CHAPTER 1

INTRODUCTION

I. The Background: Every real empirical or experiential science (against the ideal science such as mathematics or logic which derives implications from a closed system of axioms) is an organized description of empirical inter-relations among the variables of the system of its concern. For example, physics investigates into the relationship between electrical conductivity of a metallic wire and its length and thickness, and other such relations; chemistry purports to establish the relationship between the rate of various types of chemical reactions and the temperature and pressure that the reactants are subject to, and so on; botany studies the relationship between the growth of the biomass of a species of plants and the nutritional contents of the soil and the physico-chemical properties of the environment that support plant life, and such issues. Similarly, economics is concerned with the relationship between consumption expenditure of households on a specific commodity, income of the household and its size and the price of the commodity, and the likes. On constructive thinking and empirical evidence each branch of real science describes such relationships of its concern and puts them in an organic order to make an ever-growing edifice of its literature.

Real sciences originate from and feed on the empirical experiences accumulated either by observation or by experimentation. Growth of a real science is squarely based on constructive thinking that suggests the scientists to gather information on relevant variables and the suitable methods to process the information so accumulated to establish the relationship among them. Therefore, a real science progresses with a hypothesis in
which the values of a dependent variable is associated with the values of a single or several explanatory variables and it is held that variations in (the value of) the dependent variable is associated with the variations in (the values of) the explanatory variables in a particular manner. Here ‘variation’ is very crucial. It is the cornerstone of all empirical investigations. All empirical knowledge emanate from the analysis of variation. Therefore, the scientist collects information on the dependent variable (say, y) and the explanatory variables (say, x₁, x₂, x₃) taking on several values. Our set of observations may thus be described as:

\[
\begin{bmatrix}
y_1, x_{11}, x_{12}, x_{13} \\
y_2, x_{21}, x_{22}, x_{23} \\
y_3, x_{31}, x_{32}, x_{33} \\
\vdots, \vdots, \ldots, \ldots \\
y_n, x_{n1}, x_{n2}, x_{n3}
\end{bmatrix}
\] ... (1.1).

The scientist holds or conjectures that \( y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 \) (where \( a_j \) for \( j = 0, 1, 2 \) and 3 are non-zero numerical constants) is the most possible relationship between y and x’s. However, the correct relationship could be \( y = f(\chi_1, \chi_2, \chi_4, \chi_5, \omega_1, \omega_2, \zeta_1, \zeta_2) \), where \( f(.) \) is a nonlinear function. In the scientist’s scheme, y, x₁ and x₂ are the most appropriate representatives that measure \( \upsilon, \chi_1 \) and \( \chi_2 \) respectively. One must note that all hypotheses in a real science relate concepts and only few concepts can be perfectly measured by a number (denial of strict arithmomorphism as argued by N. Georgescu-Roegen, 1971). Therefore, empirically obtained numerical measurements (data) on any conceptual entity is only a representative of that entity. Now, the scientist has mis-specified the correct relationship by deviating from the fact in many possible ways; first, he has not been able to measure the concepts perfectly accurately – he is working only
with the best possible numerical measures of the concepts to be correlated; secondly he has not been able to incorporate all relevant explanatory variables that explain \( y \) — he has failed to include representative variables that measure \( \chi_4, \chi_5, \chi_6, \omega_1, \omega_2, \zeta_1 \) and \( \zeta_2 \). Thirdly, he has included \( x_3 \) among the explanatory variables while it is none of them in fact; and fourthly (may not be lastly), he has thought of \( y = f(.) \) a linear function of \( x \) while in fact it is not. It is also not unlikely that he has exchanged a dependent variable for an explanatory variable or that his dependent variable and one of the explanatory variables are co-effects of the rest of the explanatory variables. It is also plausible that explaining \( y = f(x) \) is incomplete: \( y = f(x) \) and \( z = \psi(x,q) \) jointly would be more appropriate. And all these do not exhaust the list of mis-specifications.

The scientist should therefore compensate for all such possible errors by adding an error (disturbance or residual) term to his hypothesis and express his model as:

\[
y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + \varepsilon \quad \ldots \quad (1.2)
\]

where the last term captures the effects of all mis-specifications committed by him.

Since the error is due to the deviation of the explicitly expressed model and the true relationship and there is no one-to-one correspondence between the scientist's committed specification and that omitted by him, error vector varies from sample to sample even if the values of the variables specified in the model are fixed. This also amounts to say that \( y \) varies from sample to sample for fixed values of \( x \)'s. More explicitly, in the subcribed matrices \([ ]_1, [ ]_2, [ ]_3\) and \([ ]_4\) detailed out below,
x's are fixed and yet the y's are different across the matrices. That is to say that the matrices \([ \mathbf{Y}_1 \), \([ \mathbf{Y}_2 \), \([ \mathbf{Y}_3 \) and \([ \mathbf{Y}_4 \) differ from one another in that their y columns are not identical (though their x columns are identical). This is so because for the four samples even though x's are identical, hidden \((X_4, X_5, X_6, \omega_1, \omega_2, \zeta_1, \zeta_2)\) are different giving rise to different error vectors (e's). This gives rise to stochasticity in the error, e. The error is a random variable following a particular probability distribution. The scientist may have some speculation, conjecture or hypothesis regarding the probability distribution of the error term, e.

The method of Least Squares: The scientist aims at obtaining the best values of \(a = [a_0, a_1, a_2, a_3]^T\). Since no value of a can satisfy all equations, he has to define the meaning of the qualification 'best'. It is natural to define 'best' \(\hat{a} \) - call it \(\hat{a} \) - such that it yields an error vector, e with the minimum norm. If the scientist’s choice is to minimize the Euclidean norm of the error vector, it is tantamount to minimize the squared Euclidean norm of the error vector. Therefore, he chooses to minimize

\[
S = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - (a_0 + a_1 x_{i1} + a_2 x_{i2} + a_3 x_{i3}))^2.
\]

To minimize S he differentiates S partially with respect to \(a_i\) (i=0, 1, 2, 3) and sets those partial derivatives equal to zero. This gives the so-called 'normal equations'.
This system of normal equations in this example will have four unknowns in four equations. Assuming linear independence of these equations, the system may be solved for \( a \). This method of obtaining 'best' \( a \) \( (= \hat{a}) \) is called the Principle of Least Squares.

Under the Gauss-Markov assumptions \( \hat{a} = (X'X)^{-1} X'y \) is the best linear unbiased estimator of \( a \) in \( y = Xa + e \). These assumptions are: (i) \( E(e_i) = 0 \Rightarrow E(y_i) = \sum_{j=0}^{n} a_j x_{ij} \), \( \forall i, i = 1, 2, ..., n \). In general, \( E(e) = 0 \Rightarrow E(y_i) = Xa \). Here \( E(.) \) means the expectation of the random variable (.). (ii) \( E(ee') = \sigma^2 I \) or geometrically, the spherical disturbances. (iii) The matrix of sample values of the explanatory variables, \( X \), is a fixed non-stochastic matrix of full rank.

However, reality often breaks away with the Gauss-Markov assumptions. Very often disturbances are non-spherical. Occasionally, in single equation models, the assumption regarding the non-stochasticity of the sample values of the explanatory variables may not be satisfied. That makes the (ordinary) Least Squares estimator, \( \hat{a} \), an inconsistent estimator. The violation of the first Gauss-Markov assumption introduces bias into the Least Squares estimator. The estimator continues to be biased even if the sample size is increased indefinitely.

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 income distribution bears testimony to distribution of the error term with infinite variance. Works of Meyer & Glauber (1964), Fama (1965) and Mandelbroth (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 tail" which implies that large values or 'outliers' are present. Sporadic errors in X also are frequently present.

We recall that in defining the 'best' solution the scientist's choice was to minimize the Euclidean norm of the error vector, tantamount to minimization of the squared Euclidean norm of the error vector. That choice yielded the Least Squares estimator. However, this method places a relatively heavy weight on outliers and, therefore, in the presence of outliers, the method can lead to extremely sensitive estimates. This implies that in repeated sampling from a distribution where outliers are prevalent, LS estimates will vary considerably and will fail to establish reliability, becoming extremely sample dependent. Moreover, it is also not possible to obtain a meaningful variance estimate in the cases where the variance does not exist and the LS method in such cases will cease to possess its minimum variance properties. Thus, it becomes necessary to look for an alternative, a more 'robust' estimator, which gives relatively less weight to outliers and hence is not affected by their presence or the distribution that the error term characterizes.

Therefore, one parts with the convention of minimizing the Euclidean norm of the disturbance vector and, instead, defines the 'best' solution, \( \hat{a} \), of \( \mathbf{a} \) in \( \mathbf{y} = \mathbf{Xa} + \mathbf{e} \) that minimizes the absolute norm of \( \mathbf{e} \) (or yields minimal \( \sum_{i=1}^{n} |e_i| = \sum_{i=1}^{n} |y_i - \sum_{j=0}^{n} \hat{a}_j x_{ij}| \)). It may be noted that the objective of the scientist is to obtain the best solutions - the solution that best represents the factual relationship among the variables - the solution that is least affected by the sample aberrations - and not to stick to any convention or prefer one norm to the others. Evidently, since the 'least absolute norm' estimator minimizes the sum of absolute residuals, it is less influenced by outliers and, in the presence of outliers, its sampling variability is less than that of the LS estimator. This estimator is called by
various names such as LAD (Least Absolute Deviation), LAR (Least Absolute Residual) and MAD (minimum Absolute Deviation) estimator.

The idea is not new. Among the multitude of measures of central tendency, the arithmetic mean is obtained by minimizing the Euclidean norm of errors and therefore, it is an estimator of the population mean that minimizes the error variance (or the squared expected Euclidean norm of deviations about the sample mean). However, the median is obtained by minimizing the absolute norm of deviations from the measure of the central tendency. Thus we define arithmetic mean, $\hat{a}$, and median, $\tilde{a}$ in

$$y_i = x_i a$$

... (1.4)

(where $x$ is a unitary vector or $x_i = 1$ for $i = 1, 2, ..., n$) such that $\hat{a}$ minimizes $S_2 = \sum_{i=1}^{n} |y_i - x_i \hat{a}|^2 = \sum_{i=1}^{n} (y_i - x_i \hat{a})^2$ for the choice of $\hat{a}$ as an estimator of $a$ and $\tilde{a}$ minimizes $S_1 = \sum_{i=1}^{n} |y_i - x_i \tilde{a}|$ as an estimator of $a$ in (1.4) above. It is well known that $\tilde{a}$ (the median of $y$) is unaffected by extremal values of the variate $y$ and therefore, the outliers in $y$, while $\hat{a}$ (the arithmetic mean of $y$) is oversensitive to the extremal values and outliers. Considering in a more general framework, $\tilde{a}$ and $\hat{a}$ are obtained by minimization of the Minkowski norm $L_p$ for $p=1$ and $p=2$ respectively.

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

Investigations into the performance of L₁ estimator of single equation models clearly establish its superiority to the conventional L₂ (that is, Least Squares) estimator when errors contain outliers and hence are thick-tailed. In this regard, Bassett and Koenker (1978) deserve a special mention. They analytically established that L₁ estimator has strictly smaller asymptotic ellipsoids than L₂ estimator for linear models from any F for which the sample median is a more efficient estimator of location than the sample mean. Side by side Monte Carlo experiments also were conducted to compare L₁ and L₂ estimators. These studies also strongly indicated the superiority of L₁ to L₂ estimator for single equation linear models with disturbances infested with outliers. Pollard (1991), Phillips (1991), Chen (1996), Hitomi & Kagihara (2001), etc. are some recent works on LAD estimation and its extension.
Multi-equation linear econometric models, described as \( YA + XB + \varepsilon = 0 \), were first estimated by minimization of \( L_2 \) norm. These estimators are collectively called the k-class estimators (Indirect Least Squares – ILS, Two-Stage Least Squares - 2-SLS and Limited Information Max Likelihood - LIML). The ILS has very limited application due to its applicability in case of exactly identified equations only. In 2-SLS (suitable to estimating exactly as well as over-identified equations), 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.


Amemiya (1982) is certainly a landmark in the history of investigation into the possibility of applying \( L_1 \) estimation method to multi-equation linear models. He conjectured that if the disturbances in the structural as well as reduced form equations of a multi-equation model follow a non-normal distribution, it would be better to apply LAD to
the reduced form equation as well as to the structural equation to be estimated. Based on this conjecture he developed D2SLAD (Double Two-Stage Least Absolute) estimator. Thus 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 suggested estimation by D2SLAD (Double 2-Stage LAD) in case of full non-normal and outlier infested errors. Amemiya suggested Monte Carlo experiments to be carried out in order to study the properties of 2SLAD estimator vis-à-vis those of D2SLAD and 2-SLS.

What has been described above is in essence an investigation into an estimator parallel to the Basmann-Theil estimator (2-SLS) of a multi-equation system, where instead of the Euclidean norm the absolute norm is minimized. The idea of estimation in two stages is maintained. However, there is another parallel to 2-SLS suggested by Khazzoom (1976) that is more akin to the ILS.

Khazzoom investigated into generalization of ILS for an over-identified equation. He estimated 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, he applied generalized inverse of the relevant sub-matrix of reduced form coefficients to obtain the structural coefficients. This may be called the GILS (Generalized Indirect Least Squares) estimator.

It is natural to conjecture that since Khazzoom minimizes the Euclidean norm of errors, his method of estimation may share the shortcomings of ILS and 2-SLS, especially when the error vectors contain outliers. It is plausible that in presence of outliers an
estimator parallel to that of Khazzoom (but minimizes the absolute norm of errors) is more effective.

II. Objectives of the Study: The description in the preceding section suggests investigation in several lines. In the two-stage estimation procedure one may replace OLS by LAD

(i) at the first stage only (to obtain reduced form coefficients) and apply OLS at the second stage, or

(ii) at the second stage only (to obtain the structural coefficients while the reduced form coefficients at the first stage are obtained by OLS), or

(iii) at the first stage to obtain reduced form coefficients and consequently at the second stage also to obtain the structural coefficients. Suggestions given by Amemiya need implementation and empirical/experimental corroboration.

Similarly, in case of the Khazzoom estimator that applies OLS to obtain the reduced form coefficients, one may apply LAD instead for the same purpose. The present study basically aims at investigating into these possibilities. In particular, the generic objectives of the present study are given as hereunder:

(i) To study the relative performance of the LAD estimator (vis-à-vis OLS) in case of single equation models containing error vectors of different statistical distributions infested with outliers. It is required since single equation estimation is the basic building block in all methods that concern us.
(ii) To study the relative performance of the LAD estimator applied at different stages (vis-à-vis the standard 2-SLS estimator) in estimating the multi-equation models containing error vectors of different statistical distributions infested with outliers.

(iii) To study the relative performance of the LAD-based Khazzoom type estimator vis-à-vis the original Khazzoom estimator as well as various types of two-stage estimators applying OLS or LAD estimator in estimating the multi-equation models containing error vectors of different statistical distributions infested with outliers.

III. The Methodology: Owing to very poor mathematical tractability of LAD estimator we have relied on the Monte Carlo method to study the properties of various methods of estimation that we purport to investigate into. The Monte Carlo method of investigating into the properties of a mathematical system is basically experimental and numerical in nature. This method may be loosely described as a statistical simulation method, where statistical simulation is defined in quite general terms to be any method that utilizes sequences of random numbers to perform the simulation. The main idea behind the method is either to construct a stochastic model that is in agreement with the whole problem analytically or to simulate the whole problem directly. In both cases, an element of randomness has to be introduced according to some well-defined rules. Then a large number of trials are performed, the results are observed, and finally a statistical analysis is undertaken in the usual way. The advantages of the method are that even very difficult or analytically intractable problems can often be treated very easily, and desired modifications can be applied without much trouble. Nevertheless, the conclusions emanating from Monte Carlo studies are relatively less (vis-à-vis analytical methods)
accurate and the large number of trials that are necessary. Now, when very fast computers are readily available, the cumber of calculation in large number of trials is no longer a serious matter. This facility may be utilized to increase the number of trials sufficiently large so as to attain an acceptable degree of precision. Thus the trade-off between cumber and precision goes in favour of the Monte Carlo method.

In this investigation we have limited ourselves to deal with linear econometric models. Our single equation models are described as \( y = Xa + e \). The multi-equation models are described as \( YA + XB + E = 0 \). The disturbance term (\( e \) or \( E \) as the case may be) is stochastic and follows either of the five distributions: (i) Normal, (ii) Cauchy, (iii) Beta_1, (iv) Beta_2 and (v) Gamma. We have experimented with the cases when the disturbance term is infested with outliers, different in numbers and sizes. In the simulation we have experimented with the models of various sizes (different number of variables or equations) as well as with various sample sizes implying \( Y \) and \( X \) of varying dimensions.

In case of single equation models described as \( y = Xa + e \), (i) first \( X \) of the desired dimension is generated. The coefficients, \( a \), are assumed. These \( X \) and \( a \) remain fixed in a single bout of experiment. (ii) Then \( e \) of desired distribution is generated and \( y = Xa + e \) is obtained. The desired number of outliers (of sizes within some desired range) are generated and added to \( y \). Next, the desired method(s) of estimation is (are) applied to obtain \( \hat{a} \) which is the estimator of \( a \). The step \# ii is repeated a large, say \( m \), number of times. This gives \( m \) number of \( \hat{a} 's \) which are subjected to statistical analysis to assess the properties of the method(s) of estimation.

In case of multi-equation models described as \( YA + XB + E = 0 \) : (i) first \( X \) of the desired dimension is generated and the coefficients \( A \) and \( B \) are assumed. From \( A \) and \( B \)
we obtain $\Pi = -BA^{-1}$. (ii) Then $E$ of desired distribution is generated and $Y = X\Pi - EA^{-1}$ is obtained. The desired number of outliers (of sizes within some desired range) are generated and added to $Y$. Next, the desired method(s) of estimation is (are) applied to obtain $\hat{A}$ and $\hat{B}$ which are the estimators of $A$ and $B$. The step # ii is repeated a large, say $m$, number of times. This gives $m$ number of $\hat{A}$'s and $\hat{B}$'s which are subjected to statistical analysis to assess the properties of the method(s) of estimation.

IV. Organization of the Work: The subsequent part of this work is organized as follows: First, in chapter-2, the framework of linear econometric models is presented. Presently, estimation of econometric models by minimization of the Euclidean norm of errors is the most popular practice and hence, most of the estimation techniques are based on the Least Squares principle directly or indirectly. Details on these techniques are given therein.

In chapter-3 we present a literature survey on LAD estimation and the relevant numerical algorithms for that purpose. Estimation by minimization of absolute errors is mostly intractable by analytical methods and its performance (vis-à-vis LS-based methods) can in general be assessed only by simulation-based empirical methods such as the Monte Carlo method. In chapter-4, we present the salient features of the Monte Carlo method and give an outline that paves a way to using this simulation method to assess the relative performance of LAD-based estimators in single as well as multi equation econometric models. Since the Monte Carlo method is based on generating random numbers, we also present a description of some relevant statistical distributions that random numbers can follow and the computer-based techniques that may be used to generate random variates following such statistical distributions.
Next, in chapter-5, we conduct Monte Carlo experiments to compare the performance of LAD-based and LS-based methods of estimation of single equation econometric models. The residual terms in these models follow five different statistical distributions viz. normal, Cauchy, Gamma, Beta1 and Beta2. The residual terms may contain outliers different in number and magnitude.

Chapter-6 is an extension of the earlier chapter to the case of multi-equation econometric models. However, unlike in case of single-equation econometric models, we have several LS-based methods of estimation in case of the multi-equation econometric models. That is so because these models are estimated in two stages. Consequently, it is possible to replace LS by LAD at any one or both of these stages, giving rise to many possible combinations and possibilities of a comparative study in their performance. This has been done in Chapter-6.

We present a summary of our results in chapter-7. We have also visualized the lines of future research in application of LAD estimators of econometric models.