Research Article / Araştırma Makalesi
AN ALTERNATIVE APPROACH TO SOLVE THE LAD-LASSO PROBLEM

Esra EMİROĞLU*

Istanbul University, Faculty of Science, Department of Mathematics, Beyazit-ISTANBUL

ABSTRACT

The least absolute deviation (LAD) regression is more robust alternative to the popular least squares (LS) regression whenever there are outliers in the response variable, or the errors follow a heavy-tailed distribution. The least absolute shrinkage and selection operator (LASSO) is a popular choice for shrinkage estimation and variable selection. By combining these two classical ideas, LAD-LASSO is an estimator which is able to perform shrinkage estimation while at the same time selecting the variables and is resistant to heavy-tailed distributions and outliers. The aim of this article is to reformulate LAD-LASSO problem to solve with the Simplex Algorithm, which is an area of Mathematical Programming.
Keywords: Regression, LAD, LASSO, LAD-LASSO, Mathematical Programming, Simplex Algorithm.

OPTİMİZASYON UYGULAMASI OLARAK REGRESYON PARAMETRELERİNİN TAHMİNİ

ÖZ

Yanıt değişkeninde aykırı değerler olduğunda ya da hatalar uzun-kuyruklu dağılımda olduğunda En Küçük Mutlak Sapma (Least Absolute Deviation, LAD) regresyonu popular En Küçük Kareler (Least Squares, LS) regresyonuna güçlü bir alternatifdir. En Küçük Mutlak Daralma ve Seçim Operatoru (Least Absolute Shrinkage and Selection Operator, LASSO) değişken seçimi ve parametre tahmini için populer bir seçimdir. Bu iki klasik yöntemin birleştirilmesiyle elde edilen En Küçük Mutlak Sapma ve En Küçük Mutlak Daralma ve Seçim Operatörü (Least Absolute Deviation and Least Absolute Shrinkage and Selection Operator, LAD-LASSO) değişken seçimini ve parametre tahminini aynı anda yapabilen, uzun kuyruklu dağılımlara ve aykırı değerlere dirençli olan bir tahmin edicidir. Bu tezin amacı, LAD-LASSO problemini yeniden formüle etmek ve yeniden formüle ettiğimiz LAD-LASSO problemini Matematik Programlamanın çözüm yöntemlerinden Simpleks Yöntem ile çözmektir.
Anahtar Sözcükler: LAD, LASSO, LAD-LASSO, matematik programlama simpleks yöntem.

1. INTRODUCTION

Regression analysis is a statistical process for estimating the relationships among variables in many engineering applications. It includes many techniques for modelling and analyzing several

---

* Corresponding Author/Sorumlu Yazar: e-mail/e-ileti: eertan@istanbul.edu.tr, tel: (212) 455 57 00 / 15340

variables, when the focus is on the relationship between a dependent variable and one or more independent variables. Let us consider the linear regression model which is described as follows;

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1.1}$$

where $\mathbf{y}$ is an $n \times 1$ vector of the observations, $\mathbf{X}$ is an $n \times p$ matrix of the levels of the

regressor variables, $\boldsymbol{\beta} = \left(\beta_0, \beta_1, ..., \beta_{p-1} \right)'$ is a $p \times 1$ vector of the unknown coefficients,

and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of the random errors satisfying $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $V(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$.

In regression analysis, the most important aim is the estimation of the unknown parameters. The most popular method is the Least Squares (LS) method. The LS estimator is a solution to the problem

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{p-1} x_{ij}\beta_j \right)^2. \tag{1.2}$$

According to the Gauss-Markov theorem, the LS estimator is the best linear unbiased estimators of $\boldsymbol{\beta}$. When the error ε_i are normally distributed, the LS estimator is a good parameter estimation procedure in the sense that it produces an estimator of the parameter vector $\boldsymbol{\beta}$ that has good statistical properties (Montgomery *et al.*, 2001). On the other hand, there are many situations where the distribution of the errors is nonnormal. In case of nonnormal distributions, particularly heavy-tailed distributions, the LS estimator no longer has these desirable properties. These heavy-tailed distributions tend to generate outliers, and these outliers may have an improper effect on the LS estimates (Montgomery *et al.*, 2001).

Another important problem of regression analysis is multicollinearity. When there are near-linear dependencies among the regressors, the problem of multicollinearity occurs. For multicollinearity, several alternative estimation techniques are proposed, but Ridge regression estimator, proposed by Hoerl and Kennard (1970), is one of the most widely used estimators.

The Ridge regression estimator $\hat{\boldsymbol{\beta}}_R$ is a solution to the problem

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{p-1} x_{ij}\beta_j \right)^2$$
$$\sum_{j=0}^{p-1} \left(\beta_j \right)^2 \leq s \tag{1.3}$$

which makes explicit the size constraint on the parameters. Another point of view, ridge coefficients minimize a penalized residual sum of squares,

$$\hat{\boldsymbol{\beta}}_R = \min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{p-1} x_{ij}\beta_j \right)^2 + k \sum_{j=0}^{p-1} \left(\beta_j \right)^2 \right\} \tag{1.4}$$

where $k \geq 0$ is a complexity parameter that controls the amount of shrinkage: the larger the value of *k*, the greater the amount of shrinkage. There is a one-to-one correspondence between

the parameters k in (1.4) and s in (1.3) (Friedman *et al.*, 2001). Writing the criterion in (1.4) in matrix form, the ridge regression solutions are easily seen to be

$$\hat{\boldsymbol{\beta}}_R = \left(\mathbf{X'X} + k\mathbf{I} \right)^{-1} \mathbf{X'y}, \quad k \geq 0 \tag{1.5}$$

where $\mathbf{I}$ is the $p \times p$ identity matrix. Note that when $k = 0$, the ridge estimator is the LS estimator (Montgomery *et al.*, 2001).

When there are outliers, robust regression methods, which are more powerful than the LS method, are recommended as alternatives (Huber, 1981). One of these robust estimation methods is the Least Absolute Deviation (LAD) method, where the regression parameters are estimated through the minimizing of the sum of absolute value of the errors as follows

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left| y_i - \sum_{j=0}^{p-1} x_{ij} \beta_j \right|. \tag{1.6}$$

In applications, one can frequently face with *x*-space and/or *y*-space outliers in the data sets. It is known that the LS estimator is unsuccessful in producing a reliable result under these circumstances, and that the LAD estimator is better in the case of *y*-space outliers (Arslan, 2011). However, there are some difficulties present in the calculations as the number of regressor increases.

Variable selection is another important subject in regression analysis. A large number of regressors are usually introduced at the initial stage of the regression model to decrease possible modelling biases. However, including unnecessary regressors can reduce the efficiency of the resulting estimation procedure and yields less accurate predictions. On the other hand, omitting important regressors may produce biased parameter estimates and prediction results. Therefore, selecting the significant regressors is an important task of regression analysis.

The problem of selecting a model under suitable conditions for the remainder is studied extensively in the literature. Some of the recommended and best applied methods are the Akaike Information Criterion (AIC) (Akaike 1973), the Bayes Information Criterion (BIC) (Schwarz 1978), and the Mollows-Cp statistic. Theoretically speaking there is no confirmed knowledge as to which criterion will be better (Shi and Tsai, 2002).

In order to eliminate this insufficiency, Tibshirani (1996) proposed the following the Least Absolute Shrinkage and Selection Operator (LASSO) which is minimized the penalized LS regression

$$\sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{p-1} x_{ij} \beta_j \right)^2 + n\lambda \sum_{j=0}^{p-1} \left| \beta_j \right| \tag{1.7}$$

where $\lambda > 0$ is the tuning parameter. The LASSO can effectively select significant regressors and estimate the regression parameters simultaneously (Tibshirani, 1996).

Minimizing criterion in (1.7) is equal to

$$\min_{\boldsymbol{\beta}} \quad \sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{p-1} x_{ij} \beta_j \right)^2 \tag{1.8}$$

$$subject\ to \quad \sum_{j=0}^{p-1} \left| \beta_j \right| \leq s$$

where $s \geq 0$ is tuning parameter selected by the analyst.

The finite-dimensional performance of the LASSO estimator under standard errors was shown by Tibshirani (1996) and its statistical properties were studied by Knight and Fu (2000), Fan and Li (2001), Rosset and Zhu (2004) and Zhau and Yu (2006).

However, when errors in (1.1) are distributed in a heavy-tailed manner, the performance of the LASSO which is a particular case of penalized LS regression becomes weaker due to its sensitivity to the heavy-tailed error distributions and outliers. Due to this sensitivity, the LAD regression which is resistant to outliers and heavy-tailed errors is combined with the LASSO.

The obtained LAD-LASSO is successful in simultaneously estimating robust regression and selecting variables, and therefore used as an alternative to the LASSO. When the LAD and the LAD-LASSO are compared, the LAD-LASSO is seen to be able to perform parameter estimation while at the same time for selecting the model. Also the LAD-LASSO is resistant to heavy-tailed and outliers than the LASSO. The aim of this article is to reformulate LAD-LASSO and solve the reformulated LAD-LASSO with the Simplex algorithm.

The rest of the article is organized as follows. In Section 2, we introduce the reformulated LAD-LASSO. Also, it has been shown that LAD-LASSO is a Mathematical Programming problem. A real data example is given in Section 3. The paper is finalized with a discussion section.

2. THE LAD-LASSO

For simultaneous parameter estimation and variable selection, the LAD-LASSO is obtained by minimizing the penalized LAD regression criterion as follows

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left| y_i - \sum_{j=0}^{p-1} x_{ij}\beta_j \right| + n\lambda \sum_{j=0}^{p-1} \left| \beta_j \right| \tag{2.1}$$

where $\lambda > 0$ is the tuning parameter. In studies of Wang, Li and Jiang (2007), the parametres are estimated by minimizing the following objective function

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left| y_i - \sum_{j=0}^{p-1} x_{ij}\beta_j \right| + n\sum_{j=0}^{p-1} \lambda_j \left| \beta_j \right| \tag{2.2}$$

by using the different tuning parameters for different regression coefficients. Without loss of generality, Wang, Li and Jiang (2007) assumed that $\beta_0 = 0$.

They considered an augmented dataset $\left\{ \left(y_i^*, \mathbf{x}_i^* \right) \right\}$ with $i = 1, 2, ..., n + p - 1$, where $\left(y_i^*, \mathbf{x}_i^* \right) = \left(y_i, \mathbf{x}_i \right)$ for $1 \leq i \leq n$, $\left(y_{n+j}^*, \mathbf{x}_{n+j}^* \right) = \left(0, n\lambda_j \mathbf{e}_j \right)$ for $1 \leq j \leq p - 1$, and $\mathbf{e}_j$ is a $\left(p - 1 \right) -$ dimensional vector with the jth component equal to 1 and all others equal to 0. It can be verified that

$$\text{LAD-LASSO} = \sum_{i=1}^{n+p-1} \left| y_i^* - \mathbf{x}_i^{*\prime} \boldsymbol{\beta} \right|. \tag{2.3}$$

This is just a traditional LAD criterion, obtained by treating $\left(y_i^*, \mathbf{x}_i^* \right)$ as if they were the true data. Consequently, any standard unpenalized LAD program (rq in the QUANTREG package of R) can be used to find the LAD-LASSO estimator.

In our study, we find that the LAD-LASSO estimator of $\boldsymbol{\beta}$ is obtained by

$$\min_{\boldsymbol{\beta}} \quad \sum_{i=1}^{n} |d_i|$$

$$subject\ to \quad \sum_{j=0}^{p-1} |\beta_j| \leq t \tag{2.4}$$

$$\mathbf{d}, \boldsymbol{\beta} \quad unrestricted\ in\ sign$$

where $t \geq 0$ is tuning parameter and d_i is defined as $d_i = y_i - \sum_{j=0}^{p-1} x_{ij} \beta_j$. Minimizing (2.4) is equal to

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left| y_i - \sum_{j=0}^{p-1} x_{ij} \beta_j \right| + \lambda \sum_{j=0}^{p-1} |\beta_j| \tag{2.5}$$

For estimation of β_j parameter in problem (2.5), LAD-LASSO is reformulated as follows

$$\min_{\boldsymbol{\beta}} \quad \sum_{i=1}^{n} |d_i| + \lambda \sum_{j=0}^{p-1} |\beta_j|$$

$$subject\ to \quad \mathbf{X}\boldsymbol{\beta} + \mathbf{d} = \mathbf{y} \tag{2.6}$$

$$\mathbf{d}, \boldsymbol{\beta} \quad unrestricted\ in\ sign$$

Minimizing (2.6) is equal to minimizing

$$\min_{\boldsymbol{\beta}} \quad \sum_{i=1}^{n} |d_i|$$

$$subject\ to \quad \mathbf{X}\boldsymbol{\beta} + \mathbf{d} = \mathbf{y}$$

$$\sum_{j=0}^{p-1} |\beta_j| \leq t \tag{2.7}$$

$$\mathbf{d}, \boldsymbol{\beta} \quad unrestricted\ in\ sign$$

Note that $|d_i| = d_{1i} + d_{2i}$ and $d_i = d_{1i} - d_{2i}$ where d_{1i} and d_{2i} are nonnegative and $|\beta_j| = \beta_{1j} + \beta_{2j}$ and $\beta_j = \beta_{1j} - \beta_{2j}$ where β_{1j} and β_{2j} are nonnegative. We can reformulate the problem as:

$$\min_{\boldsymbol{\beta}} \qquad \sum_{i=1}^{n} d_{1i} + \sum_{i=1}^{n} d_{2i}$$

$$subject \ to \ \mathbf{X}\boldsymbol{\beta}_1 - \mathbf{X}\boldsymbol{\beta}_2 + \mathbf{d}_1 - \mathbf{d}_2 = \mathbf{y}$$

$$\sum_{j=0}^{p-1} \beta_{1j} + \sum_{j=0}^{p-1} \beta_{2j} \le t \qquad\qquad (2.8)$$

$$\mathbf{d}_1, \mathbf{d}_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \ge \mathbf{0}$$

Definition: Any $\left(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{d}_1, \mathbf{d}_2\right)$ satisfying $\mathbf{X}\boldsymbol{\beta}_1 - \mathbf{X}\boldsymbol{\beta}_2 + \mathbf{d}_1 - \mathbf{d}_2 = \mathbf{y}$ is called a solution to (2.8).

Let $\begin{pmatrix} \mathbf{X}_{n\times p} & -\mathbf{X}_{n\times p} & \mathbf{I}_{n\times n} & -\mathbf{I}_{n\times n} \\ \mathbf{1}'_{1\times p} & \mathbf{1}'_{1\times p} & \mathbf{0}_{1\times n} & \mathbf{0}_{1\times n} \end{pmatrix}$ be denoted by the matrix $\mathbf{A}$ of order

$n+1 \times 2p+2n$, $\left(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{d}_1, \mathbf{d}_2\right)$ be denoted by the vector $\mathbf{W}'$ of order $1 \times 2p+2n$

and $\begin{pmatrix} \mathbf{y}_{n\times 1} \\ t_{1\times 1} \end{pmatrix}$ be denoted by the vector $\mathbf{P}$ of order $n+1 \times 1$. Any $\mathbf{W}$ satisfying

$\mathbf{AW} \begin{Bmatrix} = \\ \le \end{Bmatrix} \mathbf{P}$ is a solution to (2.8). Let $\mathbf{C}'$ be the vector $\left(\mathbf{0}_{1\times p}, \mathbf{0}_{1\times p}, \mathbf{1}'_{1\times n}, \mathbf{1}'_{1\times n}\right)$ where

$\mathbf{0} = (0, 0, \dots 0)$ and $\mathbf{1}' = (1, 1, \dots 1)$. Then $\mathbf{C}'\mathbf{W}$ is called the objective function of problem

(2.8). Any solution $\mathbf{W}$ to (2.8), if it further satisfies $W_j \ge 0$, $j = 1, 2, \dots, 2p+2n$, we call it a feasible solution to problem.

Thus, LAD-LASSO is translated into a mathematical programming problem and can be solved with Simplex Algorithm, which is given by Arthanari and Dodge (1993).

One of the differences between (2.3) and (2.8) is that the formulation in (2.3) includes the tuning parameters in augmented $\mathbf{X}$ matrix. But in this study, the tuning parameter is in augmented observation vector $\mathbf{y}$. In additon to, if we take $\lambda_j = \lambda$ in (2.3), dimension of the augmented $\mathbf{X}$ matrix is larger than the $\mathbf{X}$ matrix in (2.8). Therefore, the solution of (2.8) is easier than the problem in (2.3). Another difference is that the formulation in (2.8) determines a suitable range to do variable selection for tuning parameter t. Because t is chosen larger than

$\sum_{j=0}^{p-1} \left| \hat{\beta}_j^{LAD} \right|$, the LAD-LASSO estimates are equal to $\hat{\beta}_j^{LAD}$ in formulation in (2.8).

3. EXAMPLE

To illustrate parameter estimation by using LAD-LASSO, we consider Hald data, which is used widely in literature. Hald (1952) present data concerning the heat evolved in calories in calories per gram of cement $\left(\mathbf{y}\right)$ as a function of the amount of each of four ingredient in the

mix: tricalcium aluminate $(\mathbf{x}_1)$, tricalcium silicate $(\mathbf{x}_2)$, tetracalcium alumino ferrite $(\mathbf{x}_3)$, and dicalcium silicate $(\mathbf{x}_4)$. The data are shown in Table 3.1. This data has outliers and serious multicollinearity. VIF values of explanatory variables are given as follows:

$$VIF_1 = 38.496$$
$$VIF_2 = 254.423$$
$$VIF_3 = 46.868$$
$$VIF_4 = 282.513$$

We will use these data to illustrate the all-possible-regressions approach to variable selection.

Table 3.1. Hald Cement Data

Observation i	y_i	x_{i1}	x_{i2}	x_{i3}	x_{i4}
1	78.5	7	26	6	60
2	74.3	1	29	15	52
3	104.3	11	56	8	20
4	87.6	11	31	8	47
5	95.9	7	52	6	33
6	109.2	11	55	9	22
7	102.7	3	71	17	6
8	72.5	1	31	22	44
9	93.1	2	54	18	22
10	115.9	21	47	4	26
11	83.8	1	40	23	34
12	113.3	11	66	9	12
13	109.4	10	68	8	12

It is instructive to examine the pairwise correlations between $\mathbf{x}_i$ and $\mathbf{x}_j$ and between $\mathbf{x}_i$ and $\mathbf{y}$. These simple correlations are shown in Table 3.2. Note that the pairs of explanatory variables $(\mathbf{x}_1, \mathbf{x}_3)$ and $(\mathbf{x}_2, \mathbf{x}_4)$ are higly correlated since $r_{13} = -0.824$ and $r_{24} = -0.973$.

Table 3.2. Matrix of Simple Correlations for Hald's Data

	x_1	x_2	x_3	x_4	y
x_1	1.0				
x_2	0.229	1			
x_3	-0.824	-0.139	1		
x_4	-0.245	-0.973	0.030	1	
y	0.731	0.816	-0.535	-0.821	1

On the other hand, statistics for detecting outliers for the Hald cement data set is given in Table 3.3. Based on the result of Table 3.3, e_6, e_8 and e_{13} residual seems suspiciously large. Therefore, we can say that Hald data has *y*-direction outliers. On the other hand, according to the leverage (h_{ii}), Cook's distance and DFITS values, it seems that there is no *x*-direction outliers in Table 3.3. In this situation, LAD regression is much more powerful estimation method than LS regression.

Table 3.3. Statistics for detecting influential observations for the Hald cement data

ID	y	$\hat{y}$	e_i	h_{ii}	Cook's Distance	DFITS
1	78.5	78.495	0.005	0.473	0	0.006
2	74.3	72.789	1.511	0.256	0.057	0.755
3	104.3	105.971	-1.671	0.500	0.301	-2.279
4	87.6	89.327	-1.727	0.218	0.059	-0.724
5	95.9	95.649	0.251	0.281	0.002	0.140
6	109.2	105.275	3.925	0.047	0.083	0.556
7	102.7	104.149	-1.449	0.290	0.064	-0.840
8	72.5	75.675	-3.175	0.332	0.394	-2.193
9	93.1	91.722	1.378	0.217	0.038	0.575
10	115.9	115.619	0.282	0.623	0.021	0.658
11	83.8	81.809	1.991	0.349	0.171	1.475
12	113.3	112.327	0.973	0.186	0.015	0.347
13	109.4	111.694	-2.294	0.227	0.110	-1

In Table 3.4, the parameter estimates based on LS and LAD estimates are given. In Table 3.5-3.7, the parameter estimates based on Ridge, LASSO and LAD-LASSO are given, respectively.

Table 3.4. Estimates of the Hald coefficients under different estimations methods

Method	β_0	β_1	β_2	β_3	β_4
LS	62.41	1.55	0.51	0.11	-0.14
LAD	-13.337	2.354	1.28	1.007	0.601

Table 3.5. Ridge estimates at Various Values of k

k	0	0.001	1.61	2.16	3.058	4.690	4.691	10	18.579
β_0	62.40	34.291	0.09	0.082	0.072	0.0623	0.063	0.053	0.049
β_1	1.55	1.840	2.170	2.163	2.151	2.130	2.130	2.066	1.978
β_2	0.510	0.800	1.157	1.159	1.162	1.167	1.167	1.182	1.202
β_3	0.102	0.398	0.741	0.737	0.728	0.713	0.0713	0.668	0.607
β_4	-0.144	0.140	0.489	0.490	0.492	0.495	0.495	0.505	0.518

Table 3.6. LASSO estimates at Various Values of S

S	0	0.001	1.61	2.16	3.058	4.690	4.691	10	18.579	64.424
β_0	0	0	0	0	0	0.10	0.10	5.64	13.99	62.41
β_1	0	0	0	0	1.01	2.19	2.19	2.13	2.05	1.55
β_2	0	0	1.58	1.55	1.42	1.15	1.15	1.10	1.01	0.51
β_3	0	0	0	0	0	0.76	0.76	0.70	0.61	0.11
β_4	0	0	0.03	0.61	0.06	0.44	0.44	0.43	0.35	0.14
MS_{Res}	9314.	9314.	637.08	108.81	29.53	6.60	6.60	6.56	6.40	6.07

In Table 3.6, LASSO estimates based on various s values are given. According to the result which are obtained in Table 3.6, variable selection is done between 0 and 4.690. In this range, a model which has less parameter, is obtained for a suitable s value. Therefore; the obtained model is less affected from multicollinearity. After this point which variables selection stops, the LASSO estimates continues to be approach to LS estimates. Notice that if S is chosen larger than $\sum_{j=0}^{p-1}\left|\hat{\beta}_j^{LS}\right|$, the LASSO estimates are equal to $\hat{\beta}_j^{LS}$. On the other hand, until 4.690, while S increases, MS_{Res} decreases. Therefore the best point S is previous point from 4.690.

Table 3.7. LAD-LASSO estimates at Various Values of t

t	0	0.001	1.61	2.16	3.058	4.690	4.691	10	18.579	64.424
β_0	0	0	0	0	0	0	-0.0003	-5.098	-13.337	-13.337
β_1	0	0	0	0.008	1.008	2.213	2.213	2.267	2.354	2.354
β_2	0	0.001	1.609	1.491	1.437	1.145	1.145	1.196	1.280	1.280
β_3	0	0	0	0	0.0003	0.865	0.865	0.920	1.007	1.007
β_4	0	0	0.001	0.661	0.612	0.468	0.468	0.518	0.601	0.601
MS_{Res}	9314.	10080.	638.75	126.54	33.63	6.68	7.51	7.51	7.64	7.64

In Table 3.7, LAD-LASSO estimates based on various t values are given. According to the result which are obtained in Table 3.7, variable selection is done between 0 and 4.691. In this range, a model which has less parameter, is obtained for a suitable t value. Therefore the obtained model is less affected from multicollinearity and outliers. After this point which variables selection stops, the LAD-LASSO estimates continues to be approach to LAD estimates. Notice that if t is chosen larger than $\sum_{j=0}^{p-1}\left|\hat{\beta}_j^{LAD}\right|$, the LAD-LASSO estimates are equal to $\hat{\beta}_j^{LAD}$. On

the other hand, until 4.691, while t increases, $MS_{\mathrm{Re}s}$ decreases. Therefore the best point t is previous point from 4.691.

4. DISCUSSION

In this study, the tuning parameter is in augmented observations vector in our approach but in study of Wang, Li and Jiang (2007) the different tuning parameters are in augmented regressor variables matrix for different regressor coefficients. Therefore the dimension of matrix is larger and using Simplex Algoritm is more difficult than ours. The other difference is the range of tunig parameter is known in our approach. Finally based on the analysis result of Hald Data, by using the reformulated LAD LASSO, it is shown that a regression model, which is less affected from multicollinearity and outliers, can be obtained for suitable t value. Also, we can obtain a sparse model in a suitable range, which does variable selection, for the tuning parameter t.

REFERENCES / KAYNAKLAR

[1]   Akaike, H. (1973) Information Theory and an Estimation of the Maximum Likelihood Principle. In 2nd International Symposium on Information Theory, eds. B. N. Petrov and F.Csaki, Budapest:Akademia Kiado, pp. 267-281.

[2]   Arslan, O. (2011) Weighted LAD-LASSO Method for Robust Parameter Estimation and Variable Selection in Regression. Computational Statistics and Data Analysis 56, 1952-1965.

[3]   Arthanari, T. S., Dodge, Y., (1993) Mathematical Programming in Statistics, John Wiley&Sons Inc., New York, USA.

[4]   Fan and Li, (2001), Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties, Journal of the American Statistical Association, 96, 1348-1360.

[5]   Friedman, J., Hastie, T., Tibshirani, R. (2001) The elements of statistical learning. New York: Springer Series in Statistics, 2001.

[6]   Hoerl, A. E. and Kennard, R. V. (1970) Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55-67.

[7]   Huber, P. J. (1981) Robust Statistics. John Wiley & Sons Inc., New York, USA.

[8]   Knight, Fu, (2000), Asymptotic for Lasso-Type Estimators, The Annals of Statistics, 28, 1356-1378.

[9]   Montgomery, D. C., Peck, E. A., Vinig, G. G. (2001) Introduction to Linear Regression Analysis, 3th. Ed. John Wiley & Sons, Inc., USA.

[10]   Robert Tibshirani, (1996), Regression Shrinkage and Selection via the Lasso, Journal of the Royal Statistical Society, Series B Vol. 58, No. 1, pp 267-288.

[11]   Rosset, S., Zhu, J. (2004), Least Angle Regression: discussion. The Annals of Statistics 32, 469-475.

[12]   Schwarz, G. (1978), Estimating the Dimension of a Model. The Annals of Statistics, 6, 461-464.

[13]   Shi, P. and Tsai, C. L. (2002), Regression Model Selection a Residual Likelihood Approach. Journal of the Royal Statistical Society: Series B 64.2: 237-252.

[14]   Tibshirani, 1996, Regression Shrinkage and Selection via the LASSO, Journal of the Royal Statistical Society, Ser.B, 58, 923-941.

[15]   Wang, H., Li, G. and Jiang, G. (2007). Robust Regression Shrinkage and Consistent Variable Selection Through The LAD_LASSO. Journal of Business & Economic Statistics 25, 347-355.

[16]   Zhao, P. and Yu, B. (2006), On Model Selection Consistency Of Lasso. J. Mach. Learn. Res, 7, 2541-2563.