Eksakta : Berkala Ilmiah Bidang MIPA
VOLUME 25 NO 02 2024, pp 219-230
ISSN : Print 1411-3724 — Online 2549-7464
DOI : https://doi.org/10.24036/eksakta/vol25-iss02/515

Eksakta
Berkala Ilmiah Bidang MIPA
http://www.eksakta.ppj.unp.ac.id/index.php/eksakta

Article

Robust Linear Discriminant Analysis with Modified One-Step M-Estimator Qn Scale for Classifying Financial Distress in Banks: Case Study

Nabila Putri[1*], Kankan Parmikanti[1], Nurul Gusriani[1]

[1]Department of Mathematics, Faculty of Mathematics and Natural Science (FMIPA), Universitas Padjadjaran, Bandung, Indonesia

Article Info

Article history :

Received May 16, 2024
Revised May 30, 2024
Accepted June 04,2024
Published June 30, 2024
(In Press)

Keywords :
Robust linear discriminant analysis, nodified one-step M-estimator, Qn estimator

Abstract. The COVID-19 pandemic has significantly disrupted the banking sector, leading to a decline in profit growth as an indicator of financial distress. Bank financial health can be evaluated using the RGEC (Risk Profile, Good Corporate Governance, Earnings, Capital) analysis. While Linear Discriminant Analysis (LDA) ideally requires normality and homogeneity of covariance matrices, financial data often fail to meet these assumptions. Therefore, this study employs robust linear discriminant analysis using the Modified One-Step M-Estimator with Qn scale estimator (MOM-Qn) to classify 'distress' and 'non-distress' bank conditions. Given these challenges, this study acts as a preventive measure for banks to evaluate financial health simultaneously. The objective is to provide a robust discriminant function for more accurate and stable classification, particularly in the presence of outliers. It focuses on conventional private banks listed on the Indonesia Stock Exchange (IDX) during December 2021-2022. The results show a classification accuracy of 69.23% and a Press's Q value of 11.53846, indicating the method's effectiveness in classifying real financial data.

Corresponding Author :
Nabila Putri
Department of Mathematics, Faculty of Mathematics and Natural Science (FMIPA),
Universitas Padjadjaran, Bandung, Indonesia
Email : nabila20014@mail.unpad.ac.id

1. Introduction

In 2020, the emergence of the COVID-19 pandemic in Indonesia has significantly impacted one of the country's vital sectors, the economy [1]. The impact resulted in an economic rate slowdown and an increase in non-performing loans. The growing mistrust in the financial system due to bad loans, potentially hindered credit distribution during the pandemic [2]. Consequently, the financial performance of Indonesia's banking sector deteriorated. Prolonged declines in financial performance can lead to financial distress, where company's revenue drop disproportionately to its obligations [3]. This condition can serve as a warning sign of potential bankruptcy. Evaluating a bank's financial condition can be conducted through financial rasio analysis to identify indications of financial distress. According to Bank Indonesia Regulation Number 13/1/PBI/2011, banks are required to assess financial health based on the Risk-based Bank Rating (RBBR), encompassing factors such as risk profile, good corporate governance, earnings, and capital (RGEC) [4]. Financial rasio analysis results reveal the financial health status of each bank, categorizing them as either 'distress' or 'non-distress' to maintain banking sector stability of the banking sector.

Multivariate analysis can be employed to classify the financial condition banks simultaneously by examining the relationships between several independent variables and one or more dependent variables [5]. Common multivariate analyses used in classification include linear discriminant analysis (LDA), logistic regression, and machine learning algorithms like K-NN, decision trees, and random forests [6]. These methods are also employed in predicting financial distress or bankruptcy classifications [7-8]. However, LDA is highly sensitive to outliers, which can distort classification accuracy since the classical mean vector and covariance matrix are unreliable in the presence of outliers [9-11]. Therefore, robust estimators are needed to mitigate the impact of outliers, ensuring optimal classification performance.

Yahaya et al. (2016) introduced the use of the robust Automatic Trimmed Mean estimator, also known as the Modified One-step M-estimator (MOM), in discriminant analysis to address the effects of outliers that violated normality and covariance matrix homogeneity assumptions. Additionally, Melik et al. (2018) advanced the use of the Modified One-step M-estimator by integrating the robust scale estimator Qn (MOM-Qn) in discriminant analysis. Both studies demonstrated that the robust MOM and MOM-Qn estimators yielded better results compared to their counterparts [12-13].

This research employs robust linear discriminant analysis using the Modified One-step M-estimator with Qn scale (MOM-Qn) method. The objective is to develop a robust MOM-Qn linear discriminant function with optimal classification accuracy and precision for classifying financial distress among conventional private banks listed on the Indonesia Stock Exchage (IDX) during 2021-2022. This robust method offers an alternative approach for assessing and predicting banks financial health.

2. Materials and Methods

2.1. Materials

The data utilized in this research consists of secondary data related to the financial reports of conventional private banks listed on the Indonesia Stock Exchange (IDX) during the years 2021-2022. This data was sourced from official websites of the Indonesia Stock Exchange (IDX) and *Otoritas Jasa Keuangan* (OJK). The dependent variable is the indication of financial distress, based on each bank's net profit growth. A bank is indicated as distressed if it has positive net profit growth, whereas a bank with negative net profit growth is indicated as non-distressed. The independent variables used in this research are financial ratios based on the RGEC factors. However, this study focuses only on the factors of risk profile, earnings, and capital. Additional details regarding the data used can be found in Table 1.

Table 1. Data Description

Variable	Description	Factor	Denomination
Y	Indication of Financial Distress (Net Profit Growth)		Percentage (%)
X_1	Non-Performing Loan (NPL)	Risk Profile	Percentage (%)
X_2	Loan to Deposit Rasio (LDR)	Risk Profile	Percentage (%)
X_3	Net Interest Margin (NIM)	Earnings	Percentage (%)
X_4	*Biaya Operasional terhadap Pendapatan Operasional* (BOPO)	Earnings	Percentage (%)
X_5	Capital Adequacy Ratio (CAR)	Capital	Percentage (%)

2.2. Linear Discriminant Analysis

Discriminant analysis is a multivariate statistical method used to determine the dependency relationship between a categorical dependent variable and several numerical independent variables, which is employed to classify an observation into one of two or more groups. Linear discriminant analysis is the most commonly used form of discriminant analysis. It has several assumptions that need to be met: no multicollinearity, multivariate normal distribution, and homogeneity of the covariance matrix [14]. The linear discriminant function can be expressed as follows:

$$D(\mathbf{X}) = b_0 + b_1X_1 + b_2X_2 + \cdots + b_gX_g, \tag{1}$$

where

$D(\mathbf{X})$: discriminant score $\quad$ b_g : discriminant coeffiecient of variable-g

b_0 : constant coefficient $\quad$ X_g : independent variable; $g = 1,2,\dots,p$

The assignment of an observation to one of the groups can be based on the discriminant score as follows:

if: $$\left(\boldsymbol{\mu}^{(l)} - \boldsymbol{\mu}^{(m)}\right)^T \mathbf{S}_{pooled}^{-1}\left[\mathbf{x}_0 - \frac{1}{2}\left(\boldsymbol{\mu}^{(l)} - \boldsymbol{\mu}^{(m)}\right)\right] \geq ln\left(\frac{P(m)}{P(l)}\right), l \neq m, \forall l, m \in k \tag{2}$$

then : $\mathbf{x}_0 \in$ group l

otherwise : $\mathbf{x}_0 \in$ group m,

where $P(k)$ denotes the prior probability for each group; $k = l, m, \dots, K$.

2.3. Multicollinearity

Multicollinearity occurs when there is a significant linear relationship between independent variables. It is typically tested using the Variance Inflation Factor (VIF) [15]. However, multicollinearity can also be assessed by examining the collinearity between variables through the correlation matrix and measured using the Spearman rank correlation coefficient (ρ). If the correlation coefficient between two variables approaches 1, specifically more than ± 0.7, indicates the presence of multicollinearity [16].

2.4. Multivariate Normality

In discriminant analysis, it is assumed that the independent variables follow a multivariate normal distribution [14]. The multivariate normality test can be conducted using various methods, including the Henze-Zirkler method and the Q-Q plot. The Henze-Zirkler method employs a non-negative functional distance. On the other hand, the Q-Q plot method shows that data follow a multivariate normal distribution if the plot of chi-square scores for each object forms a straight line [17-18]. The Henze-Zirkler test statistic can be expressed as follows:

$$HZ = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} e^{\left(-\frac{\beta^2}{2}D_{ij}\right)} - 2(1+\beta^2)^{-\frac{p}{2}}\frac{1}{n}\sum_{i=1}^{n} e^{\left(-\frac{\beta^2}{2(1+\beta^2)}D_i\right)} + (1+2\beta^2)^{-\frac{p}{2}} \tag{3}$$

$$\beta = \frac{1}{\sqrt{2}}\left(\frac{n(2p+1)}{4}\right)^{\frac{1}{p+4}} \tag{4}$$

$$D_{ij} = \left(\mathbf{x}_i - \mathbf{x}_j\right)^T \mathbf{S}^{-1}\left(\mathbf{x}_i - \mathbf{x}_j\right) \tag{5}$$

$$D_i = (\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{S}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}). \tag{6}$$

The Henze-Zirkler test statistic follows a log-normal distribution with mean (μ_{HZ}) and variance (σ^2_{HZ}), if the data exhibit a multivariate normal distribution. To test the significance of multivariate normality, the log-normalized parameters of the HZ statistic, namely (μ_{lnHZ}) and (σ_{lnHZ}) [19]. The Wald test statistic for multivariate normality can be determined using the following equation:

$$z = \frac{ln(HZ) - \mu_{lnHZ}}{\sigma_{lnHZ}}. \tag{7}$$

If the $p - value > \alpha$, then the assumption of multivariate normality is satisfied.

2.5. Homogenity of Variance Covariance Matrices

Box's M test, using a chi-square distribution approach, can be employed to test the assumption of homogeneity of covariance matrices [18]. The Box's M test statistic can be expressed as follows:

$$\ln M = \frac{1}{2}\sum_{k=1}^{K}(n^{(k)} - 1)\ln|\mathbf{S}^{(k)}| - \frac{1}{2}\left(\sum_{k=1}^{K}(n^{(k)} - 1)\right)\ln|\mathbf{S}_{\text{pooled}}| \tag{8}$$

$$c_1 = \left[\sum_{k=1}^{K}\frac{1}{(n^{(k)} - 1)} - \frac{1}{\sum_{k=1}^{K}(n^{(k)} - 1)}\right]\left[\frac{2p^2 + 3p - 1}{6(p+1)(k-1)}\right] \tag{9}$$

$$u = -2(1 - c_1)\ln M \sim \chi^2\left[\frac{1}{2}(k-1)p(p+1)\right], \tag{10}$$

where

$n^{(k)}$: number of observations in the k^{th} group

$\mathbf{S}^{(k)}$: covariance matrix of the k^{th} group

$\mathbf{S}_{pooled}$: pooled covariance matrix; $k = 1,2,3,\dots,K$.

The null hypothesis H_0 is accepted if $u \le \chi^2_{\alpha;\frac{1}{2}(k-1)p(p+1)}$, indicating that the assumption of homogeneity of covariance matrices is satiesfied.

2.6. Robust Linear Discriminant Analysis using MOM-Qn

Robust linear discriminant analysis is a concept that replace classical measures of location and scale with robust versions [9], [13], that are resistant to outliers and violations of assumptions. One such robust estimator that can be used is the Modified One-step M-estimator integrated with the Qn scale estimator (MOM-Qn). The Modified One-step M-estimator (MOM) is a measure of location or central tendency with high breakdown points and efficiency. MOM is flexible as it determines trimming criteria based on the tail distribution of observations. Additionaly, MOM is simply the average of the remaining observations after outliers have been eliminated [20-21], making the result of MOM a new robust mean vector. Mathematically, Wilcox (2003) defines MOM as follows:

$$\hat{\theta}_g^{(k)} = \frac{\sum_{i=t_1+1}^{n_g^{(k)}-t_2} x_{ig}^{(k)}}{n_g^{(k)} - t_1 - t_2}, \tag{11}$$

where

$\hat{\theta}_g^{(k)}$: Modified One-step M-estimator for each variable X_g in k^{th} group

$x_{ig}^{(k)}$: the i^{th} observation for variable X_g in k^{th} group

t_1 : number of observations $x_{ig}^{(k)}$ that meet $\left(x_{ig}^{(k)} - \hat{M}_g^{(k)}\right) < -2.24$ (scale estimator)

t_2 : number of observations $x_{ig}^{(k)}$ that meet $\left(x_{ig}^{(k)} - \hat{M}_g^{(k)}\right) > 2.24$ (scale estimator)

$\hat{M}_g^{(k)}$: median of $x_{ig}^{(k)}$ for each variable X_g.

Essentially, the scale estimator used in the MOM criteria is the mean absolute deviation (MAD_n). However, this study employs the Qn scale estimator in the MOM criteria. The Qn scale estimator is suitable as a substitute for standard deviation for skewed and heavy tailed [22-23]. Rousseeuw dan Croux (1993) define the Qn scale estimator as follows:

$$Q_{n,g}^{(k)} = d\left\{\left|x_{ig}^{(k)} - x_{jg}^{(k)}\right|; i < j\right\}_{(q)};\ q = \binom{s}{2}, s = \left\lfloor\frac{n}{2}\right\rfloor + 1 \tag{12}$$

where d is a constant value adjusted according to the data distribution: $d = 2.2219$ for normally distributed data; $d = 1.2071$ for heavy-tailed distributed data; dan $d = 3.4760$ for asymmetric distributed data [23].

The robust covariance matrix can be constructed by multiplying robust components [24-26], such as the Spearman rank correlation (ρ) and the Qn scale estimator.

$$\hat{\boldsymbol{\Sigma}}^{(k)} = \rho_{gh}^{(k)} \times Q_{n,g}^{(k)} \times Q_{n,h}^{(k)} \tag{13}$$

2.7. Evaluation of Accuracy and Precision in Classification

There are two methods that can be used to assess the performance of the discriminant function, Apparent Error Rate (APER) test and Press's Q test. The APER test is used to determine the probability of classification errors [17]. It can be easily calculated based on the confusion matrix shown in Table 2, which provides information regarding the classification of actual and predicted groups.

Table 2. Confusion Matrix

Actual group classification	Predicted group classification		Number of observations (n)
	D	ND	
D	$n_{D,D}$	$n_{D,ND}$	n_D
ND	$n_{ND,D}$	$n_{ND,ND}$	n_{ND}

After constructing the confusion matrix, the APER value can be calculated. The APER test statistic is expressed as follows:

$$Apparent\ Error\ Rate\ (APER) = \frac{n_{D,ND} + n_{ND,D}}{n_D + n_{ND}}, \tag{14}$$

Meanwhile, The Press's Q statistic is expressed as follows [27]:

$$Press's\ Q = \frac{\left[N - \left((n_{D,D} + n_{ND,ND})K\right)\right]^2}{N(K-1)}, \tag{15}$$

where

N : total number of observations

K : total number of groups.

If Press's $Q > \chi^2_{(\alpha,1)}$, it can be concluded that the classification is accurate and the formed discriminant function is stable [28].

3. Results and Discussion

3.1. Descriptive Statistics

A general description of each data category related to the financial ratios of conventional private banks listed on the Indonesia Stock Exchange during the period 2021-2022 can be seen in Table 3.

Table 3. Descriptive Statistics

Categories	Statistics	Variable				
		NPL	LDR	NIM	BOPO	CAR
Distress (D)	Min	0.000	12.35	0.690	83.10	11.13
	Median	3.330	78.23	3.410	98.83	35.87
	Mean	3.841	80.03	4.213	106.70	42.56
	Max	14.090	146.06	13.830	171.20	169.92
Non-Distress (ND)	Min	0.010	29.67	1.170	31.05	13.69
	Median	2.390	78.83	4.660	82.44	27.30
	Mean	2.724	89.88	4.850	93.23	43.13
	Max	9.450	355.00	15.870	287.86	283.84

Based on Table 3, it can be observed that banks as indicated distressed have an average capital adequacy ratio (CAR) of 42.56%, ranging from 11.13% to 169.92%, while non-distressed banks have an average CAR of 43.13%, ranging from 13.69% to 283.84%. The same things applied to the financial ratios of NPL, LDR, NIM, and BOPO. Additionally, the data description for each variable within each category can be visually represented using histograms in Figure 1.

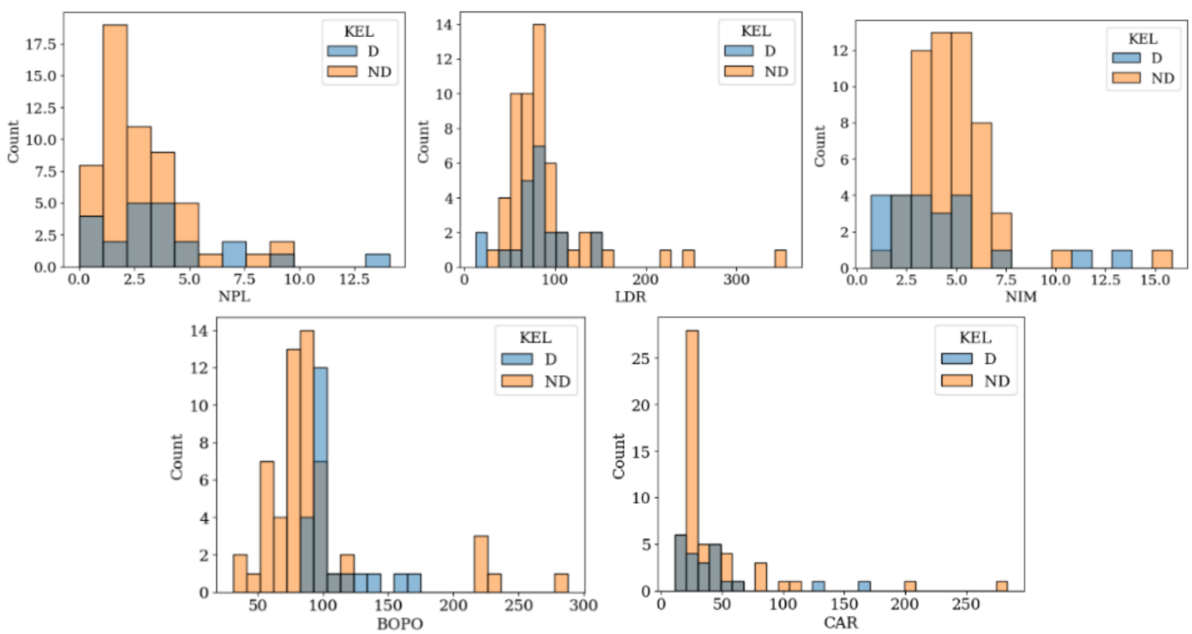

Figure 1. Histogram of indicated distress and non-distress categories on general conventional private banks listed on the Indonesia Stock Exchange in 2021-2022

Figure 1 shows that the distribution of data for each variable is asymmetrical with a longer tail to the right, which is known as positively skewed or skewed right. However, for the LDR in the indicated distressed category, it can be seen that the data tend to be symmetrically distributed, although this does not necessarily indicate that the data follow a normal distribution.

3.2. Checking Assumptions

The first assumption test conducted is the multicollinearity test, which can be performed by calculating the correlation values between independent variables. Collinearity assessment uses the Spearman rank correlation. The correlation results between variables are presented in matrix form in Table 4.

Table 4. Spearman rank correlation matrix

	NPL	LDR	NIM	BOPO	CAR
NPL	1.00000000	-0.12097070	-0.1597238	0.37880800	-0.04939521
LDR	-0.12097070	1.00000000	0.2350144	-0.03741828	0.17236038
NIM	-0.1597238	0.2350144	1.00000000	-0.44936517	0.24159190
BOPO	0.37880800	-0.03741828	-0.44936517	1.00000000	0.02472227
CAR	-0.04939521	0.17236038	0.24159190	0.02472227	1.00000000

Table 4 shows that the correlation coefficients between each pair of independent variables do not exceed -0.7 or 0.7. Therefore, indicating that there is no multicollinearity among the variables.

Next, the multivariate normality test, using the Henze-Zirkler statistic as described in equations (3-7) and the Q-Q plot generated using Rstudio. The output values of $HZ = 6.166332$ and $p - value = 0.0001113 < 0.05$. Thus, we reject H_0, indicating that the data do not follow a multivariate normal distribution.

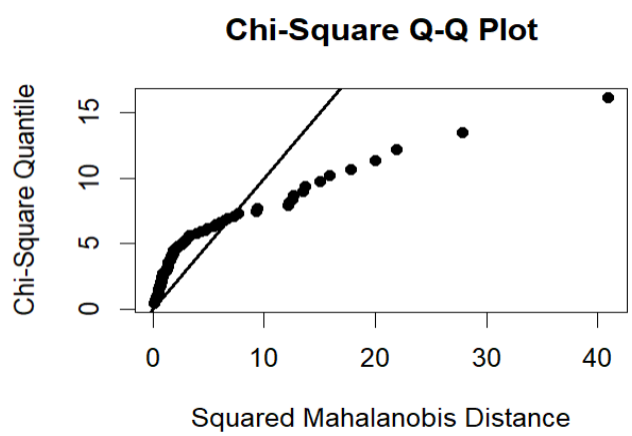

Figure 1. Q-Q plot of research data

Based on Figure 2, it can be seen that the resulting plot curves away from the model (straight line), confirming that the data do not follow a multivariate normal distribution. Thus, the assumption of multivariate normality is violated.

Finally, the equality of covariance matrices was examined using Box's M test according to equations (8-10). The results of Box's M test showed $u = 43.979 > \chi^2_{0.05;15} = 24.996$ and $p - value = 0.0001108 < 0.05$. Consequently, we reject H_0, indicating that the covariance matrices between groups are different or heterogeneous. Thus, the assumption of homogeneity of covariance matrices is violated.

3.3. Result of Robust Discriminant Linear Analysis using MOM-Qn

First, it is required to calculate the Qn scale estimator based on equation (12) and constant value adjusted according to each data distribution based on Figure 1. The Qn values for each group and variable are presented in Table 5.

Table 5. Result of Qn scale estimator

Variable (p)	Categories (k)	
	$Q_{n,g}^{(D)}$	$Q_{n,g}^{(ND)}$
NPL	4.17120	2.64176
LDR	13.12118	43.76284
NIM	4.41452	2.71128
BOPO	17.41476	33.89100
CAR	34.86428	16.51100

Based on Table 5, The Qn scale estimator values are used to determine the lower and upper bounds for the MOM trimming criteria. The upper and lower bounds based on the MOM-Qn criteria are used to detect observations considered upper outliers if they exceed the upper bound and lower outliers if they fall below the lower bound. The number of lower outliers (t_1) and upper outliers (t_2) are presented in Table 6.

Table 6. Selection of elimination based on the MOM-Qn criteria

Variable (p)	Categories (k)			
	Distress (D)		Non-Distress (ND)	
	t_1	t_2	t_1	t_2
NPL	0	1	0	2
LDR	2	2	0	3
NIM	0	1	0	1
BOPO	0	2	0	5
CAR	0	2	0	8

Table 6 illustrates the number of observations classified as lower and upper outliers based on the MOM-Qn criteria. As shown in Figure 1, most of the financial ratio data distributions are positively skewed, leading the MOM-Qn criteria to trim most observations on the right tail. Consequently, the remaining data, referred to as the MOM sample ($\hat{x}_{ig}^{(k)}$), will be used to calculate the robust mean vector. The MOM-Qn robust mean vector based on equation (11) is obtained as follows:

$$\hat{\boldsymbol{\theta}}^{(D)} = \begin{bmatrix} 3.35286 \\ 79.77000 \\ 3.75524 \\ 100.89250 \\ 31.94800 \end{bmatrix} \quad \hat{\boldsymbol{\theta}}^{(ND)} = \begin{bmatrix} 2.48130 \\ 79.54334 \\ 4.64927 \\ 78.83373 \\ 29.45688 \end{bmatrix}$$

In addition to the robust mean vector, a robust covariance matrix is required. This can be obtained by multiplying the Spearman rank correlation matrix with the Qn scale estimator values for each distress and non-distress group. To construct the robust linear discriminant function, the inverse form of the pooled robust covariance matrix is required, as follows:

$$\hat{\boldsymbol{\Sigma}}_{pooled}^{-1} = \begin{bmatrix} 0.119400 & 1.179748e-03 & 0.000394 & -0.004280 & 9.965157e-04 \\ 0.001180 & 7.462040e-04 & -0.001583 & -0.000011 & -5.661717e-05 \\ 0.000394 & -1.582795e-03 & 0.126158 & 0.005076 & -4.810909e-03 \\ -0.004280 & -1.083900e-05 & 0.005076 & 0.001465 & -2.042952e-04 \\ 0.000997 & -5.661717e-05 & -0.004811 & -0.000204 & 2.091901e-03 \end{bmatrix}$$

Once all these components are obtained, the resulting robust linear discriminant function is formulated as:

$$D_{MQ}(\mathbf{X}) = -2.45482 + 0.012053\, X_1 + 0.002232\, X_2 - 0.012825\, X_3 + 0.023545\, X_4 + 0.005862\, X_5 \quad (16)$$

Based on equation (16), NPL (X_1), LDR (X_2), BOPO (X_4), and CAR (X_5) positively affect the discriminant score, while NIM (X_3) negatively affects the discriminant score. Typically, CAR indicates a bank's ability to protect against various financial risks and would generally lower the discriminant score, indicating better financial health. However, in this context, an increase in CAR might imply that the bank needs to increase its capital as a precaution against elevated risks during the COVID-19 pandemic. Other variables are consistent with Bank Indonesia's guidelines: high NPL indicates poor financial health, high LDR suggests potential liquidity issues, and high BOPO indicates high operational costs, all indicating financial distress. Additionally, a lower NIM is favorable as it shows the bank's efficiency in managing its assets and generating interest income, reinforced by [29-30].

3.4. Result of Classification

Classification is conducted by calculating the discriminant score using the robust linear discriminant function from equation (16). To assign an observation to one of the groups using equation (2), a cutoff value based on the prior probability of each group is required. The calculated cutoff value is $\ln\left(\frac{P(ND)}{P(D)}\right) = \ln\left(\frac{0.7179487}{0.2820513}\right) = 0.9343092$. Thus, if $D_{MQ}(\mathbf{X}) \geq 0.9343092$, the observation is predicted to be indicated distressed; otherwise, if $D_{MQ}(\mathbf{X}) < 0.9343092$, the observation is predicted to be non-distressed. As an example, the classification results for several banks based on the discriminant fuction are summarized in Table 7.

Table 7. Result of classification based on discriminant score

No.	Bank	Year	Discriminant Score	Actual categories	Prediction categories
1	AGRS	2021	0.2991943	D	ND*
2	AMAR	2021	0.1636601	D	ND*
⋮	⋮	⋮	⋮	⋮	⋮
77	PNBN	2022	-0.3563344	ND	ND
78	SDRA	2022	-0.4631161	ND	ND

3.5. Result of Actual and Predicted Group Classification Evaluation

The evaluation of the robust linear discriminant function in classifying banks as indicated distressed and non-distressed can be done using the APER and Press's Q tests. First, the classification results for actual and predicted groups can be shown in the confusion matrix in Table 8.

Table 8. Result of classification using confusion matrix

Actual group classification	Predicted group classification		Number of observations (n)
	D	ND	
D	$n_{D,D} = 5$	$n_{D,ND} = 17$	$n_D = 22$
ND	$n_{ND,D} = 7$	$n_{ND,ND} = 49$	$n_{ND} = 56$

Based on Table 8, 5 out of 22 banks were correctly classified as indicated distressed, and 49 out of 56 banks were correctly classified as non-distressed. The precision of the robust linear discriminant function using the APER test shows a misclassification rate of 30.77% and classification accuracy of 69.23%. Additionally, the Press's Q test was used to assess the accuracy and stability of the discriminant function, resulting in a Press's Q value of 11.53846. With $\alpha = 0.05$ and (df) = 1, the critical value is $\chi^2_{(0.05;1)} = 3.841459$. According to the established criteria, we reject H_0 because the value of Press's Q $> \chi^2_{(0.05;1)}$. Therefore, the classification performed using the robust linear discriminant function demonstrates a good level of accuracy, precision, and stability.

4. Conclusion

Based on the research conducted using robust linear discriminant analysis with the Modified One-step M-estimator Qn scale (MOM-Qn), the robust linear discriminant function in equation (16) was obtained, with coefficients indicating that NPL, LDR, BOPO, and CAR positively affect the discriminant score, while NIM negatively affects it. This discriminant function was used to classify conventional private banks listed on the Indonesia Stock Exchange (IDX) as either indicated distressed or non-distressed. The results predicted 12 banks as indicated distress and 66 banks as indicated non-distressed, with a classification accuracy of 69.23%. According to [13], [25] higher outlier contamination increases misclassification rates in linear models. Even with robust methods, significant outlier contamination still raises misclassification rates. Furthermore, the Press's Q test results concluded that the classification using the robust linear discriminant function is reasonably accurate and stable. Overall, this study demonstrates the effectiveness of the MOM-Qn method in improving classification accuracy and provides an alternative for financial health assessment in the banking sector.

5. Acknowledgement

The author expresses sincere gratitude to Kankan Parmikanti, Dra., M.Stat., and Nurul Gusriani, S.Si., M.Si., along with the Department of Mathematics, Universitas Padjadjaran for their invaluable assistance in the completion of this research.

References

[1] Nazarudin, J., Gusriani, N., Parmikanti, K., & Susanti, S. (2023). Application of Threshold Generalized Autoregressive Conditional Heteroscedastic (TGARCH) Model in Forecasting the LQ45 Stock Price Return. *Eksakta: Berkala Ilmiah Bidang MIPA,* 24(2), 271–284.

[2] Darjana, D., Wiryono, S. K., & Koesrindartoto, D. P. (2022). The COVID-19 Pandemic Impact on Banking Sector. *Asian Economics Letters,* 3(3).

[3] Sutra, F. M., & Mais, R. G. (2019). Faktor-Faktor yang Mempengaruhi Financial Distress dengan Pendekatan Altman Z-Score pada Perusahaan Pertambangan yang Terdaftar di Bursa Efek Indonesia Tahun 2015-2017. *Jurnal Akuntansi dan Manajemen,* 16(1), 34–72.

[4] Bank Indonesia. (2011). Peraturan Bank Indonesia Nomor: 13/1/2011 tentang Penilaian Tingkat Kesehatan Bank Umum.

[5] Purnamasari, D., Suwandi, T., Rahayu, E. P., Kesehatan Hangtuah Pekanbaru, I., & Rahayu, E. P. (2020). Multivariate Analysis on The Determinants of Work Fatigue Factors for Nurses Inpatient Care at RSUD Arifin Achmad Hospital Pekanbaru. *Eksakta: Berkala Ilmiah Bidang MIPA,* 21(1), 59–69.
[6] Hartono, A., Dewi, L. A., Yuniarti, E., Putri, S. T. H., Harahap, T. S., & Hartono, A. (2023). Machine Learning Classification for Detecting Heart Disease with K-NN Algorithm, Decision Tree and Random Forest. *Eksakta: Berkala Ilmiah Bidang MIPA,* 24(4), 513–522.
[7] Tahmasebi, R., Rostamy, A. A. A., Khorshidi, A., & Sadeghi Sharif, S. J. (2020). A data mining approach to predict companies' financial distress. *International Journal of Financial Engineering,* 7(3), 2050031.
[8] Wieprow, J., & Gawlik, A. (2021). The use of discriminant analysis to assess the risk of bankruptcy of enterprises in crisis conditions using the example of the tourism sector in Poland. *Risks*, 9(4).
[9] Tjahaya, M. S., & Tinungki, G. M. (2022). Analisis Diskriminan Linear Robust Dengan Metode Winsorized Modified One-Step M-Estimator. *Journal of Statistics and Its Application,* 3(1), 1–13.
[10] Alrawashdeh, M. J., Radwan, T. R., & Abunawas, K. A. (2018). Performance of Linear Discriminant Analysis Using Different Robust Methods. *European Journal of Pure and Applied Mathematics,* 11(1), 284–298.
[11] Auguin, N., Morales-Jimenez, D., & McKay, M. R. (2019). Robust Linear Discriminant Analysis using Tyler's Estimator: Asymptotic Performance Characterization. *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* 5317–5321.
[12] Yahaya, S., Lim, Y. F., Ali, H., & Omar, Z. (2016). Robust Linear Discriminant Analysis with Automatic Trimmed Mean. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC),* 8(10), 1–3.
[13] Melik, H. N., Ahad, N. A., & Yahaya, S. S. S. (2018). Modified one-step M-estimator with Robust Scale Estimator for Multivariate Data. *Journal of Engineering and Applied Sciences,* 13(24), 10396–10400.
[14] Backhaus, K., Erichson, B., Gensler, S., Weiber, R., & Weiber, T. (2023). Multivariate Analysis an Application-Oriented introduction. *Springer Gabler.*
[15] Dewi, K., Gusriani, N., & Parmikanti, K. (2023). Factors Affecting the Number of Infant Morality Cases in West Java for the 2019-2020 Period using Generalized Poisson Regression (GPR). *Eksakta: Berkala Ilmiah Bidang MIPA,* 24(2), 259–270.
[16] Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2011). Statistics for Business and Economics (11th ed.). *South Western College Pub.*
[17] Johnson, R. A., & Wichern, D. W. (2007). Applied Multivariate Statistical Analysis. *Pearson Prentice Hall.*
[18] Achiar, A. L. M., Aidi, M. N., & Kurnia, A. (2023). Identification of Atherosclerosis Based on The Differences in Cholesterol and Creatinine in Indonesia with Multivariate Analysis of Variance. *Eksakta: Berkala Ilmiah Bidang MIPA,* 24(3), 315–329.
[19] Korkmaz, S., Goksuluk, D., & Zararsiz, G. (2014). MVN: An R Package for Assessing Multivariate Normality. *R Journal,* 6(2), 151–162.
[20] Wilcox, R. R. (2003). Applying Contemporary Statistical Techniques. *Academic Press.*
[21] Abdul Rahman, A., Syed Yahaya, S. S., & Atta, A. M. A. (2020). Robustification of CUSUM control structure for monitoring location shift of skewed distributions based on modified one-step M-estimator. *Communications in Statistics - Simulation and Computation,* 49(11), 3001–3018.
[22] Adekeye, K. S., Adewara, J. A., Aako, O. L., & Olaomi, J. O. (2021). Performance of median absolute deviation and some alternatives to median absolute deviation control charts for skewed and heavily tailed process. *Quality and Reliability Engineering International,* 37(8), 3431–3440.

[23] Rousseeuw, P., & Croux, C. (1993). Alternatives to the Mean Absolute Deviation. *Journal of the American Statistical Association,* 88(424), 1273–1283.
[24] Pang, Y. S., Ahad, N. A., & Yahaya, S. S. S. (2022). Robust Linear Discriminant Rule Using Double Trimming Location Estimator with Robust Mahalanobis Squared Distance. *Pertanika Journal of Science & Technology,* 30(4), 2393–2406.
[25] Ahad, N. A., Pang, Y. S., Yahaya, S. S. S., & Abdullah, S. (2023). Robust multiple discriminant rule with Hodges-Lehmann in handling equal proportion of cellwise-casewise outliers. *AIP Conference Proceedings,* 2896(1).
[26] Lim, Y. F., Yahaya, S. S. S., & Ali, H. (2016). Winsorization on linear discriminant analysis. In The 4th International Conference on Quantitative Sciences and Its Applications. *American Institute of Physics Inc.*
[27] Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). Multiple Discriminant Analysis. In Multivariate Data Analysis (7th ed., p. 267). *Pearson Prentice Hall.*
[28] Destriana, M., Gusriani, N., & Irianingsih, I. (2018). Klasifikasi Status Kinerja Bank yang Terdaftar di BEI dengan Pendekatan Winsorized Modified One-Step M-Estimator. *Jurnal Matematika Integratif,* 14(2), 133–140.
[29] Fakhruddin, A. H., & Aji, T. S. (2024). Risk-Based Bank Rating (RBBR) Method Using Risk Profile, Good Corporate Governance, Earning, and Capital (RGEC) Factors as an Assessment of Bank Health: Case Study of PT Bank Pembangunan Daerah Jawa Timur Tbk Period 2018-2022. *International Journal of Social Science and Human Research*, 7(1), 151–159.
[30] Abdullah, J., Hasan, W., & Dusa, S. Y. (2021). Analysis of Capital Adequacy Ratio (CAR), Non-Performing Loan (NPL), and Net Interest Margin (NIM) in Predicting Financial Distress in Financial Reports of PT. Bank Rakyat Indonesia (Persero) Tbk. *Journal of Economics, Finance and Accounting Studies (JEFAS),* 3(2), 81–90.