A cross model study of corporate financial distress prediction in Taiwan: Multiple discriminant analysis, logit, probit and neural networks models

Tzong-Huei Lin *

Department of Accounting, National Kaohsiung University of Applied Sciences, No. 415, Jiangong Road, Sanmin District, Kaohsiung City 80778, Taiwan

ARTICLE INFO

Available online 11 June 2009

Keywords:
Financial distress prediction
Corporate governance

ABSTRACT

In 2008, financial tsunami started to impair the economic development of many countries, including Taiwan. The prediction of financial crisis turns to be much more important and doubtlessly holds public attention when the world economy goes to depression. This study examined the predictive ability of the four most commonly used financial distress prediction models and thus constructed reliable failure prediction models for public industrial firms in Taiwan. Multiple discriminate analysis (MDA), logit, probit, and artificial neural networks (ANNs) methodology were employed to a dataset of matched sample of failed and non-failed Taiwan public industrial firms during 1998–2005. The final models are validated using within sample test and out-of-the-sample test, respectively. The results indicated that the probit, logit, and ANN models which used in this study achieve higher prediction accuracy and possess the ability of generalization. The probit model possesses the best and stable performance. However, if the data does not satisfy the assumptions of the statistical approach, then the ANN approach would demonstrate its advantage and achieve higher prediction accuracy. In addition, the models which used in this study achieve higher prediction accuracy and possess the ability of generalization than those of [Altman, Financial ratios—discriminant analysis and the prediction of corporate bankruptcy using capital market data, Journal of Finance 23 (4) (1968) 589–609, Ohlson, Financial ratios and the probability prediction of bankruptcy, Journal of Accounting Research 18 (1) (1980) 109–131, and Zmijewski, Methodological issues related to the estimation of financial distress prediction models, Journal of Accounting Research 22 (1984) 59–82]. In summary, the models used in this study can be used to assist investors, creditors, managers, auditors, and regulatory agencies in Taiwan to predict the probability of business failure.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Predicting corporate failure has been an important research topic in accounting, auditing, and finance for the last three decades [6,16,23,26]. In 2008, financial tsunami started to impair the economic development of many countries, including Taiwan. The prediction of financial crisis turns to be much more important and doubtlessly holds public attention when the world economy goes to depression. As a result, this study examines the predictive ability of the four most commonly used financial distress prediction models.

Since 1997, the Asian financial crisis has deeply affected the development of emerging markets in Asia. In the second half of 1998, there were some “land mine stock” broke out in Taiwan. Even in the US and Europe, a number of corporate scandals have shaken the investors’ confidence in the global financial market in 2001 and 2002. Taiwan, which had been admired as the economic miracle of the world, plays an important role in the global supply chain of electronics products and is an important capital market to global investors as well. Previous literature documented that there existed serious shortcomings of corporate governance in East Asia [8,15,19,28]. For example, in contrast to the Europe and America, the ownership structure is less dispersed and the ultimate controllers often increase their control power by means of pyramid structure and cross shareholdings. In addition, it is common for directors of Taiwanese companies to pledge their shareholdings as the collateral of the loan. If the share price falls, the directors have to increase the collateral. As a result, the financial distress prediction model constructed with developed countries may not apply to the developing countries. This study uses multiple discriminant analysis, logit, probit and artificial neural networks (ANNs) methodology to construct financial distress prediction model and compare the performance of above models with Taiwan data. Besides, I also compare the prediction
ability of the three most commonly used models, multiple discriminant analysis model (MDA, [2]), logit model [22], and probit model [30] with Taiwan data.

By using the data from Taiwan companies, this study documents that the probit, logit, and ANN models which used in this study achieve higher prediction accuracy and possess the ability of generalization. The probit model possesses the best and stable performance. However, if the data does not satisfy the assumptions of the statistical approach, then the ANN approach would demonstrate its advantage and achieve higher prediction accuracy. In addition, the models which used in this study achieve higher prediction accuracy and possess the ability of generalization than those of [2,22,30].

The contributions of this research are documented as the following. First, this study uses multiple discriminant analysis, logit, probit and artificial neural networks methodology to construct financial distress prediction models and compare the performance of above four models with Taiwan data. By doing this, it can help understand and predict financial condition of firms and prevent the firms from insolvency. Second, previous literature documented that the financial distress prediction models of [2,22,30] perform well when using America or Europe data. However, this study finds that the predictive accuracy of the above three most commonly used financial distress prediction models is lower with Taiwan data. And the models which used in this study achieve higher prediction accuracy and possess the ability of generalization than those of [2,22,30]. It can offer deeper understanding about the applicability of the three models mentioned above to the East Asia that has unique social and corporate governance condition. Finally, the models used in this study could be used to assist investors, creditors, managers, auditors and regulatory agencies in Taiwan to predict the probability of business failure, and the evidence of the research can offer the policy maker to evaluate the policy implication of the corporate governance mechanisms and formulate future policy.

The remainder of this paper is organized as follows. Section 2 discusses some literature related to this study, Section 3 discusses sample selection and statistical approach, Section 4 gives the empirical results, and Section 5 provides a summary and conclusion.

2. Literature review

There have been a fair number of previous studies in the field of predicting corporate failure; the more notable published contribution are Beaver [4], Altman [2], Deakin [10], Blum [5], Ohlson [22] and Zmijewski [30].

Before [2], there are several studies devoted to the analysis of a firm’s condition prior to financial difficulties. In almost every case, the methodology was essentially univariate in nature and emphasis was placed on individual signal of impending problems. However, ratio analysis presented in this fashion is susceptible to faulty interpretation and is potentially confusing. As a result, Altman [2] utilized the multiple discriminant analysis in the field of prediction of company financial distress. The author collects 33 failed firms and 33 non-failed firms as the sample, matched by industry and size from 1946 to 1965. Finally, Altman [2] constructs the discriminant function as follows:

\[
Z = 1.2 \times X_1 + 1.4 \times X_2 + 3.3 \times X_3 + 0.6 \times X_4 + 1.0 \times X_5
\]  

(1)

where \(X_1\) is the Working capital/Total assets, \(X_2\) the Retained Earnings/Total assets, \(X_3\) the Earnings before interest and taxes/Total assets, \(X_4\) the Market value of equity/Book value of total debt, \(X_5\) the Sales/Total assets, and \(Z\) the Overall Index.

The variable \(Z\) is named as Altman Z score. Firms will be classified as non-failed group if \(Z\) score is greater than 2.99. If the \(Z\) score is less than 1.81, then the firms will be classified as failed group. However, the area between 1.81 and 2.99 will be defined as the “gray area” because of the susceptibility to error classification. Hence, [2] uses 2.675 as the \(Z\) value that discriminates best between the bankrupt firms and non-bankrupt ones.


Blum [5] used 115 failed firms and 115 non-failed firms as the sample from 1954 to 1968. The failed firms group and non-failed firms group are matched by industry, sales, employees, and fiscal year. Ref. [5] used discriminant analysis and 12 variables to build the financial distress prediction model. The results showed that the correct classification rates are above 70%.

Zmijewski [30] pointed out researchers typically estimate financial distress prediction models on non-random samples. The first bias, a choice-based sample bias, is produced when a researcher first observes the dependent variable and then selects a sample based on that knowledge, that is, the probability of a firm entering the sample depends on the dependent variable’s attributes. The second bias, a sample selection bias, is produced when only observations with complete data are used to estimate the model and when incomplete data observations occur non-randomly. Estimating models on such samples can result in biased parameter and probability estimates if appropriate estimation techniques are not used.

In [30], the author examined the choice-based sample bias by comparing probit estimates for a financial distress model to estimates from an adjusted probit assessment of the model (weighted exogenous sample maximum likelihood probit) across alternative samples designed to induce increasing amounts of the bias. The results clearly demonstrated the existence of a bias for choice-based samples when unadjusted probit is used, decreases in the bias as the sample composition approaches the population composition, and the elimination of the bias using the adjustment procedure. However, the bias did not affect the statistical inferences or the overall classification rates for the financial distress model and the samples tested. In addition, the sample selection bias is examined by comparing probit estimates of a financial distress model conditional on complete data to estimates from a bivariate probit assessment of the model which incorporates the probability of an observation having complete data into the estimation of the financial distress model parameters. The results were qualitatively similar to the choice-based sample results in that a bias was clearly shown to exist, but, in general, it did not appear to affect the statistical inferences or overall classification rates. Overall, these estimation techniques did not appear to provide different qualitative results from the results provided by techniques that assume random sampling. Only the individual group error rates were significantly affected.

Ohlson [22] uses logistic regression approach to construct a financial distress prediction model. The author uses 105 failed firms and 2058 non-failed firms as his sample. The sample period spans from 1970 to 1976. The results show that the logit model predicting corporate failure well. Lo [20] built the financial distress prediction model with multiple discriminant analysis and logit regression. The results showed that if the data satisfied the normal distribution assumption, the predict ability of multiple discriminant analysis model will be higher than logit model, and vice versa.

Coats et al. [9] use the data which were collected from the Standard & Poor’s COMPUSTAT financial database covering the period from 1970 to 1989 to compare the performance of multiple
discriminant analysis and neural network (NN). Although the MDA model produces excellent results for the year of the going-
concern opinion, the NN approach, cascade–correlation (Cascor),
does better by comparison in the earlier years’ classifications,
sustaining reliability. Wilson et al. [24] also find that neural
networks outperformed discriminant analysis in classification
accuracy, especially in the prediction of bankrupt firms. Odom and
Sharda [21] compare the predictive ability of a neural network and
multivariate discriminant analysis models in bankruptcy risk
prediction. They also find that the neural network proved to be
more robust than the discriminant analysis method on reduced
sample sizes.

Altman et al. [3] analyzed the comparison between traditional
statistical methodologies for distress classification and prediction,
that is, linear discriminant analysis (LDA) or logit analysis, with an
artificial intelligence algorithm known as neural networks. The
study analyzed well over 1,000 healthy, vulnerable, and unsound
industrial Italian firms from 1982 to 1992. The results indicate a
balanced degree of accuracy and other beneficial characteristics
between LDA and NN. However, the study points out the problems
of the “black box” NN systems, including illogical weightings of
the indicators and over fitting in the training stage, both of which
negatively impact predictive accuracy.

In Taiwan, Huang et al. [13] use the financial variable and
operating efficiency to predict the business crisis. They find that
the model structured by logit model is better than the one
structured by artificial neural network approach. Lin et al. [18]
uses the methods of DEA-DA, neural network and logistic
regression to establish the prediction models of financial distress.
The results showed that the DEA-DA approach had the highest
prediction ability.

Chiu et al. [7] investigates the performance of enterprise
distress diagnosis by integrating the artificial neural networks
with discriminant analysis technique. They find that the proposed
combined approach predict much accurate and converge much
faster than that the conventional neural network approach. Cheng
et al. [6] adopted the radial basis function network (RBFN) to
construct the prediction model. The performance of the proposed
RBFN is compared to the traditional logit analysis and a back-
propagation neural network and demonstrates superior results to
both the counterparts in predictive accuracy for unseen data. Wu
et al. [26] employed a real-valued genetic algorithm (GA) to
optimize the parameters of support vector machine (SVM) for
predicting bankruptcy. Additionally, the proposed GA-SVM model
was tested on the prediction of financial crisis in Taiwan to
compare the accuracy of the proposed GA-SVM model with that of
other models in multivariate statistics (DA, logit, and probit) and
artificial intelligence (NN and SVM). Experimental results show that
the GA-SVM model performs the best predictive accuracy.

Abdullah et al. [1] compares three methodologies for identifying
financially distressed companies, multiple discriminant
analysis, logistic regression and hazard model. They find that
when the holdout sample is included in the sample analyzed,
MDA has the highest accuracy rate. Lennox [17] examines the
causes of bankruptcy for a sample of 949 UK listed companies
from 1987 to 1994. The paper argues that well-specified logit
and probit models can identify failing companies more accurately than
discriminant analysis (DA).

3. Research design

3.1. Sample selection

This study uses financial statement data from the database
compiled by the Taiwan Economic Journal (TEJ) Data Bank. The
sample periods spans from 1998 to 2005. The financial institutions
are excluded from the dataset because they have special
operating environment and are regulated by the special laws.

Besides, those observations with missing values are also deleted.

This study identifies the financial distress sample from the TEJ,
Taiwan Stock Exchange (TSE) and the Gre Tai Securities Market
(GTSM). Since this study uses matched pair research design, the
failed firms group and non-failed firms group are matched by year,
industry, and size. In addition, this study conducts within sample
prediction and out-of-the-sample prediction, the sample is further
divided into two subsamples: while in the within sample
prediction section, this study uses all sample to construct
financial distress prediction models and compares the perfor-
ance of the estimated models, while in the out-of-the-
sample prediction section, this study uses the data from 1998 to
2003 to construct financial distress prediction models and uses
the data from 2004 to 2005 to compare the performance of the
estimated models.

As for the definition of “failure”, this paper adopts the
definition of Beaver [4] and Blum [5], and so on, that defines
“failure” as the inability of a firm to pay its financial obligations as
they mature. Operationally, a firm is said to have been failed when
any of the following events have occurred: bankruptcy, bond
default, an overdrawn bank account, events signifying an inability
to pay debts as they come due, entrance into a bankruptcy
proceeding, an explicit agreement with creditors to reduce debts,
or be classified as “full delivery stock” by TSE or GTSM. Finally,
the sample is composed of 96 failed firms and 158 non-failed
firms.

3.2. Statistical approach

This study uses multiple discriminant analysis, logit, probit
and artificial neural networks methodology to construct financial
distress prediction model and compare the performance of above
four models. I also compare the prediction ability of the three
most commonly used models, multiple discriminant analysis
model [2], logit model [22], and probit model [30] with Taiwan
data. The statistical approach about multiple discriminant ana-
lysis, logit, probit, and neural networks will be briefly introduced as
follows:

Multiple discriminant analysis is a statistical technique used to
classify an observation into one of several a priori groupings
dependent upon the observation’s individual characteristics. The
primary objective of MDA is to classify and/or make predictions in
problems where the dependent variable appears in qualitative
form, e.g., male or female, bankrupt or non-bankrupt. The primary
advantage of MDA in dealing with classification problems is the
potential of analyzing the entire variable profile of the object
simultaneously rather than sequentially examining its individual
characteristics. Another advantage of MDA is that it can decrease
the analyst’s space dimensionality from the number of different
independent variables to K−1 dimension(s), where K equals
the number of original a priori groups. In bankruptcy prediction
models, researchers are concerned with two groups, consisting of
bankrupt firms and non-bankrupt firms. Therefore, the analysis is
transformed into one dimension.

Assuming there are g groups, then the discriminant function of
the group i is as follows:

$$d_i = w_1x_1 + w_2x_2 + \cdots + w_gx_g + c_i, \quad i = 1, \ldots, g$$

where $x_j$ is the discriminant variable, $w_j$ the discriminant
coefficients, and $c_i$ the constant term.

1 Please see Altman [2] for more discussion.
evaluate the performance, and the train goal is 0.0001. By means of above setting, the model will revise continuously until arriving at the goal. When arriving at the goal, the train process would stop and can be used to predict. In addition, the train epoch is 50 (Fig. 1).

4. Results and discussions

The study employs a relatively large number of financial ratios, proved to be successful in predicting bankruptcy in prior studies. As for the issue of variables used in financial distress prediction model, I first collect 44 financial ratios from the financial statements of sample firms. An important initial step is the identification of any possible differences between the two groups of companies. To achieve this, several main descriptive statistics were calculated (i.e. mean, median, standard deviation, minimum and maximum). Then I perform a univariate logistic regression for each ratio in turn. In order to find the best combination of financial variables for predicting company failure, I further employ stepwise logistic regression analysis. This study also refers to the related literature, considers the attributes of the financial ratios and coefficients of correlation matrix. Finally, this study uses 20 variables to construct financial distress prediction model. The definition of variables is presented in Table 1.

As stated above, this study uses multiple discriminant analysis, logit, probit and artificial neural networks methodology to construct financial distress prediction model, conducts within sample prediction and out-of-the-sample prediction, and compares the performance of four models above. In the within sample prediction section, this study uses all sample to construct financial distress prediction models and compares the performance of the estimated models, in the out-of-the-sample prediction section, this study uses the data from 1998 to 2003 to construct financial distress prediction models and uses the data of 2004 and 2005 to compare the performance of the estimated models.

A widely used measure of the predictive accuracy is the percentage of correct classification of the examples to financial distress or healthy firms [6]. Two kinds of misclassification are Type I error (i.e., a financial distress firm incorrectly classified as a healthy firm) and Type II errors (i.e., a healthy firm being classified as a financial distress firm). The classification of a firm is determined by a cutoff value which is set to balance Type I and Type II errors. A firm with a predicted value greater than this cutoff value is considered as a financial distress firm, otherwise a healthy firm.

4.1. Within sample prediction section

In the within sample prediction section, the estimation period and prediction period are entire sample period, that is, year 1998–2005. This study estimates the financial distress prediction model with the financial data of one year prior to the financial distress and of three year prior to the financial distress, respectively. The estimated financial distress prediction models are presented below:

4.1.1. Using data from one year prior to the financial distress

In this section, this study estimates the financial distress prediction model with the financial data of one year prior to the financial distress. The estimation period and prediction period are entire sample period, that is, year 1998–2005.

(i) Multiple discriminant analysis

I put the data which is one year prior to the financial distress into MDA model and thus the model of financial distress prediction is built. The estimated financial distress prediction models of two groups are presented as below:

(ii) Non-failed companies group:

The estimated financial distress prediction model is presented as below:

\[
Z_0 = -12.067 + 33.314 \times TL/TA - 0.0315 \times MV/TL + 3.998 \times SALES/TA + 1.759 \times CA/CL + 0.01529 \times ROA - 1.148 \times RE/TA + 0.151 \times GP/SALES - 0.0005257 \times IBT/SALES + 0.0008219 \times BD/SALES + 0.01626 \times CFO/CL + 0.002393 \times COD - 0.001310 \times GROGP + 0.0002317 \times GROIBT + 0.02959 \times GROE + 0.005270 \times GRODA + 0.0004530 \times IC/IBI - 0.001941 \times DEBT/EQUITY - 0.006130 \times CONL/EQUITY - 0.01656 \times SALES/AVR + 0.04291 \times COGS/AVI
\] (11)
and MDA model is the worst. When using data from one year prior to the financial distress (grids a and c), the probit model and logit model have the highest average predictive accuracy. In addition, all models have similar predictive accuracy when using data from three years prior to the financial distress (grids b and d). Furthermore, the results of out-of-the-sample prediction showed that the probit, logit, and ANN models performed as good as (or outperformed) the results of within sample prediction. It means that the probit, logit, and ANN models in this study may be generalized to other cases.

In addition, the results of Panel B showed that ANN model achieves higher predictive accuracy in out-of-the-sample prediction. It implied that the ANN model in this study does not have the problem of over fitting and possess the ability of generalization.

However, in the using data from three years prior to the financial distress section of panel B section (grid d), it shows that ANN model performs better than probit and logit model. After screening the data of year 2004 and 2005, it showed that the variance inflation factors (VIFs) of some variables are greater than 10 and violated the assumption of probit and logit model. It may lower the prediction accuracy of probit and logit model. On the other hand, the ANN approach does not require data to meet the assumption and this demonstrates its advantage.

Furthermore, this study further conducts another set of experiments. The approach this paper adopts is to conduct the out-of-the-sample prediction uses the MDA, probit, and logit model with the variables used in Altman [2], Ohlson [22], and Zmijewski [30] in order. The results are presented in Table 3.

The empirical results of Table 3 are reported as follows, the prediction accuracy of financial distress prediction models with the financial data of the year prior to the financial distress is better than that of financial distress prediction models with the data from three years prior to the financial distress. Overall, the probit model achieves highest prediction accuracy. Compare the results of Table 3 with the results of Panel B of Table 2, it could be found that the financial distress prediction models proposed in this study perform better.

To sum up, the probit, logit, and ANN models which used in this study achieve higher prediction accuracy and possess the ability of generalization. The probit model possesses the best and stable performance. However, if the data does not satisfy the assumptions of the statistical approach, then the ANN approach would demonstrate its advantage and achieve higher prediction accuracy. In addition, the models which used in this study achieve higher prediction accuracy and possess the ability of generalization than those of [2,22,30]. In summary, the models used in this study can be used to assist investors, creditors, managers, auditors and regulatory agencies in Taiwan to predict the probability of business failure.

5. Conclusion

Predicting corporate failure has been an important research topic in accounting, auditing, and finance for the last three decades [6,16,23,26]. In 2008, financial tsunami started to impair the economic development of many countries, including Taiwan. The prediction of financial distress or bankruptcy has received considerable attention for many user groups such as investors, creditors, regulators, and auditors. As a result, this study examines the predictive ability of the four most commonly used financial distress prediction models.

Since 1997, the Asian financial crisis has deeply affected the development of emerging markets in Asia. In the second half of 1998, there were some “land mine stock” broke out in Taiwan. In 2001 and 2002, a number of corporate scandals in the US and European have shaken the investors’ confidence in the global financial market.

This study examined the predictive ability of the four most commonly used financial distress prediction models and thus constructed reliable failure prediction models for public industrial firms in Taiwan. Multiple discriminate analysis, logit, probit, and artificial neural networks methodology were employed to a dataset of matched sample of failed and non-failed Taiwan public industrial firms during 1998–2005. The final models are validated using within sample test and out-of-the-sample test, respectively. The results indicated that the probit, logit, and ANN models which used in this study achieve higher prediction accuracy and possess the ability of generalization.

This study finds that the predictive accuracy of the three most commonly used financial distress prediction models is lower with Taiwan data. In the future, it could integrate the corporate governance variables with financial variables into the financial

---

*Note: a. The estimation period is year 1998–2003, the prediction period is year 2004 and 2005. b. The hit ratio (%) means that the correct prediction rate. That is, the number of sample companies which are correctly classified/the number of sample companies. Type I error is the misclassification of a failed firm as non-failed and the Type II error is the misclassification of a non-failed firm as failed. c. MDA, multiple discriminant analysis.*
distress prediction models and compare the predictive ability with this study. In addition, it could use different sample to evaluate the relative performance of MDA, probit, logit, and ANNs.

Appendix A

The variables used in the Altman [2], Ohlson [22], and Zmijewski [30].

<table>
<thead>
<tr>
<th>Variables</th>
<th>MDA</th>
<th>RE/TA</th>
<th>WC/TA</th>
<th>MV/TL</th>
<th>ROA</th>
<th>SALES/TA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probit</td>
<td>TL/TA</td>
<td>CA/CL</td>
<td>ROA</td>
<td>CL/CA</td>
<td>ROA</td>
<td></td>
</tr>
<tr>
<td>Logit</td>
<td>TL/TA (OZ)</td>
<td>Total debt/Total assets</td>
<td>MV/TL (A)</td>
<td>Market value of equity/Book value of total debt</td>
<td>CA/CL (OZ)</td>
<td>Current assets/Current liabilities</td>
</tr>
</tbody>
</table>

References


Tzong-Huei Lin is assistant professor of Accounting at the National Kaohsiung University of Applied Sciences (Taiwan). He completed his degree of Doctor of Accounting in 2004 at the National Taiwan University. His research interests include corporate governance, financial accounting, auditing, and accounting and taxation of derivatives.