KSCE Journal of Civil Engineering (2018) 22(12):4750-4759

DOI 10.1007/s12205-017-2129-x

Construction Management

pISSN 1226-7988, eISSN 1976-3808
www.springer.com/12205

A Study of Integrating Support-Vector-Machine (SVM) Model and Market-based Model in Predicting Taiwan Construction Contractor Default

Han-Tang Huang* and Hui-Ping Tserng**

Received August 31, 2017/Revised 1st: December 18, 2017, 2nd: February 7, 2018/Accepted May 4, 2018/Published Online October 18, 2018

Abstract

The construction industry is a high debt ratio, high operating risk and high financial leverage business. The financial instability causes a chain reaction among funds transferred among companies, therefore restricting competitiveness within the industry. The industry's character and accounting principles of firm-years are different from that of other industries. Creating a hypothetical model of a financial crisis within the construction industry is therefore necessary. Application of this model to real scenarios involving relevant parties can help to forecast a financial crisis in the future. This study applied the market-based model, accounting-based model, hybrid models to predict a financial crisis. These models were then compared to find which can best predict a company that will default. Also, in this paper choosing variables for the Hybrid and Accounting-based models can promote their performance. Finally, the best can be selected for predicting stability.

Keywords: *market-based model, hybrid model, Artificial Neural Network Model (ANN), Support Vector Machine (SVM), Enforced Support Vector Machine-based model (ESVM)*

1. Introduction

In the construction industry, evaluation of financial risk is important for the government, proprietors, banks, insurance companies, and contractors. If a construction project requires extensive funds and a long execution cycle, the proprietor will typically avoid a contractor with high financial risk. Therefore, it is important to distinguish companies with high financial risk from those that are well-managed. This distinction will also predict a company's likelihood of bankruptcy through an assessment of its financial status. The Financial risk assessment provided in the selection of construction contractor pre-qualification and qualification reference and reduce risk of construction project.

Factors taken into consideration include the following: predicting a crisis within the construction industry, improving upon the sample selecting bias of the Support Vector Machine (SVM), and finding the most suitable model.

The definition of finance crisis from Beaver (1966) is as follows: when an enterprise encounters one of the three following situations: it cannot pay special share dividends, accounts are overdrafts, or it has defaulted and is bankrupt, we define this enterprise as a financial crisis company. From Deakin's research, in 1972, a collapsed, bounced, bankrupt, insolvent or liquidated company is a financial crisis company. Ohlson (1980) and Zmijewski (1984) give a definition of bankruptcy as the act of filing the petition for bankruptcy. Pastena and Ruland (1986) define a financial crisis company as one where net assets are neglected or the company cannot pay its debts.

The following are the features of construction industry:

1. The high risk of operation: the causalities in situ, weather effect, the extra loss caused by nature disaster (e.g., typhoon, floods, earthquakes). The serious financial crisis may lead to bankrupt.
2. The high financial leverage: Construction industry are a high capital investment, but less of private enterprise's own capital (Owned funds) caused the high financial leverage. Usually, the builder will be financing after bought the building land and it also occur in construction.
3. The high debt ratio: As the business cycle goes by the growth and decline of the interest and profit makes a challenge for financial control.

To develop predictions for financial crisis, Beaver first uses Univariate discriminant analysis. Then Altman uses Multivariate Discriminant Analysis predicts the financial failure of the company.

Halon's research compares 105 financial crisis enterprises with 2085 non-default enterprises in 1970-1976. In the research, Conditional Logistic Analysis and nine financial indicators are used to build three models for forecasting financial crisis with 96.12%, 95.55%, and 92.84% accuracy. Based on this research, four indicators including company size, financial structure, performance of the business and the current ratio are significantly correlated with the probability of a financial crisis occurring.

*Ph.D. Candidate, Dept. of Civil Engineering, National Taiwan University, Taipei 10617, Taiwan (E-mail: d03521010@ntu.edu.tw)
**Professor, Dept. of Civil Engineering, National Taiwan University, Taipei 10617, Taiwan (Corresponding Author, E-mail: hptserng@ntu.edu.tw)

1. Merton model: The main market-based model is the Merton option pricing model, the option pricing theory (BS model) published by Black and Scholes (1973). To measure the credit risk, setting that company running by raising the loan, the asset of the company as the call option and the exercise price of the debt.
2. Barrier model: It is just like the traditional stock model and European call options, which say that a company can only default when either the option or the loan repayment is due. In reality, a company may default before the due date. Therefore, the option to default should allow for a company being unable to pay a debt at any time. The barrier option defines default as when a company's value is under a special value, and it is more close to reality.

To forecast financial crisis, the market-based model assumes an efficient stock market where all information about stock value is available. However, in reality there is not entirely efficient stock market within any capitalist system.

Predicting ability of the accounting-based model is influenced by window dressing. Because within a firm-year a large amount of human error can occur, a company may not immediately show signs of a financial crisis.

The Merton model has been shown to reflect market information better than the accounting-based model. According to a study by Sobehart, Stein, Mikityanskaya and Li (2000), the hybrid model is based on the logistic model, which uses the Merton model, accounting ratio, market information, and general economic principles as independent variables. 14,447 samples of typical situations and 923 samples of financial crisis were included in the research. The study's results show the hybrid model to be better than the Merton model.

2. Methodology

In this research, five models are used to determine the probability of a financial crisis within a publicly held company in Taiwan. In the first model, the market-based Merton model, stockholders buy a European call option as the company's asset and then determine specifics of the debt. When the market value is lower than the debt value, stockholders will default. The probability of this situation is the probability of a financial crisis.

The second model, the accounting-based model, is the most widely used logistic model. It has a binary classification, default or non-default, which is easy to analysis. The third model, the enforced Support Vector Machine-based model (ESVM), remedies the problems associated with uneven amounts of default and non-default samples in the Support Vector Machine-based model (SVM). The ESVM helps to control for imperfect samples.

The fourth and fifth models, one market-based and one accounting-based, combine to make the hybrid model. This new model sets the financial crisis predicted by the Merton model as an independent variable that is applied to the accounting-based model to forecast a financial crisis.

The flowchart below (Fig. 1) includes a determination of the stock price, debt, risk-free interest rate and the choosing of an accounting ratio. Each model is used to calculate the probability of a financial crisis, and then all results are compared with each other to determine the most suitable model.

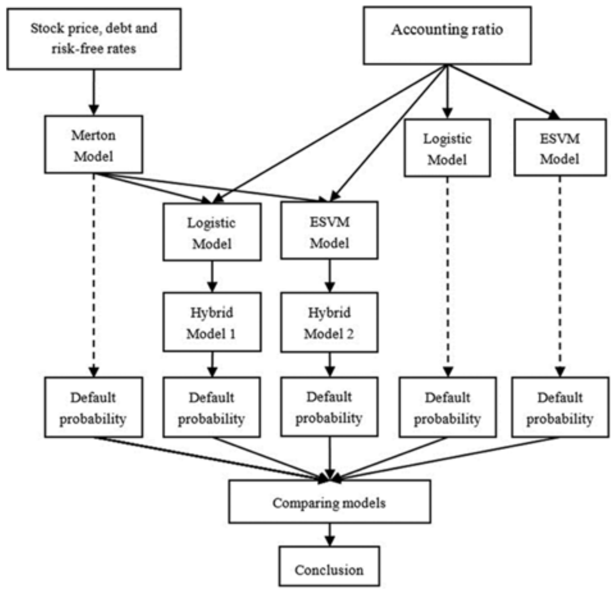

Fig. 1. Flowchart of Methodology

2.1 Support Vector Machine

Support Vector Machine (SVM) published by Vapnik *et al*. In the early 1990s is used for classification and regression analysis. Compared to most methods, the SVM has outstanding performance in pattern discrimination, regression analysis, time series prediction of finance, marketing classification, estimating the volume of production, and medical diagnosis. The main theory of the SVM uses statistics to minimize structural risk. Specific nonlinear mapping transforms data to high dimensional feature space for classification.

2.2 Enforced Support Vector Machine-based Model: ESVM

This research brings forward the enforced Support Vector Machine-based Model (ESVM) to forecast a financial crisis within the construction industry. Effective data from a firm-year is used in the ESVM model to improve the situation that unbalances sampling classification of the Support Vector Machine-based Model (SVM).

2.2.1 The Training Process of ESVM Model

According to Zmijewski (1984), the relatively small numbers of defaulting companies and bankrupted enterprises will raise the sampling error. Unless a model is built for whole samples, the estimating parameter and the predicted results will be unreliable. To avoid sample selection bias, many studies use the company's quarterly financial statements or firm-years to build a predicting model and improve accuracy and increase the predictive ability.

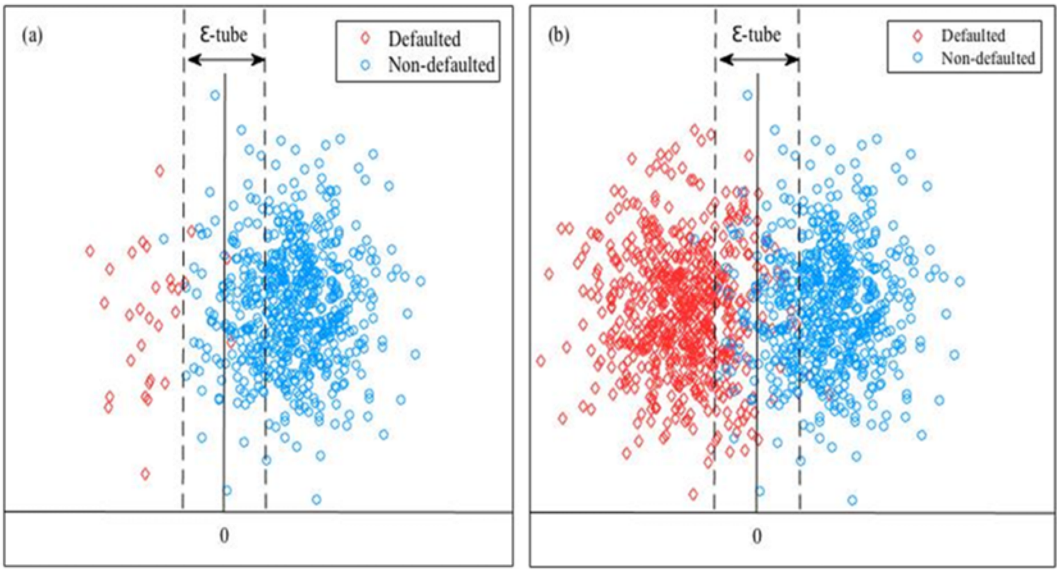

Fig. 2. The Separating Hyperplane of: (a) SVM, (b) ESVM

In this study, we applied firm-years in the ESVM model to predict the probability of financial crisis with the empirical approach. The number of defaulted and non-defaulted companies has a large imbalance called the imbalance of classification. To explain the influence of the imbalance of classification, there is a figure showing the nonlinear phenomenon. The Fig. 2. Shows a 2-D training group and two classifications corresponding.

There are 30 default companies and 600 non-default companies in Fig. 2(a) to present the unbalanced training situation.

The Separating Hyperplane could not distinguish defaulted samples from non-defaulted samples because the number of defaulted samples is much smaller than that of non-defaulted samples. The Support Vector Machine-based model has similar an artificial neural network that only analyses the main input points and ignores the small ones. In actuality, this situation is rare, but can play a key role in the prediction of a financial crisis.

This study poses a simple solution to this problem. Enhancing specific samples in SVM makes it easier to draw conclusions; one way to do this is to copy the default sample 20 times until it is the same size as the non-default samples. After compiling the training sample, the Separating Hyperplane of ESVM can be more clearly and effectively show the difference between the defaults and the non-defaults (see Fig. 2(b)). The sample selection bias associated with SVM has been eliminated.

The result shows that the ESVM model can better address sample selection bias; therefore, its ability to predict a financial crisis is better and more accurate.

2.3 Logistic Regression Model

Logistic Regression Model is published by Berkson in 1944. It is a regression analysis for cumulative incident probability to solve the problems of dichotomy.

It is assumed that the cumulative incident probability of the residual is a logistic distribution and the estimated value (P) is between 0 to 1. The following is the estimating function:

$$P(Y=1|x_i) = \frac{exp(z)}{1+exp(z)} \tag{1}$$

$$Z = \beta_0 + \sum_{i=1}^{k} \beta_i x_i \tag{2}$$

Where, x_i = Explaining variable
Y = Dependent variable, the default company, Y = 1, non-default company, Y = 0
β_i = The coefficient of explaining variable

The maximize likelihood method be used to estimate the parameter (β_i) in Logistic regression model.

$$L = \sum_{i=1}^{n} Y_i \, log \, p_i + (1-Y_i) log(1-p_i) \tag{3}$$

and

$$p_i = P(Y=1|x_i) \tag{4}$$

2.4 Merton Model

Merton (1974) based on Black and Scholes's (1973) research, the option pricing theory, assumes the value of the company as a European call option to evaluate the credit risk.

Merton thinks that company operating by raising the debt. Stockholders buy a European call option as the asset of the company and the strike price of debt. When the market value lower than debt value, stockholders will claim limited liability and choose to default for only loss of the invested capital. According to this concept, Merton estimates company's assets and its variation and combined with a number of liabilities to estimate the probability of default to measure the level of risk.

The model of option pricing theory (Black and Scholes, 1973):

$$V_E = V_A N(d_1) - De^{-rT} N(d_2) \tag{5}$$

Where,

$$d_1 = \frac{\ln(V_A/X)+\left(r+\frac{1}{2}\sigma_A^2\right)^T}{\sigma_A\sqrt{T}}, d_2 = d_1 - \sigma_A\sqrt{T} \quad (6)$$

Notations:

D = The value of the company's debt
r = Risk-free rate
T = The due date of debt
V_A = The value of company's asset
V_E = The value of company's equity
σ_A = The fluctuation ratio of the asset

Crosbie and Bohn (2003) found that, since the long-term nature of some liabilities eases the payment pressure of firms, the assets value at which the firm will default generally lies somewhere between total liabilities and short-term liabilities. Therefore, the option model used in this study defines the strike price X as the sum of short-term liabilities and one-half of long-term liabilities.

To calculate σ_A, by Crosbie and Bohn's (2003) iterative procedure. The procedure uses daily V_E from the past 12 months to obtain an estimate of the volatility of equity, that becomes an initial estimate of σ_A. Putting this initial estimate of σ_A into the Black-Scholes equation derives daily estimates of V_A, and the standard deviation of those V_A become the new estimate of σ_A. The procedure is repeated until the value of σ_A converges to 0.001. Once the converged value of σ_A is obtained, the daily V_A can be solved through Eq. (5).

In the Market-based model framework, the default probability (DP as shown in Eq. (7)) is the probability that the market value of a company's assets will be less than the face value of the company's liabilities at time T, and the μ is a function of the actual return on company's assets, that drift μ instantaneous can be computed daily values of V_A.

$$DP = N\left(-\frac{\ln(V_{A,t}/X)+\left(\mu-\frac{1}{2}\sigma_A^2\right)T}{\sigma_A\sqrt{T}}\right) \quad (7)$$

2.5 Hybrid Model

The result of the accounting-based model was more accurate for short-term forecasting, while the market-based model was more suitable for medium-term and long-term forecasting. If accounting and marketing information clouds are included in one model, the prediction of financial crisis will have better accuracy. Therefore, there are two hybrid models that have been developed in this study. The default probability given by the market-based model will become a variable that give input to the accounting-based model to integrate accounting and marketing information to predict a financial crisis.

2.6 Analytical Method

2.6.1 Receiver Operating Characteristics Curve

The Receiver Operating Characteristics curve was the main method of research, derived from Sobehart *et al.* (2000). The information from the ROC analysis can be a reference for a user to make a decision or can be applied in problem classification to evaluate criteria.

2.6.1.1 Analysis Principle of ROC

The data were divided into two classes, the "target class" or "true class," and the "non-target" class," or "false class." In this classification, the rate of truly classed samples was called True Positive (TP) and the rate of falsely classed samples was called False Positive (FT).

In general, the TP was called sensitivity, 1-FP was called specificity and the ROC curve was the relationship of sensitivity and specificity. The Area Under the Curve (AUC) is the standard of discriminatory power. A larger AUC is more powerful, and vice versa.

2.6.1.2 The Application of the ROC Curve in Predicting Financial Crisis

There are two possible biases associated with determining a financial crisis in construction industry with the default forecast model:

1. The default enterprise was classified as a non-default enterprise, known as Type I error.
2. The non-default enterprise was classified as a default enterprise, known as Type II error.

Misclassification increases adverse effects. Assuming there is a threshold (S) with which to validate the result, the default company's evaluation will be higher than S. The following are four results of decisions:

True Positive (TP), correctly identified
True Negative (TN), correctly rejected
False Negative (FN), incorrectly rejected; Type I error
False Positive (FP), incorrectly identified; Type II error

2.6.1.3 AUROC (Area under ROC)

The Area Under ROC (AUROC) is used to judge the model. The random model corresponds to the diagonal (AUC = 0.5) and the perfect model corresponds to the maximum of AUROC (AUC = 1). Therefore, the ROC curve closer to quarter arc has the power to distinguish a default enterprise from a non-default enterprise. The following is the function of the Area Under ROC (AUROC):

$$AUROC = \int_0^1 HR(FAR)d(FAR), 0.5 < AUROC < 1 \quad (8)$$

Table 1. The Result of Financial Crisis Decision

Threshold(S)		Non-default company	Default company
Evaluation	Higher than S: default	FP (Type-II error)	TP
Model	Lower than S: normal	TN	FN (Type-I error)

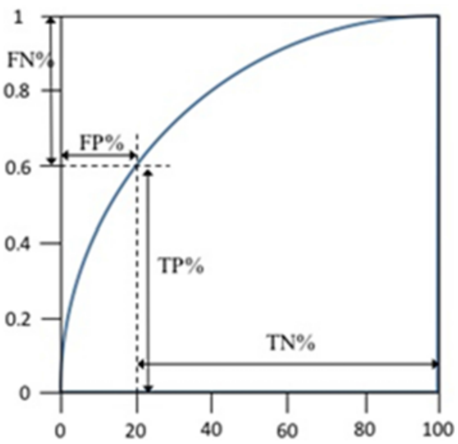

Fig. 3. ROC Curve (Stein, 2005)

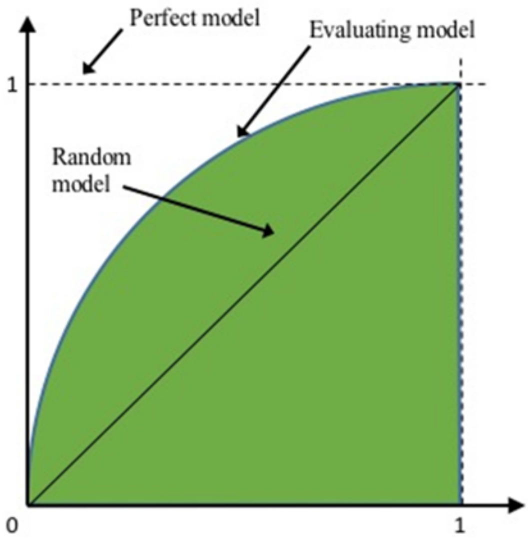

Fig. 4. ROC Curve of Perfect Model and Random Model (Sobehart *et al.*, 2000)

To assess which default-predicting model has the best performance for contractor default risk, this study employs the Receiver Operating Characteristics curve (ROC curve) to measures the extent at which the model can differentiate firms that are more likely to default from firms that are less likely to default. A model that assigns all defaulting firms larger probabilities of default than non-defaulting firms would be perfectly discriminating.

2.6.2 Cross Validation

Based on Lanchenbruch and Mickey's research on Cross Validation (CV), a statistical analysis to validate the efficiency of classification, we grouped the dataset. Some of them are in the training set, and the others are validation set. First, we applied to the training set to train the model and adopted the validation set to test the model that we gained from training. The test results were an indicator for evaluating model's classification ability.

The characteristics and the accounting principles of construction industry differ from those of other industries. Given the current economic climate, there is an urgent need for further research on the prediction of financial failure, particularly in the construction industry. Therefore, this study attempts to establish a financial failure forecasting model on public construction companies in Taiwan. By enhancing the accuracy of the forecasting model, it is anticipated that the established model can provide financial distress early warning of construction companies, thereby helping construction industry business owners and managers, financial institutions, insurance companies, investors, companies in related industries to accurately identify which construction industry firms are likely to be at risk of financial failure.

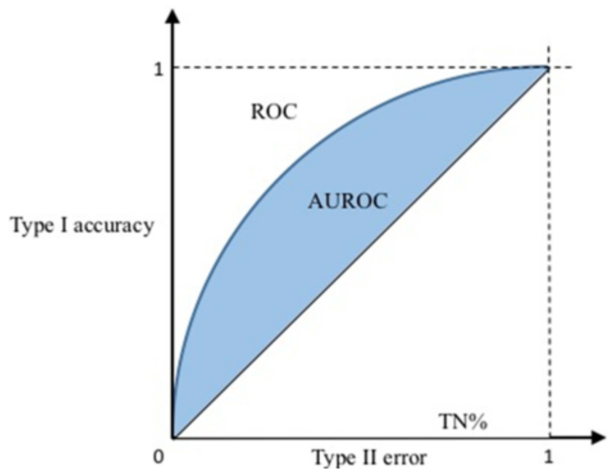

Fig. 5. AUROC Curve (Sobehartet *et al.*, 2000)

Previous research has suggested that there are some problems with the use of the Support Vector Model (SVM) and the Artificial Neural Network (ANN) model in forecasting financial failure with sample matching. Because neither SVM nor ANN is capable of distinguishing between default and non-default, this will lead to bias in sample selection.

2.6.3 Summary

Previous research has suggested that there are some problems with the use of the Support Vector Model (SVM) and the Artificial Neural Network (ANN) model in forecasting financial failure with sample matching. Because neither SVM nor ANN is capable of distinguishing between default and non-default, this will lead to bias in sample selection.

This study employs an Enforced Support Vector Machine-based Model (ESVM) to forecast financial failure in the construction industry, and uses firm-year data to solve the problem of unbalanced samples. The traditional logistic regression model can provides as a benchmark to evaluate the ESVM model's prediction ability. After selected by stepwise logistic regression method, the variables were imported into the models for comparison.

The Default Possibility (DP) calculate of those construction companies based on the Merton model from Eq. (7) (the variable of DP). After calculate the DP, that DP as a variable and combine the previous 20 factors into 21 factors. In the other set, combine the DP and 4 factors into 5 factors. Therefore, we develop the

Hybrid 1 that using these 21 factors and 5 factors and re-run the logistic model. Also, developing the Hybrid 2 that using 21 factors and 5 factors and re-run the ESVM model. Finally, compare these five models (SVM, Logistic, ESVM, Hybrid 1 and Hybrid 2) Table 6.

Five predictive models are applied in this study as to identify the best model for predicting financial failure in the Taiwanese construction industry, including the market model (Merton), two accounting models (Logistic and ESVM), Hybrid Model 1 (an integrated model with the Merton market model and the Logistic accounting model), Hybrid Model 2 (an integrated model with the Merton market model and the ESVM accounting model).

3. Data Collection and Empirical Results

In this chapter, the process of data collection and empirical results are included. Section 3.1 describes the source of data used in this study as well as the selection criteria. Section 3.2 includes the variable Selection and 3.3 shows the validating predictive ability of model and their comparisons.

3.1 Data Collection

The firm-years of study included the default or non-default companies, collected by Taiwan Economic Journal (TEJ) from 1995 to 2009. The empirical investigation of this study considers a cross-section of Taiwanese construction contractors. This study collects data from Taiwan Economic Journal (TEJ). This study restricts its attention to listed and delisted construction contractors with December fiscal year-ends by choosing firms with TEJ codes 25 (construction contractors).

Include 46 non-defaulted (currently listed) and 39 defaulted

Table 2. Non-defaulted Construction Company

Code Firms			
1436 HUA YU LIEN	2534 HSC	4416 SHANYUAN	5525 SWEETEN
2501 CATHAY RED	2535 DA-CIN	5508 YCDC	5530 LUNG YEN
2504 GBM	2536 Hong Pu	5511 TE CHANG	5531 Shining
2509 CHAINQUI	2542 High wealth	5512 RICH	5533 FOUNDING
2511 PHD	2543 HWANG CHANG	5514 SUN FON	5534 CHONG HONG
2514 LONG BON	2545 HUANG HSIANG	5515 CHIEN KUO	6177 DA-LI
2515 BES	2546 KEDGE	5516 SUN-SEA	6212 L.M.
2516 New Asia	2547 RADIUM	5519 LD	6219 FULL WANG
2520 KINDOM	2548 HUAKU	5520 Lithia	9933 CTCI
2524 KTC	2841 TLDC	5521 KSECO	9945 RUENTEX DEVELOP
2526 CEC	3052 APEX	5522 Far glory	
2527 Hung Ching	3056 Zongtai	5523 FONG CHIEN	

Table 3. Default Construction Company

Code Firm	Default Date	Code Firm	Default Date	Code Firm	Default Date
1107 CHIEN TAI	2000/8/16	2529 JEN HSIANG	1998/12/29	5503 ADC	2001/7/18
1442 ADVANCETEK	1999/2/2	2530 DELPHA	2001/6/2	5504 HSIN NAN	2000/9/17
1805 KPT	2009/3/31	2533 YUH CHEN	2004/2/9	5505 HOWARM	2001/4/2
1808 ALE	1999/5/24	2537 WE & WIN	2001/1/12	5506 EGC	2008/10/27
2505 ky	1999/3/20	2538 Kee Tai	2002/4/29	5518 GREATSUN	2001/8/23
2506 PCC	2001/10/16	2539 SAKURAD	1999/3/22	5529 Well Glory	2008/12/31
2512 BCG	2002/4/16	2540 Kim Sang-chang	2000/11/10	5532 JCC	2006/10/14
2517 CHANG KU	2000/11/30	2553 CHIEF	1999/4/18	8710 YISHIN ENG	1999/8/26
2518 EVERFORTUNE	2000/9/6	2594 d-Life	2001/9/6	8716 TOP	1999/1/7
2521 H.C.C	2000/7/29	5213 YeaShin	2005/10/28	8719 HUFU	1998/11/20
2523 DP	2006/4/28	5395 ELEMENTS	2006/1/16	8725 SUN SPLENDOR	1999/9/28
2525 Pao Shiang	2002/6/30	5501 PACCO	1998/3/1	8907 SCIENTRADE	1999/12/21
2528 CROWELL	2000/4/28	5502 KINGLAND	2001/8/28	5503 ACELAND-DYNASTY	2001/7/18

(once listed, and as defined by TEJ) construction companies, which are shown in Table 2 and Table 3, respectively. Combining firms with the sample period, the total number of firm-years is 692, with 39 defaulted samples and 653 non-defaulted samples. All firm-years between 1995 and 2009 with TEJ code 25 were included, with a few exceptions due to the absence of information provided on TEJ. The 4-digit numbers shown in the non-defaulted list is the stock code of these companies.

This study uses a definition of default according to TEJ. If the TEJ's data on a sample was incomplete, that sample was removed. The following is the definition of the Taiwan Cooperate Credit Risk Index (TCRI) from TEJ: bounced check, bankruptcy, operation difficulties, reorganization, financial crisis, management takeover, delisted (or temporarily delisted) from market, suspension of work, negative or low net worth, stolen assets, deflation of bank, major loss, or value depreciation. The default date, as defined by TEJ, is the date when any of the above situation occurred for the first time.

3.2 Variable Selections

This study was based on the characteristics of the construction industry for literature review to select primary representative variables. Then, a Logistic regression test was used to determine variable significance for distinguishing between defaulted and non-defaulted enterprises, and deciding which variable was best able to do so. The following is a summary of literature reviews from Chin (2009). Table 4 is an arrangement of financial ratio research in the construction industry. After the literature review and discussion, there are four scopes and 20 financial ratios (see Table 5. for calculation function description).

(1) Mason and Harris (1979) Proceedings, Institution of Civil Engineering
(2) Halpin (1985) John Wiley & Sons Inc.
(3) Kangari, Farid and Elgharib (1992) Journal of CEM
(4) Severson, Jaselskis and Russell (1993) Journal of CEM
(5) Langford, Iyagba and Komba (1993) Journal of CME
(6) Severson, Russell and Jaselskis (1994) Journal of CEM
(7) Abidali and Harris (1995) Journal of CME
(8) Russell and Zhai (1996) Journal of CEM
(9) Kangari and Bakheet (2001) Journal of CEM

There are four groups of accounting ratios, including Liquidity, Leverage, Activity, Profitability. Any ratios that did not easily fall into one of these categories were classified as "Others."

The accounting-based model is based on three criteria to choose the financial ratio:

(1) The ratio has been cited more than twice.

(2) The ratio corresponds with the characteristics of the construction industry.

(3) The ratio should completely correspond with the forecast of a financial crisis.

According to these criteria, 20 primary variables were chosen, and–defined as R1, R2, R3…… R20, and the definitions are given in Table 5.

Table 4. Ratio Summary of Prior Research (Chin, 2009)

Category	Accounting Ratio	Reference
Activity	Debtor/Creditor	[1]
	The number of days in arrears	[1]
	Claims trend	[1]
	Total Assets/Income	[3]
	Income/Working Capital	[2], [3]
	Accounts Receivable Turnover	[2], [3], [9]
	Accounts Payable Turnover	[4], [9]
	Under billings	[4]
	Cost of Sales	[4]
	Under billings /Net Sales	[6]
	Net Sales/Net Assets	[7]
	Net Sales/Net Worth	[2], [9]
	Asset Turnover	[2]
	Inventory Quality	[2]
	Fixed Assets to Income Ratio	[2]
	Fixed Assets to Net Assets Ratio	
Leverage	Total Liabilities/Net Assets	[2], [3], [9]
	Retained earnings /Net Sales	[4], [6]
	Net Assets/Fixed Assets	[5]
	Net Assets/Total Liabilities	[5]
	Net Assets/Total Assets	[8]
	Net Assets/Inventory	[9]
	Debt ratio	[2]
	Times Interest Earned	[2]
Liquidity	Quick ratio	[2], [5]
	Current ratio	[2], [3], [5], [9]
	Total Liabilities/Net Sales	[4]
	Current Liabilities/Net Sales	[6]
	Short-term loans/Net Profits+ Tax + Interests	[7]
	Short-term loans trends	[8]
	Current Assets-Current Liabilities/ Total Assets	[5], [8]
	Current Assets/Net Assets	[7]
	Current Assets-Current Liabilities/Inventory	[9]
	Fixed Assets accounting for Net Assets ratio	[2], [9]
Profitability	Net Profit +Tax +Interest/Net Assets	[1]
	(Net Profit +Tax +Interest)/ (Fixed Assets +Working Capital)	[1]
	ROA	[2], [3]
	ROE	[2], [3], [5], [9]
	Gross Profit	[4]
	ROS	[2], [6], [9]
	Net Profit after Interest and Tax/ (Fixed Assets +Working Capital)	[7]
	Tax Trades	[7]
	Net Profit after Interest and Tax Trades	[7]
	Gross Profit /Total Assets	[8]
	Gross Profit /Net Sales	[9]
	Profit/Working Capital	[2]
Others	Stock Trend	[4]
	Cost Monitoring	[6]
	Interest Rate Trend	[8]
	The Present Value of New Construction	[8]
	Current Expenditure/Net Sales	[9]
	Current Expenditure/Net Assets	[9]

Table 5. The Variable Selection of Logistic Regression

Category	Accounting Ratios
Activity	1. Revenues to Net Working Capital 2. Sales to Net Worth 3. Turnover of Total Assets 4. Accounts Receivable Turnover 5. Accounts Payable Turnover 6. Turnover of Total Assets 7. Revenues to Fixed Assets
Leverage	8. Total Liabilities to Net Worth 9. Debt Ratio 10. Times Interest Earned 11. Retained Earnings to Sales
Liquidity	12. Current Ratio 13. Quick Ratio 14. Net Working Capital to Total Assets 15. Current Assets to Net Assets 16. Fixed Assets to Net Worth
Profitability	17. Return on Assets (ROA) 18. Return on Equity (ROE) 19. Return of Sales (ROS) 20. Profits to Net Working Capital

The first step of the analysis is to calculate the significant values of the 20 variables (see Table 5). If a value is smaller than 0.05, that variable will be chosen and the others will be called non-chosen variables.

The second step is to add one of the non-chosen variables to the chosen variable and plug them into the Logistic regression model each time. One must choose the variables whose significant values are smaller than 0.1 from each Logistic regression model, and add the chosen variables repeatedly until there is no significant value is smaller than 0.1. After Logistic regression model selecting, the last variables are accounts payable turnover ratio (5), total liabilities/net assets (8), retained earnings / net sales (11), and ROA return on assets (17), as shown in Table 5. The Logistic regression model selected variable flow chart is shown in Fig. 6.

3.3 Validating the Predictive Ability of Model

In this study, the Area Under the ROC curve (AUC) is the criterion used to present the ability of the model. A larger AUC is more powerful, and vice versa. Using the listed construction company stock prices, debt, risk-free interest rate, and firm-year accounting ratio items, among other things, the AUC is plotted in the different models to compare the predictive abilities of each financial crisis probability mode and to identify which is most appropriate.

Comparing AUC with Hybrid model, Accounting-based model, Market-based model (Shown as Table 6). The performance of ESVM model is the best.

4. Conclusions

The study illustrates that choosing the best variable will

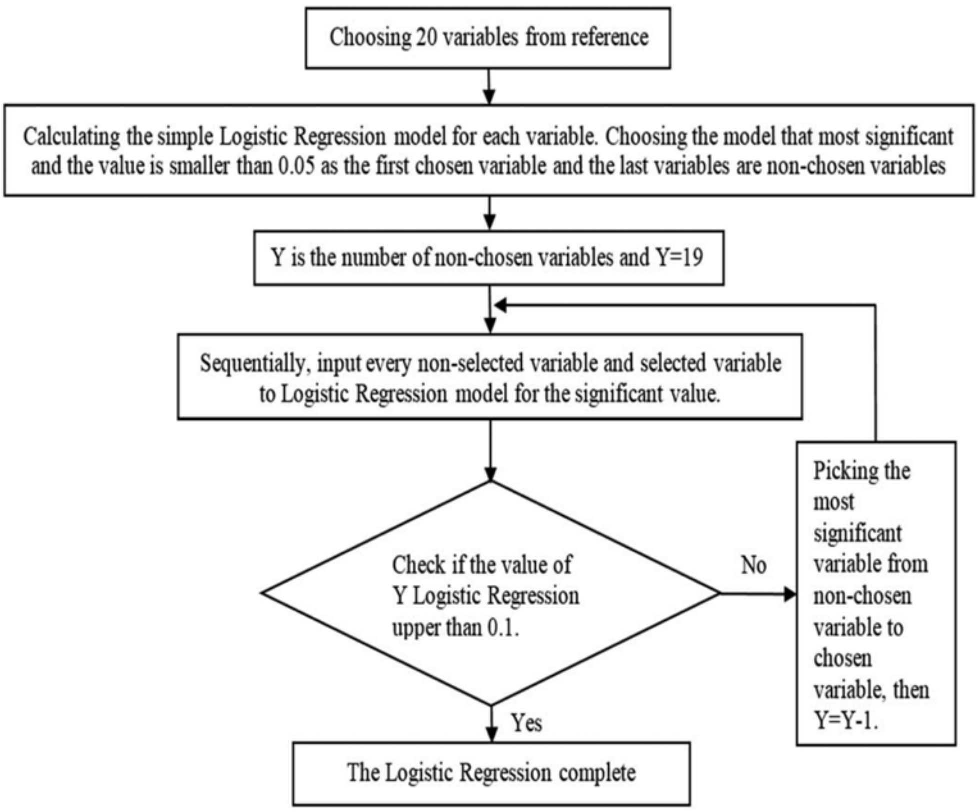

Fig. 6. Flowchart of Selecting Variables with the Logistic Regression

Table 6. Comparing AUC of Hybrid, Accounting-based Model and Market-based Mode

	Model	Variables	AUC
Hybrid model	Hybrid model 1 (Logistic+ Merton DP)	(21)	0.7459
		(5)	0.8220
	Hybrid model 2 (ESVM+ Merton DP)	(21)	0.8189
		(5)	0.8252
Accounting-based model	Logistic model	(20)	0.7425
		(4)	0.8154
	ESVM model	(20)	0.7974
		(4)	0.8258
Market-based model	Merton model		0.7197

Note:
DP: Default probability

promote the performance of the Accounting-based and hybrid models. Because of the trading environment of Taiwan, the performance of the Market-based model (Merton model) is not significant.

The performance of the hybrid model is based on that of the Market-based model (Merton model) and the Accounting-based model (logistic model and ESVM model). Comparing the predictive abilities of Hybrid 2 model and ESVM model, the ESVM model is slightly better. The main reason for this result is that the Taiwanese stock market consists of smaller listed construction companies dominated by SMEs, showing the Taiwanese market's information reliability, transparency, low efficiency and its vulnerability to man-made factors.

To detect a model's predictive capability relative to other models using the area under the ROC curve (AUC) for the evaluation criteria, one can compare five models to obtain the following results:

The most suitable variable will enhance the models' ability to detect a financial crisis.

1. The Market-based model (Merton model) for predicting financial crisis is not particularly useful in Taiwan.
2. The Hybrid model is superior to the single model.
3. The ESVM model is more stable than others.
4. The ESVM model is the best able to predict a financial crisis.
5. Taiwanese market data is not reliable, transparent, or efficiency. It is easily affected by human factors.

In conclusion, this study presents a model to predict a financial crisis in a construction company to lessen risks associated with signing a contract and to reduce disputes. The agency guarantees construction firms a more reliable and objective assessment of the financial aspects of their company. According to Basel (2006), lending institutions are allowed to credit risk of the portfolio in terms of capital expenditure, which is suitable for the development of a methodology within lending institutions. This study provides an assessment of the credit risk of financial lending institutions, the construction management company's internal credit assessment base calculation. Identification and avoidance of potential financial crisis prevent construction firms from having a high default risk, a way to co-subcontractor. In addition, construction firms that evaluate the probability of financial risk and identify deteriorating conditions can avoid bankruptcy.

In the future research program to help owners through the Artificial Intelligence methods based on this research of financial prediction to select better construction contractors.

References

Abidali, A. F. and Harris, F. C. (1995). "A methodology for predicting company failure in the construction industry." *Journal of Construction Management and Economics*, Vol. 13, No. 3, pp. 189-196, DOI: 10.1080/01446199500000023.

Agarwal, V. and Taffer, R. (2008). "Comparing the performance of market-based and accounting-based bankruptcy prediction models." *Journal of Banking & Finance*, Vol. 32, No. 8, pp. 1541-1551, DOI: 10.1016/j.jbankfin.2007.07.014.

Altman, E. I. (1968). "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy." *Journal of Finance,* Vol. 23, No. 4, pp. 589-609, DOI: 10.1111/j.1540-6261.1968.tb00843.x.

Altman, E. I., Haldeman, R., and Narayanan, P. (1977). "ZETA analysis: A new model to identity bankruptcy risk of corporations." *Journal of Banking & Finance*, Vol. 1, No. 1, pp. 29-54, DOI: 10.1016/0378-4266(77)90017-6.

Altman, E. I., Marco, G., and Varetto, F. (1994). "Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks." *Journal of Banking & Finance,* Vol. 18, No. 3, pp. 505-529, DOI: 10.1016/0378-4266(94)90007-8.

Beaver, W. H. (1966). "Financial ratios as predictors of failure." *Journal of Accounting Research*, Vol. 4, pp. 71-111, DOI: 10.2307/2490171.

Berkson, J. (1944). "Application of the logistic function to bio-assay." *Journal of the American Statistical Association,* Vol. 39, No. 227, pp. 357-365, DOI: 10.1080/01621459.1944.10500699.

Black, F. and Scholes, M. (1973). "The pricing of options and corporate liabilities." *Journal of Political Economy*, Vol. 81, No. 3, pp. 637-654, http://www.jstor.org/stable/1831029.

Brockman, P. and Turtle, H. J. (2003). "A barrier option framework for corporate security valuation." *Journal of Financial Economics*, Vol. 67, No. 3, pp. 511-529, DOI: 10.1016/S0304-405X(02)00260-X.

Chava, S. and Jarrow, R. A. (2004). "Bankruptcy prediction with industry effects." *Review of Finance*, Vol. 8, No. 4, pp. 537-569, DOI: 10.1093/rof/8.4.537.

Chin, H. W. (2009). *A barrier option framework for predicting insolvency of construction firms*, M.Sc. Thesis, Department of Civil Engineering, College of Engineering, National Taiwan University, Taiwan.

Cortes, C. and Vapnik, V. (1995). "Support-vector networks." *Machine Learning,* Vol. 20, No. 3, pp. 273-297, DOI: 10.1023/A: 1022627411411.

Crosbie, P. and Bohn, J. (2003). "Modeling default risk." *Moody's KMV*, Moody's KMV Company, <http://www.moodyskmv.com/research/files/wp/Modeling DefaultRisk.pdf.viewed in 2003> or <https://business.illinois.edu/gpennacc/MoodysKMV.pdf viewed in 2018>.

Daniel, W., Halpin, John Wiley., and Sons (1985). "A Review of: Financial and cost concepts for construction management." *The Engineering Economist*, Vol. 32, No. 1, pp. 71-72, DOI: 10.1080/00137918608928120.

Deakin, E. B. (1972). "A discriminant analysis of predictors of business failure." *Journal of Accounting Research*, Vol. 10, No. 1, pp. 167-179, DOI: 10.2307/2490225.

Gharghori, P., Chan, H., and Faff, R. (2006). "Investigating the performance of alternative default-risk models: Option-Based versus accounting-based approaches." *Australian Journal of Management*, Vol. 31, No. 2, pp. 207-234, DOI: 10.1177/031289620603100203.

Gu, G., Pan, H., Lin, Z., Li, Y., and Yuan, Z. (2017). "Data-driven support vector machine with optimization techniques for structural health monitoring and damage detection." *KSCE Journal of Civil Engineering*, Vol. 21, No. 2, pp. 523-534, DOI: 10.1007/s12205-017-1518-5.

Kangari, R. and Bakheet, M. (2001). "Construction surety bonding." *Journal of Construction Engineering and Management*, Vol. 127, No. 3, pp. 232-238, DOI: 10.1061/(ASCE)0733-9364(2001) 127:3(232).

Kangari, R., Farid, F., and Elgharib, H. M. (1992). "Financial performance analysis for construction industry." *Journal of Construction Engineering and Management,* Vol. 118, No. 2, pp. 349-361, DOI: 10.1061/(ASCE)0733-9364(1992)118:2(349).

Langford, D., Iyagba, R., and Komba, D. M. (1993). "Prediction of solvency in construction companies." *Journal of Construction Management and Economics*, Vol. 11, No. 5, pp. 317-325, DOI: 10.1080/01446199300000036.

Mason, R. J. and Harris, F. C. (1979). "Predicting company failure in the construction industry." *Proceeding*, *Institution of Civil Engineering*, Vol. 66, Part I, pp. 301-307, DOI: 10.1680/iicep.1979.2356.

Merton, R. C. (1974). "On the pricing of corporate debt: The risk structure of interest rates." *Journal of Finance*, Vol. 29, No. 2, pp. 449-470, DOI: 10.2307/2978814.

Odom, M. D. and Sharda, R. (1990). "A neural network model for bankruptcy prediction." *IJCNN International Joint Conference*, Vol. 2, pp. 163-168, DOI: 10.1109/IJCNN.1990.137710.

Ohlson, J. A. (1980). "Financial ratios and the probabilistic prediction of bankruptcy." *Journal of Accounting Research*, Vol. 18, No. 1, pp. 109-131, DOI: 10.2307/2490395.

Pastena, V. and Ruland, W. (1986)."The Merger/Bankruptcy Alternative." *Journal of the American Accounting Association,* Vol. 61, No. 2, pp. 288-301, http://www.jstor.org/stable/247259.

Rashidi, A., Sigari, M. H., Maghiar, M., and Citrin, D. (2016). "An analogy between various machine-learning techniques for detecting construction materials in digital images." *KSCE Journal of Civil Engineering*, Vol. 20, No. 4, pp. 1178-1188, DOI:10.1007/s12205-015-0726-0.

Russell, J. S. and Zhai, H. (1996). "Predicting contractor failure using stochastic dynamics of economic and financial variables." *Journal of Construction Engineering and Management*, Vol. 122, No. 2, pp. 183-191, DOI: 10.1061/(ASCE)0733-9364(1996)122:2(183).

Severson, G. D., Jaselskis, E. J., and Russell J. S.(1993). "Trends in construction contractor financial data." *Journal of Construction Engineering and Management*, Vol. 119, No. 4, pp. 854-858, DOI: 10.1061/(ASCE)0733-9364(1993)119:4(854).

Severson, G. D., Russell, J. S., and Jaselskis, E. J. (1994). "Predicting contract surety bond claims using contractor financial data." *Journal of Construction Engineering and Management*, Vol. 120, No. 2, pp. 405-420, DOI: 10.1061/(ASCE)0733-9364(1994)120:2(405).

Sobehart, J. R., Stein, R., Mikityanskaya, V., and Li, L. (2000). *Moody's public risk firm risk model: A hybrid approach to modeling short term default risk*, Report, No. 53853, Moody's Investor Service, Global Credit Research, Rating Methodolgy, NewYork, USA.

Stein, Roger M. (2005). "The relationship between default prediction and lending profits Integrating ROC analysis and loan pricing." *Journal of Banking & Finance*, Vol. 29, No. 5, pp. 1213-1236, DOI: 10.1016/j.jbankfin.2004.04.008.

Taiwan Economic Journal Data Bank (TEJ) (2011). Online, http://www.finasia.biz/ensite/.

Whitaker, R. B. (1999). "The early stages of financial distress." *Journal of Economics and Finance*, Vol. 23, No. 5, pp. 123-133, DOI: 10.1007/BF02745946.

Zmijewski, M. E. (1984). "Methodological issues related to the estimation of financial distress prediction models." *Journal of Accounting Research,* Vol. 22, pp. 59-82, DOI: 10.2307/2490859.