Institutional repository of
Jönköping University

<http://www.publ.hj.se/diva>

This is an Author's Original Manuscript of an article whose final and definitive form, the Version of Record, has been published in the Journal of Applied Statistics 17 Sep 2012 copyright Taylor & Francis, available online at:

<http://www.tandfonline.com/doi/abs/10.1080/02664763.2012.724663>.

Access to the published version may require subscription.

Citation for the published paper:

Zeebari, Z. (2012). Developing ridge estimation method for median regression.
Journal of Applied Statistics, 39(12), 2627-2638

Published with permission from: Taylor & Francis

Developing Ridge Estimation Method for Median Regression

By

Zeebari, Zangin

Department of Economics and Statistics,

Jönköping International Business School, Sweden

P.O Box 1026, SE 55111, Jönköping, Sweden

zangin.zeebari@jibs.hj.se

ABSTRACT

In this paper the ridge estimation method is generalized to the median regression. Though the Least Absolute Deviations (LAD) estimation method is robust in the presence of non-Gaussian or asymmetric error terms, it can still deteriorate into a severe multicollinearity problem when non-orthogonal explanatory variables are involved. The proposed method increases the efficiency of the LAD estimators by reducing the variance inflation and giving more room for the bias to get a smaller Mean Squared Error (MSE) of the LAD estimators. The paper includes an application of the new methodology and a simulation study, as well.

JEL Classification: C15, C18 , C21

Key-words: Lest Absolute Deviations, Shrinkage, Multicollinearity, Efficiency, Robustness

1. INTRODUCTION

It is quite possible, in multiple linear regression analysis, to have non-orthogonal explanatory variables with a high level of multicollinearity that inflates the variability of the Least Squares (LS) estimates. In such cases, the ridge estimation method is one way of dealing with the problem of multicollinearity, as proposed by Hoerl and Kennard (1970a,b).

A huge body of research shows the efficiency of the ridge regression and shrinkage methods in general. As an alternative to the LS method, the vast majority of the published research still deals with the LS context, whereas only a relatively tiny minority of the study concerns the shrinkage methods for the Least Absolute Deviations (LAD) estimates. In the LS context, it has been shown that the ridge estimation method is equivalent to special cases of each of the mixed (augmented) method (Marquardt 1970, Banerjee and Carr 1971), minimax method (Strawderman 1978), Bayesian method (Rao 1976), penalized splines (Gruber 2010), and the Least Absolute Shrinkage and Selection Operator (LASSO) method (Grandvalet 1998).

Their research claimed as the entrance into the LAD context of the ridge regression, Pfaffenberger and Dielman (1989) found the ridge biasing parameter k by the use of the LAD estimation of parameters and the variance of the residuals from the LAD fit. Denoted as the Ridge Least Absolute Value (RLAV) regression estimation, their estimation is still in the LS context, through defining only a new biasing parameter for the LS ridge regression. In a further development, named LAD-LASSO by the authors Wang, Li and Jiang (2007), a penalized LAD estimation method combines the LAD estimation with the LASSO method. The LAD-LASSO does parameter estimation and variable selection simultaneously, and is proposed for the cases with asymmetric error terms. Of its differences from the ridge is the selection of some variables and the removal of some others from the regression model.

The aim of this paper is to generalize the ridge estimation method in the LAD context in order to be able to deal, when the error terms are asymmetric or heavy-tailed, with the problem of multicollinearity. The LAD ridge regression methodology and the idea behind it are discussed in Section 2. The relative efficiency of the LAD ridge regression method compared with the methods of Ordinary Least Squares (OLS) ridge and the LAD regression, through a Monte Carlo simulation experiment, is checked in Section 3. An empirical example is used in Section 4 to assess the applicability of this new method and to compare it with what was done previously. Finally, a brief summary of conclusions is presented in Section 5.

2. METHODOLOGY

A simple way of expressing the ridge regression is through its equivalence to a special case of the mixed (augmented) regression model. A full rank mixed linear model (Theil and Goldberger, 1961) is

$$\begin{bmatrix} \mathbf{Y} \\ \mathbf{r} \end{bmatrix} = \begin{bmatrix} \mathbf{X} \\ \mathbf{R} \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{\varepsilon} \\ \mathbf{v} \end{bmatrix}, \quad (2.1)$$

where $\mathbf{Y}$ and $\boldsymbol{\varepsilon}$ are $n \times 1$ random vectors, $\mathbf{X}$ is a known $n \times p$ matrix and $\boldsymbol{\beta}$ is a $p \times 1$ unknown parameter vector, whereas $\mathbf{r}$ and $\mathbf{v}$ are $m \times 1$ random vectors and $\mathbf{R}$ is a known $m \times p$ matrix, with $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \boldsymbol{\Sigma}$, $E(\mathbf{v}) = \mathbf{0}$, $E(\mathbf{v}\mathbf{v}') = \boldsymbol{\Omega}$ and $E(\boldsymbol{\varepsilon}\mathbf{v}') = \mathbf{0}$. As a special case, Aitken's Generalized Least Squares (GLS) estimation of $\boldsymbol{\beta}$ is equivalent to the ordinary ridge regression estimation if $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_n$, $m = p$, $\mathbf{r} = \mathbf{0}$ and $\sigma^2 \mathbf{R}' \boldsymbol{\Omega}^{-1} \mathbf{R} = k \mathbf{I}$. The simplest case is when $\boldsymbol{\Omega} = \sigma^2 \mathbf{I}_p$ and $\mathbf{R} = \sqrt{k} \mathbf{I}_p$ for $k > 0$ (with $\boldsymbol{\Omega} = \sigma^2 \mathbf{I}_p$, it can be like $\mathbf{R} = \sqrt{k} \mathbf{U}$, where $\mathbf{U}$ is an orthogonal matrix). The simple mixed model becomes

$$\begin{bmatrix} \mathbf{Y} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{X} \\ \sqrt{k} \mathbf{I} \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{\varepsilon} \\ \mathbf{v} \end{bmatrix}. \quad (2.2)$$

The OLS estimation of $\boldsymbol{\beta}$ of the mixed model minimizes the objective function

$$\sum_{i=1}^n (y_i - \mathbf{x}_i' \boldsymbol{\beta})^2 + k \boldsymbol{\beta}' \boldsymbol{\beta}. \quad (2.3)$$

The idea behind the LAD version of the ridge estimation method adopted in this paper is to employ the same special form of mixed model (2.2) and, instead of using the OLS method, apply the LAD estimation method for the mixed model. The LAD estimation of $\boldsymbol{\beta} = (\beta_o, \dots, \beta_p)'$, based on the mixed model, minimizes the objective function

$$Q(\boldsymbol{\beta}) = \sum_{i=1}^n |y_i - \mathbf{x}_i' \boldsymbol{\beta}| + \sqrt{k} \sum_{j=0}^p |\beta_j|. \quad (2.4)$$

The LAD estimation of $\boldsymbol{\beta}$ in the mixed model (2.2) is considered as the LAD ridge estimation of $\boldsymbol{\beta}$, for the purposes of this paper. This is analogous to the OLS ridge estimation of $\boldsymbol{\beta}$, which is the OLS estimation based on the first augment and stemming from the mixed model. Throughout the paper, the OLS, the OLS ridge, the LAD and the LAD ridge parameter estimators are respectively denoted as, $\mathbf{b}$, $\mathbf{b}_r$, $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_r$.

In the realm of ridge regression analysis, extensive research is devoted to finding a proper value of the biasing parameter k . Standardization of all the variables, due to the fact that the value of k found from the sample is affected drastically by the scales, is the starting point of the process recommended by many researchers. However, if some researchers prefer to explain the fit model based on the original unstandardized variables, it is still possible to get back unstandardized LAD estimates from the standardized LAD estimates (see 2.7 below). It is also worth mentioning that the LAD fit does not necessarily pass through the sample mean point, as is the case with the OLS fit, i.e., standardization of the variables should not force the LAD fit to pass through the origin.

Consider the linear regression model

$$Y_i = \beta_o + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} + \varepsilon_i, \quad (2.5)$$

where $\boldsymbol{\beta} = (\beta_o, \beta_1, \dots, \beta_p)'$ is the unknown parameter vector and $\{\varepsilon_i\}$ are unobservable iid random variables with median 0. The standardized LAD parameter estimation $\hat{\boldsymbol{\beta}}^* = (\hat{\beta}_o^*, \hat{\boldsymbol{\beta}}^*)'$ is obtained as,

$$\begin{aligned} \hat{\boldsymbol{\beta}}^* &= \arg \min_{\boldsymbol{\beta}^* = (\beta_o^*, \boldsymbol{\beta}^*)'} \sum_{i=1}^n \left| (y_i - \bar{y})/s_Y - [1 \quad (\mathbf{x}_i - \bar{\mathbf{x}})' \mathbf{W}_\mathbf{x}^{-1}] \begin{pmatrix} \beta_o^* \\ \boldsymbol{\beta}^* \end{pmatrix} \right| \\ &= \arg \min_{\boldsymbol{\beta}^* = (\beta_o^*, \boldsymbol{\beta}^*)'} \sum_{i=1}^n \left| y_i - [1 \quad \mathbf{x}_i]' \left[\begin{pmatrix} s_Y & -s_Y \bar{\mathbf{x}}' \mathbf{W}_\mathbf{x}^{-1} \\ \mathbf{0} & s_Y \mathbf{W}_\mathbf{x}^{-1} \end{pmatrix} \begin{pmatrix} \beta_o^* \\ \boldsymbol{\beta}^* \end{pmatrix} + \begin{pmatrix} \bar{y} \\ \mathbf{0} \end{pmatrix} \right] \right|, \end{aligned} \quad (2.6)$$

where $\mathbf{W}_\mathbf{x}$ is the diagonal matrix of the sample standard deviations of the independent variables and s_Y is the sample standard deviation of the dependent variable. Based on Lemma 3 of Bassett and Koenker (1978), our conclusion is that between standardized LAD parameter estimator $\hat{\boldsymbol{\beta}}^* = (\hat{\beta}_o^*, \hat{\boldsymbol{\beta}}^*)'$ and unstandardized LAD estimator $\hat{\boldsymbol{\beta}}$ we have the relationship

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} s_Y & -s_Y \bar{\mathbf{x}}' \mathbf{W}_\mathbf{x}^{-1} \\ \mathbf{0} & s_Y \mathbf{W}_\mathbf{x}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\beta}_o^* \\ \hat{\boldsymbol{\beta}}^* \end{pmatrix} + \begin{pmatrix} \bar{y} \\ \mathbf{0} \end{pmatrix}, \quad (2.7)$$

which is similar to the functional relationship between the standardized and unstandardized OLS estimator, with a slight difference that, in the OLS method, $\hat{\beta}_o^*$ in (2.7) is replaced by 0.

Henceforth, we suppose that all the observable variables are standardized to guarantee that $\mathbf{X}'\mathbf{X}$ and $\mathbf{X}'\mathbf{Y}$ are of the correlation form (Gibbons, 1981). To deal with the standardized

variables when finding the LAD ridge estimates, we can simply add a column of ones along the whole design matrix of the mixed model (2.2) which is composed of a pile of the standardized design matrix $\mathbf{X}$ and $\sqrt{k} \mathbf{I}_p$. Other than to compare the effect on the OLS ridge and the LAD ridge estimates of retaining the square root of k (see the Appendix, for instance), there is no reason for keeping the square root of k in (2.4) and avoiding a simpler notation.

Analogous to the OLS ridge estimation of the canonical form of a regression model, the LAD version of the ridge estimation can also be defined in terms of the canonical form. Let the singular value decomposition of a standardized design matrix $\mathbf{X}$ be $\mathbf{X} = \mathbf{S}' \mathbf{\Lambda}^{1/2} \mathbf{U}'$, then the singular value decomposition of $\mathbf{X}'\mathbf{X}$ is $\mathbf{X}'\mathbf{X} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}'$, where $\mathbf{S}$ is a $p \times n$ orthogonal matrix, $\mathbf{\Lambda}$ is a $p \times p$ diagonal matrix of the eigenvalues of $\mathbf{X}'\mathbf{X}$ and $\mathbf{U}$ is a $p \times p$ orthogonal matrix of the eigenvectors of $\mathbf{X}'\mathbf{X}$. The canonical form of a regression model is $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{S}' \mathbf{\Lambda}^{1/2} \mathbf{U}' \boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{S}' \mathbf{\Lambda}^{1/2} \boldsymbol{\alpha} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\alpha} = \mathbf{U}' \boldsymbol{\beta}$. The OLS estimation of $\boldsymbol{\alpha}$ is

$$\mathbf{g} = \mathbf{\Lambda}^{-1/2} \mathbf{S}' \mathbf{Y}. \quad (2.8)$$

The LAD estimation of $\boldsymbol{\alpha}$ (after adding a column of ones to the design matrix of the canonical model) is

$$\hat{\boldsymbol{\alpha}}^* = \arg \min_{\boldsymbol{\alpha}^* = (\alpha_o, \boldsymbol{\alpha}')'} \sum_{i=1}^n \left| y_i - [1 \quad \mathbf{x}_i' \mathbf{U}] \begin{bmatrix} \alpha_o \\ \boldsymbol{\alpha} \end{bmatrix} \right| = \arg \min_{\boldsymbol{\alpha}^* = (\alpha_o, \boldsymbol{\alpha}')'} \sum_{i=1}^n \left| y_i - [1 \quad \mathbf{x}_i'] \begin{bmatrix} \alpha_o \\ \mathbf{U} \boldsymbol{\alpha} \end{bmatrix} \right|, \quad (2.9)$$

where $\hat{\boldsymbol{\alpha}}^* = (\hat{\alpha}_o, \hat{\boldsymbol{\alpha}}')'$.

As for the OLS estimation $\mathbf{g} = \mathbf{U}' \mathbf{b}$, where $\mathbf{b}$ is the OLS estimation of $\boldsymbol{\beta}$, from (2.9) and based on Lemma 3 of Bassett and Koenker (1978), we see that for the LAD estimation a similar association between $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\alpha}}$ holds, i.e., $\hat{\boldsymbol{\alpha}}^* = (\hat{\alpha}_o, \hat{\boldsymbol{\alpha}}' \mathbf{U}')' = (\hat{\beta}_o, \hat{\boldsymbol{\beta}}')'$, hence $\hat{\boldsymbol{\alpha}} = \mathbf{U}' \hat{\boldsymbol{\beta}}$. Briefly, $\hat{\boldsymbol{\alpha}}' \hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\beta}}' \hat{\boldsymbol{\beta}}$ like $\mathbf{g}' \mathbf{g} = \mathbf{b}' \mathbf{b}$. Thus, when the k of Hoerl, Kennard and Baldwin (1975) is $k_0 = p \hat{\sigma}^2 / \mathbf{g}' \mathbf{g}$, the k of Pfaffenberger and Dielman (1989) takes the form $k_0 = p \hat{\sigma}_m^2 / \hat{\boldsymbol{\alpha}}' \hat{\boldsymbol{\alpha}}$, where $\hat{\sigma}_m^2$ is obtained from the residuals of the LAD fit of standardized variables.

On the other hand, as we know, $\mathbf{g}_r = \mathbf{U}' \mathbf{b}_r$, where $\mathbf{g}_r$ is the OLS ridge estimation of the canonical form and $\mathbf{b}_r$ is the OLS ridge estimation of the original model. Generally, such an argument does not hold for the LAD ridge estimation method, i.e., it is generally the case that $\hat{\boldsymbol{\alpha}}_r \neq \mathbf{U}' \hat{\boldsymbol{\beta}}_r$, where $\hat{\boldsymbol{\alpha}}_r^* = (\hat{\alpha}_{or}, \hat{\boldsymbol{\alpha}}_r')'$ is the LAD ridge estimation of the canonical form and

$\hat{\boldsymbol{\beta}}_r^* = (\hat{\boldsymbol{\beta}}_{or}, \hat{\boldsymbol{\beta}}_r)'$ is the LAD ridge estimator of the original form. This is due to the fact that for $\mathbf{U}$ an orthogonal matrix, replacing $\sqrt{k}\mathbf{I}$ in model 2.1 by $\sqrt{k}\mathbf{U}'$ can result in different LAD estimates, whereas in the LS context the results are the same.

To study the asymptotic properties of the LAD ridge parameter estimators of the model (2.5) some regularity conditions well-known in the literature, for instance assumptions A1 and A2 of Knight (1998), are necessary. The assumptions are as follows.

A1: The error terms ε_i are iid random variables with mean 0 and distribution $F(\cdot)$ differentiable at median with positive derivative $f(0)$.

A2: For some positive definite matrix $\mathbf{C}$, $\lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{X}_n' \mathbf{X}_n = \mathbf{C}$, meaning that $\frac{1}{n} \max_{1 \leq i \leq n} \mathbf{x}_i' \mathbf{x}_i \rightarrow 0$.

Theorem: Under the regularity conditions A1 and A2, and the condition that $n^{-1/2} \sqrt{k} \rightarrow 0$, the LAD ridge estimator $\hat{\boldsymbol{\beta}}_r$ satisfies that $\sqrt{n}(\hat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}) \xrightarrow{d} N_{p+1} \left(\mathbf{0}, \frac{\mathbf{C}^{-1}}{[2f(0)]^2} \right)$ as $n \rightarrow \infty$.

Proof: The convex minimization problem of the convex function $Q(\boldsymbol{\beta})$ in 2.4 is written as

$$Z_n(\mathbf{u}) = Q(\boldsymbol{\beta} + n^{-1/2} \mathbf{u}) - Q(\boldsymbol{\beta}) = Z_n^{(1)}(\mathbf{u}) + Z_n^{(2)}(\mathbf{u}), \quad (2.10)$$

where

$$Z_n^{(1)}(\mathbf{u}) = \sum_{i=1}^n \left\{ |y_i - \mathbf{x}_i' \boldsymbol{\beta} - n^{-1/2} \mathbf{x}_i' \mathbf{u}| - |y_i - \mathbf{x}_i' \boldsymbol{\beta}| \right\} \quad (2.11)$$

and,

$$Z_n^{(2)}(\mathbf{u}) = \sqrt{k} \sum_{j=1}^p \left\{ |\beta_j + n^{-1/2} u_j| - |\beta_j| \right\}. \quad (2.12)$$

We have $Z_n^{(1)}(\mathbf{u}) \xrightarrow{d} -\mathbf{u}' \mathbf{W} + f(0) \mathbf{u}' \mathbf{C} \mathbf{u}$, with $\mathbf{W} \sim N_{p+1}(\mathbf{0}, \mathbf{C})$, as it is shown in Knight (1998).

On the other hand, $Z_n^{(2)}(\mathbf{u})$ converges in probability to zero, since

$$|Z_n^{(2)}(\mathbf{u})| = \left| \sqrt{k} \sum_{j=1}^p \left\{ |\beta_j + n^{-1/2} u_j| - |\beta_j| \right\} \right| \leq n^{-1/2} \sqrt{k} \sum_{j=1}^p |u_j| \rightarrow 0. \quad (2.13)$$

Then, Corollary 2 of Knight (1998) completes the proof. ■

The above theorem indicates that for $k < \infty$ the LAD ridge estimators are asymptotically distributed as the LAD estimators.

3. SIMULATION

The relative efficiency of the LAD ridge estimates, compared to each of the LAD estimates and the LS ridge estimates, is evaluated in this section. The linear scale heteroscedasticity model of Koenker and Bassett (1982),

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + (\mathbf{x}_i' \boldsymbol{\gamma} + 1) \varepsilon_i, \quad (3.1)$$

is employed, with iid $\varepsilon_i \sim \text{lognormal}(0, \sigma^2)$. The conditional quantile function of Y is $Q_Y(\tau | \mathbf{x}) = \mathbf{x}'(\boldsymbol{\beta} + \boldsymbol{\gamma} Q_\varepsilon(\tau)) + Q_\varepsilon(\tau)$, whereas the conditional mean function of Y becomes $E(Y | \mathbf{x}) = \mathbf{x}'(\boldsymbol{\beta} + \boldsymbol{\gamma} E(\varepsilon)) + E(\varepsilon)$. When $Q_\varepsilon(\tau)$ and $E(\varepsilon)$ are not equal to zero, due to the inclusion of the intercept they will be added freely to the intercept. Let us choose $\boldsymbol{\beta} = (1, 1, \dots, 1)'$ and $\boldsymbol{\gamma} = (0, 1, \dots, 1)'$ to get the real parameter for the conditional quantile function $\boldsymbol{\beta}(\tau) = (1 + Q_\varepsilon(\tau)) \cdot \mathbf{1}_p$ at $\tau = 0.5$ and for the conditional mean $\mathbf{B} = (1 + E(\varepsilon)) \cdot \mathbf{1}_p$.

Samples of size 30, 100 and 1000 are generated as follows. Models with 2 and 5 predictors are generated from multivariate normal distributions with means 10 and difference variances, but such a covariance matrix that gives us all mutual correlations equal, for each of the correlation levels 0.75, 0.9 and 0.99 between the predictors. While keeping the generated design matrix fixed, iid errors terms $\varepsilon_i \sim \text{lognormal}(0, \sigma^2)$, with $\sigma^2 = 0.2$ and 0.5 are generated and consequently the values of the dependent variable y_i are generated according to (3.1). Then, the process of generating the values of the error and the dependent variable, along with the fixed predictors (for each sample size) is repeated 5000 times.

At each of the 5000 replications, the predictors and the dependent variables are standardized so as to guarantee that $\mathbf{X}'\mathbf{X}$ and $\mathbf{X}'\mathbf{Y}$ are in the correlation form. Standardized variables firstly are put into the mixed model (2.1) in such a way that for the LAD ridge method the first column of the design matrix is a column of ones, while for the OLS ridge such a column is not needed. Then, the original parameter estimates are calculated from the standardized parameter estimates through (2.6). Finally, as the LAD ridge and the OLS ridge estimations are calculated for a relatively large range of k , the MSE of their estimators are graphed to keep track of the relative efficiency of the LAD ridge estimations along different values of k .

From the graphs shown in the Appendix, through keeping the same sample size and level of skewness (affected by σ^2 of lognormal distribution), it is noticed that the relative efficiency

of the LAD ridge estimators to the efficiency of the LAD estimators increases, with an increase in the level of correlation between the predictors. Additionally, keeping the same level of correlation and sample size, the relative efficiency of the LAD ridge estimators to the efficiency of the OLS ridge estimators increases with a corresponding increase in the level of skewness of the errors. When more predictors are involved, the gaps between the MSEs increase and the value of k corresponding to the minimum MSE for each of the LAD ridge and the OLS ridge estimators shifts to the right. With each of the LAD and the OLS methods, the MSE gap of the ridge and the non-ridge methods asymptotically vanishes.

4. EMPIRICAL EXAMPLE

As an example of skewness and multicollinearity, the data of 27 statewide observations on SIC33, the primary metals industry, used in Example 5.2 of Greene (2008), are taken to assess the applicability of the LAD ridge method. Results of estimating the parameters of a Cobb-Douglas production function of labor and capital are shown in Table 4.1.

Table 4.1: The OLS, the OLS Ridge, the LAD and the LAD Ridge Estimates with standard

Estimations (Standard Error)	OLS	OLS Ridge	LAD	LAD Ridge
Constant	1.170644 (0.326782)	1.201748 (0.317093)	1.001181 (0.412609)	0.941091 (0.397104)
Labor Coefficient	0.602999 (0.125954)	0.598332 (0.114160)	0.843942 (0.200973)	0.826310 (0.175700)
Capital Coefficient	0.375710 (0.085346)	0.375146 (0.077354)	0.208077 (0.118923)	0.227958 (0.098843)
R-Squared	0.943463	0.943462	0.934115	0.935979

For the estimations in Table 4.1, Hoerl-Kennard-Baldwin's $k= 0.235196$ and Pfaffenberger-Dielman's $k= 0.292122$ are used for both the OLS Ridge estimation and the LAD Ridge estimation, respectively. The standard errors of the LAD and the LAD ridge are calculated by bootstrapping the variables with 1000 repetitions. The residuals corresponding to the two observations 22 and 26 are outliers. Their exclusion from the data set affects the estimates of the OLS and the OLS ridge considerably, and those of the LAD method only slightly, while for 7 decimal places it does not affect the LAD ridge estimates at all. Additionally, the standard errors of the LAD ridge estimates are less than their corresponding LAD estimates.

5. CONCLUSIONS

As a robust method, the LAD ridge estimation is an attempt to make a harmony of two other similarly robust methods; the LAD estimation and the ridge estimation method. The problem of outliers or asymmetry in the data set, beset by multicollinearity at the same time, is dealt with by this new methodology. With multicollinear data, the relative efficiency of the LAD ridge over the OLS ridge method increases as the level of skewness of the error terms increases. Additionally, with asymmetric data and outliers, for any increase in the multicollinearity between the predictors, the relative efficiency of the LAD ridge over the LAD estimation method also increases.

References

- [1] Banerjee, K. S., and Carr, R. N. 1971. A Comment on Ridge Regression. Biased Estimation for Non-Orthogonal Problems, *Technometrics*, Vol. 13, No 4: 895-898.
- [2] Bassett, G., and Koenker, R. 1978. Asymptotic Theory of Least Absolute Error Regression, *Journal of the American Statistical Association*, Vol. 73, No. 363: 618-622.
- [3] Gibbons, D. G. 1981. A Simulation Study of Some Ridge Estimators, *Journal of the American Statistical Association*, Vol. 76, No. 373: 131-139.
- [4] Grandvalet, Y. 1988. Least Absolute Shrinkage is equivalent to quadratic penalization. In: Niklasson, L., Boden M., and Zimeske, T., editors, *Proceedings for the Eighth International Conference on Artificial Neural Networks ICANN'98*. Vol. 1 of *Perspectives in Neural Computing*, pages 201-206. Springer.
- [5] Greene, W. H. 2008) *Econometric Analysis*, 6th Edition, Prentice Hall, Upper Saddle River, New Jersey.
- [6] Gruber, Marvin H. J. 2010. *Regression Estimators: a Comparative Study*, 2nd edition, Baltimore: The Johns Hopkins University Press, Baltimore.
- [7] Hoerl, A. E., and Kennard, R. W. 1970a. Ridge Regression: Biased Estimation for Nonorthogonal Problems, *Technometrics*, Vol. 12, No. 1: 55-67.
- [8] Hoerl, A. E., and Kennard, R. W. 1970b. Ridge Regression: Application to Nonorthogonal Problems, *Technometrics*, Vol. 12, No.1: 69-82.
- [9] Hoerl, A. E., Kennard, R. W., and Baldwin, K. F. 1975. Ridge Regression: Some Simulations, *Communications in Statistics*, 4: 105-123.
- [10] Knight, K. 1998. Limiting Distributions for L1 Regression Estimators under General Conditions, *The Annals of Statistics*, Vol 26, No. 2: 755-770.
- [11] Koenker, R. and Bassett, G. 1982. Robust Tests for Heteroscedasticity Based on Regression Quantiles, *Econometrica*, Vol. 50, No. 1: 43-61.
- [12] Lawless, J. F., and Wang, P. 1976. A Simulation Study of Ridge and Other Regression Estimators, *Communications in Statistics*, 5:307-323.
- [13] Marquardt, D. W. 1970. Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation, *Technometrics*, Vol 12, No. 3: 591-612.
- [14] McDonald, G. C., and Glarneau, D. I. 1975. A Monte Carlo Evaluation of Some Ridge-Type Estimators, *Journal of the American Statistical Association*, Vol. 70, No. 350: 407-416.
- [15] Montgomery, D. C., Peck, E. A., and Vining, G. G. 2004. *Introduction to Linear Regression Analysis*, 3rd edition, Singapore: John Wiley & Sons, Inc.
- [16] Pfaffenberger, R. C., and Dielman, T. E. 1989. A comparison of regression estimators when both multicollinearity and outliers are present, In: Lawrence, K.D. and Arthur, J.L., editors, *Robust Regression: Analysis and Applications*, New York: Marcel Dekker, Inc. pp. 243-270.
- [17] Rao, C. R. 1976) Estimation of Parameters in a Linear Model, *The Annals of Statistics*, Vol. 4, No. 6: 1023-1037.
- [18] Strawderman, W. E. 1978) Minimax Adaptive Generalized Ridge Regression Estimators, *Journal of the American Statistical Association*, Vol. 73, No. 363: 623-627.
- [19] Theil, H., and Goldberger, A. S. 1961. On pure and mixed estimation in economics, *International Economic Review*, Vol. 2, No. 1: 65-78.
- [20] Wang, H., Li, G., and Jiang, G. 2007. Robust Regression Shrinkage and Consistent Variable Selection through the LAD-Lasso, *Journal of Business & Economic Statistics*, Vol. 25, No 3: 347-355.

APPENDIX

Figure A.1: Sample size=30, No. of predictors =2, $\sigma^2=0.2$.

Figure A.2: Sample size=30, No. of predictors =2, $\sigma^2=0.5$

Figure A.3: Sample size=100, No. of predictors =2, $\sigma^2=0.2$

Figure A.4: Sample size=100, No. of predictors =2, $\sigma^2=0.5$.

Figure A.5: Sample size=1000, No. of predictors =2, $\sigma^2=0.2$.

Figure A.6: Sample size=1000, No. of predictors =2, $\sigma^2=0.5$.

Figure A.7: Sample size=30, No. of predictors =5, $\sigma^2=0.2$.

Figure A.8: Sample size=30, No. of predictors =5, $\sigma^2=0.5$.

Figure A.9: Sample size=100, No. of predictors =5, $\sigma^2=0.2$.

Figure A.10: Sample size=100, No. of predictors =5, $\sigma^2=0.5$

Figure A.11: Sample size=1000, No. of predictors =5, $\sigma^2=0.2$

Figure A.12: Sample size=1000, No. of predictors =5, $\sigma^2=0.5$