LEAST ABSOLUTE ERROR REGRESSION

by

YI-YUAN TUNG
B.Sc., National Chung-Hsing University, 1990

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Mathematics and Statistics

We accept this thesis as conforming
to the required standard

[redacted]
Dr. Roger R. Davidson, Supervisor (Department of Mathematics & Statistics)

[redacted]
Dr. William J. Reed, Department Member (Department of Mathematics & Statistics)

[redacted]
Dr. Jane J. Ye, Department Member (Department of Mathematics & Statistics)

[redacted]
Dr. John H. Drew, External Examiner (College of William & Mary, Williamsburg Va.)

Supervisor: Dr. Roger R. Davidson

ABSTRACT

Regression analysis is usually based on the method of least squares. However, the least squares estimator is highly sensitive to outliers and can be far from optimal when the errors are non-normal in distribution. As an alternative to the least squares estimator, the least absolute error estimator has attracted much attention during the past thirty years. With the development of several efficient algorithms based on its connections to linear programming, least absolute error regression can be put into practical use.

In this thesis, we review some basic theoretical properties of least absolute error regression, examine the asymptotic efficiency of the least absolute error estimator relative to the least squares estimator for a general family of error distribution, and demonstrate its use for practical data analysis in a variety of situations using computer routines written in S-Plus.

Examiners:

[redacted]

Dr. Roger R. Davidson, Supervisor (Department of Mathematics and Statistics)

[redacted]

Dr. William J. Reed, Department Member (Department of Mathematics and Statistics)

[redacted]

Dr. Jane J. Ye, Department Member (Department of Mathematics and Statistics)

[redacted]

Dr. John H. Drew, External Examiner (College of William & Mary, Williamsburg Va.)

Contents

List of Tables

List of Figures

Acknowledgement

I am deeply indebted to my academic supervisor, Dr. Roger R. Davidson, for his endless patience, generous guidance and great interest during my thesis work. I will never forget what he has given to me.

I would also like to thank Dr. William J. Reed, Dr. Jane J. Ye and Dr. John H. Drew for devoting time to examine my thesis and for providing me with their comments and suggestions.

Finally, special thanks are due to my dear parents and my lovely wife. I could not have even started my master program without their support and encougement.

Tony Y. Tung

Chapter 1

Introduction

In the late nineteenth century, the English scientist Sir Francis Galton introduced the word "*regression*" in a study which demonstrated that the sizes of offspring do not tend toward those of the parents, but rather toward the average as compared to the parents. For instance, he found that the sons of tall fathers tended to be tall, but not to the same extent. Galton termed this pattern of inheritance "*regression to the mean.*"

Regression analysis is one of the most widely used statistical techniques for analyzing multifactor data. The term regression analysis describes a collection of statistical techniques used to examine relationships among variables in a scientific system. Its broad appeal results from the simple concept of using an equation to express the relationship among a set of variables.

As a result of its simple and well developed theory, the method of least

squares estimation has dominated the statistical literature in regression analysis for a long time. Although credit for discovery of the method of least squares generally is given to Gauss, who used the procedure in the early part of the nineteenth century, Legendre published the first work on its use in 1805. The priority between Gauss and Legendre over this discovery is discussed by Stigler (1981). Also the issue of Gauss' contribution is discussed in a recent article by Celmiņš (1998).

In spite of its mathematical beauty and computational simplicity, the least squares estimator is highly sensitive to nonconforming observations. Indeed, one single outlier can have an arbitrarily large effect on the least squares estimator. Moreover, the least squares estimator is the maximum likelihood estimator when the errors are independent and normally distributed, but can be very different from the maximum likelihood estimator for non-normal error distributions, especially those with longer tails.

The investigation of alternative criteria to least squares goes back to Fourier in 1820. He proposed an iterative procedure to minimize the sum of absolute errors which is similar to the simplex method of linear programming. Edgeworth (1887,1888) suggested the same criterion and provided an awkward geometric algorithm. Since the introduction of linear programming techniques by Dantzig

(1955), several efficient algorithms have been developed for least absolute error estimation: see for example Wagner (1959), Barrodale & Young (1966) and Barrodale & Roberts (1973). With the development of these algorithms for least absolute error estimation and a theoretical basis for statistical inference provided by Bassett & Koenker (1978), least absolute error regression can be applied to practical data analysis as an attractive alternative to least squares regression.

Least absolute error regression has been studied in many contexts and identified with different names, such as L_1-norm, minimum (least) sum of absolute errors and minimum (least) absolute deviations (errors or values). Moreover, a variety of abbreviations can be found: MSAE, LSAE, MAD, MAE, LAD, LAV, LAE, etc. The term *least absolute error* (**LAE**) regression will be used throughout this thesis. For the purpose of comparison, the method of least squares will be referred to as *least squared error* (**LSE**) regression.

The principal objectives of this thesis are to investigate and summarize the basic theoretical properties of LAE regression, to demonstrate its use for practical data analysis, and to provide a set of S-Plus computer routines to obtain the estimated regression and perform standard inference.

Chapter 2 starts with the location problem and introduces L_p^n estimation

with emphasis on three special cases: least absolute error ($p = 1$), least squared error ($p = 2$) and minimax error ($p = \infty$). This class of estimators is then extended to the multiple regression setting. Since obtaining the LAE regression equation is equivalent to solving a linear programming problem, both the primal and dual formulations for LAE regression are examined. In that the LAE regression equation need not be unique, some important properties of the set of solutions for LAE regression are presented. At the end of Chapter 2, two real data examples are given together with the results of the computer analysis.

Chapter 3 deals with statistical aspects of LAE regression. The general form of the *Laplace-p* distributions is introduced with emphasis on three special cases: the double exponential ($p = 1$), the normal ($p = 2$) and the uniform ($p = \infty$) distributions. When the error terms in a multiple linear regression model follow the *Laplace-p* distribution, it is shown that the L_p^n estimators are maximum likelihood estimators. In particular, the LAE estimator is the maximum likelihood estimator when the errors follow the double exponential distribution. The fundamental asymptotic theory developed by Bassett & Koenker (1978) is discussed in detail. The asymptotic theory is then used to obtain the asymptotic relative efficiency of LAE to LSE estimation and to construct asymptotic standard errors and confidence intervals for the regression coefficients and for the

mean response. Once again, a detailed real data example is given.

Chapter 4 presents three utilizations of LAE regression, namely model selection, polynomial models and autoregressive models. In the absence of a property analogous to the partitioning identity in LSE regression, two criteria are suggested for model selection in LAE regression, which are the *Akaike's Information Criterion* (**AIC**) and the *Schwarz's Bayes Criterion* (**SBC**). Each of these criteria is based on the maximized likelihood combined with a penalty for model complexity. Two examples of model selection are presented. For each of the two remaining topics, a brief discussion together with a real data example is given.

The last chapter contains a summary and provides some suggestions on further research for LAE regression. Appendix A contains the four data sets used in the examples, and Appendix B contains listings of the S-Plus programs developed for this thesis.

Chapter 2

LAE Estimation and its Basic Properties

In this chapter, section 2.1 introduces the location problem and the L_p^n estimators. Section 2.2 considers LAE estimation for the general multiple regression model and the linear programming setup. Section 2.3 discusses some basic properties of the LAE estimators and section 2.4 contains two examples.

2.1 Location problem

Definition 2.1 *A real parameter θ is said to be a* **location parameter** *for the distribution of a random variable X if its distribution depends only on $x - \theta$, namely, if X has distribution function $F(x;\theta) = F_0(x-\theta)$, where $F_0(\cdot) = F(\cdot;0)$.*

For example, if a random variable X is distributed $N(\theta, 1)$, then θ is a location parameter for the distribution of X since $X - \theta$ is distributed $N(0,1)$ for all θ.

Let us consider a set of observations $x_1, \cdots, x_n$ from a distribution with location parameter θ, where we assume

$$x_i = \theta + \varepsilon_i \qquad i = 1, \cdots, n$$

In this representation, ε has distribution $F_0(\cdot)$ when X has distribution $F(\cdot; \theta)$. The following definition introduces a class of estimators of the location parameter θ for which three important special cases ($p = 1$, 2 and ∞) will be examined in turn.

Definition 2.2 *The estimator $\hat{\theta}$ of θ for which* $\| \mathbf{x} - \hat{\theta} \cdot \mathbf{1} \|_p = (\sum_{i=1}^{n} | x_i - \hat{\theta} |^p)^{\frac{1}{p}}$ *is a minimum is called the* $\boldsymbol{L_p^n}$ **estimator.**

When $p = 1$, the L_1^n estimator $\hat{\theta}$ which minimizes $\sum_{i=1}^{n} | x_i - \hat{\theta} |$ is the median of $\{x_i\}$, as can be easily seen by the following argument:
First, suppose n is an odd number, say $2m - 1$, so the *order statistic* of $\{x_i\}$ can be represented as $\{x_{(1)}, \cdots, x_{(m)}, \cdots, x_{(n)}\}$ and the median of $\{x_i\}$ is $x_{(m)}$. If we start with $\hat{\theta} = x_{(m)}$ and move $\hat{\theta}$ to the right by a small amount δ, i.e. $\hat{\theta} = x_{(m)} + \delta$, then we will increase m components of $\sum_{i=1}^{n} | x_i - \hat{\theta} |$ by δ and decrease only $m - 1$ components by the same amount, so that the summation $\sum_{i=1}^{n} | x_i - \hat{\theta} |$ will be increased by δ. Similarly, if we move $\hat{\theta}$ to the left of $x_{(m)}$ by δ, we have the same conclusion. Moreover, if we start at any $\hat{\theta} \neq x_{(m)}$ and

move $\hat{\theta}$ towards $x_{(m)}$ then the summation will decrease. Thus $\sum_{i=1}^{n} \mid x_i - \hat{\theta} \mid$ reaches its global minimum when $\hat{\theta} = x_{(m)}$.

If n is an even number, say $2m$, then the median of $\{x_i\}$ is $\frac{x_{(m)}+x_{(m+1)}}{2}$. By a similar argument as above, $\hat{\theta} = \frac{x_{(m)}+x_{(m+1)}}{2}$ is the L_1^n estimator. Note that if $\hat{\theta}$ equals any value between $x_{(m)}$ and $x_{(m+1)}$, then $\sum_{i=1}^{n} \mid x_i - \hat{\theta} \mid$ remains the same, so that the minimizing value is non-unique here.

When $p = 2$, the L_2^n estimator which minimizes $(\sum_{i=1}^{n} \mid x_i - \hat{\theta} \mid^2)^{\frac{1}{2}}$ is $\bar{x}$, the mean of $\{x_i\}$, as can be easily shown by the usual calculus method.

When $p = \infty$, the L_∞^n estimator which minimizes

$$\| \mathbf{x} - \hat{\theta} \cdot \mathbf{1} \|_\infty = \lim_{p\to\infty} (\sum_{i=1}^{n} \mid x_i - \hat{\theta} \mid^p)^{\frac{1}{p}} = \max_{1\le i\le n} \mid x_i - \hat{\theta} \mid$$

is the midrange of $\{x_i\}$, namely $\frac{x_{(1)}+x_{(n)}}{2}$.

Note that if $\hat{\theta} = \frac{x_{(1)}+x_{(n)}}{2}$, we have

$$\max_{1\le i\le n} \mid x_i - \hat{\theta} \mid = \frac{x_{(n)} - x_{(1)}}{2}$$

but for $\hat{\theta} \neq \frac{x_{(1)}+x_{(n)}}{2}$,

$$\max_{1\le i\le n} \mid x_i - \hat{\theta} \mid > \frac{x_{(n)} - x_{(1)}}{2}$$

$$(\text{either } \mid x_{(1)} - \hat{\theta} \mid \text{ or } \mid x_{(n)} - \hat{\theta} \mid > \frac{x_{(n)}-x_{(1)}}{2})$$

Thus $\hat{\theta} = \frac{x_{(1)}+x_{(n)}}{2}$ minimizes $\| \mathbf{x} - \hat{\theta} \cdot \mathbf{1} \|_\infty$.

2.2 Linear programming in LAE estimation

Consider the situation where a response variable Y is to be described in terms of one or more explanatory variables $X_1, X_2, \cdots, X_k$ on the basis of n observations $(x_{i1}, x_{i2}, \cdots, x_{ik}; y_i), i = 1, \cdots, n$. The linear regression model for Y on $X_1, X_2, \cdots, X_k$ is of the form :

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i \qquad i = 1, \cdots, n \tag{2.1}$$

where β_j, $j = 0, \cdots, k$, are the regression parameters and ε_i are independent and identically distributed random errors with location zero.

The observational equations can be written in matrix form as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2.2}$$

where

$$\begin{aligned} \mathbf{y} &= [y_1, y_2, \cdots, y_n]' \\ \mathbf{X} &= \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix} \\ \boldsymbol{\beta} &= [\beta_0, \beta_1, \cdots, \beta_k]' \\ \boldsymbol{\varepsilon} &= [\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n]' \end{aligned}$$

Note that the case of a single explanatory variable X is referred to as simple linear regression.

The L_p^n estimator of $\boldsymbol{\beta}$ is the value $\mathbf{b} = [b_0, b_1, \cdots, b_k]'$ which minimizes

$$\| \boldsymbol{\varepsilon} \|_p = \| \mathbf{y} - \mathbf{X}\mathbf{b} \|_p = (\sum_{i=1}^{n} | y_i - \hat{y}_i |^p)^{\frac{1}{p}}$$

When p=2, the L_2^n estimator of $\boldsymbol{\beta}$ is the **least squared error (LSE)** estimator. In the case where the columns of $\mathbf{X}$ are linearly independent, we obtain

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{2.3}$$

which is the unique solution to the *normal equation.* The properties of LSE estimators are discussed in detail in any introductory text on linear regression.

When p=1, the L_1^n estimator of $\boldsymbol{\beta}$ is the value $\mathbf{b} = [b_0, b_1, \cdots, b_k]'$ which minimizes $\sum_{i=1}^{n} | \varepsilon_i | = \sum_{i=1}^{n} | y_i - \hat{y}_i |$. The L_1^n estimator of $\boldsymbol{\beta}$ is referred to as the **least absolute error (LAE)** estimator. Because of the absolute value function, we cannot use conventional multivariable calculus methods as in the case of least squares. Instead, the minimization problem can be reformulated as an equivalent linear programming problem.

In order to convert the absolute value function to an equivalent linear form, we let

$$\varepsilon_i = \varepsilon_i^+ - \varepsilon_i^- \qquad \text{where } \varepsilon_i^+, \varepsilon_i^- \geq 0$$

Then the primal problem can be set up as follows (see Wagner (1959)):

$$\text{Minimize} \qquad \sum_{i=1}^{n} \varepsilon_i^+ + \sum_{i=1}^{n} \varepsilon_i^- \tag{2.4}$$

subject to

$$\sum_{j=0}^{k} x_{ij} b_j + \varepsilon_i^+ - \varepsilon_i^- = y_i \qquad i = 1, \cdots, n$$

$$\varepsilon_i^+, \varepsilon_i^- \geq 0$$

$$b_j \text{ unrestricted in sign}$$

The above linear programming formulation involves determining the values of the $k+1$ regression parameters $b_0, \cdots, b_k$ and the n pairs of residuals ε_i^+, ε_i^- which minimize the objective function subject to the n observational equations and the $2n$ nonnegativity requirements.

Note that ε_i^+ and ε_i^- can be considered as the positive residuals (above the regression line) and the negative residuals (below the regression line) respectively. For any optimal solution, ε_i^+ and ε_i^- cannot have positive values at the same time, otherwise we can reduce both ε_i^+ and ε_i^- by the minimum of ε_i^+ and ε_i^- to obtain a solution with a smaller value of the objective function.

Barrodale & Roberts (1973) introduced a very efficient algorithm for the above primal problem. Their algorithm, which combines several iterations of the simplex method into one, has become a standard feature of the algorithms

since then. The algorithm of Barrodale & Roberts is included in the library of Fortran procedures accessed by S-Plus.

The algorithm developed by Barrodale & Roberts identifies a vertex of the feasible region at which the optimum occurs. Such a vertex corresponds to an extreme LAE hyperplane, i.e. an LAE hyperplane passing through $k+1$ observations (see Appa & Smith (1973)). If there is only one vertex at which the optimum occurs, then the corresponding LAE hyperplane is unique. If the optimum occurs at more than one vertex of the feasible region, then all points in the convex hull of these vertices yield the optimum. Each feasible solution at which the optimum occurs corresponds to an LAE hyperplane which is a convex combination of the extreme LAE hyperplanes, i.e. whose coefficient vector is a convex combination of those of the extreme LAE hyperplanes. So, unlike the LSE criterion, the LAE criterion need not result in unique optimal solutions, an observation first made by Turner (1887).

Although the method of obtaining an LAE regression equation is to solve the corresponding primal problem using the algorithm of Barrodale & Roberts, an examination of the dual linear programming problem is of value in developing the properties of the LAE regression.

The dual problem can be formulated using dual variables u_i, $i = 1, \cdots, n$,

as follows:

$$\text{Maximize} \qquad \sum_{i=1}^{n} y_i u_i \tag{2.5}$$

subject to

$$\sum_{i=1}^{n} u_i = 0 \quad , \quad \sum_{i=1}^{n} x_{ij} u_i = 0 \qquad j = 1, \cdots, k$$

$$u_i \leq 1 \quad \text{(corresponding to } \varepsilon_i^+ \geq 0\text{)}$$

$$u_i \geq -1 \quad \text{(corresponding to } \varepsilon_i^- \geq 0\text{)}$$

$$u_i \text{ unrestricted in sign}$$

The above dual formulation involves determining the values of the n dual variables $u_1, \cdots, u_n$ which maximize the objective function subject to the $k+1$ equality and $2n$ inequality constraints. It then follows from the duality theorem of linear programming that both the primal and dual problems have finite optimal solutions which result in the same optimal value of their respective objective functions.

In the setting of LAE regression, the duality theory for linear programming has two important consequences which will be helpful in understanding the properties of solutions to LAE regression discussed in the next section (see Gaver and Thompson (1973) pp.194-8).

1. *For any LAE hyperplane, either a nonnegative primal variable takes the value zero or the corresponding dual constraint is satisfied as equality.*

   i.e.

   $$(i) \quad either \quad \varepsilon_i^+ = 0 \quad or \quad u_i = 1$$

   $$(ii) \quad either \quad \varepsilon_i^- = 0 \quad or \quad u_i = -1$$

   This is a direct consequence of the *complementary slackness theorem* which in the context of the primal and dual systems (2.4) and (2.5) states that

   $$\varepsilon_i^+(u_i - 1) = 0 \quad \text{and} \quad \varepsilon_i^-(u_i + 1) = 0 \quad \text{for all } i = 1, \cdots, n.$$

2. *In the case of multiple LAE hyperplanes where a nonnegative primal variable takes a positive value for some LAE solution, the corresponding dual constraint is satisfied as equality for all LAE solutions.*

   i.e.

   (i) *when* $\varepsilon_i^+ > 0$ *for some solution, then* $u_i = 1$ *for all solutions*

   (ii) *when* $\varepsilon_i^- > 0$ *for some solution, then* $u_i = -1$ *for all solutions*

   This is a direct consequence of the *extended complementary slackness theorem* which states that if $\{\varepsilon_i^{+(1)}, \varepsilon_i^{-(1)}\}$, $\{u_i^{(1)}\}$ and $\{\varepsilon_i^{+(2)}, \varepsilon_i^{-(2)}\}$, $\{u_i^{(2)}\}$ are any two optimal solutions to the primal and dual systems then

   $$\varepsilon_i^{+(j)}(u_i^{(h)} - 1) = 0 \quad \text{and} \quad \varepsilon_i^{-(j)}(u_i^{(h)} + 1) = 0 \quad \text{for } j, h = 1, 2.$$

2.3 Properties of solutions in LAE regression

In order to fully understand LAE regression, a number of authors have studied the properties of its solution space; see for example Appa & Smith (1973) and Gentle, Sposito & Kennedy (1977). Since many useful properties of LAE estimators depend on the presence of a constant term in the model, we will assume the regression model (2.2) in the following discussion.

Property 2.1 *There exists an LAE hyperplane passing through k+1 observations.*

This follows from the fact that the objective function of the primal problem achieves its minimum value at a vertex of the feasible region. Such a vertex corresponds to an LAE hyperplane which passes through $k+1$ observations.

Property 2.2 *Let n^*, n^+, n^- be the number of observations respectively lying on, above, below a given hyperplane. Then a necessary condition for the hyperplane to be LAE is that $\mid n^+ - n^- \mid \leq n^*$.*

This property follows from the first consequence of the complementary slackness theorem mentioned at the end of section 2.2; see Appa & Smith (1973) section 3.2.3 for details. Note that by Property 2.1, there exists an LAE hyperplane for which $n^* = k+1$.

The following two properties are direct consequences of Property 2.2.

Property 2.3 *An LAE hyperplane passing through none of the observations must have* $n^+ = n^-$.

Property 2.4 *For* n *odd, every LAE hyperplane passes through at least one observation.*

The next property is central to characterizing the set of LAE hyperplanes when the optimal solution is not unique.

Property 2.5 *No observation lies between two LAE hyperplanes, i.e. no observation point has a positive deviation with respect to one LAE hyperplane and a negative deviation with respect to another.*

According to the second consequence of the complementary slackness theorem mentioned at the end of section 2.2, we know that if the dual variable u_i associated with observation i equals 1 (-1) for any LAE hyperplane, then it equals 1 (-1) for all LAE hyperplanes. If observation i lies between two LAE hyperplanes, then the corresponding dual variable u_i will have value 1 with respect to one LAE hyperplane and value -1 with respect to another, which is a contradiction.

As a consequence of Properties 2.4 and 2.5, we now obtain the following property.

Property 2.6 *For n odd, there exists an observation which lies on every LAE hyperplane.*

In the case n odd, assume that there are two LAE hyperplanes, each of which passes through at least one observation, but such that no observation lies on both. According to Property 2.5, no observation can lie between these two LAE hyperplanes. However, every strictly convex combination of these two LAE hyperplanes is an LAE hyperplane, and hence will not pass through any observation, a conclusion which violates Property 2.4.

The following properties hold only for the simple linear model.

Property 2.7 *There are at most two extreme LAE lines that pass through any given observations.*

Property 2.8 *For n odd, there are at most two extreme LAE lines.*

Property 2.9 *There are at most four extreme LAE lines.*

For Property 2.7, suppose that an LAE line passes through a given observation point $\mathbf{p}$ (see Figure 2.1). In order that the value of the objective function not change, the LAE line can rotate in at most two directions about the given point, and according to Property 2.5 the rotation cannot pass beyond any other

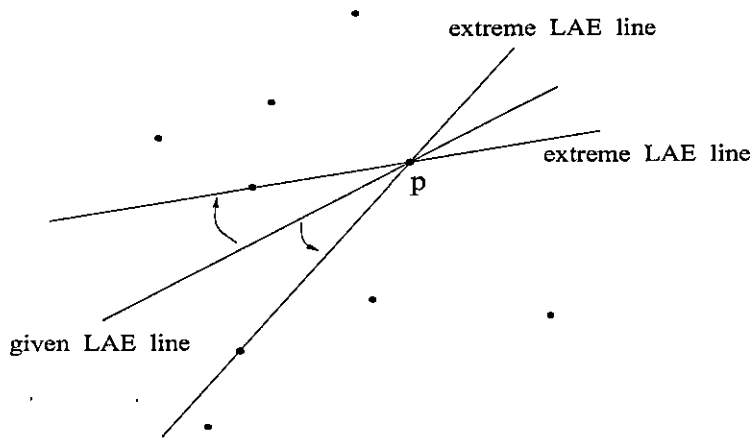

Figure 2.1: At most two extreme LAE lines passing through a given point

observation. Hence there are at most two extreme LAE lines passing through the given point and Property 2.7 follows.

Property 2.8 is a direct consequence of Properties 2.6 and 2.7.

For Property 2.9, we consider two cases: one with two parallel extreme LAE lines and the other with two intersecting extreme LAE lines. By using Property 2.5, it can be shown that there are at most two other extreme LAE lines for either case, so that at most four extreme LAE lines can exist. A detailed argument can be found in Gentle, Kennedy, and Sposito (1977).

For any data configuration with four extreme LAE lines, there are precisely four observations each of which lies on a distinct pair of extreme LAE lines. Such a configuration can only occur when the difference of the x-values for the left pair of these observations equals that for the right pair. In this situation,

any additional observations at distinct x-values must be equally divided and symmetrically located within each of the upper and lower convex hulls (see Figure 2.2 for an example).

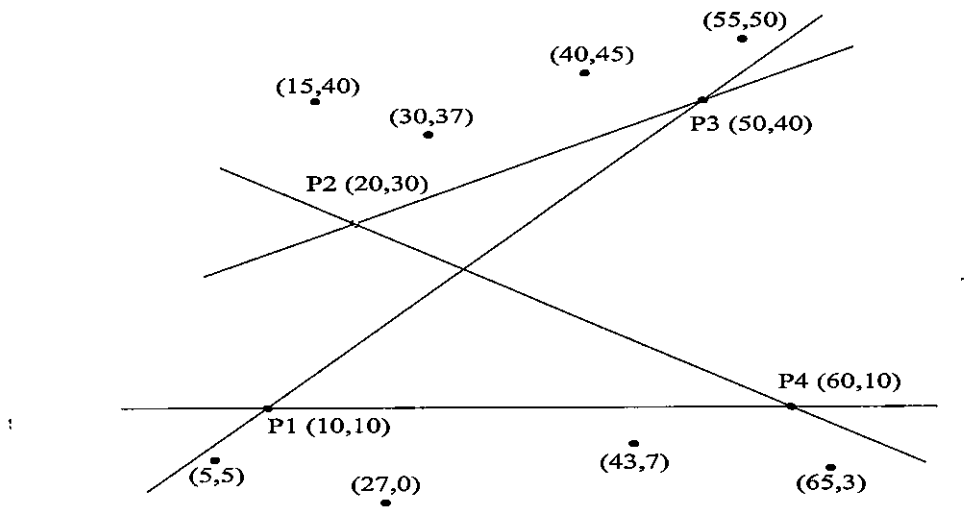

Figure 2.2: Data configuration with four extreme LAE lines

In the case of two or more explanatory variables, the number of extreme LAE hyperplanes can increase with n, the number of observation points. When n is odd, there exist data configurations for which there are $n-1$ extreme LAE hyperplanes; when n is even, there exist data configurations for which there are $n+2$ extreme LAE hyperplanes. Moreover, it is possible to have more than two extreme LAE hyperplanes passing through a given observation point. Gentle, Kennedy, & Sposito (1977) have given some examples to demonstrate these possibilities.

2.4 Examples

Example 2.1 Intra-ocular pressure data

The table in Appendix A.1 gives the mean intra-ocular pressure for twenty subjects at 16 successive equally spaced times. By using the S-Plus program *LAE.reg* in Appendix B.1, we obtain an LAE regression line relating intra-ocular pressure to time as

$$\hat{y} = 13.255 - 0.135x$$

with the 16 realized errors

0.6799994	0.5150008	0.0000000	0.0350003	-0.1300002	-0.2450004
-0.2099997	0.2249993	-0.2399999	-0.1550001	-0.3200012	0.0149987
0.0000000	0.1850002	-0.1300014	0.6549990		

and the minimum sum of absolute errors MinSAE = 3.74. Note that this LAE regression line is unique, and that it passes through two (3^{rd} and 13^{th}) observation points. We can also obtain the LSE regression line by the S-Plus program *LSE.reg* in Appendix B.2 as

$$\hat{y} = 13.394 - 0.145x$$

with the minimum sum of squared errors MinSSE = 1.4835. The plot of the observation points with the two regression lines is shown on Figure 2.3.

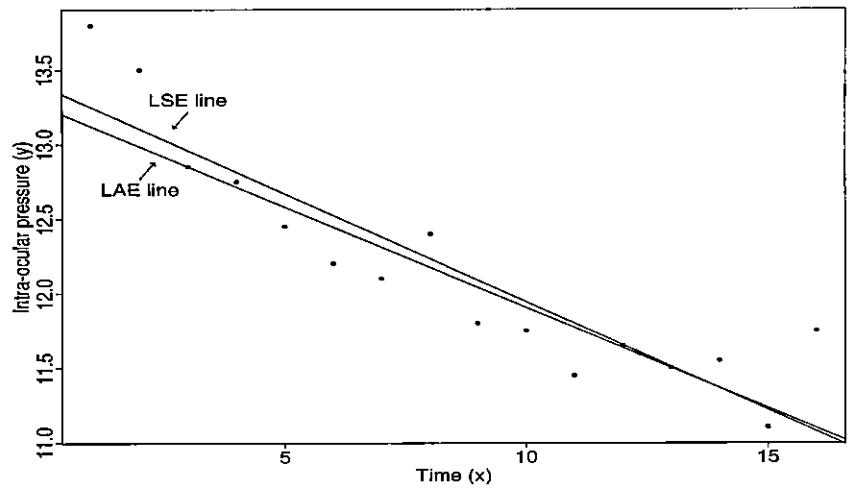

Figure 2.3: The LAE and LSE regression lines for the intra-ocular data

Example 2.2 Property valuation data

Consider the property valuation data in Appendix A.2 which gives 24 observations on the sale price of a house and 9 explanatory variables to which the sale price is potentially related. Since our purpose in this example is to provide a real data case with non-unique solutions, we will only use the first two explanatory variables (x_1 = taxes, x_2 = number of baths) in our analysis here. The multiple linear regression model is given by

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i \qquad i = 1, \cdots, 24$$

For LAE regression, the program *LAE.reg* gives the fitted model as

$$\hat{y} = 8.699961 + 2.372404 x_1 + 8.888674 x_2$$

with the 24 realized errors

-3.3551693	0.0000000	-0.4662301	-2.5003920	0.3077126	3.0803406
-0.6810724	-1.9833503	4.4845233	1.3369129	-1.4664559	-0.8262637
0.4307260	-4.6967626	3.9792790	0.0000000	4.8686328	5.7020907
-2.0261173	1.6014953	1.4354581	-2.9679289	0.0000000	2.0794618

and the minimum sum of absolute errors MinSAE = 50.27637. Notice that the LAE hyperplane passes through the 2^{nd}, 16^{th} and 23^{rd} observation points. Upon running the program, the message "Non-unique solution possible" is produced, and hence we know that there exists another set of estimates which yield the same MinSAE. In particular, the hyperplane

$$\hat{y} = 5.82905 + 2.372403x_1 + 11.75959x_2$$

which passes through the 2^{nd}, 21^{st} and 23^{rd} observation points is a second extreme LAE solution. Note that these two extreme LAE hyperplanes intersect in a line which is determined by the 2^{nd} and 23^{rd} observations. The program *LAE.reg* can then be used to establish that the extreme LAE hyperplanes given above are the unique LAE hyperplanes passing through the 16^{th} and 21^{st} observations respectively. Hence for this example the set of LAE solutions consists of all convex combinations of these two extreme LAE hyperplanes.

Chapter 3

Statistical Properties in LAE Regression

In this chapter, section 3.1 introduces the *Laplace-p* distribution. Section 3.2 develops the maximum likelihood estimators for the parameters of the *Laplace-p* distribution. Section 3.3 discusses some properties and presents the asymptotic theory for LAE regression. Section 3.4 derives the asymptotic standard errors and confidence intervals for the LAE regression parameters.

3.1 Laplace-p distribution

During the late 1700's, Laplace investigated families of symmetric distributions, to be used to describe errors of measurement, for which the maximum likelihood estimates of the location parameter are the median and the arithmetic mean. The resulting distributions have been referred to as the first and second laws

of Laplace, but are more commonly known as the double exponential (Laplace) distribution and the normal (Gaussian) distribution respectively.

By generalizing Laplace's approach to finding distributions for errors of measurement, a location-scale family of distributions, which will be referred to as the *Laplace-p* distributions, can be obtained.

Definition 3.1 *A continuous random variable X is said to have a Laplace-p distribution with location parameter θ, scale parameter $\alpha > 0$ and given $p > 0$, if the density of X is*

$$f_p(x;\theta,\alpha) = c(p)\frac{1}{\alpha}\exp\{-\frac{1}{p} \mid \frac{x-\theta}{\alpha} \mid^p\} \qquad \textit{where } c(p) = \frac{1}{2p^{\frac{1}{p}}\Gamma(\frac{1}{p}+1)}\,.$$

Note that the *maximum likelihood estimator* (**MLE**) of the location parameter θ is the value $\hat{\theta}$ which minimizes $(\sum_{i=1}^{n} \mid x_i - \hat{\theta} \mid^p)^{\frac{1}{p}}$, so that the MLE $\hat{\theta}$ is an L_p^n estimator. In addition, it can be shown that the MLE of the scale parameter α is $\hat{\alpha} = (\frac{\sum_{i=1}^{n}|x_i-\hat{\theta}|^p}{n})^{\frac{1}{p}}$. We will establish these results in the regression setting in the next section.

There are three important special cases of the family of the *Laplace-p* (θ,α) distributions :

(i) When $p = 1$, we obtain the double exponential distribution with density

$$f_1(x;\theta,\alpha) = \frac{1}{2\alpha}\exp\{-\frac{|\,x-\theta\,|}{\alpha}\}$$

The MLE of θ is the L_1^n estimator $\hat{\theta}$ = median of $\{x_i\}$ and the MLE of α is $\hat{\alpha} = \frac{\sum_{i=1}^{n}|x_i-\hat{\theta}|}{n}$.

(ii) When $p = 2$, we obtain the normal distribution with density

$$f_2(x;\theta,\alpha) = \frac{1}{\sqrt{2\pi}\alpha}\exp\{-\frac{(x-\theta)^2}{2\alpha^2}\}$$

The MLE of θ is the L_2^n estimator $\hat{\theta}$ = mean of $\{x_i\}$ and the MLE of α is $\hat{\alpha} = \left(\frac{\sum_{i=1}^{n}(x_i-\hat{\theta})^2}{n}\right)^{\frac{1}{2}}$.

(iii) When $p = \infty$, i.e. the limit as $p \to \infty$, we obtain the uniform distribution with density

$$f_\infty(x;\theta,\alpha) = \frac{1}{2\alpha} \qquad x \in [\theta-\alpha, \theta+\alpha]$$

The MLE of θ is the L_∞^n estimator $\hat{\theta}$ = midrange of $\{x_i\}$ and the MLE of α is the half range $\hat{\alpha} = \frac{x_{(n)}-x_{(1)}}{2}$.

3.2 Maximum likelihood estimation

The linear regression model (2.2) introduced in section 2.2 can be written as

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i \qquad i = 1, \cdots, n$$

with

$$\mathbf{x}_i' = [\, 1 \;\; x_{i1} \;\; \cdots \;\; x_{ik} \,]$$

where ε_i are independent and identically distributed random errors with location zero. We now add the assumption that the errors ε_i are distributed as *Laplace-p* (0,α). Equivalently, $y_1, \cdots, y_n$ are independent and the distribution of y_i given $\mathbf{x}_i$ is *Laplace-p* ($\mathbf{x}_i'\boldsymbol{\beta}$, α) with density

$$f_p(y_i \mid \mathbf{x}_i;\, \boldsymbol{\beta}, \alpha) = c(p)\frac{1}{\alpha}\exp\{-\frac{1}{p} \mid \frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}}{\alpha} \mid^p\}$$

The resulting likelihood function is

$$L(\boldsymbol{\beta}, \alpha) = c^n(p)\frac{1}{\alpha^n}\exp\{-\frac{1}{p}\sum_{i=1}^{n} \mid \frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}}{\alpha} \mid^p\}$$

and the corresponding log-likelihood function is

$$\ln L(\boldsymbol{\beta}, \alpha) = n \ln c(p) - n \ln \alpha - \frac{1}{p}\sum_{i=1}^{n} \mid \frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}}{\alpha} \mid^p$$

For any choice of $\alpha > 0$, $\ln L(\boldsymbol{\beta}, \alpha)$ is maximized by the value $\mathbf{b}$ of $\boldsymbol{\beta}$ which

minimizes $\sum_{i=1}^{n} \mid y_i - \mathbf{x}_i'\mathbf{b} \mid^p$, and hence the MLE $\mathbf{b}$ is also an L_p^n estimator of $\boldsymbol{\beta}$. Then the MLE of α is the value a which maximizes

$$\ln L(\mathbf{b}, \alpha) = n \ln c(p) - n \ln \alpha - \frac{1}{p\alpha^p} \sum_{i=1}^{n} \mid y_i - \mathbf{x}_i'\mathbf{b} \mid^p$$

Taking the derivative of $\ln L(\mathbf{b}, \alpha)$ with respect to α, we get

$$\frac{d \ln L(\mathbf{b}, \alpha)}{d \alpha} = -\frac{n}{\alpha} + \frac{1}{\alpha^{p+1}} \sum_{i=1}^{n} \mid y_i - \mathbf{x}_i'\mathbf{b} \mid^p$$

and by setting this derivative equal to zero, we obtain the MLE of α

$$a = \left(\frac{\sum_{i=1}^{n} \mid y_i - \mathbf{x}_i'\mathbf{b} \mid^p}{n} \right)^{\frac{1}{p}}$$

Note that $a^p = \frac{\sum_{i=1}^{n} |e_i|^p}{n}$ is the arithmetic average of the p^{th} power of the absolute values of the realized residuals $e_i = y_i - \mathbf{x}_i'\mathbf{b}$, $i = 1, \cdots, n$. Also note that a depends directly on $\sum_{i=1}^{n} \mid y_i - \mathbf{x}_i'\mathbf{b} \mid^p$, the minimum value under the L_p^n criterion.

Observe that as a result of the invariance property of MLE, the maximum likelihood estimatior of any function of $\boldsymbol{\beta}$ and α is obtained by evaluating the function at $\mathbf{b}$ and a. For example, $\hat{y} = \mathbf{x}_0'\mathbf{b}$ is the MLE of $E(y \mid \mathbf{x}_0) = \mathbf{x}_0'\boldsymbol{\beta}$ for any value $\mathbf{x}_0$.

3.3 Asymptotic theory for LAE regression

For the general linear regression model, the LSE estimator $\mathbf{b}$ given in (2.3) is a linear function of $\mathbf{y}$, and hence it follows under standard assumptions that $\mathbf{b}$ has an asymptotic multivariate normal distribution. As the LAE estimator does not possess this linearity, it is not easy to derive the asymptotic distribution of the LAE estimator directly from its formulation. By using a connection between the LAE estimator and the LSE estimator, Bassett & Koenker (1978) proved analytically that in a general linear model with independent and identically distributed errors having distribution function F, the LAE estimators are consistent and asymptotically normal with covariance matrix that depends on the asymptotic variance of the sample median when sampling from the distribution F. Before presenting details of this asymptotic theorem, let us investigate some preliminary results.

Property 3.1 *If* $\mathbf{X}$ *is of full rank, then every extreme LAE hyperplane can be represented as the LSE hyperplane for the subset of the* $(k+1)$ *observations through which the extreme LAE hyperplane passes. i.e. if* h *is the index set of the observations which lie on the extreme LAE hyperplane,* $\mathbf{X}(h)$ *and* $\mathbf{y}(h)$ *are the corresponding submatrices of* $\mathbf{X}$ *and* $\mathbf{y}$ *respectively, then the coefficient*

vector of the LAE hyperplane can be written

$$\mathbf{b} = \mathbf{X}(h)^{-1}\mathbf{y}(h)$$

In addition, the set of LAE hyperplanes is the convex hull of all solutions with this form.

From Property 2.1, there exists an LAE hyperplane passing through $(k+1)$ observation points, and hence the optimal hyperplane obtained by the LSE criterion on these $(k+1)$ observation points is the same as the LAE hyperplane. This property gives us a bridge by which to connect the asymptotic properties of LAE hyperplanes with those of LSE hyperplanes.

For the remaining properties, we need to introduce some notation and the directional derivative function. Let $S(\mathbf{b}) = \sum_{i=1}^{n} \mid y_i - \mathbf{x}_i'\mathbf{b} \mid$ be the sum of absolute residuals associated with the hyperplane $y = \mathbf{x}'\mathbf{b}$. Note that $S(\mathbf{b})$ is a convex surface consisting of connected planar segments in $(k+2)$ dimensional space, so that the minimum of $S(\mathbf{b})$ will occur on a region which is a point, a line, or a closed polygon. A point represents an unique solution to the estimation problem, while a line or closed polygon represents multiple solutions. Although $S(\mathbf{b})$ is not differentiable, the directional derivative function of $S(\mathbf{b})$ in arbitrary

direction $\mathbf{w}$ in the $(k+1)$ dimensional space for $\mathbf{b}$ is given by

$$\psi(S(\mathbf{b});\mathbf{w}) \;=\; -\sum_{i=1}^{n} \mathrm{sgn}^*(y_i - \mathbf{x}_i'\mathbf{b}\,,\, -\mathbf{x}_i'\mathbf{w})\,\mathbf{x}_i'\mathbf{w}$$

where

$$\mathrm{sgn}^*(u,v) = \begin{cases} \mathrm{sgn}(u) & \text{if } u \neq 0 \\ \mathrm{sgn}(v) & \text{otherwise} \end{cases}$$

In addition, let $\mathbf{B} = \{\, \mathbf{b} : \mathbf{b} \text{ minimizes } S(\mathbf{b}) \,\}$ be the LAE solution set.

Property 3.2 (i) *Vector* $\mathbf{b} \in \mathbf{B}$ *if and only if* $\psi(S(\mathbf{b}),\mathbf{w}) \geq 0$ *for all* $\mathbf{w} \neq \mathbf{0}$.

(ii) *Vector* $\mathbf{b}$ *is the unique solution if and only if* $\psi(S(\mathbf{b}),\mathbf{w}) > 0$ *for all* $\mathbf{w} \neq \mathbf{0}$.

This property states that the existence and uniqueness conditions of the LAE estimator are equivalent to the nonegativity and strict positivity of the directional derivative function. By the nature of the directional derivative function and the fact that $S(\mathbf{b})$ is convex, this property can be easily established.

For the explanation of the next two properties, which are the equivariance properties of the LAE estimator, we use $\mathbf{b}(\mathbf{y},\mathbf{X})$ to denote an LAE estimator based on the observations $\mathbf{y}$ and $\mathbf{X}$.

Property 3.3 *If* $\mathbf{b}(\mathbf{y},\mathbf{X}) \in \mathbf{B}(\mathbf{y},\mathbf{X})$ *then LAE solutions to the three transformed problems specified below satisfy the stated conditions:*

(i) $\mathbf{b}(\lambda\mathbf{y}, \mathbf{X}) = \lambda\,\mathbf{b}(\mathbf{y}, \mathbf{X})\,, \quad \lambda \in \mathbf{R}$

(ii) $\mathbf{b}(\mathbf{y} + \mathbf{X}\boldsymbol{\gamma}, \mathbf{X}) = \mathbf{b}(\mathbf{y}, \mathbf{X}) + \boldsymbol{\gamma}\,, \quad \boldsymbol{\gamma} \in \mathbf{R}^{k+1}$

(iii) $\mathbf{b}(\mathbf{y}, \mathbf{XA}) = \mathbf{A}^{-1}\,\mathbf{b}(\mathbf{y}, \mathbf{X})\,,$ $\mathbf{A}$ *nonsingular* $(k+1) \times (k+1)$ *matrix*

These conditions follow directly from the following equalities :

(i) $\mid \lambda \mid S(\mathbf{b}(\mathbf{y}, \mathbf{X})) = S(\lambda\,\mathbf{b}(\lambda\mathbf{y}, \mathbf{X}))$

(ii) $S(\mathbf{b}(\mathbf{y}, \mathbf{X})) = S(\mathbf{b}(\mathbf{y} + \mathbf{X}\boldsymbol{\gamma}, \mathbf{X}) + \boldsymbol{\gamma})$

(iii) $S(\mathbf{b}(\mathbf{y}, \mathbf{X})) = S(\mathbf{A}^{-1}\mathbf{b}(\mathbf{y}, \mathbf{XA}))$

where $S(\mathbf{b}(\mathbf{y}, \mathbf{X})) = \sum_{i=1}^{n} \mid y_i - \mathbf{x}_i'\mathbf{b} \mid$ is the sum of absolute residuals associated with the hyperplane generated by $\mathbf{b}(\mathbf{y},\mathbf{X})$. Note that estimators with conditions (i) and (ii) are termed scale and shift equivariant, respectively; estimators with condition (iii) are termed equivariant to linear transformation of $\mathbf{X}$. Also note that conditions (i)-(iii) are shared by the least squares estimator; but the next property, which is the invariance property of LAE estimation, does not hold for the least squares estimator.

Property 3.4 *If* $\mathbf{b}(\mathbf{y}, \mathbf{X}) \in \mathbf{B}(\mathbf{y}, \mathbf{X})$ *then LAE solutions satisfy the following condition:*

$$\mathbf{b}(\mathbf{X}\,\mathbf{b}(\mathbf{y}, \mathbf{X}) + \mathbf{D}\,\mathbf{e}, \mathbf{X}) = \mathbf{b}(\mathbf{y}, \mathbf{X})$$

where $\mathbf{D}$ *is a* $n \times n$ *diagonal matrix with nonnegative elements;* $\mathbf{e} \equiv \mathbf{y} - \mathbf{X}\,\mathbf{b}(\mathbf{y}, \mathbf{X})$ *is the vector of residuals.*

According to Property 3.2, $\mathbf{b}(\mathbf{y},\mathbf{X}) \in \mathbf{B}(\mathbf{y},\mathbf{X})$ implies

$$-\sum_{i=1}^{n} \operatorname{sgn}^*(y_i - \mathbf{x}_i'\mathbf{b}(\mathbf{y},\mathbf{X})\,,\ -\mathbf{x}_i'\mathbf{w})\,\mathbf{x}_i'\mathbf{w} \geq 0 \qquad \text{for all } \mathbf{w} \neq \mathbf{0}\,.$$

Note that

$$\operatorname{sgn}^*(\mathbf{x}_i'\mathbf{b}(\mathbf{y},\mathbf{X}) + d_i(y_i - \mathbf{x}_i'\mathbf{b}(\mathbf{y},\mathbf{X})) - \mathbf{x}_i'\mathbf{b}(\mathbf{y},\mathbf{X})\,,\ -\mathbf{x}_i'\mathbf{w})\,\mathbf{x}_i'\mathbf{w}$$

$$\leq \operatorname{sgn}^*(y_i - \mathbf{x}_i'\mathbf{b}(\mathbf{y},\mathbf{X})\,,\ -\mathbf{x}_i'\mathbf{w})\,\mathbf{x}_i'\mathbf{w} \qquad \text{for } d_i \geq 0\,,\ i = 1,\cdots,n\,,$$

so that

$$-\sum_{i=1}^{n} \operatorname{sgn}^*(\mathbf{x}_i'\mathbf{b}(\mathbf{y},\mathbf{X}) + d_i e_i - \mathbf{x}_i'\mathbf{b}(\mathbf{y},\mathbf{X})\,,\ -\mathbf{x}_i'\mathbf{w})\,\mathbf{x}_i'\mathbf{w} \geq 0 \qquad \text{for all } \mathbf{w} \neq \mathbf{0}\,,$$

and property 3.4 follows.

Property 3.4 has the following geometric interpretation. Image a plot of sample observations in $\mathbf{R}^{k+1}$ together with a particular LAE hyperplane slicing through the observation points. Now consider the effect of moving observations in the y direction. The result states that as long as these movements leave observations on the same side of the LAE hyperplane, the solution is unaffected. This is a generalization of the invariance property for the location problem, namely that a sample median is unaffected by moving observations on either side of the median.

The regression model (2.2) includes a constant term β_0 and hence without loss of generality the error distribution can be assumed to have median zero.

By using the previously stated properties, Bassett & Koenker (1978) proved the following asymptotic theorem for LAE estimation.

Theorem 3.1 *Let* $\{\mathbf{b_n},\ n = 1, 2, \cdots\}$ *denote a sequence of unique LAE estimators for the model* (2.2)*, and assume that*

(i) *the distribution function* F *of the random errors is continuous and has continuous and positive density* f *at the median, and*

(ii) $\lim_{n\to\infty} n^{-1}\mathbf{X}'\mathbf{X} = \mathbf{Q}$*, a positive definite matrix.*

Then $\sqrt{n}(\mathbf{b_n} - \boldsymbol{\beta})$ *converges in distribution to a* $(k+1)$ *dimensional Gaussian random vector with mean* 0 *and covariance matrix* $\omega^2\mathbf{Q}^{-1}$*, where* $n^{-1}\omega^2$ *is the asymptotic variance of the sample median for random samples from distribution* F*; i.e.,* $\omega = [2f(0)]^{-1}$.

This result implies that for any error distribution for which the median is superior to the mean as an estimator of location, the LAE estimator is preferable to the LSE estimator in the general linear model. Before expanding this connection, we now investigate the *asymptotic relative efficiency* (**ARE**) of the median to the mean under the *Laplace-p* distributions.

The density of random errors distributed with *Laplace-p*$(0, \alpha)$ is

$$f_p(x) = c(p)\frac{1}{\alpha}\exp\{-\tfrac{1}{p} \mid \frac{x}{\alpha} \mid^p\} \quad \text{where } c(p) = \frac{1}{2p^{\frac{1}{p}}\Gamma(\frac{1}{p}+1)},$$

so that the asymptotic variance of sample median is

$$\frac{\omega^2}{n} = \frac{[2f_p(0)]^{-2}}{n} = \frac{\alpha^2}{n} p^{\frac{2}{p}} \Gamma^2(\tfrac{1}{p}+1) ,$$

and the asymptotic variance of sample mean is

$$\frac{\sigma^2}{n} = \frac{1}{n}\int_{-\infty}^{\infty} x^2 f_p(x)dx = \frac{\alpha^2}{n} \frac{p^{\frac{2}{p}-1}\Gamma(\frac{3}{p})}{\Gamma(\frac{1}{p}+1)} .$$

Then the ARE of the median to the mean is given by

$$\text{ARE} = \frac{\sigma^2}{\omega^2} = \frac{p^2\Gamma(\frac{3}{p})}{\Gamma^3(\frac{1}{p})} .$$

From Figure 3.1, we can see that ARE decreases from 2 to $\frac{1}{3}$ as p increases from 1 to ∞. For the three special cases of the *Laplace-p* distributions, i.e. when $p = 1$, 2 and ∞, the ARE's are 2, $\frac{2}{\pi}$ and $\frac{1}{3}$ respectively. Also note that ARE$= 1$ when $p = 1.4074261$.

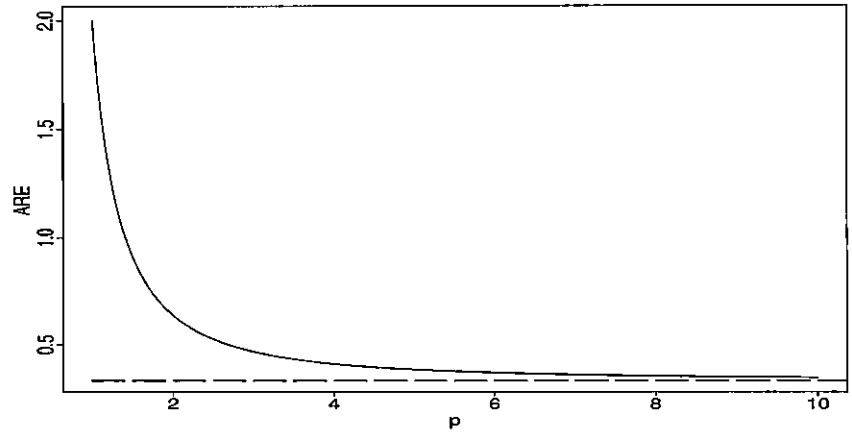

Figure 3.1: ARE of median to mean under *Laplace-p* distributions

Theorem 3.1 states that $\sqrt{n}(\mathbf{b_n}-\boldsymbol{\beta})$ converges in distribution to $N(0,\omega^2\mathbf{Q}^{-1})$, a $(k+1)$ dimensional normal distribution. According to the continuity theorem for convergence in distribution (Rao (1973) p.124), it then follows that the quadratic form

$$n(\mathbf{b_n}-\boldsymbol{\beta})'\mathbf{Q}(\mathbf{b_n}-\boldsymbol{\beta})/\omega^2$$

converges in distribution to the chi-square distribution with $(k+1)$ degrees of freedom. Since it has been assumed that $\lim_{n\to\infty} n^{-1}\mathbf{X}'\mathbf{X}=\mathbf{Q}$, the quadratic form

$$(\mathbf{b_n}-\boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\mathbf{b_n}-\boldsymbol{\beta})/\omega^2$$

differs from that above by an amount that converges to zero in probability and hence has the same limiting distribution (Rao (1973) p.122).

As a result, an $100(1-\gamma)\%$ asymptotic confidence region for $\boldsymbol{\beta}$ based on the LAE estimator is given by

$$(\mathbf{b}-\boldsymbol{\beta})'(\mathbf{X}'\mathbf{X})(\mathbf{b}-\boldsymbol{\beta}) \leq \omega^2\boldsymbol{\chi}^2_{\gamma;k+1}$$

where $\boldsymbol{\chi}^2_{\gamma;k+1}$ is the γ-critical value of the chi-square distribution with $(k+1)$ degree of freedom. Similarly, it follows from the corresponding limit theorem for least squares estimation that an $100(1-\gamma)\%$ asymptotic confidence region for $\boldsymbol{\beta}$ based on the LSE estimator is given by

$$(\mathbf{b}-\boldsymbol{\beta})'(\mathbf{X}'\mathbf{X})(\mathbf{b}-\boldsymbol{\beta}) \leq \sigma^2\boldsymbol{\chi}^2_{\gamma;k+1}$$

Hence we can conclude that when the ARE of the median to the mean exceeds 1, i.e. $\sigma^2 > \omega^2$, then LAE estimation will result in smaller asymptotic confidence ellipsoids and hence is preferable to LSE estimation in the general linear model.

It also follows from Theorem 3.1 above together with the corresponding limit theorem for the least squares estimator of $\boldsymbol{\beta}$ that the asymptotic relative efficiency of the LAE estimator of $\eta = \mathbf{c}'\boldsymbol{\beta}$, an arbitrary linear function of $\boldsymbol{\beta}$, to the LSE estimator of η, is the same as the ARE of the sample median to the sample mean, namely $\frac{\sigma^2}{\omega^2}$.

3.4 Standard errors and confidence intervals

Under the conditions stated in Theorem 3.1, the asymptotic variance-covariance matrix of $\mathbf{b}$, the LAE estimator of $\boldsymbol{\beta}$, is given by

$$\mathrm{Var}(\mathbf{b}) = \omega^2(n\mathbf{Q})^{-1} \quad .$$

Since $\lim_{n\to\infty} n^{-1}\mathbf{X}'\mathbf{X} = \mathbf{Q}$, so we can estimate $\mathbf{Q}$ by $\widehat{\mathbf{Q}} = n^{-1}\mathbf{X}'\mathbf{X}$. Moreover, since for a location-scale family ω is proportional to the scale parameter α, we will estimate ω by the value $\hat{\omega}$ obtained by estimating α by

$$\hat{\alpha} = \frac{1}{n-k-1}\sum_{i=1}^{n} \mid y_i - \hat{y}_i \mid \quad .$$

In particular, when the errors are distributed *Laplace-p*$(0, \alpha)$, we estimate ω by

$$\hat{\omega} = \frac{\hat{\alpha}}{2\,c(p)} = p^{\frac{1}{p}}\Gamma(\tfrac{1}{p}+1)\frac{1}{n-k-1}\sum_{i=1}^{n} \mid y_i - \hat{y}_i \mid \quad .$$

By using the estimates $\widehat{\mathbf{Q}}$ and $\hat{\omega}$ above, the estimated asymptotic variance-covariance matrix of $\mathbf{b}$ is given by

$$\widehat{\mathrm{Var}}(\mathbf{b}) = \hat{\omega}^2(\mathbf{X}'\mathbf{X})^{-1} \quad .$$

It then follows from the previous section that an appoximate $100(1-\gamma)\%$ confidence region for $\boldsymbol{\beta}$ is given by

$$(\mathbf{b}-\boldsymbol{\beta})'(\mathbf{X}'\mathbf{X})(\mathbf{b}-\boldsymbol{\beta}) \leq \hat{\omega}^2\boldsymbol{\chi}^2_{\gamma;k+1}$$

provided that the sampling variation associated with the estimation of ω can be assumed to be negligible.

The estimated standard error of b_j, the estimated coefficient of a specific X_j, is given by

$$s_{b_j} = \sqrt{\hat{\omega}^2(\mathbf{X}'\mathbf{X})^{-1}_{jj}}$$

where $(\mathbf{X}'\mathbf{X})^{-1}_{jj}$ denotes the j^{th} diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$, $j = 0, \cdots, k$. Hence, an approximate $100(1-\gamma)\%$ confidence interval for β_j can be based on the normal distribution, namely

$$(b_j - z_{\frac{\gamma}{2}}\, s_{b_j}\,,\; b_j + z_{\frac{\gamma}{2}}\, s_{b_j}) \quad .$$

For any specified $\mathbf{x}_0 = [1, x_{01}, \cdots, x_{0k}]'$, the estimated variance of $\hat{y}_{\mathbf{x}_0}$, the predicted value at $\mathbf{x}_0$, can be obtained as

$$\widehat{\mathrm{Var}}(\hat{y}_{\mathbf{x}_0}) = \mathbf{x}_0' \widehat{\mathrm{Var}}(\mathbf{b}) \mathbf{x}_0 = \hat{\omega}^2 \mathbf{x}_0' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0$$

so that an approximate $100(1-\gamma)\%$ confidence interval for $E(y \mid \mathbf{x}_0)$, the mean response at $\mathbf{x}_0$, is given by

$$(\hat{y}_{\mathbf{x}_0} - z_{\frac{\gamma}{2}}\, s_{\hat{y}_{\mathbf{x}_0}}\,,\, \hat{y}_{\mathbf{x}_0} + z_{\frac{\gamma}{2}}\, s_{\hat{y}_{\mathbf{x}_0}})$$

where $s_{\hat{y}_{\mathbf{x}_0}} = \sqrt{\hat{\omega}^2 \mathbf{x}_0' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0}$ is the estimated standard error of $\hat{y}_{\mathbf{x}_0}$. In particular, we can construct approximate $100(1-\gamma)\%$ confidence intervals for the mean response at each $\mathbf{x}_i$, $i = 1, \cdots, n$, as

$$(\hat{y}_i - z_{\frac{\gamma}{2}}\, s_{\hat{y}_i}\,,\, \hat{y}_i + z_{\frac{\gamma}{2}}\, s_{\hat{y}_i})$$

where $s_{\hat{y}_i} = \sqrt{\hat{\omega}^2 h_{ii}}$ is the estimated standard error of $\hat{y}_i$ and h_{ii} is the i^{th} diagonal element of the *hat* matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

Moreover, in accordance with the *Working-Hotelling* approach (see Myers (1992)), an approximate $100(1-\gamma)\%$ simultaneous confidence band on $E(y \mid \mathbf{x})$ for all $\mathbf{x}$ is given by

$$(\hat{y}_{\mathbf{x}} - \sqrt{\chi^2_{\gamma;k+1}}\, s_{\hat{y}_{\mathbf{x}}}\,,\, \hat{y}_{\mathbf{x}} + \sqrt{\chi^2_{\gamma;k+1}}\, s_{\hat{y}_{\mathbf{x}}}) \quad .$$

Example 3.1 Heat Transfer Data

Consider the heat transfer data in Appendix A.3 used to study the effect of certain variables in a chemical experiment dealing with heat transfer in a shallow fluidized bed. The data consists of 20 observations on the heat transfer coefficient and 4 explanatory variables. Only the first three variables (x_1 = fluidizing gas flow rate, x_2 = supernatant gas flow rate and x_3 = supernatant gas inlet nozzle opening) will be used in this example; the justification for this restriction will be given in section 4.2. The regression model to be analyzed is assumed to be

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i \qquad i = 1, \cdots, 20$$

The program *LAE.std* in Appendix B.3 gives several of the statistics based on the limiting distribution given in Theorem 3.1. These statistics are summarized in Table 3.1 and Table 3.2.

Parameter	Estimate	Standard error	95% Confidence limits	
β_0	-48.96038	20.99038	-90.10152	-7.81924
β_1	0.56005	0.07676	0.40959	0.71051
β_2	0.74109	0.07362	0.59679	0.88539
β_3	-1.29398	0.19387	-1.67397	-0.91400

Table 3.1: Heat transfer data — parameter estimates, asymptotic standard errors and 95% confidence limits for the model parameters.

Obs	y_i	$\hat{y}_i$	Standard error	95% Confidence limits	
1	41.852	58.44056	8.35775	42.05937	74.82176
2	155.329	152.72835	7.70357	137.62936	167.82734
3	99.628	99.62796	7.82799	84.28511	114.97082
4	49.409	20.68503	7.76255	5.47043	35.89963
5	72.958	72.95799	7.57277	58.11537	87.80061
6	107.702	115.66089	3.37930	109.03746	122.28433
7	97.239	113.80883	3.37228	107.19917	120.41850
8	105.856	114.55733	3.37215	107.94793	121.16674
9	99.348	72.11809	7.75597	56.91639	87.31979
10	111.907	143.09258	4.99558	133.30125	152.88392
11	100.008	115.86463	3.37659	109.24652	122.48275
12	175.380	184.17898	7.64773	169.18942	199.16853
13	117.800	116.12486	3.37778	109.50441	122.74531
14	217.409	161.04872	7.40302	146.53881	175.55864
15	41.725	43.37249	7.99514	27.70202	59.04297
16	151.139	152.05351	7.64728	137.06484	167.04218
17	200.630	207.75726	7.62799	192.80640	222.70811
18	131.666	131.66599	7.72598	116.52307	146.80891
19	80.537	80.53699	8.04261	64.77348	96.30050
20	152.966	149.78507	6.17117	137.68958	161.88055

Table 3.2: Heat transfer data — observed and fitted values, asymptotic standard errors and 95% confidence limits for the mean responses.

Chapter 4

Utilizations of LAE Regression

This chapter contains three utilizations of LAE regression: procedures for model selection, for polynomial regression models and for autoregressive models in time series analysis. For each of these, one or two real data examples will be given to illustrate the procedures.

4.1 Model selection

In many cases of multiple regression analysis, observations are available on a large number of potential explanatory variables. In identifying subsets of these variables which contribute significant information about the response, we need to consider the precision and the accuracy of a candidate model, as well as the purpose of the case being analyzed. Under the least squares criterion, standard

methods of variable selection are based on the partitioning identity

$$\text{SSTotal} = \text{SSReg} + \text{SSError}$$

or

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

which arises since the fitted values $\{\hat{y}_i\}$ and the realized residuals $\{e_i\}$ are orthogonal. The most common model selection criteria in least squares regression include the *adjusted-R^2*, the *MSE*, and the *Mallows' C_p*, each of which incorporates a penalty based on the sample size n and the number of variables $(k+1)$ in the candidate model.

In the case of L_p^n estimation, there is no counterpart to the partitioning identity when $p \neq 2$. As a result, we need to identify alternative criteria for model selection. Two methods that can be used when the form of the error distribution is assumed to be known are *Akaike's Information Criterion* (**AIC**) and *Schwarz's Bayes Criterion* (**SBC**). These methods, which are based on the maximized likelihood with a penalty based on n and/or k, are discussed by Schwarz (1978). The model preferred by the AIC and SBC is the one that reaches the minimum AIC and SBC values respectively.

When the errors are distributed *Laplace-p* $(0, \alpha)$, in which case the L_p^n esti-

mator of $\boldsymbol{\beta}$ is the maximum likelihood estimator, the AIC and SBC are

$$\begin{aligned} \text{AIC} &= -2\ln L(\mathbf{b}, a) + 2(k+1) \\ \text{SBC} &= -2\ln L(\mathbf{b}, a) + (k+1)\ln n \end{aligned}$$

where $L(\mathbf{b}, a)$ is the maximized likelihood obtained by evaluating $L(\boldsymbol{\beta}, \alpha)$ at the MLE's $\mathbf{b}$ and a (see section 3.2).

For the case $p = 1$, which corresponds to the double exponential distribution, the maximized log-likelihood is

$$\ln L(\mathbf{b}, a) = -n(1 + \ln 2) - n \ln \frac{\text{MinSAE}}{n}$$

where $\text{MinSAE} = \sum_{i=1}^{n} \mid y_i - \mathbf{x}_i'\mathbf{b} \mid$ is the minimum sum of absolute errors. Apart from a constant which depends only on n, the AIC and SBC can be written as

$$\begin{aligned} \text{AIC} &= 2n \ln \frac{\text{MinSAE}}{n} + 2(k+1) \\ \text{SBC} &= 2n \ln \frac{\text{MinSAE}}{n} + (k+1)\ln n \end{aligned}$$

The AIC and SBC can be used to select a model in the following manner. We first identify the optimal LAE regression model for each size k on the basis of the values of MinSAE for models of that size, and then select a suitable model from these models on the basis of the AIC and/or SBC values.

Note that the AIC and SBC can also be used for the case $p = 2$, which corresponds to the normal distribution. Analogous to the case $p = 1$, the AIC

and SBC can be written as

$$\begin{aligned} \text{AIC} &= n \ln \frac{\text{MinSSE}}{n} + 2(k+1) \\ \text{SBC} &= n \ln \frac{\text{MinSSE}}{n} + (k+1) \ln n \end{aligned}$$

where $\text{MinSSE} = \sum_{i=1}^{n} (y_i - \mathbf{x}_i' \mathbf{b})^2$ is the minimum sum of squared errors.

Example 4.1 Heat transfer data

Consider the heat transfer data in Appendix A.3. By applying the AIC and SBC discussed in the previous section, the program *LAE.best* in Appendix B.4 produces the best model for each size according to the value of the sum of the absolute residuals, and its corresponding MinSAE, AIC and SBC values. The results are summarized in Table 4.1. According to both the AIC and SBC, the model with input variables X_1, X_2 and X_3 has minimum values AIC = 107.5403 and SBC = 111.5232 respectively. Thus we will adopt it as the optimal model

Inputs				MinSAE	AIC	SBC
			2	505.7560	133.2129	135.2043
		1	2	374.0436	123.1456	126.1328
	1	2	3	240.8655	107.5403	111.5232
1	2	3	4	238.8813	109.2094	114.1880

Table 4.1: Best models by LAE estimation for the heat transfer data.

in this case. Also notice that the MinSAE values for the three and four variable models in Table 4.1 are very close, so we don't get much information by adding the variable X_4 to the three variable model. This explains the reason for using only the input variables X_1, X_2 and X_3 in Example 3.1.

For the purpose of comparison, we summarize the results generated from the program *LSE.best* for the heat transfer data in Table 4.2. Here we can see that the results by LSE estimation are in agreement with those obtained using LAE estimation.

Inputs				MinSSE	AIC	SBC
			2	30415.259	150.5393	152.5308
		1	2	17247.640	141.1490	144.1812
	1	2	3	6822.937	124.6463	128.6292
1	2	3	4	6784.906	126.5345	131.5131

Table 4.2: Best models by LSE estimation for the heat transfer data.

Example 4.2 Property valuation data

Consider the property valuation data in Appendix A.2. Table 4.3 presents the results generated by the program *LAE.best*. Here we can see that for the AIC the model with the six input variables X_2, X_3, X_5, X_6, X_8, and X_9 is optimal with minimum AIC = 37.88271. But for the SBC, the model with the four variables X_2, X_3, X_5, and X_7 turns out to be optimal with minimum SBC =

Inputs									MinSAE	AIC	SBC
								1	54.89574	43.71433	46.07044
							1	2	50.27637	41.49511	45.02927
						1	2	3	48.59923	41.86659	46.57880
					2	3	5	7	43.82432	38.90248	44.79275
				1	2	3	5	7	42.52568	39.45860	46.52693
			2	3	5	6	8	9	39.47274	37.88271	46.12908
		1	2	3	5	6	8	9	38.44535	38.61682	48.04125
	1	2	3	4	5	6	8	9	38.16454	40.26494	50.86743
1	2	3	4	5	6	7	8	9	38.11334	42.20051	53.98105

Table 4.3: Best models by LAE estimation for the property valuation data.

44.79275. Note that the four variable model is third best by the AIC and the difference in AIC values between the four and six variable models is smaller than that in SBC values. This suggests that the model with the input variables X_2, X_3, X_5, and X_7 is the most appropriate choice when considering both the AIC and SBC. In addition, some attention could be given to the model with two variables X_1 and X_2, since it is the second best model by the SBC.

Now let us look at Table 4.4 which contains the results generated by the program *LSE.best*. The optimal models suggested by the AIC and SBC are the four variable model with minimum AIC = 51.73869 and the two variable model with minimum SBC = 55.61576 respectively. Note that the best four variable model using LSE estimation has a different set of input variables than the one using LAE estimation. Since the two variable model is the second best by the

Inputs									MinSSE	AIC	SBC
								1	192.8906	54.01766	56.37377
							1	2	163.7139	52.08160	55.61576
						1	2	9	153.7997	52.58234	57.29455
					1	2	5	7	136.6149	51.73869	57.62896
				1	2	3	5	7	130.7970	52.69422	59.76254
			1	2	4	5	7	9	126.1405	53.82422	62.07060
		1	2	4	5	7	8	9	122.8959	55.19881	64.62324
	1	2	3	4	5	7	8	9	122.0067	57.02453	67.62702
1	2	3	4	5	6	7	8	9	121.7482	58.97362	70.75416

Table 4.4: Best models by LSE estimation for the property valuation data.

AIC and the four variable model is the fourth best by the SBC, the model with the input variables X_1 and X_2 will be chosen when considering both the AIC and SBC. Moreover, this model is a suitable choice for either LAE or LSE estimation.

4.2 Polynomial models

In regression analysis with only one explanatory variable, a simple linear regression model may not adequately represent the underlying relationship between the response and explanatory variables. In such situations the *k^{th}-order polynomial regression model* given by

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_k x_i^k + \varepsilon_i \qquad i = 1, \cdots, n \qquad (4.1)$$

may prove to be more appropriate. If we set $X_j = X^k,\ j = 1, 2, \cdots, k$, then (4.1) becomes a multiple linear model in the k regressors $X_1, X_2, \cdots, X_k$. Thus a polynomial model of order k may be estimated using LAE and LSE regression as discussed previously. In addition the order of the polynomial model to be used can be decided upon by using the AIC and SBC introduced in section 4.1.

It should also be noted that regression models with two or more explanatory variables can be easily generalized to polynomial models of order k which may then be reexpressed as multiple linear models with larger numbers of regressors. For instance a second order polynomial model with two explanatory variables X_1 and X_2 can be represented as a multiple linear model with five regressors X_1, X_2, X_1^2, X_2^2 and X_1X_2.

The following example illustrates some of the analyses typically associated with fitting a polynomial model in one variable.

Example 4.3 Intra-ocular pressure data

The intra-ocular pressure data in Appendix A.1 contains only one explanatory variable X. By letting $X_1 = X$, $X_2 = X^2$, $X_3 = X^3$ and $X_4 = X^4$, the degree of a polynomial regression model can be decided upon using the AIC and/or SBC. Table 4.5 gives the values of MinSAE, AIC and SBC by LAE estimation for the four polynomial models. According to each criterion the gap between the

Inputs				MinSAE	AIC	SBC
			1	3.740000	-42.51210	-40.96692
		1	2	1.987545	-60.74204	-58.42427
	1	2	3	1.890461	-60.34457	-57.25422
1	2	3	4	1.724184	-61.29071	-57.42776

Table 4.5: Polynomial models by LAE estimation for the intra-ocular data.

linear and quadratic models is large and we don't get significant inprovement by adding the cubic and quartic terms, and so the quadratic model will be chosen as the most appropriate one. The fitted quadratic model using LAE estimation is $\hat{y} = 14.1936 - 0.4093x + 0.0158x^2$.

The results for LSE estimation, which are summarized in Table 4.6, are similar to those for LAE estimation. Once again the quadratic model is suggested and the fitted LSE equation is $\hat{y} = 14.0473 - 0.3627x + 0.0128x^2$.

Inputs				MinSSE	AIC	SBC
			1	1.483492	-34.05103	-32.50585
		1	2	0.545425	-48.06045	-45.74268
	1	2	3	0.537118	-46.30603	-43.21567
1	2	3	4	0.424438	-48.07326	-44.21032

Table 4.6: Polynomial models by LSE estimation for the intra-ocular data.

4.3 Autoregressive models

In the analysis of time series, a process $\{X_t\}$ with location parameter θ is said to be an *autoregressive process* of order l if

$$X_t - \theta = \beta_1(X_{t-1} - \theta) + \beta_2(X_{t-2} - \theta) + \cdots + \beta_l(X_{t-l} - \theta) + \varepsilon_t \qquad (4.2)$$

where $\{\varepsilon_t\}$ is a purely random process with location zero. In the context of multiple regression, the model (4.2) can be written as

$$X_t = \beta_0 + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + \beta_l X_{t-l} + \varepsilon_t \qquad (4.3)$$

where $\beta_0 = \theta(1 - \sum_{i=1}^{l} \beta_i)$. In order that this model correspond to a stationary autoregressive process of order l, it suffices that the roots of

$$\beta_1 m + \beta_2 m^2 + \cdots + \beta_l m^l = 1$$

lie outside the unit circle. The principle difference between the model (4.3) and the conventional multiple regression model is that the variable X_t is regressed on lagged values of the series $\{X_t\}$.

Provided the time series process is stationary, the model parameters in (4.3) can be estimated either by the LAE or LSE criteria. An alternative estimator which is asymptotically equivalent to the LSE estimator can be obtained by substituting the *sample autocorrelations* into the *Yule-Walker equations* and

solving for the parameters of the autoregressive model (see Chatfield (1989)). However, there does not exist an alternative to LAE estimation that is analogous to the *Yule-Walker* method.

By examining a real data example, we will demonstrate the use of LAE estimation for autoregressive time series processes.

Example 4.4 U.S. Unemployment Rate Data

Consider the data in Appendix A.4 which gives the quarterly seasonally adjusted United States unemployment rate from 1948 to 1972. The plot of data is displayed on Figure 4.1. In discussing the analysis of these data, Fuller (1996)

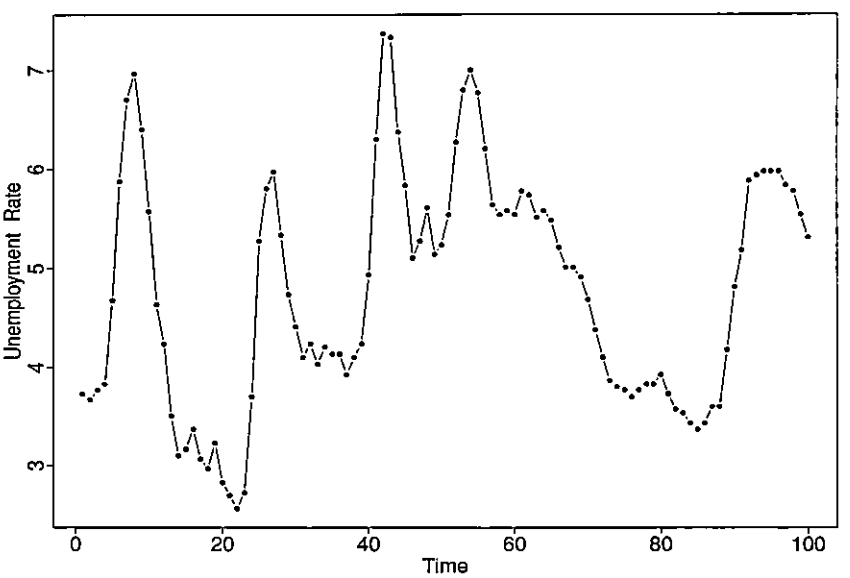

Figure 4.1: Quarterly seasonally adjusted U.S. unemployment rate 1948-1972

notes that the first few sample *autocorrelations* of the unemployment rate time series are in good agreement with those generated by a second order autoregressive process. Hence we will adopt a second order model and estimate the autoregressive parameters using both the LAE and LSE criteria.

Using LAE estimation, the second order autoregressive model corresponding to (4.3) obtained by conditioning on the first two observations and regressing X_t on X_{t-1} and X_{t-2} is given by

$$\hat{X}_t = 0.44734 + 1.55988X_{t-1} - 0.66301X_{t-2}$$

with MinSAE = 23.02817. Since the roots of $1.55988m - 0.66301m^2 = 1$ are $1.176 \pm 0.353i$ which lie outside the unit circle, thus we can conclude that this second order time series process appears to be stationary. The location parameter θ is then estimated by

$$\begin{aligned} \hat{\theta} &= \hat{\beta}_0(1 - \hat{\beta}_1 - \hat{\beta}_2)^{-1} \\ &= 0.44734(1 - 1.55988 + 0.66301) \\ &= 4.33763 \end{aligned}$$

which is somewhat smaller than the mean and median of the series of 100 observations, i.e. $\bar{x} = 4.77$ and $\tilde{x} = 4.765$ respectively.

Using LSE estimation, the second order model is given by

$$\hat{X}_t = 0.62812 + 1.56792X_{t-1} - 0.69862X_{t-2}$$

with MinSSE = 9.97354. It can be shown that the stationary condition is satisfied and the estimated location is $\hat{\theta} = 4.80581$ which is closer to the mean and median of the data than that for LAE estimation.

Chapter 5

Summary and Suggested Further Work

In our discussion of LAE estimation, we have demonstrated that estimating the model coefficients for LAE regression is equivalent to solving a linear programming problem. With the development of several efficient algorithms and computer codes, LAE regression has been brought into practical use as can be seen from the two real data examples given at the end of Chapter 2.

It has been shown that the LAE estimators are maximum likelihood and hence asymptotically efficient when the errors follow the double exponential distribution. In addition, Theorem 3.1 points out that the LAE estimator is more efficient than the LSE estimator in the linear regression model for any error distribution for which the sample median is a more efficient estimator of location than the sample mean. These results strengthen the basis of LAE regression as a suitable alternative to LSE regression.

The AIC and SBC, which are based on the maximized likelihood, provide methods for identifying the most appropriate submodel for LAE regression. From the formulation of AIC and SBC, it was seen that the SBC tends to prefer models with fewer variables than does AIC. This follows from the difference in the penalty terms for model complexity.

Since LAE regression may result in nonunique solutions, it would be desirable to be able to specify a compromise solution for this situation. Harter (1977) considered the problem of nonuniqueness of LAE regression for the simple linear model. He has shown that in the case of two extreme LAE lines, the best compromise solution in accordance with Edgeworth (1888) is not the line which bisects the angle between them but the line which is the locus of points whose vertical distances from them are equal. Harter also proposed a compromise solution for the case of four extreme LAE lines, but had no proposal for the case of three extreme LAE lines. For the case with three extreme LAE lines, there are exactly three observations each of which is located at the intersection of two extreme LAE lines, and all LAE lines pass through two of the three line segements determined by these three observations. Hence a possible compromise solution can be obtained by joining bisectors of the two line segments through which all LAE lines pass. Similarly, for the case with four extreme LAE lines, there are exactly four observations each of which is located at the intersection of

two extreme LAE lines, and all LAE lines pass through two opposite sides of the quadrilateral determined by these four observations. Hence the same criterion can be used to find a compromise solution. For the multiple linear regression model, we have not yet found a satisfactory method to specify a compromise solution. It would be useful to investigate further possible criteria for specifying a compromise solution for LAE regression.

Bassett (1992) introduced the Gauss Markov property for the median in the location problem. His result establishes that the median is the best estimator of location in a class of median-unbiased estimators, just as the sample mean is the best estimator in a class of mean-unbiased estimators. The generalization of this Gauss Markov property to LAE regression still needs to be explored.

Small sample properties of the LAE estimators were investigated using Monte Carlo studies by a number of authors in the 1960's and 1970's. These studies confirm that the LAE estimators have little or no bias. However, the relationship of the small sample properties revealed through Monte Carlo simulations and the large sample properties based on the results of Bassett & Koenker (1978) has not been adequately established. Further work needs to be devoted to Monte Carlo simulations in order to establish the practical value of the asymptotic results.

Bibliography

[1] Appa, G. & Smith, C. (1973). On L_1 and Chebyshev estimation. *Mathematical Programming* **5**, 73-87.

[2] Barrodale, I. & Roberts, F.D.K. (1973). An improved algorithm for discrete L_1 linear approximation. *S.I.A.M. Journal on Numerical Analysis* **10**, 839-848.

[3] Barrodale, I. & Roberts, F.D.K. (1974). Solution of an overdetermined system of equations in the L_1 norm. *Communications of the ACM* **17**, 319-320.

[4] Barrodale, I. & Young, A. (1966). Algorithms for best L_1 and L_∞ linear approximations on a discrete set. *Numerische Mathematik* **8**, 295-306.

[5] Bassett, G.W. (1992). The Gauss Markov property for the median. In *L_1-Statistical Analysis and Related Methods*, edited by Y. Dodge, Elsevier, 23-30.

[6] Bassett, G.W. & Koenker, R. (1978). Asymptotic theory of least absolute error regression. *Journal of the American Statistical Association* **73**, 618-622.

[7] Celmiņš, A. (1998). The method of Gauss in 1799. *Statistical Science* **13**, 123-135.

[8] Chatfield, C. (1989). *The Analysis of Time Series: an Introduction*, Chapman & Hall.

[9] Dantzig, G. (1955). Upper bounds, secondary constraints and block triangularity in linear programming. *Econometrica* **23**, 174-183.

[10] Edgeworth, F.Y. (1887). On observations relating to several quantities. *Philosophical Magazine (5th Series)* **24**, 222-223.

[11] Edgeworth, F.Y. (1888). On a new method of reducing observations relating to several quantities. *Philosophical Magazine (5th Series)* **25**, 184-191.

[12] Fuller, W.A. (1996). *Introduction to Statistical Time Series*, 2nd edition, John Wiley & Sons.

[13] Gaver, D.P. & Thompson, G.L. (1973). *Programming and Probability Models in Operations Research*, Brooks Cole.

[14] Gentle, J.E. & Kennedy, W.J. & Sposito, V.A. (1977). On least absolute values estimation. *Communications in Statistics* A**6**, 839-845.

[15] Harter, H.L. (1977). Nonuniqueness of least absolute values regression. *Communications in Statistics* A**6**, 829-838.

[16] Myers, H.M. (1990). *Classical and Modern Regression with Applications*, 2nd edition, PWS-Kent.

[17] Narula, S.C. & Wellington, J.F. (1982). The minimum sum of absolute errors regression: a State of the Art Survey. *International Statistical Review* **50**, 317-326.

[18] Rao, C.R. (1973). *Linear Statistical Inference and its Applications*, 2nd edition, John Wiley & Sons.

[19] Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* **6**, 461-464.

[20] Stigler, S.M. (1981). Gauss and the invention of least squares. *Annals of Statistics* **9**, 465-474.

[21] Turner, H.H. (1887). On Mr. Edgeworth's method of reducing observations relating to several quantities. *Philosophical Magazine (5th Series)* **24**, 466-470.

[22] Wagner, H.M. (1959). Linear programming techniques for regression analysis. *Journal of the American Statistical Association* **54**, 206-212.

Appendix A

Data Sets

A.1 Intra-ocular pressure data

A.2 Property valuation data

A.3 Heat transfer data

A.4 U.S. unemployment rate data

A.1 Intra-ocular pressure data

The following table gives the mean intra-ocular pressure for twenty subjects at 16 successive equally spaced times in an experiment to determine the variation of intra-ocular pressure with time.

Time x	Intra-ocular pressure y
1	13.80
2	13.50
3	12.85
4	12.75
5	12.45
6	12.20
7	12.10
8	12.40
9	11.80
10	11.75
11	11.45
12	11.65
13	11.50
14	11.55
15	11.10
16	11.75

x: Time period
y: Intra-ocular pressure

Source: Department of Opthamology, University of British Columbia, Vancouver, British Columbia. Also see "An Introduction to Regression and Correlation," by K. W. Smillie, the Ryerson Press and Academic Press, pp. 15, 1966.

A.2 Property valuation data

The following table presents data for 24 houses sold in Erie, Pennsylvania.

obs	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
1	25.9	4.9176	1.0	3.4720	0.9980	1.0	7	4	42	0
2	29.5	5.0208	1.0	3.5310	1.5000	2.0	7	4	62	0
3	27.9	4.5429	1.0	2.2750	1.1750	1.0	6	3	40	0
4	25.9	4.5573	1.0	4.0500	1.2320	1.0	6	3	54	0
5	29.9	5.0597	1.0	4.4550	1.1210	1.0	6	3	42	0
6	29.9	3.8910	1.0	4.4550	0.9880	1.0	6	3	56	0
7	30.9	5.8980	1.0	5.8500	1.2400	1.0	7	3	51	1
8	28.9	5.6039	1.0	9.5200	1.5010	0.0	6	3	32	0
9	35.9	5.8282	1.0	6.4350	1.2250	2.0	6	3	32	0
10	31.5	5.3003	1.0	4.9883	1.5520	1.0	6	3	30	0
11	31.0	6.2712	1.0	5.5200	0.9750	1.0	5	2	30	0
12	30.9	5.9592	1.0	6.6660	1.1210	2.0	6	3	32	0
13	30.0	5.0500	1.0	5.0000	1.0200	0.0	5	2	46	1
14	36.9	8.2464	1.5	5.1500	1.6640	2.0	8	4	50	0
15	41.9	6.6969	1.5	6.9020	1.4880	1.5	7	3	22	1
16	40.5	7.7841	1.5	7.1020	1.3760	1.0	6	3	17	0
17	43.9	9.0384	1.0	7.8000	1.5000	1.5	7	3	23	0
18	37.5	5.9894	1.0	5.5200	1.2560	2.0	6	3	40	1
19	37.9	7.5422	1.5	5.0000	1.6900	1.0	6	3	22	0
20	44.5	8.7951	1.5	9.8900	1.8200	2.0	8	4	50	1
21	37.9	6.0831	1.5	6.7265	1.6520	1.0	6	3	44	0
22	38.9	8.3607	1.5	9.1500	1.7770	2.0	8	4	48	1
23	36.9	8.1400	1.0	8.0000	1.5040	2.0	7	3	3	0
24	45.8	9.1416	1.5	7.3262	1.8310	1.5	8	4	31	0

y: Sale price of the house/1000
x_1: Taxes/1000
x_2: Number of baths
x_3: Lot size (sq ft $\times$ 1000)
x_4: Living space (sq ft $\times$ 1000)
x_5: Number of garage stalls

x_6: Number of rooms
x_7: Number of bedrooms
x_8: Age of the home (years)
x_9: Number of fireplaces

Source: "Prediction, Linear Regression and Minimum Sum of Relative Errors," by S. C. Narula and J. F. Wellington, *Technometrics*, **19**, pp. 185-90, 1977. "Letter to the Editor," *Technometrics*, **22**, pp. 452, 1980. Also see "Introduction to Linear Regression Analysis," 2nd edition, by D. C. Montgomery and E. A. Peck, John Wiley & Sons, Table B.4, 1992.

A.3 Heat transfer data

The following data are from a chemical engineerring experiment dealing with heat transfer in a shallow fluidized bed.

obs	y	x_1	x_2	x_3	x_4
1	41.852	69.69	170.83	45	219.74
2	155.329	113.46	230.06	25	181.22
3	99.628	113.54	228.19	65	179.06
4	49.409	118.75	117.73	65	281.30
5	72.958	119.72	117.69	25	282.20
6	107.702	168.38	173.46	45	216.14
7	97.239	169.85	169.85	45	223.88
8	105.856	169.85	170.86	45	222.80
9	99.348	170.89	173.92	80	218.84
10	111.907	171.31	173.34	25	218.12
11	100.008	171.43	171.43	45	219.20
12	175.380	171.59	263.49	45	168.62
13	117.800	171.63	171.63	45	217.58
14	217.409	171.93	170.91	10	219.92
15	41.725	173.92	71.73	45	296.60
16	151.139	221.44	217.39	65	189.14
17	220.630	222.74	221.73	25	186.08
18	131.666	228.90	114.40	25	285.80
19	80.537	231.19	113.52	65	286.34
20	152.966	236.84	167.77	45	221.72

y: Heat transfer coefficient
x_1: Fluidizing gas flow rate (lb/hr)
x_2: Supernatant gas flow rate (lb/hr)
x_3: Supernatant gas inlet nozzle opening (mm)
x_4: Supernatant gas inlet temperature (^{o}F)

Source: Statistical Consulting Center, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 1982. Also see "Classical and Modern Regression with Applications," 2nd edition, by R. H. Myers, PWS-Kent Publishing Company, pp. 200, 1990.

A.4 U.S. unemployment rate data

The following table gives the quarterly seasonally adjusted United States unemployment rate from 1948 to 1972. Here the quarterly data are the average of monthly data.

	Quarterly Unemployment Rate			
Year	Quarter1	Quarter2	Quarter3	Quarter4
1948	3.37	3.67	3.77	3.83
1949	4.67	5.87	6.70	6.97
1950	6.40	5.57	4.63	4.23
1951	3.50	3.10	3.17	3.37
1952	3.07	2.97	3.23	2.83
1953	2.70	2.57	2.73	3.70
1954	5.27	5.80	5.97	5.33
1955	4.73	4.40	4.10	4.23
1956	4.03	4.20	4.13	4.13
1957	3.93	4.10	4.23	4.93
1958	6.30	7.37	7.33	6.37
1959	5.83	5.10	5.27	5.60
1960	5.13	5.23	5.53	5.27
1961	6.80	7.00	6.77	6.20
1962	5.63	5.53	5.57	5.53
1963	5.77	5.73	5.50	5.57
1964	5.47	5.20	5.00	5.00
1965	4.90	4.67	4.37	4.10
1966	3.87	3.80	3.77	3.70
1967	3.77	3.83	3.83	3.93
1968	3.73	3.57	3.53	3.43
1969	3.37	3.43	3.60	3.60
1970	4.17	4.80	5.17	5.87
1971	5.93	5.97	5.97	5.97
1972	5.83	5.77	5.53	5.30

Source: "Business Statistics," Biennial edition, pp. 68 and 233, 1971. "Survey of Current Business," January 1972 and January 1973. Also see "Introduction to Statistical Time Series," 2nd edition, by W.A. Fuller, John Wiley & Sons, pp. 337, 1996.

Appendix B

S-plus Program Listings

B.1 LAE.reg — Fitting an LAE regression model

B.2 LSE.reg — Fitting an LSE regression model

B.3 LAE.std — Estimating standard errors in LAE

B.4 LAE.best — Selecting a suitable model in LAE

B.5 LSE.best — Selecting a suitable model in LSE

B.6 Subset — Producing the subset indices

B.7 LAE.tsar — Fitting an LAE autoregressive model

B.1 Program *LAE.reg*

```
#####################################################################
#                                                                   #
# Input:                                                            #
#                                                                   #
#   1. x : matrix of values of the explanatory variables.           #
#                                                                   #
#   2. y : vector of values of the response variable.               #
#                                                                   #
# Output:                                                           #
#                                                                   #
#   1. coefficients : estimates of the intercept and the            #
#                     regression coefficients.                      #
#                                                                   #
#   2. residuals : the realized residuals .                         #
#                                                                   #
#   3. minsae : the minimum sum of absolute residuals.              #
#                                                                   #
# Description:                                                      #
#                                                                   #
#       This program is a modification of the S-Plus function       #
#  "l1fit" which produces the minsae value in addition to the       #
#  coefficients and residuals.  It performs an LAE regression       #
#  of the response on the set of explanatory variables.  Note       #
#  that  the Fortran procedure based on  a specialized linear       #
#  programming  algorithm by  Barrodale & Roberts (1978)  has       #
#  been used in the program.                                        #
#                                                                   #
#####################################################################

function(x, y, intercept = T, print.it = T)
{
     xn <- if(is.matrix(x)) dimnames(x)[[2]] else "X"
     x <- as.matrix(x)
     vars <- ncol(x)
     if(length(xn) == 0)
         xn <- paste("X", 1:vars, sep = "")
     if(intercept)
     {
         vars <- vars + 1
         xn <- c("Intercept", xn)
     }
     obs <- nrow(x)
     if(obs != length(y))
         stop("x and y have different numbers of
```

```
            observations")
    if(obs <= vars)
        stop("More variables than observations")
    storage.mode(y) <- "single"
    ans <- .Fortran("l1fit",
            as.single(x),
            obs,
            obs,
            ncol(x),
            y,
            coefficients = single(vars),
            residuals = y,
            as.logical(intercept),
            minsae = single(1),
            rank = integer(1),
            code = integer(1))
    msg <- character(0)
    if(ans$code == 0)
    {
        msg <- "Non-unique solution possible"
        if(print.it)
            warning(msg)
    }
    else if(ans$code != 1)
        stop("Premature termination")
    if(ans$rank != vars)
    {
        msg <- c(msg, paste("Matrix not of full rank,
                apparent rank", ans$rank))
        if(print.it)
            warning(msg[length(msg)])
    }
    ans <- ans[c("coefficients", "residuals", "minsae")]
    names(ans$coefficients) <- xn
    if(length(msg))
        c(ans, message = list(msg))
    else ans
}

#################### End of program LAE.reg ####################
```

B.2 Program *LSE.reg*

```
#####################################################################
#                                                                   #
# Input:                                                            #
#                                                                   #
#   1. x : matrix of values of the explanatory variables.           #
#                                                                   #
#   2. y : vector of values of the response variable.               #
#                                                                   #
# Output:                                                           #
#                                                                   #
#   1. coefficients : estimates of the intercept and the            #
#                     regression coefficients.                      #
#                                                                   #
#   2. residuals : the realized residuals .                         #
#                                                                   #
#   3. minsse : the minimum sum of squared residuals.               #
#                                                                   #
# Description:                                                      #
#                                                                   #
#       This program is a modification of the S-Plus function       #
#  "lsfit" which produces the minsse value in addition to the       #
#  coefficients and residuals.  It performs an LAE regression       #
#  of the response on the set of explanatory variables.             #
#                                                                   #
#####################################################################

function(x, y, intercept = T)
{
     x  <- as.matrix(x)
     dy <- c(length(y), 1)
     dx <- dim(x)
     n  <- dx[1]
     dn <- dimnames(x)
     xn <- dn[[2]]
     if(length(xn) == 0)
         xn <- paste("X", 1:dx[2], sep = "")
     if(n != dy[1])
         stop("Number of observations in x and y
               not equal")
     if(intercept)
     {
         dx <- dx + c(0, 1)
         xn <- c("Intercept", xn)
         x <- array(c(rep(1, n), x), dx)
```

```
    }
    dimnames(x) <- list(dn[[1]], xn)
    p <- dx[2]
    q <- 1
    tolerance <- 1e-07
    z <- .Fortran("dqrls",
          qr = x,
          as.integer(dx),
          pivot = as.integer(1:p),
          qraux = double(p),
          y,
          as.integer(dy),
          coefficients = double(p * q),
          residuals = y,
          qt = y,
          tol = as.double(tolerance),
          double(2 * p),
          rank = as.integer(p))
    z <- z[match(c("coefficients", "residuals"),
                 names(z), 0)]
    z$minsse <- sum(z$residuals^2)
    ans <- z[c("coefficients", "residuals", "minsse")]
    names(ans$coefficients) <- xn
    return(ans)
}

#################### End of program LSE.reg ####################
```

B.3 Program *LAE.std*

```
#####################################################################
#                                                                   #
# Input:                                                            #
#                                                                   #
#   1. x : matrix of values of the explanatory variables.           #
#                                                                   #
#   2. y : vector of values of the response variable.               #
#                                                                   #
# Output:                                                           #
#                                                                   #
#   1. coefficients  : estimates of the intercept and the           #
#                      regression coefficients.                     #
#                                                                   #
#   2. sterror.coef  : asymptotic standard errors of the            #
#                      estimated coefficients.                      #
#                                                                   #
#   3. lcl.coeff     : lower limit of 95% confidence                #
#                      interval for the coefficients.               #
#                                                                   #
#   4. ucl.coeff     : upper limit of 95% confidence                #
#                      interval for the coefficients.               #
#                                                                   #
#   5. fitted.values : the fitted values.                           #
#                                                                   #
#   6. sterror.fits  : asymptotic standard errors of the            #
#                      fitted values.                               #
#                                                                   #
#   7. lcl.fits      : lower limit of 95% confidence                #
#                      interval for the fitted values.              #
#                                                                   #
#   8. ucl.fits      : lower limit of 95% confidence                #
#                      interval for the fitted values.              #
#                                                                   #
# Description:                                                      #
#                                                                   #
#      By using  the program LAE.reg and  the discussion in         #
#   section 3.4, this program produces the coefficients and         #
#   the fitted values,  and their corresponding  asymptotic         #
#   standard errors and 95% confidence limits.                      #
#                                                                   #
#####################################################################
```

```
function(x, y)
{
     laeout <- lae(x, y)
     coeff  <- laeout$coefficients
     minsae <- laeout$minsae
     x <- as.matrix(x)
     alpha <- minsae/(length(y) - ncol(x) - 1)
     xaug  <- cbind(1, x)
     sterr <- sqrt(diag(solve(t(xaug) %*% xaug)) * alpha^2)
     lcl <- laeout$coefficients - 1.96 * sterr
     ucl <- laeout$coefficients + 1.96 * sterr
     yfits <- as.vector(xaug %*% coeff)
     hatdiag <- diag(xaug %*% solve(t(xaug) %*% xaug)
                     %*% t(xaug))
     seyfit  <- sqrt(hatdiag * alpha^2)
     lclfits <- yfits - 1.96 * seyfit
     uclfits <- yfits + 1.96 * seyfit
     summary <- list(coeff, sterr, lcl, ucl, yfits, seyfit,
                     lclfits, uclfits)
     names(summary) <- c("coefficients", "sterror.coef",
                         "lcl.coef", "ucl.coef", "fitted.values",
                         "sterror.fits", "lcl.fits","ucl.fits")
     names(summary$coefficients) <- names(coeff)
     names(summary$sterror.coef) <- names(coeff)
     return(summary)
}

#################### End of program LAE.std ####################
```

B.4 Program *LAE.best*

```
#####################################################################
#                                                                   #
# Input:                                                            #
#                                                                   #
#   1. x : matrix of values of the explanatory variables.           #
#                                                                   #
#   2. y : vector of values of the response variable.               #
#                                                                   #
# Output:                                                           #
#                                                                   #
#      A table of best models for each subset size with             #
#   the MinSAE, AIC and SBC values.                                 #
#                                                                   #
# Description:                                                      #
#                                                                   #
#      Based on the sum of absolute errors,  this program           #
#   determines  the best model for  each subset size  and           #
#   gives the corresponding values of MinSAE, AIC and SBC.          #
#                                                                   #
#####################################################################

function(X, y)
{
     nobs <- nrow(X)
     nvar <- ncol(X)
     nmax <- min(nvar, 9)
     min  <- vector("numeric", nmax)
     aic  <- vector("numeric", nmax)
     sbc  <- vector("numeric", nmax)
     ind  <- vector("character", nmax)
     for(size in 1:nmax)
     {
         set <- subset(nvar, size)
         minsae <- vector(mode = "numeric", nrow(set))
         for(i in 1:nrow(set))
             minsae[i] <- lae(X[, c(set[i,  ])], y)$minsae
         ord <- order(minsae)
         minsae <- minsae[ord]
         setindex  <- apply(set, 1, paste, collapse = " ")
         min[size] <- minsae[1]
         aic[size] <- 2 * nobs * log(min[size]/nobs)
                      + 2 * (size + 1)
         sbc[size] <- 2 * nobs * log(min[size]/nobs)
                      + (size + 1) * log(nobs)
```

```
        setindex  <- setindex[ord]
        ind[size] <- setindex[1]
    }
    best <- data.frame(Inputs = ind, MinSAE = min,
                       AIC = aic, SBC = sbc)
    return(best)
}

##################  End of program LAE.best  ###################
```

B.5 Program *LSE.best*

```
#####################################################################
#                                                                   #
# Input:                                                            #
#                                                                   #
#   1. x : matrix of values of the explanatory variables.           #
#                                                                   #
#   2. y : vector of values of the response variable.               #
#                                                                   #
# Output:                                                           #
#                                                                   #
#      A table of best models for each subset size with             #
#   the MinSSE, AIC and SBC values.                                 #
#                                                                   #
# Description:                                                      #
#                                                                   #
#      Based on the sum of  squared errors,  this program           #
#   determines  the best model for  each subset size  and           #
#   gives the corresponding values of MinSSE, AIC and SBC.          #
#                                                                   #
#####################################################################

function(X, y)
{
     nobs <- nrow(X)
     nvar <- ncol(X)
     min  <- vector("numeric", nvar)
     aic  <- vector("numeric", nvar)
     sbc  <- vector("numeric", nvar)
     ind  <- vector("character", nvar)
     for(size in 1:nvar)
     {
         set <- subset(nvar, size)
         minsse <- vector(mode = "numeric", nrow(set))
         for(i in 1:nrow(set))
             minsse[i] <- lse(X[, c(set[i,  ])], y)$minsse
         ord <- order(minsse)
         minsse <- minsse[ord]
         setindex  <- apply(set, 1, paste, collapse = " ")
         min[size] <- minsse[1]
         aic[size] <- nobs * log(min[size]/nobs) +
                      2 * (size + 1)
         sbc[size] <- nobs * log(min[size]/nobs) +
                      (size + 1) * log(nobs)
         setindex  <- setindex[ord]
```

```
        ind[size] <- setindex[1]
    }
    best <- data.frame(Inputs = ind, MinSSE = min,
                       AIC = aic, SBC = sbc)
    return(best)
}

################## End of program LSE.best ##################
```

B.6 Program *Subset*

```
#####################################################################
#                                                                   #
# Input:                                                            #
#                                                                   #
#   1. n : a positive integer.                                      #
#                                                                   #
#   2. r : a positive integer between 1 and n.                      #
#                                                                   #
# Output:                                                           #
#                                                                   #
#   A matrix of indices for all subsets of size r.                  #
#                                                                   #
# Description:                                                      #
#                                                                   #
#      This program is a procedure called in  the progrms           #
#   LAE.best  and  LSE.best.  It generates  the matrix of           #
#   indices for all subsets of size r from a given number           #
#   n of explanatory variables.                                     #
#                                                                   #
#####################################################################

function(n, r)
{
    curvect   <- 1:r
    endvect   <- curvect + (n - r)
    indmatrix <- matrix(1:r, 1, r)
    while(curvect[1] != endvect[1])
    {   for(i in r:1)
        {   if((ii <- curvect[i]) < endvect[i])
            {   for(j in i:r)
                    curvect[j] <- ii + j - i + 1
                    break
            }
            else curvect[i] <- i
        }
        indmatrix <- rbind(indmatrix, indvect <- curvect)
    }
    return(indmatrix)
}


###################  End of program Subset  ####################
```

B.7 Program *LAE.tsar*

```
#####################################################################
#                                                                   #
# Input:                                                            #
#                                                                   #
#   1. x : vector of values of time series process.                 #
#                                                                   #
#   2. k : a positive integer represents the order of               #
#          the desired autoregressive model.                        #
#                                                                   #
# Output:                                                           #
#                                                                   #
#   1. coefficients : estimates of the intercept and the            #
#                     regression coefficients.                      #
#                                                                   #
#   2. residuals : the realized residuals .                         #
#                                                                   #
#   3. minsae : the minimum sum of absolute residuals.              #
#                                                                   #
# Description:                                                      #
#                                                                   #
#     This program fits an autoregressive model of order k          #
#   by transforming the time series data into the required          #
#   form and  then using the program LAE.reg.  The results          #
#   are the same as those in LAE.reg.                               #
#                                                                   #
#####################################################################

function(x, k)
{
        y <- x[(k + 1):length(x)]
        xlag <- matrix(0, (length(x) - k), k)
        for(i in (1:k))
        {
            xlag[, i] <- x[(k - i + 1):(length(x) - i)]
        }
        ans <- lae(xlag, y)
        ans
}


##################  End of program LAE.tsar  ##################
```

VITA

Surname: TUNG Given Names: YI-YUAN

Place of Birth: TAIWAN, R.O.C.

Educational Institutions Attended:

University of Victoria	1994 to 1998
National Chung-Hsing University	1987 to 1990

Degree Awarded:

B.Sc.	National Chung-Hsing University	1990

Honours and Awards:

University of Victoria Fellowship	1996-1997

PARTIAL COPYRIGHT LICENSE

Title of Thesis: **LEAST ABSOLUTE ERROR REGRESSION**

Author 

September 17, 1998