are proportional to the 3rd moments of independent variables. The moment problem in the L_1 norm is discussed by Hobby and Rice (1965).

Dupacova (1987a,b) used the tools of nondifferentiable calculus and epi-convergence to find the asymptotic properties of restricted L_1 norm estimates. Asymptotic interesting properties of Boscovich's estimator which is L_1 norm minimization of errors subject to zero mean of residuals constraint may be found in Koenker and Bassett (1985). L_1 norm fit for censored regression (or censored "Tobit") models has been introduced by Powell (1984,86). Paarsch (1984) by Monte Carlo experiments showed that the Powell estimator is neither accurate nor stable.

Gross and Steiger (1979) used an L_1 norm analogue of L_2 norm estimator for the parameters of stationary, finite order autoregressions. This estimator has been shown to be strongly consistent. Their evidences are based on Monte Carlo experiments (see also, Bloomfield and Steiger (1983) for more discussions).

7.2 Statistical inference

The asymptotic distribution of the three L_1 norm statistics (Wald, likelihood ratio and Lagrange multiplier tests) of linear hypothesis for general linear model have been discussed in Koenker and Bassett (1982a). They derived the asymptotic distribution for a large class of distributions. It has been shown that these tests under mild regularity conditions on design and error distribution have the same limiting chi- square behavior. Comparison of these tests based on Monte Carlo experiments is given in Koenker (1987). Since the L_1 norm estimator asymptotically follows a normal distribution, Stangenhaus and Narula (1987) by using Monte Carlo method determined the sample size at which normal distribution approximation can be used to construct the confidence intervals and test of hypothesis on the parameters of the L_1 norm regression. Comparison methods for studentizing the sample median which can be extended to L_1 norm regression is discussed by McKean and Sheather (1984); and accordingly, testing and confidence intervals are compared by Sheather and McKean (1987).

Two coefficients of determination for L_1 norm regression are given by McKean and Sievers (1987). A class of tests for heteroskedasticity based on the regression quantiles is given in Koenker and Bassett (1982b). More recent works on L_1 norm statistical inference and analysis of variance may be found in Armstrong et al (1977), Siegel (1983) Sheather (1986), McKean and Shrader (1987), Shrader and McKean (1987), Stangenhaus (1987), Brown and Hettmansperger (1987), Tracy and Khan (1987), Vajda (1987), Sheather (1987), Fedorov (1987). For other characterization see, Fichet (1987a), LeCalve (1987).

7.3 Multivariate statistics

In usual clustering method, Euclidian metric or distance as an appropriate real valued function for constructing dissimilarity criterion is used (see also, Bidabad (1983a)). Spath (1976) used L_1 metric as a criterion for clustering problem. More modification and extension may be found in Spath (1987). Kaufman and Rousseeuw (1987) introduced an L_1 norm type alternative approach, used in k-medoid method, that minimizes the average dissimilarity of the all objects of the data set to the nearest medoid. Trauwaert (1987) and Jajuga (1987) applied the L_1 metric in fuzzy clustering method of ISODATA (Iterative Self Organizing Data Analysis Technique (A)). Trauwaert (1987) showed that in the presence of outliers or data errors, L_1 metric has superiority over L_2 distance.

An L_1 norm similar version of multidimensional scaling is presented by Heiser (1988) (see also, Critchley (1980)) and of correspondence analysis by Heiser (1987). Robust L_p norm discrimination analysis is discussed by Haussler (1984) and Watson (1985a). L_1 norm estimation of principal components considered by GaL$_p$in and Hawkins (1987).

7.4 Nonparametric density estimation

L_1 norm has also been used in nonparametric statistics and density estimation. The procedure of density estimation is done via the Parzen kernel function. Abou-Jaoude (1976a,b,c), Devroye and Wagner (1979,80) give the conditions for the L_1 norm convergence of kernel density estimates. Devroye (1983,85) gives the complete characterization of the L_1 norm consistency of Parzen-Rosenblatt density estimate. Devroye concludes that all types of the L_1 norm consistencies are equivalent. Gyorfi (1987) proves the L_1 norm consistency of kernel and histogram density estimates for uniformly and strong mixing samples. Devroye and Gyorfi (1985) give a complete explanation of the L_1 norm nonparametric density estimation. The central limit theorems of L_p norms for kernel estimators of density and their asymptotic normality in different conditions of unweighted and weighted L_p norm of naive estimators, and under random censorship are discussed in Csorgo and Horvath (1987,88), Horvath (1987), Csorgo and Gombay and Horvath (1987). Bandwidth selection in nonparametric regression estimation is shown by Marron (1987). Via an example he concludes that it is an smoothing problem. Welsh (1987) considers simple L_1 norm kernel estimator of the sparsity function and investigates its asymptotic properties. L_1 and L_2 norms cross-validation criteria are studied for a wide class of kernel estimators by Rossi and Brunk (1987,88). Gyorfi and Van der Meulen (1987) investigate the density- free convergence properties of various estimators of Shannon entropy and prove their L_1 norm consistency. Munoz Perez and Fernandez Palacin (1987) consider the estimating of the quantile function by using Bernstein polynomials and examine its large sample behavior in the L_1 norm. For comparison of the L_1 and L_2 norms estimators of Weibull parameters see Lawrence and Shier (1981) and for a nonparametric approach on quantile regression see Lejeune and Sarda (1988).

7.5 Robust statistics

One of the most important properties of the L_1 norm methods is resistivity to outliers or wild points. This property makes it one of the most important techniques of robust statistics. Huber (1987) pointed out that the L_1 norm method serves in two main areas of robust estimation. Sample median plays an important role in robust statistics. The sample median is the simplest example of an estimate derived by minimizing the L_1 norm of deviations. Thus, L_1 norm minimizes the maximum asymptotic bias that can be caused by asymmetric contamination. Therefore, it is the robust estimate of choice in cases where it is more important to control bias than variance of the estimate. Next, the L_1 norm method is the simplest existing high-breakdown estimator. Thus it can be a good starting point for iterative estimators which give nonsense solution if they started with a bad initial point and since it is resistant to outliers, may be used as an starting point for trimming the wild points (see also, Taylor (1974), Holland and Welsch (1977), Harvey (1977,78), Armstrong and Frome and Sklar (1980), Antoch et al (1986), Antoch (1987), Portnoy (1987), Bassett (1988b)). This technique for polynomial regression with a test about the degree of polynomial and for regression quantiles is considered in Jureckova (1983,84), Jureckova and Sen (1984). The same thing for nonlinear regression is devised by Prochazka (1988). Ronchetti (1987) reviews the basic concepts of robust statistics based on influence function and also in relation with L_1 norm (see also GaL$_p$in (1986)). For computational algorithms in bounded influence regression see Marazzi (1988). Ekblom (1974) discusses the statistical goodness of different methods when applied to regression problem via Monte Carlo experiments and in Ekblom (1987) he shows the relationship of L_1 norm estimate as limiting case of an L_p norm or Huber estimates. Haussler (1984) and Watson (1985a) considered the robust L_p norm discrimination analysis problem. Robust estimates of principal components (see, Bidabad (1983c)) based on the L_1 norm formulation are discussed by GaL$_p$in and Hawkins (1987). The asymptotic distributional risk properties of pre-test and shrinkage L_1 norm estimators are considered by Saleh and Sen (1987). L_1 norm estimator is also a member of M and R estimators (see, Bloomfield and Steiger (1983) for more discussions).

8. Application

L_1 norm methods has been extensively developed in various fields of sciences and work as strong analytical tools in analyzing human and natural phenomena. Many branches of sciences in applied mathematics, statistics and data analysis like econometrics, biometrics, psychometrics, sociometrics, technometrics, operation research, management, physic, chemistry, astronomy, medicine, industry, engineering, geography and so forth are heavily dependent to this method.

The assumption of normally distributed errors does not always hold for economic variables as well as other data and variables and so we are not confronted with finite variance anywhere. An infinite variance means thick tail errors distribution with a lot of outliers. Since least squares gives a lot of weights to outliers, it becomes extremely sample dependent. Thus in this case least squares becomes a poor estimator. Of course, the observed distributions of economic or social variables will never display infinite variances. However, as discussed by Mandelbrot (1961,63) and herein before, the important issue is not that the second moment of the distribution is actually infinite, but the interdecile range in relation to the interquartile range is sufficiently large that one is justified in acting as though the variance is infinite. Thus, in this context, an estimator which gives relatively little weight to outliers, such as L_1 norm estimator is clearly preferred.

Distribution of personal income has been known to have this characteristic since the time of Pareto -1896. Ganger and Orr (1972) give some evidences on time series characteristics of economic variables which have this property. Many other economic variables such as security returns, speculative prices, stock and commodity prices, employment, asset sizes of business firms, demand equations, interest rate, treasury cash flows, insurance and price expectations all fall in the category of infinite variance error distribution (see, Goldfeld and Quandt (1981), Nyquist and Westlund (1977), Fama (1965), Sharpe (1971)).

Arrow and Hoffenberg (1959) used L_1 norm in the context of interindustry demand. Meyer and Glauber (1964) compare L_1 and L_2 norms directly. They estimated their investment models on a sample by both estimators and then examined them by forecasting ex-post sample. They concluded that, with very few exceptions, the L_1 norm estimation outperformed the L_2 norm estimators, even with criteria such as sum of the squared forecast errors which least squares is ordinarily thought to be minimal. Sharpe (1971) compares L_1 and L_2 norms estimators for securities and portfolios. A similar discussion has been given by Cornell and Dietrich (1978) on capital budgeting. Affleck-Graves and Money and Carter ( ) did the same research by applying L_p norm and with emphasis on factors affecting the estimation of coefficients of an individual security model. Kaergard (1987) compares L_1, L_2 and L_∞ norms estimators for Danish investments via their power to predict the even years from estimation over odd years for a long period. Hattenschwiler (1988) uses goal programming technique in relation with L_1 norm smoothing functions on several large disaggregate linear programming models for Switzerland food security policy (see, Bidabad (1984a) for description of goal programming relevance). Other applications of the L_1 norm smoothing functions on the models for planning alimentary self-sufficiency, food rationing and flux- and balancing model for feeding stuffs are referenced by Hattenschwiler (1988).

Wilson (1979) used L_1 norm regression for statistical cost estimation in a transport context. Chisman (1966) used L_1 norm estimator to determine standard times for jobs in which work-elements are essentially the same for all jobs except that the quality of each type of the work-element used may vary among jobs. Frome and Armstrong (1977) refer to this estimator for estimating the trend-cycle component of an economic time series.

Charnes and Cooper and Ferguson (1955) give optimal estimation of executive compensation of employees by solving L_1 norm problem via the technique of linear programming. Application of the L_1 norm in location theory is of special interest; because by this metric the rectangular distance of two points in two dimensional Cartesian coordinates can be considered very well (see, Cabot et al (1970), Wesolowsky and Love (1971,72), Drezner and Wesolowsky (1978), Ratliff and Picard (1978), Morris and Verdini (1979), Megiddo and Tamir (1983), Calamai and Conn (1987); see also, the bibliography of Domschke and Drext (1984)). Farebrother (1987a) applies L_1 norm to committee decision theory. Mitchell (1987) uses L_1 norm to find a shortest path for a robot to move among obstacles. L_1 norm has been applied to chemistry by Fausett and Weber (1978); in geophysic by Dougherty and Smith

(1966), Claerbout and Muir (1973), Taylor and Banks and McCoy (1979); in astronomy by Rousseeuw (1987), in physical process and pharmacokinetic by Frome and Yakatan (1980), Gonin and Money (1987a). For mechanical representation of L_1 norm see, Farebrother (1987d). Application of the L_1 norm in power systems for static state estimation is given by Kotiuga and Vidyasagar (1982). Anderson (1965) suggests using L_1 norm estimation in order to assure non negative coefficient in linear time equations. For application on the data of orbital measurement see Mudrov et al (1968).

9. Other variants

Narula and Wellington (1977a) propose the minimization of sum of weighted absolute errors. That is, minimizing the expression $\Sigma w_i |u_i|$. An algorithm for this problem is introduced. Narula and Wellington (1977b) proposd a special case of the above formulation by the name, "minimum sum of relative errors". In this problem w_i are set equal to $1/|y_i|$ (see also comment of Steiger and Bloomfield (1980)).

Narula and Wellington (1977c) give an algorithm for L_1 norm regression when the model is restricted to pass through the means of each of the variables (see, Farebrother (1987c) for a remark). In the case of restricted L_1 norm estimation some algorithms presented by Young (1971), Armstrong and Hultz (1977), Barrodale and Roberts (1977,78), Bartels and Conn (1980a,b), Armstrong and Kung (1980). An algorithm for L_1 norm regression with dummy variables is given by Armstrong and Frome (1977). Womersley (1986) introduces a reduced gradient algorithm for censored linear L_1 norm regression.

In the context of stepwise regression and variable selection, there are also special algorithms for the case of L_1 norm (see, Roodman (1974), Gentle and Hansen (1977), Narula and Wellington (1979,83), Wellington and Narula (1981), Dinkel and Pfaffenberger (1981), Armstrong and Kung (1982a)).

An algorithm for regression quantiles is given by Narula and Wellington (1984). Computation of best one-sided L_1 norm regression, that is finding an approximation function which is everywhere below or above the function is given by Lewis (1970). For numerical techniques to find estimates which minimize the upper bound of absolute deviations see Gaivoronski (1987).

Arthanari and Dodge (1981) proposed a convex combination of L_1 and L_2 norms objective functions to find new estimator for linear regression model. Dodge (1984) extends this procedure to a convex combination of Huber M-estimator and L_1 norm estimator objective functions. Dodge and Jureckova (1987) showed that the pertaining convex combination of L_1 and L_2 norms estimates can be adapted in such a way that it minimizes a consistent estimator of the asymptotic variance of the new produced estimator. In Dodge and Jureckova (1988) it is discussed that the adaptive combination of M-estimator and L_1 norm estimator could be selected in an optimal way to achieve the minimum possible asymptotic variance.

Instead of minimizing the absolute deviations, Nyquist (1988) minimized absolute orthogonal deviations from the regression line. In this paper, computational aspects of this estimator is considered and a connection to the projection pursuit approach to estimation of multivariate dispersion is pointed out. Spath and Watson (1987) also introduce orthogonal linear L_1 norm approximation method. Application of orthogonal distance criterion for L_2 and general L_p norms may be found in Spath (1982,86b), Watson (1982b), Wulff (1983).

Rousseeuw (1984) proposes a new method of estimation called by "least median of squares" regression. This estimator is derived by minimizing the expression, $\text{med}(u_i^2)$ fou β. The resulting estimator can resist the effect of nearly 50% of contamination in the data. For an applied book on this topic see Rousseeuw and Leroy (1987). Computational algorithms of this estimator may be found in Souvaine and Steele (1987), Steele and Steiger (1986).

When the number of observations in comparison with the number of unknowns is large, it ought to be better to split the observations into some unknown clusters and look for corresponding regression vectors such that the average sum of the L_p norm of the residual vector attains a minimum. This combination of clustering and regression is called clusterwise regression. A case study and numerical comparison for clusterwise linear L_1 and L_2 norms regressions are given by Spath (1986a).

For clusterwise linear L_1 norm regression algorithms see Spath (1986c), Meier (1987), and for presentation of clusterwise regression see Spath (1985,87).

Application of the L_1 norm to one and two-way tables is given by Armstrong and Frome (1976b,79), Buckley and Kvanli (1981) (see also, Bloomfield and Steiger (1983) for general discussions).

There are other applications of L_1 norm in U-statistics by Chun (1987), Baysian approach by Militky and Cap (1987), isotonic regression by Menendez and Salvador (1987), sample allocation by Melaku and Sadasivan (1987) and method of averages by Kveton (1987).

10. Bibliography

- N.N. Abdelmalek (1971) Linear approximation for a discrete point set and L_1 solutions of overdetermined linear equations. J. ACM, 18, 41-47.
- N.N. Abdelmalek (1974) On the discrete linear L_1 approximation and L_1 solutions of overdetermined linear equations. J. of Approx. Theory, 11, 38-53.
- N.N. Abdelmalek (1975a) An efficient method for the discrete L_1 approximation problem. Math. Comput., 29, 844-850.
- N.N. Abdelmalek (1975b) Chebyshev solution of overdetermined system of linear equations. BIT, 15, 117-129.
- N.N. Abdelmalek (1976) A computer program for the Chebyshev solution of overdetermined system of linear equations, Inter.J. Numer. Meth. in Engin. 10, 1197-1202.
- N.N. Abdelmalek (1977a) Computing the strict Chebyshev solution of overdetermined linear equations, Math. Comp., 31, 974-983.
- N.N. Abdelmalek (1977b) The discrete linear restricted Chebyshev approximation. BIT, 17, 249-261.
- N.N. Abdelmalek (1980a) L_1 solution of overdetermined systems of linear equations. ACM Trans. Math. Soft., 6, 220-227.
- N.N. Abdelmalek (1980b) A Fortran subroutine for the L_1 solution of overdetermined systems of linear equations. ACM Trans. Math. Soft., 6, 228-30.
- S. Abou-Jaoude (1976a) Sur une condition necessaire et suffisante de L_1-convergence pres que complete de l'estimateur de la partition fixe pour une densite. C. R. Acad. Sci. Paris, Ser. A 283, 1107-1110.
- S. Abou-Jaoude (1976b) Sur la convergence L_1 at Ll de l'estimateur de la partition aleatoire pour une densite. Ann. Inst. Henri Poincare, 12, 299-317.
- S. Abou-Jaoude (1976c) Conditions necessaires et suffisantes de convergence L_1 en probabilite de l'histogramme pour une densite. Ann. Inst. Henri Poincare, 12, 213-231.
- J.F. Affleck-Graves, A.H. Money, K. Carter ( ) An evaluation of an alternative methods of estimating the beta coefficient in the market model. Univ. of CapeTown, South Africa.
- T. Amemiya (1982) Two stage least absolute deviations estimators. Econometrica, 50,3, 689-711.
- T.W. Anderson (1962) Least squares and best unbiased estimates. Ann. of Math. Stat., 33, 266-272.
- D.W. Anderson (1965) Linear programming time estimating equations. J. of Indus. Engin. 16, 136-138
- D.H. Anderson (1975) Linear programming and the calculation of maximum norm approximations. Ph.D. thesis, Australian National Univ..
- D.H. Anderson, M.R. Osborne (1976) Discrete linear approximation problems in polyhedral norms. Numer. Math. 26, 179-189.
- D.H. Anderson, M.R. Osborne (1977a) Discrete nonlinear approximation problems in polyhedral norms, a Levenberg-like algorithm. Numer. Math. 28, 157-170.
- D.H. Anderson, M.R. Osborne (1977b) Discrete nonlinear approximation in polyhedral norms. Numer. Math., 28, 143-156.
- D. Anderson, W.L. Steiger (1982) A comparison of methods for discrete L_1 curve-fitting. Tech. Rep. DCS-TR-96. Dept. of Comp. Sci., Hill center for the Math. Sci. Busch campus, New Brunswick, N.J.
- F.J. Anscombe (1976) Topics in the investigation of linear relations fitted by the method of least square (with discussion), J. of Royal Stat. Soc. B series, 1-52.
- J. Antoch, A. Bartkowiak, J. Pekalska (1986) Computing L_1 norm, a-trimmed and a-winsorized regression on the ODRA 1305 computer. Rep. N-159, Institute of Computer Science, Wroclaw Univ., Poland.
- J. Antoch (1987) Variable selection in linear model based on trimmed least squares estimator. In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and related methods. NHPC. 231-246.
- H. Anton, C.S. Duris (1973) On an algorithm for best approximate solutions to Av=b in normed linear spaces. J. Approx. Theory 8, 133-141.
- M. Aoki (1965) Successive generation of Chebyshev approximate solution. J. Basic Engin., 87, 17-22.
- G. Appa, C. Smith (1973) On L_1 and Chebyshev estimation.Math. Prog. 5, 73-87.
- R.D. Armstrong, J.J. Elam, J.W. Hultz (1977) Obtaining least absolute value estimates for a two-way classification model.Comm. Stat., B6, 365-81.
- R.D. Armstrong, E.L. Frome (1976a) A comparison of two algorithms for absolute deviation curve fitting. JASA, 71, 328-330.
- R.D. Armstrong, E.L. Frome (1976b) The calculation of least absolute value estimates for two-way tables. Proc. of the statistical computing section, ASA, Washington D.C., 101-106.
- R.D. Armstrong, E.L. Frome (1977) A special purpose linear programming algorithm for obtaining least absolute value estimates in a linear model with dummy variables. Comm. Stat., B6, 383-98.
- R.D. Armstrong, E.L. Frome (1979) Least-absolute-value estimators for one-way and two-way tables. Naval Res. Log. Quart., 26, 79-96.
- R.D. Armstrong, E.L. Frome, D.S. Kung (1979) A revised simplex algorithm for the absolute deviation curve fitting problem. Comm. Stat B8, 175-190.
- R.D. Armstrong, E.L. Frome, M.G. Sklar (1980) Linear programming in exploratory data analysis. J. of Educ. Stat., 5, 293-307.
- R.D. Armstrong, J. Godfrey (1979) Two linear programming algorithms for the discrete L_1 problem. Math. Comput., 33, 289-300.
- R.D. Armstrong, J.W. Hultz (1977) An algorithm for a restricted discrete approximation problem in L_1 norm. SIAM J. Numer. Anal., 14, 555-565.
- R.D. Armstrong, D.S. Kung (1978) AS132: Least absolute value estimates for a simple linear regression problem. Appl. Stat., 27, 363-366.

- R.D. Armstrong, D.S. Kung (1979) AS135: Mini-max estimates for a linear multiple regression problem. Appl. Stat., 93-100.
- R.D. Armstrong, D.S. Kung (1980) An algorithm for a least absolute value regression problem with bounds on the parameters. Appl. Math. Comput. 7, 267-279.
- R.D. Armstrong, D.S. Kung (1982a) An algorithm to select the best subset for a least absolute value regression problem. TIMS studies in the management sciences, 19, 67-80.
- R.D. Armstrong, D.S. Kung (1982b) A dual algorithm to solve linear least absolute value problems. J. Oper. Res. Soc., 33, 931-936.
- K.J. Arrow, M. Hoffenberg (1959) A time series analysis of interindustry demands. NHPC, Amsterdam. T.S. Arthanari, Y. Dodge (1981) Mathematical programming in statistics. John Wiley, Interscience division, New York.
- W.C. Ashar, T.D. Wallace (1963) A sampling of minimum absolute deviations estimators. Oper. Res., 11, 747-752.
- P. Assouad (1977) Un espace hypermetric non plongeable dans un espace L_1. C. R. Acad. Sc. Paris, T. 285, Serie A, 361-363.
- S. Baboolal, G.A. Watson (1981) Computational experience with an algorithm for discrete L_1 approximation. Computing, 27, 245-252.
- S.C. Banks, H.L. Taylor (1980) A modification to the discrete L_1 linear approximation algorithm of Barrodale and Roberts.
- SIAM J. on Scientific and Stat. Comput. 1, 187-190. G.D.I. Barr, J.F. Affleck-Graves, A.H. Money, M.L. Hart (1980a) Performance of a generalized algorithm for L_p-norm regression estimates. Tech. Rep. no. ALS-4, Aug., Univ. of CapeTown, Dept. of Math. Stat. South Africa.
- G.D.I. Barr, J.F. Affleck-Graves, A.H. Money, M.L. Hart (1980b) L_p norm estimation and the choice of p, a practical approach. Univ. of CapeTown, Dept. of Math. Stat., Tech. Rep. no. ALS-3, July.
- G.D.I. Barr, A.H. Money, J.F. Affleck-Graves, M.L. Hart (1980) L_p-norm estimation of a symmetric distribution. Univ. of CapeTown, Dept. of Math. Stat., Tech. Rep. no. ALS-5, Oct..
- G.D.I. Barr, A.H. Money, J.F. Affleck-Graves, M.L. Hart (1981a) Estimation of location for skewed data sets: a comparative study utilizing the data set published by Stigler. Univ. of CapeTown, Dept. of Math. Stat., Tech. Rep. no. ALS-7, may.
- G.D.I. Barr, A.H. Money, J.F. Affleck-Graves, M.L. Hart (1981b) Estimation of location for skewed data sets. Univ. of CapeTown, Dept. of Math. Stat., Tech. Rep. no. ALS-6, April.
- Barrodale (1967) Approximation in the L_1 and Ll norms by linear programming. Ph.D. thesis, Univ. of Liverpool, Liverpool, England.
- Barrodale (1968) L_1 approximation and analysis of data. Appl. Stat. 17,51-57.
- Barrodale (1970) On computing best L_1 approximations. In A. Talbot, Approximation theory Academic Press, New York, 205-215.
- Barrodale, C. Phillips (1975) Solution of an over-determined system of linear equations in Chebyshev norm. ACM Trans. on Math. Soft. 264-70.
- Barrodale, M.J.D. Powell, F.D.K. Roberts (1972) The differential correction algorithm for rational Ll approximation. SIAM J. Num. Anal., 9, 493-503.
- Barrodale, F.D.K. Roberts (1970) Application of mathematical programming to L_p approximation. In J.B. Rosen,O.L. Mangasarian, K. Ritter, Nonlinear programming Academic Press, New York, 447-64.
- Barrodale, F.D.K. Roberts (1973) An improved algorithm for discrete L_1 linear approximation. SIAM J. Numer. Anal., 10,839-848.
- Barrodale, F.D.K. Roberts (1974) Algorithm 478: Solution of an overdetermined system of equations in the L_1 norm. Comm. ACM, 17, 319-320.
- Barrodale and F.D.K. Roberts (1977) Algorithms for restricted least absolute value estimation. Comm. Stat. B6(4), 353-363.
- Barrodale and F.D.K. Roberts (1978) An efficient algorithm for discrete L_1 linear approximation with linear constraints. SIAM J. Numer. Anal., 15, 603-611.
- Barrodale and F.D.K. Roberts, C.R. Hunt (1970) Computing best L_p approximations by functions nonlinear in one parameter. Comp. J., 13, 382-386.
- Barrodale, C. Phillips (1974) An improved algorithm for discrete Chebyshev linear approximation. Proc. 4th Manitoba conf. on numer. math. Univ. of Manitoba, Winnipeg, Manitoba, 177-190.
- Barrodale, A. Young (1966) Algorithms for best L_1 and Ll linear approximations on a discrete set. Numer. Math., 8, 295-306.
- R.H. Bartels, A.R. Conn (1977) LAV regression: A special case of piecewise linear minimization. Comm. Stat., B6, 329-340.
- R.H. Bartels, A.R. Conn (1980a) Linearly constrained discrete L_1 problems. ACM Trans. Math. Soft. 6, 594-608.
- R.H. Bartels, A.R. Conn (1980b) Algorithm 563, A program for linearly constrained discrete L_1 problems. ACM trans. Math. Soft., 6, 609-614.
- R.H. Bartels, A.R. Conn (1982) An approach to nonlinear L_1 data fitting. In J.P. Hennart (ed.), Numerical analysis, Cocoyoc, Springer Verlag, 45-58.
- R.H. Bartels, A.R. Conn, C. Charalambous (1978) On Cline's direct method for solving overdetermined linear system in the Ll sense, SIAM J. Numer. Anal., 15, 255-270.
- R.H. Bartels, A.R. Conn, Y. Li (1987) Primal methods are better than dual methods for solving overdetermined linear systems in the Ll sense? Res. Rep. CS-87-12, Univ. of Waterloo, Comp. Sci. Dept.
- R.H. Bartels, A.R. Conn, J. Sinclair (1976) The L_1 solution to an overdetermined linear system. Proc. 9th Ann. Symp. Interface Stat. In C.D.C. Hoaglin (ed.) Boston, Prindle, Weber and Schmidt Inc., 120-7.
- R.H. Bartels, A.R. Conn, J. Sinclair (1978) Minimization technique for piecewise differentiable functions: The L_1 solution to an overdetermined linear system. SIAM J. Numer. Anal. 15, 224-241.
- R.H. Bartels, G.H. Golub (1968a) Algorithm 328: Chebyshev solution to an overdetermined linear system. Comm. ACM 11, 428-430. R.H. Bartels, G.H. Golub (1968b) Stable numerical methods for obtaining the Chebyshev solution to an overdetermined system of equations, Comm. ACM, 11, 401-406.
- R.H. Bartels, G.H. Golub (1969) The simplex method of linear programming using LU decomposition. Comm. ACM, 12, 266-268.
- G.W. Bassett (1973) Some properties of the least absolute error estimator. Ph.D. thesis. Dept. of Econ., Univ. of Michigan.
- G.W. Bassett, Jr. (1987) A p-subset property of L_1 and regression quantile estimates. Work. Pap., Dept. of Economics, Univ. of Illinois-Chicago.
- G.W. Bassett, Jr. (1988a) A p-subset property of L_1 and regression quantile estimates. CSDA, 6(3), 297-304.
- G.W. Bassett, Jr. (1988b) A property of the observations fit by the extreme regression quantiles. CSDA, 6(4), 353-360.
- G.W. Bassett, Jr., R. Koenker (1978) Asymptotic theory of least absolute error regression. JASA, Sep., 73, no. 363, 618-22.
- G.W. Bassett, Jr., R. Koenker (1982) An empirical quantile function for linear models with iid errors. JASA, 77, no. 378, 407-415.
- J. Bejar (1956) Regression en mediana y la programacion lineal, Trabajos de Estadistica 7, 141-58.
- J. Bejar (1957) Calculo practico de la regression en mediana Trabajos de Estadistica, 8, 157-173.
- B. Bidabad (1984a) Goal programming, optimal decision process with different priorities and its computer program. Plan and Budget

Ministry, Tehran, Iran.
- B. Bidabad (1984b) Determining industrial location for 1992 by using goal programming method. Plan and Budget Ministry, Tehran, Iran.
- B. Bidabad (1984c) Zero-one programming, optimal decision making with bivalent variables and its computer program. Plan and Budget ministry, Tehran, Iran.
- B. Bidabad (1987a) Least absolute error estimation. Submitted to the First International Conference on Statistical Data Analysis Based on the L_1-norm and Related Methods, Neuchatel, Switzerland.
- B. Bidabad (1987b) Least absolute error estimation, part II. Submitted to the First International Conference on Statistical Data Analysis Based on the L_1-norm and Related Methods, Neuchatel, Switzerland.
- B. Bidabad (1988a) A proposed algorithm for least absolute error estimation. Proc. of the Third Seminar of Mathematical Analysis. Shiraz Univ., 24-34, Shiraz, Iran.
- B. Bidabad (1988b) A proposed algorithm for least absolute error estimation, part II. Proc. of the Third Seminar of Mathematical Analysis, Shiraz Univ., 35-50, Shiraz, Iran.
- B. Bidabad (1989a) Discrete and continuous L_1 norm regressions, proposition of discrete approximation algorithms and continuous smoothing of concentration surface, Ph.D. thesis, Islamic Azad Univ., Tehran, Iran.
- B. Bidabad (1989b) Discrete and continuous L_1 norm regressions, proposition of discrete approximation algorithms and continuous smoothing of concentration surface, Ph.D. thesis, Islamic Azad Univ., Tehran, Iran. Farsi translation.
- R. Blattberg, T. Sargent (1971) Regression with non-Gaussian stable disturbances: some sampling results, Econometrica, May, 501-510.
- P. Bloomfield, W. Steiger (1980) Least absolute deviations curve fitting. SIAM J. Sci. Stat. Comput. 1, 290-301.
- P. Bloomfield, W. Steiger (1983) Least absolute deviations: theory, applications and algorithms. Birkhauser, Boston.
- P.T. Boggs (1974) A new algorithm for Chebyshev solution of overdetermined linear system. Math. Comp., 28, 203-218.
- M.S. Borowsky (1976) Algorithms for solving the dual problem for Av=b with varying norms. J. Approx. Theory, 17, 107-114.
- R.J. Boscovich (1757) De litteraria expeditione per pontificiam ditionem, et synopsis amplioris operis..., 'Bononiensi' scientiarum et artum instituto atque academia commetarii, vol.4, 353-396. Reprinted with a Serbo-Croatian translation by N. Cubranic, Institute of higher geodesy, Univ. of Zagreb 1961.
- R.J. Boscovich (1760) De recentissimis graduum dimensionibus et figura, ac magnitudine terrae inde derivanda. Philosophiae recentioris, a benedicto stay in Romano archigynasis publico eloquentare professore, vesibus traditae, libri X, cum adnotianibus et supplementas P. Rugerii Joseph Boscovich, S.J., 2, 406-426.
- A.L. Bowley (1902) Methods of representing the statistics of wages and other groups not fulfilling the normal law of error, II: applications to wage statistics and other groups. J. of the Roy. Stat. Soc., 65, 331-54.
- A.L. Bowley (1928) F.Y. Edgeworth's contributions to mathematical statistics. London, Roy. Stat. Soc..
- G.E.P. Box, G.C. Tiao (1962) A further look at robustness vi Bayes' theorem. Biometrika, 419-432.
- D. Bradu (1987a) L_1 fit, median polish and conjugate gradients. CSIR Tech. Rep. TWISK 509, National Res. Inst. for Math Sci. CSIR, Pretoria.
- D. Bradu (1987b) An n-median polish algorithm. CSDA, 5, 327-336.
- M. Brannigan, S.A. Gustafson (1987) H-sets in convex programming and constrained approximation. Work. Pap., Rogaland Univ., Stavanger.
- H.D. Brecht (1976) Regression methodology with observation errors in the explanatory variables. Decis. Sci. 7, 57-65.
- J.J. Brennan, L.M. Seiford (1987) Linear programming and L_1 approximation using the method of vanishing Jacobians. CSDA, 5, 263-276.
- B.M. Brown (1980) Median estimates in a simple linear regression. Australian J. of Stat., 22, 154-165.
- B.M. Brown, T.P. Hettmansperger (1987) Invariant tests in bivariate models and the L_1-criterion. In Y. Dodge (ed.) Statistical data analysis based on the L_1-norm and related methods, NHPC, 333-344.
- C. Bruen (1938) Methods for the combination of observations modal points or most lesser-deviations, mean loci or least squares, and mid point of least range or least greatest-deviation. Metron 13, 61-140.
- J.J. Buckley, A.H. Kvanli (1981) The MAD estimator: when is it non-linear? Applications to two-way design models. Comm. Stat. A10, 2581-2590.
- Burgoyne (1965) Polynomial approximation by Edgeworth's method. Univ. of London.
- S. Busovaca (1985) Handling degeneracy in a nonlinear L_1 algorithm. Ph.D. thesis, Univ. of Waterloo, Waterloo, Ontario.
- V.A. Cabot, R.L. Francis, M.A. Stary (1970) A network flow solution to a rectilinear distance facility location problem. AIIE trans., vol. 2.
- P.H. Calamai, A.R. Conn (1987) A projected Newton method for L_p norm location problems. Math. Prog. 38, 75-109.
- G. Le Calve (1987) L_1-embeddings of data structure (I,D). In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and related methods. 195-202., NHPC.
- J.M. Chambers (1977) Computational methods for data analysis Wiley, New York.
- D.G. Champernowne (1953) A model of income distribution. Economic J., 63, 318-351.
- Charnes, W.W. Cooper, R.O. Ferguson (1955) Optimal estimation of executive compensation by linear programming. Manag. Sci. 1, 138-151.
- Charalambous, A.R. Conn (1978) An efficient method to solve the minimax problem directly. SIAM J. Numer. Anal., 15, no.1, 162-187.
- Charalambous (1979) On conditions for optimality of nonlinear L_1 problem. Math. Prog. 17, 123-135.
- E.W. Cheney (1966) Introduction to approximation theory McGraw-Hill, New York.
- W. Cheney, A.A. Goldstein (1958) Note on a paper by Zuhovickii concerning the chebysheff problem for linear equations. SIAM J. Numer. Anal. 6, 233-239.
- J.A. Chisman (1966) Using linear programming to determine time standards, J. Ind. Engin. 17, 189-191.
- Su Chun (1987) Some results on the L_p-convergence (pr1) of U-statistics. CSDA, 5, 321-326.
- D.I. Clarke (1981) Finite algorithms for linear optimization problems. Ph.D. thesis, Australian National Univ., Canberra.
- F.H. Clarke (1983) Optimization and nonsmooth analysis. Wiley, New York.
- J.F. Claerbout, F. Muir (1973) Robust modeling with erratic data. Geophysics 38, 826-844.
- A.K. Cline (1970) Uniform approximation as a limit of approximations. Ph.D. thesis, Univ. of Michigan, Ann, Arbor, Michigan.
- A.K. Cline (1972) Rate of convergence of Lawson's algorithm, Math. of Comp., 26, 167-176.
- A.K. Cline (1976) A descent method for the uniform solution of overdetermined system of linear equations. SIAM J. Numer. Anal., 13, 293-309.

- K.O. Cogger (1979) Time-series analysis and forecasting with an absolute error criterion. TIMS Studies in Manag. Sci., S. Markidakis, S.C. Wheelwright (eds.) NHPC, 189-201.
- T.F. Coleman (1978) A note on 'New algorithms for constrained minimax optimization', Math. Prog., 15, 239-242.
- A.R. Conn (1975) Optimization of microwave networks. IEEE Tran. on Microwave Theory and Techniques, Oct., 834-838.
- A.R. Conn (1976) linear programming via a nondifferentiable penalty function. SIAM J. Numer. Anal., 13, 145-154.
- A.R. Conn (1984) Nonlinear programming, exact penalty functions and projection techniques for non-smooth functions.
- Tech. Rep. no. CS-84-26. Dept. of Comp. Sci., Univ. of Waterloo, Ontario, Canada.
- A.R. Conn, N.I.M. Gould (1987) An exact penalty function for semi-infinite programming. Math. Prog., 37, 19-40.
- B. Cornell, J.K. Dietrich (1978) Mean-absolute-deviation versus least-squares regression estimation of beta coefficients. J. of Financial and Quantitative analysis 13, 123-131.
- F. Critchley (1980) Optimal norm characterizations of multidimensional scaling methods and some related data analysis problem. In E. Diday et al (eds.) Data analysis and informatics, NHPC, 209-229.
- D.C. Crocker (1969) Linear programming technique in regression analysis, the hidden danger. A.I.E.E. Trans., 1, 112-126.
- M. Csorgo, L. Horvath (1987) Asymptotics for L_p-norms of naive estimators of densities. Tech. Rep. series of Lab. Res. Stat. Prob. no. 96, Carleton Univ., Ottawa, Canada.
- M. Csorgo, L. Horvath (1988) Asymptotics for L_p-norms of kernel estimators of densities. CSDA, 6(3), 241-250.
- M. Csorgo, E. Gombay, L. Horvath (1987) Asymptotics for L_p-norms of kernel estimators of density under random censorship. Tech. Rep. series of Lab. Res. Stat. Prob., Carleton Univ., Ottawa, Canada.
- M. Davies (1976) Linear approximation using the criterion of least total deviations. J. Roy. Stat. Soc. B29, 101-109.
- J. Descloux (1963) Approximation in L_p and Chebyshev approximations. SIAM J. 11, 1017-1026.
- L. Devroye (1983) The equivalence of weak, strong and complete convergence in L_1 for kernel density estimates. Ann. of Stat., 11, 896-904.
- L. Devroye (1985) A note on the L_1 consistency of variable kernel estimates. Ann. Stat., 13, 1041-1049.
- L. Devroye, L. Gyorfi (1985) Non parametric density estimation, the L_1 view, Wiley, New York.
- L. Devroye, T.J. Wagner (1979) The L_1 convergence of kernel density estimates. Ann. Stat., 7, 1136-1139.
- L. Devroye, T.J. Wagner (1980) On the L_1 convergence of kernel estimators of regression functions with applications in discrimination. Z. Wahrsch, verw. Gebiete, 51, 15-25.
- T.E. Dielman (1984) Least absolute value estimation in regression models: An annotated bibliography. Comm. Stat. 13, 513-41.
- T. Dielman, R. Pfaffenberger (1982) LAV (Least Absolute Value) estimation in linear regression: A review, TIMS studies in the Manag. Sci., 19, 31-52.
- T. Dielman, R. Pfaffenberger (1984) Computational algorithms for calculating least absolute value and Chebyshev estimates for multiple regression. Amer. J. Math. Manag. Sci., 4, 169-197.
- J.J. Dinkel, R.C. Pfaffenberger (1981) Constrained L_1 estimation via geometric programming. European J. Oper. Res. 7, 299-305.
- P.J. Dhrymes (1978) Mathematics for econometrics. Springer-Verlag, New York.
- Y. Dodge (1984) Robust estimation of regression coefficients by minimizing a convex combination of least squares and least absolute deviations. Comp. Stat. Quart., 1, 139-153.
- Y. Dodge (1987) An introduction to statistical data analysis L_1-norm based. In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and related methods. NHPC. Reprinted in CSDA, 5, 239-254.
- Y. Dodge, J. Jureckova (1987) Adaptive combination of least squares and least absolute deviations estimators. In Y. Dodge (ed.) Statistical data analysis based on the L_1-norm and related methods. 275-287, NHPC.
- Y. Dodge, J. Jureckova (1988) Adaptive combination of M-estimator and L_1-estimator. In Y. Dodge, V.V. Fedorov, H.P. Wynn (eds.) Optimal design and analysis of experiments, 167-176, NHPC.
- W. Domschke, A. Drext (1984) An international bibliography for location-and layout-planning. Instut fuer Unternehmensforschung und Informatik, Hochschule der Eundeswehr, postach 700822 D-2000 Hamburg 70.
- E.L. Dougherty, S.T. Smith (1966) The use of linear programming to filter digitized map data. Geophysics, 31, 253-259.
- Z. Drezner, G.O. Wesolowsky (1978) A trajectory method for the optimization of multi-facility location problem with L_p distances. Manag. Sci., 24, no. 14, 1507-1514.
- A.F. Dufton (1928) Correlation. Nature, 121, 866.
- J. Dupacova (1987a) Asymptotic properties of restricted L_1-estimates of regression. Feb., WP-87-18, Work. Pap. Inter. Inst. for App. Sys. Analysis, A-2361 Laxenburg, Austria.
- J. Dupacova (1987b) Asymptotic properties of restricted L_1-estimates of regression. In Y. Dodge (ed.) statistical data analysis based on the L_1-norm and related methods. 263-274, NHPC.
- C.S. Duris, V.P. Sreedharan (1968) Chebyshev and L_1 solutions of linear equations using least squares solutions. SIAM J. Num. Anal., 5, 491-505.
- C.S. Duris, M.G. Temple (1973) A finite step algorithm for determining the "strict" Chebyshev solution to Ax=b. SIAM J. Num. Anal. 10, 690-699.
- R. Dutter (1977) Numerical solution of robust regression problems, computational aspects, a comparison. J. of Stat. Computation and Simulation, 5, 207-238.
- R. Dutter (1987) A fortran program for robust and bounded influence regression. In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and related methods. 139-144. NHPC.
- F.Y. Edgeworth (1883) The method of least squares. Philosophical Magazine, 16, 360-375.
- F.Y. Edgeworth (1887a) On observations relating to several quantities. Hermathena, 6, 279-285.
- F.Y. Edgeworth (1887b) A new method of reducing observations relating to several quantities. Philosophical Magazine, 24, 222-223.
- F.Y. Edgeworth (1888) On a new method of reducing observation relating to several quantities. Philosophical Magazine, 25, 184-191.
- F.Y. Edgeworth (1902) Method of representing statistics of wage and other groups not fulfilling the normal law of error, I: mathematical considerations. J. Roy. Stat. Soc., 65, 325-331.
- F.Y. Edgeworth (1923) On the use of medians for reducing observations relating to several quantities. Philosophical Magazine, 6th series, 46, 1074-1088.
- Eisenhart (1961) Boscovich and the combination of observations. Ch. 9 of Whyte (1961, 200-212) reprinted in Kendall and Plackett (1977) studies in the history of statistics and probability, vol.II, Charles Griffin and Co. Ltd., High Wycombe 88-100.
- H. Ekblom (1973a) A note on nonlinear median estimators. JASA, 68, 431-2.

- H. Ekblom (1973b) Calculation of linear best L_p-approximations. BIT 13, 292-300.
- H. Ekblom (1974) L_p methods for robust regression. BIT 14, 22-32.
- H. Ekblom (1987) The L_1 estimate as limiting case of an L_p-or Huber estimate. In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and related methods. NHPC. 109-116.
- H. Ekblom, S. Henriksson (1969) L_p-criteria for the estimation of location parameters. SIAM J. Appl. Math., 17, 1130-1141.
- R.A. El-Attar, M. Vidyasagar, S.R.K. Dutta (1976) Optimality conditions for L_1-norm minimization Proc. 19th midwest symp. on circuits and systems, 272-275.
- R.A. El-Attar, M. Vidyasagar, S.R.K. Dutta (1979) An algorithm for L_1-norm minimization with application to nonlinear L_1-approximation. SIAM J. Numer. Anal., 16, 70-86.
- J.E. Estienne (1926-28) Introduction a une theorie rationnelle des erreurs d'observation. Revue d'artillerie 97(1926), 421-441; 98(1928), 542-562; 100(1927), 471-487.
- R.C. Fair (1974) On the robust estimation of econometric models. Ann. Econ. Soc. Measurement, 3, 667-77.
- Fama (1965) The behavior of stock market prices. J. Bus. 38, 34-105.
- E.F. Fama, R. Roll (1971) Parameter estimates for symmetric stable distributions. JASA 66, 331-338.
- R.W. Farebrother (1985) Unbiased L_1 and Ll estimation. Comm. Stat., A, 14,1941-1962.
- R.W. Farebrother (1987a) The theory of committee decisions and the double median method. CSDA, 5, 437-442.
- R.W. Farebrother (1987b) The historical development of the L_1 and Ll estimation procedures. In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and related methods. NHPC. 37-64.
- R.W. Farebrother (1987c) A simple recursive procedure for the L_1 norm fitting of a straight line. Work. Pap., Dept. of Econometrics and Social Stat. Univ. of Manchester, Manchester, M13 9PL, UK.
- R.W. Farebrother (1987d) Mechanical representation of the L_1 and L_2 estimation problems. In Y. Dodge (1987) Statistical data analysis based on the L_1 norm and related methods. NHPC. 455-464.
- R.W. Farebrother (1987e) A remark on AS108: multiple linear regression with minimum sum of absolute errors. Appl. Stat., 36, no. 1, 118.
- D.W. Fausett, J.H. Weber (1978) Mass spectral pattern recognition via techniques of mathematical programming. Analytical Chemistry, 50, 722-731.
- V.V. Fedorov (1987) Various discrepancy measures in model testing (two competing regression models). In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and related methods. NHPC, 357-366.
- B. Fichet (1987a) The role played by L_1 in data analysis. In Y. Dodge (ed.), Statistical data analysis based on the L_1 norm and related methods. NHPC, 185-194.
- B. Fichet (1987b) L_p-space in data analysis. First Conference of the International Federation of Classification Societies. Aachen.
- W.D. Fisher (1961) A note on curve fitting with minimum deviations by linear programming. JASA, 11, 359-362.
- R. Fletcher (1981) Numerical experiments with an exact L_1 penalty function method. In O.L. Mangasarian, R.R. Meyer, S.M. Robinson (eds.) Nonlinear programming 4. Academic Press, New York, 99-129.
- R. Fletcher (1984) An L_1 penalty method for nonlinear constraints. Rep. NA/81, Dept. of Math. Sci., Univ. of Dundee, Scotland.
- R. Fletcher, J.A. Grant, M.D. Hebden (1971) The calculation of best L_p approximations. Comp. J., 14, 276-279.
- R. Fletcher, J.A. Grant, M.D. Hebden (1974a) Linear minimax approximation as the limit of best L_p approximation. SIAM J. Numer. Anal., 11, 123-136.
- R. Fletcher, J.A. Grant, M.D. Hebden (1974b) The continuity and differentiability of parameters of best linear L_p approximations. J. Approx. Theory, 10, 69-73.
- A.B. Forsythe (1972) Robust estimation of straight line regression coefficients by minimizing pth power deviations. Technometrics, 14, 159-166.
- C.R. Forth (1974) Robust estimation techniques for population parameters and regression coefficients. M.S. thesis. Air Force Institute of Technology, Wright-Patterson,AFB, Ohio.
- R. Fourer (1985a) A simplex algorithm for piecewise-linear programming I: derivation and proof. Math. Prog., 33, 204-233.
- R. Fourer (1985b) A simplex algorithm for piecewise-linear programming II: Finiteness, feasibility and degeneracy. Tech. Rep., 85-03 (revised), Dept. of Ind. Engin. and Manag. Sci., The Tech. Inst., Northwestern Univ., Evanston, Illinois.
- R. Fourer (1986) A simplex algorithm for piecewise-linear programming III: Computational analysis and applications. Tech. Rep., 86-03, Dept. of Ind. Engin. and Manag. Sci., The Tech. Inst., Northwestern Univ., Evanston, Illinois.
- J.B.I. Fourier (1824) Solution d'une question particuliere au calcul des inegalites, second extrait. Histoire de l'academie des sciences pour 1824, 47-55. Reprinted in oeuvres de Fourier, 2. Paris, 1980, Gauthier-Villars, 325-328.
- E.L. Frome, R.D. Armstrong (1977) A robust procedure for estimating the trend-cycle component of an economic time series. In D. Hogben (ed.) Proc. of the tenth symposium on the interface. Gaithersburg: National Bureau of Standards.
- E.L. Frome, G.J. Yakatan (1980) Statistical estimation of the pharmokinetic parameters in the one compartment open model. Comm. Stat. B9, 201-222.
- Gaivoronski (1987) Numerical techniques for finding estimates which minimize the upper bound of absolute deviations. In Y. Dodge (ed.) Statistical data analysis based on the L1 norm and related methods. NHPC. 247-262.
- Galilei (1632) Dialogo dei massimi sistemi. J.S. GaLpin (1986) Robust and bounded influence regression. National Res. Inst. for Math. Sci. WNNR, CSIR, TWISK, Pretoria, South Africa.
- J.S. Ga$L_p$in, D.M. Hawkins (1987) Methods of L_1 estimation of a covariance matrix. CSDA, 5, 305-319.
- C.W. Ganger, D. Orr (1972) Infinite variance and research strategy in time series analysis. JASA 67, 275-285.
- C.B. Garcia, F.G. Gould (1983) An application of homotopy to solving linear programs. Math. Prog. 27, 263-282.
- C.F. Gauss (1809) Theoria motus corporum coelestium. In F. Perthes, I.H. Besser, Sectionbus conicis solem ambientium, Hamburg. Reprinted in his werke, vol. 7, F. Pethes, Gotha 1871. English translation by C.H. Davis, Little, Brown and Co., Boston, 1857. Reprinted by Dover Pub. New York, 1963.
- J.E. Gentle (1977) Least absolute value estimation: an introduction. Comm. Stat., B6, 313-28.
- J.E. Gentle, T.A. Hansen (1977) Variable selection under L_1. Proceedings of the statistical computing section A.S.A., 228-230.
- J.E. Gentle, W.J. Kennedy, V.A. Sposito (1976) Properties of the L_1 estimate space. Proc. Stat. Comp. section A.S.A. 163-164.
- J.E. Gentle, W.J. Kennedy, V.A. Sposito (1977) On least absolute values estimation. Comm. Stat., A6, 839-845.
- J.E. Gentle, S.C. Narula, V.A. Sposito (1987) Algorithms for unconstrained L_1 linear regression. In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and related methods. NHPC. 83-94.

- J.E. Gentle, V.A. Sposito (1976) On the invariance of certain estimators. Bull. Austral. Math. Soc., vol.14, 405-408.
- J.E. Gentle, V.A. Sposito, W.J. Kennedy (1977) On some properties of L_1 estimators. Math. Prog., 12, 139-140.
- J.E. Gentle, V.A. Sposito, S.C. Narula (1988) Algorithms for unconstrained L_1 simple linear regression. CSDA, 6(4), 335-340.
- W.M. Gentleman (1965) Robust estimation of multivariate location by minimizing p-th power deviations. Ph.D. thesis, Princeton Univ., New Jersey.
- J. Gilsinn, K. Hoffman, R.H.F. Jackson, E. Leyendecker, P. Saunder, D. Shier (1977) Methodology and analysis for comparing discrete L_1 approximation codes., Comm. Stat., B6, 399-413.
- F.R. Glahe, J.G. Hunt (1970) The small sample properties of simultaneous equation least absolute estimators vis-a-vis least squares estimators. Econometrica, 38, 742-753. K. Glashoff, R. Schultz (1979) Uber die genaue Berechnung von besten L_1-approximierenden. J. Approx. Theory 25, 280-293.
- S.M. Goldfeld, R.E. Quandt (1981) Econometric modelling with non-normal disturbances. J. of Econometrics, Nov., 17(2), 141-55.
- A.A. Goldstein, W. Cheney (1958) A finite algorithm for the solution of consistent linear equations and inequalities and for Tchebycheff approximation of inconsistent linear equations. Pacific J. Math., 8, 415-427.
- R. Gonin (1983) A contribution to the solving of nonlinear estimation problems. Ph.D. thesis, Univ. of CapeTown.
- R. Gonin (1986) Numerical algorithms for solving nonlinear L_p-norm estimation problems: part I; a first-order gradient algorithm for well-conditioned small residual problems. Comm. Stat. B, 15(3), 801-813.
- R. Gonin, A.H. Money (1985a) Nonlinear L_p-norm estimation: part I, on the choice of the exponent, p, where the errors are additive Comm. Stat. A14, 827-840.
- R. Gonin, A.H. Money (1985b) Nonlinear L_p-norm estimation: part II, asymptotic distribution of the exponent, p, as a function of the sample kurtosis. Stat. A14, 841-849.
- R. Gonin, A.H. Money (1987a) Outliers in physical processes: L_1-or adaptive L_p norm estimation. In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and related methods. NHPC. 477-454.
- R. Gonin, A.H. Money (1987b) A review of computational methods for solving the nonlinear L_1 norm estimation problem In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and related methods. NHPC. 117-130.
- R. Gonin, A.H. Money (1987c) Nonlinear L_p-norm parameter estimation. Draft manuscript, Marcel Dekker, New York.
- R. Gonin, S.H.C. du Toit (1987) Numerical algorithms for solving nonlinear L_p-norm estimation problem, part II-a mixture method for large residual and ill-conditioned problems. Comm. Stat. A16, no. 4.
- S. Gross, W.L. Steiger (1979) Least absolute deviation estimates in autoregression with infinite variance. J. Appl. Prob., 16, 104-116.
- o Groucher (1971) Best L_1 and Ll approximations. M.Sc. thesis, Birkbeck College, London Univ., London, England.
- S.A. Gustafson, K.O. Kortanek, W. Rom (1970) Non-Chebyshevian moment problems. SIAM J. Numer. Anal., vol. 7, no. 3, 335-342.
- L. Gyorfi (1987) Density estimation from dependent sample. In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and related methods. NHPC. 393-404.
- L. Gyorfi, E.C. Van der Meulen (1987) Density-free convergence properties of various estimators of entropy. CSDA, 5(4), 425-436.
- E.E. Hagen (1975) The economics of development. Irwin inc, Illinois.
- J. Hald (1981a) A 2-stage algorithm for nonlinear L_1 optimization. Rep. no. 81-03, Numerisk Instut. Danmark Tekniske Hojskole, 2800 Lyngby, Denmark.
- J. Hald (1981b) A two stage algorithm for linearly constrained nonlinear L_1 optimization. Methods of Oper. Res., 43, 87-103.
- J. Hald, K. Madsen (1985) Combined L_p and quasi-Newton methods for nonlinear L_1 optimization. SIAM J. Numer. Anal., 22, no.1, 68-80.
- M.L. Hand, V.A. Sposito (1980) Using the least squares estimator in the Chebyshev estimation. Comm. Stat., B9(1), 43-49.
- W. Hardle (1987) XploRe, a computing environment for exploratory regression. In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and related methods. North-Holland. 163-174.
- T.E. Harris (1950) Regression using minimum absolute deviations. Am. Stat., 4, 14-15.
- H.L. Harter (1974a) The method of least squares and some alternative, I. Int. Stat. Rev., 42, 147-174.
- H.L. Harter (1974b) The method of least squares and some alternative, II. Int. Stat. Rev., 42, 235-264.
- H.L. Harter (1975a) The method of least squares and some alternative, III. Int. Stat. Rev., 43, 1-44.
- H.L. Harter (1975b) The method of least squares and some alternative, IV Int. Stat. Rev., 43, 125-190, 273-278.
- H.L. Harter (1975c) The method of least squares and some alternative, V. Int. Stat. Rev., 43, 269-272.
- H.L. Harter (1976) The method of least squares and some alternative, VI. Int. Stat. Rev., 44, 113-159.
- H.L. Harter (1977) Nonuniqueness of absolute value regression. Comm. Stat. B6, 829-838.
- H.L. Harter (1981), Method of least p-th powers. In Encyclopedia of statistical science, 5, 464-467.
- A.C. Harvey (1977) A comparison of preliminary estimators for robust regression. JASA, 72, 910-13.
- A.C. Harvey (1978) On the unbiasedness of robust regression estimators. Comm. Stat., A7, 779-783.
- P. Hattenschwiler (1988) Goal programming becomes most useful using L_1-smoothing functions CSDA, 6(4), 369-384.
- W.M. Haussler (1984) Computational experience with an eigen vector algorithm for robust L_p-discrimination. Com. Stat. Q. 1, 233-244.
- W.J. Heiser (1987) Correspondence analysis with least absolute residuals CSDA, 5, 337-356.
- W.J. Heiser (1988) Multidimensional scaling with least absolute residuals. To appear in H.H. Boc (ed.), Classification and related methods of data analysis, (IFCS'87). NHPC.
- S. Henriksson (1972) On a generalization of L_p-approximation and estimation. Thesis, Dept. of Computer Sci., Lund Univ., Sweden.
- R.W. Hill, P.W. Holland (1977) Two robust alternatives to least squares regression. JASA, 72, 828-833.
- C.R. Hobby, J.R. Rice (1965) A moment problem in L_1 approximation. Proc. Amer. Math. Soc., 16, 665-670.
- K.L. Hoffman, D.R. Shier (1980a) A test problem generator for discrete linear L_1 approximation problems. ACM Trans. Math. Soft., 6, 587-593.
- K.L. Hoffman, D.R. Shier (1980b) A test problem generator for discrete linear L_1 approximation problems. ACM Trans. Math. Soft., 6, 615-617.
- W.W. Hogan (1976) Norm minimizing estimation and unbiasedness. Econometrica, vol. 44, no.3, May.
- P.W. Holland, R.E. Welsch (1977) Robust regression using iteratively reweighted least-squares. Comm. Stat., A6, 813-827.
- L. Horvath (1987) Asymptotic normality of L_p-norms of density estimators. Tech. Rep. series of Lab Res. Stat. Prob., no.3, Carleton Univ., Ottawa, Canada.
- Th.V. Hromadka II, Ch.Ch. Yen, G.F. Pinder (1987) The best approximation method. An introduction. Springer-Verlag, Berlin.
- P.J. Huber (1987) The place of the L_1-norm in robust estimation. In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and

related methods. NHPC. Reprinted in CSDA, 5, 255-262.
- C.R. Hunt (1970) Best L_p approximation by certain nonlinear functions. M.Sc. thesis, Univ. of Victoria, Univ. of Victoria, B.C. Canada.
- J.G. Hunt, J.M. Dowling, F.R. Glahe (1974) L_1-estimation in small samples with Laplace error distributions Decision Sci., 5, 22-29.
- o Imai, K. Kato, P. Yamamoto (1987) A linear-time algorithm for linear L_1-approximation of points. Tech. Rep. CSCE-87-C30. Dept. of Comp. Sci. and Comm. Engin., Kyushu Univ. 36, Fukuoka 812, Japan.
- K. Jajuga (1987) A clustering method based on the L_1-norm. CSDA, 5, 357-371.
- K. Jittorntrum, M.R. Osborne (1980) Strong uniqueness and second order convergence in nonlinear discrete approximation. Numer. Math., 34, 439-455.
- Joe, R. Bartels (1983) An exact penalty method for constrained, discrete linear L_l data fitting. SIAM J. Sci. Stat. Comput., vol.4, no.1, 69-84.
- L.A. Josvanger, V.A. Sposito (1983) L_1-norm estimates for the simple regression problem. Comm. Stat. B12, 215-21.
- Jureckova (1983) Trimmed polynomial regression. Commentationes Mathematicae Universitatis Carolinae, 24, 4, 597-607.
- Jureckova (1984) Regression quantiles and trimmed least squares estimator under a general design. Kybernetika, vol.20, no.5, 345-357.
- Jureckova, P.K. Sen (1984) On adaptive scale-equivalent M-estimators in linear models. Stat. and Decision supplement issue, no.1, 31-46.
- K.R. Kadiyala (1972) Regression with non-Gaussian stable disturbances: some sampling results. Econometrica, July.
- N. Kaergard (1987) Estimation criterion, residuals and prediction evaluation. CSDA, 5, 443-450.
- S.W. Kahng (1972) Best L_p approximation. Math. of Comp., 26, 505-508.
- L.V. Kantorovich, G.P. Akilov (1964) Functional analysis in normed spaces. Pergamon Press, Oxford England.
- L.A. Karlovitz (1970a) An algorithm for the best L_p approximation. J. Approx. Theory. 3, 123-127.
- L.A. Karlovitz (1970b) Construction of nearest points in the L_p, even and L_l norms. J. Approx. Theory, 3, 123-127. O.J.
- Karst (1958) Linear curve fitting using least deviations. JASA, 53, 118-132.
- Kaufman, P.J. Rousseeuw (1987) Clustering by means of medoids. In Y. Dodge (ed.) Statistical data analysis based on the L1-norm and related methods. NHPC. 405-416.
- Y. Kawara (1979) Straight line fitting by minimizing the sum of absolute deviations. J. of the Japan Stat. Soc., 9, 47-64.
- J.E. Kelley (1958) An application of linear programming to curve fitting. SIAM J. Appl. Math., 6, 15-22.
- J.H.B. Kemperman (1984) Least absolute value and median polish. In Y.L. Tong (ed.), Inequalities in statistics and probability (IMS Lecture notes monograph series, vol.5), Inst. of Math. Stat., Hayward, CA, 84-113.
- J.H.B. Kemperman (1987) The median of a finite measure on a Banach space. In Y. Dodge (ed.) Statistical data analysis based on the L_1-norm and related methods. NHPC. 217-230.
- W.J. Kennedy, J.E. Gentle (1977) Examining rounding error in least absolute values regression computations. Comm. Stat., B6, 415-420.
- W.J. Kennedy, J.E. Gentle (1978) Comparisons of algorithms for minimum L_p norm linear regression. Proc. of Computer Sci. and Stat., Tenth Annual Symposium on Interface. D. Hogben (ed.), U.S. government printing office. Washington D.C., 373-378.
- W.J. Kennedy, J.E. Gentle (1980) Statistical computing. New York, Marcel Dekker.
- W.J Kennedy, J.E. Gentle, V.A. Sposito (1977) Comparisons of algorithms for L_1 estimation in the linear model. Paper presented at Midwestern Regional Meeting of IMS, Madison, WI. (Available from the second author).
- W.J Kennedy, J.E. Gentle, V.A. Sposito (1977) A computer oriented method for generating test problems for L_1 regression. Comm. Stat., B6, 21-27.
- B. Kim ( ) L_p norm estimation procedures and an L_1-norm algorithm for unconstrained and constrained estimation for linear models. VPI & SU, Blacksurg, VA 24061 USA.
- E.A. Kiountouzis (1971) Optimal L_p approximation techniques and data analysis. Bull. Soc. Math. Greece, 12, 191-206.
- E.A. Kiountouzis (1972) Mathematical programming and best linear L_p approximation. Extrait du Bull. de la Soc. Math. de Grece Novelle Serie, Tom 13, Fassc. 1, 46-57.
- E.A. Kiountouzis (1973) Linear programming techniques in regression analysis. Appl. Stat., 22, 69-73.
- Klingman, J. Mote (1982) Generalize network approaches for solving least absolute value and Tchebycheff regression problems. TIMS studies in the Mana. Sci., 19.
- J. Kmenta (1986) Elements of econometrics. MacMillan, New York.
- R. Koenker (1987) A comparison of asymptotic testing methods for L_1-regression. In Y. Dodge (ed.) Statistical data analysis based on the L_1-norm and related methods. North-Holland. 287-295.
- R. Koenker, G. Bassett, Jr. (1978) Regression quantile. Econometrica, vol. 46, no. 1 (Jan.).
- R. Koenker, G. Bassett, Jr. (1982a) Test of linear hypothesis and L_1 estimation. Econometrica, vol.50, no. 6, 1577-1583.
- R. Koenker, G. Bassett, Jr. (1982b) Robust tests for heteroskedasticity based on the regression quantiles. Econometrica, vol. 50, no. 1, Jan., 43-61.
- R. Koenker, G. Bassett, Jr. (1984) Four (pathological) examples in asymptotic statistics. The Amer. Statistician, Aug., vol.38, no.3, 209-212.
- R. Koenker, G. Bassett, Jr. (1985) On Boscovich's estimator. Ann. of Stat., 13, 1625-1628.
- W.W. Kotiuga (1982) Power system state estimation using least absolute value techniques. Ph.D. thesis, Univ. of Waterloo.
- W.W. Kotiuga, M. Vidyasagar (1982) Bad data rejection properties of weighted least absolute value techniques applied to static state estimation. IEEE Trans. on Power Apparatus and Systems. PAS-101, 844-853.
- B.R. Kripke, T.J. Rivlin (1965) Approximation in the metric of $L_1(X,u)$. Trans. Amer. Math. Soc., 119, 101-22.
- K. Kveton (1987) Method of averages as an alternative to L_1-and L_2-norm methods in special linear regression problems. CSDA, 5, 407-414.
- P.S. Laplace (1793) Sur quelques points du system du monde. Memoires de l'Academie Royale des Science de Paris. Annee 1789, 1-87. Reprinted in Oeuvres completes de Laplace II. Paris, Gauthier-Villars, 1985, 477-558.
- P.S. Laplace (1799) Traite des mecanique celeste, 2. Paris; J.B.M. Depart. Reprinted as oeuvres completes de Laplace, 2. Paris; Gauthier-Villars 1878, 116-165.
- P.S. Laplace (1812) Theorie analytique des probabilites, Mme courcier Paris 1820 Reprinted in his oeuvres, vol.7, Imprimerie Royale, Paris, 1847, and Gauthier-Villars et fils, Paris 1886.
- P.S. Laplace (1818) Duexieme supplement to Laplace (1812).
- J.L. Larson, A.H. Sameh (1980) Algorithms for round of error analysis a relative error approach. Computing 24, 275-297.
- K.D. Lawrence, D.R. Shier (1981) A comparison of least square and least absolute deviation regression models for estimation Weibull parameters. Comm. Stat., B10, 315-326.

- C.L. Lawson (1961) Contribution to the theory of linear least maximum approximations. Ph.D. thesis, Univ. of California, Los Angeles, California.
- Lazarski (1975a) Approximation of continuous functions in the space L_1. Automatika, 487, 85-93. E. Lazarski (1975b) The approximation of the continuous function by the polynomials of power functions in L_1 space. Automatika, 487, 95-106.
- Lazarski (1975c) On the necessary conditions of the uniqueness of approximation by the polynomials of power functions in L_1 space. Automatika, 487, 107-117.
- Lazarski (1977) Approximation of continuous functions by exponential polynomials in the L_1 space. Automatika, 598, 82-87.
- M.G. Lejeune, P. Sarda (1988) Quantile regression: a non parametric approach. CSDA, 6(3) 229-240.
- J.T. Lewis (1969) Approximation with convex constraints. Doctoral thesis, Brown Univ., Providence, R.I.
- J.T. Lewis (1970) Computation of best one-sided L_1 approximation. Math. Comp., 24, 529-536.
- R.F. Love (1974) The dual of a hyperbolic approximation to the generalized constrained multi-facility location problem with L_p distances Manag. Sci., vol. 21, 22-23.
- R.F. Love, J.G. Morris (1975) The use of nonlinear programming for solving facility location problems involving L_p distances using convex programming. Oper. Res., vol. 23, no.3, 581-588.
- G.S. Maddala (1977) Econometrics. McGraw-Hill.
- K. Madsen (1975) An algorithm for minimax solution of overdetermined systems of nonlinear equations. J. Inst. Math. and Appl., 321-328.
- K. Madsen (1985) Minimization of non-linear approximation functions. Copenhagen.
- B. Mandelbrot (1960) The Pareto-Levy law and the distribution of income. Inter. Econ. Rev., 1, 79-106.
- B. Mandelbrot (1961) Stable Paretian random functions and multiplicative variation of income. Econometrica, 29, 517-543.
- B. Mandelbrot (1963) New methods in statistical economics. J. of Political Economy, Oct.,421-440.
- Marazzi (1987) Solving bounded influence regression with ROBSYS. In Y. Dodge (ed.) Statistical data analysis based on the L1 norm and related methods. NHPC. 145-163.
- Marazzi (1988) Algorithms for the computation of weights in bounded influence regression. CSDA 6(3), 251-276.
- Marazzi, A. Randriamiharisoa (1985) ROBETH-ROBSYS: a software for robust statistical computing. Doc. no. 0,1,2,3,4,6. Institut Universitaire de Medecin Sociale et Preventive Lausanne, Switzerland.
- J.S. Marron (1987) What does optimal bandwidth selection mean for non parametric regression estimation. In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and related methods. NHPC, 379-392.
- C.L. Mathieu (1816) Sur les experiences du pendule, faites par les navigateurs espagnol, en differens points du globe. Connaissance des tems, 314-332.
- J.W. McKean, R.M. Schrader (1984) A comparison of the methods for studentizing the sample median. Comm. Stat., B 13(16), 751-773.
- J.W. McKean, R.M. Schrader (1987) Least absolute errors analysis of variance. In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and related methods. North-Holland. 297-306.
- J.W. McKean, G.L. Sievers (1987) Coefficients of determination for least absolute deviation analysis. Stat. and Prob. Letters, 5, 49-54.
- R.A. McLean, G.A. Watson (1980) Numerical methods for nonlinear discrete L_1 approximation problems. In L. Collatz,
- Meinardus, H. Werner (eds.) Numerical methods of approximation theory. ISNM 52, Birkhauser Verlag, Basel.
- C.R. McConnell (1987) On computing a best discrete L_1 approximation using the method of vanishing Jacobians. CSDA, 5, 277-288.
- G.F. McCormick, V.A. Sposito (1975) A note on L_1 estimation based on the median positive quotient. Appl. Stat., 24, 347-350.
- G.F. McCormick, V.A. Sposito (1976) Using the L_2-estimator in L_1-estimation. SIAM J. Numer. Anal., 13, 337-343.
- Megiddo, A. Tamir (1983) Finding least-distances line. SIAM J. Alg. Disc. Meth., 4, no. 2, 207-211.
- J. Meier (1987) A fast algorithm for clusterwise linear absolute deviations regression. OR Spektrum, 9, 187-189.
- M.S. Meketon (1986) Least absolute value regression. Work. Pap., AT&T Bell Laboratories, Holmdel, N.J.
- Melaku, G. Sadasivan (1987) L1-norm and other methods for sample allocation in multivariate stratified surveys. CSDA, 5, 415-424.
- J.A. Menendez, B. Salvador (1987) An algorithm for isotonic median regression. CSDA, 5, 399-406.
- Merle, H. Spath (1974) Computational experiences with discrete Lp-approximation. Computing, 12, 315-321.
- J.R. Meyer, R.R. Glauber (1964) Investment decisions, Economic forecasting and public policy. Harvard Business School Press, Cambridge, Massachusetts.
- Militky, J. Cap (1987) Application of Bayes approach to adaptive Lp nonlinear regression. CSDA, 5, 381-390.
- J.S.B. Mitchell (1987) Shortest rectilinear paths among obstacles. School of Oper. Res. and Ind. Engin. College of Engin. Cornell Univ.. Ithaca, New York, 14853.
- M.J. Mojarrad (1977) The application of comparative Monte Carlo methods to econometrics: an efficient Monte Carlo study of finite sample properties of iterative instrumental variables estimation. Ph.D. Diss., Univ. of Pennsylvania.
- Mond, M. Schechter (1976) A programming problem with an L_p norm in the objective function. J. Austral. Math. Soc., Ser., B, 19, 333-342.
- A.H. Money, J.F. Affleck-Graves, M.L. Hart (1978a) A review of some alternatives to least squares regression. Tech Rep. no. ALS-1, Sep., Univ. of CapeTown, South Africa.
- A.H. Money, J.F. Affleck-Graves, M.L. Hart (1978b) Least squares and some alternative: a simulation study. Tech. Rep. ALS-2. Univ. of CapeTown, South Africa.
- A.H. Money, J.F. Affleck-Graves, M.L. Hart, G.D.I. Barr (1982) The linear regression model: L_p norm estimation and the choice of p. Comm. Stat., 11, 89-109.
- R.M. Moroney (1961) The Haar problem in L_1. Proc. Amer. Math. Soc., 12, 793-795.
- J.G. Morris, W.a Verdini (1979) Minisum L_p distance location problems solved via a perturbed problem and Weiszfeld's algorithm. Oper. Res., 27, 1180-1188.
- Munoz Perez, A. Fernandez Palacin (1987) Estimating the quantile function by Bernstein polynomials. CSDA, 5, 391-398.
- V.I. Mudrov, V.L. Kushko, V.I. Mikhailov, E. Osvitskii (1968) Some experiments on the use of the least-modul: method in processing orbital data. Cosmic Res., 6, 421-431.
- W. Murray, M. Overton (1981) A projected Lagrangian algorithm for nonlinear L_1 optimization. SIAM J. Sci. Stat. Comput., 207-224.
- S.C. Narula (1987) The minimum sum of absolute errors regression. J. Quality Tech. 19, 37-45.
- S.C. Narula, J.F. Wellington (1977a) An algorithm for the minimum sum of weighted absolute errors regression. Comm. Stat., B(6), 341-352.
- S.C. Narula, J.F. Wellington (1977b) Prediction, linear regression and minimum sum of relative errors. Technometrics, 19, 185-190.

- S.C. Narula, J.F. Wellington (1977c) AS108, multiple linear regression with minimum sum of absolute error. Appl. Stat., 26, 106-111.
- S.C. Narula, J.F. Wellington (1979) Selection of variables in linear regression using the minimum sum of weighted absolute errors criterion. Technometrics, 21, no.3 Aug.
- S.C. Narula, J.F. Wellington (1982) The minimum sum of absolute errors regression, a state of the art survey. Inter. Stat. Rev., 50, 317-326.
- S.C. Narula, J.F. Wellington (1983) Selection of variables in linear regression, a pragmatic approach. J. of Stat. Comput. and Simul., 17, 159-172.
- S.C. Narula, J.F. Wellington (1985) Interior analysis for the minimum sum of absolute errors regression. Technometrics, 27, 181-188.
- S.C. Narula, J.F. Wellington (1987) An efficient algorithm for the MSAE and MMAE regression problems. Work. Pap., Virginia CommonWealth Univ., Richmond, VA 23284.
- H. Nikaido (1970) Introduction to sets and mappings in modern economics. North Holland, Amsterdam.
- H. Nyquist (1980) Recent studies on L_p-norm estimation. Ph.D. thesis, Univ. of Umea, Sweden.
- H. Nyquist (1983) The optimal L_p-norm estimator in linear regression models. Comm. Stat. A12, 2511-2524.
- H. Nyquist (1988) Least orthogonal absolute deviations. CSDA, 6(4), 361-368.
- H. Nyquist, A. Westlund (1977) L_1-versus L_2-norm estimation in interdependent systems when residual distributions are stable. Dept. of Stat., Univ. of Umea Presented at European Meeting of the Econometric Soc., Vienna, 5-9 Sep.
- W. Oberhofer (1982) The consistency of nonlinear regression minimizing the L_1 norm. Ann. of Stat., 10, 316-319.
- W. Oettli (1975) Symmetric duality, and a convergent subgradient method for discrete linear, constrained approximation problems with arbitrary norms appearing in the objective function and in the constraints. J. Approx. Theory, 14, 43-50.
- M.R. Osborne (1980) An algorithmic approach to nonlinear approximation problems. Approx. Theory III, 705-710.
- M.R. Osborne (1985) Finite algorithms in optimization and data analysis. Wiley, Chichester.
- M.R. Osborne (1987) The reduced gradient algorithm. In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and related methods. NHPC. 95-108.
- M.R. Osborne, S.A. Pruess, R.S. Womersley (1986) Concise representation of generalized gradients. J. of Austra. Math. Soc., Ser. B, 28, 57-74.
- M.R. Osborne, G.A. Watson (1967) On the best linear Chebyshev approximation. Computer J., 10, 172-177.
- M.R. Osborne, G.A. Watson (1971) On an algorithm for discrete nonlinear L_1 approximation. Computer J., 14, 184-188.
- M.R. Osborne, G.A. Watson (1978) Nonlinear approximation problems in vector norms. In Dundee, G.A. Watson (eds.) Numerical analysis. Springer Verlag.
- M.R. Osborne, G.A. Watson (1985) An analysis of the total approximation problem in separable norms, and an algorithm for the total L_1 problem. SIAM J. Sci. Stat. Comp., 6, 410-424.
- M. Overton (1982) Algorithms for nonlinear L_1 and Ll fitting. In M.J.D. Powell (ed.) Nonlinear optimization. Academic Press, London, 91-101
- R.M. Oveson (1968) Regression parameter estimation by minimizing the sum of absolute errors. Doctoral dissertation Harvard Univ., Cambridge, Massachusetts.
- H.J. Paarsch (1984) A Monte Carlo comparison of estimates for censored regression models. J. of Econometrics, 24, 197-213.
- M.J. Panik (1976) Classical optimization: foundation and extensions. NHPC.
- U. Peters, C. Willms (1983) Up-and down-dating procedures for linear L_1 regression. OR Spektrum 5, 229-239.
- R.C. Pfaffenberger, J.J. Dinkel (1978) Absolute deviations curve fitting: an alternative to least squares. In H.A. David (ed.) Contributions to survey sampling and applied statistics. Academic Press, New York, 279-294.
- Pilibossian (1987) A direct solving algorithm for a linear regression according to L_1-norm criteria. Work. Pap., L.S.T.A. Universite, Paris VI
- M.A. Porter, D.J. Winstanley (1979) Remark ASR29. Remarks on AS110: L_p norm fit of a straight line. Appl. Stat., 28, 112-113.
- S. Portnoy (1987) Using regression fractiles to identify outliers. In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and related methods. NHPC, 345-356.
- J.L. Powell (1983) The asymptotic normality of two-stage least absolute deviation estimators. Econometrica, 51, 1569-1575.
- J.L. Powell (1984) Least absolute deviations estimation for the censored regression model. J. of Econometrics, 25, 303-325.
- J.L. Powell (1986) Censored regression quantiles. J. of Econometrics, 32, 143-155.
- M.J.D. Powell, Y. Yuan (1984) Conditions for super linear convergence in L_1 and Ll solutions of overdetermined nonlinear equations. IMAJ Num. Anal., 4, 241-251.
- Prochazka (1988) Regression quantiles and trimmed least squares estimator in nonlinear regression model. CSDA, 6(4), 385-392.
- R. Prony (1804) Recherches physico-mathematiques sur la theorie des eaus courantes. Paris, l'imprimerie imperiale.
- V. Ptak (1958) On approximation of continuous functions in the metric uatb3x(t)3dt Czechoslovak Math. J. 8(83), 267-273.
- Rabinowitz (1968) Application of linear programming to numerical analysis. SIAM Rev., 10, 121-159.
- Rabinowitz (1970) Mathematical programming and approximation. In A. Talbot (ed.) Approximation Theory. Academic Press, 217-231.
- Ralston, P. Rabinowitz (1985) A first course in numerical analysis. Wiley, New York.
- J.O. Ramsay (1977) A comparative study of several robust estimates of slopes, intercept, and scale in linear regression. JASA, 72, 608-615
- M.R. Rao, V. Srinivasan (1972) A note on Sharpe's algorithm for minimum sum of absolute deviations in a simple regression problem. Manag. Sci., 19, 222-225.
- R.H. Rasche, J. Gaffney, A.Y.C. Koo, N. Obst (1980) Functional forms for estimating the Lorenz curve. Econometrica, 48, 1061-1062.
- H.D. Ratliff, J.C. Picard (1978) A cut approach to rectilinear distance facility location problem. Oper. Res., 26, 422-433.
- W. Rey (1975) On the least pth power methods in multiple regressions and location estimations. BIT, 15, 174-185.
- E.C. Rhodes (1930) Reducing observations by the method of minimum deviations. Philo. Mag., 7th series, 9, 974-92.
- J.R. Rice (1964a) On computation of L_1 approximations by exponentials, rationals, and other functions. Math. Comp., 18, 390-396.
- J.R. Rice (1964b) On nonlinear L_1 approximation. Arch. Rational Mech. Anal., 17 61-66.
- J.R. Rice (1964c) The approximation of functions, vol. I, linear theory. Reading Mass:, Addison-Wesley.
- J.R. Rice (1969) The approximation of functions, vol. II, linear theory. Reading Mass:, Addison-Wesley.
- J.R. Rice (1985) Numerical methods, software, and analysis. McGraw-Hill, ch. 11.
- J.R. Rice, J.S. White (1964) Norms for smoothing and estimation. SIAM Rev., 6, 243-256.
- P.D. Robers, A. Ben-Israel (1969) An interval programming algorithm for discrete linear L_1 approximation problem. J. Approx. Theory, 2,

323-336.

- P.D. Robers, S.S. Robers (1973) Algorithm 458: discrete linear L_1 approximation by interval linear programming. Comm. ACM, 16, 629-633.
- Ronchetti (1987) Bounded influence in regression: a review. In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and related methods. NHPC, 65-80.
- A.E. Ronner (1977) P-norm estimators in a linear regression model. Doctoral thesis, Rijkuniversiteit te Groningen.
- A.E. Ronner (1984) Asymptotic normality of p-norm estimators
- in multiple regression. Z. Wahrschein lickeitstheorie verw. Gebiete 66, 613-620.
- G. Roodman (1974) A procedure for optimal stepwise MSAE regression analysis. Oper. Res., 22, 393-399.
- Rosenberg, D. Carlson (1977) A simple approximation of the sampling distribution of least absolute residuals regression estimates. Comm. Stat., B6, 421-437.
- Rossi, H.D. Brunk (1987) L_1 and L_2 cross-validation for density estimation with special reference to orthogonal expansions. Tech. Rep. 120, Dept. of Stat., Oregon State Univ..
- Rossi, H.D. Brunk (1988) L_1 and L_2 cross-validation for density estimation with special reference to orthogonal expansions. CSDA, 6(3), 203-228.
- P.J. Rousseeuw (1984) Least median of squares regression. J. Amer. Stat. Asso., 79, 871-80.
- P.J. Rousseeuw (1987) An application of L_1 to astronomy. In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and related methods. NHPC, 437-446.
- P.J. Rousseeuw, A. Leroy (1987) Robust regression and outlier detection. Wiley-Interscience, New York.
- A.N. Sadovski (1974) AS74: L_1-norm fit of a straight line. Appl. Stat. 23, 244-248.
- A.K.Md.E. Saleh, P.K. Sen (1987) On the asymptotic distributional risk properties of pre-test and shrinkage L_1 estimators. CSDA, 5, 289-300.
- J.P. Schellhorn (1987) Fitting data through homotopy methods In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and related methods. NHPC. 131-138.
- E.J. Schlossmacher (1973) An iterative technique for absolute deviations curve fitting. JASA 68, 857-865.
- R.M. Schrader, J.W. McKean (1987) Small sample properties of least absolute errors analysis of variance. In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and related methods. NHPC. 307-322.
- Seneta (1983) The weighted median and multiple regression. Austral. J. Stat., 25(2), 370-377.
- Seneta, W.L. Steiger (1984) A new LAD curve-fitting algorithm: slightly overdetermined equation system in L_1. Discrete App. Math., 7, 79-91.
- Shanno, R.L. Weil (1970) Linear programming with absolute value functionals. Oper. Res., 19, 120-124.
- W.F. Sharpe (1971) Mean-absolute deviation characteristic lines for securities and portfolios. Manag. Sci., 18, B1-B13.
- S.J. Sheather (1986) A finite sample estimate of the variance of the sample median. Stat. and Prob. Letter., 4, 337-342.
- S.J. Sheather (1987) Assessing the accuracy of the sample median: estimated standard errors versus interpolated confidence intervals. In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and related methods. North-Holland. 203-216.
- S.J. Sheather, J.W. McKean (1987) A comparison of testing and confidence interval methods for the median. Stat. and Prob. Letter, 6, 31-36.
- H.D. Sherali, B.O. Skarpness, B. Kim (1987) An assumption-free convergence analysis for a perturbation of the scaling algorithm for linear programs, with application to the L_1 estimation problem. Dept. of Ind. Engin. and OR, Virginia Polytechnic Inst. and State Univ., Blacksburg, Virginia.
- O.B. Sheynin (1973) R.J. Boscovich's work on probability. Archive for history of exact sciences, vol. 9, 306-324, and vol. 28, 173.
- D.R. Shier, C.J. Witzgall (1978) Norm approximation problems and norm statistics., J. Res. Nat. Bur. Standards, 83, 71-74.
- O. Shisha (1974) A remark on an algorithm for best L_p approximation. J. Approx. Theory, 11, 283-284.
- R.I. Shrager, E. Hill (1980) Nonlinear curve-fitting in the L_1 and Ll norms. Math. Comput., 34, 529-541.
- A.F. Siegel (1983) Low median and least absolute residual analysis of two-way tables. JASA, 78, 371-374.
- R.L. Sielken, H.O. Hartley (1973) Two linear programming algorithms for unbiased estimation of linear models. JASA, 68, 639-.
- H.A. Simon (1955) On a class of skew distribution functions. Biometrika, 42, 425-440. Reprinted in H.A. Simon (1957) Models of man. New York, Wiley.
- R.R. Singleton (1940) A method for minimizing the sum of absolute values of deviations. Ann. of math. Stat., 11, 301-310.
- M.G. Sklar, R.D. Armstrong (1982) Least absolute value and Chebyshev estimation utilizing least squares results. Math. Prog., 24, 346-352.
- V.K. Smith, T.W. Hall (1972) A comparison of maximum likelihood versus BLUE estimators. Rev. Econ. Stat., 54, 186-190.
- S.A. Soliman, G.S. Christensen, A. Rouhi (1988) A new technique for curve fitting based on minimum absolute deviations. CSDA, 6(4), 341-352.
- D.L. Souvaine, J.M. Steele (1987) Time-and space-efficient algorithms for least median of squares regression. JASA, 82, no. 399, 794-801.
- Spath (1976) L_1 cluster analysis. Computing, 16, 379-387.
- Spath (1982) On discrete linear orthogonal L_p-approximation. Numerische Analysis, ZAMM 62, T354-T355.
- Spath (1985) Cluster dissection and analysis. Horwod, Chichester.
- H. Spath (1986a) Clusterwise linear least squares versus least absolute deviations regression, a numerical comparison for a case study. In W. Gaul, M. Schader (eds.) Classification as a tool of research. Elsevier, Amsterdam.
- H. Spath (1986b) Orthogonal least squares fitting with linear manifolds. Numer. Math., 48, 441-445.
- H. Spath (1986c) Algorithm, Clusterwise linear least absolute deviations regression. Computing, 37, 371-378.
- H. Spath (1987) Using the L_1 norm within cluster analysis. In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and related methods. NHPC. 427-434.
- H. Spath, G.A. Watson (1987) On orthogonal linear L_1 approximation. Numer. Math., 51, 531-543.
- V.A. Sposito (1976) A remark on algorithm AS74, L_1 norm fit of a straight line. Appl. Stat., 25, 96-97.
- V.A. Sposito (1982) On the unbiasedness of the L_p-norm estimators. JASA, 77, 652-653.
- V.A. Sposito (1987a) On median polish and L_1 estimators. CSDA, 5, 155-162.
- V.A. Sposito (1987b) Some properties of L_p-estimators in robust procedures. Marcel Dekker. In print.
- V.A. Sposito, M.L. Hand (1980) Optimal L_p estimators for symmetric distributions. Proc. of ASA, Stat. Comp. Sec.

- V.A. Sposito, M. Hand, G. McCormick (1977) Using an approximate L_1 estimator. Comm. Stat., B6, 263-268.
- V.A. Sposito, M.L. Hand, B. Skarpness (1983) On the efficiency of using the sample kurtosis in selecting optimal L_p estimators. Comm. Stat., B12, 265-272.
- V.A. Sposito, W.J. Kennedy, J.E. Gentle (1977) AS110: L_p norm fit of a straight line. Appl. Stat., 26, 114-118.
- V.A. Sposito, W.J. Kennedy, J.E. Gentle (1980) Useful generalized properties of L_1 estimators. Comm. Stat., A9, 1309-1315.
- V.A. Sposito, G.F. McCormick, W.J. Kennedy (1975) L_1 estimation strategies based on the simplex algorithm. In Proc. of the eighth symposium on the interface, J.W. France (ed.) Health science computing facility. UCLA, Los Angeles.
- V.A. Sposito, W.C. Smith (1976) On a sufficient and necessary condition for L_1 estimation. Appl. Stat., 25, 154-157.
- V.A. Sposito, W.C. Smith, G.F. McCormick (1978) Minimizing the sum of absolute deviations. J. of Appl. Stat. and Econometrics, Vandenhoeck and Ruprecht in Gottingen and Zurich series 12.
- V.A. Sposito, M. Tvejte (1984) The estimation of certain parameters used in L_1 interface. Proc. of Stat. Comp. Sec. of ASA, 267-270.
- Spyropoulos, E. Kiountouzis, A. Young (1973) Discrete approximation in the L_1 norm. Comp. J., 16, 180-186.
- V.P. Sreedharan (1969) Solution of overdetermined linear equation which minimize error in an abstract norm. Numer. Math., 13, 146-151.
- V.P. Sreedharan (1971) Least squares algorithms for finding solutions which minimize error in an abstract norm. Numer. Math., 17, 387-401.
- G. Stangenhaus (1987) Bootstrap and interface procedures for L_1 regression. In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and related methods. NHPC. 323-332.
- G. Stangenhaus, S.C. Narula (1987) Inference procedures for the L_1 regression. Work. Pap., Universidade Estadual de Compainas, Brasil.
- J.M. Steele, W.L. Steiger (1986) Algorithms and complexity for least median of squares regression. Discrete Appl. Math., 14, 39-100.
- Stiefel (1960) Note on Jordan elimination, linear programming and Tchebycheff approximation. SIAM J. Numer. Math., 2, 1-17.
- Steindl (1965) Random processes and growth of the firms. London, Griffin.
- W.L. Steiger (1980) Linear programming via L_1 curve fitting beats simplex. Abstracts, AMS, 80T-C26, 385-386.
- W. Steiger, P. Bloomfield (1980) Remark on a paper of Narula and Wellington. Technometrics, 22, 450.
- S.M. Stigler (1981) Gauss and invention of least squares. Ann. of Stat., 9, 465-474.
- S.M. Stigler (1984) Studies in the history of probability and statistics XL, Boscovich, Simpson and a 1760 manuscript note on fitting a linear relation. Biometrica, 71, 3, 615-620.
- Svanberg (1805) Exposition des operations faites en lappnie pour la determination d'un arc du meridien en 1801, 1802 et 1803,... Stockholm.
- Taguchi (1972a) On the two-dimensional concentration surface and extensions of concentration coefficient and Pareto distribution to the two dimensional case-I. Ann. of the Inst. of Stat. Math., vol. 24, no.2, 355-381.
- Taguchi (1972b) On the two-dimensional concentration surface and extensions of concentration coefficient and Pareto distribution to the two dimensional case-II. Ann. of the Inst. of Stat. Math., vol. 24, no.3, 599-619.
- T. Taguchi (1972c) Concentration polyhedron, two dimensional concentration coefficient for discrete type distribution and some new correlation coefficients etc. The Inst. of Stat. Math., 77-115.
- T. Taguchi (1973) On the two-dimensional concentration surface and extensions of concentration coefficient and Pareto distribution to the two dimensional case-III. Ann. of the Inst. of Stat. Math., vol. 25, no.1, 215-237.
- T. Taguchi (1974) On Fechner's thesis and statistics with norm p. Ann. of the Inst. of Stat. Math., vol. 26, no.2, 175-193.
- T. Taguchi (1978) On a generalization of Gaussian distribution. Ann. of the Inst. of Stat. Math., vol. 30, no.2, A, 211-242.
- T. Taguchi (1981) On a multiple Gini's coefficient and some concentrative regressions. Metron, vol. XXXIX -N.1-2, 5-98.
- T. Taguchi (1983) Concentration analysis of bivariate Paretoan distribution. Proc. of the Inst. of Stat. Math., vol. 31, no.1, 1-32.
- T. Taguchi (1987) On the structure of multivariate concentration. Submitted to the First International Conference on Statistical Data Analysis Based on the L_1 Norm and Related Methods, Neuchatel, Switzerland.
- T. Taguchi (1988) On the structure of multivariate concentration -some relationships among the concentration surface and two variate mean difference and regressions. CSDA, 6, 307-334.
- Takayama (1974) Mathematical economics. The Dryden Press, Illinois.
- L.D. Taylor (1974) Estimating by minimizing the sum of absolute errors. In P. Zarembka (ed.) Frontiers in econometrics. Academic Press.
- H.H. Taylor, S.C. Banks, J.F. McCoy (1979) Deconvolution with the L_1 norm. Geophysics, 44, 39-52.
- H. Theil (1965) The analysis of disturbances in regression analysis. JASA, 67, 1067-1079.
- H. Theil (1971) Principles of econometrics. Wiley.
- A Tishler, L. Zang (1982) An absolute deviations curve fitting algorithm for nonlinear models. TIMS studies in Manag. Sci., 19.
- D.S. Tracy, K.A. Khan (1987) MRPP tests in L_1 norm. CSDA, 5, 373-380.
- Trauwaert (1987) L_1 in fuzzy clustering. In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and related methods. NHPC. 417-426.
- J.W. Tukey (1977) Exploratory data analysis. Reading, Mass. Addison-Wesley.
- H.H. Turner (1887) On Mr. Edgeworth's method of reducing observations relating to several quantities. Phil. Mag. (5th series), 24, 466-470.
- K.H. Usow (1967a) On L_1 approximation: computation for continuous functions and continuous dependence. SIAM J. of Numer. Anal., 4, 70-88.
- K.H. Usow (1967b) On L_1 approximation: computation for discrete functions and discretization effect. SIAM J. Numer. Anal., 4, 233-244.
- Vajda (1987) L_1-distances in statistical inference: comparison of topological, functional and statistical properties. In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and related methods. NHPC. 177-184.
- C.W. Valentine, C.P. Van Dine (1963) An algorithm for minimax polynomial curve fitting for discrete data. J. ACM, 10, 283-290.
- J.F. Van Beeck-Calkoen (1816) Ver de theoric der Gemiddelde Waardij. Verhandlingen der K. Nederlandandsch Instituut Can Wetenschappen, 2, 1-19.
- Veidinger (1960) On the numerical determination of the best approximation in the Chebychev sense. Numer. Math., 2, 1-17.
- B.A. Von Lindenau (1806) Uber den Gebrauch der Gradmessungen zur bestimmung der gestalt der erde. Monatliche correspondenz zur befar derung der Erd-und Himmels-kunde, 14, 113-158.
- B.Z. Vulikh (1976) A brief course in the theory of functions of a real variable. Mir Publishers, Moscow.
- H.M. Wagner (1959) Linear programming technique for regression analysis. JASA, 54, 202-212.

- H.M. Wagner (1962) Nonlinear regression with minimal assumption. JASA, 57, 572-578.
- G.A. Watson (1973) On the best linear Chebyshev approximation. J. Approx. Theory, 7, 48-58.
- G.A. Watson (1973) The calculation of best linear one-side L_p approximations. Math. Comp. 27, 607-620.
- G.A. Watson (1977) On two methods for discrete L_p-approximation. Computing, 18, 263-266.
- G.A. Watson (1978) A class of programming problems whose objective function contains a norm. J. approx. Theory, 23, 401-411.
- G.A. Watson (1980) Approximation theory and numerical methods. Wiley, New York.
- G.A. Watson (1981) An algorithm for linear L_1-approximation of continuous functions. IMA J. Num. Anal., 1, 157-167.
- G.A. Watson (1982a) A globally convergent method for (constrained) nonlinear continuous L_1-approximation problems. In Numerical methods of approximation theory. ISNM59, Birkhauser Verlag.
- G.A. Watson (1982b) Numerical methods for linear orthogonal L_p approximation. IMA J. of Numer. Anal., 2, 275-287.
- G.A. Watson (1984a) Discrete L_1-approximation by rational functions. IMA J. Num. Anal., 4, 275-288.
- G.A. Watson (1984b) The numerical solution of total L_p approximation problems. In D.F. Griffiths (ed.) Numerical analysis. Dundee 1983, Lecture notes in mathematics, 1066, Springer Verlag, 221-238.
- G.A. Watson (1985a) On the convergence of eigenvector algorithms for robust L_p-discrimination. Comp. Stat. Quart., 4, 307-314.
- G.A. Watson (1985b) On a class of algorithms for total approximation. J. Approx. Theory, 45, no.3, 219-231.
- G.A. Watson (1986) Methods for best approximation and regression problems. Rep. NA/96, Dept. of Math. Sci., Univ. of Dundee, DD1 4hn, Scotland, UK.
- G.A. Watson (1987) Data fitting by sums of exponentials using the L_1-norm. Inter. Series of Numer. Math., 81, 246-261.
- J.F. Wellington, S.C. Narula (1981) Variable selection in multiple linear regression using the minimum sum of weighted absolute errors criterion. Comm. Stat., B10, 641-648.
- J.F. Wellington, S.C. Narula (1984) An algorithm for regression quantiles. Comm. Stat., B13(5), 683-704.
- A.H. Welsh (1987) Kernel estimates of the sparsity function. In Y. Dodge (ed.) Statistical data analysis based on the L_1-norm and related methods. NHPC. 369-378.
- G.O Wesolowsky (1981) A new descent algorithm for least absolute value regression problem. Comm. Stat., B10, 479-491.
- G.O. Wesolowsky, R.F. Love (1971) The optimal location of new facilities using rectangular distances. Oper. Res., Jan-Feb.
- G.O. Wesolowsky, R.F. Love (1972) A nonlinear approximation method for solving a generalized rectangular distance Weber problem. Manag. Sci., 18, 56-63.
- H.C. Wilson (1978) Least squares versus minimum absolute deviations estimation in linear models. Dec. Sci., 322-335.
- H.G. Wilson (1979) Upgrading transport costing methodology. Transportation J., 18, 49-55.
- C.S. Withers (1986) The bias and skewness of L_1-estimates in regression. CSDA, 5, 301-303.
- J.M. Wolfe (1979) On the convergence of an algorithm for a discrete L_p-approximation. Numer. Math., 32, 439-459.
- R.S. Womersley (1986) Censored discrete linear L_1-approximation. SIAM J. Sci. Comput., 7, no.1, 105-122.
- Y. Wu (1988) Strong consistency and exponential rate of the minimum L_1-norm estimates in linear regression models. CSDA 6(3), 285-296.
- Wulff (1983) Numerische verfahren zur linearen orthogonalen Lp-regression. Diplomarbeit, Universitat Oldenburg.
- J.D. Young (1971) Smoothing data with tolerances by use of linear programming. J. Inst. Math. Appl., 8, 69-79.
- R. Zeckhauser, M. Thompson (1970) Linear regression with non-normal error terms. Rev. Econ. Stat., 52, 280-286.