國立臺灣大學工學院土木工程學系

碩士論文

Department of Civil Engineering

College of Engineering

National Taiwan University

Master Thesis

灰色理論結合強化樣本技術應用於營建公司破產幾率預測之研究

Predicting default probability in construction industry basing on Over-Sampling Technique for Grey SystemTheory

張克戰

Truong Khac Chien

指導教授：曾惠斌教授

Advisor: Prof. Tserng, Hui-Ping

中華民國103年6月

June, 2014

ACKNOWLEDGEMENT

This paper could not have been completed without the help, encouragement and support from a number of people who all deserve my sincerest gratitude and appreciation.

First of all, I would like to thank Prof. Tserng, Hui-Ping, my supervisor. I'm indebted to his inspiration, scholarly supervision and intellectual support throughout the course of writing this graduation paper. His continual encouragement, careful reading, critical comments and patient guidance made my work more enjoyable and easier.

Next, I owe sincere and earnest thankfulness to professors and staff members of not only Civil Engineering Department, National Taiwan University but also National University of Civil Engineering in Vietnam, for their supports during my study. Special thank is due to many graduates I have worked with, especially Po-Cheng Chen for his suggestion, assistance to complete my thesis. It is my pleasure to thank those who made the research possible.

I would also like to show my gratitude to my wife, Nguyen Thi Lien, for her love, encouragement, spiritual supports through the years. This dissertation would not have been possible without her supports.

Last but not least, I am truly indebted and thankful to my parents and my older brother. Their unconditional support, love, patience and understanding bring me overcome the challengers.

ABSTRACT

Bankruptcy Prediction has been a hotly-debated topic among many people in business area. The fact is that once the firm goes bankrupt, it will be disastrous to not only firm itself but also other stakeholders. Many available methods have been applied to predict the possibility of business collapse; almost all of them were based on financial ratio analysis. Grey System Theory, used in the previous thesis for predicting default probability of construction firms, has brought some feasible results, by relying on the 19 initial financial ratios.

This study, with the aim of enhancing the Grey Theory application, employs Over-sampling technique before applying Grey System Theory. The results of this study are then compared with those of the prior research. Furthermore, replication and Synthetic Minority Over-sampling Technique (SMOTE), two over-sampling techniques are proposed to resolve the imbalance problem in data set.

The results reveal that over-sampling techniques could improve the predicting performance of Grey System theory. Additionally, between these two kinds of over-sampling techniques, SMOTE surpasses Replication in terms of prediction capability.

Key words: Over-sampling, SMOTE, Grey System Theory, Synthetic Degree Incidences, financial ratio, ROC curve, construction industry.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Background

Bankruptcy Prediction has inherently been a common topic in business area. The economic crisis over the last few years has made this topic more urgent than ever before. A business is declared bankrupt when it is unable to pay off its debts. Once the firm goes bankrupt, it will cause intense damages to both the firm itself and its shareholders alike.

No parts of business are immune to bankruptcy. Bankruptcy can happen in any aspects of business, ranging from monetary industries such as banking, financial institutions to non-related monetary sectors such as manufacture, construction and so on. However, almost all previous studies touching bankruptcy prediction overlook the specific area; they solely focus on general business area.

The construction industry indeed plays an important role in enhancing the economic performance and the national welfare of a country by transforming various resources to construct economic and social facilities (Basir, 2000). The construction project is usually a prolonged and a costly process. Construction enterprises, due to high competitiveness, have to reluctantly reduce their profit to win the bid, which exposes them to the default risk. Besides, the cycle of construction projects are relatively long, so the income of the construction firms might be easily affected by the price changes in materials, manpower, machine’s expenditure, as well as the adjustment of legislation and policies. Therefore, regular evaluation of financial performance should be priority of construction firms so that they are able to detect any potential company failures at the earliest opportunity, from which timely and appropriate strategies can be put in place to help them to recover. The early warning of construction contractor failure is an important issue for government organizations, construction owners lending institutions, surety underwriters, and contractors (Tsai et al., 2011).

1.2 Motivation and Problem statements

The construction industry always plays a crucial role in a country's economic flourish. It is the cornerstone and connecting factor of the other industries. However, the contractors are confronting with lots of difficulties in an increasingly competitive environment nowadays. The past data collected from the U.S. market indicated that the failure rate among construction firms has reached a critical level (Kangari et al 1992; Ekanayake 2008). Meanwhile, some highlighted characteristics of the construction industry and construction project such as unique, long term investment, big invested capital, etc. make the financial characteristics of construction industry different from those of other industries.

A lot of models anticipating business malfunction have been proposed.(e.g. Bevear, 1966; Altman, 1968; Ohlson, 1980; Shin, 2001; Chen 2004; Chien, 2005). Nevertheless, most of available researches focus on default prediction of business in banking and finance sectors; quite few researches work on the default probability prediction of construction firms.

Earlier statistical models focused on anticipating bankruptcy utilizing statistical discrimination methods. For example, Beaver (1966) studied univariate discrimination models. His model emphasized individual signals of impending matters without revealing the interaction of variables. Altman (1968) established a popular multivariate prediction framework known as the Z- score model. Then, 1977, Altman et al, developed a more comprehensive discriminant model, the Zeta model. Ohlson (1980) used the maximum likelihood estimation of conditional logit model in developing the capacity of bankruptcy prediction. Despite some significant contributions, the aforementioned researches have a number of limitations: Firstly, they only used financial ratios did not consider other factors such as economy, operation, and management; therefore, did not meet the total relationship between the causes and effects of business failure. Secondly, static models ignore the time –series effects of a firm's financial and operational performances on the risk of business bankruptcy. Lastly, these models just looked at industries in general; less attention was paid to construction industry.

In conclusion, understanding the mechanism of failure is a surefire way to business breakdown avoidance. Furthermore, little attention has been given to predicting the probability of construction company failure. Due to the high risk default probability of construction industry and its different characteristics from others'; there should be study to predict default probability of the construction contractors.

The previous thesis of Le Quyen showed some fairly accurate results in the application of Grey System Theory on the default probability of the construction firms. However, some shortcomings were inevitable. The number of non-defaulted samples greatly exceeds the defaulted samples. Consequently, to tackle imbalance data problem, in this research, two different over-sampled approaches namely Enforced Training and Synthetic Minority Over-Sampling Technique were used.

1.3 Research objectives

Apply the Grey system theory for analyzing the impact of financial ratios on the default probability of the construction firms.

Apply over-sampling technique before using Grey System to resolve imbalance data problem (the number of non-defaulted samples greatly exceeded the number of defaulted samples). The present research aims to achieve two principal objectives:

1) Compare with previous research which method gives higher efficiency.

   Give the result that: whether Grey system Theory has the imbalance problem or not.

2) Enhance the usefulness of Grey Theory

1.4 Research scope and limitations

This research only used financial ratios. Other factors such as economy, operation, and management were not considered in this study; therefore, the total relationship between the causes and effects of business failure was not met.

The data used in this thesis are obtained from US stock market in the period of 1970-2006, and concentrate on construction firms which have different characteristics from other industries'. Therefore, this study only concentrates on construction contractor. It means that the results should be cross-checked in data of other industries.

Additionally, the accuracy of any financial ratio-based predictive model largely depends on the reliability of accounting data source. In some real cases, financial data might be manipulated, which leads to failure of model. Besides, the data used in this research are all available data firm years and the bankrupt sample is the data from the last financial statement issued before the firms declared bankruptcy. That is, this model is not able to assure for predicting insolvency more than one year prior to bankruptcy.

1.5 Thesis structure

The remainder of this research is organized as follow:

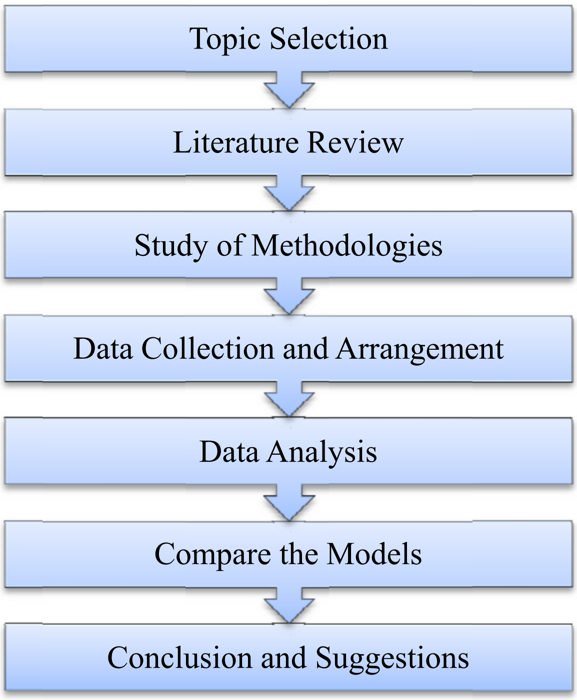


Figure 1.1: The produce of research

Figure 1-1 shows procedure of this research. Base on the above steps, the framework of thesis is divided in five chapters as follows:

Chapter 1: Introduce the background, motivation, purposes and contributions, and thesis structure.

Chapter 2: Review the literature, regard to the previous default prediction models which applied for non-construction and construction companies and related with the present research.

Chapter 3: Introduce the methodologies of default prediction, Grey System Theory, how to apply Over-sampling technique to resolve imbalance data problem, and how to use ROC curve (Receiver Operating Characteristic curve) to evaluate the prediction power of this model.

Chapter 4: Presents the standard of data collection.

Chapter 5: Analyzing and validating input data.

Chapter 6: Present the conclusions and suggestions.

CHAPTER 2: LITERATURE

This chapter presents a brief overview of the development of bankruptcy prediction model in both general industry and construction industry.

2.1 Default prediction researches

There have existed numerous approaches to the corporate failure prediction in general business area so far. Beaver was the first scholar using financial ratio in predicting bankruptcy. After Beaver, a lot of other researchers employed different methods such as multivariate discriminant analysis (Altman, 1968), logit (Ohlson, 1980) or probit (Zmijewski, 1984). These models were then developed in accordance with the information form financial statements to evaluate strengths and shortcomings of a company's financial status.

The application of financial ratio models to determine a business' profitability, and hence chances of its survival, attracted many researchers' concern in both general business (Altman, 1967; Beaver, 1968; Taffler, 1983; Robertson (1984); Keasey and Watson, 1986) and the construction domain alike (Mason and Harris, 1979; Kangari, 1988; Abidali, 1990; Russell and Jaselski, 1992; Langford et al., 1993; Ramsey Dawber, 1993)

1. **Fitz Patrick (1932)**

The first author to use financial ratio model was Fitz Patrick (Fitz Patrick, 1932). He studied 19 bankrupt firms and 19 non-bankrupt ones. The research found 3 years before bankruptcy, the financial ratios were significantly changed.

2. **Beaver (1967)**

William H. Beaver is one of the first developers of business failure prediction using quantitative method. The methodology adopted by him is discriminate analysis (DA). He collected data from Moody's Industrial Manual – 79 firms that collapsed during 1954 to 1964. None of these companies was construction firm; they mostly were in manufacturing sector. Non-defaulted firms in the same industry and asset size were also assembled to distinguish and discriminate against the distressed firm.

There were 30 ratios commonly used in the financial literature in Beaver's study. His research aims to discover the capability of these ratios to differentiate the bankrupted group from the non-bankrupted one. In his study, various methods such as mean values, mean asset size, dichotomous classification tests and analysis of likelihood ratios were integrated to analyze the ratios. After finalizing the calculated process of these ratios for the period of 5 years prior to bankrupt, the result of Beaver's study indicated that the cash flow – total debt ratio was the overall best predictor. Another contributor of his research was the affirmation of using accounting data in the forecast of business bankrupt. What could be drawn from his study was that using financial ratio could give the early warning of 5 years before bankruptcy.

3. **Altman (1968)**

Beaver's research has certain values when pointing out the importance of using financial ratio in predicting default; however the discriminatory power of the independent ratio makes this research not a perfect one. To address this drawback, E.Altman (1968) developed Beaver's method and established an innovative model named mutlti-variate approach. In this research, Altman also reconfirmed the principal role of using financial ratios as a predictor of corporate. Altman used 33 failed manufacturing firms matching up with the same number of non-failed firms in the period of 1946 to 1965. The criteria of matching were the same industry and roughly similarity in asset size. Using the Multiple Discriminant Analysis (MDA) technique in the study, the 22 financial ratios served as input variables to analyze. The stepwise method then applied to choose an optimal combination of five variables from the 22 ratios initially selected. Finally, Altman's model proposed a following linear function using five variables

$$Z = 1.2X_1 + 1.4X_2 + 3.3X_3 + 0.6X_4 + 1.0X_5 \quad (1)$$

Whereas:

X_1: Working Capital/Total Assets

X_2: Retained Earnings since Inception/Total Assets

X_3: Earnings before Taxes and Interest/Total Assets

X_4: Market Value of Equity/Book Value of Total Debt

X_5: Turnover/Total Assets.

Accordingly, businesses were classified as follows:

- Z - score less than 1.8 implied certainty of imminent failure;
- Z - score between 1.8 and 2.7 revealed the "zone of ignorance" or 'grey area' , where companies were deemed to be at risk; and
- Z - score greater than 2.7 (initially 2.9), indicated a potential for long term solvency.

The model correctly classified 95 percent of the total 66 sample firms (correctly classifying 94 percent as bankrupt firms and 97 percent as non-bankrupt firms) one-year prior to bankruptcy. The percentage of the accuracy fell down with increasing number of years before bankruptcy. This technique has a strong reputation in the history of corporate bankruptcy models until the 1980s and is weidely applied as the baseline for comparative studies.

4. **Ohlson (1980)**

Ohlson (198) is another researcher using financial ration as predictors of bankruptcy to develop his bankruptcy prediction model. His research analyzed nine financial ratios of firms' size, leverage, liquidity and performance, using logistic regression. Data were gathered from 1970 to 19976 and included 105 defaulted and 2,058 non-defaulted industrial enterprises. His model's kernel function was built as follows:

$$\text{Probability of default} = \frac{e^z}{1+e^z} = \frac{1}{1+e^{-z}}$$

Z = -1.3-0.4X1 + 6.0X2 – 1.4X3 + 0.1X4 – 2.4X5 -1.8X6 + 0.3X7 – 1.7X8 -0.5X9

Where:

X1= Log (Total Assets / GNP Price-level Index)

X2 = Total Liabilities / Total Assets

X3 = Working Capital / Total Assets

X4 = Current Liabilities / Current Assets

X5 =1 if total liabilities exceed total assets, o if otherwise

X6 = Net Income / Total Assets

X7 = Funds provided by Operation / Total Liabilities

X8 = 1 if net income was negative for the last two years, 0 if otherwise

X9 = Measure of Change in Net Income

5. Taffler (1983)

In the UK, another two authors Taffler and Tishaw (1977) adopted a similar methodology as other seniors', basing on a sample of 92 manufacturing companies. The resulting Z score equation was based on a combination of four categories ratios; however, undisclosed coefficients:

$$Z = c_0 + c_1X_1 + c_2X_2 + c_3X_3 + c_4X_4 \qquad (2)$$

Where:

X_1: Profit before Tax/Current Assets (53%)

X_2: Current Assets / Current Liabilities (13%)

X_3: Current Liabilities/Total Assets (18%)

X_4: No Credit Interval (16%)

The percentages give guidance to the relative weightings of the ratios. Taffler and Tishaw declared a 99% successful classification based on the original 92 companies from which the model was conducted. However, this success assurance lost its value when the model was re-tested by Taffler (1983) with a sample includes 825 companies.

The two models were developed by Altman and Taffler both bolstered the significance of the ratio variable of turnover to total assets as a positive indicant that contributed to corporate bankruptcy.

6. Robertson (1984)

In an effort to address the question of a theory on corporate failure, Robertson (1984) developed a ratio model that worked with general applicability to all industries. He declared that there were a priori determinants of corporate failure from their financial ratios. Robertson suggested ratios expressing trading stability, declining profits, declining working capital and increases in borrowing as predictive characteristics. Instead of the simple turnover to total assets utilized by Altman (1983) and Taffler (1983); Robertson (1984) utilized the ratio of turnover less total assets to turnover, to display the importance of trading stability in his model. The outcome model combined five ratio variables were presented as Equation 3:

$$Z = 0.3X_1 + 3.0X_2 + 0.6X_3 + 0.3X_4 + 0.3X_5 \qquad (3)$$

Whereas:

X_1: (Turnover – Total Assets)/Turnover;

X_2: Profit before Tax/Total Assets;

X_3: (Current Assets – Total Debt)/Current Liabilities;

X_4= (Equity – Total Borrowings)/Total Debt; and

X_5= (Liquid Assets – Bank Overdraft)/Creditors.

2.2 Default prediction researches in construction industry

According to S. Thomas NG, "Pertinent forecasting techniques for construction company failures include the (1) ratio analysis; (2) multiple discriminant analysis; (3) conditional probability models; and (4) subjective assessment". In completing the present study, the author read several relevant studies in the construction industry as follow:

1. **Mason and Harris (1979)**

Mason and Harris (1979) developed a six-variable model to assess construction organizations in UK. In this study, a sample of 20 bankruptcy and 20 non- bankruptcy firms was selected. Basing on the MDA), the discriminant function was developed as below equation. A positive Z-score indicated a long-term solvency, while a negative value was classified as a potential failure. Their model was conducted with a multiple regression approach and presented as:

$$Z = 25.4 - 51.2X_1 + 87.8X_2 - 4.8X_3 - 14.5X_4 - 9.1X_5 - 4.5X_6 \quad (4)$$

Where:

X_1: Profit Before Tax and Interest/Opening Balance Sheet Net Assets;

X_2: Profit before Tax/Opening Balance Sheet Net Capital Employed;

X_3: Debtors/Creditors;

X_4: Current liabilities/Current assets;

X_5:Log10 (days debtors); and

X_6: Creditors Trend Measurement.

While a positive Z-score indicated a long-term solvency, a negative value revealed a potential bankruptcy. The variable profit before tax and interest to opening balance sheet net assets (X_1) was indicated as a negative sign in the research. This implies that a higher value of a return on net assets produces a greater tendency for bankruptcy, which is rather unconvincing.

2. Abidali, 1990

Abidali also developed a Z-score model used in vetting construction companies on the tender lists. Using multivariate discriminant analysis to produce a predictive model including seven variables, 31 different variables were initially adopted. The best discriminating variable is selected according to Wilks Lambda criteria. The Z-score model is shown below in following equation:

$$Z = 14.6 + 82.0X_1 - 14.5X_2 + 2.5X_3 - 1.2X_4 + 3.55X_5 - 3.55X_6 - 3.0X_7 \quad (5)$$

Where:

X_1 =Profit after Tax and Interest/Net Capital Employed;

X_2 = Current Assets /Net Assets;

X_3 =Turnover/Net Assets;

X_4 =Short Term Loans/Profit before Tax and Interest;

X_5 =Tax Trend over three years;

X_6 =Profit after Tax Trend over three years; and

X_7 =Short Term Loan Trend over three years.

3. Russel (1988) and Jaselskis (1992)

While some previous researchers only focused on analyzing financial variables, Russell and Skibniewski (1988) deepened their research by presenting all the factors involved in the construction contractor prequalification decision-making process, which are closely related to contractor default risk. Beside financial soundness, management capability, and economic condition as well as technical expertise are also essential factors to construction contractors' success. Their research model was introduced with 5 variables, cooperating 4 financial ratios and one management related variable:

$$Y = 2.27 - 7.72 X_1 + 45.05 X_2 + 13.94 X_3 - 13,24 X_4 - 34.42 X_5 \qquad (6)$$

Where:

X_1: Cost Monitoring (not performed = 0; performed = 1);

X_2: Under-Billings to Sales;

X_3: Total Current Liabilities to Sales;

X_4: Retained Earnings to Sales; and

X_5: Net Income before Tax to Sale.

The predictability of the model is rather well: among forty sample companies, there were only 12.5 % misclassified. However, many of introduced factors are qualitative and largely depend on human judgment; incorporating them into the default prediction model with bias is potential.

4. Kangari et al. (1992)

By using multiple regression method, Kangari developed a performance index to grade a company by regressing 6 financial ratios. The researcher used the financial report of 126 construction companies and divided them into 6 groups. Financial ratios were used in the research as following:

- Current ratio.

- Total liabilities to net worth.

- Total asset to revenue.

- Revenues to net working capital.

- Return on total assets.

- Return on net worth.

2.3 Grey system model in prediction bankruptcy probability

In 1982, Professor Deng Ju-Long published “Control Problems of Grey Systems”, which signaled the coming of a new theory: the grey systems theory which managed to rapidly develop and even to impose. Grey systems theory is highly valued because of its practical applicability and been widely applied in analysis, modeling, prediction, control, decision making, in almost all areas: social, economic, mechanical and technical science, agriculture, industry, transport, petrology, meteorological, ecological, hydrological, geological, financial, medical, military, and others (Liu and Lin, 2005). The main characteristic of grey system theory is that it manages to achieve good performance in analysis based on a small range of data and on a large number of variables.

1. **Ping, J. & Kejia, C. (2005)**

Ping, J. and his colleague, Kejia, C. (2005) claimed that the theory of grey systems focuses more on the output of the systems rather than their structure and input. Moreover the theory allows grey quantity and grey relationship within them. Two scholars applied grey system analysis to design an economic cycle monitor and early warning index system. Among many kinds of degrees of grey incidences, just the absolute degree of grey incidences was shown in their research.

2. **Cheng, J. *et al* (2009)**

In 2009, Cheng, J. *et al.* conducted a hybrid model which enabled the prediction of failure firms based on their past financial performance data, combining grey prediction and rough set approach. They used 14 financial ratios considered cover all the categories suggested by previous studies, including:(a) Solvency: current ratio; quick ratio; liabilities/assets ratio; times interest earned ratio;(b) Managerial performance: average collection turnover,(c) Profitability : return on total assets; return on shareholders' equity; operating income to paid-in capital; profit before tax to paid-in capital; earnings per share, (d) Financial structure: shareholder's equity/total assets ratio, and (e) Cash flow: cash flow ratio; cash flow adequacy ratio; cash flow reinvestment ratio.

Cheng, J. *et al.*(2009) computed prediction value of the fourth year, the fifth year, the sixth year and the seventh year history data respectively for announcement and comparing to history data. The final result was that grey prediction business failure prediction models in the 4th and the 5th dimensions had better performance than history business failure prediction models. Specially, accuracy of grey prediction business failure forecasting model in the 5th dimension has the best performance. The general results are very encouraging, compared with original rough set, and prove the usefulness and strengthen the effectiveness of the proposed method for company failure prediction.

3. Delcea, C. &Scarlat, E

Basing on grey system theory, Delcea, C. and Scarlat, E determined a "matrix of symptoms" which represented by economic- financial ratios, usually used by analysts to make predictions and suggestions. The ability to create such a matrix of symptoms implies that given level of symptom's intensity, they could determinate if the analyzed firm presents some "diseases". In their analysis, the researchers introduced the existence of 9 symptoms as it follows:

- S1: Solvability (positive symptom - it shows the capacity of a firm to pay its debt within the time prescribed – as the firm is solvent, its financial situation is better)
- S2: Quick Ratio (positive symptom)
- S3: Working Capital (positive symptom)
- S4: EBIT-Yield (positive symptom)
- S5: Interest Cover Ratio (positive symptom)
- S6: Profit Margin (positive symptom)
- S7: Return On Equity (positive symptom)
- S8: Return On Total Assets (positive symptom)
- S9: Gearing (negative symptom)

This analysis found a way by which a possible "disease" or bankruptcy can be anticipated, and found a way to highlight the occurrence of such a phenomena. However, due to the fact that the accuracy of the methodology was not proved, the doubtfulness of the method's reliability is inevitable.

2.4 Summary

Financial ratio is regarded as one of the most popular methods to determine the profitability and the potential turndown of a business. The mentioned researches proposed many financial ratios, which are generally classified in five groups: (1) Liquidity; (2) Profitability; (3) Leverage; (4) Solvency; and (5) Activity. The way these ratios reflect the firm financial situation was also be displayed. Truthfully, Grey system theory was rapidly improved because of its high value in application and the application of grey system theory to deal with the prediction firm default probability problem was very effective. This sparks my interest in improving the effectiveness of Grey System Theory in the default probability of the construction firms.

CHAPTER 3: METHODOLOGY

3.1 Grey system theory

Widespread divisions in the activities of scientific research and the technological advancement have led to a tendency in the modern spectrum of science and technology. This tendency is indicated by the rapid rise of many cross-disciplinary research activities as well as appearance of many important theories. Grey systems theory is one of such significant cross disciplinary theories. The release of “The Control Problems of Grey Systems” by Professor Deng Ju Long (1982) of China marked an important and fruitful area of research with strong and successful practical applications. As mention above, grey systems theory, because of its efficacy, has been popularly applied in analysis, modeling, prediction, control, decision making in almost all areas: social, economic, mechanical and technical science, agriculture, industry, transport, petrology, meteorological, ecological, hydrological, geological, financial, medical, military, and others.

Among probability and statistics, fuzzy mathematics, and grey systems theory - three most-often applied theories and methods employed in studies of non-deterministic systems, the last one have proved to be the most effective method. Grey theory addresses the obstacles encountered in the utilizing of probability theory and statistical methods (the need of reasonable size samples and determination of certain distributions to draw a valid inferences) and those of fuzzy mathematics (which deals well with the study of problems with cognitive uncertainty phenomena, using so-called “membership functions”, based on experience). The main valued characteristic of grey system theory is that it manages to achieve good performance in analysis conducted on a small range of data and on a large number of variables.

3.1.1. Methods of Grey Numbers' Generation Based on Average

The shortage of data poses researchers a lot of problems. This is exclusively true in any economic analysis. There have many cases that the incomplete collected data set of an observed economic system causes the researchers many difficulties in the undertaken analysis. Also, on the collected data, it may happened that in the initial data set, some values to be abnormal, much higher or much lower than the other values of the series, and thus, make an analysis based on such a data set lead to erroneous results. For the abnormal values' existence, a probability can be to identify and to filter them out from the data set, and then it returns to the case where we have blanks in the data set. Nevertheless, grey system theory gives us a method to solve this problem, namely, generate grey sequence method to give birth to new values for filling gaps in data sequence.

Consider the data sequence analyzed contains "empty" information which denoted with φ (k), in which k represents the position in the data sequence. In this case, the data sequence X indicates as follows (*Liu, S.F., Lin, Y. (2006). Grey Information: Theory and Practical Applications. Springer, London*):

$$X = (x(1), x(2),\ldots , x(k-1),\varphi (k), x(k+1),\ldots, x(n))$$

The number value φ (k) is in the range delimited by x(k −1) and x(k +1), and the two values stand for the lower and the upper limit of the unknown value.

$$x^*(k) = 0.5x(k-1) + 0.5x(k+1) \qquad (3.1)$$

$X^*(k)$ is called a generated mean value of consecutive neighbors as being a generated average value based on two non-consecutive neighborhood values.

In grey systems modeling (GM), the mean generation of consecutive neighbors is often used. This method based on the raw sequence of data to build new sequences in order to reveal the particular trend, if any. In the firm analysis, the method of generating numbers based on average may be utilized if the objective of the research is to observe the evolution in time of a particular variable, and for certain periods of time, for various reasons, those are unknown to the researcher.

Take an example; we take into account the sequence, which represents the quantity of products by the firm Y (expressed in pieces) during a year, with monthly record:

$C = C(c(1), c(2), c(4), c(6), c(9), c(10),c(11), c(12))$

$= (350,590,510,420,580,720,810,790)$

As it can be seen, from a total of 12 months, we only know the quantities sold in 8 months, and for the remaining months, we can estimate the quantities by using the method of generating numbers based on average:

$C(3) = 0.5*c(2) + 0.5*c(4) = 0.5*590 + 0.5*510 = 550$

$C(5) = 0.5*c(4) + 0.5*c(6) = 0.5*510 + 0.5*420 = 465$

$C(8) = 0.5*c(6) + 0.5*c(10) = 0.5*420 + 0.5*720 = 570$

$C(7) = 0.5*c(6) + 0.5*c(8) = 0.5*420 + 0.5*570 = 495$

Follow these above steps; we gained the consequence of the quantities sold in 12 months:

C = C(c(1), c(2), c(3), c(4), c(5), c(6), c(7), c(8),c(9), c(10),c(11) c(12))

= (350,590,550,510,465,420,495,570,580,720,810,790).

3.1.2. Method of Grey Incidence Analysis

Grey incidence analysis method is one of the essential and principal sectors of grey system theory. "The fundamental idea of grey incidence analysis is that the closeness of a relationship is judged based on the similarity level of the geometric patterns of sequence curves. The more similar the curves are, the higher the degree of incidence between sequences, and vice versa" (Liu.S & Lin.Y, 2006). Choosing the right sequence of characteristic data to describe the system's behavior is the most important step when analyzing an abstract system or phenomenon. This sequence of data is called a mapping quantity of the special system's behavior. Solving the problem of diagnosis of "diseases" that can appear in the company operation, depending on the identified symptoms.

Most of the models used in diagnosis stand on a ground of on a set of data collected at firm level (sometimes processed data), in a given period of time. Normally, researchers consider the values of symptoms at the level of year t, in most of the cases this is the year prior to a company's distress, without attempting to link these values with the changes that have occurred in the indicator over a greater period, for example 3-5 years. In the previous research, the researcher proposed 3 models of which 3, 4 and 5 years, and gave the conclusions that with the 4 years analysis, the result is the best. Therefore, in this research, the Grey System with 19 financial ratios in the model of 4 years is employed to get the closest comparison. (If data are partially missing, they can be calculated using methods of grey numbers' generation based on average).

This research aims to identify which one of the considered symptoms (financial ratios) should be a sequence characteristic describe system's behavior and manifest the biggest influence in each firm as well as to establish a hierarchy of them, in order to build a matrix of symptoms at the level of all the firms considered. The author introduce 4 years analysis model as an example. Firm analysis utilizes the correlation matrix between the level of each symptom of a firm and the corresponding year, for a universe of time equal to 4 years, is noted as follows (*Liu, S.F., Lin, Y. (2006). Grey Information: Theory and Practical Applications. Springer, London*):

$$X_{\mathrm{ij}} = \begin{bmatrix} X_{11} & X_{12} & X_{13} & \cdots & X_{1n} \\ X_{21} & X_{22} & X_{23} & \cdots & X_{2n} \\ X_{31} & X_{32} & X_{33} & \cdots & X_{3n} \\ X_{41} & X_{42} & X_{43} & \cdots & X_{4n} \end{bmatrix}$$

In the previous research, the author proposed 19 financial ratios as major symptoms related to firm's financial statement.

The author will try to build an incidence level matrix of each symptom at the firms:

$$\rho_{\mathrm{ij}} = \begin{bmatrix} \rho_{1,1} & \rho_{1,2} & \rho_{1,3} & \cdots & \rho_{1,19} \\ \rho_{2,1} & \rho_{2,2} & \rho_{2,3} & \cdots & \rho_{2,19} \\ \rho_{3,1} & \rho_{3,2} & \rho_{3,3} & \cdots & \rho_{3,19} \\ \rho_{m,1} & \rho_{m,2} & \rho_{m,3} & \cdots & \rho_{m,19} \end{bmatrix}$$

Whereas: m is the number of firms.

To establish an incidence level of matrix includes m firms, the author calculate the absolute degree of grey incidence, the relative degree of grey incidence and combine the two to get the synthetic degree of grey incidence, which will determine which firm has a bad performance business.

Absolute degree incidence ε_{ij} is only related to the geometric shapes of X_i and X_j, and has nothing to do with the spatial positions of X_i and X_j. The more X_i and X_j are geometrically similar, the greater ε.

The sequence to compute the absolute matrix of incidence as follow:

- Firstly: For each behavioral sequence, we compute its image of zeroing starting point:

$$X_j = (x_{1j} - x_{1j}, x_{2j} - x_{1j}, x_{3j} - x_{1j}, x_{4j} - x_{1j}, x_{5j} - x_{1j})$$

Whereas j = 1, ... , n. n= number of symptom. In this research, n = 4.

- Secondly: Find $|s_0|$, $|s_j|$, and $|s_j - s_0|$

$$|s_0| = \left| \sum_{k=2}^{4} x_0^0(k) + \frac{1}{2} x_0^0(4) \right| \quad (3.2)$$

$$|s_j| = \left| \sum_{k=2}^{4} x_j^0(k) + \frac{1}{2} x_j^0(4) \right| \quad (3.3)$$

$$|s_j - s_0| = \left| \sum_{k=2}^{4} \left[x_j^0(k) - x_0^0(k) \right] + \frac{1}{2} \left[x_j^o(4) - x_0^0(4) \right] \right| \quad (3.4)$$

- Lastly, attain the absolute degree of grey incidences:

$$\varepsilon_{0j} = \frac{1+|s_0|+|s_j|}{1+|s_0|+|s_j|+|s_j - s_0|} \quad (3.5)$$

1. *Relative Degree of Grey Incidence*

The relative degree of grey incidence is obtained using the following relations:

- First, compute the initial images of X_0 and X_j

$$X_0' = \left(\frac{x_0\,(1)}{x_0\,(1)},\frac{x_0\,(2)}{x_0\,(1)},\frac{x_0\,(3)}{x_0\,(1)},\frac{x_0\,(4)}{x_0\,(1)},\frac{x_0\,(4)}{x_0\,(1)}\right) \quad (3.6)$$

$$X_j' = \left(\frac{x_j\,(1)}{x_j\,(1)},\frac{x_j\,(2)}{x_j\,(1)},\frac{x_j\,(3)}{x_j\,(1)},\frac{x_j\,(4)}{x_j\,(1)},\frac{x_j\,(4)}{x_j\,(1)}\right) \quad (3.7)$$

- Compute the images of zero starting points of X_0' and X_j'

$$X_0'^0 = \left(x_0'^0(1) - x_0'^0(1), x_0'^0(2) - x_0'^0(1), \dots, x_0'^0(4) - x_0'^0(1)\right) \quad (3.8)$$

$$X_j'^0 = \left(x_j'^0(1) - x_j'^0(1), x_j'^0(2) - x_j'^0(1), \dots, x_j'^0(4) - x_j'^0(1)\right) \quad (3.9)$$

$$|s_0'| = \left|\sum_{k=2}^{4} x_0'^0(k) + \frac{1}{2}x_0'^0(4)\right| \quad (3.10)$$

$$|s_j'| = \left|\sum_{k=2}^{4} x_j'^0(k) + \frac{1}{2}x_j'^0(4)\right| \quad (3.11)$$

$$|s_j' - s_0'| = \left|\sum_{k=2}^{4}(x_j'^0(k) - x_0'^0(k)) + \frac{1}{2}(x_j'^0(4) - x_0'^0(4))\right| \quad (3.12)$$

- Compute the relative degree of incidence

$$r_{0j} = \frac{1+|s'_0|+|s'_j|}{1+|s'_0|+|s'_j|+|s'_j - s'_0|} \quad (3.13)$$

2. *Compute the synthetic degree of incidence*

$$\rho_{1j} = \theta\, \varepsilon_{1j} + (1-\theta) r_{1j} \quad (3.14)$$

Whereas:

ρ_{1j}: The synthetic degree of incidence

ε_{1j}: The absolute degree of incidence

r_{1j}: The relative degree of incidence

With j = 2... n, (n = number of symptom. This research, n = 4) θ □ [0,1]

The synthetic degree of grey incidence is a numerical index that well describes the overall relationship of closeness between sequences. It reflects the similarity between the zigzagged lines X_i and X_j , and also demonstrates the degree of closeness of the individual rates of change of X_i and X_j with respect to their initial points. In general, previous researchers usually take $\theta = 0.5$. If we are more interested in the relationship between some absolute quantities, some greater value can be used as θ. In the case we are putting more emphasis on rates of change, some smaller value can be employed for θ. (Liu, S. and Lin, Y.). In the previous study scope, the researcher firstly proposed θ = 0.5 as fix value, so in this thesis scope, $\theta = 0.5$ is chosen to compare with the same condition.

The synthetic degree of grey incidence is based on the absolute and relative degrees of grey incidence obtained earlier. The size of synthetic degree of grey incidence obtained: ρ_{12}, ρ_{13}… ρ_{1n}, determine the degree in which each symptom influence the firm and conduct it to a bankrupt one. As the synthetic degree of grey incidence is higher, its corresponding variable (financial ratio) is more important.

The analysis was carried out at a single firm level then the same process with each of the consider firms will be taken into analysis. Combine the synthetic degree incidence with the sign (+/-) of each symptom, the hierarchy default probability of the sample firms are identified. About a variable j we are saying that it has a positive sign (greater is better) as it takes a higher value, the analyzed firm presents a better financial situation. Otherwise, the sign is negative (less is better). The hypothesis of variable sign will detail analyzed in the next chapter, chapter 4- Data collection). Once the default intensity of firms was determined, ROC curves (receiver operating characteristic curves) will be used to calculate the accuracy rate of prediction.

3.2 Over-sampling technique

3.2.1 Between-class Imbalance Problem in Date Set

Previous researches normally focused on “sample-matching” method to build the sample sets. In this method, each defaulted contractor was matched with one or two non-defaulted contractors at the year of default. However, Zmijewski (1984) pointed out that this sample-matching method led to choice-based biases and sample selection biases. If model is not built basing on entire population, the estimated coefficients will be biased. Additionally, the ultimate outcome predictions will not be guaranteed.

To avoid these biases, all available firm-quarters or firm-years for the sample period have been used in recent studies to construct the default prediction models. Hence, these models improve the accuracy of the coefficient estimates and increase the prediction power of the models compare to prior studies (Brockman and Turtle, 2003; Bharath and Shumway, 2004; Hillegeist et al., 2004; Gharghori et al. 2006; Reisz and Perlich, 2007; Agarwal and Taffler, 2008, Tserng et al., 2011, Tsai et al., 2011, Tserng et al., 2012). Therefore, this thesis was also used all available firm-years data to develop the default prediction model.

After putting in all firm-years data, a new problem generated, that is, there was a huge discrepancy in sample size between defaulted and non-defaulted firms. It means that the number of non-defaulted samples greatly surpassed that of defaulted samples, which is referred to as between-class imbalance issue (He and Garcia, 2009). It might lead the model discriminate inaccurately healthy and non-healthy group (Chang, 2007).

In order to solve the imbalance problem, I used two over-sampling techniques in this study: replication (Japkowicz, 2000; Chawla et al., 2008) and

Synthetic Minority Over-sampling Technique (SMOTE) developed by Chawla et al. (2002)

3.2.2 Replication

To tackle imbalance problem, He and Garcia (2009) put forward the idea that important information should be priority. The simple method is replication. In this research, great emphasis is put on the default samples. In other words, the default samples will be replicated several times.

In the experiment of my study, default samples in original training set were multiplied until they equal to non-default samples. It takes about 35 times replication in this research. However, to examine the models and see how good it is, I would like to replicate 1-70 times of default samples.

3.6.3 Synthetic Minority Over-sampling Technique (SMOTE)

Another way to increase number of default samples is Synthetic Minority Over-sampling Technique (SMOTE) method that was suggested by Chawla et al. (2002). The scholars proposed an over-sampling approach in which the minority class is over-sampled by creating "synthetic" examples rather than by over-sampling with replication. This approach is inspired by a technique that proved successful in handwritten character recognition.

Firstly, the minority class is over-sampled by taking each minority class sample and introducing synthetic examples along the line segments joining any/all of the k minority class nearest neighbors. Secondly, depending upon the amount of over-sampling required, neighbors from the k nearest neighbors are randomly chosen. For example, if we use nine nearest neighbors and the amount of over-sampling needed is 200%. There are only two neighbors from the nine nearest neighbors are chosen and one sample is generated in the direction of each. In this study, k is 25 for all group.

The synthetic samples are generated in following way: take the difference between the original minority sample and its nearest neighbor. Then, multiply this difference by a random number between 0 and 1, and add it to the original sample. This causes the selection of a random point along the line segment between two specific features. In this study, the minority class in the training set was over sampled at 100%, 200% until 5400% of its original size. The algorithm will be described in Figure 3.3 as follows:

Algorithm SMOTE (T; N; k)

Input: Number of minority class samples T ; Amount of SMOTE N%; Number of nearest neighbors k

Output: (N/100) * T synthetic minority class samples

1. (∗ If N is less than 100%, randomize the minority class samples as only a random percent of them will be SMOTEd. ∗)

2. if N < 100

3. then Randomize the T minority class samples

4.T = (N/100) ∗ T

5.N = 100

6. End if

7. N = (int)(N/100) (∗ The amount of SMOTE is assumed to be in integral multiples of100. ∗)

8. k = Number of nearest neighbors

9. numattrs = Number of attributes

10. Sample[ ][ ]: array for original minority class samples

11. newindex: keeps a count of number of synthetic samples generated, initialized to 0

12. Synthetic[ ][ ]: array for synthetic samples(∗ Compute k nearest neighbors for each minority class sample only. ∗)

13. for i ← 1 to T

14. Compute k nearest neighbors for i, and save the indices in the nnarray

15. Populate(N, i, nnarray)

16. End for Populate(N, i, nnarray) (∗ Function to generate the synthetic samples. ∗)

17. While N = 0

18. Choose a random number between 1 and k, call it nn. This step chooses one of the k nearest neighbors of i.

19. for attr ← 1 to numattrs

20. Compute: dif = Sample[nnarray[nn]][attr] − Sample[i][attr]

21. Compute: gap = random number between 0 and 1

22. Synthetic[newindex][attr] = Sample[i][attr] + gap ∗ dif

23. endfor

24. newindex++

25.N = N − 1

26. endwhile

27. return (∗ End of Populate. ∗)

Figure 3.3 Synthetic Minority Over-sampling Technique

Chawla et al. (2002) argued that the outcome of SMOTE was better than the outcome of replication. In this research, I would like apply those methods in construction prediction model and compare the performance of each method.

3.3 ROC Curve

3.3.1 Concept and methodology of ROC curve

Assessment of predictive accuracy is an important aspect of evaluating and comparing models, algorithms or technologies that produce the predictions. Receiver operating characteristic (ROC) curves are common, widely applicable method which useful for assessing the accuracy of tests because they provide a comprehensive and visually attractive way to summarize the accuracy of predictions. So, in this research's scope, the author proposes applying ROC curves to assess and compare the accuracy rate of default probability predictions which were applied by grey system analysis.

ROC curves, generalize contingency table analysis by providing information on the performance of a model at any cut-off that might be chosen (Green and Swets, 1966; Hanley,1989; Pepe, 2002; Swets, 1988; Swets, 1996). In the simplest case, the model produces only two ratings (Bad/Good) which are shown along with the actual outcomes (default/no default) in tabular form. The cells in the table indicate the number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN), respectively. FN represents a Type I error and FP represents Type II error. These fractions are presented in table 3.

- TP: a predicted default that actually occurs,
- TN: a predicted non-default that actually occurs
- FP: a predicted default that does not occur and,
- FN: is a predicted non-default where the company actually defaults.

Table 3-1: Types of error of ROC

Test	Default Probability				Total
	Present	n	Absent	N	
Positive	True Positive(TP)	a	False Positive(FP)	C	a + c
Negative	False Negative(NP)	b	True Negative(TN)	D	b + d
Total		a + b		c + d	

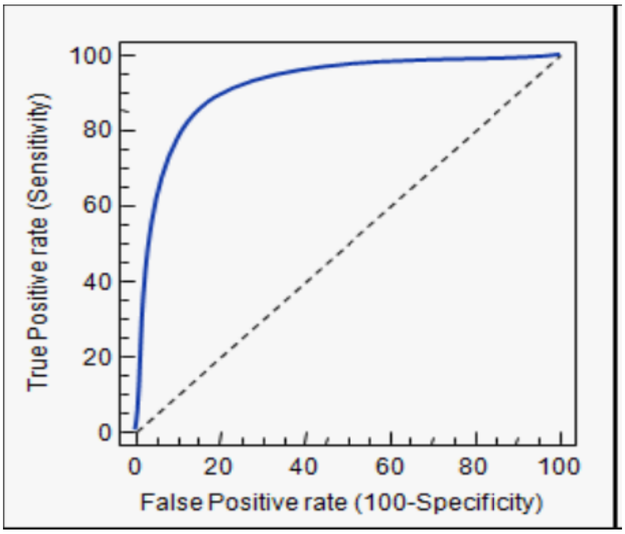


Figure 3.1: An example of ROC curve *(Zweig & Campbell, 1993)*

The figure3.1 shows how all four quantities of a contingency table can be identified on a ROC curve. The true positive rate (Sensitivity) is plotted in function of the false positive rate (100-Specificity) for different cut-off points. “Each point on the ROC curve represents a sensitivity/specificity pair corresponding to a particular decision threshold. A test with perfect discrimination (no overlap in the two distributions) has a ROC curve that passes through the upper left corner (100% sensitivity, 100% specificity). Therefore, the closer the ROC curve is to the upper left corner, the higher the overall accuracy of the test is” (Zweig & Campbell, 1993).

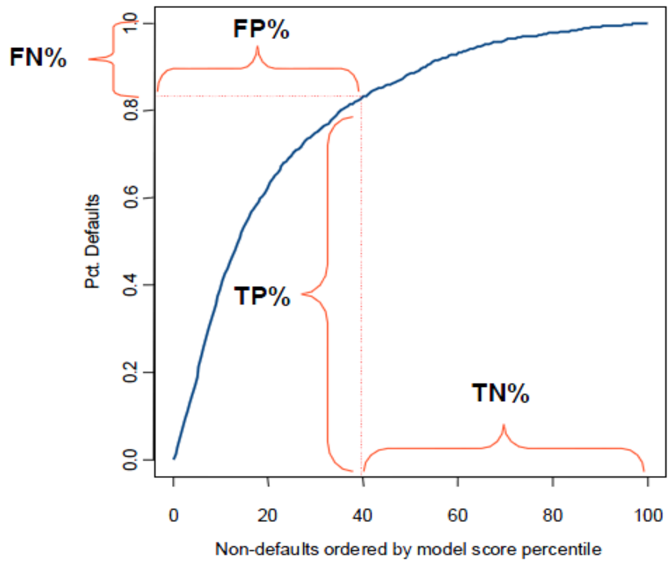


Figure 3.2: Schematic of a ROC

3.3.2 Utilizing ROC curve to validate the model

One useful characteristic of the ROC curve is the area under the curve (AUC) clearly reflects how good the test is at distinguishing between firms with disease and those without disease. The AUC serves as a single measure, independent of prevalence that summarizes the discriminative ability of a prediction across the full range of cut-off points. The greater value of the AUC, the better the prediction is. A perfect discrimination test will have an AUC of 1.0, while a completely useless test (one whose curve falls on the diagonal line) has an AUC of 0.5. A test with an area greater than 0.9 has high accuracy, while 0.7–0.9 indicates moderate accuracy and 0.5–0.7 indicates low accuracy.

Corresponding to different circumstances of period of time (4 year data consequence) as well as X_0 (sequence of characteristic data to describe the system's behavior), the value of AUC will be calculated. And this value is the basis for evaluate and compare the accuracy of the Grey System Theory application with the different number of variables.

3.3 Summary

In this chapter, the author introduces the main methodology adopted in the research – grey system analysis and the way to apply this method to forecast a company default probability. By calculating synthetic degree incidence of considered firms and combine these values, the default probability of firms were identified. Beside, Over-sampling technique is used before applying Grey Theory, to address imbalance data problem. After different models are calculated, ROC curves are used to compare with the previous study.

CHAPTER 4: DATA COLLECTION

4.1 Data collection

4.1.1 Source and validity of data

Data in this research was gathered from COMPUSTAT Industrial File (Wharton Research Data Services) as well as the Center for Research in Securities Prices (CRSP) for construction companies of the U.S. My research concentrated on construction contractors with December fiscal year-ends by choosing firms with SIC codes between 1,500 and 1,799. Similar to the researches of Severson et al (1993) and Russell and Zhai (1996), the sample contractors include three construction categories:

- ***Major Group 15***: Building construction, general contractors and operative builders. The construction of buildings subsection includes establishments involved in constructing residential, industrial, commercial, and institutional buildings
- ***Major Group 16***: Heavy construction other than building construction contractors. The heavy and civil engineering subsection includes establishments involved in infrastructure projects.
- ***Major Group 17***: Construction special trade contractors. The specialty trade contractors engaged in activities such as plumbing, electrical work, masonry, carpentry, and roofing that are generally needed in the construction of all building types.

4.1.2 Principles of collecting data

The selection of firm is confined in construction industry only. 92 companies were selected as participant of the research, among which 24 were defaulted. The observed period was 1972-2008. According to Chin (2009), Tserng et al. (2008), Tserng et al. (2009), there are two main criteria in data collection principle to select samples:

1. Companies which do not have financial statement for at least 5 years will be taken out of the sample.
2. Default firms are defined by CRSP delisting code of 400 and 550 to 585, which correspond to the delisting reason concerned with company failures such as bankruptcy, liquidation of poor performance.

The chosen firms must have at least five years' data in Compustat Industrial File to ensure that all the unhealthy firms will be excluded in the population of study as well as to consider the impact of market factors to these companies in a long term.

4.1.3 Summary of the input data

This thesis have a total of 24 failed companies among 92 construction companies which were identified during the year of determination. These firm were chosen because they are suitable to two criteria above. The number of firm may be 50 but 92 firm is big enough to improve the exact of study. Besides, to find out exact number of firm, it is beyond the thesis' limit.

Table 4.1 disclosed the name of failed firms.

Table 4-1: Information of the defaulted companies

ORD	CODE	COMPANY'S NAME	DEFAULTED YEAR	OBSERVED FIRM-YEARS
1	60409	AMERICAN MEDICAL BLDGS INC	1989	1978 -1989
2	85607	ATKINSON (G F) CO/CA	1997	1985 - 1997
3	63095	BANK BUILDING &EQUIP CORP AM	1989	1972 - 1989
4	11901	ENTRX CORP	2004	1988 - 2004
5	86933	COMSTOCK GROUP INC	1988	1984 - 1988
6	22382	CERBCO INC -CL A	2000	1981 - 2000
7	55079	MORRISON KNUDSEN CORP OLD	1995	1972 - 1995
8	58641	CANISCO RESOURCES INC	1998	1982 - 1998
9	10036	NEUROTECH DEVELOPMENT CORP	1990	1986 - 1990
10	29621	DEVCON INTERNATIONAL CORP	2007	1987 - 2007
11	11109	CEC INDUSTRIES CORP	1994	1987 - 1994
12	80220	ABLE TELCOM HOLDING CORP	1999	1994 - 1999
13	76432	RYAN MURPHY INC	1994	1990 - 1994
14	76796	BUILDING MATERIALS HLDG CP	2007	1991 - 2007
15	77334	SHOLODGE INC	2004	1992 - 2004
16	77831	XXSYS TECHNOLOGIES INC	1998	1992 - 1998
17	79017	TRANSCOR WASTE SERVICES INC	1997	1993 - 1997
18	10227	KIMMINS CORP	1998	1986 - 1998
19	79815	COFLEXIP SA	2000	1993 - 2000
20	79958	DAW TECHNOLOGIES INC	2000	1993 - 2000
21	82829	NESCO INC	2000	1996 - 2000
22	82731	CHINA CONVERGENT CORP LTD	2000	1996 - 2000
23	85606	ENCOMPASS SERVICES CORP	2001	1997 - 2001

24	88642	DISTRIBUTED ENERGY SYS CORP	2007	2003 - 2007

4.2 Data classification

4.1.4 4.2.1 Collection of Financial ratios data

Theodossiou (1991) claimed that the selection of the independent variables for a bankruptcy prediction model is the most toughing aspect of every bankruptcy because financial theory does not indicate which variables should be included in the. The forward stepwise statistical procedure has been recognized as the most popular method used in previous studies for the development of bankruptcy prediction models. Due to some specific properties of construction finance, this research's financial ratios are collected following prior researches (Mason and Harris(1979) ; Abidali (1990); Russel and Jaselskis (1992); Cheng, J. et al (2009); Delcea, C. &Scarlat, E) which concerned to the prediction of the probability of construction firms. Besides, the selected financial ratios must involve all the aspects of a contractor finance situation and has to include the liquidity, profitability, leverage, activity of a firm and even refer to the market factor. The last principle to select financial ratios is all of these ratios must have a predicted relationship with the default risk.

4.2.2 Clacification of selected financial ratios

19 single financial ratios developed from financial data from 92 construction firms across a 37-year period (1972-2008) were taken into account. These ratios are classified into 4 categories of ratios (liquidity, leverage, profitability, activity) which are typically used in analyzing financial position:

Table 4-2: Selected ratios' classification

1. Liquidity Ratios		
No.	***Symbol***	***Ratio***
1	VAR1	Current Ratio
2	VAR2	Quick Ratio
3	VAR3	Net Working Capital to Total Assets
4	VAR4	Current Assets to Net Assets

2. Leverage Ratios		
No.	***Symbol***	***Ratio***
5	VAR5	Total Liabilities to Net Worth
6	VAR6	Retained Earnings to Sales
7	VAR7	Debt Ratio
8	VAR8	Times Interest Earned

3. Activity Ratios		
No.	***Symbol***	***Ratio***
9	VAR9	Revenues to Net Working Capital
10	VAR10	Accounts Receivable Turnover
11	VAR11	Accounts Payable Turnover
12	VAR12	Sales to Net Worth
13	VAR13	Quality of Inventory
14	VAR14	Turnover of Total Assets
15	VAR15	Revenues to Fixed Assets

4. Profitability Ratios

No.	Symbol	Ratio
16	VAR16	ROA
17	VAR17	ROE
18	VAR18	ROS
19	VAR19	Profits to Net Working Capital

4.3 Financial ratios' definition

The definition and the sign of 19 major represented variables in table 4-3 below:

Table 4-3: Definition and usage ratios

Var.	Ratio	Definition	Usage	Sign
1	Current Ratio	Current assets/ Current liabilities	A liquidity ratio that measures a company's ability to pay short-term obligations.	+
2	Quick Ratio	(Current Assets - Inventory)/ Current liabilities	An indicator of a company's short-term liquidity measures a company's ability to meet its short-term obligations with its most liquid assets	+
3	Net Working Capital to Total Assets	(Current assets - Current liabilities)/Total assets	Measures both a company's efficiency and its short - term financial health	+
4	Current Assets to Net Assets	Current assets/(Total assets -Current liabilities)	Indicates how effectively a company is using its assets to generate cash before contractual obligations must be paid	+
5	Total Liabilities to Net Worth	Total liabilities / Net worth	Indicates the extent to which a company is utilizing its re-investment	-
6	Retained Earnings to Sales	Retained earnings/ Net Sales	Indicates how effectively reinvested into the company.	+
7	Debt Ratio	Total liabilities / Total assets	Indicates what proportion of debt a company has relative to its assets.	-
8	Times Interest Earned	Earnings before interest and taxes / Interest charges	Measures a company's ability to honor its debt payments.	+

9	Revenues to Net Working Capital	Net Sales/ (Average Current Assets - Average Current Liabilities)	Measures a company's ability to honor its debt payment	+
10	Accounts Receivable Turnover	Annual credit sales / Average account receivable	Measures the number of times that accounts receivable amount is collected throughout the year.	+
11	Accounts Payable Turnover	Net Sales/ Average accounts payable	A short-term liquidity measure used to quantify the rate at which a company pays off its suppliers.	-
12	Sales to Net Worth	Net sales / average net worth	Measures the number of times working capital turns over annually in relation to net sales.	+
13	Quality of Inventory	Revenue / Inventory	Show intensity with which the firm uses assets in generating product. The inventory quality ratio is the ratio of the active inventory dollars to total inventory dollars	+
14	Turnover of Total Assets	Net sales / average total assets	Simply compares the turnover with the assets that the business has used to generate that turnover.	+
15	Revenues to Fixed Assets	Net sales / Average fixed assets	Measures a company's ability to generate net sales from fixed-asset investments.	+
16	ROA	Profit before interest and taxes / Total assets	An indicator of how profitable a company is relative to its total assets. ROA gives an idea as to how efficient management is at using its assets to generate earnings.	+
17	ROE	Profit before interest and taxes / Equity	Measures a corporation's profitability by revealing how much profit a company generates with the money shareholders have invested. Indicates performance and potential for growth.	+
18	ROS	Profit before interest and taxes / Total sales	Evaluates a company's operational efficiency.	+
19	Profits to Net Working Capital	Net profit after interest and tax /(Current assets - current liabilities)	Evaluates the efficiency of a company's investment.	+

Table 4.3 illustrates 19 chosen financial ratios which are relatively classified in to 4 groups: liquidity, leverage, activity, profitability. The table also depicts the expected dependence between the accounting ratio and the default probability, in which symbol (+) means the bigger the ratio's value, the healthier financial statement of company (a decrease in the default probability) and symbol (-) signifies an increase in the default probability given a decrease in the explanatory variable.

The first group includes four financial ratios - Current ratio (VAR1), Quick ratio (VAR2), Net working capital to total assets (VAR3) and Current assets to net assets (VAR4). These ratios were used as variables for liquidity. Liquidity measures a firm's ability to meet its short-term obligations. Moyer and Chatfield (1983) propose a negative effect of liquidity on bankruptcy because high liquidity implies a low level of short-term obligations and low default risk. Therefore, in this research liquidity group symbolize as positive symptom – as the firm is solvent, its financial situation is better.

Leverage ratios, as categorized as the second group, measure the extent to which a company has been financed by debt and shareholder funds. This kind of ratio reflects the cooperative's ability to meet both short-term and long-term debt obligations. Leverage ratios are computed either by comparing earnings from the income statement to interest payments or by relating the debt and equity items from the balance sheet. In this category, Total liabilities to net worth ratio (VAR 5) is known as the higher this ratio, the less protection there is for creditors. If total liabilities exceed net worth then creditors have more at stake than stockbrokers. The debt ratio (VAR7) can help investors determine a company's level of risk. A debt ratio of greater than 1 indicates that a company has more debt than assets. VAR 5 and VAR 7 present to negative symptom – as the higher value they are, the worse financial situation of the firm is. In this group, two remained ratios are Retained earnings to sales (VAR6) and Times interest earned (VAR8) are considered as positive symptom because retained earnings are typically reinvested into the company, so, the higher value it is, the better. VAR8 is a great tool when measuring a company's ability to meet its debt obligations. When the interest coverage ratio is smaller than 1, the company is not generating enough cash from its operations EBIT to meet its interest obligations, and it is a warning sign when interest coverage falls below 2.5x.

The third group, profitability ratios: measure the overall performance, or returns, which management has been able to achieve. Profit is a crucial goal of a firm, so poor performance of a firm signals an imminent collapse. In this profitability group, Accounts payable turnover (VAR 11) shows investors how many times per period the company pays its average payable amount. If the turnover ratio is decreasing from one period to another, this is a sign that the company is taking longer time to pay off its suppliers than it was before. The opposite is true when the turnover ratio is rising, which means that the company is paying of suppliers at a faster rate. A higher value of the ratio implies a greater exposure to financial risk, therefore, VAR 11 is considered as negative symptom. Other ratios in this group stand for positive symptoms; for example: A high accounts receivable turnover ratio (VAR10) indicates a tight credit policy, meanwhile, low or declining accounts receivable turnover ratio indicates a collection problem, part of which may be due to bad debts. A higher fixed-asset turnover ratio (VAR15) shows that the company has been more effective in using the investment in fixed assets to generate revenues.

The last category, activity ratios: measure the intensity with which the firm uses assets in generating sales and show how well a company has been using its resources. These ratios indicate whether the firm's investment in current and long-term assets is too large, too small, or just right. If too large, funds may be tied up in assets that could be used more productively. There are two basic approaches to the computation of activity ratios: the first looks at the average performance of the firm over the year and the second uses year-end balances in the calculations. As can be seen in the table 4.2, four selected ratios represent activity ratios are ROA, ROE, ROS and Profits to net working capital. Improvement in the ROE ratio (VAR17) implies improved marketing, improved productivity, or improvement in both, and thus requires further investigation. Improvement in the ROA ratio (VAR16) implies a strengthening of marketing effectiveness. Whereas, ROS (VAR 18) is helpful to management, providing insight into how much profit is being produced per dollar of sales. An increasing ROS indicates the company is developing more efficient, while a decreasing ROS could signal looming financial troubles. Four represented ratios of profitability category are recognized as positive symptom.

4.4 Data analysis process

In order to add time series information to the model of business failure prediction, Visual Basic Application (VBA) embedded in Microsoft EXCEL 2003 was utilized to build model of grey analysis prediction. Prediction value of the four year and history data were computed and plotted in to the ROC curve. After that, by comparing the area under ROC curve with the area under ROC of the previous research which used the same history data, some comparisons, conclusions and announcements are pointed out respectively based on the work results.

4.5 Summary

This chapter illustrated the input data collection process. Firstly, the financial statements and the actual default or non- default situation of 92 construction companies are collected according to Standard Industry Classification. Secondly, the 19 selected financial ratios are taken into account based on some special principles and are divided into four typical groups rely on its characteristics. The procedure of data analysis is also depicted to demonstrate clearly all key steps.

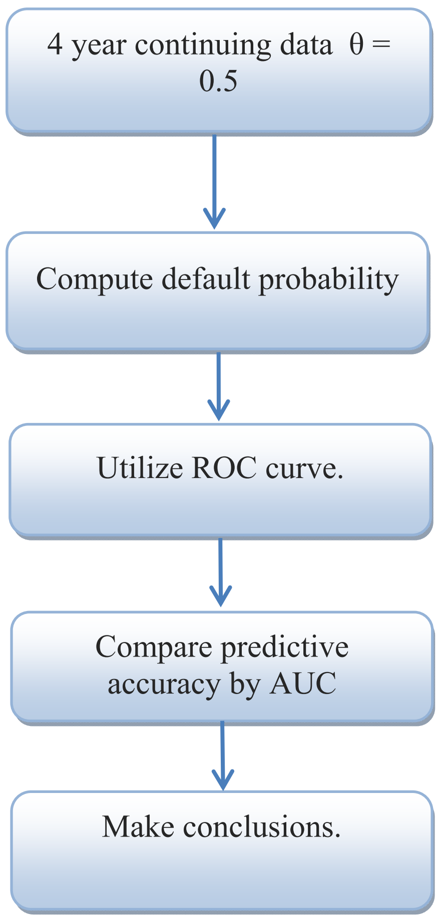


Figure 4.1: The algorithm chart of the data analysis process

CHAPTER 5: DATA ANALYSIS AND RESULTS

5.1 Data analysis

5.1.1 Example analysis

As mentioned in the chapter 3- Methodology, the analysis was conducted at a single firm level then the same proceed with each of the consider firms will be taken into analysis. In order to better understand the proposed model, the writer develops a numerical example below with an assumption that analyze a set of F = 15 firms, for five years. In the analysis, the existence of nineteen symptoms as it follows:

Table 5.1: Selected variables and their default probability correlation

No.	Symbol	Ratio	Sign
1	VAR 1	Current Ratio	+
2	VAR 2	Quick Ratio	+
3	VAR 3	Net Working Capital to Total Assets	+
4	VAR 4	Current Assets to Net Assets	+
5	VAR 5	Total Liabilities to Net Worth	-
6	VAR 6	Retained Earnings to Sales	+
7	VAR 7	Debt Ratio	-
8	VAR 8	Times Interest Earned	+
9	VAR 9	Revenues to Net Working Capital	+
10	VAR 10	Accounts Receivable Turnover	+
11	VAR 11	Accounts Payable Turnover	-
12	VAR 12	Sales to Net Worth	+

13	VAR 13	Quality of Inventory	+
14	VAR 14	Turnover of Total Assets	+
15	VAR 15	Revenues to Fixed Assets	+
16	VAR 16	ROA	+
17	VAR 17	ROE	+
18	VAR 18	ROS	+
19	VAR 19	Profits to Net Working Capital	+

5 year history data consequence of one sample firm are arrange into the table 5.2. The data of another 14 companies are show later in the list of figure.

Table 5.2: 5 year history data of firm No.1

Firm No.1 (ACMTA)		Year I	Year II	Year III	Year IV	Year V
Actual Year		1972	1973	1974	1975	1976
VAR1	X1	3.4330	3.5270	1.4545	2.5931	4.1683
VAR2	X2	2.8325	2.9562	1.2634	2.3115	3.4158
VAR3	X3	0.6487	0.5266	0.2332	0.3637	0.3850
VAR4	X4	1.2481	0.9285	1.5324	0.7672	0.5765
VAR5	X5	0.3963	1.0242	3.5241	1.6003	0.9285
VAR6	X6	0.2081	0.2121	0.0398	0.0607	0.1558
VAR7	X7	0.2838	0.5060	0.7790	0.6154	0.4815
VAR8	X8	37.7037	13.8194	-9.5519	1.1197	1.4519
VAR9	X9	3.6344	3.4677	3.5581	4.1834	2.1663
VAR10	X10	4.9930	3.3645	2.2740	2.0995	1.9437
VAR11	X11	26.4538	17.4515	11.1902	9.9414	10.1431
VAR12	X12	3.4485	3.4444	3.7805	4.1834	1.8246
VAR13	X13	16.7764	11.9245	14.4978	10.9723	6.8751
VAR14	X14	2.5708	1.9802	1.3505	1.1765	0.8079
VAR15	X15	40.3190	9.9322	5.2065	3.7887	1.8149
VAR16	X16	0.1528	0.0954	-0.2322	0.0503	0.0517
VAR17	X17	0.2030	0.1702	-1.2452	0.0080	0.0192
VAR18	X18	0.0704	0.0540	-0.2042	0.0019	0.0107
VAR19	X19	0.2242	0.1597	-1.1804	0.0085	0.0259

1. *Absolute Degree of Grey Incidence*

The sequence to compute the absolute matrix of incidence as the consequence X0 = X1 for firm 1 as follow:

- Step 1: Compute image of zeroing starting point:

$X_j = (x_{1j}- x_{1j}, x_{2j} - x_{1j}, x_{3j} - x_{1j}, x_{4j} - x_{1j}, x_{5j} - x_{1j})$ (5.1)

Whereas j = 1, ... , n. n = number of symptom. In this example, n = 19.

For example:

X1 = {x(1); x(2); x(3); x(4); x(5)}

= (3.4330; 3.5270; 1.4545; 2.593; 4.1683)

$X_1 = (x_{1,1}- x_{1,1}; x_{2,1} - x_{1,1}; x_{3,1} - x_{1,1};x_{4,1} - x_{1,1}; x_{5,1} - x_{1,1})$

= (0; 0.094; -1.978; -0.840; 0.735)

X2 = {x(1); x(2); x(3); x(4); x(5)}

= (2.8325; 2.9562; 1.2634; 2.3115; 3.4158)

$X_2 = (x_{1,2}- x_{1,2}; x_{2,2} - x_{1,2}; x_{3,2} - x_{1,2}; x_{4,2} - x_{1,2}; x_{5,2} - x_{1,2})$

= (0; 0.124; -1.569; -0.521; 0.583; -1.675)

This process will be continued until X19.

- Step 2: Find $|s_0|$, $|s_j|$, and $|s_j - s_0|$

For X0 (= X1):

$$|s_0| = |s_1| = \left|\sum_{k=2}^{4} x_0^0\,(k) + \tfrac{1}{2}x_0^0(5)\right| \qquad (5.2)$$

$$= \left|(0.094 + (-1.978) + (-0.840) + \tfrac{1}{2}(0.735)\right| = 2.357$$

For *j = 2, 3, …, 19.* Take j = 2 as an example, the same formula will be accounted for other value of j.

$$|s_2| = \left|\sum_{k=2}^{4} x_2^0\,(k) + \tfrac{1}{2}x_2^0(5)\right| \qquad (5.3)$$

$$= \left|0,124 + (-1569) + (-0..521) + \frac{1}{2}(0.583)\right| = 1.675$$

$$|s_2 - s_0| = \left|\sum_{k=2}^{4}[x_1^0(k) - x_0^0(k)] + \frac{1}{2}[x_1^o(5) - x_0^0(5)]\right| \quad (5.4)$$

$$= \left|\sum[(0.124 - 0.094) + ((-1.569) - (-1.978)) + ((-0,521) - (-0.840))] + \frac{1}{2}(0.583 - 0.735)\right| = 0.682$$

- Step 3: Attain the absolute degree of grey incidences:

$$\varepsilon_{0i} = \frac{1+|s_0|+|s_i|}{1+|s_0|+|s_i|+|s_i - s_0|} \quad (5.5)$$

$$\varepsilon_{01} = \frac{1+|s_0|+|s_1|}{1+|s_0|+|s_1|+|s_1 - s_0|} = 1$$

$$\varepsilon_{02} = \frac{1+|s_0|+|s_2|}{1+|s_0|+|s_2|+|s_2 - s_0|} = \frac{1+2.357+1.675}{1+2.357+1.675+|1.675-2.357|}$$

$$= 0.881$$

Table 5.3: The absolute ε_j value of firm No.1

Firm 1 (ACMTA)	Year I	Year II	Year III	Year IV	Year V	lSjl	$|S_j - S_0|$	ε_{1j}
X_1^0	0	0.094	-1.978	-0.840	0.735	2.357	0.000	1.000
X_2^0	0	0.124	-1.569	-0.521	0.583	1.675	0.682	0.881
X_3^0	0	-0.122	-0.416	-0.285	-0.264	0.955	1.402	0.755
X_4^0	0	-0.320	0.284	-0.481	-0.672	0.852	1.505	0.737
X_5^0	0	0.628	3.128	1.204	0.532	5.226	7.583	0.531
X_6^0	0	0.004	-0.168	-0.147	-0.052	0.338	2.019	0.647
X_7^0	0	0.222	0.495	0.332	0.198	1.148	3.505	0.562
X_8^0	0	-23.884	-47.256	-36.584	-36.252	125.850	123.493	0.511
X_9^0	0	-0.167	-0.076	0.549	-1.468	0.428	1.929	0.662
X_{10}^0	0	-1.629	-2.719	-2.894	-3.049	8.766	6.409	0.654
X_{11}^0	0	-9.002	-15.264	-16.512	-16.311	48.934	46.577	0.529
X_{12}^0	0	-0.004	0.332	0.735	-1.624	0.251	2.608	0.580
X_{13}^0	0	-4.852	-2.279	-5.804	-9.901	17.885	15.529	0.578
X_{14}^0	0	-0.591	-1.220	-1.394	-1.763	4.087	1.730	0.811
X_{15}^0	0	-30.387	-35.112	-36.530	-38.504	121.282	118.925	0.512
X_{16}^0	0	-0.057	-0.385	-0.103	-0.101	0.596	1.761	0.692

X_{17}^{0}	0	-0.033	-1.448	-0.195	-0.184	1.768	0.589	0.897
X_{18}^{0}	0	-0.016	-0.275	-0.068	-0.060	0.389	1.968	0.656
X_{19}^{0}	0	-0.064	-1.405	-0.216	-0.198	1.784	0.573	0.900

2. Relative Degree of Grey Incidence

The relative degree of grey incidence is obtained using the following relations:

- Step 1: Compute the initial images of X_0 and X_j

$$X_0' = \left(\frac{x_0\,(1)}{x_0\,(1)}, \frac{x_0\,(2)}{x_0\,(1)}, \frac{x_0\,(3)}{x_0\,(1)}, \frac{x_0\,(4)}{x_0\,(1)}, \frac{x_0\,(5)}{x_0\,(1)}\right) \qquad (5.6)$$

$$X_0' = X_1' = \left(\frac{3.4330}{3.4330}, \frac{3.5270}{3.4330}, \frac{1.4545}{3.4330}, \frac{2.593}{3.4330}, \frac{4.1683}{3.4330}\right)$$

= (3.4330; 3.5270; 1.4545; 2.593; 4.1683)

$$X_j' = \left(\frac{x_j\,(1)}{x_j\,(1)}, \frac{x_j\,(2)}{x_j\,(1)}, \frac{x_j\,(3)}{x_j\,(1)}, \frac{x_j\,(4)}{x_j\,(1)}, \frac{x_j\,(5)}{x_j\,(1)}\right) j = [2,19]. \qquad (5.7)$$

Take j = 2 for example:X2 = (2.8325; 2.9562; 1.2634; 2.3115; 3.4158)

$$X_2' = \left(\frac{x_2(1)}{x_2\,(1)}, \frac{x_2\,(2)}{x_2(1)}, \frac{x_2\,(3)}{x_2(1)}, \frac{x_2\,(4)}{x_2\,(1)}, \frac{x_2(5)}{x_2\,(1)}\right)$$

$$= \left(\frac{2.8325}{2.8325}, \frac{2.9562}{2.8325}, \frac{1.2634}{2.8325}, \frac{2.3115}{2.8325}, \frac{1.4158}{2.8325}\right)$$

= (1; 1.044; 0.446; 0.816; 1.206)

- Step 2: Compute the images of zero starting points of X_0' and X_j'

$$X_0'^0 = \left(x_0'^0(1) - x_0'^0(1), x_0'^0(2) - x_0'^0(1), \dots, x_0'^0(5) - x_0'^0(1)\right) \qquad (5.8)$$

$$X_0'^0 = \{(1-1); (1.027-1); (0.424-1); (0.755-1); (1.214-1)\}$$

= (0; 0.027; - 0.576; -0.245; 0.214)

$$X_j'^0 = \left(x_j'^0(1) - x_j'^0(1), x_j'^0(2) - x_j'^0(1), \dots, x_j'^0(5) - x_j'^0(1)\right) \qquad (5.9)$$

$$X_2'^0 = \left(x_2'^0(1) - x_2'^0(1), x_2'^0(2) - x_2'^0(1), \dots, x_2'^0(5) - x_2'^0(1)\right)$$

$$X_2'^0 = \{(1-1); (1.044-1); (0.446-1); (0.816-1); (1.206-1)\}$$

$= (0; 0.044; -0.554; -0.184; 0.206)$

- Step 3: Compute $|s_0'|$, $|s_j'|$ and $|s_j' - s_0'|$

$$|s_0'| = \left|\sum_{k=2}^{4} x_0'^0(k) + \frac{1}{2} x_0'^0(5)\right| \quad (5.10)$$

$$|s_j'| = \left|\sum_{k=2}^{4} x_j'^0(k) + \frac{1}{2} x_j'^0(5)\right| \quad (5.11)$$

$$|s_j' - s_0'| = \left|\sum_{k=2}^{4}(x_j'^0(k) - x_0'^0(k)) + \frac{1}{2}(x_j'^0(5) - x_0'^0(5))\right| \quad (5.12)$$

- $|s_0'| = \left|(0.027) + (-0.576) + (-0.245) + \frac{1}{2}0.214\right| = 0.6865$
- $|s_2'| = \left|(0.044) + (-0.554) + (-0.184) + \frac{1}{2}0.206\right| = 0.5915$
- $|s_2' - s_0'| = \left|(0.044 - 0.027) + ((-0.554) - (-0.576)) + (0.184 - (-0.245)) + \frac{1}{2}(0.206 - 0.214)\right| = 0.095$

Table 5.4: The initial images value of firm No.1

Firm No.1 (ACMTA)	X1'	X2'	X3'	X4'	X5'
$X_1^{0'}$	1	1.027	0.424	0.755	1.214
$X_2^{0'}$	1	1.044	0.446	0.816	1.206
$X_3^{0'}$	1	0.812	0.359	0.561	0.593
$X_4^{0'}$	1	0.744	1.228	0.615	0.462
$X_5^{0'}$	1	2.585	8.893	4.038	2.343
$X_6^{0'}$	1	1.020	0.191	0.292	0.749
$X_7^{0'}$	1	1.783	2.745	2.168	1.696
$X_8^{0'}$	1	0.367	-0.253	0.030	0.039
$X_9^{0'}$	1	0.954	0.979	1.151	0.596
$X_{10}^{0'}$	1	0.674	0.455	0.420	0.389
$X_{11}^{0'}$	1	0.660	0.423	0.376	0.383
$X_{12}^{0'}$	1	0.999	1.096	1.213	0.529
$X_{13}^{0'}$	1	0.711	0.864	0.654	0.410
$X_{14}^{0'}$	1	0.770	0.525	0.458	0.314
$X_{15}^{0'}$	1	0.246	0.129	0.094	0.045
$X_{16}^{0'}$	1	0.624	-1.520	0.329	0.339
$X_{17}^{0'}$	1	0.838	-6.133	0.040	0.095
$X_{18}^{0'}$	1	0.768	-2.903	0.028	0.152
$X_{19}^{0'}$	1	0.712	-5.266	0.038	0.115

- Step 4: Compute the relative degree of incidence

$$r_{0j} = \frac{1+|s'_0|+|s'_j|}{1+|s'_0|+|s'_j|+|s'_j - s'_0|} \qquad (5.13)$$

- $r_{01} = \frac{1+|s'_0|+|s'_1|}{1+|s'_0|+|s'_1|+|s'_1 - s'_0|} = 1$
- $r_{02} = \frac{1+|s'_0|+|s'_2|}{1+|s'_0|+|s'_2|+|s'_2 - s'_0|} = \frac{1+0.6865+0.5915}{1+0.6865+0.5915+0.095} = 0.8001$

Table 5.5: The relative r_j value of firm No.1

Firm No.1	$X_1' - X_1'$	$X_2' - X_1'$	$X_3' - X_1'$	$X_4' - X_1'$	$X_5' - X_1'$	$\|S_j\|$	$\|S_j - S_1\|$	r_{1j}
$X_1^{0'}$	0	0.027	-0.576	-0.245	0.214	0.686	0.000	1.000
$X_2^{0'}$	0	0.044	-0.554	-0.184	0.206	0.591	0.095	0.960
$X_3^{0'}$	0	-0.188	-0.641	-0.439	-0.407	1.471	0.785	0.801
$X_4^{0'}$	0	-0.256	0.228	-0.385	-0.538	0.683	0.004	0.998
$X_5^{0'}$	0	1.585	7.893	3.038	1.343	13.188	13.874	0.517
$X_6^{0'}$	0	0.020	-0.809	-0.708	-0.251	1.623	0.937	0.779
$X_7^{0'}$	0	0.783	1.745	1.168	0.696	4.044	4.731	0.548
$X_8^{0'}$	0	-0.633	-1.253	-0.970	-0.961	3.338	2.651	0.655
$X_9^{0'}$	0	-0.046	-0.021	0.151	-0.404	0.118	0.569	0.760
$X_{10}^{0'}$	0	-0.326	-0.545	-0.580	-0.611	1.756	1.069	0.763
$X_{11}^{0'}$	0	-0.340	-0.577	-0.624	-0.617	1.850	1.163	0.752
$X_{12}^{0'}$	0	-0.001	0.096	0.213	-0.471	0.073	0.759	0.699
$X_{13}^{0'}$	0	-0.289	-0.136	-0.346	-0.590	1.066	0.380	0.879
$X_{14}^{0'}$	0	-0.230	-0.475	-0.542	-0.686	1.590	0.903	0.784
$X_{15}^{0'}$	0	-0.754	-0.871	-0.906	-0.955	3.008	2.322	0.669
$X_{16}^{0'}$	0	-0.376	-2.520	-0.671	-0.661	3.897	3.210	0.635
$X_{17}^{0'}$	0	-0.162	-7.133	-0.960	-0.905	8.708	8.021	0.564
$X_{18}^{0'}$	0	-0.232	-3.903	-0.972	-0.848	5.532	4.845	0.598
$X_{19}^{0'}$	0	-0.288	-6.266	-0.962	-0.885	7.958	7.271	0.570

3. *Compute the synthetic degree of incidence:*

$$\rho_{1j} = \theta.\ \varepsilon_{1j} + (1-\theta).r_{1j} \qquad (5.14)$$

Whereas:

ρ_{1j}: The synthetic degree of incidence

ε_{1j}: The absolute degree of incidence

r_{1j}: The relative degree of incidence

With j = 2... n, (n = number of symptom. This example, n = 19).$\theta \in [0,1]$

Table 5.6: The synthetic ρ_j value of firm No.1

Firm No.1	ε_{1j}	r_{1j}	ρ_{1i}
X1	1	1	1
X2	0.8806	0.9599	0.9202
X3	0.7546	0.8009	0.7778
X4	0.7367	0.9984	0.8675
X5	0.5309	0.5174	0.5242
X6	0.6466	0.7794	0.7130
X7	0.5624	0.5478	0.5551
X8	0.5113	0.6546	0.5829
X9	0.6624	0.7603	0.7114
X10	0.6542	0.7630	0.7086
X11	0.5289	0.7525	0.6407
X12	0.5804	0.6985	0.6395
X13	0.5777	0.8788	0.7282
X14	0.8114	0.7839	0.7977
X15	0.5117	0.6691	0.5904
X16	0.6918	0.6349	0.6633
X17	0.8969	0.5644	0.7307
X18	0.6556	0.5984	0.6270
X19	0.8997	0.5701	0.7349

As mentioned in the chapter 3, generally researchers take $\theta = 0.5$ to calculate the value of ρ. In this research, firstly, the author will take $\theta = 0.5$ follow by previous researcher to consider the effect of other factor like as X0, the reasonable number of collected data year. After that, the value of θ will be taking into account. So, the formula (5.14) can be converted as follow:

$$\rho_{1j} = 0.5\,\varepsilon_{1j} + 0.5 r_{1j}$$

According to grey analysis theory's principle, the greater value of ρ_i the larger effect of variable X_i into the firm's default (or non-default) characteristics. Therefore, the result are presented in above table 5.6, the VAR2 play the most important role to the firm No.1's default characteristics when X0 = X1.

After the analysis was conducted at a single firm level, the same process with each of the 15 firms (the first 15 firms) will be taken into analysis. By using an excel worksheet, we obtain a matrix of synthetic degree of grey incidence:

$$\rho = \begin{bmatrix} \rho_{11} & \cdots & \rho_{1n} \\ \cdots & \cdots & \cdots \\ \rho_{F1} & \cdots & \rho_{Fn} \end{bmatrix}$$

While as: N = the total number of symptoms manifested at the level of considered firms (N = 19), and F = the total number of firms (F = 15).

	1	0.92	0.78	0.87	0.52	0.71	0.56	0.58	0.71	0.71	0.64	0.64	0.73	0.8	0.59	0.66	0.73	0.63	0.73
	1	0.86	0.75	0.62	0.58	0.67	0.65	0.51	0.58	0.65	0.79	0.53	0.52	0.53	0.51	0.69	0.55	0.66	0.82
	1	0.81	0.86	0.98	0.96	0.8	0.83	0.54	0.66	0.75	0.61	0.75	0.61	0.7	0.84	0.65	0.62	0.66	0.61
	1	0.96	0.73	0.65	0.83	0.65	0.69	0.59	0.55	0.66	0.81	0.81	0.56	0.66	0.69	0.65	0.63	0.62	0.6
	1	0.96	0.99	0.94	0.99	0.9	1	1	0.69	0.73	0.67	0.69	0.59	0.88	0.66	0.93	0.91	0.9	0.9
	1	0.98	0.67	0.61	0.54	0.65	0.54	0.52	0.53	0.54	0.53	0.54	0.53	0.54	0.59	0.54	0.54	0.54	0.53
ρ =	1	0.98	0.64	0.64	0.62	0.58	0.56	0.63	0.52	0.58	0.68	0.53	0.52	0.54	0.81	0.69	0.71	0.72	0.76
	1	0.98	0.72	0.55	0.54	0.74	0.55	0.52	0.53	0.54	0.79	0.53	0.53	0.54	0.54	0.54	0.54	0.54	0.54
	1	0.65	0.78	0.83	0.9	0.64	0.75	0.64	0.61	0.65	0.67	0.8	0.94	0.95	0.71	0.65	0.69	0.64	0.62
	1	1	0.7	0.58	0.57	0.7	0.58	0.56	0.56	0.75	0.54	0.56	0.6	0.58	0.56	0.59	0.59	0.68	0.59
	1	0.99	0.9	0.73	0.53	0.65	0.91	0.53	0.59	0.61	0.56	0.81	0.59	0.61	0.6	0.61	0.53	0.62	0.58
	1	0.8	0.88	0.7	0.72	0.83	0.91	0.6	0.64	0.71	0.67	0.69	0.62	0.94	0.65	0.94	0.94	0.86	0.87
	1	0.9	0.79	0.57	0.54	0.65	0.55	0.67	0.56	0.64	0.78	0.57	0.73	0.68	0.82	0.57	0.57	0.58	0.56
	1	0.85	0.79	0.85	0.76	0.64	0.67	0.62	0.72	0.63	0.59	0.87	0.82	0.85	0.66	0.73	0.7	0.67	0.71
	1	0.93	0.87	0.75	0.87	0.91	0.96	0.55	0.65	0.68	0.85	0.65	0.63	0.69	0.67	0.71	0.66	0.7	0.61

The figures above show the matrix of synthetic degree (ρ) with the number of symptoms is 19. This is the number of symptom in the previous thesis.

An essential emphasis point in the default analysis process is that different financial ratios which have contrary effect on firms' financial statement. In the case X0 = Xi, while Xi stands for a positive symptom, it means that the smaller the level of an aggregated intensity, the more likely is that the firm to become bankrupt, and vice versa when Xi stands for a negative symptom.

The next step after obtaining a matrix of synthetic degree of grey incidence, we will establish intensity levels for each firm. By aggregation, we obtain a matrix of intensity level for each symptom and firms, in the form:

$$Q = \begin{bmatrix} q_{11} & \cdots & q_{1n} \\ \cdots & \cdots & \cdots \\ q_{F1} & \cdots & q_{Fn} \end{bmatrix}$$

The intensity of each firm was proposed to conduct form the synthetic degree of grey incidence as follow: in the case a symptom is positive, we attribute an intensity level q, equal to ρ; if negative symptom, the award will be in reverse order, q = 1 – ρ.

	var 2	var 3	var 4	var 5	var 6	var 7	var 8	var 9	var 10	var 11	var 12	var 13	var 14	var 15	var 16	var 17	var 18	var 19
	+	+	+	-	+	-	+	+	+	-	+	+	+	+	+	+	+	+
	1	0.78	0.87	0.48	0.71	0.44	0.58	0.71	0.709	0.359	0.639	0.728	0.798	0.59	0.663	0.731	0.627	0.735
	1	0.75	0.62	0.42	0.67	0.35	0.51	0.58	0.646	0.211	0.535	0.517	0.53	0.506	0.693	0.554	0.663	0.822
	1	0.86	0.98	0.04	0.8	0.17	0.54	0.66	0.749	0.392	0.746	0.607	0.696	0.835	0.654	0.623	0.656	0.611
	1	0.73	0.65	0.17	0.65	0.31	0.59	0.55	0.661	0.194	0.81	0.562	0.663	0.694	0.646	0.631	0.625	0.598
	1	0.99	0.94	0.01	0.9	0	1	0.69	0.735	0.331	0.687	0.591	0.883	0.663	0.928	0.915	0.903	0.903
	1	0.67	0.61	0.46	0.65	0.46	0.52	0.53	0.544	0.465	0.539	0.528	0.541	0.594	0.537	0.536	0.54	0.533
Q =	1	0.64	0.64	0.38	0.58	0.44	0.63	0.52	0.579	0.317	0.534	0.522	0.538	0.809	0.693	0.713	0.723	0.762
	1	0.72	0.55	0.46	0.74	0.45	0.52	0.53	0.542	0.214	0.535	0.528	0.538	0.536	0.543	0.541	0.542	0.537
	1	0.78	0.83	0.1	0.64	0.25	0.64	0.61	0.645	0.326	0.799	0.941	0.949	0.711	0.652	0.691	0.639	0.623
	1	0.7	0.58	0.43	0.7	0.42	0.56	0.56	0.745	0.463	0.561	0.601	0.579	0.559	0.587	0.588	0.677	0.591
	1	0.9	0.73	0.47	0.65	0.09	0.53	0.59	0.608	0.442	0.813	0.585	0.613	0.602	0.613	0.525	0.62	0.575
	1	0.88	0.7	0.28	0.83	0.09	0.6	0.64	0.712	0.332	0.694	0.619	0.941	0.651	0.941	0.94	0.859	0.874
	1	0.79	0.57	0.46	0.65	0.45	0.67	0.56	0.638	0.219	0.57	0.729	0.683	0.82	0.57	0.565	0.576	0.559
	1	0.79	0.85	0.24	0.64	0.33	0.62	0.72	0.634	0.406	0.87	0.823	0.851	0.656	0.734	0.703	0.671	0.706
	1	0.87	0.75	0.13	0.91	0.04	0.55	0.65	0.676	0.147	0.646	0.626	0.693	0.666	0.711	0.662	0.696	0.611

On our numerical example, the default characteristics of firm in this case was presented by X0 = X1, the current ratio, it means that the smaller the level of intensity, the more likely is that the firm to become bankrupt. Among the considered firms, we can easily see that the firm nr.2, firm nr.6 and firm nr.8 or may be firm nr.4 record low level of intensity for most of the symptoms,comparatively, while firm nr.5 presents the best from this regard. This leads to a point that the firm nr.2, firm nr. 6 and firm nr.8 are the most weakness companies prone to default risk. Form the chart, it do not shows exactly default firm, it just shows which firm may be bankrupt. In this case, the firm nr.2, firm nr.6 and firm nr.8 were predicted bankruptcy.

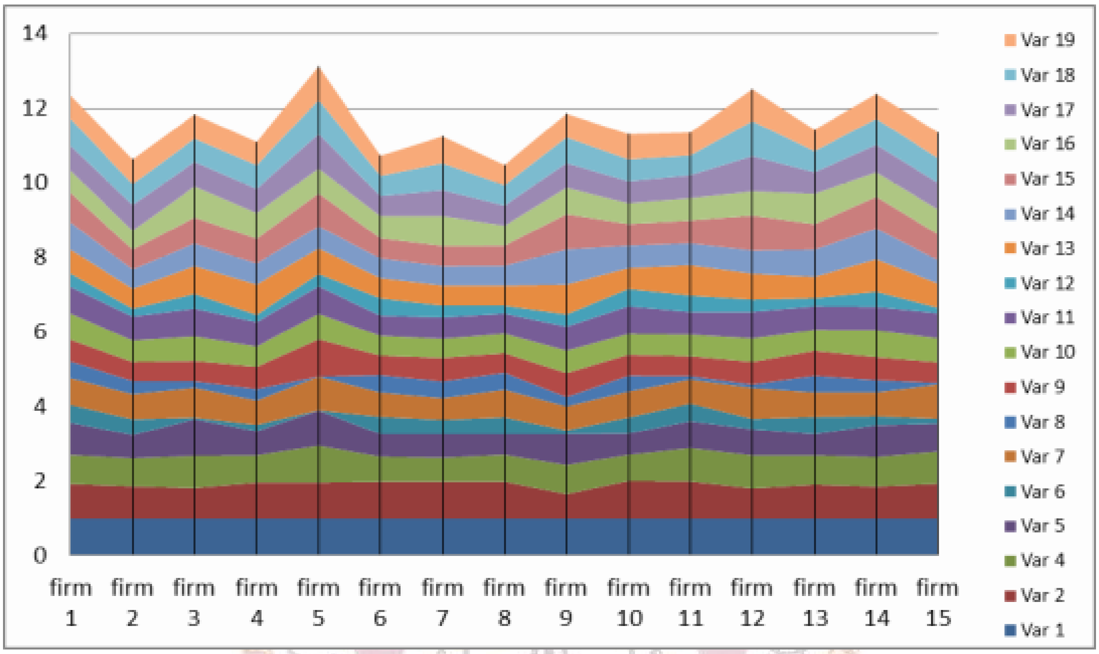


Fig. 5.1: Sum of the registered intensity levels calculated based on matrix Q

5.2 Results

5.2.1 Reasonable data consequence

As mentioned in chapter 3, by measuring the area under the ROC curve, the model's predictive ability can be quantified quickly. The larger the area, the better the diagnostic test is. Prediction value of the four years history data were computed and plotted the ROC curve. The results are illustrated in the table 5.7 below:

Table 5.7: AUC value (846 samples, θ = 0.5, 4 years, training time 35)

Group	X0	Replication	SMOTE
(Liquidity	X1	0.6457	0.6512
	X2	0.659	0.6625
	X3	0.6467	0.6512
	X4	0.6423	0.6534
Leverage	X5	0.7798	0.7835
	X6	0.7081	0.717
	X7	0.7983	0.7992
	X8	0.5975	0.6041
Activity	X9	0.7356	0.7498
	X10	0.6572	0.6619
	X11	0.6781	0.682
	X12	0.7659	0.7719
	X13	0.646	0.6521
	X14	0.7405	0.7559
	X15	0.6492	0.6548
Profitability	X16	0.7582	0.7638
	X17	0.7959	0.7991
	X18	0.7871	0.7919
	X19	0.789	0.7973

5.2.2 Results of previous study (Le Quyen's thesis)

With the same data input and the same method, by applying Grey Theory to compute the synthetic degree of incidence, the prediction value with 19 initial financial ratios were computed and plotted the ROC curve. The AUC results of Le Quyen's thesis are summarized in the following table:

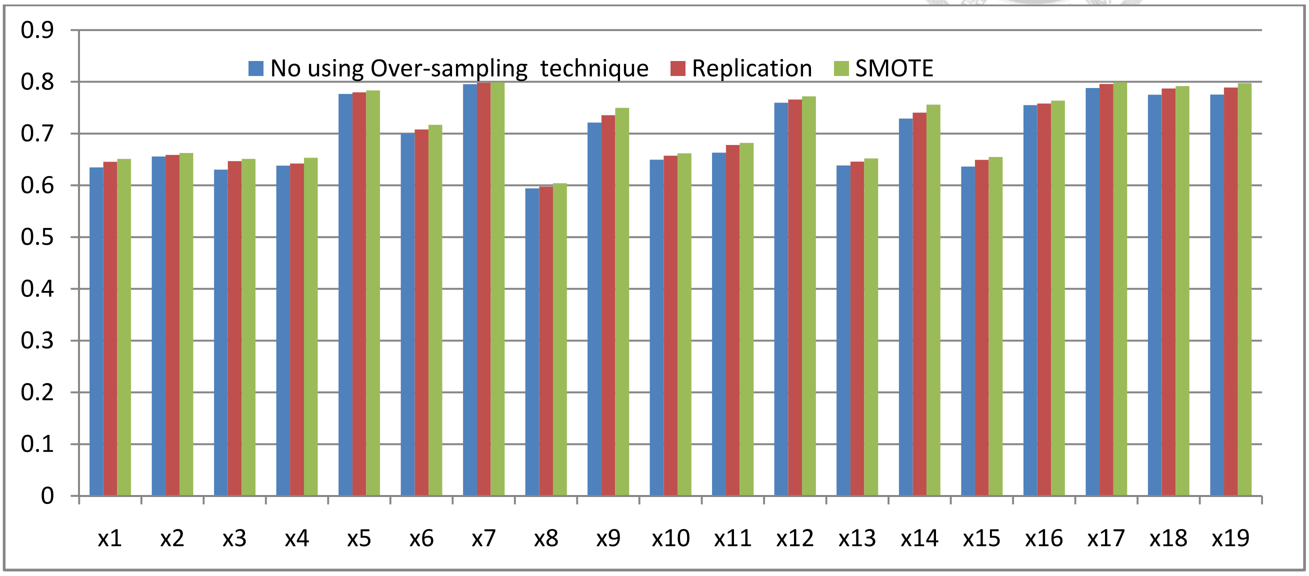
Table 5.8: AUC value (846 samples, θ = 0.5, 19 initial var.)

Group	X0	5 year	4 year	3 year
Liquidity	X1	0.6257	0.6347	0.6516
	X2	0.6344	0.6557	0.653
	X3	0.6451	0.6305	0.6906
	X4	0.6713	0.6382	0.6689
Leverage	X5	0.7548	0.7767	0.7538
	X6	0.6880	0.7001	0.7018
	X7	0.7747	0.7954	0.7337
	X8	0.6286	0.5942	0.5477
Activity	X9	0.6486	0.7212	0.7473
	X10	0.6795	0.6497	0.5912
	X11	0.6319	0.6632	0.5288
	X12	0.7156	0.7515	0.7621
	X13	0.6388	0.6385	0.6548
	X14	0.7118	0.729	0.731
	X15	0.6173	0.6363	0.6396
Profitability	X16	0.7609	0.7549	0.6892
	X17	0.774	0.788	0.7382
	X18	0.7551	0.7752	0.7212
	X19	0.7458	0.7753	0.7125

The table 5.8 above illustrates the prediction value of the three year, the four year and the five year history data were computed and plotted the ROC curve corresponding to each circumstance. And the power of prediction when X0 = X7 (debt ratio) and X0 = X17(ROE) are higher than other variables if 19 initial variables were utilized. And the 4 years history data give the highest accuracy.

By comparison with the best results of the previous thesis, (the same 4 years history data and the same θ = 0.5) the author made the following comments: VAR.7 (Dept Ratio) and VAR.17 (ROE) variables also bring the highest value.

5.2.3 Comparisons



Summary of AUC value (846 samples, □ = 0.5, 19 initial var, 4 year.)

When using Over-sampling technique the value of AUC is bigger than it when no using Over-sampling technique. It can be seen from the chart that AUC with Replication and SMOTE is little bigger than AUC with No using Over-sampling technique. For example, in case of X1, value of AUC of No using over-sampling technique, Replication and SMOTE is 0.6347; 0.6457; 0.6512 respectively.

Between two kinds of over-sampling technique, the SMOTE outperforms the failure prediction compared to Replication. The results of this research are similar to those of Chawla et al. (2002) in a way that SMOTE shows better predicting ability than replication. For instance, in case of x1: value of AUC of Replication is smaller than value of AUC of SMOTE (0.6475 compare with 0.6512).

5.3 Summary

In this chapter, the author develops a numerical example to explain how to apply grey incidence analysis in forecast default construction firms. Applying Over-sampling technique before using Grey system put forward some results. Firstly, when using Over-sampling technique, the value of AUC is bigger in comparison with when not using Over-sampling technique. Secondly, between the two kinds of over-sampling technique, the SMOTE outperforms the failure prediction compared to Replication.

CHAPTER 6: CONCLUSION

By analyzing the historical data, the author found out that the usefulness of Grey Systems Theory in bankruptcy prediction of construction companies was proven by several studies. As presented, the purposes of this research are: Apply over-sampling technique before using Grey System to resolve imbalance data problem. Then, the author want to compare with previous research to find out which method gives higher efficiency and improve the usefulness of Grey system Theory.

Some main conclusions drawn in this research through empirical results are:

1. When using Over-sampling technique the value of AUC is bigger in comparison with when not using Over-sampling technique.
2. Between these two kinds of over-sampling technique, the SMOTE outperforms the failure prediction compared to Replication.

In practice, there are many methods for the prediction default probability. Each method could provide different selected data and then lead to different results in prediction process. Furthermore, the author wish to combine Grey method with the other methods to assess the accuracy of prediction default probability.

REFERENCES

1. Altman, E.I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance, 23, (4), 589-609.*

2. A.Ohlson, J. (1980). Financial ratios and the probabilistic prediction of Bankruptcy. *Journal of Accounting reseach , Vol.18, 109-131*

3. Beaver, W.H. (1967). Financial ratios predictors of failure. Empirical research in accounting: selected studies 1966.*Journal of Accounting Research 4(suppl.), 71–111.*

4. Cheng, J. et al (2009). Business Failure Prediction Model based on Grey Prediction and Rough Set Theory. *WSEAS transactions on information science and applications, Vol. 6.*

5. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P. (2002). "SMOTE: Synthetic Minority Over-sampling Technique," Journal of Artificial Intelligence Research, Vol. 16, pp. 321-357.

6. David S. C (2000). Financial management and ratio analysis for corporative enterprises. *Rural Business-Cooperative Service, U.S. Department of Agriculture. Research Report 175.*

7. Deng, Ju long (1982). Control problems of Grey Systems*, S*ystem and *Control letters, Volume 1, number 5, 288-94.*

8. Delcea, C. & Scarlat, E. The diagnosis of firm's "Diseases" using the grey systems theory methods. *Springer (106-120).*

9. Edum-Fotwe, F., Price, A. & Thorpe, A. A review of financial ratio tools for predicting contractor insolvency. *Construction Management and Economics (1996) 14, 189-19.*

10. Jame M.W Wong & S. Thomas. NG,(2010). Company failure in the construction industry: A critical review and a future research agenda.

11. Kagari, R., Farid, F., & Elgharib, H. (1992). Financial performance analysis for construction industry. *Journal of Construction Engineering and Management, 118, (2), 349 – 361.*

12. Kangari, R. (1988). Business failure in construction industry. Journal of construction Engineering and management , 172-190.

13. Kim, H. & Gu, Z. (2006). Predicting Restaurant Bankruptcy: A Logit Model in Comparison with a Discriminant Model. *Journal of Hospitality & Tourism Research Vol. 30, No. 4, November 2006, 474-493.*

14. Liu, S.F., Lin, Y. (2006). Grey Information: Theory and Practical Applications. *Springer, London.*

15. Ping, J & Kejia, C. (2005). Application of Grey Incidence Analysis to Economic Index Time Difference Analysis. *IEEE*

16. Russell, J. & Zhai, H.(1996). Predicting contractor failure using stochastic dynamics of economic and financial variables. *Journal of Construction Engineering and Management, 122, (2), 183 – 191.*

17. Stein. R.M (2002). Benchmarking default prediction models: Pitfalls and remedies in model validation. *Technical Report #020305.*

18. Taffler, R.J. (1982). Forecasting company failure in the UK using discriminant analysis and financial ratio data. *Journal of the Royal Statistical Society 145, 342–358 (part 3).*

19. Tserng, H.P., Liao, H.H., Tsai, L.K., Chen, P.C. (2011). "Predicting construction contractor default with option-based credit models-models' performance and

comparison with financial ratio models." *Journal of Construction Engineering and Management, Vol. 137, No. 6, pp. 412-420*

20. Tserng, H.P., Lin, G.F, Tsai, L.K., Chen, P.C. (2011). “An enforced support vector machine model for construction contractor default prediction,” *Automation in Construction*, *Vol. 20, No. 8, pp. 1242-1249*

21. Zweig MH, Campbell G (1993). Receiver-operating characteristic (ROC) plots a fundamental evaluation tool in clinical medicine.

APPENDICES

A.1 Data collection of construction firms

GVKEY: Permanent number of Compustat; PERMO: Permanent of CRSP; SIC: standard industry classification code; Default: 0 = non- default, 1 = default.

Information of the 92 sample companies

No.	Company symbol	GVKEY	SIC	PERMO	Year	Default-1
1	ACMTA	1097	1540	10664	1972 - 1978	0
2	6989B	1506	1540	60409	1978-1989	1
3	ATKQ	1835	1600	85607	1985- 1997	1
4	BNC	2013	1600	50390	1970 -1998	0
5	3BBEQ	2015	1540	63095	1972- 1989	1
6	BLAK	2260	1623	18391	1979- 1990	0
7	ENTZ	2340	1700	11901	1988 - 2004	1
8	MTZ	2497	1623	19880	1972 - 2008	0
9	CTX	2845	1531	53831	1972- 1987	0
10	CERB	2889	1700	22382	1981- 2000	1
11	CSTK	3356	1731	86933	1984 - 1988	1
12	DANC.	3737	1520	28442	1972- 1976	0
13	DBNN	3825	1520	29130	1972- 1977	0
14	DCA.2	3901	1531	51449	1971- 1985	0
15	DHM.	3965	1600	45989	1971- 1981	0
16	DONOA.	4042	1600	30518	1972 - 1980	0
17	DY	4115	1623	12008	1984 - 2008	0
18	EDWRQ	4231	1531	31705	1975- 1982	0
19	OHB	4515	1531	54439	1972 - 1979	0
20	FIS.1	4746	1731	28441	1971 - 1989	0
21	FLR	4818	1600	88853	2001 - 2008	0
22	FWLT	4864	1600	18112	1971 - 2002	0
23	GV	5218	1623	32299	1974 - 2008	0
24	GL.2	5308	1600	51836	1971 - 1984	0
25	STRL	5559	1600	76927	1991 - 2008	0
26	HWOD	5689	1531	42412	1972 - 1978	0
27	HBLD	5779	1520	42869	1972 - 1977	0
28	INSU	5978	1623	44274	1982 - 2008	0

29	JEC	6216	1600	52329	1972 - 2008	0
30	EME	6223	1731	82694	1996 - 2008	0
31	5627C	6302	1700	63650	1975- 1987	0
32	FRM	6331	1700	42067	1971 - 2008	0
33	3KCPYB	6411	1531	48717	1971- 1988	0
34	MKE.	7168	1520	27131	1971 - 1977	0
35	MKN.1	7170	1531	54025	1972 - 1980	0
36	HMS.	7399	1531	56039	1972 - 1983	0
37	3MRNKQ	7567	1540	55079	1972 - 1995	1
38	MYR.	7635	1623	66640	1984- 1999	0
39	4657B	7871	1731	58528	1972 - 1985	0
40	CANR10	8028	1700	58641	1982 - 1998	1
41	3PSYE	8451	1600	58763	1976 - 1987	0
42	TPC	8486	1540	50550	1971 - 2008	0
43	PU	8820	1600	14066	1971 - 1979	0
44	PHM	8823	1531	54148	1972 - 1987	0
45	3DVCIE	8900	1731	62375	1980 - 1991	0
46	RII.	8968	1600	27406	1971 - 1982	0
47	RYN.1	9297	1531	52062	1971 - 1986	0
48	RYL	9302	1531	62383	1972- 1987	0
49	4238B	9333	1700	58544	1975 - 1984	0
50	SHA.1	9634	1531	52193	1971- 1982	0
51	SHO.1	10031	1531	53399	1971- 1982	0
52	2976B	10590	1540	76356	1975- 1980	0
53	TUR.1	10770	1540	54287	1976 - 1987	0
54	7621B	11137	1531	53997	1972- 1989	0
55	WBB	11328	1531	39562	1971-1987	0
56	EROQ	12567	1700	10676	1986 - 1993	0
57	3NEKDA	12595	1540	10036	1986- 1990	1
58	NALR	12891	1700	10964	1987 - 1992	0
59	INSMA	13005	1700	11035	1987- 1994	0
60	DEVC	13311	1600	29621	1987- 2007	1
61	3CECN	13471	1531	11109	1987- 1994	1
62	KMMS	14162	1623	10227	1986 - 1998	0
63	UTLX	14337	1623	11853	1988 - 1999	0
64	3ABTE	17209	1731	80220	1994 - 1999	1
65	3NYMRE	19593	1700	76432	1990 - 1994	1
66	USHS	19807	1520	75755	1990 - 2008	0
67	GVA	21429	1600	76135	1990 - 2008	0
68	IWSI	21952	1700	76170	1990 - 1995	0
69	OFP	22838	1623	76245	1990 - 1994	0
70	MTRX	23195	1700	76279	1990 - 2007	0

71	3BLGM	24415	1700	76796	1991- 2007	1
72	ACX.1	24749	1731	77210	1991- 2000	0
73	LODG	24954	1520	77334	1992 - 2004	1
74	ICA	25160	1600	77545	1992 - 2008	0
75	TRCW	27987	1700	79017	1997 - 2004	1
76	CXIPY	29235	1623	79815	1993 - 2000	1
77	DAWKQ	29317	1540	79958	1993 - 2000	1
78	ALNK	30589	1623	80794	1994 - 1998	0
79	NESCQ	31594	1600	82829	1996 - 2000	1
80	MVCO	61425	1600	82507	1995 - 2007	0
81	CVNGY	62197	1540	82731	1996 - 2000	1
82	WG	63495	1623	83834	1996 - 2008	0
83	CBI	64549	1700	84651	1997 - 2008	0
84	FIX	64997	1700	85059	1998 - 2008	0
85	ENGEF	65513	1531	85465	1997 - 2002	0
86	3ESVNQ	65795	1731	85606	1997 - 2001	1
87	NOBLQ	65881	1700	85622	1997 - 2007	0
88	IESC	66371	1731	85768	1998 - 2008	0
89	PWR	66446	1731	85792	1998 - 2008	0
90	HOFF	109043	1623	85986	1998 - 2006	0
91	DESCQ	140071	1600	88642	2003- 2007	1
92	LSMJ	25625	1600	77831	1992 - 1998	1

A.2 Data collection of 15 firms for example mathematics

Actual Year	Firm 1 (ACMTA)	VAR1 X1	VAR2 X2	VAR3 X3	VAR4 X4	VAR5 X5	VAR6 X6	VAR7 X7	VAR8 X8	VAR9 X9	VAR10 X10	VAR11 X11	VAR12 X12	VAR13 X13	VAR14 X14	VAR15 X15	VAR16 X16	VAR17 X17	VAR18 X18	VAR19 X19
1972	Year I	3.43	2.83	0.65	1.25	0.40	0.21	0.28	37.70	3.63	4.99	26.45	3.45	16.78	2.57	40.32	0.15	0.20	0.07	0.22
1973	Year II	3.53	2.96	0.53	0.93	1.02	0.21	0.51	13.82	3.47	3.36	17.45	3.44	11.92	1.98	9.93	0.10	0.17	0.05	0.16
1974	Year III	1.45	1.26	0.23	1.53	3.52	0.04	0.78	-9.55	3.56	2.27	11.19	3.78	14.50	1.35	5.21	-0.23	-1.25	-0.20	-1.18
1975	Year IV	2.59	2.31	0.36	0.77	1.60	0.06	0.62	1.12	4.18	2.10	9.94	4.18	10.97	1.18	3.79	0.05	0.01	0.00	0.01
1976	Year V	4.17	3.42	0.38	0.58	0.93	0.16	0.48	1.45	2.17	1.94	10.14	1.82	6.88	0.81	1.81	0.05	0.02	0.01	0.03

Actual Year	Firm 2 (6989B)	VAR1 X1	VAR2 X2	VAR3 X3	VAR4 X4	VAR5 X5	VAR6 X6	VAR7 X7	VAR8 X8	VAR9 X9	VAR10 X10	VAR11 X11	VAR12 X12	VAR13 X13	VAR14 X14	VAR15 X15	VAR16 X16	VAR17 X17	VAR18 X18	VAR19 X19
1972	Year I	1.13	0.98	0.06	1.07	5.29	-1.18	0.84	-3.66	35.65	14.60	8.82	11.56	29.29	1.66	3.06	-0.03	-0.23	-0.02	-0.59
1973	Year II	1.24	1.16	0.13	1.53	3.79	-0.76	0.79	1.73	34.25	10.87	9.00	18.41	37.80	3.40	8.92	0.03	0.06	0.00	0.10
1974	Year III	1.10	1.06	0.06	2.32	2.38	-0.54	0.70	-75.13	45.82	8.36	7.97	16.06	106.0	4.21	15.11	-0.25	-0.87	-0.07	-3.99
1975	Year IV	0.70	0.66	-0.39	-2.92	-4.16	-0.69	1.32	-34.22	-50.11	7.19	6.59	49.84	117.3	4.47	22.89	-0.83	2.71	-0.13	2.19
1976	Year V	0.65	0.58	-0.50	-2.15	-3.30	-0.82	1.43	2.44	-15.12	8.67	7.58	-18.15	74.09	6.59	90.75	0.06	-0.08	0.00	-0.07

Actual	Firm 3	VAR1	VAR2	VAR3	VAR4	VAR5	VAR6	VAR7	VAR8	VAR9	VAR10	VAR11	VAR12	VAR13	VAR14	VAR15	VAR16	VAR17	VAR18	VAR19
Year	(ATKQ)	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19
1985	Year I	1.61	1.26	0.28	1.37	1.27	0.19	0.56	5.58	7.32	4.96	9.36	5.05	12.00	2.30	8.80	0.05	0.08	0.02	0.12
1986	Year II	1.55	1.23	0.28	1.55	1.39	0.20	0.57	3.64	7.00	4.03	7.96	4.58	10.77	1.95	8.07	0.04	0.05	0.01	0.07
1987	Year III	1.73	1.25	0.32	1.35	1.46	0.18	0.58	2.15	7.01	5.00	8.10	5.18	10.09	2.09	9.01	0.02	0.01	0.00	0.01
1988	Year IV	1.95	1.44	0.39	1.36	1.07	0.19	0.52	2.16	6.18	5.28	7.32	4.98	9.40	2.17	9.78	0.02	0.01	0.00	0.01
1989	Year V	1.55	0.99	0.29	1.73	1.54	0.15	0.61	-9.57	6.43	5.28	6.52	4.98	8.40	2.20	11.48	-0.12	-0.33	-0.06	-0.45

Actual	Firm 4	VAR1	VAR2	VAR3	VAR4	VAR5	VAR6	VAR7	VAR8	VAR9	VAR10	VAR11	VAR12	VAR13	VAR14	VAR15	VAR16	VAR17	VAR18	VAR19
Year	(BNC)	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19
1994	Year I	1.18	0.91	0.11	1.81	1.95	0.07	0.66	3.15	14.94	4.86	8.85	4.97	13.90	1.82	5.54	0.03	0.04	0.01	0.13
1995	Year II	1.34	1.07	0.19	1.81	2.49	0.06	0.71	1.04	11.85	4.96	8.19	6.00	11.22	1.85	7.26	0.01	0.00	0.00	0.01
1996	Year III	1.51	1.22	0.27	1.69	2.19	0.06	0.69	4.28	8.04	5.20	9.78	6.24	11.10	1.86	8.60	0.04	0.08	0.01	0.09
1997	Year IV	1.50	1.12	0.26	1.66	1.97	0.07	0.66	1.90	6.26	4.36	9.21	5.15	8.70	1.67	8.10	0.02	0.01	0.00	0.01
1998	Year V	1.67	1.33	0.34	1.74	1.63	0.09	0.62	3.09	5.98	4.68	9.07	5.03	8.98	1.81	10.04	0.02	0.04	0.01	0.04

Actual	Firm 5	VAR1	VAR2	VAR3	VAR4	VAR5	VAR6	VAR7	VAR8	VAR9	VAR10	VAR11	VAR12	VAR13	VAR14	VAR15	VAR16	VAR17	VAR18	VAR19
Year	(3BBEQ)	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19
1972	Year I	1.42	1.41	0.25	2.02	1.48	0.08	0.60	999.0	13.21	4.59	7.81	8.02	262.1	3.46	19.29	0.08	0.19	0.02	0.31
1973	Year II	1.43	1.41	0.25	2.05	1.51	0.08	0.60	999.0	13.89	4.67	7.75	8.65	268.2	3.47	21.41	0.08	0.21	0.02	0.33
1974	Year III	1.40	1.37	0.24	2.09	1.57	0.09	0.61	999.0	14.65	4.96	8.58	9.15	195.8	3.60	22.49	0.08	0.22	0.02	0.35
1975	Year IV	1.46	1.43	0.26	1.88	1.36	0.09	0.58	999.0	14.86	5.31	9.71	9.16	169.1	3.72	21.93	0.08	0.19	0.02	0.30
1976	Year V	1.39	1.34	0.23	1.93	1.47	0.11	0.60	999.0	13.23	4.76	9.25	7.76	107.0	3.21	17.35	0.07	0.17	0.02	0.31

Actual	Firm 6	VAR1	VAR2	VAR3	VAR4	VAR5	VAR6	VAR7	VAR8	VAR9	VAR10	VAR11	VAR12	VAR13	VAR14	VAR15	VAR16	VAR17	VAR18	VAR19
Year	(BLAK)	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19
1979	Year I	4.20	4.01	0.43	0.65	0.16	0.34	0.13	481.7	4.09	7.59	31.12	2.13	57.14	1.78	4.44	0.13	0.15	0.08	0.31
1980	Year II	6.16	5.80	0.54	0.72	0.12	0.37	0.11	212.3	3.69	8.02	29.43	2.05	46.77	1.80	4.60	0.10	0.11	0.06	0.19
1981	Year III	8.24	7.85	0.62	0.78	0.09	0.47	0.09	92.8	2.49	7.05	25.98	1.61	33.85	1.45	4.54	0.08	0.09	0.06	0.13
1982	Year IV	8.00	7.50	0.65	0.82	0.10	0.52	0.09	175.1	2.16	7.00	23.26	1.51	28.69	1.37	5.01	0.09	0.09	0.06	0.13
1983	Year V	8.82	8.53	0.69	0.85	0.10	0.55	0.09	127.8	1.95	6.09	20.40	1.43	31.96	1.30	5.39	0.07	0.08	0.06	0.11

Actual	Firm 7	VAR1	VAR2	VAR3	VAR4	VAR5	VAR6	VAR7	VAR8	VAR9	VAR10	VAR11	VAR12	VAR13	VAR14	VAR15	VAR16	VAR17	VAR18	VAR19
Year	(ENTZ)	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19
2000	Year I	1.18	1.14	0.07	0.71	0.71	-3.46	0.42	-5.62	58.35	5.33	11.59	3.24	61.72	1.79	3.16	-0.13	-0.27	-0.10	-2.26
2001	Year II	4.11	4.03	0.69	1.17	0.30	-3.02	0.23	-10.39	2.77	5.50	12.41	1.80	60.20	1.26	4.72	-0.08	-0.11	-0.08	-0.12
2002	Year III	2.57	2.34	0.42	0.93	0.70	-4.16	0.41	-43.44	1.78	5.50	16.43	1.49	26.02	1.04	5.86	-0.47	-0.81	-0.35	-1.15
2003	Year IV	1.53	1.42	0.18	0.75	1.59	-5.15	0.61	-16.48	4.16	5.12	15.73	2.59	22.28	1.29	3.25	-0.33	-0.90	-0.24	-1.98
2004	Year V	1.74	1.61	0.21	0.71	0.92	-5.01	0.48	2.39	6.80	6.14	24.75	2.91	29.92	1.34	2.71	0.10	0.11	0.05	0.27

Actual	Firm 8	VAR1	VAR2	VAR3	VAR4	VAR5	VAR6	VAR7	VAR8	VAR9	VAR10	VAR11	VAR12	VAR13	VAR14	VAR15	VAR16	VAR17	VAR18	VAR19
Year	(MTZ)	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19
1972	Year I	1.76	1.64	0.19	0.58	0.66	0.15	0.40	24.77	13.05	6.82	17.92	3.84	52.12	2.32	4.23	0.11	0.17	0.06	0.53
1973	Year II	1.93	1.76	0.21	0.58	0.59	0.17	0.37	21.35	10.25	6.86	19.40	3.38	46.30	2.08	3.73	0.12	0.18	0.06	0.52
1974	Year III	4.05	3.78	0.35	0.53	0.45	0.24	0.31	5.12	5.11	5.61	18.75	2.19	34.51	1.44	2.65	0.05	0.04	0.02	0.09
1975	Year IV	3.83	3.61	0.34	0.52	0.43	0.33	0.30	1.84	3.04	4.97	20.16	1.51	30.21	1.05	1.95	0.02	0.01	0.01	0.02
1976	Year V	4.26	3.98	0.36	0.52	0.40	0.31	0.29	4.15	3.42	6.19	21.44	1.68	34.82	1.19	2.21	0.04	0.03	0.02	0.06

Actual	Firm 9	VAR1	VAR2	VAR3	VAR4	VAR5	VAR6	VAR7	VAR8	VAR9	VAR10	VAR11	VAR12	VAR13	VAR14	VAR15	VAR16	VAR17	VAR18	VAR19
Year	(CTX)	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19
1983	Year I	1.58	0.36	0.21	0.87	1.59	0.28	0.61	4.41	5.84	10.61	5.35	3.06	2.30	1.13	2.59	0.07	0.12	0.04	0.23
1984	Year II	1.84	0.31	0.32	1.11	1.80	0.21	0.64	4.69	4.98	12.45	5.25	3.41	2.37	1.27	3.33	0.07	0.13	0.03	0.15
1985	Year III	1.81	0.35	0.29	0.99	1.79	0.21	0.64	7.65	5.40	16.70	5.72	4.53	2.75	1.62	4.84	0.07	0.15	0.03	0.18
1986	Year IV	2.04	0.30	0.29	0.79	1.74	0.26	0.63	10.99	4.75	16.64	4.92	3.78	2.53	1.37	3.44	0.05	0.12	0.03	0.15
1987	Year V	1.75	0.24	0.26	0.95	1.85	0.25	0.65	4.28	5.20	20.18	5.04	4.01	2.72	1.44	3.52	0.03	0.07	0.02	0.09

Actual	Firm 10	VAR1	VAR2	VAR3	VAR4	VAR5	VAR6	VAR7	VAR8	VAR9	VAR10	VAR11	VAR12	VAR13	VAR14	VAR15	VAR16	VAR17	VAR18	VAR19
Year	(CERB)	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19
1981	Year I	1.95	1.95	0.46	1.80	0.92	0.05	0.48	8.49	6.84	3.51	40.06	6.08	999.0	3.20	48.25	0.13	0.20	0.04	0.22
1982	Year II	1.73	1.73	0.40	2.06	1.20	0.08	0.54	11.94	7.87	3.68	36.73	6.87	999.0	3.30	53.84	0.16	0.30	0.05	0.35
1983	Year III	3.39	3.39	0.67	1.32	0.39	0.08	0.28	19.40	4.92	3.94	17.31	4.53	999.0	2.96	56.20	0.09	0.11	0.04	0.12
1984	Year IV	3.56	3.56	0.65	1.21	0.34	0.09	0.25	70.88	3.89	3.99	14.20	3.51	999.0	2.57	35.57	0.09	0.11	0.03	0.13
1985	Year V	1.97	1.97	0.33	1.00	0.81	0.14	0.45	12.26	4.24	3.93	15.11	3.09	999.0	1.95	8.24	0.08	0.12	0.04	0.21

Actual	Firm 11	VAR1	VAR2	VAR3	VAR4	VAR5	VAR6	VAR7	VAR8	VAR9	VAR10	VAR11	VAR12	VAR13	VAR14	VAR15	VAR16	VAR17	VAR18	VAR19
Year	(CSTK)	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19
1984	Year I	1.68	1.68	0.34	1.64	2.16	0.08	0.68	5.31	11.23	5.14	24.53	11.16	999.0	3.52	18.84	0.08	0.17	0.02	0.16
1985	Year II	1.42	1.42	0.23	1.65	2.51	0.06	0.72	0.73	12.62	5.12	21.89	11.70	999.0	3.51	17.04	0.03	0.00	0.00	-0.01
1986	Year III	1.81	1.81	0.36	1.41	3.75	0.06	0.79	-1.87	8.16	3.73	11.65	9.77	999.0	2.40	10.85	-0.02	-0.28	-0.03	-0.17
1987	Year IV	2.19	2.18	0.48	1.48	3.46	0.01	0.78	-1.97	6.10	3.85	11.28	11.80	733.4	2.57	16.03	-0.05	-0.37	-0.03	-0.17
1988	Year V	1.89	1.86	0.44	1.80	38.03	-0.08	0.97	-2.97	4.75	3.01	9.26	17.44	180.8	2.17	22.39	-0.14	-7.42	-0.09	-0.44

Actual	Firm 12	VAR1	VAR2	VAR3	VAR4	VAR5	VAR6	VAR7	VAR8	VAR9	VAR10	VAR11	VAR12	VAR13	VAR14	VAR15	VAR16	VAR17	VAR18	VAR19
Year	(DANC)	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19
1972	Year I	1.24	0.59	0.12	1.24	1.23	0.09	0.55	19.00	32.95	17.47	16.95	9.75	13.23	4.30	11.68	0.08	0.16	0.02	0.59
1973	Year II	1.09	0.56	0.05	1.65	1.59	0.09	0.61	21.71	49.00	14.74	14.21	9.57	11.43	3.95	10.97	0.07	0.16	0.02	1.24
1974	Year III	1.24	0.50	0.14	1.63	1.67	0.07	0.63	8.59	45.38	17.31	15.64	11.64	11.17	4.42	13.88	0.08	0.17	0.02	0.48
1975	Year IV	1.35	0.47	0.18	1.41	1.34	0.07	0.57	10.96	28.87	21.82	15.74	11.32	9.95	4.54	14.99	0.09	0.17	0.02	0.41
1976	Year V	1.51	0.66	0.23	1.28	1.10	0.08	0.52	20.42	23.33	23.90	17.50	10.66	10.99	4.81	15.59	0.08	0.15	0.01	0.30

Actual	Firm 13	VAR1	VAR2	VAR3	VAR4	VAR5	VAR6	VAR7	VAR8	VAR9	VAR10	VAR11	VAR12	VAR13	VAR14	VAR15	VAR16	VAR17	VAR18	VAR19
Year	(DBNN)	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19
1972	Year I	3.20	1.86	0.57	1.13	0.35	0.30	0.26	48.56	3.44	5.74	14.02	2.63	4.51	1.98	11.03	0.06	0.07	0.03	0.10
1973	Year II	2.08	1.05	0.42	1.34	0.69	0.32	0.41	29.96	3.44	4.75	12.31	2.56	3.69	1.68	9.48	0.06	0.10	0.04	0.14
1974	Year III	1.63	0.83	0.32	1.65	1.06	0.23	0.51	2.97	5.04	5.47	12.29	3.44	3.99	1.84	10.05	0.05	0.05	0.02	0.08
1975	Year IV	2.15	1.26	0.41	1.20	0.71	0.24	0.41	11.82	5.31	6.29	10.69	3.61	4.33	1.93	9.42	0.11	0.16	0.05	0.22
1976	Year V	2.26	1.49	0.42	1.14	0.75	0.32	0.43	19.43	3.78	6.10	9.38	2.72	4.37	1.57	6.63	0.07	0.11	0.04	0.15

Actual	Firm 14	VAR1	VAR2	VAR3	VAR4	VAR5	VAR6	VAR7	VAR8	VAR9	VAR10	VAR11	VAR12	VAR13	VAR14	VAR15	VAR16	VAR17	VAR18	VAR19
Year	(DCA.2)	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19
1971	Year I	3.38	1.59	0.61	1.17	2.32	0.15	0.70	18.92	2.03	26.81	17.18	3.80	1.61	1.10	7.46	0.09	0.27	0.11	0.13
1972	Year II	2.90	1.01	0.44	0.87	2.13	0.13	0.68	11.09	2.04	17.90	20.70	3.29	1.69	1.03	4.04	0.10	0.25	0.10	0.18
1973	Year III	2.61	0.80	0.39	0.84	2.31	0.17	0.70	6.93	2.14	11.26	19.52	2.85	1.53	0.88	2.53	0.07	0.18	0.07	0.14
1974	Year IV	2.66	0.70	0.44	0.95	2.35	0.14	0.70	-0.85	1.70	6.88	16.04	2.34	1.11	0.70	2.10	0.00	-0.16	-0.06	-0.11
1975	Year V	2.44	0.46	0.39	0.91	1.39	0.22	0.58	2.50	1.66	7.64	19.92	1.97	0.94	0.69	2.17	0.07	0.07	0.04	0.08

Actual	Firm 15	VAR1	VAR2	VAR3	VAR4	VAR5	VAR6	VAR7	VAR8	VAR9	VAR10	VAR11	VAR12	VAR13	VAR14	VAR15	VAR16	VAR17	VAR18	VAR19
Year	(DHM)	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19
1971	Year I	1.45	1.08	0.10	0.40	2.24	0.18	0.69	1.12	11.19	5.09	12.44	3.23	11.56	1.03	1.55	0.03	0.02	0.01	0.06
1972	Year II	1.46	1.11	0.09	0.37	2.10	0.18	0.67	1.70	10.79	5.32	12.87	3.27	11.92	1.03	1.48	0.04	0.05	0.01	0.17
1973	Year III	1.35	1.06	0.09	0.46	2.02	0.18	0.67	2.98	12.88	5.71	11.99	3.59	14.25	1.17	1.72	0.05	0.08	0.02	0.29
1974	Year IV	1.41	1.04	0.11	0.51	2.16	0.16	0.68	3.04	13.73	5.83	11.75	4.23	14.19	1.36	2.12	0.06	0.09	0.02	0.27
1975	Year V	1.43	1.07	0.13	0.60	2.11	0.15	0.67	2.63	12.65	5.82	12.05	4.67	13.13	1.48	2.46	0.06	0.09	0.02	0.23