

Research Article

Robust AIC with High Breakdown Scale Estimate

Shokrya Saleh

Institute of Mathematical Sciences, University of Malaya, 50603 Kuala Lumpur, Malaysia

Correspondence should be addressed to Shokrya Saleh; f84fh@hotmail.com

Received 25 May 2014; Revised 20 August 2014; Accepted 20 August 2014; Published 8 September 2014

Academic Editor: Bernard J. Geurts

Copyright © 2014 Shokrya Saleh. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Akaike Information Criterion (AIC) based on least squares (LS) regression minimizes the sum of the squared residuals; LS is sensitive to outlier observations. Alternative criterion, which is less sensitive to outlying observation, has been proposed; examples are robust AIC (RAIC), robust Mallows Cp (RCp), and robust Bayesian information criterion (RBIC). In this paper, we propose a robust AIC by replacing the scale estimate with a high breakdown point estimate of scale. The robustness of the proposed methods is studied through its influence function. We show that, the proposed robust AIC is effective in selecting accurate models in the presence of outliers and high leverage points, through simulated and real data examples.

1. Introduction

Akaike Information Criterion (AIC) [1] is a powerful technique for model selection, and it has been widely used for selecting models in many fields of study.

Consider a multiple linear regression model:

$$y_i = \mu + \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i, \quad (1)$$

where $\mathbf{x}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{ip})^T$ is a vector containing p explanatory variables, y_i is the response variable, $\boldsymbol{\beta}$ is a vector of p parameters, μ is the intercept parameter, and ϵ_i is the error component, which is independent and identically distributed (iid), with mean 0 and variance σ^2 . The classical AIC is defined as

$$\text{AIC} = \ln \left(\frac{\text{SSE}}{n} \right) + 2p, \quad (2)$$

where $\text{SSE} = \sum r_i^2$, with $r_i = y_i - \mu_{\text{LS}} - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{\text{LS}}$.

Since the LS estimator is vulnerable in presence of outliers, it is not surprising that AIC inherits this problem. Several robust AIC alternatives to the AIC have been proposed in the literature (see [2–4]). For an example Ronchetti [3] has proposed and investigated the properties of a robust version of AIC, with respect to M -estimation. A similar idea was used by Martin [2] for autoregressive models. Furthermore, very recently Tharmaratnam and Claeskens [4] have proposed

a robust AIC with respect to S -estimation and MM -estimation, to generalize the information criteria, using the full likelihood models. In spite of all these complicated approaches of deriving robust AIC, we introduce a straightforward approach to derive robust AIC, which focuses on modifying the estimate of the scale.

To set the idea, the influence of outlier on AIC is illustrated through the presence of outliers in the Y -direction (called a vertical outlier) or in the X -direction (called a leverage point). For this, a point with coordinates $(0, y_{10})$ is added, where the value of y ranges between $(-1.5$ and $3)$. A similar approach is done for leverage points, by replacing the value \mathbf{x} with $(x_{10}, 0)$ (Figure 1). Table 1 and Figure 2 show that the value of AIC increases as the size of contamination in (x_{10}, y_{10}) increases, as expected, and if x or y is extremely large, the AIC is unbounded; that is, it will tend to infinity.

The remainder of the paper has been organized as follows. Section 1.1 reviews some robust regression estimate methods. In Section 1.2 we review a robust version of AIC; we discuss the robustness problem from the viewpoint of model selection and point out the sensitivity of robust AIC, based on M -estimator to leverage points. We derive the influence function of AIC and study its properties in Section 2. The performance of robust AIC is evaluated and compared to the commonly used classical AIC in Section 3. Finally, concluding remark is presented in Section 4.

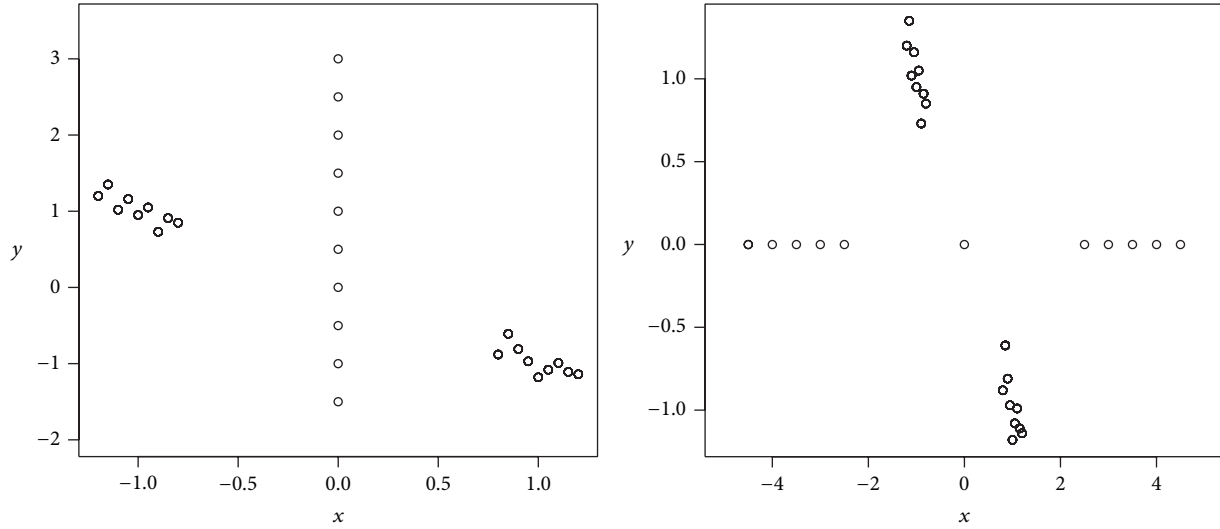


FIGURE 1: Data and positions for y_{10} and x_{10} points.

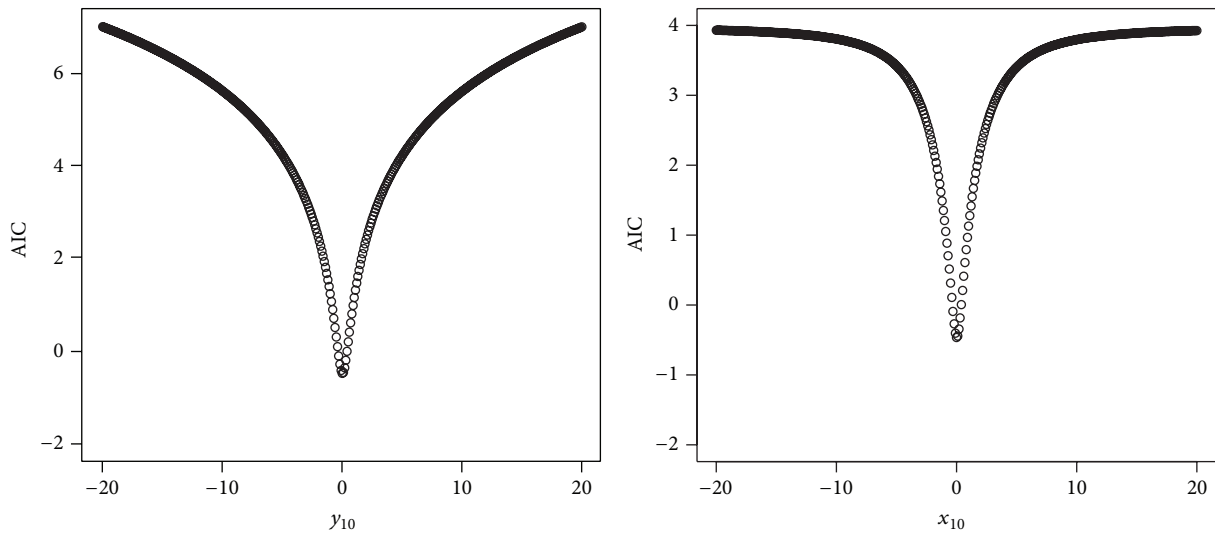


FIGURE 2: Effect of adding one observation $(0, y_{10})$ or $(x_{10}, 0)$ on the values of AIC.

TABLE 1: AIC for different values of y_{10} and x_{10} .

y_{10}	AIC	x_{10}	AIC
3	3.2	4.5	3.3
2.5	2.9	4	3.2
2	2.4	3.5	3.0
1.5	1.9	3	2.8
1	1.2	2.5	2.6
0.5	0.2	0	-0.5
0	-0.5	-2.5	2.6
-0.5	0.3	-3	2.9
-1	1.2	-3.5	3.1
-1.5	1.9	-4	3.2

1.1. *A Robust Regression Estimate.* The M -estimator [5] $\hat{\beta}_M$ is the value of β that minimizes the following function:

$$\sum_{i=1}^{i=n} \rho(y_i - \mu - \mathbf{x}_i^T \beta), \tag{3}$$

where ρ is symmetric and nondecreasing on $[0, \infty)$. Furthermore, $\rho(0) = 0$, and ρ is almost continuously differentiable, anywhere. Furthermore, ρ is a function, which is less sensitive to outliers than squares, yielding the estimating equation:

$$\sum_{i=1}^{i=n} \psi(y_i - \mu - \mathbf{x}_i^T \hat{\beta}) \mathbf{x}_i = 0, \tag{4}$$

where $\psi = \rho'$. If we choose ρ function in (3) as Tukey Biweight function

$$\rho(r; d) = \begin{cases} 3\left(\frac{r}{d}\right)^2 - 3\left(\frac{r}{d}\right)^4 + \left(\frac{r}{d}\right)^6, & \text{if } |r| \leq d, \\ 1, & \text{elsewhere,} \end{cases} \quad (5)$$

where $d = 1.5476$ yields $b = E_{\Phi}[\rho(Z; d)] = 2(1 - F_0(d))$, with Φ the standard normal cumulative distribution function and $Z \sim N(0, 1)$, the resulting estimator is then Biweight S-estimator. M -estimators are efficient and highly robust to unusual value of Y , but one rogue leverage point can break them down completely. For this reason, generalized M -estimators were introduced, which solve

$$\sum_{i=1}^{i=n} w(\mathbf{x}_i) \psi(y_i - \mu - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) \mathbf{x}_i = 0, \quad (6)$$

where w is a weight function [6].

In recent years, a good deal of attention in the literature has been focused on high-breakdown methods; that is, one method must be resistant to even multiple severe outliers. Many methods are based on minimizing a more robust scale estimate than sum of squared residuals. For example, Rousseeuw [7] proposed LMS, a high-breakdown method that minimizes the median of the squared residuals, rather than their sum. In addition, Rousseeuw [8] proposed least trimmed squares (LTS), which minimizes the sum of the H smallest squared residuals, defined as

$$\hat{\boldsymbol{\beta}}_{LTS,H,N} = \arg \min \sum_{i=1}^{i=H} r_i^2(\boldsymbol{\beta}), \quad (7)$$

based on the ordered absolute residuals $|r_{(1)}| \leq |r_{(2)}| \leq \dots \leq |r_{(n)}|$. LTS converges at the rate of $n^{1/2}$ with the same asymptotic efficiency under normality, as Huber's skip estimator. The convergence rate of LMS is $n^{1/3}$, and its objective function is less smooth than LTS.

1.2. *The Robust Version of AIC.* Consider scale estimate of errors defined by

$$S = \frac{SSE}{(n-p)}, \quad (8)$$

with $SSE = \sum_{i=1}^{i=n} r_i^2$ and $r_i = y_i - \hat{\mu} - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}$. By replacing the value of SSE in (2) in terms of S , AIC in (2) can be expressed as follows:

$$AIC_S = \ln\left(\frac{(n-p)S^2}{n}\right) + 2p. \quad (9)$$

Notice that the largest values of AIC indicate that the model (and hence the explanatory variables) is less successful in explaining the variations in the response, while a small value of AIC indicates an excellent fit (i.e., model) to the response data.

Ronchetti [3] proposed a robust counterpart of the AIC statistic. The extension of AIC to RAIC is inspired

by the extension of maximum likelihood estimation to M -estimation. The author derived RAIC for an error distribution with density function $f(\epsilon) = \exp(-\rho(\epsilon))$. For a given constant α and a given function ρ , the author chose the model that minimizes

$$RAIC(p, \alpha, \rho) = \sum_{i=1}^{i=n} \rho\left(\frac{(y_i - \mu - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_M)}{\hat{\sigma}}\right) + \alpha p, \quad (10)$$

where $\hat{\sigma}$ is some robust estimate of σ and $\hat{\boldsymbol{\beta}}_M$ is the M -estimator defined as in (3). Huber [9] suggested $\alpha = 2E[\psi^2(((y_i - \mu - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})/\hat{\sigma}))]/(E[\psi'(((y_i - \mu - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})/\hat{\sigma}))])$ and $k = 1.345$.

We introduce an alternative robust version of AIC, by replacing S in (8) to the robust estimator of scale which attains a 50% breakdown point. When $H = n/2$, (8) finds the estimates corresponding to the half samples, having the smallest sum of squares of residuals. As such, as expected, the breakdown point is 50%, and the estimated scale from LTS is

$$\hat{\sigma}_{(LTS,H,N)}^2 = \frac{1}{H} \sum_{i=1}^{i=H} r_{(i)}^2(\hat{\boldsymbol{\beta}}_{(LTS,H,N)}). \quad (11)$$

For other robust estimations, the M -estimator and the Biweight S-estimator are compared to least trimmed of squares. Based on the results shown in Table 2, it is evident that the M -estimator is much more robust than LS but suffers from leverage points. The Biweight S-estimator and LTS show robust behavior: the AIC_{BS} is stable, even though the size of the outliers increases. In the next section, we generalize these findings, by computing the associated influence functions.

2. Influence Function

Consider the linear model in (1), for $i = 1, \dots, n$. Assume the distribution of errors satisfying $F_{\sigma}(\mathbf{x}) = F_0(\mathbf{x}/\sigma)$, where σ is the residual scale parameter and F_0 is symmetric with valid probability 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 , as follows:

$$T(H) = \begin{bmatrix} a(H) \\ \mathbf{b}(H) \end{bmatrix} = \begin{bmatrix} \mu \\ \boldsymbol{\beta} \end{bmatrix}. \quad (12)$$

For a Fisher-consistent scale estimator, $F_{\sigma}(\mathbf{x}) = F(\mathbf{x}/\sigma)$, for all $\sigma > 0$. In general, the influence function of T at the distribution F is defined as

$$IF((\mathbf{x}, y), T, H) = \lim_{\epsilon \rightarrow 0} \frac{T((1-\epsilon)H + \epsilon\Delta_{(\mathbf{x},y)}) - T(H)}{\epsilon} = \frac{\partial}{\partial \epsilon} (T(\Delta_{(\mathbf{x},y)})), \quad (13)$$

where $T(H)$ is the functional defined as the solution of the objective model and $\Delta_{(\mathbf{x},y)}$ is the distribution which contains

TABLE 2: Robust AIC for different values of y_{10} (vertical) and x_{10} (leverage).

y_{10}	AIC _M	AIC _{BS}	AIC _{LTS}	x_{10}	AIC _M	AIC _{BS}	AIC _{LTS}
3	3.7	3.9	3.4	4.5	3.9	3.9	3.4
2.5	3.7	3.9	3.4	4	3.7	3.9	3.4
2	3.7	3.9	3.4	3.5	3.6	3.9	3.4
1.5	3.7	3.9	3.4	3	3.6	3.9	3.4
1	3.7	3.9	3.4	2.5	3.7	3.9	3.4
0.5	3.7	3.9	3.4	0	3.5	3.6	3.2
0	3.6	3.6	3.2	-2.5	3.7	3.9	3.4
-0.5	3.7	3.9	3.4	-3	3.7	3.9	3.4
-1	3.7	3.9	3.4	-3.5	3.9	3.9	3.4
-1.5	3.7	3.9	3.4	-4	4	3.9	3.4

outliers. The following theorem gives the influence function of AIC with any scale S .

Theorem 1. *Let H be some distribution other than F . Take $(\mathbf{X}, Y) \sim H$ and denote by ϵ the error term of the model. Assume that S has the property that is differentiable with partial derivatives equal to zero at the origin $(0,0)$. Then,*

$$\begin{aligned} IF((\mathbf{x}, y), AIC_S, H) \\ = \frac{2n}{(n-p)} IF\left(\frac{r_{ip}}{\sigma_p}, S, F_0\right), \end{aligned} \quad (14)$$

where $r_{ip} = y_i - \hat{\mu} - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_p$ and $\hat{\sigma}_p^2 = SSE/(n-p)$ are computed from sum model. (The proof is in the Appendix.)

It is clear that the influence function in (13) is bounded if the $IF(r_{ip}/\sigma_p, S, F_0)$ is also bounded. It is evident that AIC is nonrobust, since the LS estimate has unbounded influence function. The influence function of M -estimation with respect to Y is bounded by choice of ψ , but it is unbounded with respect to the X direction. That is,

$$\begin{aligned} IF((\mathbf{x}, y), T, F) \\ = M^{-1}(\psi, F) \mathbf{x} \psi(y - \mathbf{x}^T T(F)), \end{aligned} \quad (15)$$

where $\psi = \rho'$ and M is a certain $p \times p$ matrix given by

$$\begin{aligned} M(\psi, F) \\ = \int \psi'(y - \mathbf{x}^T T(F)) \mathbf{x} \mathbf{x}^T dF(\mathbf{x}, y). \end{aligned} \quad (16)$$

The influence function of the AIC using LTS estimators, following Theorem 1, takes the form

$$\begin{aligned} IF((\mathbf{x}, y), AIC_{LTS}, H) \\ = \frac{2n}{(n-p)} \cdot IF\left(\frac{r_{ip}}{\sigma_p}, \hat{\sigma}_{LTS}^2, F_0\right). \end{aligned} \quad (17)$$

We have noted that the influence function of AIC_{LTS} is bounded in both Y and X directions, as $IF(r_{ip}/\sigma_p, \hat{\sigma}_{LTS}^2, F_0)$

is bounded. Moreover, we conclude that AIC with high breakdown point estimator provides reliable fits to the data in presence of leverage points.

3. Numerical Examples

In this section, AIC_{LTS} , AIC_{LMS} , and AIC_{BS} are compared with AIC and RAIC. In this study, 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 $\epsilon_i \sim N(0, 9)$ were generated. The true model is given by $y_i = \mathbf{x}_{i1} + \mathbf{x}_{i2} + \epsilon_i$, for $i = 1, \dots, 50$, using two variables \mathbf{x}_{i1} and \mathbf{x}_{i2} . We have computed the following AIC, based on the respective estimation criterion: (i) LS estimation; (ii) M -estimator; (iii) BS estimator; (iv) LTS estimator; and (v) LMS estimator. In order to illustrate the robustness to outliers, we consider the following cases:

- vertical outliers (outliers in the y only),
- good leverage points (outliers in the y and \mathbf{x}),
- bad leverage points (outliers in the \mathbf{x} only).

For all situations, we randomly generate (0%, 5%, 10%, 20%, 30%, and 40%) of outliers from $N(50, 0.1^2)$ and $N(100, 0.5^2)$ distributions, respectively. For each of these settings we simulate 1000 samples.

3.1. Simulation Result. The resulting fit to the data is classified as one of the following:

- correct fit (true model);
- overfit (models containing all the variables in the true model plus other variables that are redundant);
- under fit (models with only a strict of the variables in true model);
- wrong fit (model that are neither of the above).

Tables 3, 4, and 5 show detailed simulation result for different versions of AIC methods. For uncontaminated datasets, the classical AIC performs best, compared to robust AIC. By introducing vertical outliers, the classical AIC selects a large proportion of wrong fit models and, as expected, the robust AIC will usually (i.e., with higher proportion) select the correct model.

For bad leverage points, we observe that AIC tends to produce overfit and with high level of contamination it takes a higher proportion of wrong fit. However, AIC_M tend to produce either an under fit or wrong fit model. However, the robust estimate produces comparable power in the presence of bad leverage points.

For good leverage points, AIC tends, also, to produce overfit. On the other hand, the robust AIC tend to produce either correct fit or under fit model.

3.2. Example 2 (Stack Loss Data). Stack loss data was presented by [10]. This data set consists of 21 observations on three independent variables, and it contains four outliers

TABLE 3: Percentage of select models from classical AIC, robust RAIC, AIC_{LTS} , AIC_{LMS} , and AIC_{BS} , with vertical outliers.

ϵ %		AIC	RAIC	AIC_{LTS}	AIC_{LMS}	AIC_{BS}
0	Correct fit	84.6%	54.4%	57.6%	65.2%	45.2%
	Overfit	15.4%	0%	0%	0%	0%
	Under fit	0%	43.6%	41.2%	34%	54.6%
	Wrong fit	0%	2.0%	1.2%	0.8%	0.2%
5	Correct fit	2.8%	49.6%	56.8%	62.6%	45.4%
	Overfit	2%	0%	0%	0%	0%
	Under fit	22%	51%	42.8%	36.6%	54.6%
	Wrong fit	73.2%	1.4%	0.4%	0.8%	0%
10	Correct fit	3.2%	45.0%	51%	56%	39.8%
	Overfit	0.8%	0%	0%	0%	0%
	Under fit	20.2%	52.4%	48%	43.8%	55.2%
	Wrong fit	75.8%	2.6%	1.0%	0.2%	0%
20	Correct fit	4.2%	30.8%	49.4%	58%	34.2%
	Overfit	0.2%	0%	0%	0%	0%
	Under fit	22.8%	67.8%	49.6%	41.4%	65.8%
	Wrong fit	72.8%	1.4%	1.0%	0.6%	0%
30	Correct fit	1.6%	0%	44.0%	51.2%	23.2%
	Overfit	0.2%	0%	0%	0%	0%
	Under fit	22.6%	73.8%	55.4%	48.2%	76.8%
	Wrong fit	72.8%	26.2%	0.6%	0.6%	0%
40	Correct fit	2.2%	0%	37.6%	41.2%	12.4%
	Overfit	0.6%	0%	0%	0%	0%
	Under fit	22.8%	70.8%	62.4%	58.4%	87.4%
	Wrong fit	74.4%	29.2%	0%	0.4%	0.2%

(cases 1, 3, 4, and 21) and high leverage points (cases 1, 2, 3, and 21). The data are given in Table 6.

We applied the traditional and robust versions of AIC methods on the data. Table 7 shows that the classical method selects the full model, and robust RAIC method ignored one of the important variables (\mathbf{x}_{i2}), whereas robust AIC methods, based on high break down points estimators, agreed with the importance of the two variables, \mathbf{x}_{i1} and \mathbf{x}_{i2} .

4. Conclusions

The least trimmed squares (LTS) and the least median squares (LMS) are robust regression methods, frequently used in practice. Nevertheless, they are not commonly used for selecting models. This paper introduced the Akaike Information Criterion (AIC) based on LTS and LMS scales, which are robust against outliers and leverage point. Our simulation result illustrated excellent performance of AIC_{LTS} and AIC_{LMS} for contaminated data sets. This paper focused on the AIC variable selection criteria; one might be interested in extending other robust model selection criteria to advanced robust breakdown point estimation methods, such as LTS,

LMS, or BS estimators. In addition, this paper has considered regression model with continuous variables; however, future studies might consider mixed variables (i.e., continuous and dummy).

Appendix

Proof of Theorem 1. Consider

$$\begin{aligned}
 IF((\mathbf{x}, y), AIC, H) &= \frac{\partial}{\partial \epsilon} \left(AIC(\Delta_{(\mathbf{x}, y)}) \right) \Big|_{\epsilon=0} \\
 &= \frac{\partial}{\partial \epsilon} \left(\ln \frac{n-p}{n} \cdot S_p^2(H_\epsilon) \right) + 2p \\
 &= \frac{n}{(n-p) \cdot S_p^2(H)} \cdot \frac{\partial}{\partial \epsilon} \left(S_p^2(H_\epsilon) \right) \Big|_{\epsilon=0} \tag{A.1}
 \end{aligned}$$

$$= \frac{n}{(n-p) \sigma_p^2} \cdot 2\sigma_p \frac{\partial}{\partial \epsilon} \left(S_p(H_\epsilon) \right) \Big|_{\epsilon=0}, \tag{A.2}$$

TABLE 4: Percentage of select models from classical AIC, robust RAIC, and robust AIC_{LTS}, AIC_{LMS}, and AIC_{BS} with bad leverage points.

ϵ %		AIC	RAIC	AIC _{LTS}	AIC _{LMS}	AIC _{BS}
5	Correct fit	0%	0%	54.6%	60.8%	43.8%
	Overfit	70.4%	0%	0%	0%	0%
	Under fit	0%	64.4%	44%	38.4%	55.4%
	Wrong fit	29.6%	35.6%	1.4%	0.8%	0.8%
10	Correct fit	0%	0%	63.8%	67.8%	51.0%
	Overfit	63%	0%	0%	0%	0%
	Under fit	0%	54.6%	34.6%	31.4%	48.8%
	Wrong fit	37%	42.4%	1.6%	0.8%	0.2%
20	Correct fit	0%	0%	56.6%	63.4%	49.2%
	Overfit	54.8%	0%	0%	0%	0%
	Under fit	0.2%	60.8%	42.4%	35.8%	50.6%
	Wrong fit	44.8%	39.2%	1.0%	0.8%	0.2%
30	Correct fit	0%	0%	56.6%	61.4%	46%
	Overfit	37.6%	0%	0%	0%	0%
	Under fit	0.25%	29.6%	42.6%	36%	53.8%
	Wrong fit	60.4%	42.6%	0.8%	2.4%	0.2%
40	Correct fit	1.0%	0%	55.2%	64.6%	51.4%
	Overfit	13.8%	0%	0%	0%	0%
	Under fit	1.2%	54.4%	43.8%	33.8%	48.6%
	Wrong fit	81%	45.4%	1.0%	1.6%	0%

TABLE 5: Percentage of select models from classical AIC, robust RAIC, and robust AIC_{LTS}, AIC_{LMS}, and AIC_{BS}, with good leverage points.

ϵ %		AIC	RAIC	AIC _{LTS}	AIC _{LMS}	AIC _{BS}
5	Correct fit	0.2%	47.0%	53.6%	58.4%	42.6%
	Overfit	99.8%	0%	0%	0%	0%
	Under fit	0%	50.4%	46%	41.2%	57.4%
	Wrong fit	0%	2.6%	0.4%	0.4%	0%
10	Correct fit	0%	44.2%	54.6%	59.4%	40.4%
	Overfit	99.6%	0%	0%	0%	0%
	Under fit	0.2%	53.8%	44.6%	39.8%	59.6%
	Wrong fit	0.2%	2.0%	0.8%	0.8%	0%
20	Correct fit	0.8%	32.4%	50.4%	56.2%	33.2%
	Overfit	97.6%	0%	0%	0%	0%
	Under fit	0.8%	66.6%	48.4%	43%	66.4%
	Wrong fit	0.8%	1.0%	1.2%	0.8%	0.2%
30	Correct fit	1.8%	0%	46.0%	50.6%	27.0%
	Overfit	97.8%	96.8%	0%	0%	0%
	Under fit	2.8%	2.8%	53.6%	49.2%	72.8%
	Wrong fit	0.6%	0.4%	0.4%	0.2%	0.2%
40	Correct fit	0.2%	0%	37.4%	37%	13.8%
	Overfit	97.4%	100%	0%	0%	0%
	Under fit	2.2%	0%	62.6%	62.6%	86.2%
	Wrong fit	0.2%	0%	0%	0.4%	0%

TABLE 6: Stack loss data set.

\mathbf{x}_1	\mathbf{x}_2	\mathbf{x}_3	y
80	27	89	42
80	27	88	37
75	25	90	37
62	24	87	28
62	22	87	18
62	23	87	18
62	24	93	19
62	24	93	20
58	23	87	15
58	18	80	14
58	18	89	14
58	17	88	13
58	18	82	11
58	19	93	12
50	18	89	8
50	18	86	7
50	19	72	8
50	19	79	8
50	20	80	9
56	20	82	15
70	20	91	15

TABLE 7: The result variable selection of stack loss data.

Selected variables	AIC	RAIC	AIC _{LTS}	AIC _{LMS}	AIC _{BS}
\mathbf{x}_1	6.7	8.0	4.7	4.3	5.5
\mathbf{x}_2	7.1	6.5	5.9	5.4	7.0
\mathbf{x}_3	8.4	7.3	7.0	6.4	7.3
$\mathbf{x}_1, \mathbf{x}_2$	8.2	9.0	5.5	4.7	6.9
$\mathbf{x}_1, \mathbf{x}_3$	8.7	8.9	6.9	6.3	7.6
$\mathbf{x}_2, \mathbf{x}_3$	9.1	9.0	8.1	7.3	8.8
$\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$	4.7	10.6	7.6	6.7	9.1

where

$$\begin{aligned}
 \text{AIC}(\Delta_{(\mathbf{x}, y)}) &= \ln\left(\frac{n-p}{n} \cdot S_p^2(H_\epsilon)\right) + 2p, \\
 \text{AIC}(H) &= \ln\left(\frac{n-p}{n} \cdot S_p^2(H)\right) + 2p, \\
 S_p^2(H) &= \sigma_p^2, \\
 \left. \frac{\partial}{\partial \epsilon} S_p(H_\epsilon) \right|_{\epsilon=0} &= \left. \frac{\partial}{\partial \epsilon} S\left((1-\epsilon)F_{\sigma_p} + \epsilon\Delta_{r_{ip}}\right) \right|_{\epsilon=0} \\
 &= \sigma_p \left. \frac{\partial}{\partial \epsilon} S\left((1-\epsilon)F_0 + \epsilon\Delta_{r_{ip}/\sigma_p}\right) \right|_{\epsilon=0} \\
 &= \sigma_p IF\left(\frac{r_{ip}}{\sigma_p}, S, F_\sigma\right),
 \end{aligned} \tag{A.3}$$

where

$$r_{ip} = y_i - \hat{\mu} - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_p. \tag{A.4}$$

Inserting (A.2) into (A.1) yields

$$IF((\mathbf{x}, y), \text{AIC}, H) = \frac{2n}{(n-p)} IF\left(\frac{r_{ip}}{\sigma_p}, S, F_0\right). \tag{A.5}$$

□

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The author would like to thank Professor Nor Aishah Hamzah and Dr. Rossita M Yunus for their support to complete this study. The author is also grateful to the anonymous reviewers for their valuable comments on earlier draft of this paper. This research has been funded by University of Malaya, under Grant no. RG208-11AFR.

References

- [1] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Breakthroughs in Statistics*, pp. 610–624, Springer, 1992.
- [2] R. D. Martin, "Robust estimation of autoregressive models," in *Directions in Time Series*, pp. 228–254, 1980.
- [3] E. Ronchetti, *Robust testing in linear models: The infinitesimal approach [doctoral, thesis]*, University of Geneva, Geneva, Switzerland, 1982.
- [4] K. Tharmaratnam and G. Claeskens, "A comparison of robust versions of the AIC based on M-, S- and MM-estimators," *Statistics*, vol. 47, no. 1, pp. 216–235, 2013.
- [5] P. J. Huber, "Robust estimation of a location parameter," *Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.
- [6] W. S. Krasker and R. E. Welsch, "Efficient bounded-influence regression estimation," *Journal of the American statistical Association*, vol. 77, pp. 379–595, 1982.
- [7] P. J. Rousseeuw, "Least median of squares regression," *Journal of the American Statistical Association*, vol. 79, no. 388, pp. 871–880, 1984.
- [8] P. Rousseeuw, "Multivariate estimation with high breakdown point," in *Mathematical Statistics and Applications*, W. Grossmann, G. Pflug, I. Vinceze, and W. Wertz, Eds., pp. 283–297, Reidel, Dordrecht, The Netherlands, 1985.
- [9] P. J. Huber, *Robust Statistics*, Springer, New York, NY, USA, 2011.
- [10] K. A. Brownlee, *Statistical Theory and Methodology in Science and Engineering*, John-Wiley, New York, NY, USA, 2009.