CHAPTER – IV

RESEARCH METHODOLOGY

4. RESEARCH METHODOLOGY

4.1. Rationale of the Study

Since there has been very limited research in the field of wilful default in India and more specifically related to prediction of wilful default, the present study is a modest attempt to highlight the problem and explore solutions which are scientifically relevant. An extensive study of literature across academic and research articles along with review of several bankruptcy predication models from across the World, few areas which required immediate attention are highlighted in the study.

The use of existing models under Indian circumstances with reference to Wilful Default requires an examination to verify the accuracy of the models. Further, new models should be constructed that can suit the condition of Indian Wilful Default companies. Based on survey, Econometric Techniques should be considered as it has an advantage of maximum transparency as well as Artificial Neural Network with maximum accuracy rate. It is an interesting and important proposal to study Logistic Regression, Multivariate Discriminant Analysis, Decision Tree and Artificial Neural Network models as it overcomes the weakness of each other.

4.2. Objective of the Study

1. To build and compare bankruptcy prediction models for Wilful Default public limited companies listed after the year 2000 by using Logistic Regression, Multivariate Discriminant Analysis, Decision Tree and Artificial Neural Network.
2. To use existing bankruptcy models to predict Wilful Default public limited companies listed after the year 2000.

4.3. Type of Research

The objective of the research is to understand and demystify the phenomena of Wilful Default through bankruptcy prediction amongst public limited companies. The research primarily is categorized as analytical; it focuses on the problem of default that negatively affects the financial sector; lending institutions, banks, and economy at large. The research study is based on the financial performance data of the companies. It critically examines and tries to draws relationship amongst variables. It is primarily a statistical compilation and follows vague computations. It is also an applied research since it focuses on the banking business related problems. Research work is quantitative in nature; it takes only the financial performance data. The present research can be termed as empirical research since is based on past data, *ex post.* The approach is deductive where conclusions drawn by using various statistical techniques.

The research is based on public limited listed companies in India; more than 5000 companies are listed on the major stock exchanges of India; National Stock Exchange and Bombay Stock Exchange. The research primarily focuses on the companies which are declared as Wilful Default by Indian Banks. However, the scope of research is limited to the companies which came up with IPO's and got listed after the year 2000. Since the enactment and amendments in Securities and Exchange Board of India (SEBI) Disclosure, Investors and Protection) Regulation, 2000 free market mechanism was followed to decide the price of IPO by the issuer. Before 2000, Capital Controller of India (CCI) controlled the IPO pricing and premium, which was against the market principles and led to undervaluation of the stocks. Between 2000 and 2017, over 900 IPOs hit the primary markets, out of which, 106 have been declared as Wilful Defaulters, the list is prepared from National Stock Exchange's website using that as the base to search Wilful Default published by TransUnion Cibil Pvt. Ltd. These 106 companies constitute the entire population of the study.

4.4. Scope of Research

The scope of research is narrowed to the Wilful Default Public Limited Companies. According to Companies Act, 2013 Section 14, all public limited companies are mandated to disclose its final accounts. The assurance of data availability help in narrowing down the focus to public limited companies especially the Wilful Default declared by the Banks as defined by RBI. As stated above year 2000 is considered as a reference point and the companies listed after 2000 are considered for the research.

4.5. Sample Design

Sample design is the description of the samples shortlisted for the research; it includes the description of the entire Population, sampling Unit, the Source of data, Sampling Method, Size of the sample and Parameters of Interest.

4.5.1. Universe/Population

It comprises of all the Wilful Default Public Limited listed companies in India after 2000 constituting the population is106 listed companies.

4.5.2. Sampling unit

It is same as the population since the study focuses on all 106 listed Wilful Default companies in India.

4.5.3. Data Source

The source of data includes websites of Credit Information Company (CIC)like TransUnion Cibil, Equifax, CRF High Mark and Experian. According to the notification of RBI, CICs have to publish the list of Wilful Default as reported by respective banks and financial institutions on quarterly basis. The objective of displaying names in the public domains is to alert other lenders, investors and public at large. The list of IPOs is obtained through the website of SEBI and National Stock Exchange. It provides the data regarding IPOs with dates, amount, number of shares, Book Running Lead Manager, etc. While data related to financial performance of the companies under study is taken from EBSCO listed database called Ace Equity.

4.5.4. Sampling Method

The method of sampling is Purposive since the sample is selected based on the pre-decided defined objective.

4.5.5. Sample Size

The sample size is limited to 106 public limited wilful default companies. However, for the construction of Bankruptcy Prediction Model, an equal number of Non-Default Companies is also considered. The top 106 companies from BSE 200 Index are shortlisted as non-default companies based on market capitalization excluding the banking and financial sector, because the assets of banks and financial sector constitutes of Loans and Advance which take years to get recovered. The data related to these companies is taken as on 24th July 2019.

4.5.6. Parameters of Interest

In order to eliminate the magnitude, the scale of the company, the data on financial performance is taken in form of ratios from the year 2000 to 2018.

The types of ratios considered for the study are presented in the below Table 4-1. The ratios are classified under five categories including liquidity, profitability, capital structure, valuation, cash flow and miscellaneous. Total number of ratios taken is 21 across these categories.

Table 4-1 List of Ratios

No.	Types of Ratio
1	**Liquidity Ratios**

1.1	Current Ratio
1.2	Net Working Capital/ Total Assets
2	**Profitability Ratios**
2.1	Net Profit Margin
2.2	Operating Profit Margin
2.3	PBIT Margin
2.4	Return on Assets
2.5	Return on Shareholders' Fund
2.6	Return on Capital employed
3	**Solvency and Valuation Ratios**
3.1	Interest Service Coverage Ratio
3.2	Total Debt/Total Assets
3.3	Retained Earnings/Total Assets
3.4	EBIT/Total Assets
3.5	Sales/ Total Assets
3.6	Debt/Enterprise Value
3.7	Profit After Tax/ Enterprise Value
3.8	Enterprise Value / Total Assets
4	**Cashflow Ratios**
4.1	Increase(Decrease) Loan Funds/Cashflow from Loan
4.2	Cashflow Financing/Cashflow Investing
5	**Miscellaneous**
5.1	Market Capitalization/Outstanding Debt
5.2	Sales/Capital Employed
5.3	Minority Interest/PAT

Rationale of various categories of ratios like liquidity, solvency, valuation, cashflow, profitability and miscellaneous are discussed further.

4.5.6.1. Liquidity Ratio:

Liquidity Ratios are indicator of short term solvency of the company. They provide insights on the working capital of the business. Liquidity is the primal pre-requisite of survival for any business house. It indicates how well the company is functioning on day to day basis and whether the funds are utilized appropriately. In case of liquidity crunch, the company may not be able to pay off its instalment, interest on loans, employee salary, payment to suppliers, etc. Current Ratio indicates the capacity to pay short term liabilities through short term assets; cash, bank, inventory, bills receivables and debtors. The study covers the Current Ratio and Working Capital to Asset ratio to understand the pattern of liquidity of Wilful Default and non-default companies in India. It has been used by many previous researchers (Smith & Winakor, 1935) (Merwin, 1942), (Altman, 1968), (Libby, 1975). Current Ratio is the comparison between Current Assets and Current Liabilities. Short term liquidity problems can become the reason for major failures. Net Working Capital to Total Assets provides the view on composition of working capital against total assets. A company with continuous decline in the ratio can bring out the pattern of financial instability.

4.5.6.2. Profitability

The ultimate objective of the firm is reflected through its profits. Profits decide the future of the company. Profit is the result of Sales, operational efficiency, capital employed, etc. It has always been an important yardstick for analysis. A plethora of researches have used profitability related variable for the bankruptcy models (Jardin P. D., 2009), (Ahmadi A. P., Soleimani, Vaghfi, & Salimi, 2012), (Aliakbari, 2016), (Nouri & Soltani, Designing a bankruptcy prediction model based on account, market and macroeconomic variables, 2016). Primarily, Net Profit Margin, Operating Profit Margin, Return on Assets, Return on Capital Employed, Return on Net-worth, EBIT Margin and Minority Profit to PAT have been used for building up the model. Most of the model covers profit related parameters invariably.

4.5.6.3. Solvency and Valuation Ratios

The most critical of all is the long term solvency ratio, it indicates whether the borrower has maintained good repayment provisions and the trend provides a clear idea of the direction. It includes Interest Service Coverage Ratio, Total Debt/Total Assets, Retained Earnings/Total Assets, EBIT/Total Assets, Sales/ Total Assets, DEBT/Enterprise Value, Profit After Tax/ Enterprise Value and Enterprise Value / Total Assets.

Interest service coverage ratio signifies the capacity to cover the interest component against Earnings before Interest and Taxes. Total Debt to Assets signifies the amount of debt used to finance total assets. Retained Earnings to Assets shows the proportion of reserves and surplus till date against assets. EBIT and Sales to Asset indicates how efficiently the company is using its assets to generate sales and earn EBIT. Rest of the ratios is related to the valuations and it is important in quantifying the worth of the Company or Enterprise. The ratio related to the debt, PAT and Assets are compared with Enterprise Value. It provides a significant insight on where the company stands and helps the bankers to evaluate and approve loan accordingly (Jardin P. D., 2009) (Altman, Financial Ratios, Discriminant Analysis And The Prediction Of Corporate Bankruptcy, 1968) (Libby, 1975) (Ohlson, Financial Ratios and the Probabilistic Prediction of Bankruptcy, 1980).

4.5.6.4. Cash-flow Ratios

Cash flow is the lifeblood of all business. Hence, the financial analysis should be done considering the liquidity position in the business. Increase(Decrease) Loan Funds/Cash flow from Loan signify whether the company is repaying the loan in normal course or borrowing from somebody else to pay off its debt. Cash flow from Financing Activity by Cash flow for Investing Activity shows the significance of loan taken against the investment made. This also indicates whether the long term borrowings are used for investments in assets or the long term loan is used for operating purpose.

4.5.6.5. Miscellaneous

Other ratios include Market Capitalization to Outstanding Debt and Sales to Capital Employed; the former is applicable to a listed company only. Since the study is based on listed company, it would be absolutely relevant and important to consider this ratio. It signifies the market value of the company against the debt. Market capitalization is the perceived market value by all the investors in the market. It was used by Altman to develop Z score model. Another ratio on Sales to Capital Employed shows how significant is the sales against total owners' fund and borrowed fund. It is through the trend one can measure the efficient use of funds.

The ratios stated in Table 4-1 are to be used as independent variables for the short listed models. With reference to objective regarding usage of existing models and based on the literature survey Ohlson O Score and Altman Z Score model is used. Details of existing models are discussed further.

4.5.6.6. Existing models: Ohlson O-Score and Altman Z-Score

4.5.6.6.1. Ohlson O-Score

Ohlson O-Score model is very famous and used under various contexts. It is renowned for its usage of not just the financial ratios but also macroeconomic indicator like Gross National Product. It considers the effect of state of the business in the economy for prediction of bankruptcy. In 1980, Dr. James Ohlson from the New York University developed an alternative to Altman Z-Score method using Multivariate Discriminate analysis for Bankruptcy prediction models. Interestingly, macroeconomic factor like Gross National Product was incorporated to provide a benefit of doubt for the failed businesses due to business cycle impact (Ohlson, Financial Ratios and the Probabilistic Prediction of Bankruptcy, 1980).

$$T=-1.32-0.407\log(TA_t/GNP)+6.03TL_t/TA_t+0.0757CL_t/CA_t-1.72X-2.37NI_t/TA_t-1.83FFO/TL_t+0.285Y-0.521\ (NI_t-NI_{t-1})/\ |\ NI_t|+|\ NI_{t-1}|$$

1. TA = total assets
2. GNP = Gross National Product price index level
3. TL = total liabilities
4. WC = working capital
5. CL = current liabilities
6. CA = current assets
7. X = 1 if TL > TA, 0 otherwise
8. NI = net income
9. FFO = funds from operations
10. Y = 1 if a net loss for the last two years, 0 otherwise

The score less than 0.5 indicate high probability of bankruptcy while more than 0.5 is considered to be safe.

4.5.6.6.2. Altman Z-Score

Altman Z-Score model is one of the most widely used models across the countries and in different contexts. Edward Altman further refined the model and made Zeta Score but it is not available in the public domain as it has been under the intellectual proprietary. It gives only 5 financial ratios covering liquidity, profitability, efficiency and market value of the company. The use is particularly limited to listed company as X4 variables contain Market Capitalization of the company. Following is the formula

$$Z = 0.012X1 + 0.014X2 + 0.033X3 + 0.006X4 + 0.999X5$$

1. X1 = Working Capital / Total Assets (WCTA) (%),
2. X2 = Retained Earnings / Total Assets (RETA) (%),
3. X3 = Earnings Before Interest And Tax / Total Assets (EBITTA) (%),
4. X4 = Market Value Of Equity / Book Value Of Total Liabilities (VETL) (%),
5. X5 = Sales / Total Assets (STA) (number of times)

The final result is measured through a score further bifurcated in 3 categories; more than 3.00 as the safest or green, between 1.81 and 2.99 as cautious or amber, and lastly less than 1.8 would be very risky or red flag.

4.6. Data Collection

The data is extracted from the Software Ace Equity powered by Accord FinTech. The software is recognized and listed in EBSCO. Statistical computations are on IBM SPSS and Microsoft Excel.

4.7. Data Size

The data size is presented using various variables like Companies, Number of Years, Company Years (which comprises of number years the data for the company was available), Financial Ratios, Wilful Default Years and finally all Companies with all years and parameters. The details are presented in Table 4-2 Details of Data for the research**Error! Reference source not found.**.

Table 4-2 Details of Data for the research

Sr. No	Parameter	Total Observations
1	Total companies	106+106=212 (Default and Non Default)
2	Number of Years	18 (2000-2018)
3	Company Years	3319
4	Total Parameters	Total 36 Variables 21 (Financial Ratios for Model Construction Objective) 15 (Usage of Existing Models)
5	Wilful Defaults-Years and Parameters	28917

6	Non Defaults-Years and Parameters	40782
7	All Companies, All Years, All Parameters	69699

There are total 212 companies out of which 106 are Wilful Default and the remaining 106 are top companies listed on BSE 100 based on market capitalization.

The total period of study is 18 years from 2000 to 2018. The total number of observations under Company Years is 3319. Total Parameters taken for the objectives are 36, it includes 21 financial ratios for model construction and rest 15 variables are for existing model usage. Total observations for Wilful Default across all the years and parameters were 28,917 while Non-Wilful Default is 40,782. By the end of financial year 2018 only 56 out of 106 Wilful Default companies were listed. The entire data set has 69699 observations.

A company is not instantly declared a wilful default, as per the literature review and similar studies performed earlier, it takes a period of 2 years of non-payment of interest or principal, post which the company is declared to be a wilful default company. In the present study a company is considered wilful default as 'Yes' if for the previous two years the company has defaulted on repayment of interest or principal. While the status of company will remain 'No' if it has not defaulted or not listed under wilful default list. For a total of 315 company years the status is considered as 'Yes' and for 1062 company years the status is considered as 'No' for the purpose of present study. However, for a total of 2932 company years including both the wilful default and non-default the status is 'No'.

4.8. Data Analysis

The available data is analysed using the two prominent bankruptcy prediction models; Altman Z-Score Model and Ohlson O-Score Model. The results are useful in meeting the listed research objectives. The results were also used to segregate the default companies under different sectors and industries to give a better understanding of sector wise wilful default. Among other analysis, the study also attempted to find if the default companies received advances subsequent to being predicted to be defaulted or on a high risk of being default. Such outcomes raise serious questions regarding the quality of credit appraisal and lending efficiency of commercial banks.

Data is used to formulate new models and compare the accuracy between Logistic Regression, Multivariate Discriminant Analysis, Neural Network and Decision Tree method. The significant variables out of the total variables are to be short listed and compared across the methods. Special focus is on the accuracy as the final product of the models lies in accurate predictability. Further discussed four model construction techniques; Logistic Regression, Multivariate Discriminant Analysis, Neural Network and Decision Tree method.

4.8.1. Logistic Regression

Logistic Regression is a statistical technique which uses logistic function for dichotomous dependent variables. It is mainly used where the outcome is binary; 0 or 1 and in this case default or not, buy or not, etc. The logistic regression computes the probability that the binary response is as a function of a set of predictor variables and regression coefficients presented as:

$X=[X_0, X_1, X_2, X_3 ... X_n]^T$-Independent Variables

$B=[\beta_0+\beta_1+\beta_2+..... \beta_n]^T$ – Regression co-efficient

P=probability

β_0-Constant

The outcome of logistic regression function provides the probability and to ensure the probability is between 0 and 1; the function is divided with 1+ function.

$$\ln\left(\frac{p}{1+p}\right) = \beta 0 + X_1 \beta_1 + X_2 \beta_2 ...... X_n \beta n$$

With the use of SPSS statistical software, the database selected outcomes to be focused on:

The method selected includes the command to display each step of optimal model with classification plots, Hosmer-Lemeshow goodness of-fit, Correlation of variables, Iteration History and Constant in the model. By default, Classification cut off is 0.5 and Maximum iteration is 20 while Probability step-wise has Entry of 0.05 and removal of 0.10. The results are computed using forward method; forward method starts model construction with a step-wise method where one variable and keeps adding significant variables and eliminates insignificant variables.

4.8.2. Multivariate Discriminant Analysis

Discriminant Analysis searches a set of prediction equations; it helps to classify individuals into groups through independent variables. It helps in better understanding of the relationship amongst the variables. It helps to find relationship through mathematical expressions. Discriminant analysis is used to determine the minimum number of dimensions needed to describe these differences. A distinction is sometimes made between descriptive discriminant analysis and predictive discriminant analysis.

MDA analysis is done through a structured method which includes Eigen Values, Canonical Correlation which is expected to be more than 0.5, Wilks' Lamda and provides Standardized Canonical Discriminant Function Coefficients for those variables which are significant to model construction. Wilks' lambda is a measure of how well each function separates cases into groups.

4.8.3. Artificial Neural Network (ANN)

Artificial Neural Network is an attempt to replicate the brain's neural network. It has been used extensively for programming of Artificial Intelligence Software. As per the literature review, this method has the highest accuracy rate. Results from SPSS software to be considered are:

ANN case processing is divided into training and testing data in the proportion of 70:30. Total number of hidden layers is 1, units in hidden layers are 6 and activation function is Hyperbolic Tangent. For output layer, activation function is based on Softmax and Error Function is Cross-Entropy.

4.8.4. Decision Tree

Decisions Tree creates classification and helps in better identifying groups, discover relationships between groups and predict future events. Visual diagrams enable to present categorical results in an intuitive manner; it can clearly explain the results to non-technical audiences. The trees explore results and visually determine model flows. Visual results can help to find specific subgroups and relationships that might not uncover using more traditional statistics. Because classification trees break the data down into branches and nodes, one can easily see where a group splits and terminates.

The process of Decision Tree includes finding the Risk resulting in estimation and standard error by using CHAID growing method. CHAID builds all possible cross tabulations under category predictor until the best outcome and there is no need for further splitting. It creates the visuals of the relationships between the split variables and the associated related factor within the tree.

Based on the outcome of all 4 types, significant variables are shortlisted and accuracy of the method is summarized.