Extracting Common Factors to Classify Companies Listed in the Stock Exchange of Thailand by Using an Accounting Based Model

Krisada Khruachalee
Faculty of Business Administration,
Siam Technology College, Thailand
krisada.stc@gmail.com

Abstract - **The objective of this study is to classify the company listed in the Stock Exchange of Thailand by applying the Principle Component Analysis (PCA) to overcome the multicollinearity problem in the purpose of extracting the common factors used in the Logistic regression model. The data set are gathered from the companies' annual financial statements registered in the Stock Exchange of Thailand since 2002 to 2016. The data set consists of 101 companies which there are only 29 distressed companies that have been defaulted on corporate bonds and there are 72 non-distressed companies that have not been defaulted on corporate bonds. The log-odds score are calculated based on the predicted default probability generated from the Logistic regression whose variables are integrated with the 4 statistically significant common factors. These scores are used to classify the company's risk level. The overall accuracy of the predicted model is 92.3% which we can group the companies' default probability into 5 different categories as very good, good, rather good, rather not good, and not good with a credit score below 475, 475-500, 500-535, 535-570, and above 570 respectively in each of which the percentages of companies contained in each categories are 58.42, 12.87, 4.95, 8.91, and 14.85 respectively.**

Keywords - **Principle Component Analysis, Financial Ratio, Extracting, Classifying**

I. INTRODUCTION

The exposure to credit risk is considered as a significant source of serious problems for both financial institutions and companies especially when economic recession has been discovered. With the development of methodologies used in order to capture the volatility of the companies' assets value to the prospective losses, the vital protective strategies to mitigate risk has been promptly implemented to minimize the harmfulness of company failure in the future.

The general information to be used as the evaluated risk factors is normally the company financial position which usually gathered from their financial statements. Despite the important of company financial statements, there are many studies applied the financial ratios as the predetermined factors in order to evaluate the company's default probability, this is widely known as "the accounting-based model". However, there are also some crucial limitations that the accounting-based model cannot fully capture the total variation of company's assets value. In addition, the financial ratios used in the accounting-based model are internally related with each other. Therefore, it will cause a multicollinearity problem when performing the analysis of regression. With these issues, the evaluations of the accounting-based model are treated as an inferior model. In addition, the model will also provide spurious results.

In this study, we try to improve the performance of the accounting-based model in assessing the default probability by applying the Principal Component Analysis (PCA) to overcome the fundamental problem of multicollinearity in the variable selection process. To customize the general users, we will convert the default probability calculated from the accounting-based model models to a credit score by applying the log-odds scoring process.

II. LITERATURE REVIEWS

With the important of accounting information regarding the measurement of company's market exposure, [1] reviewed the relationship between the accounting information and the market risk exposure. They concluded that the accounting information gathering from the accounting statements formatively related to the market risk indicators. Therefore, the use of accounting statements could estimate the exposure of unlisted securities as well as rating company default probability.

The evolution of credit scoring model obviously started from the developing of [2] that applied the financial ratios calculated from the accounting statements of both the bankruptcy companies and the normally operated companies. These companies information were gathered from Moody's Industrial Manual between the years of 1954 to 1964. The average of each financial ratio had been compared between normal and bankruptcy firms by applying the univariate analysis. These samples are no difference in the companies' characteristic in terms of size and industry. The overall results revealed that the ratio analysis provided a great contribution to explain the company's failure 5-year in advance. However, each financial ratio was unequally able to illustrate the company's failure. They found that the ratio of cash flow to total liabilities performed the best in describing the company's failure.

We obviously saw from many literatures that the study of [3] has often been used to develop the classification model which usually applied the techniques of multiple discriminant analysis. It is typically called "Z-Score Model" that the Logistics regression analysis was implemented. This model mainly used the financial accounting ratios as a tool to classify the company default probability. This model accurately forecasted the company's failure 1 year in advance with 95% of total sample. However, the accuracy would die down to 83% when the model forecasted the company's failure 2 year in advance. The framework of [3] has often been developed in order to match with the general environment of company's failure and countries' specific. In Thailand, the empirical study of [4] tried to examine the model of [5] whose model used to characterize the companies' failure in an emerging market. In addition, the estimated default probabilities from the logistic regression equation were used to calculate the log-odds credit score in order to classify the company's risk level. The results revealed that there were only two variables applying in the Altman's model, where there were the ratios of retain earnings to total assets and the ratio of earnings before interest and tax to total assets. These two variables were statistically significant at 0.05. The overall accuracy of the model was 93.84% by classifying the companies into 5 groups with credit scores as very good, good, rather good, rather not good, and not good respectively.

Unfortunately, the study of [6] claimed that the financial ratios used in the accounting-based model internally related with each other which it will seriously cause a multicollinearity problem. Therefore, the evaluated results from accounting-based model were considered as spurious results which it was treated as an inferior model. There are many studies that proposed the method to solve the multicollinearity problem for example the study of [7] that detected the multicollinearity by applying the method of observing correlation matrix, variance influence factor, and eigen-values of the correlation matrix. Their results found that the method of Principle Component Analysis (PCA) could be used to amend the multicollinearity problem.

III. RESEARCH OBJECTIVES

The intents of this study are to construct a classification model and prescience the company's default probability applying the extension of logistic regression analysis technique proposed by [3]. In addition, we will employ a large amount of the financial accounting ratios that calculated from the company's accounting statements. These ratios are covered the ratios measuring the company's liquidity, efficiency, profitability and leverage.

IV. CONCEPTUAL FRAMEWORK

As we focus closely more on revamping the accounting-based model in assessing the default probability of the companies listed in the Stock Exchange of Thailand, we would gather the accounting information from the company's accounting statements and calculate the financial ratios used in the Logistic regression model.

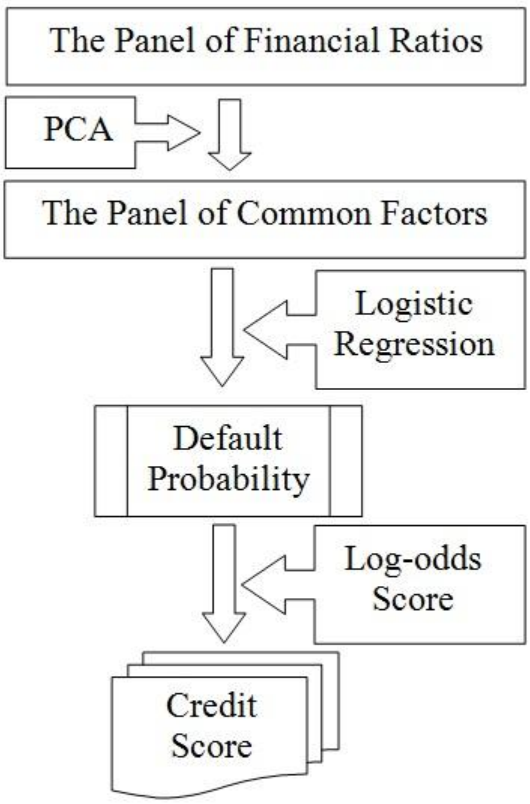

Fig. 1 The Conceptual Framework of Classifying the Companies' Default Probability

In order to get rid of the multicollinearity problem, the Principle Component Analysis (PCA) has been used to extract the common components in the accounting-based model.

- **Dependent Variable:** The event where the company has been default ($y=1$) on the corporate bond and the event where the company normally operates ($y=0$).

- **Independent Variables:** The financial ratios calculated from the company's accounting statements which each value was ideally assumed to capture the behaviour of company's liquidity, efficiency, profitability and leverage.

V. DATA

This empirical study has been conducted to cover the period of January 2002 to December 2016 which is over 14 years of study. The population of this study is the companies listed on the Stock Exchange of Thailand which used to raise fund by issuing a corporate bond since January 2002 to December 2016. However, there are only 29 companies during this period that have been default on corporate bonds. Therefore, we would consider all the defaulted companies in the model construction process for increasing the precision.

In order to construct the classification model, the normally operated companies have been randomly selected from list of the Stock Exchange of Thailand posted on the website by using a simple random sampling. As results, the selected sample comprised of 101 companies listed in the stock exchange of Thailand which particularly consists of 29 distressed companies and 72 non-distressed companies. The monthly company's annual financial statements and their key financial ratios measuring the company's liquidity, efficiency, profitability and leverage have been calculated to be the independent variables for the predictive Z-Score model. The data set has mainly downloaded from the Datastream program and the website of the Securities and Exchange Commission, Thailand (SEC) which they are all publicly available.

VI. RESEARCH METHODOLOGY

In order to extract the common factors, let the variable x denotes the panel of financial ratios changes which $Cov(x) = xx'$ denotes the covariance matrix of the financial ratios changes. The common factors extracted from the panel of financial ratios can be calculated from

$$x = v\lambda^{1/2}F = \Phi F$$

where;

- v denoted the matrix whose columns are the eigenvectors of $Cov(x)$.
- $\lambda^{1/2}$ denotes the diagonal matrix whose elements are the square-roots of the eigenvalues.
- Φ denotes the vectors of principal component loadings, or factors loadings which can also be illustrated as $\Phi = v\lambda^{1/2}$.
- F denotes the vector of principal components.

Estimating the parameters to assess the default probability, the logistic regression analysis has been employed incorporating with the common factors extracted from the data set of company financial ratios.

$$g(x) = \beta_0 + \beta_1 F_1 + \beta_2 F_2 + ... + \beta_n F_n + \varepsilon$$

The number of common factors is based on the first common factors that can explain at least 90 percentage of the total variation of the panel of all financial ratios.

The estimation of company's default probability can be calculated as follow;

$$P\left[y = 1 \middle| g(x)\right] = \frac{e^{g(x)}}{1 + e^{g(x)}}$$

To friendly customize the general users, the log-odd score has been employed to convert the default probability to be a credit score.

$$Log-odds\ Score = \ln\left(\frac{P\left(y=1\middle|g(x)\right)}{P\left(y=0\middle|g(x)\right)}\right)$$

Ranking the score based on the default probability, we can calculate as follow;

$$Credit\ Score = \left(Log-odds\ Score \times 100\right) + 500$$

The company whose credit score is largely greater than 500 is theoretically considered as a highest default probability. For those company whose credit score largely less than 500 is theoretically considered as a lowest default probability (Good Company).

VII. EMPRIRICAL RESULTS

A. Extracting the Common Factors from the Data Set of Accounting Ratios

Applying the principle component analysis to extract the common factors from a data set of various types of the financial ratios, we selectively have 10 common factors that can capture at least 90 percent of the total variation of a data set, which the percentage of the first 10^{th} eigenvalue that can explain the variation of the data set are sorted in descending order shown in the Table I, where the percentage of explanation can be calculated as follow;

$$Explained = \left(D \Big/ \sum_{i=1}^{n} D_i\right) \times 100$$

TABLE I
THE PERCENTAGE EXPLAINED BY THE FIRST 10^{TH} COMMON FACTORS

Common Factor	Explained (%)	Common Factor	Explained (%)
1	25.02414	6	5.953759
2	14.06980	7	5.284338
3	10.85254	8	4.585825
4	9.828711	9	4.301915
5	6.796513	10	3.621327

B. Estimating the Parameters Used in the Logistic Regression Model

After we applied the first 10^{th} common factors and the determined events in the logistic regression model to estimate the parameters, we found that only the first 4 common factors, which are the 1st, 2nd, 3rd, and 10th, are statistically significant at 0.05. Therefore, we selectively applied only the first 4 common factors in the classification model to classify the company's probability of default which can be illustrated as.

$$\hat{g}(x) = -1.230 + 21.257F_1 + 2.379F_2 - 4.703F_3 - 2.639F_{10}$$

Analysing the classification model with 4 statistically significant common factors, we found that the Negelkerke R-Square and the Cox & Snell R-Square are 90.7% and 66.8% respectively, which they are considerably high level of explanation. Accordingly, we can imply that 66.8% of the total variation of the financial ratios' data set can be explained by the Logistic regression with 4 statistically significant common factors. In addition, the Chi-square of Hosmer and Lemeshow Test, which it is used to test the suitability of the predicted model, showed the P-value at 0.983. This value is considerably greater than the statistical level of significant 0.05 which we can imply that the predicted model with the 4 statistically significant common factors suitably classify the default probability for companies listed in the Stock Exchange of Thailand.

According to the model's suitability results tested above, we can calculate the company's default probability as the following equation;

$$P\left[y=1\middle|\hat{g}(x)\right] = \frac{e^{\hat{g}(x)}}{1+e^{\hat{g}(x)}}$$

Following the procedures in ranking the company based on their default probability, we can obtain the company's credit score from these equations;

$$\hat{Log} - odds\ Score = \ln\left(\frac{P\left(y=1\middle|\hat{g}(x)\right)}{P\left(y=0\middle|\hat{g}(x)\right)}\right)$$

$$\hat{Credit}\ Score = \left(\hat{Log} - odds\ Score \times 100\right) + 500$$

TABLE II
NUMBER OF COMPANY CLASSIFIED BY THE LOGISTICS REGRESSION MODEL BASED ON THE COMPANY STATUS

Observed Company's Status	Predicted Company's Status*	
	Default	Non-Default
Default	22	3
Non-Default	2	38
Overall Percentage	92.3	

*cut off point at 0.5

True positive rate (recall) = 0.88
True negative rate = 0.95
False positive rate = 0.12
False negative rate = 0.05

Analyzing the model accuracy, we found that the overall accuracy of the predicted model is 92.3% (60/65). Moreover, the observed 25 companies that used to default on corporate bond can be correctively classified by the predictive model with 88% (22/25) of accuracy (22 companies). On the other hands, the observed 40 companies that normally oblige on corporate bond can be correctively classified by the predictive model with 95% (38/40) of accuracy (38 companies).

TABLE III
THE PERCENTAGE AND NUMBER OF COMPANY LISTED IN THE STOCK EXCHANGE OF THAILAND CLASSIFIED BY THE CREDIT SCORE

Score Range	Criteria	No. of Company	%
> 570	Not good	15	14.85
535 - 570	Rather not good	9	8.91
500 - 535	Rather good	5	4.95
475 - 500	Good	13	12.87
< 475	Very good	59	58.42
Total		**101**	

To rate the company based on the credit scoring process, we can group the companies into 5 different categories which are defined the credit scores as very good, good, rather good, rather not good, and not good with credit score below 475, 475-500, 500-535, 535-570, and above 570 respectively in each of which the percentages of companies contained in each categories are 58.42, 12.87, 4.95, 8.91, and 14.85 respectively.

VIII. DISCUSSION AND CONCLUSION

This study examines the implication of the Logistic regression model whose variables are applying the benefit of the Principle Component Analysis (PCA) in an extracting the common factors from a large data set of the financial ratios. We found from the empirical results that there are only 4 statistically significant common factors that can explain at least 90 percent of the total variation of the data set which comprised of both defaulted and non-defaulted companies listed in the Stock Exchange of Thailand. Applying these 4 statistically significant common factors and the determined events to estimate the company's credit score via the logistic regression model, the overall accuracy of the predicted model is 92.3% which we found that the total observed 25 companies, who used to default on corporate bond, can be correctively classified by the predictive model with 88% of accuracy (22 companies). Moreover, the observed 40 companies that normally oblige on corporate bond can be correctively classified by the predictive model with 95% of accuracy (38 companies). By applying the credit scoring process calculated by the log-odds score, we can group the companies' default probability into 5 different categories as very good, good, rather good, rather not good, and not good with a credit score below 475, 475-500, 500-535, 535-570, and above 570 respectively in each of which the percentages of companies contained in each categories are 58.42, 12.87, 4.95, 8.91, and 14.85 respectively. With these constructive results, the extracting method which is the principle component analysis intuitively provides a great contribution to a classification model which is the Logistic regression model in classifying the company's default probability for the companies listed in the Stock Exchange of Thailand.

REFERENCES

(Arranged in the order of citation in the same fashion as the case of Footnotes.)

[1] Watts, R.L. and Zimmerman, J.L. (1986). "Positive Accounting Theory". New Jersey: Prentice-Hall.

[2] Beaver, W.H. (1966). "Financial Ratio as Predictor of Failure". Journal of Accounting Research, Vol. 4(3), pp. 71-111.

[3] Altman, E. (1968). "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy". Journal of Finance, Vol. 23(4), pp. 589-609.

[4] Krisada, K., Vichit, L., and Jirawan, J. (2001). "Classification of Companies Listed in Stock Exchange of Thailand by Using Logistic Regression Analysis". Burapha Journal of Business Management, Vol. 5(2), pp. 19-32.

[5] Altman, E., Hartzell, J., and Peck, M. (1995). "A Scoring System for Emerging Market Corporate Bonds: Saloman Brothers High Yield Research". New York.

[6] Shumway, T. (2001). "Forecasting bankruptcy more accurately: A simple hazard model". Journal of Business, 74, pp. 101-124.

[7] Alibuhtto and Peiris (2015). "Principle Component Regression for solving Multicollinearity Problem". Conference Proceeding on the 5th International Symposium 2015, SEUSL.