

PREDICTING THE PROBABILITY OF DEFAULT OF LONG TERM INDIAN CORPORATE BONDS USING A LOGISTIC REGRESSION MODEL

Aditya Bihani, Arushi Sahay, Aliza Khan, Divyang Sharma, Kanishk Shetty, Vaishaki Chowta,
Vani Viswanathan

Third year, Bachelor's of Science in Economics, NMIMS University, Mumbai

DOI: 10.46609/IJSSER.2020.v05i11.002 URL: <https://doi.org/10.46609/IJSSER.2020.v05i11.002>

ABSTRACT

This paper aims to present methods for directly estimating corporate probability of default (PD) using financial variables. A logistic regression model is employed to directly estimate the probability of default. The financial variables used in the model have been shortlisted on the basis of literature reviewed on this topic. We look at companies, both listed and unlisted, whose bonds defaulted during FY 2018 and 2019. The results are satisfactory with four out of five explanatory variables having statistically significant coefficients and with the expected signs. Whilst work on this field has been done for markets like the US, study on Indian bond defaults is relatively scarce and thus this paper seeks to help fill this void.

Keywords: India, Bonds, Modeling, Emerging markets, Defaults

1. INTRODUCTION

Besides all the major types of risks, corporate liabilities have a significant amount of default risk. There is always a chance that a corporate borrower will not meet their contractual obligations and may renege from paying the principal and the interest due. Even for the typical high-grade borrower, this risk is there even though it may be small, perhaps 1/10 of 1 percent per year. Although these risks do not seem large, they are in fact highly significant. They can even increase quickly and with little warning. Further, the margins in corporate lending are very tight, and even small miscalculations of default risks can undermine the profitability of lending. But most importantly, many lenders are themselves borrowers, with high levels of leverage. Unexpected realizations of default risk have destabilized, decapitalized, and destroyed many internationally active lending institutions.

Prediction of default probability (PD) for each borrower or group of borrowers is a key input for the estimation of regulatory capital as well as economic capital for banks. It is also equally important for the banking industry and financial institutions to discriminate the good borrowers (non-defaulting) from the bad borrowers (defaulting). Thus, the calculation of probability of default of corporates on the basis of certain variables (in our case, financial variables) will not only help them in taking lending decisions but also practicing better pricing strategies to cover against the counter party risk. While, internationally, considerable research has been made to predict corporate default, very few attempts have been done for Emerging Markets like India. To this end, this paper attempts to estimate the probability of default (PD) for firms by making use of a logistic regression.

The rest of the paper is structured as follows. In the next section, we review literature which in turn provides us with a basis for selecting five financial variables as the explanatory variables in our logistic model. The third section discusses about the data which is followed by the fourth section, which elaborates the definitions used and the construction of variables and hypotheses. The fifth section presents the results and methodology of logistic analysis to predict corporate bond defaults. The last section discusses the main conclusions.

The codes for the model estimation as well as for the plots are given in the appendix.

2. LITERATURE REVIEW

Wilcox (1971) tries to frame a theoretical financial ratio with an improved predictability of failure so as to better explain the findings of Beaver et.al (1970) where Beaver found out that several financial ratios which were easily available, were better indicators of financial failure of a company as compared to widely used ratios like current ratio as current ratio can be manipulated by the managements of firms so as to maintain credit availability.

Based on the theoretical model used for determining the probability of failure, the study constructs a more realistic model where he considers the firms wealth to be a function of assets and liabilities and tries to find the relationship between the probability of a gain or loss and the drift rate of the firm's wealth. By using Binomial theorem, the study is able to express the probability of a company's failure in terms of its drift rate and the firm's wealth.

The ratio from the realistic model is compared with ratios found by Beaver and Tinsley. It is found that the study's results were corresponding to the ratios found by Beaver and Tinsley.

Thus, the study expects major improvement in the discriminant ability between high risk and low risk firms would come through refinements in measures of firm wealth, gains and the drift rate.

For, example the various classes of assets and liabilities, both can be estimated depending on the time horizon of the estimate. Measures of gain and drift rate might be improved by basing them on exponential weighted averages of past performance rather than on the maximum likelihood estimates from a fixed sample.

In conclusion, the study states that three components need to be focused on in order to improve total risk prediction, the variance of the cash inflows, the variance of the cash outflows and the covariance of inflows and outflows.

Scott, J. (1981) integrated the development of empirical models to predict corporate bankruptcy and several bankruptcy theories. His review shows a considerable amount of overlap between the two, although the overlap is not perfect. His study shows optimistic results – bankruptcy prediction is both empirically and theoretically explainable. The models which suggests that bankruptcy prediction is possible are not based on explicit theory and its success suggested strong underlying regularity. There are earnings and cash flow variables appear in all of the models while in some there is appearance of debt.

Largely bankruptcy-prediction models are derived using a paired-sample technique. The data in the sample is part, firms that failed and the other part is firms that do not fail. The models calculate a number of traditional financial ratios which are calculated from financial statements that were published before the failure of the firms. The researcher looks for a formula, that is either based on a single ratio or combination of ratios, that best discriminates between firms that eventually failed and firms that remained solvent. The implication of the empirical work in the paper make it possible to improve the predictive accuracy by using the variables and functional forms suggested by bankruptcy theory. And can also determine which bankruptcy theories are most powerful.

The author mentions many ideas, while two of them describe the best results available. First of all, the gambler's ruin model assumes that the company is experiencing losses on the sale of assets. On the other side of the scope, the perfect-access model assumes that while the firm may incur losses in the sale of assets, it may sell either debt or equity in a more efficient, uncomplicated security market. Scott then went on to present examples from the category of dissent between the gambler's ruin and the perfect-access models. In these assumptions, firms have limited access to security markets, and the bankruptcy forecast that has resulted in both the liquidation or accounting equity of the shareholder and the market value of the equity.

The point made by the author is that understanding the determinants of corporate bankruptcy is important whatever the size of the bankruptcy costs. There are numerous practical examples of this importance. Bankruptcy prediction models can help auditors decide whether or not a corporation is a 'going concern'. Understanding corporate bankruptcy is important in theory as

well. For example, under risk neutrality, the probability of bankruptcy is an important parameter in valuation formulas for debt and equity.

In conclusion, the study states that most of the other studies use the term bankruptcy by it is somewhat misleading, most of the firms in theory are dealing with the failure to meet its financial obligations and these failures don't always lead to bankruptcy. It is seen that creditors most times forgive small failures of payment rather than bear large bankruptcy costs.

Gentry, Newbold and Whitford (1985) studied a cash-based funds flow model developed in 1972 by Helfert to test if such a model can adequately classify failed and non-failed companies and serve as an alternative to financial ratios in accrual accounting. In the model developed by Helfert and redesigned in the paper, eight major components were used- Funds from Operations, Working Capital, Financial, Fixed Coverage Expenses, Capital Expenditures, Dividends, Other Asset and Liability Flows, and the Change in Cash and Marketable Securities. For the analysis, MDA, probit and logit techniques were used. Only logit results were reported as it gave the best results. However, use of probit or MDA did not alter the results. They also made use of a secondary sample of 23 companies rated financially weak to test the abilities of the model and matched them to 23 non-weak companies of same industry and same asset sizes. S&P's Compustat 1981 Industrial Annual Research File and Compustat Industrial Files were used to determine failed companies between 1970 & '81. Other sources included F&S Index of Corporate Change, Financial Stock Guide Service and Wall Street Journal Index. Secondary data was taken from the CreditWatch list by Wells Fargo Bank and various financial services.

The data showed a significant difference between the means of failed and non-failed companies. The standard deviations of components for failed companies were larger compared to non-failed companies. Aside from the dividend component, the distributions of the components for the failed components were broadly disbursed across a wide range of possible outcomes. Distribution of dividends were skewed to the left for both failed and non-failed companies. Not-failed ranged from - 0 to -40%, where six companies paid no dividend. Meanwhile, failed companies ranged from - 0 to -25%, and 23 paid no dividends. The logit model classified 77 to 83% companies correctly.

All three models show similar results for misclassified companies. Dividend component was significant at 5% for both time periods. Smaller the dividend component, the higher the probability of failure.

For the secondary data, using one-year and three-year mean funds flow information, 70% and 78% respectively of weak companies were correctly classified while 74% and 70% respectively for non-weak were accurate.

The study concluded dividend funds flow component in a logit model was significant in distinguishing between failed and non-failed companies. The smaller the dividend component, the higher the probability of failure. A firm will experience a shortfall of funds from operations and, subsequently, observe a reduction in dividend payments. Logit findings suggest that cash flow from operations do not improve the classification results of failed and non-failed companies. This may be due to larger variances of CFO for both kinds of companies.

Casey and Bartczak (1985) assess whether operating cash flow and other relevant measures lead to more accurate predictions of bankrupt and non-bankrupt firms. The focus of the study is on the marginal predictive content of the operating cash flow ratios.

Sixty firms were selected that had petitioned for bankruptcy during the period 1971-82. The bankrupt firms were a subset of the 105 failed firms used in a previous study (Ohlson [1980]) and included firms listed by Dun and Bradstreet and the Wall Street Journal Index. A sample of 230 non-failed firms was chosen from the Compustat Industrial Tape. The financial data for the non-failed firms were taken from time periods contemporaneous with the failed firm.

The operating cash flow variables examined in this study were CFO, as previously defined, CFO divided by current liabilities, or CFCL, and CFO divided by total liabilities, or CFTL. Multivariate models (described below) that did not include operating cash flow ratios were employed as convenient standards for assessing the marginal discriminatory power of the operating cash flow variables. These models are identical to the accrual-based models employed in our previous study in which the related classification accuracy proved significantly greater than accuracy based on the univariate operating cash flow ratio. Eight MDAs were run using the Cooley and Lohnes [1971] program. Classifications were performed using a chi-square procedure, as well as one based on a Euclidean distance measure. The classification results were essentially insensitive to the choice of the classification procedure, so only the chi-square classifications are reported. The eight MDA models were run using the six accrual-based ratios alone and using the six ratios plus one or more of the operating cash flow ratios. For the first year prior to failure, the average annual first differences in the cash flow ratios for the preceding four years were also included. All MDA models were statistically significant ($\alpha < .05$) for the first three years prior to bankruptcy. The univariate F-ratios for the operating cash flow variables did not exhibit a consistent pattern. CFCL was statistically significant ($\alpha < .05$) for the first three years, CFTL for the first two years, and CFO for the first, fourth, and fifth years prior to bankruptcy. The group means for CFO, CFCL, and CFTL are displayed in figure 1. None of the differences between group means of the average annual first differences in the operating cash flow variables was statistically significant. Results suggest that operating cash flow data do not provide incremental predictive power over accrual-based ratios.

Zavgren (1985) uses the logit technique to develop and test a new bankruptcy model which enumerates the signs of financial ill health for a five-year period prior to failure. The paper also develops a methodology for evaluating the significance of probabilities of financial risk.

The importance of individual variables is realistically assessable in this study for the first time. The variables included in the models are determined empirically so as not to omit any important financial attributes. The results obtained highlight the dimensions of financial data to which researchers and practitioners should devote attention should they wish to understand the impact of managerial activities on the firm's financial risk.

The profitability measure proves insignificant in any year, while the turnover ratios are significant for the long run. The significance of the liquidity measure in earlier years and the negative sign of its coefficient indicates that the most successful firms devote their resources to productive capital rather than liquid assets. The acid test ratio proved to be highly significant with a negative coefficient in the first three years prior to failure. This points to the ability to meet maturing obligations as an important factor in avoiding bankruptcy. The ratio of long term debt to invested capital was found to be highly significant in all years because the most reliable indication of a firm's health or lack of health is its use of debt.

The models proved highly significant with reference to both the R2 and likelihood ratio tests in detecting ailing firms up to five years prior to their failure. By applying the entropy concept to the probabilities from the models, a methodology is developed which objectively measures the quantity of information in these signals. The information available from these models is significant even five years prior to failure, and it increases up to the year immediately prior to failure. This indicates a decrease in uncertainty over the fate of these firms over this lead time.

Opler & Titman (1994) examined the relationship between financial distress and corporate performance by identifying industries that had experienced economic distress and investigating whether firms in those industries with high financial leverage prior to the distressed period fared differently than their more conservatively financed counterparts. It is expected that if financial distress is costly, then more highly leveraged firms will have the greatest operating difficulties in a downturn. Alternatively, financial distress benefits firms by focusing efficient operating changes then more highly leveraged firms will perform better than less leveraged firms.

Sales growth, stock returns and changes in operating income relative to industry averages were used to measure firm performance by the authors. Sales growth as a measure of firm performance was used because it is the most direct measure of consumer driven losses in sales. But the authors also wanted to estimate the extent to which losses in sales translate into lost profits and value which is why stock returns and operating income were also used. The authors made use of form level

data from the 1992 Standard and Poor's COMPUSTAT PST, FC and research files. These files contain 105074 firm-years of data on income statement and balance sheet items in the 1972 to 1991 period. After excluding certain firms and industries due to non-feasibility and insufficiency of data, 46799 firm-years of data was retained for empirical analysis.

The authors then made use of OLS regression to predict firm level-sales growth, stock returns and profitability growth as a function of size and profitability controls, industry condition and ex ante leverage ratios. Each measure of firm performance was regressed on log of sales, industry-adjusted profitability, industry-adjusted investment/assets, industry-adjusted assets sale rate, distressed industry dummy, high leverage dummy and interaction of distressed industry dummy and high leverage dummy. It was found that the coefficient estimates of the average effect of higher leverage and performance in all industries and the coefficient estimates of the additional effect of leverage on performance in periods of industry distress were negative and statistically significant in all cases. Moreover, a negative relationship was found between sales growth and high leveraged firms, suggesting that highly leveraged firms lose market share to their more conservatively financed counterparts even during normal times. This implies that firms anticipating sales growth buildup financial slack, in order to fund increased investment, rather than increase leverage to signal their prospects. It was also found that the coefficient for the high leverage and distressed industry interaction dummy was negative and statistically significant. Industry adjusted sales growth was found to be 13.6% lower for highly leveraged firms in distress industries than for less leveraged firms. This indicates that leveraged firms lose significant market share during economic downturns. For operating income as a measure of financial performance, the coefficients were not statistically significant, therefore, results remain inconclusive. In the case of stock return regressions, the coefficient of that leverage interaction variable is statistically and economically significant and negative, therefore, in keeping with what was expected.

Lennox (1999) studied a sample of 949 UK listed companies between 1987 to 1994 to identify causes of bankruptcy. The author identified certain problems with the DA approach which make the probit and logit models more suitable to the study. Under DA, explanatory variables are assumed to have multivariate normal distribution while the variables used in bankruptcy studies are rarely normally distributed. The matched pair technique employed in these studies also violates the assumption of a randomly chosen sample. Moreover, DA studies employ a linear classification rule. Thus, probit and logit models are better suited for bankruptcy studies. The study then employs a logit model and a homoscedastic probit model.

The data for the study was taken from Datastream and Stock Exchange Yearbook. After checking for heteroscedasticity in all models and accounting for non-linearity of gross-cash flows (GCF) and leverage (CAPG) variables, the final logit and probit models are used for comparison. The

LM1 test reveals no problems with omitted variable bias in the models. Once the non-linear effects of cashflows and leverage have been taken care of, null hypothesis of homoscedasticity can no longer be rejected in the models. It was seen there were very little difference in the logit and probit models.

The results showed that bankruptcy was more likely when the economy moved from a boom to a recession. A company is also less likely to go bankrupt in the future if the economy is currently in a recession. CBI indicator showed a negative coefficient indicating that improvement in business confidence is correlated with a fall in probability of bankruptcy. A company is also less likely to go bankrupt if economy is expected to recover in 12-18 months shown by the signs on CBI and F. the coefficient of number of employees is negative indicating that a smaller company is more likely to go bankrupt. Chances of failure are higher for companies in Construction or financial services. Cashflow difficulties, shown by debt turnover ratio and cash ratio and return on capital are also negatively related to probability of bankruptcy.

Well-specified non-linear probit models and logit models were superior in predictive accuracy compared to DA models.

Westgaard & Wijst (2001) studied the expected default frequency for Norwegian limited liability companies with total assets more than NOK 100,000 over the period 1995-1999. The financial ratios included in the study are PBITD/Debt, financial coverage, current ratio, and equity to total capital ratio. Other non-financial factors like age and size of the company, region, and industry were also included. As the objective was to estimate default probabilities, a logit model was employed. The results indicate that there is a strong relationship between bankruptcy and the regressors in the model. All the financial ratios were found to be negative and significant. The odds ratio indicated that the predicted odds of default in the real estates and services industry are

0.386 times that of other industries. Additionally, the model is able to discriminate between the bankrupt and non-bankrupt firms, as the average default rate is 11 times higher for defaulting companies than non-defaulters. Hence, the study suggests that that Norwegian banks should directly estimate expected default frequencies to minimize associated credit risks.

Grunert et al. (2005) hypothesized that the inclusion of a combination of financial and non-financial factors will result in a more precise prediction of default compared to a model which involves only either of the factors. Banks assign internal credit ratings to their borrowers, and these ratings have gained momentum since 1990s. There has been considerable research on assessing the borrower's insolvency probability based on financial factors, however there has been no research on the effect of non-financial factors like management quality and industry perspectives. The study is the first to explore the role of non-financial factors on internal credit

ratings.

The data consisted of two samples; the first constituted credit file information from six major German banks across 1992-1996 for 240 borrowers, and the second included borrowers in financial distress over the same time period. To ensure comparability of ratings across banks, a meta rating scale with grades 1-6 was created, with grade 1 meaning very high and grade 6 meaning highly distressed or defaulted. The study observed the borrowers financial, non-financial, and overall rating along with their default status in the succeeding year. The financial factors comprised of logarithm of total assets, equity-to-assets ratio, current ratio, cash flow-to-net liabilities, capital intensity ratio, and return on assets. The non-financial factors comprised of management quality and market position.

The study estimated three probit regression models in which the default status was the dependent variable and the explanatory variables were the financial rating in model 1, non-financial rating in model 2, and both factors in model 3. Sampling weights are used to correct for potentially biased coefficients. Additionally, the study utilized McFadden's R^2 , Brier Score, type I and type II error rates, and percentage of correctly classified observations as the evaluation criteria. The regression results yielded positive and significant coefficients of all the rating variables, as was expected. Overall, model 3 is found to be a better model. The type I error rate is lower in model 3 than in model 1. On employing a bootstrap methodology to overcome the problem of biased probit estimates, results found that for model three, 997 out of 1000 cases had a better goodness of fit, which directly supports the hypothesis of the study. Additionally, model 3 showed better McFadden's R^2 value and Brier Score than the other two models. Hence, a 'mixed' model comprising of both financial and non-financial factors leads to a more precise prediction of the default statuses.

3. DATA

The information on defaulted and solvent firms is collected from India's Research and Ratings agency (a Fitch Group subsidiary) which releases annual data on long term corporate bonds which have defaulted. We use data for fiscal years 2018 and 2019. The list released by the agency consists of both listed and non-listed companies which have defaulted on debt for the said years. These companies are then matched by their asset size, year, and industry affiliation. Thus via random selection, we isolate a sample of 43 solvent firms and 33 defaulted firms. As far as defaulted firms are concerned, a majority of them were unlisted and data on unlisted companies is difficult to come across, as such we have had to stock a sample size of 33 defaulted firms. As far as non-defaulting firms were concerned, via random selection from a set of listed and unlisted companies we picked 43 companies which belong to the diversified manufacturing sector and have an asset size between

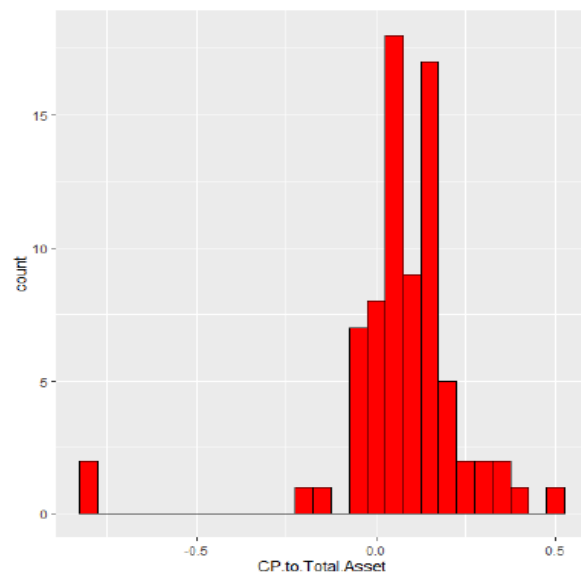
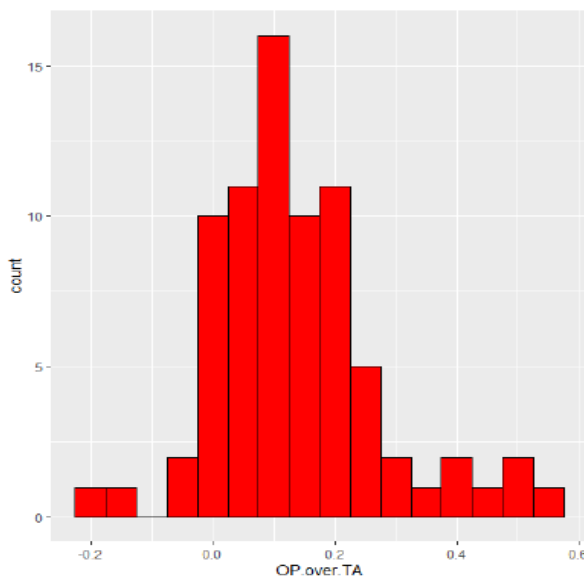
the maximum and minimum assets size of the defaulted company sample that we have.

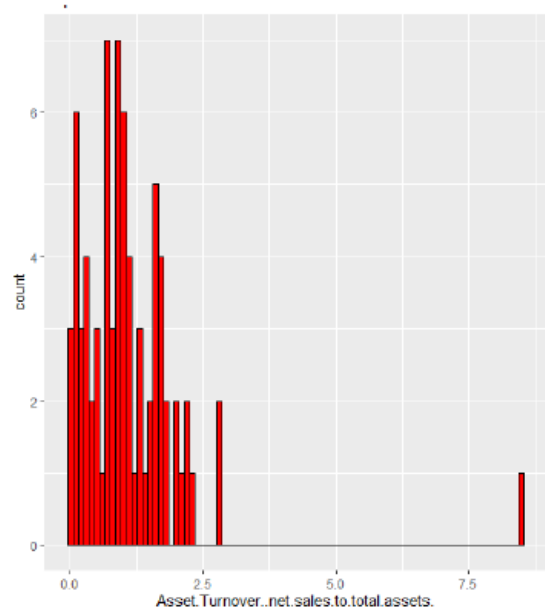
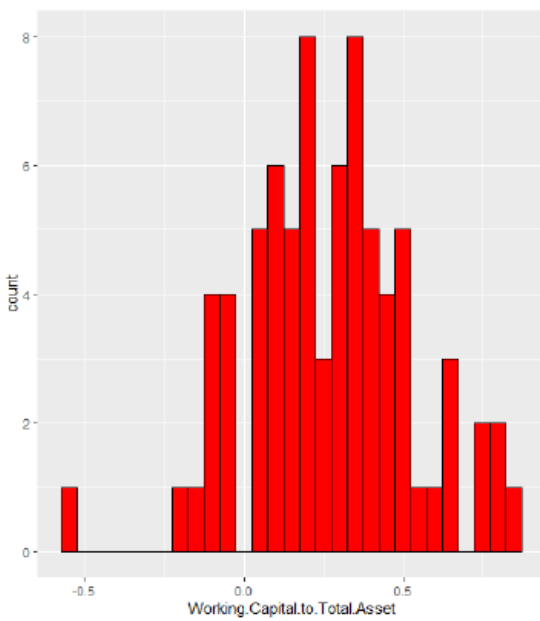
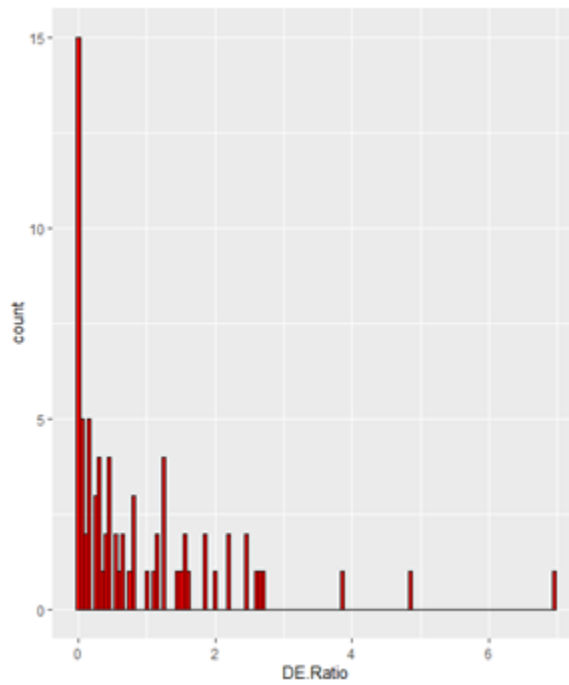
The financial information of these 77 companies over the five years prior to default is obtained from CAPITALINE plus database, which contains information on all listed companies, as well as on a few unlisted companies. CAPITALINE plus is a premium financial data repository.

Initially we had a sample of around 40 defaulted firms for which data was readily available on CAPITALINE plus, however a few of these firms had extreme outlier values which tend to results in warning messages on R, when the logic model is run. For instance, while the debt equity ratio can range between a range of +5 to -5 in extreme cases, one of the firms had a 33 debt to equity ratio. Such extreme values prevent the logic model to run smoothly on R, and as a consequence we had removed such companies from our sample. Thus, after such data cleaning, we were left with a sample of 33 defaulted companies, along with which we randomly selected 43 non-defaulted companies as per the criteria mentioned earlier.

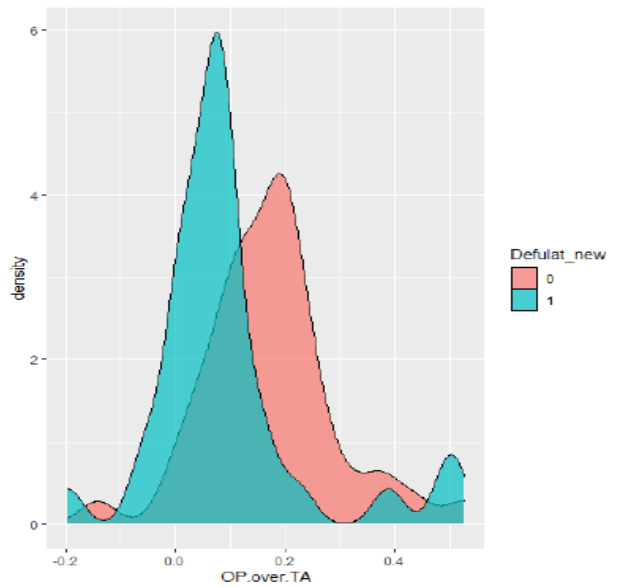
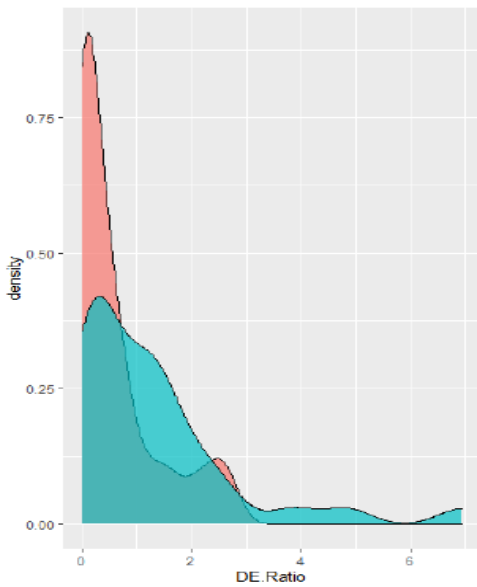
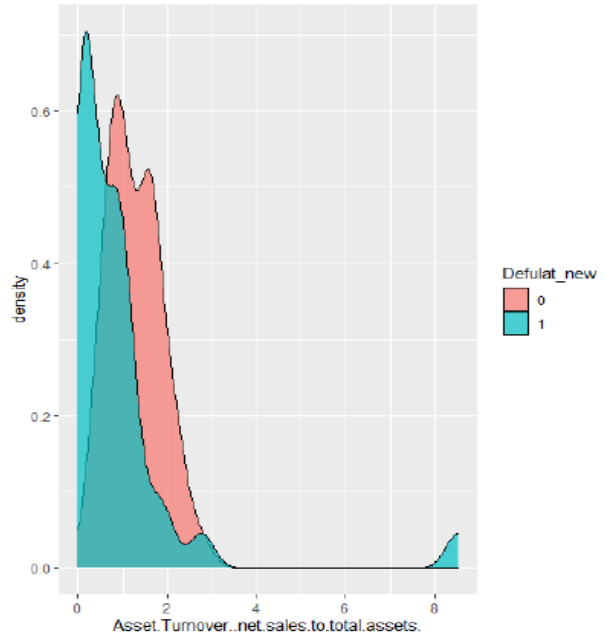
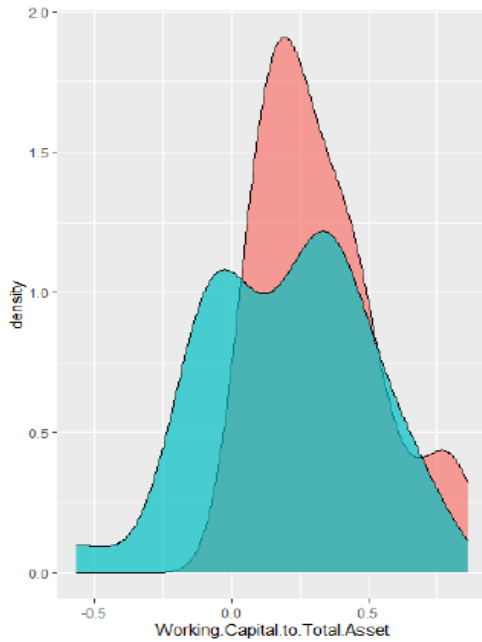
In order to have a quick glance at the data we have plotted a histogram for each variable and have also plotted a Y on X plot for each to get an idea of the data for default and non-default separately.

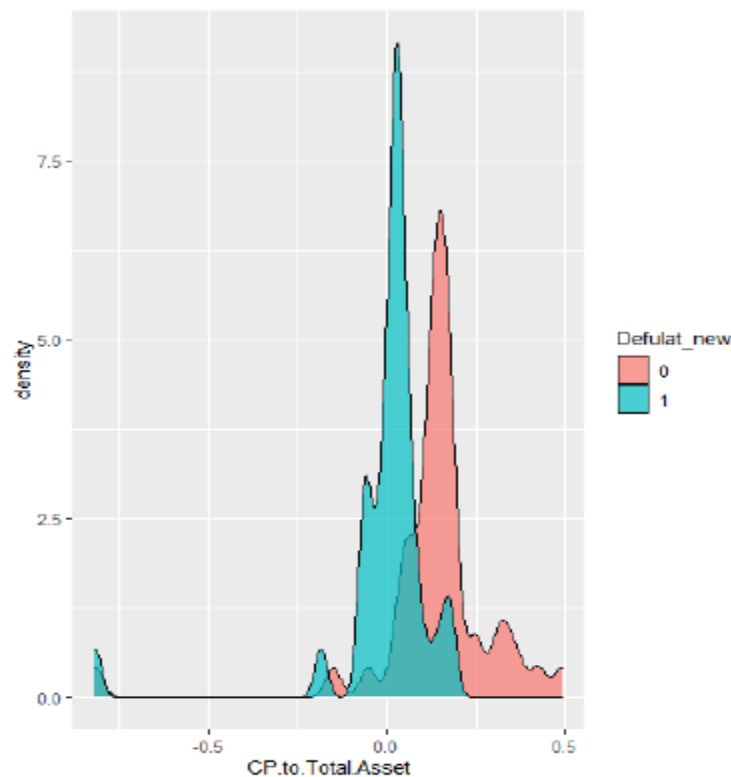
Plotting a histogram for each explanatory variable:





Plotting Y on X to get an idea of the data for default and non-default separately:





4. VARIABLE DEFINITIONS AND HYPOTHESES

The variables which should be used as regressors in the logit model should be able to relate the properties of the cash flow in combination with the debt obligations and the movement in the asset value of the firm. As was mentioned earlier, the variables are selected on the basis of the literature review conducted.

Bandyopadhyay (2006) uses an MDA technique to successfully draw the best combination of key ratios from a large set of financial ratios, which help explain risk of default. The paper initially started with many financial ratios and finally arrived at five key ratios (in the MDA model used) that best discriminate between a sample of good and bad firms. In order to pick up the best ratios, the author looked at i) F and Wilk's Lambda statistics to check the statistical significance of each individual ratio, including determination of relative magnitude of each independent variable (i.e. standardized values of their coefficients), ii) Within sample discriminatory power of these ratios' best combinations, iii) Chi-square statistic as check for the overall significance of various discriminant functions and, iv) the authors own analytical judgment.

The five key ratios that Bandyopadhyay (2006) narrowed down on were i) Working Capital to total assets, ii) Cash Profit to total assets, iii) Solvency Ratio (Debt-Equity Ratio), iv) Operating Profit over total assets, and, v) Asset Turnover Ratio.

Our logit model makes use of these five variables as well.

The definition and hypothesis with regards to these financial ratios is explained below:

1. Working capital over total assets - This a measure of the net liquid assets of the firm relative to the total capitalization. Working capital is defined as the difference between current assets and current liabilities. Hence the ratio is a proxy for the short-term liquidity condition of the firm. A firm which has defaulted or rather is about to default will tend to have lower (possibly negative) values for this financial ratio.

2. Cash profit over total assets - This is a measure of cash flow of the firm. Cash profit is obtained by adding the non-cash charges such as depreciation and amortization to the profit after tax (or net profit).

A firm which has defaulted or rather is about to default will tend to have lower (possibly negative) values for this financial ratio.

3. Solvency ratio (in our case the debt-equity ratio) - This ratio judges the long-term solvency of a firm. The debt-to-equity (D/E) ratio is calculated by dividing a company's total liabilities by its shareholder equity.

Lower the debt equity ratio, better is the ability of the firm to meet key term obligations and lower will be the probability of default. In other words, the higher is the ratio, the worse it is for the company, thus a firm which has defaulted or rather is about to default will tend to have higher values for this financial ratio.

4. Operating profit over total assets - This is a measure of the true productivity of the firm's assets. It measures the firm's earning capability.

The higher is the ratio, the better for the company, thus a firm which has defaulted or rather is about to default will tend to have lower (possibly negative) values for this financial ratio.

5. Asset Turnover Ratio - This is a standard financial ratio (also used by Altman in his original 1968 model) illustrating the sales generating ability of the firm's assets. It is the ratio of total net sales to total assets. This ratio gives an indication as to how efficiently a company is utilizing its assets.

The higher is the ratio, the better for the company, thus a firm which has defaulted or rather is about to default will tend to have lower (possibly negative) values for this financial ratio.

5. LOGIT MODEL : ESTIMATION PROCEDURE

In this paper we conduct a Logistic regression analysis to investigate the relationship between binary or ordinal response probability and explanatory variables. The method used to calculate the estimates is the Maximum Likelihood Method. To do the same we use R program. The glm package in R is used to run the logit model. The code for the same is in the appendix.

The logit equation we have estimated takes the following form:

$$PD = F(Z)$$

$$= \frac{1}{1+e^{-z}} = 1/1 + e^{-(b_0+b_1X_1+b_2X_2+b_3X_3+\dots+b_kX)}$$

Where, F(Z) is the cumulative logistic distribution.

The dependent variable is the binary variable ‘Default’ which takes the value of ‘1’ if a company has defaulted and takes the value of ‘0’ if the company has not defaulted. The explanatory variables are all quantitative in nature and are the five financial ratios which are mentioned in the previous section. In the logit regression, our purpose is to evaluate the role of balance sheet variables (financial variables) in predicting corporate bond default and to arrive at an estimate of probability of default for a firm using them.

Before we discuss our results, let us first look at the descriptive statistics of the variables used in the logistic regressions. Table I, gives us some descriptive statistics about the sample of firms used in the logistic analysis. It is evident from the descriptive statistics table that all the financial ratios for solvent group of firms on average look relatively better than their defaulted counterparts.

Table I : Descriptive Statistics of Financial Ratios for Companies that Defaulted

	Defaulted	n = 33	
	Mean	Median	SD
Working Capital to Total Asset	0.196146667	0.2181	0.2905902
CP to Total Asset	-0.003211782	0.0276	0.1621398

DE Ratio	1.285151515	1.02	1.5138075
OP over TA	0.093419476	0.07301	0.1413655
Asset Turnover (net sales to total assets)	0.88646603	0.5326	1.5094764

Table II : Descriptive Statistics of Financial Ratios for Companies that did not Default

	Non - Defaulted	n = 43	
	Mean	Median	SD
Working Capital to Total Asset	0.323733097	0.281	0.2187349
CP to Total Asset	0.135694904	0.15	0.1880909
DE Ratio	0.589534884	0.27	0.7809486
OP over TA	0.173184194	0.168	0.1200147
Asset Turnover (net sales to total assets)	1.266930775	1.104	0.5846742

Table III : t statistic for Financial Ratios

	t-stat for diff
Working Capital to Total Asset	-2.105621053
CP to Total Asset	-3.451809977
DE Ratio	2.405464979
OP over TA	-2.600880483

Asset Turnover (net sales to total assets)	-1.371136407
--	--------------

Furthermore, a simple t-stat test to see if the differences in the values of the financial ratios is statistically significant is conducted. The t-stat values are given in the table above. The value for critical t is 2.0369 for a 95% confidence interval at 32 degrees of freedom (we take 32 df, since in a conservative approach the df would be one less than the sample size of the smaller sample). As is seen, the t-stat values for all the financial ratios except for the asset turnover is greater than the critical t value, thus indicating that the statistics for the solvent firms are statistically different from the defaulted ones.

Below we have correlation matrix, which was attained via R. This was calculated in order to an idea of the existence of multicollinearity before running the logit estimation. As is evident, none of the variables show an excessive correlation, thus multicollinearity is not an issue.

Table IV: Correlation Matrix

	Working Capital to Total Asset	CP to Total Asset	DE Ratio	OP over TA	Asset Turnover (net sales to total assets)
Working Capital to Total Asset	1	0.29006	0.09187	0.02345	0.09114
CP to Total Asset	0.29006	1	-0.08702	0.19171	0.13742
DE Ratio	0.09187	-0.08702	1	-0.13926	-0.23895
OP over TA	0.02345	0.19171	-0.13926	1	0.07036
Asset Turnover (net sales to total assets)	0.09114	0.13742	-0.23895	0.07036	1

6. LOGIT MODEL : ESTIMATION RESULTS

The logit model is run with the binary dependent variable of Default on five financial ratios as regressors as has been explained in the previous sections. The code used to run the same in R is as follows:

```
Model <- glm (Default_new ~ OP.over.TA+ DE.Ratio+ Working.Capital.to.Total.Asset +
CP.to.Total.Asset + Asset.Turnover..net.sales.to.total.assets. , data = data, family =
binomial('logit'))
```

After running the model, a Likelihood ratio test was also run on the same model. The code for the same is: lrtest (Model)

The results obtained on R are shown in the figures below.

```
Coefficients:
                                Estimate Std. Error z value Pr(>|z|)
(Intercept)                    0.9418    0.6928    1.359    0.1740
OP.over.TA                      -4.3859    2.5223   -1.739    0.0821 .
DE.Ratio                        0.5412    0.3147    1.720    0.0855 .
Working.Capital.to.Total.Asset -2.2296    1.2848   -1.735    0.0827 .
CP.to.Total.Asset               -4.8298    2.5802   -1.872    0.0612 .
Asset.Turnover..net.sales.to.total.assets. -0.1708    0.2371   -0.720    0.4713
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 104.039  on 75  degrees of freedom
Residual deviance:  78.767  on 70  degrees of freedom
AIC: 90.767

Number of Fisher Scoring iterations: 5
```

Figure 1: Results from Logit Regression

```
Likelihood ratio test

Model 1: Defulat_new ~ OP.over.TA + DE.Ratio + Working.Capital.to.Total.Asset +
CP.to.Total.Asset + Asset.Turnover..net.sales.to.total.assets.
Model 2: Defulat_new ~ 1
#Df  LogLik Df  Chisq Pr(>Chisq)
1    6 -39.384
2    1 -52.019 -5 25.271  0.0001235 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 2: Results from Likelihood ratio test

Figure 1 above shows that the financial ratios working capital over total assets, CP to Total Asset, OP over Ttotal assets, and, DE Ratio are statistically significant as the 10% confidence interval. The former three have a negative coefficient and the last has a positive coefficient as was hypothesized and expected. Furthermore, the LR test run shows that the model as a whole is statistically significant as can be seen in the Figure 2 above (in other words, the hypothesis that the joint effect of the variables is zero can be rejected).

Interpreting the coefficients obtained from the logit regression:

If the value of Working Capital over TA increases by one unit, then the ln (odds ratio in favour of defaulting) reduces by 2.2296. If the CP to TA increases by one unit, then the ln (odds ratio in favour of defaulting) reduces by 4.8298. If the OP over TA increase by one unit, then the ln (odds ratio in favour of defaulting) reduces by 4.3859. Finally, if the DE ratio increases by the one unit then the ln (odds ratio in favour of defaulting) increases by 0.5412.

In order to interpret these results better, we take the anti-log of the estimated coefficients, which then gives the factor change of the odds ratio. The anti-log values are given below.

Table V: Anti-log Values of Coefficients

	Coefficients	Anti-log of Coeff.	% Change
OP over TA	-4.3859	0.012451677	-98.75%
DE Ratio	0.5412	1.718067306	71.81%
Working Capital to Total Asset	-2.2296	0.10757145	-89.24%
CP to Total Asset	-4.8298	0.007988119	-99.20%
Asset Turnover (net sales to total assets)	-0.1708	0.842990155	-15.70%

As can be seen from the given table, if the value of OP over TA increases by one unit, then the odds ratio in favour of defaulting reduces by 98.75%. If the Working Capital to TA increases by one unit, then the odds ratio in favour of defaulting reduces by 89.24%. If the CP over TA increase

by one unit, then the odds ratio in favour of defaulting reduces by 99.20%. Finally, if the DE ratio increases by the one unit then the odds ratio in favour of defaulting increases by 71.81%.

However, since the values of such ratio don't tend to change by one unit entirely, thus to get a more realistic interpretation we see how the odd ratio in favour of defaulting changes when each of these ratios change by 0.1 units (which is a more realistic amount by which such ratios tend to change on a yearly basis). The table below shows the percentage changes for the same.

Table VI: Percentage Change in the odds ratio with a 0.1 change in financial ratios

	Amount of change in the fin ratio	% change in odds ratio in favour of defaulting
OP over TA	0.1	-9.88%
DE Ratio	0.1	7.18%
Working Capital to Total Assets	0.1	-8.92%
CP to Total Assets	0.1	-9.92%
Asset Turnover (net sales to total assets)	0.1	-1.57%

Thus, it is evident that if the OP over TA ratio increases by 0.1, then our model predicts that the odds in favour of a company defaulting on its long-term corporate bond reduces by 9.88%. Similarly, if the Working Capital to TA increases by 0.1 unit, then the odds ratio in favour of defaulting reduces by 8.92%. If the CP over TA increase by 0.1 unit, then the odds ratio in favour of defaulting reduces by 9.92%. Finally, if the DE ratio increases by the one unit then the odds ratio in favour of defaulting increases by 7.18%.

The VIF Test

After running the model, a VIF test was conducted using the car package in R, and as is evident from the results in the figure below, multicollinearity is not an issue between the variables in our model.

Code for VIF: car::vif(model)

OP. over. TA	DE. Ratio	Working. Capital. to. Total. Asset
1.053093	1.077722	1.067528
CP. to. Total. Asset	Asset. Turnover. . net. sales. to. total. assets.	
1.043168	1.039664	

Figure 3: VIF Test Results

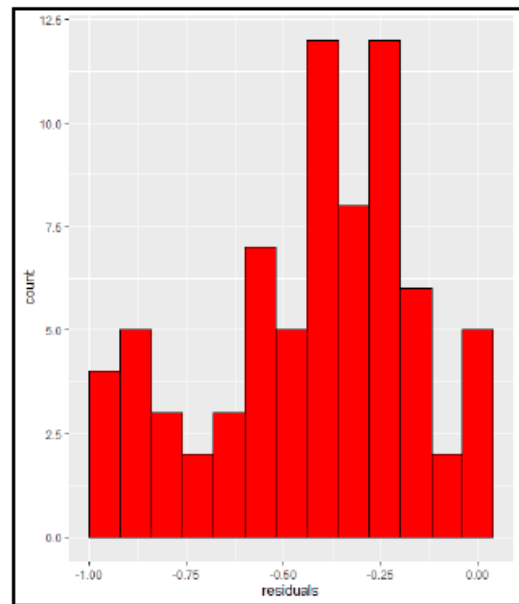
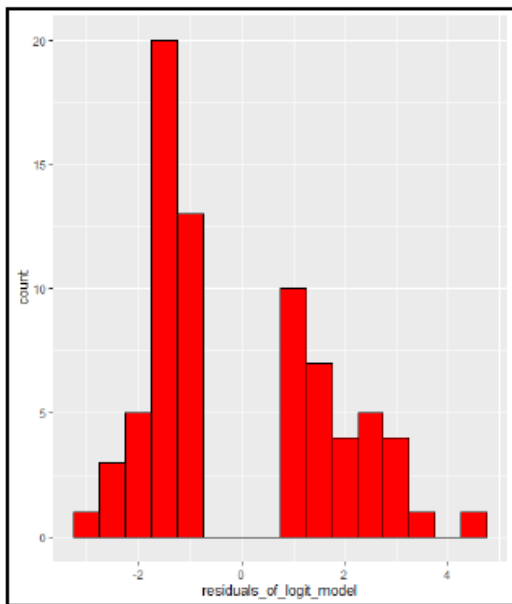
Residuals Patterns and Plots

The residuals of the model and the predicted Y values are plotted below in order to get an idea about their distribution.

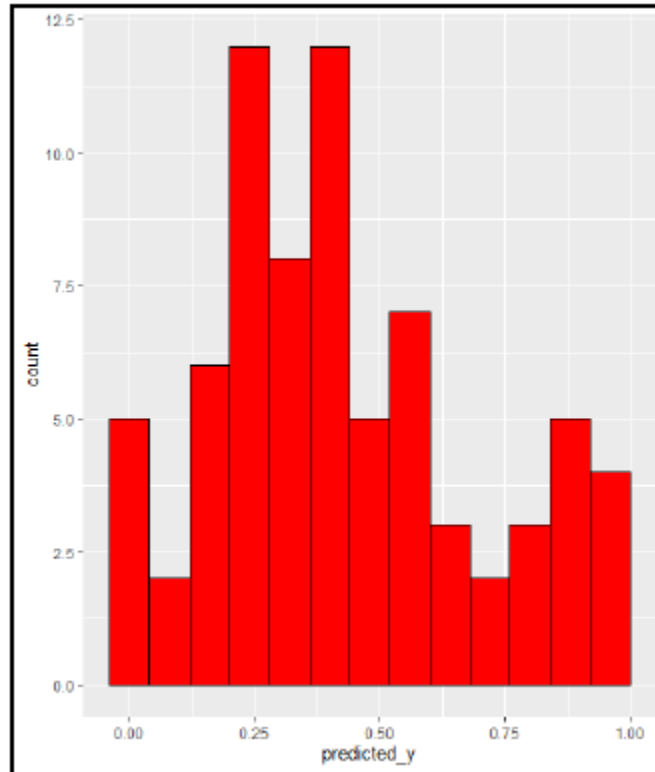
Residuals are of two types:

1) the direct residuals of the logit estimation model, values

2) the difference between actual Y and predicted Y



Predicted Y Values Plot:



7. CONCLUSION

Using a sample of 77 listed and unlisted corporations, we have developed a logit model which predicts the probability of default of a corporate bond of a firm given the values of certain financial ratios. The results are satisfactory and all the coefficients have the expected signs.

The logit results show that probability of default is a decreasing function of i) cash profit over total assets, ii) working capital to assets, iii) total sales relative to total assets, and, is an increasing function of the debt to equity ratio. The coefficient for the variable operating profit to total assets was found to be statistically insignificant.

Our analysis however excludes the use of non-financial ratios. This is a major limitation of our model and presents as a scope for future research. Ratios which can be accounted for in future research are-

1. Age of the firm - A relatively young firm will probably show a low retained earnings/total assets (RE/TA) ratio because it has not had time to build up its cumulative profits (Altman,

2000).

2. Group ownership - Studies covering various countries have found that firms associated with top business groups have greater stability in the cash flows and show better productivity as well as risk sharing than unaffiliated firms (Gangopadhyay et al., 2001).

3. ISO Quality Certification (ISOD) - This dummy is taken as a product market signal about the firm that it maintains a quality management system and is concerned with customer expectations and satisfactions. It has been empirically observed that ISO certified firms are successful in the product market (Bandyopadhyay and Das, 2005).

REFERENCES

Arindam Bandyopadhyay (2006), "Predicting probability of default of Indian corporate bonds: logistic and Z-score model approaches", *Journal of Risk Finance* Vol. 7 No. 3, 2006 pp. 255-272.

Casey, C. and Bartczak, N. (1985), "Using operating cash flow to predict financial distress: some extensions", *Journal of Accounting Research*, Vol. 23 No. 1, pp. 384-401.

Gentry, J.A., Newbold, P. and Whitford, D.T. (1985), "Classifying bankrupt firms with funds flow components", *Journal of Accounting Research*, Vol. 23 No. 1, pp. 146-59.

Grunert, J., Norden, L. and Weber, M. (2005), "The role of non-financial factors in internal credit ratings", *Journal of Banking and Finance*, Vol. 29, pp. 509-31.

Lennox, C. (1999), "Identifying failing companies: a re-evaluation of the logit, probit, and DA approaches", *Journal of Economics and Business*, Vol. 51, pp. 347-64.

Opler, T.C. and Titman, S. (1994), "Financial distress and corporate performance", *Journal of Finance*, Vol. 49, pp. 1015-40.

Scott, J. (1981), "The probability of bankruptcy: a comparison of empirical predictions and theoretical models", *Journal of Banking and Finance*, Vol. 5, pp. 317-44.

Westgaard, S., & Wijst, N. v. (2001). Default probabilities in a corporate bank portfolio: A logistic model approach. *European Journal of Operational Research*, 338-349.

Wilcox, J.W. (1971), "A simple theory of financial ratios as predictors of failure", *Journal of Accounting Research*, Vol. 9 No. 2, pp. 389-95.

Zavgren, C.V. (1985), "Assessing the vulnerability to failure of American industrial firms: a logistic analysis", *Journal of Business Finance and Accounting*, Vol. 12 No. 1, pp. 19-45.

APPENDIX

Codes Used in R:

Importing Data

```
data <- read.csv("Eco Project 3.csv")
```

```
str(data)
```

```
data$Defulat_new <- as.factor(data$Defulat)
```

#Plotting the diff variables and plotting X on Y

```
ggplot(data=data,aes(x=Working.Capital.to.Total.Asset))+geom_histogram(binwidth=0.05,color="black", fill="red") summary(data_1$age)
```

```
ggplot(data = data, aes(x=CP.to.Total.Asset)) + geom_histogram(binwidth=0.05,color="black", fill="red")+ summary(data_1$income)
```

```
ggplot(data = data, aes(x=DE.Ratio)) + geom_histogram(binwidth=0.05,color="black", fill="red") summary(data_1$share)
```

```
ggplot(data=data,aes(x=Asset.Turnover..net.sales.to.total.assets.))+geom_histogram(binwidth=0.1,color="black", fill="red") summary(data_1$expenditure)
```

```
ggplot(data=data,aes(x=OP.over.TA))+geom_histogram(binwidth=0.05,color="black",fill="red")
```

```
summary(data_1$dependents)
```

#Plotting the variables X on Y

```
ggplot(data=data,aes(x=Working.Capital.to.Total.Asset,fill=Defulat_new))+ geom_density(alpha = 0.7)
```

```
ggplot(data = data, aes(x = CP.to.Total.Asset, fill = Defulat_new)) + geom_density(alpha = 0.7)
```

```
ggplot(data = data, aes(x = DE.Ratio, fill = Defulat_new)) + geom_density(alpha = 0.7)
```



```
ggplot(data = data, aes(x = Asset.Turnover..net.sales.to.total.assets., fill = Defulat_new)) +  
geom_density(alpha = 0.7)
```

```
ggplot(data = data, aes(x = OP.over.TA, fill = Defulat_new)) + geom_density(alpha = 0.7)
```

```
### Var-Covariance Matrix
```

```
a <- data$Working.Capital.to.Total.Asset
```

```
b <- data$CP.to.Total.Asset
```

```
c <- data$DE.Ratio
```

```
d <- data$Asset.Turnover..net.sales.to.total.assets. e <- data$OP.over.TA
```

```
M <- cbind(a,b,c,d,e) cov(M)
```

```
cor(M) library(lmtest)
```

```
# Running The regression
```

```
model_a<-glm(Defulat_new~OP.over.TA+DE.Ratio+Working.Capital.to.Total.Asset+
```

```
CP.to.Total.Asset+Asset.Turnover..net.sales.to.total.assets.,data=data,family=binomial('logit'))
```

```
summary(model_a)
```

```
lrtest(model_a)
```

```
car::vif(model_a)
```

```
#McFaddenR2
```

```
model_1<-
```

```
glm(card_new~age+income+dependents+owner_new+selfemp_new,data=data_1,family=binomial('logit'))
```

```
nullmod<-glm(card_new~1,data=data_1,family=binomial('logit'))
```

```
1-logLik(model_1)/logLik(nullmod)
```

```
#Seetheresidualsandplottingthesame
```

```
data_2<-NULL
data_2$predicted_y<-model_a$fitted.values
y<-ifelse(data$Defulat=='yes',1,0)
data_2$y<-y
data_2$residuals<-y-data_2$predicted_y
residuals_of_logit_model<-model_a$residuals
data_2$residuals_of_logit_model<-residuals_of_logit_model
data_2<-as.data.frame(data_2)
ggplot(data=data_2,aes(x=residuals_of_logit_model))+
geom_histogram(binwidth=0.5,color="black",fill="red")
ggplot(data=data_2,aes(x=predicted_y))+geom_histogram(binwidth=0.08,color="black",fill="red")
ggplot(data=data_2,aes(x=residuals))+geom_histogram(binwidth=0.08,color="black",fill="red")
```