STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **7**, December 2019, pp 816–825.
Published online in International Academic Press (www.IAPress.org)

Alternative Robust Variable Selection Procedures in Multiple Regression

Shokrya Saleh [1,*] Ali Abuzaid [2]

[1] *Department of Mathematics, Jazan University, Kingdom of Saudi Arabia*
[2] *Department of Mathematics, Al-Azhar University-Gaza, Palestine*

Abstract Most of the commonly used linear regression variable selection techniques are affected in the presence of outliers and high leverage points and often could produce misleading conclusions. This article proposes robust variable selection methods, where the suspected outliers and high leverage points are identified by regression diagnostics tools and then the best variables are selected after diagnostic checking. The performance of the proposed methods is compared with the classical non-robust criteria and the existing criteria via simulations. Furthermore, Hawkins-Bradu-Kass data set was analyzed for illustration.

Keywords Model selection criteria, Regression diagnostics, Robust variable selection, Breakdown point.

AMS 2010 subject classifications 62J05, 62J20.

DOI: 10.19139/soic-2310-5070-642

1. Introduction

Variable selection is one of the important topics in regression modeling, it gains the interest of many authors. Beside the common stepwise deletion and subset selection others proposed penalized likelihood approach ([1]). Recently, [2] considered the problem of variable selection for ultrahigh-dimensional additive models, and [3] employed the firefly algorithm to select variables in count data regression.

For multiple linear regression model which is given in the form:

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i, \tag{1}$$

where, $\mathbf{x}_i = (\mathbf{x}_{i1}, ..., \mathbf{x}_{ip})^T$ is a vector, containing p explanatory variables, and y_i is the response variable, $\boldsymbol{\beta}$ is a vector of p parameters, and ε_i is the error component, that is independent and identically distributed (iid), with mean 0 and variance σ^2. There are different criteria for model selection, the most common are Mallows' Cp [4], Schwartz criterion (SIC) [5] and Akaike information criterion (AIC) [6], which are defined as the following form:

$$Z = G(SSE) + c. \tag{2}$$

Here the $G(SSE)$ is a function in the term of sum of square error, $SSE = \sum_{i=1}^{n} r_i^2$, with residual $r_i = y_i - \mathbf{x}_i^T \boldsymbol{\beta}$ and c is a constant. The $G(SSE)$ value equals SSE/σ^2, $\log(SSE/n)$ and $\ln(SSE/n)$ for the Cp, SIC and AIC respectively, where n is the sample size. In the classical criteria, the $\hat{\boldsymbol{\beta}}$ is the OLS-estimator corresponding to the

*Correspondence to: Shokrya Saleh (Email: salshekak@jazanu.edu.sa). Department of Mathematics, Jazan University, Jazan, Kingdom of Saudi Arabia

ISSN 2310-5070 (online) ISSN 2311-004X (print)

traditional square function. It is defined by

$$\hat{\boldsymbol{\beta}}_{OLS} = \arg\min \sum_{i=1}^{n} (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2. \tag{3}$$

Indirect approaches to robust model selection procedures are using a residual, r_i from robust fit. In robust model selection of [7], [8] and [9] where robust versions of AIC ,SIC and Cp ($RAIC$, $RSIC$ and RCp) are proposed considering the residual r_i following robust M-estimator, by choosing $\hat{\boldsymbol{\beta}}$ to minimize $\sum \rho(r_i)$, where, $\boldsymbol{\beta}$ is a vector of p parameters in linear regression model and $\rho(.)$ is a function less sensitive to outliers than squared, yielding the estimating equation $\sum \psi(r_i/\sigma^2)\mathbf{x}_i = 0$, where $\psi(.) = \rho'(.)$. M-estimators are efficient and highly robust to unusual values of y, but one rogue leverage point can break them down completely.

Whereas in recent years a good deal with outlier identification on direct approaches focused on the use of single-case diagnostic (see [10] [11], and [12]). A general idea to outlier diagnostic is to form a clean subset of data that is free of outliers. Let R be the set of indexes of the observations in the clean subset, y_R and $\mathbf{x}_R$ be the subsets of observations indexed by R, $\hat{\boldsymbol{\beta}}_R$ be estimated regression coefficients computed from fitting the model to the set R. And let SSE_R be the corresponding sum of squares residual that finds the estimates corresponding to the clean samples having the smallest sum of squares of residuals. As such, as expected, the breakdown point is 50%. When $R = n$, $\hat{\boldsymbol{\beta}}_R = \hat{\boldsymbol{\beta}}_{OLS}$. This study suggests using SSE_R in different model selection criteria.

The paper is organized as follows. In section 2, we indicate the extreme sensitivity of the exiting robust model to the leverage point, and discuss the robust diagnostic-variable selection criteria in Section 3. In Section 4 we review a popular regression diagnostic tool and its breakdown point which leads to a robust selection criteria. Section 5 presents the result for our simulation study and real data sets, while section 6 concludes the study.

2. Robust model selection criteria

The influence of leverage point on $RAIC$, $RSIC$ and RCp are illustrated through the presence of outliers in the X-direction. For simplicity, a set of independent random uniform variable $\mathbf{X}$ on [-2,2] was generated according to the simple regression model given as follows:

$$y_i = \mathbf{X}_i + \varepsilon_i, i = 1, ..., 19 \tag{4}$$

where, the ε_i are iid, normally distributed with expectation 0 and variance (0.1^2). The data has been presented in Table 1 and Figure 1.

For this, a point with coordinates $(0, x_{10})$ is added, Figure 2 shows the situation. Figure 3 shows that, the value of different criteria increases as the size of contamination in x_{10} increases, as expected, and if $\mathbf{x}$ are extremely large, then the values of $RIAC$, $RSIC$ and RCp change dramatically, and that mean reject the straight line. This shows the high sensitivity of $RIAC$, $RSIC$ and RCp selections procedures. Indeed, a very small change in the observations at $\mathbf{x} = 0$ has already the effect of changing the selected model.

Table 1. The data set

y_i	1.2	1.35	1.02	1.16	0.95	1.05	0.73	0.91	0.85	0	-0.88	-0.61	-0.81	-0.97	-1.18	-1.08	-0.99	-1.11	-1.14
$\mathbf{X}_i$	-1.2	-1.15	-1.1	-1.05	-1	-0.95	-0.9	-0.85	-0.8	x_{10}	0.8	0.85	0.9	0.95	1	1.05	1.1	1.15	1.2

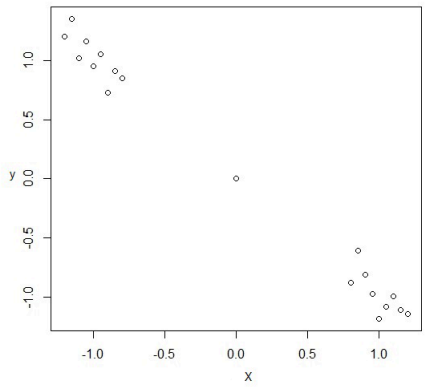

Figure 1. Scatter diagram of y versus X.

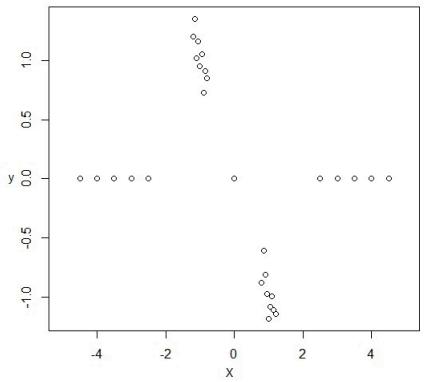

Figure 2. Data and positions for x_{10}

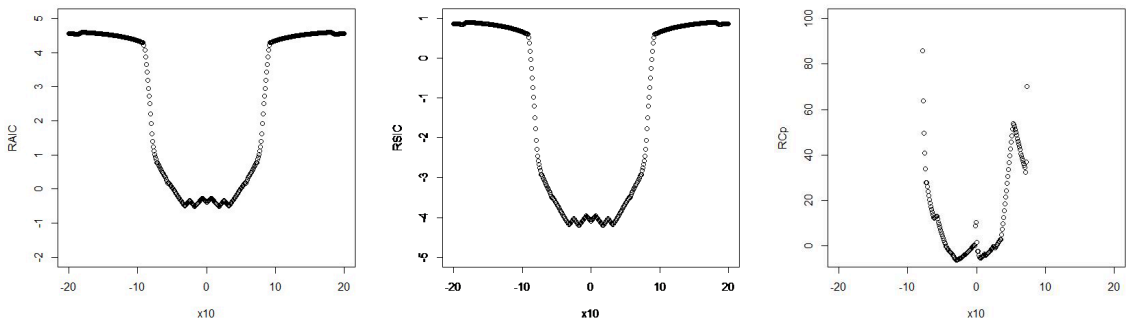

Figure 3. $RAIC$, $RSIC$ and RCp criteria for different value of x_{10}

3. The diagnostic version of model selection criteria

Consider the diagnostic sum of squares error SSE_R, by replacing the value of SSE in equation (2) in terms of SSE_R , the criteria in equation (2), can be expressed as follows:

$$Z_R = G(SSE_R) + c, \tag{5}$$

Then the new robust methods, AIC_R, Cp_R and SIC_R, can be expressed as follows:

$$AIC_R = \ln(SSE_R/R) + 2p, \tag{6}$$

$$Cp_R = SSE_{R_p}/(\hat{\sigma}^2_{full} - R + 2p), \tag{7}$$

$$SIC_R = \log(SSE_{R_p}/R) + (p\log(R))/R. \tag{8}$$

where SSE_R is compute from the diagnostic-OLS (OLS_R) estimator defined as:

$$\hat{\boldsymbol{\beta}}_{OLS_R} = \arg\min \sum_{i=1}^{R} (r^2_R(\boldsymbol{\beta}_R))_i. \tag{9}$$

Therefore, OLS_R corresponds to find the clean subset of R observations whose least squares fit produces the lowest sum of squared residuals, and has a high breakdown point. It is resistant to outliers, including leverage points. In equation (5), the estimates corresponding to the R samples are having the smallest sum of residuals. This would be the most direct implementation of the idea that one wants to find the model which fits best for the majority of the data. However, the distributional properties of OLS residuals are much better understood.

3.1. Breakdown point

Definition breakdown point of an estimate $\hat{\boldsymbol{\beta}}$ of the parameters $\boldsymbol{\beta}$ is the largest amount of contamination that the data may contain and even turn over some information about $\hat{\boldsymbol{\beta}}$. In other words, the breakdown point of an estimate $\hat{\boldsymbol{\beta}}$ shows the effects of replacing several data values by outliers [17]. Hence, the breakdown point for the regression estimator $\hat{\boldsymbol{\beta}}$ of the sample $Z = (\mathbf{X}, y)$ can be defined as:

$$\varepsilon^\star(\hat{\boldsymbol{\beta}}; Z) = \min\{m/n : \sup_{\tilde{Z}} \|\hat{\boldsymbol{\beta}}(Z')\|_2 = \infty\}, \tag{10}$$

where $\tilde{Z}$ are contaminated data obtained from Z by replacing m of the original n data be outliers. We obtained the following result from the breakdown point of the OLS_R estimator.

Remark 1. The breakdown point of the diagnostic-OLS estimator $\hat{\boldsymbol{\beta}}_{OLS_R}$ with subset size $R \leq n$ is given by

$$\varepsilon^\star(\hat{\boldsymbol{\beta}}_{OLS_R}; Z_R) = (n - R + 1)/n. \tag{11}$$

Applying Remark 1 to the OLS ($n - R = 0$) yields a finite sample breakdown point of

$$\varepsilon^\star(\hat{\boldsymbol{\beta}}_{OLS}; Z) = 1/n. \tag{12}$$

However, only one outlier can already let the OLS tend to infinity, this classical OLS comes from the use of squared residuals. Using other convex loss function, as we have done in the diagnostic-squared residuals, does solve the problem and also results in a breakdown point of $(n - R + 1)/n$. With smaller values of R, the breakdown point will be higher, and by making R small enough (k big enough), it may result in a breakdown point larger than 50%. Instead, we suggest to use this method with data containing no more than 25% outliers, then take a value of R equal to 0.75% of the sample size. Such that the diagnostic-estimate is based on a sufficiently large number of observations. This will be high efficiency, as will be shown in the numerical example (simulations).

4. Techniques used to identify vertical and leverage

In this section we introduce a different way of finding the clean subset R. The residual mean square, $\hat{\sigma}^2 = (y - \hat{y})/(n - p)$, the ordinary residual vector is defined as:

$$\hat{\varepsilon} = y - \hat{y} = (1 - H)y, \tag{13}$$

where H is the hat or leverage matrix that is considered as a symmetric matrix and contains the information on the influence of the response value y on the corresponding fitted values $\hat{y}_i = H_i^T$. In this equation, H_i^T is the ith row of H matrix and h_{ii} are the diagonal elements of H.

[10] suggested using the least median of squares (LMS) estimator to detect regression outliers. This method began by computing the residuals associated with LMS regression:

$$s = 1.4826 \left(1 + \frac{5}{(n - p - 1)}\right) \sqrt{M_r}, \tag{14}$$

where, M_r is the median of $r_1^2, ..., r_n^2$, and p is the number of predictors. However, a regression outlier is ith vector that satisfies $(r_i/s) > 2.5$.

Later, [11] introduced potentials as a single leverage deleted measure:

$$p_{ii} = \mathbf{x}_i^T (\mathbf{X}_{(i)}^T \mathbf{X}_{(i)})^{-1} \mathbf{x}_i, \tag{15}$$

where $X_{(i)}$ is the data matrix with the ith row deleted. A cut-off point for p_{ii} is $Median(p_{ii}) + 3MAD(p_{ii})$, where MAD is a median absolute deviation.

Well-known Mahalanobis (MD_i) distances are also suggested to apply as measures of leverage points in the literature (e.g. [12]). Another study [14] reviewed different types of residuals for the diagnostic purpose of which the most commonly used is Studentized residuals, define as:

$$t_i = \frac{y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}^{(-i)}}{\hat{\sigma}_{(i)} \sqrt{1 - h_{ii}}}, \tag{16}$$

an observation i is termed as an outlier if $|t_i| > c$, where c is a constant value $2 \leq c \leq 3$. According to [15], $DFFITS$ are introduced as:

$$DFFITS_i = \sqrt{\frac{h_{ii}}{1 - h_{ii}}} t_i. \tag{17}$$

Further, the authors recommended considering observations as influential if $|DFFITS_i| \geq 2\sqrt{p/n}$. However, the quantity $DFFITS$ is closely related to the well-known Cook's distance [16]. Cook's distance is defined as

$$CD_i = \frac{(\hat{\boldsymbol{\beta}}^{(-i)} - \hat{\boldsymbol{\beta}})^T (\mathbf{X}^T \mathbf{X})^{-1} (\hat{\boldsymbol{\beta}}^{(-i)} - \hat{\boldsymbol{\beta}})}{p\hat{\sigma}^2}. \tag{18}$$

The relationship between CD_i and $DFFITS_i$ is given as:

$$CD_i = \frac{\hat{\sigma}_{(i)}}{p\hat{\sigma}^2} DFFITS_i^2. \tag{19}$$

[13] suggested that h_{ii} with values less than 0.2 appear to be safe, values between 0.2 and 0.5 as being risky and values greater than 0.5 if possible, are better to be avoided by controlling the design matrix.

5. Numerical examples

A simulation study is carried out to investigate the performance of the AIC_R, Cp_R, and SIC_R statistic for detecting best variables in the regression model in equation (1) based on equations (6), (7) and (8).

The simulation study aims to compare the performance of the three proposed robust measures, namely AIC_R, Cp_R and SIC_R with nonCrobust measures AIC, Cp and SIC as well as robust measures based on M-estimation $RAIC$, RCp and $RSIC$, respectively.
In this simulation, 50 independent replicates of 3 independent uniform random variables on [-1,1] of $\mathbf{x}_{i1}$, $\mathbf{x}_{i2}$ and $\mathbf{x}_{i3}$, and 50 independent normally distributed errors $\varepsilon_i \sim N(0,1)$ were generated. The true model is given by $y_i = \mathbf{x}_{i1} + \mathbf{x}_{i2} + \varepsilon_i$, for $i = 1, ..., 50$ using two variables $\mathbf{x}_{i1}$ and $\mathbf{x}_{i2}$. In order to illustrate the robustness to outliers, we consider the following cases:

1. Vertical outliers (outliers in the y only),
2. Bad leverage points (outliers in some $\mathbf{x}$ only).
3. Good leverage points (outliers in both y and $\mathbf{x}$).

For all cases we introduce outliers into the data such that the percentages of contamination used are varied (0%, 5%, 10%, and 20%) of outliers from $N(50, 0.1^2)$ distribution. For each of these setting we simulate 1000 samples. We use the LMS given in equation (14) and potentials given in equation (15), to identify the verticals outliers and leverage points, respectively. The simulations were performed in R. To run the simulations, the R package rlm was used for the LMS ($lmsreg$).

Tables 2 to 4 shows several results as follows:
1. The classical methods AIC, Cp, and SIC work better than the robust methods for the data without outliers.
2. When the percentage of vertical increases from 5% to 20%, the classical methods tend to under fit (x_{i1} or x_{i2}). By contrast, the variable selection methods with the diagnostic tool (AIC_R, Cp_R, and SIC_R) perform well with reasonably good power. Whilst, the robust methods based on M-estimation ($RAIC$, RCp, and $RSIC$) continues to perform well until 20% contamination. Nonetheless, the proposed methods AIC_R, Cp_R, and SIC_R perform well compared to the $RAIC$, RCp, and $RSIC$ in both the uncontaminated and contaminated regression data.
3. In the presence of bad leverage point, the model selection criteria based on OLS and M-estimation often over fit or wrong fit in this case. Interestingly, the diagnostic tool based methods tend to correctly fit the true model more often.

The simulation results above illustrates that the performance of the proposed method (AIC_R, Cp_R, and SIC_R) yields a comparable power of selection, correct fit of those obtain in classical or $RAIC$, RCp, and $RSIC$ approaches for both cases in presence of vertical and leverage points.

5.1. Example

In this example, Hawkins-Bradu-Kass dataset is used. This data available from the R library wle as data (artificial). Artificial data set containing 75 observations with 10 outliers (cases 1 to 10) and 14 high leverage points (cases 1 to 14). Scatter plots of Y on each three $\mathbf{X}'s$ as shown in Figure 4,clearly separate 10 high leverage outliers, 4 high leverage points and 61 clean observations. The robust regression model based on M-estimator was fitted to the data set. The parameter estimates are given by, intercept = -0.7848, $\boldsymbol{\beta}_{Hawkins} = 0.1791$, $\boldsymbol{\beta}_{Bradu} = 0.0062$, $\boldsymbol{\beta}_{Kass} = 0.2715$. Further, the correlation matrix of data is given by

$$\begin{pmatrix} 1 & 0.9450 & 0.9606 \\ 0.9450 & 1 & 0.9786 \\ 0.9606 & 0.9786 & 1 \end{pmatrix},$$

which suggest that the data seem to be highly concentrated. All 2^3 possible models fitted with a combination of any of these covariates and computed several model selection methods values for each model (see Tables 5 to 7).

Table 2. Percentage of different variable select criterion from simulated data, with vertical outliers

ϵ (%)	No. of Verticals	Set of Variables	AIC_{OLS}	AIC_M	AIC_R	Cp_{OLS}	Cp_M	Cp_R	SIC_{OLS}	SIC_M	SIC_R
0	0	$Intercept$	0	0	0	0	0	0	0	0	0
		x_1	0.4	21.2	3.4	0.6	8.0	3.8	2.4	10.4	5.6
		x_2	1.2	21.6	3.6	1.2	8.8	0.0	3.4	11.2	7.4
		x_3	0.2	2.0	0.2	0.2	3.2	0.4	0.2	3.8	0.4
		x_1, x_2	**82.2**	**54.4**	**59.0**	**81.0**	**36.6**	**62.4**	**88.2**	**40.4**	**63.8**
		x_1, x_3	0.0	0.0	2.0	0.0	6.6	2.2	0.0	5.8	3.4
		x_2, x_3	0.2	0.0	2.2	0.2	8.0	0.0	0.2	6.6	1.8
		x_1, x_2, x_3	15.8	0.0	29.4	16.8	28.8	31.2	5.6	21.8	17.6
5	2	$Intercept$	0	0	0	0	0	0	0	0	0
		x_1	6.6	23.2	1.8	6.8	5.8	2.2	16.6	7.8	6.6
		x_2	7.2	25	2.0	8.2	8.8	0.0	15.4	12.6	6.0
		x_3	0.2	2.0	0.4	0.4	2.8	1.2	0.4	3.6	1.0
		x_1, x_2	**70.8**	**49.8**	**63.2**	**70.4**	**34.8**	**64.4**	**63.2**	**39.6**	**67.4**
		x_1, x_3	1.2	0.0	1.0	1.2	7.6	1.0	0.8	6.8	1.0
		x_2, x_3	0.4	0.0	1.4	0.2	8.6	0.0	0.2	7.4	2.2
		x_1, x_2, x_3	12.4	0.0	29.8	12.8	31.6	31.2	3.4	22.2	15.8
10	5	$Intercept$	0	0	0	0	0	0	0	0	0
		x_1	17.8	24.8	3.0	26.0	7.2	3.4	35.2	11.6	7.6
		x_2	20.2	26.2	3.0	30.6	5.4	0.0	37.8	9.0	6.2
		x_3	7.0	1.6	0.2	15.2	0.2	0.6	17.6	0.2	0.6
		x_1, x_2	**15.2**	**47.4**	**67.0**	**15.2**	**70.4**	**69.0**	**6.0**	**69.2**	**69.2**
		x_1, x_3	6.4	0.0	2.0	6.6	1.2	2.0	1.6	0.6	2.0
		x_2, x_3	3.8	0.0	1.6	4.0	1.8	0.0	1.4	1.8	2.4
		x_1, x_2, x_3	2.4	0.0	23.2	2.4	13.8	25.0	0.4	7.6	12.0
20	10	$Intercept$	0	0	0	0	0	0	0	0	0
		x_1	16.6	36.6	4.0	27.4	11.2	4.6	32.4	14.4	9.8
		x_2	21.0	34.2	5.0	35.4	11.0	0.0	41.0	14.6	10.0
		x_3	7.4	1	0.0	17.2	0.2	0.2	20.0	0.6	0.4
		x_1, x_2	**8.4**	**28.2**	**68.4**	**8.4**	**70.4**	**73.2**	**3.2**	**67.0**	**67.4**
		x_1, x_3	4.4	0.0	2.0	4.6	0.8	1.8	1.4	0.6	1.6
		x_2, x_3	4.8	0.0	1.8	5.0	0.4	0.0	1.4	0.4	1.4
		x_1, x_2, x_3	2.0	0.0	18.6	2.0	6.0	20.2	0.6	2.4	9.4

Table 3. Percentage of different variable select criterion from simulated data, with bad leverage points

ϵ (%)	No. of Verticals	Set of Variables	AIC_{OLS}	AIC_M	AIC_R	Cp_{OLS}	Cp_M	Cp_R	SIC_{OLS}	SIC_M	SIC_R
5	2	$Intercept$	0	0	0	0	0	0	0	0	0
		x_1	5.4	31.8	2.8	8.2	12.0	3.2	17.4	17.2	6.2
		x_2	3.8	35.2	3.4	9.4	11.6	0.0	19.4	18.4	7.6
		x_3	0.8	33.0	0.2	7.0	14.0	0.4	15.6	20.2	0.4
		x_1, x_2	**0.8**	**0.0**	**60.4**	**1.6**	**2.4**	**63.8**	**0.4**	**1.0**	**65.0**
		x_1, x_3	32.8	0.0	2.0	1.6	2.6	1.6	0.4	1.0	1.4
		x_2, x_3	36.4	0.0	2.4	1.6	2.8	0.0	0.6	1.8	2.4
		x_1, x_2, x_3	15.4	0.0	28.6	70.6	54.6	31.0	46.2	40.4	17.0
10	5	$Intercept$	0	0	0	0	0	0	0	0	0
		x_1	2.2	34.4	4.4	6.6	12.2	4.2	15.6	18.6	8.4
		x_2	3.0	31.6	5.0	7.2	13.0	0.0	16.4	18.6	10.0
		x_3	1.4	34.0	0.6	8.0	12.0	1.2	17.4	18.6	1.2
		x_1, x_2	**0.0**	**0.0**	**57.4**	**0.8**	**2.0**	**62.0**	**0.2**	**1.4**	**62.6**
		x_1, x_3	36.6	0.0	2.0	1.4	1.8	2.2	1.0	0.8	1.2
		x_2, x_3	36.6	0.0	3.4	1.4	2.2	0.0	0.0	1.6	3.4
		x_1, x_2, x_3	15.8	0.0	27.2	74.6	56.8	30.4	49.4	40.4	13.2
20	10	$Intercept$	0	0	0	0	0	0	0	0	0
		x_1	2.2	33.0	4.4	5.8	11.4	8.6	15.6	17.6	8.8
		x_2	2.4	34.4	3.2	10.6	12.2	0.0	19.6	19.2	8.6
		x_3	1.6	32.6	2.2	8.4	14.6	5.4	17.0	19.8	6.8
		x_1, x_2	**0.2**	**0.0**	**20.0**	**2.0**	**1.4**	**22.6**	**0.8**	**0.8**	**20.4**
		x_1, x_3	38.0	0.0	25.6	1.4	2.4	28.2	0.2	1.2	25.6
		x_2, x_3	35.0	0.0	20.0	1.2	2.4	0.0	0.2	1.2	19.8
		x_1, x_2, x_3	14.6	0.0	18.8	70.6	55.6	35.2	46.6	40.2	10.0

The best three selected models based on each version of AIC, Cp, and SIC methods are given in Table 8.

We observe from the results that all of the commonly used measures of selection model fail to focus on best variables. Tables 5 to 7 present the commonly used model selection AIC, Cp, and SIC together with robust $RAIC$, RCp, $RSIC$ methods and AIC_R, Cp_R, and SIC_R. It is clear from the results presented in this table that variable selected by the classical selection methods are not correct enough. Though the robust model selection

Table 4. Percentage of different variable select criterion from simulated data, with good leverage points

ϵ (%)	No. of Verticals	Set of Variables	AIC_{OLS}	AIC_M	AIC_R	Cp_{OLS}	Cp_M	Cp_R	SIC_{OLS}	SIC_M	SIC_R
5	2	$Intercept$	0	2.6	0	0	0	0	0	0	0
		x_1	0.0	25.3	3.4	0	1.0	3.2	0	1.2	6.4
		x_2	0.0	25.1	1.8	0	0.0	0.0	0	1.0	3.8
		x_3	0.0	2.0	0.0	0	0.2	0.2	0	0.4	0.6
		x_1, x_2	**85.6**	**47.0**	**64.2**	**0**	**0.2**	**65.8**	**0**	**0.0**	**70.4**
		x_1, x_3	0.2	0.0	0.8	0	0.2	1.0	0	0.2	1.8
		x_2, x_3	0.2	0.0	1.6	0	0.0	0.0	0	0.0	1.6
		x_1, x_2, x_3	14.0	0.0	28.2	100	98.4	29.8	100	97.2	15.4
10	5	$Intercept$	0	2.0	0	0	0	0	0	0	0
		x_1	0.0	26.6	2.4	0	0.0	2.6	0	0.0	5.0
		x_2	0.0	26.3	2.8	0	0.0	0.0	0	0.2	5.2
		x_3	0.0	0.0	0.0	0	0.2	0.2	0	0.2	0.2
		x_1, x_2	**81.0**	**44.2**	**62.8**	**0**	**0.0**	**67.2**	**0**	**0.0**	**70.2**
		x_1, x_3	0.2	0.0	1.2	0	0.0	1.0	0	0.0	1.4
		x_2, x_3	0.0	0.0	0.4	0	0.0	0.0	0	0.2	1.2
		x_1, x_2, x_3	18.8	0.0	30.2	100	99.8	29.0	100	99.4	16.8
20	10	$Intercept$	0	0	0	0	0	0	0	0	0
		x_1	0.0	30.0	2.6	0	0.0	2.6	0	0.0	6.0
		x_2	0.0	36.6	1.8	0	0.2	0.0	0	0.6	5.6
		x_3	0.0	1.0	0.2	0	0.0	0.4	0	0.0	0.6
		x_1, x_2	**85.2**	**32.4**	**66.8**	**0**	**0.0**	**70.2**	**0**	**0.0**	**71.6**
		x_1, x_3	0.0	0.0	1.2	0	0.0	1.2	0	0.0	1.2
		x_2, x_3	0.0	0.0	2.6	0	0.0	0.0	0	0.0	2.4
		x_1, x_2, x_3	14.8	0.0	24.6	100	99.8	25.6	100	99.4	12.6

based on M-estimation is also sensitive to high leverage points, the table shows that they fail to choose the first variable (Hawkins). Robust model selection based on the diagnostic tool suggests that first observation (Hawkins) is best variable. When we apply the diagnostic checking based on LMS and hat matrix cases 1 to 14 return to the contamination subset and thus the AIC_R, and Cp_R finally identify the first variable as best variable. And SIC_R tends to chose Kass as best variable.

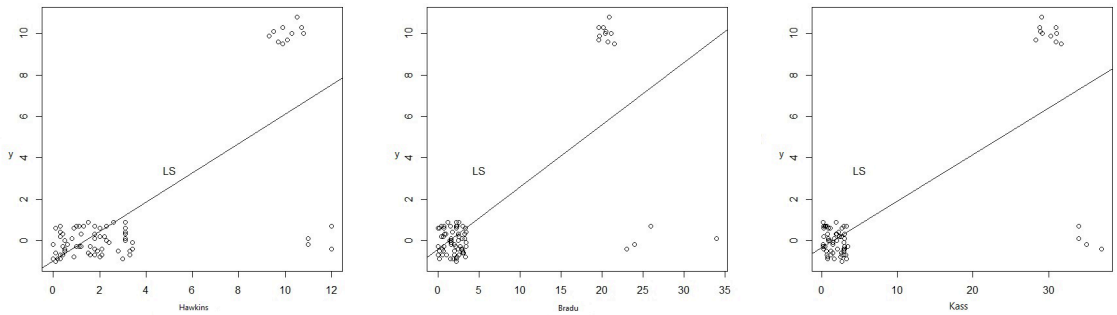

Figure 4. $RAIC$, $RSIC$ and RCp criteria for different value of x_{10}

Table 5. Values of the classical AIC, and robust $RAIC$, and AIC_R statistics for Hawkins-Bradu-Kass data

Selected Variables	AIC	$RAIC$	AIC_R
(y, Hawkins)	5.68	4.81	**2.74**
(y, Bradu)	5.79	4.14	2.80
(y, Kass)	**5.63**	**3.62**	2.77
(y, Hawkins, Bradu)	7.68	5.62	4.73
(y, Hawkins, Kass)	7.62	5.79	4.69
(y, Bradu, Kass)	7.57	5.67	4.75
(y, Hawkins, Bradu, Kass)	9.56	7.76	6.66

Table 6. Values of the classical Cp, and robust RCp, and Cp_R statistics for Hawkins-Bradu-Kass data

Selected Variables	Cp	RCp	Cp_R
(y, Hawkins)	5.30	130.77	**1.19**
(y, Bradu)	8.90	32.55	2.26
(y, Kass)	17.93	**-9.72**	3.16
(y, Hawkins, Bradu)	**2.93**	-7.53	2.03
(y, Hawkins, Kass)	6.68	4.13	3.14
(y, Bradu, Kass)	10.84	-4.38	4.26
(y, Hawkins, Bradu, Kass)	4.00	4.00	4.00

Table 7. Values of the classical SIC, and robust $RSIC$, and SIC_R statistics for Hawkins-Bradu-Kass data

Selected Variables	SIC	$RSIC$	SIC_R
(y, Hawkins)	1.79	0.9	-1.04
(y, Bradu)	1.90	0.26	-1.02
(y, Kass)	**1.75**	**-9.72**	**-1.063**
(y, Hawkins, Bradu)	1.85	-0.20	-0.97
(y, Hawkins, Kass)	1.80	-0.03	-1.01
(y, Bradu, Kass)	**1.75**	-0.15	-0.99
(y, Hawkins,Bradu, Kass)	1.79	-0.15	-0.94

Table 8. Hawkins-Bradu-Kass, the selected best variables from best three models based on different classical criteria, robust criteria with M-estimation, and robust criteria using a deletion estimate of the scale

Criteria	Selected variables		
	Best model	Second best model	Third best model
AIC	Kass	Hawkins	Bradu
$RAIC$	Kass	Bradu	Hawkins
AIC_R	Hawkins	Kass	Bradu
Cp	Hawkins, Bradu	Hawkins, Bradu, Kass	Hawkins
RCp	Kass	Hawkins, Bradu	Bradu, Kass
Cp_R	Hawkins	Hawkins, Bradu	Bradu
SIC	Kass	Bradu, Kass	Hawkins
$RSIC$	Kass	Hawkins, Bradu	Bradu, Kass
SIC_R	Kass	Hawkins	Bradu

6. Conclusion

Diagnostic regression measures are robust regression methods, which are frequently used in practice. Nevertheless, they are not commonly used in selection models. This paper had introduced variable selection criteria based on a diagnostic scale, which are robust against outliers and leverage points. The simulation results had illustrated good performance of the proposed diagnostic variable selection criteria.

REFERENCES

1. J. Fan, and R. Li, *Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties*, J.Amer. Statist. Assoc, vol. 96, no. 452, pp. 1348-1360, 2001.
2. M. Kazemi, D. Shahsavani, and M. Arashi, *Variable Selection and structure identification for ultrahigh-dimensional partially linear additive models with application to cardiomyopathy microarray data*, Statistics, Optimization & Information Computing, vol. 6, no. 452, pp. 373C-382, 2018.
3. Z. Y. Algamal, *Variable Selection in Count Data Regression Model based on Firefly Algorithm*, Statistics, Optimization & Information Computing, vol. 7, pp. 520-529, 2019.
4. Mallows, Colin L, *Some comments on C p*, Taylor & Francis Group, vol. 15, no. 4, pp. 661–675, 1973.
5. Schwarz, Gideon, *Estimating the dimension of a model*, The annals of statistics, vol. 6, no. 2, pp. 461–464, 1978.
6. Akaike, Hirotogu, *Information theory and an extension of the maximum likelihood principle*, Springer, 199–213, 1998.
7. Rousseeuw and Peter, *Multivariate estimation with high breakdown point*, Reidel, 1985.
8. Machado, Jose AF, *Robust model selection and M-estimation*, Econometric Theory, vol. 9, no. 03 pp. 478–493, 1993.
9. Ronchetti, Elvezio and Staudte, Robert G, *A robust version of Mallows's Cp*, Journal of the American Statistical Association, vol. 89, no. 426, pp. 550–559, 1994.
10. Rousseeuw and Zomeren, Bert C, *Unmasking multivariate outliers and leverage points*, Journal of the American Statistical Association, vol. 85, no. 411 pp. 633–639, 1990.
11. Hadi, Ali S, *A new measure of overall potential influence in linear regression*, Computational Statistics & Data Analysis, vol. 14, no. 1 pp. 1–27, 1992.
12. Rousseeuw, Peter J and Leroy, Annick M, *Robust regression and outlier detection*, John Wiley & Sons, vol. 589, 2005.
13. Huber, Peter J, *Robust statistics*, Springer & Data Analysis, 2011.
14. Ryan, Thomas P, *Modern regression methods*, John Wiley & Sons, vol. 655, 2008.
15. Belsley, David A and Kuh, Edwin and Welsch, Roy E, *Regression diagnostics: Identifying influential data and sources of collinearity*, John Wiley & Sons, vol. 571, 2005.
16. Cook, R Dennis, *Detection of influential observation in linear regression*, Technometrics, pp. 15–18, 1977.
17. Maronna, Ricardo and Martin, Douglas and Yohai, Victor, *Robust statistics*, John Wiley & Sons, Chichester. ISBN, 2006.
18. Serneels, Sven and Filzmoser, Peter and Croux, Christophe and Van Espen, Pierre J, *Robust continuum regression*, Chemometrics and Intelligent Laboratory Systems, vol. 76, no. 2 pp. 197–204, 2005.
19. Lee, Jong Soo and Cox, Dennis D, *Robust smoothing: Smoothing parameter selection and applications to fluorescence spectroscopy*, Computational statistics & data analysis, vol. 54, no. 12 pp. 3131–3143, 2010.
20. Leung, Denis Heng-Yan, *Cross-validation in nonparametric regression with outliers*, Annals of Statistics, pp. 2291–2310, 2005.
21. Morell, Oliver and Otto, Dennis and Fried, Roland, *On robust cross-validation for nonparametric smoothing*, Computational Statistics, vol. 28, no. 4 pp. 1617–1637, 2013.