INTEGRATION OF ACCOUNTING-BASED AND OPTION-BASED MODELS TO PREDICT CONSTRUCTION CONTRACTOR DEFAULT

Lung-Ken Tsai¹, Hui-Ping Tserng², Hsien-Hsing Liao³, Po-Cheng Chen², and Wen-Pei Wang²

Key words: construction contractor, default prediction, option-based model, logit model.

ABSTRACT

This paper aims to predict construction contractor default, which is excluded by most extant studies, due to the distinct characteristics of construction industry. Default predicting models developed in past literatures are mostly built by accounting information, yet accounting sheets have innate flaws. To calculate default probability, several recent studies applied the option pricing theory, which presumes that the stock market is efficient. This presumption isn’t always true in real life. In this paper, a hybrid model is proposed. It combines information from both models by inputting the default probability from the option-based model into the accounting-based model. As the measure of models’ predicting performance, the Area Under the receiver operating characteristic Curve (AUC) is used. Empirical results show that the hybrid model (AUC: 0.8732) outperforms both the accounting-based model (AUC: 0.7519) and the option-based model (AUC: 0.8581). This result shows that accounting or stock market information alone is not sufficient to explain real-world behavior. It is suggested that the hybrid model be used as an alternative prediction model of construction contractor default.

I. INTRODUCTION

Financial distress early warning is highly concerned in every industry. Past researches on financial default early warning models aimed at the whole industry rather than at single industries. However, [11] pointed out that different industries face different competing environments and use different accounting principles, thus their bankruptcy probabilities are also different, even when bearing the same balance sheet.

The construction industry differs from other industries in many ways: owner-dominant trade, long duration in completing products, complicated production process, unpredictable fluctuations in construction volume, and high uncertainty and risk involved. [17] stated that the construction industry has high default probability compared to other industries. Past researches on bankruptcy prediction models, such as [3, 4, 9], and [30], mostly excluded the construction industry from their sample. Yet evaluating the financial failure probability of the construction industry is a critical issue in successfully completing a project. It has always been an important issue for governmental organizations, construction owners, lending institutions, surety underwriters, and contractors. Thus, this paper aims to measure and predict the construction contractor default risk.

The financial distress early warning models developed in past literatures are in large built by historical accounting information. They supposed that there may be different patterns between defaulters and non-defaulters in historical accounting information, and tried to find out these patterns by some regression or data mining analysis, such as the univariate ratio analysis model [4], the multivariate ratio analysis model [3, 8, 14, 36] the LPM model [27], the logit model [29], the probit model [39], and Artificial Neural Network Models [12, 28, 35]. Although the above accounting-based models have considerable predicting abilities, accounting sheets are subject to manipulation and unable to show immediate default symptoms (because the information is announced only 4 times in a year).

Among innovative approaches to forecast corporate defaults, [26] applied the option pricing theory derived by [7] to calculate default probability. There is an essential difference between Merton’s [26] option-based model and accounting-based models: The option-based model does not employ information from data mining, but depicts a company’s default by using option-pricing equations. That means it need not find out the “default pattern” from huge firms’ historical account-
ing data; instead, only the target firm’s stock price, liability, and risk-free interest rate are necessary to predict its credit risk. In an efficient market, the company’s stock price not only reflects accounting and economic information but also reflects qualitative factors such as management and technique, which are also essential to construction contractors’ success. Although many scholars such as [2, 6, 9, 10, 13, 18, 30, 38] used the option-based model on evaluating default probability, none employed this approach to contractor failure prediction.

Option-based models have a limitation, that is, they are based on the presumption that all information is reflected in the company’s stock price, yet this is not always true because stock market is sometimes not efficient. Due to the fact that accounting information is lagged and the option-based model may be suffered from market inefficiency, this paper proposes a hybrid default prediction model that uses a simple way to combine information from both accounting-based and option-based models – inputting the default probability from the option-based model into the accounting-based model as an input variable.

This paper empirically validates the predicting performance of the hybrid model in construction contractor default. The option-based model and the logistic regression model (an accounting-based default prediction model) are provided as benchmarks for assessing the results of the hybrid model’s forecasting ability.

The rest of this paper is divided into four sections: Section 2 introduces the methodology of default prediction; Section 3 presents how this paper applies the prediction models to predict construction contractor default, this part includes our data set, sample selection criteria, and input variable selection; Section 4 reports the assessment criteria of the models’ predicting ability and compares models’ predicting performances; finally, Section 5 provides concluding comments.

II. METHODOLOGY

1. The Option-Based Model

According to [26], the equity of a levered firm can be viewed as a European call option on the market value of the firm’s assets with the book value of total liabilities as the strike price. Equity holders exercise their option on the firm’s assets and the firm continues to exist if the market value of assets is greater than the level of liabilities at maturity. On the other hand, equity holders do not exercise their option on the firm’s assets and the firm defaults if the market value of assets is less than the level of liabilities.

In the Black-Scholes-Merton framework [7, 25, 26], the market value of a firm’s equity, \( V_e \), as a European call option on the value of the firm’s assets, can be written as Eq. (1):

\[
V_e = V_A N(d_1) - X e^{-rT} N(d_2)
\]

where

\[
d_1 = \frac{\ln(V_A / X) + (r + \frac{1}{2} \sigma^2)T}{\sigma \sqrt{T}},
\]

\[
ds_2 = d_1 - \sigma \sqrt{T}
\]

\( V_A \) is the firm’s assets value, with an instantaneous volatility \( \sigma \). \( X \) is the book value of liabilities maturing at time \( T \). \( r \) is the risk-free rate, and \( N \) is the cumulative density function of the standard normal distribution.

[13] found that the assets value at which the firm will default do not lie at the book value of their total liabilities, rather, it generally lies somewhere between total liabilities and short-term liabilities. Following [13] and [38], the option-based model used in this paper defines the strike price \( X \) as the sum of short-term liabilities and one-half of long-term liabilities.

An iterative procedure to calculate \( \sigma \) was suggested by [13]: daily \( V_e \) from the past 12 months is used to obtain an estimate of the volatility of equity \( \sigma \), which becomes an initial estimate of \( \sigma \). One can solve the Black-Scholes equation using this initial estimate to obtain daily estimates of \( V_A \) and then compute the standard deviation of those \( V_A \)’s daily return, which becomes the new estimate of \( \sigma \) for the next iteration. This procedure is repeated until the value of \( \sigma \) converges to 0.0001 and the daily \( V_A \) can be solved through Eq. (1).

Under the Option-based model framework, the Default Probability (DP) is defined as the probability that the market value of a firm’s assets will be less than the face value of the firm’s liabilities at time \( T \). The option-based models assumes that the firm’s asset returns is Normally distributed, thus the default probability can be defined as Eq. (2):

\[
DP = N\left( -\frac{\ln(V_A / X) + (\mu - \frac{1}{2} \sigma^2)T}{\sigma \sqrt{T}} \right)
\]

(2)

Note that the value of the call option in Eq. (1) is derived under the assumption of risk-neutrality, that is, all assets are expected to grow at the risk-free rate. However, the default probability depends upon the actual distribution of asset values, which is a function of the actual return on assets, \( \mu \). The drift \( \mu \) can be computed from daily values of \( V_A \).

2. Logistic Regression Model

[29] is the first scholar to predict business bankruptcy by the logistic regression model. Along with [19, 31, 34] have successfully built their logistic regression models to predict contractor performance. This paper also employs the logistic regression model as the representative of our accounting-based model and a comparison method to Option-based model.

The logistic regression model is defined as a statistical modeling technique seeking the relationship between a binary dependent variable and other selected independent variables [22]. Let \( y_i \in \{0,1\} \) for all \( i = 1 \) to \( n \) (\( n \) is the number of samples), logistic regression model estimates the probability that the label is 1 for a given example \( x \) using the model [5]:
\[ P(y = 1 \mid x) = \frac{1}{1 + \exp(-\mathbf{w} \cdot \mathbf{x} - \beta_0)} \quad (3) \]

Parameters \( \mathbf{w} \) and \( \beta_0 \) can be estimated using the maximum likelihood procedure to maximize the log-likelihood function, with respect to \( \mathbf{w} \) and \( \beta_0 \),

\[ L(x_i, \ldots, x_n \mid \mathbf{w}, \beta_0) = \sum_{i=1}^{n} y_i \log p_i + (1 - y_i) \log(1 - p_i) \]

where

\[ p_i = P(y = 1 \mid x_i) \quad (4) \]

III. DATA AND VARIABLE SELECTON

1. Data

This paper collects Data from Compustat Industrial file and the Center for Research in Securities Prices (CRSP). This paper uses the accounting data from the Compustat annual file for calculations of accounting-based model. As for the option-based model, the market value of each contractor’s equity, \( V_E \), is computed from the CRSP database as the product of share price at each trading day and number of shares outstanding. For the risk free rate, \( r \), this paper uses the 1-Year Treasury Constant Maturity Rate obtained from the Board of Governors of the Federal Reserve System.

This paper focuses on construction contractors by choosing firms with Standard Industrial Classification (SIC) codes between 1,500 and 1,799. The sample contractors include three construction categories: Major Group 15 (Building construction, general contractors, and operative builders), Major Group 16 (Heavy construction other than building construction contractors), and Major Group 17 (Construction special trade contractors).

The samples period from 1970 to 2006 is selected with two criteria: First, The contractors must belong to both Compustat and CRSP for five consecutive years. Second, following [9, 15], this paper defines default firms as those with CRSP delisting code between 550 and 585, which represent delisted companies due to bankruptcy and other poor performance. As a result, 29 contractors were identified failed.

To avoid sampling error due to “selecting” a group of non-defaulters on which to perform the analysis, this paper uses every firm-year for which data are available. Finally, the combined sample of solvent and defaulted contractors consists of 1,560 firm-year observations representing 121 individual contractors.

2. Variable Selection for Accounting-Based Model

Selecting the accounting variables is the first stage in deriving an accounting-based default prediction model. This paper selects twenty variables for analyses which were commonly used in previous studies regarding contractor default prediction models including [1, 20, 21, 23, 24, 32-34]. These variables are shown in Table 1.

These ratios describe a contractor’s liquidity, leverage, activity, and profitability, and encompass a broad cross-section of accounting ratios.

While each of these variables provide important perspectives on a contractor’s condition, so many variables may yield a model that is “over-fitted.” In other words, the model performs very well in-sample on the data used to develop the model, but performs poorly on out-of-sample data [16]. This paper uses a forward stepwise logistic method to select a limited number of variables that yield a powerful model to avoid building an “over-fitted” model. The variables selected by forward stepwise logistic method are shown in Table 2.

In the following sections, this paper will put all 20 variables and the selected 4 variables for comparison in the accounting-based model.

| Table 1. Accounting variables chosen for this paper. |
|---|---|
| 1. Current Ration | 2. Quick Ratio |
| 5. Fixed Assets to Net Worth | 6. Total Liabilities to Net Worth |
| 7. Retained Earnings to Sales | 8. Debt Ratio |
| 11. Accounts Receivable Turnover | 12. Accounts Payable Turnover |
| 13. Sales to Net Worth | 14. Quality of Inventory |
| 15. Turnover of Total Assets | 16. Revenues to Fixed Assets |
| 17. Return on Assets (ROA) | 18. Return on Equity (ROE) |
| 19. Return on Sales (ROS) | 20. Profits to Net Working Capital |

| Table 2. Definition of variables selected by forward step-wise logistic method. |
|---|---|
| Variables | Description |
| ROA | (Net Profit After Interest and Taxes + Interest Expense)/Total Assets |
| Fixed Assets to Net Worth | Fixed Assets/Net Worth |
| Debt Ratio | Total Liabilities/Total Assets |
| Accounts Receivable Turnover | Net Sales/Average Receivables |
Table 3. Coefficient estimates for the Hybrid model.

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>S.E.</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-5.953</td>
<td>0.671</td>
<td></td>
</tr>
<tr>
<td>$X_1$</td>
<td>3.952</td>
<td>0.612</td>
<td>1.145</td>
</tr>
<tr>
<td>$X_2$</td>
<td>-2.394</td>
<td>0.820</td>
<td>1.094</td>
</tr>
<tr>
<td>$X_3$</td>
<td>0.007</td>
<td>0.007</td>
<td>1.014</td>
</tr>
<tr>
<td>$X_4$</td>
<td>1.488</td>
<td>0.902</td>
<td>1.066</td>
</tr>
<tr>
<td>$X_5$</td>
<td>0.000</td>
<td>0.000</td>
<td>1.007</td>
</tr>
</tbody>
</table>

3. Input Variables of the Hybrid Model

Option-based models are based on the presumption that all information is reflected on the company’s stock price, yet it is not always true in real life. On the other hand, though accounting information is lagged, it may be able to reflect the financial health of a company from a long-term perspective. Thus, this paper proposes a hybrid default prediction model that combines information from both accounting-based and option-based models by inputting the default probability from the option-based model into the logistic model as an input variable. Finally, the hybrid model is used as shown in Eq. (5). The coefficient estimates for the hybrid model are shown in Table 3.

$$DP = P(y=1 \mid x) = \frac{1}{1 + \exp(-W \cdot x - \beta_0)} \quad (5)$$

where

$$W \cdot x = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5$$

$X_1$ = Default Probability from the Option-based Model
$X_2$ = ROA
$X_3$ = Fixed Assets to Net Worth
$X_4$ = Debt Ratio
$X_5$ = Accounts Receivable Turnover
$y = 1$, if the observation goes into default and $y = 0$, if not.

From Table 3, the values of VIF show that these 5 variables do not have the problem of multicollinearity, which demonstrates the appropriateness of the hybrid model used in this paper.

IV. VALIDATION PROCESS AND RESULT

1. The Measure of Models’ Predicting Performance

This paper utilizes the Receiver Operating Characteristic curve (ROC curve) to evaluate the performance of models. ROC curve shows a graphic analysis of the trade-offs between type I error and type II error regarding to different cut-off points. The x-axis is shown by the percentile in ranking the non-defaulters from riskiest to safest, and the y-axis is the percentile of defaults excluded. Also, this paper applies the Area Under the ROC Curve (AUC) as the measure of models’ predicting performance, where a higher AUC is desired. In a perfect model, the value of AUC equals 1, whereas in a random model, the value of AUC equals 0.5.

2. Cross-Validation

The key assessment criterion for the accounting-based model is the out-of-sample performance, thus the whole sample is generally separated into two groups: training and testing groups in previous studies. The training group data is used to construct the models, while the testing group data is used to examine the performance of the models. Different selections of training data and testing data yield different results and sometimes lead to different conclusions. To avoid this problem, this paper conducts cross-validations. One firm-year observation is kept out-of-sample, and the remaining firm-years observations are used as training data to build the model. The observation kept out-of-sample is then put back into the pool and replaced by a second observation, and this process is repeated until every firm-year observation in the whole sample is tested.

The validation result set is the collection of all the out-of-sample model predictions, and can then be used to analyze the performance of the model. Note that the option-based model is based on a physical framework — it does not require any priors on whether a firm subsequently defaults.

3. Validation Result

The validation results of the option-based model, the logistic regression model and the hybrid model are summarized in Fig. 1 and Table 4.

The results shows that (1) the test set AUC of the option-based model is always higher than the accounting-based
logistic regression models. The option-based model’s predicting performance (AUC = 0.8581) is much higher than that of the 20-variable logistic regression model (AUC = 0.6066). Though the 4-variable logistic model’s predicting performance improved (AUC = 0.7519), the option-based model still outperforms the logistic regression model. This result is consistent with our contention that the information carried in accounting sheets is lagged and may be manipulated and the option-based models reflect timely and more comprehensive information.

(2) The predicting performance of the hybrid model (AUC = 0.8732) is higher than the option-based model or the logistic model alone. This result is consistent with our contention: the financial statements only provide information about a firm’s past performance and financial soundness, thus accounting-based model is limited in that it cannot provide information about a contractor’s future and qualitative factors relative to its success. The option-based model solves the above problem. However, option-based model has its limitations in application. In particular, it relies heavily on the condition that the market is efficient. Since most firms’ assets and liabilities do not possess the idealized characteristics and liquidity required by option-based models, there are lots of value uncertainty and potential arbitrage situations. By combining the option-based approach with accounting variables, they produce a new model that outperforms both accounting-based model and option-based model in the construction contractor default prediction.

V. CONCLUSION

Although several recent papers have used the option-based models to assess the likelihood of corporate failure, the construction industry is usually excluded in their empirical validation. Yet the construction industry has a relatively high default probability. This paper aimed to measure the construction contractor default risk using several methods.

A hybrid default prediction model is proposed by combining information from both accounting-based and option-based models by inputting the default probability from the option-based model into the logistic model as an explanatory variable. This was done in account for the limit of option-based models, which are based on the presumption that all information is reflected on the company’s stock price. Since this is not completely consistent with real life, accounting information can reflect information not shown in the stock price.

According to the empirical results, there were three major conclusions in this paper. First, after the forward stepwise logistic method to select input variables, the predicting performance of the logistic regression model improved. Too many input variables add training time to the models, yet don’t always improve the predicting performance. Sometimes they are even a disturbance and lower the model’s predicting ability.

Second, option-based models outperform enhanced accounting-based models in classifying defaulted and non-defaulted contractors. This result is consistent with our primary contention that accounting sheets are subject to manipulation and unable to show immediate default symptoms.

Third, since stock price cannot completely reflect all information of the companies in the real world, accounting information should be combined to default prediction. The hybrid model showed better predicting ability than both the accounting-based model and the option-based model.

The proposed modeling technique is useful to improve the construction contractor default forecasting, thus this paper recommends the proposed hybrid default prediction model as an alternative to the existing models.

REFERENCES


