LEAST ABSOLUTE DEVIATION ESTIMATION OF MULTI-EQUATION LINEAR ECONOMETRIC MODELS

A STUDY BASED ON MONTE CARLO EXPERIMENTS

Prof. SK Mishra
Ms M Dasgupta
Dept. of Economics
NEHU, Shillong, India

Introduction: Ideal properties of the Ordinary Least Squares (OLS or L_2 norm) estimator of single-equation linear econometric model y = X**a** + e, while e obeys Gauss-Markov conditions, are well known. Additionally, if e is normally distributed, OLS estimator of **a** is also the Maximum Likelihood (ML) estimator. However, when e is non-normally distributed, hyper-kurtic or infested with sizeable outliers, OLS estimator fails to perform. It has been observed that in such cases ***Least Absolute Deviation*** (LAD or L_1 norm) estimator performs very well. Sporadic errors in X (where sample X is true X + Ξ, and Ξ is a sparse matrix with nonzero elements substantial in size) also vitiate OLS estimation. There too, LAD performs well.

The real world data often consist of disturbances that are non-normally distributed, some of which permit the variate to take on non-negative values only. It has been known since Pareto, that distribution of income bears testimony to distribution of the error term with infinite variance. Works of **Meyer** & **Glauber** (1964), **Fama** (1965) and **Mandlebroth** (1967), among others, confirm that economic data series like prices in commodity and financial markets present a class of distribution with infinite variance. An infinite variance means 'thick tails, which implies that large values or 'outliers' are present. Sporadic errors in X also are frequently present.

The recommendations for using LAD estimator may be traced back to **Gauss** and **Laplace** (1818) as mentioned by **Taylor** (1974). **Edgeworth** (1887, 1888, 1923), **Rhodes** (1930) and **Singleton** (1940) investigated into this method of estimation. But at those times, computational difficulties involved with the method went in its disfavour. However, development of linear programming (LP) and fast computing machines intensified the interest of researchers in estimation of regression models by minimization of L_1 norm. **Charnes, Cooper** & **Ferguson** (1955) were the first to transform the problem of estimation of parameters of a linear (regression) model into an LP problem. Later, iterative and search methods to obtain the solution were also discovered. **Fisher** (1961), **Ashar** & **Wallace** (1963**), Meyer** & **Glauber** (1964), **Rice** & **White** (1964), **Usow** (1967), **Oveson** (1968), **Robers** & **Ben-Israel** (1969**), Abdelmalek** (1971, 1974), **Blattberg** & **Sargent** (1971), **Smith** & **Hall** (1972), **Barrodale** & **Roberts** (1973), **Schlossmacher** (1973), **Fair** (1974), **Taylor** (1974), **Nyquist** & **Westland** (1977), **Bassett** & **Koenkar** (1978), **Powell** (1984), **Pollard** (1991), **Phillips** (1991), **Chen** (1996), **Hitomi** & **Kagifara** (2001), etc. intensively worked on the L_1 estimator of single equation (regression) models.

Multi-equation linear econometric models, described as YA+XB+E, were first estimated by minimization of L_2 norm. These estimators are collectively called the k-class estimators (Two-Stage Least Squares - 2-SLS and Limited Information Max Likelihood - LIML). In 2-SLS estimation, OLS is used to estimate the matrix of Reduced Form Coefficients (P) at the first stage, which also gives estimated Y ($\hat{Y}$ = XP). In the second stage, Y is replaced by $\hat{Y}$ if and only if it appears as an explanatory variable in any structural equation. After this replacement, OLS is applied to estimate the structural equations, one at a time. Thus, OLS is applied twice, once at each stage. The 2-SLS is also an Instrumental Variable method of estimation. In spite of OLS – the basic building block of 2-SLS – being an ideal estimator if the required conditions for its application are met, 2-SLS is ordinarily a biased but consistent estimator. It was found that L_2-based k-class estimator performs extremely poorly when non-normally distributed, hyper-kurtic or outlier-infested errors are met with.

Glahe & **Hunt** (1970) were the first to apply LAD estimation method to multi-equation linear models. **Amemiya** (1982) extended LAD (L_1 estimator) to multi-equation models. He generalized LAD to include 2-SLS as its special case. Amemiya's work is theoretical and his conclusions relate to asymptotic properties (consistency) of LAD in estimating the multi-equation model. He derived consistency of 2-SLAD expressed as the minimization problem of $\sum_{i=1}^{n}\left|ky_{ij}+(1-k)\hat{y}_{ij}-[\hat{Y}_{ij}\mid X_{ij}][\frac{a_j}{b_j}]\right|^L$ for j[th] structural equation, j=1,m. For L=2 and $0 \le k \le 1$ it gives 2-SLS, but for L=1 and $0 \le k \le 1$ it gives 2-SLAD. He found that for 0 < k < 0.5, 2-SLAD performs better than 2-SLS if errors are mixed normal. He suggested estimation by D2SLAD (Double 2-Stage LAD) in case of full non-normal and outlier infested errors. **Newey** (1985), **Pagan** (1986), **Fair** (1994) and **Kim** & **Muller** (2000) are some important studies on 2-SLAD estimation.

In real life, small or medium size samples are important. Small or medium sample properties of an estimator are difficult to obtain by analytical methods. Therefore, Amemiya suggested Monte Carlo experiments to assess the performance of LAD-based estimators of multi-equation models with small or medium size samples. Our literature survey suggests that perhaps no study was conducted in this line. Possibly, frequent application of single equation estimation by LAD attracted intensive research as it is in the domain of statistics in which users from many disciplines are interested while multi-equation estimation is limited to econometric models only.

Monte Carlo Method of Simulation Experiments:: The Monte Carlo method as a numerical simulation technique was initiated by **S. Ulam** & **N. Metropolis** (1949). Working with **John von Neumann** they developed algorithms for computer implementations, as well as methods of transforming non-random problems into random forms that would facilitate their solution via statistical sampling. Monte Carlo method is now used routinely in many diverse fields, from the simulation of complex physical phenomena such as radiation transport in the earth's atmosphere and the simulation of the esoteric sub-nuclear processes in high energy physics experiments to the mundane problems such as the simulation of simple games. The primary components of a Monte Carlo simulation method includes: *(i) Probability distribution functions (pdf's)* - the physical (or mathematical) system must be described by a set of pdf's; *(ii) Random number generator* - a source of random numbers uniformly distributed on the unit interval must be available; *(iii) Sampling rule* - a prescription for sampling from the specified pdf's, assuming the availability of random numbers on the unit interval, must be given; *(iv) Scoring (or tallying)* - the outcomes must be accumulated into overall tallies or scores for the quantities of interest. Additionally, an estimate of the statistical error (variance) as a function of the number of trials and other quantities have to be determined. To enhance the speed at which the experiments are carried out, methods for reducing the variance in the estimated solution (resulting into reduction of the computational time for Monte Carlo simulation) are applied. For complicated and large systems, the *Parallelization* *and vectorization*-based algorithms are used such as to allow the Monte Carlo method to be implemented efficiently on advanced computer architectures.

Monte Carlo Method to study the Properties of an Estimator: We find a good account of Monte Carlo experiments conducted for studying the properties of various k-class estimators in **Intriligator** (1978, pp. 416-420). The method starts by postulating a specific model and assign numerical values to all parameters in the model. In a single equation model such as y = X**a** + e, X and **a** are assigned specific numerical values. The dimension of X(n,m) is fixed. Properties of e (such as its pdf or distribution and the parameters of the distribution) are specified. For fixed X**a**, disturbances (e) are generated with the given specification and, using the candidate estimator, **a** is estimated to give $\hat{a}$ repeatedly for a large (say, r) number of times. This gives an array of estimated parameters, say $\hat{a}(r)$. Then mean($\hat{a}(r)$), variance($\hat{a}(r)$) and RMS($\hat{a}(r)$) = $\{\frac{1}{r}\sum_{i=1}^{r}(\hat{a}_i - a)^2\}^{1/2}$ are obtained. Bias is defined as B= {mean($\hat{a}(r)$)-**a**} and variance($\hat{a}(r)$) = V = $\frac{1}{r}\sum_{i=1}^{r}(\hat{a}_i - \frac{1}{r}\sum_{i=1}^{n}\hat{a}_i)^2$. RMS($\hat{a}(r)$) = $\sqrt{V + B^2}$. If needed, higher order moments of $\hat{a}(r)$ may be obtained to determine its distribution. Performance of the candidate estimator is judged on these criteria. Smaller are B^2 and V, better is the performance of the candidate estimator. Smaller RMS may be used as a comprehensive measure of better performance. This exercise makes a unit experiment. Such experiments are carried out repeatedly. Conclusions are derived from the results of these experiments.

In case of a multi-equation models such as YA+XB+E=0, the experiment starts with given (pre-assigned) X(n,k), A(m,m) and B(k,m) matrices as well as the specifications regarding the structural disturbances, E. Then the matrix of sample disturbances, E, is generated with the pre-specified statistical properties. With the given X, $\Pi = -XBA^{-1}$ and $U = -EA^{-1}$, Y = XΠ + U is obtained. The candidate estimator is now applied to obtain $\hat{A}$ and $\hat{B}$. The process of generation of E, Y = XΠ + U and estimation of $\hat{A}$ and $\hat{B}$ is repeated for a large number of times and finally bias, variance and RMS are obtained as in case of single equation models. This exercise makes a unit experiment. Such experiments are carried out repeatedly. Conclusions are derived from the results of these experiments.

In Monte Carlo experiments parameters such as sample size (n), model size (m in single equation models, and m and k in multi-equation models) and distribution of E and the parameters of chosen distribution are variables that may influence the performance of a candidate estimator. Therefore, one may vary some or all of these parameters to study the performance of the candidate estimator.

It is also possible to study the performance of a candidate estimator in presence of outliers. Then experiments are carried out with y = X**a** + e+ O or YA+XB+E+O = 0, where O is the matrix of outliers. Usually O has a sparse structure, a few of its elements being non-zero and the rest zero. Non-zero elements are usually randomly spaced and random in magnitude, though within a given range. Large and numerous outliers have a great impact on the performance of LS-based estimators while LAD-based estimators are usually immune to their number and size.

Sporadic errors in X (where sample X is true X + Ξ, and Ξ is a sparse matrix with nonzero elements substantial in size - not to be confused with stochastic X) also destabilize LS-based estimators. Monte Carlo experiments may be carried out to study the performance of an estimator in presence of sporadic errors or perturbations in X.

Objectives: We aim at investigating into the performance of some LAD-based estimators vis-à-vis that of the LS-based estimators for linear econometric models of various specifications. In particular, we aim at comparing the performance of the said estimators in case of multi-equation linear econometric models. We envisage that in case of linear models with non-normal disturbances or outlier-infested disturbances, LAD-based estimators will outperform the LS-based estimators.

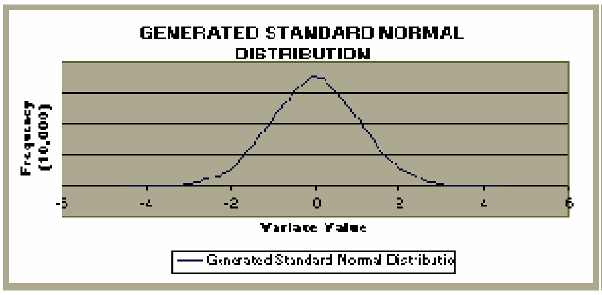
The Candidate Estimators: The 2-SLS (or LS-LS) is the standard LS-based estimator of multi-equation linear econometric model. It is a member of the k-class estimator with k =1. The performance of 2-SLS is the reference for the candidate (LAD-based) estimators. Among the LAD-based estimators we have included (i) LS-LAD, (ii) LAD-LS, (iii) LAD-LAD, (iv) LS-GILN and (v) LAD-GILN estimators.

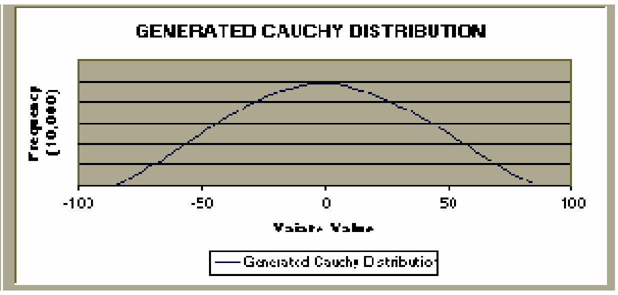
LS-LAD estimator obtains P (the matrix of Reduced form Coefficients) by OLS, but the structural coefficients (a_j and b_j) by LAD estimator. LAD-LS estimator obtains P by LAD, but the structural coefficients (a_j and b_j) by OLS. It may be considered as a variant of IV (Instrumental Variable) estimators where LAD-estimated $\hat{Y}$ are used as the instrumental variables for Y. LAD-LAD estimator is D2SLAD of **Amemiya** (1982), which obtains P as well as the structural coefficients (a_j and b_j) by LAD estimator. The LS-GILN is the Khazzoom estimator (**Khazzoom**, 1976) that estimates reduced form equations of a multi-equation linear econometric model by OLS but (in the second stage) instead of estimating the (modified) structural equations by OLS (or the Instrumental variable method) as done by the 2-SLS, it applies generalized inverse of the relevant submatrix of reduced form coefficients to obtain the structural coefficients. More explicitly, for the model YA+XB+E=0 (the reduced form equations being $Y = X\Pi + U$, $\Pi = -BA^{-1}$ and $P = \hat{\Pi}$), in the relationship $Pa_j = -b_j$ for any (j^{th}) structural equation, we have

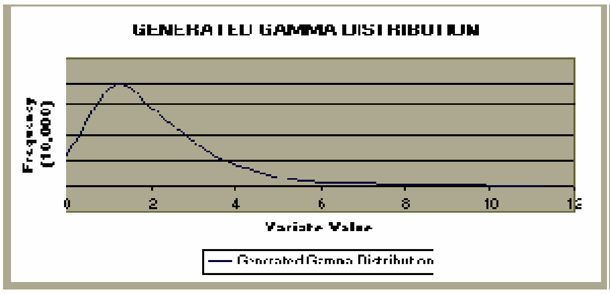
$\begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = -\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$ where a_1 and b_1 are unknown structural coefficients, a_2 = (0 0 … 0 -1)' and b_2 = (0 0 … 0 0)'. From this we obtain $\hat{a}_1 = -P_{21}^{-g}(b_2 - P_{22}a_2)$ and $\hat{b}_1 = -(P_{11}\hat{a}_1 + P_{12}a_2)$. This is GILN (Generalized Indirect Least Norm) estimator because ILS (Indirect Least Squares) estimator, which is applicable only in case of an exactly identified structural equation, is defined as $\hat{a}_1 = -P_{21}^{-1}(b_2 - P_{22}a_2)$ and $\hat{b}_1 = -(P_{11}\hat{a}_1 + P_{12}a_2)$. If P_{21} is a square matrix of full rank, $P_{21}^{-g} = P_{21}^{-1}$. Therefore, GILN is applicable in case of any structural equation exactly or over identified. In LS-based GILN (Khazzoom estimator) $P = (X'X)^{-1}X'Y$ is the matrix of Reduced Form coefficients estimated by OLS. However, if P is estimated by LAD, the GILN is LAD-GILN, which may perform better than LS-GILN if the structural disturbances are non-normal or infested with outliers.

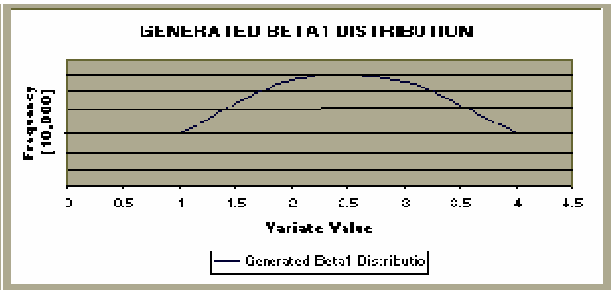
Distribution of Disturbance: LS-based estimators of linear econometric model with normal disturbance have several merits. But for non-normal disturbances they may lose these merits. Therefore, in this study we compare the performance of candidate (LAD-based) estimators with LS-based estimators for the linear models with normally as well as non-normally distributed disturbances. Among the non-normal distributions we have chosen (i) $Beta_1$, (ii) $Beta_2$ (iii)

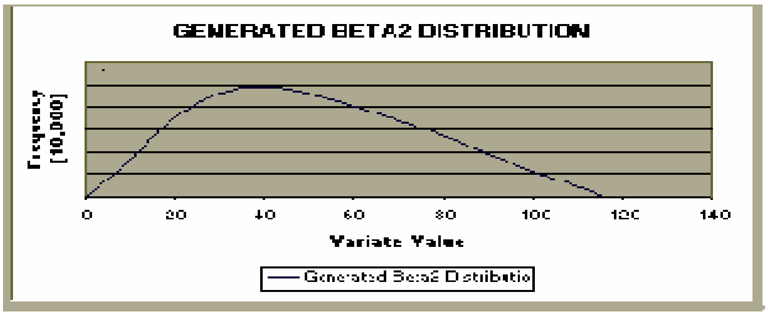
Cauchy, and (iv) Gamma distributions. Among these, Gamma is a very skewed distribution for small shape and scale parameters. Gamma variates are non-negative. Beta variates are non-negative and largely plato-kurtic. Notoriety of Cauchy variates are well known. The 2nd and higher moments of Cauchy distributed variates do not exist (in the population), though in samples they may be computed, but are unpredictable as they vary wildly from sample to sample. Yet, Cauchy distributed variates have zero mean and it is a symmetric distribution. Graphic presentations of these (generated) distributions give a fairly representative view of their nature.











The Process of Generating Random Variates with different Distributions: Variates following these distributions are computer generated. The program runs the following process (and exploits the relevant theorems regarding their nature). First, uniformly distributed random numbers are generated by some suitable method. We have used the Power Residue method (**Krishnamurthy & Sen**, 1976, pp. 303-307) using a 16-bit processor. Uniformly distributed random numbers may be transformed into normally distributed random numbers, N(0,1), by the transformation $x = \sqrt{\{-2\ln(u_1)\}}\{Cos(2\pi u_2)\}$ where u_1 and u_2 are uniformly distributed independent random numbers lying between (0,1) and x is the standard normal variate (**Knuth** (1969), **Texas Instruments Inc** (1979), p 54). Alternatively, one may generate N(0,1) from uniformly distributed u(0,1) numbers, by using the Central Limit Theorem (**Gillett** (1979) p. 519, **Kapur &**

Saxena (1982, p. 386). However, this method is less accurate and time consuming than Knuth's method. Normally distributed variate, x, may be used to generate Gamma distributed variate, g, since, if x is a standard normal variate, then g $=\frac{x^2}{2}$ is a Gamma variate with parameter $\frac{1}{2}$ (**Kapur** & **Saxena**, 1982; p 288). Due to the additive property of Gamma variates, if x_i (i=1,2,…,n) are n independent normal variates with means m_i and standard deviations σ_i then g $= \frac{1}{2}\sum_{i=1}^{n}\frac{(x_i - m_i)^2}{\sigma_i^2}$ is a Gamma variate with parameter $\frac{1}{2}n$ (**Kapur** & **Saxena**, 1982; p. 289). From two independent normally distributed variates x_1 and x_2 we may obtain a Cauchy distributed variate, since, if x_1 and x_2 are independent normal variates with means m_1 and m_2 and variances σ_1^2 and σ_2^2 then the variate $c = \frac{x_1 - m_1}{x_2 - m_2}$ is Cauchy distributed (**Kapur** & **Saxena**, 1982; p. 427). In particular, the quotient of two independent standard normal variates is Cauchy distributed. From two independent Gamma variates, g_1 and g_2 with parameters *l* and *m* respectively, we may obtain $v_1 = \frac{g_1}{g_1 + g_2}$, which is a $\beta_1(l,m)$ distributed variate, and $v_2 = \frac{g_1}{g_2}$, which is a $\beta_2(l,m)$ distributed variate (**Kapur** & **Saxena**, 1982; p. 292). In general, starting from uniformly distributed variates, we may obtain a variate with almost any kind of distribution by a sequence of suitable transformations.

The Design of Monte Carlo Experiments with Single Equation Linear Models: A single equation linear econometric model is y = X**a** + e + O, where e is the disturbance term with a specified distribution and O is the sparse vector of outliers with a specified number of nonzero elements lying within a specified range and spaced randomly over the sample observations. X(n,m) is the (fixed) non-stochastic matrix, n referring to the sample size and m referring to the number of explanatory variables (model size). We have experimented with two sample sizes; n = 20 and n = 50. Models are of three different sizes; m =2 (Model_1) , m = 4 (Model_2) and m = 6 (Model_3). Disturbances (e) are of five different distribution types; normal, Cauchy, Gamma(2), Beta_1(2,2) and Beta_2(2,2). Disturbances have standard deviations = 0.1 relative to the mean magnitude of the dependent variable, y. The numbers of outliers (non-zero elements in O) are: 0 (i.e. absence of outliers), 1, 3 and 5. The ranges of magnitudes of outliers (when present) are (0, 0.5) , (0, 1) and (0, 2) times the mean value of the dependent variable, y, in a particular model.

To obtain $\tilde{a}$ by minimizing $S_1 = \sum_{i=1}^{n}\left| y_i - \sum_{j=1}^{m} a_j x_{ij} \right|$,where $\tilde{a}$ is the LAD estimator of **a** in y = X**a** + e + O, we have used **Fair** (1974) algorithm. The algorithm is iterative in nature and exploits the Brouwer-Kakutani fixed point theorem . We set the upper limit of iteration (to reach convergence) to 100. Accuracy in estimation is limited to at least 0.0001. For all computations we have used double precision arithmetic in FORTRAN 77 (16 bit arithmetic).

As a digression though, it is pertinent to mention as to the reason for choosing the Fair algorithm. Minimization of S_1 was tried with a number of non-linear optimization methods such as Hooke-Jeeves, Nelder-Mead, Powell and Rosenbrock algorithms. These are principal derivative free multi-variable non-linear search algorithms. FORTRAN source codes of these algorithms are available in **Kuester** & **Mize** (1973). The Random Walk algorithm (**Rao**, 1978, pp. 252-257) also was programmed in FORTRAN 77 and run. However, these algorithms did not succeed in searching the optimal values of **a** (in $\sum_{i=1}^{n}\left|y_i - \sum_{j=1}^{m} a_j x_{ij}\right|$) that minimized S_1. Fair's algorithm performed much better. However, when Fair's algorithm yielded near-optimal values of **a**, the Random Walk method often improved them (brought the solutions further near to the optimal values). This exercise gave us two successful methods of optimization of S_1 : (i) Fair's algorithm (LAD) and (ii) Fair-Random Walk algorithm (LADRW).

For every experiment hundred (100) replicates were used – that is, with fixed X**a**, disturbances, e and outliers, O were generated 100 times. Based on these 100 replicates, RMS_1 = $[\frac{1}{100}\sum_{k=1}^{100}\{\sum_{j=1}^{m}(\tilde{a}_{kj} - a_j)^2\}]^{0.5}$, $\tilde{a}$ being the LAD estimator of **a**, was obtained. The OLS estimators of **a** (that is, $\hat{a}$) was obtained as $\hat{a}$=(X'X)$^{-1}$X'y for each of the 100 replicates and from this RMS_2 = $[\frac{1}{100}\sum_{k=1}^{100}\{\sum_{j=1}^{m}(\hat{a}_{kj} - a_j)^2\}]^{0.5}$ was obtained. The quotient of the two RMS or ρ = $\frac{RMS_2}{RMS_1}$ was obtained. If, $\rho > 1$, it implies that LAD (or LADRW as the case may be) has outperformed OLS and vice versa. Such 300 experiments were carried out with different parameters. The distribution of the number of experiments according to various parameters of experiments are given below. The figures within the parentheses are the number of experiments carried out.

Size of Model: Model#1 (100), Model#2 (100), Model#3 (100)
Sample Size: Small i.e. 20 observation (150), Larger, i.e. 50 observations (150)
Distributions: Normal (60), Cauchy (60), Gamma (60), B_1 (60), B_2 (60)
Number of Outliers: None (120), One (60), Three (60), Five (60)
Size of Outliers: None/zero (120), Small (90), Large (90).

Findings of Monte Carlo Experiments with Single Equation Linear Models: First we present the frequency distribution of success of different candidate estimators in dominating OLS as tabulated in the tables B.1 through B.5. The entries in the cells of these tables are the number of success (out of the total number of experimental trials or cases - given in the rightmost columns of the tables) and where the ratio of the number of success to the number of trials (cases) is larger than 0.5, the entry has been highlighted. We observe the following:

(1). Among the different disturbance distributions only Cauchy and B_2 favour LAD (or LADRW) estimators against the OLS estimator. Among the distributions, Cauchy is absolutely dominating.
(2). No outlier or single outlier cases give no edge to the LAD estimator against the OLS estimator, but for larger number (3 or 5) of outliers LAD outperforms OLS. However, LADRW performs better than OLS in presence of outliers (1, 3 or 5) while it has no edge over OLS if no outlier is present. LAD and LADRW outperform OLS for whatever size of outliers.

We conclude therefore that Cauchy or B_2 distributed errors favour LAD estimator against the OLS estimator. Performance of LAD (or LADRW) in case of Cauchy distributed disturbances is spectacular. We observe that out of 60 experiments with Cauchy distributed disturbance term LAD outperforms OLS in 59 cases. LADRW outperforms OLS in 60 out of 60 experiments.

TABLE B1. EFFECT OF DISTRIBUTION

Sl. No.	DISTBUTION	LAD	LADRW	CASES
1	n	18	23	60
2	c	**59**	**60**	60
3	g	21	24	60
4	v	10	14	60
5	b	**53**	52	60

TABLE B2. EFFECT OF SAMPLE SIZE

Sl. No.	SAMPLE SIZE	LAD	LADRW	CASES
1	s	73	81	150
2	l	**88**	**92**	150

TABLE B3. EFFECT OF MODEL SIZE

Sl. No.	MODEL SIZE	LAD	LADRW	CASES
1	2	49	**58**	100
2	4	**56**	**58**	100
3	6	**56**	**57**	100

TABLE B4. EFFECT OF NO. OF OUTLIERS

Sl. No.	NO. OF OUTLIERS	LAD	LADRW	CASES
1	0	41	41	120
2	1	30	**32**	60
3	3	**42**	**49**	60
4	5	**48**	**51**	60
Total (AT LEAST ONE OUTLIER)		**120**	**132**	180

TABLE B5. EFFECT OF SIZE OF OUTLIERS

Sl. No.	OUTLIER SIZE	LAD	LADRW	CASES
1	0	10	10	30
2	a	11	10	30
3	b	10	10	30
4	c	10	11	30
Total (ZERO OUTLIERS)		41	41	120
5	**s**	**43**	**51**	90
6	**b**	**77**	**81**	90

NOTE: In 0,a,b,and c categories SIZE of OUTLIERS = 0

Probit Analysis: For a comprehensive analysis of the results of our experiments we construct a variable, R, which takes on only two values, zero or unity. R_i =1 if ρ (= $\frac{RMS_2}{RMS_1}$)>1, zero otherwise. In other words, R takes on a value of unity if LAD (or LADRW) outperforms OLS, else R = 0. We go in for the *Probit analysis* of R. Since we have carried out 300 experiments, R is a vector of 300 elements. This vector may be considered as a sample R from the population ℝ since one may carry out many more experiments to obtain a still larger number of ρ to construct R with many more elements. Then we postulate that the value of R_i is conditioned by the model specification in terms of sample size, model size, distributional specifications of the disturbances and the number and size of outliers that make up the dependent variable y in y = X**a** + e + O. Therefore, in the Probit analysis, we have the following explanatory variables:

N (Normal Distribution): A binary variable, taking on a value of unity if the disturbance term follows a normal distribution, zero otherwise.

G (Gamma Distribution): A binary variable, taking on a value of unity if the disturbance term follows a Gamma distribution, zero otherwise.

B_1 ($Beta_1$ Distribution): A binary variable, taking on a value of unity if the disturbance term follows a $Beta_1$ distribution, zero otherwise.

B_2 ($Beta_2$ Distribution): A binary variable, taking on a value of unity if the disturbance term follows a B_2 distribution, zero otherwise.

NOBSRV (Sample size): A binary variable, taking on a value of unity if the number of observations has a larger size = 50, zero otherwise.

COEFF (Size of the model): This variable takes on the value of 2 for a model with two explanatory variables, 4 for a model of four explanatory variables and 6 for a model with 6 explanatory variables.

NOUT (No. of Outliers): This variable takes on the value of 1 if the disturbance contains one outlier, 3 if the disturbance contains three outliers, 5 if the disturbance contains five outliers, and 0 if the disturbance contains no outlier.

OSIZE (Outliers of a large size): To recapitulate (what has been mentioned earlier), in the experiments we have used four different factors (call k) for generating outliers of different sizes. For small samples we use two values of k (0.5 and 1.0). For large samples we use k=1.0 and k= 2.0. The size of outliers is k*mean(y) or k times the mean of the dependent variable. The specified number of outliers are added to different observations of the dependent variable (y) at randomly selected locations (cases or observation). The value of the decision variable OSIZE takes on these values of k. Thus OSIZE takes on three different values – 0.5, 1.0 and 2.0 depending on the specified conditions.

Binary coding of distribution variables necessitate that any four (out of 5 namely, N, C, G, B_1 and B_2) of them at most can at a time be included in the list of explanatory variables, since all five, if included, make up for a perfect multicollinearity. In view of the spectacular performance of LAD and LADRW for Cauchy distributed disturbance term (exhibiting 59 and 60 successes respectively out of 60 experiments) we have chosen to drop out Cuachy from the list of explanatory variables. Accordingly, our general Probit Regression model may be described as follows:

$$R_i = k_0 + a_1 N + a_2 G + a_3 B_1 + a_4 B_2 + b_1 NOBSRV + b_2 COEFF + c_1 NOUT + c_2 OSIZE + error$$

where R_i (binary measure of the success/failure of one of the estimators, namely LAD or LADRW).

The results of Probit Analysis are tabulated in the following tables (S1 through S6). It is especially to mention that inclusion of N (normal disturbance) and B_1 ($Beta_1$ disturbance) together as the explanatory variables in the Probit regression makes the estimation procedure unstable due to ill conditioned variance-covariance matrix of explanatory variables. Hence, only one of them (at a time) may be included in the list of explanatory variables. Accordingly, we find that N, B_1 and G have depressing effect on LAD and LADRW estimators. B_2 gives an edge to LAD (as well as LADRW) estimator over OLS estimator. Presence and size of outliers favour LAD (as well as LADRW) estimator. Sample size (n = NOBSRV) and model size (m = COEFF) do not matter.

Results of Probit Analysisof Performance of LAD (LADRW) Estimator of Single Equation Linear Econometric Models

Table S1: Binary LAD Probit regression; **(No of 0's:139 No. of 1's:161: R=0.62236)**								
	Const	N	G	B2	NOBSRV	COEFF	NOUT	OSIZE
Estimate	-0.7442	-0.9511	-0.7797	1.0905	0.0684	0.0575	0.1326	0.8837
Std.Err.	0.2699	0.2415	0.2295	0.2651	0.1750	0.0504	0.0557	0.1981
t(292)	-2.76	-3.94	-3.40	4.11	0.39	1.14	2.38	4.46
p-level	0.006	0.000	0.001	0.000	0.696	0.255	0.018	0.000

Table S2: Binary LADRW Probit regression; **(No of 0's:139 No. of 1's:161: R=0.62692)**								
	Const	N	G	B2	NOBSRV	COEFF	NOUT	OSIZE
Estimate	-0.3368	-0.8693	-0.8114	0.9751	0.0266	-0.0176	0.1980	0.8430
Std.Err.	0.2551	0.2448	0.2403	0.2790	0.1710	0.0513	0.0637	0.1949
t(292)	-1.32	-3.55	-3.38	3.49	0.16	-0.34	3.11	4.32
p-level	0.188	0.000	0.001	0.001	0.876	0.732	0.002	0.000

Table S3: Binary LAD Probit regression; **(No of 0's:139 No. of 1's:161: R=0.68728)**								
	Const	G	B1	B2	NOBSRV	COEFF	NOUT	OSIZE
Estimate	-0.6871	-0.9962	-1.7798	0.9574	0.0430	0.0747	0.1781	0.8717
Std.Err.	0.2574	0.2217	0.2885	0.2437	0.1694	0.0497	0.0568	0.1707
t(292)	-2.67	-4.49	-6.17	3.93	0.25	1.50	3.13	5.11
p-level	0.008	0.000	0.000	0.000	0.800	0.134	0.002	0.000

Table S4: **Binary LADRW Probit regression (No of 0's:127 No. of 1's:173: R=0.70076)**								
	Const	G	B1	B2	NOBSRV	COEFF	NOUT	OSIZE
Estimate	-0.17200	-1.27654	-2.00777	0.77401	-0.06756	-0.00513	0.26018	1.02743
Std.Err.	0.24452	0.31956	0.39401	0.25780	0.16956	0.05177	0.07238	0.21530
t(292)	-0.70	-3.99	-5.10	3.00	-0.40	-0.10	3.59	4.77
p-level	0.482	0.000	0.000	0.003	0.691	0.921	0.000	0.000

Table S5: Binary LAD Probit regression; **(No of 0's:139 No. of 1's:161: R=0.68417)**						
	Const	G	B1	B2	NOUT	OSIZE
Estimate	-0.36403	-0.97404	-1.74530	1.01170	0.16984	0.87214
Std.Err.	0.13320	0.21575	0.27821	0.25773	0.05404	0.16102
t(294)	-2.73	-4.51	-6.27	3.93	3.14	5.42
p-level	0.007	0.000	0.000	0.000	0.002	0.000

Table S6: **Binary LADRW Probit regression (No of 0's:127 No. of 1's:173: R=0.70055)**					
	Const	G	B1	B2	NOUT
Estimate	-0.22111	-1.27045	-2.00753	0.76828	0.26540
Std.Err.	0.13276	0.30780	0.38318	0.24641	0.07077
t(294)	-1.67	-4.13	-5.24	3.12	3.75
p-level	0.097	0.000	0.000	0.002	0.000

The Design of Monte Carlo Experiments with Multi-Equation Linear Models: A multi-equation linear econometric model is YA + XB + E + O = 0 where E is the disturbance matrix with a specified distribution and O is the sparse matrix of outliers with a specified number of nonzero elements lying within a specified range and spaced randomly over the sample observations of different structural equations. Y(n,m) is the matrix of n sample observations on m endogenous variables and X(n,k) is the (fixed) non-stochastic matrix of n sample observations on k pre-determined (exogenous) variables. The number of equations in the model, m, determines the model size. We have experimented with two sample sizes; n = 20 and n = 50. Models are of three different sizes; m =3 ($Model_1$ with 10 parameters) , m = 5 ($Model_2$ with 24 parameters) and m = 7 ($Model_3$ with 54 parameters). Disturbances (E) are of five different distribution types; normal, Cauchy, Gamma(2), $Beta_1(2,2)$ and $Beta_2(2,2)$. Disturbances have two alternative standard deviations, very small (0.001) and sufficiently large (0.3) relative to the mean magnitude of the dependent variable in a particular (say j^{th}) structural equation, y_j. The numbers of outliers (non-zero elements in O_j) are: 0 (i.e. absence of outliers), 1, 2 and 3. The ranges of magnitudes of outliers (when present) are (0, 0.5) , (0, 1) and (0, 1.5) times the mean value of the dependent variable, y_j, in the j^{th} equation of the model.

For every experiment fifty (50) replicates were used. Such 630 experiments were carried out with different parameters. **Fair**'s algorithm with double precision arithmetic (16 bit processor) was programmed in FORTRAN 77. The distribution of the number of experiments according to various parameters of experiments are given below. The figures within the parentheses are the number of experiments carried out.

Size of Model: Model#1 (210), Model#2 (210), Model#3 (210)
Sample Size: Small i.e. 20 observation (315), Larger, i.e. 50 observations (315)
Distributions: Normal (126), Cauchy (126), Gamma (126), B_1 (126), B_2 (126)
Standard Deviation of Disturbance Vector: Small (180), Large (450)
Number of Outliers: None (360), One (90), Two (90), Three (90)
Size of Outliers: None/zero (360), Small (90), Medium (90), Large (90).

Candidate Estimators: We have six candidate estimators, LS-LS (2-SLS of Theil-Basmann), LS-LAD, LAD-LS, LAD-LAD (D2SLAD of Amemiya), LS-GILN (of Khazzoom) and LAD-GILN. The 2-SLS is the referent to compare the performance of the other five estimators.

Findings of Monte Carlo Experiments with Multi-Equation Linear Models: First we present the frequency distribution of success of different candidate estimators in dominating 2-SLS as tabulated in the tables C1 through C6. The entries in the cells of these tables are the number of success (out of the total number of experimental trials or cases - given in the rightmost columns of the tables) and where the ratio of the number of success to the number of trials (cases) is larger than 0.5, the entry has been highlighted. We observe the following:

Effect of Model Size: For M_3 (7-equations model) LS-LAD and LAD-LAD estimators outperform 2-SLS.
Effect of Sample Size: No estimator outperforms 2-SLS.
Effect of Distribution of Disturbance term: For Cauchy and B_2 (named as B) distributed disturbance terms LAD-LAD and LAD-GILN outperform 2-SLS.
Effect of Size of Standard Deviation of Disturbance term: No estimator outperforms 2-SLS.
Effect of No. of Outliers: For 2 and 3 outliers LAD-LAD outperforms 2-SLS.
Effect of Size of Outliers: For medium and large size outliers LAD-LAD estimator outperforms 2-SLS.

Table C1: Effect of Model Size

	Ls_Lad	Lad_Ls	Lad_Lad	LS_Giln	Lad_Giln	Cases
M_1	00	04	47	23	61	210
M_2	12	04	62	40	94	210
M_3	**123**	100	**153**	42	78	210

Table C2: Effect of Sample Size

	Ls_Lad	Lad_Ls	Lad_Lad	LS_Giln	Lad_Giln	Cases
S	91	82	145	59	90	315
L	44	26	117	46	143	315

Table C3: Effect of Distribution

	Ls_Lad	Lad_Ls	Lad_Lad	LS_Giln	Lad_Giln	Cases
N	20	25	32	18	13	126
C	41	21	**81**	22	**85**	126
G	16	22	29	29	14	126
V	23	19	34	18	29	126
B	35	21	**86**	18	**92**	126

Table C4: Effect of Standard Deviation

	Ls_Lad	Lad_Ls	Lad_Lad	LS_Giln	Lad_Giln	Cases
S	35	14	56	31	70	180
L	100	94	206	80	163	450

Table C5: Effect of No. of Outliers

	Ls_Lad	Lad_Ls	Lad_Lad	LS_Giln	Lad_Giln	Cases
E	35	14	56	29	70	180
0	49	41	57	28	52	180
1	20	19	38	05	32	90
2	14	16	**50**	17	41	90
3	17	18	**51**	26	38	90

Table C6: Effect of Size of Outliers

	Ls_Lad	Lad_Ls	Lad_Lad	LS_Giln	Lad_Giln	Cases
0	21	12	24	07	26	60
1	12	07	19	06	21	60
2	12	08	21	10	19	60
3	17	09	21	14	22	60
4	11	11	20	11	16	60
5	11	08	18	09	18	60
S	19	19	26	06	29	90
M	16	18	**47**	14	40	90
B	16	16	**66**	28	42	90

Entries are the frequencies while relative norms are greater than unity in magnitude.

Probit Analysis: For a comprehensive analysis of the results of our experiments, we construct a variable, R, which takes on only two values, zero or unity. $R_C = 1$ if ρ ($= \frac{RMS_{2-SLSC}}{RMS_C}$)>1, zero otherwise. In other words, R takes on a value of unity if a particular candidate estimator (LS-LAD, LAD-LS, LAD-LAD, LS-GILN and LAD-GILN) outperforms 2-SLS, else R = 0. We go in for the *Probit analysis* of R_C. Since we have carried. out 630 experiments, R is a vector of 630 elements for each candidate estimator. This vector may be considered as a sample R from the population R since one may carry out many more experiments to obtain a still larger number of ρ to construct R with many more elements. Then we postulate that the value of R_i is conditioned by the model specification in terms of sample size, model size, distributional specifications of the disturbances and the number and size of outliers that make up the dependent variables Y in YA + XB + E + O.

Thus, we have five dependent variables, each binary in nature, representing the success (1) or failure (0) of the estimator in outperforming 2-SLS. On the other hand, specifications may be used as the independent variables that explain the variations in the dependent variables. Accordingly, we have used the following independent (explanatory) variables that represent the model specifications:

N (Normal Distribution): A binary variable, taking on a value of unity if the disturbance term follows a normal distribution, zero otherwise.

C (Cauchy Distribution): A binary variable, taking on a value of unity if the disturbance term follows a Cauchy distribution, zero otherwise.

G (Gamma Distribution): A binary variable, taking on a value of unity if the disturbance term follows a Gamma distribution, zero otherwise.

B_1 ($Beta_1$ Distribution): A binary variable, taking on a value of unity if the disturbance term follows a $Beta_1$ distribution, zero otherwise.

B_2 ($Beta_2$ Distribution): A binary variable, taking on a value of unity if the disturbance term follows a B_2 distribution, zero otherwise.

SSD (Disturbance term with a small Standard Deviation): A binary variable, taking on a value of unity if the disturbance term has a small standard deviation = 0.001, zero otherwise.

BSD (Disturbance term with a larger Standard Deviation): A binary variable, taking on a value of unity if the disturbance term has a larger standard deviation = 0.3, zero otherwise.

OSS (Outliers of a small size): A binary variable, taking on a value of unity if the disturbance term has a small size outlier = 0.50 times the mean of the dependent variable, zero otherwise.

OSM (Outliers of a medium size): A binary variable, taking on a value of unity if the disturbance term has a medium size outlier = 1.00 times the mean of the dependent variable, zero otherwise.

OSB (Outliers of a large size): A binary variable, taking on a value of unity if the disturbance term has a medium size outlier = 1.50 times the mean of the dependent variable, zero otherwise.

MODEL (Size of the model): This variable takes on the value of 3 for a model with three equations, 5 for a model of 5 equations and 7 for a model with 7 equations.

OBSERV (Sample size): A binary variable, taking on a value of unity if the number of observations has a larger size = 50, zero otherwise.

NOUT (No. of Outliers): This variable takes on the value of 1 if the disturbance contains one outlier, 2 if the disturbance contains two outliers, 3 if the disturbance contains three outliers, and 0 if the disturbance contains no outlier.

Of distribution variables any four (out of 5 namely, N, C, G, B_1 and B_2) at most can at a time be included in the list of explanatory variables, since all five, if included, make up for a perfect multicollinearity. Similarly, either SSD or BSD can be included among the explanatory variables since they together make up for a perfect multicollinearity. Accordingly, our general Probit Regression model may be described as follows:

$$R_i = k_0 + a_1 N + a_2 C + a_3 G + a_4 B_1 + a_5 B_2 + b_1 SSD + b_2 BSD + c_1 OSS + c_2 OSM + c_3 OSB + r_1 MODEL + r_2 OBSERV + r_3 NOUT + error$$

where R_i (binary measure of the success/failure of one of the estimators, namely LS-LAD, LAD-LS, LAD-LAD, LS-GILN or LAD-GILN). There would be some zero restrictions on the coefficients (implying absence of some particular explanatory variables), These restrictions are: (i) at least one of the a_j will be zero and (ii) either b_1 or b_2 (or both) will be zero. Other coefficients may be zero or non-zero.

The results of Probit Analysis indicate that except LAD-GILN and LAD-LAD, other estimators do not outperform 2-SLS in any significant manner. The probability of success for LS-LAD, LAD-LS and LS-GILN are 135/630 (21.43%), 108/630 (17.14%) and 111/630 (17.62%) respectively. Even in the experiments with outliers (270 in number or 42.86% of the total number of experiments), the probability of their success is less than 0.5. However, LAD-GILN and LAD-LAD have probability of success 233/630 (36.98%) and 262/630 (41.59%) respectively. This indicated that in presence of outliers or non-normal disturbance, they outperform 2-SLS in the majority number of cases as we have seen in the tables above.

We report here the most significant tables and results only. The summary table indicates that of eleven LAD-LAD Probit models, #6 and and #7 are the best (followed by LAD-GILN Probit model #5 – among the seven alternative models). *LAD-LAD outperforms 2-SLS for Cauchy and B_2 distributed disturbance term with small/large SD containing small/medium/large outliers for smaller as well as larger models of whatever sample size*. This result is consistent with the performance of LAD in estimation of the parameters of single equation models. LAD-GILN is the second best estimator.

Probit Analysis of Performance of the Most Powerful Estimators of Multi-Equation Linear Models

LAD-GILN MODEL - 5

(Final loss: 88.972972053 R=.62772 Variance explained: 39.403%)

	Const.B0	C	B2	OSM	OSB	OBSERV
Estimate	-3.65119	2.370894	2.585693	.439568	.503179	.048871
Std.Err.	.57311	.331435	.356232	.190725	.200297	.010075
t(624)	-6.37082	7.153420	7.258445	2.304728	2.512164	4.850489
p-level	.00000	.000000	.000000	.021509	.012251	.000002

A GIST OF LAD-GILN MODELS

[s (x) = Significant (insignificant) at 5% - or less level of Significance]

MODEL	**1**	**2**	**3**	**4**	**5***	**6**	**7**
N	-	-	-	-	-	-	-
C	s	s	s	s	S	S	S
G	x	x	x	-	-	-	-
B1	s	s	s	-	-	-	-
B2	s	s	s	s	S	S	S
SSD	-	-	-	-	-	S	-
BSD	s	s	-	-	-	-	S
OSS	x	-	-	-	-	X	X
OSM	s	s	s	s	S	S	S
OSB	s	s	s	s	S	S	S
MODEL	x	x	x	x	-	X	X
OBSERV	s	s	s	s	S	S	S
NOUT	-	-	-	-	-		

LAD-LAD MODEL - 6

(Final loss: 79.626177524 R=.69261 Variance explained: 47.971%)

	Const.B0	C	B2	SSD	OSS	OSM	OSB	MODEL	OBSERV
Estimate	-3.51815	1.719273	1.910865	-0.34022	-0.52395	0.746808	1.759795	0.587351	-0.01815
Std.Err.	0.37768	0.20953	0.222927	0.14911	0.20522	0.19051	0.239213	0.063254	0.00431
t(621)	-9.31523	8.205394	8.571704	-2.28173	-2.55314	3.920048	7.356608	9.285591	-4.20674
p-level	0	0	0	0.02284	0.01091	0.000098	0	0	0.00003

LAD-LAD MODEL - 7

(Final loss: 79.626177524 R=.69261 Variance explained: 47.971%)

	Const.B0	C	B2	BSD	OSS	OSM	OSB	MODEL	OBSERV
Estimate	-3.85837	1.719272	1.910865	0.34022	-0.52395	0.746808	1.759795	0.587351	-0.01815
Std.Err.	0.41113	0.20955	0.222953	0.149115	0.20522	0.190503	0.239212	0.063265	0.00431
t(621)	-9.3849	8.204589	8.570716	2.281593	-2.55303	3.920198	7.356643	9.284008	-4.20676
p-level	0	0	0	0.022851	0.01092	0.000098	0	0	0.00003

A GIST OF LAD-LAD MODELS

[s (x) = Significant (insignificant) at 5% - or less level of Significance]

MODEL	1	2	3	4	5	6*	7*	8	9	10	11
N	-	-	-	-	x	-	-	-	-	-	-
C	s	s	s	s	s	s	s	s	s	s	s
G	x	x	x	x	-	-	-	x	x	x	-
B1	x	-	-	x	-	-	-	x	x	x	-
B2	s	s	s	s	s	s	s	s	s	s	s
SSD	s	s	-	-	s	s	-	-	-	-	-
BSD	-	-	s	s	-	-	s	x	x	-	-
OSS	s	s	x	s	s	s	s	-	-	-	-
OSM	s	s	x	s	s	s	s	s	s	s	s
OSB	s	s	s	s	s	s	s	s	s	s	s
MODEL	s	s	s	s	s	s	s	s	s	s	s
OBSERV	s	s	x	s	s	s	s	s	s	s	s
NOUT	-	-	-	-	-	-	-	x	-	-	-

- ***means a variable not appearing in the model.***

A Summary Table on Performance of Estimators of Multi-Equation Models with their Explanatory Variables [*s (x) = Significant (insignificant) at 5% - or less level of Significance*]			
Estimator	**LAD-LAD**		**LAD-GILN**
Criteria/Variables	**Model 6***	**Model 7***	**Model 5 ***
N	-	-	-
C	s	s	s
G	-	-	-
B1	-	-	-
B2	s	s	s
SSD	s	-	-
BSD	-	s	-
OSS	s	s	-
OSM	s	s	s
OSB	s	s	s
MODEL	s	s	-
OBSERV	s	s	s
NOUT	-	-	-
Note: - means a variable not appearing in the model.			

Appendix

Sporadic errors in X (where sample X is true X + Ξ, and Ξ is a sparse matrix with nonzero elements substantial in size) also vitiate OLS estimation. There too, LAD performs well. Since it does not make a part of our stated objectives, but we have carried out some experiments on this too, we find it appropriate to relegate this finding to the appendix. To describe the problem we explain it with an example. Let true Y (=Y*) be $Y^* = 19.7 + 9.6\ X^*_1 + 11.1X^*_2 + 6.2X^*_3$; $Y = Y^* + E$, where E is drawn from a normal population with mean = 0 and $\sigma = 10$ Table A1). The regression coefficients (OLS and LAD estimated) are given in table A2. Four experiments are done by adding Ξ_1 through Ξ_4 to X* such that $X = X^* + \Xi_k$ (k=1,2,3,4). Results of OLS and LAD estimation for the four experiments are given in table A4. We find that LAD outperforms OLS.

Table A1: Basic Variables in the Regression Model						
Sl. No.	True X* Matrix			True Y	Error	Stochastic Y
	X^*_1	X^*_2	X^*_3	Y*	E	Y=Y*+E
1	60	21	75	1293.8	-15.58	1278.22
2	47	45	80	1466.4	-3.534	1462.866
3	54	35	90	1484.6	-5.238	1479.362
4	22	0	27	398.3	13.635	411.935
5	72	48	85	1770.7	-14.191	1756.509
6	9	44	93	1171.1	-15.17	1155.93
7	9	24	42	632.9	1.419	634.319
8	67	0	15	755.9	2.603	758.503
9	34	69	7	1155.4	17.965	1173.365
10	51	38	97	1532.5	16.363	1548.863
11	17	96	49	1552.3	4.337	1556.637
12	64	43	71	1551.6	-10.142	1541.458
13	3	86	39	1244.9	-5.09	1239.81
14	90	77	25	1893.4	-8.986	1884.414
15	86	36	10	1306.9	9.858	1316.758
16	22	81	72	1576.4	3.446	1579.846
17	56	19	7	811.6	11.419	823.019
18	76	58	36	1616.3	-11.32	1604.98
19	21	80	61	1487.5	-5.25	1482.25
20	59	14	96	1336.7	13.457	1350.157

Table A2: Regression Coefficients of $Y = b_0 + b_1X^*_1 + b_2X^*_2 + b_3X^*_3$							
	OLS Estimates					True Coeff	LAD Estimates
	$\hat{b}$	S.Err	Beta	't' value	Sig.	b*	$\tilde{b}$
X_0	35.842	8.564	-	4.185	.001	19.70	33.144
X_1	9.506	.095	.667	100.218	.000	9.60	9.481
X_2	10.993	.089	.819	123.871	.000	11.10	10.986
X_3	6.072	.077	.505	78.756	.000	6.20	6.082

Table A3: Sporadic Disturbance Matrix (Ξ Matrix) Added to X*												
Sl No.	Ξ_1 added to			Ξ_2 added to			Ξ_3 added to			Ξ_4 added to		
	X^*_1	X^*_2	X^*_3	X^*_1	X^*_2	X^*_3	X^*_1	X^*_2	X^*_3	X^*_1	X^*_2	X^*_3
1	0	0	0	20	0	0	0	0	0	0	0	0
2	10	0	0	0	0	0	0	0	0	70	0	0
3	0	0	0	0	0	0	40	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	100	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0
8	0	20	0	0	0	0	0	0	0	0	0	0
9	30	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	60	0	0	0	0	0	0
11	0	0	0	0	0	0	70	0	0	0	0	0
12	0	0	0	0	0	0	0	0	100	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0
14	0	30	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	100	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	200	0

Table A4: Regression Coefficients of $Y = (X^*+\Xi_i)B + E$					
Sporadic Disturbance in X	LAD and OLS Estimators	Regression Coefficients Estimated by LAD and OLS with $X = X^* + \Xi$			
	Coefficients (True)	B0 (19.70)	B1 (9.60)	B2 (11.10)	B3 (6.20)
Ξ_1	LAD	24.367	8.791	10.762	6.759
	OLS	75.506	7.461	10.695	6.001
Ξ_2	LAD	44.068	9.445	10.975	5.928
	OLS	278.410	9.649	7.708	3.054
Ξ_3	LAD	55.953	9.221	11.105	5.671
	OLS	344.804	5.628	10.517	3.034
Ξ_4	LAD	44.440	9.422	10.948	5.962
	OLS	670.396	5.190	2.943	3.942

Bibliography and References

Abdelmalek, N N (1974). On the Discrete Linear L_1 Approximation and L_1 Solution of Over- determined Linear Equations, *Journal of Approximation Theory,* Vol. 11, pp. 35 – 53.

Amemiya, T (1982). The Two-Stage Least Absolute Deviation Estimators, *Econometrica*, Vol. 50 (1-3), pp. 689-711.

Ashar, V C & T D Wallace (1963). A Sampling Study of Minimum Absolute Deviation Estimators, *Operations Research*, Vol. 11, pp. 747-758.

Barrodale, I (1968). L_1 Approximation and the Analysis of Data, Applied. Statistics, Vol. 17, pp. 51-57.

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

Basmann, R L (1975). A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation, *Econometrica*, Vol. 25, pp. 77-84.

Basset, G Jr. & R. Koenkar (1978). Asymptotic Theory of Least Absolute Error Regression, *Journal of American Statistical Association*, Vol. 73, pp. 618-622.

Blattberg, R & T Sargent (1971). Regression with Non-Gaussian Stable Disturbances: Some Sampling Results, *Econometrica*, Vol. 39, pp. 501 – 510.

Charnes, A & C E Lemke (1954). Computational Theory of Linear Programming: The Bounded Variables Problem, *Graduate School of Industrial. Administration, Carnegie Inst. of Technol*., Pittsburg, Pennsylvania.

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

Chen, X (1996). On a Problem of Strong Consistency of Least Absolute Deviation Estimates. *Statistical Sinica*, Vol. 6, pp. 481-489.

Cragg, J (1967). On The Relative Small-Sample Properties of Several Structural Equation Estimators, *Econometrica,* Vol. 35, pp. 89 – 110.

Dasgupta, M (1992). *Degree of Over-Identification and its Effects on the Performance of Estimators*. M Phil. Dissertation, Department of Economics, NEHU, Shillong (unpub).

Edgeworth, F Y (1887). On Discordant Observations, *Philosophical Magazine*, Vol. 5(24), pp. 364 - 375.

Edgeworth, F Y (1888). On a New Method of Reducing Observations Relating to Several Quantities", *Philosophical Magazine*, Vol. 5 (25), pp. 185 – 191.

Edgeworth, F Y (1923). On the Use of Medians for Reducing Observations Relating to Several Quantities , *Philosophical Magazine,* 6^{th} Series. Pp. 1074 – 1088.

Fair, R C (1974). On Robust Estimation of Econometric Models, *Annals of Economic and Social Measurement*, Vol. 3,pp. 667 – 678.

Fair, R C (1994). *Estimating and Testing the US Model*, Yale Univ.

Fama, E F (1965). The Behaviour of Stock Market Prices, *Journal of Business*, Vol. 38, pp. 34 – 105.

Fisher, W D (1961). A Note on Curve Fitting with Minimum Deviation by Linear Programming", *Journal of American Statistical Association*, Vol. 56, pp. 359 – 361.

Froberg, C E (1964). *Introduction to Numerical Analysis*. Addison- Wesley Publishing Co. Inc. London.

Gillett, B E (1979). *Introduction to Operations Research*. Tata McGraw Hill, New Delhi.

Glahe, F R & J G Hunt (1970). The Small Sample Properties of Simultaneous Equation Least Absolute Estimators vis-a-vis Least Squares Estimators, *Econometrica,* Vol. 5(38), pp. 742– 53.
Haavelmo, T (1943). The Statistical Implications of a System of Simultaneous Equations, *Econometrica*, Vol. 11, pp. 1–12.
Haavelmo, T (1947). Methods of Measuring the Marginal Propensity to Consume. *Journal of American Statistical Association*, Vol. 42, pp. 105-22.
Hammersley, J M & D C Handcomb (1964). Monte Carlo Methods, John Wiley, New York.
Hill, R W & W P Holland (1977). Two Robust Alternatives to Least SquaresRegression, *Journal of the American Statistical Association*, Vol. 72, pp. 828-833.
Hitomi, K & M Kagihara (2001). Calculation Method for Non-Linear Dynamic Least Absolute Deviation Estimator. *J. of Japan Statist. Society*, Vol. 31(1), pp. 39-51.
Hood, W C & T C Koopmans (Ed.) (1953). Studies in Econometric Method, Cowels Foundation Monograph 14, John Wiley, New York.
Huber, P J (1964). Robust Estimation of a Location Parameter, *Annals of Mathematical Statistics*, 35 (73 – 101).
Intriligator, M D (1978). *Econometric Models, Techniques and Applications*, Prentice Hall, Englewood Cliffs, New Jersey.
Johnston, J (1963). *Econometric Methods,* McGraw-Hill, New York.
Judge, G G, W E Griffiths, R C Hill & T Lee (1980). *The Theory and Practice of Econometrics*. John Wiley, New York.
Kapur, J N & H C Saxena (1982). *Mathematical Statistics*. S Chand & Co. New Delhi.
Khazzoom, J D (1976). An Indirect Least Squares Estimator for Over identified Equations, *Econometrica*, Vol. 44 (4) , pp. 741–750.
Kim, T H & C Muller (2000). Two-Stage Quantile Regression, *Discussion Paper in Economics* #00/1, Nottingham University, January.
Knuth, D E (1969). *The Art of Computer Programming*. Addison Wesley.
Koopmans, T C, H Rubin & R B Leipnik (1950). Measuring the Equation Systems of Dynamic Economics. in T. C. Koopmans (ed.) (1950) *Statistical Inference in Dynamic Economic Models*, Wiley, New York, pp. 53 – 237.
Krishnamurthy, E V & S K Sen (1976). *Computer-Based Numerical Algorithms***.** Affiliated East-West Press, New Delhi.
Kuester, J L & J H Mize (1973). *Optimization Techniques with Fortran*. McGraw-Hill Book Company, New York.
Mandlebroth, B B (1967). The Variation of Certain Speculative Prices, *Journal of Business*, Vol. 36, pp. 394 – 419.
Medhi, N (1992). *A Study on the Behaviour of the Minimum Absolute Deviation (MAD) Regression Estimator*, M. Phil. Dissertation. Dept. of Economics, NEHU., Shillong (unpub).
Metropolis, N & S Ulam (1949). The Monte Carlo method, *Journal of the American Statistical Association*, Vol. 44 (247), pp. 335-341.
Meyer, J R & R R Glauber (1964). *Investment Decisions : Economic Forecasting and Public Policy*, Harvard Business School Press, Cambridge, Massachusetts.
Mishra. S K (1997). Generalisation of Indirect Least Squares to Estimation of Over-identified Equations of a Multi-Equation Linear Econometric Model, *Assam Statistical Review*, Vol. 9(1), pp. 57 – 66.
Nagar, A L (1960). A Monte Carlo Study of Alternative Simultaneous Equation Estimators, *Econometrica,* Vol. 28, pp. 573 – 590.

National Bureau of Standard (1951). *Monte Carlo Method*. Applied Math Serial 12, Washington.
Naylor, T H (Ed.) (1969). *The Design of Computer Simulation Experiments*. Duke Univ. Press, Durham, N. C.
Newey, W K (1985). Semi-Parametric Estimation of Limited Dependent Variables Models with Endogenous Explanatory Variables. *Annales de'l INSEE*, Vol. 59-60, pp. 219-235.
Nyquist, H & A Westlund (1977). L_1Versus L_2 Norm Estimation in Interdependent Systems when Residual Distributions are Stable, Mimeo, Dept. of Statistics, Univ. of Umea.
Oveson, R M (1969). Regression Parameter Estimation by Minimizing the Sum of Absolute Errors, Doctoral Dissertation, Harvard University, Cambridge, Massachusetts (unpub).
Pagan, A R (1986). Two-Stage and Related Estimators and their Applications. *Review of Economic Studies*, Vol. 57, pp. 517-538.
Phillips P C B (1991). A Shortcut to LAD Estimator Asymptotics. Econometric Theory, Vol. 7(4), pp. 450-463.
Pollard, D (1991). Asymptotics for Least Absolute Deviation Regression Estimators. *Econometric Theory*, Vol. 7(2), pp. 186-199.
Portnoy S & R Koenker (1997). The Gaussian Hare and the Laplacian Tortise: Computability of Squared-error versus Absolute-error Estimators. *Statistical Science*, Vol. 12, pp. 279-300.
Powell, J L (1984). Least Absolute Deviations Estimation for the Censored Regression Model. *J. of Econometrics*, Vol. 25, pp. 303-325.
Rhodes, E C (1930). Reducing Observations by the Method of Minimum Deviations, Phil. Magazine 7th Series. Pp. 974 – 992.
Rao, C R & S K Mitra (1971). *Generalised Inverse of Matrices and its Applications*. John Wiley, New York.
Rao, S S (1978). *Optimization: Theory and Applications*. Wiley Eastern Limited, New Delhi.
Rice, J R & J S White (1964). Norms for Smoothing and Estimation, *SIAM Review*, Vol. 6, pp. 243 – 256.
Robers, P D. & **A Ben-Israel** (1969). An Interval Programming Algorithm for Discrete Linear L_1 Approximation Problems, *Journal of Approximation Theory*, Vol. 2, pp. 323 – 336.
Schlossmacher, E J (1973). An Alternative Technique for Absolute Deviations Curve Fitting, *Journal of the American Statistical Association*, Vol. 68, pp. 857 – 859).
Simmonard, M (1966). *Linear Programming*, Prentice Hall, Englewood Cliffs, New Jersey.
Singh. L K (1993). *Performance of Generalised Indirect Least Squares in Over-Identified Equation Estimation: A Monte Carlo Experiment*, M Phil. Dissertation, Department of Economics, NEHU, Shillong (unpub).
Singleton, R R (1940). A Method for Minimizing the Sum of Absolute Errors, Am. Math. Statist. Vol. 11, pp. 301 – 310.
Smith, V K (1973). *Monte Carlo Methods*, Lexington Books, Lexington.
Smith, V K & T W Hall (1972). A Comparison of Maximum Likelihood versus BLUE Estimators, *The Review of Economics and Statistics*, Vol. 54, pp. 186 – 190.
Spyropoulous, K, E Kiountouzis & A Young (1973). Discrete Approximation in the L_1 Norm, *Computer Journal*, Vol. 16, pp. 180 – 186.
Texas Instruments Inc (1979). *TI Programmable 58C/59 Master Library*, Texas Instruments Inc. Texas.
Theil H (1953). *Repeated Least Squares Applied to Complete Equation Systems*, Central Planning Bureau (Mimeo), The Hague.

Theil, H (1957). Specification Errors and the Estimation of Economic Relationships, R*eview of International Statistical Institute*, Vol. 25, pp. 41- 51.
Theil, H (1961). *Economic Forecasts and Policy*, North Holland Publishing Co., Amsterdam.
Theil, H (1962). Three Stage Least Squares: Simultaneous Estimation of Simultaneous Equations, *Econometrica*, Vol. 30, pp. 54 – 78.
Theil, H (1971). *Principles of Econometrics*, John Wiley and Sons, Inc. New York.
Thomas, R L (1985). *Introductory Econometrics Theory and Application*, Longman Group Limited. New York.
Tylor, L D (1974). Estimation by Minimizing the Sum of Absolute Errors, in Zerembka, P (ed.), *Frontiers of Econometrics*, Academic Press, New York.
Usow, K H (1967). On L_1 Approximation for Discrete Functions and Discretization Effects, *SIAM Journal of Numerical Analysis*, Vol. 44, pp. 233 –244.
Wagner, H M (1958). A Monte Carlo Study of Estimates of Simultaneous Linear Structural Equations, *Econometrica*, Vol. 26, pp. 117 – 133.
Wagner, H M (1959). Linear Programming Techniques for Regression Analysis, *Journal of American Statistical Association*, Vol. 56, pp. 206 –212.
Weiss, A A (1991). Estimating Non-linear Dynamic Models using Least Absolute Error Estimation. *Econometric Theory*, Vol. 7, pp. 46-68.