Mathematical Statistician and Engineering Applications
ISSN: 2326-9865
https://doi.org/10.17762/msea.v71i1.5

ROBUST VARIABLE SELECTION BASED ON SCHWARZ INFORMATION CRITERION FOR LINEAR REGRESSION MODELS

SHOKRYA SALEH

Department of Mathematics, Jazan University, Saudia Arabia, Jazan

E-mails: salshekak@jazanu.edu.sa.

Abstract

Schwarz information criterion (SIC) is a popular tool to select the best variables in regression data sets. However, SIC defined using an unbounded estimator (Least Squares (LS)) which is very sensitive to the presence of outlying observations, especially bad leverage points. Thus, robust variable selection based on SIC for linear regression models is in need. This paper study the robust properties of SIC derives its influence function and proposes robust SIC based on the MM-estimation scale, aim to produce criterion which is effective in selecting accurate models in the presence of vertical outliers and high leverage points. The advantages of the proposed robust SIC is demonstrated through simulation study and analysis of a real data set.

Keywords: Robust variable selection, robust regression, Influence function, Schwarz information criterion.

1 Introduction

This paper considers the problem of robust and selection variables for linear regression models. In modern regression data sets, outliers are commonly encountered in applications, which may appear either in response variables (vertical outliers) or in the predictors (leverage points). In this type of data set, it is difficult to select the best variables using criteria based on the classical estimator (LS). Traditional selection criteria have a bad behavior with regards to robustness when vertical outliers in the data sets (see [1] and [2]). Moreover, they cannot be selected appropriate models for data with leverage points. Thus, robust variable selection methods for regression data are in need. Robust variable selection is one of the important topics in regression modeling; it gains the interest of many authors. For an instant, robust Mallow's Cp (RCp) proposed by [5], robust Akaike information criterion ($RAIC$) proposed [6] and robust R-squared proposed by [7].
The Bayesian information criterion (BIC) proposed by [3] is one of the commonly used

Mathematical Statistician and Engineering Applications
ISSN: 2326-9865
https://doi.org/10.17762/msea.v71i1.5

criteria in model selection in linear regression. For a more general situation, [4] uses a Bayesian approach with a penalty term of the form $(p\log(n))/n$, where n is the sample size, and p is the model dimension. Consider a linear regression model of the form

$$y_i = \mu + \mathbf{X}_i^T \boldsymbol{\beta} + \epsilon_i, \tag{1}$$

where μ is the intercept parameter, $\mathbf{X}_i = (\mathbf{x}_{i1}, ..., \mathbf{x}_{ip})^T$ is a vector contains p explanatory variables, 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. The classical SIC based on LS estimate is defined as

$$SIC_{LS} = \log\left(SSE_p/n\right) + \left(p\log(n)\right)/n, \tag{2}$$

where $SSE_p = \sum_{i=1}^n r_i^2$, is the sum of squares error for sub model with p variables and the residual $r_i = y_i - \hat{\mu}_{LS} - \mathbf{X}_i^T \hat{\boldsymbol{\beta}}_{LS}$. Therefore, models with values of SIC_{LS} small will be preferred. Since the LS estimator is vulnerable in the presence of outliers, it is not surprising that SIC_{LS} inherits this problem. However, a robust version of SIC based on M-estimators ([8]) proposed by [9], in this method replaced the squared residuals with a robust function ρ and subsequently derived, $SIC_M = \sum_{i=1}^n \rho\left(r_i/\sigma\right) + \left(p\log(n)\right)/n$, where ρ is a known function. Unfortunately, this criterion is not robust concerning contaminations in the predictor variables.

[10] proposed a robust version of SIC based on Least Trimmed Squares estimator (LTS) ([11]), named SIC_{LTS} criterion. In a simulation study, [10] show that the SIC_{LTS} can be robust for contamination in both the response and predictor variables. [10] discussed the influence of outliers on SIC criterion, but the LTS is highly inefficiency estimator when all the observations satisfy the regression model with normal errors.

[12] purpose MM-estimator of regression which having simultaneously, high breakdown point and high efficiency under normal errors; this estimator robust in a variety of contamination scenarios. However, MM-estimation is a combination of high breakdown value estimation and efficient estimation. MM-estimator does not use in SIC criterion for variable selection aim. The purpose of this paper is to present SIC_{MM} criterion for a robust variable selection criterion based on MM-scale estimates. The robust SIC_{MM} allows to choose the best models, which fit the majority of the data by taking into account the presence of outliers and possible departures from the normality assumption on the error distribution.

The paper is organized as follows: Section 2 reviews the definition and some of the most important properties of MM-estimator in regression models. Section 3 define SIC_{MM} criterion, study their robust properties, and describe an algorithm to compute SIC_{MM} criterion. A simulation is conducted to study the performance of the proposed robust criterion in Section 4. Section 5 applies the robust criterion to the real data set. Finally, the concluding remark is present in Section 6.

2 MM-estimates

MM-estimators proposed by [12] has become increasingly popular and one of the most commonly employed robust regression techniques. The MM-estimators reach a high level of robustness as well as high efficiency, by combining the properties of M-estimators ([8]) and S-estimators ([13]). The MM-estimators defined in three stages as follows:
Stage 1: take an initial estimate $\hat{\boldsymbol{\beta}}_0$ of $\hat{\boldsymbol{\beta}}$ in Equation (1) with a high breakdown point, possibly 0.5. The LTS estimation can be selected of $\hat{\boldsymbol{\beta}}_0$.
Stage 2: compute the residuals, $r_i = y_i - \hat{\mu}_0 - \hat{\boldsymbol{\beta}}_0 \mathbf{X}_i^T$ and compute the M-scale $\sigma(r_i(\hat{\boldsymbol{\beta}}_0))$, defined as the value of σ which is the solution of

$$\frac{1}{n}\sum_{i=1}^{n} \rho_0 \left((r_i(\hat{\boldsymbol{\beta}}_0) \right) / \sigma) = b,$$

where b is constant defined by $E_\Phi \left(\rho(r_i(\hat{\boldsymbol{\beta}}_0) \right) = b$, where Φ stands for the standard normal distribution. Using a function ρ_0 where satisfying following assumption (A_1): $\rho_0(0) = 0$, $\rho_0(-u) = \rho_0(u)$, for $0 \leq u \leq v$ implies $\rho_0(u) \leq \rho_0(v)$, ρ_0 is continuous , if $a = \sup \rho_0(u)$, then $0 \leq a \leq \infty$, if $\rho_0(u) < a$ and $0 \leq u < v$, then $\rho_0(u) \leq \rho_0(v)$. Using a constant b such that $b/a = 0.5$, this implies that this scale estimate has breakdown point equal to 0.5.
Stage 3: Let ρ_1 be another function satisfying: assumption (A1), $\rho_1(u) \leq \rho_0(u)$ and $\sup \rho_1(u) = \sup \rho_0(u) = a$. However, if $\psi_1 = \rho_1'$, then the MM-estimate ($\hat{\boldsymbol{\beta}}_{MM}$) is defined as any solution of $\sum_{i=1}^{n} \psi_1 \left(r_i/\sigma \right) \mathbf{X}_i = 0$,. $\hat{\boldsymbol{\beta}}_{MM}$ obtained with iteratively reweighted least squares (IRWLS). [12] proved that MM-estimators are strongly consistent for $\hat{\boldsymbol{\beta}}_0$, besides, MM-estimator has simultaneously the two following properties:

1. Normal asymptotic efficiency.
2. Breakdown point greater than or equal to that of the initial estimator.

However, MM-estimator have the highest possible breakdown point equal to 50% (see [14]).

3 SIC_{MM} criterion for variable selection in linear regression

This section discusses the possibility of extending the idea of using robust MM-estimators in the SIC. The SIC method is expressed in terms of the variance, which are computed in LS or robust method such as M- or LTS- estimation. [10] showed by derived the influence function of the SIC criterion that, the robustness of the SIC criterion will depend heavily on the robustness of the scale. In this study, instead of working with theses scales, a high breakdown point, and efficient MM-estimators for the SIC criterion will use. This, in turn, reduces the effect of outliers and leverage points. Given scale

ISSN: 2326-9865
https://doi.org/10.17762/msea.v71i1.5

Table 1: The simulated data set.

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

estimate of errors defined by $S = SSE_p/(n-p)$, with $r_i = y_i - \hat{\mu}_{MM} - \mathbf{X}_i^T \hat{\boldsymbol{\beta}}_{MM}$, then SIC_{MM} criterion define as

$$SIC_{MM} = \log\left(\frac{(n-p)S^2}{n}\right) + \frac{p\log(n)}{n}. \tag{3}$$

The small value of SIC_{MM} reveals that the explanatory variables adequately explain the distribution of y. Following same as experiment in [10], a set of independent random uniform variable $\mathbf{X}$ on [-2,2] was generated according to the simple regression model, $y_i = \mathbf{X}_i + \epsilon_i$, $i = 1, ..., 19$, where, ϵ_i are iid, normally distributed with expectation 0 and variance (0.1^2), the data has been presented in Table 1. The purpose of using this experiment to show the influence of an outlier on SIC_{MM}, this is illustrated through the presence of outliers in the Y -direction (vertical outlier) or in the X-direction (leverage point). For this, a point with coordinates $(0, y_{10})$ is added, where the values of y range between (-1.5,3). A similar approach is performed for leverage points, that is, replacing the value x with $(0, x_{10})$, Figure 1 shows the situations of y_{10} and x_{10}.
Figure 2 shows the results where the SIC_{MM} shows a very robust behavior; there is only a slight loss in criteria, becoming constant when the outlier moves further away from the origin. Based on these results, it is evident that SIC_{MM} show robust behavior in the presence of verticals or leverage point. In Section 4 simulation study and real data set illustrations clearly behavior of the proposed SIC_{MM}.

3.1 Properties of the proposed robust SIC_{MM} criterion

3.1.1 Influence function

Consider the linear regression model in Equation (1) and assume that the distribution of errors are satisfying $F_\sigma(\mathbf{X}) = F_0(\mathbf{X}/\sigma)$, where σ is the residual scale parameter, and F_0 is symmetric, with a strictly positive density function.

Let $\mathbf{X}$ and y be independent stochastic variables with distribution H. The functional T is Fisher-consistent for the parameters $(\mu, \boldsymbol{\beta})$ at the model distribution H, which is as

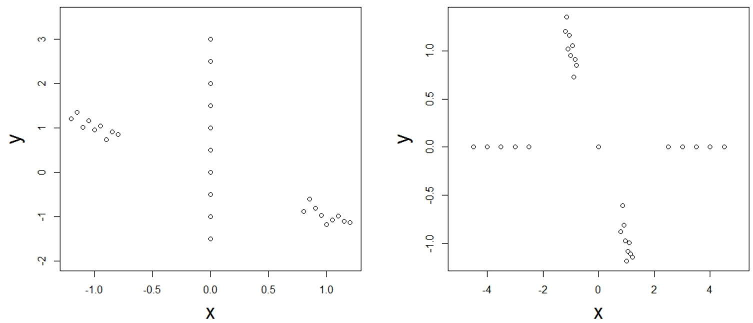

Figure 1: Data and positions for y_{10} (left) and x_{10} (right).

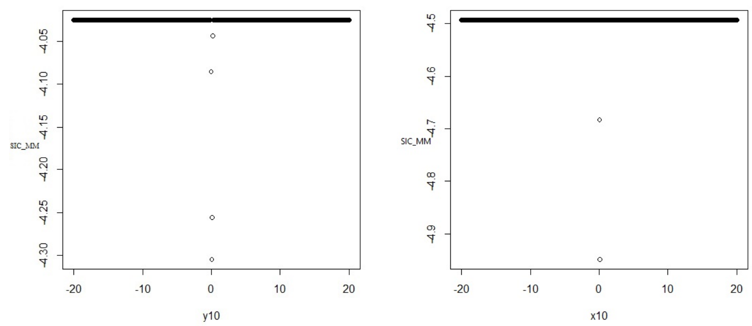

Figure 2: Effect of adding one observation y_{10} (left) and x_{10} (right) to the values of SIC_{MM}.

follows:

$$T(H) = \left[\begin{array}{c} a(H) \\ \mathbf{b}(H) \end{array} \right] = \left[\begin{array}{c} \mu \\ \boldsymbol{\beta} \end{array} \right]. \tag{4}$$

For a Fisher-consistent scale estimator, $T(F_\sigma) = \sigma$, for all $\sigma > 0$. [14] defined the influence function of T at the distribution F as,

$$IF((\mathbf{X}, y), T, H) = \lim_{\epsilon \to 0} \frac{T((1-\epsilon)H + \epsilon\Delta_{(\mathbf{X},y)}) - T(H)}{\epsilon} = \frac{\partial}{\partial\epsilon}(T(\Delta_{(\mathbf{X},y)})). \tag{5}$$

where $T(H)$ is the function defined in the solution of the objective model and $\Delta_{(\mathbf{X},y)}$ is the distribution contains outliers. The influence function measures the effect of possible outliers in the SIC_{MM} criterion. It gives the amount of change in the model selection criterion estimator, caused by an infinitesimal amount of contamination at $(\mathbf{X}, y)$. Theorem 1 in [10] derived the influence function of SIC based on scale estimates S as follows:

$$IF((\mathbf{X}, y), SIC_S, H) = 2n/(n-p)IF(r_i/\sigma_S, \hat{\sigma}_S^2, F_0), \tag{6}$$

which is bounded in both Y and X directions, as $IF(r_i/\sigma, \hat{\sigma}_S^2, F_0)$ is bounded. Follows immediately from (6), the influence function of SIC_{MM} is

$$IF((\mathbf{X}, y), SIC_{MM}, H) = 2n/(n-p)\psi_1(r_i)\mathbf{X}_i\sigma_0^2(B(\psi_1, F_0)V)^{-1}, \tag{7}$$

ISSN: 2326-9865
https://doi.org/10.17762/msea.v71i1.5

where $V = E_{G_0}(\mathbf{X}_i\mathbf{X}_i^T)$ with G_0 has second moment, $B(\psi_1, F_0) = E_F\left(\psi_1(\frac{r_i}{\sigma_0})\right)$ and F is the distribution of the error r_i. Whereas, the influence function for the proposed criterion is bounded and note that a large zone of vertical outliers have zero influence, even when they are bad leverage points.

3.1.2 The gross-error sensitivity of SIC_{MM} criterion:

[15] defined the gross-error sensitivity of an estimator T at a distribution F by

$$\gamma^\star = \sup_{\mathbf{X}} |IF(\mathbf{X}; T, F)|.$$

By taking the supreme over all $\mathbf{X}$ for which the $IF(\mathbf{X}; T, F)$ exists, gross-error sensitivity measures the worst possible influence on an estimator by an arbitrary infinitesimal contaminant. If the gross-error sensitivity is unbounded, $\gamma^\star = \infty$, then the estimator is completely intolerant of outliers; a single outlier can ruin the estimator.
According to this definition, the gross-error sensitivity of the SIC_{MM} criterion is defined as the supreme influence that observation can have. If $\hat{\boldsymbol{\beta}}_{MM} = 0$, then $IF = 0$, so it is assumed that $\hat{\boldsymbol{\beta}}_{MM} \neq 0$ then, if $\mathbf{X}$ tend to ∞, the gross-error sensitivity of will turn into:

$$\gamma^\star(SIC_{MM}, F) = \sup(\mathbf{X}, y) IF((\mathbf{X}, y), SIC_{MM}, H) = 2n(n-p)E_{F_0}[\rho_1(\epsilon)\epsilon] \cdot \rho_1(\infty). \quad (8)$$

Briefly, if $\mathbf{X}$ tends to infinity, both LS and M-estimators gain ρ function yields high gross-error sensitivity. On the other hand, MM-estimator compute with ρ function which yields the lowest $\gamma^\star$.

4 Simulation study

4.1 Settings

A simulation study was carried out to investigate the performance of the proposed robust SIC_{MM} criterion. Furthermore, to compare this criterion with existing robust criteria, SIC_{LTS} and SIC_M, and classical SIC_{LS} criterion. For simplicity, considering the case when $p = 3$, hence, the following set of parameters have to be estimated: $(\mu, \beta_1, \beta_2, \beta_3)$ and the set of different correlated random errors ϵ_i from the independent Normal distribution with mean 0 and variances $\sigma^2 = 0.7$.
The regression variables $\mathbf{x}_{i1}$, $\mathbf{x}_{i2}$ and $\mathbf{x}_{i3}$ are generated in two different cases:
Case 1: independent uniform random variables on [-1, 1] .
Case 2: correlated multivariate normal distribution, $N(0, \Sigma_r)$, for some $r \geq 0$, the variance matrix of the variables is defined by $\Sigma_{r,i,j} = r^{|i-j|}$ for $1 \leq i$, $j \leq 3$, $r = 0.03, 0.1, 0.5$. Then the true model is given by: $y_i = \mu + \mathbf{x}_{i1} + \mathbf{x}_{i2} + \epsilon_i$. We then introduce vertical and leverage outliers into the data such that the percentages of contamination used are c%= 10%, 20%, 30% and 40% from two different sample sizes, namely n= 50 and 100 . To investigate the robustness of the criteria against vertical and leverage outliers, the following scenarios were considered:

(a) no contamination,
(b) vertical outliers (outliers in some y_i only),
(c) good leverage points (outliers in the y_i and $\mathbf{X}$),
(d) bad leverage points (outliers in some $\mathbf{X}$ only).
For vertical outliers, randomly generated different percentage of outliers from $N(50, 0.1^2)$ for each of the simulated cases. For a good leverage point, considered the different percentages of outliers on the variables $\mathbf{x}_1$ and $\mathbf{x}_2$ are generated from $N(100, 0.5^2)$ distribution, then generated y. For bad leverage points, different percentages of outliers on the variables $\mathbf{x}_1$ and $\mathbf{x}_2$ have generated from $N(100, 0.5^2)$ distribution.
The performance of the criteria was then determined by assessing summary of the percentage over a simulation of selected following models : (i) correct fit (true model); (ii) over fit (models containing all the variables in the true model plus other variables that are redundant $\mathbf{x}_1$,$\mathbf{x}_2$, and $\mathbf{x}_3$); (iii) under fit (models with only a strict of the variables in true model); (iv) wrong fit (the model that are neither of the above). The simulations were performed by the statistical software R based on s = 1000 Monte Carlo trials, the function *rlm* and *ltsreg* from the library (robust) was used for M- and LTS-estimation, respectively, and function *lmrob* from library (robustbase) used for MM-estimation.

4.2 Results and discussion

First, consider the data without outliers, Table(2) shows detailed simulation results for two cases of simulation setting with all different SIC criterion. The proposed SIC_{MM} selects nearly 70% to 80% proportion of correct fit models, while the classical SIC_{LS} performed better compared to robust SIC with a high percentage (94% t0 96%), However, as the percentage of outliers increased (see, Table (3)), SIC_{LS} selected a larger proportion of wrong fit models than other criteria, this holds for both cases 1 and 2. While the SIC_M continues to yield a higher percentage of correct fit and these results hold as the percentage of vertical outliers increased to 20%, then it tends to under fit. Thus, SIC_M method ignored some of the important variables in the model. A higher proportion of over fit and correct fit models are select by SIC_{LTS}. As expected, the percentage of the true model in all cases of SIC_{MM} was always large in the presence of vertical outliers, this result holds for both cases and with a high contamination level of vertical outliers. Table (4) Shows the situation where the data was contaminated with good leverage points, the results it can be concluded that good leverage points do not have much effect on all different SIC criteria. The presence of bad leverage points changes the picture dramatically. It can be observed from Tables (5) SIC_{LS} and SIC_M select a higher proportion of wrong fit than the SIC based on LTS-estimators, SIC_{LTS} tended to produce either correct fit or over fit model and the proposed criterion performed better when the bad leverage points are presents in the data.
In general, robust SIC criteria with M- and LTS-estimation are robust in the presence of outliers in the response variable. However, in the presence of bad leverage point, the value of these criteria will be affected and differs significantly from the true fit as the percentage of bad leverage point increases. But, SIC_{MM} criterion less affected in all cases in the presence of outliers in X and Y -directions.

ISSN: 2326-9865
https://doi.org/10.17762/msea.v71i1.5

Table 2: Percentage of selected models from different criteria for data with no contamination.

		n	SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}	n	SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}
Case 1	Correct fit	50	70.6	36.3	55.6	94.8	100	79.3	48.3	64.1	96.6
	Under fit		4.2	2.8	10.0	0.1		0.2	0.1	2.7	0.0
	Over fit		23.6	57.1	27.1	5.1		20.5	50.3	31.0	3.4
	Wrong fit		1.6	3.8	7.3	0.0		0.0	1.3	2.2	0.0
Case 2, r=0.03	Correct fit	50	75.2	39.8	63.8	94	100	78.7	46.1	66.5	97.1
	Under fit		0.0	0.1	0.4	0		0.0	0.0	0.0	0.0
	Over fit		24.6	59.8	35.1	6		21.3	53.9	33.5	2.9
	Wrong fit		0.2	0.3	0.7	0		0.0	0.0	0.0	0.0
Case 2, r=0.1	Correct fit	50	75.8	36.9	64.0	93.6	100	78.9	49.1	65.8	97.7
	Under fit		0.2	0.4	0.9	0.0		0.0	0.0	0.0	0.0
	Over fit		24.0	62.6	34.6	6.4		21.1	50.9	34.2	2.3
	Wrong fit		0.0	0.1	0.5	0.0		0.0	0.0	0.0	0.0
Case 2, r=0.5	Correct fit	50	71.8	38.2	62.5	94.5	100	81.0	47.7	68.8	96.9
	Under fit		0.3	0.0	2.0	0.0		0.1	0.2	0.2	0.0
	Over fit		27.4	60.8	33.5	5.5		18.9	51.9	30.5	3.1
	Wrong fit		0.5	1.0	2.0	0.0		0.0	0.2	0.5	0.0

5 Practical Example

Hawkins-Bradu-Kass Data: This data has been generated by [16] for illustrating some of the merits of robust technique, the full data set is given in Table (6). They pointed out that the first 10 observations are bad leverage points; i.e. the first 10 observations are outliers and the next 4 observations are good leverage points. Figure (3) showed the regression plot of y_i via different variables.
Table (7) shows the values of different criteria for Hawkins-Bradu-Kass data for different set of variables, where the small values of criteria are considered to show the best model. SIC_{MM} agree on the importance of all three variables, which appears in a low value of SIC_{MM}. And the values of the other criteria are small with under fit values.

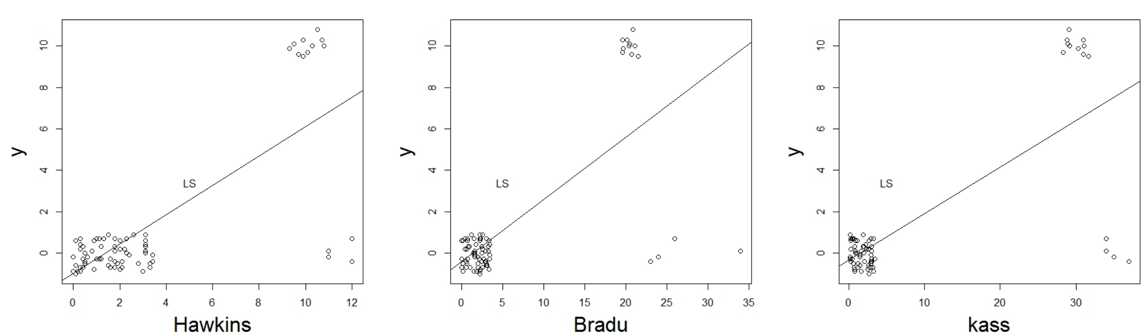

Figure 3: The regression plot of y via Hawkins, Bradu, and Kass

6 Conclusion

In this article the SIC criterion considered to be used with a high breakdown, efficient, and bounded influence scale estimators. The influence function of the criterion for the

ISSN: 2326-9865
https://doi.org/10.17762/msea.v71i1.5

linear regression model based on the MM-scale approach was discussed. The simulation study and the application on real data set suggest that, at least for the scenarios considered, the proposed SIC_{MM} criterion provide the best select the correct model for uncontaminated data sets, and stability in the presence of outliers.

Acknowledgments

This research has been funded by Jazan University, The author would like to thank Jazan University for their support to publish this study.

References

[1] Alshqaq, Shokrya Saleha A. "Robust Variable Selection in Linear Regression Models." PhD diss., Institut Sains Matematik, Fakulti Sains, Universiti Malaya, 2015.

[2] Saleh, Shokrya, and Ali Hassan Abuzaid. "Alternative Robust Variable Selection Procedures in Multiple Regression." Statistics, Optimization & Information Computing 7, no. 4 (2019): 816-825.

[3] Akaike, Hirotugu. "A Bayesian analysis of the minimum AIC procedure." In Selected Papers of Hirotugu Akaike, pp. 275-280. Springer, New York, NY, 1998.

[4] Schwarz, G. "Estimating the dimension of a model The Annals of Statistics 6 (2), 461–464." URL: http://dx. doi. org/10.1214/aos/1176344136 (1978).

[5] Ronchetti, Elvezio, and Robert G. Staudte. "A robust version of Mallows's Cp." Journal of the American Statistical Association 89, no. 426 (1994): 550-559.

[6] Saleh, Shokrya. "Robust AIC with high breakdown scale estimate." Journal of Applied Mathematics 2014 (2014).

[7] Saleh, Shokrya. "Model selection via robust version of r-squared." Journal of Mathematics and Statistics 10, no. 3 (2014): 414-420.

[8] Huber, Peter J. "Robust regression: asymptotics, conjectures and Monte Carlo." The Annals of Statistics 1, no. 5 (1973): 799-821.

[9] Machado, Jose AF. "Robust model selection and M-estimation." Econometric Theory 9, no. 3 (1993): 478-493.

[10] Saleh, Shokrya, Nor Aishah Hamzah, and Rossita M. Yunus. "A robust version of Schwarz information criterion based on LTS." In AIP Conference Proceedings, vol. 1605, no. 1, pp. 967-972. American Institute of Physics, 2014.

[11] Rousseeuw, Peter J. "Least median of squares regression." Journal of the American statistical association 79, no. 388 (1984): 871-880.

[12] Yohai, Victor J. "High breakdown-point and high efficiency robust estimates for regression." The Annals of Statistics (1987): 642-656.

[13] Rousseeuw, Peter, and Victor Yohai. "Robust regression by means of S-estimators." In Robust and nonlinear time series analysis, pp. 256-272. Springer, New York, NY, 1984.

[14] Hampel, Frank R., Elvezio M. Ronchetti, Peter J. Rousseeuw, and Werner A. Stahel. Robust statistics: the approach based on influence functions. Vol. 196. John Wiley & Sons, 2011.

[15] Hampel, Frank R. "Contribution to the theory of robust estimation." Ph. D. Thesis, University of California, Berkeley (1968).

[16] Hawkins, Douglas M., Dan Bradu, and Gordon V. Kass. "Location of several outliers in multiple-regression data using elemental sets." Technometrics 26, no. 3 (1984): 197-208.

Table 3: Percentage of selected models from different criteria for data with vertical outliers.

5% verticals	n		SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}	n	SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}
Case 1	50	Correct fit	72.9	39.3	53.7	0.8	100	82.6	49.6	64.3	0.5
		Under fit	5.3	3.2	11.9	67.6		0.2	0.4	3.1	69.3
		Over fit	19.9	54.4	27.1	0.1		17.1	49.3	31.0	0.0
		Wrong fit	1.9	3.1	7.3	31.5		0.1	0.7	1.6	30.2
Case 2, r=0.03	50	Correct fit	78.1	38.4	62.7	0.7	100	80.3	48.2	67.4	1.2
		Under fit	0.1	0.1	1.2	70.6		0.0	0.0	0.0	77.7
		Over fit	21.8	61.2	35.5	0.2		19.7	51.8	32.6	0.0
		Wrong fit	0.0	0.3	0.6	28.5		0.0	0.0	0.0	21.1
Case 2, r=0.1	50	Correct fit	78.6	40.8	65.2	1.2	100	83	47.7	66.8	1.2
		Under fit	0.0	0.1	0.5	71.4		0	0.0	0.0	77.5
		Over fit	21.4	59.1	34.0	0.3		17	52.3	33.2	0.0
		Wrong fit	0.0	0.0	0.3	27.1		0	0.0	0.0	21.3
Case 2, r=0.5	50	Correct fit	79.2	41.1	61.4	1.4	100	82.4	49.9	67.3	0.4
		Under fit	0.5	0.6	1.9	70.9		0.0	0.0	0.0	78.8
		Over fit	19.8	57.6	33.8	0.1		17.6	50.1	32.7	0.0
		Wrong fit	0.5	0.7	2.9	27.6		0.0	0.0	0.0	20.8
10% verticals	n		SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}	n	SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}
Case 1	50	Correct fit	77.8	40.7	57.3	0.8	100	82.2	51.6	65.6	0.3
		Under fit	3.2	3.4	9.1	66.6		0.0	0.8	1.9	67.9
		Over fit	18.0	52.1	27.9	0.0		17.7	47.1	30.7	0.0
		Wrong fit	1.0	3.8	5.7	32.6		0.1	0.5	1.8	31.8
Case 2, r=0.03	50	Correct fit	83.4	44.3	64.6	0.8	100	84.5	52	66.5	0.8
		Under fit	0.0	0.1	0.6	69.0		0.0	0	0.0	68.1
		Over fit	16.6	55.4	34.6	0.0		15.5	48	33.5	0.0
		Wrong fit	0.0	0.2	0.2	30.2		0.0	0	0.0	31.1
Case 2, r=0.1	50	Correct fit	80.7	41.4	64.8	0.9	100	85.3	55.1	67.2	0.3
		Under fit	0.2	0.3	0.9	68.3		0.0	0.0	0.0	73.4
		Over fit	19.0	58.0	33.6	0.2		14.7	44.9	32.8	0.1
		Wrong fit	0.1	0.3	0.7	30.6		0.0	0.0	0.0	26.2
Case 2, r=0.5	50	Correct fit	80.0	42.4	62.7	0.7	100	85	51.2	68.7	0.7
		Under fit	0.1	0.4	1.9	67.0		0	0.0	0.1	72.1
		Over fit	19.8	56.2	33.4	0.0		15	48.7	31.1	0.0
		Wrong fit	0.1	1.0	2.0	32.3		0	0.1	0.1	27.2
20% verticals	n		SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}	n	SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}
Case 1	50	Correct fit	85.5	48.0	54.9	0.7	100	88.5	58.0	66.8	0.6
		Under fit	4.0	4.9	13.4	64.8		0.0	0.1	3.5	69.0
		Over fit	9.9	43.6	22.8	0.0		11.4	41.1	28.1	0.0
		Wrong fit	0.6	3.5	8.9	34.5		0.1	0.8	1.6	30.4
Case 2, r=0.03	50	Correct fit	88.1	51.2	67.6	0.4	100	90.4	59.6	66.7	0.3
		Under fit	0.0	0.1	1.1	63.0		0.0	0.0	0.0	69.6
		Over fit	11.8	48.5	30.2	0.1		9.6	40.4	33.3	0.0
		Wrong fit	0.1	0.2	1.1	36.5		0.0	0.0	0.0	30.1
Case 2, r=0.1	50	Correct fit	86.3	51.0	67.8	0.4	100	89.5	63	68.1	0.5
		Under fit	0.0	0.3	1.8	68.5		0.0	0	0.0	69.2
		Over fit	13.6	48.6	29.4	0.1		10.5	37	31.8	0.0
		Wrong fit	0.1	0.1	1.0	31.0		0.0	0	0.1	30.3
Case 2, r=0.5	50	Correct fit	86.2	49.9	63.9	0.7	100	89.5	62.6	67.7	0.1
		Under fit	0.2	0.5	3.4	64.6		0.0	0.0	0.3	69.4
		Over fit	13.3	48.6	29.4	0.1		10.5	37.3	31.9	0.0
		Wrong fit	0.3	1.0	3.3	34.6		0.0	0.1	0.1	30.5
30% vertical	n		SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}	n	SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}
Case 1	50	Correct fit	89.2	57.0	4.7	0.2	100	94.1	69.7	8.3	0.3
		Under fit	3.0	4.4	60.0	66.0		0.1	0.4	60.3	68.9
		Over fit	7.0	37.0	1.7	0.0		5.8	29.8	0.7	0.0
		Wrong fit	0.8	1.6	33.6	33.8		0.0	0.1	30.7	30.8
Case 2, r=0.03	50	Correct fit	92.7	61.2	7.7	0.2	100	93.7	71.7	15.6	0.4
		Under fit	0.0	0.1	63.3	67.0		0.0	0.0	58.8	68.1
		Over fit	7.3	38.7	1.7	0.1		6.3	28.3	2.2	0.0
		Wrong fit	0.0	0.0	27.3	32.7		0.0	0.0	23.4	31.5
Case 2, r=0.1	50	Correct fit	93.1	59.8	8.3	0.6	100	94	69.8	18.2	0.0
		Under fit	0.0	0.1	60.2	66.4		0	0.0	57.8	67.2
		Over fit	6.9	40.0	1.3	0.0		6	30.2	1.0	0.0
		Wrong fit	0.0	0.1	30.2	33.0		0	0.0	23.0	32.8
Case 2, r=0.5	50	Correct fit	91.6	58.3	7.5	0.4	100	94.4	73.1	15.4	0.2
		Under fit	0.3	0.3	65.5	63.5		0.0	0.0	60.5	66.1
		Over fit	8.0	40.7	1.6	0.2		5.6	26.9	1.4	0.1
		Wrong fit	0.1	0.7	25.4	35.9		0.0	0.0	22.7	33.6
40% verticals	n		SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}	n	SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}
Case 1	50	Correct fit	86.6	69.4	5.3	0.4	100	97.1	81.7	1.1	0.1
		Under fit	10.9	5.2	61.9	68.0		0.1	0.2	69.1	67.5
		Over fit	1.8	24.4	3.6	0.0		2.8	18.1	0.9	0.0
		Wrong fit	0.7	1.0	29.2	31.6		0.0	0.0	28.9	32.4
Case 2, r=0.03	50	Correct fit	97.3	74.0	2.6	0.1	100	98.4	82.7	3.8	0.1
		Under fit	0.0	0.1	68.2	66.1		0.0	0.0	77.6	67.4
		Over fit	2.7	25.9	2.9	0.0		1.6	17.3	0.5	0.0
		Wrong fit	0.0	0.0	26.3	33.8		0.0	0.0	18.1	32.5
Case 2, r=0.1	50	Correct fit	98.8	76.2	3.8	0.8	100	97.9	81.7	2.4	0.2
		Under fit	0.0	0.1	67.8	65.2		0.0	0.0	77.0	67.3
		Over fit	1.2	23.7	2.4	0.0		2.1	18.3	0.3	0.0
		Wrong fit	0.0	0.0	26.0	34.0		0.0	0.0	20.3	32.5
Case 2, r=0.5	50	Correct fit	98.1	74.9	3.4	1.2	100	98.1	82.5	2.0	0.9
		Under fit	0.1	0.4	68.8	64.8		0.0	0.0	78.7	67.6
		Over fit	1.8	24.5	3.6	0.3		1.9	17.5	1.0	0.0
		Wrong fit	0.0	0.2	24.2	33.7		0.0	0.0	18.3	31.5

Table 4: Percentage of selected models from different criteria for data with good leverage points

5% good leverage	n		SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}	n	SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}
Case 1	50	Correct fit	93.4	67.5	76.0	100	100	97.3	78.1	84.9	100
		Under fit	2.1	2.6	3.4	0		0.2	0.2	0.2	0
		Over fit	3.9	26.9	19.0	0		2.5	21.2	14.7	0
		Wrong fit	0.6	3.0	1.6	0		0.0	0.5	0.2	0
Case 2, r=0.03	50	Correct fit	96.0	69.4	78.9	100	100	97.4	82.5	87.3	100
		Under fit	0.0	0.6	1.0	0		0.0	0.0	0.1	0
		Over fit	3.8	29.1	19.7	0		2.6	17.5	12.5	0
		Wrong fit	0.2	0.9	0.4	0		0.0	0.0	0.1	0
Case 2, r=0.1	50	Correct fit	94.6	70.9	76.2	100	100	97.5	79.1	88.3	100
		Under fit	0.1	0.3	1.6	0		0.0	0.0	0.0	0
		Over fit	4.8	27.6	21.0	0		2.5	20.9	11.7	0
		Wrong fit	0.5	1.2	1.2	0		0.0	0.0	0.0	0
Case 2, r=0.5	50	Correct fit	94.0	70.7	78.6	100	100	97.6	76.9	85.6	100
		Under fit	0.3	0.4	1.3	0		0.0	0.0	0.0	0
		Over fit	5.3	28.1	19.6	0		2.4	23.1	14.4	0
		Wrong fit	0.4	0.8	0.5	0		0.0	0.0	0.0	0
10% good leverage	n		SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}	n	SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}
Case 1	50	Correct fit	97.2	81.2	86.0	100	100	100	94.0	96.4	100
		Under fit	0.8	2.3	2.5	0		0	0.4	0.6	0
		Over fit	1.5	14.1	9.4	0		0	5.6	2.9	0
		Wrong fit	0.5	2.4	2.1	0		0	0.0	0.1	0
Case 2, r=0.03	50	Correct fit	98.7	85.6	90.2	100	100	99.7	94.3	95.9	100
		Under fit	0.0	0.0	0.0	0		0.0	0.0	0.0	0
		Over fit	1.3	14.4	9.8	0		0.3	5.7	4.1	0
		Wrong fit	0.0	0.0	0.0	0		0.0	0.0	0.0	0
Case 2, r=0.1	50	Correct fit	98.5	83.8	90.1	100	100	100	94.8	96.9	100
		Under fit	0.0	0.0	0.0	0		0	0.0	0.0	0
		Over fit	1.5	16.1	9.8	0		0	5.2	3.1	0
		Wrong fit	0.0	0.1	0.1	0		0	0.0	0.0	0
Case 2, r=0.5	50	Correct fit	98.0	82.6	88.9	100	100	99.9	94	95.1	100
		Under fit	0.3	0.5	1.1	0		0.0	0	0.1	0
		Over fit	1.5	15.7	9.1	0		0.1	6	4.8	0
		Wrong fit	0.2	1.2	0.9	0		0.0	0	0.0	0
20% good leverage	n		SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}	n	SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}
Case 1	50	Correct fit	98.6	94.7	92.8	100	100	99.9	94	95.1	100
		Under fit	0.8	1.2	3.7	0		0.0	0	0.1	0
		Over fit	0.1	1.1	0.6	0		0.1	6	4.8	0
		Wrong fit	0.5	3.0	2.9	0		0.0	0	0.0	0
Case 2, r=0.03	50	Correct fit	100	98.7	99.3	100	100	100	99.9	100	100
		Under fit	0	0.0	0.2	0		0	0.0	0	0
		Over fit	0	1.2	0.5	0		0	0.1	0	0
		Wrong fit	0	0.1	0.0	0		0	0.0	0	0
Case 2, r=0.1	50	Correct fit	100	98.5	99.2	100	100	100	99.9	100	100
		Under fit	0	0.0	0.1	0		0	0.0	0	0
		Over fit	0	1.4	0.4	0		0	0.1	0	0
		Wrong fit	0	0.1	0.3	0		0	0.0	0	0
Case 2, r=0.5	50	Correct fit	99.8	97.3	97.2	100	100	100	99.7	99.9	100
		Under fit	0.1	0.5	1.2	0		0	0.0	0.1	0
		Over fit	0.0	1.1	0.2	0		0	0.1	0.0	0
		Wrong fit	0.1	1.1	1.4	0		0	0.2	0.0	0
30% good leverage	n		SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}	n	SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}
Case 1	50	Correct fit	99.0	96.9	94.3	100	100	100	99.8	99.2	100
		Under fit	0.5	0.6	2.4	0			0.1	0.4	0
		Over fit	0.0	0.1	0.0	0		0.0	0.0	0	
		Wrong fit	0.5	2.4	3.3	0		0	0.1	0.4	0
Case 2, r=0.03	50	Correct fit	100	99.9	99.5	100	100	100	100	100	100
		Under fit	0	0.0	0.0	0		0	0	0	0
		Over fit	0	0.0	0.0	0		0	0	0	0
		Wrong fit	0	0.1	0.5	0		0	0	0	0
Case 2, r=0.1	50	Correct fit	100	99.4	99.2	100	100	100	100	100	100
		Under fit	0	0.1	0.1	0		0	0	0	0
		Over fit	0	0.1	0.0	0		0	0	0	0
		Wrong fit	0	0.4	0.7	0		0	0	0	0
Case 2, r=0.5	50	Correct fit	99.8	97.9	95.9	100	100	100	99.7	99.4	100
		Under fit	0.0	0.1	2.3	0		0	0.1	0.3	0
		Over fit	0.0	0.0	0.0	0		0	0.0	0.0	0
		Wrong fit	0.2	2.0	1.8	0		0	0.2	0.3	0
40% good leverage	n		SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}	n	SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}
Case 1	50	Correct fit	98.8	96.0	93.5	100	100	100	99.7	99.0	100
		Under fit	0.6	1.2	4.1	0		0	0.0	0.7	0
		Over fit	0.0	0.0	0.0	0		0	0.0	0.0	0
		Wrong fit	0.6	2.8	2.4	0		0	0.3	0.3	0
Case 2, r=0.03	50	Correct fit	99.8	99.4	98	100	100	100	100	100	100
		Under fit	0.1	0.1	1	0		0	0	0	0
		Over fit	0.0	0.0	0	0		0	0	0	0
		Wrong fit	0.1	0.5	1	0		0	0	0	0
Case 2, r=0.1	50	Correct fit	99.8	99.3	98.9	100	100	100	100	99.8	100
		Under fit	0.1	0.4	0.4	0		0	0	0.1	0
		Over fit	0.0	0.0	0.0	0		0	0	0.0	0
		Wrong fit	0.1	0.3	0.7	0		0	0	0.1	0
Case 2, r=0.5	50	Correct fit	98.9	97.3	94.9	100	100	100	100	99.9	100
		Under fit	0.5	0.7	2.8	0		0	0	0.1	0
		Over fit	0.0	0.0	0.0	0		0	0	0.0	0
		Wrong fit	0.6	2.0	2.3	0		0	0	0.0	0

ISSN: 2326-9865
https://doi.org/10.17762/msea.v71i1.5

Table 5: Percentage of selected models from different criteria for data with bad leverage points

5% bad leverage	n		SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}	n	SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}
Case 1	Correct fit	50	71.8	36.3	19.6	1.8	100	80.9	47.1	15.2	1.1
	Under fit		3.8	3.0	30.7	49.0		0.0	0.8	35.2	49.2
	Over fit		21.2	55.9	9.3	0.0		19.0	51.6	8.4	0.0
	Wrong fit		3.2	4.8	40.4	49.2		0.1	0.5	41.2	49.7
Case 2, r=0.03	Correct fit	50	75.5	39.6	17.1	4.0	100	83.2	51.4	19.4	3.7
	Under fit		0.1	0.3	32.4	50.9		0.0	0.0	31.7	48.3
	Over fit		24.4	60.1	8.7	0.1		16.8	48.6	7.8	0.0
	Wrong fit		0.0	0.0	41.8	45.0		0.0	0.0	41.1	48.0
Case 2, r=0.1	Correct fit	50	77.5	40.1	18.6	4.0	100	79.9	50.6	14.4	2.7
	Under fit		0.0	0.2	31.5	47.0		0.0	0.0	35.4	45.0
	Over fit		22.5	59.7	9.9	0.1		20.1	49.4	9.7	0.1
	Wrong fit		0.0	0.0	40.0	48.9		0.0	0.0	40.5	52.2
Case 2, r=0.5	Correct fit	50	80.1	37.9	9.5	0.4	100	80.5	50.4	6.1	0.1
	Under fit		0.4	0.6	16.6	2.9		0.0	0.0	12.2	0.4
	Over fit		19.5	60.9	19.4	4.1		19.5	49.6	22.6	4.9
	Wrong fit		0.0	0.6	54.5	92.6		0.0	0.0	59.1	94.6
10% bad leverage	n		SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}	n	SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}
Case 1	Correct fit	50	72.8	41.1	17.1	2.0	100	85.5	53.2	15.8	0.5
	Under fit		4.5	3.7	31.6	49.2		0.3	0.9	34.0	49.1
	Over fit		18.5	50.4	7.6	0.0		14.1	45.2	7.1	0.1
	Wrong fit		4.2	4.8	43.7	48.8		0.1	0.7	43.1	50.3
Case 2, r=0.03	Correct fit	50	80.7	44.6	16.7	4.7	100	84	54.3	15.8	2.7
	Under fit		0.1	0.2	32.4	47.7		0	0.0	34.0	49.8
	Over fit		19.0	54.9	8.7	0.2		16	45.7	7.2	0.0
	Wrong fit		0.2	0.3	42.2	47.4		0	0.0	43.0	47.5
Case 2, r=0.1	Correct fit	50	78.5	42.3	18.2	5.1	100	84.5	53.7	14.6	3.0
	Under fit		0.1	0.2	31.4	44.3		0.0	0.0	31.8	45.3
	Over fit		21.3	57.1	7.1	0.1		15.5	46.3	8.8	0.0
	Wrong fit		0.1	0.4	43.3	50.5		0.0	0.0	44.8	51.7
Case 2, r=0.5	Correct fit	50	80.7	42.6	9.1	0.7	100	81.8	52.6	5.2	0.0
	Under fit		0.6	0.7	16.6	2.8		0.0	0.0	12.2	0.0
	Over fit		18.1	56.0	20.2	3.8		18.2	47.4	23.9	3.9
	Wrong fit		0.6	0.7	54.1	92.7		0.0	0.0	58.7	96.1
20% bad leverage	n		SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}	n	SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}
Case 1	Correct fit	50	54.4	37.8	17.1	2.0	100	70.9	52.8	16.4	0.5
	Under fit		15.5	4.5	31.7	48.5		8.0	3.4	32.5	52.4
	Over fit		9.3	44.0	6.7	0.1		9.5	35.7	5.0	0.0
	Wrong fit		20.8	13.7	44.5	49.4		11.6	8.1	46.1	47.1
Case 2, r=0.03	Correct fit	50	87.0	47.6	17.0	4.2	100	87.4	60.5	13.7	2.4
	Under fit		0.5	0.5	32.9	47.6		0.0	0.0	34.3	47.1
	Over fit		11.8	50.9	8.5	0.1		12.6	39.5	7.6	0.0
	Wrong fit		0.7	1.0	41.6	48.1		0.0	0.0	44.4	50.5
Case 2, r=0.1	Correct fit	50	87.5	52.5	17.3	4.7	100	90.4	60.9	16.0	2.4
	Under fit		0.6	0.4	32.8	45.9		0.0	0.0	31.1	43.6
	Over fit		11.6	46.7	7.6	0.2		9.6	39.1	7.2	0.2
	Wrong fit		0.3	0.4	42.3	49.2		0.0	0.0	45.7	53.8
Case 2, r=0.5	Correct fit	50	87.9	50.0	7.9	0.4	100	89.2	59.2	5.9	0.2
	Under fit		0.2	0.4	17.9	2.9		0.0	0.0	11.5	0.0
	Over fit		11.3	48.2	20.2	3.0		10.8	40.8	20.8	3.3
	Wrong fit		0.6	1.4	54.0	93.7		0.0	0.0	61.8	96.5
30% bad leverage	n		SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}	n	SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}
Case 1	Correct fit	50	43.7	29.6	17.3	2.6	100	44.3	36.7	16.1	0.7
	Under fit		20.0	10.9	30.7	46.9		15.9	10.3	32.1	49.9
	Over fit		31.0	34.2	7.5	0.1		36.5	26.8	7.9	0.0
	Wrong fit		5.3	25.3	44.5	50.4		3.3	26.2	43.9	49.4
Case 2, r=0.03	Correct fit	50	75.7	53.3	18.4	4.5	100	91.4	69.6	17.5	2.3
	Under fit		6.6	1.5	29.2	46.6		0.9	0.2	31.2	50.2
	Over fit		7.9	40.7	7.9	0.1		5.7	29.5	7.9	0.0
	Wrong fit		9.8	4.5	44.5	48.8		2.0	0.7	43.4	47.5
Case 2, r=0.1	Correct fit	50	79.2	54.7	15.9	3.0	100	92.8	70.0	17.6	1.9
	Under fit		6.2	1.8	29.3	44.4		1.1	0.3	32.4	41.9
	Over fit		7.3	39.3	8.6	0.2		4.6	29.0	7.6	0.0
	Wrong fit		7.3	4.2	46.2	52.4		1.5	0.7	42.4	56.2
Case 2, r=0.5	Correct fit	50	83.6	56.1	8.4	0.4	100	94.0	69.1	5.4	0.0
	Under fit		1.1	0.6	16.3	2.6		0.1	0.2	12.2	0.2
	Over fit		8.1	39.9	17.4	3.1		5.1	30.6	20.5	3.3
	Wrong fit		7.2	3.4	57.9	93.9		0.8	0.1	61.9	96.5
40% bad leverage	n		SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}	n	SIC_{MM}	SIC_{LTS}	SIC_M	SIC_{LS}
Case 1	Correct fit	50	46.3	19.5	15.1	3.3	100	47.1	22.0	17.3	1.4
	Under fit		12.5	14.3	29.6	48.6		11.4	15.2	30.9	51.2
	Over fit		36.7	33.1	11.7	0.0		39.1	24.4	7.7	0.0
	Wrong fit		4.5	33.1	43.6	48.1		2.4	38.4	44.1	47.4
Case 2, r=0.03	Correct fit	50	44.7	42.1	15.5	4.3	100	62.9	17.4	16.0	1.9
	Under fit		33.6	7.8	29.4	48.3		5.0	34.7	32.8	46.1
	Over fit		4.9	31.2	8.9	0.1		21.5	3.8	6.7	0.0
	Wrong fit		16.8	18.9	46.2	47.3		10.6	44.1	44.5	52.0
Case 2, r=0.1	Correct fit	50	45.4	42.2	16.8	4.4	100	65.9	22.4	15.5	3.0
	Under fit		18.8	8.3	25.5	42.6		4.5	30.6	30.5	41.6
	Over fit		31.1	32.1	9.6	0.1		18.7	2.7	7.9	0.0
	Wrong fit		4.7	17.4	48.1	52.9		10.9	44.3	46.1	55.4
Case 2, r=0.5	Correct fit	50	43.0	15.1	7.2	0.1	100	65.8	19.7	5.5	0.0
	Under fit		1.9	9.4	16.0	2.1		0.9	3.8	11.6	0.1
	Over fit		37.0	11.7	18.7	3.2		22.5	11.4	20.4	2.6
	Wrong fit		18.1	63.8	58.1	94.6		10.8	65.1	62.5	97.3

Mathematical Statistician and Engineering Applications
ISSN: 2326-9865
https://doi.org/10.17762/msea.v71i1.5

Table 6: Hawkins-Bradu-Kass Data

Obs. No.	Hawkins	Bradu	kass	y
1	10.1	19.6	28.3	9.7
2	9.5	20.5	28.9	10.1
3	10.7	20.2	31.0	10.3
4	9.9	21.5	31.7	9.5
5	10.3	21.1	31.1	10.0
6	10.8	20.4	29.2	10.0
7	10.5	20.9	29.1	10.8
8	9.9	19.6	28.8	10.3
9	9.7	20.7	31.0	9.6
10	9.3	19.7	30.3	9.9
11	11.0	24.0	35.0	-0.2
12	12.0	23.0	37.0	-0.4
13	12.0	26.0	34.0	0.7
14	11.0	34.0	34.0	0.1
15	3.4	2.9	2.1	-0.4
16	3.1	2.2	0.3	0.6
17	0.0	1.6	0.2	-0.2
18	2.3	1.6	2.0	0.0
19	0.8	2.9	1.6	0.1
20	3.1	3.4	2.2	0.4
21	2.6	2.2	1.9	0.9
22	0.4	3.2	1.9	0.3
23	2.0	2.3	0.8	-0.8
24	1.3	2.3	0.5	0.7
25	1.0	0.0	0.4	-0.3
26	0.9	3.3	2.5	-0.8
27	3.3	2.5	2.9	-0.7
28	1.8	0.8	2.0	0.3
29	1.2	0.9	0.8	0.3
30	1.2	0.7	3.4	-0.3
31	3.1	1.4	1.0	0.0
32	0.5	2.4	0.3	-0.4
33	1.5	3.1	1.5	-0.6
34	0.4	0.0	0.7	-0.7
35	3.1	2.4	3.0	0.3
36	0.1	2.2	2.7	-1.0
37	0.1	3.0	2.6	-0.6
38	1.5	1.2	0.2	0.9
39	2.1	0.0	1.2	-0.7
40	0.5	2.0	1.2	-0.5
41	3.4	1.6	2.9	-0.1
42	0.3	1.0	2.7	-0.7
43	0.1	3.3	0.9	0.6
44	1.8	0.5	3.2	-0.7
45	1.9	0.1	0.6	-0.5
46	1.8	0.5	3.0	-0.4
47	3.0	0.1	0.8	-0.9
48	3.1	1.6	3.0	0.1
49	3.1	2.5	1.9	0.9
50	2.1	2.8	2.9	-0.4
51	2.3	1.5	0.4	0.7
52	3.3	0.6	1.2	-0.5
53	0.3	0.4	3.3	0.7
54	1.1	3.0	0.3	0.7
55	0.5	2.4	0.9	0.0
56	1.8	3.2	0.9	0.1
57	1.8	0.7	0.7	0.7
58	2.4	3.4	1.5	-0.1
59	1.6	2.1	3.0	-0.3
60	0.3	1.5	3.3	-0.9
61	0.4	3.4	3.0	-0.3
62	0.9	0.1	0.3	0.6
63	1.1	2.7	0.2	-0.3
64	2.8	3.0	2.9	-0.5
65	2.0	0.7	2.7	0.6
66	0.2	1.8	0.8	-0.9
67	1.6	2.0	1.2	-0.7
68	0.1	0.0	1.1	0.6
69	2.0	0.6	0.3	0.2
70	1.0	2.2	2.9	0.7
71	2.2	2.5	2.3	0.2
72	0.6	2.0	1.5	-0.2
73	0.3	1.7	2.2	0.4
74	0.0	2.2	1.6	-0.9
75	0.3	0.4	2.6	0.2

ISSN: 2326-9865
https://doi.org/10.17762/msea.v71i1.5

Table 7: The values of different criteria for Hawkins-Bradu-Kass data for different set of variables

Set of variables	SIC_{MM},	SIC_{LTS}	SIC_M	SIC_{LS}
(y,Hawkins)	-0.3877	-0.6233	0.8786	1.7553
(y,Bradu)	-0.4056	-0.6277	0.2157	1.8609
(y,Kass)	-0.3982	-0.5835	-0.3089	1.7077
(y,Hawkins,Bradu)	-0.3525	-0.5904	-0.2360	1.7898
(y,Hawkins,Kass)	-0.3999	-0.6514	**-0.1147**	1.7358
(y,Bradu,Kass)	-0.3766	**-0.5719**	-0.2206	**1.6839**
(y,Hawkins,Bradu,Kass)	**-0.4062**	-0.6816	-0.1385	1.7077