Hacettepe Journal of Mathematics & Statistics

Hacet. J. Math. Stat.
Volume 50 (6) (2021), 1773 – 1792
DOI : 10.15672/hujms.842589

Research Article

High leverage points and vertical outliers resistant model selection in regression

Kundalik S. Shende*, Dattatraya N. Kashid

Department of Statistics, Shivaji University, Kolhapur, India (MS)

Abstract

It is necessary to consider only relevant predictor variables for prediction purpose because irrelevant predictors in the regression model will tend to misleading inference. There are so many model selection methods available in the literature; among these, some methods are resistant to vertical outliers, but still, the problem of the presence of vertical outliers and leverage points is not well studied. In this article, we have modified the S_p statistic using the generalized M-estimator for robust model selection in the presence of vertical outliers and high leverage points. The proposed model selection criterion selects only relevant predictor variables by probability one for a large sample size. We found the equivalence of this criterion and the existing C_p and S_p criteria. The superiority of a proposed criterion is demonstrated using simulated and real data.

Mathematics Subject Classification (2020). 62J05, 62F35

Keywords. GM-estimator, adaptive S_p statistic, S_p statistic, C_p statistic, consistency, robust model selection

1. Introduction

The prime intention behind the use of a regression model is to predict the unknown response variable for the given values of predictor variables. The prediction of the response variable depends on the predictor variables in the model. The relevant predictor variables in the model give accurate predictions. Model selection methods consider only relevant predictor variables. The general form of a multiple linear regression model is

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

where y is a vector of n observations on a response variable, $\beta = (\beta_0, \beta_1, \beta_2, ..., \beta_{k-1})'$ is a vector of unknown k regression coefficients, X is a matrix of size $n \times k$ of observations on $(k-1)$ predictor variables $X_1, X_2, X_3, ..., X_{(k-1)}$ with 1's in the first column and ε is a vector of errors with $E(\varepsilon) = 0$ and $Var(\varepsilon) = \sigma^2 I_n$. The full model (1.1) can be written as

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon, \tag{1.2}$$

where X and β are partitioned as $X = [X_1 : X_2]$ and $\beta' = [\beta_1' : \beta_2']$. X_1 is $n \times p$ matrix of observations on $(p-1)$ predictor variables with 1's in the first column and β_1 is a vector

*Corresponding Author.
Email addresses: shendeks.stat99@gmail.com (K.S. Shende), dnk_stats@unishivaji.ac.in (D.N. Kashid)
Received: 18.12.2020; Accepted: 16.07.2021

of corresponding unknown regression coefficients. Similarly, X_2 is $n \times (k-p)$ matrix of observations on remaining predictor variables and β_2 is a vector of corresponding unknown regression coefficients. Now, consider a submodel based on $p-1$ ($p < k$) predictor variables,

$$y = X_1\beta_1 + \varepsilon. \tag{1.3}$$

We can select an appropriate model by testing the null hypothesis $H_0 : \beta_2 = 0$ or using criterion function available in the literature. Many model selection criteria have been proposed based on the classical least squares (LS) estimator. Among these, Mallows's C_p [22] is a well-known model selection technique. It is defined as

$$C_p = \frac{RSS_p}{\sigma^2} - (n - 2p), \tag{1.4}$$

where RSS_p is residual sum of squares of the submodel which has $(p-1)$ predictor variables. σ^2 is an unknown error variance, and it can be replaced by its suitable estimate like residual mean squares of the full model. The LS estimator is used to calculate RSS_p and residual mean squares in Mallows's C_p. It is well-known that the LS estimator is optimal and strongly efficient when assumptions of regression are satisfied. However, an outlier in the data destroys the LS estimator, and consequently, it is also affecting on Mallows's C_p and other LS estimator based methods.

Several robust estimators alternative to the LS estimator have been proposed in the literature to overcome this issue. The M-estimator [17] is prominent broad class of robust estimators, and it reduces the effect of outliers by assigning low weights to the outliers. Researchers have suggested robust model selection methods based on this M-estimator. Rao et al. [25] have reviewed many non-robust and robust model selection criteria. Ronchetti and Staudte [28] have proposed a robust version of Mallows's C_p based on the M-estimator. It is defined as

$$RC_p = \frac{W_p}{\hat{\sigma}^2} - (U_p - V_p), \tag{1.5}$$

where W_p is the weighted residual sum of squares of the submodel, U_p and V_p are constants depends on weight function and the number of parameters in the corresponding model, and $\hat{\sigma}^2$ is a robust and consistent estimate of error variance based on a full model. The RC_p criterion selects the model whose RC_p value is close to V_p. Further, Kashid and Kulkarni [18] proposed a more simple robust model selection S_p criterion based on the M-estimator. It is defined as

$$S_p = \frac{\|\hat{y}_k - \hat{y}_p\|^2}{\sigma^2} - (k - 2p), \tag{1.6}$$

where $\|\cdot\|$ represents L_2 norm, $\hat{y}_k$ and $\hat{y}_p$ are vector of predicted values of y based on full model and submodel respectively. The M-estimator is used to calculate these predicted values. An unknown error variance σ^2 is replaced by its suitable robust estimate. The S_p criterion selects the model whose S_p value is close to the number of unknown parameters (p) in the submodel.

Kim and Hwang [19] proposed a method based on Mallows's C_p called as $C_p(d)$ by deleting d outlying observations to select relevant predictor variables. A robust version of Akaike information criterion (RAIC) [27] and robust version of Bayesian information criterion (RBIC) [21] are also available in the literature to select a model in presence of outliers. Tharmaratnam and Claeskens [32] have compared the classical AIC criterion with a robust version of AIC based on different robust estimators in the presence of outliers. André et al. [2], Croux and Dehon [10], Maronna et al. [23], Renaud and Victoria-Feser [26] proposed different robust coefficient of determination based on robust estimators to identify appropriate predictor variables and assess the quality of the model. All the aforesaid robust model selection methods are robust to vertical outliers.

Generally, three types of outliers namely, vertical outliers, bad leverage points and good leverage points are considered in the regression analysis . Vertical outliers and bad leverage points are outlying only in Y-space and X-space respectively, and these types of outliers are located far away from the regression line. Good leverage points are outlying in both space, and they are found near to the regression line. The vertical outliers and bad leverage points are significantly affecting the estimated regression parameters, while good leverage points are not affecting the estimated regression parameters [11]. All these outlying points can be identified by using methods available in the literature [1,3,4,6,8,12,29,30].

The M-estimator is an one of the widely used robust estimator in regression, but it fails to account the high leverage point in the parameter estimation [33] and consequently in model selection methods. Almost all the above-cited methods are introduced to curb the effect of vertical outliers but are not performing well in the presence of leverage points. In response to this problem, Mallows and Schweppe [20,33] have suggested a generalized M (GM) estimator as an alternative to classical M-estimator. The GM-estimator $\hat{\beta}$ of β for a linear regression model is the solution of an equation,

$$\sum_{i=1}^{n} \eta \left(X_i, \frac{y_i - X_i'\beta}{\sigma} \right) X_i = 0. \tag{1.7}$$

In general, the function η can be represented as

$$\eta(X, r) = \omega(X)\psi\left(r\nu(X)\right),$$

where ω and ν are weight functions such that, $\omega : \mathbb{R}^k \to \mathbb{R}^+$, $\nu : \mathbb{R}^k \to \mathbb{R}^+$, $\psi : \mathbb{R} \to \mathbb{R}$, and r is standardized residual [13]. Mallows and Schweppe [20,33] have recommended the function η choosing $\nu(X) = 1$ and $\nu(X) = \frac{1}{\omega(X)}$ weight functions respectively. Thus, the function η for Schweppe type estimator is defined as

$$\eta(X, r) = \omega(X)\psi\left(\frac{r}{\omega(X)}\right),$$

where $\psi(\cdot)$ is odd, bounded, uniformly continuous, non-decreasing and $\psi(u) > 0$ for $u > 0$ [24]. Hill [14] pointed out that, Mallows and Schweppe estimators are more efficient than several other estimators with Schweppe's method having an advantage. The main advantage of GM-estimator with the Schweppe's weighting scheme is that it assigns a weight to the high leverage point considering its distance from the regression line (i.e. assigns low weight to the leverage point if it has high residual) [9, 13]. The diagonal values (h_{ii}) of the hat matrix $H = X(X'X)^{-1}X'$ are used to identify leverage points. Generally, $h_{ii} > 2k/n$ (or $3k/n$) indicates that the corresponding observation is a high leverage point [6, 20, 29]. It is necessary to assign a low weight to this observation in the parameter estimation to reduce the effect of leverage point on regression parameters. This can be achieved by the weight function $\omega(X_i) = \sqrt{1 - h_{ii}}$ because $\omega(X_i) < \omega(X_j)$ for $h_{ii} > h_{jj}, i \neq j = 1, 2, , n$. Therefore, in this article we propose a model selection criterion based on the GM-estimator with Schweppe's weight function, $\nu(X_i) = \frac{1}{\omega(X_i)}$ and $\omega(X_i) = \sqrt{1 - h_{ii}}$, $i = 1, 2, ..., n$. Equation (1.7) can be solved using iterative method, and at convergence the GM-estimator is given as,

$$\hat{\beta} = X(X'WX)^{-1}X'Wy,$$

where the final diagonal weight matrix W, $W_{ii} = \frac{\eta(X_i, r_i)}{r_i}, i = 1, 2, ..., n$ is obtained by using $\omega(X)$ and $\psi(\cdot)$ function.

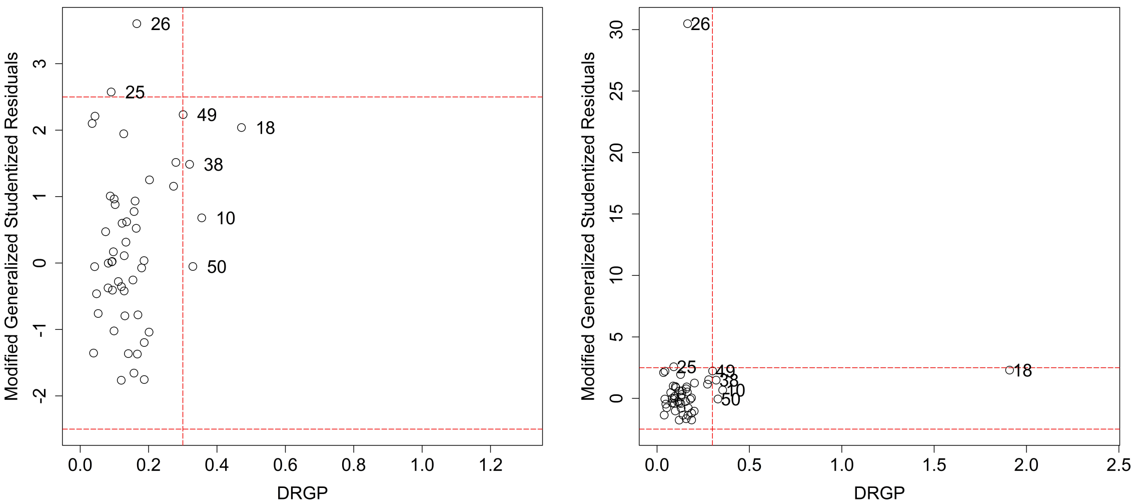
The remaining article has organized as follows: Section 2 explains the problem of the existence of a vertical outlier and a leverage point in the data and evaluates the performance of existing methods. In Section 3, the new criterion is proposed based on the GM-estimator to combat the simultaneous occurrence of vertical outliers and high leverage points. Also, the consistency property of the proposed method has established. An extensive simulation study is carried out in Section 4 to illustrate the performance of the proposed criterion and compared with existing methods through the real data. The article ends with a discussion given in Section 5.

2. The problem

In this section, we consider an example to illustrate the effect of the simultaneous occurrence of a vertical outlier and a leverage point in the data. Consider the regression model

$$y_i = 5 + 3.5X_{i1} + 6X_{i2} + 0X_{i3} + 0X_{i4} + \varepsilon_i,$$

where $\varepsilon_i, i = 1, 2, ..., 50$ are independent and identical errors generated from a standard normal distribution. The predictor variables X_j (j=1,2,3,4) are generated from standard uniform distribution and using the above model, we generate the response y. A vertical outlier has introduced in the data multiplying by three to a response variable corresponding to the highest absolute residual. The leverage point has introduced in the data multiplying by three to a row of the X matrix (excluding 1's column) corresponding to the highest leverage($\max(h_{ii})$). The plot of Modified Generalized Studentized Residuals (MGti) versus the Diagnostic Robust Generalized Potential (DRGP) is used to identify the outliers in the data [1].



(a) (b)

Figure 1. (a) MGt-DRGP plot for original data, (b) MGt-DRGP plot in the presence of a vertical outlier and a leverage point.

In the simulated data from the model under consideration, 26^{th} observation has the largest absolute standardized residual and MGt value, and 18^{th} observation has the largest leverage and DRGP value. To make these observations as a vertical outlier and a high leverage point multiplying by three to y_{26} and 18^{th} row of an X matrix (excluding 1's column) respectively. Thus, the modified data has influential vertical outlier as well as mild bad leverage point (Figure 1 (b)).

We have evaluated the performance of non-robust and robust model selection methods using the above simulated data. We compute C_p, S_p, RC_p and S_p (based on GM-estimator) for all possible submodels using simulated data and compare by plotting values of these statistics (see Figure 2). The classical C_p chooses different subset of predictor variables

(X_1, X_2, X_3), (X_1, X_2, X_4) and (X_1, X_2, X_3, X_4) according to 'close to p' criterion. Hence, C_p select overfitted model means a model having all relevant predictor variables and at least one irrelevant predictor variable. However, the value of the C_p statistic corresponds to correct subset X_1, X_2 is very small as compared to p and this incline to select overfitted model. This indicates that, C_p statistic is sensitive to vertical outliers and leverage points. The RC_p and S_p are M-estimator based robust model selection criteria choose same subset of predictor variables X_1, X_2, X_3. This model has extra irrelevant variable X_3 and is overfitted.

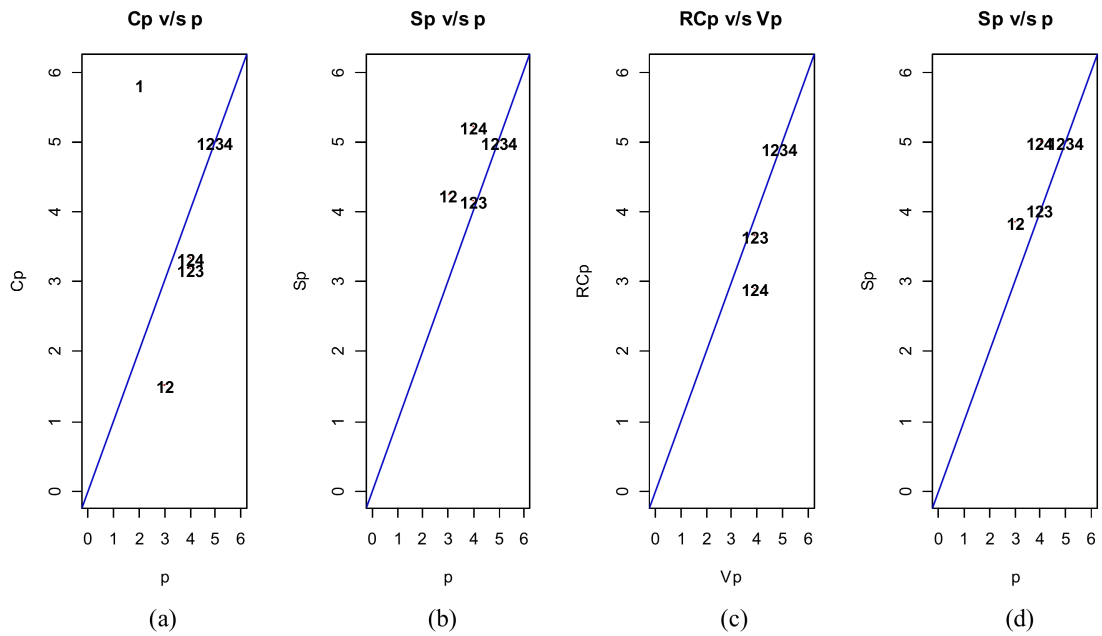


Figure 2. a) C_p versus p, (b) S_p versus p, (c) RC_p versus V_p, (d) S_p (based on GM) versus p.

In the light of the above discussion, model selection methods based on LS estimator and M-estimator fails to select a proper model in the presence of both a vertical outlier and a high leverage point. Also, S_p statistic based on GM-estimator fails to select the correct model for this situation. Hence, it is not enough to change the estimator only; it is also needed to modify the form of criterion. In next Section, we have proposed adaptive S_p statistic based on GM estimator to tackle the same problem.

3. Adaptive S_p (AS_p)

Let $\hat{\beta}$ be the GM-estimator of β and $\hat{y}_k$ be the vector of predicted values of y based on the full model (1.1) using the GM-estimator. The vector of predicted values $\hat{y}_k$ is given by $\hat{y}_k = X\hat{\beta} = Hy$, $H = X(X'WX)^{-1}X'W$ is a hat or projection matrix, W is a diagonal matrix of non-negative weights. Let $\hat{\beta}_1$ be the GM-estimator of β_1 and $\hat{y}_p$ be the vector of predicted values of y based on the submodel (1.3) using the GM-estimator. The fitted equation based on submodel is $\hat{y}_p = X_1\hat{\beta}_1 = H_p y$, $H_p = X_1(X_1'W_1X_1)^{-1}X_1'W_1$ is a hat or projection matrix, W_1 is a diagonal matrix of non-negative weights. A good model selection criterion is one which considers the goodness of fit as well as the complexity of the model [7]. Therefore, we propose adaptive S_p statistic (AS_p) based on predicted values $\hat{y}_k$, $\hat{y}_p$ and model complexity measure $C(n,p)$. It is defined as

$$AS_p = \frac{\|\hat{y}_k - \hat{y}_p\|^2}{\sigma^2} + C(n,p), \tag{3.1}$$

where σ^2 is an unknown error variance which can be replaced by using suitable estimate $\hat{\sigma} = 1.48\times$Median(largest (n-k+1) absolute residuals of full model), the constant term 1.48 is used to achieve consistency of scale parameter at normal distribution [33]. The term $\frac{\|\hat{y}_k-\hat{y}_p\|^2}{\sigma^2}$ is a measure of the discrepancy between full model and submodel. The value of this term is large for the wrong model as compared to the correct model, and it is a good measure to detect the correct models. However, the smallest value of this term indicates that the submodel is closer to the full model, and its value is zero when the submodel itself is the full model. Thus, minimization of $\frac{\|\hat{y}_k-\hat{y}_p\|^2}{\sigma^2}$ is not a proper model selection criterion because this can't accomplish principle of parsimony. Hence, the measure of discrepancy is necessary but not sufficient in model selection.

Most of the model selection criterion expressed as a goodness of fit term plus a measure of complexity. The dimension of the model (p) is a trivial measure of the complexity of the model. Any increasing function of p can be viewed as a complexity measure of a model. However, it ignores the sample size, hence making the resulting criterion inconsistent for instance AIC [16]. Hence, a good complexity measure should be an increasing function of both dimension of the model, p and the sample size, n. To make a good criterion, we consider an increasing function of n and p, $C(n,p)$ as a measure of the complexity of the model. The model having smaller value of AS_p will be the best model for prediction. Hence, we select the model having smaller AS_p.

The AS_p criterion based on LS estimator and M-estimator with certain $C(n,p)$ are equivalent to Mallows's C_p and S_p criteria respectively. The AS_p criterion can be viewed as a generalization of C_p and S_p statistic. The C_p, S_p (or RC_p) criterion selects the model according to 'close to p (or V_p)' rule, whereas the AS_p criterion selects the model for which its AS_p value is minimum. Hence, the model selection using C_p, S_p and RC_p are more intricate as compared to AS_p.

Proposition 3.1. *Least squares estimator based AS_p criterion with $C(n,p) = 2p - k$ is equivalent to Mallows's C_p.*

Proof. The AS_p criterion based on the LS estimator is defined as

$$AS_p = \frac{\|\hat{y}_k - \hat{y}_p\|^2}{\sigma^2} + C(n,p),$$

where LS estimator is used to estimate $\hat{y}_k$ and $\hat{y}_p$ values, and

$$\begin{aligned}\|\hat{y}_k - \hat{y}_p\|^2 &= \|(\hat{y}_k - y) - (\hat{y}_p - y)\|^2 \\ &= RSS_k + RSS_p - 2(\hat{y}_k - y)'(\hat{y}_p - y).\end{aligned}$$

After simplification we get,

$$\|\hat{y}_k - \hat{y}_p\|^2 = RSS_k + RSS_p - 2RSS_k = RSS_p - RSS_k \text{ [18]}.$$

Hence, AS_p criterion with penalty $C(n,p) = 2p - k$ is

$$AS_p = \frac{RSS_p - RSS_k}{\sigma^2} + 2p - k.$$

Since σ^2 is unknown, it replaced by its suitable estimator $\hat{\sigma}^2 = \frac{RSS_k}{n-k}$. Thus,

$$\begin{aligned}AS_p &= \frac{RSS_p}{\hat{\sigma}^2} - (n - 2p) \\ &= C_p.\end{aligned}$$

□

Proposition 3.2. *M-estimator based AS_p criterion with $C(n,p) = 2p - k$ is equivalent to S_p.*

Proof. The AS_p criterion based on the M-estimator with a penalty $C(n,p) = 2p - k$ is given as

$$\begin{aligned} AS_p &= \frac{\|\hat{y}_k - \hat{y}_p\|^2}{\sigma^2} - (k - 2p) \\ &= S_p. \end{aligned}$$

□

Following [16], consider the selected model referred by M_α, α denotes the set of selected predictor variables including the intercept. The class of models having all relevant predictor variables is denoted by the class of correct models ($\mathcal{M}_c$), and the class of wrong models ($\mathcal{M}_w$) is a class of models in which at least one relevant predictor is missing. A model having only relevant predictors is known as the optimal model, and it is denoted by M_{α^o}.

Proposition 3.3. *For any correct model,* $E(AS_p) = tr[(H - H_p)^{'}(H - H_p)] + C(n,p)$.

Proof. The expected value of AS_p is given by,

$$E(AS_p) = E(\frac{1}{\sigma^2}[y^{'}(H - H_p)^{'}(H - H_p)y] + C(n,p)).$$

Since, $(H - H_p)^{'}(H - H_p)$ is symmetric matrix and $C(n,p)$ is a constant,

$$\begin{aligned} E(AS_p) &= \frac{1}{\sigma^2}E[y^{'}(H - H_p)^{'}(H - H_p)y] + C(n,p) \\ &= \frac{1}{\sigma^2}\{\sigma^2 tr[(H - H_p)^{'}(H - H_p)] + \beta^{'}X^{'}(H - H_p)^{'}(H - H_p)X\beta\} + C(n,p). \end{aligned}$$

But, for any correct model, $\beta^{'}X^{'}(H - H_p)^{'}(H - H_p)X\beta = 0$. Hence,

$$E(AS_p) = tr[(H - H_p)^{'}(H - H_p)] + C(n,p).$$

□

Under following mild regulatory conditions, Theorem 3.6 exhibits consistency property of the proposed AS_p criterion for fixed number of predictor variables.

Condition 3.4. For any wrong model $M_\alpha \in \mathcal{M}_w$, $\liminf_{n\to\infty} \frac{\|X\beta - X_\alpha\beta_\alpha\|^2}{n\sigma^2} > 0$.

Whenever $M_\alpha \in \mathcal{M}_w$, it is prospect that the difference $\|X\beta - X_\alpha\beta_\alpha\|^2$ is large and hence the assumption is justifiable.

Condition 3.5. $C(n,p_\alpha) = o(n)$ and $C(n,p_\alpha) \to \infty$ as $n \to \infty$, p_α is a cardinality of a set α ($|\alpha|$).

The following theorem indicates that, if we choose a model having small AS_p value from all possible models, then asymptotically the selected model is an optimal model.

Theorem 3.6 (Consistency Property). *Under Conditions 3.4 and 3.5,* AS_p *selects the optimal model with probability one, i.e.* $\lim_{n\to\infty} Pr(M_\alpha = M_{\alpha^o}) = 1$.

Proof. The proof is divided into two parts. In first part, we show that AS_p value of any wrong model is greater than any correct model and in second part, AS_p value of optimal model M_{α^o} is smaller among the class of correct models $\mathcal{M}_c$. For any model M_α, $\hat{y}_{p_\alpha}$ is a vector of predicted values of y and hence

$$
\begin{aligned}
\|\hat{y}_k - \hat{y}_{p_\alpha}\|^2 &= \left\|X\hat{\beta} - X_\alpha\hat{\beta}_\alpha\right\|^2 \\
&= \left\|X\hat{\beta} - X\beta\right\|^2 + \left\|X_\alpha\hat{\beta}_\alpha - X\beta\right\|^2 - 2(X\hat{\beta} - X\beta)'(X_\alpha\hat{\beta}_\alpha - X\beta) \\
&= \left\|X\hat{\beta} - X\beta\right\|^2 + \left\|X_\alpha\hat{\beta}_\alpha - X_\alpha\beta_\alpha\right\|^2 + \|X\beta - X_\alpha\beta_\alpha\|^2 \\
&\quad - 2(X_\alpha\hat{\beta}_\alpha - X_\alpha\beta_\alpha)'(X\beta - X_\alpha\beta_\alpha) - 2(X\hat{\beta} - X\beta)'(X_\alpha\hat{\beta}_\alpha - X_\alpha\beta_\alpha) \\
&\quad + 2(X\hat{\beta} - X\beta)'(X\beta - X_\alpha\beta_\alpha) \\
&= \left\|X\hat{\beta} - X\beta\right\|^2 + \left\|X_\alpha\hat{\beta}_\alpha - X_\alpha\beta_\alpha\right\|^2 + \|X\beta - X_\alpha\beta_\alpha\|^2 + \zeta_1 + \zeta_2 + \zeta_3. \qquad (3.2)
\end{aligned}
$$

Since, $\hat{\beta}_\alpha - \beta_\alpha = o_p(1)$ [24] and by law of large numbers, $\zeta_1 = o_p(n), \zeta_2 = o_p(n)$ and $\zeta_3 = o_p(n)$. Thus,

$$
\|\hat{y}_k - \hat{y}_{p_\alpha}\|^2 = \left\|X\hat{\beta} - X\beta\right\|^2 + \left\|X_\alpha\hat{\beta}_\alpha - X_\alpha\beta_\alpha\right\|^2 + \|X\beta - X_\alpha\beta_\alpha\|^2 + o_p(n). \qquad (3.3)
$$

Similarly, under some regulatory conditions, we have $\left\|X_\alpha\hat{\beta}_\alpha - X_\alpha\beta_\alpha\right\|^2 = O_p(1)$ [9,20,24]. For any $M_\alpha \in \mathcal{M}_c$, $X\beta = X_\alpha\beta_\alpha$ and by Cauchy-Schwartz inequality

$$
\begin{aligned}
\|\hat{y}_k - \hat{y}_{p_\alpha}\|^2 &= \left\|X\hat{\beta} - X\beta\right\|^2 + \left\|X_\alpha\hat{\beta}_\alpha - X_\alpha\beta_\alpha\right\|^2 - 2(X\hat{\beta} - X\beta)'(X_\alpha\hat{\beta}_\alpha - X_\alpha\beta_\alpha) \\
&\leq \left\|X\hat{\beta} - X\beta\right\|^2 + \left\|X_\alpha\hat{\beta}_\alpha - X_\alpha\beta_\alpha\right\|^2 + 2\left\|X\hat{\beta} - X\beta\right\|^2 \left\|X_\alpha\hat{\beta}_\alpha - X_\alpha\beta_\alpha\right\|^2 \\
&= O_p(1). \qquad (3.4)
\end{aligned}
$$

Let AS_{p_α} and $AS_{p_{\alpha^*}}$ are values of the AS_p of any wrong model $M_\alpha \in \mathcal{M}_w$ and correct model $M_{\alpha^*} \in \mathcal{M}_c$ respectively. By Condition 3.4, 3.5 and combining (3.3), (3.4),

$$
\begin{aligned}
\liminf_{n\to\infty} Pr\left(AS_{p_\alpha} > AS_{p_{\alpha^*}}\right) &= \liminf_{n\to\infty} Pr\left(\frac{\|X\beta - X_\alpha\beta_\alpha\|^2}{\sigma^2} + o_p(n) > 0\right) \\
&\geq Pr\left(\liminf_{n\to\infty} \frac{\|X\beta - X_\alpha\beta_\alpha\|^2}{\sigma^2} + o_p(n) > 0\right) \\
&= 1. \qquad (3.5)
\end{aligned}
$$

Thus, the value of AS_p of a wrong model is greater than any correct model belongs to the class of correct models for large sample size. In the light of Equation (3.5), it is sufficient to show that the AS_p value of the optimal model is small among the class of correct models $\mathcal{M}_c$ to complete the proof of consistency property. By Condition 3.5, $C(n, p_\alpha) \to \infty$ as $n \to \infty$ and $C(n, p_{\alpha^*}) - C(n, p_{\alpha^o}) \geq 0$ because $p_{\alpha^*} \geq p_{\alpha^o}$ for any correct model. From Equation (3.4), we have

$$
\begin{aligned}
AS_{p_{\alpha^*}} - AS_{p_{\alpha^o}} &= \frac{\|\hat{y}_k - \hat{y}_{p_{\alpha^*}}\|^2}{\sigma^2} - \frac{\|\hat{y}_k - \hat{y}_{p_{\alpha^o}}\|^2}{\sigma^2} + C(n, p_{\alpha^*}) - C(n, p_{\alpha^o}) \\
&= O_p(1) + C(n, p_{\alpha^*}) - C(n, p_{\alpha^o}) \\
&\geq 0. \qquad (3.6)
\end{aligned}
$$

This proves that, the AS_p is a consistent criterion. □

4. Performance of AS_p criterion

In this section, we evaluate the performance of AS_p through a simulation study by considering three regression models and six penalty functions (Table 1). We have generated data and introduced vertical outliers and leverage points in this data using the procedure given in the Section 2. The good leverage point does not affect estimator of regression parameters and the AS_p criterion based on this estimator. Therefore, throughout the entire simulation study, bad leverage points are considered and are mentioned as leverage points. The different combinations of the number of vertical outliers and the number of leverage points have considered in this simulation study. The Huber $\psi(\cdot)$ function is widely used for robust parameter estimation in linear regression, and the tunning constant 1.345 achieves high efficiency over LS estimator in normal case [20]. The AS_p statistic for all possible submodels calculated using Huber's $\psi(\cdot)$ function with tuning constant 1.345.

Table 1. Models and Penalty functions.

Sr. No.	Model	Penalty function $C(n, p)$
1.	$M_1 : y = 5 + 2X_1 + 3X_2 + 0X_3 + \varepsilon$	P1= $3p$
2.	$M_2 : y = 4 + 3X_1 - 2X_2 + 7X_3 + 0X_4 + 0X_5 + \varepsilon$	P2= $2p\log(p)$
3.	$M_3 : y = 3 + 2.5X_1 + 1.7X_2 - 6X_3 + 8X_4 + 0X_5 + 0X_6 + 0X_7 + \varepsilon$	P3= $p\log(n)$
		P4= $p(\log(n) + 1)$
		P5= $6p\log(\log(n))$
		P6= $p\sqrt{n}$

4.1. AS_p criterion with different penalties

In Table 2 and 3, the percentage of optimal model selection based on 1000 runs for different combinations of penalties, sample sizes, and models has recorded. For model M_1, all penalties work well in case of clean data (0-Vertical outliers and 0-Leverage points) and select an optimal model with at least 75% when the sample size is 50. After adding vertical outliers/and leverage points in the data still, AS_p work well and selects an optimal model with up to 81.4% in the presence of vertical outlier only, 65.6% in the presence of leverage point only, 62.7% in the presence of both vertical outlier and leverage point for sample size 50. Moreover, the AS_p works satisfactorily for other combinations of the number of vertical outliers and leverage points. These percentages of optimal model selection increase with increasing sample size. The similar results have obtained from model M_2.

For model M_3, the AS_p criterion selects the optimal model with at least 59% for clean data and sample size 50; but as sample size increases this percentage increases. AS_p performs better in the presence of vertical outliers/and leverage points, and select optimal model with more than 90% for large sample size. It is observed that, the performance of the AS_p criterion is based on ratio k/n. The percentage of optimal model selection increases as k/n decreases.

Hence, for small sample size and in the presence of vertical outliers/and leverage points, AS_p performs satisfactorily. As the sample size increases, the optimal model selection percentage increases to 100% for P3, P4, P5 and P6 penalties. Thus, the simulation results show that the proposed criterion is consistent. Also, the AS_p criterion with P1 and P2 penalties select the optimal model preciously in the presence of vertical outliers/and leverage points for large sample size. Thus, the simulation study indicates that the proposed criterion performs well and selects the optimal model more preciously for a large sample size.

4.2. Comparative study of AS_p with C_p, S_p and RC_p

The comparative study of AS_p and other criteria (C_p, S_p and RC_p) in the presence of vertical outlier and leverage point is carried out through model selection ability for same simulation design and M_1, M_2 models. The model selection ability of these criteria has presented in Figure 3 by using 100 simulated datasets. In Figure 3, P1, P2, P3, P4, P5 and P6 refers to AS_p criterion with P1, P2, P3, P4, P5 and P6 penalty function respectively. This simulation study shows that the simultaneous occurrence of a vertical outlier and a leverage point effect on the performance of C_p, S_p, RC_p criteria, and most of the times C_p selects wrong or overfitted model. Mostly S_p select overfitted model, but the number of the wrong model selected by S_p is small as compared to C_p. Another one robust criterion RC_p select optimal model more preciously as compare to C_p and S_p, but it performs poorly as compared to AS_p. The difference between overfitted models selected by RC_p and AS_p is notable. Thus, the AS_p criterion increases the optimal model selection percentage by at least 50% relative to others. As the sample size increases, the probability of selecting the wrong model using the above criteria approaches to zero. However, the AS_p selects an optimal model with a large percentage as compared to C_p, S_p and RC_p. In conclusion, the AS_p has more model selection ability as compared to others.

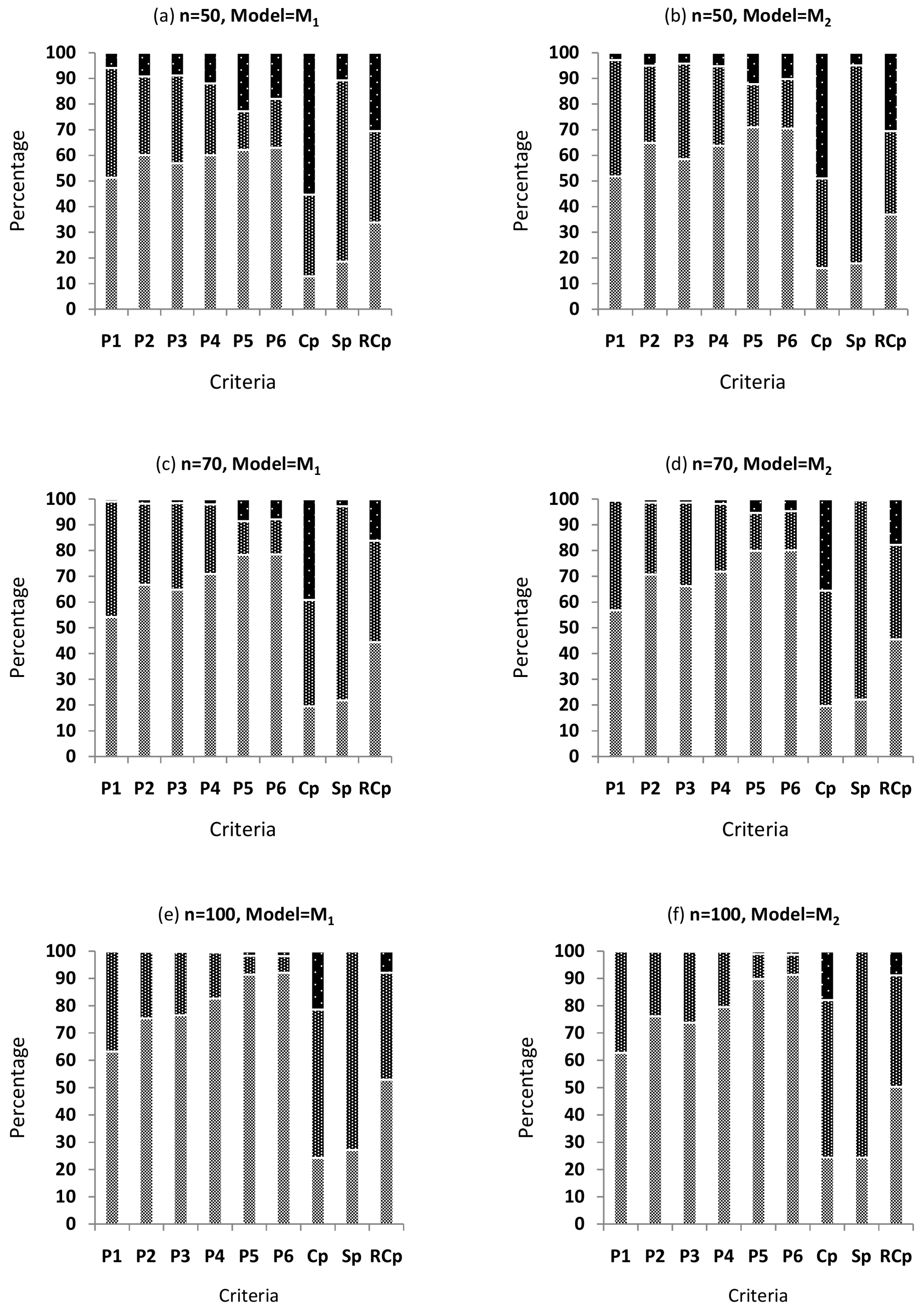
Table 2. Percentage of selecting optimal model (1000 runs).

Model	n	0-Vertical outliers and 0-Leverage points						1-Vertical outlier and 0-Leverage points					
		P1	P2	P3	P4	P5	P6	P1	P2	P3	P4	P5	P6
M_1	50	75.4	85.4	81.8	85.6	84.7	86.1	68.5	79.9	75.6	80.2	79.9	81.4
	70	77.7	87.8	86.7	89.8	91.9	92.2	74.6	86.8	85.3	89.1	91.7	91.8
	100	78.2	89.4	89.8	93.0	97.1	97.4	74.0	86.8	87.2	90.7	97.0	97.2
	200	81.1	89.4	92.9	96.0	99.3	99.8	79.0	90.3	93.5	95.9	99.5	99.8
	300	79.3	91.1	94.6	97.2	99.6	100.0	79.6	90.5	95.0	96.5	99.4	99.8
M_2	50	75.0	86.1	81.6	85.2	85.2	86.0	65.5	80.4	74.1	79.5	81.2	82.5
	70	78.1	89.4	87.1	89.7	93.3	93.3	75.1	87.3	84.0	88.2	92.0	92.3
	100	77.7	91.2	89.7	93.3	97.7	98.4	73.0	89.3	87.3	91.9	96.8	97.5
	200	81.3	92.2	92.8	95.6	99.1	99.9	79.7	92.1	93.0	95.4	98.8	99.8
	300	80.0	91.9	94.0	96.6	99.4	100.0	79.4	91.5	94.3	96.2	99.4	99.8
M_3	50	59.1	72.5	66.7	70.4	67.4	69.7	55.0	70.3	64.0	68.0	65.0	67.2
	70	66.1	84.0	77.8	83.0	82.3	82.8	61.0	80.0	73.1	79.1	81.1	81.5
	100	70.0	88.6	85.1	89.3	93.8	92.3	66.6	86.2	81.7	86.9	93.8	92.6
	200	73.3	91.7	91.0	95.0	99.0	99.8	70.7	90.1	89.3	93.9	98.6	99.7
	300	72.3	91.7	92.7	95.7	99.3	99.8	74.5	92.0	92.9	95.5	99.4	99.8
Model	n	0-Vertical outliers and 1-Leverage point						1-Vertical outlier and 1-Leverage point					
		P1	P2	P3	P4	P5	P6	P1	P2	P3	P4	P5	P6
M_1	50	54.6	64.1	60.9	63.9	64.7	65.6	49.0	58.3	54.8	58.6	62.7	61.6
	70	57.9	69.8	68.1	73.4	79.8	79.3	55.4	69.6	68.2	72.2	79.7	79.8
	100	63.3	76.5	76.9	81.9	90.7	91.2	58.8	71.5	72.4	78.3	87.9	88.9
	200	71.8	82.4	86.4	90.7	96.8	99.1	69.0	80.4	85.7	89.9	96.2	98.6
	300	73.4	84.7	89.3	93.1	98.6	99.9	74.5	85.6	91.8	94.9	98.1	99.7
M_2	50	52.7	65.7	59.7	65.1	71.9	70.6	45.8	60.8	53.0	59.6	68.1	64.9
	70	56.1	71.0	66.2	72.4	82.1	81.4	55.7	69.3	66.0	71.2	82.4	82.0
	100	60.1	77.7	75.0	80.9	90.9	91.8	59.8	77.4	74.7	80.4	90.5	91.4
	200	68.6	85.1	86.4	90.8	96.9	99.0	69.8	85.0	86.8	90.2	96.8	99.1
	300	73.0	88.9	91.7	94.5	98.6	99.8	74.3	86.4	90.1	92.4	97.7	99.9
M_3	50	34.8	49.0	41.8	46.5	50.4	50.6	33.7	47.1	41.0	44.9	45.7	47.0
	70	49.7	67.0	60.1	65.5	69.9	69.4	42.1	62.0	53.3	60.6	65.2	64.9
	100	49.6	71.2	65.7	71.6	82.4	82.7	47.2	69.4	63.9	70.1	80.4	81.6
	200	61.2	83.7	82.6	88.2	95.2	98.1	60.2	83.0	82.3	87.6	95.6	98.2
	300	64.3	84.3	85.5	89.7	97.1	99.8	63.3	86.5	87.6	90.4	96.7	99.5

Table 3. Percentage of selecting optimal model (1000 runs).

Model	n	0-Vertical outliers and 2-Leverage points						2-Vertical outliers and 0-Leverage points					
		P1	P2	P3	P4	P5	P6	P1	P2	P3	P4	P5	P6
M_1	50	52.3	58.0	55.2	55.6	51.8	53.7	82.8	87.3	85.0	86.7	79.3	82.7
	70	62.7	71.0	69.9	72.2	72.5	72.3	85.7	91.5	90.5	91.6	89.7	90.6
	100	69.0	77.7	77.9	81.2	86.5	86.0	88.3	94.3	94.3	96.3	97.5	96.8
	200	73.6	81.1	84.4	87.6	94.8	97.6	88.5	93.5	95.1	96.8	99.0	99.9
	300	81.5	89.2	92.9	94.4	98.2	99.7	90.2	94.9	96.9	98.6	99.9	100.0
M_2	50	44.0	54.5	50.4	53.3	59.6	59.6	69.2	79.4	75.5	78.3	76.0	78.4
	70	46.7	60.2	55.6	61.1	71.7	70.9	75.4	88.1	84.7	87.4	88.3	88.7
	100	51.6	67.5	64.7	71.0	82.6	84.6	78.8	91.7	89.4	94.0	97.4	97.3
	200	63.4	79.2	81.0	85.7	93.1	96.3	79.0	91.9	92.9	95.8	99.6	100.0
	300	66.1	82.7	86.7	90.3	96.8	98.9	81.1	93.6	95.5	97.6	99.5	100.0
M_3	50	30.4	36.7	33.5	35.6	33.9	34.7	62.0	66.9	66.0	65.9	54.3	58.9
	70	33.6	46.9	42.8	46.0	48.9	49.4	66.3	81.0	75.3	80.3	76.1	77.0
	100	42.5	62.2	57.7	62.5	71.1	71.3	73.3	90.6	87.2	90.3	92.3	90.4
	200	51.1	73.5	72.8	77.8	90.5	95.1	71.9	91.8	91.1	94.5	98.9	99.4
	300	54.0	76.9	78.7	83.0	94.9	98.7	74.3	92.3	92.9	95.6	99.3	100.0
Model	n	1-Vertical outlier and 2-Leverage points						2-Vertical outliers and 1-Leverage point					
		P1	P2	P3	P4	P5	P6	P1	P2	P3	P4	P5	P6
M_1	50	52.1	58.3	55.5	57.0	53.2	55.3	65.9	71.8	69.7	70.7	68.8	70.5
	70	60.5	68.0	66.4	67.4	68.2	68.4	77.0	84.0	82.0	84.5	84.5	85.0
	100	65.7	75.9	76.1	80.0	85.5	86.0	78.5	85.4	85.4	88.7	93.4	94.0
	200	76.0	84.4	87.4	89.8	95.3	97.9	84.1	91.2	93.1	95.5	98.4	99.5
	300	79.7	87.6	91.4	93.6	97.4	99.2	84.2	91.4	95.1	96.8	98.8	99.8
M_2	50	39.8	54.5	48.0	53.5	61.9	59.7	54.5	66.9	61.0	65.9	67.8	68.0
	70	44.7	59.6	54.7	60.9	70.3	70.0	60.1	73.3	69.1	74.2	81.5	81.2
	100	51.6	69.1	65.8	72.2	82.5	84.4	64.2	81.0	78.7	84.0	91.8	93.2
	200	61.9	78.5	80.4	84.3	93.8	96.5	72.0	86.5	87.5	90.6	97.2	99.3
	300	68.2	82.5	85.1	88.1	96.5	99.6	75.4	89.5	92.2	94.6	98.9	99.9
M_3	50	26.6	36.1	31.8	35.1	34.0	35.3	36.4	47.6	42.4	45.5	43.7	44.6
	70	35.3	48.2	43.3	47.5	49.6	50.3	47.7	62.8	57.2	61.8	62.7	63.0
	100	40.3	58.7	53.9	59.3	68.1	69.2	56.0	74.8	69.9	75.5	82.7	82.1
	200	50.4	73.6	73.0	78.8	90.5	94.4	62.6	84.9	84.3	89.5	96.6	98.1
	300	53.1	77.9	79.8	84.7	94.7	99.0	64.8	85.3	86.9	90.2	97.4	99.5

Model	n	2-Vertical outliers and 2-Leverage points					
		P1	P2	P3	P4	P5	P6
M_1	50	45.6	51.6	48.6	49.2	48.2	49.3
	70	61.1	67.2	64.8	67.1	68.2	68.5
	100	66.8	75.3	75.6	78.4	83.3	83.4
	200	77.1	83.5	86.8	89.3	95.6	98.1
	300	80.4	88.1	92.3	94.1	97.6	99.7
M_2	50	38.4	52.1	44.9	51.1	58.5	58.1
	70	44.7	58.3	53.5	59.1	69.9	69.5
	100	49.8	64.9	62.2	69.8	81.1	81.8
	200	59.9	74.5	76.2	81.3	90.8	95.2
	300	63.1	79.5	84.3	87.9	95.9	99.0
M_3	50	24.6	32.4	28.2	30.5	33.6	33.4
	70	33.8	48.0	42.3	45.5	50.1	50.6
	100	39.0	56.3	51.9	57.1	67.5	68.3
	200	50.2	72.6	71.8	78.1	90.1	95.2
	300	55.8	79.0	81.2	85.4	95.5	99.1


(a) n=50, Model=M_1
(b) n=50, Model=M_2
(c) n=70, Model=M_1
(d) n=70, Model=M_2
(e) n=100, Model=M_1
(f) n=100, Model=M_2
Percentage
0
10
20
30
40
50
60
70
80
90
100
P1
P2
P3
P4
P5
P6
Cp
Sp
RCp
Criteria

Continue..

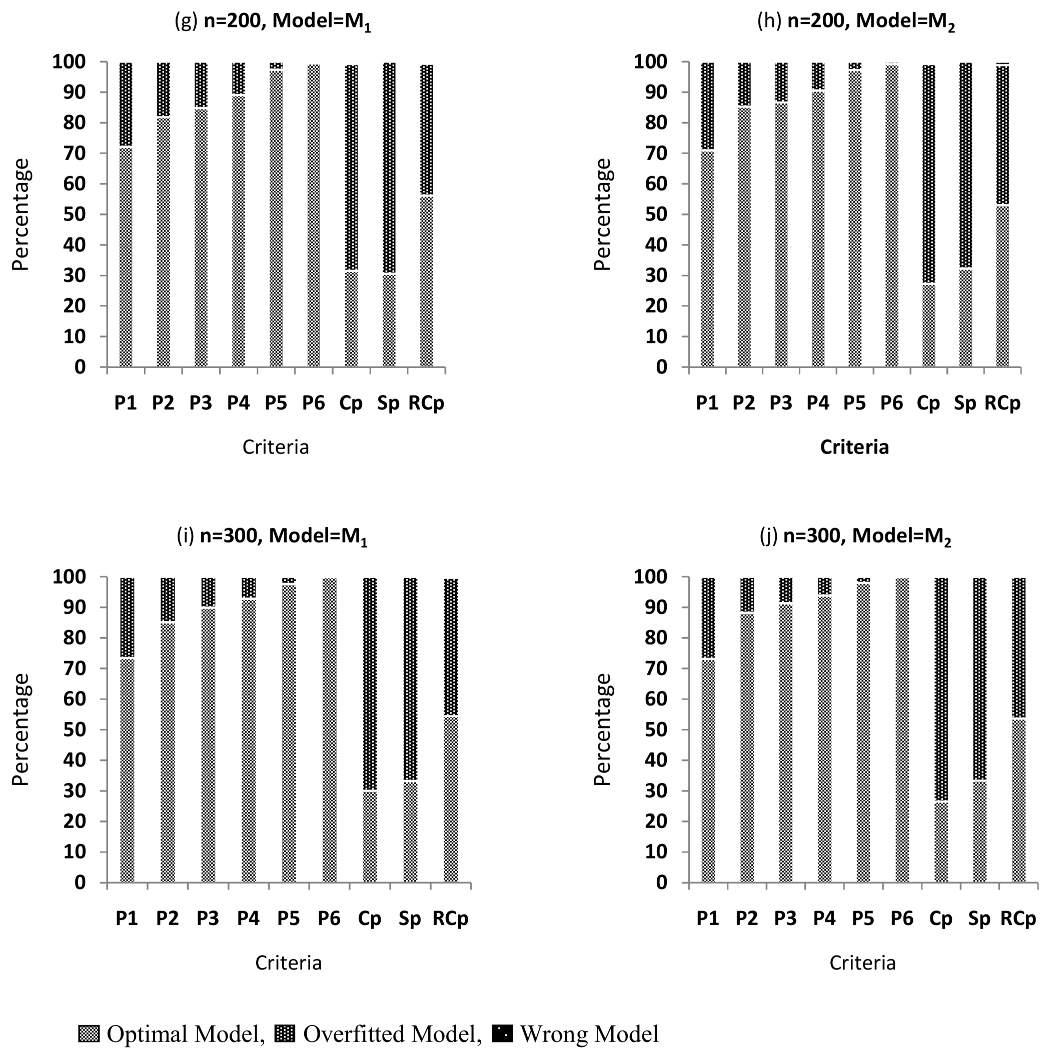


Figure 3. (a)-(j) Model selection ability of AS_p, C_p, S_p and RC_p.

4.3. Example: Agglomeration in Bayer precipitation data

Sommer and Staudte [31] and Bab-Hadiashar and Suter [5] analyzed Bayer precipitation data on the agglomeration of aluminium trihydroxide ($Al(OH)_3$) crystals by means of the Bayer precipitation process. Data has 51 observations on the response variable y (An agglomeration of crystals with size exceeding 45 microns, i.e. the difference between the percentage of crystals exceeding this size, which leave the agglomerator tank and those which enter it) and nine predictor variables $X_1 - X_9$ [5,31]. Figure 4 shows that, the data has 11 vertical outliers, 6 bad leverage points and 5 good leverage points.

For this data, we compared the performance of non-robust as well as robust criteria. The C_p and RC_p recommend several submodels with three or more variables, and S_p suggests model containing at least five variables (Figure 5). Whereas, the AS_p would consider five predictor variables X_1, X_3, X_4, X_5, X_8. Thus, the model having X_1, X_3, X_4, X_5, X_8 variables seems to be "good", according to AS_p criterion with different penalties (Table 4). Sommer and Staudte [31] suggest $X_1, X_3, X_4, X_5, X_7, X_8$ predictor variables using RC_p with Mallows weight, and Bab-Hadiashar and Suter [5] suggest three best subsets $(X_1, X_3, X_4, X_5, X_8), (X_1, X_3, X_4, X_5, X_7, X_8)$ and $(X_1, X_3, X_4, X_5, X_7, X_8, X_9)$ for the same data.

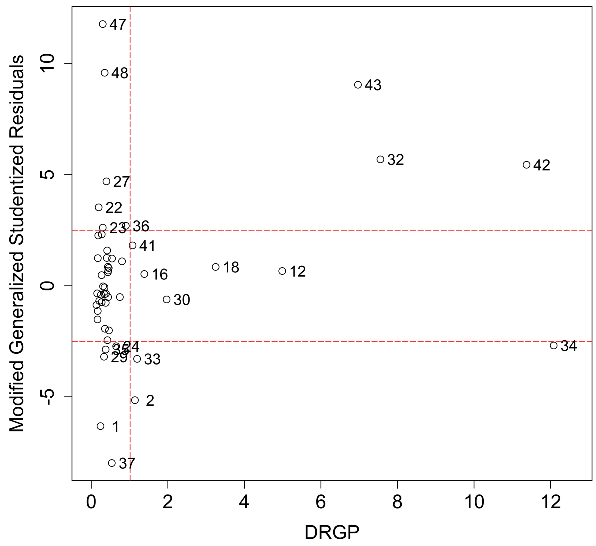


Figure 4. MGt-DRGP plot for Bayer Precipitation data.

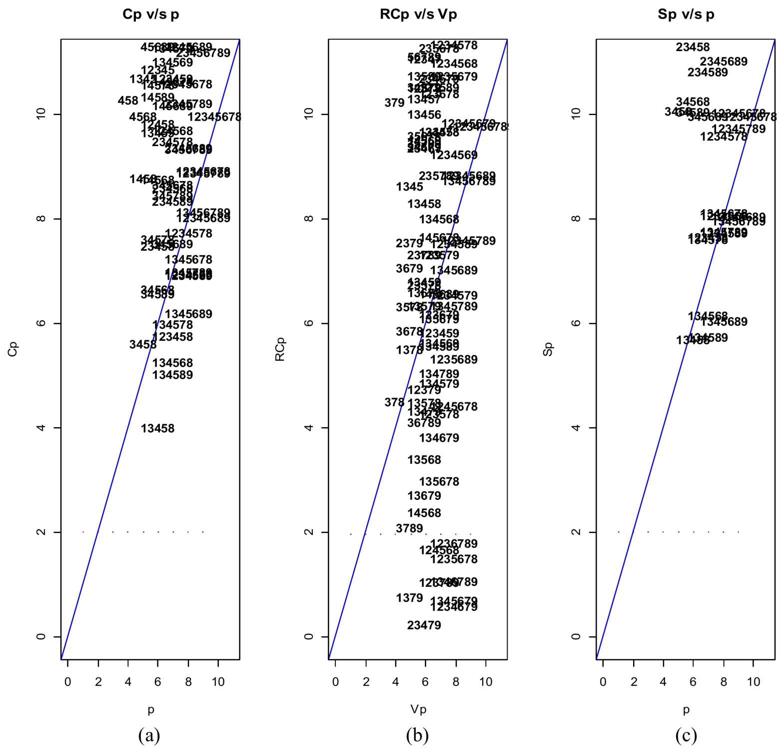


Figure 5. C_p, RC_p and S_p plots for Bayer Precipitation data: (a) C_p versus p, (b) RC_p versus V_p, (c) S_p versus p.

For a real-life data set, it is difficult to show which submodel is good. Generally, the prediction error increases due to an addition of irrelevant variables in the model [34]. Consider three models suggested by [5] for further study, and compare the prediction error of these three models using GM-estimator and k-fold cross-validation (k=10). The prediction error of $(X_1, X_3, X_4, X_5, X_8)$, $(X_1, X_3, X_4, X_5, X_7, X_8)$ and $(X_1, X_3, X_4, X_5, X_7, X_8, X_9)$ are

57.8597, 69.4523 and 67.7013 respectively. The k-fold cross-validation study indicates that the model selected by AS_p has a small prediction error as compared to other, and the prediction error of the model increases due to an addition of X_7 or X_7 and X_9. Thus, a subset of predictor variables namely X_1, X_3, X_4, X_5, X_8 is appropriate for fitting the model.

Table 4. AS_p values and selected variables corresponding to different penalties.

Penalty	Selected Variables	AS_p (minimum)
P1	X_1, X_3, X_4, X_5, X_8	24.9337
P2		28.4348
P3		30.5247
P4		36.5247
P5		56.2214
P6		49.7823

4.4. The performance of AS_p criterion for different $\psi(\cdot)$ functions

We have compared the performance of the AS_p criterion based on GM-estimator and M-estimator with different $\psi(\cdot)$ functions in this section. The data was generated using the model and procedure explained in Section 2. The performance of three types of $\psi(\cdot)$ functions [15] i) Monotonic- Huber, Fair, ii) Soft Redescending- Cauchy, Welsch and iii) Hard Redescending- Tukey's bisquare (or biweight), Talwar has examined. The percentages of optimal model selection are calculated using 1000 simulated datasets, and the results are recorded in Tables 5-6. It is observed that, the hard and soft rescending functions are working well as compared to the monotonic functions. The performance of the M-estimator with redescending function is notable, and among redesceding functions, the Tukey's bisquare function has better performance. Nevertheless, the AS_p criterion based on the GM-estimator selects the optimal model more precisely as compared to that based on the M-estimator. The performance is dependent on the choice of penalty, and the penalty function satisfying Conditions 3.4, 3.5 are selecting the optimal model with a high percentage. Overall, the criterion based on GM-estimator with the redescending function will be a better choice for small as well as large n.

4.5. The performane of AS_p criterion for difference $\omega(\cdot)$ weight functions

In the previous simulation study, the weight function $\omega(X_i) = \sqrt{1 - h_{ii}}$ based on the hat matrix has considered to compute the AS_p criterion value. The hat matrix is hampered by a masking effect, and consequently, the diagonal values of the hat matrix do not always detect leverage points [30]. To overcome this problem, we can use alternative $\omega(\cdot)$ weights based on robust measures. The Robust Mahalnobis Distance (RMD) based on Minimum Volume Ellipsoid (MVE) can be used to identify leverage points [11,29,30]. The Generalized Potential (GP) is an another measure used to identify the leverage point, and the observation is said to be a high leverage point if corresponding GP value is greater than the threshold value $Median(GP_i) + 3MAD(GP_i)$ [1].

In this section, the superiority of the AS_p criterion based on GM-estimator with different $\omega(\cdot)$ weights (Table 7) and $\psi(\cdot)$ functions is checked for severe cases: 5% vertical outliers and 5% leverage points, 10% vertical outliers and 10% leverage points, 15% vertical outliers and 15% leverage points. The data was generated using the model explained in the Section 2. Vertical outliers were included in the data multiplying by three to the response variable corresponding to maximum absolute residual, and bad leverage points were included in the data multiplying by three to response variable and row of the matrix X (excluding 1's column) corresponding to maximum leverage (h_{ii}). The results are reported in the Table 8. A weight function ω_1 is unable to detect all leverage points for large n and consequently, it effect on the performance of the AS_p criterion in presence of multiple leverage points. It is observed that, AS_p criterion based on GM-estimator with robust weights ω_3, ω_4 and ω_5 perform better in presence of multiple outliers. Also, the performance of ω_2 is remarkable. The AS_p criterion with hard rescending functions and robust weights perform well. Hence, the AS_p criterion with hard rescending function with robust weight is a better choice in presence of multiple outliers.

Table 5. Performance of AS_p criterion for GM-estimator and M-estimator with different $\psi(\cdot)$ functions.

n	Estimator	$\psi(\cdot)$ function	1-Vertical outlier and 1-Leverage point						2-Vertical outliers and 1-Leverage point					
			P1	P2	P3	P4	P5	P6	P1	P2	P3	P4	P5	P6
50	GM estimator	Huber	52.90	64.30	60.80	66.90	80.40	75.80	49.30	62.40	58.10	65.10	77.10	73.80
		Fair	43.00	53.80	49.70	55.60	70.40	65.00	40.10	51.80	46.80	53.80	66.90	63.50
		Cauchy	59.60	71.10	66.70	73.30	84.50	81.90	55.80	68.80	64.10	71.00	83.20	79.40
		Welsch	61.30	74.10	69.30	76.20	86.60	83.50	59.30	70.00	66.10	72.30	84.20	81.30
		Bisquare	61.50	74.00	69.50	76.50	86.90	83.50	59.30	69.70	66.40	72.50	84.50	81.20
		Talwar	60.00	71.80	67.60	73.70	83.10	80.70	58.50	69.10	65.90	71.20	82.80	80.40
	M estimator	Huber	29.40	38.80	36.40	41.00	50.60	47.60	29.20	38.20	35.40	40.20	49.40	46.20
		Fair	23.30	31.70	28.10	32.90	41.60	40.10	22.30	31.00	28.50	32.90	42.60	40.00
		Cauchy	38.60	47.80	44.70	50.70	60.50	57.70	36.60	48.00	44.20	50.00	61.20	57.60
		Welsch	44.80	54.70	52.00	56.60	65.60	63.60	43.00	53.40	50.20	55.40	65.70	63.00
		Bisquare	43.60	54.40	51.30	56.30	65.20	63.00	43.20	53.60	49.90	55.50	65.50	63.00
		Talwar	30.20	37.60	35.20	39.00	45.30	44.10	29.40	37.50	34.10	38.70	45.70	43.90
70	GM estimator	Huber	60.10	71.90	70.40	75.60	86.70	85.90	56.00	69.70	67.70	74.10	86.30	85.30
		Fair	48.50	61.10	59.30	64.90	78.20	77.60	47.50	58.80	57.50	62.30	77.80	76.70
		Cauchy	64.60	77.10	75.60	81.50	89.80	89.50	60.70	74.20	72.50	78.50	89.50	88.60
		Welsch	66.80	79.50	78.10	83.60	91.40	90.90	62.50	76.30	75.00	80.70	92.00	91.30
		Bisquare	66.60	79.10	77.90	83.70	91.50	91.20	63.40	76.60	75.40	81.10	92.00	91.60
		Talwar	66.40	77.80	76.60	81.30	89.50	89.30	63.70	77.70	75.80	81.40	90.70	90.50
	M estimator	Huber	38.10	49.60	47.90	53.10	64.60	64.50	38.00	49.20	46.70	52.70	65.30	63.80
		Fair	30.90	40.60	39.60	44.30	54.80	53.70	31.20	40.00	39.00	44.50	55.40	54.90
		Cauchy	50.60	62.70	61.40	67.30	78.00	77.30	47.40	60.00	57.70	64.50	76.90	75.90
		Welsch	57.30	70.30	68.10	74.00	84.80	84.40	53.90	65.90	63.80	70.60	82.40	81.40
		Bisquare	57.40	70.30	68.30	74.40	85.00	84.50	54.80	66.80	64.40	71.10	82.70	81.70
		Talwar	47.00	57.40	56.00	60.80	68.30	67.90	46.30	56.40	54.60	59.60	69.10	68.80
100	GM estimator	Huber	65.00	75.50	75.90	80.30	91.20	92.50	63.50	73.50	74.00	79.20	91.90	92.80
		Fair	54.70	66.60	67.40	73.60	86.30	87.50	52.20	63.90	64.90	70.60	84.90	87.00
		Cauchy	68.90	79.50	80.20	84.20	94.00	95.30	66.60	79.40	80.30	85.20	95.10	95.80
		Welsch	70.70	82.20	82.60	86.80	95.30	96.90	69.50	80.90	81.50	86.50	95.80	96.50
		Bisquare	71.20	82.70	82.90	87.00	95.20	96.80	69.80	81.30	82.10	87.30	95.90	96.50
		Talwar	71.30	81.60	82.40	86.70	95.90	96.50	69.90	81.30	81.80	87.60	95.60	96.50
	M estimator	Huber	49.40	60.30	60.70	65.60	79.00	81.50	49.50	59.80	59.90	64.30	78.90	81.10
		Fair	40.10	50.10	50.40	54.80	68.40	70.70	37.90	48.40	48.60	53.40	68.50	70.50
		Cauchy	58.80	69.40	70.10	76.30	87.40	89.20	58.10	70.20	70.50	75.20	86.70	88.70
		Welsch	65.60	77.20	77.70	82.70	92.30	93.60	63.30	75.80	76.30	81.30	91.80	93.60
		Bisquare	65.30	77.00	77.80	83.30	92.30	93.80	64.50	76.20	77.00	82.30	92.20	93.40
		Talwar	61.10	72.00	72.60	77.00	86.80	87.80	61.70	72.40	72.70	78.30	87.00	88.30
200	GM estimator	Huber	68.70	81.40	85.10	89.30	95.90	98.40	67.20	81.20	85.20	88.90	96.70	98.60
		Fair	59.80	73.40	79.60	84.40	92.90	97.40	60.20	72.80	77.80	83.10	92.80	98.20
		Cauchy	72.80	85.10	89.30	92.30	97.30	99.50	72.60	84.90	88.40	91.50	97.60	99.50
		Welsch	74.40	87.00	89.70	92.40	98.30	99.60	74.90	86.50	90.10	94.10	98.20	99.70
		Bisquare	74.40	87.20	89.70	92.40	98.30	99.60	75.00	86.60	90.40	94.10	98.30	99.60
		Talwar	74.60	87.50	90.20	93.30	97.80	99.50	74.70	87.40	90.60	93.80	98.40	99.70
	M estimator	Huber	61.40	74.70	79.20	84.70	93.20	97.20	61.10	74.00	79.10	83.30	93.20	97.70
		Fair	50.40	64.90	69.60	75.80	87.40	93.30	50.20	64.90	69.30	74.20	86.40	92.50
		Cauchy	68.50	81.10	86.00	89.70	95.60	98.50	68.50	81.40	85.10	88.30	96.60	99.00
		Welsch	71.10	85.60	89.10	91.60	97.60	99.20	71.90	84.60	88.30	92.20	98.00	99.60
		Bisquare	71.90	86.10	89.10	91.90	97.70	99.20	73.10	85.40	89.20	92.90	98.00	99.60
		Talwar	71.60	85.00	88.70	91.90	97.50	99.10	73.00	85.10	87.90	91.60	97.60	99.20
300	GM estimator	Huber	73.60	85.40	90.60	93.10	98.20	100.00	72.70	85.40	91.20	93.60	98.40	99.80
		Fair	66.90	80.90	87.00	89.70	96.40	99.90	67.60	80.10	86.30	89.60	96.80	99.50
		Cauchy	76.20	89.60	93.30	95.10	98.60	99.90	77.00	88.80	92.30	95.90	98.90	100.00
		Welsch	78.20	89.60	94.10	95.90	98.70	99.90	77.90	89.20	93.90	96.60	99.00	100.00
		Bisquare	78.30	89.70	94.20	95.90	98.80	99.90	77.50	89.40	93.60	96.70	99.00	100.00
		Talwar	78.90	89.70	94.20	96.00	98.90	99.70	78.00	88.30	93.30	96.30	99.00	100.00
	M estimator	Huber	69.00	81.20	88.00	90.50	97.00	99.90	69.20	82.20	87.60	91.90	97.00	99.50
		Fair	61.00	75.00	81.70	85.90	94.90	99.30	62.40	75.20	81.80	85.80	93.90	98.70
		Cauchy	73.90	87.20	92.00	94.30	98.20	99.80	74.40	87.30	91.30	94.70	98.40	99.70
		Welsch	76.70	89.00	93.40	95.80	98.60	99.70	76.10	88.40	92.70	96.00	98.70	99.90
		Bisquare	77.30	89.10	93.90	95.80	98.60	99.70	76.40	88.20	92.70	96.40	98.90	100.00
		Talwar	77.30	88.80	93.30	95.30	98.50	99.70	76.70	87.30	92.50	95.90	99.10	100.00

Table 6. Performance of AS_p criterion for GM-estimator and M-estimator with different $\psi(\cdot)$ function.

n	Estimator	$\psi(\cdot)$ function	1-Vertical outlier and 2-Leverage points						2-Vertical outliers and 2-Leverage points					
			P1	P2	P3	P4	P5	P6	P1	P2	P3	P4	P5	P6
50	GM estimator	Huber	30.90	41.40	37.40	44.10	58.50	53.60	31.70	42.10	38.60	44.70	59.70	55.70
		Fair	21.50	31.90	27.00	33.80	46.40	43.10	23.90	33.80	30.70	35.50	47.20	43.30
		Cauchy	40.80	53.00	48.20	55.30	68.10	63.40	39.50	51.80	47.50	53.20	66.60	63.60
		Welsch	47.30	58.90	55.00	61.00	72.90	69.20	46.50	58.10	54.00	60.00	71.90	67.90
		Bisquare	47.50	58.60	55.50	60.60	72.90	69.70	48.10	59.00	55.10	60.90	72.90	69.40
		Talwar	38.40	46.90	43.60	48.20	57.50	54.90	38.70	48.30	45.10	49.60	59.80	57.60
	M estimator	Huber	13.70	19.20	16.60	20.10	29.70	28.00	15.50	20.80	18.90	21.70	29.40	27.70
		Fair	12.60	18.20	16.00	19.10	24.40	22.90	14.20	18.90	16.90	19.20	25.40	23.60
		Cauchy	18.10	24.10	21.70	25.80	35.20	33.00	17.20	23.50	21.10	24.90	34.50	30.90
		Welsch	24.10	31.40	28.30	33.10	41.30	38.60	23.50	30.90	28.70	32.80	41.00	38.90
		Bisquare	24.70	32.00	29.00	33.80	41.70	38.60	23.70	30.90	28.90	32.80	41.80	39.80
		Talwar	12.90	17.00	15.20	17.90	24.00	22.10	16.70	20.60	18.50	21.40	26.60	24.90
70	GM estimator	Huber	39.50	49.50	48.10	53.40	68.70	67.00	39.60	52.00	49.80	55.90	71.10	69.50
		Fair	29.10	39.40	38.00	42.70	56.50	55.90	30.70	40.50	38.90	44.70	57.40	56.60
		Cauchy	51.00	60.60	58.90	65.60	79.20	78.10	50.90	63.10	61.00	66.80	80.50	80.10
		Welsch	56.10	67.10	65.00	72.20	83.90	82.90	58.70	68.70	67.30	73.10	84.20	83.30
		Bisquare	56.30	67.50	65.90	72.10	83.70	83.20	59.10	69.70	67.90	73.60	85.30	84.00
		Talwar	50.40	60.30	58.40	64.30	76.20	75.80	53.60	64.40	62.50	67.80	79.50	78.70
	M estimator	Huber	21.00	28.60	27.40	31.70	42.70	42.10	23.90	31.10	30.10	34.20	46.00	45.00
		Fair	15.40	22.80	21.20	24.50	33.40	33.00	19.10	25.90	25.20	27.40	36.10	35.30
		Cauchy	32.00	41.10	40.00	44.60	55.50	54.70	31.90	40.30	39.70	43.00	58.60	56.80
		Welsch	42.80	52.20	50.90	56.60	68.80	68.30	44.10	57.00	54.60	60.30	71.20	70.90
		Bisquare	43.40	52.10	50.90	56.80	68.40	68.10	44.40	56.60	54.70	59.90	71.20	70.90
		Talwar	26.70	33.40	32.20	36.00	43.30	43.00	29.90	36.10	34.80	38.80	47.50	47.00
100	GM estimator	Huber	43.60	54.90	55.40	61.20	74.80	76.70	40.00	53.10	53.60	59.60	74.60	76.40
		Fair	32.80	44.60	45.30	50.60	64.30	66.80	31.80	42.70	43.10	48.10	62.00	63.90
		Cauchy	52.10	65.10	66.30	72.90	85.90	87.90	49.90	64.80	65.00	72.30	85.30	86.80
		Welsch	59.40	73.10	73.80	78.80	90.30	91.50	57.50	70.80	71.60	77.10	89.40	90.90
		Bisquare	59.70	73.90	74.80	79.30	90.40	91.80	57.70	71.20	71.80	77.20	89.50	90.70
		Talwar	55.90	69.90	70.10	74.60	86.40	87.70	56.50	69.30	69.80	75.20	86.60	88.00
	M estimator	Huber	28.40	38.20	38.50	43.40	57.10	59.70	27.70	37.10	37.30	42.20	54.70	57.70
		Fair	21.60	29.10	29.30	33.70	46.20	49.90	20.30	27.20	27.70	33.00	45.20	47.50
		Cauchy	39.80	50.40	50.70	54.50	68.40	70.70	37.10	48.40	49.40	54.90	68.60	70.60
		Welsch	49.40	63.80	64.80	71.00	83.00	84.50	49.60	62.10	62.80	68.40	81.50	83.90
		Bisquare	50.60	65.50	66.40	71.40	83.30	84.80	50.80	62.80	63.70	69.50	82.60	84.10
		Talwar	41.50	53.20	53.50	58.10	68.20	70.00	42.00	53.60	53.90	60.10	69.80	71.40
200	GM estimator	Huber	55.30	68.50	72.70	77.20	88.30	94.10	56.90	67.60	72.60	77.20	89.70	94.80
		Fair	43.10	57.80	60.40	65.80	78.50	86.10	45.30	57.80	61.80	66.50	79.70	88.00
		Cauchy	65.10	79.60	84.00	87.00	95.50	97.30	66.20	78.80	82.90	87.00	95.10	98.40
		Welsch	70.10	82.50	87.30	91.50	96.80	98.10	71.00	83.00	87.90	90.50	96.50	99.30
		Bisquare	70.40	82.40	87.50	91.40	97.10	98.20	71.70	83.80	88.00	90.90	96.90	99.40
		Talwar	69.30	81.60	86.80	91.00	96.50	97.90	71.90	83.90	87.60	90.80	96.60	98.60
	M estimator	Huber	44.90	58.10	63.10	67.60	80.40	88.40	47.50	60.80	64.10	68.40	82.00	88.80
		Fair	34.00	45.10	49.80	55.00	68.70	77.50	35.70	47.50	51.90	56.70	70.30	79.40
		Cauchy	57.80	72.30	77.40	81.50	91.00	95.50	59.80	71.90	75.60	80.80	90.50	95.80
		Welsch	67.00	79.90	84.80	89.30	95.70	97.50	67.20	80.60	84.80	87.90	95.50	98.20
		Bisquare	67.50	79.50	85.10	89.40	95.90	97.60	68.20	81.80	85.70	89.30	96.10	98.60
		Talwar	65.50	78.90	83.80	87.90	94.60	96.80	68.90	80.70	84.50	87.70	93.60	96.80
300	GM estimator	Huber	63.10	76.10	82.50	86.50	95.60	98.70	62.60	76.00	82.70	86.90	94.90	98.70
		Fair	52.80	66.50	72.90	77.80	89.40	96.60	53.90	64.60	72.70	77.70	89.00	97.00
		Cauchy	70.20	83.60	90.10	93.00	97.60	99.40	72.10	84.70	90.10	92.90	97.80	99.70
		Welsch	73.40	86.10	92.10	94.60	98.20	99.90	75.10	87.10	91.20	93.70	98.60	99.90
		Bisquare	73.70	86.70	92.20	94.50	98.50	99.90	74.90	87.70	91.30	94.00	98.90	99.90
		Talwar	74.30	86.90	91.50	94.80	98.40	99.80	75.60	86.80	92.00	94.10	98.70	99.90
	M estimator	Huber	57.00	70.00	77.50	82.10	91.90	97.40	57.80	68.60	77.30	81.40	91.60	98.00
		Fair	45.30	58.90	66.90	71.10	83.10	93.70	46.70	58.40	65.40	69.90	83.40	92.70
		Cauchy	66.20	80.30	86.50	90.00	96.60	98.80	66.90	79.60	86.40	90.80	96.70	98.90
		Welsch	71.40	84.90	91.40	93.40	98.00	99.70	73.20	86.00	90.30	93.00	98.20	99.70
		Bisquare	72.30	84.80	91.60	93.40	98.10	99.80	73.50	86.30	90.30	93.10	98.40	99.70
		Talwar	72.40	85.20	90.50	93.60	98.10	99.50	73.70	85.30	91.00	93.10	98.50	99.80

Table 7. List of $\omega(\cdot)$ weight functions.

Sr. No.	$\omega(\cdot)$ weight function
1.	$\omega_1(X_i) = \sqrt{1 - h_{ii}}$
2.	$\omega_2(X_i) = \begin{cases} 1, & \text{if } h_{ii} \le \frac{2p}{n} \\ 0, & \text{otherwise} \end{cases}$
3.	$\omega_3(X_i) = \begin{cases} 1, & \text{if } RMD_i \le Median(RMD_i) + 3MAD(RMD_i) \\ 0, & \text{otherwise} \end{cases}$
4.	$\omega_4(X_i) = \begin{cases} 1, & \text{if } RMD_i \le \sqrt{\chi^2_{p,0.95}} \\ 0, & \text{otherwise} \end{cases}$
5.	$\omega_5(X_i) = \begin{cases} 1, & \text{if } GP_i \le Median(GP_i) + 3MAD(GP_i) \\ 0, & \text{otherwise} \end{cases}$

Table 8. Performance of AS_p criterion for different $\omega(\cdot)$ weight functions.

n	$\omega(\cdot)$ weight	$\psi(\cdot)$ function	5%-Vertical outliers and 5%-Leverage points						10%-Vertical outliers and 10%-Leverage points						15%-Vertical outliers and 15%-Leverage points					
			P1	P2	P3	P4	P5	P6	P1	P2	P3	P4	P5	P6	P1	P2	P3	P4	P5	P6
100	$\omega_1(X_i)$	Huber	8.40	11.50	11.70	14.10	24.10	25.90	1.70	3.10	3.20	4.00	7.50	8.60	1.00	1.90	1.90	2.20	5.00	6.00
		Fair	6.80	10.10	10.60	11.70	19.10	20.90	2.60	3.70	3.80	5.00	8.40	9.70	1.80	2.70	2.70	3.70	9.20	10.90
		Cauchy	16.00	22.60	23.00	27.00	37.00	39.30	1.90	3.10	3.10	4.10	8.00	8.60	0.60	1.30	1.30	1.70	3.60	4.30
		Welsch	37.50	47.50	47.80	53.50	65.20	66.50	6.20	8.20	8.30	10.20	13.80	14.70	1.70	2.50	2.60	3.00	4.60	5.40
		Bisquare	38.30	48.70	48.80	54.60	67.00	68.20	6.60	9.60	9.70	11.40	15.70	17.10	2.20	2.80	2.80	3.00	6.00	6.10
		Talwar	22.00	28.90	29.30	32.10	40.30	41.50	3.00	4.10	4.10	4.80	7.80	8.20	1.50	2.30	2.60	3.10	4.80	5.80
	$\omega_2(X_i)$	Huber	44.50	57.40	57.70	63.60	77.00	80.10	20.10	31.40	31.90	37.00	52.70	56.50	10.00	16.10	16.50	21.70	35.30	38.20
		Fair	38.40	50.50	51.00	56.00	71.30	74.60	11.80	19.90	20.20	25.60	39.80	43.50	5.50	10.40	9.90	12.80	21.90	24.50
		Cauchy	47.80	58.80	58.90	66.10	80.30	81.90	25.80	38.60	38.70	44.70	60.90	64.80	21.30	30.30	30.30	36.20	50.70	54.20
		Welsch	49.40	61.00	61.50	67.50	81.40	83.00	31.70	43.90	44.40	49.40	65.30	67.90	30.70	40.90	41.80	47.30	63.60	66.50
		Bisquare	50.00	61.50	61.60	67.80	81.90	82.90	31.70	44.10	44.30	49.80	65.70	69.00	30.50	42.10	42.70	48.40	63.60	67.80
		Talwar	50.50	62.60	63.30	68.80	82.00	85.10	31.80	44.50	45.20	50.20	66.40	69.70	30.90	42.00	43.30	49.50	63.60	66.90
	$\omega_3(X_i)$	Huber	46.50	57.40	58.00	63.80	76.70	80.40	21.60	33.20	33.50	38.80	54.20	57.50	12.80	22.20	22.30	28.10	43.70	46.50
		Fair	39.30	52.00	52.40	57.00	71.30	74.80	14.60	22.70	23.00	28.30	42.80	46.20	6.40	14.20	13.90	17.50	29.60	32.70
		Cauchy	47.90	59.10	59.00	65.90	79.40	81.20	26.30	39.00	39.40	45.10	61.60	65.00	25.00	34.80	34.90	39.40	57.20	60.10
		Welsch	49.20	60.70	61.30	67.50	81.10	83.00	31.50	44.20	44.80	49.70	65.30	67.70	30.60	40.50	41.50	48.00	63.70	66.50
		Bisquare	49.70	61.10	61.30	67.80	81.60	83.00	31.70	44.50	44.80	50.00	65.30	68.60	30.60	41.70	42.40	48.50	63.90	67.80
		Talwar	50.50	62.00	62.70	68.00	81.70	84.70	31.90	44.70	45.30	50.50	66.30	69.90	31.50	42.50	43.80	50.30	64.60	67.50
	$\omega_4(X_i)$	Huber	44.50	55.80	56.70	62.70	76.20	79.70	23.00	33.00	33.60	38.60	54.50	58.10	14.00	24.20	24.40	29.70	44.70	48.20
		Fair	38.10	51.10	51.80	56.30	70.70	74.10	16.00	24.30	24.20	29.80	43.50	46.60	7.70	15.00	14.90	18.80	31.10	33.90
		Cauchy	47.60	58.90	58.90	64.50	79.70	80.90	27.40	38.80	39.40	45.30	62.00	64.90	25.40	34.90	35.10	40.40	57.50	59.90
		Welsch	49.70	60.30	60.60	66.70	81.10	83.50	30.90	44.20	44.80	50.10	65.20	67.50	30.70	41.10	41.90	48.30	63.80	66.50
		Bisquare	49.80	60.20	60.80	67.40	81.70	83.90	31.40	44.50	45.00	50.20	65.20	68.40	30.70	42.10	42.60	48.80	64.00	68.00
		Talwar	49.90	62.00	62.80	68.20	82.20	84.90	31.60	44.90	45.30	50.90	66.60	69.90	32.20	43.30	44.50	50.90	65.20	68.00
	$\omega_5(X_i)$	Huber	45.20	56.70	57.10	62.90	76.70	80.00	22.80	33.50	33.60	39.00	54.90	58.60	13.60	23.50	23.50	29.00	46.30	48.60
		Fair	38.50	49.50	49.90	56.00	72.10	75.50	15.60	24.30	24.60	29.80	44.00	47.10	7.90	15.20	15.20	19.10	30.90	34.00
		Cauchy	47.90	58.60	58.90	65.70	80.00	81.60	27.00	38.90	39.20	45.30	61.40	64.50	25.40	34.90	35.20	41.10	58.20	60.50
		Welsch	49.40	60.60	61.30	66.60	81.20	82.90	31.30	43.90	44.30	49.60	65.30	67.80	30.20	40.60	41.60	48.40	64.30	67.10
		Bisquare	49.70	60.90	61.20	66.80	81.50	82.90	31.40	44.10	44.20	49.80	65.50	68.80	30.20	41.70	42.50	48.80	64.20	68.10
		Talwar	50.60	62.00	62.80	68.00	82.10	85.30	31.60	44.40	45.10	50.50	66.60	69.90	31.10	42.80	44.10	50.80	65.40	68.40
200	$\omega_1(X_i)$	Huber	1.20	2.30	2.80	4.40	8.60	12.40	0.00	0.00	0.00	0.00	0.90	1.70	0.00	0.20	0.20	0.30	0.70	1.20
		Fair	1.20	1.70	2.10	2.90	5.90	9.50	0.00	0.10	0.10	0.10	0.70	2.40	0.30	0.30	0.50	0.60	1.00	2.50
		Cauchy	5.50	8.20	10.20	11.70	18.30	24.80	0.00	0.00	0.10	0.10	0.30	1.60	0.00	0.00	0.10	0.20	0.30	0.70
		Welsch	40.50	50.20	54.50	58.50	69.80	78.30	2.90	4.00	4.90	5.40	8.00	10.40	0.20	0.40	0.40	0.40	0.70	1.70
		Bisquare	43.10	53.30	58.00	63.30	74.30	83.10	5.80	8.00	9.10	10.10	14.20	17.80	0.30	0.40	0.50	0.50	1.80	2.20
		Talwar	27.70	34.80	38.80	41.50	50.70	56.90	0.70	0.90	0.90	1.30	2.50	3.50	0.20	0.30	0.40	0.40	0.80	1.40
	$\omega_2(X_i)$	Huber	43.60	56.60	62.60	68.80	82.60	91.70	22.80	33.80	38.20	43.40	59.90	72.80	9.80	17.30	20.20	25.40	41.50	54.40
		Fair	37.00	48.70	55.00	60.40	78.40	87.50	12.40	20.80	25.30	30.80	47.10	59.00	3.00	7.20	9.10	11.90	22.60	35.40
		Cauchy	47.70	59.70	64.70	70.50	85.00	92.90	30.90	42.10	46.70	51.70	67.00	78.60	21.30	33.50	37.80	43.50	60.10	72.20
		Welsch	49.90	61.80	67.00	72.20	84.80	93.60	33.80	46.10	50.70	54.80	69.60	81.00	33.90	43.50	46.70	52.20	67.80	78.40
		Bisquare	49.70	62.00	67.30	72.30	84.80	93.80	33.10	46.20	50.80	55.40	70.10	81.40	34.10	44.10	47.70	52.50	67.80	79.00
		Talwar	50.10	62.90	68.30	73.40	86.00	93.70	33.50	46.60	51.20	56.30	70.70	82.00	34.10	43.70	47.50	52.60	68.40	78.70
	$\omega_3(X_i)$	Huber	43.30	57.40	63.20	68.50	82.90	92.20	23.20	34.80	39.60	44.50	60.70	72.20	12.70	22.70	25.90	29.60	46.30	59.50
		Fair	36.80	49.40	55.20	60.50	78.70	87.50	12.40	21.30	25.30	30.60	47.30	60.30	4.70	9.60	12.00	15.30	27.90	40.70
		Cauchy	47.50	59.90	64.50	70.50	84.90	93.00	31.00	42.10	47.30	52.00	67.30	78.90	25.00	37.50	40.60	46.40	61.90	74.10
		Welsch	49.90	61.80	67.00	72.20	84.80	93.80	34.00	46.30	50.70	55.10	69.60	81.30	34.10	43.50	47.60	52.30	68.10	78.50
		Bisquare	49.60	62.00	67.30	72.20	85.00	93.90	33.30	46.30	50.90	55.50	70.00	81.40	34.20	44.10	48.40	52.90	68.50	79.10
		Talwar	50.30	62.90	68.30	73.40	86.00	93.90	33.90	46.70	51.20	56.50	70.70	82.10	34.20	44.10	47.70	53.30	69.10	79.20
	$\omega_4(X_i)$	Huber	43.40	57.80	62.70	68.70	82.70	92.00	22.60	34.20	39.30	44.40	61.00	72.60	13.40	22.80	25.80	28.90	46.20	59.60
		Fair	37.20	49.60	55.50	60.40	78.30	87.70	12.80	20.90	25.40	30.70	47.20	60.80	4.70	9.60	12.00	16.20	27.70	40.80
		Cauchy	46.60	59.80	64.60	70.60	84.70	93.00	31.00	41.50	46.90	52.00	67.50	78.90	24.50	37.00	41.30	46.60	61.50	74.80
		Welsch	49.50	61.60	67.00	71.70	85.10	93.60	33.90	46.50	50.80	55.20	69.40	81.20	34.10	43.40	47.50	52.30	68.00	78.70
		Bisquare	49.10	62.00	67.10	72.00	85.10	93.80	33.40	46.40	51.00	55.60	69.80	81.40	34.30	44.00	48.40	52.80	68.30	79.10
		Talwar	49.80	63.10	68.40	73.10	86.30	93.70	34.00	46.80	51.30	56.40	70.60	81.90	34.30	44.10	47.90	53.50	69.00	79.20
	$\omega_5(X_i)$	Huber	43.20	57.30	62.60	68.90	82.80	91.90	23.00	34.00	39.50	44.60	61.00	72.50	13.50	21.70	25.00	29.00	46.80	60.00
		Fair	37.10	49.40	55.10	59.70	78.10	87.60	12.80	20.70	25.20	30.50	47.20	61.00	4.60	9.30	12.30	16.10	27.20	41.60
		Cauchy	47.10	59.70	64.30	70.40	84.90	92.90	30.70	42.00	47.10	51.90	67.50	78.60	25.00	36.50	41.10	45.80	61.40	74.70
		Welsch	49.60	61.80	66.70	71.90	84.70	93.60	33.80	46.20	50.80	55.30	69.60	81.20	34.20	43.40	47.50	52.30	68.10	78.50
		Bisquare	49.30	62.10	67.00	72.10	84.80	93.80	33.40	46.20	51.00	55.50	70.20	81.30	34.20	44.10	48.40	52.80	68.30	78.90
		Talwar	50.00	62.80	68.20	73.30	86.00	93.80	33.80	46.60	51.20	56.50	70.80	82.00	34.30	44.20	47.90	53.50	69.10	79.10
300	$\omega_1(X_i)$	Huber	0.30	0.40	0.60	0.60	1.30	4.30	0.00	0.00	0.00	0.00	0.00	0.40	0.00	0.00	0.00	0.00	0.00	0.30
		Fair	0.00	0.20	0.20	0.40	0.70	2.60	0.00	0.00	0.00	0.00	0.20	0.60	0.00	0.00	0.00	0.00	0.10	0.60
		Cauchy	1.20	2.50	2.90	3.30	5.50	12.10	0.00	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.00	0.00	0.00	0.20
		Welsch	34.30	44.90	52.20	56.60	71.50	80.80	1.30	1.80	2.00	2.50	4.50	6.70	0.00	0.00	0.00	0.00	0.10	0.20
		Bisquare	38.90	52.10	58.60	64.00	78.20	87.20	6.50	9.10	10.20	11.50	15.50	20.70	0.20	0.20	0.30	0.30	0.30	0.80
		Talwar	29.50	38.40	43.30	47.70	56.90	66.40	0.20	0.50	0.50	0.60	0.90	2.30	0.00	0.00	0.00	0.10	0.10	0.60
	$\omega_2(X_i)$	Huber	40.40	55.60	64.60	69.50	83.70	93.30	23.70	35.30	43.70	49.40	66.90	82.80	7.50	16.40	22.60	26.30	42.10	62.40
		Fair	31.50	46.70	56.00	62.50	78.10	91.30	12.40	20.70	28.90	32.90	48.80	71.50	2.10	5.90	9.10	11.40	22.20	41.50
		Cauchy	44.50	58.60	66.00	71.70	85.90	95.40	31.20	44.60	53.50	59.00	73.00	87.70	21.40	31.60	38.80	44.60	60.80	76.60
		Welsch	46.50	61.20	67.90	73.10	87.20	95.90	38.00	50.80	57.00	62.70	76.80	89.20	30.40	41.70	49.20	54.80	68.60	83.50
		Bisquare	46.80	61.60	67.80	73.50	87.50	96.30	38.00	50.90	57.60	63.50	76.90	89.90	31.30	42.20	50.00	54.90	69.10	83.80
		Talwar	48.80	62.50	68.90	74.60	86.90	96.10	39.30	51.50	59.00	64.20	78.60	90.50	32.50	42.90	50.20	55.20	69.40	83.80
	$\omega_3(X_i)$	Huber	40.20	56.00	65.00	69.50	83.40	93.40	24.40	36.50	43.60	49.90	68.10	83.80	9.10	18.40	24.60	30.20	45.30	65.60
		Fair	30.90	47.10	56.10	62.40	77.80	91.20	13.20	22.40	29.60	34.50	50.50	72.40	2.30	6.80	10.70	13.20	24.90	45.60
		Cauchy	44.40	58.80	65.80	71.50	85.90	95.40	32.10	45.20	53.60	59.40	73.80	87.30	22.40	32.60	40.70	46.80	62.50	78.20
		Welsch	46.50	61.30	68.00	72.90	87.30	95.90	38.10	50.90	57.20	62.80	77.00	89.30	31.90	42.80	49.90	55.00	69.00	83.60
		Bisquare	46.80	61.60	67.80	73.50	87.60	96.30	38.10	51.00	57.60	63.60	77.00	89.90	31.80	43.00	50.20	54.90	69.40	83.50
		Talwar	48.70	62.50	68.90	74.60	87.00	96.10	39.30	51.30	58.70	64.00	78.60	90.50	33.10	43.50	50.40	55.50	69.60	84.10
	$\omega_4(X_i)$	Huber	40.20	56.00	64.80	69.40	83.50	93.50	24.20	35.80	43.90	49.90	67.90	84.10	9.60	17.90	24.90	29.50	44.40	66.00
		Fair	30.90	47.30	55.90	62.60	77.50	90.90	13.80	22.90	29.80	34.20	50.00	72.60	2.80	7.70	12.00	15.00	24.80	46.20
		Cauchy	44.70	59.20	66.00	71.50	85.60	95.50	32.40	45.30	53.60	59.40	73.60	87.30	22.00	32.60	40.60	47.20	61.70	78.60
		Welsch	46.50	61.40	67.90	72.90	87.10	95.90	38.20	51.00	57.30	62.90	77.00	89.30	31.90	42.90	50.10	54.90	69.10	83.40
		Bisquare	46.90	61.70	68.00	73.50	87.40	96.30	38.40	51.20	57.70	63.80	77.10	89.90	31.80	43.10	50.30	54.80	69.40	83.50
		Talwar	48.50	62.50	69.20	74.90	87.00	96.10	39.40	51.30	58.80	64.10	78.70	90.50	33.10	43.60	50.50	55.40	69.70	84.20
	$\omega_5(X_i)$	Huber	39.90	55.40	65.00	69.50	83.40	93.40	24.10	35.70	43.80	49.60	68.10	84.00	9.20	16.80	24.70	28.70	45.30	65.90
		Fair	31.00	46.70	55.70	62.00	77.70	91.10	12.70	23.10	29.40	34.70	50.20	71.60	2.80	7.20	10.80	14.90	25.10	45.80
		Cauchy	44.30	58.90	65.70	71.60	85.50	95.40	32.50	45.10	53.10	59.40	73.30	87.20	21.80	32.40	40.80	46.70	62.10	78.30
		Welsch	46.50	61.50	67.90	73.20	87.00	95.90	38.40	50.80	57.20	62.70	77.00	89.30	31.80	43.20	50.10	54.80	69.20	83.50
		Bisquare	46.70	61.70	67.90	73.60	87.40	96.20	38.30	51.00	57.60	63.50	77.00	89.90	31.70	43.10	50.40	54.70	69.40	83.60
		Talwar	48.70	62.60	68.90	74.80	87.00	96.00	39.30	51.20	58.80	64.00	78.70	90.50	33.10	43.60	50.50	55.50	69.80	84.30

5. Discussion

We have suggested the AS_p model selection criterion based on GM-estimator. It can be viewed as a generalization of C_p and S_p criteria. The proposed criterion takes into account a discrepancy between the full model and submodel as well as complexity in the model. This criterion has consistency property and selects an optimal model with a high probability for large n. The simulation study reveals that the proposed criterion works well in the four different situations: clean data, vertical outliers, leverage points, and both vertical outliers and leverage points. The proposed criterion performs well and simple as compared to C_p, S_p and RC_p. The AS_p criterion based on the GM-estimator with hard redescending $\psi(\cdot)$ function and robust weight can be a better choice.

Acknowledgments

We sincerely thank the Editor, Associate Editor and anonymous reviewers for their careful review, constructive comments and valuable suggestions that resulted in the significant improvement of the original manuscript. We would also like to thanks DST, New Delhi and UGC, New Delhi for financial support under DST-FIST and UGC-SAP scheme, respectively.

References

[1] M. Alguraibawi, H. Midi and A.H.M. Imon, *A new robust diagnostic plot for classifying good and bad high leverage points in a multiple linear regression model*, Math. Probl. Eng., Doi:10.1155/2015/279472, 2015.

[2] C.D.S. André, S.N. Elian, S.C. Narula and R.A. Tavares, *Coefficients of determinations for variable selection in the msae regression*, Comm. Statist. Theory Methods **29** (3), 623-642, 2000.

[3] B. Asikgil and A. Erar, *Research into multiple outliers in linear regression analysis*, Hacet. J. Math. Stat. **38** (2), 185-198, 2009.

[4] A.C. Atkinson, *Fast very robust methods for the detection of multiple outliers*, J. Amer. Statist. Assoc. **89** (428), 1329-1339, 1994.

[5] A. Bab-Hadiashar and D. Suter, *Data Segmentation and Model Selection for Computer Vision: A Statistical Approach*, Springer, 2000.

[6] D.A. Belsley, E. Kuh and R.E. Welsch, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, Wiley, 1980.

[7] H. Bozdogan and D.M.A. Haughton, *Informational complexity criteria for regression models*, Comput. Statist. Data Anal. **28** (1), 51-76, 1998.

[8] S. Chatterjee and A.S. Hadi, *Influential observations, high leverage points and outliers in linear regression*, Statist. Sci. **1** (3), 379-416, 1986.

[9] C.W. Coakley and T.P. Hettmansperger, *A bounded influence, high breakdown, efficient regression estimator*, J. Amer. Statist. Assoc. **88** (423), 872-880, 1993.

[10] C. Croux and C. Dehon, *Estimators of the multiple correlation coefficient: Local robustness and confidence intervals*, Statist. Papers **44** (3), 315-334, 2003.

[11] C. Dehon, M. Gassner and V. Verardi, *Beware of 'Good' outliers and overoptimistic conclusions*, Oxf. Bull. Econ. Stat. **71** (3), 437-452, 2009.

[12] A.S. Hadi and J.S. Simonoff, *Procedures for the identification of multiple outliers in linear models*, J. Amer. Statist. Assoc. **88** (424), 1264-1272, 1993.

[13] F. R. Hampel, E.M. Ronchetti, P.J. Rousseeuw and W.A. Stahel, *Robust Statistics: The Approach based on Influence Functions*, Wiley, 1986.

[14] R.W. Hill, *When there are outliers in the carriers: The univariate case*, Comm. Statist. Theory Methods **11** (8), 849-868, 1982.

[15] P.W. Holland and R.E. Welsch, *Robust regression using iteratively reweighted least-squares*, Comm. Statist. Theory Methods **6** (9), 813-827, 1977.

[16] B. Hu and J. Shao, *Generalized linear model selection using R^2*, J. Statist. Plann. Inference **138** (12), 3705-3712, 2008.
[17] P.J. Huber, *Robust estimation of a location parameter*, Ann. Math. Stat. **35** (1), 73-101, 1964.
[18] D.N. Kashid and S.R. Kulkarni, *A more general criterion for subset selection in multiple linear regression*, Comm. Statist. Theory Methods **31** (5), 795-811, 2002.
[19] C. Kim and S. Hwang, *Influence subsets on the variable selection*, Comm. Statist. Theory Methods **29** (2), 335-347, 2000.
[20] W.S. Krasker and R.E. Welsch, *Efficient bounded-influence regression estimation*, J. Amer. Statist. Assoc. **77** (379), 595-604, 1982.
[21] J.A.F. Machado, *Robust Model Selection and M-estimation*, Econometric Theory **9** (3), 478-493, 1993.
[22] C. Mallows, *Some comments on C_p*, Technometrics **15** (4), 661-675, 1973.
[23] R.A. Maronna, R.D. Martin and V.J. Yohai, *Robust Statistics: Theory and Methods*, Wiley, 2006.
[24] R.A. Maronna and V.J. Yohai, *Asymptotic behavior of general M-estimates for regression and scale with random carriers*, Probab. Theory Related Fields **58** (1), 7-20, 1981.
[25] C.R. Rao, Y. Wu, S. Konishi and R. Mukerjee, *On model selection*, Lecture Notes in Monograph Series **38**, 1-64, 2001.
[26] O. Renaud and M.P. Victoria-Feser, *A robust coefficient of determination for regression*, J. Statist. Plann. Inference **140** (7), 1852-1862, 2010.
[27] E. Ronchetti, *Robust model selection in regression*, Statist. Probab. Lett. **3** (1), 21-23, 1985.
[28] E. Ronchetti and R.G. Staudte, *A robust version of Mallows's C_p*, J. Amer. Statist. Assoc. **89** (426), 550-559, 1994.
[29] P.J. Rousseeuw and A.M. Leroy, *Robust Regression and Outlier Detection*, Wiley, 2003.
[30] P.J. Rousseeuw and B.C. Van Zomeren, *Unmasking multivariate outliers and leverage points*, J. Amer. Statist. Assoc. **85** (411), 633-639, 1990.
[31] S. Sommer and R.G. Staudte, *Robust variable selection in regression in the presence of outliers and leverage points*, Aust. N. Z. J. Stat. **37** (3), 323-336, 1995.
[32] K. Tharmaratnam and G. Claeskens, *A comparison of robust versions of the AIC based on M, S and MM-estimators*, Statistics **47** (1), 216-235, 2013.
[33] R. Wilcox, *Introduction to Robust Estimation and Hypothesis Testing*, Elsevier, 2012.
[34] R. Wilcox, *Modern Statistics for the Social and Behavioral Sciences: A Practical Introduction*, CRC Press, 2012.