

Article

Properties of Correlation Determination Based on Qn Estimator for Standard Logistic Distribution with Contamination

Nor Adilla Che Ahmad¹, Lim Fong Peng¹, Yow Kai Siong¹ and Yap Hong Keat²

¹ Department of Mathematics and Statistics, Faculty of Science, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor Darul Ehsan, Malaysia.

² Department of Mathematical and Actuarial Sciences, Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Bandar Sungai Long, Cheras, 43000 Kajang, Selangor Darul Ehsan, Malaysia.

Correspondence should be addressed to:

Lim Fong Peng; fongpeng@upm.edu.my

Article Info

Article history:

Received: 26 April 2026

Accepted: 15 October 2026

Published: 30 October 2026

Academic Editor:

Shahrina Ismail

Malaysian Journal of Science,
Health & Technology

MJoSHT2026, Volume 12, Issue No. 2
eISSN: 2601-0003

<https://doi.org/10.33102/mjosht.574>

Copyright © Nor Adilla Che Ahmad et al. This is an open access article distributed under the Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract— Robust statistical methods have become increasingly important in the presence of contaminated data, particularly in regression-based goodness of fit testing. The classical coefficient of determination is highly sensitive to deviations from assumptions such as normality and the absence of contaminated data. To address this limitation, this study investigates the properties of a robust coefficient of determination derived from the Qn estimator, which is a high-breakdown-point scale estimator. This research focuses on the behaviour of the sampling distribution under the standard logistic distribution with varying contamination structures. A comprehensive simulation study is conducted under different sample sizes, contamination levels and distributions of the contaminants. Bootstrap bias-corrected and accelerated (BCa) confidence intervals are constructed to evaluate the estimator's statistical properties in the presence of contaminated data. Results demonstrate that contamination significantly influences skewness and variability, although the estimator remains bounded and robust.

Keywords—Qn estimator; goodness of fit; correlation of determination; contamination, robust estimator

I. INTRODUCTION

Goodness of fit test is essential in statistical modelling to determine whether observed data conform to a specified theoretical distribution [1]. Regression-based approaches are commonly used due to their simplicity and interpretability, where the relationship between ordered observations and theoretical quantiles is measured using correlation-based metrics, such as the coefficient of determination.

Due to its widespread use, the classical coefficient of determination is highly sensitive to outliers and contamination. It can be significantly distorted even by a small proportion of

contaminated data. This limitation has motivated the development of robust correlation measures that remain stable under deviations from classical assumptions [2].

The Qn estimator, a highly robust and efficient statistical measure of scale proposed by Rousseeuw and Croux [3], provide an effective alternative due to their high breakdown point and resistance to outliers. It has a 50% breakdown point, high Gaussian efficiency (82%), and does not require a prior location estimate, making it effective for skewed data. Furthermore, the Qn estimator, designed as a superior alternative to the Median Absolute Deviation (MAD), is

particularly suitable for constructing robust measures of scale and correlation.

Recent studies have further expanded the field of robust correlation. For instance, robust inference methods under heterogeneous data conditions have been explored in econometric contexts [4], while high-dimensional robust correlation estimation has gained attention in modern statistical applications [5]. Additionally, real-time and streaming data scenarios have motivated the development of fast robust correlation estimators [6]. Researchers are also interested in investigating the sampling properties, as well as the confidence intervals of the coefficient of determination and the adjusted coefficient of determination [7] – [16].

Even with these advances, the behaviour of the coefficient of determination based on the Q_n estimator under contaminated distributions remains insufficiently explored. Hence, this study aims to fill this gap by analysing the sampling distribution of the coefficient of determination based on the Q_n estimator under various contamination scenarios, such as different sample sizes, contamination levels and the distribution of the contaminants. Bootstrapping BCa is used to investigate whether the estimator remains bounded and stable under contamination [17] – [19]. In this study, the logistic distribution is considered as it is commonly utilised in statistical modelling due to its flexibility and heavier tails compared to the normal distribution. Its probability density function is symmetric, but the presence of heavier tails makes it more sensitive to extreme observations. This property makes the logistic distribution particularly suitable for studying the robustness of statistical estimators [20] – [21]. Since the standard logistic distribution is foundational for logistic regression, modelling, and growth processes, it serves as an appropriate framework for evaluating the performance of robust estimators.

II. METHODOLOGY

This section briefly introduces the standard logistic distribution and presents the generation of the coefficient of determination based on Q_n for the standard logistic distribution. It is then followed by the bootstrapping of confidence intervals for the mean and standard deviation of $R_{Q_n}^2$.

A. Standard Logistic Distribution

The logistic distribution is widely applied in growth modelling and serves as the foundation for logistic regression. In probability theory and statistics, it is classified as a continuous probability distribution without a shape parameter. Consequently, its probability density function (pdf) always takes a bell-shaped curve. This curve resembles the normal distribution but is distinguished by its heavier tails. Let x be a random variable drawn from a logistic distribution. The probability density function (pdf) of the logistic distribution is expressed as:

$$f(x) = \frac{e^w}{\sigma(1 + e)^{w^2}} \quad ; \quad -\infty < x < \infty, -\infty < \mu < \infty, \sigma < 0$$

where $w = \frac{x-\mu}{\sigma}$, μ = location parameter and σ = scale parameter; its cumulative distribution function (cdf) is

$$F(x) = \frac{e^w}{1 + e^w}.$$

Thus, the pdf and cdf of the standard logistic distribution are

$$f(x) = \frac{e^x}{(1 + e)^{x^2}}$$

and

$$F(x) = \frac{e^x}{1 + e^x}$$

respectively.

B. Correlation Coefficient Based on Q_n Estimator, r_{Q_n}

Let $X = \{x_1, x_2, \dots, x_n\}$. This section introduces a correlation-based counterpart to the work of Bickel and Lehmann [22], obtained by substituting pairwise averages with pairwise distances, resulting in:

$$\text{med}\{|x_i - x_j| : i < j\}.$$

The Q_n estimator can be defined as

$$Q_n = d \cdot \text{med}\{|x_i - x_j| : i < j\}$$

where d is a normalisation constant, which depends on sample size and asymptotically $d \approx 2.2219$.

Furthermore, Zakaria et al. (2017) define r_{Q_n} as

$$r_{Q_n} = \frac{Q_n^2 u - Q_n^2 v}{Q_n^2 u + Q_n^2 v} \quad (1)$$

where

$$u = \frac{x_i - \text{med}(x_i)}{\sqrt{2}Q_n(x_i)} + \frac{h_i - \text{med}(h_i)}{\sqrt{2}Q_n(h_i)}$$

and

$$v = \frac{x_i - \text{med}(x_i)}{\sqrt{2}Q_n(x_i)} - \frac{h_i - \text{med}(h_i)}{\sqrt{2}Q_n(h_i)}.$$

Hence,

$$R_{Q_n}^2 = r_{Q_n}^2.$$

Zakaria et al. [23] verified that it is evidently not true that $|r_{Q_n}| \leq 1$. Thus, r_{Q_n} may sometimes be beyond +1 or -1.

C. Generation of $R_{Q_n}^2$ for the Standard Logistic Distribution

Let $X_{(i)}$ denote the i^{th} order statistic and let T_i be a set of constants for $i = 1, \dots, n$. A regression test is a graphical technique related to probability plots, where the order statistics $X_{(i)}$ are plotted on the vertical axis against T_i on the horizontal axis. A straight line is then fitted to the plotted points, and the coefficient of determination R^2 is used to assess the strength of the relationship between $X_{(i)}$ and T_i . Refer to Maarof et al. [24] – [26] for the details of the regression test based on R^2 . In this study, a similar regression test is conducted by using the same approximate value of T_i for standard logistic distribution given by D'Agostino and Stephens [27], that is

$$H_i = \log\{i/(n + 1)\}. \quad (2)$$

Then, each pair (X_i, H_i) is substituted into equation (1) for standard logistic distribution in the previous section.

The study advanced by simulating data to derive the sampling distribution of $R_{Q_n}^2$ under contamination from the standard logistic distribution. Specifically, the sampling distribution of $R_{Q_n}^2$ is generated using logistic contaminants (logistic (0, 0.2), logistic (2, 0.2), and logistic (2, 1)) as well as a normal contaminant (normal (3, 0.2)). Contamination levels of 5%, 25%, and 50% are considered, with sample sizes $n = 20, 40$, and 100 . For example, in the case of $n = 20$ with 5% logistic (2, 0.2) contamination, the sample consisted of 95% observations from the standard logistic distribution and 5% from the logistic (2, 0.2) distribution. The same procedure is applied across all contamination scenarios.

In this study, introducing contamination during sample generation is crucial for examining how sample size, contamination proportion, and contaminant distribution influence the confidence intervals for the mean and standard deviation of $R_{Q_n}^2$. The logistic contaminants (logistic (0, 0.2), logistic (2, 0.2), and logistic (2, 1)) are selected to assess the behaviour of confidence intervals as the location and scale parameters of the logistic distribution increased. For the normal contaminants, parameter values are chosen to remain reasonably close to those of the logistic contaminants.

For cases without outliers, the following procedure is used to calculate $R_{Q_n}^2$ for the standard logistic distribution:

Step 1. Using R, a random sample of order statistics ($X_1, X_2, \dots, X_n$) is generated from the standard logistic distribution with parameters $\mu = 0$ and $\sigma = 1$, for a sample of size n .

Step 2. The approximate values of T_i , denoted as H_i , are computed using equation (2).

Step 3. The X_i 's are regressed on the H_i 's, and r_{Q_n} is calculated using equation (1). This is followed by the computation of $R_{Q_n}^2$.

Step 4. Steps 1 – 3 are repeated 5000 times to generate 5000 values of $R_{Q_n}^2$ for the standard logistic distribution.

Step 5. Steps 1 – 4 are repeated for different sample sizes ($n = 20, 40, 100$).

For cases with outliers, the order statistics used in calculating $R_{Q_n}^2$ are drawn from contaminated logistic distributions: logistic (0, 0.2), logistic (2, 0.2), logistic (2, 1), and normal (3, 0.2). The same simulation procedure (Steps 2 – 5) is applied, and 5000 values of $R_{Q_n}^2$ are obtained for contamination levels of 5%, 25%, and 50%, with sample sizes $n = 20, 40, 100$.

D. Bootstrap Confidence Intervals for the Mean and Standard Deviation of $R_{Q_n}^2$

In this study, bias-corrected and accelerated (BCa) confidence intervals are obtained through simulation using sample sizes of 20, 40, and 100, as this method provides greater accuracy compared with other bootstrap interval approaches. As noted by Efron [28], BCa intervals achieve second-order accuracy, whereas the percentile (P), bias-corrected (BC), and asymptotic normal intervals attain only first-order accuracy.

The bootstrap procedure can be outlined as follows. Let $\hat{\theta}^{*(\alpha)}$ denote the $100 \cdot \alpha^{th}$ percentile of the B bootstrap estimates $\hat{\theta}^{*(1)}, \hat{\theta}^{*(2)}, \dots, \hat{\theta}^{*(B)}$. Then, the percentile interval $(\hat{\theta}_{lo}, \hat{\theta}_{up})$ corresponding to a coverage of $1 - 2\alpha$ is given by

$$(\hat{\theta}_{lo}, \hat{\theta}_{up}) = (\hat{\theta}^{*(\alpha)}, \hat{\theta}^{*(1-\alpha)}).$$

A similar process is repeated 1000 times ($B = 1000$), after which the means and standard errors for both the mean and standard deviation of $R_{Q_n}^2$ are computed. In addition, 95% confidence intervals for these quantities are constructed.

The statistic $R_{Q_n}^2$, used in the bootstrap procedure, is generated using the method described in the previous section. All bootstrap estimates of the mean, standard deviation, and the corresponding associated confidence intervals for the various scenarios considered in this study are obtained via R software.

III. RESULTS

This section discusses the results of the bootstrap estimates for the mean and standard deviation of the sampling distribution of $R_{Q_n}^2$, simulated under varying sample sizes, contamination levels, contaminant distributions, and the location and scale parameters of the generating distributions.

Firstly, simulation studies analyse the sampling distribution of the correlation measure under contamination, sample size variations, and different contamination levels. In cases without outliers, the distribution becomes more skewed as sample size increases. Meanwhile, contamination affects the mean, standard deviation, and skewness of the sampling distribution. In this context, logistic contaminants (5%, 25%, 50%) and normal contaminants influence the distribution's properties. Logistic contaminants with increased location and scale impact the confidence interval ranges. Meanwhile, normal contaminants tend to stabilise the distribution but still affect the properties depending on contamination level.

In addition, this study examines how different types and levels of contaminants affect the sampling distribution of a statistic derived from contaminated logistic data, using the mean, standard deviation and bootstrap confidence intervals. As the percentage of logistic contaminants increases, the mean of the sampling distribution decreases from approximately 0.96 to 0.86. Standard deviation increases from about 0.03 to 0.04 with higher contamination levels. Graphs become more skewed to the left as sample size increases, especially with higher location parameters. Generally, skewness decreases when sample size and percentage of contaminants increase but increases with higher distribution location and scale. Confidence intervals for the mean and standard deviation widen as contamination percentage and distribution location or scale increase.

On the other hand, normal contaminants tend to improve symmetry and reduce variability as contamination percentage increases. The mean remains around 0.90 except at 50% contamination, where it drops. Meanwhile, standard deviation increases from 0.03 to 0.07 as sample size grows. Distribution graphs become more symmetric with increased sample size and contamination percentage. The skewness decreases, indicating improved symmetry, while kurtosis stabilises. Besides, confidence intervals for the mean and standard deviation narrow with larger sample sizes but widen with higher contamination levels.

Finally, this study demonstrates the use of bootstrap BCa confidence intervals for assessing the bounds of the mean and

standard deviation under various contamination scenarios. As expected for the cases with uncontaminated data, Table I indicates that the range of the confidence intervals for the mean and standard deviation of $R_{Q_n}^2$ narrow as sample size increases, with ranges decreasing from about 0.0054 to 0.0010. On the other hand, for the cases with contaminated data, Tables II – V show similar trends to those for cases with uncontaminated data. However, the range of the confidence intervals for the mean and standard deviation of $R_{Q_n}^2$ are wider, especially with higher contamination percentages.

Table I. 95% Confidence intervals of $R_{Q_n}^2$ for cases without contamination data

n		Observed	BCa (95%)	Range
20	mean	0.0762	(0.0748, 0.0802)	0.0054
	standard deviation	0.0972	(0.0939, 0.1011)	0.0072
40	mean	0.0355	(0.0325, 0.0351)	0.0026
	standard deviation	0.0475	(0.0470, 0.0490)	0.0020
100	mean	0.0128	(0.0123, 0.0133)	0.0010
	standard deviation	0.0179	(0.0171, 0.0189)	0.0018

Table II. 95% Confidence Intervals of $R_{Q_n}^2$ For Cases Logistic (0, 0.2) Contaminants

n		Contamination	Observed	BCa (95%)	Range
20	mean	5%	0.9633	(0.9626, 0.9641)	0.0015
		25%	0.9503	(0.9492, 0.9513)	0.0021
		50%	0.9375	(0.9363, 0.9387)	0.0024
	standard deviation	5%	0.0268	(0.0254, 0.0285)	0.0031
		25%	0.0384	(0.0369, 0.0404)	0.0035
		50%	0.0438	(0.0421, 0.0456)	0.0035
40	mean	5%	0.9805	(0.9801, 0.9809)	0.0008
		25%	0.9674	(0.9668, 0.9681)	0.0013
		50%	0.9564	(0.9557, 0.9571)	0.0014
	standard deviation	5%	0.0133	(0.0126, 0.0141)	0.0015
		25%	0.0232	(0.0223, 0.0242)	0.0019
		50%	0.0273	(0.0264, 0.0283)	0.0019
100	mean	5%	0.9920	(0.9919, 0.9922)	0.0003
		25%	0.9789	(0.9785, 0.9792)	0.0007
		50%	0.9668	(0.9663, 0.9672)	0.0009
	standard deviation	5%	0.0047	(0.0045, 0.0049)	0.0004
		25%	0.0133	(0.0129, 0.0138)	0.0009
		50%	0.0160	(0.0156, 0.0164)	0.0008

Table III. 95% Confidence Intervals of $R_{Q_n}^2$ For Cases Logistic (2, 0.2) Contaminants

n		Contaminant	Observed	BCa (95%)	Range
20	mean	5%	0.9635	(0.9627, 0.9642)	0.0015
		25%	0.9319	(0.9303, 0.9335)	0.0032
		50%	0.8704	(0.8681, 0.8725)	0.0044
	standard deviation	5%	0.0282	(0.0264, 0.0304)	0.0040
		25%	0.0592	(0.0567, 0.0619)	0.0052
		50%	0.0804	(0.0783, 0.0828)	0.0045
40	mean	5%	0.9810	(0.9807, 0.9814)	0.0007
		25%	0.9497	(0.9486, 0.9508)	0.0022
		50%	0.8687	(0.8670, 0.8704)	0.0034
	standard deviation	5%	0.0123	(0.0117, 0.0130)	0.0013
		25%	0.0390	(0.0375, 0.0406)	0.0031
		50%	0.0624	(0.0610, 0.0639)	0.0029
100	mean	5%	0.9923	(0.9921, 0.9924)	0.0003
		25%	0.9635	(0.9628, 0.9642)	0.0014
		50%	0.8597	(0.8585, 0.8609)	0.0024
	standard deviation	5%	0.0042	(0.0041, 0.0044)	0.0003
		25%	0.0237	(0.0229, 0.0246)	0.0017
		50%	0.0431	(0.0421, 0.0441)	0.0020

Table IV. 95% Confidence Intervals of R_{Qn}^2 For Cases Logistic (2, 1) Contaminants

n		Contamination	Observed	BCa (95%)	Range
20	mean	5%	0.9643	(0.9634, 0.9649)	0.0015
		25%	0.9639	(0.9630, 0.9646)	0.0016
		50%	0.9639	(0.9631, 0.9646)	0.0015
	standard deviation	5%	0.0268	(0.0252, 0.0284)	0.0032
		25%	0.0287	(0.0267, 0.0311)	0.0044
		50%	0.0268	(0.0255, 0.0284)	0.0029
40	mean	5%	0.9815	(0.9812, 0.9819)	0.0007
		25%	0.9813	(0.9809, 0.9816)	0.0007
		50%	0.9815	(0.9811, 0.9818)	0.0007
	standard deviation	5%	0.0122	(0.0115, 0.0129)	0.0014
		25%	0.0124	(0.0117, 0.0133)	0.0016
		50%	0.0115	(0.0110, 0.0121)	0.0011
100	mean	5%	0.9927	(0.9926, 0.9928)	0.0002
		25%	0.9924	(0.9923, 0.9925)	0.0002
		50%	0.9924	(0.9922, 0.9925)	0.0003
	standard deviation	5%	0.0041	(0.0039, 0.0043)	0.0004
		25%	0.0044	(0.0042, 0.0046)	0.0004
		50%	0.0043	(0.0042, 0.0045)	0.0003

Table V. 95% Confidence Intervals of R_{Qn}^2 For Cases Normal (3, 0.2) Contaminants

n		Contamination	Observed	BCa (95%)	Range
20	mean	5%	0.9625	(0.9617, 0.9632)	0.0015
		25%	0.9124	(0.9096, 0.9148)	0.0052
		50%	0.7337	(0.7297, 0.7378)	0.0081
	standard deviation	5%	0.0283	(0.0269, 0.0298)	0.0029
		25%	0.0920	(0.0881, 0.0961)	0.0080
		50%	0.1444	(0.1418, 0.1470)	0.0052
40	mean	5%	0.9810	(0.9806, 0.9813)	0.0007
		25%	0.9384	(0.9367, 0.9400)	0.0033
		50%	0.6863	(0.6830, 0.6896)	0.0066
	standard deviation	5%	0.0124	(0.0118, 0.0131)	0.0013
		25%	0.0606	(0.0576, 0.0637)	0.0061
		50%	0.1191	(0.1170, 0.1211)	0.0041
100	mean	5%	0.9921	(0.9919, 0.9922)	0.0003
		25%	0.9619	(0.9611, 0.9627)	0.0016
		50%	0.6320	(0.6299, 0.6342)	0.0043
	standard deviation	5%	0.0045	(0.0043, 0.0047)	0.0004
		25%	0.0291	(0.0278, 0.0308)	0.0030
		50%	0.0938	(0.0923, 0.0953)	0.0030

IV. CONCLUSIONS

The study confirms the robustness of the correlation measure based on the Qn estimator under contamination. It highlights the influence of contamination type, percentage, and sample size on distribution properties and confidence intervals. The results obviously show that the presence of logistic and normal contaminants significantly impacts the distribution and confidence intervals of the statistic. Logistic contaminants increase skewness and decrease the mean with higher contamination. Meanwhile, normal contaminants improve symmetry and reduce variability as contamination increases. Confidence intervals become narrower with larger sample sizes but are sensitive to contamination levels. The results obtained are also used to verify that the bounds of the correlation measure are always less than 1. Overall, the results

support the robustness of the estimator and the effectiveness of bootstrap methods in inference. Future research may explore further contamination scenarios and refine bootstrap techniques for improved accuracy.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this paper.

ACKNOWLEDGEMENT

We would like to thank the Ministry of Higher Education Malaysia (Grant no: GP-IPS/2025/9824100) for funding this study.

DECLARATION OF AI-ASSISTED TOOLS USAGE

AI-assisted tools were used solely for language editing to improve grammar, clarity, and readability. An AI-detection assessment reported a score below 15%. The authors have critically reviewed and verified all revisions and accept full responsibility for the accuracy, originality, integrity, and entire content of the manuscript.

REFERENCES

- [1] G. L. Shevlyakov and P. O. Smimov, "Robust estimation of the correlation coefficient: An attempt of survey," *Austrian Journal of Statistics*, vol. 40, no. 1–2, pp. 147–156, 2011.
- [2] G. L. Shevlyakov and P. O. Smimov, "Robust estimation of the correlation coefficient: An attempt of survey," *Austrian Journal of Statistics*, vol. 40, no. 1–2, pp. 147–156, 2011.
- [3] P. J. Rousseeuw and C. Croux, "Alternatives to the median absolute deviation," *Journal of the American Statistical Association*, vol. 88, no. 424, pp. 1273–1283, Dec. 1993.
- [4] L. Giraitis, Y. Li, and P. C. B. Phillips, "Robust inference on correlation under general heterogeneity," *Journal of Econometrics*, vol. 240, no. 1, Art. no. 105691, Mar. 2024.
- [5] J. Qiu, Z. Li, and J. Yao, "Robust estimation for number of factors in high-dimensional factor modeling via Spearman correlation matrix," *Journal of the American Statistical Association*, vol. 120, no. 550, pp. 1139–1151, 2025.
- [6] S. Jang and C. Choi, "FORCE: Fast outlier-robust correlation estimation via streaming quantile approximation for high-dimensional data streams," *Mathematics*, vol. 14, no. 1, Art. no. 191, Jan. 2026.
- [7] J. Koerts and A. P. J. Abrahamse, "The correlation coefficient in the general linear model," *European Economic Review*, vol. 1, pp. 401–427, 1970.
- [8] A. C. Rencher and F. C. Pun, "Inflation of R^2 in best subset regression," *Technometrics*, vol. 22, no. 1, pp. 49–53, 1980.
- [9] J. S. Cramer, "Mean and variance of R^2 in small and moderate samples," *Journal of Econometrics*, vol. 35, no. 2–3, pp. 253–266, 1987.
- [10] M. L. Carrodus and D. E. A. Giles, "The exact distribution of R^2 when regression disturbances are autocorrelated," *Economics Letters*, vol. 38, no. 4, pp. 375–380, 1992.
- [11] G. Meepagala, "The small sample properties of R^2 in a misspecified regression model with stochastic regressors," *Economics Letters*, vol. 40, no. 1, pp. 1–6, 1992.
- [12] K. Ohtani and H. Hasegawa, "On small sample properties of R^2 in a linear regression model with multivariate t errors and proxy variables," *Econometric Theory*, vol. 9, no. 4, pp. 504–515, 1993.
- [13] K. Ohtani, "The density function of R^2 and adjusted R^2 , and their risk performance under asymmetric loss in mis specified linear regression models," *Economic Modelling*, vol. 11, no. 4, pp. 463–471, 1994.
- [14] A. K. Srivastava and A. Ullah, "The coefficient of determination and its adjusted version in linear regression models," *Econometric Reviews*, vol. 14, no. 2, pp. 229–240, 1995.
- [15] K. Ohtani, "Bootstrapping R^2 and adjusted R^2 in regression analysis," *Economic Modelling*, vol. 17, no. 4, pp. 473–483, 2000.
- [16] I. S. Helland, "On the interpretation and use of R^2 in regression analysis," *Biometrics*, vol. 43, no. 1, pp. 61–69, 1987.
- [17] Q. Heng and K. Lange, "Bootstrap estimation of the proportion of outliers in robust regression," *Statistics and Computing*, vol. 35, no. 1, Art. no. 3, 2025.
- [18] M. Usman, M. Kashif, M. Z. Iqbal, and G. Farid, "Evaluating robust process capability indices for Weibull process in presence of outliers and their bootstrap confidence intervals," *Journal of Statistical Computation and Simulation*, vol. 95, no. 16, pp. 3497–3525, 2025.
- [19] H. Choi and K. Jung, "Impact of data distribution and bootstrap setting on anomaly detection using isolation forest in process quality control," *Entropy*, vol. 27, no. 7, Art. no. 761, 2025.
- [20] C. Croux, C. Flandre, and G. Haesbroeck, "The breakdown behavior of the maximum likelihood estimator in the logistic regression model," *Statistics & Probability Letters*, vol. 60, no. 4, pp. 377–386, 2002.
- [21] H. D. Bondell and B. J. Reich, "Simultaneous regression shrinkage, variable selection, and supervised clustering of predictors with OSCAR," *Biometrics*, vol. 64, no. 1, pp. 115–123, 2008.
- [22] P. J. Bickel and E. L. Lehmann, "Descriptive statistics for non-parametric models. IV. Spread," in *Contributions to Statistics: Jaroslav Hájek Memorial Volume*, J. Jurečková, Ed. Prague, Czech Republic: Academia, 1979, pp. 33–40.
- [23] N. A. Zakaria, S. Abdullah, and N. A. Ahad, "Robust correlation coefficient based on estimator," *AIP Conference Proceedings*, pp. 1–6, 2017.
- [24] F. Maarof, L. F. Peng, and N. A. Ibrahim, "Bootstrapping the confidence intervals for samples from contaminated standard logistic distribution," *Pertanika Journal of Science & Technology*, vol. 18, no. 1, pp. 209–221, 2010.
- [25] F. P. Lim, I. Mohamed, N. Daud, and S. L. Goh, "Comparison of outlier detection methods in standard 2×2 crossover design," *Sains Malaysiana*, vol. 45, no. 3, pp. 499–506, 2016.
- [26] F. P. Lim, I. Mohamed, A. I. N. Ibrahim, S. L. Goh, N. A. Mohamed, and A. Rahman, "Outlier detection in 2×2 crossover design using Bayesian framework," *Sains Malaysiana*, vol. 48, no. 4, pp. 893–899, 2019.
- [27] R. B. D'Agostino and M. A. Stephens, *Goodness-of-Fit Techniques*. New York, NY, USA: Marcel Dekker, 1986.
- [28] B. Efron, "Better bootstrap confidence intervals," *Journal of the American Statistical Association*, vol. 82, no. 397, pp. 171–200, 1987.