

**A NOVEL RESAMPLING METHOD FOR VARIABLE
SELECTION IN ROBUST REGRESSION**

Zafar Mahmood¹ 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

Variable selection in regression analysis is of vital importance for data analyst and researcher to fit the parsimonious regression model. With the inundation of large number of predictor variables and large data sets requiring analysis and empirical modeling, contamination becomes usual problem. Accordingly, robust regression estimators are designed to easily fit contaminated data sets. In the last three decades much work have been done regarding various robust regression methods to dealt the data sets contaminated with outliers, relatively less attentions was given to construct a best subset of the predictor variables in robust regression model. We initially considered cross-validation resampling technique working well for variable selection in linear regression models; see Zafar and Salahuddin (2009, 2011). It turned out that the usual prediction errors inflated by outlier are not the reliable measure for robust model selection. Ultimately, a novel resampling procedure is proposed by introducing alternative and robust prediction error based on Winsor principle in the contaminated model. We demonstrate that superior results for robust model selection are obtainable by relaxing the requirement for the absolute minimum Winsorized prediction error while using our proposed optimum choice of the tuning constant. The simulation study reveals that the proposed technique is working well.

1. INTRODUCTION

In many research areas such as biological, medical, public health, social, and agricultural sciences, there may be large number of predictor variables that could be used to predict a regression model. The important aspect of regression analysis is selecting the subset of significant predictor variables whose model fit the data fairly well. We need a search strategy and computational algorithms for fitting the best subset model. The aim is to fit the model that leads to accurate prediction with minimum variance of prediction. In regression problem for developing a concise and best model, different computer intensive variable selection procedures (including all possible regression, best subset regression model, forward selection, backward elimination and stepwise variable selection

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

procedures) have been commonly used. Selection criteria are often based upon R^2 , Adjusted R^2 , C_p Statistic and F-test Statistics etc. Unfortunately, Miller (1990) demonstrated that these methods are biased and not recommended for variable selection. Breiman (1995) suggested that some measure of prediction error resulting from the bootstrap and cross-validation resampling method should be preferred for variable selection. This article aims to cope with the variable selection problem for robust regression estimators designed to easily fit contaminated data sets due to potential outliers.

Many robust regression procedures have been proposed in the last 3 decades but little guidance is available for model selection in robust regression. Usually, robust regression variable selection methods are based on robust version of the general linear test that uses the asymptotic covariance matrix; see Hampel et al. (1986). Markatou and He (1994) and Hertier and Ronehetti (1994) suggested the Wald (similar to t-test) and drop-in-dispersion tests (similar to F-test) to Generalized M and compound estimators. Field (1997) and Field and Welsh (1998) proposed saddle point approximations of tail area probabilities for robust regression hypothesis testing as improvement to the asymptotic approach. Ronchetti and Staudte (1994) proposed a robust version of Mallows's C_p . Their method multiplies the squared residuals by the final weights from a robust regression and two additional quantities are also added to the residual sum of squares that are a function of the number of parameters and the selected robust estimator. The robust C_p seems working well for their three examples, but no simulation results are reported.

Davison and Hinkley (1997) discussed the applications of resampling techniques in robust regression. They suggested removing gross outliers from the analysis because too many outliers could appear in the resample data leading to inefficiency and breakdown and then applying any of the least squares prediction error method to robust regression. Wilcox (1998) proposed a bootstrap resampling scheme for variable selection in robust regression. He used a percentile bootstrap approach to find critical values for the joint confidence region on the Mahalanobis distance for the model parameters. Wilcox stated that there is a room for improvement with this method because the probability of Type-I error can be substantially less than nominal levels in many circumstances. He cautions that this approach does not work well with least squares and hence correction factors through simulation are required to achieve the right coverage probabilities.

Wisnowski et al. (2003) provide a new resampling variable selection method by introducing alternative estimates of prediction error. They proposed, relaxing the requirement for the absolute minimum prediction error and selecting a model with the fewest number of predictor variables and a low (not necessarily minimum) prediction error. They are not advocating using a specific percentage for the suggested change in prediction errors to be used for variable selection in robust regression. Muller and Welsh (2005, 2009) proposed a bootstrap criterion for robustly selecting linear and generalized linear models. They combined a robust penalized measure of fit to the sample with a robust measure of out of sample predictive ability that is estimated using a post-stratified m-out-of-n bootstrap. Salibian-Barrera and Van Aelst (2008) also considered a robust selection criteria based on fast bootstrap procedure.

It is clear from the above review that little guidance is available for variable selection in robust regression especially using robust cross validation estimates of prediction error in robust regression.

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. An attempt has been made to apply resampling variable selection method in robust regression with alternative estimates of prediction error based on Winsor's principle. Furthermore, our resampling optimum choice of the tuning constant for M-estimators and a robust initial resistance fit has been considered.

2. ROBUST REGRESSION, M-ESTIMATORS AND RESAMPLING PREDICTION ERROR

The term robust regression is to make use of a fitting criterion that is not as vulnerable as least squares to contaminated data with heavy tailed errors. Robust regressions are alternative procedures and have been developed to modify the least squares methods so that outliers have least influence on the final estimates. The most familiar common method of robust regression is M-estimation. The name M-estimators can be regarded as a generalization of Maximum Likelihood Estimation (MLE). 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:

$$\text{Minimize}_{\beta} \sum_{i=1}^n \rho(r_i) = \sum \rho(y_i - x_i^T \beta). \quad (1)$$

where, ρ is a symmetric function with a sole minimum at zero and gives the contribution of each residual to the objective function. The function ρ is linked to the likelihood function for a suitable choice of error distribution. The majority of the M-estimators can be solved with the help of iteratively reweighted least square method. The procedure can be found in the Birch (1997) and Simpson et al. (1998).

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

$$\sum \psi(y_i - x_i^T \beta) x_i^T = 0 \quad (2)$$

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:

$$\sum \psi \left(y_i - x_i^T \beta / s \right) x_i^T = 0. \quad (3)$$

Divide the ψ -function by residual which gives us the weight function that can be defined by $w(u) = \psi(u) / u$, $u = y_i - x_i^T \beta / s$ and let the weights $w_i = w(u_i)$, and then the estimating equations may be written as:

$$\sum w_i \left(y_i - x_i^T \beta \right) x_i^T = 0. \quad (4)$$

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

To apply IRLS, we require computing initial fit using least-squares residuals (LSR) or least absolute residuals (LAR). However, these two methods do not protect against high-leverage observations and a robust regression might have complexity in recovering from poor preliminary fit. Therefore, suitable preliminary resistant fit, regression by medians, suggested by Andrews (1974) can be preferred to arrive at a good parsimonious robust regression model. 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 as their ψ -functions equal zero for sufficiently large $|u|$ that is observations having large residuals will receive zero weights.

Cross-validation and bootstrapping are the resampling methods, presently opted as a measure of prediction error for variable selection in linear regression. The regression model is fit with training data set and consequently used to predict the response variable for assessment data set. Cross-validation is impetuously an attractive method to calculate the prediction error either by leaving one observation out called leave-one-out cross-validation (LOOCV) or a subset of observation out called K-fold cross-validation (KFCV) and then the response value(s) is predicted. The predicted sum of square is then calculated by;

$$PRESS = \sum \left(Y_i - \hat{Y}_{(i)} \right)^2 \quad (5)$$

Shao (1993) exposed with asymptotic simulations results that the regression model is usually over specified with the minimum LOOCV prediction error and recommended KFCV as a measure of prediction error. Davison and Hinkley (1997) recommended $K = \min(n/2, 10)$ in practice and Zafar and Salahuddin (2009) proposed a procedure by introducing alternative estimates of boosted prediction sum of squares to choose the appropriate tuning parameter that is K in K-fold cross-validation for each data set.

The KFCV procedure decreases the variance of prediction error over that of the LOOCV but at the expense of increased bias. Shao (1993) also suggested that the smaller

the training set, the better the K-fold model selection estimate and demonstrate that both the leave-one-out and K-fold cross-validation procedures have a slight probability of selecting an underspecified model for least-squares estimators. The challenge is avoiding an over-fitted model.

Bootstrap resampling produce many samples from the original data set using sampling with replacement and for the variable selection problem, the estimate of the average prediction error for the Bth bootstrap sample. Zafar and Salahuddin (2009) used the sum all the partial square of predicted values for each bootstrap sample and computed the average of the boosted PRESS;

$$PRESS_{boot} = \frac{\sum_{b=1}^B \left[\sum_{i=1}^n \left(Y_i - \hat{Y}_{(i)} \right)^2 \right]}{B} \tag{6}$$

A Proposed Variable Selection Criterion in Robust Regression

There are many variable selection methods in regression analysis to select best possible subset regression model. The most common are the automated model selection methods (forward, backward, stepwise), subset regression and all possible regression. These variable selection methods are based on R², Adjusted-R², F-test statistics (F-to-enter and F-to-remove), AIC, Akaike, Mallows-Cp, and BIC criteria, see Schwarz (1978). But these methods based on least square regression parameters lose power in the presence of unusual points. Breiman (1995) preferred some measure of prediction error for variable selection in regression based on resampling methods of cross validation and bootstrapping. Cross validation and bootstrap methods are well reputed in least-square variables selection but cannot be directly applied to contaminated data sets using robust regression models. Wisnowski et al. (2003) proposed, relaxing the requirement for the absolute minimum prediction error and selecting a model with the fewest number of predictor variables and a low (not necessarily minimum) prediction error.

We applied a new resampling variable selection method by introducing alternative estimates of prediction error based on Winsor’s principle.

The cross-validation scheme for choosing subset predictor variables in robust regression is as follows:

Let $\hat{\beta}_{(i)}$ be a robust estimate of parameters obtained from data with the ith observation deleted, then a subset predictor variable(s) in robust regression is a good choice that corresponds to minimum occurrence of Winsorized Prediction Error (WPE) computed as the average of Winsorized Prediction Sum of Squares (PRESSWinsor).

$$PRESS_{Winsor} = \sum \left[\left(Y_i - \hat{Y}_{(i)} \right)_{Winsor}^2 \right]$$

$$WPE = PRESS_{Winsor}/n \tag{7}$$

The idea is to choose the subset model that produces the near minimum Winsorized PRESS and then a final robust model with the selected predictors variables are fitted for full data set. The suggested algorithm is given below:

1. Read input data consisting of n -rows (data points) and m -columns (predictors and a response variable).
2. Apply the Cross-validation to choose appropriate tuning constant value for any M-estimation robust regression.
3. Initialize cross-validation (LOOCV) by deleting first case or K-fold cross validation (KFCV) by deleting first group of observation.
4. Run a robust regression for a subset of predictor variables using the selected tuning constant value and compute the predicted residual for the deleted case.
5. Return to step 3 and repeat the whole procedure by deleting instead case 2, case 3 and so on or group 2, group 3 and so on until all the observations has been deleted once and calculate Winsorized predicted residuals sum of squares (PRESSWinsor).
6. Compute the Winsorized Predicted Error (WPE).
7. Return to step 3 and repeat the procedure for all possible subset of predictor variables.
8. Select the subset of predictor variables that produces the minimum WPE. For superior results we further suggest relaxing the requirement for strict minimum WPE and selecting a model with a fewest number of predictor variables corresponding to near minimum WPE. A reasonable strategy involves observing a line connected scatter plot (screeplot), the subset model with the fewest number of predictor variables where the screeplot levels off is selected.
9. Use the selected subset of predictor variables and compute the robust estimates 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 the first three predictor variables of the data frame, X1 (Flow of cooling air), X2 (Cooling water inlet temperature), X3 (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 first applying the efficient K-fold cross-validation resampling technique in robust regression for selecting the appropriate tuning constant for Tukey’s bi-square M-estimators. Table 1 shows the summary of the resultant robust regression models using the selected tuning constant.

Table 1
Robust Regression Models for Tukey’s Bi-Square M-Estimators
Using the Selected Tuning Constant for Stack-Loss Data

M-estimator	Initial Fit	Model Coefficients				Tuning Constant
		Intercept	X1	X2	X3	
Bisquare M-estimator	Med	-37.0661	0.8208	0.5127	-0.0733	3.00
	LS	-36.7170	0.8490	0.4336	-0.0764	2.60
LS(full data)	-	-39.9197	0.7156	1.2953	-0.1521	-
LS(reduced data)	-	-37.6525	0.7977	0.5773	-0.0671	-

Evaluating model coefficients in Table 1, it is obvious that the robust regression model coefficients have close concurrence with the coefficients of least square fit without unusual observations.

Table 2
Cross-Validated Wensorized Prediction Error (WPE) as a Function of the
Number of Variable(s) in the Model using Tukey’s Bi-Square
Estimator for Stack Loss Data

Model Order	Model Parameters	WPE	
		LOOCV	KFCV
1	β_0	140.670	149.916
2	β_1	3.316	4.332
3	β_2	71.686	80.471
4	β_3	8.692	16.846
5	$\beta_1\beta_2$	1.014	3.193
6	$\beta_1\beta_3$	5.333	5.988
7	$\beta_2\beta_3$	37.251	80.755
8	$\beta_1\beta_2\beta_3$	5.545	4.962

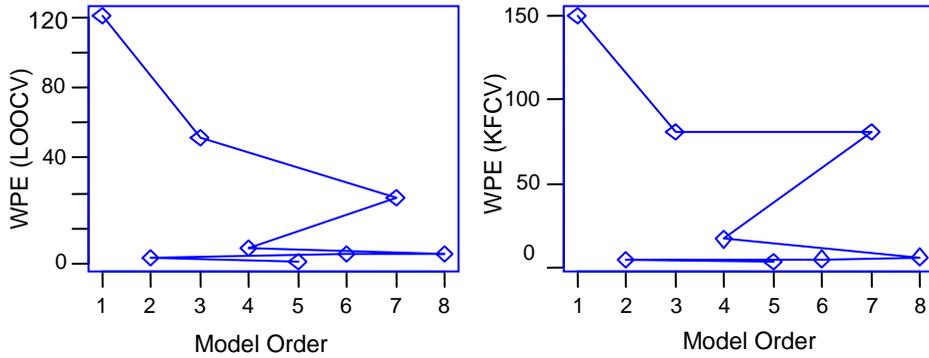


Figure 1: Screenshot of Winsorized Prediction Error

The procedure selects the subset model having predictor variables X_1 and X_2 that correspond to minimum Winsorized Prediction Error, WPE (=1.041 using LOOCV and =3.193 using KFCV). We believe that superior results for robust model selection are obtainable by relaxing the requirement for the absolute minimum Winsorized prediction error. The strategy is to observe a screeplot (line connected scatter plot of subset model order versus Winsorized prediction error as shown in Figure 1). The subset model with fewest number of predictor variable(s) that is X_1 corresponding to near minimum WPE, where the screeplot levels of, is selected. The final robust regression for full data is,

$$\hat{Y} = -39.88 + 0.95X_1 \quad (8)$$

For the reduced data set that is without observations 1, 3, 4 and 21 a least square fit for predictor variable X_1 in the model yields,

$$\hat{Y} = -40.03 + 0.95X_1 \quad (9)$$

Comparing equations 8 and 9, we see that these two models are quite close.

THE SIMULATION EXPERIMENT

For the better understanding of the proposed resampling method performance, we run 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 $Z \sim N(0,1)_{40}$, β is the vector of the known parameters [2, 3, 6, 0, 0] and $\varepsilon \sim N(0,1)$. 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 . From the previous knowledge and pilot studies, the following factors are included.

Percentage of Outliers: This important factor reveals the number of outliers in the data set. In our generated data sets, the outliers' density levels are 10% and 20%.

Outlying Distance: This factor shows how many standard deviations (SD) from the means make the outlier in residuals. Our generated data sets contain residuals outliers at distance of 10 standard deviations and 15 standard deviations.

Cross-validation Assessment Size: This factor is most important factor to correctly identify important predictor variables. The levels are 1 (leave-one-out cross-validation), 5 (K-fold cross-validation).

To illustrate the methodology, the Winsorized predicted error for models with increasing number of predictor variables are observed. The cross-validation design is a full factorial 23. The results in Table 3 are the proportions of times that each model is selected out of 350 replicates. Note that very little additional information is gained if the number of replicates is increased. The shaded column in Table 3 reports the proportion of times that each cross-validation method using various criteria is able to find the correct model. The most striking results are the general failure of the minimum prediction error criterion in all cases. It incorrectly selects the largest model ($p=5$) in most of the cases instead of selecting the correct model. The proposed minimum Winsorized prediction error criterion and that of near minimum Winsorized prediction error criterion significantly outperforms the minimum prediction error criterion in robust regression. Clearly, the best method is the near minimum Winsorized prediction error applied from the K-fold cross-validation.

Design experiment results for cross-validation methods using Winsorized residuals to compute prediction error in robust regression models are given in Table 3.

The top values in each cell of last four columns are the proportion of time that a model is selected out of 350 replications using minimum prediction error by least square method. The middle values are the proportion selected using minimum Winsorized prediction error in robust regression models and the bottom values are the proportion selected using near minimum Winsorized prediction error in robust regression models.

Table 3
Results for cross-validation methods, using Winsorized residuals,
to compute prediction error in robust regression models

% Outliers	Outlier Distance	CV Size	$\beta_0 - \beta_1$	$\beta_0 - \beta_2$	$\beta_0 - \beta_3$	$\beta_0 - \beta_4$
10	10	1	0.000	0.291	0.120	0.589
			0.000	0.623	0.106	0.271
			0.000	0.674	0.294	0.026
20	10	1	0.000	0.406	0.100	0.494
			0.000	0.529	0.274	0.194
			0.000	0.649	0.177	0.174
10	15	1	0.000	0.306	0.097	0.597
			0.000	0.497	0.229	0.274
			0.000	0.720	0.180	0.100
20	15	1	0.000	0.205	0.389	0.406
			0.000	0.520	0.089	0.391
			0.000	0.723	0.034	0.243
10	10	5	0.000	0.377	0.126	0.497
			0.000	0.723	0.100	0.177
			0.000	0.891	0.083	0.026
20	10	5	0.000	0.591	0.194	0.215
			0.000	0.631	0.363	0.005
			0.000	0.783	0.211	0.005
10	15	5	0.000	0.217	0.294	0.489
			0.000	0.637	0.094	0.269
			0.000	0.780	0.103	0.117
20	15	5	0.000	0.294	0.277	0.429
			0.000	0.623	0.106	0.271
			0.000	0.729	0.069	0.203

CONCLUSION

With the inundation of large data sets, unusual observations are common place. A good approach to deal with contaminated data is applying robust estimates instead of using least square estimates. In regression analysis, robust regression methods are designed to reduce the impact of unusual observations by reducing the weights given to large residuals. The study extended several methodologies for variable selection in robust regression with contaminated data sets. The proposed novel method select the parsimonious model while choosing appropriate tuning constant, initial robust resistance fit and near minimum alternative prediction error from a resampling procedure (leave-one-out cross-validation and K-fold cross-validation).

We have analyzed the illustrious Brownlee's Stack Loss Plant Data by applying K-fold resampling technique first to choose the tuning constant and then select subset

model in robust regressions. We find that K-fold cross-validation choice of tuning constant provides robust regression estimates that have close agreement with the least square coefficients when the outliers are deleted from the data set.

The study proposes a resampling variable selection strategy by introducing alternative estimates of prediction error based on Winsor's principle for contaminated data sets. We demonstrate that, although robust estimation and resampling techniques are computationally complex procedures, we can combine these procedures for better results. We find that a subset predictor variable(s) in robust regression is a good choice that corresponds to minimum occurrence of Winsorized Prediction Error (WPE) computed as the average of Winsorized Prediction Sum of Squares (PRESSWinsor). We further suggest that superior results are obtainable by relaxing the requirement for strict minimum WPE and selecting a model with a fewest number of predictor variables corresponding to near minimum WPE. A reasonable strategy involves observing a line connected scatter plot (screepplot), the subset model with the fewest number of predictor variables where the screeplot levels off is selected.

The simulation study 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. *Tachometrics*, 16, 523-531.
2. Birch, J.B. (1997). *Exploratory and Robust data analysis*, Unpublished lecture notes.
3. Breiman, L. (1995). Better Subset Regression Using the Nonnegative Garrote. *Tachometrics*, 37, 373-384.
4. Brownlee, K.A. (1960). *Statistical Theory and Methodology in Science and Engineering*. New York: Wiley. (2nd ed. 1965) 491-500.
5. Davison, A.C. and Hinkley, D.V. (1997). *Bootstrap Method and Their Application*. Cambridge University Press, Cambridge, UK.
6. Field, C.A. (1997). Robust Regression and Small Samples Confidence Intervals. *J. Statist. Plann. Inference*, 57, 39-48.
7. Field, C.A. and Welsh, A.H. (1998). Robust Regression Confidence Intervals for Regression Parameters. *Austral. N.Z.J. Statist.*, 40, 55-64.
8. Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York.
9. Hertier, S. and Ronchetti, E.M. (1994). Robust Bounded-Influence Tests in General Parametric Models. *J. Amer. Statist. Assoc.*, 89, 897-904.
10. Holland, P.J. and Welsch, R.E. (1977). Robust Regression Using Iteratively Reweighted Least Square. *Communications in Statistics*, A6, 813-827.

11. Huber, P.J. (1981). *Robust Statistics*. New York, Wiley.
12. Mahmood, Z. and Salahuddin (2009). On the Use of K-Fold Cross-Validation to Choose Cutoff Values and Assess the Performance of Predictive Models in Stepwise Regression. *The Internl. J. of Biostat.*, 5(1), Art. 25 Available at: <http://www.bepress.com/ijb/vol5/iss1/25>©2009 Berkeley Electronic Press.
13. Mahmood, Z. and Salahuddin (2011). A Cross-Validation Approach to Optimize Unequal Cutoff Values in Stepwise Regression. *Pak. J. Statist.*, 27(2), 197-211.
14. Markatou, M. and He, X., (1994). Bounded Influence and High Breakdown Point Testing Procedures in Linear Models. *J. Amer. Statist. Assoc.*, 89, 543-549.
15. Miller, A.J. (1990). *Subset Selection in Regression*. Chapman & Hall, London, 1990
16. Muller, S. and Welsh, A.H. (2005). Outlier Robust Model Selection in Linear Regression. *J. Amer. Statist. Assoc.*, 100, 1297-1310.
17. Ronchetti, E. and Staudte, R.G. (1994). A Robust Version of Mallows's Cp. *J. Amer. Statist. Assoc.*, 89, 550-559.
18. Rousseeuw, P.J. (1984). Least Median of Square Regression. *J. Amer. Statist. Assoc.*, 79, 871-880.
19. Rousseeuw, P.J. and Leroy, A. (1987). *Robust Regression and Outlier Detection*. New York, John Wiley.
20. 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*. Springer, Berlin, 26, 256-272.
21. Salibian-Barrera and Van Aelst. (2008). Robust Model Selection Using Fast and Robust Bootstrap. *Computational Statistics and Data Analysis*, 52, 5121-5135.
22. Shao, J. (1993). Linear Model Selection by Cross-Validation. *J. Amer. Statist. Assoc.*, 88, 486-494.
23. Shao, J. (1996). Bootstrap Model Selection. *J. Amer. Statist. Assoc.*, 91, 655-665.
24. Simpson, D.G., He, X. and Liu, Y.T. (1998). Comment on Edge-Preserving Smoothers by Chu, Glad, Godtliebsen, and Marron. *J. Amer. Statist. Assoc.*, 93(442), 544-548.
25. Wilcox, R.R. (1998). The goals and strategies of robust methods. *British J. Math. Statist. Psychol.*, 51, 1-39.
26. 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.
27. Yohai, V.J. (1987). High Breakdown-Point and High Efficiency Robust Estimates for Regression. *Ann. Statist.*, 15(2), 642-656.