PREDICTION OF DEFAULT PROBABILITY FOR CONSTRUCTION FIRMS USING THE LOGIT MODEL

H. Ping TSERNG[a], Po-Cheng CHEN[a], Wen-Haw HUANG[b], Man Cheng LEI[a], Quang Hung TRAN[a]

[a]*Department of Civil Engineering, National Taiwan University, No. 1 Roosevelt Rd., Sec. 4, Taipei, Taiwan*
[b]*Long Reign Development Co., 16F-2, No. 76, Sec. 2 Dunhua S. Road, Taipei, Taiwan*

Abstract. Recently, the high incidence of construction firm bankruptcies has underlined the importance of forecasting defaults in the construction industry. Early warning systems need to be developed to prevent or avert contractor default; additionally, this evaluation result could facilitate the selection of firms as collaboration or investment partners. Financial statements are considered one of the key basic evaluation tools for demonstrating firm strength. This investigation provides a framework for assessing the probability of construction contractor default based on financial ratios by using the Logit model. A total of 21 ratios, gathered into five financial groups, are utilized to perform univariate logit analysis and multivariate logit analysis for assessing contractor default probability. The empirical results indicate that using multivariate analysis by adding market factor to the liquidity, leverage, activity and profitability factors can increase the accuracy of default prediction more than using only four financial factors. While considering the market factor in the multivariate Logit model, clear incremental prediction performance appears in 1-year evaluation. This study thus suggests that the market factor comprises important information to increase the prediction performance of the model when applied to construction contractors, particularly in short-term evaluation.

Keywords: default probability, financial ratios, Logit model, bankruptcy prediction.

Introduction

The construction industry is always the vanguard of national economic development. This industry is the foundation and connector between other industries. However, contractors are facing numerous difficulties and a highly competitive environment. Historical U.S. data indicates that the failure rate among construction firms has reached a critical level (Kangari *et al.* 1992). Numerous risks occur during the life cycle of construction objects (Zavadskas *et al.* 2010). Meanwhile, some highlighted characteristics of the construction industry and construction projects, such as uniqueness, long term investment, large investment capital, etc. result in the industry having unique financial characteristics. One of the required conditions of a competent construction contractor is the use of proper processes and construction project completion (Plebankiewicz 2010). High default risk may prevent contractors from completing a construction project. Tserng *et al.* (2011a) also indicated the importance of identifying potentially failing contractors; thus enabling clients to avoid awarding contracts to contractors likely to default. In order to monitor the financial risk or any adverse effects on joint venture projects or cooperative projects, prime construction contractors are concerned about the financial health of their sub-contractors and vice versa (Tserng *et al.* 2011b). Recently, interest has grown in predicting the default likelihood of contractors.

Contractor financial ratios, which can be calculated from information contained in financial statements, are useful data for predicting company default probability. Financial statements summarize the value of a company at the end of a specific period, as well as assessing company operations, indicating and highlighting performance in all related sectors (Ross *et al.* 2010). Financial ratios serve as the fundamental basis for evaluating company financial capabilities, and provide useful information for predicting firm likelihood of default. The usages of financial ratios in construction companies can provide an early warning mechanism that offers an effective monitoring tool to avoid continuing poor corporate performance or insolvency. Beaver (1966) and Altman (1968) were the first to use financial ratios to analyse and predict the default probability, and their work was subsequently continued by Abidali and Harris (1995), and Edum-Fotwe *et al.* (1996). These studies focused on applying statistical methods to financial ratios analysis to determine the likelihood of corporate default. Furthermore, most of the previous investigations examined industries in general, rather than specifically focusing on construction industry concerns.

Corresponding author: H. Ping Tserng
E-mail: hptserng@ntu.edu.tw

Logit regression analysis performs well in identifying the likelihood of an outcome belonging to one of the two discrete classes of failure versus non-failure. Additionally, no assumption is made regarding the distribution of the independent variables, and no normally distributed requirement of multivariate variables. Owing to less demanding conditions, the logit regression is more useful in practice. The Logit model is widely used in numerous domains, especially in relation to credit risk and health sciences. In credit risk analysis, the Logit model is a common technique used in credit scoring to determine default probability. Financial institutions build up the credit scoring model based on consumer application and credit reference agency data (Ohlson 1980; Bellotti, Crook 2009). In health sciences, the Logit model is a multivariable analysis tool for modelling dichotomous outcomes. The Logit model is widely used and appropriate for models of disease state (diseased/healthy) and decision making (yes/no) (Bagley *et al.* 2001). Moreover, the strong development of information technology and the wide application of the Internet simplify the collection of contractor financial data from the biggest market, the U.S., which has a long history and modern, convenient systems of storing data. This study uses the Logit model to analyse the relationship of financial ratios to the default probability of construction firms. This investigation estimates and quantifies the default risk of construction firms by applying the Logit model to analyse historical data from the financial statements of firms participating in the US market.

The remainder of this investigation is divided into six sections: Section 1 reviews the literature on construction industry characteristics and research on default probability prediction; Section 2 then introduces the research methods; next, Section 3 describes data set and input selections; Section 4 presents single variable analysis and empirical results; Section 5 then describes multivariate analysis and the empirical results; finally, the final section presents conclusions.

1. Literature review

1.1. General and financial characteristics of construction industry

The construction industry differs from other industries in its organization and products, stakeholders, large projects, processes, and operating environment. Construction industry activities include the building of new structures, including site preparation, as well as additions and modifications to existing ones. The construction industry also includes maintenance, repair, and improvements of these structures. Since the construction industry includes numerous sub-sectors, it is difficult to clearly define the construction industry, despite general agreement about its characteristics. Barrie and Paulson (1992) identified the general characteristics of the construction industry as being that most of construction projects are unique, have long life time expectancy, and involve a long cycle from design to production.

Due to the distinctive operational behaviours of the construction industry, its financial characteristics also differ from other industries. Some of these characteristics are as follows: (a) The construction industry requires large amounts of cash; (b) Contractors must always implement several projects simultaneously, causing them to tend to decrease the payback period, meaning short term finance is invariably required; (c) Inventories of construction firms occupy a large proportion of total assets since they include construction in progress and materials. As a consequence, construction firms have high current ratio and low quick ratio (Barrie, Paulson 1992); (d) Contractors always have a lot of valuable machines and equipment, for example, cranes, boats, shaped steel store, that cause the book value of total asset to be very high, and usually under the affection of depreciation.

1.2. Recent research on default probability prediction

Since the construction industry has its own general characteristics and financial risks, recently numerous investigations have used financial ratios to predict construction company failure. Several approaches have been adopted, including multiple discriminant analysis (MDA) (e.g. Mason, Harris 1979; Severson *et al.* 1994; Abidali, Harris 1995), ratio analysis (e.g. Chan *et al.* 2005; Huang 2009), and logit regression (e.g. Severson *et al.* 1994; Russell, Zhai 1996; Tserng *et al.* 2011a). Among these methods, the logit regression analysis performs well in identifying the likelihood of an outcome being in one of the two discrete classes: failure versus non-failure. Besides, no assumptions are made with respect to the distribution of the independent variables, and the multivariate variables are not required to be normally distributed. These less demanding condition increase the usefulness of the logit regression.

2. Methodology

This study performs the univariate and multivariate ratio analysis by using the Logit model to predict construction contractor default. The financial ratios are classified into five groups, and the best ratio of each financial group is selected for combination into four multivariate ratio analyses. In the multivariate ratio analysis process, the market to book ratio is especially considered to assess the impact of the market factor on the probability of construction firm insolvency. Three maturities of the forecasting time are considered in default prediction, namely one, two, and three years before the default. To avoid over-fitting in the result, an iterative method called leave-one-out cross-validation (LOOCV) is adopted in the modelling. For model evaluation, the receiver operation characteristic (ROC) curve is utilized to determine the goodness of the Logit models via single variables or the combinations of multivariables.

2.1. The Logit model

Ohlson (1980) pioneered the approach of using the logistic regression model to predict business bankruptcy. Later, researchers pursued using the logistic regression

model to predict construction contractor performance (Jaselskis, Ashley 1991; Russell, Jaselskis 1992; Severson *et al.* 1994). A binary logistic regression model can effectively show the correlation between independent variables of binary response and a group of explanatory variables (Luo, Lei 2008). The output of the logistic regression model is between 0 and 1, indicating two separate events (default and non-default). The model is suitable for demonstrating the likelihood of occurrence of an event by probabilities. The probability *P* of y = 1 is the research object, while the independent variables X_1, X_2 … X_k are the explanatory variables of firm default probability (Luo, Lei 2008). Finally, Eqn (1) shows the logistic function:

$$P = (y = 1 \mid \textit{Explanatory variables}) = \frac{1}{1+e^{-z}}, \quad (1)$$

where: *P* denotes the default probability; and *Z* represents the linear regression of explanatory variables. $Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + ... + \beta_k X_k$.

Generally, parameters **β** can be estimated using the maximum likelihood method to maximize the log-likelihood function:

$$\log L(\beta) = \sum_{i=1}^{n} y_i \log p_i + (1 - y_i) \log(1 - p_i), \quad (2)$$

where: p_i: depends on the covariates X_i and a vector of parameters β through the Logit transformation of equation $z_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + ... + \beta_k X_{ki}$; y_i: a qualitative variable comprising one of two values, with 0,1 representing the non-default and default event, respectively; X_i: the explanatory variables of firm default probability.

2.2. Model validation

The derived logit model has to be validated to strictly test its prediction ability. A validation process called leave-one-out cross-validation (**LOOCV**) is used to verify the fitness of the derived model in out-of sample forecasts, and the model prediction ability is evaluated using ROC curve.

2.2.1. ROC curve

Assessing the model misclassification rate by setting a specific cut-off point is a traditional method of model validation in default prediction. The cut-off point setting depends on the assumed misclassification costs of Type I error (the actual is default and the prediction is non-default) and Type II error (when the actual is non-default and the prediction is default). However, it is difficult to assume the misclassification costs when making an assessment, and the cut point setting varies among different models creating potential difficulties in comparison (Bellotti, Crook 2009). To fully perform model prediction, this study adopts a receiver operation characteristic (ROC) curve for the comparison. Table 1 lists the different fractions of the ROC curve.

Table 1. The fractions of ROC curve

Prediction	Actual outcome	
	Default	Non- Default
Positive (default)	True Positive (TP)	False Positive** (FP)
Negative (non-default)	False Negative* (FN)	True Negative (TN)

Note: *False Negative is known as type I error.
  **False Positive is known as type II error.

For every possible cut-off point, positive means the prediction model is identified as the default, while negative indicates that the prediction is non-default. Furthermore, True Positive (TP) shows that the actual default is correctly classified as positive, while the actual non-default correctly classified as negative is termed the True Negative (TN). Meanwhile, False Negative (FN), which is generally known as type I error, occurs when the actual default is classified as negative. In contrast to False Negative, False Positive (FP) is termed as the type II error that shows the actual non-default mistakenly classified as default (Akobeng 2007).

In the ROC curve, the true positive rate (Sensitivity) is plotted as a function of the false positive rate (1-Specificity) for the full range of possible cut-off points. That is, each point on the ROC plots the type II error versus one minus the type I error corresponding to each possible cut-off point.

The ROC graph is applied in the area under the ROC curve, called AUC, which is calculated as the proportion of the area below the ROC relative to the total area of the unit square. A model with prefect discrimination power has a ROC that passes through the upper left corner, resulting in an AUC equalling 1; while a ROC that is a diagonal line from the bottom left corner to the upper right corner indicates a random model, resulting in an AUC of 0.5. Therefore, the closer the AUC is to 1, the higher the differentiation ability of the model. The AUC value can help to verify the fit of the model results to the actual event, and therefore is utilized as the selection criteria for determining the importance of the single variables, and for showing the prediction performance of the multivariable models.

2.3. Cross-validation

One of the purposes of this study is to establish a model that relies on the Logit model to predict the default probability of the construction firms. Constructing a prediction model requires the cross-validation step to avoid the over-fitting problem. The over-fitting problem indicates that the established model only perform well for in-sample data, but fails to make accurate predictions when using out-of-sample data. This investigation uses a crucial process, leave-one-out cross-validation (**LOOCV**) for the model construction. Leave-one-out cross-validation (**LOOCV**) involves using a single observation from the original sample as the validation data, and the remaining observations as the training data. This process is repeated until each observation in the sample has been

used once as the validation data. The model validation result is taken as the average of the (Area Under Curve) AUC in each round.

3. Data collection

This study empirically investigates a large cross-section of construction contractors. Data is collected from the Compustat Industrial files and the Center for Research on Securities Prices (CRSP). This investigation emphasizes on construction contractors with December fiscal year-ends by choosing firms with SIC codes ranging between 1,500 and 1,799. Similar to the studies of Severson *et al.* (1994) and Russell and Zhai (1996), the sample contractors are drawn from three construction categories:

- Major Group 15: Building construction, general contractors, and operative builders;
- Major Group 16: Heavy construction other than building construction contractors;
- Major Group 17: Construction special trade contractors.

The study population comprises 119 construction contractors, of which 29 defaulted. The observation period ranges from 1970 to 2006. The following are principles of data screening:

1) To consider the long term impact of the financial ratios, the selected construction contractors require at least five years of continuous data in Compustats;

2) Using a broad definition of bankruptcy, default events are defined by CRSP delisting codes of 400, 550 and 585, which are referred to as the reasons of bankruptcy, liquidity or poor performance, respectively;

3) The selected financial ratios can encompass all aspects of a contractor finance situation, including the

liquidity, profitability, leverage, activity of a firm and even the market factor. Furthermore, these ratios have been used in at least two studies dealing with construction finance.

The final sample consists of 87 contractors, 29 of which defaulted while 58 are non-defaulted, including 1560 firm-year observations. Twenty-one financial ratios of each observation are collected and used to implement the Logit models. Table 2 lists the definitions and summary statistics of these ratios.

4. Single variable analysis and empirical result

Through applying the Logit model for each single variable with validation process, the estimated default probabilities of each sample can be assessed corresponding to each univariate Logit models. The ROC curve is utilized to demonstrate the prediction performance of 21 univariate logit analyses constructed by different single variables. Table 3 lists the AUC of each maturity of 21 single variables. AUCs exceeds 0.5 (AUC = 0.5 indicates a random model) in the case of six variables (Var 17, Var 11, Var 7, Var 3, Var 15, Var 1), where AUCs exceeds 0.5 in years 1, 2 and 3. The AUCs of return on assets ratio (ROA) exceeds 0.7; therefore, the ROA performs best in default prediction in univariate logit analysis.

For the multivariate analysis, the variable selection criteria are based on the AUC results of 21 variables. The results of the AUC of single financial ratios that exceed 0.5 are considered for multivariate analysis, and thus Var 1 (current ratio) and Var 3 (net working capital to total asset) in the liquidity group are selected. The selected financial ratio of the leverage group is Var 7 (debt ratio). The Var 11 (accounts payable turnover ratio) and

Table 2. Statistical characteristics of the selected financial ratios

Group	Ratios	Sign	Mean	Standard deviation
Liquidity measurement ratios	Current ratio	Var 1	3.414	5.047
	Quick ratio	Var 2	1.396	1.648
	Net working capital to total asset	Var 3	0.357	0.260
	Current asset to net assets	Var 4	1.183	1.099
Financial leverage ratios	Total liabilities to net worth	Var 5	2.958	26.215
	Retained earnings to sales	Var 6	0.863	36.550
	Debt ratio	Var 7	0.609	0.193
	Times interest earned ratio	Var 8	50.808	315.941
Asset utilization or turnover ratios	Revenue to networking capital	Var 9	5.215	75.651
	Accounts receivable turnover	Var 10	68.785	249.275
	Accounts payable turnover	Var 11	25.774	83.978
	Sales to net worth	Var 12	8.446	130.550
	Quality of inventory	Var 13	19.629	54.336
	Fixed assets to net worth	Var 14	1.249	10.384
	Turnover of total assets	Var 15	1.571	0.995
	Revenue to fixed assets	Var 16	10.712	22.687
Profitability indicator ratios	Return on assets (ROA)	Var 17	0.040	0.120
	Return on equity (ROE)	Var 18	−0.053	3.758
	Return on sales (ROS)	Var 19	1.178	35.831
	Profits to networking capital	Var 20	0.122	7.919
Market value ratios	Book to market ratio	Var 21	0.946	0.321

Table 3. The result of AUC of 21 univariate Logit models in 1, 2, 3 year ahead bankruptcy prediction

Rank	Var	Ratios	AUC		
			1 year	2 year	3 year
1	Var 17	Return on assets (ROA)	0.7830	0.7667	0.7388
2	Var 11	Accounts payable turnover	0.6922	0.7005	0.7119
3	Var 7	Debt ratio	0.6529	0.6546	0.6194
4	Var 3	Net working capital to total asset	0.6123	0.6065	0.5942
5	Var 15	Turnover of total assets	0.5989	0.5865	0.5843
6	Var 1	Current ratio	0.5832	0.6107	0.5951
7	Var 2	Quick ratio	0.5600	0.5687	0.4990
8	Var 21	Book to market ratio	0.4806	0.4845	0.4809
9	Var 18	Return on equity (ROE)	0.3845	0.4231	0.4299
10	Var 4	Current asset to net assets	0.3730	0.4529	0.5076
11	Var 13	Quality of inventory	0.2741	0.2631	0.2431
12	Var 8	Times interest earned ratio	0.2739	0.3423	0.2365
13	Var 12	Sales to net worth	0.2607	0.4043	0.4171
14	Var 6	Retained earnings to sales	0.2515	0.2484	0.2415
15	Var 19	Return on sales (ROS)	0.1901	0.3229	0.3781
16	Var 10	Accounts receivable turnover	0.1620	0.1780	0.1791
17	Var 14	Fixed assets to net worth	0.1618	0.1864	0.1963
18	Var 5	Total liabilities to net worth	0.1573	0.2040	0.2185
19	Var 9	Revenue to networking capital	0.1350	0.2127	0.0821
20	Var 20	Profits to networking capital	0.1235	0.1393	0.1022
21	Var 16	Revenue to fixed assets	0.0302	0.2739	0.3838

Var 15 (the turnover of the total assets) of the turnover group are also valid, and the last selected financial ratio is Var 17 (return on asset-ROA) of the profitability group. Besides, Var 21 (book to market ratio) is also considered to make a general assessment, and also to investigate the influence of market factor for default prediction. Table 4 lists the selected financial ratios used in the multivariable analysis.

Table 4. The selected financial ratios for multivariable logit analysis

Groups	Var	AUC		
		1 year	2 year	3 year
Liquidity group	Var 1	0.5832	0.6107	0.5951
	Var 3	0.6123	0.6065	0.5942
Leverage group	Var 7	0.6529	0.6546	0.6194
Turnover group	Var 11	0.6922	0.7005	0.7119
	Var 15	0.5989	0.5865	0.5843
Profitability group	Var 17	0.7830	0.7667	0.7388
Market group	Var 21	0.4806	0.4845	0.4809

5. Multivariate analysis and the empirical result

To perform the multivariate analysis, this study first combines only one single ratio of each financial group (Liquidity group, Leverage group, turnover group, and Profitability group) to establish two multivariate Logit models (models 1 and 3). Including one single ratio of each financial group in a combination can help avoid repetition and close correlation between ratios in the same group. Additionally, market factor is considered and combined with other financial groups to create two multivariate Logit models (models 2 and 4). Therefore, four combinations are selected to perform the multivariate analysis with logit function. Table 4 lists the combinations of four models.

For the multivariate analysis, the value of a combo, Z, is calculated by multiplying the regression coefficient and ratio value correlatively, then transferred into default probability using the Logit model. Table 6 lists the coefficient estimates for the logit regressions.

Table 6 shows that Var 7 is statistically significant ($p < 0.01$) in model 2, model 3 and model 4; meanwhile, Var 17 and Var 21 are significant ($p < 0.01$) in all of the 4 models. Thus, these three ratios have better default prediction ability.

One of the major characteristic of logit function is that default probability increases with Z value. Therefore, the ratio with a positive coefficient with Z value also exhibits with a positive correlation with default probability. In Table 6, the estimated coefficients are consistent with the common understanding that high debt ratio (Var 7) and high book to market value (Var 21) increase default probability (with a positive sign), whereas the other five ratios (Var 1, Var 3, Var 11, Var 15, Var 17) exhibits an opposite effect.

ROC curve area and accuracy

The assessed default probabilities with each combo in years 1, 2 and 3 are used in the model validation step. This study employs ROC curve to evaluate the discriminatory ability of established multivariate Logit models. Discriminatory ability means how effectively the models can rank companies correctly from the riskiest one to the safest according to the model's default probabilities.

Table 5. The combinations of models in the multivariate logit analysis

	Combo 1		Combo 3	
	Var3-Net working capital to total asset		Var1-Current ratio	
	Var7-Debt ratio		Var7-Debt ratio	
	Var17-Return on asset		Var17-Return on asset	
	Var11-Accounts payable turnover		Var15-Turnover of total assets	
Variables	Combo 2		Combo 4	
	Var3-Net working capital to total asset		Var1-Current ratios	
	Var7-Debt ratio		Var7-Debt ratio	
	Var17-Return on asset		Var17-Return on asset	
	Var11-Accounts payable turnover		Var15-Turnover of total asset	
	Var21-Book to market ratio		Var21-Book to market ratio	

Table 6. The coefficient estimates for the logit regressions

Panel A: Prior one-year:

	Var1	Var3	Var7	Var11	Var15	Var17	Var21	Constant
Model 1		−0.636**	2.575	−0.042***		−2.433***		−4.958
Model 2		−0.522	4.169***	−0.029		−3.504***	2.297***	−8.629***
Model 3	−0.108		3.210***		−0.438**	−3.106***		−5.241***
Model 4	−0.109		4.512***		−0.333	−3.915***	2.216***	−8.570***

Panel B: Prior two-year:

	Var1	Var3	Var7	Var11	Var15	Var17	Var21	Constant
Model 1		−0.366	2.666***	−0.037*		−3.339***		−4.410***
Model 2		−0.433	3.591***	−0.030*		−4.173***	1.782***	−6.909***
Model 3	−0.109		2.994***		−0.328**	−4.113***		−4.486***
Model 4	−0.118*		3.767***		−0.267*	−4.797***	1.709***	−6.801***

Panel B: Prior three-year:

	Var1	Var3	Var7	Var11	Var15	Var17	Var21	Constant
Model 1		−0.209	1.574*	−0.055***		−3.755***		−3.012***
Model 2		−0.292	2.047***	−0.051***		−4.184***	1.140***	−4.488***
Model 3	−0.084**		1.978***		−0.268**	−4.793***		−3.456***
Model 4	−0.088**		2.393***		−0.228*	−5.167***	1.143***	−4.913***

***significant at the 1% level (two side test). **significant at the 5% level (two side test). *significant at the 10% level (two side test).

Table 7 lists the AUC of four multivariate Logit models with different years. All AUCs of the four multi-variable Logit models exceed 0.7 in construction contractor default prediction. Besides, the ROC curves of the four multivariable Logit models are shown in Figures 1, 2, and 3.

Table 7. Result of AUC of four multivariate Logit models

Multivariate	AUC		
Logit models	1 year	2 year	3 year
Model 1	0.7107	0.7475	0.7630
Model 2	0.7843	0.7846	0.7654
Model 3	0.7433	0.7698	0.7503
Model 4	0.7918	0.7951	0.7514

Several findings can be obtained by comparing the results among the models: First, as Table 7 shows, the AUCs of model 2 (0.7843; 0.7846; 0.7654) outperform those of model 1 (0.7107; 0.7475; 0.7630) in default prediction over a one to three year timeframe. Similarly, the AUCs of model 4 (0.7918; 0.7951; 0.7514) also exceed the those of model 3 (0.7433, 0.7698, 0.7503), indicating that combining the book to market ratio (Var 21) with the financial factors (Var 3, Var 7, Var 17, Var 11 or Var 1, Var 7, Var 17, Var 15) for construction contractor default prediction increases the model performance.

Second, the improvements obtained from adding the book to market ratio (Var 21) in multivariate logit analysis are obvious in one year prediction (from AUC = 0.7107 for Model 1 to AUC = 0.7843 for Model 2, and from AUC = 0.7433 for Model 3 to AUC = 0.7918 for Model 4). These differences can also be observed in Figures 1, 2, and 3. Because of Models 2 and 4 considering the book to market ratio (Var 21), which is an indicator of present firm situation, Model 2 and Model 4 offer a means of making fast and precise short-term predictions.

Third, model 3 slightly outperforms Model 1. The AUC of Var 3 and Var 11 being slightly higher than those of Var 1 and Var 15 in the univariate analyses, demonstrating that the Var 1 and Var 15 adopt and provide a better general view of the activity and directly relate to the short term debt of a firm in multivariate analysis. The empirical result of model 3 also indicates that 2-year prediction is the most suitable of predicting default probability of construction firms.

Fourth, Model 4 is the best model in this study that could be used to predict default probability. Like Models 1

and 2, Model 4 is based on the ratios of the Model 3 but with the addition of the market factor (Var 21). The AUC of Model 4 performs stably in one-year and two-year prediction (AUC = 0.7918 and 0.7951), suggesting that Model 4 obtains a better default prediction in the short term.
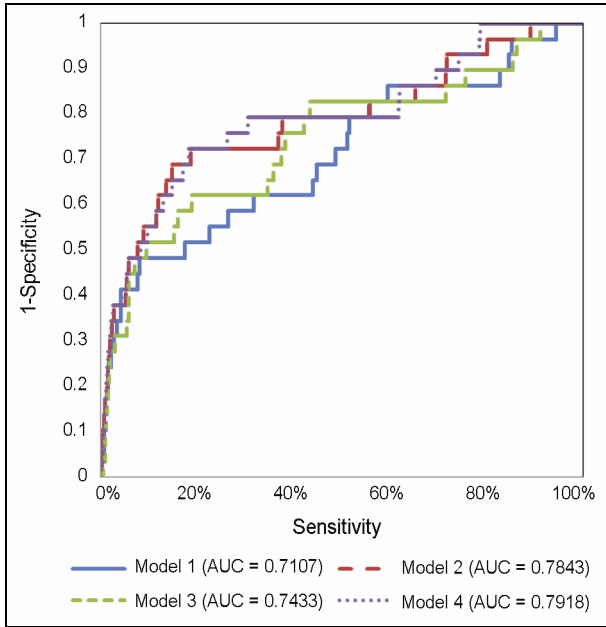

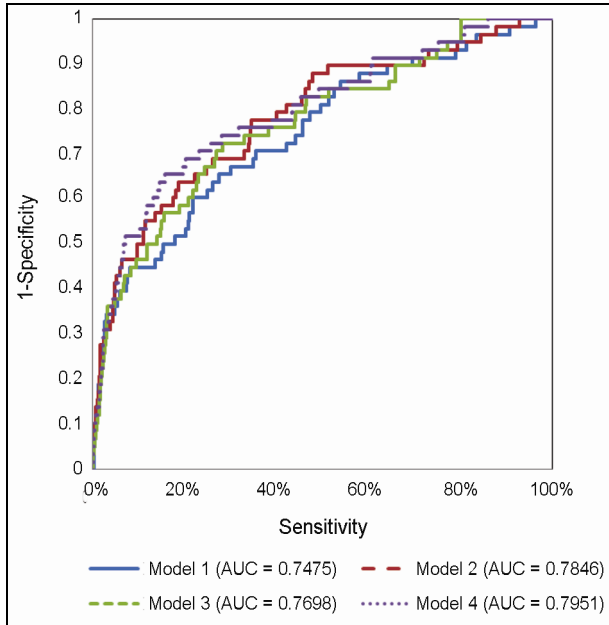Fig. 1. ROC curves for multivariate Logit models on prior one-year

Fig. 2. ROC curves for multivariate Logit models on prior two-year

Conclusions and suggestions

Employing the Logit model to analyse construction firm financial ratios in the U.S. market, and to quantitatively measure firm default probability based on actual firm financial situation can provide extremely useful information for relevant stakeholders, such as clients, lending institutions, and surety underwriters. The models presented in this
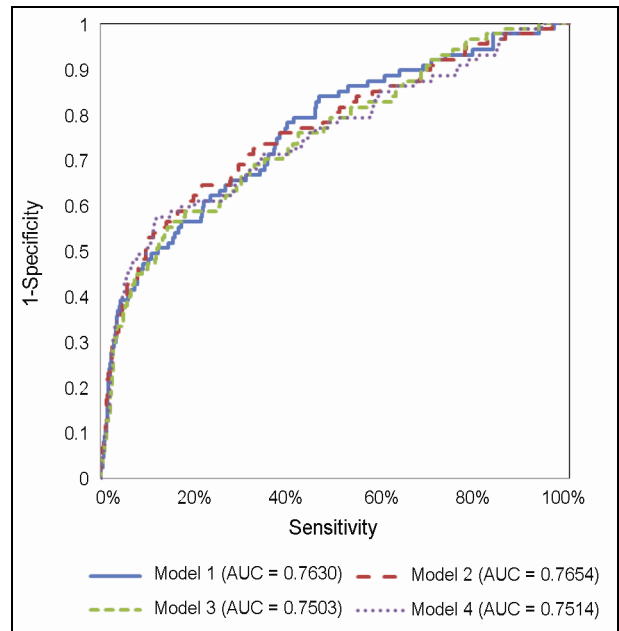
Fig. 3. ROC curves for multivariate Logit models on prior three-year

study are validated to verify their effectiveness by comparison with actual default events.

The empirical result of the single variable model was used to demonstrate the impact of twenty-one financial ratios on the construction company insolvency. The result indicates that liquidity plays an important role in the predicting of the default probability for construction firms. Besides, for those ratios directly related to short term and long term company debt, such as debt ratio (Var 7) and accounts payable ratio (Var 11), which also strongly affect the default probability. The validation process also indicates that the return on assets ratio (Var 17) is the most suitable variable for univariate analysis. However, numerous reasons may exist for a default, reflected in various financial ratios. Therefore, basing such a prediction on just one financial ratio may be insufficient to fully capture the default phenomena. Most of the 21 variables have AUCs below 0.5. This indicates that using the univariate logit model for default prediction cannot fully satisfy the need for predicting default by construction contractors. Notably, an unexpected finding of univariate logit analysis was that using ROA (return on assets ratio) alone can achieve decent performance in default prediction for the US contractors. ROA is a key profit indicator for investors. In the United States, accounting principles and asset recognition criteria are well developed, and ROA is considered to more accurately reflect the real situation of a firm rather than other countries. Since this investigation aims to develop a general estimation method for predicting construction contractor default, observing only the profitability of the construction contractors for default prediction may lack consideration of other aspects of financial abilities. Therefore, multivariate logit analyses with different financial ratios are required to enable a whole financial indicator to predict construction contractor default.

This study is the first to use multivariate analysis to assess the simultaneous impacts of the four financial aspects, namely liquidity, leverage, activity and profitability, on construction firm default probability. Subsequently, the market factor (book to market ratio) is taken into account. The empirical results indicate that the proposed method is most accurate when combining multiple ratios in the analysis. Using a combination of valid ratios can help capture all aspects of the risk of bankruptcy. Notably, including market factor enhanced the prediction performance when it was considered along with other financial factors.

It is noteworthy that this paper places much emphasis on the data collection and screening parts to obtain available construction contractors samples in the United States for providing an overall investigation in default prediction. All adopted financial data follow a unified set of standards and principles – the Generally Accepted Accounting Practice (GAAP), and the default events are defined by CRSP publicly. Besides, using multivariate logit analyses with five financial aspects (liquidity, leverage, turnover, profitability and market) enable a whole financial indicator to predict construction contractor default. For construction contractor default prediction in other world markets, this paper could serve as a basis for investigation. Further research could also collected financial data that follow the same accounting principles and asset recognition criteria, i.e. international financial reporting standards (IFRS). With the same accounting standards, the collected data are consistent thus providing representative results of the regions. Furthermore, the performance of financial ratios in predicting construction default among different countries or regions would also be an interesting global topic for further research. Finally, this study only surveyed public listed construction contractors with SIC codes between 1,500 and 1,799. For further study, this study recommends expanding the sample to include non-publicly listed contractors and broadening the research domain to include other financial ratios, particularly market-related ones.

Acknowledgement

References

Abidali, A. F.; Harris, F. C. 1995. A methodology for predicting company failure in the construction industry, *Construction Management and Economics* 13(3): 189–196. http://dx.doi.org/10.1080/01446199500000023

Akobeng, A. K. 2007. Understanding diagnostic tests 3: receiver operating characteristic curves, *Acta Paediatrica* 96(5): 644–647. http://dx.doi.org/10.1111/j.1651-2227.2006.00178.x

Altman, E. I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *Journal of Finance* 23(4): 589–609. http://dx.doi.org/10.1111/j.1540-6261.1968.tb00843.x

Bagley, S. C.; White, H.; Golomb, B. A. 2001. Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain, *Journal of Clinical Epidemiology* 54(10): 979–985 http://dx.doi.org/10.1016/S0895-4356(01)00372-9

Barrie, D. S.; Paulson, B. C. 1992. *Professional construction management: including CM, design-construct, and general contracting*. 3rd ed. McGraw-Hill.

Beaver, W. H. 1966. Financial ratios as predictors of failure, *Journal of Accounting Research* 4: 71–111. http://dx.doi.org/10.2307/2490171

Bellotti, T.; Crook, J. 2009. Support vector machines for credit scoring and discovery of significant features, *Expert Systems with Applications* 36(2): 3302–3308. http://dx.doi.org/10.1016/j.eswa.2008.01.005

Chan, J. K. W.; Tam, C. M.; Cheung, R. K. C. 2005. Construction firms at the crossroads in Hong Kong, *Engineering, Construction and Architectural Management* 12(2): 111–124. http://dx.doi.org/10.1108/09699980510584476

Edum-Fotwe, F.; Price, A.; Thorpe, A. 1996. A review of financial ratio tools for predicting contractor insolvency, *Construction Management and Economics* 14(3): 189–198. http://dx.doi.org/10.1080/014461996373458

Huang, Y. 2009. Prediction of contractor default probability using structural models of credit risk: an empirical investigation, *Construction Management and Economics* 27(6): 581–596. http://dx.doi.org/10.1080/01446190902960474

Jaselskis, E. J.; Ashley, D. B. 1991. Optimal allocation of project management resources for achieving success, *Journal of Construction Engineering and Management* 117(2): 321–340. http://dx.doi.org/10.1061/(ASCE)0733-9364 (1991)117:2(321)

Kangari, R.; Farid, F.; Elgharib, H. M. 1992. Financial performance analysis for construction industry, *Journal of Construction Engineering and Management* 118(2): 349–361. http://dx.doi.org/10.1061/(ASCE)0733-9364(1992)118: 2(349)

Luo, J.-H.; Lei, H.-Y. 2008. Empirical study of corporation credit default probability based on logit model, in *Proc. of the 4th International Conference on Wireless Communications, Networking and Mobile Computing*: 1–8.

Mason, R. J.; Harris, F. C. 1979. Predicting company failure in the construction industry, *Proceedings Institution of Civil Engineers* 66(2): 301–307. http://dx.doi.org/10.1680/iicep.1979.2356

Ohlson, J. A. 1980. Financial ratios and the probabilistic prediction of bankruptcy, *Journal of Accounting Research* 18(1): 109–131. http://dx.doi.org/10.2307/2490395

Plebankiewicz, E. 2010. Construction contractor prequalification from Polish clients' perspective, *Journal of Civil Engineering and Management* 16(1): 57–64. http://dx.doi.org/10.3846/jcem.2010.05

Ross, S. A.; Westerfield, R. W.; Jordan, B. D. 2010. *Essentials of corporate finance*. 7th ed. The Mcgraw-Hill/Irwin.

Russell, J. S.; Jaselskis, E. J. 1992. Predicting construction contractor failure prior to contract award, *Journal of Construction Engineering and Management* 118(4): 791–811. http://dx.doi.org/10.1061/(ASCE)0733-9364(1992)118:4(791)

Russell, J. S.; Zhai, H. 1996. Predicting contractor failure using stochastic dynamics of economic and financial variables, *Journal of Construction Engineering and Management* 122(2): 183–191. http://dx.doi.org/10.1061/(ASCE)0733-9364(1996)122:2(183)

Severson, G. D.; Russell, J. S.; Jaselskis, E. J. 1994. Predicting contract surety bond claims using contractor financial data, *Journal of Construction Engineering and Management* 120 (2): 405–420. http://dx.doi.org/10.1061/(ASCE)0733-9364(1994)120:2(405)

Tserng, H. P.; Liao, H. H.; Tsai, L. K.; Chen, P. C. 2011a. Predicting construction contractor default with option-based credit models – models' performance and comparison with financial ratio models, *Journal of Construction Engineering and Management* 137(6): 412–420. http://dx.doi.org/10.1061/(ASCE)CO.1943-7862.0000311

Tserng, H. P.; Lin G.-F.; Tsai, L. K.; Chen, P.-C. 2011b. An enforced support vector machine model for construction contractor default prediction, *Automation in Construction* 20(8): 1242–1249. http://dx.doi.org/10.1016/j.autcon.2011.05.007

Wharton Research Data Services. 2009. [online] Wharton School of the University of Pennsylvania, Philadelphia [cited 1 July 2009]. Available from Internet: http://wrds.wharton.upenn.edu

Zavadskas, E. K.; Turskis, Z.; Tamosaitiene, J. 2010. Risk assessment of construction projects, *Journal of Civil Engineering and Management* 16(1): 33–46. http://dx.doi.org/10.3846/jcem.2010.03

H. Ping TSERNG is a Full Professor at the Department of Civil Engineering of NationalTaiwan University. He is also a Corresponding Member of Russian Academy of Engineering. He has a PhD degree in Construction Engineering and Management from University of Wisconsin-Madison and he is an Official Reviewer or Editorial Board Member of several international journals. His research interests include advanced techniques for project management, construction finance, knowledge management, management information system, GPS/wireless sensor network, and automation in construction.

Po-Cheng CHEN is a PhD candidate in the Department of Civil Engineering at National Taiwan University. His research interests include contractor default prediction, and construction finance.

Wen-Haw HUANG is the CEO for Long Reign Development Company. He has a MBA degree from Loyola Marymount University. Currently, he is a part-time PhD student in the Department of Civil Engineering at National Taiwan University. His research interests include contractor prequalification, contractor default prediction, financial management, and construction management.

Man Cheng LEI is a Master of Science in Civil Engineering from National Taiwan University. Her research interests include construction finance, construction management, and project performance evaluation.

Quang Hung TRAN is a Master of Science in Civil Engineering from National Taiwan University, Taiwan. His research interests include construction finance, risk management, and construction prequalification.