Robust Linear Model Selection Using Paired Bootstrap

Fazli Rabbi^{1*}, Salahuddin², Alamgir¹ and Najma³

¹Department of Statistics, University of Peshawar, Khyber Pakhtunkhwa, Pakistan; fazlirabbi1@gmail.com ²CECOS University of IT and Emerging Sciences, Hayatabad, Peshawar, Khyber Pakhtunkhwa, Pakistan; salahuddin_90@yahoo.com ³Shaheed Benazir Bhutto Women University, Peshawar, Khyber Pakhtunkhwa, Pakistan; najma.fwu@gmail.com

Abstract

Objectives: To develop a robust paired bootstrap criterion for linear model selection and to compare the performance of the proposed criterion across different error distributions. **Methods/Analysis**: Our proposed robust paired bootstrap criterion is based on a robust conditional expected prediction loss function. We estimate the robust conditional expected prediction loss by using the *m*-out-of-*n* stratified bootstrap approach. The *m*-out-of-*n* bootstrap procedure is considered to obtain the asymptotic consistency. The effects of large residuals are reduced by using a robust ρ function. Model with a minimum robust prediction loss is used as a selection criterion. The usefulness of our proposed robust model selection procedure is investigated through real data set and Monte Carlo simulations under a variety of contamination and error structures. **Findings:** The conventional least squares selection procedures generally fail in the existence of outliers or in heavy-tailed error distributions. The stratified bootstrap selection procedure has shown good results as compared to simple bootstrap procedure. The proposed robust method has shown good robustness features with contaminated normal and heavy-tailed distributions. The proposed criterion outperforms the alternative procedure in both situations, i.e. in contamination-free data as well as in contaminated data. **Applications:** The proposed criterion can be used in both cases, i.e. in contamination-free data as well as in contaminated data, to select a model.

Keywords: MM Estimation, Outliers, Stratified Bootstrap, Out-Of-Bag Bootstrap, Robust Expected Prediction Loss

1. Introduction

Model selection procedures involve fitting a set of competing models and then selecting the model by comparing their prediction loss. Specifically, the robust estimators are used, when the data quality is questionable (i.e., assumptions about error distribution are not fulfilled). Various studies on model selection procedures depend on maximum likelihood-type or least squares approach such as^{1,2}, and³. All these criteria are severely affected when dataset contains outliers. Some other criteria for model selection are based on minimizing the expected squared prediction loss. The prediction loss

is estimated by re-sampling techniques such as cross-validation or the bootstrap⁴⁻⁶. However, these criteria are susceptible because of using the unbounded loss function when computing the prediction loss.

To cope with these problems to model selection, the researchers' proposed different approaches, some of these are robust versions of well-known standard criteria¹⁻³ or on various re-sampling techniques, like bootstrap or cross-validation⁷⁻²¹.

The rest of the paper is organized as: Section 2 describes paired bootstrap method to calculate prediction error, Section 3 discusses the stratified bootstrap procedure, Section 4 presents a RoBust Paired bootstrap

Criterion (RBPC) for model selection, Section 5 reports the simulation results, Section 6 demonstrates the data example and Section 7 provides the conclusion with a brief discussion.

2. Bootstrap Measures of Prediction Error

The re-sampling procedures such as the bootstrap and cross-validation are recommended to estimate the prediction error for variable/model selection. Bootstrap procedures involve obtaining multiple independent samples by sampling with replacement of the original data set. For paired bootstrap, sampling with replacement is done to form bootstrap samples from the original sample $\{(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)\}^{22}$ The bootstrap sample (y_i^*, x_i^*) for *i*=1,2,...,*n* may contain an observation from the original sample once or many times. The bootstrap estimate of β from paired bootstrap samples is:

$$\hat{\beta}^{\star} = \left(X^{\star^T} X^{\star}\right)^{-1} X^{\star^T} Y^{\star}.$$

The estimate of the prediction error for the k^{th} bootstrap sample from the original sample (y_i, x_i) , is given by:

$$\wedge_{k} = \frac{\sum_{i=1}^{n} \left(y_{i} - x_{i} \hat{\beta}_{k}^{*} \right)^{2}}{n} \tag{1}$$

Where β_k is the bootstrap estimator based on the k^{th} bootstrap sample. The unbiased estimator of prediction error for the k^{th} bootstrap sample is suggested by²³ as:

$$\hat{A}_{k} = \frac{\sum_{i=1}^{n} \left(y_{i} - x_{i} \hat{\beta}_{k} \right)^{2} + \sum_{i=1}^{n} \left(y_{i} - x_{i} \hat{\beta}_{k} \right)^{2} + \sum_{i=1}^{n} \left(y_{i}^{*} - x_{i}^{*} \hat{\beta}_{k} \right)^{2}}{n}$$
(2)

Where y_i is the vector of responses for the *i*th observation in the k^{th} bootstrap sample.Author⁵ showed that a selection procedure based on $\hat{\Lambda}_k$ is inconsistent? To obtain asymptotic consistency, an *m*-out-of-**n** bootstrap procedure used by⁵ for an appropriately chosen m < n.

Suppose that we have a column vector of *n* responses $Y = (y_1, y_2, ..., y_n)^T$ and *X* is an *n* x *p* design matrix. Let α represent a subset of size p_α from {1,2,...,p}, and X_α is an *n* x p_α matrix. Let $x_{\alpha i}^T$ represent the *i*th row vector of the design matrix X_α The linear regression model is given by

$$y_i = x_{ai}^T \beta_a + \sigma_a \epsilon_{ai}, \quad i = 1, 2, \dots, n$$
(3)

Where, $\sigma_{\alpha} > 0$, X_{α} and $\epsilon_{\alpha i} = (\epsilon_{\alpha i}, \epsilon_{\alpha 2}, ..., \epsilon_{\alpha n})^T$ are independent, and the errors $\epsilon_{\alpha i} \sim N(0, 1)$, β_{α} , is an unknown p_{α} -vector of regression coefficients. Let \mathcal{A} represent a collection of candidate models.

The interest here is to select a model α from \mathcal{A} that fits the data well. So, the model is indexed by $\alpha \in \mathcal{A}$ and β_{α} is estimated by the estimator β_{α} . A good model should have the ability to predict future observations with great accuracy, so for this purpose, one can use the conditional expected prediction error. For a given nonnegative loss function $\rho(.)$, the robust conditional expected prediction error for model α is measured as

$$M^{PE}(a) = \frac{\sigma^2}{n} E\left[\sum_{i=1}^{n} \rho\left(\frac{z_i - x_{ai}^T \hat{\beta}_a}{\sigma}\right) | y, X\right]$$
(4)

Where β_{α} is the estimator of β_{α} , σ is the measure of spread for a given data, and $z = (z_1, z_2, ..., z_n)^T$ is a vector of future responses at X, independent of y. Initially, the measure of conditional expected prediction error with $\rho(x) = \frac{x^2}{2}$ is considered by⁵ as as election criterion. Following⁵, the *m*-out-of-*n* stratified bootstrap procedure is used by¹⁴ to estimate the conditional expected prediction error. The estimated robust expected prediction error is given by:

$$M_{m,n}^{PE}(a) = \frac{\overset{\wedge}{\sigma}^{2}}{n} E_{\star} \left[\sum_{i=1}^{n} \rho \left(\frac{y_{i} - x_{ai}^{T} \overset{\wedge}{\beta}_{a,m}}{\breve{\sigma}} \right) \right]$$
(5)

Where $\hat{\beta}_{a,m}$ is the bootstrap estimate of $\hat{\beta}_a$ and E_* represents expectation with respect to the bootstrap distribution. The interest here is to choose a model $\alpha \in \mathcal{A}$ that minimizes $M_{mn}^{PE}(\alpha)$ i.e.:

$$\overline{a}_{m,n} = \arg\min_{a \in \mathcal{A}} M_{m,n}^{PE}(a)$$
(6)

Ignoring the stratification, the bootstrap estimator given in eq. (6) becomes a robust form of the estimator suggested by⁵.

3. The Stratified Bootstrap

Here we explain the main steps in applying the stratified bootstrap procedure. Following^{14,15}, we construct a stratified bootstrap sample of size *m* from the set of *n*

original observations { (y_i, x_i) : *i*=1,2..., *n*} by using the following steps:

- 1. Calculate the residuals i.e., $r_{ia} = y_i X_{ia}^T \hat{\beta}_a$ and then arrange them to obtain the order statistics.
- 2. Determine the number of strata, say *S* at between 3 and 8.
- 3. For each stratum set boundaries as 1/S, 2/S ... , (S-1)/S.
- 4. Allocate observations to the stratum *k* in which the residuals $r_{i\alpha}$ lie, so that $\sum_{k=1}^{s} n_k = n$, where n_k is the number of observation in stratum *k*.
- 5. From stratum *k*, sample rows of (*y*, *X*) independently with replacement and determine $m_k = n_k m/n$ (rounded number), such that $m = \sum_{k=1}^{s} m_k$.

In the simple expression, one can make an assumption about m_k , say these are integers. Then the bootstrap observations are denoted as: (y_{ki}^*, x_{ki}^*) , $i = 1, 2, ..., m_k$.

4. The Proposed Robust Model Selection Criterion

In this section, a RBPC is proposed for model selection. The proposed criterion is based on Robust Expected Prediction Loss (REPL). The robust expected prediction loss is estimated by using an m-out-of-n paired bootstrapping procedure. To obtain the bootstrap estimate of robust expected prediction loss, we follow the following steps:

- 1. From the full model calculate and arrangePearson residuals.
- Fix the number of strata S. The number of strata should be in between 3 to 8 as suggested by²⁴.
- 3. Allocate observations into different strata so that observations in the extreme tail are kept in lower or upper tail strata and other strata comprising the remaining observations.
- 4. From each stratum, sample rows of (y, X) independently with replacement so that total bootstrap sample of size is $m \le n$.
- 5. Construct the estimator $\beta_{a,m}$ from data obtained in step (d).
- 6. Repeat the steps (d) and (e) K independent times.

The modified robust expected prediction loss is estimated by $M_{m,n}^{PE^*}(a)$, where:

$$M_{m,n}^{PE^{\star}}(\alpha) = \frac{\overset{\wedge}{\sigma}^{2}}{n} E_{\star} \left[\sum_{i=1}^{n-m} \rho \left(\frac{y_{i[-m]} - x_{ai[-m]}^{*} \overset{\wedge}{\beta}_{a,m}}{\overset{\wedge}{\sigma}} \right) \right]$$
(7)

Where $\beta_{a,m}$ is the bootstrap estimate of $\beta_{a,m}$, E_* represents expectation with respect to the bootstrap distribution and *m* is the number of distinct observations in the boots trapsample and [-m] denotes the *m* observations are excluded when calculating $M_{m,n}^{PE^*}(a)$. To use eq. (7), we have to specify $\rho(.)$ and σ . We take to be bounded because our interest here is to fit and predict the core observations rather than those observations that lie in the tails. Here we prefer trimming, which means that for large $|\mathbf{x}|$, $\rho(.)$ is constant. Following¹⁴, we use a function which is just like a quadratic about the origin and becomes constant when it is away from the origin, such a $\rho(.)$ function is given by:

$$\rho(x) = \min(x^2, b^2). \tag{8}$$

The values of *b* can be varied, but b = 2 is reasonable in our simulation study. For simplicity, σ is measured by the MAD from the median and is given by:

$$\hat{\sigma} = 1.483 \operatorname{med}_{1 \le i \le n} \left| e_i - \operatorname{med}_{1 \le j \le n} \left(e_j \right) \right|$$

Where $e_i = y_i - x_{ai}^T \hat{\beta}_{a,i}, e_j = y_j - x_{aj}^T \hat{\beta}_{a,j}$ and $\hat{\beta}_a$ is the estimator for β_a .

With the proposed robust paired bootstrap criterion, we select a model $\alpha \in \mathcal{A}$ that minimizes $M_{m,n}^{PE^*}(\alpha)$, i.e.:

$$\hat{a}_{m,n} = \arg\min_{a \in \mathcal{A}} M_{m,n}^{PE^*}(a)$$
(9)

The optimal *m* depends on the true model, one should use $0.25n \le m \le 0.5n$ for moderate *n* (i.e., $50 \le n \le 200)^{14,15}$.

5. Simulation Study

We carry out two simulation studies to evaluate the performance of our proposed RPBC. The first one is designed to compare the behavior of our proposed procedure for contamination free dataset (simulation setting 1) with that of classical procedures (i.e. the AIC and BIC). The second simulation study demonstrates the utility of our RPBC in handling contaminated data (simulation setting 2).

5.1 Simulation Setting 1

To compare the performance of our proposed bootstrap model selection criterion in the no-outlier case, the following regression model is considered:

$$Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \epsilon_i, \qquad i = 1, 2, \dots, 40$$

where $\epsilon_i \sim N(0,1), XI$, is the column of 1's and X_2, X_3, X_4 , and X_5 are taken from the solid waste data of $\frac{25}{25}$, the same as those used in $\frac{4.5,13,14,16.26,27}{4}$. The comparison of our proposed criterion $\hat{\alpha}_{m,n}$ (eq. (9)) is made with a robust version of $\bar{\alpha}_{m,n}$ (eq. (6)). In the case of zero contamination, the LS estimator is used to fit the regression models.

The estimated selection probabilities for our proposed criterion $\hat{\alpha}_{m,n}$ and $\bar{\alpha}_{m,n}$ using the LS estimator and robust loss function $\rho(x) = \min(x^2, b^2)$ are given in Table 1. The results presented in Tables 1 are obtained for L=1,000 MC simulations using simple bootstrap and K=100 bootstrap replications for *m*=15, 20, 30. For comparison purpose, the classical procedures such as AIC and BIC are also included.

The results based on our simulation study are summarized as follows:

- The proposed selection criterion $\hat{\alpha}_{m,n}$ performs very well as compared to $\bar{\alpha}_{m,n}$ and other classical criteria. For example, for $\beta_2 = (2,0,0,4,8)$ we see that the modified criterion $\hat{\alpha}_{20,40}$ selects the optimal model 95.1% $(sd_{0.951} = 0.007), \bar{\alpha}_{20,40}$ selects the optimal model 86.0% $(sd_{0.860} = 0.011)$, the AIC selects the optimal model 69.6% $(sd_{0.696} = 0.015)$ and the BIC selects the optimal model 83.5% $(sd_{0.835} = 0.012)$.
- The proposed procedure clearly improves for smaller *m*. For example, the percent correct is 84.6% of the time for *m* = 30, which is much less than the 97.1% for *m* =15, using *ρ*(*x*) = min (*x*², *b*²) and β₁ = (2,0,0,4,0)
- The proposed criterion $\hat{\alpha}_{m,n}$ is less dependent on m(bootstrap sample size) as compared to $\bar{\alpha}_{m,n}$.
- If the optimal model is full model, thenoutperforms our proposed criterion

Table 1. Selection probabilities of $\hat{\alpha}_{m,n}$ and $\bar{\alpha}_{m,n}$ based on the LS estimator and $\rho(x) = \min(x^2, b^2)$

True β	Model	m=	=15	m=20		m=30		AIC	BIC
		$\bar{\alpha}_{15.40}$	$\hat{\alpha}_{_{15,40}}$	$\bar{\alpha}_{_{20,40}}$	â _{20,40}	$\bar{\alpha}_{_{30,40}}$	â _{30,40}		
(2,0,0,4,0)	1,4*	0.897	0.971	0.800	0.927	0.559	0.846	0.590	0.778
	1,4,5	0.028	0.010	0.053	0.021	0.106	0.046	0.104	0.058
	1,3,4	0.029	0.010	0.071	0.018	0.139	0.054	0.102	0.066
	1,2,4	0.042	0.009	0.060	0.031	0.109	0.046	0.106	0.066
	1,3,4,5	0.001	0.000	0.006	0.001	0.027	0.003	0.027	0.004
	1,2,4,5	0.002	0.000	0.005	0.002	0.025	0.002	0.027	0.013
	1,2,3,4	0.001	0.000	0.005	0.000	0.021	0.003	0.024	0.009
	1,2,3,4,5	0.000	0.000	0.000	0.000	0.014	0.000	0.020	0.006
(2,0,0,4,8)	1,4,5*	0.934	0.981	0.860	0.951	0.695	0.893	0.696	0.835
	1,3,4,5	0.023	0.010	0.065	0.021	0.139	0.052	0.122	0.068
	1,2,4,5	0.043	0.009	0.068	0.028	0.124	0.049	0.136	0.075
	1,2,3,4,5	0.000	0.000	0.007	0.000	0.042	0.006	0.046	0.022
(2,9,0,4,8)	1,2,5	0.000	0.005	0.000	0.001	0.000	0.001	0.000	0.003
	1,3,4,5	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.001
	1,2,4,5*	0.979	0.989	0.932	0.972	0.825	0.939	0.829	0.905
	1,2,3,4,5	0.021	0.006	0.068	0.027	0.174	0.059	0.171	0.091
(2,9,6,4,8)	1,3,4,5	0.008	0.036	0.002	0.010	0.001	0.005	0.000	0.001
	1,2,4,5	0.001	0.004	0.000	0.001	0.000	0.001	0.000	0.001
	1,2,3,5	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
	1,2,3,4,5*	0.991	0.959	0.998	0.989	0.999	0.994	1.000	0.998

Note: (*) denotes the optimal model.

5.2 Simulation Setting 2

Another simulation study iscarried out to show the performance of our proposed RPBC in handling contaminated data. Furthermore, the use of stratified bootstrap is explored when the atypical observations are present in the data. For this purpose the following regression model is considered

$$y_i = 2 + 2x_{i1} + 0x_{i2} + \epsilon_i,$$
 $i = 1, 2, \dots, 64$

where the design matrix X had columns generated as uniform on [-1, 1]; and X is kept constant for all simulation replications. We choose five different error distributions that are deviated from normality:

- 1. ϵ_1 is [3/8] outliers (i.e., 62.5% [5/8] from a standard normal and 37.5% [3/8] from a normal with, $\mu = 30 2 2x$, and $\sigma = 1$);
- 2. ϵ_2 is [1/4] outliers (i.e., 75% [3/4] from a standard normal and 25% [1/4] from a normal with, $\mu = 30 2$ -2 x_1 and $\sigma = 1$);

- 3. ϵ_3 is [1/8] outliers (i.e., 87.5% [7/8] from a standard normal and 12.5% [1/8] from a normal with, $\mu = 30 2 2x_1$ and $\sigma = 1$);
- 4. ϵ_4 Is the Cauchy distribution; and
- 5. ϵ_{5} is the slash distribution

In Table 2 the following possible models are considered:

- Model (1) means, amodel with intercept only;
- Model (1, 2) means a model having intercept and X₁;
- Model (1,3) means a model having intercept and X₂; and
- Model (1, 2, 3) means the full model.

The MM-estimator of²⁸ is used to fit the robust regression models. For such purpose, the rlm () function in R is used for estimating the regression parameters. The selection probabilities of the criterion $\bar{\alpha}_{m,n}$ and our proposed criterion $\hat{\alpha}_{m,n}$ on the basis of un-stratified bootstrap and stratified bootstrap (using four strata) are computed. The selection probabilities areobtained

Table 2. Selection probabilities of and based on the simple bootstrap and the stratified bootstrap usingMM-estimator and simple bootstrap using LS-estimator

Errors	Model		MM-e	LS-estimator				
		Simple E	Bootstrap	Stratified	bootstrap	Simple Bootstrap		
		$\bar{\alpha}_{24.64}$	<i>α</i> _{24,64}	$\bar{\alpha}_{24,64}^{SB}$	$\hat{\alpha} _{24,64}^{SB}$	$\bar{\alpha}_{24.64}$	$\hat{\alpha}_{_{24,64}}$	
ϵ_1	1	0.368	0.392	0.166	0.189	0.756	1.000	
	1,3	0.000	0.000	0.000	0.000	0.244	0.000	
	1,2*	0.290	0.362	0.721	0.753	0.000	0.000	
	1,2,3	0.342	0.246	0.113	0.058	0.000	0.000	
ϵ_{2}	1	0.000	0.000	0.000	0.000	0.999	1.000	
	1,3	0.000	0.000	0.000	0.000	0.000	0.000	
	1,2*	0.935	0.968	0.905	0.966	0.001	0.000	
	1,2,3	0.065	0.032	0.095	0.034	0.000	0.000	
$\epsilon_{_3}$	1	0.000	0.000	0.000	0.000	1.000	1.000	
	1,3	0.000	0.000	0.000	0.000	0.000	0.000	
	1,2*	0.894	0.952	0.887	0.956	0.000	0.000	
	1,2,3	0.106	0.048	0.113	0.044	0.000	0.000	
$\epsilon_{_4}$	1	0.008	0.016	0.006	0.013	0.770	0.823	
	1,3	0.000	0.000	0.000	0.000	0.001	0.001	
	1,2*	0.954	0.974	0.937	0.966	0.227	0.175	
	1,2,3	0.038	0.010	0.057	0.021	0.002	0.001	
$\epsilon_{_{5}}$	1	0.062	0.139	0.047	0.097	0.828	0.882	
	1,3	0.004	0.005	0.007	0.003	0.005	0.002	
	1,2*	0.885	0.842	0.889	0.877	0.164	0.116	
	1,2,3	0.049	0.014	0.057	0.023	0.003	0.000	

Note: (*) denotes the optimal model. For both selection criteria.

by using 1,000 MC simulations and *K*=100 bootstrap replications. These results are mentioned in Table 3.

From the simulation outputs in Table 3, it is clear that for all error distributions our proposed robust criterion $\hat{\alpha}_{mn}$ performs very well as compared to $\bar{\alpha}_{mn}$. However, in the case of Slash error distribution, our modified procedure does not perform well as compared to. The output indicates that the proposed model selection procedure using the robustfunction in (8) and MM-estimator is robust in the presence of highly contaminated data. For example, the percent correct is 36.2% for un-stratified bootstrap, whereas the percent correct is75.3% for stratified bootstrap for the contaminated normal situation ϵ_1 . Furthermore, we observe that, in the presence of outliers and heavytailed error distributions, the conventional least squares selection procedures are not performing well. For example, under ϵ_4 error distribution, the percent correct is 96.6% for our proposed robust criterion using MM-estimator, whereas the percent correct is 17.5% for the least squares procedure. Similarly, under \in error distribution, our proposed procedure with MM-estimator is 87.7% correct, whereas the least squares procedure is11.6% correct.

These outputs show that the proposed robust method has good robustness features with contaminated normal and heavy-tailed distributions, while the least squares procedure performs very poorly in both situations. This clearly proves the lack of robustness of the least squares procedure in the presence of outliers and heavy-tailed distributions.

5.3 Data Example (Stack Loss Data)

In this section, we analyze the Stack loss data presented by^{29} . This dataset consists of three explanatory variables,

		-		
Selected variables	$\bar{\alpha}_{10.21}$	$\hat{\alpha}_{10,21}$	AIC	BIC
X	2.97	3.50	4.61	4.71
X ₂	3.02	3.59	2.19	2.28
X ₃	3.48	4.20	3.21	3.31
X ₁ , X ₂	1.70	2.55	1.55	1.70
X ₁ , X ₃	2.38	3.42	2.69	2.84
X ₂ , X ₃	2.09	3.49	1.47	1.62
X ₁ , X ₂ , X ₃	1.81	3.05	1.34	1.54

 Table 3. Selected best model for the stack loss data

 using a range of model selection procedures

and it contains four outliers, namely observations 1,3,4, and 21. The response is the Stack loss (y) observed on n=21 observations. The explanatory variables are theFlow of cooling air (X₁), Cooling Water Temperature (X₂), and Concentration of acid (X₃). We applied our robust method, the existing methods, and the traditional methods on the data. Table 2 presents a summary of selected best models. Table 2 shows that the classical methods select the full model whereas robust criteria agreed with the importance of the two variables, X₁ and X₂. The best model according to our criterion includes X₁ and X₂.

6. Conclusion

A RPBC is proposed to select the best subset of variables in linear regression models. The proposed method is based on robust conditional expected prediction loss and a robust -function. A stratified bootstrap procedure is used to estimate the expected prediction loss function.We recommended a robust -function to decrease the effect of large residuals. From the simulation outputs, it is clear that in both cases, i.e. in contamination-free data as well as in contaminated data, the proposed selection criterion performed well. The results show that the proposed procedure has good robustness features with contaminated normal and heavy-tailederrors distributions, while the least squares method performed very poorly in both situations. This clearly proved the lack of robustness of the least squares procedure in the presence of outliers and heavy-tailed distributions.Furthermore, our proposed criterion is less dependent on a bootstrap sample of sizemas compared to the robust version of⁵. In conclusion, our proposed criterion is superior and will do better when the data generating model is small.

7. References

- Akaike H. A new look at the statistical model identification, IEEE Transactions on Automatic Control. 1974; 19(6): 716-23. https://doi.org/10.1109/TAC.1974.1100705.
- Mallows CL. Some comments on CP, Technometrics. 1973; 15(4):661-75. https://doi.org/10.2307/1267380, https://doi. org/10.1080/00401706.1973.10489103.
- Schwarz G. Estimating the dimension of a model, The Annals of Statistics. 1978; 6(2):461-64. https://doi.org/10.1214/ aos/1176345415, https://doi.org/10.1214/aos/1176344136.

- Shao J. Linear model selection by cross-validation, Journal of the American Statistical Association. 1993; 88(422):486-94. https://doi.org/10.1080/01621459.1993.10476299.
- Shao J. Bootstrap model selection, Journal of the American Statistical Association. 1996; 91(434):655-65. https://doi. org/10.1080/01621459.1996.10476934.
- Breiman L. Better subset regression using the nonnegative garrote, Technometrics. 1995; 37(4):373-84. https://doi.org /10.1080/00401706.1995.10484371.
- Machado JA. Robust model selection and M-estimation, Econometric Theory. 1993; 9(3):478-93. https://doi.org/ 10.1017/S0266466600007775.
- Sommer S, Staudte RG. Robust variable selection in regression in the presence of outliers and leverage points, Australian Journal of Statistics. 1995; 37(3):323-36. https:// doi.org/10.1111/j.1467-842X.1995.tb00663.x.
- Ronchetti E, Field C, Blanchard W. Robust linear model selection by cross-validation, Journal of the American Statistical Association. 1997; 92(439):1017-23. https://doi. org/10.1080/01621459.1997.10474057.
- Ronchetti E, Staudte RG. A robust version of Mallows's CP, Journal of the American Statistical Association. 1994; 89(426):550-59. https://doi.org/10.1080/01621459.1994.10 476780, https://doi.org/10.2307/2290858.
- Sommer S, Huggins RM. Variables selection using the Wald test and a robust CP, Applied Statistics. 1996; 15-29. https:// doi.org/10.2307/2986219.
- Agostinelli C. Robust model selection in regression via weighted likelihood methodology, Statistics and Probability Letters. 2002; 56(3):289-300. https://doi.org/10.1016/S0167-7152(01)00193-6.
- Wisnowski JW, Simpson JR, Montgomery DC, Runger GC. Resampling methods for variable selection in robust regression, Computational Statistics and Data Analysis. 2003; 43(3): 341-55. https://doi.org/10.1016/S0167-9473(02)00235-9.
- Müller S, Welsh A. Outlier robust model selection in linear regression, Journal of the American Statistical Association. 2005; 100(472):1297-310. https://doi.org/10. 1198/016214505000000529.
- 15. Müller S, Welsh A. Robust model selection in generalized linear models, Statistica Sinica. 2009; 1155-70.
- Salibian-Barrera M, Van Aelst S. Robust model selection using fast and robust bootstrap, Computational Statistics and Data Analysis. 2008; 52(12):5121-35. https://doi. org/10.1016/j.csda.2008.05.007.

- Tharmaratnam K, Claeskens G. A comparison of robust versions of the AIC based on M, S-and MM-estimators. Statistics. 2013; 47(1):216-35. https://doi.org/10.1080/0233 1888.2011.568120.
- Sakate D, Kashid D. A new robust model selection method in GLM with application to ecological data, Environmental Systems Research. 2016; 5(1):9. https://doi.org/10.1186/ s40068-016-0060-7.
- Mahmood Z, Salahuddin. A novel resampling method for variable selection in robust regression, Pakistan Journal of Statistics. 2015; 31(3):327-38.
- 20. Robust model selection between population growth and multiple mergers coalescents. Date accessed: 18/06/2018. https://arxiv.org/abs/1806.06746.
- 21. Robust Bayesian Model Selection for Variable Clustering with the Gaussian Graphical Model. Date accessed: 15/06/2018. https://arxiv.org/abs/1806.05924.
- 22. The jackknife, the bootstrap and other resampling plans. Date accessed: 1980. https://statistics.stanford.edu/ research/jackknife-bootstrap-and-other-resampling-plans.
- Efron B. Estimating the error rate of a prediction rule: Improvement on cross-validation, Journal of the American Statistical Association. 1983; 78(382):316-31. https://doi. org/10.1080/01621459.1983.10477973.
- 24. Sampling Technique. Date accessed: 1977. https:// hwbdocuments.env.nm.gov/Los%20Alamos%20National%20 Labs/General/14447.pdf.
- Regression Analysis and its Applications: A Data-Oriented Approach. Date accessed: 01/07/1980. https://www.crcpress. com/Regression-Analysis-and-its-Application-A-Data-Oriented-Approach/Gunst-Mason/p/book/9780824769932.
- Shao J. An asymptotic theory for linear model selection, Statistica Sinica. 1997; 221-42.
- 27. Wu Y. An M-estimation-based model selection criterion with a data-oriented penalty, Journal of Statistical Computation and Simulation. 2001; 70(1):71-87. https://doi.org/10.1080/00949650108812108.
- Yohai VJ. High breakdown-point and high efficiency robust estimates for regression, The Annals of Statistics. 1987; 642-56. https://doi.org/10.1214/aos/1176350366.
- 29. Statistical Theory and Methodology in Science and Engineering. Date accessed: 15/01/1965. https://www. amazon.com/Statistical-Theory-Methodology-Science-engineering/dp/0471113557.