In the Name of God,
the Merciful, the Compassionate

They said: Glory be to thee ! we have no knowledge but that which Thou has taught us; surely Thou art the Knowing, the Wise.

(Koran; Sura: The Cow; Verse: XXXII)

DISCRETE AND CONTINUOUS L_1 NORM REGRESSIONS

PROPOSITION OF DISCRETE APPROXIMATION ALGORITHMS AND CONTINUOUS SMOOTHING OF CONCENTRATION SURFACE

BY

BIJAN BIDABAD

THESIS

SUBMITTED TO THE FACULTY OF ECONOMICS IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Ph.D

IN

ECONOMICS

ISLAMIC AZAD UNIVERSITY

TEHRAN, IRAN

WITH COOPERATION OF THE UNIVERSITY OF NEUCHATEL OF SWITZERLAND

APPROVED AND EVALUATED AS *Excellent*

M.J. MOJARRAD, PROFESSOR OF ECONOMICS ..... (SUPERVISOR)

A. MONAJJEMI, PROFESSOR OF STATISTICS ..... (ADVISOR)

Y. DODGE, PROFESSOR OF STATISTICS AND OPERATIONS RESEARCH, UNIVERSITY OF NEUCHATEL, SWITZERLAND ..... (ADVISOR)

A. KOMIJANI, PROFESSOR OF ECONOMICS ..... (MEMBER OF COMMITTEE)

M.R. SHARIFAZADEH, PROFESSOR OF ECONOMICS ..... (DEAN OF THE FACULTY OF ECONOMICS)

COMMITTEE IN CHARGE

DECEMBER 1989

" Other uses and qualifications of the method [of plural medians] would doubtless come to light if experts would turn their attention to the subject .... An unquestioning acquiescence in the orthodox method of least squares is [not justified] .... Orthodoxy should not put tradition before inspiration ".

F.Y. Edgeworth 1888

(Farebrother 1987)

A B S T R A C T

In this thesis a rather complete survey of literature on data analysis based on the L_1 norm and related methods is given which is hoped to fill the empty place of such a review in the area of L_1 norm estimation. Different descent algorithms for L_1 norm estimation of regression parameters for simple and multiple models are proposed which benefit from accuracy and efficiency among existing algorithms. With a digression to concentration surface and developing continuous L_1 norm smoothing, estimation of proposed functional forms of Lorenz curve is considered. In this respect continuous probability density functions information which is of continuous type data are used.

This thesis has also been translated to Farsi language

A C K N O W L E D G M E N T S

This acknowledgments will never compensate my debts to persons who helped me during the course of this study. However, I should remember their assistances that without them, I could never be successful to complete the requirements of Ph.D degree in economics.

At the first my thanks go to professor A. Jassebi, the president of Islamic Azad university and professor M.R. Sharifazadeh, the head of faculty of economics who granted me a full four years scholarship to study in this university and other fellowships to participate in national and international conferences. My sincere appreciation is presented to my major advisor professor M.J. Mojarrad for his guidance and encouragement at each stage of this endeavor who allocated many times to ripe this work, and also to my second advisors professors A. Monajjemi and Y. Dodge of Neuchatel University of Switzerland for their valuable comments and suggestions and to the members of the justice committee professor A. Komijani for their constructive critics and comments. I am also indebted to the professors of the faculty of economics who spent a lot of time in relation to my educative problems.

I am grateful to professor Y. Dodge of Neuchatel University and organizer of the "First International Conference of Statistical Data Analysis Based on the L_1 Norm and Related Methods" for his hospitality during my participation in the conference and allocating time to present my papers and professor K. Sadighi of Shiraz University who invited me to have two lectures in the "Third Seminar of Mathematical Analysis" and also professor M. McAleer of Australian National University for devoting time to present my paper in "Australasian Meeting of Econometric Society" and to professors H. Imai and M. Iri of Chou University for inviting me to "13th International Symposium on Mathematical Programming".

My thanks also go to professor C.R. Rao for his constructive critics, guidance and encouragement; and to professors F. Ahrabi, P. Bloomfield, R.W. Farebrother, R. Koenker, S.C. Narula, M.R. Osborne, S. Portnoy, H. Spath, V. Sposito who criticized the two manuscripts of my papers presented at the Neuchatel conference.

Finally I am thankful to the following professors from different universities all over the world who sent me their books, articles and papers related to my work, J. Antoch, P. Bloomfield, R.H. Bartels, H. Brunk, G. Bassett, D. Bradu, M. Csorgo, P.H. Calamai, S.S. Chiu, A.R. Conn, T.M. Cavalier, Y. Dodge, J. Dupacova, R. Dutter, J.R. Eriksson, A.K.md. Ehsanes Saleh, H. Ekblom, R.M. Freund, R. Fourer, R. Fletcher, V.V. Fedorov, R.W. Farebrother, R. Gonin, J. Galpin, D. Goldfarb, S.A. Gustafson, E. Gorecki, J.L. Goffin, C.C. Gonzaga, W. Heiser, J. Hald, M.L. Hart, W. Hardle, T. Hettmansperger, M. Iri, H. Imai, J.P. Ignizio, K. Jittorntrum, K. Jafarpour, J. Jureckova, L. Kaufman, K.O. Kortanek, N. Kaergard, M. Kojima, J.S. Marron, A. Money, A. Marazzi, N. Megiddo, C. McConnell, S Mizuno, J. Mitchell, J. McKean, J.A. Menendez, S.C. Narula, M.R. Osborne, U. Peters, P. Pilibossian, C. Roos, P. Rousseeuw, F. Reza, J. Renegar, R. Rossi, S.M. Stigler, D.L. Souvaine, G. Stangenhaus, V. Sposito, Su. Chun, R.I. Shrager, A.F. Siegel, H.D. Sherali, H. Spath, A.H. Seheult, E. Seneta, S.J. Sheather, A. Schrijver, L. Seiford, W. Steiger, I.M. Steele, R.M. Schrader, D.S. Tracy, T. Taguchi, E. Trauwaert, J.A. Tomlin, K. Turner, M.J. Todd, S.H.C. du Toit, J.P. Vial, R.J. Vanderbai, A.H. Welsh, G. Wesolowsky, C.S. Withers, G.A. Watson, R.S. Womersley, U. Zimmerman.

I express my sincere and wholehearted thanks to my dear parents Mr. Mohammad Ali Bidabad and Mrs. Maleehe Rabeti who despite their numerous problems and difficulties of life offered the best continuous devotion to my education. And I truly wish that God will bless them for my family. I also deeply thank my dear wife Mrs. Nadereh Rastin for having provided proper conditions at home and accepted everyday life

obligations in the very harsh situation, so that I could keep up to my Ph.D researches. Meanwhile I thank my dear older brother professor Behrouz Bidabad for his worthy consultations. At the same time I thank my dear younger brother and sister Pirouz Bidabad and Faranak Bidabad who were nice to me and regret that because of the heavy work of my researches I had very little time to help them with their difficulties.

It is proper to thank Mr. M. Roghani Zanjani the director, Mr. M.J. Eslami the deputy director of regional affairs and Mr. M.H. Fouladi the chief of regional planning bureau of plan and budget organization for having put the computer facilities at my disposal. I am grateful to professor M. Nayerivand the educational and student affairs deputy, Mr. M.H. Sadehinia the financial and administrative deputy, Mr. M. Familly responsible for the administrative office, Mr. H. Ansari Amoli responsible for the accountancy, Mr. H. Akhlaghiyan personnel of the administrative office, Mrs. F. Bannazadeh responsible for the computer center, Miss. L. Khodabakhsh Moghadam chief librarian, Mr. A. Rezai, Miss. F. Sharilli, Miss. A. Vaseghi library staffs, Mr. B. Gandomkar, Mr. P. Taheri, Mr. K. Maleki responsibles for technical services, and Miss. N. Sarshar, Miss. N. Abreeshamkar, R. Amooi secretariats and Mr. Gh.A. Amani, Mr. Gh. Ghayeghi working staffs and all other staffs of the Advanced Research Center (Ph.D) of the Islamic Azad University every one of them in a way have been deeply concerned with my researches and helped me a lot. I finally thank Miss. R. Naghavi Jooshari and Mr. F. Farashi Oghani who typed the Farsi version of this thesis.

Bijan Bidabad

T A B L E O F C O N T E N T S

CHAPTER I

GENERAL INTRODUCTION

1. Introduction

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 (see, chapter II) 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.

Boscovich (1757) was the first who proposed the least absolute errors estimator. Laplace (1779) solved the problem for the simple one parameter model. Edgeworth (1888) introduced an algorithm for this minimization problem. Karst (1958), Sharpe (1971), Rao and Srinivasan (1972), Appa and Smith (1973), Sposito and Smith (1976), Josvanger and Sposito (1983) developed iterative algorithms for least absolute estimation of simple linear regression parameters.

In the case of multiple linear regression, Rhodes (1930) and Singleton (1940) extended the proposal of Edgeworth. However, these procedures become extremely unwieldy as the number of variables increases. Charnes and Cooper and Ferguson (1955) pointed out that least absolute errors regression is essentially a linear programming problem and Wagner (1959) formulated it as a linear programming problem. The modified simplex method of Barrodale and Young (1966), descent method of Usow (1967) and interval programming procedure of Robers and Ben-Israel (1969) followed in quick succession. Abdelmalek (1971) obtained this estimate by solving a sequence of nonlinear problems, whereas Schlossmacher (1973) found it by iterative reweighted least squares procedure. Barrodale and Roberts (1973) introduced a very efficient special purpose algorithm. Their algorithm integrates several simplex iterations into one. Armstrong and Frome and Kung (1979) used LU (Lower-Upper triangular

matrices) decomposition to introduce a revised simplex algorithm for this problem. Bloomfield and Steiger (1980) and Wesolowsky (1981) have proposed two descent algorithms.

There are few comparative studies which examine the performance of different algorithms with respect to different characterization of the problem as perfect multicolinearity, various specification of error distributions and computational efficiency of required space and time for computer execution. However, the results show that each algorithm has a good performance in certain conditions and there is no unique algorithm which has all the best properties. It can be concluded that there exists a computational gap between least squares and least absolute estimators and attempt to eliminate this gap is worthy.

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 ( ), 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 force 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 computation 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. One of the goals of this research is to give some improvements on these problems.

Since the algebraic closed form of the L_1 norm estimator has not been derived yet, it is 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, four algorithms are proposed 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.

However, we do not stop the work on the discrete L_1 norm estimation and try to generalize the above elaborations to the continuous case of the L_1 norm estimation. Although, the

continuous case of the L_1 norm estimation has not been ever used in econometrics, we apply this technique to solve one of the famous problem of econometrics which is Lorenz curve estimation.

Distribution of personal income is often portrayed on a Lorenz curve. The Lorenz curve which is a simple case of concentration surface is the most commonly used method to describe the income distribution. Income distribution is one of the earliest economic phenomena that economists have tried to describe it by mathematical modelling. Cramer (1973) in his book discusses this problem under the heading of "the distribution of personal income ... an unsolved problem". By clarifying the skewness of income distribution , he argues that this characteristic is persistently exhibited for different populations and in different times. But despite of this known condition, there is not a generally acceptable model and this subject is still controversial. It is discussed that the log-normal and Pareto distributions are two rival functions to explain this phenomenon. Both these distributions may be fitted to income data of any population to derive the corresponding distribution parameters and use these parameters as measures of inequality. However, Cramer (1973) concludes that the Pareto distribution fits the higher income earners, but the log-normal is a good distribution for full income range.

Pareto distribution does not always have finite variance. Certain values of Pareto parameter make the variance of distribution infinite. As we discussed before, any observed distribution of income has a finite variance. We need not worry about the sample, the question remains whether, the sample comes from a distribution with finite or infinite variance. Mandelbrot (1960,61) picks this point of Pareto distribution up and tries to justify the meaning of infinite variance by the asymptotic shape of the upper tails of observed income distributions. That is, the interdecile range to interquartile range is sufficiently large that we

may say the variance is infinite. However, he tries to find a stable distribution (see, Kendall and Stuart (1977) for properties of this type of distributions) with finite mean, infinite variance and maximum skewness toward the right. His approach to emphasize on the stable distributions is the nice properties of this type of distributions. That is, the probability distribution of linear combination of n independent random variables with a specified probability distribution be as the same type of distribution of the random variables. Total income is the sum of all incomes from various sources and since all the income gradients have the same type of distribution as total income, stable distributions seem to be suitable for this case. However, Mandelbrot (1960,61) found that the Pareto-Levy distribution meets his three requirements. Unfortunately, Pareto-Levy distribution function and also its density function can not be written in simple analytic form and should be presented by Laplace transform or characteristic function.

There are also other approaches to formulate the distribution of income which finally lead to Pareto distribution. In this context, we may refer to Champernowne (1953) model which uses Markov stochastic process (see, Bidabad and Bidabad (1989c,d) for some clarifications on this process). Simon (1955) and Steindle (1965) consider a discrete random variable having a steady-state distribution with Pareto tail. Herein before, we emphasized on the distribution of personal income, it is also true for other size distributions which may be a stock variable as assets and wealth or a flow variable as output and sales for any microeconomic unit as individual or firm (see, Cramer (1973)).

Concentration surface technique is a method to analyze the skew distributions and in this respect it links to Pareto and log-normal distributions. A simple concentration surface is called Lorenz curve which is specially devised to explain income distribution inequality. As it is discussed by Kendall

and Stuart (1977), there is a relation between the area under the Lorenz curve and the corresponding probability distribution function of the statistical population. That is, when the probability distribution function is known, we may find the corresponding Gini coefficient as the measure of inequality. But as discussed by Hagen (1975) and Bidabad and Bidabad (1989a,b), two different Lorenz curves with different curvature can yield the same Gini ratio. Therefore, it is possible to happen some redistribution of income that does not change the Gini ratios. This point weaken the comparability of two different Gini coefficients as measures of inequality. Thus, we should take back and use the functional forms of the Lorenz curves to interpret the inequality of income distribution and its changes.

The estimation of the functional form of the Lorenz curve is also confronted with some difficulties. For this estimation, we should define an appropriate functional form which can accept different curvatures. This point is discussed by Bidabad and Bidabad (1989a,b). In these articles, autoregressive nature of the errors in the necessary cumulative data for estimating the Lorenz curve is clarified too. There is another problem that is, to create the necessary data set for estimating the corresponding parameters of the Lorenz curve, a large amount of computation on raw sample income data is inevitable. Obviously, these problems despite of their computational difficulties, make the significance of the estimated parameters poor. To avoid this, we try to estimate the functional form of the Lorenz curve by using continuous data. In this research we use the probability density function of income of the population to derive the estimated values of the Lorenz function parameters. The continuous L_1 norm smoothing method which will be developed for estimating the regression parameters is used to solve this problem. However, we concentrate on two rival probability density functions of Pareto and log-normal. Since, the former is simply integrable, there is no general

problem to derive the corresponding Lorenz function and the function is uniquely derived. But in the latter case, the log-normal density function which has better performance for full income range than Pareto distribution, is not integrable and we can not determine its related Lorenz function. In this regard we should solve the problem by defining a general Lorenz curve functional form and applying the L_1 norm smoothing to estimate the corresponding parameters.

The main parts of this research are based on the following scheme. This chapter proceeds with giving an introduction on the normed space and their properties as the starting point. The purpose of this section is to familiarize the reader with characteristics and notations of the norms. The relationship of norm and regression is the subject of the next section. In this section with reference to L_1 norm regression as a special case of L_p norm regression, its position in norm regressions is considered from a statistical viewpoint. In the next section the important properties of L_1 norm estimation are discussed. Mathematical properties of least absolute errors regression such as invariance property, convexity, optimality condition, uniqueness and non-uniqueness cases of solution and so forth are noted.

Chapter II is devoted to the survey of literature on L_1 norm data analysis and related methods with special attention on descent algorithms. Since there is no general review on this topics, it is hoped that this chapter becomes an original source for others who are interested in L_1 norm. With referring to more than 600 original papers and books the theoretical extension of the L_1 norm in various fields of statistics and econometrics such as statistical inference, asymptotic and small sample properties, sampling distribution, simultaneous system of equations and application of the L_1 norm in sciences with special emphasis on economics, since Galilei (1632) up to now will be surveyed.

In chapter III new descent algorithms are proposed.

In the previous studies of Bidabad (1987a,b,88a,b) preliminary algorithms for L_1 norm regression have been introduced. We shall try to expand them and apply an special type of steepest descent method which primarily stated in the above papers to find the least absolute errors estimates of linear regression parameters. Namely, 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 L_1 norm objective function. In this section the new convergence theorems related to the proposed algorithms are proved and their properties are discussed. Step by step clarification of the proposed algorithms is one of the subjects of this chapter.

Comparison of the proposed algorithms with some of the existing ones are made in chapter IV. Comparisons are mainly based on various specification of errors distributions. Space and the execution time required for selective algorithms in comparison to the proposed ones are reported. The results are compared and conclusion is made.

In chapter V the problem of continuous L_1 norm estimation is considered. By developing a similar method proposed in chapter III for discrete case, continuous L_1 norm estimation is derived. The method is applied to estimation of the Lorenz curve functional forms which have been proposed by Bidabad and Bidabad (1989a,b) and Gupta (1984).

Chapter VI is devoted to summary, conclusions and suggestions for further research. Appendix of computer programs and bibliography are the last segments of this research.

2. Normed space

Through the entire of this text we should use some conventional notations and definitions. So, it will be suitable to have a look at the notations, definitions and properties of normed space. Although, the normed space is a long course in mathematics, this section is restricted to just a few pages as remembrance.

Given two sets X and Y, consider an ordered pair (x,y) where $x \in X$ and $y \in Y$. The collection of all these ordered pairs is called the Cartesian product of X and Y and is denoted by XxY. The common two-dimensional plane can be written as RxR or R^2, where R is the real numbers set. Given the sets $X_1, \ldots, X_n$ we may consider the Cartesian product as:

$$\underset{i=1}{\overset{n}{\times}} X_i \qquad (1)$$

If we write $x=(x_1, \ldots, x_n) \in \underset{i=1}{\overset{n}{\times}} X_i$, then x_i is called the i^{th} coordinate of the point x.

Given any two arbitrary members $u=(u_1, \ldots, u_n)$ and $w=(w_1, \ldots, w_n)$ of R^n, we may define a rule for multiplication of u and w as follows:

$$u.w = \sum_{i=1}^{n} u_i w_i \qquad (2)$$

u.w is a real number and computation by the above rule is called the inner product in R^n. In general, given an arbitrary linear space U (over a real field), inner product is defined as a real-valued function defined on Cartesian product UxU (denoted by u.w or $\langle u,w \rangle$ where $u \in U$ and $w \in W$), which satisfies the following properties. Let, $u,w,z \in U$ and $\alpha, \beta \in R$ then:

i-1) $u.w = w.u$

i-2) $(\alpha u + \beta w).z = \alpha(u.z) + \beta(w.z)$ (3)

i-3) $u.u \geq 0$ and $u.u = 0$ if and only if $u=0$

A linear space with an inner product defined above is called an inner product space. Clearly, the inner product defined above for R^n satisfies the above axioms and is called as usual (Euclidian) inner product.

Given a point $u=(u_1,\ldots,u_n)$ in R^n, the distance between u and the origin can be computed by,

$$d(u,0) = \sqrt{(\sum_{i=1}^{n} u_i^2)} = \sqrt{(u.u)} = \sqrt{\langle u,u\rangle} \qquad (4)$$

Similarly, given any two arbitrary points $u=(u_1,\ldots,u_n)$ and $w=(w_1,\ldots,w_n)$ in R^n, the distance between u and w can be computed by,

$$d(u,w) = \sqrt{[\sum_{i=1}^{n} (u_i-w_i)^2]} = \sqrt{[(u-w).(u-w)]} = \sqrt{\langle (u-w),(u-w)\rangle} \qquad (5)$$

The distance defined above is called Euclidian distance. This distance satisfies the following properties,

ii-1) $d(u,w)=0$ if and only if $u=w$

ii-2) $d(u,w)+d(w,z) \geq d(u,z)$

ii-3) $d(u,w) \geq 0$ for all u and w (6)

ii-4) $d(u,w)=d(w,u)$

Given an arbitrary set X, if a function d maps XxX into R and if d satisfies the above conditions, then X is called a metric, and d(u,w) is called the distance between two points u and w in XxX. The metric space is denoted by (X,d).

Given a point u in R^n, the Euclidian distance between u and origin, d(u,0) is also called the Euclidian norm of u and is denoted by $||u||$. Therefore d(u,w) can be denoted by $||u-w||$. Given arbitrary points $u,w \in R^n$ and scalar $\alpha \in R$ we may specify the properties of the corresponding norm as,

iii-1) $||u|| \geq 0$ and $||u||=0$ if and only if $u=0$

iii-2) $||u+w|| \leq ||u|| + ||w||$ (7)

iii-3) $||\alpha u|| = |\alpha|.||u||$

Given an arbitrary vector space X (not necessarily R^n), if we define the real-valued function, called a norm, which satisfies the above three properties, then we call X a normed vector space or a normed linear space. Clearly, every normed linear space is a metric space with respect to the induced metric defined by $d(u,w)=||u-w||$; (see, Takayama (1974)).

Given again the two points $u,w \in R^n$ in vector space X, the metric d as a function from XxX into R which is called Minkowski metric is defined as follows,

$$d_p(\mathbf{u},\mathbf{w}) = [\sum_{i=1}^{n} |u_i - w_i|^p]^{1/p} \qquad p \geq 1 \tag{8}$$

Now, the properties of i-1 to i-4 are all satisfied, **X** is a metric space and $d_p(\mathbf{u},\mathbf{w})$ is called Minkowski distance between two points **u** and **w** in **X**x**X**. Triangular inequality ii-2 for any $p \geq 1$ is called Minkowski inequality which always holds. The proof may be found in Vulikh (1976) and Spiegel (1968). Note that Minkowski's inequality fails to hold for $p < 1$. Thus, $d_p(\mathbf{u},\mathbf{w})$ is no more a distance for $p < 1$, since ii-2 does not hold. The derived metric space $(\mathbf{X}, d_p)$ is denoted by L_p space. The L_p space is a linear and also Banach space for $p \geq 1$ (see, Vulikh (1976), Kantorovich and Akilov (1964)). The Euclidian distance is a special case of Minkowski distance where p=2.

Given a point **u** in $\mathbf{R}^n$ again, the Minkowski distance between **u** and origin, $d_p(\mathbf{u},\mathbf{0})$ is called the Minkowski norm of **u** and is denoted by $||\mathbf{u}||_p$. Thus, $d_p(\mathbf{u},\mathbf{w})$ can be denoted as $||\mathbf{u}-\mathbf{w}||_p$. Properties iii-1 to iii-3 hold for $||\mathbf{u}||_p$. However, Minkowski norm or L_p norm can be written as the following expression,

$$||\mathbf{u}||_p = d_p(\mathbf{u},\mathbf{0}) = [\sum_{i=1}^{n} |u_i|^p]^{1/p} \tag{9}$$

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)).

There are important relations between inner product and norm (see, Nikaido (1970), Spiegel (1968)) as follows,

iv-1) Cauchy-Schwarz inequality:

$$|\langle \mathbf{u},\mathbf{w} \rangle|^2 \leq ||\mathbf{u}||^2 . ||\mathbf{w}||^2 \tag{10}$$

The equality holds if and only if $(u_i/w_i)=(u_j/w_j)$ for all $i,j=1,\ldots,n$.

iv-2) Holder's inequality:

$$|\langle \mathbf{u},\mathbf{w} \rangle| \leq ||\mathbf{u}||_p . ||\mathbf{w}||_q \tag{11}$$

where $(1/p)+(1/q)=1$, $p,q>1$. The equality holds if and only if $(|u_i|^{p-1}/|w_i|)=(|u_j|^{p-1}/|w_j|)$ for all $i,j=1,\ldots,n$. When p=q=2, it reduces to Cauchy-Schwarz inequality iv-1.

iv-3) Minkowski inequality:

$$||u+w||_p \leq ||u||_p + ||w||_p \qquad (12)$$

where p>1. The equality holds if and only if $(|u_i|/|w_i|) = (|u_j|/|w_j|)$ for all i,j=1,..,n. When p=2 it reduces to triangular inequality iii-2.

For more relations and inequalities on norm see, Hromadka II et al (1987). With the above definition of norm in the next section we proceed by application of Minkowski or L_p norm in regression analysis.

3. L_p norm and regression analysis

The following overdetermined system of equations is given,

$$y = X\beta + u \qquad (13)$$

where y is a nx1 vector of dependent variables, X, a nxm matrix of independent or explanatory variables with n>m, β, a mx1 vector of unknown parameters and u is a nx1 vector of random errors. The problem is to find the unknown vector β such that the estimated value of 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 $||u||_p$ is minimized to find the β vector,

$$\min_{\beta} S = \min_{\beta} ||u||_p = \min_{\beta} ||y-X\beta||_p =$$

$$\min_{\beta} \left[\sum_{i=1}^{n} |y_i - x_i\beta|^p\right]^{1/p} = \min_{\beta} \left[\sum_{i=1}^{n} |y_i - \sum_{j=1}^{m} \beta_j x_{ij}|^p\right]^{1/p} \Longrightarrow$$

$$\min_{\beta} \sum_{i=1}^{n} |y_i - x_i\beta|^p = \min_{\beta} \sum_{i=1}^{n} |y_i - \sum_{j=1}^{m} \beta_j x_{ij}|^p \qquad (14)$$

where y_i is the i^{th} element of y and x_i is the i^{th} row of the matrix X. Any value of $p\in[1,\infty]$ may be used to find β in (14) (see, Money et al (1978a), Rice (1983)), 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 (13), $\mathbf{X}$ is nonstochastic and $E(\mathbf{u})=\mathbf{0}$, $E(\mathbf{uu}^T)=\sigma^2\mathbf{I}$, and $\mathbf{u}$ distributed with $f(u)=h.\exp(-k|u|^p)$, where h and k are constants and $p\in[1,\infty]$; then the "best" $\boldsymbol{\beta}$ with maximum likelihood properties is a vector which comes from minimization of (14).

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 (14) 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

$$f(u)=(1/2\theta).\exp(-|u|/\theta) \qquad (15)$$

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

4. 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.

4.1 Invariance property

An estimator $\beta\hat{}(y,X)$ of population parameter β is invariant if,

$$\beta\hat{}(\theta y,X) = \theta\beta\hat{}(y,X), \qquad \theta\in[0,\infty) \tag{16}$$

Gentle and Sposito (1976), Koenker and Bassett (1978) have proved that the L_p norm estimator of β 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 $\beta\hat{}$ solution function.

4.2 Transformation of variables

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

$$\beta\hat{}(y+X\theta,X) = \beta\hat{}(y,X) + \theta \tag{17}$$

If A is a mxm nonsingular matrix, transformation of X to XA premultiplies optimal $\beta\hat{}$ by inverse of A (see, Taylor (1974), Koenker and Bassett (1978), Bassett and Koenker (1978)).

$$\beta\hat{}(y,XA) = A^{-1}\beta\hat{}(y,X) \tag{18}$$

4.3 convexity of the objective function

To show the convexity of S in (14), Suppose m=1; the objective function (14) reduces to

$$S = \sum_{i=1}^{n} |y_i - \beta_1 x_{i1}| = \sum_{i=1}^{n} S_i \tag{19}$$

where $S_i=|y_i-\beta_1 x_{i1}|$. If we plot S_i as a function of β_1 then we will have a broken line in $Sx\beta_1$ 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 $-|x_{i1}|$ and $|x_{i1}|$ 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 (14) when m=2,

$$S = \sum_{i=1}^{n} |y_i - \beta_1 x_{i1} - \beta_2 x_{i2}| = \sum_{i=1}^{n} S_i \qquad (20)$$

Where $S_i = |y_i - \beta_1 x_{i1} - \beta_2 x_{i2}|$. 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.

4.4 Zero residuals in optimal solution

L_1 norm regression hyperplane always passes through r of the 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 (14). 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.

$$y_i - \sum_{j=1}^{m} \beta_j x_{ij} = 0 \qquad i \in \{1, \ldots, n\} \qquad (21)$$

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)).

4.5 Optimality condition

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

$$\sum_{i \in A} \alpha_i x_i + \sum_{i \in I} \operatorname{sgn}(y_i-x_i\beta^*)x_i = 0 \qquad (22)$$

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

$$|n^+ - n^-| \leq m \qquad (23)$$

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

4.6 Unique and non-unique solutions

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. In 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)).

4.7 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).

CHAPTER II

SURVEY OF LITERATURE ON THE L_1 NORM DATA ANALYSIS AND RELATED METHODS

1. Introduction

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 chapter. 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. Though some topics gathered in this chapter are less related to the subject of the following chapters, because of lack of a general survey on this area, we did not delete them and hope to be more helpful for others who are interested in this field.

The structure of this chapter is as follows. The chronology and historical development of L_1 norm for period of 1632 to 1928 is subject of the forthcoming section. The main contributions since Galilei are mentioned. The devised algorithms in this period are not so well organized to handle the general L_1 norm regression. In the next section we will discuss the computational algorithms which belong to the times of after 1928 in different manner. This period along 1928 to now is the time of sophisticated algorithms which are classified into three categories of direct descent, simplex type and other algorithms. Since the emphasis of the next chapters are based on new proposition of descent methods, the section 3.1 is rather more expository. This section tries to summarize the important descent methods in well defined steps. Algorithms which are based on simplex method of linear programming are discussed in section 3.2. The next section is devoted to review the algorithms which do not belong to the above two categories. Initial value problem for starting different algorithms are discussed in section 3.4. Section 3.5 refers to available computer programs and packages. Section 3.6 reviews the comparative studies of previous researchers on relative efficiency of existing algorithms.

Computational methods for nonlinear L_1 norm regression and references to the case of L_p norm are the subject of the next sections. In section 4 application of L_1 norm to simultaneous equations system is under consideration. In the next part statistical aspects of L_1 norm are briefly noted and a glance on L_1 norm sampling distribution, statistical inference, multivariate statistics, nonparametric density estimation and robust statistics will be made. Application of L_1 norm with more emphasis on economics is the subject of section 6. In section 7, other variants of L_1 norm are briefly cited.

2. 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,

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

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 chapter 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 $\bar{x}_1$ and $\bar{y}$ be the means of x_{i1} and y_i then,

$$\beta_0 = \bar{y} - \beta_1 \bar{x}_1 \qquad (2)$$

Value of β_1 is found by,

$$\min_{\beta_1}: S = \sum_{i=1}^{n} |y_i^{\sim} - \beta_1 x_{i1}^{\sim}| \qquad (3)$$

where $x_{i1}^{\sim}$ and $y_i^{\sim}$ are deviations of x_{i1} and y_i from their means respectively. By rearranging the observations in descending order of $y_i^{\sim}/x_{i1}^{\sim}$ values, Laplace notes that S is infinite when β_1 is infinite and decreases as β_1 is reduced. β_1 reaches the critical value $y_t^{\sim}/x_{t1}^{\sim}$ when it again begins to increase. This critical value of β_1 is determined when,

$$\sum_{i=1}^{t-1} |x_{i1}^{\sim}| < \tfrac{1}{2} \sum_{i=1}^{n} |x_{i1}^{\sim}| \leq \sum_{i=1}^{t} |x_{i1}^{\sim}| \qquad (4)$$

This procedure to find β_1 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 (4) 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.

3. Computational algorithms

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,

i) Direct descent algorithms

ii) Simplex type algorithms
iii) Other algorithms
which will be discussed in the following sections sequentially.

3.1 Direct descent algorithms

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;

Step 1) Select m-1 equations arbitrarily.
Step 2) Solve these equations for m-1 parameters.
Step 3) Estimate the remaining m^{th} parameter by the Laplace's method.
Step 4) Recognize the resulting equation in step 3 and add it to the m-1 equations in step 1.
Step 5) If the set of m equations has recurred m times then stop; otherwise discard the oldest equation and go to step 2.

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. Although the paper was not clearly written, the following steps summarize his algorithm.

Step 1) Select a point $\beta_j^{(0)}$, $j=1,\ldots,m$ arbitrarily.

Step 2) Determine the gradient,
$g_j^{(0)}=-\sum_{i=1}^{n} \operatorname{sgn}(u_i \mid \beta_j^{(0)}, j=1,\ldots,m)x_{ij}$.

Step 3) Compute, $w_i^{(0)}=\sum_{j=1}^{m} x_{ij}g_j^{(0)}$, $z_i^{(0)}=y_i-\sum_{j=1}^{m} x_{ij}\beta_j^{(0)}$.

Step 4) Determine the value of $t^{(0)}$ as weighted median of $\sum_{i=1}^{n} |w_i t-z_i|$ by Laplace's method. The t value is the length of movement along direction of the gradient.

Step 5) Compute $\beta_j^{(1)}=\beta_j^{(0)}+g_j^{(0)}t^{(0)}$.

Step 6) Test the optimality condition. Singleton gives a condition to stop, but it is not quite clear. In this step any other criterion relevant to L_1 norm function may be displaced.

Step 7) This step is not well defined by Singleton. In this phase he tries to choose the best gradient among usable gradients. Without this step algorithm is still operational, because, the steps are all standards of Cauchy steepest descent method and instead of choosing the best gradient we can proceed by going to step 2.

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). Given the function $f(x)$ defined on a finite subset $X=\{x_1,\ldots,x_n\}$ of an interval on the real line and also linearly independent continuous functions $\Phi_1(x),\ldots,\Phi_m(x)$ where $m<n$; consider the "polynomial", $L(\beta,x)=\sum_{j=1}^{m}\beta_j\Phi_j(x)$. In Usow (1967b) the following function is to be minimized.

$$\min_{\beta}: S(\beta)=\sum_{i=1}^{n}|L(\beta,x_i)-f(x_i)|=\sum_{i=1}^{n}\left|\sum_{j=1}^{m}\beta_j\Phi_j(x_i)-f(x_i)\right| \quad (5)$$

where β is vector of size m. The above form is general; if we let $f(x_i)\equiv y_i$ and $\Phi_j(x_i)\equiv x_{ij}$ the function S becomes the standard linear regression objective function. Now let the set K be:

$$K=\{(\beta,d)\,|\,(\beta,d)\in E^{m+1},S\leq d\} \quad (6)$$

Then K is a convex polytope, the vertices of which occur only when $L(\beta,x)-f(x)$ is zero at m or more points of X. The Usow's algorithm is to descend on K 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 (β^*,d^*) is reached. To clarify the algorithm assume that we are at the vertex (β^k,d^k) on K and the polynomial $L(\beta^k,x)$ interpolates m points of X denoted by $U^k=\{u_1{}^k,\ldots,u_m{}^k\}$. Thus,

$$L(F^k,x)=\sum_{i=1}^{m}f(u_i{}^k)\pi_i(x) \quad (7)$$

where,

$$F^k=(f(u_1{}^k),\ldots,f(u_m{}^k)) \quad (8)$$

$$\pi_i(x)=\sum_{j=1}^{m}a_{ji}\Phi_j(x) \quad i=1,\ldots,m \quad (9)$$

The m coefficients a_j^i are calculated as follows. Form the matrix $(\Phi_j(u_i^k))$ for i,j=1,...,m. Let $\pi(x)$ and $\Phi(x)$ be two mx1 vectors for any $x \in X$ whose elements are $\pi_i(x)$ and $\Phi_i(x)$ for i=1,...,m respectively. Hence, we can derive,

$$\pi(x)=[(\Phi_j(u_i^k))^T]^{-1}\Phi(x) \qquad (10)$$

a_j^i's are the elements of the i^{th} row of the matrix $[(\Phi_j(u_i^k))^T]^{-1}$, where the superscript T denotes transposition. Let e_i be a zero vector of size m where its i^{th} element is equal to one. Then if for some δ, $S(F^k-\delta e_i)<S(F^k)$, there is a T_j such that $T_j\delta>0$ and $S(F^k-T_j e_i)<S(F^k)$. Also,

$$S(F^k - T_j e_i) = \min_t \{S(F^k - t e_i)\} \qquad (11)$$

and $((F^k-T_j e_i),S(F^k-T_j e_i))$ is a vertex. On the other hand, a point $u_i^k \in U^k$ may be replaced by a point $u_i^{k+1} \in \{X-U^k\}$ such that the polynomial $L(\beta^{ki},x)$ interpolating $U_i^k=\{u_1^k,\ldots,u_i^{k+1},\ldots,u_m^k\}$ and $S(\beta^{ki})<S(\beta^k)$. $S(\beta^{Ki})$ is the minimum of all norms obtained if u_i^k were replaced by the different points of the set $\{X-U^k\}$, as indicated by the above relation.

In going from vertex $(\beta^k,S(\beta^k))$ to $(\beta^{ki},S(\beta^{ki}))$, one or more vertices on K might have been by-passed. The nearest vertex to $(F^k,S(F^k))$ and below it on the edge parallel to the i^{th} parameter space coordinate axis, say the vertex $((F^k-t_r e_i),S(F^k-t_r e_i))$, is obtained from:

$$|t_r| = \min_s \{\frac{|L(F^k,x_s)-f(x_s)|}{|\pi_i(x_s)|}\} \qquad x_s \in \{X-U^k\} \qquad (12)$$

The point x_r is characterized by,

$$\mathrm{sgn}[L(F^k,x_l) - f(x_l)] = \mathrm{sgn}[L(F^{ki},x_l) - f(x_l)],$$
$$x_l \in \{X-U^k-x_r\} \qquad (13)$$

Now, if there is not any δ such that $S(F^k-\delta e_i)<S(F^k)$, then $S(F^k)$ could not be reduced by moving on K along the edge parallel to the i^{th} parameter space coordinate axis and u_i^k should not be replaced by another point from the set $\{X-U^k\}$. This iteration is repeated m times, once for each point in U^k in succession. The whole cycle is then repeated a finite number of times until the solution (β^*,d^*) is reached (see

also, Abdelmalek (1974)).

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. He argues that for the simple model with the objective function,

$$S = \sum_{i=1}^{n} |y_i - (\beta_0 + \beta_1 x_{i1})| \tag{14}$$

it must be possible to assign half of the points above and half below the regression line. With any given value of $\beta\tilde{}_1$, we can derive β_0 as the median of $\beta_{0i}=y_i-\beta\tilde{}_1 x_{i1}$. Now segregate the points, such that:

$$S=\sum_{i_{above}}^{n_{above}} [y_i-(\beta_0+\beta_1 x_{i1})] - \sum_{i_{below}}^{n_{below}} [y_i-(\beta_0+\beta_1 x_{i1})] \tag{15}$$

Rearranging the terms and note that $n_{above}=n_{below}$ gives,

$$S = k_1 + k_2\beta_1 \tag{16}$$

where,

$$k_1 = \sum_{i_{above}}^{n_{above}} y_i - \sum_{i_{below}}^{n_{below}} y_i \tag{17}$$

$$k_2 = \sum_{i_{above}}^{n_{above}} -x_{i1} + \sum_{i_{below}}^{n_{below}} x_{i1} \tag{18}$$

The overall solution strategy may now be formulated as follows. Any value of β_1 may be chosen at the outset. By computing β_{0i}, segregate the points above and below the line. By using equations (16) through (18) the corresponding segment of the S and β_1 relation is calculated. Sign of k_2 indicates the appropriate direction for the next iteration. If k_2 is positive, only smaller values of β_1 need be considered. If k_2 is negative, only larger values of β_1 need

be considered. If k_2 is zero, the initial value of β_1 is a solution.

When a border line (median of β_{0i}) and the direction of search is determined, the nearest intersection of the present border line with another should be found. The calculations can be reduced by comparing slopes (x_{i1} values) to determine whether or not two lines intersect in the region of interest (thus avoiding needless division operations). The new values k_1 and k_2 must be computed. If this alteration causes k_2 to change sign, the solution has been obtained and algorithm stops.

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. The algorithm may be reduced as follows.

Step 0) Select an arbitrary point β.

Step 1) i) Identify the index set,

$I = \{i_1,\ldots,i_m\} = \{i \mid x_i^T\beta - y_i = 0\}$.

Let I^{th} row of X be $X_I=[x_{i_1},\ldots,x_{i_m}]$ where $i_j \in I$ and the nullspace $N=N(X_I^T)=\{\delta \mid x_i\delta=0, i\in I\}$. Let the orthogonal projector onto N be denoted by P_N.

ii) Let I^c be the complement of the set I. Compute the vector $h = \sum_{i\in I^c} \mathrm{sgn}(x_i^T\beta - y_i)x_i$.

iii) Compute $p=-P_Nh$ that is the projection of $-h$ onto the nullspace of X_I, as long as this projection is non zero. If $p\neq 0$, let $g=h$ and go to step 2.

iv) compute $\mathbf{w}$ according to $\mathbf{h}=\mathbf{X}_I\mathbf{w}=\sum_{j=1}^{m} w_j x_{i_j}$, $i_j \in I$.

v) if $|w_j| \leq 1$ for all $j=1,\ldots,m$ stop. In this case β is optimal.

vi) Find $i_{j_0} \in I$ such that $|w_{j_0}|>1$.

vii) Change I to $I-\{i_{j_0}\}$ and make corresponding changes to $\mathbf{X}_I$ and $\mathbf{N}$. Compute $\mathbf{p}=-\mathrm{sgn}(w_{j_0})P_N x_{i_{j_0}}$ and let $\mathbf{g}=\mathbf{h}-\mathrm{sgn}(w_{j_0})x_{i_{j_0}}$.

Step 2) Determine $A=\{\alpha_l \mid l \in I^c, \alpha_l=(x_l{}^T\beta-y_l)/x_l{}^Tp, \alpha_l>0\}$ elements and order them such that $0<\alpha_{l_1}<\alpha_{l_2}<\ldots<\alpha_{l_t}$. Let $\tau=1$.

Step 3) If $p^Tg \geq 2\mathrm{sgn}(x_{l_\tau}{}^T\beta-y_{l_\tau})p^Tx_{l_\tau}$, then go to step 5.

Step 4) Change $\mathbf{g}$ to $g-2\mathrm{sgn}(x_{l_\tau}\beta-y_{l_\tau})x_{l_\tau}$ and τ to $\tau+1$ and go to step 3.

Step 5) Replace β by $\beta+\alpha_{l_\tau}\mathbf{p}$ and go to step 1.

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. Let $\mathbf{U}$ and $\mathbf{v}$ be of sizes pxm and px1 respectively. Consider the function:

$$\Phi(\beta) = \mathbf{h}^T\beta + \Sigma_i |y_i - x_i{}^T\beta| + \Sigma_j \max(0, v_j - u_j{}^T\beta) \qquad (19)$$

Where $y_i-x_i{}^T\beta$ represents the i^{th} element of the residual vector $\mathbf{y}-\mathbf{X}\beta$ and $v_j-u_j{}^T\beta$ represents the j^{th} element of the residual vector $\mathbf{v}-\mathbf{U}\beta$. To minimize Φ with respect to β the following steps should be taken.

Step 0) Start with arbitrary $\beta^{(k)}$.

Step 1) Find $\delta^{(k)}$ so that $\Phi(\beta^{(k)}+\theta\delta^{(k)}) \leq \Phi(\beta^{(k)})$ for all $\theta>0$ small enough.

Step 2) Choose $\theta^{(k)} \geq 0$ to obtain the largest possible decrease in Φ.

Step 3) Let $\beta^{(k+1)}=\beta^{(k)}+\theta^{(k)}\delta^{(k)}$.

When $\mathbf{h}=\mathbf{0}$ and the sum on j is vacuous in the $\Phi(\beta)$ function (19), the resulting simplification of the above steps corresponds precisely to the algorithm proposed by Bartels and Conn and Sinclair (1978). Other related problems can be solved by modification of $\Phi(\beta)$ by a quantity μ to obtain a

parameterized family of piecewise linear functions $\Phi_\mu(\beta)$, and take following steps to find the minimum solution.

Step 0) Set $\mu>0$; select any $\beta=\beta^{(0)}$.

Step 1) Minimize $\Phi_\mu(\beta)$ with respect to β according to the above procedure.

Step 2) Stop if a prescribed terminating condition on β is met; otherwise set $\mu=\mu/10$ and go to step 1.

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.

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.

Denote the $x_1^T,\ldots,x_m^T$ as rows of the independent variables design matrix which correspond to the current set of points $1,\ldots,m$; and x_m^T for replacement. Set,

$$y_i = x_i^T\beta \qquad i=1,\ldots,m-1 \tag{20}$$

Redefine β as,

$$\beta = \beta^0 + t\delta \tag{21}$$

Where β^0 is arbitrary member of the set and δ vector obeys the following system,

$$x_i^T\delta = 0 \qquad i=1,\ldots,m-1 \tag{22}$$

Given this set of points, optimum value of S may be found by minimizing the following expression with respect to the scalar t.

$$\sum_{i=1}^{n} |y_i - x_i^T(\beta^0 + t\delta)| \qquad (23)$$

Rearranging the terms, leads to:

$$\sum_{i=1}^{n} |w_i||r_i - t| \qquad (24)$$

where $w_i = x_i^T\delta$ and $r_i = (y_i - x_i^T\beta^0)/(x_i^T\delta)$. Value of t may be found by Laplace's weighted median method. Bloomfield and Steiger propose a weighted modification of partial quicksort procedure of Chamber (1971) to find the t value efficiently. The data point corresponding to weighted median replaces x_k^T. By using the y_i from (20) and the additional m^{th} equation, value of β^0 is computed. Vector δ is determined up to scalar multiples of (22). The new set of parameters values are computed by (21).

Now a point should be deleted. Bloomfield and Steiger do not give an assured way to identify this point. They propose a heuristic method based on gradients and use the following quantity:

$$\text{rho} = \frac{\left| \sum_{i:r_i<0} w_i - \sum_{i:r_i>0} w_i \right| - \sum_{i:r_i=0} w_i}{\sum_{i=1}^{n} w_i} \qquad (25)$$

Once rho is calculated for each candidate point for deletion, and delete the point for which rho is largest.

To start the algorithm, any set of m rows of X may be chosen, with the appropriate β^0. Add variables stepwise until a fit for β^0 and corresponding set of m points are derived. At each intermediate step, the fit involves k variables where $0 \le k < m$ and corresponding set of k data points with zero residuals. Improving the fit by entering a new variable, thus increasing k to k+1. At each stage, it is the measure rho that determines whether we set up to a larger model or improve the current one without setting up. Suppose at the current stage we are dealing with k variables. If the value of rho is largest at the variable p, which p is not in the set of k variables above, k is increased to k+1 by entering the variable p and improving the fit with k+1 variables. If p

is in the current set of k variables and rho is in largest value, corresponding point to p is deleted and replaced in a manner described before. 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. Given $(x_i, y_i) \in R^{m+1}$, $i=1,\ldots,n$ and $k=n-m$, $X=(x_1,\ldots,x_n)^T$, their algorithm may be described as follows:

Step 1) Renumber the rows of $(X|y)$ such that X_N the bottom m rows of X, is invertible. Solve the k linear equations system $X_N{}^T N=-X_T{}^T$ for N, where X_T denotes the top k rows of X.

Step 2) Let $D=(A|c)$ where $A=(I|N)$ is of size kxn and $c=Ay$.

Step 3) Let $\sigma=(1,\ldots,k)$, $\sigma^c=(k+1,\ldots,n)$.

Step 4) Set $r_{\sigma(i)}=b_i$ for $i=1,\ldots,k$; $r_{\sigma^c(i)}=0$ for $i=k+1,\ldots,n$.

Step 5) Let $I=\{i \mid 1\le i\le k, c_i=0\}$.

Step 6) Do loop for $j=1$ to m:

Let $v_i=D_{i\sigma^c(j)}$ for $j=1,\ldots,k$. Let $M=\{i \mid \operatorname{sgn}(c_i)\neq\operatorname{sgn}(v_i)\}$ and $J=\{1,\ldots,k\}\backslash N\backslash I$. Let

$$\beta_j = \frac{|\Sigma_M |v_i| - \Sigma_J |v_i| | - 1 - \Sigma_I |v_i|}{1+\sum_{i=1}^{k} |v_i|}$$

End loop.

Step 7) Set $S=\{1,\ldots,m\}$.

Step 8) Choose β_q as $\max_s\{\beta_j\}$.

Step 9) If $\beta_q>0$ go to step 11.

Step 10) If $\prod_{i=1}^{k} r_{\sigma(i)}=0$, the problem is degenerate; and stop. Otherwise, solve, $y_{\sigma^c(i)}=\sum_{j=1}^{m} x_{\sigma^c(i)j}\theta_j$ for θ, stop.

Step 11) Let $v_i=D_{i,\sigma^c(q)}$ for $i=1,\ldots,k$.

Step 12) Compute $t^\wedge$=weighted median of $c_1/v_1,\ldots,c_k/v_k,0$ with weights $|v_1|,\ldots,|v_k|,1$.

Step 13) If $t^\wedge=c_p/v_p\neq 0$ go to step 16.

Step 14) Let $S=S\backslash\{q\}$; if $S=\emptyset$ go to step 10.

Step 15) Go to step 8.

Step 16) Divide row p of **D** by $D_{p\sigma c(q)}$.

Step 17) For $i \neq p$ in D, let (row i)=(row i)-(row p)*$D_{i\sigma c(q)}$.

Step 18) Commute for pair $\sigma(p)$ and $\sigma^c(q)$.

Step 19) Let $r_{\sigma(i)}=b_i$ for i=1,...,k; and set $r_{\sigma c(q)}=0$.

Step 20) Go to step 5.

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. 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. Select a set of m points $(x_{j1}{}^I,\ldots,x_{jm}{}^I,y_j{}^I)$. The following system of equations can be solved for a unique set of coefficients.

$$y_j{}^I - \sum_{h=1}^{m} \beta_h x_{jh}{}^I = 0, \quad j=1,\ldots,m \tag{26}$$

An edge is formed of any subset J consisting of m-1 equations. To minimize along an edge, set:

$$\mathbf{Y}_I = \begin{bmatrix} y_1{}^I \\ \vdots \\ y_m{}^I \end{bmatrix}, \quad \mathbf{X}_I = \begin{bmatrix} x_{11}{}^I & \cdots & x_{1m}{}^I \\ \vdots & & \vdots \\ x_{m1}{}^I & \cdots & x_{mm}{}^I \end{bmatrix},$$

$$\mathbf{Y}_J = \begin{bmatrix} y_1{}^J \\ \vdots \\ y_{m-1}{}^J \end{bmatrix}, \quad \mathbf{X}_J = \begin{bmatrix} x_{11}{}^J & \cdots & x_{1m}{}^J \\ \vdots & & \vdots \\ x_{m-1,1}{}^J & \cdots & x_{m-1,m}{}^J \end{bmatrix} \tag{27}$$

Let $\boldsymbol{\beta}_p$ be formed from $\boldsymbol{\beta}$ by deleting β_p and let $\mathbf{x}_p$ be the p^{th} column in $\mathbf{X}_J$ and let $\mathbf{X}_{pJ}$ be formed by removing $\mathbf{x}_p$ from $\mathbf{X}_J$. Then for a given β_p it can be shown that,

$$\boldsymbol{\beta}_p = \mathbf{X}_{pJ}{}^{-1}\mathbf{Y}_J - \beta_p \mathbf{X}_{pJ}{}^{-1}\mathbf{x}_p \tag{28}$$

Let the elements of $\boldsymbol{\beta}_p$ be

$$\beta_q = r_q - s_q\beta_p \quad \text{for } q \neq m \tag{29}$$

Substitution in the L_1 norm objective function gives,

$$\min_{\beta_1} \sum_{i=1}^{n} |x_{ip} - \sum_{q\neq p}^{m} s_q x_{iq}| \left| \frac{y_i - \sum_{q\neq p}^{m} r_q x_{iq}}{x_{ip} - \sum_{q=2}^{m} s_q x_{iq}} - \beta_1 \right| \qquad (30)$$

$$\sum_{q\neq p}^{m} s_q x_{iq} \neq x_{ip}$$

Now the following steps should be taken.

Step 1) Set k=1, l=0, choose initial values for $\beta_1,\ldots,\beta_m$. Least squares values are one possibility. Let $I^{(1)}=\{j_1^{(1)},\ldots,j_m^{(1)}\}$ be a set of m data points chosen in sequence as follows. The point with the smallest squared residual is chosen each time subject to the condition that $X_I^{(1)}$ is nonsingular. Find $\beta_q^{(1)}$ for q=1,...,m, by solving $Y_I^{(1)}=X_I^{(1)}\beta$. Set $I^{(k)}=(j_1^{(k)},\ldots,j_m^{(k)})$; $J=(j_2^{(k)},\ldots,j_m^{(k)})$.

Step 2) Set k=k+1. Obtain β_p for the smallest p from (30) by using the weighted median procedure. Set $\beta_p^{(k)}=\beta_p$ and let i be the index which defines the lower weighted median β_p in (30) for the lowest possible p.

Step 3) a) If $\beta_p^{(k)}-\beta_p^{(k-1)} = 0$ and if $l>m$, go to step 4. Otherwise, set $I=(j_2^{(k-1)},\ldots,j_m^{(k-1)},i)$ and l=l+1, $\beta_q^{(k)}=\beta_q^{(k-1)}$ for all q and go to step 2.

b) If $\beta_p^{(k)}-\beta_p^{(k-1)}\neq 0$, set l=0. Calculate $\beta_q^{(k)}$ for $q\neq p$ from (29). Set $I^{(k)}=(j_2^{(k-1)},\ldots,j_m^{(k-1)},i)$ and go to step 2.

Step 4) Calculate $\beta_q^{(k)}$ for $q\neq p$ from (29); set $\beta^*=\beta^{(k)}$ and stop.

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 $\beta_1^{(k-1)}$ and larger or equal values above it, then recheck the inequalities (4) 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 nonsmooth 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 $\beta_0 x \beta_1$ 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 descent methods for the simple and multiple L_1 norm regressions. These algorithms with many improvements will be discussed in the next

chapter and a new modified algorithm will be proposed.

3.2 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,

$$\begin{aligned} \min_{\beta}: \ & 1_n^T(w+v) \\ \text{s.to}: \ & X\beta+I_n(w-v)=y \\ & w,v\geq 0 \\ & \beta \text{ unrestricted in sign} \end{aligned} \qquad (31)$$

where 1_n is a vector of size nx1 of 1's and 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 related 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 β as a difference of two nonnegative vectors **c** and **d**, their formulation can be stated as follows,

$$\begin{aligned} &\min_{\mathbf{c},\mathbf{d}}: \ 1_n^T(\mathbf{w}+\mathbf{v}) \\ &\text{s.to}: \ \mathbf{X}(\mathbf{c}-\mathbf{d})+I_n(\mathbf{w}-\mathbf{v})=\mathbf{y} \\ &\qquad \mathbf{w},\mathbf{v},\mathbf{c},\mathbf{d}\geq 0 \end{aligned} \tag{32}$$

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 nonbasic 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 according 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 nonpositive (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).

3.3 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 β 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 pseudoinverse or least squares approximation as $\beta\hat{}=(X^TX)^{-1}X^Ty$. Calculate the residuals as $u=|y-X\beta\hat{}|$; where u is nx1 column 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,

$$u=\begin{bmatrix}u\hat{}\\ \hline u\tilde{}\end{bmatrix}, \quad X=\begin{bmatrix}X\hat{}\\ \hline X\tilde{}\end{bmatrix}, \quad y=\begin{bmatrix}y\hat{}\\ \hline y\tilde{}\end{bmatrix}.$$

Solve $y\hat{}=X\hat{}\beta\hat{}$ for the top partitions as $\beta\hat{}=X\hat{}^{-1}y$. 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 ϵ-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)).

3.4 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.

3.5 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.

3.6 Comparison of the algorithms

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. 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.

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

ref.	compared with	m range	n range	time/ performance
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	100-1000	W is faster for larger n and smaller m
SS	BS	4-34	10-50	SS is faster for m near n
AK	S	2	50-500	AK is faster
JS	Ak	2	10-250	JS is faster

n ≡number of observations.
m ≡number of parameters.
BCS≡Bartels,Conn,Sinclair(1978).
BR ≡Barrodale,Roberts(1973,74).
AK ≡Armstrong,Kung(1978).
S ≡Sadovski(1974).
AFK≡Armstrong,Frome,Kung(1979).
A ≡Abdelmalek(1980a,b).
BS ≡Bloomfield,Steiger(1980).
W ≡Wesolowsky(1981).
JS ≡Josvanger,Sposito(1983).
SS ≡Seneta,Steiger(1984).

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.

Table 2. Array storage requirement for selected algorithms.

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=number 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 BIG=1.0E+15
BLAD1	BS	4n+2m(n+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

See table 1 for abbreviations.
Source: Gentle, Narula, Sposito (1987).

In their study the problem consists of uniform (0,1) random values for X and normal (0,3) variates for random error term. The value of dependent variable y computed as sum of the independent variables and error term. Summary of the results is shown in tables 3 and 4 for simple and multiple

regressions respectively. Values in the cells are the CPU time averages of 100 replications and the values in the parentheses are corresponding maximum CPU time of the 100 replications. 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.

Table 3. CPU time for simple model.

n	AK	JS	A	AFK	BS
100	0.021 (0.03)	0.023 (0.04)	0.094 (0.21)	0.034 (0.06)	0.023 (0.04)
500	0.193 (0.38)	0.302 (0.61)	1.434 (3.13)	0.287 (0.49)	0.145 (0.26)
1000	0.544 (1.36)	0.971 (2.16)	4.775 (10.60)	0.784 (1.76)	0.422 (1.19)
5000	1.262 (24.58)	2.837 (48.88)	211.23* (----)	1.614 (31.22)	+ +

See table 1 for abbreviations.
* Average of three runs.
+ Failed to produce correct answers.
Source: Gentle, Narula, Sposito (1987).

Table 4. CPU time for multiple model (m=5,15).

n	m	A	AFK	BS
100	5	0.331 (0.53)	0.149 (0.23)	0.114 (0.17)
100	15	1.976 (2.73)	1.313 (1.70)	0.933 (1.38)
500	5	3.686 (5.47)	1.120 (1.81)	0.829 (1.22)
500	15	17.876 (23.4)	7.808 (10.1)	7.294 (9.13)
1000	5	13.211 (18.3)	2.930 (4.38)	+
1000	15	49.866 (72.7)	17.901 (24.0)	+
5000	5	248.91* (----)	34.311 (51.8)	+
5000	15	687.31* (----)	140.321 (160.1)	+

See table 1 for abbreviations.
* Average of three runs.
+ Failed to produce correct answers.
Source: Gentle, Narula, Sposito (1987).

They concluded that BS program performs quite well on smaller problems but in larger cases, because of accumulated roundoff 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.

By considering all aspects, they concluded that AFK seems to be the best.

3.7 Nonlinear form computational methods

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

$$y_i = f_i(x_i,\beta) + u_i \qquad i=1,\ldots,n;\ n\geq m \tag{33}$$

where f_i is the response function and x_i is the i^{th} row of X. L_1 norm regression parameters are derived by minimizing the following sum:

$$\min_{\beta}: \sum_{i=1}^{n} |y_i - f_i(x_i,\beta)| \tag{34}$$

The function (34) can be reformulated as a nonlinear programming problem as,

$$\begin{aligned} \min_{\beta}: & \sum_{i=1}^{n} w_i \\ \text{s.to: } & y_i - f_i(x_i,\beta) - w_i \leq 0 \\ & -y_i + f_i(x_i,\beta) - w_i \leq 0 \\ & w_i \geq 0 \\ & i=1,\ldots,n \end{aligned} \tag{35}$$

Over the last two 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).

3.8 L_p norm computation

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

$$\min_{\beta} \sum_{i=1}^{n} |y_i - \sum_{j=1}^{m} \beta_j x_{ij}|^p \qquad (36)$$

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 $u=w-v$; $w,v \geq 0$. The L_p norm approximation problem reduces as follows (see, Kiountouzis (1972)),

$$\begin{aligned} \min_{\beta}: & \sum_{i=1}^{n} (w_i^p + v_i^p) \\ \text{s.to}: & \; w_i - v_i + \sum_{j=1}^{m} \beta_j x_{ij} = y_i \\ & \; w_i, v_i \geq 0 \\ & \; \beta_j \text{ unrestricted in sign} \\ & \; i=1,\ldots,n; \; j=1,\ldots,m \end{aligned} \tag{37}$$

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 $f_i(\mathbf{x}_i, \beta)$ instead. That is,

$$\begin{aligned} \min_{\beta}: & \sum_{i=1}^{n} (w_i^p + v_i^p) \\ \text{s.to}: & \; w_i - v_i + f_i(\mathbf{x}_i, \beta) = y_i \\ & \; w_i, v_i \geq 0 \\ & \; \beta_j \text{ unrestricted in sign} \\ & \; i=1,\ldots,n; \; j=1,\ldots,m \end{aligned} \tag{38}$$

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)).

4. 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,

$$y = Y\theta + X_1\beta + u = [Y \mid X_1]\begin{bmatrix}\theta\\ \hline \beta\end{bmatrix} + u \equiv Z\alpha + u \qquad (39)$$

where y is a vector of dependent endogenous, Y, matrix of independent endogenous, X_1, matrix of exogenous variables; θ and β are vectors of regression parameters and u is random error vector. The reduced form for Y is given by,

$$Y = X\pi + v \qquad (40)$$

Direct and indirect least absolute deviations (DLAD, IDLAD) analogues of direct and indirect least squares (DLS, IDLS) may be applied to the systems (39) and (40) respectively. The

L_1 norm objective function analogue of two stage least squares (2SLS) for estimation of α may be defined as,

$$\min_{\alpha}: \sum_{i=1}^{n} |y_i - P_i^T Z\alpha| \tag{41}$$

where y_i is the i^{th} element of $\mathbf{y}$, P_i^T is the i^{th} row of $P=(X^TX)^{-1}X^T$ (see, Fair (1974)). Amemiya (1982) by comparing the problem (41) with Theil's interpretation of 2SLS,

$$\min_{\alpha}: \sum_{i=1}^{n} (y_i - P_i^T Z\alpha)^2 \tag{42}$$

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

$$\min_{\alpha}: \sum_{i=1}^{n} (P_i^T y - P_i^T Z\alpha)^2 \tag{43}$$

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

$$\min_{\alpha}: \sum_{i=1}^{n} |P_i^T y - P_i^T Z\alpha| \tag{44}$$

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

$$\min_{\alpha}: \sum_{i=1}^{n} |qy_i + (1-q)P_i^T y - P_i^T Z\alpha| \tag{45}$$

where q is a parameter to be determined by researcher. When q=0, problem (45) is equivalent to (44) and yields the estimator which is asymptotically equivalent to 2SLS. When q=1 then (45) is equivalent to (41). For any value of $q\in[0,\infty)$ 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 $\mathbf{u}$ and $\mathbf{v}$. 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 $\hat{\pi}$ is computed. Then by minimizing the following expression,

$$\min_{\beta}: \sum_{i=1}^{n} |y_i - X_i^T \hat{\pi}\theta - X_{1i}^T \beta| \tag{46}$$

$\theta\hat{}$ and $\beta\hat{}$ 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.

5. 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.

5.1 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 (1977) showed that the error in the L_1 norm estimation is approximately,

normally distributed with mean zero and variance covariance matrix $\delta^2(X^TX)^{-1}$, 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 $\delta^2(X^TX)^{-1}$, 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).

5.2 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).

5.3 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 Galpin and Hawkins (1987).

5.4 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).

5.5 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 Galpin (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 Galpin 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).

6. 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), 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).

7. 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) propose 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, $med(u_i^2)$ for 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).

CHAPTER III

INTRODUCING NEW ALGORITHMS

1. Introduction

In the previous chapter, various aspects of the L_1 norm regression was reviewed and we mentioned the works that have been done on this area. We observed that the L_1 norm criterion is going to find its place in scientific analysis. Since it is not computationally comparable with other criteria such as L_2 norm, it needs more work to make it a hand tool. The closed form of the solution of the L_1 norm estimator has not been derived yet, and therefore, makes further inferences of the properties of this estimator difficult. Any attempt to give efficient computational algorithms which may introduce significant insight into the different characteristics of the problem is desirable. In this regard, we shall try to give a general procedure in this chapter to solve L_1 norm linear regression problem. The proposed algorithms are based on a special descent method and use discrete differentiation technique. Primary designs of the algorithms have been discussed by Bidabad (1987a,b,88a,b). By manipulating the algorithms, more efficient ones are introduced which will be shown to have better performance than other existing algorithms.

Consider the following regression model,

$$y_i = \sum_{j=1}^{m} \beta_j x_{ji} + u_i \qquad i=1,\ldots,n \tag{1}$$

where βj, $j=1,\ldots,m$ are population parameters to be estimated, y_i, x_{ji} and u_i are dependent, independent and random variables respectively. In this chapter x_{ji} is used instead of x_{ij} for convenience. We wish to estimate β_j's by minimizing the expression,

$$S = \sum_{i=1}^{n} |u_i\hat{}| = \sum_{i=1}^{n} |y_i - \sum_{j=1}^{m} \beta_j\hat{} x_{ji}| \tag{2}$$

Suppose m=1, then expression (2) reduces to,

$$S = \sum_{i=1}^{n} |y_i - y_i\hat{}| = \sum_{i=1}^{n} |y_i - \beta\hat{} x_i| = \sum_{i=1}^{n} S_i \tag{3}$$

A typical element, $S_i=|y_i-\beta\hat{} x_i|$ can be viewed as a broken line in the $Sx\beta\hat{}$ plane composed of two half-lines. S_i attains

its minimum which is zero at,

$$\beta_i\hat{} = y_i/x_i \qquad (4)$$

The slopes of the half-lines to the left and right of $\beta\hat{}_i$ are $-|x_i|$ and $|x_i|$ respectively. So S_i's are all convex and hence their sum S is also convex with a slope at any $\beta\hat{}$ equal to sum of the slopes of S_i's at that value of $\beta\hat{}$. Since the slope of each S_i changes only at the corresponding $\beta_i\hat{}$, the minimum of S will lie on one of the $\beta_i\hat{}$. Thus, the regression line will pass through origin with the slope equal to $\beta_i\hat{}$ which minimizes S. On the other hand, to find the L_1 norm estimate of β we need to find only one observation (see, Karst (1958), Taylor (1974)). Furthermore, Taylor (1974) concludes that: "This implies, of course, that the regression line must pass through the observation corresponding to the minimizing i. The regression line, therefore, is determined by the point of origin and the observation associated with minimizing $\beta_i\hat{}$". But he did not continue this approach, that is minimizing S with respect to subscript i. In this chapter as Bidabad (1987a,88a) a first endeavor is to develop this point of view. In the next section after rewriting (3) in a suitable fashion, the value of i is determined by using discrete differentiation.

By a similar discussion it can be concluded that when the number of parameters is m, m observations must lie on the regression hyperplane. On the other hand, m equations of the form given by (5) are necessary to specify the regression hyperplane (see also chapter I).

$$y_i - \sum_{j=1}^{m} \beta_j\hat{}\, x_{ji} = 0 \qquad (5)$$

Now, let us proceed with the simplest regression model.

2. Restricted simple linear regression

For model (1) consider the case of one independent variable with no intercept, namely, $y_i=\beta x_i+u_i$. To find the L_1 norm estimate of β, the following procedure can be suggested.

$$S = \sum_{i=1}^{n} |u_i| = \sum_{i=1}^{n} |y_i - \beta x_i| = \sum_{i=1}^{n} (y_i - \beta x_i)\,\mathrm{sgn}(y_i - \beta x_i) =$$

$$\sum_{i=1}^{n} |x_i|(y_i/x_i - \beta)\,\mathrm{sgn}(y_i/x_i - \beta) \qquad (6)$$

Let $z_i = y_i/x_i$ and sort z_i in descending order. Rename the resultant ordered z_i, $i=1,\ldots,n$ to z_h, $h=1,\ldots,n$. Elements of z_h should have the following property,

$z_h > z_l$ if $h < l$ for all $h,l=1,\ldots,n$.

Rewrite (6) with ordered observations as,

$$S = \sum_{h=1}^{n} |x_h|(z_h - \beta)\,\mathrm{sgn}(z_h-\beta) \qquad (7)$$

Let us denote the observation which will be on the regression line by (x_{t+1}, y_{t+1})- that is the $(t+1)^{th}$ observation in the z_h array. Value of z_h is the slope of a ray passing through the origin and the h^{th} observation. Therefore,

if, $h < t+1$ then, $z_h > \beta$ and $u_h > 0$

if, $h = t+1$ then, $z_h = \beta$ and $u_h = 0$

if, $h > t+1$ then, $z_h < \beta$ and $u_h < 0$

We can now rewrite (7) as follows,

$$S = \sum_{h=1}^{t} |x_h|(z_h - \beta) - \sum_{h=t+1}^{n} |x_h|(z_h - \beta) \qquad (8)$$

To find the minimum of S, we need to differentiate it with respect to β and subscript t, because both β and t are unknowns. Note that β has continuous domain and t has discrete domain. Thus,

$$\frac{\delta S}{\delta \beta} = -\sum_{h=1}^{t} |x_h| + \sum_{h=t+1}^{n} |x_h| \qquad (9)$$

The differentiation of S with respect to subscript t must be discrete derivative (see, Bender and Orszag (1978), Clarke (1983)):

$$\frac{\mathrm{Delta}(S)}{\mathrm{Delta}(t)} = \lim_{\mathrm{Delta}(t)\longrightarrow 1} \frac{S[t+\mathrm{Delta}(t)]-S(t)}{\mathrm{Delta}(t)} = S(t+1)-S(t) =$$

$$\sum_{h=1}^{t+1} |x_h|(z_h-\beta) - \sum_{h=t+2}^{n} |x_h|(z_h-\beta) -$$

$$\sum_{h=1}^{t} |x_h|(z_h-\beta) + \sum_{h=t+1}^{n} |x_h|(z_h-\beta) = 0$$

or;

$$2|x_{t+1}|(z_{t+1}-\beta) = 0$$

thus,

$$\beta^{\wedge} = z_{t+1} = y_{t+1}/x_{t+1} \qquad (10)$$

It should be noted that if $x_{t+1}=0$ then the value of $z_{t+1}=\infty$, and when the array z is sorted, infinite values of z locate at extremes of two tails of the array and make no problem in (10). Remember that equation (10) is the same as relation (4). Equations (9) and (10) are two equations with two unknowns t and β. The variable t can be found by rewriting (9) as follows,

$$D_k \equiv \sum_{h=1}^{k} |x_h| - \sum_{h=k+1}^{n} |x_h| \qquad k=1,\ldots,n \qquad (11)$$

It is obvious that D_k is an increasing function of k. When k increases from one to n, D_k attains different values which increases from negative to positive. So, initially k is set equal to one and D_k is computed accordingly. If D_k is negative, k is increased by one and the procedure is repeated until D_k becomes positive. When D_k reaches the first positive value, then, t+1=k. By this procedure, value of t+1 is found. The observation corresponding to this subscript (t+1=k) is selected (x_{t+1},y_{t+1}). The L_1 norm estimate of β is found by substituting the values of x_{t+1} and y_{t+1} into (10), (see, Bidabad (1987a,88a)).

The presented procedure is essentially the weighted median procedure of Laplace (1818) which was discussed in the previous chapter. Namely, this solution is the steps that Laplace took. Main difference is the mathematical derivation. Laplace found this solution by analyzing algebraic characteristic of the L_1 norm objective function; while the above procedure exactly differentiates the objective function. Another new contribution of this procedure is application of the partial discrete differentiation over subscript accompanying with conventional differentiation on a regular variable with continuous domain. It should be remembered that Boscovich (1757) solved this problem by a geometrical procedure and Karst (1958) by an analytical

method.

2.1 Computation of weighted median

Two series of computations are necessary to compute weighted median. One sorting algorithm is essential to sort the z array and restoring the corresponding subscripts to compute the second part of calculation to find the optimal value of k by using D_k defined by (11).

Efficient sorting algorithms exist for the first part of the computation. The algorithms 'quicksort' of Hoare (1961,62), 'quickersort' of Scowen (1965) and 'sort' of Singleton (1969) have desirable performances and efficiencies. For the second part of the computation there is no special purpose procedure, but Bloomfield and Steiger (1980a) used the partial sorting of Chambers (1971) to give an efficient way to combine the two steps of sorting the z array and finding the value of k. The superiority of this procedure is in sorting the smaller segments of z array rather than all its elements. With some modification, this procedure is used for our proposed algorithms in the forthcoming sections. This procedure can be stated as the following function.

FUNCTION LWMED (n,ys,w,l)

Step 0) Initialization.
Real: ys(n), w(n).
Integer: l(n), hi.
Set: ii=0, shi=0, slo=0, sz=0, sp=0, sn=0.

Step 1) Compute left, middle and right sum of weights.
Do loop for i=1,n: w(i)=|w(i)|; if ys(i)<0, then sn=sn+w(i), if ys(i)>0, then sp=sp+w(i), if ys(i)=0 then sz=sz+w(i); end do.
If shi≤slo then go to step 2.b, otherwise go to step 2.a.

Step 2) Assign subscripts for arrays.
a. Let: shi=0.

Do loop for i=1,n: if ys(i)≤0 go to continue, otherwise ii=ii+1, l(ii)=i, continue, end do.
Go to step 2.c.

b. Let: slo=0.
Do loop for i=1,n: if ys(i)>0 go to continue, otherwise ii=ii+1, l(ii)=i, continue, end do.

c. Let: lo=1, hi=ii.

Step 3) Check for solution.
If hi>lo+1 then go to step 4, otherwise lwmed=l(lo).
If lo=hi return, otherwise if ys(l(lo))≤ys(l(hi)) go to step 3.a, otherwise lt=l(lo), l(lo)=l(hi), l(hi)=lt, lwmed=l(lo).

a. If shi+w(l(hi))>slo+w(l(lo)) then set lwmed=l(hi), otherwise return.

Step 4) Divide the string into two halves then sort.
Set: mid=(lo+hi)/2, lop=lo+1, lt=l(mid), l(mid)=l(lop), l(lop)=lt.

a. If ys(l(lop))≤ys(l(hi)) then go to step 4.b, otherwise lt=l(lop), l(lop)=l(hi), l(hi)=lt.

b. If ys(l(lo))≤ys(l(hi)) then go to step 4.c, otherwise lt=l(lo), l(lo)=l(hi), l(hi)=lt.

c. If ys(l(lop))≤ys(l(lo)) then go to step 5, otherwise lt=l(lop), l(lop)=l(lo), l(lo)=lt.

Step 5) Compute the accumulation of weights.
Let: lwmed=l(lo), i=lop, j=hi, xt=ys(lwmed), tlo=slo, thi=shi.

a. Set: tlo=tlo+w(l(i)), i=i+1.
If ys(l(i))<xt then go to step 5.a, otherwise go to step 5.b.

b. Let: thi=thi+w(l(j)), j=j-1.
If ys(l(j))>xt then go to step 5.b, otherwise if j≤i then go to step 6, otherwise lt=l(i), l(i)=l(j), l(j)=lt, go to step 5.a.

Step 6) Test for solution.
Let: test=w(lwmed).
If i≠j then go to step 6.a, otherwise

test=test+w(l(i)), i=i+1, j=j-1.

a. If test≥|thi-tlo| then return, otherwise, if tlo>thi then step 6.b, otherwise slo=tlo+test, lo=i, go to step 3.

b. Let: shi=thi+test, lo=lop, hi=j.
Go to step 3.

END

2.2 Discussion of the m parameter model

The procedure applied to the simple one parameter model can not be simply generalized to the m parameter model. This difficulty is due to the fact that we can not reorder observations in such a way that their corresponding residuals are increasingly or decreasingly ordered. On the other hand, to apply the discrete differentiation technique, primarily (2) should be rewritten as follows:

$$S = \sum_{h=1}^{t} (y_h - \sum_{j=1}^{m} \beta_j x_{jh}) - \sum_{h=t+1}^{n} (y_h - \sum_{j=1}^{m} \beta_j x_{jh}) \qquad (12)$$

Expression (12) which is free of absolute value sign is generalization of (8) for m parameters. But our first problem is to find a logic which enables us to form (12). On the other hand, we need to reorder observations in such a way that when h is less, equal or greater than t+1, u_h be greater, equal or less than zero respectively; and as h increases from one to n, the corresponding residual u_h decreases accordingly.

If (2) could be written as (12), we could again differentiate it with respect to β_j for all j=1,...,m and t. Differentiating S with respect to β_j for all j=1,...,m would give,

$$\frac{\delta S}{\delta \beta_j} = -\sum_{h=1}^{t} x_{jh} + \sum_{h=t+1}^{n} x_{jh} \qquad j=1,\ldots,m \qquad (13)$$

In order to differentiate S with respect to discrete domain variable t we would have,

$$\frac{\text{Delta}(S)}{\text{Delta}(t)} = \lim_{\text{Delta}(t)\longrightarrow 1} \frac{S[t+\text{Delta}(t)]-S(t)}{\text{Delta}(t)} = S(t+1)-S(t) =$$

$$\sum_{h=1}^{t+1}\Big(y_h - \sum_{j=1}^{m}\beta_j x_{jh}\Big) - \sum_{h=t+2}^{n}\Big(y_h - \sum_{j=1}^{m}\beta_j x_{jh}\Big) -$$

$$\sum_{h=1}^{t+1}\Big(y_h - \sum_{j=1}^{m}\beta_j x_{jh}\Big) + \sum_{h=t+1}^{n}\Big(y_h - \sum_{j=1}^{m}\beta_j x_{jh}\Big) =$$

$$2\Big(y_{t+1} - \sum_{j=1}^{m}\beta_j x_{j,t+1}\Big) = 0 \qquad (14)$$

m distinct values for t can be computed from (13) by using the following D_{kj} for each explanatory variable. The procedure to compute t is similar to what we did in (11) for the one parameter model.

$$D_{kj} \equiv \sum_{h=1}^{k} x_{jh} - \sum_{h=k+1}^{n} x_{jh} \qquad j=1,\ldots,m;\ i=1,\ldots,n \qquad (15)$$

It should be noted that these m values for t, all give a unique value for S in (12), because all of the errors corresponding to these t's have zero values. This becomes clear by zero residuals property of the L_1 norm regression discussed in chapter I, that there exist m points on the regression hyperplane which have zero residuals. In (12), these m points locate sequentially, because we had reordered observations in such a way that, when h increases from one to n, the errors decrease from greatest positive to lowest negative values. So these m zero errors will locate one after another nearly in the middle of n elements sequence of (12). Corresponding to these m values of t, m observations are recognized. Substitute the values of these m observations into (14), m equations are found which can be solved simultaneously for m unknown β_j's. These last m equations again confirm the above cited property of the L_1 norm regression.

Since for the m parameter model, the ordering of residuals before computing optimal values of β_j's is not possible, another device should be adopted to compute the L_1 norm estimates of the multiple linear regression parameters.

3. Unrestricted simple linear regression

To find the L_1 norm estimates of $\{\beta_j, j=1,\ldots,m\}$ for

model (1), we shall try to propose algorithms to search for those observations on the optimal regression hyperplane. For $m=2$ and $x_{1i}=0$, $i=1,\ldots,n$; (1) reduces to restricted simple linear regression with only one parameter. The L_1 norm estimate of the slope $\beta 2$ for simple model can be obtained by the algorithm given in the previous sections. Let us now consider a simple unrestricted linear model in which $m=2$ and $x_{1i}=1$ for all $i=1,\ldots,n$; namely,

$$y_i = \beta_1 + \beta_2 x_{2i} + u_i \tag{16}$$

3.1 Algorithm 1

For the model given in (16), the L_1 norm objective function S to be minimized will be,

$$S = \sum_{i=1}^{n} |y_i - \beta_1 - \beta_2 x_{2i}| \tag{17}$$

Let k_1 denote a subscript which belongs to the range of one to n, and assume that $k1^{th}$ observation (y_{k1}, x_{2k1}) is a candidate to be on the regression line. If this is the case, then $u_{k1}=0$ and we can transfer the origin of YxX_2 coordinates to the point (y_{k1}, x_{2k1}) without any loss. For this, we should rewrite all observations as deviates from the point (y_{k1}, x_{2k1}),

$$\begin{aligned} y_i^{k1} &= y_i - y_{k1} && i=1,\ldots,n \\ x_{2i}^{k1} &= x_{2i} - x_{2k1} && i=1,\ldots,n \end{aligned} \tag{18}$$

Where the terms with superscripts k1 denote deviated values. Rearrange the terms;

$$\begin{aligned} y_i &= y_i^{k1} + y_{k1} && i=1,\ldots,n \\ x_{2i} &= x_{2i}^{k1} + x_{2k1} && i=1,\ldots,n \end{aligned} \tag{19}$$

Now, substitute (19) into (17), then, our L_1 norm minimization problem can be redefined as,

$$\min_{\beta_1,\beta_2}: S_{k1} = \sum_{i=1}^{n} |y_i^{k1} - \beta_2 x_{2i}^{k1} + (y_{k1} - \beta_1 - \beta_2 x_{2k1})| \tag{20}$$

Since, we assumed that the $k1^{th}$ observation is on the regression line, the expression $y_{k1}-\beta_1-\beta_2 x_{2k1}=0$. Thus (20) reduces to,

$$\min_{\beta_2}: S_{k1} = \sum_{i=1}^{n} |y_i^{k1} - \beta_2 x_{2i}^{k1}| \tag{21}$$

Solution of the optimization problem (21), is the same as the one described before for one parameter linear model. Note that when $u_{k1}=0$, then the minimum value of (21) is equal to the minimum value of (17).

Let β_2 derived by minimizing (21) be denoted by β_2^{k1}. By changing k1 from one to n and minimizing (21), $\beta_2^1,\ldots,\beta_2^n$ are attained accordingly. Now the question is: what value of k1 minimizes (17) ?. In other words, which observations are on the optimal regression line ?. Note that in the two parameter L_1 norm linear model, there exists at least two observations with zero errors. Transferring of YxX_2 coordinates to both these points does not change the minimum of (17). Suppose p^{th} and q^{th} are those observations on the regression line, thus,

$$u_p = y_p - \beta_1 - \beta_2 x_{2p} = 0 \qquad (22)$$
$$u_q = y_q - \beta_1 - \beta_2 x_{2q} = 0$$

Rewriting (18) and (19) for k1=p,q and substituting them into (21), yields,

$$S_p = \sum_{i=1}^{n} |y_i^p - \beta_2 x_{2i}^p| \qquad (23)$$

$$S_q = \sum_{i=1}^{n} |y_i^q - \beta_2 x_{2i}^q| \qquad (24)$$

Using (18) and rewriting (23) and (24), it is shown that S_p is equal to S_q,

$$S_p = \sum_{i=1}^{n} |y_i-\beta_2 x_{2i} - (y_p-\beta_2 x_{2p})| \qquad (25)$$

$$S_q = \sum_{i=1}^{n} |y_i-\beta_2 x_{2i} - (y_q-\beta_2 x_{2q})| \qquad (26)$$

S_p is equal to S_q if and only if the two parentheses inside the absolute value signs of (25) and (26) are equal. This can be concluded by solving the two equations of (22) for β_1. That is,

$$\beta_1 = y_p - \beta_2 x_{2p} = y_q - \beta_2 x_{2q} \qquad (27)$$

The equality of S_q and S_p is because the points p and q are on the regression line and therefore $u_p=u_q=0$. Thus, $S_p=S_q$, so, the values of β_2^p and β_2^q derived by minimizing either S_p

or S_q must be equal. This gives a criterion to find the desired β_2 from all β_2^{k1} for k1=1,...,n. That is, when $\beta_2^p=\beta_2^q$. The estimated value of β_2 is denoted by $\beta_2\hat{}$. Value of $\beta_1\hat{}$ is simply computed by (27). Now let us summarize the whole procedure for finding the values of $\beta_1\hat{}$ and $\beta_2\hat{}$ for the model given by (16).

Crude algorithm 1

Step 0) Set k1=1.

Step 1) Compute (18).

Step 2) Minimize (21) by using weighted median procedure and find β_2^{k1}.

Step 3) Check if $\beta_2^{k1}=\beta_2^{k1-k}$ for $0<k<k1$, then set $\beta_2\hat{}=\beta_2^{k1}$ and $\beta_1\hat{}=y_{k1}-\beta_2\hat{}x_{2k1}$ then stop.

Step 4) Increase k1 by one and go to step 1.

3.2 Algorithm 2

The crude algorithm 1 for finding $\beta_1\hat{}$ and $\beta_2\hat{}$ is not computationally efficient because it usually requires testing the majority of observations. Therefore, in order to make this algorithm efficient, some elaborations are necessary.

Instead of setting k1=1, let us set k1 equal to an arbitrary integer value such as a where a is any integer from one to n. Suppose now, $u_a=y_a-\beta_1-\beta_2x_{2a}=0$, and rewrite (21) as,

$$\min_{\beta_2}: S_a = \sum_{i=1}^{n} |y_i^a - \beta_2 x_{2i}^a| \qquad (28)$$

Minimizing (28) gives β_2^a which is equal to

$$\beta_2^a = \frac{y_b^a}{x_{2b}^a} = \frac{y_b - y_a}{x_{2b} - x_{2a}} \qquad (29)$$

On the other hand, by pivoting on the a^{th} observation another point such as b is found where b refers to that observation which has zero error in the minimum solution of (28), so, $u_b=y_b-\beta_1-\beta_2x_{2b}=0$. Let us denote the minimum of S_a in (28) as S_a^*, then,

$$S_a^* = \sum_{i=1}^{n} |y_i^a - \beta_2^a x_{2i}^a| \qquad (30)$$

Note that,

$$\beta_1 = y_a - \beta_2{}^a x_{2a} = y_b - \beta_2{}^a x_{2b} \qquad (31)$$

This comes from multiplying (29) by its denominator and rearranging the terms. Substitute (31) in (17),

$$S = \sum_{i=1}^{n} |y_i - y_a - \beta_2{}^a (x_{2i} - x_{2a})| = \sum_{i=1}^{n} |y_i - y_b - \beta_2{}^a (x_{2i} - x_{2a})|$$

or,

$$S = \sum_{i=1}^{n} |y_i{}^a - \beta_2{}^a x_{2i}{}^a| = \sum_{i=1}^{n} |y_i{}^b - \beta_2{}^a x_{2i}{}^b| \qquad (32)$$

Using (30) the first sum in (32) is S_a^* and the second sum is S_b evaluated at $\beta_2=\beta_2{}^a$. Thus, it can be concluded that,

$$S_a^* = S_b\Big|_{\beta_2=\beta_2{}^a} \qquad (33)$$

S_a^* is at minimum but $S_b\Big|_{\beta_2=\beta_2{}^a}$ still can be minimized for other values of β_2. Therefore, an important result is derived, that is,

$$S_a^* > S_b^* \qquad (34)$$

The inequality of (34) guarantees that if we choose an arbitrary point to transfer the origin of coordinates to it and minimizing the objective function (21) another point is found, then transferring the origin to this newly found point decreases the total sum of absolute errors. Therefore, at each transference we get closer to the minimum of S. By a similar discussion, this conclusion can be generalized as,

$$S_a^* > S_b^* > S_c^* > S_d^* \ldots \qquad (35)$$

Note that a is an arbitrary starting point. The point b is derived by minimizing S_a, c is derived by minimizing S_b and d is derived by minimizing S_c and so on.

Now, the question is, when the minimum value of S is reached ?. Suppose $S^*=S_f^*$; by transferring the origin of coordinates to the point f and minimizing S_f, the g^{th} observation is derived. When $S^*=S_f^*$ by minimizing S_g, f^{th} observation is again found, because $S_g^*=S_f^*=S^*$ and the g^{th} and f^{th} observations are both on the L_1 norm regression line. This conclusion can be a criterion to stop the procedure. Hence,

$$S_a^* > S_b^* > S_c^* > S_d^* \ldots > S_f^* = S_g^* = S^* \qquad (36)$$

It should be noted that if the minimum solution of S is not unique, that is function S has a horizontal segment, the procedure stops when it reaches the first minimum solution.

Now, let us introduce the whole stages of this algorithm to find the L_1 norm estimates of β_1 and β_2 in the simple linear model (16).

Algorithm 2

Step 0) Select an arbitrary integer value a between one to n and set k1=a.

Step 1) Compute (18) with k1=a.

Step 2) Minimize (21) by using weighted median procedure and find the appropriate observation on the line, namely observation b.

Step 3) Compute (18) with k1=b.

Step 4) Minimize (21) and find that observation on the line; observation c.

Step 5) Check if c=a, then $\beta_2\hat{}=y_b{}^c/x_{2b}{}^c$, $\beta_1\hat{}=y_c-\beta_2\hat{}x_{2c}$ and stop.

Step 6) Set a=b and go to step 1.

Operationally, the following steps actually are taken. The following program will be used in chapter IV in comparison to other algorithms for the two parameter linear regression model.

PROGRAM BL1S

Step 0) Initialization.
Parameter: n.
Real: y(n), x2(n), z(n), w(n).
Integer: l(n).
Set: k1=arbitrary, k1r=0, k1s=0, iter=0.
Read (y(i), x2(i), i=1,n)

Step 1) Compute weights and ratios.
Do loop for i=1,k1-1: w(i)=x2(i)-x2(k1), z(i)=(y(i)-y(k1))/w(i), end do.
Set: w(k1)=0, z(k1)=0.

```
        Do    loop    for    i=k1+1,n:    w(i)=x2(i)-x2(k1),
        z(i)=(y(i)-y(k1))/w(i) end do.
        Set: iter=iter+1.
Step 2) Compute weighted median.
        Let: lm=LWMED(n,z,w,1).
Step 3) Check for optimality.
        Set: k1s=k1r, k1r=k1.
        If lm=k1s then go to step 4, otherwise k1=lm.
        Go to step 1.
Step 4) Compute the solution.
        Let b2=z(lm), b1=y(k1)-b2*x2(k1).
        Print b1, b2, k1, lm, iter.
        stop.
END
```

4. Generalization to m parameters

Now we extend the above procedure to the restricted two parameter model, namely,

$$y_i = \beta_2 x_{2i} + \beta_3 x_{3i} + u_i \tag{37}$$

Let,

$$S = \sum_{i=1}^{n} |y_i - \beta_2 x_{2i} - \beta_3 x_{3i}| \tag{38}$$

S can be written as,

$$S = \sum_{i=1}^{n} |x_{2i}||y_i/x_{2i} - \beta_2 - \beta_3 x_{3i}/x_{2i}| =$$

$$\sum_{i=1}^{n} |x_{2i}||y_i{}^{s1} - \beta_2 - \beta_3 x_{3i}{}^{s1}| \tag{39}$$

where,

$$\begin{aligned} y_i{}^{s1} &= y_i/x_{2i} && i=1,\ldots,n \\ x_{3i}{}^{s1} &= x_{3i}/x_{2i} && i=1,\ldots,n \end{aligned} \tag{40}$$

Minimization of (39) is similar to that of simple linear model explained in the previous section. An important distinction for solving (39) in comparison to (17) is the expression $|x_{2i}|$ which has been multiplied to $|y_i{}^{s1}-\beta_2-\beta_3 x_{3i}{}^{s1}|$. This multiplication does not make any problem when (39) is minimized, because if we deviate $y_i{}^{s1}$

and x_{3i}^{s1} from the point k1 as,

$$\begin{aligned} y_i^{s1k1} &= y_i^{s1} - y_{k1}^{s1} && i=1,\ldots,n \\ x_{3i}^{s1k1} &= x_{3i}^{s1} - x_{3k1}^{s1} && i=1,\ldots,n \end{aligned} \tag{41}$$

then we can rewrite (39) similar to (21), so,

$$S_{k1} = \sum_{i=1}^{n} |x_{2i}||y_i^{s1k1} - \beta_3 x_{3i}^{s1k1}| \tag{42}$$

To minimize (42), we should use the following expression,

$$S_{k1} = \sum_{i=1}^{n} |x_{2i}x_{3i}^{s1k1}||y_i^{s1k1}/x_{3i}^{s1k1} - \beta_3| \tag{43}$$

According to Bidabad (1987a,88a), in applying discrete derivative to (43) which was described in the previous section the term in the first absolute value sign is used to find the subscript of that point which locates on the regression line. In comparison to the simple unrestricted linear model (16) this is the main difference.

In the case of inclusion of an intercept in the model given by (37), we have,

$$S = \sum_{i=1}^{n} |y_i - \beta_1 - \beta_2 x_{2i} - \beta_3 x_{3i}| \tag{44}$$

Let k1 be an arbitrary subscript then, transference of the origin of coordinates to the k1th point is done by deviating all observations from this point, namely,

$$\begin{aligned} y_i^{k1} &= y_i - y_{k1} && i=1,\ldots,n \\ x_{2i}^{k1} &= x_{2i} - x_{2k1} && i=1,\ldots,n \\ x_{3i}^{k1} &= x_{3i} - x_{3k1} && i=1,\ldots,n \end{aligned} \tag{45}$$

Rearrange the terms of (45) and substitute in (44), then we have,

$$S = \sum_{i=1}^{n} |y_i^{k1} - \beta_2 x_{2i}^{k1} - \beta_3 x_{3i}^{k1} + (y_{k1} - \beta_1 - \beta_2 x_{2k1} - \beta_3 x_{3k1})| \tag{46}$$

If k1th observation is on the regression plane, then

$$y_{k1} - \beta_1 - \beta_2 x_{2k1} - \beta_3 x_{3k1} = 0 \tag{47}$$

So, instead of minimizing (44), the following function is minimized:

$$S_{k1} = \sum_{i=1}^{n} |y_i^{k1} - \beta_2 x_{2i}^{k1} - \beta_3 x_{3i}^{k1}| \tag{48}$$

Minimization of (48) is completely similar to that of (38)

and can be proceeded as follows,

$$S_{k1} = \sum_{i=1}^{n} |x_{2i}^{k1}| |y_i^{k1}/x_{2i}^{k1} - \beta_2 - \beta_3 x_{3i}^{k1}/x_{2i}^{k1}| =$$

$$\sum_{i=1}^{n} |x_{2i}^{k1}| |y_i^{s1} - \beta_2 - \beta_3 x_{3i}^{s1}| \qquad (49)$$

where,

$$y_i^{s1} = y_i^{k1}/x_{2i}^{k1}$$
$$x_{3i}^{s1} = x_{3i}^{k1}/x_{2i}^{k1} \qquad (50)$$

Now, again, transfer the origin of the two dimensional $Y^{s1}X_3^{s1}$ coordinates to an arbitrary point k2 by deviating y_i^{s1} and x_{3i}^{s1} from y_{k2}^{s1} and x_{3k2}^{s1} as follows,

$$y_i^{s1k2} = y_i^{s1} - y_{k2}^{s1} \qquad i=1,\ldots,n$$
$$x_{3i}^{s1k2} = x_{3i}^{s1} - x_{3k2}^{s1} \qquad i=1,\ldots,n \qquad (51)$$

By rearranging the terms of (51) and substituting them into (49) and assuming that the point k2 is on the regression plane, we can rewrite (49) as,

$$S_{k1k2} = \sum_{i=1}^{n} |x_{2i}^{k1}| |y_i^{s1k2} - \beta_3 x_{3i}^{s1k2}| \qquad (52)$$

or,

$$S_{k1k2} = \sum_{i=1}^{n} |x_{2i}^{k1} x_{3i}^{s1k2}| |y_i^{s1k2}/x_{3i}^{s1k2} - \beta_3| \qquad (53)$$

The objective function (53) can be minimized as suggested before. Now, the procedure from (49) to (53) can be repeated with different values of k2 as in algorithm 2 proposed for the simple linear model. When the last point (M) in the process of minimizing (53) is reached, the origin of three dimensional YxX_2xX_3 coordinates (k1) is transferred to this newly found point M and the procedure from (45) to (53) is again repeated with the exception that instead of assigning an arbitrary value to k2, we set k2 equal to the previous value of k1. This procedure continues until the point found by minimizing (53) is equal to previous value of k1. The values of $\beta_1\hat{}$, $\beta_2\hat{}$ and $\beta_3\hat{}$ can be computed according to the following formulas,

$$\beta_3\hat{} = y_M^{s1k2}/x_{3M}^{s1k2}$$
$$\beta_2\hat{} = y_{k2}^{s1} - \beta_3\hat{} x_{3k2}^{s1} \qquad (54)$$
$$\beta_1\hat{} = y_{k1} - \beta_2\hat{} x_{2k1} - \beta_3\hat{} x_{3k1}$$

It should be noted that at each step it can be proved that we are getting closer to the minimum of S in (44). The proof is similar to that of the two parameter model given by (16).

It can be shown that for any two arbitrary points k1 and k2, value of S_{k1k2} given by (53) for any commutation for k1 and k2 remains unchanged. This will be shown in the next sections. However,

$$S_{k1k2} = S_{k2k1} \tag{55}$$

Now, following the procedure of the proof given by (28) through (36) we can write,

$$S_{k1a}^* = S_{k1b}\Big|_{\beta_3=\beta_3^{k1a}} \tag{56}$$

where S_{k1a}^* denotes the optimal value of S_{k1a} with respect to β_3 of (53) for any arbitrary integer a from one to n, namely,

$$S_{k1a}^* = \min_{\beta_3} (S_{k1a}) \tag{57}$$

Subscript b in (56) denotes the observation which is found by minimizing S_{k1a}. Since the a^{th} and b^{th} observations both are on the regression hyperplane, by a similar discussion stated by (28) through (32) we can conclude that S_{k1a}^* is equal to S_{k1b} which evaluated at that value of β_3 which is found by minimizing S_{k1a}. Since the left hand side of (56) is minimum and its right hand side can be decreased by other values of β_3, it can be concluded that,

$$S_{k1a}^* > S_{k1b}^* \tag{58}$$

According to (55) we can write,

$$S_{k1a}^* > S_{k1b}^* = S_{bk1}^* \tag{59}$$

Since in each step we discard an observation from the basis and replace it with the newly found one, we have,

$$S_{k1a}^* > S_{k1b}^* = S_{bk1}^* > S_{bc}^* = S_{cd}^* \tag{60}$$

or generally,

$$S_{k1a}^* > S_{k1b}^* > S_{bc}^* > S_{cd}^* \tag{61}$$

The minimum solution S^* is reached when by entering the f^{th} observation in the basis we can not reduce the objective function value. That is the previous observation is again reached, namely,

$$S_{k1a}^* > S_{ab}^* > S_{bc}^* > S_{cd}^* > \ldots > S_{fg}^* = S_{gf}^* = S^* \quad (62)$$

The relation (62) guarantees that at each step we are descending down the objective function surface.

4.1 Algorithm 3

To generalize the above algorithm to the m parameter linear model (1), we should reduce the number of parameters in the same fashion as in the case of three parameter model explained above. If the model is restricted, we can make it unrestricted by dividing all dependent and independent variables to one of independent variables as follows.

$$S = \sum_{i=1}^{n} |y_i - \sum_{j=2}^{m} \beta_j x_{ji}| = \sum_{i=1}^{n} |x_{2i}| \left| y_i/x_{2i} - \beta_2 - \sum_{j=3}^{m} \beta_j x_{ji} \right| \quad (63)$$

If the model is unrestricted, we can make it restricted by deviating all observations from an arbitrary one observation; namely,

$$S = \sum_{i=1}^{n} |y_i - \beta_1 - \sum_{j=2}^{m} \beta_j x_{ji}| = \sum_{i=1}^{n} |y_i^{k1} - \sum_{j=2}^{m} \beta_j x_{ji}^{k1}| \quad (64)$$

where,

$$\begin{aligned} y_i^{k1} &= y_i - y_{k1} && i=1,\ldots,n \\ x_{ji}^{k1} &= x_{ji} - x_{jk1} && i=1,\ldots,n;\ j=2,\ldots,m \end{aligned} \quad (65)$$

Therefore, according to the transformations (63), (64) and (65) any m parameter model can be reduced to a simple restricted one parameter model and then be estimated as a weighted median problem. To conduct this transformation, if the model is unrestricted, we should deviate all observations from an arbitrary one and make the model restricted. Then divide all dependent and independent variables to an arbitrary independent variable. This makes the model unrestricted. At this step we have reduced the number of parameters by one. By continuing this process, we can reduce any m parameter model to a one parameter model. If the model is restricted we should start by dividing all of the variables by one of the independent variables and make the model unrestricted. Now, we can again reduce one of the parameters by applying the procedures explained above in

order to transform unrestricted model to restricted form. By solving the one parameter model, we can then solve for the two parameter and then three parameter models and so on.

The most delicate part of this algorithm is that, at the starting point, once k1,k2,...,k(m-1) are selected arbitrarily, then the algorithm assigns the best possible values to the integers k1,k2,...,k(m-1). To explain the procedure, let us deal with the four parameter unrestricted linear model. Once, value of k1 is arbitrarily selected. Deviating all observations from $k1^{th}$ observation. In this way, the model reduces to a three parameter restricted model. By dividing all of the variables to one of the independent variables, our model becomes completely similar to (44). By minimizing (44) according to the algorithm previously explained for the three parameter model, subscript M corresponding to M^{th} point is derived. This is the newly found point which its subscript (M) is assigned to k1. The previous value of k1 is assigned to k2 and the previous value of k2 is assigned to k3 and the whole procedure is repeated again. The procedure stops when the value of M is equal to k1.

The above important assigning technique which is essential for pivoting on the origins of different size coordinates can be extended to more parameters models as we did before. Again, we should note that at each succeeding step we get closer to the minimum of S in (1). The proof is completely similar to the three parameter model explained before. The whole procedure is as follows.

Algorithm 3

Step 0) Select arbitrary points k1,k2,...,k(m-1).

Step 1) Set counter=1.

Step 2) Proceed the transformations (63) through (65) to reduce model to a one parameter transformed restricted linear model.

Step 3) Apply weighted median procedure to find the point M.

Step 4) If M is equal to previous value of k(counter) go to step 5; otherwise set k(counter)=M and go to step 1.

Step 5) If counter = m-1 go to step 6; otherwise let counter=counter+1 and go to step 4.

Step 6) Set k(m)=M. By using the final values of k1,...,k(m) solve the following system of equations for $\beta_j\hat{}$, for j=1,...,m;

$$\sum_{j=1}^{m} \beta_j\hat{}\, x_{jki} = y_{ki} \qquad i=1,\ldots,m$$

Stop.

4.2 Algorithm 4

In algorithm 3 to solve any m parameter model, the reduced m-1 parameter model should be solved first and therefore makes the algorithm rather costly. Because to solve a m parameter model we should reduce the objective function value of the reduced m-1 parameter model down to its minimum value and then going back to the m parameter model. In interior steps for m-1,m-2,...,1 this procedure repeats. In order to make this algorithm more efficient another assigning technique may be adopted.

Once m-1 observations are selected arbitrarily. The S function is reduced to a one parameter weighted median problem as we did in algorithm 3 using (63) and (64). By solving the weighted median problem a new observation M is found. Now, the newly found point is replaced by the most previous observation which entered the basis. The procedure is repeated until we can not find any new observation out of the current set of basis points and the deleted point in the previous iteration. On the other hand, in this algorithm m-1 point is selected as a basic feasible solution. By pivoting on this solution a new point is found to enter the basis and should be replaced by another one which was previously in the basis. The procedure ends when we can not enter any new point to reduce the objective function value. In this context it is similar to simplex procedure of linear programming technique.

Algorithm 4 is also a descent method. Because in each iteration the value of S is decreased. Proof is completely similar to that of algorithm 3 stated before. The main difference of this algorithm with algorithm 3 is in choosing the point which should be deleted from the basis. Therefore, in each step we need not to keep some points in the basis to solve smaller size models. On the other hand, in each m iterations all m points in the basis are discarded and new points enter. In algorithm 3 we kept some points in the basis until we reach an optimal point for smaller size reduced model. It should be noted that one point may be discarded and reenter the basis through m iterations. This is not contradictory and actually occurs specially when we are near the optimal solution.

As we will see in the forthcoming section, due to different properties of this algorithm many manipulations in computational steps may be adopted to start calculations. Before making this algorithm more sophisticated it is suitable to summarize its general steps as follows.

Algorithm 4

Step 0) Select arbitrary points k1,...,k(m-1).

Step 1) Proceed the transformations (63) through (65) to reduce model to one parameter transformed restricted linear model.

Step 2) Apply weighted median procedure to find the point M.

Step 3) Assign M to one of the k1,...,k(m-1) which is oldest in the basis.

Step 4) Check that if the newly found point M which enters the basis discard the point which entered in previous iteration repeats m-1 times then go to step 5; otherwise go to step 1.

Step 5) Set k(m)=M. By using the final values of k1,...,k(m), solve the following system of equations for $\beta_j\hat{}$, for j=1,...,m;

$$\sum_{j=1}^{m} \beta_j\hat{}\, x_{jki} = y_{ki} \qquad i=1,...,m$$

Stop.

After discussing the properties of algorithm 4, we will give more expository steps of computation for this algorithm. By using these properties we can make this algorithm more efficient.

4.3 properties

Before discussing the properties of the proposed algorithms specifically algorithm 4, let us consider the following four parameter model L_1 norm objective function,

$$S = \sum_{i=1}^{n} |y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} - \beta_3 x_{3i}| \qquad (66)$$

Some steps should be taken to reduce (66) to a weighted median problem by entering three arbitrary integers k1, k2, k3 between one and n as follows. Note that function (1) has been slightly redefined to include intercept as a simple term, in equation (66). Similar to the transformations (63) through (65) the following steps are taken sequentially.

Deviate all observations from $k1^{th}$ observation,

$$y_i^{k1} = y_i - y_{k1} \qquad i=1,\ldots,n \qquad (67)$$

$$x_{ji}^{k1} = x_{ji} - x_{jk1} \qquad i=1,\ldots,n;\ j=1,2,3 \qquad (68)$$

The objective function becomes,

$$S_{k1} = \sum_{i=1}^{n} |y_i^{k1} - \beta_1 - \beta_2 x_{2i}^{k1} - \beta_3 x_{3i}^{k1}| \qquad (69)$$

Divide y_i^{k1} and x_{ji}^{k1} for j=2,3 by x_{1i}^{k1} for all i=1,...,n;

$$y_i^{s1} = y_i^{k1}/x_{1i}^{k1} \qquad i=1,\ldots,n \qquad (70)$$

$$x_{ji}^{s1} = x_{ji}^{k1}/x_{1i}^{k1} \qquad i=1,\ldots,n;\ j=2,3 \qquad (71)$$

The objective function becomes,

$$S_{k1} = \sum_{i=1}^{n} |x_{1i}^{k1}||y_i^{s1} - \beta_1 - \beta_2 x_{2i}^{s1} - \beta_3 x_{3i}^{s1}| \qquad (72)$$

Deviate y_i and x_{ji}^{s1} for j=2,3 from $k2^{th}$ observation,

$$y_i^{k2} = y_i^{s1} - y_{k2}^{s1} \qquad i=1,\ldots,n \qquad (73)$$

$$x_{ji}^{k2} = x_{ji}^{s1} - x_{jk2}^{s1} \qquad i=1,\ldots,n,\ j=2,3 \qquad (74)$$

The objective function reduces to,

$$S_{k1k2} = \sum_{i=1}^{n} |x_{1i}^{k1}||y_i^{k2} - \beta_2 x_{2i}^{k2} - \beta_3 x_{3i}^{k2}| \quad (75)$$

Divide y_i^{k2} and x_{ji}^{k2} for j=3 by x_{2i}^{k2} for all i=1,..,n;

$$y_i^{s2} = y_i^{k2}/x_{2i}^{k2} \qquad i=1,\ldots,n \quad (76)$$

$$x_{ji}^{s2} = x_{ji}^{k2}/x_{2i}^{k2} \qquad i=1,\ldots,n;\ j=3 \quad (77)$$

The objective function forms as,

$$S_{k1k2} = \sum_{i=1}^{n} |x_{1i}^{k1} x_{2i}^{k2}||y_i^{s2} - \beta_2 - \beta_3 x_{3i}^{s2}| \quad (78)$$

Deviate y_i^{s2} and x_{ji}^{s2} from k3th observation,

$$y_i^{k3} = y_i^{s2} - y_{k3}^{s2} \qquad i=1,\ldots,n \quad (79)$$

$$x_{ji}^{k3} = x_{ji}^{s2} - x_{jk3}^{s2} \qquad i=1,\ldots,n,\ j=3 \quad (80)$$

Rewrite the objective function as,

$$S_{k1k2} = \sum_{i=1}^{n} |x_{1i}^{k1} x_{2i}^{k2} x_{3i}^{k3}||y_i^{s3} - \beta_3 x_{3i}^{k3}| \quad (81)$$

Divide y_i^{k3} by x_{3i}^{k3} for all i=1,...,n;

$$y_i^{s3} = y_i^{k3}/x_{3i}^{k3} \qquad i=1,\ldots,n \quad (82)$$

Finally the objective function becomes,

$$S_{k1k2k3} = \sum_{i=1}^{n} |x_{1i}^{k1} x_{2i}^{k2} x_{3i}^{k3}||y_i^{s3} - \beta_3| \quad (83)$$

Rewrite (83) as,

$$S_{k1k2k3} = \sum_{i=1}^{n} |w_i^{k1k2k3}||r_i^{s1s2s3} - \beta_3| \quad (84)$$

where,

$$w_i^{k1k2k3} = x_{1i}^{k1} x_{2i}^{k2} x_{3i}^{k3} \quad (85)$$

$$r_i^{s1s2s3} = y_i^{s3} \quad (86)$$

The objective function (84) is a weighted median problem which may easily be minimized to find β_3. Equation (84) for the general m parameter model

$$y_i = \sum_{j=0}^{m-1} x_{ji}\beta_j + u_i \qquad i=1,\ldots,n \quad (87)$$

$$x_{0i} = 1 \qquad i=1,\ldots,n \quad (88)$$

with the corresponding objective function,

$$S = \sum_{i=1}^{n} |y_i - \sum_{j=0}^{m-1} x_{ji}\beta_j| \quad (89)$$

is,

$$S_{k1k2\ldots k(m-1)} =$$

$$\sum_{i=1}^{n} |w_i{}^{k1k2\ldots k(m-1)}|\,|r_i{}^{s1s2\ldots s(m-1)} - \beta_{m-1}| \qquad (90)$$

where,

$$w_i{}^{k1k2\ldots k(m-1)} = |x_{1i}{}^{k1} x_{2i}{}^{k2} \ldots x_{(m-1)i}{}^{k(m-1)}| \qquad (91)$$

$$r_i{}^{s1s2\ldots s(m-1)} = y_i{}^{s(m-1)} \qquad (92)$$

Now, we proceed to explain the properties of algorithm 4 by using the formulation described above.

Property 1

If we start the computation with m-1 points which do not belong to the original set of observations, value of S decreases to its global minimum at each iteration. Suppose the points $(y_{n+1},\ldots,x_{m-1,n+1}),\ldots,(y_{n+m-1},\ldots,x_{m-1,n+m-1})$ do not belong to our sample points. If n+1,n+2,...,n+m-1 are assigned to k1,k2,...,k(m-1), then minimizing the function in (90) leads to finding a new point M which belongs to the sample observations. Replacing k1 by M and deleting the $(n+1)^{th}$ point from the basis and minimizing (90) leads to finding another point M' in the sample which will be replaced by the $k2^{th}$ observation which is out of our sample. Thus, in each iteration one of the out of sample points is deleted and optimization procedure will be done on the sample points after m-1 iterations.

This property has a good performance on ill-conditioned data. That is if for example rounding error causes an observation to get incorrect rounded value, then this point in the next iteration will be replaced by a sample point and thus redirect the path of descending to the right points.

Property 2

Reduction of m-1 parameter model to fewer parameters model can be done by using the computation of m-1 parameter model and conversely. Thus stepwise and all possible regressions may be computed efficiently.

To show this property let us deal with equation (86). Value of $r_i{}^{s1s2s3}$ may be computed by substituting (67), (68),

(70), (71), (73), (74), (76), (77), (79), (80) and (82) into (86). The result is,

$$r_i^{s1s2s3} =$$

$$\begin{vmatrix} \dfrac{y_i-y_{k1}}{x_{1i}-x_{1k1}} - \dfrac{y_{k2}-y_{k1}}{x_{1k2}-x_{1k1}} & \dfrac{y_{k3}-y_{k1}}{x_{1k3}-x_{1k1}} - \dfrac{y_{k2}-y_{k1}}{x_{1k2}-x_{1k1}} \\ \dfrac{x_{2i}-x_{2k1}}{x_{1i}-x_{1k1}} - \dfrac{x_{2k2}-x_{2k1}}{x_{1k2}-x_{1k1}} & \dfrac{x_{2k3}-x_{2k1}}{x_{1k3}-x_{1k1}} - \dfrac{x_{2k2}-x_{2k1}}{x_{1k2}-x_{1k1}} \end{vmatrix} \quad (93)$$

$$\begin{vmatrix} \dfrac{x_{3i}-x_{3k1}}{x_{1i}-x_{1k1}} - \dfrac{x_{3k2}-x_{3k1}}{x_{1k2}-x_{1k1}} & \dfrac{x_{3k3}-x_{3k1}}{x_{1k3}-x_{1k1}} - \dfrac{x_{3k2}-x_{3k1}}{x_{1k2}-x_{1k1}} \\ \dfrac{x_{2i}-x_{2k1}}{x_{1i}-x_{1k1}} - \dfrac{x_{2k2}-x_{2k1}}{x_{1k2}-x_{1k1}} & \dfrac{x_{2k3}-x_{2k1}}{x_{1k3}-x_{1k1}} - \dfrac{x_{2k2}-x_{2k1}}{x_{1k2}-x_{1k1}} \end{vmatrix}$$

By inspection, it is obvious that the contents of rectangle grid lines in (93) are r_i^{s1}, r_i^{s1s2} and r_i^{s1s2s3}. Generally, $r_i^{s1s2\ldots s(m-1)}$ has the following scheme,

$$r_i^{s1\ldots s(m-1)} = \begin{vmatrix} r_i^{s1} & & & \\ & r_i^{s1s2} & & \\ & & \ddots & \\ & & & r_i^{s1\ldots s(m-1)} \end{vmatrix} \quad (94)$$

A similar discussion may also be performed for $w_i^{k1k2\ldots k(m-1)}$. Rewrite (91) by substitution of (68), (74) and (80) in (91),

$$w_i^{k1k2k3} = \begin{vmatrix} (x_{1i}-x_{1k1})\left(\dfrac{x_{2i}-x_{2k1}}{x_{1i}-x_{1k1}} - \dfrac{x_{2k2}-x_{2k1}}{x_{1k2}-x_{1k1}}\right) & \\ \left(\dfrac{\dfrac{x_{3i}-x_{3k1}}{x_{1i}-x_{1k1}} - \dfrac{x_{3k2}-x_{3k1}}{x_{1k2}-x_{1k1}}}{\dfrac{x_{2i}-x_{2k1}}{x_{1i}-x_{1k1}} - \dfrac{x_{2k2}-x_{2k1}}{x_{1k2}-x_{1k1}}} - \dfrac{\dfrac{x_{3k3}-x_{3k1}}{x_{1k3}-x_{1k1}} - \dfrac{x_{3k2}-x_{3k1}}{x_{1k2}-x_{1k1}}}{\dfrac{x_{2k3}-x_{2k1}}{x_{1k3}-x_{1k1}} - \dfrac{x_{2k2}-x_{2k1}}{x_{1k2}-x_{1k1}}}\right) & \end{vmatrix} \quad (95)$$

By inspection of (93), (94) and (95) it is obvious that calculation of $r_i^{s1s2\ldots sh}$ and $w_i^{s1s2\ldots sh}$ where $0<h\leq m-1$ which are necessary for h parameter regression may be derived from the m parameter model steps of computation.

Property 3

By any commutation for kj for j=1,...,m-1, values of $r_i^{s1s2...s(m-1)}$ defined by (92) for all i≠k1,k2,...,k(m-1) remain unchanged. This calculation can be made by deleting the suitable terms in (93) and (94) after taking common denominator. Because of this property, we can commute for the values of k1,...,k(m-1) in the process of computation without changing the value of $r_i^{s1...s(m-1)}$.

Property 4

By any commutation for kj for j=1,...,m-1, values of $w_i^{k1...k(m-1)}$ for all i=1,...,n defined by (91) remain unchanged. This conclusion is made by multiplication of sequential parentheses of (97) and deleting the suitable terms. This property accompanying with property 3 guarantees that any commutation for k1,...,k(m-1) does not change the necessary items of (91) and (92) in computation of weighted median problem (90).

Property 5

Value of $r_{k(m-1)}^{s1...s(m-1)}$ is equal to zero. By substituting i=k(m-1) in (93) or (94) this property is concluded. Since, this term is equal to zero we should be aware of divide check (dividing by zero) in the process of computation of $r_{k1}^{s1}, r_{k2}^{s1s2}, \ldots, r_{k(m-1)}^{s1s2...s(m-1)}$.

Property 6

Values of $w_h^{k1k2...k(m-1)}=0$ for all h=k1,k2,...,k(m-1). If h is equal to one of the k1,k2,...,k(m-1), then one of the parentheses of (95) is equal to zero and this property is assured. According to this property to avoid divide check, without computing $w_h^{k1k2...k(m-1)}$ for h=k1,k2,...,k(m-1) we can set these terms equal to zero.

Property 7

By any commutation for kj for j=1,...,(m-1), the Mth

point derived by minimizing (89) remains unchanged. Because, by this commutation values of $r_i^{s1s2...s(m-1)}$ and $w_i^{k1k2...k(m-1)}$ remain unchanged (properties 3,4,5 and 6).

Property 8

Updating procedure to evaluate different subsets of sample observations can be simply done by assigning zero values to their $w_h^{k1k2...k(m-1)}$ in (90) for deleting points. This makes those unwanted points h's be unaffected in calculation.

Property 9

According to properties 6 and 8, since the value of $w_h^{k1k2...k(m-1)}=0$ for all h=k1,k2,...,k(m-1); the point M which derives from solving weighted median problem (90) will not be equal to k1,k2,...,k(m-1). This property guarantees that when we enter an observation into the basis we do not find this point again in the same iteration.

Property 10

If all kj for j=1,...,m-1 are equal in starting time, no problem occurs in the process of computation, but the execution time may be increased. Conversely if the kj points are selected in such a way that they are near the minimum of S, the algorithm converges to the optimal solution with less iterations.

Property 11

Values of $w_i^{k1k2...k(m-1)}$ and $r_i^{s1s2...s(m-1)}$ which are essential to solve the weighted median problem of (90) can be retrieved from their values in the previous iteration. That is when the point kh is discarded from the basis and replaced by kj, values of $w_i^{k1k2...k(h-1)}$ and $r_i^{s1s2...s(h-1)}$ do not change and need not be computed again. Applying this property to algorithm 4 makes it more efficient.

Property 12

Since at each iteration, values of $r_i^{s1s2\ldots s(m-1)}$ and $w_i^{k1k2\ldots k(m-1)}$ should be computed from the source input data **y** and **X**, rounding error does not accumulate and convergence does not perturb. This property makes the algorithm safe from accumulation of rounding error which exists in simplex-type algorithms.

property 13

When m optimal points are reached, the optimal values of $\beta_j\hat{}$ for j=1,...,m can be computed as a recursive system of equations efficiently. Values of $\beta_j\hat{}$ will be,

$$\begin{bmatrix} \beta_{m-1}\hat{} \\ \beta_{m-2}\hat{} \\ \vdots \\ \beta_2\hat{} \\ \beta_1\hat{} \\ \beta_0\hat{} \end{bmatrix} = \begin{bmatrix} y_M^{s(m-1)} \\ y_{k(m-1)}^{s(m-2)} \\ \vdots \\ y_{k3}^{s2} \\ y_{k2}^{s1} \\ y_{k1} \end{bmatrix} - \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 & 0 \\ x_{(m-1)k(m-1)}^{s(m-2)} & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ x_{(m-1)k3}^{s2} & x_{(m-2)k3}^{s2} & \cdots & 0 & 0 & 0 \\ x_{(m-1)k2}^{s1} & x_{(m-2)k3}^{s1} & \cdots & x_{2k2}^{s1} & 0 & 0 \\ x_{(m-1)k1} & x_{(m-2)k1} & \cdots & x_{2k1} & x_{1k1} & 0 \end{bmatrix} \begin{bmatrix} \beta_{m-1}\hat{} \\ \beta_{m-2}\hat{} \\ \vdots \\ \beta_2\hat{} \\ \beta_1\hat{} \\ 0 \end{bmatrix} \quad (96)$$

The recursive system of equations (96) can be obtained as follows. Solve (67) and (68) for y_i and x_{ji}, j=1,2,3; and substitute them into (66), gives

$$S_{k1} = \sum_{i=1}^{n} \quad (97)$$

$$|y_i^{k1} - \beta_1 x_{1i}^{k1} - \beta_2 x_{2i}^{k1} - \beta_3 x_{3i}^{k1} - (y_{k1} - \beta_0 - \beta_1 x_{2k1} - \beta_3 x_{3k1})|$$

Since we force the regression to pass through the point k1, the expression inside the parentheses of (97) is equal to zero. Hence,

$$y_{k1} - \beta_0 - \beta_1 x_{2k1} - \beta_3 x_{3k1} = 0 \quad (98)$$

We can solve (73) and (74) again for y_i^{s1} and x_{ji}^{s1}, where, j=2,3 and substitute them into (72) which leads to the following objective function,

$$S_{k1k2} = \sum_{i=1}^{n} |x_{1i}{}^{k1}|$$

$$|y_i{}^{k2}-\beta_2 x_{2i}{}^{k2}-\beta_3 x_{3i}{}^{k2}-(y_{k2}{}^{s1}-\beta_1-\beta_2 x_{2k2}{}^{s1}-\beta_3 x_{3k2}{}^{s1})| \quad (99)$$

By a similar discussion on (97) we have,

$$y_{k2}{}^{s1}-\beta_1-\beta_2 x_{2k2}{}^{s1}-\beta_3 x_{3k2}{}^{s1} = 0 \quad (100)$$

Similarly we can derive,

$$S_{k1k2k3} = \sum_{i=1}^{n} |x_{1i}{}^{k1} x_{2i}{}^{k2}|$$

$$|y_i{}^{k3}-\beta_3 x_{3i}{}^{k3}+(y_{k3}{}^{s2}-\beta_2-\beta_3 x_{3k3}{}^{s2})| \quad (101)$$

and based on (101),

$$y_{k3}{}^{s2}-\beta_2-\beta_3 x_{3k3}{}^{s2} = 0 \quad (102)$$

Finally, by solving weighted median problem (83) we find a point M which gives,

$$\beta_3 = y_M{}^{s3} \quad (103)$$

Equations (103), (102), (100) and (98) can be solved sequentially for $\beta_3\hat{},\ldots,\beta_0\hat{}$. In a similar fashion the general solution for m parameter model is given by (96).

Now let us use these properties and give the computational steps of algorithm 4 more expository in the following BL1 program. This program in the next chapter will be used in comparison to other algorithms for general linear L_1 norm regression.

PROGRAM BL1

Step 0) Initialization.

Parameter: n, m, m1=m-1, m2=m-2.

Real: y(n), x(n,m1), xsk(m1), yw(n), xkw(n), w(n), ys(n), xs(n,m1), b(m), xw(n,2:m1), yso1(m1), xso1(m1,m1).

Integer: 1(n), kk(m1).

Common: /c1/i1,i2.

Read: (y(i),(x(i,j),j=1,m1),i=1,n).

Let: iter=0, kr=0, mm=1, (kk(j)=arbitrary,j=1,m1).

Step 1) Refill working arrays.

Do loop for i=1,n: ys(i)=y(i), do loop for j=1,m1: xs(i,j)=x(i,j), end do, end do.

Step 2) Store weights and ratios for next iteration.
Do loop for i=1,n: w(i)=xkw(i), ys(i)=yw(i), do loop for j=1,m1: xs(i,j)=xw(i,j), end do, end do.

Step 3) Compute the arguments for weighted median.

a. Set: jj=mm, k=kk(jj), ysk=ys(k), i1=1, i2=k-1.
Do loop for j=jj,m1: xsk(j)=xs(k,j), end do.

b. Do loop for j=jj,m1: call COL1(xsk(j),xs(1,j)) end do.
Call COL2(ysk,jj,w,ys,xs(1,jj)).
If i2=n go to step 3.c; otherwise set: i1=k+1, i2=n, go to step 3.b.

c. Set: w(k)=0.
If jj=m1 go to step 4; otherwise i1=1, i2=k-1, go to step 3.d.

d. Do loop for j=jj+1,m1: call COL3(xs(1,j),xs(1,jj)), end do.
If i=n go to step 3.e; otherwise i1=k+1, i2=n, go to step 3.d.

e. If jj≠mm jj=jj+1, go to step 3, otherwise do loop for i=1,n: xkw(i)=w(i), yw(i)=ys(i); do loop for j=jj+1,m1: xw(i,j)=xs(i,j), end do; end do.
Set: jj=jj+1, go to step 3.

Step 4) Compute the weighted median.
Set: ys(k)=0, iter=iter+1, lm=LWMED(n,ys,w,l).

Step 5) Test for optimality.
If lm=kr go to step 5.b; otherwise iopt=0 go to step 5.a.

a. If mm=m1 set mm=1, kr=kk(mm), kk(mm)=lm, go to step 1; otherwise set mm=mm+1, kr=kk(mm), kk(mm)=lm, go to step 2.

b. Set: iopt=iopt+1.
If iopt≠m1 go to step 5.a, otherwise go to step 6.

Step 6) Compute the solution.
Set: b(m)=ys(lm).
Do loop for i=1,m1: ysol(i)=y(kk(i)); do loop for j=1,m1: xsol(i,j)=x(kk(i),j), end do; end do.

```
      Set: jj=1.
   a. Set: ysk=ysol(jj).
      Do loop for j=jj,m1: xsx(j)=xsol(jj,j), end do.
      Do loop for i=jj,m1: if i=jj go to continue;
      otherwise ysol(i)=ysol(i)-ysk, do loop for j=jj,m1:
      xsol(i,j)=xsol(i,j)-xsk(j), end do; set
      ysol(i)=ysol(i)/xsol(i,jj), continue, end do.
   b. Do loop for i=jj,m1: if i=jj go to continue,
      otherwise, do loop for j=jj+1,m1:
      xsol(i,j)=xsol(i,j)/xsol(i,jj), end do; continue;
      end do.
   c. If jj=m2 go to step 6.d; otherwise go to step 6.a.
   d. Do loop for i=1,m2: k=m-i, s=ysol(k); do loop for
      j=k,m1, s=s-b(j+1)*xsol(k,j) end do, b(k)=s, end do.
      Set: s=y(kk(1)).
      Do loop for j=1,m1: s=s-b(j+1)*x(kk(1),j), b(1)=s,
      end do.
      Print: ((b(j),j=1,m),(kk(j),j=1,m1),lm,iter).
      Stop.
END
```

The major portion of computation in this program is transformation of two dimensional arrays. Passing columns of these arrays to other subroutines which involves only one dimensional arrays saves the time of computation (see, Barrodale and Roberts (1974)). Subroutine COL1, COL2 and COL3 have been coded to do this task for subtraction, multiplication and division, and for only division respectively. Function LWMED which is used to compute weighted median has been introduced in section 2.1.

```
SUBROUTINE COL1(v1,v2)
Step 0) Initialization
        Real: v2(1).
        Common /c1/i1,i2.
Step 1) Subtraction.
```

```
        Do loop for i=i1,i2: v2(i)=v2(i)-v1, end do.
        Return.
END

SUBROUTINE COL2(ysk,jj,v1,ys,v2)
Step 0) Initialization.
        Real: v1(1),v2(1),ys(1).
        Common /c1/i1,i2.
Step 1) Compute weights and ratios.
        If jj≠1 go to step 1.a,; otherwise do loop for
        i=i1,i2: v1(i)=v2(i), ys(i)=(ys(i)-ysk)/v2(i).
        Return.
     a.   Do   loop   for   i=i1,i2:   v1(i)=v1(i)*v2(i),
        ys(i)=(ys(i)-ysk)/v2(i), end do.
        Return.
END

SUBROUTINE COL3(v1,v2)
Step 0) Initialization.
        Real: v1(1),v2(1),ys(1).
        Common /c1/i1,i2.
Step 1) Division.
        Do loop for i=i1,i2: v1(i)=v1(i)/v2(i), end do.
        Return.
END
```

4.4 Initial value problem

Selection of the arbitrary points k1,k2,...,k(m-1) at the first stage of computation has important role in number of iterations necessary to reach the optimal solution. One way to select these m-1 points may be based on applying another estimator to guess those points which their residuals are smallest in absolute values among all estimated residuals. That is one estimator for example least squares is applied on the data and smallest m-1 residuals in absolute values are selected and the corresponding points to these

residuals are nominated for starting algorithms 1, 2, 3 or 4. However, we examined this procedure for algorithms 2 and 4 and found that it makes them more efficient. Details of the experiments will not be reported here.

CHAPTER IV

COMPARISON OF THE ALGORITHMS

1. Introduction

The objective of this chapter is to compare some of the algorithms described in chapter II with those proposed in chapter III. Our point of view is to compare the accuracy and relative efficiencies of them. In this respect accuracy of the solution of the algorithms is more important than the other criteria. By the term accuracy we mean, reaching the correct solution in finite number of steps or iterations. By efficiency, we mean that the algorithm performs with smaller amount of required storage and execution time to reach accurate optimal solution.

In section 2 design of experiments is discussed. The generation of necessary data, specification of the models with corresponding number of parameters and observations and computing environment are subjects of this section. Section 3 compares L_1 norm algorithms for the simple linear regression. In this case algorithm 2 of previous chapter is compared with the algorithm of Josvanger and Sposito (1983). In section 4, L_1 norm algorithms for multiple linear regression will be considered. In this respect, algorithm 4 of chapter III is compared with algorithms of Barrodale and Roberts (1973), Armstrong and Frome and Kung (1979) and Bloomfield and Steiger (1980) which are more accurate and efficient than the other algorithms cited in chapter II. In the last section conclusions are made.

2. Design of experiments

Performance of every algorithm in any specific computing environment is different and thus makes the absolute comparison of algorithms very difficult, especially if the system uses, virtual or real storage, a cache or any array processor, mathematical co-processor and etc. As it was discussed in chapter II, many algorithms exist for L_1 norm regression with corresponding computer program and comparison of all of them is very costly. In order to reduce the number of experiments, we rely on the experience of previous

researchers which were discussed in chapter II. However, the experiments are divided into two general categories of simple and multiple linear L_1 norm regressions.

Despite of the coded computer programs, computing environment, numbers of observations and parameters of the model and "condition" of data are the major sources of comparisons for performances of algorithms. Thus different sizes problems are to be tested in this section.

To judge the superiority of algorithms there are many criteria. Accuracy and efficiency are basic ones. In the former, we are concerned with obtaining the true results in different samples and in the latter the computation time and storage requirement of the algorithms are compared.

To perform the experiments, once uniform random values selected for β_j in the following model,

$$y_i = \sum_{j=1}^{m} \beta_j x_{ij} + u_i \qquad i=1,\ldots,n \qquad (1)$$

Random values generated for x_{ij} and u_i with five specifications of distributions. Uniform and normal random generators (given by Mojarrad (1977)) used to generate three uniform and two normal sets of random data for each experiment. Generated uniform random deviates belong to the [-10,10], [-100,100] and [-1000,1000] intervals. Normal deviates have zero mean with 100 and 1000 variances. Values of y_i computed for β_j, x_{ij}, and u_i which had been generated as explained above. Values of 20, 50, 100, 500, 1000, 2000, 5000 and 10000 were used for number of observations n and values of 2, 3, 4, 5, 7 and 10 selected for number of parameters m.

Hence, for all of the five specifications of distribution of u_i, and for all m and n, a replication is done for each of the selected algorithms. Average and range of these five replications are reported for each m and n for each algorithm. In the case of simple regression number of replications is ten than five.

The programs were all complied by Fortran IV, VS

compiler, 1.3.0 level (may 1983) and 77 LANGLVL with 03 optimization level to reduce the coding inefficiencies. The programs were run on BASF 7.68 (MVS) computer. Since this machine is a multitasking system, swapping process should affect the execution time. When system is running for more than one job, this effect can not be measured and removed completely. In order to filter the swapping time, Service Request Block (SRB) time has been reduced from the total Central Processing Unit (CPU) time. However, when system is busy this may not exhaust all the swapping times. It has been tried to run all comparable algorithms simultaneously, and also in one class of input with enough initiators and the same priority level to cause similar situations for all comparable submitted jobs. The pre-execution times of compilation and linkage-editor are excluded from all tested programs.

3. Comparison of the simple regression L_1 norm algorithms

In this study comparisons are limited to the algorithm 2 of previous chapter and that proposed by Josvanger and Sposito (1983). Gentle and Narula and Sposito (1987) and Gentle and Sposito and Narula (1988) introduced the latter as the most efficient algorithm for simple linear L_1 norm regression. Their studies compare the algorithms of Karst (1958) with code of Sadovski (1974), Armstrong and Kung (1978), Armstrong and Frome and Kung (1979), Bloomfield and Steiger (1980), Wesolowsky (1981) and Josvanger and Sposito (1983).

Table 1. Array storage requirement for simple model selected algorithms

Algorithm	Program name	Storage requirement	Stopping constant(s)
JS	DESL1	5n	TOL = 1.0E-6
B(Alg.2)	BL1S	5n	----------

Js ≡ Josvanger and Sposito (1983).
B(Alg.2) ≡ Bidabad (Algorithm 2) chapter III.
n ≡ Number of observations.

The amount of array storage requirement for these two programs are shown by table 1. This table may be compared with table 2 of chapter II for other algorithms. None of the programs destroys the input data. Both programs have been coded in single precision.

Table 2 shows the results of the experiments for simple linear L_1 norm regression. The values reported in the cells of the table are the averages of ten replications CPU times in seconds with different random samples. The values in the parentheses are the corresponding minimum and maximum CPU times of the ten runs. Both algorithms converged and gave accurate results for all of the experiments.

As it is clear from table 2 in small samples the computation times are not very different, though algorithm 2 is faster. In medium samples, this difference becomes significant and in larger samples, algorithm 2 becomes strongly superior to that of Josvanger and Sposito (1983). Thus it can be concluded that algorithm 2 performs better than the other algorithms and may be used for applied work to achieve more efficiency.

Table 2. CPU times for simple model L_1 norm selected algorithms

n	JS	B(Alg.2)
20	0.096 (0.09,0.10)	0.094 (0.09,0.10)
50	0.109 (0.10,0.11)	0.106 (0.10,0.11)
100	0.141 (0.13,0.16)	0.132 (0.12,0.14)
500	0.469 (0.35,0.59)	0.360 (0.33,0.38)
1000	0.997 (0.66,1.31)	0.645 (0.61,0.69)
2000	2.770 (1.36,4.27)	1.194 (0.98,1.28)
5000	11.554 (4.58,18.91)	2.848 (2.71,3.15)
10000	42.406 (9.67,60.88)	5.823 (5.16,6.87)

See table 1 for abbreviations.

4. Comparison of the multiple regression L_1 norm algorithms

To compare algorithm 4 of chapter III with other algorithms, experiments have been limited to three algorithms which are more accurate and efficient among the others. These are algorithms of Barrodale and Roberts (1973,74) (BR), Bloomfield and Steiger (1980) (BS), Armstrong and Frome and Kung (1979) (AFK). Although, BS and AFK algorithms are faster than BR, the reason to select BR algorithm was that the other two algorithms failed to produce correct answers for larger samples (see, Gentle and Narula and Sposito (1987) and also chapter II).

The amount of array storage requirement for these programs are indicated by table 3. This table may be compared with table 2 of chapter II for other algorithms. All programs have been coded in single precision. None of the programs destroys input data.

Table 3. Array storage requirement for multiple model selected algorithms

Algorithm	Program name	Storage requirement	Stopping constant(s)
AFK	AFKL1	6n+m(n+3m/2+15/2)	ACU = 1.0E-6 BIG = 1.0E+15
BR	L1BAR	3n+m(n+5)+4	BIG = 1.0E+75
BS	BLOD1	4n+m(2n+4)	----------
B(Alg.4)	BL1	2n+m(3n+m+2)-2	----------

AFK ≡ Armstrong and Frome and Kung (1970).
BR ≡ Barrodale and Roberts (1973,74).
BS ≡ Bloomfield and Steiger (1980).
B(Alg.4) ≡ Bidabad (Algorithm 4) chapter III.
n ≡ Number of observations.
m ≡ Number of parameters.

Tables 4 through 8 report the averages of five runs CPU times for different sample sizes and parameters. The values in the parentheses are minimum and maximum CPU times of replications. For the three parameter model, as it can be seen from table 4, the algorithm 4 is superior to other algorithms. In this case the BS, AFK, and BR possess less efficiency respectively. When the sample size is small, the difference is not large. In medium sample sizes this

difference is going to increase. In larger size experiments, algorithm 4 and BS have small difference, but BR and AFK are far from them. In all cases algorithm 4 is faster than the other algorithms.

Table 4. CPU times for multiple model (m=3) selected algorithms

n	B(Alg.4)	BR	BS	AFK
20	0.098 (0.09,0.10)	0.110 (0.11,0.11)	0.104 (0.10,0.11)	0.112 (0.11,0.12)
50	0.144 (0.13,0.16)	0.148 (0.14,0.16)	0.146 (0.13,0.16)	0.146 (0.14,0.15)
100	0.182 (0.18,0.19)	0.216 (0.20,0.23)	0.194 (0.19,0.20)	0.214 (0.20,0.23)
500	0.698 (0.63,0.74)	1.116 (1.07,1.17)	0.810 (0.63,1.00)	0.986 (0.85,1.11)
1000	1.390 (1.27,1.53)	2.420 (2.16,2.76)	1.662 (1.37,2.01)	2.180 (2.01,2.36)
2000	2.812 (2.34,2.99)	5.884 (4.98,6.63)	2.932 (2.81,3.18)	4.800 (4.30,5.09)
5000	7.456 (6.82,9.03)	25.038 (22.33,27.54)	7.520 (6.16,10.12)	20.172 (16.57, 22.33)
10000	14.330 (12.45,16.61)	80.008 (73.16,87.03)	15.434 (12.88, 18.12)	59.634 (55.60, 65.80)

See table 3, for abbreviations.

In the case of four parameter model as shown by table 5, though BS algorithm is competing with algorithm 4, this ordering remains unchanged and algorithm 4 is again most efficient. The ranking of the selected algorithms are similar to that of three parameter experiments in all cases of small, medium and larger sample sizes.

When number of parameters increased to five, BS algorithm failed to produce correct answers for sample sizes of 2000 and more. Gentle and Narula and Sposito (1987) also referred to failure of BS algorithm for sample sizes of 1000 and greater for five and more parameter models and for sample size of 5000 when the number of parameters are two (see also chapter II). With reference to table 6, efficiency of algorithm 4 to others with respect to failure of BS is clear. The algorithms of AFK and BR are in the next positions

Table 5. CPU times for multiple model (m=4) selected algorithms

n	B(Alg.4)	BR	BS	AFK
20	0.112 (0.11,0.12)	0.116 (0.11,0.12)	0.116 (0.11,0.12)	0.116 (0.11,0.12)
50	0.156 (0.15,0.16)	0.168 (0.16,0.19)	0.160 (0.15,0.17)	0.160 (0.16,0.16)
100	0.284 (0.26,0.32)	0.286 (0.26,0.30)	0.286 (0.27,0.30)	0.286 (0.26,0.30)
500	1.098 (0.85,1.44)	1.596 (1.23,1.77)	1.260 (0.92,1.58)	1.394 (1.17,1.71)
1000	2.194 (2.03,2.45)	4.016 (3.35,5.05)	2.200 (0.64,3.28)	3.022 (2.73,3.21)
2000	4.650 (3.92,5.05)	10.636 (9.44,11.86)	5.430 (4.54,6.15)	7.774 (7.25,8.16)
5000	12.852 (10.00,15.23)	41.282 (32.40,47.86)	12.938 (11.62, 14.04)	32.110 (29.55, 33.17)
10000	27.720 (22.93,39.14)	119.152 (107.14, 129.70)	27.864 (23.94, 32.10)	101.472 (100.31, 103.15)

See table 3, for abbreviations.

Table 6. CPU times for multiple model (m=5) selected algorithms

n	B(Alg.4)	BR	BS	AFK
20	0.138 (0.13,0.15)	0.124 (0.12,0.13)	0.124 (0.12,0.14)	0.124 (0.11,0.13)
50	0.208 (0.18,0.24)	0.240 (0.22,0.26)	0.204 (0.20,0.21)	0.188 (0.18,0.20)
100	0.348 (0.33,0.37)	0.380 (0.34,0.42)	0.404 (0.36,0.46)	0.338 (0.32,0.36)
500	2.024 (1.57,2.40)	2.498 (2.29,2.68)	1.754 (1.34,2.01)	1.684 (1.57,1.78)
1000	3.702 (3.19,4.45)	5.684 (4.86,6.28)	3.364 (2.93,3.71)	3.876 (3.67,4.13)
2000	8.770 (7.51,9.41)	15.500 (13.04,16.58)	+ +	9.120 (7.91,9.93)
5000	24.418 (20.15,27.96)	66.394 (59.93,71.90)	+ +	36.600 (33.95, 38.67)
10000	53.924 (38.35,64.90)	244.072 (217.81, 270.06)	+ +	108.406 (99.65, 119.75)

+ Failed to compute correct answers.
See table 3, for abbreviations.

Table 7. CPU times for multiple model (m=7) selected algorithms

n	B(Alg.4)	BR	BS	AFK
20	0.160 (0.16,0.16)	0.164 (0.16,0.17)	0.156 (0.15,0.16)	0.160 (0.15,0.17)
50	0.346 (0.29,0.37)	0.302 (0.29.0.31)	0.328 (0.31,0.34)	0.282 (0.26,0.30)
100	0.706 (0.58,0.81)	0.572 (0.52,0.65)	0.540 (0.48,0.62)	0.530 (0.47,0.62)
500	3.898 (3.25,4.54)	4.486 (4.03,4.85)	3.202 (2.41,3.87)	2.990 (2.45,3.48)
1000	9.178 (7.03,10.44)	11.448 (9.88,12.71)	+ +	6.568 (5.87,7.68)
2000	19.984 (18.06,21.55)	31.872 (24.52,35.22)	+ +	14.908 (13.56, 16.23)
5000	57.286 (49.53,64.24)	141.234 (129.20, 154.03)	+ +	58.314 (49.12, 65.17)
10000	130.79 (103.20, 183.45)	475.826 (421.40, 521.39)	+ +	151.526 (144.94, 165.62)

+ Failed to compute correct answers.
See table 3, for abbreviations.

Table 8. CPU times for multiple model (m=10) selected algorithms

n	B(Alg.4)	BR	BS	AFK
20	0.276 (0.27,0.29)	0.212 (0.20,0.23)	0.218 (0.21,0.22)	0.212 (0.20,0.23)
50	0.956 (0.69,1.72)	0.502 (0.47,0.54)	0.492 (0.43,0.54)	0.398 (0.35,0.44)
100	2.414 (1.55,4.78)	1.144 (1.03,1.25)	0.998 (0.86,1.11)	0.776 (0.66,0.91)
500	13.980 (11.76,16.30)	8.446 (7.88,9.99)	5.970 (4.81,6.92)	5.210 (4.43,5.78)
1000	62.624 (22.23, 193.19)	23.506 (20.26,26.93)	+ +	11.144 (9.08,13.14)
2000	109.268 (72.20, 278.27)	62.756 (59.43,65.66)	+ +	+ +
5000	409.438 (154.64, 1010.92)	284.618 (240.02, 322.47)	+ +	+ +
10000	679.540 (283.04, 1076.04)	967.794 (770.79, 1064.43)	+ +	+ +

+ Failed to compute correct answers.
See table 3, for abbreviations.

respectively. For smaller sample size, BR, BS and AFK algorithms are competing, but the differences are very small. In larger sample size, algorithm 4 becomes strictly superior to other algorithms.

In table 7 when number of parameters is seven, BS algorithm failed to compute correct answer for sample of sizes 1000 and more. AFK is the best for smaller samples, but for large samples, algorithm 4 is again superior. BR algorithm is in the third position.

In table 8, with ten parameters, BS and AFK algorithms failed to compute correct answers for larger sample size. BR algorithm is the most efficient with respect to accuracy. Algorithm 4 remains in the second position of both computing time and accuracy, except for sample size of 10000, where algorithm 4 is the most efficient.

5. Conclusions

Since in computational algorithms, accuracy is more important than efficiency, those L_1 norm algorithms should be selected which produce correct solutions and among them, the fastest one should be selected. Algorithm 2 and algorithm of Josvanger and Sposito (1983) both computed correct answers for two parameter linear L_1 norm regression model. Algorithm 2 which is faster than JS introduced for applied works.

For multiple regression, BS and AFK algorithms failed to compute correct answers in larger models. As stated by Gentle and Narula and Sposito (1987), because of the accumulated roundoff error, algorithm of Bloomfield and Steiger (1980) was not usable in larger size problems. Coding to avoid rounding problems often increase the execution time, so it is not clear what would happen to the relative efficiency if the BS code is modified. This is also the case for algorithm of Armstrong and Frome and Kung (1979), though it is less sensitive to rounding error than BS algorithm. However, from the previous tables it may be concluded that algorithm 4 is more appropriate for models with less than ten parameters and

algorithm of Barrodale and Roberts (1973,74) for the ten parameter model. This last conclusion is not very constructive, because in the case of ten parameter model with 10000 observations algorithm 4 is highly superior to BR. However, since in applied work we are not always confronted with very large amount of data and parameters, this conclusion is poor in operational sense.

This page has been intentionally left blank

CHAPTER V

CONTINUOUS L_1 NORM SMOOTHING AND LORENZ CURVE

1. Introduction

In the previous chapters, the L_1 norm estimation of discrete data was discussed and the corresponding topics considered. In this chapter we intend to use some of the techniques described before to handle the continuous data in L_1 norm smoothing of functions. Application of this type of data in econometric estimation problem is new. However, we are going to develop this technique and apply it to estimation of simple concentration surface, namely, Lorenz curve function which has many applications in economics and other related sciences.

In section 2 we proceed by defining the L_1 norm of continuous functions and corresponding estimation or smoothing or approximation specification of the objective function of regression with stochastic term and unknown parameters. In the next sections, the L_1 norm estimation problem of one and two parameter linear models in continuous case are solved. Before going to use these techniques in estimation of Lorenz curve, it would be appropriate to have a look at the theory of concentration surfaces. Thus, in section 5 the definition of concentration surface and especially Lorenz curve will be given and its proposed functional forms will be discussed. Section 6 is devoted to continuous L_1 norm approximation of Lorenz curve. In this section, we proceed to clarify how we can use the information of the probability distribution function of income to exactly determine the L_1 norm approximation of the corresponding Lorenz curve of the statistical population under consideration.

2. L_1 norm of continuous functions

Generally, L_p norm of a function f(x) (see, Rice and White (1964)) is defined by,

$$||f(x)||_p = (\int_{x \in I} |f(x)|^p dx)^{1/p} \qquad (1)$$

where I is a closed bounded set. The L_1 norm of f(x) is simply written as,

$$||f(x)||_1 = \int_{x\in I} |f(x)|dx \qquad (2)$$

Suppose that the nonstochastic function of x, $f(x,\beta)$, is combined with stochastic term u to form y(x) as follows,

$$y(x) = f(x,\beta) + u \qquad (3)$$

where β is unknown parameters vector of the function f. Rewriting u as residual of $y(x)-f(x,\beta)$, in L_1 norm approximation of β we should find β vector such that the L_1 norm of u is minimized. That is,

$$\min_{\beta}: S=||u||_1=||y(x)-f(x,\beta)||_1=\int_{x\in I} |y(x)-f(x,\beta)|dx \qquad (4)$$

3. Linear one parameter L_1 norm smoothing

Redefine $f(x,\beta)$ as the following linear function,

$$y(x) = \beta x + u \qquad (5)$$

Expression (4) reduces to

$$\min_{\beta}: S = ||u||_1 = ||y(x)-\beta x||_1 = \int_{x\in I} |y(x)-\beta x|dx \qquad (6)$$

In chapter III the discrete analogue of (6) was described and with referencing to the Laplace's procedure of weighted, median computation of the optimal value of β was explained. In another procedure we proposed, the use of discrete and regular derivatives with a slack variable t as a point to distinguish negative and positive residuals before any further computation. A similar approach is used again in minimizing (6). To do so in this case certain Lipschitz conditions are imposed on the functions involved (see, Usow (1967a), Dieudonne (1963)). Rewrite (6) as follows,

$$\min_{\beta}: S = \int_{x\in I} |x||y(x)/x - \beta|dx \qquad (7)$$

For convenience, define I as a closed interval [0,1]. The procedure may be applied to other intervals with no major problem (see, Usow (1967a), Hobby and Rice (1965), Kripke and Rivlin (1965)). Since x belongs to closed interval I, y(x) which is a linear function of x and also y(x)/x are smooth and continuous. Thus, since y(x)/x is uniformly increasing or

decreasing function of x, a value of $t \in I$ can be found to have the following properties,

$$\begin{aligned} y(x)/x &< \beta \qquad \text{if } x < t \\ y(x)/x &= \beta \qquad \text{if } x = t \\ y(x)/x &> \beta \qquad \text{if } x > t \end{aligned} \tag{8}$$

According to (8), we can rewrite (7) as two separate definite integrals with different upper and lower bounds.

$$\min_{\beta}: S = -\int_0^t |x|(y(x)/x - \beta)dx + \int_t^1 |x|(y(x)/x - \beta)dx \tag{9}$$

To solve (9), partially differentiate it with respect to t and β variables.

$$\frac{\delta S}{\delta \beta} = \int_0^t |x|dx - \int_t^1 |x|dx = 0 \tag{10}$$

and using Liebniz' rule for differentiating the integrals with respect to their variable bounds t, yields,

$$\frac{\delta S}{\delta t} = -|t|[\frac{y(t)}{t} - \beta] - |t|[\frac{y(t)}{t} - \beta] = 0 \tag{11}$$

Since x belongs to [0,1], equation (10) can be written as,

$$\int_0^t x dx - \int_t^1 x dx = 0 \tag{12}$$

or

$$\tfrac{1}{2} x^2 \Big]_0^t - \tfrac{1}{2} x^2 \Big]_t^1 = 0 \tag{13}$$

or,

$$\tfrac{1}{2} t^2 - \tfrac{1}{2} + \tfrac{1}{2} t^2 = 0 \tag{14}$$

which yields,

$$t = \sqrt{2}/2 \tag{15}$$

Substitute for t in equation (11), yields,

$$-2|\sqrt{2}/2|[\frac{y(\sqrt{2}/2)}{\sqrt{2}/2} - \beta] = 0 \tag{16}$$

or,

$$\beta = \frac{y(\sqrt{2}/2)}{\sqrt{2}/2} \tag{17}$$

Remember that y(t) is function y(x) evaluated at x=t. Value of β given by (17) is the optimal solution of (6). The above procedure actually is the weighted median derivation of the continuous case.

Before discussing the application of this procedure in Lorenz curve, let us develop the procedure for the two parameter linear model L_1 norm estimation of continuous data.

4. Linear two parameter L_1 norm smoothing

Now, we try to apply the above technique to the linear two parameter model. Rewrite (4) as,

$$\min_{\alpha,\beta}: S=||u||_1=||y(x)-\alpha-\beta x||_1=\int_{x\in I} |y(x)-\alpha-\beta x|dx \qquad (18)$$

According to Rice (1964c), let $f(\alpha^*,\beta^*,x)$ interpolates $y(x)$ at the set of canonical points $\{x_i;i=1,2\}$, if $y(x)$ is such that $y(x)-f(\alpha^*,\beta^*,x)$ changes sign at these x_i's and at no other points in $[0,1]$, then $f(\alpha^*,\beta^*,x)$ is the best L_1 norm approximation to $y(x)$ (see also, Usow (1967a)). With the help of this rule, if we denote these points to t_1 and t_2 we can rewrite (18) for $I=[0,1]$ as,

$$S = \int_0^{t_1} [y(x)-\alpha-\beta x]dx - \int_{t_1}^{t_2} [y(x)-\alpha-\beta x]dx + \int_{t_2}^{1} [y(x)-\alpha-\beta x]dx \qquad (19)$$

Since t_1 and t_2 are also unknowns, we should minimize S with respect to α, β, t_1 and t_2. Taking partial derivative of (19) using Liebniz' rule with respect to these variables and equating them to zero, we have,

$$\frac{\delta S}{\delta \alpha} = -\int_0^{t_1} dx + \int_{t_1}^{t_2} dx - \int_{t_2}^{1} dx = 0 \qquad (20)$$

$$\frac{\delta S}{\delta \beta} = -\int_0^{t_1} x dx + \int_{t_1}^{t_2} x dx - \int_{t_2}^{1} x dx = 0 \qquad (21)$$

$$\frac{\delta S}{\delta t_1} = 2[y(t_1) - \alpha - \beta t_1] = 0 \qquad (22)$$

$$\frac{\delta S}{\delta t_2} = -2[y(t_2) - \alpha - \beta t_2] = 0 \qquad (23)$$

Equations (20) through (23) may be solved for α, β, t_1 and t_2 simultaneously. Thus we have the following system of equations,

$$2t_2 - 2t_1 - 1 = 0 \qquad (24)$$

$$t_2^2 - t_1^2 - \tfrac{1}{2} = 0 \qquad (25)$$

$$y(t_1) - \alpha - \beta t_1 = 0 \tag{26}$$

$$y(t_2) - \alpha - \beta t_2 = 0 \tag{27}$$

The solution is,

$$t_1 = 1/4 \tag{28}$$

$$t_2 = 3/4 \tag{29}$$

$$\alpha = y(3/4) - (3/4)\beta = y(1/4) - (1/4)\beta \tag{30}$$

$$\beta = 2[y(3/4) - y(1/4)] \tag{31}$$

The above procedures for simple restricted and unrestricted continuous L_1 norm regression, then, will be applied to computing the L_1 norm estimation of Lorenz curve using the probability distribution of income.

This procedure, similar to that of multiple regression model for discrete case may be expanded to include m unknown parameters. However, it is not discussed here and is left for future research. Some computational methods for solving the different cases of m parameter model are investigated by Ptak (1958), Rice and White (1964), Rice (1964a,b,c,69,85), Usow (1967a), Lazarski (1975a,b,c,77) (see also, Hobby and Rice (1965), Kripke and Rivlin (1965), Watson (1981)). Now, let us have a look at the concentration surfaces and its proposed functional forms for the simple case or Lorenz curve.

5. Concentration surface

The concentration analysis is a method for analyzing skew distributions. Since skew distributions fall within the category of non normal distributions, their links to robust statistics and L_1 norm increase. Income distribution often obeys a log-normal or Pareto like distribution (see, Cramer (1973), Bidabad (1989)) and in this respect relates to concentration surface.

The simple concentration surface or Lorenz curve for a random variable with probability density function f(v) may be defined as the ordered pair,

$$\left(P(V|V\leq v), \frac{E(V|V\leq v)}{E(V)}\right) \qquad v\in R \tag{32}$$

(Taguchi (1972a,b,c,73,81,83,87,88) multiplies the second

element of (32) by $P(V|V\leq v)$ which is not correct. His definition of (33) is equivalent to ours). For a continuous density function f(v), (32) can be written as,

$$\left(\int_{-\infty}^{v} f(w)dw, \frac{\int_{-\infty}^{v} wf(w)dw}{\int_{-\infty}^{+\infty} wf(w)dw}\right) \equiv (y(x(v)),x(v)) \qquad (33)$$

We denote (33) by (y(x(v)),x(v)) where y(x(v)) and x(v) are its elements. Therefore, y is a function which maps x(v) to y(x(v)) and x is a function which maps v to x(v). The function y(x) is simply the Lorenz curve function. For discussions on properties of Lorenz curve see Kendall and Stuart (1977).

Concentration surface has been extended to bivariate and multivariate cases which we are not going to talk about them. For more information see Taguchi (1972a,b,c,73,81,83,87,88). However, in this context we are dealing with approximation of Lorenz curve function using L_1 norm criterion which has many applications in economics.

Income distribution is often portrayed on a Lorenz curve. In recent years some of its functional forms have been introduced. These forms should satisfy some properties, and also make estimation of the function parameters by the good estimating methods simple.

Gupta (1984) defined the following properties. A function y=f(x) represents the Lorenz curve if:

$$\begin{aligned}
&\text{i)} \quad f(0)=0 \\
&\text{ii)} \quad f(1)=1 \\
&\text{iii)} \quad f'(x)\geq 0 \quad \text{for} \quad 0\leq x<1 \\
&\text{iv)} \quad f''(x)\geq 0 \quad \text{for} \quad 0\leq x\leq 1 \\
&\text{v)} \quad f(x)\leq x \quad \text{for} \quad 0<x<1 \\
&\text{vi)} \quad 0\leq \int_0^1 f(x)dx\leq \tfrac{1}{2}
\end{aligned} \qquad (34)$$

It is obvious that the property vi is redundant; when i to v are satisfied. Because by manipulating i to v we have $0\leq f(x)\leq x$. By integrating this inequality, vi is derived,

$$\int_0^1 0dx \leq \int_0^1 f(x)dx \leq \int_0^1 xdx \qquad (35)$$

or,

$$0 \leq \int_0^1 f(x)dx \leq \frac{1}{2} \qquad (36)$$

Therefore property vi is always satisfied, and we need not test it for any candidate function which satisfies i to v.

Bidabad and Bidabad (1989a,b) emphasize on two other characteristics of Lorenz curve which have been neglected. First, Lorenz curve could be asymmetric with respect to the line $y=1-x$, for $0\leq x\leq 1$. This enables that different Lorenz curves cross the others which are the same in functional form and different in parameters for $0<x<1$.

Now, Let us review the proposed functional forms by previous researchers. Kakwani and Podder (1976) proposed the function,

$$M = aN^{\lambda}(\sqrt{2}-N)^{c} \qquad (37)$$

where $M=(x-y)/\sqrt{2}$, $N=(x+y)/\sqrt{2}$, $a\geq 0$, $0\leq\lambda\leq 1$, $0\leq c\leq 1$. This form does not satisfy all the properties of (34).

Rasche and Gaffney and Koo and Obst (1980) proposed the function,

$$y=[1-(1-x)^{a}]^{1/\lambda} \qquad (38)$$

where $0\leq a\leq 1$, $0\leq\lambda\leq 1$. This form makes estimation of the parameters by the least squares method difficult.

Gupta (1984) proposed the functional form,

$$y=xA^{x-1} \qquad (39)$$

where $A>1$. This form satisfies the properties and simply can be estimated by ordinary least squares method, but by changing the parameter A (from A_i to A_j), resulted functions (y_i and y_j) will never intersect for $0<x<1$. To prove this, we can solve the following system for x and y.

$$\begin{aligned} y&=xA_i^{x-1} \\ y&=xA_j^{x-1} \end{aligned} \qquad (40)$$

Solutions are $x=y=0$ and $x=y=1$ which are not in the domain $0<x<1$.

Bidabad and Bidabad (1989a,b) suggest the following functional form which satisfies the properties i to v, and by changing its parameters, resulting curves may cross each other.

$$y=x^{B}A^{x-1} \tag{41}$$

where $B\geq 1$, $A\geq 1$ for $0\leq x\leq 1$. Properties of (34) are satisfied as follows,

$$\begin{aligned}
&\text{i)}\quad f(0)=0\\
&\text{ii)}\quad f(1)=1\\
&\text{iii)}\quad f'(x)=x^{B-1}A^{x-1}[B+x.\ln(A)]\geq 0 \quad \text{for } 0\leq x<1\\
&\text{iv)}\quad f''(x)=x^{B-2}A^{x-1}\{[B+x.\ln(A)]^{2}-B\}\geq 0 \quad \text{for } 0\leq x\leq 1\\
&\text{v)}\quad f(x)=x^{B}A^{x-1}=(x^{B}/A^{1-x})\leq x \quad \text{for } 0<x<1
\end{aligned} \tag{42}$$

Different shapes of the function as $y=x^{Bi}A_i^{x-1}$ and $y=x^{Bj}A_j^{x-1}$ may have intersection for $0<x<1$. By solving the following equations simultaneously, the locus of intersection points with different parameters values of two Lorenz curve is derived.

$$\begin{aligned}
&y=x^{Bi}A_i^{x-1}\\
&y=x^{Bj}A_j^{x-1}
\end{aligned} \tag{43}$$

The solution is,

$$\frac{x-1}{\log(x)}=\frac{B_j-B_i}{\log(A_i/A_j)} \tag{44}$$

It is obvious when (44) is satisfied, there is an intersection between two curves of (43). So if we solve the following system,

$$\begin{aligned}
&x-1=B_j-B_i\\
&\log x=\log(A_i/A_j)
\end{aligned} \tag{45}$$

we can find a relation in terms of A_i, A_j, B_i and B_j which satisfies (44). Therefore,

$$A_i/A_j-1=B_j-B_i \tag{46}$$

Hence, whenever A_i, A_j, B_i and B_j can satisfy equation (46), there is a solution (or intersection) for (43). But the intersection is inside the domain $0<x<1$ when:

$$\begin{aligned}
&0<x=B_j-B_i+1<1\\
&0<x=(A_i/A_j)<1
\end{aligned} \tag{47}$$

or,

$$\begin{aligned}
&0<B_j-B_i<1\\
&0<A_i<A_j
\end{aligned} \tag{48}$$

The second subject that has been emphasized by Bidabad and Bidabad (1989a,b) is the autoregressive nature of the

errors in the Lorenz curve data. On the other hand, when there is an error in the $(t-1)^{th}$ percent of income earners, this error completely will transfer to the next cumulative percent (t). This is because of using cumulative data to estimate the Lorenz curve. So, if we define u_t as disturbance term of the t^{th} observation (cumulative percent), autoregressive specification of the error would be,

$$u_t = u_{t-1} + w_t \tag{49}$$

where, w_t obeys classical assumptions of regression. Therefore the stochastic form of our suggested functional form could be as follow,

$$y_t = x_t^B . A^{x_t - 1} . e^{u_t} \tag{50}$$

We may write (50) for one period lag as,

$$y_{t-1} = x_{t-1}^B . A^{x_{t-1} - 1} . e^{u_{t-1}} \tag{51}$$

Divide (50) by (51) and take natural logarithm, gives,

$$\ln\left(\frac{y_t}{y_{t-1}}\right) = B.\ln\left(\frac{x_t}{x_{t-1}}\right) + \ln(A).(x_t - x_{t-1}) + u_t - u_{t-1} \tag{52}$$

Since, $u_t - u_{t-1}$ is equal to w_t of (49) the problem of autoregression has been discarded and (52) can be estimated by least squares method.

However, to estimate the above functions we should gather discrete data from the statistical population, and manipulate them to construct a relevant data set for estimation procedures. As it was cited above, many errors will enter into the data through accumulation and other necessary manipulations. To avoid these problems we may use another way to estimate the Lorenz curve of our statistical population. However, if the probability distribution of income is known, instead of gathering discrete observations, we can estimate the Lorenz curve by using the L_1 norm approximation method for continuous functions. In the following section we proceed to apply this method to estimate the parameters of the models (39) and (41) by using the information of probability density function of income.

6. Continuous L_1 norm approximation of Lorenz curve

To estimate the Lorenz curve parameters when the information of income probability density function is given, we can not always take straightforward steps. When the probability density function is easily integrable, there is no major problem in advance. We can find the functional relationship between the two elements of (33) by simple mathematical derivation. But, when integrals of (33) are not obtainable, another procedure should be adopted.

Suppose that income of a society is distributed with probability density function f(w). This density function may be a skewed function such as Pareto or log-normal,

$$f(w)=\theta k^{\theta} w^{-\theta-1} \qquad w \geq k>0,\ \theta>0 \tag{53}$$

$$f(w)=[1/w\sigma\sqrt{(2\pi)}]\exp\{-[\ln(w)-\mu]^2/2\sigma^2\} \qquad w\in(0,\infty),\ \mu\in(-\infty,+\infty),\ \sigma>0 \tag{54}$$

These two distributions have been known as good candidates for presenting distribution of personal income (see, Cramer (1973), Bidabad (1989)).

In the case of Pareto density function (53) we can simply derive the Lorenz curve function as follows. Let F(w) denote the Pareto distribution function:

$$F(w)=1-(k/w)^{\theta} \tag{55}$$

with mean,

$$E(w)=\theta k/(\theta-1), \quad \theta>1 \tag{56}$$

If we find the function y as stated by (33) as a function of x, the Lorenz function will be derived. Now, proceed as follow. Rearrange the terms of (33) as,

$$y(x(v))=\int_{-\infty}^{v} f(w)dw \tag{57}$$

$$x(v)=[1/E(x)]\int_{-\infty}^{v} wf(w)dw \tag{58}$$

Substitute Pareto distribution function,

$$y(x(v))=F(v)=1-(k/v)^{\theta} \tag{59}$$

$$x(v)=[(\theta-1)/\theta k]\int_{k}^{v} w\theta k^{\theta} w^{-\theta-1}dw \tag{60}$$

or,

$$x(v)=1-(k/v)^{\theta-1} \tag{61}$$

Now, by solving (61) for v and substituting in (59), the

Lorenz curve for Pareto distribution is derived as,

$$y=1-(1-x)^{\theta/(\theta-1)} \tag{62}$$

It is obvious that (62) satisfies the properties of Lorenz curve given by (34).

As it was shown in the case of Pareto distribution (53), formula of Lorenz curve is easily obtained. But, if we select the log-normal density function (54), the procedure may not be the same. Because the integral of log-normal function has not been derived yet. Thus, we should devise another technique to overcome this problem. One way is using L_p norm approximation or smoothing methods. In the following pages, the L_1 norm smoothing technique will be developed to estimate the parameters of given functional forms (39) and (41) by using the continuous probability density function. Now we proceed as follow.

According to (32) and (33) dependent and independent variables of (39) and (41) may be written as,

$$y(x(v))=\int_0^v f(w)dw \tag{63}$$

$$x(v)=[1/E(x)]\int_0^v wf(w)dw \tag{64}$$

Substitute (63) and (64) inside (39) and define random error term u as,

$$\int_0^v f(w)dw=[1/E(w)]\int_0^v wf(w)dw.A^{[1/E(w)]\int_0^v wf(w)dw-1}e^u \tag{65}$$

or briefly,

$$y(x)=xA^{x-1}e^u \tag{66}$$

Similarly for the model (41),

$$\int_0^v f(w)dw=\{[1/E(w)]\int_0^v wf(w)dw\}^B A^{[1/E(w)]\int_0^v wf(w)dw-1}e^u \tag{67}$$

or briefly,

$$y(x)=x^B A^{x-1}e^u \tag{68}$$

Taking natural logarithm of (66) and (68), gives,

$$\ln y(x)=\ln x + (x-1)\ln A + u \tag{69}$$

$$\ln y(x)=B.\ln x + (x-1)\ln A + u \tag{70}$$

With respect to properties i to v defined by (34) and probability density function of f(w) and equations (65) to

(68), it is obvious that x belongs to the interval [0,1]. Thus the L_1 norm objective function for minimizing (69) or (70) is given by,

$$\min: S = \int_0^1 |u|dx \tag{71}$$

Now, let us deal with L_1 norm estimation of A of Lorenz curve functional form (39) proposed by Gupta (1984) and redefined by (69). The corresponding L_1 norm objective function will be,

$$\min_A: S = \int_0^1 |\ln y(x) - \ln x - (x-1)\ln A|dx \tag{72}$$

or,

$$\min_A: S = \int_0^1 |x-1||[\ln y(x)-\ln x]/(x-1) - \ln A|dx \tag{73}$$

By a similar technique used by (9), we can rewrite (73) as,

$$\min_A: S = \int_0^t |x-1|\{[\ln y(x)-\ln x]/(x-1) - \ln A\}dx$$
$$- \int_t^1 |x-1|\{[\ln y(x)-\ln x]/(x-1) - \ln A\}dx \tag{74}$$

since, $0 \le x \le 1$ we have,

$$\min_A: S = \int_0^t [\ln y(x) - \ln x - (x-1)\ln A]dx$$
$$- \int_t^1 [\ln y(x) - \ln x - (x-1)\ln A]dx \tag{75}$$

Differentiate (75) partially with respect to t and A and equate them to zero;

$$\frac{\delta S}{\delta A} = -\int_0^t [(x-1)/A]dx + \int_t^1 [(x-1)/A]dx = 0 \tag{76}$$

$$\frac{\delta S}{\delta t} = 2[\ln y(t) - \ln t - (t-1)\ln A] = 0 \tag{77}$$

From equation (76), we have,

$$-[\tfrac{1}{2}x^2 - x]_0^t + [\tfrac{1}{2}x^2 - x]_t^1 = 0 \tag{78}$$

or,

$$t = 1 \pm \sqrt{2}/2 \tag{79}$$

Since t should belong to the interval [0,1], we accept,

$$t = 1 - \sqrt{2}/2 \tag{80}$$

Substitute (80) in (77),

$$\ln y(1-\sqrt{2}/2) - \ln (1-\sqrt{2}/2) - (1-\sqrt{2}/2-1)\ln A = 0 \qquad (81)$$

Thus the L_1 norm estimation for A is equal to,

$$A = \left[\frac{1-\sqrt{2}/2}{y(1-\sqrt{2}/2)}\right]^{\sqrt{2}} \qquad (82)$$

Now, the L_1 norm approximation procedure is applied to another Lorenz curve functional form (41) proposed by Bidabad and Bidabad (1989a,b) and redefined by (70). Rewrite L_1 norm objective function (71) for the model (70),

$$\min_{A,B}: S = \int_0^1 |\ln y(x) - B \ln x - (x-1) \ln A| dx \qquad (83)$$

or,

$$\min_{A,B}: S=\int_0^1 |x-1|\,|[\ln y(x)]/(x-1)-(\ln x)/(x-1)-\ln A| dx \qquad (84)$$

The objective function (84) - by some changing on variables - is similar to (18). Thus, by a similar procedure to those of (19) through (31) we can write S as,

$$\begin{aligned}\min_{A,B}: S = &\int_0^{t1} |x-1|\{[\ln y(x)]/(x-1)-(\ln x)/(x-1)-\ln A\}dx \\ &- \int_{t1}^{t2} |x-1|\{[\ln y(x)]/(x-1)-(\ln x)/(x-1)-\ln A\}dx \\ &+ \int_{t1}^{1} |x-1|\{[\ln y(x)]/(x-1)-(\ln x)/(x-1)-\ln A\}dx \end{aligned} \qquad (85)$$

Since $0 \le x \le 1$, then (85) reduces to,

$$\begin{aligned}\min_{A,B}: S = &- \int_0^{t1} [\ln y(x) - B \ln x - (x-1) \ln A]dx \\ &+ \int_{t1}^{t2} [\ln y(x) - B \ln x - (x-1) \ln A]dx \\ &- \int_{t2}^{1} [\ln y(x) - B \ln x - (x-1) \ln A]dx \end{aligned} \qquad (86)$$

Differentiate S partially with respect to A, B, t_1 and t_2 and equate them to zero,

$$\frac{\delta S}{\delta A} = \frac{1}{A}\left[\int_0^{t_1}(x-1)dx - \int_{t_1}^{t_2}(x-1)dx + \int_{t_2}^{1}(x-1)dx\right] = 0 \qquad (87)$$

$$\frac{\delta S}{\delta B} = \int_0^{t_1}\ln(x)dx - \int_{t_1}^{t_2}\ln(x)dx + \int_{t_2}^{1}\ln(x)dx\Big] = 0 \qquad (88)$$

$$\frac{\delta S}{\delta t_1} = -2\{\ln[y(t_1)] - B\ln(t_1) - (t_1-1)\ln(A)\} = 0 \qquad (89)$$

$$\frac{\delta S}{\delta t_2} = 2\{\ln[y(t_2)] - B\ln(t_2) - (t_2-1)\ln(A)\} = 0 \qquad (90)$$

The above system of simultaneous equations can be solved for the unknowns t_1, t_2, A and B. Equation (87) is reduced to,

$$t_1^2-t_2^2-2(t_1-t_2)-\tfrac{1}{2} = 0 \qquad (91)$$

Equation (88) can be written as,

$$\left[x(\ln x-1) \right]_0^{t_1} - \left[x(\ln x-1) \right]_{t_1}^{t_2} - \left[x(\ln x-1) \right]_{t_2}^{1} = 0 \qquad (92)$$

or,

$$t_1(\ln t_1-1)-t_2(\ln t_2-1)+t_1(\ln t_1-1)+\ln 1-1-t_2(\ln t_2-1)=0 \qquad (93)$$

or,

$$t_1(\ln t_1-1) - t_2(\ln t_2-1) - \tfrac{1}{2} = 0 \qquad (94)$$

Calculate t_1 from (91) as,

$$t_1 = 1 \pm \sqrt{(t_2^2-2t_2+3/2)} \qquad (95)$$

Since $0 \le t_1 \le 1$, we accept,

$$t_1 = 1 - \sqrt{(t_2^2-2t_2+3/2)} \qquad (96)$$

Substitute t_1 from (96) into (94) gives,

$$[1 - \sqrt{(t_2^2-2t_2+3/2)}]\{\ln[1 - \sqrt{(t_2^2-2t_2+3/2)}] - 1\}$$
$$- t_2(\ln t_2-1) - \tfrac{1}{2} = 0 \qquad (97)$$

Rearrange the terms,

$$\ln \frac{[1-\sqrt{(t_2^2-2t_2+3/2)}]^{[1-\sqrt{(t_2^2-2t_2+3/2)}]}}{t_2^{t_2}} + t_2 - 3/2$$
$$+ \sqrt{(t_2^2-2t_2+3/2)} = 0 \qquad (98)$$

The root of equation (98) may be computed by a suitable numerical algorithm. However, it has been computed and rounded for five digits floating point as,

$$t_2 = 0.40442 \qquad (99)$$

Value of t_1 is derived by substituting t_2 into (96);

$$t_1 = 0.07549 \qquad (100)$$

Values of B and A are computed from (89) and (90) using t_2 and t_1 given by (99) and (100). Thus,

$$B = \frac{(t_2-1)\ln y(t_1) - (t_1-1)\ln y(t_2)}{(t_2-1)\ln(t_1) - (t_1-1)\ln(t_2)} \qquad (101)$$

or,

$$B = -0.848571n[y(0.07549)] + 1.317221n[y(0.40442)] \qquad (102)$$

and,

$$A = [y(0.07549)]^{-1.28986}[y(0.40442)]^{3.68126} \qquad (103)$$

Now, let us describe how equation (82) for the model

(39) and equations (102) and (103) for the model (41) can be used to estimate the parameters of the Lorenz curve when the probability distribution function is known. In the model (39) we should solve (64) for $x(v)=1-\sqrt{2}/2$. On the other hand, we should find value of v such that,

$$x(v) = [1/E(w)]\int_0^v wf(w)dw = 1-\sqrt{2}/2 \qquad (104)$$

By substituting this value of v into (63), value of $y(1-\sqrt{2}/2)$ is computed. The value $y(1-\sqrt{2}/2)$ is used to compute the parameter A given by (82) for model (39).

The procedure for the model (41) is also similar, with the difference that two values of v should be computed. Once two different values of v are computed as follow,

$$x(v) = [1/E(w)]\int_0^v wf(w)dw = 0.07549 \qquad (105)$$

$$x(v) = [1/E(w)]\int_0^v wf(w)dw = 0.40442 \qquad (106)$$

Values of v are substituted in (63) to find y(0.07549) and y(0.40442). These values of y are used to compute the parameters of the model (41) by substituting them into (102) and (103).

The only problem remains is computation of related definite integrals of x(v) defined by (104), (105) and (106) which can be done by appropriate numerical methods.

CHAPTER VI

SUMMARY AND CONCLUSIONS, RECOMMENDATIONS FOR FURTHER RESEARCH

1. Summary and conclusions

In the first chapter we tried to familiarize the reader with justification and importance of L_1 norm estimator. It was discussed why least squares estimator is not reliable in non-Gaussian cases of error distributions. By a brief look at the chronology of data analysis based on the L_1 norm the difficulties corresponding to this estimator were refereed to and the purpose of this research was clarified. The major steps to be taken in forthcoming chapters was also cited. Definition and properties of normed space were summarized to prepare the reader for notations of other chapters. Position of L_1 norm estimates in the class of L_p norm estimates was considered and properties of the L_1 norm estimation were presented. In this respect invariance property, transformation on variables, convexity of objective function, zero residuals of optimal solution, optimality condition, unique and non unique solutions and interior and sensitivity analysis of the L_1 norm estimation were all discussed.

In chapter II we tried to give a rather general review of literature based on the L_1 norm criterion on the area of estimation theory and related topics with special attention on computational procedures. Since in chapter III we developed some descent algorithms for computing the L_1 norm estimation of regression parameters, the section devoted to this type of algorithms was rather more expository. However, the chronology and historical development of the L_1 norm estimation theory for the period of 1632-1928 was surveyed and the algorithms belonging to the after 1928 period were categorized into three main classes of direct descent, simplex type and other algorithms. Discussions on initial value problem for starting different algorithms, existing computer softwares, comparisons of the algorithms which have been done by previous researchers, and a glance on nonlinear form computational methods and L_p norm computation literature all were done. L_1 norm specification of simultaneous equations system which is an interested field for

econometricians was discussed. Statistical aspects of the L_1 norm, such as sampling distribution, statistical inference, multivariate, nonparametric and robust statistics were briefly cited. Application of the L_1 norm in various fields of sciences with special attentions on economic problems was the other topic in chapter II. At the end of this chapter to complete this survey, we gave an exposition of other related variants of L_1 norm.

In chapter III, we proposed some algorithms for L_1 norm computation of regression parameters, where two of them were selected as more efficient for simple and multiple regression models. However, we started with restricted simple linear regression and corresponding derivation and computation of weighted median problem. In this respect a computing function was coded. With discussion on the m parameter model we continued to expand the algorithm to include unrestricted simple linear regression and two crude and efficient algorithms were proposed. A computer program was also introduced. The procedures then generalized to the m parameter model by presenting two new algorithms, where the algorithm 4 was selected as more efficient. Various properties of these algorithms were discussed and the corresponding computer programs introduced. In the last section, we had a glance on initial value problem.

Since accuracy is more important than efficiency in computational algorithms, those algorithms should be selected which produce correct solutions and among them, the fastest one should be selected. In this regard, algorithm 2 and algorithm of Josvanger and Sposito (1983) were compared in chapter IV. In all of the experiments both algorithms computed the correct answers for the two parameter linear L_1 norm regression model. Algorithm 2 which was faster than the latter introduced for applied works. For multiple regression, algorithms of Bloomfield and Steiger (1980) and Armstrong and Frome and Kung (1979) and Barrodale and Roberts (1973) were compared with algorithm 4. The two first algorithms failed to

compute correct answers in larger models. Because of the accumulated roundoff error, these two algorithms were not usable in larger size problems. However, it concluded that algorithm 4 is more appropriate for models with less than ten parameters and algorithm of Barrodale and Roberts (1973,74) for the ten parameter model. This last conclusion is not very constructive, since in the case of the ten parameter model with 10000 observations algorithm 4 was highly superior to that of Barrodale and Roberts (1973). However, since in applied works we are not always confronted with very large models, this conclusion is poor in operational sense.

Chapter V was devoted to develop the L_1 norm estimation technique for continuous case. Two procedures for the L_1 norm smoothing of restricted and unrestricted simple linear models with continuous error functions were proposed. With a glance on concentration surface theory and the previously introduced functional forms for Lorenz curve, the estimation of their parameters was under consideration. By referring to two probability distribution functions of Pareto and log-normal, which are convenient in modeling the income distribution, the different techniques to link them to Lorenz curve were developed which can be used to other probability distributions.

2. Recommendations for further research

In recent years the L_1 norm criterion has received much attentions in different fields of sciences and therefore, seems to be necessary to turn the researchers's attention to this area. However, in the process of this study we found some points which might be leading to develop the field of L_1 norm computation and application.

1) The diagnostic techniques as sensitivity and interior analysis of L_1 norm regression need further investigation and development. For basic discussion on this topic see Narula and Wellington (1985), and also chapter I.

2) In the proposed algorithm 4, we could not find a

criterion to delete that observation from the current set of observations in the basis which reduces the objective function value more than other points. If anyone finds a criterion like the heuristic method of Bloomfield and Steiger (1980) (see also chapter II), number of iterations will decrease very much and will reduce computation time.

3) The main obstacle to apply discrete differentiation technique suggested in chapter III to operational problems, is reordering of the residuals in descending order before starting computation. If this obstacle is removed, a similar algorithm to that of simple restricted linear regression for multiple regression may be proposed.

4) New improved algorithms for linear programming, as Karmarkar projective procedure, have possibilities for application to the L_1 norm problem.

5) A similar procedure to that of multiple regression model for discrete case may be expanded to include m unknown parameters for continuous L_1 norm smoothing procedure. The cases of one and two parameter models were developed through the chapter V.

This page has been intentionally left blank

APPENDICES

Appendix A. Computer program for weighted median problem

See, chapter III, section 2.1 for description of the procedure.

```
      FUNCTION LWMED(N,YS,W,L)
      REAL YS(N),W(N)
      INTEGER L(N),HI
      II=0
      SHI=0.
      SLO=0.
      SZ=0.
      SP=0.
      SN=0.
      DO 4  I=1,N
      W(I)=ABS(W(I))
      IF(YS(I))3 ,2 ,1
1     SP=SP+W(I)
      GO TO 4
2     SZ=SZ+W(I)
      GO TO 4
3     SN=SN+W(I)
4     CONTINUE
      SHI=SP+SZ
      SLO=SN+SZ
      IF(SHI.LE.SLO) GO TO 6
      SHI=0.
      DO 5 I=1,N
      IF(YS(I).LE.0.) GO TO 5
      II=II+1
      L(II)=I
5     CONTINUE
      GO TO 8
6     SLO=0.0
      DO 7 I=1,N
      IF(YS(I).GT.0.) GO TO 7
      II=II+1
      L(II)=I
7     CONTINUE
8     LO=1
      HI=II
10    IF(HI.GT.LO+1)GO TO 30
      LWMED=L(LO)
      IF(LO.EQ.HI) RETURN
      IF(YS(L(LO)).LE.YS(L(HI))) GO TO 20
      LT=L(LO)
      L(LO)=L(HI)
      L(HI)=LT
      LWMED=L(LO)
20    IF(SHI+W(L(HI)).GT.SLO+W(L(LO))) LWMED=L(HI)
      RETURN
30    MID=(LO+HI)/2
      LOP=LO+1
      LT=L(MID)
      L(MID)=L(LOP)
      L(LOP)=LT
      IF(YS(L(LOP)).LE.YS(L(HI))) GO TO 40
      LT=L(LOP)
      L(LOP)=L(HI)
      L(HI)=LT
40    IF(YS(L(LO)).LE.YS(L(HI)))GO TO 50
      LT=L(LO)
      L(LO)=L(HI)
      L(HI)=LT
50    IF(YS(L(LOP)).LE.YS(L(LO))) GO TO 60
      LT=L(LOP)
      L(LOP)=L(LO)
      L(LO)=LT
60    LWMED=L(LO)
      I=LOP
      J=HI
```

```
      XT=YS(LWMED)
      TLO=SLO
      THI=SHI
 70   TLO=TLO+W(L(I))
      I=I+1
      IF(YS(L(I)).LT.XT) GO TO 70
 80   THI=THI+W(L(J))
      J=J-1
      IF(YS(L(J)).GT.XT) GO TO 80
      IF(J.LE.I) GO TO 90
      LT=L(I)
      L(I)=L(J)
      L(J)=LT
      GO TO 70
 90   TEST=W(LWMED)
      IF(I.NE.J) GO TO 100
      TEST=TEST+W(L(I))
      I=I+1
      J=J-1
 100  IF(TEST.GE.ABS(THI-TLO)) RETURN
      IF(TLO.GT.THI)GO TO 110
      SLO=TLO+TEST
      LO=I
      GO TO 10
 110  SHI=THI+TEST
      LO=LOP
      HI=J
      GO TO 10
      END
```

Appendix B. Computer program for algorithm 2

See, chapter III, section 3.2 for description of the procedure.

```
      PROGRAM BL1S
      PARAMETER (N=1000)
      DIMENSION Y(N),X2(N),Z(N),W(N),L(N)
      DO 10 I=1,N
10    READ(5,20) Y(I),X2(I)
20    FORMAT(2F10.3)
      K1=N/2
      K1R=0
      K1S=0
30    DO 40 I=1,K1-1
      W(I)=X2(I)-X2(K1)
40    Z(I)=(Y(I)-Y(K1))/W(I)
      W(K1)=0.
      Z(K1)=0.
      DO 50 I=K1+1,N
      W(I)=X2(I)-X2(K1)
50    Z(I)=(Y(I)-Y(K1))/W(I)
      ITER=ITER+1
      LM=LWMED(N,Z,W,L)
      K1S=K1R
      K1R=K1
      IF (LM.EQ.K1S) GOTO 60
      K1=LM
      GOTO 30
60    B2=Z(LM)
      B1=Y(K1)-B2*X2(K1)
      PRINT 70,B1,B2
70    FORMAT(1X,'B1=',F13.5,3X,'B2=',F13.5)
      STOP
      END
```

Appendix C. Computer program for algorithm 4

See chapter III, sections 4.2 and 4.3 for description of the procedures.

```
      PROGRAM BL1
      PARAMETER (N=1000,M=5,M1=M-1,M2=M-2)
      DIMENSION Y(N),X(N,M1),XSK(M1),YW(N),XKW(N)
      DIMENSION W(N),YS(N),XS(N,M1),B(M),XW(N,2:M1)
      DIMENSION L(N),KK(M1),YSOL(M1),XSOL(M1,M1)
      COMMON /C1/I1,I2
      DO 10 I=1,N
10    READ(5,20) Y(I),(X(I,J),J=1,M1)
20    FORMAT(10F10.3)
      ITER=0
      KR=0
      MM=1
      DO 30 J=1,M1
30    KK(J)=J*N/M
40    DO 50 I=1,N
      YS(I)=Y(I)
      DO 50 J=1,M1
50    XS(I,J)=X(I,J)
      GO TO 80
60    DO 70 I=1,N
      W(I)=XKW(I)
      YS(I)=YW(I)
      DO 70 J=MM,M1
70    XS(I,J)=XW(I,J)
80    JJ=MM
90    K=KK(JJ)
      YSK=YS(K)
      DO 100 J=JJ,M1
100   XSK(J)=XS(K,J)
      I1=1
      I2=K-1
110   DO 120 J=JJ,M1
120   CALL COL1(XSK(J),XS(1,J))
      CALL COL2(YSK,JJ,W,YS,XS(1,JJ))
      IF(I2.EQ.N) GO TO 130
      I1=K+1
      I2=N
      GO TO 110
130   W(K)=0.
      IF (JJ.EQ.M1) GO TO 190
      I1=1
      I2=K-1
140   DO 150 J=JJ+1,M1
150   CALL COL3(XS(1,J),XS(1,JJ))
      IF(I2.EQ.N) GO TO 160
      I1=K+1
      I2=N
      GO TO 140
160   IF(JJ.NE.MM) GO TO 180
      DO 170 I=1,N
      XKW(I)=W(I)
      YW(I)=YS(I)
      DO 170 J=JJ+1,M1
170   XW(I,J)=XS(I,J)
180   JJ=JJ+1
      GO TO 90
190   YS(K)=0.
      ITER=ITER+1
      LM=LWMED(N,YS,W,L)
      IF(LM.EQ.KR) GO TO 220
      IOPT=0
200   IF(MM.EQ.M1) GO TO 210
      MM=MM+1
      KR=KK(MM)
      KK(MM)=LM
      GO TO 60
```

```
210   MM=1
      KR=KK(MM)
      KK(MM)=LM
      GO TO 40
220   IOPT=IOPT+1
      IF (IOPT.NE.M1) GO TO 200
      B(M)=YS(LM)
      DO 230 I=1,M1
      YSOL(I)=Y(KK(I))
      DO 230 J=1,M1
230   XSOL(I,J)=X(KK(I),J)
      JJ=1
240   YSK=YSOL(JJ)
      DO 250 J=JJ,M1
250   XSK(J)=XSOL(JJ,J)
      DO 270 I=JJ,M1
      IF(I.EQ.JJ) GO TO 270
      YSOL(I)=YSOL(I)-YSK
      DO 260 J=JJ,M1
260   XSOL(I,J)=XSOL(I,J)-XSK(J)
      YSOL(I)=YSOL(I)/XSOL(I,JJ)
270   CONTINUE
      DO 290 I=JJ,M1
      IF(I.EQ.JJ) GO TO 290
      DO 280 J=JJ+1,M1
280   XSOL(I,J)=XSOL(I,J)/XSOL(I,JJ)
290   CONTINUE
      IF (JJ.EQ.M2) GO TO 300
      JJ=JJ+1
      GO TO 240
300   DO 320 I=1,M2
      K=M-I
      S=YSOL(K)
      DO 310 J=K,M1
310   S=S-B(J+1)*XSOL(K,J)
320   B(K)=S
      S=Y(KK(1))
      DO 330 J=1,M1
330   S=S-B(J+1)*X(KK(1),J)
      B(1)=S
      PRINT 340,(B(J),J=1,M)
340   FORMAT(1X,F13.5)
      PRINT 350,(KK(J),J=1,M1),LM,ITER
350   FORMAT(1X,I13)
      STOP
      END


      SUBROUTINE COL1(V1,V2)
      DIMENSION V2(1)
      COMMON /C1/I1,I2
      DO 1 I=I1,I2
    1 V2(I)=V2(I)-V1
      RETURN
      END


      SUBROUTINE COL2(YSK,JJ,V1,YS,V2)
      DIMENSION V1(1),V2(1),YS(1)
      COMMON /C1/I1,I2
      IF (JJ.NE.1) GO TO 2
      DO 1 I=I1,I2
      V1(I)=V2(I)
    1 YS(I)=(YS(I)-YSK)/V2(I)
      RETURN
    2 DO 3 I=I1,I2
      V1(I)=V1(I)*V2(I)
    3 YS(I)=(YS(I)-YSK)/V2(I)
      RETURN
      END
```

```
      SUBROUTINE COL3(V1,V2)
      DIMENSION V1(1),V2(1)
      COMMON /C1/I1,I2
      DO 1 I=I1,I2
1     V1(I)=V1(I)/V2(I)
      RETURN
      END
```

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 L_∞ 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. University 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. Annals 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 University.

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, α-trimmed and α-winsorized regression on the ODRA 1305 computer. Rep. N-159, Institute of Computer Science, Wroclaw university, 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. North-Holland. 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. Commun. Statist., 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. Proceeding 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. Commun. 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. Commun. 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. North-Holland, 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., University 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. University 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. University 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. University 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. University of CapeTown, Dept. of Math. Stat., Tech. Rep. no. ALS-6, April.

I. Barrodale (1967) Approximation in the L_1 and L_∞ norms by linear programming. Ph.D. thesis, University of Liverpool, Liverpool, England.

I. Barrodale (1968) L_1 approximation and analysis of data. Appl. Stat. 17,51-57.

I. Barrodale (1970) On computing best L_1 approximations. In A. Talbot, Approximation theory Academic Press, New York, 205-215.

I. Barrodale, C. Phillips (1975) Solution of an over-determined system of linear equations in Chebyshev norm. ACM Trans. on Math. Soft. 264-70.

I. Barrodale, M.J.D. Powell, F.D.K. Roberts (1972) The differential correction algorithm for rational L_∞ approximation. SIAM J. Num. Anal., 9, 493-503.

I. 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.

I. Barrodale, F.D.K. Roberts (1973) An improved algorithm for discrete L_1 linear approximation. SIAM J. Numer. Anal., 10, 839-848.

I. Barrodale, F.D.K. Roberts (1974) Algorithm 478: Solution of an overdetermined system of equations in the L_1 norm. Commun. ACM, 17, 319-320.

I. Barrodale and F.D.K. Roberts (1977) Algorithms for restricted least absolute value estimation. Commun. Stat. B6(4), 353-363.

I. 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.

I. 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.

I. Barrodale, C. Phillips (1974) An improved algorithm for discrete Chebyshev linear approximation. Proc. 4th Manitoba conf. on numer. math. University of Manitoba, Winnipeg, Manitoba, 177-190.

I. Barrodale, A. Young (1966) Algorithms for best L_1 and L_∞ 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. Commun. 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 L_∞ 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 L_∞ sense? Res. Rep. CS-87-12, University 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 Statist. 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. Commun. 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, Commun. ACM, 11, 401-406.

R.H. Bartels, G.H. Golub (1969) The simplex method of linear programming using LU decomposition. Commun. ACM, 12, 266-268.

G.W. Bassett (1973) Some properties of the least absolute error estimator. Ph.D. thesis. Dept. of Econ., University of Michigan.

G.W. Bassett, Jr. (1987) A p-subset property of L_1 and regression quantile estimates. Work. Pap., Dept. of Economics, University 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.

C.M. Bender, S.A. Orszag (1978) Advanced mathematical methods for scientists and engineers, McGraw-Hill.

B. Bidabad (1980) The new international scholastic order, an essay in political economy of science. Shiraz University, Shiraz, Iran.

B. Bidabad (1981) Estimation of the Engel's curves for Iran (urban & rural). M.S. thesis. Dept. of Econ., Shiraz University, Shiraz, Iran.

B. Bidabad (1983a) Analysis of taxonomy and its application in classification of regions and creating development index for regional planning. Plan and Budget Ministry, Centeral Province Plan and Budget Organization. Arak, Iran.

B. Bidabad (1983b) The discorrelation of provincial development budget allocation for 1984 and undevelopment index. Centeral Province Plan and Budget Organization, Plan and Budget Ministry, Arak, Iran.

B. Bidabad (1983c) Principal components analysis in creating development index for regional planning. Regional Planning Bureau, Plan and Budget Ministry, Tehran Iran.

B. Bidabad (1984a) Goal programming, optimal decision process

with different priorities and its computer program. Plan and Budget Ministry, Regional Planning Bureau, Tehran, Iran.

B. Bidabad (1984b) Determining industrial location for 1992 by using goal programming method. Plan and Budget Ministry, Regional Planning Bureau, Tehran, Iran.

B. Bidabad (1984c) Zero-one programming, optimal decision making with bivalent variables and its computer program. Plan and Budget ministry, Regional Planning Bureau, Tehran, Iran.

B. Bidabad (1985) Model of Iran's economic growth alternatives, 1986-2001 prospect. Plan and Budget Ministry, Regional Planning Bureau, Tehran, Iran.

B. Bidabad (1987a) A proposed algorithm for 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) A proposed algorithm for 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. Proceedings of the Third Seminar of Mathematical Analysis. Shiraz University, 24-34, Shiraz, Iran.

B. Bidabad (1988b) A proposed algorithm for least absolute error estimation, part II. Proceedings of the Third Seminar of Mathematical Analysis, Shiraz University, 35-50, Shiraz, Iran.

B. Bidabad (1989) A mathematical model for simultaneous computation of preferred prices and number of pieces of lands in non-beneficial land preparation projects. Economics and Management quarterly journal of the Islamic Azad University. Vol.2, 38-50.

B. Bidabad ( ) Dynamic economic models. vol. I, Difference equations. Unpublished.

B. Bidabad, B. Bidabad (1989a) Functional form for estimating the Lorenz curve. Submitted to the Australasian Meeting of Econometric Society, Canberra, Australia.

B. Bidabad, B. Bidabad (1989b) Functional form for estimating the Lorenz curve. To be appeared in Economics and Management quarterly journal of the Islamic Azad University.

B. Bidabad, B. Bidabad (1989c) Complex probability and Markov stochastic processes. Submitted to the 47th session of ISI, sep., Paris, France.

B. Bidabad, B. Bidabad (1989d) Complex probability and Markov stochastic process. To be appeared in Economics and Management quarterly journal of the Islamic Azad University.

B. Bidabad, M. Sabetghadam, S. Davar, G.R. Sadeghi (1988) Dynamic model of population, mathematical model for analysis and forecasting Iran population changes. Regional Planning Bureau, Plan and Budget Ministry,. Tehran, Iran.

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. Statist. 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, University 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 ϵ-median polish algorithm. CSDA, 5, 327-336.

M. Brannigan, S.A. Gustafson (1987) H-sets in convex programming and constrained approximation. Work. Pap., Rogaland University, 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, North-Holland, 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. Commun. Stat. A10, 2581-2590.

Burgoyne (1965) Polynomial approximation by Edgeworth's method. University of London.

S. Busovaca (1985) Handling degeneracy in a nonlinear L_1 algorithm. Ph.D. thesis, University 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., North-Holland.

J. Chambers (1971) Algorithm 410: partial sorting. Commun. ACM, 14, 357-358.

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.

A. Charnes, W.W. Cooper, R.O. Ferguson (1955) Optimal estimation of executive compensation by linear programming. Manag. Sci. 1, 138-151.

C. Charalambous, A.R. Conn (1978) An efficient method to solve the minimax problem directly. SIAM J. Numer. Anal., 15, no.1, 162-187.

C. 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 Tchebysheff 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 ($p \geq 1$) of U-statistics. CSDA, 5, 321-326.

D.I. Clarke (1981) Finite algorithms for linear optimization problems. Ph.D. thesis, Australian National University, 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, University 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.) North-Holland, 189-201,

Amsterdam.

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., University 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, North-Holland, 209-229.

D.C. Crocker (1969) Linear programming technique in regression analysis, the hidden danger. A.I.E.E. Trans., 1, 112-126.

J.S. Cramer (1973) Empirical econometrics. North-Holland, Amsterdam.

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 University, 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 University, 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. Annals 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. Dieudonne (1963) Foundations of modern analysis. Academic Press, New York.

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. Quarterly, 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. North-Holland. 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 L1-norm and related methods. 275-287, North-Holland.

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, North-Holland.

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 Applied 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, North-Holland.

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. North-Holland.

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.

C. 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. North-Holland. 109-116.

H. Ekblom, S. Henriksson (1969) Lp-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 symposium 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.

E. 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 L_∞ estimation. Commun. 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 L_∞ estimation procedures. In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and related methods. North-Holland. 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. University 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. North-Holland. 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. North-Holland, 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. North-Holland, 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., University 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 p^{th} 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 University, 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 University, 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.) Proceeding 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. Commun. Stat. B9, 201-222.

A. 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 L_1 norm and related methods. North-Holland. 247-262.

G. 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. Galpin, 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. Commun. 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. North-Holland. 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 University, 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., Commun. 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, University 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. North-Holland. 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. North-Holland. 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.

G. Groucher (1971) Best L_1 and L_∞ approximations. M.Sc. thesis, Birkbeck College, London University, London, England.

M.R. Gupta (1984) Functional forms for estimating the Lorenz curve. Econometrica, 52, 1313-1314.

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. North-Holland. 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. Commun. 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. Statist., 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). North-Holland.

S. Henriksson (1972) On a generalization of L_p-approximation and estimation. Thesis, Dept. of Computer Sci., Lund university, Sweden.

R.W. Hill, P.W. Holland (1977) Two robust alternatives to least squares regression. JASA, 72, 828-833.

C.A.R. Hoare (1961) Algorithm 63 partition; 64, quicksort; and 65, find., Comm. ACM, 4, July, 321-322.

C.A.R. Hoare (1962) Quicksort. Comput. J., 5, 10-15.

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 University, 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. North-Holland. Reprinted in CSDA, 5, 255-262.

C.R. Hunt (1970) Best L_p approximation by certain nonlinear

functions. M.Sc. thesis, University of Victoria, University 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.

H. 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 Commun. Engin., Kyushu University 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.

B. Joe, R. Bartels (1983) An exact penalty method for constrained, discrete linear 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.

J. Jureckova (1983) Trimmed polynomial regression. Commentationes Mathematicae Universitatis Carolinae, 24, 4, 597-607.

J. Jureckova (1984) Regression quantiles and trimmed least squares estimator under a general design. Kybernetika, vol.20, no.5, 345-357.

J. 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.

N.C. Kakwani (1980) Functional forms for estimating the Lorenz curve: a reply. Econometrica, 48, 1063-64.

N.C. Kakwani, N. Podder (1976) Efficient estimation of the Lorenz curve and associated inequality measures from grouped observations. Econometrica 44, 137-148.

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_∞ norms. J. Approx. Theory, 3, 123-127.

O.J. Karst (1958) Linear curve fitting using least deviations. JASA, 53, 118-132.

L. Kaufman, P.J. Rousseeuw (1987) Clustering by means of medoids. In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and related methods. North-Holland. 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. North-Holland. 217-230.

M. Kendall, A. Stuart (1977) The advanced theory of statistics. vol.1, Charles Griffin & Co., London.

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.

D. 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.K (1985) On Boscovich's estimator. Annals of Stat., 13, 1625-1628.

W.W. Kotiuga (1982) Power system state estimation using least absolute value techniques. Ph.D. thesis, University 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, University of California, Los Angeles, California.

E. 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.

E. 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.

E. 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 university, 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.

A. Marazzi (1987) Solving bounded influence regression with ROBSYS. In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and related methods. North-Holland. 145-163.

A. Marazzi (1988) Algorithms for the computation of weights in bounded influence regression. CSDA 6(3), 251-276.

A. Marazzi, A. Randriamiharisoa (1985) ROBETH-ROBSYS: a software for robust statistical computing. Document 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. North-Holland. 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, G. 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.

N. 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.

A. Melaku, G. Sadasivan (1987) L_1-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.

G. Merle, H. Spath (1974) Computational experiences with discrete L_p-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.

J. Militky, J. Cap (1987) Application of Bayes approach to adaptive L_p 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 University. 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., University of Pennsylvania.

B. 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., University 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. University 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.

J. 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 University, 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, University 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., University 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. North-Holland. 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 L_∞ 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 university, 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. North-Holland, Amsterdam.

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.

P. 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. North-Holland, 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 L_∞ solutions of overdetermined nonlinear equations. IMAJ Num. Anal., 4, 241-251.

B. 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 $\int_a^b |x(t)|dt$ Czechoslovak Math. J. 8(83), 267-273.

P. Rabinowitz (1968) Application of linear programming to numerical analysis. SIAM Rev., 10, 121-159.

P. Rabinowitz (1970) Mathematical programming and approximation. In A. Talbot (ed.) Approximation Theory. Academic Press, 217-231.

A. 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 p_{th} 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., 7^{th} 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.

E. Ronchetti (1987) Bounded influence in regression: a review. In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and related methods. North-Holland, 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.

B. Rosenberg, D. Carlson (1977) A simple approximation of the sampling distribution of least absolute residuals regression estimates. Comm. Stat., B6, 421-437.

R. 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 University.

R. 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. North-Holland, 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. North-Holland. 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. North-Holland. 307-322.

R.S. Scowen (1965) Algorithm 271 Quickersort. Commun. ACM, 8, 669-670.

E. Seneta (1983) The weighted median and multiple regression. Austral. J. Stat., 25(2), 370-377.

E. Seneta, W.L. Steiger (1984) A new LAD curve-fitting algorithm: slightly overdetermined equation system in L_1. Discrete Applied Math., 7, 79-91.

D. 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 University, 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 L_∞ 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. Annals of math. Stat., 11, 301-310.

R.S. Singleton (1969) Algorithm 347 Sort. Comm. ACM, 12, 185-186.

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.

H. Spath (1976) L_1 cluster analysis. Computing, 16, 379-387.

H. Spath (1982) On discrete linear orthogonal L_p-approximation. Numerische Analysis, ZAMM 62, T354-T355.

H. 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. North-Holland. 427-434.

H. Spath, G.A. Watson (1987) On orthogonal linear L_1 approximation. Numer. Math., 51, 531-543.

M.R. Spiegel (1968) Mathematical handbook. Schaum's outline series, McGraw-Hill, New York.

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.

K. 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. North-Holland. 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.

E. Stiefel (1960) Note on Jordan elimination, linear programming and Tchebycheff approximation. SIAM J. Numer. Math., 2, 1-17.

J. 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. Annals 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.

J. Svanberg (1805) Exposition des operations faites en lappnie pour la determination d'un arc du meridien en 1801, 1802 et 1803,... Stockholm.

T. Taguchi (1972a) On the two-dimensional concentration surface and extensions of concentration coefficient and Pareto distribution to the two dimensional case-I. Annals of the Inst. of Stat. Math., vol. 24, no.2, 355-381.

T. Taguchi (1972b) On the two-dimensional concentration surface and extensions of concentration coefficient and Pareto distribution to the two dimensional case-II. Annals 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. Annals 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.

A. 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.

E. Trauwaert (1987) L_1 in fuzzy clustering. In Y. Dodge (ed.) Statistical data analysis based on the L_1 norm and related methods. North-Holland. 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.

I. 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. North-Holland. 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.

L. 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., University 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. North-Holland. 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. Decision 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.

A. Wulff (1983) Numerische verfahren zur linearen orthogonalen L_p-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.