Mathematics and Statistics 11(4): 693-702, 2023
DOI: 10.13189/ms.2023.110411 http://www.hrpub.org

Hybrid Correlation Coefficient of Spearman with MM-Estimator

Siti Hajar binti Abu Bakar[1], Muhamad Safiih Bin Lola[1], Anton Abdulbasah Kamil[2,*], Nurul Hila Zainuddin[3], Mohd Tajuddin Abdullah[4]

[1]School of Informatics and Applied Mathematics, Universiti Malaysia Terengganu, Malaysia
[2]Faculty of Economics, Administrative and Social Sciences, Istanbul Gelisim University, Turkey
[3]Department of Mathematics, Faculty of Science and Mathematics, Universiti Pendidikan Sultan Idris, Malaysia
[4]Faculty of Fisheries and Food Science, Universiti Malaysia Terengganu, Malaysia

Received November 29, 2022; Revised July 3, 2023; Accepted July 19, 2023

Cite This Paper in the Following Citation Styles

(a): *[1] Siti Hajar binti Abu Bakar, Muhamad Safiih Bin Lola, Anton Abdulbasah Kamil, Nurul Hila Zainuddin, Mohd Tajuddin Abdullah , "Hybrid Correlation Coefficient of Spearman with MM-Estimator," Mathematics and Statistics, Vol. 11, No. 4, pp. 693 - 702, 2023. DOI: 10.13189/ms.2023.110411.*

(b): *Siti Hajar binti Abu Bakar, Muhamad Safiih Bin Lola, Anton Abdulbasah Kamil, Nurul Hila Zainuddin, Mohd Tajuddin Abdullah (2023). Hybrid Correlation Coefficient of Spearman with MM-Estimator. Mathematics and Statistics, 11(4), 693 - 702. DOI: 10.13189/ms.2023.110411.*

Copyright©2023 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract The Spearman rho nonparametric correlation coefficient is widely used to measure the strength and degree of association between two variables. However, outliers in the data can skew the results, leading to inaccurate results as the Spearman correlation coefficient is sensitive toward outliers. Thus, the robust approach is used to construct a robust model which is highly resistant to data contamination. The robustness of an estimator is measured by the breakdown point which is the smallest fraction of outliers in a sample data without affecting the estimator entirely. To overcome this problem, the aim of this study is two-fold. Firstly, researchers have proposed a robust Spearman correlation coefficient model based on the MM-estimator, called the MM-Spearman correlation coefficient. Secondly, to test the performance of the proposed model, it was tested by the Monte Carlo simulation and contaminated air pollution data in Kuala Terengganu, Terengganu, Malaysia. The data have been contaminated from 10% to 50% outliers. The performance of the MM-Spearman correlation coefficient properties was evaluated by statistical measurements such as standard error, mean squared error, root mean squared error and bias. The MM-Spearman correlation coefficient model outperformed the classical model, producing significantly smaller standard error, mean squared error, and root mean squared error values. The robustness of the model was evaluated using the breakdown point, which measures the smallest fraction of outliers that can be present in sample data without entirely affecting the estimator. The hybrid MM-Spearman correlation coefficient model demonstrated high robustness and efficiently handled data contamination up to 50%. However, the study has a limitation in that it can only overcome data contamination up to a maximum of 50%. Despite this limitation, the proposed model provides accurate and efficient results, enabling management authorities to make sound decisions without being affected by contaminated data. The MM-Spearman correlation coefficient model provides a valuable tool for researchers and decision-makers, allowing them to analyze data with a high degree of accuracy and robustness, even in the presence of outliers.

Keywords Coefficient Correlation Spearman, Outliers, MM-estimator, Hybrid Model, Robustness

MSC: 62G05, 62G35

1. Introduction

Non-parametric correlation coefficients are coefficients based on the rank of the observation where the distribution of functions under the observation is unknown. While the parametric correlation coefficient is the correlation

coefficient that measures the linear relationship between normally distributed variables [1].

Variables are said to be correlated with one variable if one variable is followed by another variable. However, in analyzing the relationships between variables, correlation coefficient models are often sensitive to the presence of outliers. Outliers are one or more of these rare data observations or significantly different in value from the majority of other data. According to Daszykowski [2], an estimator is considered not a robust estimator when outliers come to computing mean data. In fact, the use of too much data that involves some material between variables is a factor in controlling the presence of outliers. Materiality is a measure of the importance of information in making a decision.

The nature of the correlation coefficient is sensitive to outliers in incorrect calculations. The correlation coefficients sensitive to the presence of outliers can be improved by a robust statistical hybrid approach [3]. According to Abdullah [4], robust estimation is not easily influenced by the existence of the outliers. A robust approach is applied to produce estimators that are not easily affected by the presence of outliers. Therefore, in this study, the robustness approach will be mixed with the correlation coefficient and the study on the correlation coefficient based on the robust approach will be done by focusing on the Spearman correlation coefficient, r_s. The Pearson correlation coefficient is very efficient in the normal distribution but the efficiency statistic for Spearman correlation coefficients is over 70% for any probability of population correlation value [5]. In addition, the Spearman correlation coefficient can be used if the data is not normally distributed. In fact, the Spearman correlation coefficient is an alternative to the Pearson correlation coefficient calculated based on the rank of observation.

There have been previous studies examining Spearman correlation coefficients but studies on Spearman correlation coefficients have been found to be poorly conducted, especially with regard to the problem of controlling the existence of outliers. Most studies focus on Pearson correlation coefficients that can only be applied to normal scattering data. Given that the Spearman correlation coefficient was chosen as Pearson's correlation coefficient, the Spearman correlation coefficient was chosen to delve deeper into the effects of outliers on the correlation coefficient model that ignored the data distribution. However, Spearman's non-parametric correlation coefficient model is more sensitive to the presence of outliers than the Kendall correlation coefficient model [5]. Therefore, the study of a robust Spearman correlation coefficient model based on the MM estimator is novel because there is no study conducted using the Spearman correlation coefficient and a robust approach using the MM estimator in the literature. According to Maronna [3], the impact of outliers can be significant and in order to address the problem an approach is needed. Such a correlation coefficient model that is sensitive to data contamination should be addressed, so that the correlation coefficient model is more robust to the presence of outliers and has a high impact value especially in limited data situations and contaminated data.

The results of this study are expected to overcome the problem of the sensitivity of the model to the presence of isolates that exist. This is crucial because the accuracy of the correlation coefficient value is affected by the existence of outliers. The robust approach model especially of MM-estimator can not only give good value in the situation of the amount of contaminated data that is small but also in the situation of the existence of the amount of contaminated data approaching 50% of data contamination. Therefore, the method developed in this study is likened to an efficient approach because it is able to produce a hybrid model that is able to curb the problem of the presence of outliers that invite less accurate estimates.

2. Methodology

2.1. M-Estimation

The discussion of correlation coefficients highlights the sensitivity of the correlation coefficient model to the outliers present. Modeling is robust when insensitive to data errors. The correlation coefficient model cannot handle the influence of outliers although only one present can influence the findings of a prediction. Thus, a robust estimator-MM approach consisting of the Huber-M estimator [6] was applied to obtain the robust estimation value against the outliers.

Table 1. Huber-M weights estimator

Estimator Type	$\rho(u)$	$\psi(u)$	$w(u)$	Range for u
Huber	$\frac{u^2}{2}$	u	1	$\lvert u\rvert \le c$
	$c\lvert u\rvert - \frac{c^2}{2}$	$c\ \ sign(u)$	$c/\ \lvert u\rvert$	$\lvert u\rvert > c$

According to Table 1, the Huber-M estimator is defined as the estimator that minimizes the objective function, (ρ). Estimates can be made against M-estimators representing the objective function. Whereas c is known as the tuning constant and u is the residual, smaller values of c produce more resistance to outliers, but at the expense of lower efficiency when the errors are normally distributed. The tuning constant is generally picked to give reasonably high efficiency in the normal case; c = 1.345σ for the Huber and c = 4.685σ for the bi-square (where σ is the standard deviation of the errors) produce 95-percent efficiency when the errors are normal, and still offer protection against outliers [4].

Objective function, $\rho(u)$

$$\rho(u) = \begin{cases} u^2/2 & if \ |u| \le c \\ c|u| - c^2/2 & if \ |u| > c \end{cases} \quad (1)$$

Differences are made in objective functions to produce influence functions.

Influence function, $\Psi(u)$

$$\Psi(u) = \begin{cases} u & if \ |u| \le c \\ c \ sign \ u & if \ |u| > c \end{cases} \quad (2)$$

The function of influence and also known as the bi-square-psi function of (u) is the result of the difference in the objective function.

Weight Function, w (x)

The weighting function is generated by dividing the influence function by the value of u.

$$w(u) = \begin{cases} 1 & if \ |u| \le c \\ c \ / \ |u| & if \ |u| > c \end{cases} \quad (3)$$

In general, to obtain a weighting function, several steps need to be done by differentiating the objective function, $\rho(u)$ which is equation (1) which produces the influence function, $\Psi(u)$. Subsequently, the influence function is divided by the value of u to obtain equation (2) which is the weighting function. The weighting function is to be included in the correlation coefficient equation.

For example, if $|u| \le c$, c = 1.345, u = data value.

First step: ρ(u)= $u^2/2$

Second step : Make a differentiation against $\rho(u)$.

$$\rho'(u)=\Psi(u)$$

$$\frac{d}{du}\left(\frac{u^2}{2}\right) = u$$

Third step: After getting the value of the influence function, $\Psi(u)$, the value is divided by the value of u.

$$\frac{\Psi(u)}{u} = \ w(u)$$

$$\frac{(u)}{u} = \ 1$$

$$w(u) = \ 1 \quad (4)$$

Based on the steps shown above and the final value of $w(u)$, (3) it is shown that when the value of u is smaller or equal to the tuning constant, c = 1.345 will produce a weight value of 1.

2.2. MM-Estimation

Yohai [7] has introduced a robust estimator-MM approach. This approach combines the estimator of failure values with the M-estimator to produce a robust estimator of the M-estimator [8, 9]. The M-estimator is a simple, theoretical approach to the calculation but cannot contain the leverage point. "MM" refers to an approach that uses more than one M-estimator to determine the final estimation value with a combination of a high impact criterion 50% and high efficiency approaching 95% [8]. Therefore, the robust estimator-MM approach is used because it has adjective failure from high statistical accuracy and is more capable of suppressing the presence of the estimator than the M-estimation approach.

Given a sample of size n, $\boldsymbol{u} = u_i, u_2, u_3 \ldots . u_n)$ the estimator scale $s(\boldsymbol{u})$ is the value of s which is the solution for

$$\left(\frac{1}{n}\right)\sum_{i=1}^{n} \rho\left(\frac{u_i}{s}\right) = b \quad (5)$$

and n may be defined by $E_{\emptyset}(\rho(u) = b$ where the symbol $\emptyset$ represents the standard normal distribution.

It's easy to show if

$$c\ (\mathbf{u}) = \#\ \{i: 1 \le i \le n, u_i = 0\}\ /n < 1 - (b/a) \quad (6)$$

Then (6) has a unique solution and this solution is not 0. If $c(\boldsymbol{u}) \ge 1 - (b/a)$, then $s\ (\boldsymbol{u}) = 0$.

According to Yohai [7], MM-estimators can be determined using three steps:

(i) Select one of the estimators $T_{0,n}$ of θ_0 that have a high leakage point of 0.5.

(ii) Price is calculated by the formula

$$r_i(T_{0,n}) = y_i - T_{0,n}x_i \ , 1 \le i \le n \quad (7)$$

calculate the M-scale $s_n = s(r(T_{0,n})$ using the function ρ_0 that satisfies the assumptions in (9) and which constant b

$$\frac{b}{a} = 0.5 \quad (8)$$

$a = max \ \rho_0(u)$. As Huber [10] has proved, (8) shows that the M-estimator scale has an influx of 0.5 points.

ρ is a function that satisfies the following assumptions:

(1) $\rho(0) = 0$;
(2) $\rho(-u) = u$;
(3) $0 \le u \le v$ where $\rho(u) \le \rho(v)$;
(4) ρ is a continuous function;
(5) Let $a = \sup \rho(u)$, then $0 < a < \infty$;
(6) If $\rho(u) < a$ and $0 \le u < v$, then

$$\rho(u) \le \rho(v) \quad (9)$$

(iii) Let ρ_1 be another function that satisfies equation (8)

$$\rho(u) \le \rho(v) \quad (10)$$

$$\sup \ \rho_1(u) \le \rho_0(u) = 0 \quad (11)$$

Let $\psi_1 = \rho_1$. Thus the estimator-MM $T_{1,n}$ is defined as the solution

$$\sum_{i=1}^{n} \psi_1\left(\frac{r_1(\theta)}{s_n}\right) x_i = 0 \quad (12)$$

that fulfills the assumption

$$S(T_{1,n}) \le S(T_{0,n}) \quad (13)$$

Based on (13) shows the error value of the estimation using MM-estimator is smaller than the error value of the

estimation using the M-estimator. This proves that the MM-Spearman correlation coefficient approach is better and resolves errors by producing small values.

2.3. The MM-Spearman Correlation Coefficient

In this study, the Spearman correlation coefficient model is integrated with a robust approach. The hybrid results of the correlation coefficient model and the method form a new model, namely the MM-Spearman correlation coefficient. Calculations for the robust Spearman correlation coefficient-MM estimation model are based on sample rank $\bar{R}(x)$ and $\bar{R}(y)$ that sensitive to outliers. The 10% to 50% outliers that exists in the sample data are to test the model's robustness to the outliers. The formula (14) shows the formula of the Spearman correlation coefficient model with no robust approach whereas (15) is the robust approximation of the correlation coefficient formula which incorporates the weights in the original formula of the Spearman correlation coefficient.

$$r_s = \frac{\sum_i (R(x_i) - \bar{R}(x))\big(R(y_i) - \bar{R}(y)\big)}{\sqrt{\sum_i (R(x_i) - \bar{R}(x))^2 \sum_i (R(y_i) - \overline{R}(y))^2}}$$

Where:
$R(x_i)$ is the rank of observation x_i
$R(y_i)$ is the rank of observation y_i
$\bar{R}(x)$ is the rank of observation $x = \frac{\sum R(x_i)}{n}$
$\bar{R}(y)$ is the rank of observation $y = \frac{\sum R(y_i)}{n}$

The formula of the MM-Spearman correlation coefficient is as follow:

$$r_{smm} = \frac{\sum w_i (R(x_i) - \bar{R}(x))(R(y_i) - \bar{R}(y))}{\sqrt{\sum w_i (R(x_i) - \bar{R}(x))^2 \sum w_i (R(y_i) - \overline{R}(y))^2}} \quad (15)$$

$$w(u) = \begin{cases} 1 & jika \quad |u| \le c \\ c \ / \ |u| & jika \quad |u| > c \end{cases}$$

$R(x_i)$ is the rank of observation x_i
$R(y_i)$ is the rank of observation y_i
$\bar{R}(x)$ is the rank of observation $x = \frac{\sum w_i R(x_i)}{w_i}$
$\bar{R}(y)$ is the rank of observation $y = \frac{\sum w_i R(y_i)}{w_i}$
u is data
c is a constant

2.4. Determination of Stability Achievement

The relevance of forecasting methods to outliers can be identified by monitoring the impact of outliers on both the modeled and non-hybridized models using a robust approach. The impact of outlier influence is known based on error size. This section describes the method of determining achievement for the solid correlation coefficient model developed. Determination of firm performance is an important feature of testing correlation coefficients. The robustness of the Spearman correlation coefficient model and the MM-Spearman correlation coefficient model were compared in terms of estimation of the points involving the following criteria:

i. $Bias = \bar{\rho} - \rho$ (16)

where $\bar{\rho} =$ sample mean of correlation coefficient and ρ = correlation coefficient value.

ii. *Standard Error, (SE)*

$$SE = \sqrt{V(\hat{\rho})} \quad (17)$$

Where:

$$V(\hat{\rho}) = L^{-1} \sum(\hat{\rho}_1 - \bar{\rho}) \quad (18)$$

$\hat{\rho}_1$= expectation value of correlation coefficient
$\bar{\rho}$ = sample mean of correlation coefficient
L^{-1} = sample size

iii. *Mean Square Error, (MSE)*

$$MSE = [bias]^2 + V(\hat{\rho}) \quad (19)$$

Further study is followed by developing the simulation model. The developed method is transferred in the R-programming language to continue the study with the Monte Carlo simulation method.

3. Results and Interpretation

The Monte Carlo simulation produces the results and the probability that any choice will be made. This technique is used to determine the risk impact identified by performing repeated simulations and identifying various possible outcomes in different scenarios [11]. In this study, a Monte Carlo simulation study was performed to test the MM-Spearman correlation coefficient model developed to meet the best estimation criteria. The factors that determine the best estimation are unbiased and high efficiency in estimating the error.

In this study, data generation using Monte Carlo simulation with software R, we assume that data follow the Normal distribution [12-14]. There are several criteria set for generating data distribution. Data for normal distribution is with mean value equal to zero, variance is equal to one and sample size is n = 100. Simulation of 100 of data generated by linear relation equation $y_i = 2.0 + 1.0x_i + u_i$ [15].

Table 2. Monte Carlo Simulation of Correlation Coefficient Model

Contaminated Data	*rs*	*rsmm*
0%	0.984	0.990
10% **5%**	0.469	0.988
20%	0.088	0.985
30%	-0.241	0.984
40%	-0.307	0.976
50%	-0.518	-0.591

Table 2 shows the results of the Spearman correlation

coefficient and the MM-Spearman correlation coefficient estimation by applying Monte Carlo simulation data. The comparisons between these two correlation coefficients can be clearly seen by setting the sum of the percentages for each percentage comparison. Based on the table above the correlation coefficient values for non-contaminated data are 0.98430 for Spearman and 0.99045 for the MM-Spearman correlation coefficient. The MM-Spearman correlation coefficient gives a better correlation coefficient than the Spearman correlation coefficient for non-contaminated data.

Subsequently the Spearman correlation coefficient was based on the MM-Spearman correlation coefficient and the Spearman correlation coefficient was tested with the presence of 10% contaminated data with the presence of 10 defective data. The Spearman correlation coefficient value is 0.46921 compared to the 0.98801 coefficient-based correlation coefficient MM. The contamination level of 10% showed a significant decrease in Spearman correlation coefficient with a 47.67% decrease compared to values when no contaminated data were present. The same situation applies to the Spearman correlation coefficient based on MM-estimation which is a decrease but only a slight change with a 1% decrease compared to values on the good data. This shows that the Spearman coefficient model based on the MM amplifier is stronger and better than the Spearman correlation coefficient value. The Spearman correlation coefficient is not able to withstand the influence of the presence of 10 contaminated data because the model produces a relatively small correlation coefficient compared to no high-correlated contaminated values.

The next level of is 20% of the contaminated data. The Spearman correlation coefficient with 20% of the contaminated data was 0.08782 with an 81.28% decrease compared to the value at the 10% level of contamination. The decrease shows a Spearman correlation coefficient of moderate correlation at the 10% level of contamination and then down to the low-correlation category at the 20% level of contamination. The MM-Spearman correlation coefficient decreased slightly by only 1.00% compared to the value at the 10% pollution level of 0.98502. At the level of 20% data pollution clearly reflects the Spearman correlation coefficient based on MM-estimator more robust and consistent than the original model.

Subsequently for the level of 30%, 40% and 50%, the results of the Spearman correlation coefficient and the Spearman-MM estimator coefficient in Table 2 are plotted in Figure 1.

Figure 1 shows the plot of the influence of the Spearman correlation coefficient and the MM-Spearman correlation coefficient for 100 simulation data generated. The figure shows that the correlation coefficient values for the Spearman model decreased to a negative value of 30% of the contaminated data while the hybrid value of the solid coefficient model decreased to a negative value of 47% of the contaminated data. The Spearman correlation coefficient based on MM-estimation looks strong despite contaminated data increasing by up to 47%. The Spearman correlation model based on MM-estimator is robust due to the existence of the MM estimator — the weighting of the data values that are contaminated in the model. Generally, it can be concluded that both correlation coefficient models are undervalued but the degree of ability to counteract the presence of contaminated data is different.

However, there are several aspects of comparisons in further studies of the properties of the impulse namely Bias, standard error (SE), mean squared error (MSE) and root mean square error (RMSE). Each of the following aspects was identified 500 times with each sample size of 30, 50, and 100. According to Mundform [16] still has no empirical proposal on the number of trials that need to be made to ensure accurate results, but most studies begin with less than 1000 trials. The results for the Spearman correlation coefficient model and the MM-Spearman correlation coefficient used point estimates by [17] as in Table 3, Table 4, and Table 5. Based on the SE, MSE and RMSE error estimation tables, five decimal points are used because the error values are too small and do not fit three decimal points.

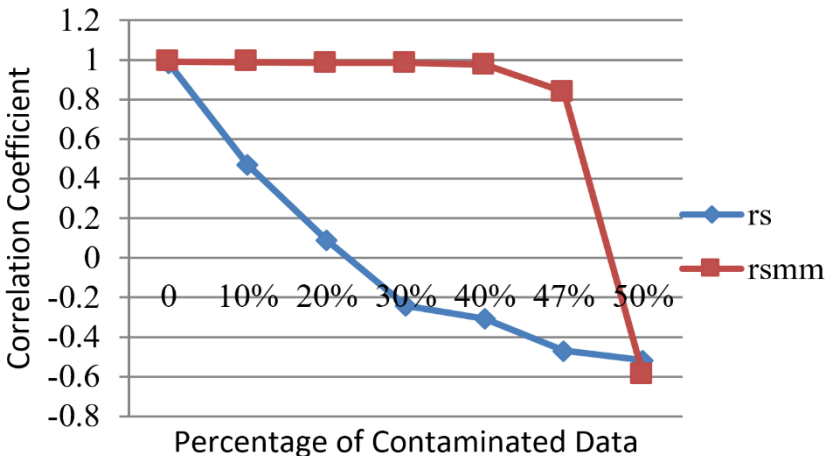


Figure 1. The plot influence of r_s and r_{smm} on the simulated data pollution

Table 3. Bias values r_s and r_{smm} for n=30, n=50 and n=100

Contaminated Data	Model	Bias value		
		n=30	n=50	n=100
0%	r_s	-0.03111	-0.02760	-0.02436
	r_{smm}	-0.01746	-0.01622	-0.01494
10%	r_s	-0.49436	-0.49104	-0.48893
	r_{smm}	-0.02344	-0.02102	-0.01857
20%	r_s	-0.86962	-0.86406	-0.86301
	r_{smm}	-0.03029	-0.02651	-0.02334
30%	r_s	-1.15266	-1.14879	-1.14741
	r_{smm}	-0.03795	-0.03277	-0.02837
40%	r_s	-1.35858	-1.35212	-1.34924
	r_{smm}	-0.06593	-0.04087	-0.03359
50%	r_s	-1.47145	-1.46931	-1.46854
	r_{smm}	-1.53619	-1.53312	-1.52380

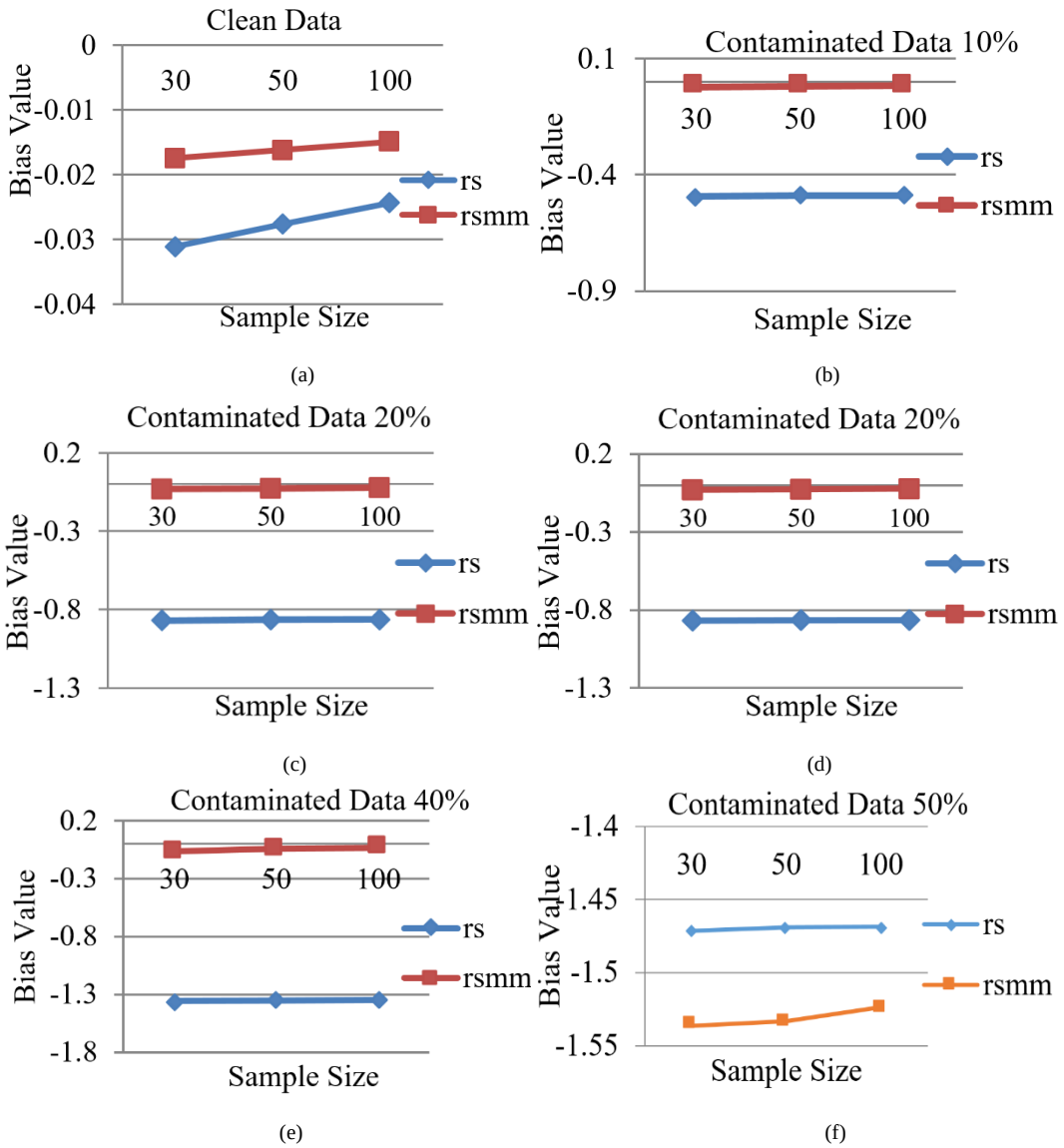


Figure 2. Achievement Plot of Bias values for r_s and $r_{smm,}$ Models

Table 4. Standard Error value r_s and r_{smm} for n=30, n=50 and n=100

Contaminated Data	Model	*SE*		
		n=30	n=50	n=100
0%	r_s	0.01503	0.00969	0.00594
	r_{smm}	0.01057	0.00708	0.00432
10%	r_s	0.06695	0.05403	0.03859
	r_{smm}	0.01276	0.00873	0.00516
20%	r_s	0.08759	0.06933	0.04804
	r_{smm}	0.01620	0.01027	0.00617
30%	r_s	0.09717	0.07467	0.05260
	r_{smm}	0.02028	0.01212	0.00714
40%	r_s	0.10169	0.07677	0.05322
	r_{smm}	0.09443	0.02233	0.00959
50%	r_s	0.10125	0.07802	0.05686
	r_{smm}	0.12115	0.08962	0.06601

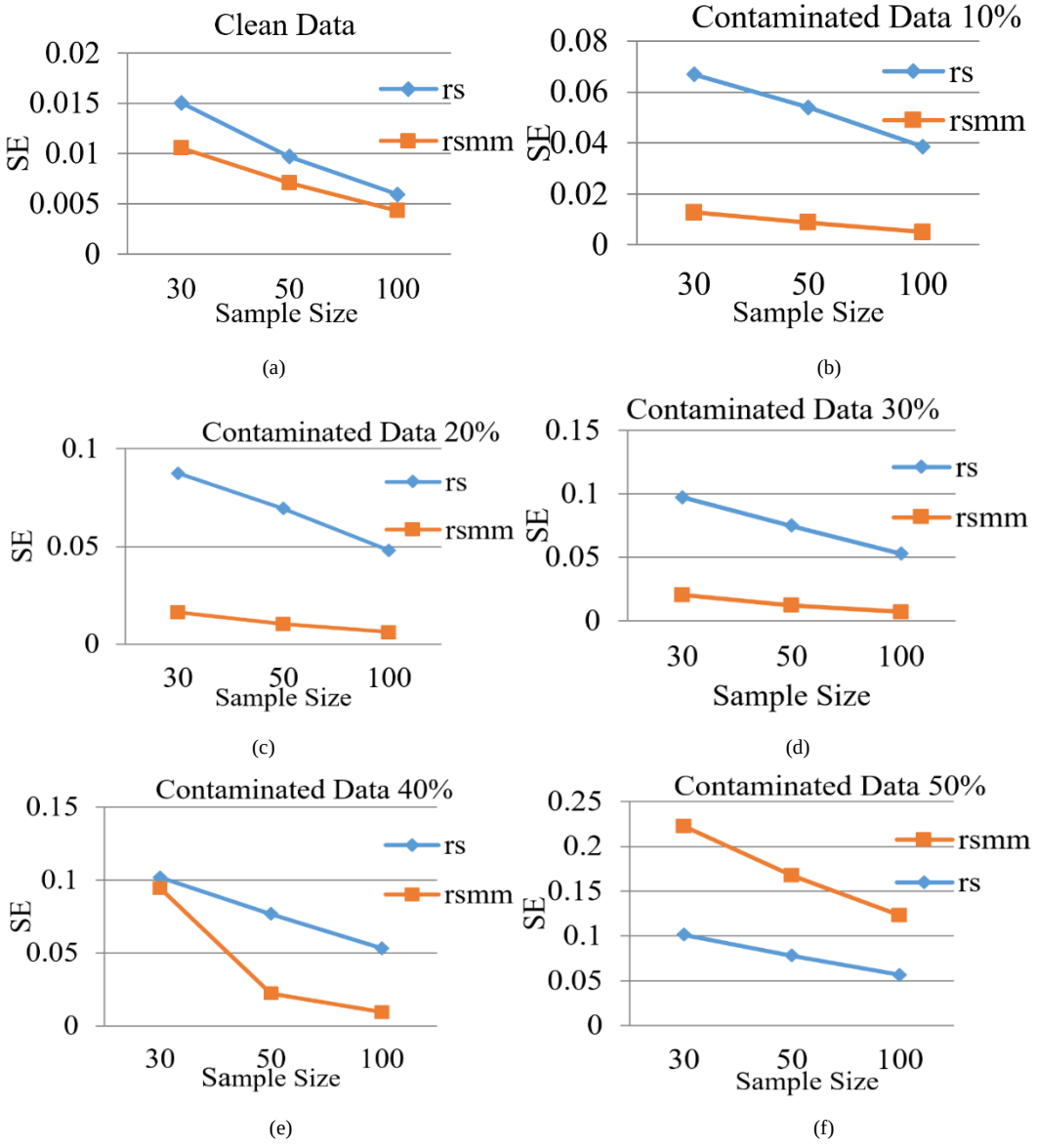


Figure 3. Achievement Plot of Error SE for r_s and $r_{smm.}$ Models

Table 5. *MSE* Values r_s and r_{smm} for n=30, n=50 and n=100

Contaminated Data	Model	*MSE*		
		n=30	n=50	n=100
0%	r_s	0.00120	0.00085	0.00062
	r_{smm}	0.00040	0.00031	0.00024
10%	r_s	0.24890	0.24437	0.24054
	r_{smm}	0.00070	0.00050	0.00037
20%	r_s	0.76390	0.75140	0.74709
	r_{smm}	0.00120	0.00080	0.00855
30%	r_s	1.33369	1.32530	1.31932
	r_{smm}	0.00190	0.00122	0.00085
40%	r_s	1.83380	1.82950	1.82307
	r_{smm}	0.01306	0.00232	0.00121
50%	r_s	2.17544	2.16496	2.15962
	r_{smm}	2.36423	2.35851	2.33659

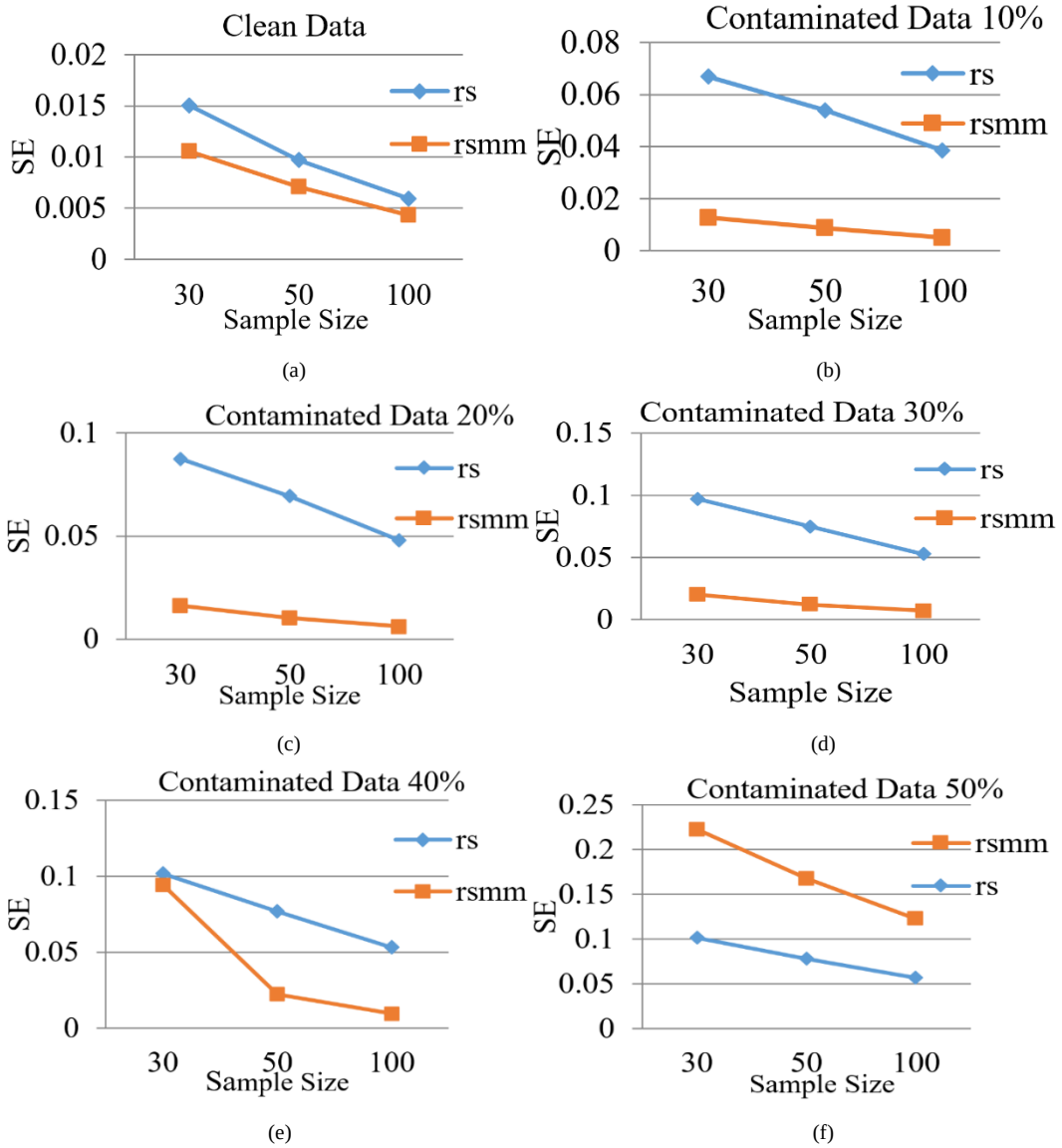


Figure 4. Achievement Plot of MSE for r_s and r_{smm} Models

Based on Figure 2 the performance of the bias estimation method of the estimation of the bias value is clearly shown that the bias value is reduced after the Spearman correlation coefficient model is integrated with a robust approach of up to 40% data pollution. In addition, the correlation coefficient bias value decreases as the number of samples increases. The next Figure 3 plot of the SE estimation method achieves a robust approach that yields good estimation errors. This is evidenced by the low error value of the Spearman correlation coefficient of a robust approach compared to the Spearman correlation coefficient of up to 40% of the contaminated data. In fact, the SE value decreases as the volume of data samples increases. The next error estimation method used to analyze the Spearman correlation coefficient and the MM-based Spearman correlation coefficient is the mean squared error (MSE) as shown in Table 5.

Table 5 is a plot of the performance of the MSE estimation method that shows the better MSE values for the robust Spearman correlation coefficients. This is evidenced by the MSE values for the low-robust Spearman correlation coefficients of up to 40% of the contaminated data and the error values decreasing with increasing amounts of data. The correlation coefficients were then further studied using the RMSE estimation method. Referring to Figure 4, the achievement plot of the RMSE estimation method shows smaller error values for the hybrid model coefficient than the original Spearman correlation coefficient model up to 40% contaminated data level. Even errors in the Spearman correlation coefficient model appear to decrease as data volume increases.

Overall, numerical estimation results for simulation data show that the r_{smm} robust approximation method provides better performance. This is evidenced by the small estimates of error and small errors for the MM-based correlation coefficients compared to the Spearman correlation coefficients. With reference to the n = 30 data sample, in the absence of contamination of the estimated damage value data, SE, MSE and RMSE for r_{smm} were -0.01746, 0.01057, 0.000416 and 0.020414. While the r_s results show lighter values, SE, MSE and RMSE are -0.03111, 0.01503, 0.00119 and 0.03455. Subsequent to 40% contamination of the estimated damage value data, SE, MSE and RMSE for r_{smm} = 0.06437, 0.09443, 0.01306, 0.11428 versus = -1.3503, 0.1017, 1.83377, 1.35416. Bias estimates, SE, MSE and RMSE at 40% for sample contaminated data n = 50, r_{smm} = -0.04264, 0.02794, 0.00262, 0.05116 compared to r_s = -1.34913, 0.07670, 1.82604 and 1.35131. Whereas n = 100 data sample results show smaller error values than n = 30 and n = 50 data samples. The estimated value of sample n = 100 with 40% contamination for bias value = -0.03359 versus value r_s = -1.24940, SE r_{smm} = 0.00959 versus SE r_s = 0.05322. MSE value r_{smm} = 1.82307 versus 0.01210 for r_s. RMSE value r_{smm} = 0.02081 versus RMSE r_s = 0.84931. The results of the comparison criterion show that the MM-Spearman correlation coefficient estimates a better value than the Spearman correlation coefficient for sample size 100 compared to data samples containing 30 and 50 sample data. This indicates the effect of sample size on the sample size because the larger the sample size the less influence the sample size and vice versa.

4. Conclusions

A simulation study of the MM-Spearman correlation coefficient model showed the best performance. This is because the robust approach to estimating MM has minimized the existence of errors by making the model less sensitive to the presence of contaminated data. These results demonstrate the robust nature of the hybrid model of a robust approach that can deliver good results for up to 50% of contaminated data. However, in order to further improve the developing hybrid models, additional research should consider some robust approaches, for example Tukey's Biweight function with having redescending psi function. Also, can use a robust multiple correlation coefficient and include (square roots of) for estimators in simulation study as competitors of estimators.

REFERENCES

[1] Renaud, O., & Victoria-Feser, M. P. "A robust coefficient of determination for regression". Journal of Statistical Planning and Inference, vol. 140, no. 7, pp. 1852-1862, 2010, https://doi.org/10.1016/j.jspi.2010.01.008.

[2] Daszykowski, M., Kaczmarek, K., Heyden, Y. V., Walczak, B. "Robust Statistics in Data Analysis: A Review Basic Concepts". Journal Chemometrics and Intelligent Laboratory Systems, vol. 85, no. 2, pp. 203–219, 2007, https://doi.org/10.1016/j.chemolab.2006.06.016

[3] Maronna, R. A., Martin, R. D., Yohai, V. J., Salibián-Barrera, M. Robust statistics: theory and methods (with R). John Wiley & Sons. 2019, pp. 17-85.

[4] Abdullah, M., Analisis Regresi (1st ed.). Kuala Lumpur: Perpustakaan Dewan Bahasa dan Pustaka. 1994, pp. 95-115.

[5] Croux, C., Dehon, C. "Influence Functions of the Spearman and Kendall Correlation Measures". Statistical Method & Applications, vol. 19, pp. 497-515, 2010, https://doi.org/10.1007/s10260-010-0142-z

[6] Huber, P. J. "Robust Estimation of Location Parameter". The Annals of Mathematical Statistics, vol. 35, No.1, pp.73-101, 1964. https://www.jstor.org/stable/2238020

[7] Yohai, V. J. High Breakdown-Point and High Efficiency Robust Estimates for Regression. The Annals of Statistics, 15, pp. 642-656, 1987. https://www.jstor.org/stable/2241331.

[8] Maru, A. P., Modi, C. K., & Nataraj, P. S. V. "Comparison of Robust MM Estimator and Robust M Estimator Based Denoising Filters for Gray Level Image Denoising," 2012

International Conference on Communication Systems and Network Technologies, Rajkot, Gujarat, India, 2012, pp. 109-113, doi: 10.1109/CSNT.2012.33.

[9] Huber, P. J. "Robust Regression: Asymptotics, Conjectures and Monte Carlo". The Annals of Statistics, vol. 1, No. 5, pp. 799-821. 1973. https://www.jstor.org/stable/2958283

[10] Huber, P. J. Robust Statistics (1st ed.). New York: Wiley Interscience Publication. 1981, pp. 199-237

[11] Lola, M. S., " Fuzzy Parametric Sample Selection Model: Monte Carlo Simulation Approach". Journal of Statistical Computation and Simulation. vol. 83, no. 6, pp. 992-1006, 2013. https://doi.org/10.1080/00949655.2011.646277

[12] Pascual, L., Romo, J., Ruiz, E. "Bootsrap Prediction for Returns and Volatile in GARCH Model". Computational Statistics and Data Analysis. vol. 50, no. 9, pp. 2293-2312, 2006. DOI: 10.1016/j.csda.2004.12.008.

[13] Zainuddin, N. H., Lola, M.S., and Nur Shazrahanim, K. "Modelling Moving Centerline Exponentially Weighted Moving Average (MCEWMA) with bootstrap approach: Case study on sukuk musyarakah of Rantau Abang Capital Berhad, Malaysia". International Journal of Applied, Business and Economic Research. vol. 14, no. 2, pp. 621-638, 2016.

[14] Zainuddin, N. H., Lola, M. S. "The Performance of BB-MCEWMA Model: Case Study on Sukuk Rantau Abang Capital Berhad, Malaysia". International Journal of Applied, Business and Economic Research. vol. 14, no. 2, pp. 63-77, 2016.

[15] Abdullah, M. "On a Robust Correlation Coefficient". Journal of the Royal Statistical Society, Series D (The Statistician). vol. 39, no. 4, pp. 455-460. 1990. https://doi.org/10.2307/2349088

[16] Mundform, D. J., Schaffer, Jay, K. M., Jin, S. D., Thongteeraparp, A., Supawan, P. "Number of Replications Required in Monte Carlo Simulation Studies: A Synthesis of Four Studies". Journal of Modern Applied Statistical Methods. vol. 10, no. 1, pp.19-28, 2011. 10.22237/jmasm/1304222580

[17] Razali, M.R., Lola, M. S., Abd Wahid, M. E.,, Zainuddin, N. H., Abdullah, M. T., K Abdul Hamid, A. A., Chandra Segaran, T., Lassaw, S., and Djauhari, M. "A Hybrid Logistic Regression Model with a Bootstrap Approach to Improve the Accuracy of the Performance of Jellyfish Collagen Data". Journal of Sustainability Science and Management. vol. 16, No. 6, pp. 191-203, 2021.