

# Regression Model Estimation Using Least Absolute Deviations, Least Squares Deviations and Minimax Absolute Deviations Criteria

Pranesh Kumar<sup>1</sup> and Jai Narain Singh<sup>2</sup>

**Abstract**— Regression models and their statistical analyses is the most important tool used by scientists in data analyses especially for modeling the relationship among random variables and making predictions with higher accuracy. A fundamental problem in the theory of errors, which has drawn attention of leading mathematicians and scientists since past few centuries, was that of fitting functions. For the pioneering work to develop procedures for fitting functions, in the eighteenth century, we refer to the research by Mayer, Boscovich, Laplace, Legendre, Simpson, Gauss, and Euler. They worked on the methods of least absolute deviations, least squares deviations and minimax absolute deviations. Today's widely celebrated procedure of the method of least squares for function fitting is credited to the published works of Legendre and Gauss [1]-[2]. The least square estimates are best linear unbiased estimates and are optimal under assumptions that the errors follow normal distributions, are free of large size outliers and satisfy the Gauss-Markov assumptions. However, the least squares based models in practice may fail to provide optimal results in non-Gaussian situations especially when the errors follow distributions with the fat tails. In this paper, we will present an overview of some important works in fitting linear relationship. We will also statistical properties of the  $L_p$ -norm based model estimation, measures of the model adequacy and also future research questions of interest.

**Keywords**— Regression model, Least squares estimates, Least absolute deviations, Minimax absolute deviations,  $L_p$ -norm.

## I. FITTING FUNCTIONS: INTRODUCTION AND REVIEW

A standard linear regression model expressing the linear relationship between the study variable and explanatory variables is defined by  $n$  linear equations:

$$y_i = \beta_0 + x_{i1}\beta_1 + \dots + x_{ip}\beta_p + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (1.1)$$

where  $y$  and  $x$  respectively represent the values of the study variable and the  $p$ -explanatory variables and  $\varepsilon$  denotes the random error term. Problem of interest is to obtain the best estimates of the unknown model parameters  $\beta$ . To study this problem, methods considered were algebraic and geometric estimation methods. Attention was restricted to the special

cases when there was either the same number of observations as there were unknowns, or, there were  $n$  discordant observations on a single unknown quantity. In [3]-[4], Boscovich outlined an objective procedure for determining suitable values for the parameters of the fitted model from a greater number of observations. He formulated the principle that the values of the unknown model parameters be chosen such that the corrections ( $\varepsilon$ ) sum to zero and have minimum absolute  $\sum|\varepsilon_i|$ . An unconstrained variant of this fitting procedure is now known as the  $L_1$ -norm or least absolute deviations procedure and serves as a robust alternative to the  $L_2$ -norm or least squares. Laplace's first variant of the Boscovich's algorithm from its geometrical framework was published in [5]. He applied this procedure to the nine arc measurements and noted his results were unsatisfactory which led him to doubt the validity of the elliptical model. In [6], Laplace made a small but significant change that instead of choosing unknown model parameters to minimize the unweighted sum of the absolute errors  $\sum|\varepsilon_i|$  subject to the unweighted adding-up constraint  $\sum\varepsilon_i = 0$ , he chose parameters to minimize the weighted absolute sum  $\sum|l_i\varepsilon_i|$  subject to the weighted constraint  $\sum l_i\varepsilon_i = 0$ , where the  $l_i$  are given weights. However, applying this method, Laplace found error to be too large to sustain the hypothesis of an elliptical form for the figure of earth. Laplace [7]-[8] developed the minimax procedure by considering the problem of correcting the elements of a formula describing the celestial longitude of Saturn. He obtained results which were comparable with those later obtained from the method of least squares. Laplace's promulgation of Boscovich's method clearly represents an important contribution to its continued development. The algebraic reformulation of the problem and introduction of weighted observations lead to Laplace's development of an asymptotic distribution theory for this procedure. Legendre's method of least squares is discussed in the first four pages of a nine page appendix attached to his work on the determination of the orbits of comets. Legendre's procedure discussed by Laplace is restricted to two unknowns but can be applied to any number of equations in any lesser number of unknowns. Legendre chose values of unknown parameters to minimize the sum of squared errors. Legendre noted that these *équations du minimum* may also be obtained by multiplying all the terms in each of the equations by the coefficients of one of the unknowns and summing the result.

<sup>1</sup> Pranesh Kumar is Professor in the Department of Mathematics and Statistics, University of Northern British Columbia, Prince George, BC, V2N 4Z9, Canada (e-mail: [pranesh.kumar@unbc.ca](mailto:pranesh.kumar@unbc.ca)).

<sup>2</sup> Jai Narain Singh is Professor in the Department of Mathematics and Computer Science, Barry University, United States (email: [jsingh@barry.edu](mailto:jsingh@barry.edu)).

There is, therefore, the same number of equations as there are unknowns. These *équations du minimum* may be solved for the unknowns by the ordinary method (which he did not describe). He asserted that the solution found by this procedure corresponds to a minimum sum of squares function. However he did not prove this result. Similarly, to determine the position of a point  $(x, y, z)$  in three dimensional space from  $n$  observations on this point,  $(a_i, b_i, c_i)$ ,  $i = 1, 2, \dots, n$ , Legendre noted that the sum of squared Euclidean distances is minimized when  $x, y$  and  $z$  are set to their respective arithmetic means.

Gauss's least sum of even powers: Legendre's general object was to choose the values of the unknowns  $x, y, z$  etc. to make the errors  $e_0, e_1, etc$  as small as possible. Without any alternative, he asserted that the sum of the squares of the errors was the simplest criterion. Gauss in [9]-[10] offered that the sum of the fourth powers  $e_0^4 + e_1^4 + e_2^4 + etc$ , the sum of the sixth powers  $e_0^6 + e_1^6 + e_2^6 + etc$ , or, generally the sum of even powers as alternatives to the sum of squared errors. Of these criteria, simplest is the sum of squared errors which Gauss used in practical calculations. Gauss continued with the observation that in the limit as the even power selected becomes infinite, we have the criterion which chooses unknowns to minimize the largest absolute error. Laplace in [7]-[8] has considered the characterization of the minimax procedure and noted: I have found from numerous examples that when this power does not exceed the square, the results differ little from those which give the systems that minimize largest errors; this is the new advantage for the method of least squares of errors of observations.

Gauss [9]-[10] noted that besides the principle which minimizes the largest error, Laplace used another principle (previously proposed by Boscovich) for resolving a system of linear equations in which the number of equations is much larger than the number of unknowns; namely that the sum of the absolute values of the differences should be minimum. It is easy to show that the system of values for the unknowns found by this principle alone will necessarily exactly satisfy the same number of equations (selected from the given equations) as there are unknowns and that the other equations are not used except in making this selection. For example, if the equation  $a_i = V_i + e_i$  is amongst the number of those equations which are not satisfied exactly, then the system of values found by this principle will not be changed if instead of  $a_i$  one observed another value  $b_i$  such that the differences  $a_i - V_i$  and  $b_i - V_i$  have the same sign. Here  $V_i$  is the fitted value. Laplace moderates this principle in several ways by adding a new condition that the sum of the differences taken with their signs should be zero. As a result the number of equations which are satisfied exactly is one less than the number of unknowns; but this observation holds even when there are two unknowns. In [11]-[12], Bowditch offered an alternative generalization of the Boscovich's procedure. Instead of restricting himself to a single adding-up constraint as Gauss had done, he suggested that one should impose  $q - 1$  such constraints to eliminate all

but the  $q$  unknowns and to reduce the problem to one which could be solved by Laplace's algebraic.

It may thus be noted that the algebraic phase of the calculus of observations reached its climax in 1805 with the publication of Legendre's description of the method of least squares. Despite the significant contributions which Gauss made to the theory of the method of least absolute errors in 1809, he probably abandoned this method in favor of more accessible method of least squares. Laplace also adopted a Bayesian (inverse probabilistic) approach to the fitting problem. He suggested six optimality criteria which may be used to derive a mean from the posterior density function. Although Laplace mentioned the principle of maximum probability, he does not seem to have taken it seriously. Laplace missed the opportunity of applying this principle to the fitting problem when the observations follow the double exponential law. Instead in his worked examples, he addressed a more complicated problem deriving the value which divides the posterior distribution into two equal halves. Laplace's own derivation of the double exponential law was by assertion a statement without any reason or support. By contrast, Gauss's derivation of the normal law from the principle of arithmetic mean and the principle of maximum probability seems well developed. In his later writings Gauss discarded this derivation of normal law and normal law itself. This proof and another related to the optimality of the arithmetic mean attracted considerable interest in the following century. Normal law is central to the asymptotic arguments of Laplace.

In [35], it is postulated that there exist limit theorems which may be relevant when considering the sum of number of components in a regression disturbance that leads to non-normal stable distribution characterized by infinite variance. An infinite variance means fat tails which indicate presence of outliers in the disturbances. Huber [25] remarked: Just a single grossly outlying observation may spoil the least squares estimate and, moreover, outliers are much harder to spot in the regression than in the simple location case. The outliers occurring with extreme values of the regressor variables can be especially disruptive. In data based model development to establish relationships, it often involves determining functional forms in addition to screening regressors for inclusion [46]. In many applied settings one has little basis for judging, a priori, whether the usual least squares assumptions on the errors are satisfied. An important problem involves the identification and handling of observations that are outliers. It is usually desirable that outliers be identified and that they not have an unduly large influence on model parameter estimates. Least squares estimation falls short on these counts. The least squares principle of minimizing the sum of squared deviations dictates that very large deviations from the regression hyper plane will be avoided. That is, the fit of the model to other points will be sacrificed in order to accommodate outliers. Since the estimation procedure avoids large deviations, it is often difficult to spot true outliers. Further, the damage is compounded because the inclusion of the outliers can lead to serious errors in the parameter estimates and difficulty in recognizing the correct model. Andrews [20] noted that even

when the errors follow a normal distribution, alternatives to least squares may be required, especially, if the form of the model is not exactly known. Least squares estimation is not very satisfactory if the quadratic loss-function is not a satisfactory measure of the loss. In [26], Meyer and Glauber stated that for at least certain economic problems absolute error may be a more satisfactory measure of loss than the squared error.

Ashar and Wallace [21] studied the statistical properties of estimators of the regression parameters by minimization of  $L_1$ -norm. In [25], Huber explored the properties of  $L_1$ -norm based regression estimates in its robustness to wild fluctuations in the magnitude of residual elements. In [22], Barrodale and Roberts presented an algorithm for  $L_1$ -norm approximation by modifying the simplex method of linear programming which is computationally superior to the algorithm given in [33] and is an improved version of the primal algorithm. They were able to significantly reduce the total number of iterations required by discovering how to pass through several neighboring simplex vertices in a single iteration. Oveson [28] gave a new thrust to the investigation into the properties and applicability of the  $L_1$ -norm estimator. It was almost fully established that in the presence of errors generated by thick tailed distribution,  $L_1$ -norm regression performed better than least squares regression. Abdelmalek [19] described an algorithm, which determines the best  $L_1$ -approximations as the limit of  $L_p$ -approximations. His technique obtains a solution to a linear problem by solving a sequence of non-linear problems. An alternative approach for studying the asymptotic theory of  $L_1$ -norm estimator in a simple regression context was given by Pollard [30]. The approach was built on the convexity of the  $L_1$ -norm criterion to construct a quadratic approximation whose minimum is close enough to the  $L_1$ -norm estimator for the latter to share the same asymptotic normal distribution. Taylor [32] gave the condition under which the  $L_1$ -norm estimator is unbiased and consistent. He discussed some of the problems encountered when trying to establish a distribution theory, under the assumptions that errors are independent, identically distributed random variables with a continuous distribution function and median zero. Phillips in [29] presented the asymptotic theory for the  $L_1$ -norm estimator of a regression model by using generalized functions of random variables and generalized Taylor series expansions. The approach was justified by the smoothing that was delivered in the limit by the asymptotic, whereby the generalized functions were forced to appear as linear functions wherein they became real valued. He studied models with fixed random regressors and auto-regressions with infinite variance errors and a unit root. He showed that the  $L_1$ -norm estimator converges at a faster rate in the unit root model for  $0 < \alpha < 2$  than the least squares estimator.

Sakata [32] proposed a general estimation principle using assumption that instrumental variables (IV) do not explain the error term in a structural equation. He opined that unlike the IV estimators such as two-stage least squares estimator, the estimators based on the proposed principle are independent of

the normalization constraint. Based on this new principle, he proposed the  $L_1$ IV estimator, which is an IV estimation counterpart of the  $L_1$ -norm estimator. He investigated the asymptotic properties of the  $L_1$ IV estimator. A consistent estimator of its asymptotic covariance matrix and a consistent specification test based on the  $L_1$ IV estimator were proposed. He also discussed the problem of identification in L1IV estimation. In [23], Bassett and Koenker developed the asymptotic theory of  $L_1$ -norm regression. They resolved a long-standing open question concerning  $L_1$ -norm estimator by establishing its asymptotic normality under general conditions, thereby extending a result of Laplace to the general linear model. The result confirmed that for the general linear model the  $L_1$ -norm estimator is a natural analog of the sample median. They proved that in the general linear model with independent and identically distributed errors, the estimator which minimizes the sum of absolute residuals is consistent and asymptotically Gaussian. This indicated that for any error distribution for which median is more efficient than the mean as an estimator of location, the  $L_1$ -norm estimator has smaller asymptotic ellipsoids than the least squares estimator and therefore is more efficient than least squares estimator. It was proved that the  $L_1$ -norm estimator is affine equivariant, scale and shift equivariant and equivariant to reparameterization of design. Though least squares estimator shares the same properties, typically robust alternatives to least squares are not equivariant in one or more of the above senses. Powell [44] proposed an alternative to maximum likelihood estimation of the parameters of the censored regression model. He generalized the  $L_1$ -norm estimation for the standard linear regression model. He showed that the censored least absolute deviation estimator is robust to heteroscedasticity and is consistent and asymptotically normal for a wide class of error distribution. Consistency of the asymptotic covariance matrix was also proved. As a consequence, tests of hypothesis concerning the unknown regression coefficient can be constructed which are valid in large samples. He also opined that the censored least absolute deviation estimator can be computed using direct search methods of nonlinear programming. In [45], Weiss established that it was possible to use the  $L_1$ -norm estimator to estimate the parameters of a nonlinear dynamic model. The nonlinear least absolute deviations (LAD) estimator was defined and proved that this estimator was consistent and asymptotically normal under certain assumptions. He showed the sufficient condition for strong consistency of the least absolute deviations estimate given in [37] fails. Breidt, Davis and Trindade [36] studied the  $L_1$ -norm estimation for All-Pass time series models. The authors opined that an approximation to the likelihood of the model in the case of Laplace (two-sided exponential) noise yields a modified absolute deviation criterion, which can be used even if the underlying noise is not Laplacian. They established the asymptotic normality for  $L_1$ -norm estimators of the model parameters under general conditions. Behaviour of the estimators in finite samples was also studied via simulation. Kim and Muller [43] presented the asymptotic properties of two-stage quantile regression estimators. In their

paper, they derived the asymptotic representation of the estimators and proved the asymptotic normality with quantile regression predictions. The asymptotic variance matrix and asymptotic bias were discussed. They also analysed the asymptotic normality and the asymptotic covariance matrix with least squares predictions.

Furno [39] compared the performance of  $L_1$ -norm and least squares in the linear regression model with random coefficient autocorrelated (RCA) errors. The presence of fat tailed error distribution led to the estimation of the RCA model by  $L_1$ -norm estimator. He proved that the  $L_1$ -norm estimator for randomly autocorrelated errors is asymptotically normal. However, in the case of constant autocorrelation model, the results confirmed that  $L_1$ -norm is not advantageous, especially in small samples, since its sampling distribution differs from the asymptotic distribution. In [41], Hitomi and Kagihara proposed a nonlinear smoothed LAD estimator that is practically computable and has the same asymptotic properties as the nonlinear LAD estimator in Weiss's nonlinear dynamic model. Two types of error distributions were considered – standard normal distribution where the nonlinear smoothed LAD estimator becomes maximum likelihood estimator (MLE) and the Laplace distribution where the nonlinear LAD estimator is MLE. The results indicate that as the sample size increases the bias becomes negligible and the difference between nonlinear smoothed LAD and nonlinear smoothed estimators ceases. No difference was found in the performance of the two estimators with respect to median and quartiles. The  $L_1$ -norm estimation entered the domain of multi-equation model in the paper by Glahe and Hunt [40]. They compared the estimated parameters with those estimated through least squares and use Monte Carlo Methods for their performance appraisal. Amemiya [34] developed the two-stage least absolute deviation estimator, which is rather analogous to two-stage least squares. He extended the method to provide it a mathematical and statistical basis in the direction of consistency and related statistical properties. He defined a class of estimators called the two-stage least absolute deviation estimators (2SLAD) and derived their asymptotic properties. The problem of finding the optimal member of the class was also considered.

Khazzoom [42] generalized indirect least squares estimator for exactly- or over- identified equations. One may conjecture that if LAD performs better than least squares in estimating the matrix of reduced form coefficients, application of generalized inverse on such matrix of reduced form coefficients would be better than the GILS suggested by Khazzoom. A more generalized name, generalized indirect least norm (GILN), may be given to the family of such methods for the minimand norm (GILN2 as suggested by Khazzoom) or absolute calling it GILN1. Dasgupta and Mishra [24] compared GILN1 with GILN2, 2SLS, LS-LAD, LAD-LS and LAD-LAD estimates of structural coefficients. The results showed that LAD-LAD estimator performs better than 2SLS if errors are non-normal or outliers are present. Fair [38] estimated the US model by 2SLS, 2SLAD, 3SLS (Three

Stage Least Squares) and FIML (Full Information Max Likelihood) methods. Median unbiased estimates were obtained for eighteen lagged dependent variable coefficients. The 2SLS asymptotic distribution was compared to the exact distribution and was found to be close. A comparative study of four sets of estimates, that is, 2SLS, 2SLAD, 3SLS and FIML was made. The results showed that the estimates are fairly close to each other with the FIML being the farthest apart. The 3SLS estimator was found to be more efficient than the 2SLS estimator. The 2SLS standard errors were on an average 28 percent larger than the 3SLS standard errors. While the 3SLS standard errors were on average smaller (19 percent) than the FIML standard errors. To compare the different sets of coefficient estimates, the sensitivity of the predictive accuracy of the model to the different sets was also examined. The RMSEs were found to be very similar across all the five sets of estimates. The author also compared the US model to the VAR5/2, VAR4 and AC models. The US model was found to do well in the tests relative to the VAR and AC models. In view of the possibilities of replacing OLS with LAD estimator at either or both stages (parallel to 2SLS) of estimation of the structural equations of a multi-equation linear model, Dasgupta and Mishra [24] conducted Monte Carlo experiments to compare 2SLS (alias LS-LS) with LS-LAD, LAD-LS and LAD-LAD estimates of structural coefficients while the disturbances in the structural equations were normal, Beta1, Beta2, Gamma and Cauchy distributed with and without the presence of outliers.

## II. $L_p$ -NORM ESTIMATORS: LEAST ABSOLUTE, LEAST SQUARES AND MINIMAX ABSOLUTE DEVIATIONS

The  $L_p$ -norm of the residual vector  $\varepsilon$  is defined by

$$\|\varepsilon\|_p = \begin{cases} (\sum |\varepsilon_i|^p)^{1/p}, & \text{for } p \in [1; \infty), \\ \max |\varepsilon|, & \text{for } p \rightarrow 1. \end{cases} \quad (2.1)$$

An estimator minimizing a  $L_p$  norm of the residual vector  $\varepsilon$  is called an  $L_p$ -norm estimator. Measuring the size of  $\varepsilon$  using the  $L_p$ -norm, we arrive at the  $L_p$ -regression problem (Kumar and Kashanchi). In regression analysis, goal is to determine  $\beta$  that attains the minimum  $L_p$ -norm for the difference between  $y$  and  $X\beta$ . Thus, the  $L_p$ -regression problem is to determine  $\beta$  such that  $\min \|X\beta - y\|_p$ .

On  $L_p$ -norm estimators in [13], Nyquist has prepared a monograph that provides relations between  $L_p$ -norm estimators and the class of maximum likelihood type estimators (M-estimators). The class of estimators that are linear combinations of order statistics (L-estimators) are given. The concept of the optimal  $L_p$ -norm estimator that possesses the smallest asymptotic variance is discussed and also, the  $L_p$ -norm based methods for estimating multicollinear regression models, with serially dependent residuals, interdependent systems. The interest in the  $L_1$ -norm estimator has its origin in the works of Laplace and Edgeworth [14]. Nyquist presents the investigations of  $L_p$ -norm estimators of linear regression models. He discussed geometrical interpretations of  $L_p$ -norm estimators and theorems on

existence, uniqueness and asymptotic distributions of these estimators.

#### A. Least Absolute Deviations Regression

The  $L_1$ -norm regression problem, by letting  $p = 1$  in (2.1) becomes  $\min \|X\beta - y\|_1$ , which is written as the linear programming (LP) problem

$$\min \sum t_i: -t_i \leq x_i^T \beta - y_i \leq t_i, \quad i = 1, \dots, n. \quad (2.2)$$

Methodology to estimate unknown parameters using least absolute deviations was first introduced by Boscovich (1757).

#### B. Least Squares Deviations Regression

The  $L_2$ -norm regression problem by setting  $p = 2$  in (2.1) becomes  $\min \|X\beta - y\|_2$ . This is equivalent to minimizing  $\min \sum (y_i - \sum x_{ij}\beta_j)^2$  with respect to  $\beta$ .

Solution of  $\beta$  for  $L_2$ -norm regression problem is commonly known as the least square estimators (LSE). It may be noted from the works of Legendre (1805) and Gauss (1809) that they proposed to minimize the sum of the squares of the measurement errors and, thereafter, the method of least square became the most popular estimating technique. The main reason for least squares popularity is presumably easy computations and due to the fact that when the residuals are independent and identically normally distributed, the least squares estimators regression model are also the best linear unbiased estimator as well as equivalent to the maximum likelihood estimator, implying the inference to be easily performed [14]. However, it has been noted that the least squares estimates are sensitive to departures from the assumptions, for example, normally distributed errors.

#### C. Least Squares Deviations Regression

Letting  $p \rightarrow \infty$  in (2.1), the  $L_\infty$ -norm regression problem becomes  $\min \|X\beta - y\|_\infty$  which can be written as the linear programming (LP) problem

$$\min t: -t \leq x_i^T \beta - y_i \leq t, \quad i = 1, \dots, n. \quad (2.3)$$

This minimization problem is often referred to as the Chebyshev approximation. Laplace (1818) and Edgeworth (1887) have shown that the  $L_1$ -norm estimator is preferable to the least squares, when estimating a simple linear regression model with fat-tailed distributed residuals.

### III. MODEL ADEQUACY: METRICS

Principle of analysis of variance partitions the total response variance into two components: variance explained by the model and variance that remained unexplained. For assessing model adequacy, one commonly used measure based on estimated residuals is the coefficient of determination  $R^2$  which is defined as the proportion of the total response variance that is explained by the model

$$R^2 = 100 \left[ 1 - \frac{\sum \varepsilon^2}{\sum (y - \bar{y})^2} \right]. \quad (3.1)$$

Another measure by Kumar and Kashanchi in [15] denoted by  $\|R^2\|_1$  based on the estimated residuals for checking model accuracy is

$$\|R^2\|_1 = 100 \left[ 1 - \frac{\sum |\varepsilon|}{\sum |y - \bar{y}|} \right]. \quad (3.2)$$

Note that the numerator  $\varepsilon$  and denominator  $\sum |y - \bar{y}|$  terms of  $\|R^2\|_1$  are  $L_1$ -norm while in case of  $R^2$ , these are  $L_2$ -norm. Either measure compares how well model fits. A higher value of  $\|R^2\|_1$  or  $R^2$  indicates a better fit. Further, sampling distribution and statistical properties of the measure  $\|R^2\|_1$  are being investigated. These results will be communicated elsewhere.

### IV. APPLICATION TO KINESIOLOGY: A SCIENTIFIC STUDY OF HUMAN MOVEMENT

#### A. Kinesiology Experiment

To study the interrelationship of the physiological processes and anatomy of the human body with respect to movement, an experimenter admitted 53 subjects in the study. Each subject performed a standard exercise at gradually increasing levels and oxygen uptake and expired ventilation (which is related to the rate of exchange of gases in the lungs) were recorded. The main objective of the study was to develop a model that relates the expired ventilation to the oxygen intake [16].

#### B. Results and Discussions

Summary statistics of expired ventilation and oxygen uptake for data ( $n = 53$ ): Minimum (16.9, 574), Maximum (144.8, 4393), Mean (60.7, 2542), Median (46.5, 2490), Standard deviation (39.9, 1224), Skewness (0.7, 0), Kurtosis (-0.7, -1), Correlation coefficient (0.9550). These statistics indicate that the distributions of both highly linearly correlated variables can be assumed to be approximately normal distributions however both distributions have large variability. The fitted models using different criteria:

$L_1$ -norm:

$$\text{Expired ventilation} = -19.756 + 0.0300 \text{ Oxygen uptake}$$

$L_2$ -norm:

$$\text{Expired ventilation} = -18.449 + 0.0311 \text{ Oxygen uptake}$$

$L_\infty$ -norm:

$$\text{Expired ventilation} = -19.648 + 0.0322 \text{ Oxygen uptake}$$

Model accuracy measure values for these fitted  $L_1$ -,  $L_2$ - and  $L_\infty$ -norm models:

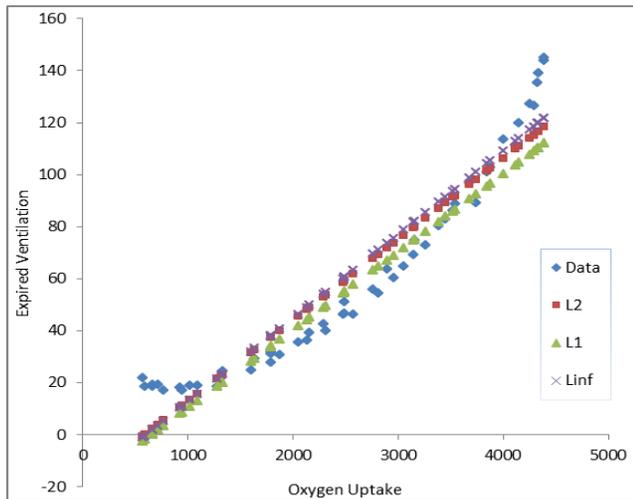


Fig.1 Data and Fitted Models.

Measure  $R^2(\%)$  are (89.94, 91.20, 90.95) and measure  $\|R^2\|_1(\%)$  are (84.22, 83.65, 82.84). Adequacy measure values, as expected, are high for all models because of high linear correlation coefficient between study variables, expired ventilator and oxygen uptake. Fig. 1 shows the graphs of the actual and predicted expired ventilation against oxygen uptake. It may be noted that the predicted values except the tails are close enough to the actual values of the expired ventilation.

Results of the residual analysis:

$L_1$ -model (min= -32.77, max= 11.25; range = 44.02)

$L_2$ -model (min= -26.45, max= 15.50; range = 41.95)

$L_\infty$ -model (min= -23.06, max= 17.03; range = 40.10).

Summary statistics indicate that the normality assumptions are approximately satisfied more closely in case of  $L_2$ -residuals followed by  $L_\infty$ - and  $L_1$ - residuals.

Absolute relative error analysis of the lower tail (10%) of the data ( $n = 5$ ) in Table I and Fig. 2 indicate the model performance (in decreasing order):  $L_2 \cong L_\infty \cong L_1$  and from Table II and Fig. 3 in the upper tail (10%) data ( $n = 5$ ):  $L_2 \cong L_\infty \cong L_2, L_1$ .

TABLE I  
ABSOLUTE RELATIVE ERROR: LOWER TAIL (10%)

$L_1$	$L_2$	$L_\infty$
1.1158	1.0262	1.0532
1.1073	1.0007	1.0315
0.9912	0.8802	0.9068
0.9867	0.8784	0.9042
0.9071	0.7963	0.8192

Absolute relative error analysis of the middle (80%) of the data ( $n = 43$ ) in Table III:

$L_1$ -model (min= 0.0015, max= 0.1956; range = 0.1941)

$L_2$ -model (min=-0.2719, max= 0.0586; range = 0.2133)

$L_\infty$ -model (min= 0.0336, max= 0.3421; range = 0.3085)

Model performance (in decreasing order):  $L_1 \cong L_\infty, L_2$ .

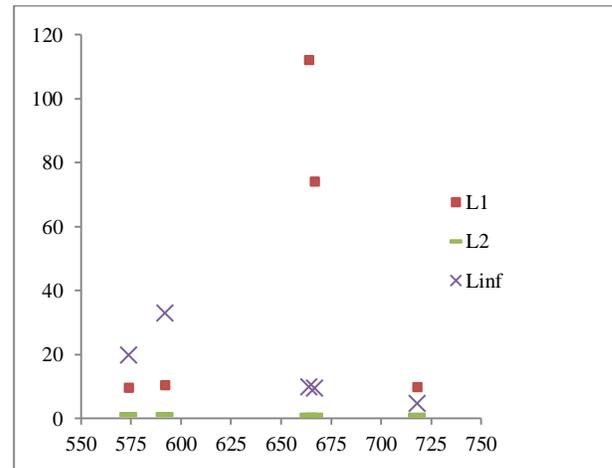


Fig. 2 Absolute Relative Error- Lower tail (10%).

TABLE II  
ABSOLUTE RELATIVE ERROR: UPPER TAIL (10%)

$L_1$	$L_2$	$L_\infty$
0.1381	0.0890	0.0626
0.1869	0.1408	0.1158
0.2066	0.1616	0.1372
0.2210	0.1770	0.1530
0.2263	0.1826	0.1588

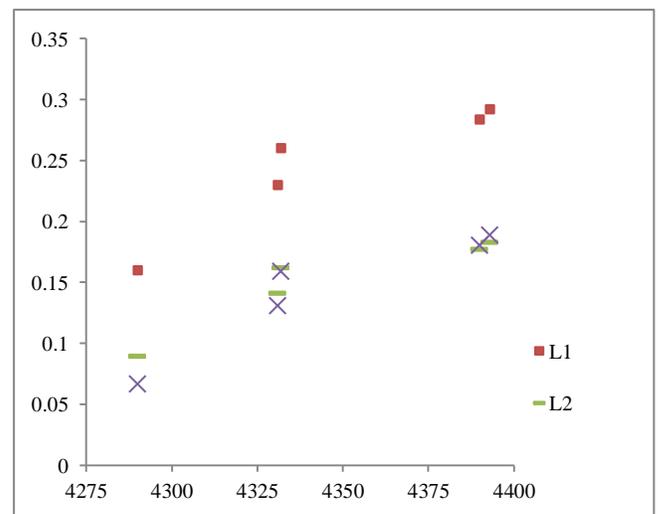


Fig. 3 Absolute Relative Error- Upper tail (10%).

### C. Remarks and Further Questions

The least squares estimation (LSE) although is simple and algebraically highly developed, studies have shown that LSE based linear regression may not be the optimal model when one or more of its assumptions fail. In our present analysis of the experimental data, we have fitted the  $L_1$ -,  $L_2$ - and  $L_\infty$ -norm models and noted that our results, like those in other studies, are in agreement and that:

- Model choice based only on  $R^2$  is not always appropriate and
- Strong indications that for making predictions especially in lower and upper tails,  $L_1$ -and  $L_\infty$ - norm models may over perform  $L_2$ -norm (least squares) regression models.

Our study raises questions to investigate:

- Sampling distributions and statistical properties of the  $L_p$ -norm regression models,
- Robustness and consistency of the  $L_p$ -norm estimates,
- Inference like interval estimation, hypothesis testing and prediction bands and
- Optimal choice of  $p$  for the  $L_p$ -norm regressions.

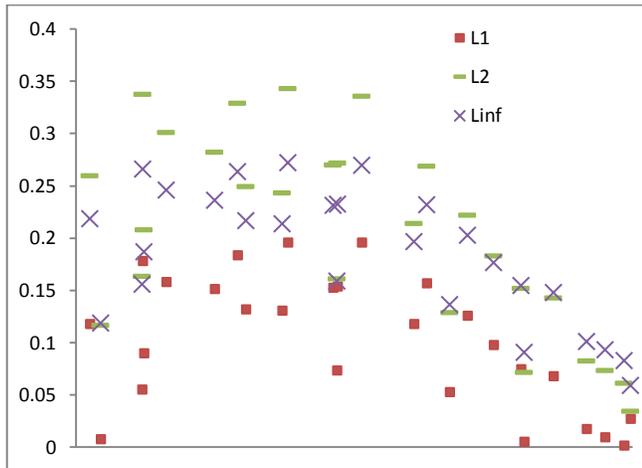


Fig. 4 Absolute Relative Error for Middle Range (80%).

#### ACKNOWLEDGMENT

This research was supported by Dr. Pranesh Kumar's Natural Sciences and Engineering Research Council Discovery Grant.

#### REFERENCES

- [1] A. M. Legendre, "Nouvelles methodes pour la determination des orbites des cometes," Mme. Courcier, Paris, 1805.
- [2] C. F. Gauss, "Theoria motus corporum coelestium in sectionibus conicis solem ambientium," Hamburg, 1809.
- [3] R. J. Boscovich, "De literaria expeditione per ponti ciam ditionem et synopsis amplioris operis," Bononiensi Scientiarum et Artium Instituto atque Academia Commentarii 4, pp. 353-396, 1757.
- [4] R. J. Boscovich, "De Solis ac Lunae Defectibus," Paris, 1779.
- [5] P. S. Laplace, "Sur quelques points du système du monde," Mme. De l'Academie Royale des Sciences de Paris, 1793.
- [6] P. S. Laplace, "Traité de Mécanique Céleste," Tome II, J. B. M. Dubrat, Paris, 1799.
- [7] P. S. Laplace, "Théorie de Jupiter et de Saturne," Mme. De l'Academie Royale des Sciences de Paris, 1788.
- [8] P. S. Laplace, "Théorie de Analytique des Probabilités," Courcier, Paris, 1812.
- [9] P. S. Laplace, "Mémoire sur la probabilités des causes par les événements," Mme. De l'Academie Royale des Sciences de Paris 6, 1788.
- [10] P. S. Laplace, "Deuxieme Supplement to Laplace (1812)," 1818.
- [11] N. Bowditch, "Observations of the comet of 1807," Mme. Amer. Acad. Arts Sci. 3, part 1, pp. 1-17, 1809.
- [12] N. Bowditch, "Elements of the comet of 18011," Mme. Amer. Acad. Arts Sci. 3, pp. 313-325, 1815.
- [13] H. Nyquist, "Recent Studies on  $L_p$ -norm Estimation," Doctoral thesis, University of Umea, 1980.
- [14] F. Y. Edgeworth, "On observations relating to several quantities," Phil. Mag., 5th Series, vol. 24, pp. 222-223, 1887.
- [15] P. Kumar and F. Kashanchi, "On the  $L_p$ -norm regression models for estimating value-at-risk," Serdica J. Computing, vol. 8, no. 3, pp. 101–114, 2014.

- [16] B. Abraham and J. Ledolter, "Introduction To Regeression Modeling," Brooks/Cole, pp. 213-214, 2006.
- [17] G. O. Young, "Synthetic structure of industrial plastics (Book style with paper title and editor)," in Plastics, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, pp. 15–64, 1964.
- [18] N. N. Abdelmalek, "Linear L1 approximation for a discrete point set and L1 solutions of overdetermined linear equations," J. Assoc. Compt. Mach., 18, pp.41-47, 1971.
- [19] D. F. Andrews, "A robust method for multiple linear regression," Technometrics, 16, pp.523-531, 1974.
- [20] V. Ashar and T. D. Wallace, "A sampling study of minimum absolute deviations estimators," Operations Research, 11, pp.747-758, 1963.
- [21] I. Barrodale and F. D. K. Roberts, "An improved algorithm for discrete L1- approximation," SIAM J. Num. Anal., 10, pp.839-848, 1973.
- [22] G. Basset ( Jr.) and R. Koenkar, "Asymptotic theory of least absolute error regression," J. Amer. Statist. Assoc., 73, pp.618-622, 1978.
- [23] M. Dasgupta and S. K. Mishra, "Least absolute deviation estimation of linear econometric models: A literature review," <http://mpr.aub.uni-muenchen.de/1781/>, MPRA Paper No. 1781, 2007.
- [24] P. J. Huber, "Robust estimation of a location parameter," Ann. Math. Statist., 35, pp.73-101, 1964.
- [25] J. R. Meyer and R. R. Glauber, "Investment Decisions: Economic Forecasting and Public Policy," Harvard Business School Press, Cambridge, Massachusetts, 1964.
- [26] S. C. Narula and J. F. Wellington, "The minimum sum of absolute errors regression: Astate of the art survey," International Statistical Review / Revue Internationale de Statistique, 50, pp.317-326, 1982.
- [27] R. M. Oveson, "Regression Parameter Estimation by Minimizing the Sum of Absolute Errors," Doctoral Dissertation, Harvard University, Cambridge, Massachusetts, 1969.
- [28] P. C. B. Phillips, "A shortcut to LAD estimator asymptotics," Econometric Theory, 7, pp.450-463, 1991.
- [29] D. Pollard, "Asymptotics for least absolute deviation regression estimators," Econometric Theory, 7, pp.186-199, 1991.
- [30] S. Sakata, "Instrumental Variable Estimation Based on Mean Absolute Deviation," Department of Economics, University of Michigan, USA, 2001 [www.econ.lsa.umich.edu/wpweb/11iv.pdf](http://www.econ.lsa.umich.edu/wpweb/11iv.pdf)
- [31] L. D. Taylor, "Estimation by minimizing the sum of absolute errors," Frontiers in Econometrics, Zarembka (Ed.), Academic Press, 1974.
- [32] K. H. Usow, "On L1 approximation for discrete functions and discretization effects," SIAM J. Num. Analysis, 44, pp.233-244, 1967.
- [33] T. Amemiya, "The Two Stage Least Absolute Deviations Estimators," Technical Report No. 297, Institute for Mathematical Studies in the Social Sciences, Stanford, 1980.
- [34] R. Bartels, "On the Use of Limit Theorem Arguments in Economic Statistics," American Statistician, 31, pp. 85-87, 1977.
- [35] F. J. Breidt, R. A. Davis and A. Trindade, "Least Absolute Deviation Estimation for All-Pass Time Series Models," Annals of Statistics, 2000. [www.stat.colostate.edu/~jbreidt/technical\\_reports.htm](http://www.stat.colostate.edu/~jbreidt/technical_reports.htm)
- [36] X. R. Chen, Z. D. Bai, L. C. Zhao and Y. H. Wu (1992). "Consistency of Minimum L1- norm Estimates in Linear Models," Development of Statistics: Recent Contributions from China, pp. 249-260, Pitman Research Notes in Mathematical Series 258, Longman Scientific and Technical, London 1992.
- [37] R. C. Fair, "Estimating and Testing the US Model," Yale Univ. USA, 1994.
- [38] M. Furno, "LAD Estimation with Random Coefficient Autocorrelated Errors," 2000. [www.eco.unicas.it/docente/furno/Text-rca.pdf](http://www.eco.unicas.it/docente/furno/Text-rca.pdf)
- [39] F. R. Glahe and J. G. Hunt. "The Small Sample Properties of Simultaneous Equation Least Absolute Estimators vis-a-vis Least Squares Estimators," Econometrica, 5: 38, pp.742-753, 1970.
- [40] K. Hitomi and M. Kagihara, "Calculation Method for Non-Linear Dynamic Least Absolute Deviation Estimator," J. of Japan Statist. Society, 31, 1, pp. 39-51, 2001.
- [41] J. D. Khazzoom, "An Indirect Least Squares Estimator for Over-identified Equations," Econometrica, 44, 4, pp.741-750, 1976.
- [42] T. H. Kim and C. Muller, "Two-Stage Quantile Regression," Discussion Paper in Economics #00/1, Nottingham University, 2000.
- [43] J. L. Powell, "Least Absolute Deviations Estimation for the Censored Regression Model," J. of Econometrics, 25, pp.303-325, 1984.
- [44] A. A. Weiss, "Estimating Non-linear Dynamic Models using Least Absolute Error Estimation," Econometric Theory, 7, pp.46-68, 1991.
- [45] H. G. Wilson, "Least squares versus minimum absolute deviations estimation in linear models," Decision Sciences 9, pp.322-335, 1978.

TABLE III  
ABSOLUTE RELATIVE ERROR: MIDDLE RANGE (80%)

$L_1$	$L_2$	$L_\infty$
0.8021	0.6728	0.6955
0.5599	0.4307	0.4425
0.4968	0.3580	0.3694
0.4292	0.2992	0.3054
0.3092	0.1746	0.1766
0.0024	0.1461	0.1544
0.1257	0.0022	0.0067
0.1811	0.0663	0.0578
0.1331	0.2589	0.2787
0.0073	0.1161	0.1345
0.0580	0.1625	0.1842
0.2166	0.3367	0.3617
0.0989	0.2070	0.2297
0.1878	0.3000	0.3256
0.1784	0.2813	0.3088
0.2245	0.3281	0.3574
0.1520	0.2485	0.2763
0.1503	0.2424	0.2712
0.2432	0.3421	0.3734
0.1795	0.2689	0.2997
0.0789	0.1603	0.1886
0.1816	0.2708	0.3017
0.2431	0.3348	0.3679
0.1331	0.2130	0.2440
0.1854	0.2683	0.3009
0.0557	0.1282	0.1576
0.1435	0.2211	0.2532
0.1081	0.1820	0.2134
0.0806	0.1514	0.1823
0.0050	0.0708	0.0996
0.0730	0.1421	0.1731
0.0177	0.0820	0.1117
0.0097	0.0729	0.1025
0.0015	0.0604	0.0899
0.0266	0.0336	0.0623
0.0648	0.0080	0.0198
0.0379	0.1004	0.1315
0.0529	0.0035	0.0320
0.0623	0.0066	0.0216
0.1155	0.0637	0.0369
0.0694	0.0156	0.0128
0.1259	0.0755	0.0488
0.1525	0.1041	0.0781