Automation in Construction 20 (2011) 1242–1249

Contents lists available at ScienceDirect

Automation in Construction

journal homepage: www.elsevier.com/locate/autcon

An enforced support vector machine model for construction contractor default prediction

H. Ping Tserng [a,1], Gwo-Fong Lin [a,2], L. Ken Tsai [b,*], Po-Cheng Chen [a,3]

[a] Dept. of Civil Engineering, National Taiwan Univ., No. 1 Roosevelt Rd., Sec. 4, Taipei, Taiwan
[b] National Council of Structural Engineers Associations, 7F, No.37, Tung Hsing Rd., Taipei, 11070, Taiwan

ARTICLE INFO

Article history:
Accepted 11 May 2011
Available online 22 June 2011

Keywords:
Contractor analysis
Default prediction
Support vector machine

ABSTRACT

The financial health of construction contractors is critical in successfully completing a project, and thus default prediction is highly concerned by owners and other stakeholders. In other industries many previous studies employ support vector machine (SVM) or other Artificial Neural Networks (ANN) methods for corporate default prediction using the sample-matching method, which produces sample selection biases. In order to avoid the sample selection biases, this paper used all available firm-years samples during the sample period. Yet this brings a new challenge: the number of non-defaulted samples greatly exceeds the defaulted samples, which is referred to as between-class imbalance. Although the SVM algorithm is a powerful learning process, it cannot always be applied to data with extreme distribution characteristics. This paper proposes an enforced support vector machine-based model (ESVM model) for the default prediction in the construction industry, using all available firm-years data in our sample period to solve the between-class imbalance. The traditional logistic regression model is provided as a benchmark to evaluate the forecasting ability of the ESVM model. All financial variables related to the prediction of contractor default risk as well as 7 variables selected by the Multivariate Discriminant Analysis (MDA) stepwise method are put in the models for comparison. The empirical results of this paper show that the ESVM model always outperforms the logistic regression model, and is more convenient to use because it is relatively independent of the selection of variables. Thus, we recommend the proposed ESVM model as an alternative to the traditionally used logistic model.

1. Introduction

Evaluating the failure probability of construction contractors is a critical issue in successfully completing a project. Project owners or managers are suggested to avoid awarding contracts to those contractors with high failure tendency. It is also an important issue for other stakeholders such as surety underwriters, creditors, auditors, investors, and contractors themselves. As a result of growth in project scale, progression of construction techniques, and complication of materials and equipments, a single construction contractor is unable to complete a project alone. Thus, a successful construction project is highly dependent on the cooperation of prime contractor and sub-contractors. On the other hand, if one of them defaults or bankrupts, the others will also be affected. Therefore, nowadays prime construction contractors are concerned about the financial health of their sub-contractors and vice versa.

There is abundant literature on the development of prediction models of corporate failure, including univariate analysis [1], Multivariate Discriminant Analysis (MDA) [2–5], Linear Probability Model (LPM) [6], logistic regression model [7], probit model, and Cumulative Sums (CUSUM) procedure [8]. Most prior studies did not pay much attention to single industries, likely due to the limitedness of defaulted samples. Yet, economic intuition suggests that industry effects should be an important component in company default prediction. Chava and Jarrow [9] suggested that different industries face different levels of competition and have different accounting conventions; therefore, the likelihood of bankruptcy can differ for firms in different industries with otherwise identical balance sheets.

As the characteristics of the construction industry are highly different from other industries, the financial risk also differs from others. First, the construction industry is easily influenced by economic situation. Contractors often adopt the strategy of providing a lower price to win bids during economic turmoil, and thus have poor financial stability. Second, construction material and constructions in progress are the common items of inventory in the construction industry. When the inventory cannot be realized into cash due to contract disputes, the contractor suffers from insufficient liquidity. Third, contractors apply for advanced payment or progress payment from the owner according to the construction progress milestones,

* Corresponding author. Tel.: +886 2 87681117; fax: +886 2 87681116.
E-mail addresses: hptserng@ntu.edu.tw (H.P. Tserng), gflin@ntu.edu.tw (G.-F. Lin), p92748014@yahoo.com.tw (L.K. Tsai), f97521714@ntu.edu.tw (P.-C. Chen).
[1] Tel.: +886 2 23644154.
[2] Tel.: +886 2 33664368.
[3] Tel.: +886 917181585.

0926-5805/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.autcon.2011.05.007