
Regularized Nonlinear Least Trimmed Squares Estimator in the Presence of Multicollinearity and Outliers

George Kemboi Kirui Keitany, Ananda Omutokoh Kube, Joseph Mutua Mutisya, Fundi Daniel Muriithi

Department of Statistics and Actuarial Science, Kenyatta University (KU), Nairobi, Kenya

Email address:

gekemboi@gmail.com (G. K. K. Keitany), oakube3@gmail.com (A. O. Kub), fund.danie@gmail.com (F. D. Muriithi), josemutisya@gmail.com (J. M. Mutisya)

To cite this article:

George Kemboi Kirui Keitany, Ananda Omutokoh Kube, Fundi Daniel Muriithi, Joseph Mutua Mutisya. Regularized Nonlinear Least Trimmed Squares Estimator in the Presence of Multicollinearity and Outliers. *American Journal of Theoretical and Applied Statistics*. Vol. 7, No. 4, 2018, pp. 156-162. doi: 10.11648/j.ajtas.20180704.14

Received: April 17, 2018; **Accepted:** May 29, 2018; **Published:** June 29, 2018

Abstract: This study proposes a regularized robust Nonlinear Least Trimmed squares estimator that relies on an Elastic net penalty in nonlinear regression. Regularization parameter selection was done using a robust cross-validation criterion and estimation through Newton Raphson iteration algorithm for the optimal model coefficients. Monte Carlo simulation was conducted to verify the theoretical properties outlined in the methodology both for scenarios of presence and absence of multicollinearity and existence of outliers. The proposed procedure performed well compared to the NLS and NLTS in a viewpoint of yielding relatively lower values of MSE and Bias. Furthermore, a real data analysis demonstrated satisfactory performance of the suggested technique.

Keywords: Elastic Net, Multicollinearity, Regularization, Nonlinear Least Trimmed Squares, Outliers

1. Introduction

The knowledge of nonlinear regression is one of the widely used models in analyzing the effect of explanatory variables on a response variable. For example, Khademzadeh et al. [9] used MapReduce to model large-scale nonlinear regression problems, Ramalho and Ramalho [13] used moment based estimation of nonlinear regression models with boundary outcomes and endogeneity to nonnegative and fractional responses. Moreover, Tabatabai et al. [15] used robust nonlinear regression to estimate drug concentration, and tumor size-metastasis.

The computational field of statistics has developed over time with the introduction of the computer which has increased statistical methods and algorithms with numerous applications capable of generating and storing significant amounts of data. Various studies in Genomics, Medicine, Epidemiology, Marketing and basic sciences meet data sets whereby the effects of collinearity arise in analysis amongst their study variables [2].

Regularization techniques have widely been used for the solution of ill-posed problems occurring in the use of maxi-

imum likelihood or least squares methods and have proved successful in several fields including regression analysis and machine learning [16]. The methodology involves the addition of restrictions or penalty to a model [14], with the objective of preventing overfitting [11]. It also improves predictive accuracy in situations where there exist many predictor variables, high collinearity, seeking for a sparse solution or accounting for variables grouping in the high-dimensional dataset [5]. Tikhonov [17] developed this mathematical technique of regularization while working on the solution of ill-posed problems. In vast literature, choices of penalty functions such as, L_0 , L_1 , L_2 or L_∞ norms are available.

Ando, Konishi and Imoto [1] introduced radial basis functions with hyperparameter and constructed nonlinear regression models with the help of regularization. Tateishi, Matsui and Konishi [16] constructed nonlinear regression models with Gaussian basis functions using weighted L_1 type regularization for analyzing data with complex structures. Farnoosh, Ghasemian and Fard [4] proposed a weighted ridge penalty on a fuzzy nonlinear regression model using fuzzy numbers and Gaussian basis functions. Jiang, Jiang and Song [7] developed weighted composite regression estimation and used the Adaptive Lasso and SCAD regularization to achieve

a simultaneous parameter model estimation and selection. Zucker et al. [19] developed an approximate version of the Stefanski-Nakamura corrected score approach, using the method of regularization to obtain an approximate solution of the relevant integral equation and Hang et al. [6] proposed a graph regularized nonlinear ridge regression (RR) model for remote sensing data analysis, including hyper-spectral image classification and atmospheric aerosol retrieval. Although these regularization methods have shown exceptional performance in various fields, it uses the least squares loss function which is influenced by outliers [12]. Outlier resistant regularized methods have been developed by replacing the least squares loss function with the robust technique [5]. Amongst these methods are Huber’s M-estimators, MM-estimators, Least Trimmed Squares, Least Median Squares estimators. Lim [10] proposed robust Ridge regression estimation procedures for nonlinear models with varying variance structure.

Regularization parameter selection is an essential problem in Lasso-type methods since effective variable selection, and model estimation depends on adjusted parameters. AIC, BIC, Cross-validation, Mallows’s C_p criterion and generalized information criterion have been suggested for choosing regularization parameters. However, the Lasso-type penalty cannot be analytically derived because they are not differentiable and local quadratic approximation, LARS, and Coordinate descent algorithms have been developed to settle this issue.

Elastic Net penalty, a regularization method, has been established to encourage grouping effect when predictors are highly correlated and also useful when there exists a large number of predictors than that of the observations in linear regression [18]. Section 2 introduces the proposed regularized robust nonlinear Least Trimmed Squares estimator with an Elastic net penalty in nonlinear regression models. In section 3, the procedure for estimation and parameter selection of the proposed methodology is derived. Lastly, the efficiency of the strategy is investigated through Monte Carlo study and real data analysis in section 4.

2. Regularized Nonlinear Least Trimmed Squares Regression

The nonlinear regression model has the form

$$y_i = f(x_i, \beta) + \varepsilon_i. \quad i = 1, 2, \dots, n \quad (1)$$

where y_i are response values, x_i covariates, $f(x_i, \beta)$ is a nonlinear function of the parameters, β is an unknown p dimensional vector of parameters and ε_i is a random error assumed to have a distribution function $F(\cdot)$ with finite variance. The most popular criterion for obtaining parameter model estimates is the Nonlinear Least Squares criterion (NLS), given by

$$\min_{\beta} \sum_{i=1}^n (y_i - f(x_i, \beta))^2. \quad (2)$$

However, the error distribution in the model (1) may exhibit heavy tails and asymmetry due to data inadequacies often caused by measurement or recording errors. Robust techniques are commonly used to accommodate violation of the assumptions that such errors cause.

Nonlinear Least Trimmed Squares estimator (NLTS) is a high breakdown regression technique which safeguards against wild observations [3]. It derives the parameters of the model (1) by minimizing the objective function of the form

$$\hat{\beta}_{NLTS} = \arg \min_{\beta} \sum_{i=1}^s r_{(i)}^2(\beta) \quad (3)$$

Where s is the breakdown point, $r_{(i)}^2$ are the order statistics of the squared residuals and $s \leq n$. Model estimates of the objective function (3) are derived using an iterative procedure because it cannot be expressed in a closed form. An approach to nonlinear modeling is to approximate $f(x_i, \beta)$ by a first order Taylor series expansion about an initial value β^0 . This is given by

$$f(x_i, \beta) = f(x_i, \beta^0) + \frac{\partial f(x_i, \beta)}{\partial \beta_j} \Big|_{\beta=\beta^0} (\beta - \beta^0)$$

and leads to the approximation

$$\hat{\beta}_{NLTS} = \arg \min_{\beta} \sum_{i=1}^s (y_i - f(x_i, \beta^0) - \frac{\partial f(x_i, \beta)}{\partial \beta} \Big|_{\beta=\beta^0} (\beta - \beta^0))^2.$$

when the Jacobian $\frac{\partial f(x_i, \beta^0)}{\partial \beta}$ is of full rank p , the step is unique and is presented as

$$(\beta - \beta^0) = \left(\frac{\partial f(x_i, \beta)}{\partial \beta} \Big|_{\beta=\beta^0} \cdot \frac{\partial f(x_i, \beta)}{\partial \beta} \Big|_{\beta=\beta^0} \right)^{-1} \cdot \frac{\partial f(x_i, \beta)}{\partial \beta} \Big|_{\beta=\beta^0} \cdot (y - f(x_i, \beta^0))$$

However, due to the presence of multicollinearity, the inversion may not be useful. The study proposes to obtain parameter model estimates to enhance the estimation performance of NLTS by minimizing the objective function

$$\hat{\beta}_{NLTS-ELnet} = \arg \min_{\beta} \left[\sum_{i=1}^s r_{(i)}^2 + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^p \beta_j^2 \right] \quad (4)$$

Where $\lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^p \beta_j^2$ is the Elastic net penalty.

The objective function (4) above combines NLTS loss function with an Elastic net penalty envisaged to produce

stable model estimates when multicollinearity is present with the existence of outliers. The NLTS-ELnet estimator is based only on s observations having small residuals to relax the effect of outliers satisfying $\frac{n}{2} < s \leq n$. The trimming constant s defines the breakdown point of the estimator, and a breakdown point of $s = 0.75n$ is considered, to include a sufficient number of observations [8].

3. Parameter Selection and Estimation of the NLTS-ELnet Estimator

The objective function (4) above can be simplified to

$$\hat{\beta}_{NLTS-ELnet} = \underset{\beta}{\operatorname{arg\,min}} \left\{ \sum_{i=1}^s r_{(i)}^2 + \lambda P_{\alpha}(\beta) \right\}. \quad (5)$$

Where,

$$P_{\alpha}(\beta) = \sum_{j=1}^p \left\{ \frac{\alpha \beta_j^2}{2} + (1-\alpha) |\beta_j| \right\}, \quad (6)$$

and $\alpha = \lambda_2 / (\lambda_1 + \lambda_2)$, $0 \leq \alpha \leq 1$ and $\lambda > 0$.

However, the penalty term (6) cannot be analytically derived since it is not differentiable at zero. The local quadratic approximation of a Lasso-type penalty has been proposed to settle this drawback. Suppose β^0 is given as an initial value of the minimizer of equation (5), then the local quadratic approximation of the derivative of the penalty term is given by

$$\left[P_{\lambda}(|\beta_j|) \right]' \approx \left\{ \frac{P'_{\lambda}(|\beta_j^0|)}{|\beta_j^0|} \right\} \beta_j \quad (7)$$

Where for $P_{\lambda}(|\beta_j|) = \lambda |\beta_j|^q$, $P'_{\lambda}(|\beta_j|) = \lambda q |\beta_j|^{q-1}$ evaluated at $\beta = \beta^0$.

The regression parameter vector $\hat{\beta}$ is derived using Newton-Raphson iterative method given by

$$\beta^{new} = \beta^{old} - \left[E \left\{ \frac{\partial^2 \hat{\beta}_{NLTS-ELnet}}{\partial \beta \partial \beta^T} \right\} \right]^{-1} \frac{\partial \hat{\beta}_{NLTS-ELnet}}{\partial \beta}. \quad (8)$$

The study proposes to use the robust Generalized Cross Validation approach to select regularization parameters of the suggested procedure. The criterion is given by

$$GCV = \frac{1}{s} \frac{\sum_{i=1}^s \{y_i - \hat{f}(\lambda, x_i)\}^2}{\left\{1 - \frac{1}{s} \operatorname{tr}H\right\}^2} \quad (9)$$

Where $y_i - \hat{f}(\lambda, x_i)$ is the predicted value and H the hat matrix. An optimal set of the regularization parameters that

minimizes equation (8) based on the estimated $\hat{\beta}$ by the NLTS-ELnet at each set of tuning parameters λ_1 and λ_2 are selected using grid search.

4. Data Analysis and Discussion

In this section, Monte Carlo study is conducted to investigate the behaviour of the proposed NLTS-ELnet estimator on simulated and real data with a comparison of its performance to NLS and NLTS. The Mean Squared Error and Bias are considered in evaluating their performance. A breakdown point of $s = 0.75n$ is considered to include a sufficient number of observations. Standard normal error distribution with 30% Uniform (-1, 4) and Student's t (3) contamination are considered. Subsection (4.1) studies the properties of NLTS-ELnet, NLTS and NLS estimators using simulated data sets. Subsection (4.2) exhibits the behaviour of NLTS-ELnet, NLS and NLTS methods on SENIC data set obtained from the Hospital Infections Program for the 1975-76 study period.

4.1. Study on the Regularized Least Trimmed Squares Estimator

In this subsection, a Monte Carlo study is carried out to evaluate the performance of the proposed nonlinear regression estimator.

Samples of size 10, 20, 40, 80 and 200 are generated from an exponential regression model (10) below adopted from [7].

$$y = \beta_0 + \beta_1 \exp(cX^T) + \varepsilon. \quad (10)$$

where $X = x_i$ for $i=1, 2, 3$, and $x_i \sim N(0, 1)$, ε is the error, β_0, β_1 and $c = (\beta_2, \beta_3, \beta_4)$ are model parameters. The true values of the parameters are $\beta_0 = 1, \beta_1 = 1.5$ and $c = (-0.6, -0.8, -0.7)$. The predictors used in model (18) are generated with Correlation between x_l and x_m components chosen as 0.99 for all $l \neq m$.

MSE and Bias of parameter estimates are computed using the following equations

$$MSE(\hat{\beta}_j) = \frac{\sum_{i=1}^{100} (\hat{\beta}_j^i - \beta_j)^2}{100},$$

and,

$$\operatorname{Bias}(\hat{\beta}_j) = \left(\frac{1}{100} \sum_{i=1}^{100} \hat{\beta}_j^i - \beta_j \right).$$

where $\hat{\beta}_j^i$ is the j^{th} estimate for the i^{th} replication and β_j is the actual value of the j^{th} coefficient.

Table 1. Parameter estimates, Mean squared error and Bias for NLS under Standard normal error distribution with 30% Uniform (-1, 4) and Student's t (3) contamination and correlation $\rho = 0.99$, where the actual value of $\beta_1 = 1.5, \beta_2 = -0.6, \beta_3 = -0.8$ and $\beta_4 = -0.7$.

Sample size	Method		N (0, 1)	U (-1, 4)	Student's t (3)
10	NLS	$\hat{\beta}_1$	1.609327	1.77371411	1.660426
		$\hat{\beta}_2$	0.243427	0.07853708	1.407617
		$\hat{\beta}_3$	-1.605003	-1.46876380	-2.121418
		$\hat{\beta}_4$	-0.734116	-0.58356018	-1.457292
		MSE	0.9595708	0.7582854	2.670094
		Bias	-0.3465914	-0.2750182	-0.3526667
		20	NLS	$\hat{\beta}_1$	1.5272375
$\hat{\beta}_2$	-0.4701652			-1.3465106	-0.2879741
$\hat{\beta}_3$	-0.6547902			-0.8149247	-1.0895364
$\hat{\beta}_4$	-0.9778032			0.1913489	-0.5474260
MSE	0.3506121			0.3859287	0.2387807
Bias	-0.3688803			-0.2088896	-0.2839435
40	NLS			$\hat{\beta}_1$	1.4959432
		$\hat{\beta}_2$	-0.5533025	-0.7946368	-0.5926128
		$\hat{\beta}_3$	-1.2039739	-0.7353629	-1.0001144
		$\hat{\beta}_4$	-0.3668179	-0.5043841	-0.4862598
		MSE	0.39657641	0.2207952	0.3115529
		Bias	-0.3820378	-0.3026186	-0.3454058
		80	NLS	$\hat{\beta}_1$	1.5291588
$\hat{\beta}_2$	-0.4599272			-0.4428682	-0.4073917
$\hat{\beta}_3$	-0.6229685			-0.5830034	-0.5286050
$\hat{\beta}_4$	-1.0102957			-1.0221216	-1.1369590
MSE	0.3573145			0.358397	0.4040612
Bias	-0.3660082			-0.3230497	-0.3472507
200	NLS			$\hat{\beta}_1$	1.4766949
		$\hat{\beta}_2$	-0.6142729	-0.6253511	-0.6158236
		$\hat{\beta}_3$	-0.7626872	-0.7067085	-0.7676446
		$\hat{\beta}_4$	-0.7299139	-0.7281655	-0.7006240
		MSE	0.2788032	0.2615671	0.2707363
		Bias	-0.3825448	-0.3334694	-0.3584172

Table 2. Parameter estimates, Mean squared error and Bias for NLTS under Standard normal error distribution with 30% Uniform (-1, 4) and Student's t (3) contamination and correlation $\rho = 0.99$, where the true value of $\beta_1 = 1.5, \beta_2 = -0.6, \beta_3 = -0.8$ and $\beta_4 = -0.7$.

Sample size	Method		N (0, 1)	U (-1, 4)	Student's t (3)
10	NLTS	$\hat{\beta}_1$	1.3437355	1.6982425	0.9987979
		$\hat{\beta}_2$	0.3507451	-1.2423452	1.7786562
		$\hat{\beta}_3$	-2.0735187	-1.7478642	-1.3560213
		$\hat{\beta}_4$	-1.7891295	0.7787984	-2.2302166
		MSE	1.444423	1.028366	3.180467
		Bias	-0.519817	-0.3532921	0.6879124
		20	NLTS	$\hat{\beta}_1$	1.2574173
$\hat{\beta}_2$	0.4424735			-0.1514221	-0.4278354
$\hat{\beta}_3$	-1.7775931			-0.1245679	-0.4368943
$\hat{\beta}_4$	-0.9978442			-1.0034554	-0.5578914
MSE	1.325282			0.3201558	0.1396193

Sample size	Method		N (0, 1)	U (-1, 4)	Student s t (3)
40	NLTS	Bias	-0.4938866	-0.2295277	-0.2465187
		$\hat{\beta}_1$	1.8534590	1.7527521	1.9904183
		$\hat{\beta}_2$	-0.7851019	-0.5478011	-0.5963147
		$\hat{\beta}_3$	-1.5693212	-0.6329218	-0.8815890
		$\hat{\beta}_4$	-0.5213991	-0.3021591	-0.5256798
80	NLTS	MSE	0.7234556	0.1396244	0.3235155
		Bias	-0.4805909	-0.1575323	-0.2282913
		$\hat{\beta}_1$	1.4587371	1.4969420	1.4997631
		$\hat{\beta}_2$	-0.6788972	-0.6690801	-0.5905346
		$\hat{\beta}_3$	-0.7653232	-0.5797892	-0.7485236
200	NLTS	$\hat{\beta}_4$	-0.6888851	-0.6091631	-0.4974361
		MSE	0.2670524	0.1780042	0.201955
		Bias	-0.393592	-0.3152725	-0.3091828
		$\hat{\beta}_1$	1.5036851	1.4975423	1.5352761
		$\hat{\beta}_2$	-0.6675032	-0.5659642	-0.5178946
		$\hat{\beta}_3$	-0.6554532	-0.6731342	-0.7986543
		$\hat{\beta}_4$	-0.6944664	-0.6656221	-0.5838793
		MSE	0.2291182	0.2243317	0.2466874
		Bias	-0.3534344	-0.3267946	-0.316288

Table 3. Parameter estimates, Mean squared error and Bias for NLTS-Elnet under Standard normal error distribution with 30% Uniform (-1, 4) and Student's t (3) contamination and correlation $\rho = 0.99$, where the true value of $\beta_1 = 1.5, \beta_2 = -0.6, \beta_3 = -0.8$ and $\beta_4 = -0.7$.

Sample size	Method		N (0, 1)	U (-1, 4)	Student s t (3)
10	NLTS-Elnet	$\hat{\beta}_1$	1.1638659	1.4980528	0.9167169
		$\hat{\beta}_2$	0.2218354	-0.8023751	1.5666763
		$\hat{\beta}_3$	-1.1224322	-1.5938046	-2.3750704
		$\hat{\beta}_4$	-1.6595342	0.4037786	-2.0503756
		MSE	0.453267	0.4723531	2.334767
20	NLTS-Elnet	Bias	-0.1990663	0.0264129	-0.3355132
		$\hat{\beta}_1$	1.4264178	1.4624274	1.5185405
		$\hat{\beta}_2$	0.2344779	-0.2758001	-0.5070444
		$\hat{\beta}_3$	-1.6575995	-0.6473159	-0.6317984
		$\hat{\beta}_4$	-0.9217019	-1.1847514	-0.9538816
40	NLTS-Elnet	MSE	0.371599	0.09120341	0.02543304
		Bias	-0.07960139	-0.01135998	0.006454027
		$\hat{\beta}_1$	1.5592783	1.5662982	1.8057587
		$\hat{\beta}_2$	-0.5759615	-0.6228295	-0.3979009
		$\hat{\beta}_3$	-1.0434944	-0.9112180	-0.7115833
80	NLTS-Elnet	$\hat{\beta}_4$	-0.4675405	-0.5589851	-0.8197677
		MSE	0.02935468	0.009292828	0.03912356
		Bias	0.01807045	0.01831639	0.1191264
		$\hat{\beta}_1$	1.4749362	1.5074405	1.5441754
		$\hat{\beta}_2$	-0.5436800	-0.5642293	-0.6305667
200	NLTS-Elnet	$\hat{\beta}_3$	-0.8648110	-0.9621981	-0.8440099
		$\hat{\beta}_4$	-0.7061941	-0.5791841	-0.6055693
		MSE	0.002009741	0.0105599	0.003434955
		Bias	-0.009937187	0.0004572395	0.01600737
		$\hat{\beta}_1$	1.4859725	1.4735395	1.5159358
		$\hat{\beta}_2$	-0.6097271	-0.6054939	-0.6252666
		$\hat{\beta}_3$	-0.7518974	-0.7423253	-0.8325394
		$\hat{\beta}_4$	-0.7443109	-0.7613827	-0.6383703
		MSE	0.001142175	0.001956137	0.001437345
		Bias	-0.004990739	-0.008915572	0.004939872

Table 4. Parameter Model estimates using NLS and NLTS-Elnet Methods for SENIC data.

Parameter	Method	Estimated Parameters
$\hat{\beta}_1$	NLS	-1.17154828
	NLTS	-1.17154881
	NLTS-Elnet	-1.20376870
$\hat{\beta}_2$	NLS	-0.28534202
	NLTS	-0.28434123
	NLTS-Elnet	-0.03864024
$\hat{\beta}_3$	NLS	-0.08243559
	NLTS	-0.08121456
	NLTS-Elnet	-0.06126209
$\hat{\beta}_4$	NLS	0.11961430
	NLTS	0.12673410
	NLTS-Elnet	0.08975999

Table 5. MSPE estimates of NLS and NLTS-Elnet Methods for SENIC data set.

METHOD	RMSPE
NLS	0.03042244
NLTS	0.03033311
NLTS-Elnet	0.03029495

+ The example for this table. The example for this table.

Tables (1), (2) and (3) displays the MSE and Bias under standard normal error distribution, and in the presence of Uniform and students t error contaminations. Moreover, parameter estimates of model (18) are displayed where the three covariates x_1 , x_2 , and x_3 are highly correlated ($\rho = 0.99$).

From table (1), the nonlinear least squares method show low estimates of MSE and bias under the considered error distributions and sample sizes. Throughout the considered error distributions, the method performed poorly under small sample sizes, producing estimates far from the true values, and estimates well with the sample of size 200.

From table (2), the nonlinear least trimmed squares estimator shows high values of MSE and bias estimates for small sample sizes and reduces significantly for larger sample sizes. Generally, the values of MSE are lower than those for NLS.

From table (3), the proposed regularized NLTS method give best estimates of MSE, bias and parameters of the model. These values became better when n increased from 10 to 40, which give true values from 80 to 200 under all the considered error distributions.

The methodologies perform poorly when n=10 and improves as the sample increases. Furthermore, NLTS-Elnet has the lowest values of MSE under all the considered error distributions. Its values also seem to decrease consistently from n=10 to n=200 except for U (-1, 4) unlike for NLS and NLTS. Moreover, the model estimates of the proposed method appear to approach a stable value faster than those for the considered methods.

4.2. Real Data Application

In this section, the NLTS-Elnet, NLS and NLTS estimators are applied to SENIC data set obtained from the Hospital Infections Program for the 1975-76 study period. Model (11) below is used to model the relationship between the infection

rate (y) and length of stay (x_1), routine culturing (x_2), number of beds (x_3) and the average daily census (x_4) for hospitals indexed 1, 2, 3, 111, 112 and 113. NLTS-Elnet, NLS methods are compared in parameter model estimation.

$$y = \frac{\exp(\sum_{i=1}^{11} \beta_i x_i)}{1 + \exp(\sum_{i=1}^{11} \beta_i x_i)} \tag{11}$$

First, Variance Inflation Factor (VIF) which measures multicollinearity in independent variables was computed using the formula,

$$VIF = \frac{1}{1 - R_i^2}$$

where R_i^2 is the multiple correlation coefficient when x_i is regressed against all the other explanatory variables in the model. The average VIF value was calculated as 10.54781 which is an indication of highly correlated variables. Mean Squared Prediction Error (MSPE), the fitted and estimated observations are considered in evaluating the performances of NLTS-Elnet, NLS and NLTS methods. The MSPE is given by

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

where y_i and \hat{y}_i denote the actual and estimated observations respectively. The results are summarized in Tables (4) and (5).

From Table (4), the estimated coefficients for $x_1 - x_3$ were negative and positive for x_4 , showing that the infection rate

increases with the average number of patients in the hospital per day and decreases with the length of stay, routine culturing and the average number of beds during the 1975-76 study period of the Hospital Infections Program. Probably, this depicts the actual scenario because a high average number of patients in the hospital increases the chances of cross infection among patients which can be minimized by decreased length of stay, routine culturing and the average number of beds. The Mean Squared Prediction Error values in Table (5) are comparably the same and lowest for the proposed methodology.

5. Conclusions

Regularized robust Nonlinear Least Trimmed Squares estimator has been proposed in this study by adding an Elastic net penalty to the NLTS loss function. Robust generalized cross-validation was used to select the regularization parameters robustly. It can be seen through Monte Carlo study and real-world data example that the proposed methodology performs well in several situations compared to the NLS and NLTS in the viewpoint of yielding relatively lower values of MSE and Bias when there is the presence of multicollinearity. A breakdown point of $0.75n$ was predetermined to enhance efficiency by including the majority of the data points. Future work remains to be done towards considering the simultaneous selection of optimal values of regularization parameters and the breakdown point. Also, variable selection behaviour of the proposed estimator in a large number of predictors could be a possible problem which needs to be explored.

References

- [1] Ando, T., Konishi, S., Imoto, S. (2008). Nonlinear regression modeling via regularized radial basis function networks, *Journal of Statistical Planning and Inference*, Volume 138, Issue 11, Pages 3616-3633.
- [2] Batterham, A., Tolfrey, K., and George, K. Nevills (1997). explanation of kleibers 0.75 mass exponent: an artifact of collinearity problems in least squares models? *Journal of Applied Physiology*, 82:693-697.
- [3] Cizek, P. (2001). Nonlinear least trimmed squares. SFB Discussion paper Humboldt University, 25/2001.
- [4] Farnoosh, R., J. Ghasemian, J., Fard, O. S. (2012). *Comp. and Appl. Math.* - no.2, 31, 323-338.
- [5] Hashem, H. (2014). Regularized and robust regression methods for high-dimensional data. Department of Mathematical Sciences, Brunel University.
- [6] Hang, R., Liu, J. J., Q., Song, H., Zhu, F., and Pei, H. (2017). Graph Regularized Nonlinear Ridge Regression for Remote Sensing Data Analysis. *IEEE*, 10: 277 – 285.
- [7] Jiang, X., Jiang, J., and Song, X. (2012). Oracle model selection for nonlinear models based on weighted composite quantile regression. *Statistica Sinica*, 22:1479-1506.
- [8] Kamal, D. and Ali, B. (2015). H. Robust linear regression using l1-penalized mm-estimation for high dimensional data. *American Journal of Theoretical and Applied Statistics*, 4(3):77-84.
- [9] Khademzadeh, A., Khademzadeh, A., D, P. C. P., D, M. S. P., and Anagnostopoulos, G. C. (2013). Large-scale non-linear regression within the mapreduce framework.
- [10] Lim, C. (2015). Robust ridge regression estimators for nonlinear models with applications to high throughput screening assay data. *Statistics in medicine*.
- [11] Ohlsson, H. (2010). Regularization for sparseness and smoothness applications in system identification and signal processing. Linkoping University Electronic Press.
- [12] Park, H. (2013). Robust regression modelling via l1 type regularization. Dept. of Mathematics Chuo University.
- [13] Ramalho, E. and Ramalho, J. (2014). Moment-based estimation of nonlinear regression models with boundary outcomes and endogeneity, with applications to nonnegative and fractional responses. CEFAE-UE Working Paper 2014/09.
- [14] Sima, D. (Apr. 2006). Regularization techniques in modeling and parameter estimation. PhD thesis.
- [15] Tabatabai, M. A., Kengwoung-Keumo, J. J., Eby, W. M., Bae, S., and Manne, U. (2014). A new robust method for nonlinear regression. *J Biomet Biostat*, 5.
- [16] Tateishi, S., Matsui, H., and Konishi, S. (2009). Nonlinear regression via the lasso-type regularization. *Journal of statistical planning and inference*.
- [17] Tikhonov, A. N. (1943). On the stability of inverse problems. *Dokl. Akad. Nauk SSSR*, 39:176-179.
- [18] Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society*, B67: 301-320.
- [19] Zucker, D. M., Gorfine, M., Li, Y., Tadesse, M., & Spiegelman, D. (2013). A Regularization Corrected Score Method for Nonlinear Regression Models with Covariate Error. *Biometrics*, 69(1), 80–90.