Pak. J. Statist.
2015 Vol. 31(3), 281-294

RESAMPLING METHOD FOR THE DATA ADAPTIVE CHOICE OF TUNING CONSTANT IN ROBUST REGRESSION

Zafar Mahmood[1] and **Salahuddin**[2*]

[1] Department of Mathematics, Statistics and Computer Science
The University of Agricultural Peshawar, Pakistan
Email: zafarjan@gmail.com

[2] Department of FAMCO, University of Dammam, Dammam,
Kingdom of Saudi Arabia. Email: salahuddin_90@yahoo.com

ABSTRACT

Robust regression estimators are designed to easily fit the data sets that are contaminated with outliers. A common problem associated with robust regression estimators to be completely specified is the proper choice of cut-off points called tuning constant. The choice of the tuning constant is somewhat arbitrary and is largely the matter of the personal preference. Several authors suggested different value of tuning constant '*c'* for various M estimators. Yohai (1974) considered a class of error distribution in the linear regression model and showed how to choose *c* for Huber's robust regression estimator so that the resulting estimator was minimax over the class of error distribution. Kelly(1992), with a different approach, showed that the choice of tuning constant is critical in the trade-off between bias and variance and suggested minimum choice of jackknife asymptotic mean-square error of the estimator to choose tuning constant. Salahuddin (1990) used leave-one-out cross-validation to choose an optimal value of tuning constant for Andrews' wave estimator. We propose K-fold cross-validation procedure for choosing optimal choice of tuning constant to minimize cross-validated absolute Median Predicted Residual. Furthermore, we investigated a suitable preliminary resistant estimator to arrive at a good robust fit. Andrew's, Tukey's, Qadir and Asad robust regression estimators are explored and compared. The study found that the proposed technique is working well.

1. INTRODUCTION

The ordinary least square method is the optimal procedure for fitting linear regression model when the necessary assumptions are fulfilled, see, Draper and smith (1998). But when the regression model does not meet the fundamental assumptions or if the data contain missing observations or outliers, the sample estimates can be misleading and the predictions of the model may become biased, see, Rousseeuw and Leroy (1987), Ho & Naugher (2000). If the basic assumption of the normality of the residuals is violated that is, the distribution of errors is heavy tailed; the least squares method may not be appropriate, see, Andrews et al. (1972). To deal with long-tail error distribution, one approach is to construct outlier diagnostics, remove the largest residuals as unusual points and still use

[*] Institute of Management and Information Sciences, CECOS,
University of IT & Emerging Sciences, Peshawar, Pakistan.

least square method. However, least squares may not be effective if there are many outliers in the data set because of the leave-one-out nature of the outlier test. Accordingly, robust regression procedures provide an alternative and have been developed to modify the least squares methods so that outliers have least influence on the final estimates.

There are several methods of robust regression, such as, M-estimators introduced by Huber (1981), least trimmed squares (LTS) introduced by Rousseeuw (1984), S-estimators introduced by Rousseeuw & Yohai (1984), least median of squares regression (LMS) introduced by Rousseeuw (1984) and MM-estimators introduced by Yohai (1987). We discuss only M-estimators for regression as these estimators are relatively simple, perform well and are easy to compute. The choice of tuning constant is somewhat arbitrary for these M-estimators and we applied K-fold cross-validation procedure with alternative estimates of prediction error (Median Prediction Error) to choose best value of tuning constant for popular and some newly established M-estimators.

1.1 M-Estimation

Robust regression procedures are concerned with modification of linear least squares when the error distribution is not normal, particularly when the errors are heavy-tailed. The term robust regression is to employ a fitting criterion that is not as vulnerable as least squares to unusual data.

The most common general method of robust regression is M-estimation. This class of estimators can be regarded as a generalization of maximum likelihood estimation, hence the term 'M' estimation.

Consider the linear model,

$$Y = X\beta + \varepsilon \tag{1}$$

Suppose the errors ε_i are independent random variables and follow a double exponential distribution:

$$f(\varepsilon_i) = \frac{1}{2}\sigma\left[\exp\left(-|\varepsilon_i|\right)/\sigma\right] \tag{2}$$

Here the estimates of β are obtained by maximum likelihood method using the likelihood function:

$$\begin{aligned} L(\beta) &= \prod_{i=1}^{n} \frac{1}{2}\sigma\left[\exp\left(-|\varepsilon_i|\right)/\sigma\right] \\ &= \left(\frac{1}{2}\sigma\right)^n \exp\left(-\sum|\varepsilon_i|/\sigma\right) \end{aligned} \tag{3}$$

Maximizing the likelihood function would involve minimizing the sum of absolute error that is $\sum|\varepsilon_i|$. Minimizing the sum of absolute error is often called L1-norm regression problem, while least square is called the L2-norm regression problem that minimizes the sum of the square of errors. Thus for heavy-tailed distribution, least square method is no longer the optimum choice and L1-norm regression is quite preferable. For more details on L1-norm regression, see Book et al. (1980), Narula and Wellington (1982), and Dodge (1987).

Similar to L1-norm regression, another approach for heavy-tailed distribution is the use of M-estimation. The basic idea of M-estimations is to minimize some function of residual rather than the sum of the square residuals. The general M-estimator minimizes the objective function:

$$\sum \rho(u_i) = \sum \rho\left(y_i - x_i^T \beta\right). \tag{4}$$

where, ρ is a symmetric function (that is $\rho(u) = \rho(-u)$ for all u) with a unique minimum at zero and gives the contribution of each residual to the objective function. The function ρ is related to the likelihood function for an appropriate choice of error distribution. For example, the least square estimation is $\rho(u_i) = u_i^2$.

Differentiating the objective function with respect to the coefficients β and setting the partial derivative equals zero provides $k+1$ estimating equations for coefficients. Let $\psi = \rho'$ (the derivative of ρ), then

$$\sum \psi\left(y_i - x_i^T \beta\right) x_i^T = 0 \tag{5}$$

The solution of this equation that minimizes the objective function is called the M-estimator of β. These M-estimators are not necessarily scale invariant that is if the error were multiplied by a constant, the new solution may not be the same as the previous one. The scale invariant version of these M-estimates equations:

$$\sum \psi\left(y_i - x_i^T \beta / s\right) x_i^T = 0 \,. \tag{6}$$

where‘s’ is a robust estimate of scale and the commonly used scale estimate is the median absolute deviation:

$$s = Med\left|u_i - Med\left(u_i\right)\right| / 0.6745 \,. \tag{7}$$

u_i denotes the residuals of initial fit and the value 0.6745 is an approximately unbiased estimator for large sample from normal distribution.

The weight function is defined $w(u) = \psi(u)/u$, $u = y_i - x_i^T \beta / s$ and let the weights $w_i = w(u_i)$, and then the estimating equations may be written as:

$$\sum w_i \left(y_i - x_i^T \beta\right) x_i^T = 0 \,. \tag{8}$$

Now solving these estimating equations for M-estimators is a weighted least-square problem that is minimizing $\sum w_i^2 u_i^2$. However, the weights depend upon the residuals, the residual depend upon the estimated coefficients, and the estimated coefficients depend upon the weight. Therefore, a convenient computational scheme is the iteratively reweighted least-squares (IRLS) method proposed by Holland and Welsch (1977). The scheme consists of the following steps:

1. Choose initial estimate $\beta^{(0)}$, such as least squares or regression by medians and calculate residuals and the scale estimate s.
2. Compute residuals $u_i^{(t-1)}$ at each iteration and associated weights $w_i^{(t-1)} = w\left(u_i^{(t-1)}\right)$ from the previous iteration.
3. Compute the new weighted least square estimates,

$$\beta^{(t)} = \left(X^T W^{(t-1)} X\right)^{-1} X^T W^{(t-1)} Y$$

This iteration procedure is continued until the estimated coefficients converge.

Several M-estimators have been proposed: the familiar least-square estimator (OLS); least absolute residuals (LAR); the Huber estimator; the Tukey bisquare (or biweight) estimator; the Andrews' wave (or sine) estimator and some other recently developed re-descending M-estimators. The objective functions, and the corresponding ψ and weight functions for these M-estimators are given in Table1. Both the least square and Huber objective functions increase without bound as residual 'u' departs from 0, but the least square objective function increases more rapidly. Least-squares assign equal weight to each observation while Huber estimator has a monotone ψ-function, and does not weight large residuals as heavily as least-squares. Huber estimator is the compromise between least-squares and least absolute residuals estimators.

The Andrews' wave estimator, Tukey's biweight estimator, Qadir objective function, Asad function and Insha's function belongs to the class of re-descending M-estimators because their ψ-functions equal zero for sufficiently large |*u*| that is observations having large residuals will receive zero weights.

Table 1
Objective Functions, and the Corresponding ψ and Weight Functions for Various M-Estimators

Method	Objective-Function ρ(u)	ψ –function ρ/(u)	Weight-function w(u)	Range of u
LS	$\frac{1}{2}u^2$	U	1	$\|u\| < \infty$
LAR	|u|	sign(u)	sign(u)/u	$\|u\| < \infty$
Huber	$\frac{1}{2}u^2$ $c\|u\|-\frac{1}{2}c^2$	U c sign(u)	1 c/|u|	$\|u\| \le c$ $\|u\| > c$
Bisquare	$c^2/6[1-\{1-(u/c)^2\}^3]$ $c^2/6$	$u[1-(u/c)^2]^2$ 0	$[1-(u/c)^2]^2$ 0	$\|u\| \le c$ $\|u\| > c$
Andrews	$c^2[1-\cos(u/c)]$ $2c^2$	c sin(u/c) 0	sin(u/c)/(u/c) 0	$\|u\| \le c\pi$ $\|u\| > c\pi$
Qadir	$u^2/96c^4(3c^4-3c^2u^2+u^4)$ $c^2/96$	$u/16c^4(c+u)^2(c-u)^2$ 0	$1/16c^4(c+u)^2(c-u)^2$ 0	$\|u\| \le c$ $\|u\| > c$
Asad	$u^2/45c^8(3u^8-10c^4u^4+15c^8)$ $8c^2/45$	$2u/3[1-(u/c)^4]^2$ 0	$2/3[1-(u/c)^4]^2$ 0	$\|u\| \le c$ $\|u\| > c$
Insha	$c^2/4[\text{Arc tan}(u/c)^2$ $+ c^2u^2/c^4+u^4]$	$u[1+(u/c)^4]^{-2}$	$[1+(u/c)^4]^{-2}$	$\|u\| \ge 0$

The constant '*c*' in these re-descending M-estimators is usually called tuning constant and determine the properties of the associated estimators (such as efficiency, influence function, and gross-error sensitivity). Smaller values of '*c*' produce more resistance to outlier, but at the expense of lower efficiency when the errors are normally distributed.

The choice of the tuning constant is generally picked arbitrarily and largely is the matter of personal preference. Several authors suggested different value of tuning constant *c* for various M estimators. For Wave function Andrew (1974) uses *c*=1.5, Gross (1976) suggest *c*=1.8 & 2.4, Hogg (1979) uses values of *c*=1.5, 2.0 and Rey (1983) suggests *c*=1.3387.

We applied the adaptive K-fold cross-validation procedure for choosing best tuning constant in these re-descending M-estimators to give reasonably high efficiency.

2. RESAMPLING CHOICE OF TUNING CONSTANT IN ROBUST EGRESSION

The robust regression estimators of re-descending M-estimations involve the use of ψ –function which replaces the derivative of the square function of the least-squares estimator. This ψ –function is not completely specified and needs the choice of tuning constant. Kelly (1996) reports the results of simulation study to investigate tuning constant c which minimize the jackknife asymptotic mean-squared error of the estimators.

We propose a resampling strategy of applying K-fold cross-validation method to choose that value of tuning constant that minimizes the Median Predicted Residual (MedPR).

Let $\hat{\beta}_{(i)}$ be a robust estimate of β obtained from the data with the i^{th} group of data eliminated. Then the best value of tuning constant {c = 1.1 (0.1)3.0(0.2)5.0} is when;

$$\text{Minimum (MedPR)} = Med\left|Y_i - \hat{Y}_{(i)}\right| = Med\left|Y_i - X_i^T\hat{\beta}_{(i)}\right| \quad (9)$$

where Med. stands for median, $X_i^T\hat{\beta}_{(i)}$ is the predicted value of observation(s) Y_i deleted from the data set. The strategy is to choose that value of '*c*' which minimizes the median predicted residual instead of PRESS in linear and generalized regression.

Solving the estimating equations for M-estimators, we start using iteratively re-weighted least-squares (IRLS) to obtain a robust fit. To initiate IRLS, we need a resistant fit (that is least-squares residuals, LSR or least absolute residuals, LAR) to compute the preliminary fit. However, these two methods do not protect against high-leverage observations and a robust regression might have difficulty in recovering from poor preliminary fit. Therefore, we need a suitable resistant fit to arrive at a good robust fit. Andrews (1974) developed a robust method to provide preliminary fit called as regression by medians. One will prefer this method as a source of initial estimates and residuals as it converge in fever iteration and suffer less computational cost.

For simple regression the review of Andrews' procedure can be described as follows:

1. Arrange the X-values of the data in ascending order.
2. Remove a certain number say np_1 of the smallest and largest X-values.
3. Remove further a certain number say np_2 of X-values immediately above and below the median.
4. Compute the medians (say $Med.X_L$ and $Med.X_H$) of the two remaining subsets corresponding to the lower and higher values of X.
5. Also compute the medians (say $Med.Y_L$ and $Med.Y_H$) for the corresponding Y-values.
6. The slope of the fit is then computed as:

$$\hat{\beta} = (Med.Y_H - Med.Y_L) / (Med.X_H - Med.X_L).$$

This estimator has a high breakdown point because half of the data on either subset can determine the fitted line. We use (np1+np2) ≈ 25% of the total number of observations.

Andrew generalized this procedure to a multiple regression case by applying a sweep operator to predictors successively and then to the outcome variable. Suppose we consider a multiple regression model with three predictor variables, X_1, X_2 and X_3, then the procedure can be summarized as follows:

The predictor variable X_1 is used to modify the variables X_2, X_3 and Y by sweeping X_1 out of these variables respectively:

$$X_{2.1} = X_2 - \hat{\beta}_1 X_1$$

$$X_{3.1} = X_3 - \hat{\beta}_2 X_1$$

$$Y_{.1} = Y - \hat{\beta}_3 X_1$$

Then the predictor variable X_2 is used to modify the variables X_3 and Y by sweeping X_2 out of these variables respectively:

$$X_{3.12} = X_{3.1} - \hat{\beta}_4 X_{2.1}$$

$$Y_{.12} = Y_{.1} - \hat{\beta}_5 X_{2.1}$$

And the predictor variable X_3 is used to modify the variable Y by sweeping X_3 out of Y:

$$Y_{.123} = Y_{.12} - \hat{\beta}_6 X_{3.12}$$

The subscript after a dot shows list of integers indicating that these predictor variables have been swept out and β's are the estimates of slope. The procedure proceeds for more predictor variables in the regression models in similar fashion. The procedure may be iterated and the total number of iteration is (p/2+2), where p is the number of predictor variables. At each iteration, the procedure is applied to a set of adjusted variables obtained at a previous iteration. The above procedure is used to compute a set of initial residuals and initiate the iterative procedure for robust fit.

K-fold cross-validation is then used to choose the tuning constant c which minimizes the median of the absolute prediction. The selected value of c is the best choice to compute robust estimates from the full data-set.

The Algorithm:

Our proposed algorithm consists of the following steps.

1. Generate or read input data for m-rows (observations) and n = p+1 columns (predictors and a response variable).
2. Initiate tuning constant $c = 1.1$ and its increment $\Delta c = 0.01$
3. Run cross-validation procedure by removing a group of observations.
4. To initiate IRLS, run least-squares method or least absolute residuals or Andrews' method of regression by medians to compute the preliminary fit. Use the residuals from this preliminary fit to compute scale estimate, s. Our algorithm runs Andrews' method of regression by medians for initial fit.
5. Run any weight function of re-descending M-estimators and compute weights W_i for residuals. Also calculate the weighted data of variables, their sum, means, and the weighted sum of squares and cross-products by matrix X^TWX, W is the diagonal matrix of order (m ─ 1) * (m ─ 1) and X is a matrix of order (m ─ 1 * n. Then compute the weighted residual sum-of-squares and cross-products from the above quantities and normalize to simple correlation matrix.
6. Compute standard regression coefficients, transfer back into original units and calculate residuals. Then compute new weights while using the calculated residuals.
7. Return to step 5 and repeat the whole procedure. This iteration procedure is continued until the estimated coefficients converge. The procedure terminates once the maximum change in the coefficients from one step to another is less than 0.1% or the number of iteration exceed 20. Fit the resultant robust estimate and compute the predicted residuals.
8. Return to step 4 and repeat the whole procedure by deleting instead group 2, group 3, and so on until all groups have been deleted once. Compute the median of the absolute values of the predicted residual.
9. Return to step 3 and increment the tuning constant c accordingly {c = 1.1 (0.1)3.0(0.2)5.0}. Repeat the whole procedure and choose the best value of c, the one that gives the smallest median predication error.
10. Use the selected value of tuning constant for the respective weight function and fit the robust regression for the complete data set.

(A computer program in R for this algorithm can be provided on demand from principle author).

3. BROWNLEE'S STACK LOSS PLANT DATA

This is an Operational data of a plant for the oxidation of ammonia to nitric acid. Stack Loss is a data frame with 21 observations on 4 variables. The Stack Loss data set has three predictor variables, X_1 (Flow of cooling air), X_2 (Cooling water inlet temperature), X_3 (Concentration of acid) and a response variable Y (Stack loss).Air Flow represents the rate of operation of the plant. Water temperature is the temperature of cooling water circulated through coils in the absorption tower. Acid concentration is the concentration of the acid circulating, minus 50, times 10: that is, 89 correspond to 58.9 percent acid. Stack loss (the dependent variable) is 10 times the percentage of the ingoing ammonia to the plant that escapes from the absorption column unabsorbed; that is, an (inverse) measure of the over-all efficiency of the plant, see Brownlee (1960)

We analyzed this data by applying K-fold cross-validation resampling technique in robust regression for choosing the appropriate tuning constant for various M-estimators. Table 2 represents the median prediction residuals for various tuning constant values resulting from K-fold cross-validation technique for Andrew's and Tukey's type M-estimators. Table 3 shows the summary of the resultant robust regression models and Table 4 represents the residuals for various fit.

Table 2
Cross-validated tuning constants and their corresponding Median Predicted Residuals (MedPR) for Andrews and Bisquare Estimators Resulting from K-fold Cross-validation

Tune. C	Andrews		Bisquare	
	Resistant Fit for Initial Residuals		Resistant Fit for Initial Residuals	
	Med	LS	Med	LS
1.10	.9263	1.9026	2.7654	2.0593
1.20	.9235	3.0018	2.8323	2.0438
1.30	1.2452	3.0577	2.9254	2.1454
1.40	1.3267	3.1049	2.9624	2.1047
1.50	1.3711	3.2163	2.5921	1.9431
1.60	1.4013	3.3016	2.6741	2.8433
1.70	1.4233	3.3849	2.3576	2.8016
1.80	1.4402	3.4854	2.1028	2.7649
1.90	1.4537	2.7804	2.0323	2.7367
2.00	1.4646	2.8314	1.9924	1.9582
2.10	1.4736	2.8557	1.9666	1.9635
2.20	1.4812	2.8696	1.9486	1.9428
2.30	1.4876	2.8809	1.9352	1.5745
2.40	1.4932	2.8905	1.9251	1.5696
2.50	1.4980	2.8985	1.9170	1.5650
2.60	1.5021	2.9053	1.9106	1.5608
2.70	1.5058	2.9112	1.9053	1.7241
2.80	1.5090	2.9162	1.9010	1.7475
2.90	1.5119	2.9207	1.8976	1.8614
3.00	1.5145	2.9246	1.8950	1.8569
3.20	1.5189	2.9389	.8704	1.8497
3.40	1.5224	2.9511	.8870	2.0459
3.60	1.5254	2.9613	.9003	2.0673
3.80	1.5279	2.9698	.9111	2.0897
4.00	1.5300	2.9770	.9200	2.8165
4.20	1.5318	2.9832	.9260	2.8919
4.40	1.5334	2.9885	.9245	2.9546
4.60	1.5347	2.9931	1.1909	3.5717
4.80	1.5359	2.9971	1.1820	3.6878
5.00	1.5369	3.0007	1.2594	3.7053

The procedure selects c = 1.20, 1.10 corresponding to the minimum Med PR = 0.9235, 1.9026 respectively when regression by median and least square method is used for initial fit in Andrews estimator. For Tukey's estimator the procedure selects c = 2.60, 3.120 corresponding to the minimum MedPR = 0.8704, 1.5608 when regression by median and least square method is used for initial fit. The resultant robust regression model for Andrews procedure on the basis of selected c = 1.20 is exactly the same equation found by Andrews (1974) using c = 1.5 and different value of np_1 and np_2 for obtaining the initial fit. We used $np_1 = 3$ and $np_2 = 3$ (total of which is approximately 30%

of the full data set) to obtain the initial fit using method of regression by medians. Our program then uses the selected tuning constant values on the whole data set and fit the robust regression model. It is clear from the residuals in Table 4 that four observations 1, 3, 4 and 21 have large residuals receiving zero weights by each M-estimator we described.

For the purpose of comparison, we have computed the least square estimator for both full and reduced data sets as shown in Table 3. For reduced data set that is without observations 1, 3, 4 and 21 a least square fit is,

$$\hat{Y} = -37.6525 + 0.7977X_1 + 0.5773X_2 - 0.0671X_3 \quad (10)$$

Comparing model coefficients in Table 3, we see that all the robust regression model coefficients has close agreement with the coefficients of least square fit without unusual observations.

Table 3
Robust Regression Models for Various M-Estimators using the Selected Tuning Constant by our Procedure to the Stack-Loss Data

M-estimator	Initial Fit	Model Coefficients				Tuning Constant
		Intercept	X1	X2	X3	
Andrew's	Med	-37.1590	0.8172	0.5225	-0.0723	1.20
	LS	-37.4178	0.7548	0.7654	-0.0849	1.10
Bisquare	Med	-37.0661	0.8208	0.5127	-0.0733	3.00
	LS	-36.7170	0.8490	0.4336	-0.0764	2.60
Qadir	Med	-37.2153	0.8148	0.5291	-0.0717	3.40
	LS	-36.7170	0.8491	0.4336	-0.0764	2.60
Asad	Med	-37.4913	0.8023	0.5645	-0.0689	3.20
	LS	-35.6166	0.8464	0.4434	-0.0897	1.80
LS(full data)	-	-39.9197	0.7156	1.2953	-0.1521	-
LS(reduced data)	-	-37.6525	0.7977	0.5773	-0.0671	-

Residuals for various M-estimators, using the selected tuning constants, are given in Table 4. It is clear from the residuals of these robust regression models that by using regression by median as an initial fit, cases 1, 3, 4 and 21 are identified as unusual points by having large residuals (see columns 2, 3, 4 & 5 of Table 4).

Table 4
Residuals for Various M-Estimators using the Selected Tuning Constant by our Procedure to the Stack-Loss Data

Cases	Residuals							
	M-Estimators (Initial fit by Med)				M-Estimators (Initial fit by LS)			
	Andrew	Tukey	Qadir	Asad	Andrew	Tukey	Qadir	Asad
1	6.1062	6.0831	6.1216	6.1912	5.9258	5.8841	5.8841	5.9187
2	1.0339	1.0098	1.0499	1.1223	0.8409	0.8077	0.8077	0.8289
3	6.3095	6.2857	6.3255	6.4008	6.3158	6.0729	6.0729	6.1272
4	8.2388	8.2487	8.2321	8.1892	7.6391	8.3152	8.3152	8.3048
5	-0.7161	-0.7259	-0.7096	-0.6818	-0.8300	-0.8177	-0.8177	-0.8085
6	-1.2386	-1.2386	-1.2388	-1.2463	-1.5954	-1.2512	-1.2512	-1.2519
7	-0.3276	-0.3116	-0.3380	-0.3976	-0.8511	-0.2264	-0.2264	-0.1569
8	0.6724	0.6884	0.6620	0.6024	0.1488	0.7735	0.7735	0.8431
9	-0.9698	-0.9555	-0.9796	-1.0369	-1.5761	-0.8549	-0.8549	-0.8662
10	0.1370	0.0950	0.1646	0.3035	0.6564	-0.2218	-0.2218	-0.2775
11	0 .7874	0.7545	0.8095	0.9233	1.4209	0.4657	0.4657	0.5300
12	0 .2377	0.1939	0.2670	0.4190	1.1014	-0.1771	-0.1771	-0.1164
13	-2.7184	-2.7584	-2.6921	-2.5587	-2.1737	-3.0691	-3.0691	-3.0981
14	-1.4461	-1.4650	-1.4331	-1.3657	-1.0047	-1.6623	-1.6623	-1.5544
15	1.3250	1.3208	1.3278	1.3421	1.4596	1.2582	1.2582	1.3013
16	0.1082	0.1009	0.1129	0.1355	0.2047	0.0290	0.0290	0.0321
17	-0.4261	-0.4377	-.4194	-0.3932	-0.7501	-0.4740	-0.4740	-0.6674
18	0.0798	0.0753	0.0822	0.0889	-0.1554	0.0607	0.0607	-0.0393
19	0.6295	0.6359	0.6247	0.5933	0.1641	0.7035	0.7035	0.6071
20	1.8709	1.8577	1.8792	1.9169	1.8050	1.7619	1.7619	1.7080
21	-8.9195	-8.9737	-8.8831	-8.6961	-7.9979	-9.4375	-9.4375	-9.3342

Thus the robust regression procedure with the K-fold cross-validated choice of tuning constant successfully and efficiently identified these four unusual points by leaving their residuals much larger and hence receiving zero weights. Starting with the ordinary least square as an initial fit, Andrews estimator shows large residuals (>3) for the four unusual points (see column 6 of Table 4) while Tukey's, Qadir and Asad estimators have large residuals for cases 1, 3, 4, 13 and 21 see columns 7, 8 & 9 of Table 4) but we found cases 1, 3, 4 and 21 are receiving zero weights while case 13 does not (that is wt> 0.0 for case 13).

THE SIMULATION EXPERIMENT

We observed the proposed methodology and its performance by a design experiment approach using Monte Carlo simulation. In this study, the data sets consist of n = 40 observations and p = 5 parameters as extensively used in Shao (1993, 1996) and Wisnowski et al. (2003). The response variable is generated as $Y = Z\beta + \varepsilon$, where $Z \overset{iid}{\sim} N(0,1)_{40}$, β is the vector of the known parameters [2, 3, 6, 0, 0] and $\varepsilon \overset{iid}{\sim} N(0,1)$. A value of 10 and/or 15 is added to create outliers for the last 4 or 8 observations. The data sets thus contain 10% and 20% residuals outliers at a distance of 10σ and /or 15σ.

The results in Table 6 are the probabilities of the K-fold cross-validated choice of tuning constant for successfully and efficiently identifying the unusual points by leaving their residuals much larger and hence receiving zero weights. The probability that each time the real unusual points were correctly identified out of 500 replicates, while using regression by median or OLS as initial fit of various M-estimators, were recorded. Note that very little additional information is gained if the number of replicates is increased or decreased. It is obvious from these results that the strategy of choosing the optimum tuning constant for M-estimators and thereafter fitting robust regression model is an efficient and adaptive resampling method.

Table 6
Probabilities of the K-Fold Cross-Validated Choice of Tuning Constant for Successfully and Efficiently Identifying the Unusual Points

M-Estimator	Initial Fit	% Outliers							
		10%				20%			
		Outlier Distance							
		10 15				10 15			
		CV Size							
		2	5	2	5	2	5	2	5
Andrew's	Med	0.942	0.965	0.956	0.943	0.967	0.944	0.943	0.916
	LS	0.916	0.903	0.944	0.941	0.924	0.915	0.903	0.911
Bisquare	Med	0.991	0.931	0.943	0.949	0.945	0.916	0.895	0.917
	LS	0.799	0.793	0.855	0.874	0.756	0.715	0.773	0.711
Qadir	Med	0.884	0.864	0.894	0.873	0.853	0.864	0.841	0.861
	LS	0.765	0.721	0.721	0.783	0.715	0.705	0.702	0.693
Asad	Med	0.890	0.854	0.895	0.867	0.815	0.839	0.810	0.821
	LS	0.711	0.694	0.784	0.756	0.715	0.695	0.702	0.685

CONCLUSION

Linear least square estimates can behave badly for contaminated data sets. One remedy is to detect and remove unusual observations from the least square fit. Another approach, termed robust regression, designed to reduce the impact of unusual observations by reducing the weights given to large residuals. This can be done by the most common general method of robust egression called M-estimations. Several M-estimators have been proposed (see Table 1), each require the analyst to specify some tuning constant. In practice, the choice of the tuning constant for these robust regressions is somewhat arbitrary and do not adapt to each particular data set. We have applied the efficient K-fold cross-validation in a robust way by computing the Median Predicted Residual (MedPR) to choose tuning constant for each data set.

We have analyzed the well-known Brownlee's Stack Loss Plant Data by applying K-fold resampling technique to choose the tuning constant and compared the robust regression model with least square model when unusual observations are removed. We find that K-fold cross-validation in robust regression has led to an appropriate value of tuning constant which deletes the unusual points by assigning them zero weights. Comparing model coefficients in Table 3, we find that all the robust regression estimates has close agreement with the coefficients of least square fit when unusual observations are deleted from the data set.

We demonstrate that, least-squares residuals (LSR) and regression by medians both can be used as a resistant fit to initiate IRLS in robust estimation. However, to arrive at a good robust fit one should prefer the later method as a source of initial estimates as it converges in fever iteration, suffer less computational cost and initiate the iterative procedure for better robust fit. The simulation experiment clearly reveals that decent results are possible with the resampling method using alternative estimate of prediction error in robust regression. We see that the proposed resampling criterion often outperform the minimum prediction error criterion in contaminated data sets.

ACKNOWLEDGEMENT

We really appreciate the referees efforts and thankful for their valuable suggestions, correction and recommendations that led the paper to improvement.

REFERENCES

1. Andrews, D.F. (1974). A Robust Method for Multiple Linear Regressions. *Technimetrics*, 16, 523-531.
2. Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H. and Tukey, J.W. (1972). Robust Estimation of Location: Survey and Advances. Princeton, New Jersey: Princeton University Press.
3. Book, D., Booker, J., Hartley, H.O. and Sielken, R.I. (1980). Unbiased L1 Estimators and Their Covariance. *ONR THEMIS Technical Report No. 64*, Institute of Statistics, Texas A & M University.
4. Brownlee, K.A. (1960).Statistical Theory and Methodology in Science and Engineering. New York: Wiley. (2nd ed. 1965) 491-500.

5. Draper, N.R. and Smith, H. (1998). Applied Regression Analysis (3rd ed.), *John Wiley*, New York.
6. Dodge, Y. (1987). *Statistical Data Analysis Based on the L_1-Norm and Related Methods*. North-Holland, Amsterdam.
7. Gross. A.M. (1976). Confidence Interval Robustness with Long-Tailed Symmetric Distributions. *J. Amer. Statist. Assoc.*, 71, 409-416.
8. Hogg, R.V. (1979). Statistical Robustness: One View of Its Use in Application Today. *The American Statisticians*, 33, 108-115.
9. Huber, P.J. (1981).*Robust Statistics*. New York, Wiley.
10. Holland, P.J. and Welsch, R.E. (1977). Robust Regression Using Iteratively Reweighted Least Square. *Communications in Statistics*, A6, 813-827.
11. Ho, K. and Naugher, J. (2000). Outliers Lie: An illustrative Example of Identifying Outliers and Applying Robust Models. *Multiple Linear Regression Viewpoints*, 26(2), 2-6.
12. Kelly, G.E. (1992). Robust Regression Estimators-The Choice of Tuning Constants. *The Statistician*, 41, 303-314.
13. Kelly, G. (1996). Adaptive Choice of Tuning Constant for Robust Regression Estimators. *The Statisticians*, 45, 35-40.
14. Narula, S.C. and Wellington, J.F. (1982). The Minimum Sum of Absolute Error Regression: A State of the Art Survey. International Statistical Review, 50, 317-326.
15. Salahuddin (1990). Cross-Validation in Regression Analysis. Ph.D. Thesis, University of Wales, Swansea.
16. Shao, J. (1993). Linear Model Selection by Cross-Validation. *J. Amer. Statist.Assoc*., 88, 486-494.
17. Shao, J. (1996). Bootstrap Model Selection. *J. Amer. Statist. Assoc*., 91, 655-665.
18. Rey, W.J.J. (1983).*Introduction to Robust and Quasi-Robust Statistical Methods*. Belin, Springer-Verlag.
19. Rousseeuw, P.J. and Leroy, A. (1987).*Robust Regression and Outlier Detection*. New York, John Wiley.
20. Rousseeuw, P.J. (1984).Least Median of Square Regression. *J. Amer. Statist. Assoc.*, 79, 871-880.
21. Rousseeuw, P.J. and Yohai, V.J. (1984). Robust Regression by Means of S-Estimators: in Franke, J., Hardle, W. and Martin, R.D. (Eds.), Robust and Nonlinear Time Series Analysis: lecture Notes in Statistics. Springe, Berlin, 26, 256-272.
22. Wisnoski, J.W., Simpson, J.R., Montgomery, D.C. and Runger, G.C. (2003). Resampling Method for Variable Selection in Robust Regression. *Compt. Statist. Data Analysis*, 43, 341-355.
23. Yohai, V.J. (1974). Robust Estimation in the Linear Model. Ann. Statist, 2, 562-567.
24. Yohai, V.J. (1987). High Breakdown-Point and High Efficiency Robust Estimates for Regression. *Ann. Statist.*, 15(2), 642-656.