ADBI Working Paper Series

A COMPREHENSIVE METHOD FOR THE CREDIT RISK ASSESSMENT OF SMALL AND MEDIUM-SIZED ENTERPRISES BASED ON ASIAN DATA

Naoyuki Yoshino and
Farhad Taghizadeh-Hesary

No. 907
December 2018

Asian Development Bank Institute

Naoyuki Yoshino is Dean of the Asian Development Bank Institute and Professor Emeritus at Keio University, Tokyo. Farhad Taghizadeh-Hesary is Assistant Professor, Faculty of Political Science and Economics, Waseda University, Tokyo.

The Working Paper series is a continuation of the formerly named Discussion Paper series; the numbering of the papers continued without interruption or change. ADBI's working papers reflect initial ideas on a topic and are posted online for discussion. Some working papers may develop into other forms of publication.

Suggested citation:

Yoshino, N. and F. Taghizadeh-Hesary. 2018. A Comprehensive Method for the Credit Risk Assessment of Small and Medium-Sized Enterprises Based on Asian Data. ADBI Working Paper 907. Tokyo: Asian Development Bank Institute. Available: https://www.adb.org/publications/comprehensive-method-credit-risk-assessment-sme-based-asian-data

Please contact the authors for information about this paper.

Email: nyoshino@adbi.org; farhad@aoni.waseda.jp

Asian Development Bank Institute
Kasumigaseki Building, 8th Floor
3-2-5 Kasumigaseki, Chiyoda-ku
Tokyo 100-6008, Japan

Tel: +81-3-3593-5500
Fax: +81-3-3593-5571
URL: www.adbi.org
E-mail: info@adbi.org

Abstract

Due to the asymmetry of information between borrowers that are small- or medium-sized enterprises (SMEs) and lenders (banks), many banks are considering this sector as a risky sector. It is crucial for banks to be able to distinguish healthy from risky companies in order to reduce their nonperforming assets in the SME sector. If they can do this, lending and financing to SMEs through banks will be easier with lower collateral requirements and lower interest rates. In this paper, we provide a scheme originally developed by Yoshino and Taghizadeh-Hesary (2014) for assigning credit ratings to SMEs by employing two statistical analysis techniques—principal component analysis and cluster analysis—applying 11 financial ratios of 1,363 SMEs in Asia. If used by the financial institutions, this comprehensive and efficient method could enable banks and other lending agencies around the world, and especially in Asia, to group SME customers based on financial health, adjust interest rates on loans, and set lending ceilings for each group.

Keywords: Asian economies, SME credit rating, SME financing

JEL Classification: G21, G24, G32

Contents

1. INTRODUCTION

Because of the significance of SMEs to Asian national economies, it is important that ways be found to provide them with stable access to inexpensive finance. Asian economies are often characterized as having bank-dominated financial systems and capital markets, in particular venture capital markets, that are not well developed. This means that banks are the main source of financing. Although the soundness of banking systems has improved significantly since the 1997/98 Asian financial crisis, banks have been cautious about lending to SMEs, even though such enterprises account for a large share of economic activity. Start-up companies and riskier SMEs, in particular, are finding it increasingly difficult to borrow money from banks because of strict Basel III capital requirements (Yoshino and Hirano 2011, 2013; Yoshino 2012). Most recently, following the subprime mortgage crisis and global financial crisis, banking sectors in developed and developing countries have become more cautious in lending to riskier sectors, including SMEs. Driver and Muñoz-Bugarin (2018) estimated the effect of external financial constraints on access to finance in manufacturing companies based on size of company. They found that only for the crisis period were financial constraints important for large firms, and then only for periods of falling business optimism. By contrast, small firms experienced continuous constraint, and many of them were going bankrupt during the crisis.

On the other hand, when looking at the nonperforming loans (NPL) structure in most Asian countries, the NPL ratio of SME loans is usually higher compared to the NPL ratio of total loans and NPL ratio of loans to larger enterprises. The main reason is that SMEs are in essence riskier investments; they have fewer assets and they usually have less credit history (Beck 2007). Most large enterprises are stock-listed companies, and hence they have to follow certain auditing rules by external auditors so that their financial statements can be seen as trustworthy. However, the majority of SMEs do not keep their financial statements updated, they are not necessarily using external auditors as per large enterprises, and many of them are keeping more than one accounting book. Therefore, when SMEs are applying for bank loans, an asymmetry of information exists between lenders (banks) and borrowers (SMEs) if the banker wants to rely just on the self-declared financial statements of the borrower. To cover the risk that is associated with SMEs, banks usually ask for collateral, in a majority of the cases in the form of real estate, and charge the SMEs higher interest rates. Many SMEs cannot afford to provide the collateral or pay high interest rates, however, so this is a major constraint for SME financing that endangers their growth. To reduce the information asymmetry between SMEs and banks, an optimal solution is to accumulate SME data in a nationwide scale database and then employ credit-rating and credit-scoring techniques on them; in this way banks could compare the status of the specific SME that asked for the loan with data from a large number of SMEs from the same industry and the same geographical location. The importance of credit ratings has increased recently after the global financial crisis and because of increased capital requirements for banks. Hence, an efficient credit-rating scheme that rates SMEs based on their financial health would help banks to lend money to SMEs in a more rational way while at the same time reduce the risk to banks.

Various credit-rating indexes such as Standard and Poor's (S&P) rate large enterprises. By looking at a large enterprise's credit rating, banks can decide to lend them up to a certain amount. For SMEs, the issue is more complicated as there are no comparable ratings. Nevertheless, there is a useful model in Japan. In a governmentsupported project, 51 credit guarantee corporations[1] collected data from Japanese SMEs.[2] These data are now stored at a private corporation called Credit Risk Database (CRD) (Kuwahara et al. 2016). If similar systems could be established in other parts of Asia to accumulate and analyze credit risk data, and to measure each SME's credit risk accurately, banks and other financial institutions could use such data to categorize their SME customers based on their financial health. SMEs would also benefit as they could both raise funds from the banks more easily and gain access to the debt market by securitizing their claims. In the absence of a nationwide comprehensive SME credit-risk database, it should be important for banks to start accumulating SME data by themselves and do credit risk assessment on them by applying credit-rating techniques. For the credit rating of SMEs, Yoshino and Taghizadeh-Hesary (2014) developed a method for the credit risk analysis using statistical analysis techniques (principal component analysis and cluster analysis) that can be helpful in facilitating bank financing. The background of this method and an empirical analysis using this method are provided in this chapter.

In Section II, we provide a literature survey on credit risk assessment of enterprises based on their sizes and most recently developed models and methods for the credit rating of SMEs. In Section III, we discuss the model developed for credit risk assessment of SMEs using Asian data. Section IV provides concluding remarks.

2. LITERATURE REVIEW

Credit ratings are opinions expressed in terms of ordinal measures, reflecting the current financial creditworthiness of issuers such as governments, firms, and financial institutions. These ratings are conferred by rating agencies—such as Fitch Ratings, Moody's, and S&P—and may be regarded as a comprehensive evaluation of an issuer's ability to meet its financial obligations in full and on time. Hence, ratings play a crucial role by providing participants in financial markets with useful information for financial planning. To conduct rating assessments of large corporations, agencies resort to a broad range of financial and nonfinancial pieces of information, including domain experts' expectations. Rating agencies usually provide general guidelines on their rating decision-making process, but detailed descriptions of the rating criteria and the determinants of banks' ratings are generally not provided (Orsenigo and Vercellis 2013). In search of more objective assessments of the creditworthiness of large corporate and financial institutions, there has been a growing body of research into the development of reliable quantitative methods for automatic classification according to institutions' financial strength.

Extensive empirical research devoted to analyzing the stability and soundness of large corporations dates back to the 1960s. Ravi Kumar and Ravi (2007) provided a comprehensive survey of the application of statistical and intelligent techniques to predicting the likelihood of default among banks and firms. Despite its obvious relevance, however, the development of reliable quantitative methods for the prediction of large corporations' credit ratings has only recently begun to attract strong interest. These studies are mainly conducted within two broad research strands focusing on

[1] Credit guarantee corporations (funds) have a cost, which is paid by SMEs in the form of a guarantee premium. Based on the credit score that the CRD gives to each SME, the credit guarantee corporation charges that SME. If the SME has lower risk, then the payable premium is lower, and if the SME is riskier, the premium rate that the SME needs to pay to be guaranteed by the credit guarantee corporation is higher (Yoshino and Taghizadeh-Hesary, 2018).

[2] See conclusion for more info.

statistical and machine-learning techniques, and may address both feature selection and classification. Poon, Firth, and Fung (1999) developed logistic regression models for predicting financial strength ratings assigned by Moody's, using bank-specific accounting variables and financial data. Factor analysis was applied to reduce the number of independent variables and retain the most relevant explanatory factors. The authors showed that loan provision information and risk and profitability indicators added the greatest predictive value in explaining Moody's ratings. Yoshino, Taghizadeh-Hesary, and Nili (2015) used two statistical analysis techniques on various financial variables taken from bank statements for the classification and credit rating of 32 Iranian banks. The underlying logic of both techniques—principal component analysis (PCA) and cluster analysis—is dimension reduction, that is, summarizing information on numerous variables in just a few variables. While the two techniques achieved this in different ways, their results both classified 32 banks into two groups and sorted them based on their credit ratings.

While the aforementioned examples are for credit ratings of large corporate and financial institutions, the story is different for SMEs because of the lack of reliable data in addition to difficulties in collecting them and the profitability of loans. The literature on credit rating and credit risk assessment of SMEs is scarce. Angilella and Mazzù (2015) mention that the obstacles for financing SMEs increase if SMEs are innovative. In this case, financial data are insufficient or even unreliable. Therefore, credit risk assessment will be mainly based on qualitative criteria (soft information). In their paper, they provided a multicriteria credit risk model named ELECTRE-TRI through Monte Carlo simulations. Li et al. (2016), based on traditional statistical methods and recent artificial intelligence (AI) techniques, proposed a hybrid model that combines the logistic regression approach and artificial neural networks (ANN) using data of Finnish SMEs. Their results suggest that the proposed ANN/logistic hybrid model is more accurate than either of the initial models (ANN or logistic regression) on its own. Fernandes and Artes (2016), from a data set with the localization and default information of 9 million Brazilian SMEs, proposed a measure of the local risk of default based on the application of ordinary kriging. They included this variable in logistic credit-scoring models as an explanatory variable. Their model has shown better performance when compared to models without this variable. Altman, Esentato, and Sabato (2018) mention that assessments of credit risk must be convincing and objective, providing complements to the traditional rating agency process. In a study on a sample of Italian SMEs, they developed a model to assess SMEs' creditworthiness and tested it on the companies that have issued mini-bonds so far. Their findings confirm that the amount of information asymmetry is still high in the market and is affecting the level of risk/return trade-off, potentially reducing the number of investors and small businesses that would be interested in using this new channel to fund their business growth. Yoshino and Taghizadeh-Hesary (2014) developed a model for credit rating of SMEs, employing two statistical analysis techniques—PCA and cluster analysis—to analyze the credit risks of a sample of Iranian SMEs by using their financial variables. The comprehensive method that they developed is novel, and their test results show that the accuracy of this model that considers different aspects of SMEs (leverage, liquidity, profitability, coverage, and activity) is higher compared to conventional probit/logit and other binary response models.

3. CREDIT RISK ANALYSIS OF SMES USING ASIAN DATA

In this section, we present an efficient and comprehensive scheme for rating the creditworthiness of SMEs that was developed by Yoshino and Taghizadeh-Hesary (2014). First, they examined various financial ratios that described the characteristics of SMEs. The model that they developed enables banks to categorize their SME customers into different groups based on their financial health. The data for their statistical analysis were provided by an Iranian bank for 1,363 SMEs.

3.1 Selection of the Variables

A large number of possible ratios have been identified as useful in predicting a firm's likelihood of default. Chen and Shimerda (1981) show that out of more than 100 financial ratios, almost 50% were found to be useful in at least one empirical study. Some scholars have argued that quantitative variables are not sufficient to predict SME defaults and that including qualitative variables—such as the legal form of the business, the region where the main business is carried out, and industry type—improves a model's predictive power (Lehmann 2003; Grunert, Norden, and Weber 2004). However, the data Yoshino and Taghizadeh-Hesary (2014) used were based on firms' financial statements, which do not contain such qualitative variables.

Altman and Sabato (2007) and Yoshino and Taghizadeh-Hesary (2014; 2015) proposed five categories to describe a company's financial profile: (i) liquidity, (ii) profitability, (iii) leverage, (iv) coverage, and (v) activity. For each of these categories, they created a number of financial ratios identified in the literature. Table 1 shows the financial ratios selected for this survey.

Table 1: Examined Variable

No.	Symbol	Definition	Category
1	Equity_TL	Equity (book value)/total liabilities	Leverage
2	TL_Tassets	Total liabilities/total assets	
3	Cash_Tassets	Cash/total assets	Liquidity
4	WoC_Tassets	Working capital/total assets	
5	Cash_Sales	Cash/net sales	
6	EBIT_Sales	Ebit/sales	Profitability
7	Rinc_Tassets	Retained earnings/total assets	
8	Ninc_Sales	Net income/sales	
9	EBIT_IE	Ebit/interest expenses	Coverage
10	AP_Sales	Account payable/sales	Activity
11	AR_TL	Account receivable/total liabilities	

Notes: Retained earnings refers to the percentage of net earnings not paid out as dividends, but retained by the company to be reinvested in its core business or to pay debt; it is recorded under shareholders' equity on the balance sheet. Ebit refers to earnings before interest and taxes. Account payable refers to an accounting entry that represents an entity's obligation to pay off a short-term debt to its creditors; the accounts payable entry is found on a balance sheet under current liabilities. Account receivable refers to money owed by customers (individuals or corporations) to another entity in exchange for goods or services that have been delivered or used, but not yet paid for; receivables usually come in the form of operating lines of credit and are usually due within a relatively short time period, ranging from a few days to 1 year.

Source: Yoshino and Taghizadeh-Hesary (2014 and 2015).

The firms considered as being non-sound in this study are those whose risk-weighted assets are greater than their shareholders' equity.

In the next stage, two statistical techniques were used: PCA and cluster analysis. The underlying logic of both techniques is dimension reduction—summarizing information on multiple variables into just a few variables—but they achieve this in different ways. PCA reduces the number of variables into components (or factors). Cluster analysis reduces the number of SMEs by placing them in small clusters. In this empirical work, Yoshino and Taghizadeh-Hesary (2014) used components (factors) that are the result of PCA and then ran the cluster analysis to group the SMEs.

3.2 Principal Component Analysis

PCA is a standard data-reduction technique that extracts data, removes redundant information, highlights hidden features, and visualizes the main relationships that exist between observations.[3] PCA is a technique for simplifying a data set by reducing multidimensional data sets to lower dimensions for analysis. Unlike other linear transformation methods, PCA does not have a fixed set of basis vectors. Its basis vectors depend on the data set, and PCA has the additional advantage of indicating what is similar and different about the various models created (Bruce-Ho and Dash-Wu 2009). Through this method, Yoshino and Taghizadeh-Hesary (2014) reduced the 11 variables listed in Table 1 to determine the minimum number of components that can account for the correlated variance among SMEs.

To examine the suitability of these data for factor analysis, the Kaiser–Meyer–Olkin (KMO) test and Bartlett's test of sphericity were performed. KMO is a measure of sampling adequacy that indicates the proportion of common variance that might be caused by underlying factors. High KMO values (larger than 0.60) generally indicate that factor analysis may be useful, which is the case in this study as the KMO value is 0.71. If the KMO value is less than 0.5, factor analysis will not be useful. Bartlett's test of sphericity indicates whether the correlation matrix is an identity matrix, indicating that variables are unrelated. A significance level less than 0.05 indicates that there are significant relationships among the variables, which is the case in this study as the significance of Bartlett's test is less than 0.001.

Next, the number of factors to be used in the analysis was determined. Table 2 reports the estimated factors and their eigenvalues. Only those factors accounting for more than 10% of the variance (eigenvalues >1) are kept in the analysis. As a result, only the first four factors were finally retained. Taken together, Z1 through Z4 explain 71% of the total variance of the financial ratios.

In running the PCA, direct oblimin rotation was used. Direct oblimin is the standard method to obtain a non-orthogonal (oblique) solution—that is, one in which the factors are allowed to be correlated. To interpret the revealed PCA information, the pattern matrix must then be studied. Table 3 presents the pattern matrix of factor loadings by the use of the direct oblimin rotation method, where variables with large loadings, absolute value (>0.5) for a given factor, are highlighted in bold.

[3] PCA can also be called the Karhunen–Loève transform (KLT), named after Kari Karhunen and Michel Loève.

Table 2: Total Variance Explained

Component	Eigenvalues	% of Variance	Cumulative Variance %
Z1	**3.30**	**30.00**	**30.00**
Z2	**2.19**	**19.90**	**49.90**
Z3	**1.25**	**11.38**	**61.28**
Z4	**1.08**	**9.78**	**71.06**
Z5	0.94	8.56	79.62
Z6	0.75	6.79	86.41
Z7	0.56	5.09	91.50
Z8	0.48	4.36	95.86
Z9	0.32	2.87	98.73
Z10	0.13	1.14	99.87
Z11	0.09	0.13	100.00

Source: Yoshino and Taghizadeh-Hesary (2014).

Table 3: Factor Loadings of Financial Variables after Direct Oblimin Rotation

Variables (Financial Ratios)	Component			
	Z1	**Z2**	**Z3**	**Z4**
Equity_TL	0.009	0.068	0.113	**0.705**
TL_Tassets	–0.032	**–0.878**	0.069	–0.034
Cash_Tassets	–0.034	–0.061	**0.811**	0.098
WoC_Tassets	–0.05	**0.762**	0.044	0.179
Cash_Sales	**–0.937**	0.021	0.083	0.009
EBIT_Sales	**0.962**	0.008	0.024	–0.004
Rinc_Tassets	0.014	**0.877**	0.015	–0.178
Ninc_Sales	**0.971**	–0.012	0.015	0.014
EBIT_IE	0.035	0.045	**0.766**	–0.098
AP_Sales	**–0.731**	–0.017	–0.037	–0.016
AR_TL	0.009	–0.041	–0.104	**0.725**

Notes: The extraction method was principal component analysis. The rotation method was direct oblimin with Kaiser normalization.

Source: Yoshino and Taghizadeh-Hesary (2014).

As can be seen in Table 3, the first component, Z1, has four variables with an absolute value (>0.5), of which two are positive (ebit/sales and net income/sales) and two are negative (cash/net sales and account payable/sales). For Z1, the variables with large loadings are mainly net income and earnings. Hence, Z1 generally reflects the net income of an SME. As this factor explains the most variance in the data, it is the most informative indicator of an SME's overall financial health. Z2 reflects short-term assets. This component has three major loading variables: (i) liabilities/total assets, which is negative, meaning that an SME has few liabilities and mainly relies on its own assets; (ii) working capital/total assets, which is positive, meaning that an SME has short-term assets; and (iii) retained earnings/total assets, which is positive, meaning that an SME has some earnings that it keeps with the company or in the bank. These three variables indicate an SME whose reliance on borrowings is small and which is rich in working capital and retained earnings, and therefore has plenty of short-term assets. Z3 reflects the liquidity of SMEs. This factor has two variables with large loadings (cash/total assets and ebit/interest expenses), both with positive values, which shows an SME that is cash-rich and has high earnings. Hence, it mainly reflects an SME's liquidity. The last factor, Z4, reflects capital. This factor has two variables with large loadings, both with positive values: equity (book value)/total liabilities and accounts receivable/total liabilities, meaning an SME with few liabilities that is rich in equity.

Table 4: Component Correlation Matrix

Component	Z1	Z2	Z3	Z4
Z1	1	0.037	–0.031	–0.005
Z2	0.037	1	0.106	0.102
Z3	–0.031	0.106	1	0.033
Z4	–0.005	0.102	0.033	1

Note: The extraction method is principal component analysis. The rotation method is direct oblimin with Kaiser normalization.

Source: Yoshino and Taghizadeh-Hesary (2014).

Table 4 shows the correlation matrix of the components and shows there is no correlation among these four components. This means a regular orthogonal rotation approach could be used to force an orthogonal rotation, although in this empirical work an oblique rotation method was used, which still provided basically an orthogonal rotation factor solution because these four components are not correlated with each other and are distinct entities.

Figure 1 shows the distribution of the four components (Z1, Z2, Z3, and Z4) for Group A, which comprises financially sound SMEs, and Group B, which comprises non-sound SMEs.

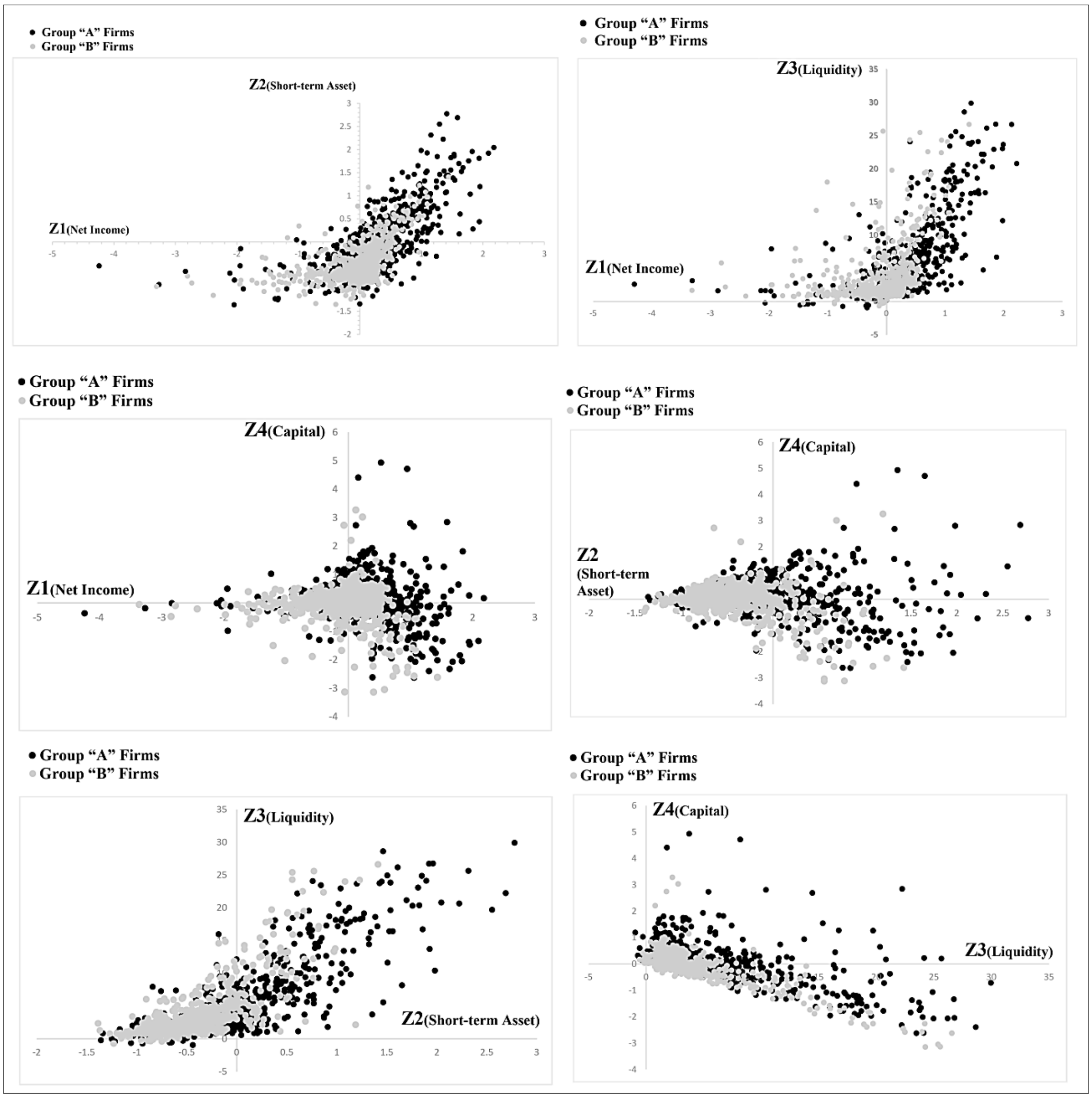
It is clear from all six graphs in this figure that Group A SMEs can generally be found in the positive areas of the graphs and Group B SMEs in the negative areas. This is evidence that these four defined components (Z1, Z2, Z3, and Z4) are able to separate SMEs, suggesting they represent a good measure for showing the financial soundness of SMEs.[4]

3.3 Cluster Analysis and Classification of SMEs

In this section we take four components obtained in the previous section and identify those SMEs that have similar traits. The next step is to generate the clusters and place the SMEs in distinct groups. To do this, cluster analysis technique is employed, which organizes a set of data into groups so that observations from a group with similar characteristics can be compared with those from a different group (Martinez and Martinez 2005). The result of the cluster analysis tells us how much each individual SME is close to others, and it looks at the distance between two companies based on their financial statements. If they are close to each other in the cluster analysis, it means their financial statements are similar; if two SMEs are different, it means their financial statements are completely different. Thus, the similarities and differences between two companies are statistically analyzed.

[4] The number of significant components is based on the data set. It means that when applying this method on another data set, perhaps two, three, or more components become statistically significant.

Figure 1: Distribution of Factors for SME Groups A and B



Group A = sound SMEs, group B = non-sound SMEs. The firms considered to be non-sound in this study have risk-weighted assets greater than their shareholders' equity.

Source: Yoshino and Taghizadeh-Hesary (2014).

In this case, SMEs were organized into distinct groups according to the four components derived from the PCA used in the previous section. Cluster analysis techniques can themselves be broadly grouped into three classes: hierarchical clustering, optimization clustering, and model-based clustering.[5] In this empirical work,

[5] The main difference between the hierarchical and optimization techniques is that in hierarchical clustering the number of clusters is not known beforehand. The process consists of a sequence of steps where two groups are either merged (agglomerative) or divided (divisive) according to the level of similarity. Eventually, each cluster can be subsumed as a member of a larger cluster at a higher level of similarity. The hierarchical merging process is repeated until all subgroups are fused into a single cluster (Martinez and Martinez 2005). Optimization methods, on the other hand, do not necessarily form hierarchical classifications of the data as they produce a partition of the data into a specified or predetermined number of groups by either minimizing or maximizing some numerical criterion (Feger and Asafu-Adjaye 2014).

hierarchical clustering was used, which is the most prevalent of the three methods cited in the literature. This produced a nested sequence of partitions by merging (or dividing) clusters. At each stage of the sequence, a new partition is optimally merged (or divided) from the previous partition according to some adequacy criterion. The sequence of partitions ranges from a single cluster containing all the individuals to a number of clusters (*n*) containing a single individual. The series can be described by a tree display called a dendrogram (Figure 2). Agglomerative hierarchical clustering proceeds by a series of successive fusions of the *n* objects into groups. By contrast, divisive hierarchical methods divide the *n* individuals into progressively finer groups. Divisive methods are not commonly used because of the computational problems they pose (Everitt, Landau, and Leese 2001; Landau and Chis Ster 2010). As described below, the average linkage method was used, which is a hierarchical clustering technique.

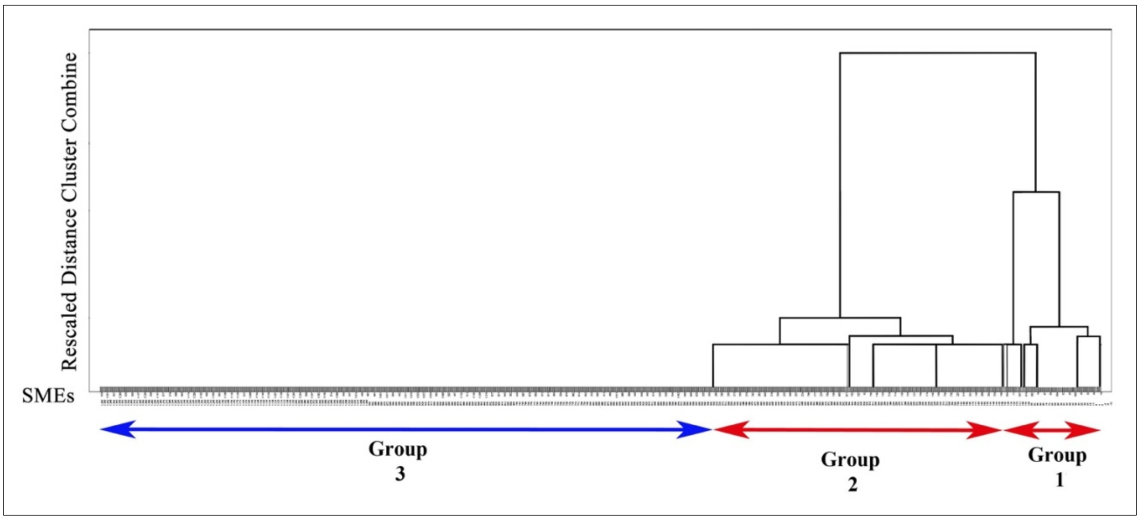
3.3.1 The Average Linkage Method

The average linkage method defines the distance between clusters as the average distance from all observations in one cluster to all points in another cluster. In other words, it is the average distance between pairs of observations, where one is from one cluster and one is from the other. The average linkage method is relatively robust and also takes the cluster structure into account (Martinez and Martinez 2005; Feger and Asafu–Adjaye 2014; Yoshino and Taghizadeh-Hesary 2014, 2015; Yoshino et al. 2016). The basic algorithm for the average linkage method can be summarized in the following manner:

- *N* observations start out as *N* separate groups. The distance matrix *D* = (*dij*) is searched to find the closest observations, for example, Y and Z.
- The two closest observations are merged into one group to form a cluster (YZ), producing *N* − 1 total groups. This process continues until all observations are merged into one large group.

Figure 2 shows the dendrogram that results from this hierarchical clustering.

Figure 2: Dendrogram Using Average Linkage



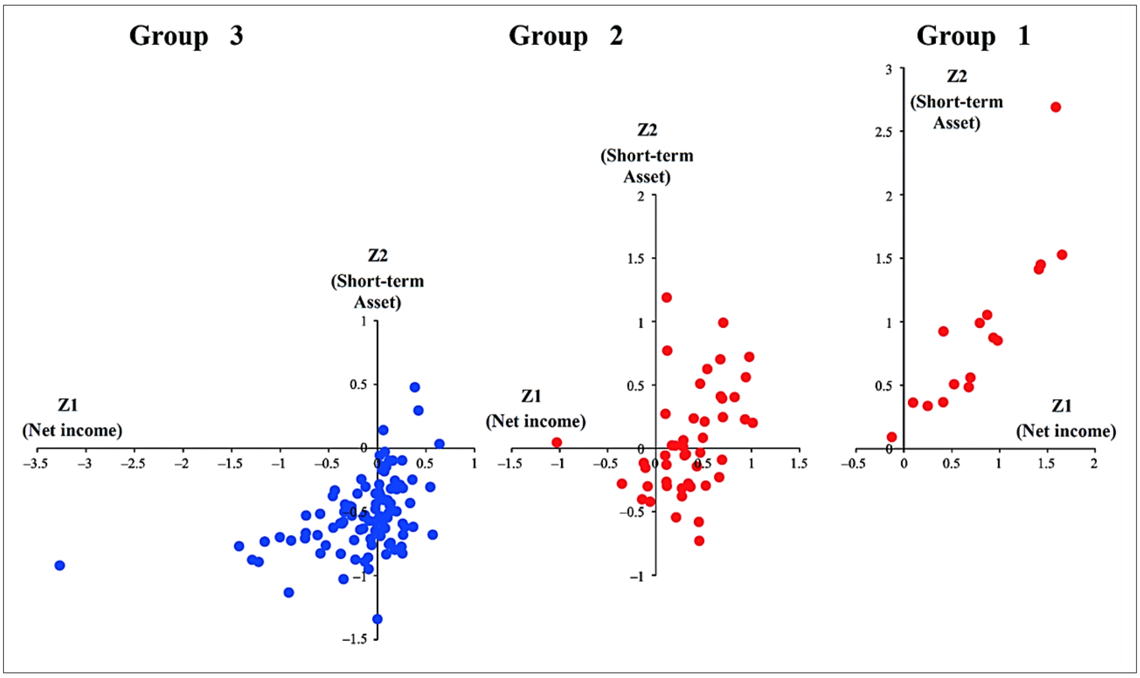
SME = small- and medium-sized enterprise.

Source: Yoshino and Taghizadeh-Hesary (2014).

The resultant dendrogram (hierarchical average linkage cluster tree) provides a basis for determining the number of clusters by sight. In the dendrogram shown in Figure 2, the horizontal axis shows 1,363 SMEs. Because of the large number of SMEs in this empirical work, they have not been identified by number in the dendrogram, although this is how they are identified in this survey. Rather, the dendrogram categorizes the SMEs in three main clusters (Groups 1, 2, and 3), but it does not show which of these three clusters contains the financially healthy SMEs, which contains non-healthy SMEs, and which contains intermediate SMEs. Hence, there is one more step to go.

Figure 2 shows the 1,363 SMEs categorized into three major clusters. Using their components, which were derived from the PCA analysis, the distribution of factors for each member of the three major clusters was plotted. Figure 3 shows the distribution of Z1–Z2 for these three cluster members separately.[6]

Figure 3: Grouping Based on Principal Component Analysis (Z1–Z2) and Cluster Analysis



Notes: Group 1 comprises the healthiest SMEs. Group 2 represents the in-between SMEs. Group 3 represents the least healthy SMEs.

Source: Yoshino and Taghizadeh-Hesary (2014).

As is clear in Figure 3, Group 1 comprises the healthiest SMEs, Group 3 the least healthy SMEs, and Group 2 the in-between SMEs. Interestingly, when we do this grouping using the other components (Z1–Z3, Z1–Z4, Z2–Z4, Z2–Z3, and Z3–Z4), the grouping is similar in most cases, which implies that this analysis is an effective way of grouping SMEs.

6 The dendrogram shows us the major and minor clusters. One useful feature of this tree is that it identifies a representative SME of most of the minor groups, which has the average traits of the other members of the group. For simplification, in Figure 3, we have only used data from these representative SMEs, which explains the whole group's traits. This is why the total number of observations in Figure 3 is lower than the 1,363 observations in this empirical work.

For a robustness check of classifications based on the aforementioned method, we performed one more step, and the results are summarized in Table 5.

Table 5: Average of Financial Ratios for Each Group of SMEs

Variables (Financial Ratios)	SME Groups		
	Group 1	**Group 2**	**Group 3**
Equity_TL	1.11	0.77	0.33
TL_Tassets	0.56	0.62	0.78
Cash_Tassets	0.08	0.03	0.05
WoC_Tassets	0.15	0.11	0.04
Cash_Sales	0.06	0.05	0.05
EBIT_Sales	0.24	0.26	0.13
Rinc_Tassets	0.28	0.17	0.06
Ninc_Sales	0.20	0.25	0.18
EBIT_IE	22.88	7.74	2.04
AP_Sales	0.49	0.50	0.60
AR_TL	0.61	0.44	0.41

Notes: Group 1 comprises the healthiest SMEs. Group 2 represents the in-between SMEs. Group 3 represents the least healthy SMEs. For the definition of each variable (financial ratios) see Table 2.

Source: Authors' calculations.

Table 5 shows the average of the 11 financial ratios based on our classifications, which categorized 1,363 SMEs into three groups. The healthiest group of SMEs (Group 1) in all ratios had a relatively better performance in comparison with the two other groups. The performance of the in-between SMEs (Group 2) in most cases was better than the least healthy SMEs (Group 3). On the other hand, 59% of firms in Group 3 are non-sound firms, which means they have risk-weighted assets greater than their shareholders' equity. This percentage is higher than the share of non-sound SMEs in either Group 1 or Group 2, demonstrating that the rationale of our method is acceptable and we can retain the results.

4. CONCLUDING REMARKS

SMEs play a significant role in all Asian economies. They are responsible for very high shares of employment and output. However, they find it difficult to borrow money from banks and other financial institutions. After the global financial crisis and implementation of the Basel III capital requirements, banks became more reluctant to lend to risky sectors. Because of the asymmetry of information existing between banks (lenders) and SMEs (borrowers), it is difficult for banks to distinguish healthy SMEs from risky ones; hence banks consider this sector to be a risky sector.

In this chapter we showed that using accumulated data of SMEs, it was possible to develop a comprehensive method for the credit risk assessment of SMEs by employing statistical analysis techniques. In the empirical part of this chapter, we created 11 financial variables of 1,363 SMEs that are customers of Asian banks and performed PCA and cluster analysis on them. The results showed that four variables (net income, short-term assets, liquidity, and capital) are the most important for describing the general characteristics of SMEs. Three groups of SMEs were then differentiated based on financial health.

The policy implications of this research are that if Asian governments can provide a comprehensive SME database—such as the CRD in Japan—and apply credit-risk assessment techniques similar to those presented in this chapter, then a comprehensive and efficient credit-rating system for SMEs can be created. Accordingly, financially healthy SMEs could borrow more money from banks at lower interest rates with lower collateral requirements because of their lower default risk, while SMEs in poor financial health would have to pay higher interest rates and have a lower borrowing ceiling with higher collateral requirements. By using such a credit-rating mechanism, banks could reduce the amount of nonperforming loans made to SMEs, which would improve the creditworthiness of the financial system and help healthy SMEs to raise money more easily from banks while contributing to economic growth.

Last but not least, there is an important point regarding SME data collection, because in many developing countries there is a lack of reliable SME data. Therefore, it might be difficult to apply the credit-rating methods that were used in this research when data are insufficient and unreliable. Japan exemplifies an efficient means of SME data collection. In that country there are 51 public credit guarantee corporations (CGCs), one for each prefecture and one in each of the cities of Kawasaki, Gifu, Nagoya, and Yokohama. CGCs are public entities that, by using the budgets from the central and local governments and also by receiving credit guarantee premiums from SMEs, provide guarantees for SME loans. The credit guarantee acts as collateral. Japan has a partial guarantee system, which covers 80% of the SME loan (Yoshino and Taghizadeh-Hesary 2018). When a SME approaches a CGC in a specific province, for example in Hokkaido, the Hokkaido CGC collects the data from the SME, which includes quantitative, qualitative, financial, and nonfinancial data. The SME needs to provide financial statements and other evidence regarding its current and historical activities. These data are accumulated within the nationwide SME database, the CRD. The CRD performs data cleansing and cross-checking and is responsible for the credit-risk assessment and scoring of SMEs. This successful experience shows that is possible to collect relevant and useful data through CGCs. To avoid biased output, the SME credit-scoring company and the CGC should be two separate and independent entities.

REFERENCES

Altman, E. I. and G. Sabato. 2007. Modelling Credit Risk for SMEs: Evidence from the US Market. *ABACUS* 43(3): 332–57.

Altman, E. I., M. Esentato, and G. Sabato. 2018. Assessing the Credit Worthiness of Italian SMEs and Mini-bond Issuers. *Global Finance Journal*, https://doi.org/10.1016/j.gfj.2018.09.003.

Angilella, S. and S. Mazzù. 2015. The Financing of Innovative SMEs: A Multicriteria Credit Rating Model. *European Journal of Operational Research* 244(2): 540–554.

Beck, T. H. L. 2007. *Financing Constraints of SMEs in Developing Countries: Evidence, Determinants and Solutions*. Tilburg University, School of Economics and Management.

Bruce-Ho, C.-T. and D. Dash-Wu. 2009. Online Banking Performance Evaluation Using Data Envelopment Analysis and Principal Component Analysis. *Computers & Operations Research* 36(6): 1835–42.

Chen, K. H. and T. A. Shimerda. 1981. An Empirical Analysis of Useful Financial Ratios. *Financial Management* 10(1): 51–60.

Driver, C. and J. Muñoz-Bugarin. 2018. Financial Constraints on Investment: Effects of Firm Size and the Financial Crisis. *Research in International Business and Finance*. DOI:10.1016/j.ribaf.2018.09.006.

Everitt, B. S., S. Landau, and M. Leese. 2001. *Cluster Analysis.* 4th ed. London: Arnold.

Feger, T. and J. Asafu-Adjaye. 2014. Tax Effort Performance in Sub-Sahara Africa and the Role of Colonialism. *Economic Modelling* 38: 163–74.

Fernandes, G. B. and R. Artes. 2016. Spatial Dependence in Credit Risk and Its Improvement in Credit Scoring. *European Journal of Operational Research* 249(2): 517–524. DOI:10.1016/j.ejor.2015.07.013.

Grunert, J., L. Norden, and M. Weber. 2004. The Role of Non-Financial Factors in Internal Credit Ratings. *Journal of Banking and Finance* 29(2): 509–31.

Kuwahara, S., N. Yoshino, M. Sagara, and F. Taghizadeh-Hesary. 2016. Role of the Credit Risk Database in Developing SMEs in Japan: Ideas for Asia. In *SMEs in Developing Asia New Approaches to Overcoming Market Failures.* P. Vandenberg, P. Chantapacdepong, and N. Yoshino, eds. Asian Development Bank Institute: Tokyo.

Landau, S. and I. Chis Ster. 2010. Cluster Analysis: Overview. *International Encyclopedia of Education.* 3rd ed. pp. 72–83. Oxford: Elsevier.

Lehmann, B. 2003. Is it Worth the While? The Relevance of Qualitative Information in Credit Rating. Working paper presented at the EFMA 2003 Meetings. Helsinki.

Li, K., J. Niskanen, M. Kolehmainen, and M. Niskanen. 2016. Financial Innovation: Credit Default Hybrid Model for SME Lending. *Expert Systems with Applications* 61(C): 343-355. DOI:10.1016/j.eswa.2016.05.029

Martinez, W. L. and A. R. Martinez. 2005. *Exploratory Data Analysis with Matlab*. Florida: Chapman and Hall/CRC Press.

Orsenigo, C. and C. Vercellis. 2013. Linear versus Nonlinear Dimensionality Reduction for Banks' Credit Rating Prediction. *Knowledge-Based Systems* 47: 14–22.

Poon, W. P. H., M. Firth, and H. G. Fung. 1999. A Multivariate Analysis of the Determinants of Moody's Bank Financial Strength Ratings. *Journal of International Financial Markets, Institutions and Money* 9: 267–283.

Ravi Kumar, P. and V. Ravi. 2007. Bankruptcy Prediction in Banks and Firms via Statistical and Intelligent Techniques–A Review. *European Journal of Operational Research* 180(1): 1–28.

Yoshino, N. 2012. Global Imbalances and the Development of Capital Flows among Asian Countries. *OECD Journal: Financial Market Trends* 2012/1.

Yoshino, N. and T. Hirano. 2011. Pro-Cyclicality of the Basel Capital Requirement Ratio and Its Impact on Banks. *Asian Economic Papers* 10(2): 22–36.

———. 2013. Counter-Cyclical Buffer of the Basel Capital Requirement and Its Empirical Analysis. In *Current Developments in Monetary and Financial Law* 6: 417–29. Washington, DC: International Monetary Fund.

Yoshino, N. and Taghizadeh-Hesary, F. 2014. Analytical Framework on Credit Risks for Financing SMEs in Asia*. Asia-Pacific Development Journal* 21(2): 1–21.

———2015. Analysis of Credit Risk for Small and Medium-Sized Enterprises: Evidence from Asia. *Asian Development Review* 32(2): 18-37.

———. 2018. Optimal Credit Guarantee Ratio for Small and Medium-sized Enterprises' Financing: Evidence from Asia. *Economic Analysis and Policy.* https://doi.org/10.1016/j.eap.2018.09.011

Yoshino, N., F. Taghizadeh-Hesary, and F. Nili. 2015. Estimating Dual Deposit Insurance Premium Rates and Forecasting Non-Performing Loans: Two New Models. *ADBI Working Paper.* No. 510. ADBI: Tokyo.

Yoshino, N., F. Taghizadeh-Hesary, P. Charoensivakorn, and B. Niraula. 2016. Small and Medium-Sized Enterprise (SME) Credit Risk Analysis Using Bank Lending Data: An Analysis of Thai SMEs. *Journal of Comparative Asian Development* 15(3): 383–406.