Mathematical Sciences and Informatics Journal
Vol. 5, No. 2, Nov. 2024, pp. 81-94
http://www.mijuitm.com.my
DOI : 10.24191/mij.v5i2.926

Predicting Default and Non-Default Firms using Discriminant Analysis: Adaptation of KMV-Merton's Default Probabilities and Financial Ratios

Nur Ain Al-Hameefatul Jamaliyatul
Universiti Teknologi MARA, Cawangan Johor, Kampus Segamat, Johor, Malaysia.
ainnurain06@gmail.com

Nurul Afiqah Zainuddin
Unit 25-01, Level 25, Menara Felda, 11 Persiaran KLCC, 50450 Kuala Lumpur, Malaysia.
afiqahzainuddin99@gmail.com

Izza Suraya Zulhazmi
Wisma United Malacca, No.61 Level 4 Jalan Melaka Raya 8, Taman Melaka Raya, 75009 Melaka, Malaysia.
izzasuraya99@gmail.com

Norliza Muhamad Yusof
Kolej Pengajian Perkomputeran, Informatik dan Matematik, Universiti Teknologi MARA, Cawangan Negeri Sembilan, Kampus Seremban, Negeri Sembilan, Malaysia.
norliza3111@uitm.edu.my

Muhamad Luqman Sapini
Kolej Pengajian Perkomputeran, Informatik dan Matematik, Universiti Teknologi MARA, Cawangan Negeri Sembilan, Kampus Seremban, Negeri Sembilan, Malaysia.
luqman0211@uitm.edu.my

Article Info

Article history:

Received Aug 12, 2024
Revised Aug 25, 2024
Accepted Sept 15, 2024

Keywords:

Discriminant analysis
KMV-Merton
Financial ratios
Default
Adaptation

ABSTRACT

The KMV-Merton model provides conceptual determinants for predicting firms' default risk, but its accuracy was tested long ago, and it contains insufficient statistics for default prediction. Therefore, previous literature adapted the KMV-Merton model into a statistical model involving financial ratios to improve its predictive capabilities. Discriminant Analysis (DA) is a widely used statistical model for predicting financial distress. The objectives of this study are to identify financial ratios significant to KMV-Merton's default probabilities using DA, to predict default and non-default firms using the DA model obtained, and to compare the performance of the KMV-Merton and DA models in predicting default risk. The study uses 11 years of data from Malaysian publicly listed firms, applying the KMV-Merton model and stepwise DA in SPSS. DA identifies the significance of selected financial ratios to firm default, with KMV-Merton's default probabilities as the dependent variable, forming a discrimination function to predict default and non-default firms. Credit ratings and Type 1 and Type II errors are used to compare model performance. The DA using SPSS reveals a discriminant function with net profit margin and return on assets significantly related to KMV-Merton's default probabilities. The DA model is more biased in predicting non-default firms due to the need for more information on default firms, yet it slightly outperforms the KMV-Merton model. This study offers guidance on adapting KMV-Merton's default probability estimates with financial ratios in the DA model and highlights the significant financial ratios related to KMV-Merton's default probabilities.

Corresponding Author:
Norliza Muhamad Yusof
Kolej Pengajian Perkomputeran, Informatik dan Matematik, Universiti Teknologi MARA Cawangan Negeri Sembilan Kampus Seremban, Negeri Sembilan, Malaysia.
Email: norliza3111@uitm.edu.my

1. Introduction

Default risk prediction is crucial as it helps to identify the financial condition of the firms [1] plus it assists the investors in acknowledging the firm's performance. Over the last decade, a variety of mathematical modelling methods have been initiated to estimate firms' default risk. Many categories of models can be separated based on the approach they take. The most well-known structural model of default risk, named KMV-Merton model, has been commonly used to forecast a firm's default. The KMV-Merton model has been used since the early period of credit risk modelling.

The KMV-Merton model is a revision of the Merton model that varies in certain aspects. It provides an assessment of the likelihood of firms to default based on the ability of the firm equity holders to pay off their debts. Recent studies on the application of the KMV-Merton model include studies on the impact of COVID-19 on default risk [2], the default risk of internet finance companies [3], and the firm's volatility estimation [4]. Other recent studies, such as [5] focused on improving the accuracy of the model. Besides the ability of the model to predict default in advance, there were discussions on the data implementation of the model in capturing enough data on default. This led to many modifications of the KMV-Merton model. It is said that the probability of default from the KMV-Merton model can only be calculated using some comparability analysis based on accounting data [6]. Others said that the KMV-Merton model is not a sufficient statistic for the probability of default [7].

Therefore, in the previous literature, the KMV-Merton model was adapted into a statistical model to form a model to predict the likelihood of firms defaulting. Most studies concluded that the combination of the KMV-Merton model with the statistical models gives better default prediction (see [8]–[13]). From this combination also, the factors that contribute to a certain event were analyzed. For example, one said that the default probability from the KMV-Merton model contributes to the firm's sector analysis [14], and tax arrears predict corporate bank loan defaults [15]. Among the studies, one of the statistical models used is the Discriminant Analysis (DA).

The DA is a statistical technique used to evaluate the differences between each observation. It is the most direct form of a linear combination of utilised variables developed in 1936 by Fisher and can be viewed as predictive [10]. [16] is the pioneer of the univariate approach of DA in bankruptcy prediction. Then, [17] expanded it to a multivariate context and developed the Z-Score model. A variety of research comes from these, making the DA one of the widely used methods in predicting the financial distress of firms [18], [19]. The direct techniques and stepwise approaches are the two most often utilised strategies for developing discriminant models ([20]).

In applying the statistical models, the authors [8]–[10], [12], [13] utilised the same independent variables encompass financial ratios to predict the probability of default of companies. Financial ratios are vital components in analysing the financial distress of firms [21]–[23]. A study found that financial ratios except for cash flow ratios have a significant impact on the firm's probability of default [24]. Meanwhile, [25] found that return on assets, current ratio, debt to total assets ratio, sales to working capital ratio and cash flow to total assets ratio are statistically significant in predicting default. Apart from that, Debt Ratio, Total Assets Turnover, and Net Profit Margin are also found significant to financial distress [26].

Nonetheless, the authors [8]–[10], [12], [13] used different approaches to define the dependent variables. Some used the information from financial statements [8], company rating [9], loan payment [10], the hybrid KMV-Merton and logistic score model [12] and a list of insolvent/distressed companies [13]. This study used a similar approach as [12], where the KMV-Merton default probabilities are used to define the dependent variables. Since we are using the DA, we are not converting the KMV-Merton default probabilities into a logistic score as [12] did.

Accordingly, this paper adapts the KMV-Merton default probabilities to be the dependent variable and selected financial ratios as the independent variables of the DA model to predict the default/non-default of Malaysian publicly listed companies. This study contributes to the obtaining of the discriminant function that can significantly differentiate between default and non-default of the firms. In addition, the probability of default of the selected firms is estimated quarterly using the KMV-Merton model, instead of yearly as mostly done in the previous literature [3], [27], [28]. Lastly, the performance of the discriminant function obtained is evaluated in our study based on Type 1 and Type II errors.

The rest of the paper is organized as follows. In Section 2, the data and the methodology used in this study to predict default and non-default firms are explained. Then in Section 3, the results are presented and discussed accordingly. Lastly, Section 4 concludes.

2. Data and Methodology

This section explains the data, mathematical model and methods used to achieve the goal of this study. This includes an explanation of how default prediction is done using the KMV-Merton model and discriminant analysis.

2.1 Data Description

This study utilised 11-year financial data from six publicly listed Malaysian firms. Half of the firms were rated consistently from AAA to AA- and the other half were rated inconsistently from AA- to D. The data obtained is in the form of quarter data from 2009 to 2019. We used about 70% of the data from 2009 to 2016 as data training to obtain a discriminant function. The discriminant function obtained is then used to predict the default and non-default firms quarterly from 2017 to 2019, using the rest 30% of the data.

The data collected for the default probabilities estimation are market capitalisation obtained from the DataStream, and short-term borrowing and long-term borrowing obtained from the quarterly report of the firms.

The second type of data is for calculating the financial ratios of firms. The data collected for calculating the financial ratios of firms involving current assets, current liabilities, inventory, total liabilities, total assets, shareholders' equity, operating profit, interest expense, account receivables, net sales, and net profit. These quarterly data are obtained from the quarterly reports of the firms.

2.2 Discriminant Analysis

Discriminant analysis (DA) derives a linear combination of independent variables that discriminates between default and non-default firms from an equation that takes the following form [17]:

$$Z = a_0 + a_1 X_1 + a_2 X_2 + \cdots + a_N X_N \qquad (1)$$

where:
Z is discriminant score,
a_0 is coefficient (discriminant) weight,
$a_1, a_2, \ldots, a_N$ are discriminant coefficients,
$X_1, X_2, \ldots, X_N$ are discriminant variables.

Based on [8], there are two methods widely used for the derivation of the discriminant model, which are the direct and stepwise methods. The difference between the two methods is based on model construction. The direct method is based on model construction, which means that the model is defined in advance and then used in DA. Meanwhile, the stepwise method is a method where a subset of variables is chosen. Five statistical methods can be chosen to undergo stepwise methods: Wilks' lambda [29], Unexplained Variance [30], Mahalanobis distance [31] and Smallest F ratio [32] [33].

Wilks' lambda test can determine whether a link exists between the dependent variable and the explanatory variables [10]. It also tests the importance of the discriminant function by measuring differences between groups [34]. Wilks' lambda test values are always between 0 and 1. A value of 1 indicates that the median is equal, and the most discriminating variable has a lambda value and significance level close to 0. A low lambda value indicates minimal intra-group Variance and, thus, substantial intergroup variation, resulting in a significant difference in class mean.

The value of standardised discriminant function coefficients interprets the discriminant function. It is vital to examine the relative importance of the variables by analysing the absolute magnitude of the coefficient. The higher the standardised coefficient value, the more significant the contribution of the respective variable to the discrimination between groups [35]. This can be interpreted from the results of the standardised canonical discriminant function coefficients. Both studies from [10] and [11] show that the variable that gives the highest discriminant function coefficient is the most crucial variable to its discriminant function.

However, the results of interpretation of the coefficient value did not tell the values of the discriminant function that discriminant between the groups. This is where the group means are important. Group means are called centroids. The different values of the group centroids visualise how the function can discriminate between groups. This is presented by a study from [10] that shows the positive values of the group centroids determine that the firm is healthy, while negative values of the group centroids show that the firm is failing. The value of group centroids depends on the means of each group. Therefore, different studies obtained different values of group centroids.

In summary, this study focuses on obtaining the discriminant function as in Equation (1) for default risk prediction. This study used Wilks' lambda which is a stepwise step to the DA done. Meanwhile, the default probability estimated from the KMV-Merton model is used to determine the Z-Score of Equation (1).

2.3 KMV-Merton's Default Probabilities

The probability of default (*PD*) is defined as the probability that the market value of an asset falls below the face value of debt at time *t*. [36] believed that the dynamic of market value (dV_A) of the underlying properties of the firm follows the Geometric Brownian Motion as follows:

$$d_A = \mu V_A dt + \sigma V_A dW \tag{2}$$

where V_A is the market value of the firm's asset, μ is the drift rate, σ is the volatility, and *dW* is the Wiener Process. Since the natural log of future asset values is distributed normally, the PD is estimated using the standard normal distribution function of inverse *d* as follows [8]:

$$\mathrm{PD} = NORMS(-d) \tag{3}$$

for *d* being as distance to default, *d* that is computed by the following equation:

$$d = \left[ln\left(V_{A,t}\right) - ln(B_t) + (\mu - 0.5\sigma^2)T\right]/\sigma\sqrt{T} \tag{4}$$

where:

B_t is the book value of a firm's liabilities at any time *t* defined as the summation of short-term and half of long-term borrowings. Here, the values of borrowings are the same each day unless there are changes in the quarterly report,

$V_{A,t}$ is the daily market value of an asset at any time *t* defined based on basic accounting definition, firm's market capitalisation plus its book value of liabilities,

μ is the expected firm's asset return calculated by finding the mean of the daily log return of the market value of the asset for each quarter,

σ is the asset volatility that calculated using the standard deviation of the daily log return of the market value of the asset. Then, the standard deviation is multiplied by the square root of the number of trading days, $\sqrt{63}$, to obtain quarterly volatility. The 63 days referred to the number of trading days in each quarter [37],

T is time denoted as one quarter.

2.4 Setting Up the Dependent and Independent Variables of the Discriminant Function

The dependent variable denoted by the Z-score given in Equation (1) is determined based on the PD estimated from the KMV-Merton model. Equations (3) and (4) are used to estimate the PD of the selected firms and are next categorised according to Table 1.

Table 1. Discriminant Score Z

Firm's category	PD	Z-score
Non-default	0-0.49	0
Default	0.50-1.00	1

Meanwhile, the independent variables known by the discriminant variables ($X_1, X_2, X_{3,}\ldots, X_N$) are represented by the financial ratios shown in Table 2.

Table 2. List of financial ratios and formulas [38]

Variables	Financial Ratios	Formula
X_1	Current ratio	Current assets / Current liabilities
X_2	Acid test ratio	(Total current assets-inventory-prepayments) / Current liabilities
X_3	Debt ratio	Total liabilities (total non-current liabilities and current liabilities / Total assets (total non-current assets and current assets)
X_4	Debt to equity ratio	Total liabilities / Shareholders' equity
X_5	Time interest earned ratio	Operating profit / Interest expense
X_6	Total assets turnover	Net sales / Total assets (non-current assets and current assets)
X_7	Account receivables turnover ratio	Trade receivables x 63 / Total net sales
X_8	Net profit margin ratio	Net profit / Total net sales or revenue
X_9	Return on assets ratio	Net profit / Total assets
X_{10}	Return on capital employed ratio	Operating profit / Capital employed

These are three criteria considered in the selection of ratios [9]:

i. The financial ratios have been identified theoretically as indicators for determining default,
ii. previously used in empirical work to predict insolvency,
iii. and can be computed and determined conveniently from the researcher's database.

2.5 Predicting the Firm's Default Risk

SPSS Discriminant Analysis was employed to run 70% of the data, that is set up as explained in sub-section 2.4 to obtain a discriminant function as in Equation (1). The method used in this study is Wilks' lambda. It is a variable selection method in the stepwise discriminant analysis that chooses variables for the discriminant function. The KMV-Merton model and the discriminant function obtained from running the SPSS discriminant analysis are used to predict the default and non-default firms using the rest of 30% of the data.

The prediction was done before and after the adaptation of the KMV-Merton model into the DA. Therefore, by using Equations (3) and (4), the PD of the firms is estimated and categorised as default or non-default according to Table 1. This represents the prediction before the adaptation was made. The next prediction is after the KMV-Merton model was adapted into the DA model, where the discrimination function Z-Score is obtained as in Equation (1). Here, the discriminant function consists of only significant independent variables. The Z-Score of the firms was calculated and classified into default and non-default groups according to one of the outputs from the SPSS discriminants analysis, which is called functions at the group centroids. Group centroids indicate the mean values for the discriminant functions (Z-score) for a group. Therefore, the Z-Score calculated from the discriminant function that is near the centroid is said to belong to that group.

After that, the performance of each prediction is determined based on Type I and Type II errors. Type I error is defined as incorrectly classified default firm as non-default, while Type II error is defined as incorrectly classified non-default firm as default [39]. All these are expressed in formulae as follows [40]:

$$\text{Accuracy (\%)} = \frac{C_d}{T_d} + \frac{C_s}{T_s} \tag{5}$$

$$\text{Type I error (\%)} = 1 - \frac{C_d}{T_d} \tag{6}$$

$$\text{Type II error (\%)} = 1 - \frac{C_s}{T_s} \tag{7}$$

where T_d is the actual number of default firms, T_s is the actual number of non-default firms, C_d is the number of correctly predicted default firms, and C_s is the number of correctly predicted non-defaulted firms. In actual cases, the default and non-default firms are determined based on the ratings available for the firms. Specifically, firms with a rating between AAA to BBB- are considered non-default, while B+ to C is considered as default.

3. Results and Discussion

This section discusses the results obtained from the study. It involves a discussion of the results from the data descriptive, discriminant analysis using SPSS, and performance before and after the combination of the KMV-Merton model in the discriminant analysis model.

3.1 Descriptive Statistics

Table 3 shows the data descriptive for estimating firms' default probabilities using the KMV-Merton Model.

Table 3. The data descriptive for the default probabilities estimation

	N	Minimum	Maximum	Mean	Standard Deviation
Market capitalisation	264	73.956	94871.726	15994.516337	23929.833735
Short term borrowing	264	0	14855.141	8584.735807	1870.898180
Long term borrowing	264	0	11386.399	1617.181595	2908.022436

Meanwhile, Table 4 shows the descriptive data of the financial ratios used in the Discriminant Analysis.

Table 4. The data descriptive for the financial ratios

	N	Minimum	Maximum	Mean	Standard Deviation
Current asset	264	10.8461	15109.329	2392.763383	339.8933931
Current liabilities	264	6.3043	10956.308	1724.949362	2757.31698
Inventory	264	.6789	1935.419	148.9530682	265.8803581
Total liabilities	264	44.0111	19652.028	3639.699565	5707.969253
Total assets	264	69.7051	56155.899	9529.74627	16315.067939
Shareholders' equity	264	24.9468	16763.5	2556.981929	3954.369202
Operating profit	264	-1651.89	2040.89	85.23968106	261.7806059
Interest expense	264	-.943	1995.714286	22.83984718	131.70429
Account receivables	264	2.393	5246.37	651.2074602	114.5255567
Net sales	264	2.981	4889.12	547.1006371	993.0591672
Net profit	264	-409.777	1695.973	87.97215492	237.2578048

In Tables 3 and 4, we have 264 samples for each type of financial data of firms. The data with the lowest and highest mean are the interest expense and market capitalization, respectively. Greater standard deviation in the data shows greater variations in the samples. This means the spreads of each data distribution from the mean are all high, with the most found in the data of market capitalization and the least in the data of account receivables.

3.2 Tests of Equality of Group Means and Stepwise Statistics

The purpose of the equality test is to prove the significant differences between non-default and default groups on each of the independent variables. The output from this test can be seen in Table 5. Stepwise statistics show the steps taken for the selection of variables that are included in the analysis. The method chosen for stepwise statistics is Wilks' Lambda. It allows one to select the variables that will be entered in the discriminant function. The result from this stepwise statistic is given in Table 6.

Table 5. Tests of Equality of Group Means

Discriminant Variables	Wilks' lambda	F	df1	df2	Sig.
X_1	0.994	1.134	1	190	0.288
X_2	0.993	1.398	1	190	0.239
X_3	0.999	0.163	1	190	0.687
X_4	0.993	1.385	1	190	0.241
X_5	1.000	0.015	1	190	0.902
X_6	1.000	0.019	1	190	0.891
X_7	0.992	1.471	1	190	0.227
X_8	0.941	11.982	1	190	0.001
X_9	0.883	25.116	1	190	0.000
X_{10}	0.883	25.112	1	190	0.000

Table 5 shows which ratios are considered the most discriminating variable between the non-default and default groups. It also tests the null hypothesis that the group means are equal across all dependent variables. If Wilk's lambda is smaller than the critical value, then the null hypothesis can be rejected. The variables X_8, X_9, and X_{10} show significant differences in Wilk's lambda and F-values, leading to the rejection of the null hypothesis. These variables present higher values of F's and lower values of Wilks's lambda. The significant value (Sig.) that is close and equal to 0 also indicates that these variables are significant to the discriminant function. This is parallel to Table 6 where both X_8 and X_9, have a high tolerance of 0.931, near 1. It proves that these variables contribute high information to the discriminant function model. Although X_{10} gives a high correlation to the discriminant function, we found that X_{10} and X_9 are detected with multicollinearity in the pooled within-groups matrices test. Therefore, one of the variables must be removed to avoid potential problems in the prediction.

Table 6. Variables in the Analysis

Step		Tolerance	F to Remove	Wilks' Lambda
1	X_9	1.000	25.116	
2	X_9	.931	16.918	.883
	X_8	.931	4.345	.941

3.3 Test of Homogeneity of Covariance Matrices

Tests of homogeneity of covariance matrices include the outputs from Tables 7 and 8. This test is conducted to show whether the covariance matrices are equivalent or not. It can be evaluated through the null hypothesis.

Table 7. Log Determinants

Default Probability	Rank	Log Determinant
Non-default	2	20.062
Default	2	22.781
Pooled within-groups	2	20.233

Table 8. Test Results

Box's M		21.683
F	Approx.	5.883
	df1	3
	df2	553.242
	Sig.	0.001

Table 7 presents the log determinants for the group's covariance matrix and the pooled within-group covariance. The Default group shows the highest log determinant value, indicating a difference in its covariance matrix. This is supported by Table 8. Table 8 tests the null hypothesis that the covariance matrices do not differ between groups formed by the dependent variables. The null hypothesis is that if the significance is greater than 0.05, the covariance matrices are equal (H0), and if the significance is less than 0.05, the covariance matrices are not equal (H1). Table 8 shows the high value of Box's M, which is 21.683, and the significance (Sig.) of F tends to be 0, which is less than 0.05. According to [10], the Box's M values must be high, and the significance of the F must be near 0 for the analysis to be valid, which is to be unequal in covariance matrices. Thus, this study rejects the null hypothesis of equal covariance matrices. In addition, the amount of "Rank" in Table 7 represents two significant independent variables that can be used for the discriminant function model.

3.4 Summary of Canonical Discriminant Function

In summary of the canonical discriminant function, five outputs are obtained, and they are presented in Tables 9-14.

Table 9. Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	0.158	100.0	100.0	0.370

Note: Function 1 canonical discriminant functions were used in the analysis.

Table 9 provides information on the discriminant functions produced. Eigenvalue indicates the measure of association between the discriminant function and the dependent variable. A higher eigenvalue (near 1) displays a stronger discriminant function model. Like the canonical correlation, the value that is near to one presents a better discriminant function model [10]. In this study, the canonical correlation is 0.370, and the eigenvalue of 0.158, which is not extremely high. However, this model is statistically significant based on Table 10. Hence, it can still be considered a good model [11].

Table 10. Wilks' lambda

Test of Function(s)	Wilks' lambda	Chi-square	df	Sig.
1	0.863	27.762	2	0.000

Table 10 shows Wilks' lambda of the discriminant function. The closer Wilks' lambda value to 0 illustrates the higher quality of the model [10]. Here, the discriminant function model could be better based on the Wilks' lambda 0.863. However, the discriminant function is said to be able to discriminate the groups based on the significant value (Sig.) at 0.000.

Table 11. Structure Matrix

Discriminant Variables	Function
	1
X_9	.914
$X_{10}{}^a$	.913
X_8	.631
$X_4{}^a$	.387
$X_7{}^a$	-.212
$X_3{}^a$	-.159
$X_1{}^a$	-.046
$X_6{}^a$	.019
$X_5{}^a$	-.015
$X_2{}^a$	-.012

Note: Pooled within-group correlations between discriminating variables and standardised canonical discriminant functions. Variables are ordered by the absolute size of correlation within the function. (a) indicates the variables that are not used in the analysis.

Table 11 provides the structure matrix, which illustrates the importance of correlations between discriminant variables and discriminant function. By ignoring the variables that are not used in the analysis, X_9 presents the highest absolute size of correlation within functions, followed by X_8. Hence, X_9 is considered the most important variable in determining the firm's default risk, followed by X_8. Thus, it is confirmed that X_8 (net profit margin ratio) and X_9 (return on assets ratio) are suitable independent variables for the discriminant function of the model.

Table 12. Canonical Discriminant Function Coefficients

Selected discriminant variables	Function
	1
X_8	.009
X_9	.001
(Constant)	-.301

Note: Unstandardised coefficients.

Table 12 presents the list of coefficients of the independent variables X_8 and X_9. These are used to form an unstandardised discriminant function expressed as:

$$Z = -0.301 + X_8 * 0.009 + X_9 * 0.001 \quad (8)$$

Equation (8) represents the discriminant function Z-score, which is the result of the adaptation of the KMV-Merton model and financial ratios into the DA. It is used in this study to make predictions on default and non-default groups.

Table 13. Functions at Group Centroids

Default Probability	Function
	1
Non-default	-.065
Default	2.420

Note: Unstandardised canonical discriminant functions evaluated at group means.

Table 13 displays the average Z value of the discriminant function. It indicates that the value of functions at group centroids for non-default firms is at -0.065, and the centroids for default firms is at 2.420. Therefore, when the Z-value in Equation (8) of the firm is negative, the firm is predicted to be a non-default firm, while when the Z-value obtained is positive, the firm is predicted to default.

Table 14. Classification Results

			Predicted Group Membership		Total
			Non-default	**Default**	
Original[a]	Count	Non-default	174	13	187
		Default	2	3	5
	%	Non-default	93.0	7.0	100.0
		Default	40.0	60.0	100.0
Cross-validated[b,c]	Count	Non-default	69	1	70
		Default	2	0	2
	%	Non-default	98.57	1.43	100
		Default	100	0	100

a. 92.2% of original grouped cases were correctly classified.
b. Cross-validation is done only for those cases in the analysis. In cross-validation, each case is classified by the functions derived from all cases other than that case.
c. 92.2% of cross-validated grouped cases are correctly classified.

Table 14 summaries the performance of the discriminant function (8) to classify the two groups, default, and non-default firms, based on the trained data. Out of 187 cases, 174 were correctly predicted as non-default, and 13 were incorrectly predicted. In the meantime, only three were correctly predicted for the default group out of 5. This produces a higher accuracy of about 93% for the non-default group compared to 60.0% for the default group.

3.5 Default Risk Prediction

The default and non-default of the selected firms are predicted at first using the KMV-Merton model and next using the discriminant function model in (8). This discriminant function represents the adaptation of the KMV-Merton model into the DA. Afterwards, the performance of each prediction is compared according to the percentage of accuracy, Type I and Type II errors calculated based on Equations (5), (6), and (7). The results are presented in Tables 15 and 16.

Table 15. Prediction Results

Model			Predicted		Total
			Non-default	**Default**	
Discriminant Function	Actual	Non-default	**47**	1	48
		Default	22	**2**	24
KMV-Merton	Actual	Non-default	**46**	2	48
		Default	24	**0**	24

Table 16. The accuracy, Type I, and Type II errors

	KMV-Merton Model	Discriminant Function
Accuracy (%)	63.89	68.06
Type I error (%)	100	91.67
Type II error (%)	4.17	2.08

Tables 15 and 16 show that both the KMV-Merton model and the discriminant function can predict the non-default group well compared to the default group. This is shown as the percentage of incorrectly predicted for the non-default group (Type II error) is extremely low than the default group (Type I error). This is contrary to the study done by [39] where Type I error is found lower than Type II error. Higher Type I error is due to the unequal data between the non-default and default groups that lead to a potential bias in the data-trained model, as shown in Table 14. Therefore, the discriminant function tends to be more biased in predicting the non-default group and causes high errors in the default group prediction. Overall, the accuracy for both the KMV-Merton model and the discriminant function is low. However, the discriminant function gives higher accuracy in prediction than the KMV-Merton model. This means that the addition of two financial ratios (X_8 and X_9) into a discriminant function may improve the KMV-Merton model performance to predict default risk.

4. Conclusion

Several studies have been conducted previously in the field of firm default prediction. In this study, the non-default and default firms are predicted quarterly. It gives a more detailed focus on the firm's financial performance as it evaluates through quarters. Essentially, the default prediction is done based on the combination of financial ratios and default probability of the KMV-Merton model in the discriminant analysis (DA) model.

This combination leads to a discriminant function with net profit margin (NPM) and return on asset (ROA) as its determinant variables. NPM and ROA are found significant (p-value = 0) in discriminating the default/non-default of the firms with Wilks's lambda = 0.941, F(1,190) = 11.982 for NPM and Wilks's lambda = 0.883, F(1,190) = 25.116 for ROA. Although the return on capital employed ratio is also found significant, it is detected with multicollinearity in the pooled within-groups matrices test.

In the discriminant analysis, the discriminant function failed to classify the default group well due to the 60% accuracy compared to the 93% accuracy of classifying the non-default group. This may be due to the limited and imbalanced data on default firms in the data training process. This makes the discriminant analysis model biased in making predictions to a majority non-default group. Plus, the eigenvalue and canonical correlation of the discriminant function are low at 0.158 and 0.370, respectively. However, the discriminant function could be better as it can discriminate the groups based on Wilks' lambda = 0.863 and p-value at 0.000.

In addition, we found that the accuracy of predicting default using the combination of the KMV-Merton model and financial ratios in the DA (68% accuracy) is slightly higher than using the KMV-Merton model alone (63% accuracy). This is also supported by the lower values of Type 1 and Type II errors for the discriminant functions compared to the KMV-Merton model. Hence, the adaptation of KMV-Merton's default probabilities and financial ratios in the DA model is said to be able to improve the firms' default prediction.

In future research, the model could be improved as more data on default firms are included in the samples. Nevertheless, data acquisition on default firms has been limited due to several factors. In the alternative, the group classification may be redefined in future works to obtain an equal distribution for each group.

Acknowledgements

The authors gratefully acknowledge the Universiti Teknologi MARA (UiTM), Negeri Sembilan branch.

Conflict of Interest

The authors declare no conflict of interest in the subject matter or materials discussed in this manuscript.

References

[1] M. S. Tosun and S. Yildiz, "How Does Aggregate Tax Policy Uncertainty Affect Default Risk?," *J. Risk Financ. Manag.*, vol. 13, no. 12, pp. 1–17, 2020, doi: 10.3390/jrfm13120319.

[2] J. Welburn and A. Strong, "Estimating the Impact of COVID-19 on Corporate Default Risk," 2020. doi: 10.7249/wra173-2.

[3] L. Zeng, W. Lau, and E. N. A. Bahri, "Can the Modified ESG-KMV Logit Model Explain the Default Risk of Internet Finance Companies?," *Front. Environ. Sci.*, vol. 10, 2022, doi: 10.3389/fenvs.2022.961239.

[4] B. Christoffersen, D. Lando, and S. B. Nielsen, "Estimating Volatility in the Merton Model: The KMV Estimate Is Not Maximum Likelihood," *Math. Financ.*, vol. 32, pp. 1214–1230, 2022, doi: 10.1111/mafi.12362.

[5] X. Cheng, Z. Sun, and W. Bao, "Study on credit risk of real estate industry based on genetic algorithm KMV model," *J. Phys. Conf. Ser.*, vol. 1629, no. 1, p. 12072, 2020, doi: 10.1088/1742-6596/1629/1/012072.

[6] S. Anthony and L. Allen, *Credit risk measurement: new approaches to value at risk and other paradigms(google pdf book)*. New York: John Wiley & Sons, Inc., 2002.

[7] S. T. Bharath and T. Shumway, “Forecasting Default with the Merton Distance to Default Model,” *Rev. Financ. Stud.*, vol. 21, no. 3, pp. 1339–1369, May 2008, doi: 10.1093/rfs/hhn044.
[8] S. Aktan, “Early Warning System for Bankruptcy: Bankruptcy Prediction,” *Thesis*, pp. 1–188, 2011.
[9] V. Gupta, “An Empirical Analysis of Default Risk for Listed Companies in India: A Comparison of Two Prediction Models,” *Int. J. Bus. Manag.*, vol. 9, no. 9, pp. 223–234, 2014, doi: 10.5539/ijbm.v9n9p223.
[10] Z. Khemais, D. Nesrine, and M. Mohamed, “Credit Scoring and Default Risk Prediction: A Comparative Study between Discriminant Analysis & Logistic Regression,” *Int. J. Econ. Financ.*, vol. 8, no. 4, p. 39, 2016, doi: 10.5539/ijef.v8n4p39.
[11] N. T. Duong, D. Thi Thu Ha, and N. Bich Ngoc, “The Application of Discriminant Model in Managing Credit Risk for Consumer Loans in Vietnamese Commercial Bank,” *Asian Soc. Sci.*, vol. 13, no. 2, p. 176, 2017, doi: 10.5539/ass.v13n2p176.
[12] P. Andrikopoulos and A. Khorasgani, “Predicting unlisted SMEs’ default: Incorporating market information on accounting-based models for improved accuracy,” *Br. Account. Rev.*, vol. 50, no. 5, pp. 559–573, 2018, doi: https://doi.org/10.1016/j.bar.2018.02.003.
[13] P. Gaba, S. Arora, A. Lall, V. Khurana, and A. Rathi, “Forecasting and Comparing the Results From Altman Z - Score and Merton Distance To Default Model in the Indian Scenario,” *Int. J. Adv. Res.*, vol. 7, no. 10, pp. 1295–1309, 2019, doi: 10.21474/ijar01/9959.
[14] V. A. Delapedra-Silva, “The Bankruptcy Risk in Infrastructure Sectors: An Analysis From 2006 to 2018,” *Ram Rev. Adm. Mackenzie*, vol. 22, no. 4, 2021, doi: 10.1590/1678-6971/eramf210104.
[15] K. Kohv and O. Lukason, “What Best Predicts Corporate Bank Loan Defaults? An Analysis of Three Different Variable Domains,” *Risks*, vol. 9, no. 29, 2021, doi: 10.3390/risks9020029.
[16] W. H. Beaver, “Financial Ratios As Predictors of Failure,” *J. Account. Res.*, vol. 4, pp. 71–111, Apr. 1966, doi: 10.2307/2490171.
[17] E. I. Altman, “Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy,” *J. Finance*, vol. 23, no. 4, pp. 589–609, Apr. 1968, doi: 10.2307/2978933.
[18] I. Wahyudi, I. Hakiman, and T. Mm, “Analysis of Factors to Predict Financial Distress,” *Int. J. Res. Stud. Publ.*, vol. 10, no. 7, 2020, doi: 10.29322/ijsrp.10.07.2020.p10372.
[19] J. Lord, A. Y. Landry, G. T. Savage, and R. Weech-Maldonado, “Predicting Nursing Home Financial Distress Using the Altman Z-Score,” *Inq. J. Heal. Care Organ. Provis. Financ.*, vol. 57, 2020, doi: 10.1177/0046958020934946.
[20] F. Sun, Y. Chen, Y. Qiu, S. Wang, and S. Liang, “Systematic vs. Stepwise Parameter Optimization for Discriminant Model Development: A Case Study of Differentiating Pinellia Ternata From Pinellia Pedatisecta With Near Infrared Spectroscopy,” *J. Near Infrared Spectrosc.*, vol. 28, no. 5–6, pp. 287–297, 2020, doi 10.1177/0967033520924579.
[21] A. K. Benson, T. Swanepoel, and H. J. van Vuuren, “Firm’s Value Sustainability via Accounting Ratios: The Case of Nigerian Listed Firms,” *J. Econ. Financ. Sci.*, vol. 14, no. 1, 2021, doi: 10.4102/jef.v14i1.529.
[22] İ. E. Ceylan, “The Impact of Firm-Specific and Macroeconomic Factors on Financial Distress Risk: A Case Study From Turkey,” *Univers. J. Account. Financ.*, vol. 9, no. 3, pp. 506–517, 2021, doi: 10.13189/ujaf.2021.090325.
[23] Y. Tong and Z. Serrasqueiro, “Predictions of Failure and Financial Distress: A Study on Portuguese High and Medium-High Technology Small and Mid-Sized Enterprises,” *J. Int. Stud.*, vol. 14, no. 2, 2021, doi: 10.14254/2071-8330.2021/14-2/1.
[24] A. M. A. Bajwa and M. H. Rashid, “Financial Ratios : A Tool for Computing Probability of Corporate Default,” *Audit Account. Rev.*, vol. 1, no. 2, 2021, doi 10.32350/aar.12.04.
[25] H. S. Shetty and T. N. Vincent, “Corporate Default Prediction Model: Evidence From the Indian Industrial Sector,” *Vis. J. Bus. Perspect.*, vol. 28, no. 3, 2021, doi: 10.1177/09722629211036207.
[26] H. Herbowo and E. R. Saputri, “Pengujian Rasio Keuangan Terhadap Financial Distress: Bukti Di Indonesia,” *Wahana J. Ekon. Manaj. Dan Akunt.*, vol. 26, no. 1, pp. 1–15, 2023, doi: 10.35591/wahana.v26i1.805.
[27] D. Malasari, M. Adam, Y. Yuliani, and A. Hanafi, “Financial Ratios and Probability of Default by Using The KMV-Merton Method in the Non-Financial Sector Listed on the Indonesia Stock Exchange,” *Financ. Theory Pract.*, vol. 24, no. 1, 2020, doi: 10.26794/2587-5671-2020-24-1-6-13.
[28] M. Mushafiq, S. A. Sami, M. K. Sohail, and M. I. Sindhu, “Merton-Type Default Risk and

Financial Performance: The Dynamic Panel Moderation of Firm Size," *J. Econ. Adm. Sci.*, vol. 40, no. 2, pp. 168–181, 2022, doi: 10.1108/jeas-09-2021-0181.

[29] X. Bai and Z. Zhao, "An Optimal Credit Scoring Model Based on the Maximum Default Identification Ability for Chinese Small Business," *Discret. Dyn. Nat. Soc.*, vol. 2022, pp. 1–14, 2022, doi: 10.1155/2022/1551937.

[30] R. B. Avery, P. S. Calem, and G. B. Canner, "Consumer credit scoring: Do situational circumstances matter?," *J. Bank. Financ.*, vol. 28, no. 4, pp. 835–856, 2004, doi: 10.1016/S0378-4266(03)00202-4.

[31] M. Rezapour, "Anomaly Detection using Unsupervised Methods : Credit Card Fraud Case Study," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 11, pp. 1–8, 2019.

[32] V. B. Djeundje, J. Crook, R. Calabrese, and M. Hamid, "Enhancing credit scoring with alternative data," *Expert Syst. Appl.*, vol. 163, p. 113766, 2021, doi: https://doi.org/10.1016/j.eswa.2020.113766.

[33] N. M. Yusof, I. Q. Alias, A. J. M. Kassim, and F. L. N. M. Zaidi, "Determining the credit score and credit rating of firms using the combination of KMV-Merton model and financial ratios," *Sci. Technol. Indones.*, vol. 6, no. 3, pp. 105–112, 2021, doi: 10.26554/sti.2021.6.3.105-112.

[34] M. Şahin and Ş. Koç, "A Monte Carlo Simulation Study Robustness of MANOVA Test Statistics in Bernoulli Distribution," *Süleyman Demirel Univ. J. Nat. Appl. Sci.*, vol. 22, no. 3, pp. 1125–1131, 2018, [Online].

[35] E. Puskarczyk, "Application of multivariate statistical methods and artificial neural network for facies analysis from well logs data: An example of Miocene deposits," *Energies*, vol. 13, no. 7, pp. 1–18, 2020, doi: 10.3390/en13071548.

[36] R. C. Merton, "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates," *J. Finance*, vol. 29, no. 2, pp. 449–470, Apr. 1974, doi: 10.2307/2978814.

[37] W. Chun-Ping, Q. Lin-Jing, and L. Jian-Wei, "Commercial bank credit risk measurement based on KMV model studies," in *Proceedings of the 2015 International Conference on Engineering Management, Engineering Education and Information Technology*, 2015, pp. 456–465.

[38] D. A. W. Akma Hidayu, A. J. Ja'izah, and W. T. Wan Mardyatul Miza, *Financial and Management Accounting*, 2nd ed. Malaysia: McGraw-Hill Education Malaysia, 2016.

[39] Y. Zizi, A. J. Alaoui, B. El Goumi, M. Oudgou, and A. El Moudden, "An Optimal Model of Financial Distress Prediction: A Comparative Study Between Neural Networks and Logistic Regression," *Risks*, vol. 9, no. 11, p. 200, 2021, doi: 10.3390/risks9110200.

[40] M. S. Çolak, "A New Index Score for the Assessment of Firm Financial Risks," Turkey, 2019.

Biography of all authors

Picture	Biography	Authorship contribution
	Nur Ain Al-Hameefatul Jamaliyatul is currently a master's student in the field of mathematical sciences at the College of Computing, Informatics, and Mathematics, Universiti Teknologi MARA, Johor Branch, Segamat Campus.	Conducting a research and investigation process, specifically collecting, implementing, analysing, and interpreting the data. Also, writing the initial draft of the article.
	Nurul Afiqah Zainuddin is currently working in the industrial field after completing her degree in mathematics at the College of Computing, Informatics, and Mathematics, Universiti Teknologi MARA (UiTM) Negeri Sembilan Branch, Seremban Campus.	Conducting a research and investigation process, specifically collecting, implementing, analysing, and interpreting the data. Also, writing the initial draft of the article.
	Izza Suraya Zulhazmi is currently working in the industrial field after completing her degree in mathematics at the College of Computing, Informatics, and Mathematics, Universiti Teknologi MARA (UiTM) Negeri Sembilan Branch, Seremban Campus.	Conducting a research and investigation process, specifically collecting, implementing, analysing, and interpreting the data. Also, writing the initial draft of the article.
	Norliza Muhamad Yusof is a senior lecturer at the College of Computing, Informatics, and Mathematics, Universiti Teknologi MARA (UiTM) Negeri Sembilan Branch, Seremban Campus. Most of her studies are in the field of financial mathematics, focusing on default risk measurement.	Design the research work and write the article.
	Muhamad Luqman Sapini is a senior lecturer at the College of Computing, Informatics, and Mathematics, Universiti Teknologi MARA (UiTM) Negeri Sembilan Branch, Seremban Campus. Most of his studies are in the field of applied mathematics, focusing on Dynamical Systems and Topological Data Analysis.	Review and edit the article.