CHAPTER IV

RESEARCH METHODOLOGY

4.1 Introduction

In the previous chapter the literature reviewed predominantly demonstrates the use of financial variables in the early warning prediction models. It is evident that though different methods and statistical tools have been used in different studies, both carried out internationally and in Indian context, but primarily financial variables have been used as potential indicators for corporate distress or bankruptcy.

The primary objective of this chapter is to describe scope of the research which also includes the age-wise, entity-type and Industry group classification of the suit filed wilful defaulters’ (SF_WD) data. The hypotheses framed for empirical examination have been dwelt upon and finally the research design and the methodology adopted in this study have been provided. The chapter also discusses as to how the data were collected and analysed for satisfying the research objective. Moreover, the data quality and its completeness are highlighted so that they would be kept in mind while drawing conclusions.

4.2 Scope of the Research

As seen in the previous chapter, the study on wilful defaults is very limited. They are either based on the general characteristics of these defaulters or partial analysis carried out pertaining to a specific geographic area or to a particular sector. Hence, there are no comparable literatures available with regard to prediction of wilful defaults. As provided in Chapter II, credit institutions were instructed by RBI in 2002 to submit the list of suit-filed accounts of wilful defaulters of Rs.25 lakh and above on a quarterly basis to CIBIL, a credit information company. This study is an initiative to explore this data. The data has been extracted from CIBIL website for the period March 2002 to March 2016. Thus, the scope of the study is quite comprehensive and covers data
reported on suit filed wilful defaulters (SF_WD) by all groups of credit institutions i.e. Co-operative banks, Financial Institutions, Nationalized Banks, Foreign Banks, Private Sector Banks and State Bank of India and its Associate banks. The annual financial indicators of the companies taken in the sample set, which have been reported as wilful by these credit institutions, pertain to both public and private (listed and unlisted) companies of India. As the data taken for study is as reported by all the branches of the credit institutions to the bureau, hence it has complete coverage geographically and has been analysed for all the States and Union Territories of India. An effort has been made to critically analyse this data extensively for the last fifteen years.

4.2.1 Analysis of Companies as per their age group

It was felt pertinent to undertake study of certain basic characteristics of the wilfully defaulted companies. The age-group of the companies taken in the sample data set were analysed. The year of incorporation ranges from 1911 to 2007. Table 13 and Figure 5 depicts the year of incorporation of these companies and the number of such companies which falls in that range.

<table>
<thead>
<tr>
<th>Year of Incorporation</th>
<th>Number of Companies</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1911-1970</td>
<td>29</td>
<td>10%</td>
</tr>
<tr>
<td>1971-1980</td>
<td>30</td>
<td>11%</td>
</tr>
<tr>
<td>1981-1990</td>
<td>114</td>
<td>41%</td>
</tr>
<tr>
<td>1991-2000</td>
<td>85</td>
<td>30%</td>
</tr>
<tr>
<td>2001-2010</td>
<td>21</td>
<td>8%</td>
</tr>
</tbody>
</table>

(Source: Developed for this thesis)
It was observed that the year of incorporation of majority of the companies’ falls between the years 1981 to 1990, i.e., the companies are middle aged. There are evidences that established companies have built strong reputation and hence are reluctant to default as that will tarnish their reputation and increase financing costs. New companies and start-ups are more dependent on bank funds hence avoid actions that could impair their relationship with the lender. As shown above, around 40% of the companies have been incorporated between the year 1981-1990 and the next highest with the range of 30% are those which have the year of incorporation post 1991-2000. These two mid decades cover almost 70% of the wilfully defaulted companies taken in the sample set.

4.2.2 Analysis of Companies as per their Entity type and Classification

This study has also taken up analysis of the companies as per their entity type. It was observed in most of the literature review that in respect of corporates those which have defaulted are predominantly the listed ones and are registered as a public limited company. The same has been observed in the Indian context also. Table 14 shows that more than 96% of the companies are
Public Limited Companies and as is evident from Table 15 around 72% of the sample set is listed in the stock exchanges.

Table 14: Classification of the data as per their Entity Type

<table>
<thead>
<tr>
<th>Entity Type</th>
<th>Number of Companies</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Private Limited Companies</td>
<td>12</td>
<td>4%</td>
</tr>
<tr>
<td>Public Limited Companies</td>
<td>267</td>
<td>96%</td>
</tr>
</tbody>
</table>

(Source: Developed for this thesis)

Table 15: Classification of the data as per their Company Identification

<table>
<thead>
<tr>
<th>Company Identification</th>
<th>Number of Companies</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listed Companies</td>
<td>201</td>
<td>72%</td>
</tr>
<tr>
<td>Unlisted Companies</td>
<td>78</td>
<td>28%</td>
</tr>
</tbody>
</table>

(Source: Developed for this thesis)

4.2.3 Analysis of Companies as per their Industry Group

The analysis carried out of the sampled companies as per the Industrial Group presented that the companies are spread across more than 76 Industrial groups. These Industries have been classified as per the National Industrial Classification Code (NIC Code) allotted by the Central Statistical Organisation of the Government of India for all the economic activities as per the latest available version of 2008. The details of the industry group are presented in Annexure V of the Appendix.

4.3 Hypotheses

In order to address the research questions, the following hypotheses have been framed for empirical examination:
4.3.1 Hypothesis I

Apparently bigger the size of advances of a credit institution, higher would be the non-performing assets and therefore the percentage of amount outstanding in respect of suit-filed wilful defaulters should also be high. In order to test this hypothesis H1 is proposed and is stated below as:

H1 – There is significant positive correlation between gross advances / non-performing assets and amount outstanding in respect of suit-filed wilful defaulters.

4.3.2 Hypothesis II

One major factor that influences the borrower to wilfully default is the amount of his outstanding debt. Increased leverage or high dependency on the debts should benefit the borrower more when defaulted. In order to test this hypothesis H2 is proposed and is stated below as:

H2 – High outstanding debt increases the probability of wilful default since the benefits from wilful default is more likely to exceed its implied costs.

4.3.3 Hypothesis III

It is felt that use of a predictive model on certain key annual financial indicators of the companies ex-ante, would facilitate the stakeholders to mitigate the credit risk. This would also improve the information on the probability of wilful defaults. However, the annual financial statements of the companies reflect the position as on a particular date, i.e., the date on which the financial year ends. This date in respect of Indian companies is March 31st of every year. Therefore, though the lending institutions predominantly depend upon these static variables for analysing the financial position of the companies, ex-ante and ex-post, keeping in view that funds flow is a better predictor of wilful defaults; it is proposed to explore the cash flow variables of the financial statements. These variables primarily reflect the cash and cash equivalents, from operations of the business and cash flowing from / to financing and investment activities. In order to test this hypothesis H3 is proposed and is stated below as:
H3 – Cash flow variables are better indicators of diversion and siphoning of funds and hence have the ability to increase the accuracy of the wilful default prediction model.

4.4 Research Design and Methodology

4.4.1. Sources of Data

This section focuses on the data set utilized for the estimation of our model. As seen in the foregoing paragraph, the credit institutions were required to submit the list of suit-filed accounts of wilful defaulters of Rs.25 lakh and above on a quarterly basis, i.e., as at the end of March, June, September and December to CIBIL. Further, dissemination of credit information in respect of these defaulters to the financial system was entrusted to CIBIL and accordingly such data is made available by the bureau in their website (www.cibil.com) since March 2002. However, it must be noted that the information disseminated, is as reported by the credit institutions to CIBIL.

Effort has been made to extract this data set of suit filed wilful defaulters (SF_WD) from the website of CIBIL for the period March 2002 to March 2016. The data is available credit institution-wise and State / Union Territory-wise on a quarterly basis. This is being reported by all groups of credit institutions i.e. Co-operative banks, Financial Institutions, Nationalized Banks, Foreign Banks, Private Sector Banks and SBI and its Associate banks. The wilful defaulters identified by the lending institutions represent public limited companies (both listed and unlisted), private limited companies, other associates, firms and individuals. This has been culled out and generated in a uniform format for all the quarters for necessary compilation. The data has been compiled in a structured format selecting majority of the details like name of the credit institution, the group they belong to, the branch name, State, borrower’s name, and the outstanding amount for each quarter.

The total number of records reported by all institutions put together for the entire fifteen year period is 198724 in respect of these wilful defaulters. “Record” as per CIBIL data implies the number of times the credit institutions report the outstanding of the borrowers, the periodicity of which is quarterly. The total number of credit institutions covered in this study is 115. As the financial statements of the companies pertaining to the financial services sector broadly differs
from those of the non-financial companies, all the companies from the financial services sector have been removed from the database. The relevant data of 558 companies (279 wilful defaulters and 279 non-defaulters) has been extracted from Prowess database of the Centre for Monitoring Indian Economy (CMIE). Thus the sample has been divided into two groups (Wilfully Defaulted companies (WDC) and Non-defaulted companies (NDC). The data set is a reasonable representation of the corporate sector of India comprising of more than 76 Industry groups.

4.4.2 Sample selection

The names of the companies culled out from CIBIL database were more than 2500 each of public limited companies and private limited companies. The names of these companies were run in the Prowess database of the CMIE. A thorough scanning of the database revealed that financial statements of around 700 companies are available. The financial statements of these 700 companies were critically examined for the entire period of 2002-2016. Out of these, 279 wilfully defaulted companies were selected based on leading and prominent financial variables selection criteria. As majority of the wilful defaulters’ are having consortium or multiple lending arrangements the data set has to a great extent covered the credit institutions from all groups having large outstanding in respect of wilful defaulters. Further, reporting to CIBIL is on quarterly basis but financial statements were available in prowess database annually. Hence, capturing the first time default had to be adjusted as of March every year, though in some cases institutions would have reported the company as wilful defaulter three quarters ahead.

In order to examine the predictive power of the selected annual indicators for wilful defaults of Indian corporates, both wilfully defaulted companies and non-defaulted companies are required. A sample that would contain only the wilfully defaulted companies would not suffice for this study as the financial indicators being selected needs to be applied in both the set of companies to compare the outcomes and thus classify them as per the findings. The main reason for comparing and contrasting the indications derived from both the set of companies is to effectively assess the differences between them.

As seen in majority of the studies, the “paired sample technique” has been extensively used for predicting the corporate distress or bankruptcy (Beaver, 1966; Altman, 1968; Ko, et. al., 2001;
Ahmadi, et. al., 2012; Gupta, 2014; Mraihi, 2015). On the same lines, selection of wilfully defaulted companies is also based on a “paired sample technique”. The “paired sample technique” compares a wilfully defaulted company to a non-defaulting one. This technique is being applied in this study as the sample set taken are primarily similar on all aspects except for the financial parameters.

Accordingly, in order to compare the financials of these 279 wilfully defaulted companies with non-stressed ones, another set of 279 standard non-financial companies and their annual financials were extracted from the Prowess database of CMIE from the BSE 500 companies for the fifteen year period from 2002 to 2016. The panel data taken for this study comprises of 4113 company year observations of financial indicators for the combined set of 558 companies. Thus, on an average, around 8 annual financials per company have been used in the final panel data. Among these, there are 78 company years classified as wilfully defaulted (Table 16). Identification and declaration of a borrower as “wilful” in a credit institution is a time consuming process (average 1-2 years). Hence, the quarter in which it is reported does not represent economic or the “real” event of default. As required by the binary logistic regression model, companies classified as wilful defaulter are denoted as “1” and non-defaulting companies as “0”.

Table 16: Classification of annual observations

<table>
<thead>
<tr>
<th>Number of NDC</th>
<th>Number of WDC</th>
<th>Total number of annual observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>4035</td>
<td>78</td>
<td>4113</td>
</tr>
</tbody>
</table>

Notes: This table reports summary statistics for the entire sample used in the construction of the wilful defaulter prediction model. NDC- Non-defaulted companies and WDC – Wilfully defaulted companies (Source: Developed for this thesis)

The other point of consideration which needs to be factored while taking up the data sampling, in a development of a prediction model is the investigation period which needs to be studied and tested in the model. The scrutiny of the literature reviewed indicated that there is no standard norm with regard to the period of investigation which should be definitely taken in the study of corporate distress / bankruptcy prediction. However it is observed that in majority of the studies taken up on predictions, right from the pivotal ones of Beaver (1966) and Altman (1968), the investigation
period is on an average five years preceding the distress period. Further the studies have also pointed out that the ability to predict diminishes as the period of warning time increases. Beaver (1966) and Hamer (1983) have stated that “financial ratios are useful in predicting failure as much as five years before the event”. In this regard, Hossari (2006) has explicitly pointed out that “The objective of any corporate collapse prediction model is to signal collapse before it happens. If the reporting period were too short, it would be too late to take corrective action and try to turn the company around. Likewise, if the reporting period were too long, then the prediction model might not detect any signs of impending collapse.” The study has further quantified that 22 of the 60 studies undertaken during the period from 1968 to 2004, have on an average considered a five year time period before the event as appropriate for examination. Hence, it can be safely assumed that five years norm is suitable for study on predictions and any large variation to this norm might not bring out the desired results in signalling the corporates’ financial distress / bankruptcy.

Taking above inputs into consideration, the annual financial data of the companies culled out from the Prowess database of the CMIE were analysed with the data which has been reported by the credit institutions to CIBIL on a quarterly basis in respect of such wilful defaulters. The investigation demonstrated that the process undertaken by the credit institutions from identification to reporting a borrower as wilful defaulter is time consuming. It is, therefore, observed that there is a considerable time gap of average 2-3 years, between the period when the company enters the state of financial distress and the quarter in which the first credit institution reports the borrower as a wilful defaulter to the bureau, CIBIL. Further, it has also been observed that majority of the suit-filed wilfully defaulted companies are in consortium or multiple banking arrangements. However, considerable time lag (average 2 to 3 years) was observed between identification and reporting amongst these lenders too. Therefore, it is established that in practice, the date of reporting of a borrower as a “wilful defaulter” to the bureau is not known to the officials who manage the credit risk of the lending institutions. Thus they work on the financials as given to them at the time of taking decisions for fresh or renewal credit propositions and accordingly assess the credit risk of the borrowers at that point in time.

Further, the examination of the sample set with the annual financials as available in the Prowess database of CMIE exhibits that majority of these companies do not submit their annual financial
report to the Ministry of Corporate Affairs as soon as they slip into the stressed category in the books of the credit institutions. Therefore, the financials of these wilfully defaulted companies, on an average, were not available in the Prowess database for 2-3 years before the first credit institution had reported them as wilful defaulter to CIBIL. Hence, financials as available prior to that have been identified and classified into two categories “Non-defaulted companies” as “0” and the “wilfully defaulted companies” as “1” and have been numbered chronologically in the data set.

Therefore, final estimation of the probability of wilful default is based on data which is in the band of three to four years prior to declaring them as wilful defaulters, which is as per the suggested norm in the earlier literatures on predictions. This is in alignment with the research objective to pre-empt identification of wilful defaults at an early stage.

**Data quality and completeness**

While scrutinizing the data there were certain conspicuous anomalies observed in the data reported by the institutions, for example – data reported in absolute numbers or in thousands or crores, instead of reporting in lakhs as stipulated by RBI, which have been corrected after comparing the same with the data reported during the previous and subsequent quarters.

Another limitation of the data was the manner in which the names of the wilful defaulters were reported by the credit institutions. It was observed that the same borrower has been spelt differently by different credit institutions or by different branches of the same credit institutions, during this entire period. Hence, it required multiple iterations of cleansing the data which were incorrect, incomplete or were improperly formatted. There were also inconsistencies observed in grouping the branches of these institutions under a particular State or Union Territory. For the sake of this study, these have been grouped together as per the current status of the States and Union Territories in India.

The data also contained additional details with regard to the Registered Office of the borrower’s company and their directors, which were not used in the present study. Further, the study has been taken up with the data received in the bureau up to March 2016. It is observed that few credit
institutions have reported / corrected the data pertaining to the earlier periods, which have not been factored into as the findings are not significantly altered.

4.4.3 Logistic Regression Model

The “Logistic Regression” can be defined as a “statistical tool used for analysing a dataset in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable, i.e., where there are only two possible outcomes”. Thus, the “Logistic Regression” predicts a probability that an observation falls into one of the two categories of a dichotomous dependent variable based on one or more independent variables.

The study has attempted to build a credit risk model using logistic regression to predict the probability of wilful defaulters and classify them with that of the non-defaulting companies. All the potential financial determinants of the wilful defaulters and the non-defaulting companies have been verified and tested in the logistic model. As is the case with many other default prediction models, the developed logistic model has been compared with other prominent work on prediction of distress. Thus, this study investigates empirically the utility of key financial variables that could help to mitigate the credit risk by improving information on the probability of wilful defaults. The detailed description of the model developed has been provided in chapter VI.

Logit Methodology

\[ Y_i = \begin{cases} 1 & \text{if the } i^{\text{th}} \text{ company is wilful defaulter} \\ 0 & \text{otherwise} \end{cases} \]

The probability distribution function of \( Y_i \) is given by

\[ P(Y_i = y_i) = \binom{n_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{(n_i-y_i)} \text{ for } y_i = 0,1, \ldots, n_i \]

The odds of the \( i^{\text{th}} \) firm are given by:

\[ \frac{\pi_i}{1 - \pi_i} = \exp(X_i'\beta) \]
Upon calculation for the probability \( \pi_i \) in the logit model the above equation become as

\[
\pi_i = \frac{\exp(X_i'\beta)}{1 - \exp(X_i'\beta)}
\]

The probability \( \pi_i \) varies from 0 to 1 i.e. on the probability scale the right-hand side of the above expression become a non-linear function of the predictor.

On differentiating with respect to \( x_{ij} \) we have

\[
\frac{d\pi_i}{dx_{ij}} = \beta_j \pi_i (1 - \pi_i)
\]

Thus, the effect of the \( J^{th} \) predictor on the probability \( \pi_i \) depends on the coefficient \( \beta_j \) and the value of the probability.

4.4.4 Discriminant Analysis Model

As stated by (Altman, 1968), “Multiple Discriminant Analysis (MDA) is a statistical technique used to classify an observation into one of several \textit{a priori} groupings dependent upon the observation’s individual characteristics”. It facilitates the experimenters to determine the finest set of financial ratios that would best classify the sample set. MDA is an appropriate statistical technique when the dependent variable \((y)\) is qualitative and the independent variables \((x)\) are quantitative. Thus, as inferred by (Altman, 2000), MDA is primarily applied when there is a requirement to classify and make predictions wherein the dependent variable is in a qualitative form, for example “bankrupt” or “non-bankrupt”.

The results derived from the logistic regression model have been compared with the outcome of the discriminant analysis model to check the robustness of the prediction. The details of the same are given in chapter VII.

4.5 Conclusion
The chapter began by discussing the scope of the study which is quite comprehensive and covers investigation of the suit-filed wilful defaulters’ data pertaining to India for a period of fifteen years. There is complete coverage of the data analysed geographically too as the reporting made by the credit institutions and its branches of entire India has been taken up for this study. The age-group, entity type and Industrial group classification have also been analysed under the scope of the study. After discussing the scope, three hypotheses have been framed for empirical examination in tune with the research questions. The chapter further elaborates on the sources of data. As stated, the data on wilful defaulters’ data has been extracted from the website of CIBIL and primarily the financials of the companies in the sample set have been culled out from the Prowess database of CMIE. The justification for the research design drawn which is based on the comparisons carried out on both these data sets have been discussed extensively. The basic framework of the research methodology which is being run on the logistic regression model and discriminant analysis has been briefly narrated and finally the chapter unfolds the data quality and completeness so that they would be kept in mind while drawing conclusions of the research.

As indicated at the beginning, the aim of this chapter was to discuss the scope of the research, the hypotheses framed, justification to the research design and methodology adopted and to discuss as to how the data were collected and analysed along with its limitations.

After collection of the data and framing the research methodology, a thorough understanding of the evolution of wilful defaults in India is felt pertinent. This can be achieved by critically examining the data of wilful defaulters and analysing their characteristics. Therefore, an analytical approach to comprehend the trend of the outstanding amount credit-institution wise, borrower-wise and States / Union Territory wise has been carried out in the next chapter.