벤처창업연구 제11권 제3호 (통권45호)
Asia-Pacific Journal of Business Venturing and Entrepreneurship Vol.11. No.3 pp.1-15

A Comparative Study on Failure Pprediction Models for Small and Medium Manufacturing Company

Hwangbo, Yun*
Moon, Jong Geon**

Abstract

This study has analyzed predication capabilities leveraging multi-variate model, logistic regression model, and artificial neural network model based on financial information of medium-small sized companies list in KOSDAQ. 83 delisted companies from 2009 to 2012 and 83 normal companies, i.e. 166 firms in total were sampled for the analysis. Modelling with training data was mobilized for 100 companies inlcuding 50 delisted ones and 50 normal ones at random out of the 166 companies. The rest of samples, 66 companies, were used to verify accuracies of the models. Each model was designed by carrying out T-test with 79 financial ratios for the last 5 years and identifying 9 significant variables. T-test has shown that financial profitability variables were major variables to predict a financial risk at an early stage, and financial stability variables and financial cashflow variables were identified as additional significant variables at a later stage of insolvency. When predication capabilities of the models were compared, for training data, a logistic regression model exhibited the highest accuracy while for test data, the artificial neural networks model provided the most accurate results.

There are differences between the previous researches and this study as follows. Firstly, this study considered a time-series aspect in light of the fact that failure proceeds gradually. Secondly, while previous studies constructed a multivariate discriminant model ignoring normality, this study has reviewed the regularity of the independent variables, and performed comparisons with the other models. Policy implications of this study is that the reliability for the disclosure documents is important because the simptoms of firm's fail woule be shown on financial statements according to this paper. Therefore institutional arragements for restraing moral laxity from accounting firms or its workers should be strengthened.

KeyWords: firm's fail prediction, multi-variate discriminant model, Logistic regression analysis model artificial neural networks model, small & medium manufacturing company,

Ⅰ. INTRODUCTION

At the time of 1997 Asian financial crisis, local companies had difficulty in financing due to sluggish exports and infeasible investment expansion. The International Monetary Fund stepped in to stabilize currencies and credit crunch, which caused many companies to face bankruptcy. Various efforts were made by the government, including retrenchment in finance, higher interest policy, a gold collection campaign, an agonizing reconstructing, and restruction and vitalization of small-medium companies and venture companies In order to overcome such situations and thereby we could be graduated from the IMF management system.

In 2008, the global financial crisis caused by the investment loss from the sub-prime mortgage in US led to increase in financing cost and conservative credit policy of financial institutions, and therefore, as the small business financial conditions were getting worse than the bigger business companies, many of them became insolvent.

Alongside the global economic deceleration resulted from European debt crisis, started from 2011, the Europe economic recession, and the slowdown in economic growth of China and Southeast Asian countries, it is anticipated that the Korean

* Associate Professor, Graduate School of Global Entrepreneurship, Kookmin University, yun88@kookmin.ac.kr
** Adjunct Professor , Graduate School of Information and Communication, Konkuk University, moonjames@empas.com
· 투고일: 2016-05-28 · 수정일: 2016-06-21 · 게재확정일: 2016-06-21

economy will fall into L-shaped recession with rapid increase in the aging population and the household debt, sluggish business caused by long lasting recession will lead to gradual deepening insolvency, and therefore, the number of bankruptcy companies will be steadily increased.

Especially, since Korean economy is highly dependent on international economy, the consequence would be serious. As for manufacturing companies, they are losing their competitiveness to China and Southeast Asian countries which are armed with low wages and rapid technology development.

According to the global manufacturing competitiveness report released by the Deloitte consulting in 2013, Korea ranked No. 5, which was lower than its rank as No. 3 in 2010. The ranking is anticipated to lag far behind much further than before, which can properly explain the emergence situation as mentioned in the foregoing(Craig et al. 2013).

Under the free market economic system, uncompetitive companies are weeded out through an insolvency process and finally go into bankrupt, while companies which have competitiveness and run the business efficiently by continuously innovating itself will grow well and be survived.

If a company is insolvent or in bankruptcy, the shareholders, employees, financial institutions, suppliers and customers would suffer hardship, and eventually, the huge losses would be brought to the national economy in the end.

In this paper, business failure prediction models have been developed for a small to medium-sized business (SMB), in particular, for those running manufacturing industry, which have natural limitations at funds, manpower and sales force when compared to a big to medium-sized business (Kwon et al. 2012).

We believe that through this paper, if any potential loss of stockholders of the SMB can be reduced in advance, and thereby, companies can minimize the loss on bankruptcy before reaching the bankruptcy or deepening into insolvency by pushing M&A ahead and fostering their own structural innovation, this kind of research may help the domestic industry and the national economy.

Various reasons have been taken into account for insolvency and bankruptcy. However, since their results are ultimately to be reflected on the financial statements, in this study, we have focused on financial characteristics of the companies.

Among the registered companies in KOSDAQ market, which have easy collectability of information and reliability of data as the subject to external auditing, sample company selection has been carried out by excluding those in financial and non-manufacturing industries, and including delisted 83 small to medium-sized manufacturing companies and 83 normal companies from 2009 to 2012, i.e. 166 companies in total have been chosen to compose paired samples so as to minimize differences between industries and to maximize homogeneity of samples.

We have selected 50 failure companies and 50 non-failure companies from the 166 samples on a random basis to be used as training data for constructing models. The remaining 66 companies (33 failure companies, 33 non-failure companies) have been used as test data, i.e. out of sample test to evaluate prediction performances of the models.

This paper has applied a Student's T-test to analyze an average difference between an failure company group and a non-failure company group by utilizing 79 financial ratios, which are typically used in the financial statement analysis, so as to find out significant financial ratio variables for a fifth consecutive year. Based on such variables, we have used a multi-variate discriminant analysis, a logistic regression analysis and a artificial neural network analysis for constructing failure prediction models.

Most of the previous studies regarding failure prediction models constructed the models by using data of 3 years prior to the failure. In order to make differences, this paper has constructed the models by using the data of 5 years prior to the failure. This is based on our considerations that insolvency of company proceeds over a considerable time, and meanings of failure prediction would be diluted if financial indicators between insolvent companies and normal companies are getting clearly distinguished from each other as the insolvent companies are getting closer to be delisted.

Ⅱ. LITERTURE REVIEW

2.1 Definition of the failure business

Business failure slowly progresses over a long period of time and its concept is difficult to be quantified. Accordingly, the measurement is also difficult. Moreover, as every country has its different economic and social environment, it is hard to define the meaning clearly. Even scholars in this field use the term in their own limited ranges restrictively.

Beaver(1966) defined "Failure" as inability of a company to pay its financial obligations as they mature, and he regarded it as a failure if any of the following events occurred: bankruptcy, bond default, an overdrawn bank account, or nonpayment of a preferred stock dividend. Deakin(1972) defined it as failure business when a company goes through liquidation in the form of bankruptcy, insolvency for debt and any else for the sake of creditors' interests on the premise that great damages will be brought out to main interest parties, i.e. creditors and stockholders.

Altman & McGough(1974) named business failure to a company under a financial circumstance of infeasible reimbursement by the settlement date, two consecutive years of deficit, inevitable reduction in labor force due to poor profitability, ongoing lawsuits or impossibility to run the business any more due to difficulties in procuring the raw material from the suppliers.

Altman & Hotchkiss(2005) divided business failure into economic failure, insolvency, default, and bankruptcy.

In Korea, Article 306 (bankruptcy cause of the corporation), Paragraph 1 of Debtor Rehabilitation and Bankruptcy Law prescribes that “The court order shall declare bankruptcy of a debtor when the debtor’s debts exceed its assets,” and in general, if the company is in the situation of company dissolution, we call it as business failure.

Hwang(1991) defined the companies having difficulties not only in finance but in adminstration due to bankruptcy or insolvency as business failure. Lee(1993) applied the same to the companies, which are out of business, of which business activities were suspended, of which liquidation was requested, of which liquidation was begun, and of which the impaired capital continues over 3 years as the subject to be classified as the company filed under Chapter 11 and applying to be delisted or has already been delisted from the stock exchange. Jeon, & Kim(2000) argued that a company theoretically could have its unlimited life even if it is a non-biological entities, but most of companies are known to dissipate within 20 years, so the birth of the company will also bring the dissipation. They also explained that companies with inefficient business are weeded out by bankruptcy and enable the survival of the fittest under the free competition of the capitalistic system so that the bankruptcy can be considered as institutional arrangements to enhance the effectiveness of the whole economy, and added that as for organisms or companies, it would not be possible to hold it merely with a simple symptomatic treatment if the root cause is not timely adjusted, and therefore, the more we drag time, the more resources will be wasted to take the company to a healthy place; subsequently, the failure will be deepened. Choi, et al.(2002) regarded vergleich, bankruptcy occurrence, designated firms as subject to administrative issues, suspension of business with banks, suspension of business, applying for liquidation and companies filed for court receivership as business failure. Jeong(2003) described bankruptcy as these situations: receiving debt reduction measurement from the creditor, receiving status management from finance companies, transferring the company ownership to the creditor, being in insolvency, occurring bankruptcy, suspending business and being out of business, and applying for liquidation. Oh(2005) regarded a company as a failure company when it applies for vergleich, formalizes insolvency and files under Chapter 11, seeing external appearance of the insolvency as the financial insolvency, and added that the consequential legal action covers the legal concept of formalized liquidation and legal management etc. Jo(2005) described the failure business as a comprehensive concept including legal bankruptcy, management failure and insolvency. Kim(2011) viewed a selected workout company defined as the same, which is delisted, of which bankruptcy occurs, of which bank transaction is suspended, which has administration issues, which is managed by a bank, filed under Chapter 11 or applying for the liquidation, business suspension, company dissolution or bankruptcy or out of business, and explained that the time of recognizing business failure is the disclosure day when the bankruptcy occurred and the transaction was suspended, and further, in case of applying for the liquidation or filed under Chapter 11, the applying date is the management designation date.

Jang(2012) argued that business failure includes management failure, legal bankruptcy and insolvency. Kwon, et al.(2012) defined the failure business such that capacity of debt repayment or cost of insolvency of a company has been weakened or attenuated. Moon & Hwangbo(2014) suggested the definition of the business failure as a corporate which KOSDAQ Listing Practical Commrttee decided to be delisted on KOSDAQ market for the study of the failure prediction model of the firms listed on the KOSDAQ.

As explained in the previous studies above, we can find that the concept of the failure business is difficult to be defined. Therefore, in order to seek convenience in studying, such as standardizing the concept and easily collecting data, in this paper, failure business is limited to the companies corresponding to the KOSDAQ Listed Regulation Article 38 (delisting).

2.2 The prediction of the insolvent company

In this section, we will review the theoretical description and previous study on the multi-variate discriminant analysis, logistic regression analysis and artificial neural network analysis.

2.2.1 Multivariate discriminant analysis

A multi-variate discriminant analysis is a statistic technique of finding a linear combination (a discriminant) of independent variables, which divide an observation group into two or more contrasting groups. For example, as shown in the below, it is to find the best linear combination of financial ratios to discriminate between the failure company group and the non-failure company group.

$$Z=\alpha+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_k X_k$$

where, Z=discriminant score

α : an intercept

$X_1, \cdots, X_k$: independent variables

$\beta_1 \cdots \beta_k$: discriminant coefficients (weight value)

In this study, we derived a linear combination function by using a stepwise input method when selecting variables, inputted the relevant independent variables to get the Z value, and then, calculated the cut off point with the average discriminant score of each group. When the discriminant score was larger than the cut off point, we defined it as non-failure business, and when it was below the cut off point, we defined it as failure business. The significance of the discriminant was verified by using Wilk's Lambda value, Chi-square value and eigenvalue. The previous studies on discriminant analysis models are shown in <Table 1>.

<Table 1> Previous study using multivariate discriminant analysis

researcher	sample size		subject	classification accuracy	forecast period	normality test
	failure	non-failure				
Altman (1968)	33	33	listed	95.5%	2 years before	none
Altman et al(1994)	34	61	listed	97.0%	3 years before	none
Issah (2012)	35	35	listed	71.4%	3 years before	none
Bhandari, Iyer(2013)	50	50	listed	83.3%	1 year before	none
Kim, et al., (1998)	72	107	unlisted	82.5%	1 year before	none
Park(2008)	51	51	listed	76.5%	3 years before	none
Jun, et al.,(2011)	31	31	listed	85.5%	3 years before	none

2.2.2 Logistic regression analysis

Logistic regression model is a statistic technique, wherein estimated probabilities of the dependent variables are restricted to [0,1] and then the relationship between the dependent variable and independent variable is measured. The logistic function with k independent variables is defined as follows:

$$E(Y_i)=\frac{1}{1+e^{-(Z=\beta_0+\beta_1 X_{1i}+\beta_2 X_{2i}+\cdots\cdots+\beta_k X_{ki})}}$$

where, $E(Y_i)$: probability for Y_i being 1

X_k : independent variable

β_k : response coefficient

$e \fallingdotseq 2.71828$

For example, if Y is a bankruptcy company, we take it as 1 whereas if contrary, we take it as 0. Assuming that the independent variables, which affect the E(Yi), are financial ratios of $X_1, \ldots, X_k$, $\beta_1, \ldots, \beta_k$ are the estimations of measuring the impact of the failure probability for the variation of the independent variables. If the Z value in the following formula takes a positive infinite value, the equation value would be close to 1, whereas if the Z value of the formula takes a negative infinite value, the equation value would be close to 0.

$$Z=\alpha+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_k X_k$$

In order to construct a failure prediction model via the logistic regression model, the Z value should be estimated. Therefore, we estimated the equation by denoting 1 to the failure company and 0 to the non-failure company. In order to test the goodness of fit for the model, we used −2Log Likelihoods ratio (-2LL), Hosmer & Lemeshow test and Chi-square value. The overall goodness of fit of the logistic regression model represents a correspondence degree between the expected value of the model and the observed value of the dependent variable, and here, the smaller the value indicates, the better it fits. Previous studies using the logistic regression analysis model are shown in <Table 2>.

<Table 2> Previous study on logistic regression analysis

Researcher	Sample size		Subject	Classification accuracy	Prediction period
	Failure	Non-failure			
Ohlson(1980)	105	2,058	listed	96.1%	1 year before
Issah(2012)	35	35	listed	71.4%	3 years before
Kim, et al., (1998)	72	107	unlisted	96.1%	1 year before
Nam(1998)	44	44	listed	81.0%	1 year before
Park(2008)	51	51	listed	77.5%	3 years before
Bae, et al., (2011)	40	40	listed	81.3%	1 year before
Kim(2011)	111	111	listed	82.4%	2 years before
Jung(2014)	144	157	unlisted	74.8%	1 year before

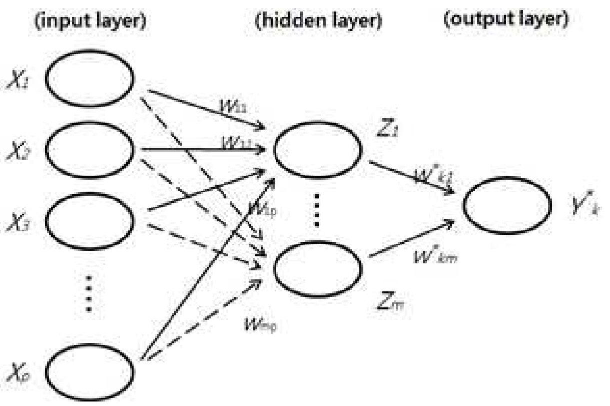
2.2.3 Artificial neural network analysis

As one field of artificial intelligence, artificial neural networks (ANNs) are characterized in that a computer thinks, infers, learns, and judges like human being, they can find out types of the data by itself without any presumption about the input data, so as to have a non-parametic feature of forming a model, and it can construct a multi-variate model with various variables and output values.

Thus, in case where distribution of the variables cannot be defined and it is not possible to presume the linear relationship among the variables, ANNs allows us to have more reasonable results than traditional multivariate linear regression analysis(Jo, et al. 1999). The ANNs comprise multiple processing elements, and the linear array of such processing elements composes a layer. The ANN model has an input layer, an output layer, and one or more hidden layers. Each processing element calculates the linear combination of the imput signal, then applied the result to a transfer function to get the output value. (Heo, et al. 2008)

Regarding indications in the neural network model, observations and data values are indicated as follows.

Here, n is the number of samples. $x_i = (x_{i1}\ x_{i2}\ \cdots\ x_{ip})$ is a multivariate (p dimension) observation variable, i.e explanatory vector, y_{ik} refers to the k th target variable value for the object i, wherein $i = 1, ..., n, k = 1, ..., c$. In case of classification, $y_{i1}, \cdots, y_{ic}$ take a dummy variable, which is 1 for object i's category, and which is 0 for the rest. In this regard, $X_1, X_2, \cdots, X_P$ are the explanatory variables in the input layer, $Z_1, Z_2, \cdots, Z_m$ are m hidden variables in the hidden layer, $Y_1, Y_2, \cdots, Y_c$ are the target variables for prediction and classification. The output variables $Z_1, Z_2, \cdots, Z_m$ are unobservable hidden variables.



<Figure 1> ANN diagram

Although there are various types of ANN, primarily, they are classified according to the number of the hidden layers between the input and the output. <Figure 1> is a neural network model with the hidden variables to be inputted to the 1 hidden layer. In general, artificial neural network models can be divided into multi-layer perceptron (MLP) and radial basis function (RBF) depending on methods of generating the hidden variables.

In this paper, we chose the MLP model and the determinant model for the hidden variables as follows:

$$Z_j = f(w_{j0} + w_{j1}X_1 + w_{j2}X_2 + \cdots + w_{jp}X_p),\ j = 1, \cdots, m$$

Here, as $Z_j = f(s_j)$ is an activation function, a logistic function (sigmoid) and Hyperbolic cotangent function are commonly used. That is,

Logistic: $f(s_j) = \dfrac{e^{s_j}}{1+e^{s_j}} = \dfrac{1}{1+e^{-s_j}}$

Hyperbolic cotangent:

$$f(s_j) = \frac{e^{s_j} - e^{-s_j}}{e^{s_j} + e^{-s_j}} = \frac{e^{2s_j} - 1}{e^{2s_j} + 1}$$

Thus, hidden variable Z_j takes a value between 0 to 1.

The discriminative model for the output variable is

$$Y_k^* = g_k(w_{k0}^* + w_{k1}^* Z_1 + \cdots + w_{km}^* Z_m),\ k = 1, \cdots, c.$$

Here, as $Y_k^* = g_k(s_k^*)$ is an output activation function, the logistic function is used for forecast and a softmax function is used for classification. That is, for forecast,

$g_k(s_k^*) = \dfrac{1}{1+e^{-s_k^*}}$ is used, and for classification,

$g_k(s_k^*) = \dfrac{e^{s_k^*}}{e^{s_1^*} + \cdots + e^{s_c^*}}$ is used, wherein $k = 1, \cdots, c$.

The previous studies on ANNs are shown in <Table 3>.

<Table 3> Previous studies on ANN

Researcher	Sample size		Subject	Classification accuracy	Prediction period
	Failure	Non-failure			
Han(2006)	22	66	listed	75.3%	4 years before
Kim(2006)	23	23	external audit	78.3%	3 years before
Kim, & Kang,(2010)	729	729	external audit	74.8%	1 year before
Odom, & Sharda (1990)	65	64	-	80.6%	1 year before
Back et al. (1996)	37	37	-	90.5%	3 years before
Sarah, & Shahriar (2011)	66	66	listed	89.1%	1 year before

Ⅲ. Model design for the empirical study

3.1 Sample selection

As it is very difficult to clearly define failure business, this study has selected the companies corresponding to the KOSDAQ Listed Regulation Article 38 (delisting), considering concept standardization, and easy collectability and reliability of data (external auditing company). This study selected 83 companies running manufacturing business while excluding those in financial and non-manufacturing industries so as to minimize differences between industries and to maximize homogeneity of the samples.

The sample company selection of normal companies, as opposite to the 83 failure companies, are those listed in the KOSDAQ market, which are also 83 companies and related to the same product or pertaining to the same industry as that of the failure companies, i.e. 166 companies in total have been chosen to compose paired samples.

We have selected 50 failure companies and 50 non-failure companies randomly from the 166 samples as training data to construct the model. The remaining 66 companies (33 failure companies, 33 non-failure companies) have been used as the test data set, i.e. out of sample test to evaluate predictive performance of the model.

The whole sample data were divided into 2 parts on random basis, designating one was for training data and the other was for test data. Training data was used to construct the model and to test the goodness of fit, test data is used to test the prediction model obtained as such.

Then, fair evaluation of the model is possible. If the prediction model were constructed by using the whole data as training data and use the results to evaluate the model, the dual use of the identical data would naturally cause an prediction error underestimated. Therefore, if the total data size is not sufficient, then We may use 60%~75% of the total data as training data and 40%~25% of the total data as test data(Heo, & Li, 2008).

In this paper, we divided the sample data at the ratio of 60% (100 companies) to 40% (66 companies) as shown in <Table 4>. The specific data has been extracted from Electronic Disclosure System of the Financial Supervisory Service and the site of Korea Exchange by searching the business reports and audit reports of 5 consecutive years of individual companies immediately before delisting.

<Table 4> The construction of the samples

Division	training data	test data	Total
Failure	50	33	83
Non-Failure	50	33	83
Subtotal	100	66	166
Proportion	60%	40%	100%

3.2 Operational definition for the variables

Prior to selecting financial variables necessary to construct a failure business prediction model, we have noted that although there are various causes of failure business, their results ultimately will come out as financial indicators, based on the hypothesis that there are significant differences between failure business indexes and non-failure business indexes, to proceed empirical research.

In this paper, we have used general financial ratios in financial statement analysis. The total of 79 financial ratios were used to test significant levels between the failure business and non-failure business of 5 consecutive years before delisting, and the financial ratios were 18 stability ratios (A group) <Table 5>, 12 profitability ratios (B group) <Table 6>, 9 activity ratios (C group) <Table 7>, 3 productivity ratios (D group) <Table 8>, 9growth ratios (E group) <Table 9>, 28 cash flow ratios(F group) <Table 10>.

<Table 5> stability ratios

Division	ratio	definition
A1	Current ratio	(Current asset/Current liabilities)*100
A2	Quick ratio	(Quick assets/Current liabilities)*100
A3	Current liabilities ratio	(Current liabilities/Owner's equity)*100
A4	No current ratio	(Non current assets/Owner's equity)*100
A5	Working capital ratio	[(Current asset-Current liabilities)/Total assets]*100
A6	Debt equity ratio	(Total debt/Total assets)*100
A7	Short-term current liabilities ratio	(Short-term financial assets/Current liabilities)*100
A8	Debt ratio	(Total debt/Owner's equity)*100
A9	Reliance of total borrowings	(Total borrowings/Total assets*100
A10	Short-term current assets ratio	(Total financial assets/Current asset)*100
A11	non-current assets to stockholders' equity and non- current liabilities	[Non-current assets/ (Non-current debts+Owner's equity)]*100
A12	**Liquidation value ratio**	(Net worth/Sales)*100
A13	Cashable assets ratio	(Cashable assets/Total assets)*100
A14	Retained earnings ratio	(Retained earnings/Owner's equity)*100
A15	Capital reserves ratio	(Capital reserves/Owner's equity)*100
A16	Accounts receivable to accounts payable ratio	(Account payable/Account receivable)*100
A17	Asset management ratio	(Current assets+Assets)/ (Total assets)*100
A18	Short-term financial assets to short-term borrowings ratio	(Short-term financial assets/Short-term borrowings)*100

<Ttable 6> profitability ratios

Division	ratio	definition
B1	Return on average equity	(Net income/Avg owner's equity)*100
B2	Return on total capital	(Net income/Avg total capital)*100
B3	Operating return on equity	(Avg operating earnings/Owner's equity)*100
B4	Operating return on total capital	(Operating earnings/Avg total capital)*100
B5	Net income to working capital	(Net income/Avg working capital)*100
B6	Operating income to working capital	(Operating profit/Avg working capital)*100
B7	Net profit ratio	(Net income/Sales*100
B8	Interest coverage ratio	(Operating profit/Interest cost)*100
B9	Financial expenses to sales ratio	(Interest cost/Sales)*100
B10	Return on net sales	(Operating profit/Sales)*100
B11	Sales to cost of judge ratio	(Cost of judge/Sales)*100
B12	Cost to revenue ratio	(Cost of sales/Revenue)*100

<Table 7> activity ratios

Division	ratio	definition
C1	Quick assets turnover ratio	(Sales/Avg quick assets)*100
C2	Inventory turnover ratio	(Cost of sales/Avg inventory)*100
C3	Owner'equity turnover ratio	(Sales/Avg owner's equity)*100
C4	Total assets turnover ratio	(Sales/Avg total assets)*100
C5	Working capital turnover ratio	[Sales/(Avg current assets+ Avg assets)]*100
C6	Non current assets turnover ratio	(Sales/Avg non current assets)*100
C7	Account receivables turnover ratio	(Sales/Account receivable)*100
C8	Account payable turnover ratio	(Cost of sales/Accounts payable)*100
C9	Tangible assets turnover ratio	(Sales/Avg tangible assets)*100

<Table 8> productivity ratios

Division	ratio	definition
D1	Added value ratio	(Value added/Sales)*100
D2	Gross value added to property, plant and equipment ratio	(Value added/Avg tangible assets)*100
D3	Productivity of capital, gross value-added to total assets	(Value added/Avg total capital)*100

<Table 9> growth ratios

Division	ratio	definition
E1	Total assets growth rate	[(Current year total assets-Prior year total assets/Prior year total assets]*100
E2	Owner's equity growth rate	[(Current year owner's equity-Prior year owner's equity)/ Prior year owner's equity]*100
E3	Total debt growth rate	[(Current year total debt-Prior year total debt)/ Prior year total debt]*100
E4	Sales growth rate	[(Current year sales-Prior year sales)/ Prior year sales]*100
E5	Operating profit growth rate	[(Current year operating profit-Prior year operating profit)/ Prior year operating profit]*100
E6	Net income growth rate	[(Current year net income-Prior year net income)/ Prior year net income]*100
E7	Tangible asset growth rate	[(Current year tangible asset-Prior year tangible asset)/Prior year tangible asset]*100
E8	Gross value added growth rate	[(Current year value added-Prior year value added)/Prior year value added]*100
E9	Current asset growth rate	[(Current year current asset-Prior year current asset)/Prior year current asset]*100

<Table 10> cash flow ratios

Division	ratio	definition
F1	Inventory retention period	Inventory/Avg cost of sales per day
F2	Account receivable collection period	Account receivable/Ave sales per day
F3	Account payable payment period	Account payable/Avg cost of sales per day
F4	Cash conversion cycle	(Inventory retention period+Account receivable collection period-Account payable payment period)
F5	Debt service coverage ratio	[(Operating income+Depreciation+Financial cost)/(Short-term borrowings+Financial cost)]*100
F6	EBITDA Interest coverage ratio	(EBITDA)*100/Interest expense
F7	EBITDA Current liabilities ratio	(EBITDA/Current liabilities)*100
F8	EBITDA Sales ratio	(EBITDA/Sales)*100
F9	EBITDA Short-term borrowings ratio	(EBITDA/Short-term borrowings)*100
F11	Operating cash flow/Short-term borrowings ratio	(Cash flow after operations/Short-term borrowings)*100
F12	Operating cash flow/ Total borrowings ratio	(Cash flow after operations/Total borrowings)*100
F13	Free cash flow to sales ratio	(Free cash flow/Sales)*100
F14	Free cash flow to short-term borrowings ratio	(Free cash flow/Short-term borrowings)*100

F15	Free cash flow to total borrowings ratio	(Free cash flow/Total borrowings)*100
F16	Operating cash flow/ current liabilities ratio	(Cash flow after operations/ Current liabilities)*100
F17	Operating cash flow to debt ratio	(Cash flow fater operations/ Debt)*100
F18	Free cash flow to current liabilities ratio	(Free cash flow/current liabilities)*100
F19	Free cash flow to debt ratio	(Free cash flow/Debt)*100
F20	Free cash flow to current liabilities coverage ratio	(Free cash flow/ Growth in current liabilities)*100
F21	Free cash flow to short-term borrowings coverage ratio	(Free cash flow/ Growth in short-term borrowings)*100
F22	EBITDA Total assets ratio	(EBITDA/Total assets)*100
F23	EBITDA Debt ratio	(EBITDA/Debt)*100
F24	EBITDA Total borrowings ratio	(EBITDA/Total borrowings)*100
F25	Operating cash flow/ Total assets ratio	(Cash flow after operations/Total assets)*100
F26	Free cash flow to total assets ratio	(Free cash flow/Total assets)*100
F27	Operating earnings to operating cash flow ratio	(Operating earnings/Operating cash)*100
F28	Operating earnings to EBITDA ratio	(Operating earnings/EBITDA)*100

Ⅳ. Empirical analysis results

4.1 T-test results of the financial variables′

The results on the 79 financial indicators of the two groups' average difference analysis (namely T-test) between the 50 failure companies and the 50 non-failure companies in 5 years before delisting are shown in <Table 11>. There are 13 indicators in the fifth year (T-5), 20 indicators in the fourth year (T-4), 26 indicators in the third year (T-3), 36 indicators in the second year (T-2), 52 indicators in the first year (T-1) before delisting. As it is getting closer to the year immediately before delisting under the significant probability 1%, the significant variables are increasing, which means that the financial ratio difference between the failure companies and non-failure companies is further clearly widening. In the process of processing the collected data, as for failure companies, as there are many extreme values of financial indicators, we have limited the upper limit and the lower limit to substitute the extreme values.

<Table 11> financial variables whose significance probability is less than 1% in past 5 years by year

year	group	variable	number of variables
T 1	A	A1,A2,A5,A6,A7,A9,A10,A13,A17,A18	52
	B	B2,B4,B5,B6,B7,B8,B9,B10,B11,B12	
	C	C1,C4,C5,C6	
	D	D1,D2	
	E	E2,E4,E8	
	F	F2,F3,F4,F5,F6,F7,F8,F9,F10,F11,F12,F13,F14,F15,F16,F17,F18,F19,F22,F23,F24,F25,F26	
T 2	A	A4,A5,A6,A8,A9,A14,A15,A17	36
	B	B2,B3,B4,B5,B6,B7,B8,B9,B11,B12	
	C	C1,C4,C6	
	D	D3	
	F	F2,F3,F5,F6,F7,F10,F13,F16,F17,F18,F22,F23,F25,F26	
T 3	A	A5,A6,A9,A11,A14,A15,A17	26
	B	B2,B3,B4,B5,B6,B7,B8,B9,B10,B11	
	C	C4,	
	F	F10,F12,F13,F16,F17,F18,F25,F26	
T 4	A	A5,,A6,A9,A11,A14,A17	20
	B	B1,B2,B3,B4,B5,B6,B7,B9,B10,B11	
	F	F10,F13,F25,F26	
T 5	A	A6,A9,A14,A15	13
	B	B2,B6,B7,B9,B10,B11,	
	F	F16,F25,F26	

When <Table 11> is referred, in the 4th year and the 5th year before the delisting significant variables are shown only on the indicators in relation to the stability, profitability and cash flow. However, the 3rd year before the delisting, the significant variables are also shown on the activity indicators, and the 2nd year and the 1st year before the delisting, the significant variables came to be shown on the productivity and growth indicators. Therefore, it can be seen that the significant variables are shown on all the indicators.

This suggests that with regard to failure business prediction, observation of stability, profitability and cash flow indicators must take precedence above all, as the meaning of early warning signs. As shown in <Table 12>, there are 9 variables of which the significant probability is less than 1% for the past 5 consecutive years. In this study, we will consider these 9 variables as the key points to be analyzed.

<Table 12> the appearance frequency of variables

appearance frequency	variables	number of variables
5times	**A6,A9,B2,B6,B7,B9,B11,F25,F26**	**9**
4times	A5,A14,A17,B4,B5,B10,F10,F13,F16	9
3times	A15,B3,B8,C4,F17,F18	6
twice	A11,B12,C1,C6,D3,F2,F3,F5,F6,F7,F12,F22,F23,	13
once	A1,A2,A4,A7,A8,A10,A13,A18,B1,C5,D1,D2,E2, E4,E8,F4,F8,F9,F11,F14,F15,F19,F24,	23

<Table 13> The significant variables for the 5 consecutive years

Division	ratio	definition
A6	Debt equity ratio	(Total debt/Total assets)*100
A9	Reliance of total borrowings	(Total borrowings/Total assets*100
B2	Return on total capital	(Net income/Avg total capital)*100
B6	Operating income to working capital	(Operating profit/Avg working capital)*100
B7	Net profit ratio	(net income/sales)*100
B9	Financial expenses to sales ratio	(Interest cost/Sales)*100
B11	Sales to cost of judge ratio	(Cost of judge/Sales)*100
F25	Operating cash flow/ Total assets ratio	(Cash flow after operations/Total assets)*100
F26	Free cash flow to total assets ratio	(Free cash flow/Total assets)*100

4.2 Multi-variate discriminant analysis model

Discriminant analysis assumes that an independent variable should satisfy normal distribution. However, Kim, & Ban(1990) claimed that at least in economics and financial management, deviations of the normal distribution seem to be regular rather than to be exceptional, and validity verification for the hypothesis of normality has been ignored for most of the existing studies, as shown in <Table 13>. Deakin(1972) argued that the deviations of the financial ratio often do not comply with the normal distribution, and even though they would follow the normality when converted to square root or log transformation, there are no guidelines for such transformation.

In this study, we performed skewness, kurtosis and Shapiro-Wilks tests for the normality test. This is because that in statistical tests, the Kolmogonov-Sminov test method is mainly used in case that the number of data is more than 2000, while the Shapiro-Wilks test method is mainly used in case that the number of data is less than 2000 (Razali, & Wah, 2011).

In <Table 14>, as a result of the tests, only the variables of the model in the 3rd year before delisting met the requirements for the normal distribution while those of the other four models did not follow the normal distribution, under the significance level of 0.01.

<Table 14> Descriptive statistics in the normality test of multi-variate discriminant analysis by year

year	variable	corporation	skewness	kurtosis	Shapiro-Wilks normality test	
					statistics	significance probability
T1	A9	non-failure	0.195	-1.255	0.922	0.003
		failure	1.016	1.565	0.937	0.010
	B2	non-failure	-2.069	10.067	0.823	0.000
		failure	-1.725	2.692	0.802	0.000
	B6	non-failure	0.739	-0.551	0.909	0.001
		failure	-2.708	13.013	0.745	0.000
	F25	non-failure	-0.259	0.344	0.976	0.391
		failure	-0.800	1.138	0.958	0.072
T2	A9	non-failure	0.330	-1.029	0.924	0.003
		failure	-0.041	-0.705	0.983	0.683
	B2	non-failure	0.612	3.130	0.924	0.003
		failure	-1.053	1.599	0.878	0.000
T3	A6	non-failure	0.666	0.322	0.959	0.082
		failure	-0.441	-0.575	0.944	0.019
	B6	non-failure	0.341	0.070	0.972	0.268
		failure	-0.356	1.882	0.949	0.032
T4	A9	non-failure	0.482	-0.987	0.884	0.000
		failure	-0.125	-0.976	0.948	0.029
	B2	non-failure	0.153	0.884	0.964	0.135
		failure	-1.492	3.350	0.891	0.000
	B11	non-failure	2.178	5.874	0.790	0.000
		failure	0.956	-0.111	0.865	0.000
T5	B9	non-failure	1.901	3.924	0.784	0.000
		failure	1.499	1.859	0.828	0.000

To verify if the discriminant is significant or not, we can use canonical correlation coefficients and wilk's lambda values. As the canonical correlation coefficient represents a correlation between the discrimination score and the group, the greater value indicates the higher explanatory power of the discriminant (Chae, 2008). In <Table 15>, the canonical correlation coefficient becomes greater and greater when it is getting closer to T-1 year from T-5 year, and it shows that the explanatory power of the discriminant is getting higher and higher when it is closer to the year before delisting.

Since the wilk's lambda is approximated with a chi-squared distribution, then we can apply this result so to find that the discriminant is significant under the 1% significance level. In addition, if a Variance Inflation Factor (VIF) had been larger than 10, and consequently more than 90% of a bias of the corresponding variables had been explained according to the other explanatory variables, it would have been taken seriously for the multi-collinearity (Kang, et al. 2010). However, there are no multi-collinearity issues in this study as the VIF of the model by year is less than 3.

<Table 15> Multi-variate discriminant analysis model and descriptive statistics by year

Division	T-1	T-2	T-3	T-4	T-5
constant	-0.891	-1.189	-0.877	-1.442	-0.866
Debt equity ratio			0.021		
Reliance of total borrowings	0.014	0.031		0.034	
Return on total capital	-0.009	-0.026		-0.017	
Operating income to working capital	-0.015		-0.052		
Financial expenses to sales ratio					0.500
Sales to cost of judge ratio				0.025	
Operating cash flow/ Total assets ratio	-0.028				
wilk's λ	0.482	0.632	0.719	0.701	0.807
chi-square	70.058	44.454	31.960	34.301	20.909
significance probability	0.000	0.000	0.000	0.000	0.000
eigenvalue	1.075	0.581	0.390	0.427	0.239
VIF	2.3	1.1	1.1	2.1	1.0
canonical correlation coefficient	0.720	0.606	0.530	0.547	0.439
the centroid of group (non-failure)	-1.026	-0.755	-0.618	-0.647	-0.484
the centroid of group(failure)	1.026	0.755	0.618	0.647	0.484
accuracy of training sample classification(%)	**86.0**	**83.0**	**79.0**	**80.0**	**74.0**
failure discrimination(%)	72.0	78.0	72.0	78.0	58.0
non-failure discrimination(%)	100.0	88.0	86.0	82.0	90.0
accuracy of verification sample classification(%)	**87.9**	**77.3**	**72.7**	**68.2**	**63.6**
failure discrimination(%)	84.8	75.8	69.7	57.6	45.5
non-failure discrimination(%)	90.9	78.8	75.8	78.8	81.8

In the T-5 year, which is the beginning of failure 5 years before delisting, the financial expenses to sales ratio (B9) are emerged as an important indicator. In the T-4 year, the reliance of the total borrowings (A9), the return on total capital (B2), the sales to cost of judge ratio (B11) are important indicators, and in the T-3 year, the debt equity ratio (A6) and the operating income to working capital (B6) are important indicators.

In the T-2 year, the rate of the reliance of total borrowing (A9) and the return on total capital ratio (B2) come out as important indicators. In the T-1 year, which is the year immediately before delisting, the rate of the reliance of total borrowing (A9), the return on total capital ratio (B2), the operating income to working capital (B6) and the operating cash flow to total assets ratio (F25) are important indicators.

In <Table 15>, each accuracy of those classifications by year of the discriminant model was 86% in the T-1 year, 83% in the T-2 year, 79% in the T-3 year, 80% in the T-4 year, and 74% in the T-5 year. Meanwhile, with regard to the samples for test data, i.e. for the out of data samples, the classification accuracy was 87.9% in the T-1 year, 77.3% in the T-2 year, 72.7% in the T-3 year, 68.2% in the T-4 year and 63.6% in the T-5 year, which exhibit lower levels than those of classification accuracies of the other models, excluding the T-1 year.

In this model, a percentage of the classification accuracy, which discriminates a failure business to a non-failure business, was significantly lower than that discriminates a non-failure business to a failure business. In other words, it shows a characteristic that the type I error was higher than the type II error.

4.3 Logistic regression analysis model

As the chi-square value of Hosmer and Lemeshow Test (HL test) presents how well the logistic regression analysis model fits the data, the whole data samples are necessary to be divided into a certain number according to the ordering, and if there is no significant difference between actually observed data and predictive data, then it is concluded that the model is well fitted(Park, 2002).

In <Table 16>, the test results showed that the model is well fitted when the significance level is over 0.05. As the Variance Inflation Factor (VIF), which is a multi-collinearity indicator, did not include any model having a value over 3, it was found that there was no issue in relation to multicollinearity between the variables.

In the T-5 year which is the beginning of failure before the delisting, the financial expenses to sales ratio (B9) and the net profit ratio (B7) are emerged as important indicators in the financial ratios. In the T-4 year, the return on total capital ratio (B2) and the financial expenses to sales (B9) are important indicators. Furthermore, in the T-3 year, the return on total capital ratio (B2) and the operating cash flow to total assets ratio (F25) are important indicators.

In the T-2 year, the return on total capital ratio (B2), the financial expenses to sales ratio (B9) and operating cash flow to total assets ratio (F25) are turned out as important indicators. In the T-1 year, the reliance of the total borrowings (A9), the return on total capital ratio (B2) and the operating cash flow to total assets ratio (F25) are important indicators.

Each accuracy of those classifications by year of the logistic regression analysis model was 96% in the T-1 year, 88% in the T-2 year, 84% in the T-3 year, 81% in the T-4 year, and 79% in the T-5 year. Meanwhile, with regard to the samples for test data, i.e. for the out of data samples, the classification accuracy was 90.9% in the T-1 year, 92.4% in the T-2 year, 77.3% in the T-3 year, 71.2% in the T-4 year and 71.2% in the T-5 year, which exhibits slightly lower levels than those of classification accuracies of the other models, excluding the T-2 year.

<Table 16> Logistic regression analysis model and descriptive statistics by year

Division	T-1	T-2	T-3	T-4	T-5
constant	-4.147	-1.483	-0.543	-0.685	-0.740
Reliance of total borrowings	0.079				
Return on total capital ratio	-0.183	-0.070	-0.099	-0.057	
Net profit ratio					-0.046
the financial expenses to sales ratio		0.551		0.401	0.616
Operating cash flow to total assets ratio	-0.113	-0.063			
Free cash flow to total assetsratio			-0.047		
HL test	5.598	6.979	4.524	13.893	10.076
chi-square	109.5	68.0	51.1	39.6	34.2
VIF	2.2	1.7	1.5	1.9	1.3
accuracy of training sample classification(%)	**96.0**	**88.0**	**84.0**	**81.0**	**79.0**
(failure discrimination)(%)	96.0	88.0	78.0	74.0	68.0
(non-failure discrimination)(%)	96.0	88.0	90.0	88.0	90.0
accuracy of verification sample classification(%)	**90.9**	**92.4**	**77.3**	**71.2**	**71.2**
(failure discrimination)(%)	93.9	93.9	81.2	63.6	60.6
(non-failure discrimination)(%)	87.9	90.9	72.7	78.8	81.8

4.4 Artificial neural network model

In this study, we used a multi-layer perceptron (MLP) model, which is composed of an input layer having 9 input nodes corresponding to the respective input variables, one hidden layer corresponding to a hidden node for each year and an output layer corresponding to a target variable.

Referring to <Table 17>, if we examine the relative importances of the input nodes on a year-on-year basis, we can notice that the return on total capital ratio (B2) was the most important in the T-1 year and the T-2 year, and the operating cash flow to total assets ratio (F25) took the second. In the T-3 year, Sales to cost of judge ratio (B11) was the most important, followed by the operating cash flow to total assets ratio (F25) and the financial expenses to sales ratio (B9). In the T-4 year and the T-5 year, the financial expenses to sales ratio (B9) was the most important.

In <Table 18>, each accuracy of those classifications by year of the training samples was 92% in T-1 year, 91% in T-2 year, 82% in T-3 year, 81% in T-4 year and 74% in T-5 year. Meanwhile, with regard to the samples for test data, the classification accuracy by year was 90.9% in T-1 year, 92.4% in T-2 year, 86.4% in T-3 year, 71.2% in T-4 year and 74.2% in T-5 year. Especially, the classification accuracies of the samples for verification were shown as higher than those of the samples for training, in T-2 year, T-3 year and T-5 year.

<Table 17> Input nodes' relative importance by year

Division	T-1	T-2	T-3	T-4	T-5	average
A6	43.8%	18.5%	9.3%	34.2%	60.1%	33.2%
A9	46.6%	16.3%	6.1%	48.7%	36.9%	30.9%
B2	100.0%	100.0%	10.7%	57.0%	7.0%	54.9%
B6	11.3%	8.8%	63.9%	34.4%	9.7%	25.6%
B7	63.0%	32.2%	12.2%	52.4%	90.9%	50.1%
B9	49.4%	88.4%	75.2%	100.%	100.%	82.6%
B11	65.3%	47.0%	100.0%	88.5%	79.2%	76.0%
F25	85.9%	84.9%	78.8%	18.4%	26.1%	58.8%
F26	13.8%	6.3%	49.1%	34.7%	38.0%	28.4%
Hidden node	6	2	6	5	5	
Hidden activation function	hyperbolic tangent	hyperbolic tangent	hyperbolic tangent	hyperbolic tangent	hyperbolic tangent	
Output activation function	Soft max	Soft max	Soft max	Soft max	Soft max	

<Table 18> Accuracy of classification in artificial neural network model by year

division	-1	T-2	T-3	T-4	T-5
accuracy of training sample classification(%)	**92.0**	**91.0**	**82.0**	**81.0**	**74.0**
(failure discrimination)	90.0	88.0	68.0	78.0	70.0
(non-failure discrimination)	94.0	94.0	96.0	84.0	78.0
accuracy of verification sample classification(%)	**90.9**	**92.4**	**86.4**	**71.2**	**74.2**
(failure discrimination)	90.9	90.9	90.9	63.6	66.7
(non-failure discrimination)	90.9	93.9	81.8	78.8	81.8

4.5 Comparison of the predictive power by the model

Among the models constructed by training data, the classification accuracy for the logistic regression analysis model is the highest with the mean percentage of 5 years as 85.6%, the second one is the artificial neural network model with 84.0%, and the multi-variate discriminant analysis model is the lowest with 80.4%. On the contrary, as for the test data, the prediction accuracy of the artificial neural network model is the highest with 83.0%, the second one is the logistic regression model with 80.6%, and the lowest value is 73.9% of the multi-variate discriminant analysis model.

As can be seen from <Table 19>, in all the models, type I error, which verifies failure business as non-failure business, occurred more than type II error, which verifies non-failure business as failure business. Practically, as the loss of the stake-holders resulted from type I error is more crucial than the loss of the opportunity profit resulted from the type II error, further research should be done in this topic so as to reduce losses of any interest parties by reducing chances of the type I error.

<Ttable 19> Comparison of accuracy of classification by analysis model by year

division		T-1	T-2	T-3	T-4	T-5	average
MDA	accuracy of training sample classification	86.0	83.0	79.0	80.0	74.0	80.4
	(failure discrimination)	72.0	78.0	72.0	78.0	58.0	71.6
	(non-failure discrimination)	100.0	88.0	86.0	82.0	90.0	89.2
	accuracy of verification sample classification	87.9	77.3	72.7	68.2	63.6	73.9
	(failure discrimination)	84.8	75.8	69.7	57.6	45.5	66.9
	(non-failure discrimination)	90.9	78.8	75.8	78.8	81.8	81.2
LRA	accuracy of training sample classification	96.0	88.0	84.0	81.0	79.0	85.6
	failure discrimination)	96.0	88.0	78.0	74.0	68.0	80.8
	(non-failure discrimination)	96.0	88.0	90.0	88.0	90.0	90.4
	accuracy of verification sample classification	90.9	92.4	77.3	71.2	71.2	80.6
	(failure discrimination)	93.9	93.9	81.8	63.6	60.6	78.8
	(non-failure discrimination)	87.9	90.9	72.7	78.8	81.8	82.4
ANN	accuracy of training sample classification	92.0	91.0	82.0	81.0	74.0	84.0
	(failure discrimination)	90.0	88.0	68.0	78.0	70.0	78.8
	(non-failure discrimination)	94.0	94.0	96.0	84.0	78.0	89.2
	accuracy of verification sample classification	90.9	92.4	86.4	71.2	74.2	83.0
	(failure discrimination)	90.9	90.9	90.9	63.6	66.7	80.6
	(non-failure discrimination)	90.9	93.9	81.8	78.8	81.8	85.4

Ⅴ. Conclusion

In this paper, we have constructed widely used models, i.e. the multivariate discriminant analysis model, the logistic regression analysis model and the artificial neural network model by using the training data for studying the business failure prediction so as to figure out their classification accuracies, and compared and evaluated their prediction capabilities.

For this study, 83 small to medium-sized manufacturing companies, which were delisted in KOSDAQ market from 2009 to 2012, with 83 normal companies, i.e. the total of 166 company samples were selected. In order to compare the classification accuracies of the models, 100 companies were randomly extracted from the samples, applying 66 companies as the samples for testing each model developed as the training data.

The result of T-test showed that while the profitability indicators were the main variables for the failure prediction at the beginning phase of the failure, the stability and cash flow indicators also became the main variables for the failure prediction at the latter phase of the failure.

When prediction capabilities of the models were compared, in case of using training data, logistic regression analysis model manifested its highest classification accuracy, and the artificial neural network was the second. The lowest one was the multi-variate discriminant analysis model. Contrarily, as for the test data, the artificial neural network showed the highest classification accuracy, the logistic regression analysis model was the second, and the multi-variate discriminant analysis model had the lowest accuracy. The prediction capabilities of the artificial neural network analysis model and the logistic regression analysis model as constructed were higher than in previous researches.

There are differences between the previous researches and this study as follows. First, in order to construct the models, we have selected, as variables, financial ratios showing significant differences between the failure business and the non-failure business in 5 consecutive years, and considered a time-series aspect in light of the fact that failure proceeds gradually.

Second, differentiated from the previous research, which predicted failure business merely with materials of 3 years before the relevant companies were delisted, the present study additionally reflected materials regarding 4th year and 5th year before delisting, taking the fact into consideration that failure business slowly proceeds over a long time, and tried to predict business failure as an early warning of failure symptoms.

Third, the basic assumption, i.e. regularity of the independent variables was ignored in the previous study during construction of the multi-variate discriminant model. However, we have reviewed the regularity of the independent variables, and performed comparisons with the other models with respect to both cases where the regularity was satisfied and the regularity was not satisfied.

Policy implications of this study is that the reliability for the disclosure documents is important because the simptoms of firm’s fail woule be shown on financial statements according to this paper. Therefore institutional arragements for restraing moral laxity from accounting firms or its workers should be strengthened.

Notwithstanding, this study has its limitations. First, as the present research limited the analyzed subject to small to medium-sized manufacturing companies listed in the KOSDAQ market, we have a view that the range of subjects to be analyzed should be broaden in the future, considering that importance of service industries such as internet-related companies has increased in recent years.

Second, as a result of analyzing the models, type I error, which verifies failure business as non-failure business, occurred more than type II error, which verifies the non-failure business verified as failure business. Practically, as the loss of the stake-holders resulted from type I error is more crucial than the loss of the opportunity profit resulted from the type II error, further research should be done in this topic so as to reduce losses of any interest parties by reducing chances of the type I error.

Third, in this paper, we utilized the multi-variate discriminant analysis model, the logistic regression analysis model and the artificial neural network model, but we acknowledged that we failed to apply other various models such as a probit model and a survival analysis model for comparison and analysis, and that non-financial factors were not considered. We leave those issues to be taken into consideration for the future research.

Nevertheless, as focusing on small to medium-sized manufacturing companies, of which the business performances are being deteriorated due to the long-term domestic and overseas economy recession and the dropped competitiveness of the manufacturing industry according to technical pursuits of China and Southeast Asian countries, this study is believed to be of help to foresee any failure at a rather early stage both in a timely manner and in a practical manner.

References

Altman, E. I.(1968), Financial Ratios, Discriminant Analysis and Prediction of Corporate Bankruptcy, *Journal of Finance,* 23(4), 589-609.

Altman, E. I., Eom, Y. H. & Kim, D. W.(1994), *Distress Classification of Korean Firms,* New York University, working paper series, FD-94-2, 1-38.

Altman, E. I. & Hotchkiss, E.(2005), Corporate Financial Distress and Bankruptcy: Predict and Avoid Bankruptcy, Analyze and Invest in Distressed Debt, *3rd Edition, Joniley & Sons Inc., Hoboken,* New Jersey, 3-6.

Altman, E. I. & McGough, T. P.(1974), *Evaluation of a Company as a Going Concern, Journal of Accounting*, 138(6), 50-57.

Back, B., Laitinen, T. & Sere, K.(1996), Neural Networks and Genetic Algorithms For Bankruptcy Predictions, *Expert Systems With Applications,* 11(4), 407-413

Bae, K. S. & Jung, S. H.(2011), An empirical study on the prediction model of the delisted firms, *Tax Accounting Research.* 30. 125-140.

Beaver, W. H.(1966), Financial Ratios as Predictors of Failure, *Journal of Accounting Research*, 4(3), 71-111.

Bhandari, S. B. & Iyer, R.(2013), Predicting Business Failure Using Cash Flow Statement Based Measures, *Managerial Finance*, 39(7), 667-676.

Chae, S. I.(2008), *Social Science Search Methodology*. Seoul : B&M Books press.

Choi, T. S., Kim, H. K. & Kim, S. H.(2002), A Comparison of the Discrimination of Business Failure prediction Models, *Journal of the Korean OR and MS Society,* 27(2), 1-13.

Craig, A., Aleda, V., Bharath, G., Atanu, C., Jack, M., Tim, H. & Michelle, D.(2013), 2013 *Global Manufacturing Competitiveness Index*, NY, Deloitte Touche Tomatsu Limited and U.S. Council on Competitiveness, Retrieved Jan 1, 2014 from https://www.deloitte.com/assets/Dcom-Mexico/Local%20Assets/Documents/mx(en-mx)Global_Manufacturing_Competitiveness2013.pdf

Deakin, E. B.(1972), A Discriminant Analysis of Predictors of Business Failure, *Journal of Accounting Research,* 10(1), 167-179.

Han, H. N.(2006), A Comparative Study on the Predictability of Bankruptcy Model : Artificial Neural Networks Model and Logistics Regression Model", Management includes, *Management Research Center of Inha University*, 11(1), 215-236.

Heo, M. H. & Lee, Y. K.(2008), *Data Mining Modeling and Case,* Seoul : Hannare Press.

Hwang, S. H.(1991), A Study on Corporate Failure Prediction, *Korean Accounting Review*, 12(7), 57-78.

Issah, M.(2012), Predicting Corporate Failure of UK's Listed Companies Comparing Multiple Discriminant Ananlysis and Logistic Regression, *Lap LaMBERT Academic Publishing*, 32-68.

Jang, Y. K.(2012), *Business Analysis*, Seoul: Trading Management Sa Press.

Jeong, Y. S.(2003), *A Neural Network Model for Bankruptcy Prediction,* Doctoral dissertation, Kyung Hee University.

Jeon, S. B. & Kim, M. C.(2000), *Reality and Theory on Business Failure*, Seoul: Dasan Publishing Company.

Jo, Q. G.(2005), *A study on the prediction models of corporate failure*, Master's thesis, Dong-Eui University.

Jo, Y. J., Heo, J. & Choi, I. K.(1999), *Data Mining Neural Network Analysis*, Seoul : SPAA academy Press.

Jun, H. W., Chung, Y. H. & Shin, D. H.(2011), A Study on the Failure Prediction Model of Delisting Firms, *Korea International Accounting Review,* 38(8), 331-362.

Jung, K. W.(2014), *A Study on the default prediction model of SMES after supporting the credit guarantee,* Masters's thesis, Hanyang University.

Kang, H. C., Choi, H. S. & Han, S. T.(2010), *Introduction to SPSS Data Analysis*. Paju: Jayou Academy.

Kim, C. K. & Kim, S. H.(1998), The Business Failure Prediction Model of Small and Medium-sized Firms, Paichai University, Journal of Social Science Research 17, 111-132.

Kim, G. C.(2011), *A study on corporate failure predictions by using audit opinions and accounting firm's characteristics,* Doctoral dissertation, Soongsil University.

Kim, M. J. & Kang, D. K.(2010), An Empirical Analysis of Boosting Neural Networks for Bankruptcy Prediction, *Journal of the Korea Institute of marine Information and Communication Engineering*, 14(1),63-69.

Kim, S. Y.(2006), Prediction of Hotel Bankruptcy Using Multi-variate Discriminant Analysis, Logistic Regression and Artificial Neural Network, *Journal of tourism sciences,* 30(2), 53-75.

Kim, W. K. & Ban, S. S.(1990), A Study on Muti-variate Firm's Fail Prediction Model: Focus on MDA, Logit and Non-parametric Model, *Journal of Management Research.* 1. 5-36.

Kueon, S. I.(2012), *Corporate Credit Analysis*, Seoul: Korea banking Institute press.

Kwon, K. S., Kim, B. J. & Ha, K. S.(2012), An Exploratory Study of Key Success Factors in Public R&D Institutions' Business Supporting Service, *Asia-Pacific Journal of Business Venturing and Entrepreneurship*, 7(1). 225-232.

Lee, K, C.(1993), A Comparative Study on the Bankruptcy Prediction Power of Statistical Model and AI Models : MDA, Inductive Learning, Neural Network, *Journal of the Korean OR and MS Society*, 18(2), 57-81.

Moon, J. G. & Hwangbo, Y.(2014), An Empirical study on a Firm's Fail Prediction Model by Considering whethere there are Embezzlement, Malpractice and the Largest Shareholder Changes or not, *Asia-Pacific Journal of Business Venturing and Entrepreneursip*, 9(1), 119-132.

Nam, J. H.(1998), The Cause of Business's Failure and the Analysis of the Bankruptcy Prediction Model, *Journal of Money & Finance* 12(2), 77-107.

Odom, M & Sharda, R.(1990), A neural networks model for bankruptcy prediction, *Proceedings of the IEEE internation Conference on Neural Networks,* 2, 163-168.

Oh, H. J.(2005), Usefulness of Credit Rating in the Bankruptcy Prediction Model, *Journal of Economic Studies,* 23(2), 173-208.

Ohlson. J. A.(1980), Financial Ratios and the Probabilistic Prediction of Bankruptcy, *Journal of Accounting Research*, 18(1), 109-131.

Park, H. J.(2008), *An empirical study on the failure prediction model of the firms listed on the kosdaq,* Doctoral dissertation, Paichai University.

Park, S. E.(2002), *Multi-variate Discriminant Analysis and Logistic Regression Analysis*, Seoul: Minyoung Sa Press.

Razali, N. M. & Wah, Y. B.(2011), Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests, *Journal of Statistical Modeling and Analytics,* 2(1), 21-33.

Sarah, A. & Shahriar, M.(2011), Compare failure prediction nodels based on feature selection technique: Empirical case from Iran, Procedia Computer Science, 3, 568-573.

중소제조기업의 부실예측모형 비교연구

황보윤 (국민대 글로벌창업벤처대학원 부교수)*
문종건 (건국대 정보통신대학원 겸임교수)**

국 문 요 약

본 연구는 코스닥 시장에 상장 폐지된 중소제조기업의 재무자료를 이용하여 다변량 판별분석모형, 로지스틱회귀분석모형 그리고 인공신경망분석모형을 구축하고 이들의 예측력을 비교분석하였다. 표본기업은 2009년에서 2012년까지 상장 폐지된 83개의 부실기업과 83개의 정상기업 총166개사로 정하였다. 166개사 중에서 무작위로 부실기업50개사와 정상기업 50개사 총100개사를 선정하여 훈련용 표본(training data)으로 모형을 구축하는데 사용하였다. 나머지 66개사는 모형의 예측성과를 평가하기 위하여 검증용 표본(test data)으로 사용하였다. 과거 5년 동안의 재무비율 79개 자료로 T-test를 실시하여 5년 연속 유의미한 변수 9개를 선정하고 각각의 모형을 구축하였다.

T-test 결과, 부실초기에는 주로 수익성지표들이 부실예측에 주요 변수로 나타났으며 부실 후반에 가면서 안정성지표와 현금흐름지표들이 추가로 유의미한 변수로 나타났다. 모형의 예측력을 비교해 보면 훈련용 표본의 경우, 로지스틱회귀분석모형이 가장 높은 분류 정확도를 보였고, 검증용 표본의 경우에는 인공신경망모형이 가장 높은 분류 정확도를 보였다.

본 연구는 첫째, 부실이 서서히 진행된다는 점을 감안하여 T-test를 실시하여 5년 연속 유의미한 변수로 모형을 구축하여 변수의 시계열적인 측면이 고려되었다는 점과, 둘째, 기존 선행 연구들이 정규성을 무시하고 판별분석모형을 구축하였으나, 본 연구가 정규성 여부를 검정하고 모형을 구축하였다는 점이 차별화된다.

본 연구에 따른 정책적 시사점은 부실기업의 징후는 본 논문에서처럼 대체로 재무제표에 나타나기 때문에 회사에 대한 공시서류의 신회성 확보가 중요하다. 따라서 이런 점에서 회계법인 혹은 세무기장 종사자들의 도덕적 해이을 억제할 수 있는 제도적 장치가 강화되어야 할 것이다.

핵심주제어: 기업부실예측, 다변량 판별모형, 로지스틱회귀분석모형, 인공신경망 모형, 중소제조기업

* 제1저자, 국민대학교 글로벌창업벤처대학원 부교수, yun88@kookmin.ac.kr
** 교신저자, 건국대학교 정보통신대학원 겸임교수, moonjames@empas.com