CHAPTER 4 - METHODOLOGY

4.0. Introduction

This section includes the data set, univariate logistic regression as well as multivariate logistic regression.

4.1 Data Set

The definition of corporate failure can have a wide spectrum of interpretations. The majority of corporate failure prediction studies defined failure legalistically while selection of failed samples was made according to the Insolvency Acts. However, the definition of failed companies in this paper are those publicly traded companies involved in property development that have defaulted on their debt obligation and/or have filed for protection under Section 176 of Company Act. The main reason for not choosing the failed companies under the Bankruptcy Act in Malaysia is due to the fact that public listed companies in Malaysia rarely file for bankruptcy as this may jeopardize minority shareholders' interests, leading to public outcry.

This paper includes a sample of 78 companies that are involved in property development from 1995-1998 (see Appendix 2). 1998 is the year of failure following the start of the financial crisis in end-1997. Some of the companies in our sample are not directly listed under the property industry under the KLSE listing classification. Nevertheless, these companies have moved into property in a big way and investment in the property sector has exceeded 40 percent of their total net assets. Within the sample, there are in total 22 failed companies. The failed companies were identified from i) companies that seeked

protection in 1998 under Section 176 of Company Act, ii) listed companies classified under Practice Note 4 of KLSE requirement in 1998. PN4 companies are listed companies in financial distress with negative shareholders' funds. The sample of firms must have had at least three years of data prior to their failure.

The financial information needed for this study was collected from the data provided by Bloomberg electronic database, companies' annual reports and company announcements to the KLSE. Information on the firms was matched by the financial year of the respective companies.

4.2 Univariate logistic regression

Univariate logistic regression is useful for the identification of the potential predictor variables that will ultimately be part of the multivariate model. The objective is to evaluate the predictive ability of each variable for one, two and three years prior to failure. The selection of financial ratios was based on the statistical significance of the estimated parameters, the sign of each variable's coefficient and the model's classification results.

Univariate logistic regression was performed for twelve variables (including three dummy variables) to identify the statistically significant ratios, based on the following equation:

$$Pjt\,(Y=1) = 1/(1 + e^{-Z}) = 1/\{1 + \exp[-(bo + b1X1]\}$$

where

Pjt (Y=1) = probability of failure (1 for failed companies and 0 otherwise) for entity j at the end of year t; exp = exponential function; b1= slope coefficient; X1 = predictor variable.

4.3 The Multivariate Logit Model

The multivariate logit model utilises the coefficient of the independent variables to predict the probability of occurrence of a dichotomous dependent variable. Specifically, the technique weights the independent variables and creates a score for each company in order to classify it as failed or non-failed.

The phenomenon of company failure is discrete, and the dependent variable describing outcomes is a binary response, which means that we equate the event of "failure " with 1 and event of "non-failure" with 0. The function considered in the logistic regression is called the logistic function and can be written as follows:

$$Pjt\,(Y=1) = 1/(1+e^{-z}) = 1/\{1+\exp[-(bo + b1X1+b2X2+ \ldots.. + bnXn)]\}$$

where

Pjt (Y=1) = probability of failure (1 for failed companies and 0 otherwise) for entity j at the end of year t; exp = exponential function; b1, b2,....., bn = slope coefficients; X1, X2, ..., Xn = predictor variables.

Variables that are statistically significant in the univariate model will not necessary enter the multivariate model due to prediction problems such as multicolinearity, autocorrelation, etc. Nearly all the possible combinations of the variables were examined. Care, however was taken for the variables' coefficients to be significant at the 10% level of significance, for the sign of each variable's coefficient to be in accordance with the one stated in the hypothesis and finally for the model to provide high classification results. Most importantly, selection conforms to a priori findings.