Firm's performance forecasting and impact of financial crisis 2008.

M.Sc. Project

Submitted in partial fulfilment of the requirements

for the award of the degree

Of

Master of Science

in

ECONOMICS

Sahil Kumar

DEPARTMENT OF HUMANITIES AND SOCIAL SCIENCES,

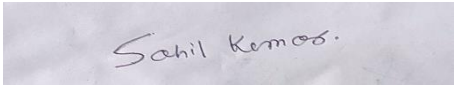
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE

ROORKEE – 247667 (INDIA) APRIL, 2022

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the project entitled " Firm's performance forecasting and impact of financial crisis 2008" in partial fulfilment of the requirements for the award of the Degree of Master of Science in Economics and submitted in the Department of Humanities and Social Sciences of the Indian Institute of Technology Roorkee, Roorkee is an authentic record of my own work carried out during a period from December ,2021 to May, 2022 under the supervision of Dr. Abhishek Samantray, Assistant professor, Department of Humanities and Social Sciences, Indian Institute of Technology Roorkee, Roorkee.

The matter presented in this project has not been submitted by me for the award of any other degree of this or any other Institution.



(SAHIL KUMAR)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

(Dr. Abhishek Samantray)

Date:

Abstract

Forecasting the firm performance has been an interesting and challenging problem for Investors and Managers based on financial information problem. The results will help Investors set a proper strategy to allocate their wealth and for managers it will help them to set goal and achieve their firm goals and this paper also tried to analysis the impact of financial crisis 2008 on firm performance at industry level and overall impact on across the all industries combined.

There is a wide range of quantitative forecasting techniques in econometric and Machine learning that have been developed to forecast and analysis. There are many parameters to measures firm's performance like ROE, Sale growth etc. but In this study, we will use Return on Assets (ROA) as a parameter to measure firm performance . Based on Historical accounting details of US firms from the Compustat(WRDS) database from financial year 2001 to 2016, we deployed the ordinary least squares (OLS) and Decision tree technique to forecast next year firms' ROA for non-banking sectors. We found that financial year, industry and current year cash flow have significant impact on next year performance but capital structure does not have significant impact in next year performance.

Financial crisis 2008 might have significant impact on firm performance. We found there is a significant difference between overall performance of all industries combined after financial crisis(2009 – 2016) and before financial crisis year (2001 – 2008). Real estate industries performance has significant difference before and after financial crisis 2008 whereas Mortgage Bankers and loan corporation industry and commercial Bank industry have no significant difference

1. Introduction

Measuring the firm's performance is a critical aspect for investors, analysts and managers. Measuring firm performance is subjective thing , parameter of measuring firm performance will vary with group of participants in economy, for an example for a customer , it would be how better product or service a firm provide to customers, for a manager it might be whether firm's sales growth or how well firm is using its resources , for an investor stock price return or how well firm is using firm's resource. Broadly firm performance measure will divide into two categories qualitative and quantitative measure. We will continue with quantitative measure of firm performance.

There are many different quantitative parameters for measuring firm performance for example ROE (Return on Equity) , stock price return, ROA (Return on Asset), Return on Sales (ROS), and Return on Investment (ROI) but in this article will use ROA (return on Asset) as proxy of firm performance because Performance efficiency is better measured by ROA, Because ROA specifically considers the assets utilised to support corporate operations. It decides whether the company can make a sufficient profit.

Previously, there have been many studies and implemented various methodologies in order to assess firms' financial performance in relation to financial

ratios While earlier studies relied heavily on traditional statistical techniques (e.g., factor analysis, linear regression, ANOVA, and so on), more recent studies used more advanced decision-making approaches. The decision tree analysis has been one of the most popular approaches, which is often preferred due to its simplicity, descriptive and predictive power. In this study, the financial performance was evaluated using linear regression and decision tree analyses, as well as several financial ratios and compare both method performance.

The basis of the global financial crisis was built behind the real estate market bubble began in 2007. Banks and credit agencies offer lower interest rates in mortgage and encouraged many home owners to take off their uninterrupted loans. Subprime Mortgage Loan Bundle was increasing at the number of overwhelming degrees, so it has been largely moved with the default payment, and credit agencies have started facing worst financial issues. This led to dire financial conditions around the world from 2008 to 2009. In this financial crisis financial institutions, real estate got impacted most. To assess the impact of financial crisis on economy and some specific industries, We will use hypothesis testing to test is there any significant difference between before financial crisis(2001- 2008) and after the financial crisis(2009- 2016) at overall across all industries combined and some specific industry level. we will also try to find what happened to those industries performance which had higher leverage(higher debt to equity ratio) in period (2001-2008) before and after financial crisis 2008.

2. Research Question

I. forecast next year firm performance.

II. Is there a significant difference between across all industries combined performance before (2001- 2008) and after (2009-2016) US financial crisis 2008 ?

III. Is there a significant difference between performance before (2001- 2008) and after (2009-2016) US financial crisis 2008 in Real Estate, Mortgage Bankers, commercial Bank Industries respectively ?

IV. Is there a significant difference between performance before (2001- 2008) and after (2009-2016) US financial crisis 2008 in those industries which higher leverage(debt to equity ratio) before US financial crisis 2008?

3. Literature review

There is many research work have done in field of measuring firm performance using ROA. The use of financial ratios to evaluate or forecast firm performance is not new. Studies frequently distinguish themselves from the rest of the papers by developing and employing different independent variables (financial ratios) and/or employing machine learning or other statistically based analysis techniques.(Horrigan 1965) , he said the development of financial ratio estimates should be a unique product of the emergence of accounting processes as well U.S. practices; it also reveals that the origins of monetary values and their initial application date back to the late 19th century. (Ross, et al. 2003)

Financial ratio estimates, calculated using the various financial variables that are commonly found in the financial statements, may provide the following benefits.

1) Measuring performance management on purpose rewards.
2) Evaluate departmental performance within multiple levels companies.
3) forecasting the future by using historical information to existing or potential investors.
4) Providing insight information to lenders and suppliers.
5) Assessing competing positions of competitors.
6) Assessing financial performance of acquisitions.

Financial estimates are frequently used to forecast future performance. They are used as ideas for practical studies, for example, or to develop models to predict financial stress or failure (Altman, 1968; Beaver, 1966). Indeed, the majority of recent research has focused on analysing and predicting bankruptcy as a means of determining the characteristics (in terms of finances) of well-performing or poor firms, as well as their potential values (Kumar & Ravi, 2007). Studies on the predictability of bankruptcy differentiate themselves from others by using a different set of financial symbols or a different set of speculative models (mathematical or machine learning based) (Alfaro et al. 2008; Holsapple & Wu, 2011; Lee et al. 1996; Martn-Oliver & Salas-Fumás, 2012; Olson et al. 2012; Wilson & Sharda, 1994). While many of these studies are successful in forecasting bankruptcy, they frequently fail to uncover and explain potential determinants like company performance decisions. (H. Youn and Z. Gu, 2010) The ratio of earnings before interests, taxes, depreciation, and amortisation to total liabilities, as well as the debt ratio, were discovered as key predictors of ROA when looking at the factors that affect Korean lodging enterprises' performance assessed in terms of return on assets (ROA). (Kim, M., & Burnie, D.2002) investigated the impact of the economic cycle on business performance using ROA as the performance metric. They discovered that small enterprises outperform large firms in good economic times, but that small firms are more vulnerable to weak economic times.(Panday,R & Diaz,J. 2019) examined some specific factors effect on Return on assets (ROA) of US technology and financial firms and found Return on equity(ROE), Long term Debt ratio(LTD) both have negative impact on ROA for both Financial and Technology firms while Return on Sales and Debt ratio have positive impact on Return on Asset(ROA) for US Technology and Financial firms.(Tailab, M.2014)investigated the impact of liquidity(current assets to total current liabilities), leverage(total liabilities to total asset) ,inventory(inventory investment to total assets), sales growth(current year sales to previous year sales), size(sales , asset size) and firm's age on financial performance as measured by return on assets (ROA) and found inventory, growth, age and leverage have a negative significant impact on ROA, while size in terms of sales and liquidity have a positive significant effect on the profitability of the U.S firms. However, an insignificant negative relationship was found between total assets size of firm and return on assets.(Delen, et al. 2013) they worked on Turkish listed public companies from 2005 to 2011.they used two parameters ROE and ROA as firm performance measure and took one parameter as firm performance measure at a time to build model, they worked with Decision tree models and found where When ROA was employed as

the dependent variable, the most relevant financial variables were Earnings before Tax-to-Equity Ratio, Net Profit Margin, Debt Ratio, and Asset Turnover Ratio, with the biggest impact on predicting ROA, accordingly. (Uyar & Okumus.2010) used financial ratings to analyse the impact of the global financial crisis in 2008 on Turkish industrial enterprises trading publicly and discovered that firms were financially weak during the crisis. Using fifty financial projections, (Wang et al.2009) developed a bagging decision tree model to predict stock recovery. Sun and Hui (2006) employed decision-making trees and genetic algorithms to anticipate the financial stress of Chinese publicly traded enterprises. (Agnihotri,A 2014) If a corporation can make an effective capital structure decision based on its strategies, competitive and market conditions, both high and low leverage will result in lower loan costs, hence improving its performance. Finally, depending on the conditions, a financial ratio can have the same or a different relationship with capital structures. For example, (Saeedi, A. and Mahmoodi,i. 2011) discovered a negative relationship between capital structure and return on assets (ROA), but no significant relationship between capital structure and return on equity (ROE). On the other hand, (Ebrati et al.2013) discovered a similar result for ROA, but a positive relationship for ROE. (Yu and Wenjuan.2010) used the decision tree to assess which financial measures had a significant impact on the growth of profitability for listed companies; use C5.0, which is one of the decision tree methods

Many studies in the field of firm pre-financial-crisis, post-financial-crisis, and post-crisis recovery (2010-2013) examined the impact of financial leverage on the performance of Canadian oil and gas companies, finding that leverage has a strong negative relationship with firm performance for all three periods, pre-crisis (2004-2006), crisis (2007-2009), and post-crisis recovery (2010-2013). (Saleh, et al. 2017) examined the financial performance of listed firms on the Australian Securities Exchange (ASX) using ROA and ROE over two sample periods before and during the global financial crisis periods (1998-2007) and 2008-2010). The results show that family firms with ownership concentration performed better than nonfamily firms with sprinkled ownership structures, and ownership concentration has a positive and significant impact during the crisis period and the other finding, financial leverage had a positive and significant effect on the performance of Australian family-owned firms during both periods.

4. Methodology and Data

I. Data

for the analysis we used US firms historical accounting details data from the Compustat (WRDS) database from the financial year 2001 to 2016.

II. Objective

Forecast the next year firm performance using current year financial data using linear regression and decision tree model. Analysis and compare the industries performance in two period 2001-2008 and 2009-2016.

III. Variables and Descriptions

Table 1

Variable Name	Symbol	Definition
Financial year	fyear	Current financial year
Standard Industrial Classification	SIC	Standard Industrial Classification Number
Short term debt to total asset	STDTA	$\frac{\text{Short term } Debt}{\text{Total asset}}$
Long term debt to total asset	LDTA	$\frac{\text{Long term Debt}}{\text{Total Asset}}$
Debt to Total Asset	TDTA	$\frac{\text{Total Debt}}{\text{Total Asset}}$
Net Income to Total liabilities	NILT	$\frac{\text{Net income}}{\text{Total liabilities}}$
Short term debt to equity	STDEQ	$\frac{\text{Short term debt}}{\text{total equity}}$
Long term debt to equity	LDEQ	$\frac{\text{Long term debt}}{\text{total equity}}$
Debt to equity	DEQ	$\frac{\text{Total debt}}{\text{total equity}}$
Return on equity	ROE	$\frac{\text{Net income before tax}}{\text{Total Equity}}$
Net profit margin	NIRV	$\frac{\text{Net income}}{\text{revenue}}$
Asset turnover ratio	RVAT	$\frac{\text{Revenue}}{\text{total asset}}$
Return on Asset	ROA	$\frac{\text{Net income before tax}}{\text{total asset}}$

Dependent variable(endogenous variable) : next year ROA(nxt_ROA)

Independent Variables(exogeneous variable) : fyear, SIC, STDTA ,LDTA, TDTA, NILT,STDEQ, DEQ,ROE, NIRV, RVAT, ROA(current year ROA).

IV. Null Hypothesis

A. H1 : there is no significant difference between the time period (2001- 2008) and the period(2009-2016) across all industries combined performance.

B. H2 : There is no significant difference between the performance on the period(2001- 2008) and the period(2009-2016) in Real Estate, Mortgage Bankers, commercial Bank Industries respectively.

C. H3 : There is no significant difference between the performance on the period(2001- 2008) and the period(2009-2016) in those industries which had higher leverage(debt to equity ratio) before US financial crisis 2008.

V. Model for forecasting next year performance

A. OLS (ordinary least square method)

Ordinary Least Squares (OLS) is a one form of linear least squares technique for estimating the unknown parameters (dependent variable) in a linear regression model. OLS chooses the parameters of a linear function of a set of explanatory variables(independent variables) by the principle of least squares method, minimizing the sum of the squares of the differences between the observed dependent variable (actual value of dependent variable) in the given dataset and those predicted by the linear function of the independent variable. Model specification is given below.

$$\text{Nxt\_ROA} = \beta 0 * fyear + \beta 1 * SIC + \beta 2 * STDTA + \beta 3 * LDTA + \beta 4 * TDTA + \beta 5 * NILT + \beta 6 * STDEQ + \beta 7 * LDEQ + \beta 8 * DEQ + \beta 9 * ROE + \beta 10 * NIRV + \beta 11 * RVAT + \beta 12 * ROA.$$

Assumption of OLS

(Gujarati .D & Porter.D.2012)OLS model is based on several assumptions(A), which are as following.

A1: The regression model is linear in the parameters.

A2: exogeneous variables are independent of the error term. Here, this means we require zero covariance between error term and each exogeneous variable.

A3: For given exogeneous variables, the mean value of error term is zero.

A4: For given exogeneous variables, the variance of error term is constant or homoscedastic.

A5: For given exogeneous variables, there is no autocorrelation, or serial correlation, between the error term.

A6: The number of observations n must be greater than the number of parameters to be estimated.

A7: There must be sufficient variation in the values of the X variables.

A8: There are no Multicollinearity between the X variables.

A9: The model is correctly specified, so there is no specification bias.

A10: The error term is normally distributed.

Error term(Ui)

It is difference between actual value and predicted value of dependent variable.

Ui = Yi - $\hat{Y}$i, where Ui = error term, Yi = Actual value, $\hat{Y}$i = predicted value.

Autocorrelation

(Gujarati .D & Porter.D.2012) it is defined as "Correlation between members of series of observations ordered in space (as in cross-sectional data) or time (as in time series data)." Since the classical linear regression model assumes that there is no autocorrelation in the error term . mathematically representation.

$$\text{Cov}(\text{Ui}, \text{Uj} \mid \text{Xi}, \text{Xj}) = \text{E}(\text{Ui}, \text{Uj}) = 0 \quad \text{where i} \neq \text{j}$$

Test for Autocorrelation :

1. **The Breusch–Godfrey (BG) Test :** This test was developed by statisticians Breusch and Godfrey to test the presence of autocorrelation.

Multicollinearity

If there is exact linear relationship between two or more exogeneous variables is called perfect Multicollinearity, mathematically if there is K exogeneous variables in model (X1,X2,X3, .., Xk) and there is K constant term (C1,C2,C3,…,Ck) there exist a linear relationship $C1 * X1 + C2 * X2 + \cdots + Ck * Xk = 0$.

such that all these constant terms are not zero simultaneously. If there is less perfect multicollinearity exist, we can estimate coefficient of model but it will have wider confident interval. To tackle the problem of perfect multicollinearity, if two exogeneous variables have correlation coefficient more than 0.75, we will drop one exogeneous variable out of these two variables. Even present of partial multicollinearity , parameter have BLUE(Best Linear unbiased Estimator) property.

Homoscedastic

(Gujarati .D & Porter.D.2012) variance in each error term ei ,conditional for given values of the explanatory variables, is some constant number. This is the assumption of homoscedasticity that is, equal variance.

Symbolically, $\text{E}(ei^2) = \sigma 2$, $\text{i} = 1, 2, 3 \ldots$ where $\sigma 2$ is constant term.

Test to detect Homoscedasticity

(Gujarati .D & Porter.D.2012) suggested some following formal test to detect homoscedasticity

a) **White's General Heteroscedasticity Test**

this model was proposed by Halbert White(1998) and does not make the normality assumption and is easy to implement. null hypothesis(Ho) that there is no heteroscedasticity and it follow Chi square distribution.

b) **Breusch–Pagan–Godfrey Test**

it was proposed by (T. Breusch and A. Pagan 1979). Null hypothesis is Homoscedasticity is present (the error terms are distributed with equal variance) and assumption is error term is normally distributed. This test follows the Chi square test.

B. Decision Tree

Decision Trees is Supervised learning method approach and it is non-parametric method. Both classification and regression problem can be solved by Decision Tree . The goal is to create a model that predicts the value of endogenous variable by learning simple decision rules inferred from the data features. tree model can be

seen as a piecewise constant approximation. Decision trees are commonly used model in data mining. decision trees method can perform both task classification tree analysis and regression tree analysis. Because they are easy to interpret and understand decision trees are becoming increasingly popular in data mining. We used Scikit-learn open-source library in python to deploy this model.

Advantages of decision trees Model (scikit-learn)

a) Easy to interpret and understand. Decision Trees' decision flow can be visualised.
b) It necessitates minimal data pre-processing. Other machine learning models frequently necessitate data normalisation, the creation of dummy variables, and the removal of blank values. However, missing values are not supported in the scikit-learn module.
c) It can handle the multi-output problems.
d) It works like white box model. When a situation can be observed in a model, Boolean logic may simply describe the decision-making process.
e) Performs well even if it's some assumptions are violated by the true model from which the data were generated.

Disadvantages of decision trees Model (scikit-learn)

a) Decision tree prone to overfitting the train data, to fix it we use pruning mechanism to predefine the maximum dept by setting the maximum depth of the tree or setting the minimum number of samples required at a leaf node are necessary to avoid this problem.
b) Predictions of decision trees are piecewise constant approximations neither smooth nor continuous. Therefore, they are not good at extrapolation.
c) Some concepts, such as parity, XOR, and multiplexer difficulties, are difficult to grasp because decision trees do not easy to describe them.
d) In classification problem Decision tree learners create biased trees if some classes dominate.

Type of Decision Tree Model : There are many specific decision tree algorithms, the C5.0, C4.5, ID3, QUEST, CHAID and C&RT algorithms are the most commonly used form of Decision tree method

ID3

Iterative Dichotomiser 3(ID3) was developed by (Quinlan, J. Ross.,1986). ID3 generates a multiway Decision tree, seeking the categorical feature in the exogenous field that will produce the biggest information gain for categorical endogenous variable for each node (i.e., in a greedy manner of searching). Trees are grown to their maximum size, and then pruned to improve the tree's robustness, allowing it to generalise to new data and avoid overfitting.

C4.5

It is the successor to ID3 and it is developed by (Quinlan, J. Ross.,1986). and It did so by dynamically defining a discrete attribute (based on numerical variables) that separated the continuous attribute value into a discrete set of intervals, removing the requirement that features be of the categorical type. The trained trees model (i.e. the

output of the ID3 algorithm) is converted into sets of if-then rules criteria using C4.5. Each rule's correctness is assessed to determine the sequence in which it should be applied. Pruning is when a rule's precondition is removed if the rule's accuracy improves without it.

C5.0

This was developed by Quinlan (1993). It is significant improved version of C4.5, it is significantly efficient than C4.5 in term of execution time and memory usages. C5.0 produces a significantly smaller decision tree than C4.5 while producing similar results; it boosts the trees, improving and increasing accuracy; it allows C5.0 to weight different attributes and misclassification types; and it automatically extracts data to help reduce noise in the exogenous variables field, improving the precision of the decision tree classification algorithm. Boosting is an integrating technology included in the C5.0 decision tree algorithm that improves classification accuracy. It starts with the root point, which is the highest level of the details, and creates the decision tree using both post-pruning and pre-pruning procedures. The set of training dataset is divided into two or more subsets, based on the outcome of a test of the value of a single attribute. The particular test is chosen by an information from past data that generally gives close to optimal partitioning criteria. This is repeated on each new subset until a subset contains only single class of dependent variable or the partitioning tree has reached a predetermined maximum depth.

CHAID

Chi-squared Automatic Interaction Detector (CHAID) is an effective statistical technique, which was created by (Kass,1980). CHAID is a programme for segmentation or tree growing. It is based on adjusted significance testing process. It can be utilised for classification and regression analysis predictions, as well as discovering interactions or associations between variables. CHAID is not a binary tree approach, as other decision tree techniques can create more than two categories at any level in the tree. Its output is simple to understand and very graphic, and it builds a larger tree than binary growth methods because it uses multi category split splits by default. Because it accepts both case weights and frequency variables, this approach works for any type of variable. Missing values are handled by CHAID by treating them all as a single valid category.

CART

Classification and Regression Trees were established by(Breiman.et al. 1984). C&RT is a binary decision tree algorithm capable for both classification and regression of endogenous variable . CART works in a recursive manner, splitting data into two subsets in order to make each subset's records more homogeneous than the preceding one. These two subsets are then split into two subsets again until the homogeneity criterion or some other stopping criteria is met. The same exogenous field might be utilised for decision criteria many times in the tree. The ultimate goal of splitting is to match the proper variable to the right threshold level in order to optimise sample subgroup homogeneity. It also deals with missing values by employing surrogate splitting to make the most of the data. CART generates a series of nested trimmed trees, each of which has the potential to be optimal. Rather than using internal data, the right

size is determined by a selected predictive performance parameter of each tree in the pruning sequence on an independent test dataset or via cross-validation (training-dataset). Following a test-data-based evaluation, the best tree is chosen. This mechanism supports missing value handling as well as optional automatic class balancing.

Since Scikit-learn uses CART version of Decision tree which is an optimised version of the CART algorithm. We will use CART version of Decision Tree algorithm. We will use decision tree model for regression analysis in this article with dependent variable next year ROA.

VI. Performance measure for model

We will us Root mean Square Error(RMSE) as performance measure of predictive ability of a model for next year ROA.

$$RMSE = \sqrt{\sum_{i=1}^{N}(X^i - X^{ei})^2/N}$$

Where, $X^i = Actual\ value$

$X^{ei} = predicted\ value$

N = Total Number of samples

VII. Standarizatoion

We transformed the exogeneous variable into normal distribution using Z-statistic method by given method to fullfill the normality assumtion of exogeneous variable in OLS and as well as for decision tree method.

$$x = (X- \mu)/std(X)$$

where x = normal form of variable X

X= value of X variable

μ= mean of X variable

std(X) = standard deviation of X variable

we selected randomly 70% data for training the model and remaining 30% data for testing purpose **.** mean of depedent variable is 0.02 for both test and train data while standard deviation is 0.09 for both train and test data.

5. Descriptive Analysis

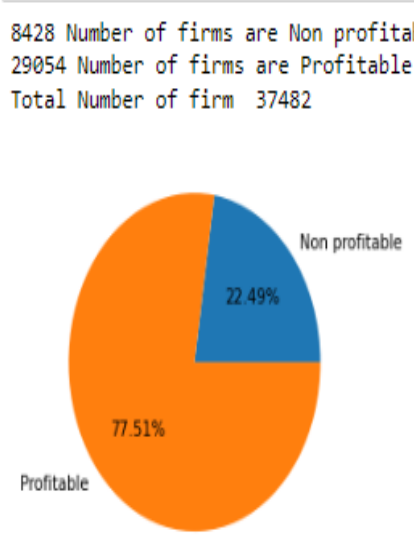
Table 2

(Variables Descriptive Statistics)

Total number of observations in the sample 319933 and total number of firms is 37482.

	STD TA	LTD TA	TDTA	NILT	STDEQ	LDEQ	DEQ	ROE	NIRV	RVAT	ROA	Nxt_ ROA
Mean	0.11	0.11	0.23	0.09	0.40	0.44	0.83	0.06	0.05	0.91	0.03	0.03
Standard Deviation	0.13	0.15	0.20	0.21	0.67	0.80	1.19	0.18	0.16	0.76	0.07	0.10
Minimum	-0.13	-0.03	0.00	-0.99	-0.29	-0.09	0.00	-0.99	-0.99	0.00	-0.95	-0.99
Maximum	4.57	4.60	4.80	0.99	9.96	9.86	9.99	0.99	0.99	9.98	0.92	0.99

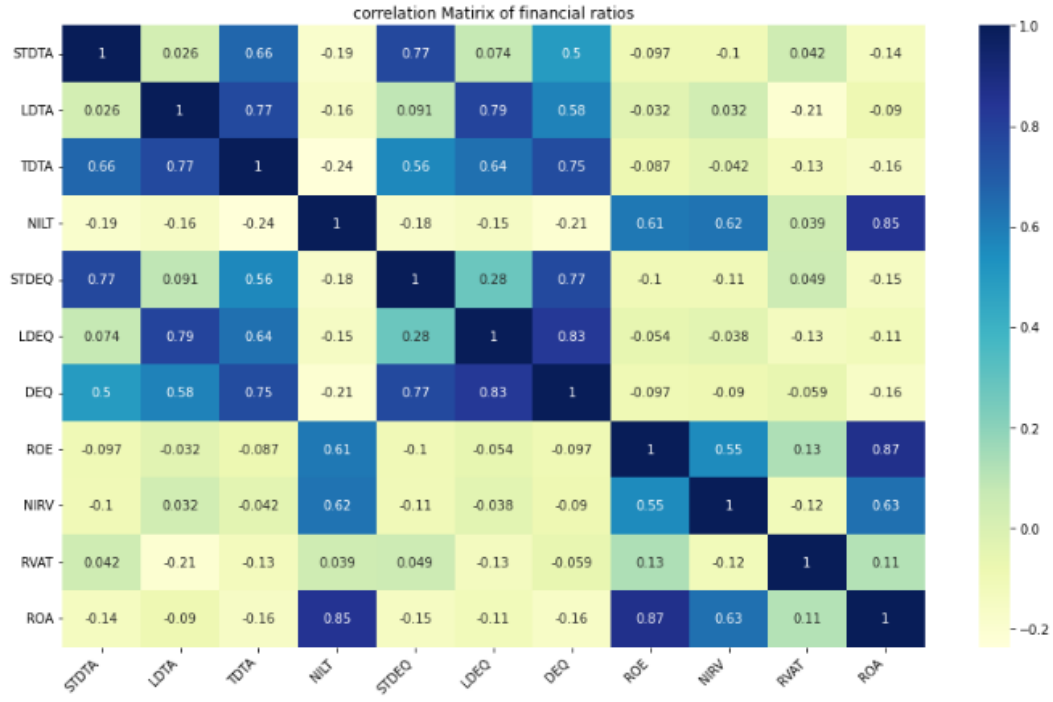
Figure 1



In the given data there are 22.49% firms are non-profitable, that is, mean of ROA of these firm in period the 2001 to 2016 is less than or equal to zero.

In Figure 2, the correlation coefficient between NILT and ROA is 0.85, between ROE and ROA is 0.87,between LDTA and TDTA is 0.77, ,STDEQ and DEQ is 0.76 , between LDEQ and DEQ is 0.84.Since in OLS method, we made assumption that there is no multicolinearity among exogeneous variables,but weak multicolinearity may be present in model.we will drop either vriable of two variables which have correlation coefficient more than 0.75. we will drop NILT, ROE, LDTA, STDEQ, LDEQ to tackle the problem of multicolinearity.

Figure 2 (Correlation matrix)



6. Empirical Analysis

I. Trend of the mean of ROA versus the financail Industries.

Figure 3 (show the trend of mean of ROA of all firm in sample of all indutries.)

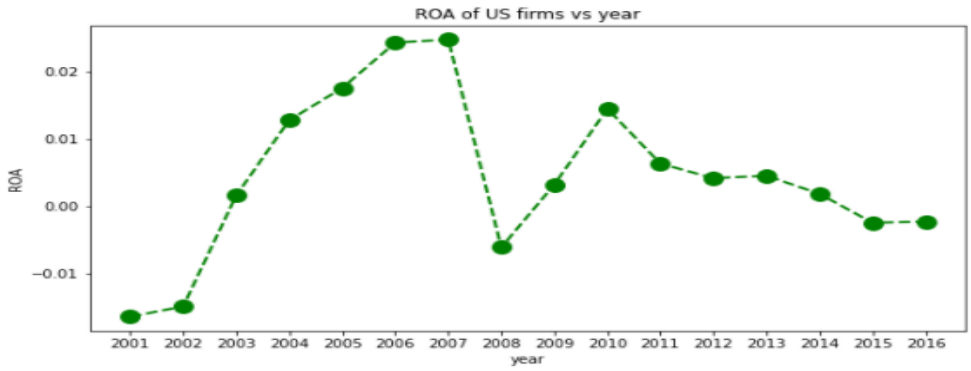


Figure 4 (show the trend of mean of ROA of all firm in sample of agriculture indutries.)

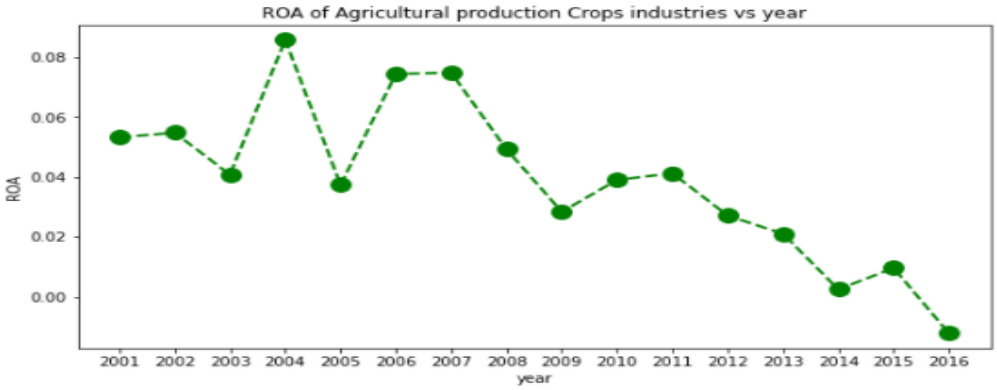


Figure 5 (show the trend of mean of ROA of all firm in sample of Commercial bank indutries.)

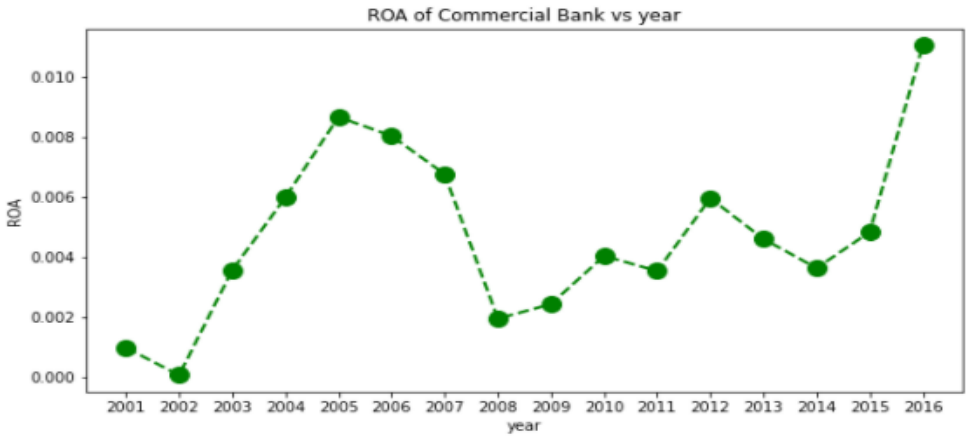


Figure 6 (show the trend of mean of ROA of all firm in sample of Jewelry and Watches indutries.)

Figure 7(show the trend of mean of ROA of all firm in sample of Mortgage Bankers and loan corr indutries.)

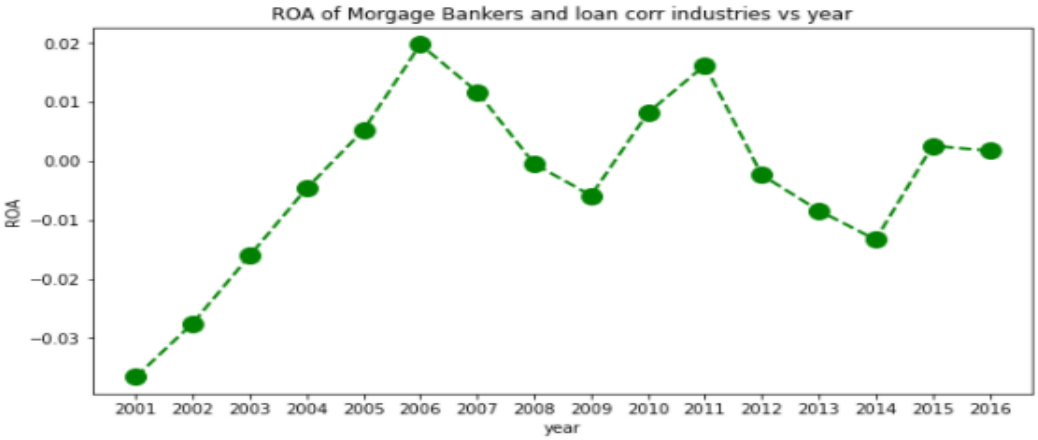
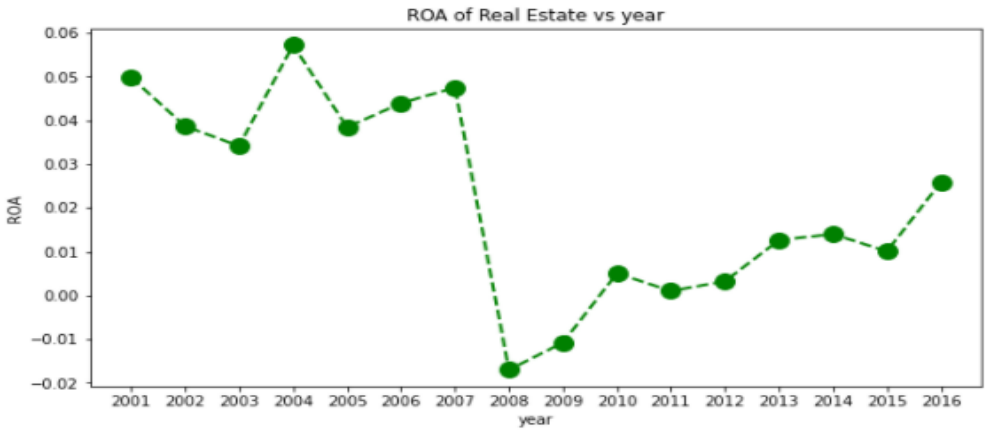


Figure 8 (show the trend of mean of ROA of all firm in sample of Real Estate Industry)



II. Hypothsis testing Result

Applied z statistic to test all three null hypothsis above mention at level of 5% level of significant. We reject Null if Z-score is greater than 1.96 0r less than -1.96.

Table 3

Null Hypothesis	**Number of Firm**	**Industry**	**Z-score**	**Is Null Hypothesis Accepted ?**
H1	37482	All industries are included in given sample	5.66	NO
H2	847	Real Estate	7.91	NO
H2	49	Morgage Bankers and loan corr	-0.50	Yes
H2	1132	commercial Bank	0.46	Yes

In Table 3, menation the Z statistics score for Null hypothesis H1 and H2. we rejected the null hypothesis when we used all firms in given sample combined. Hance, there is significant difference between the mean of ROA of US frims in period 2001-2008 and in period 2009-2016. From this result we may conclude there has been a long term impact on US economy due to global financial crisis 2008. For the real estate industry ,we rejected the null hypothesis, there is significant difference between the ROA of firms in period 2001-2008 and 2009- 2016. Hance we may conclude that the ROA in real estate industry became lower after financial crisis 2008. While Morgage Bankers and loan corr and commercial Bank industry recovered the indutry performance, which got impact severly due to financial crisis 2008.

Now, we will test H3 null hypothesis in those industries which had higher leverage(debt to equity ratio) before US financial crisis 2008. We used Debt to equity ratio as a leverage ratio and selected some industries which had higher mean of Debt-to-equity ratio(DEQ) in period 2001-2008. We reject Null hypothsis if Z-score is greater than 1.96 0r less than -1.96 at 5% level of significant level.

Table 4

Null Hypothesis	Industry	Number of Firm	Debt to equity ratio	Z-Score	Is Null Hypothesis Accepted ?
H3	REAL ESTATE DEALERS	5	8.49	-1.26	Yes
H3	FINANCE LESSORS	23	5.26	-1.19	Yes
H3	MORTGAGE BANKERS & LOAN CORR	49	8.28	-0.50	Yes
H3	FINANCE SERVICES	4	2.30	-1.23	Yes
H3	Commercial Bank	1132	3.24	0.46	Yes

In Table4 , Debt to equity ratio column show the means of Debt to equity ratio of all firms in particular industry in period 2001-2008.we found that these highly leaverge selected indutries mean performance are not different in the period 2001-2008 and 2009-2016 at 5% level of significant.this may be because of government measures taken during the finacial crisis like lowering the interest rate and increse the liquidity in economy. Hance from this sample industry and finding from Table 4, we can conclude that leaverge factors(Debt to equity ratio) are not significant to firm performance on long term horizon.

Hance from above discussion, we rejected Null Hypothsis H1, but accpeted H2 except for Real estate industry . accepted null H3 for all selected highly leaverage industries. Below mentioned the trend of ROA and Debt to equity ratio over the period 2001-2016 for thesis selected higly leverage industries.

Figure 9 (ROA and Debt to equity ratio yearly trend of commercial bank industry)

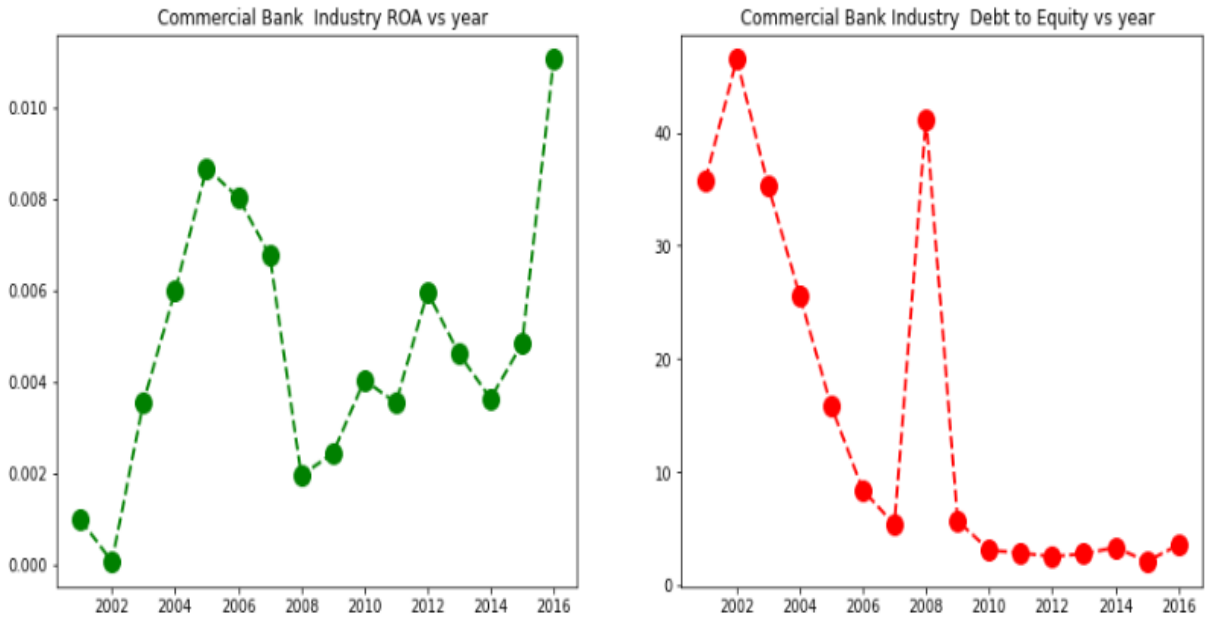


Figure 10 (ROA and Debt to equity ratio yearly trend of financial lessors industry)

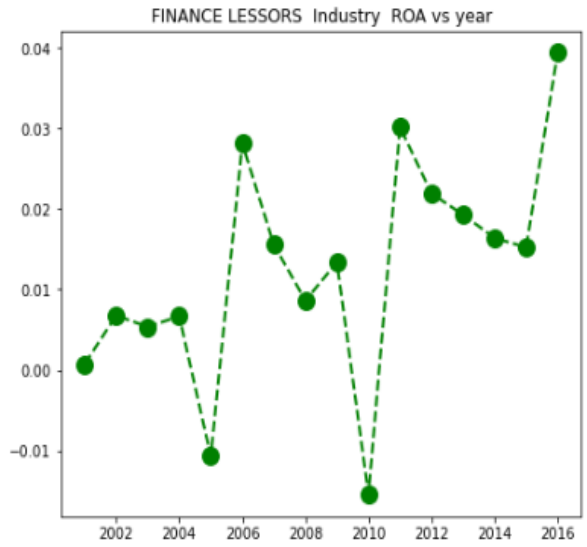
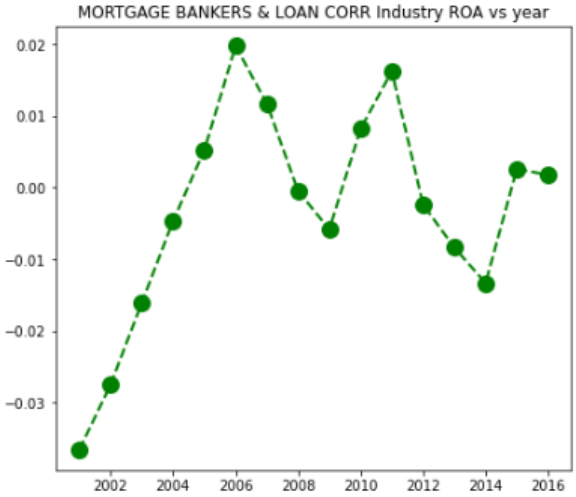


Figure 11 (ROA and Debt to equity ratio yearly trend of mortgage banker industry)



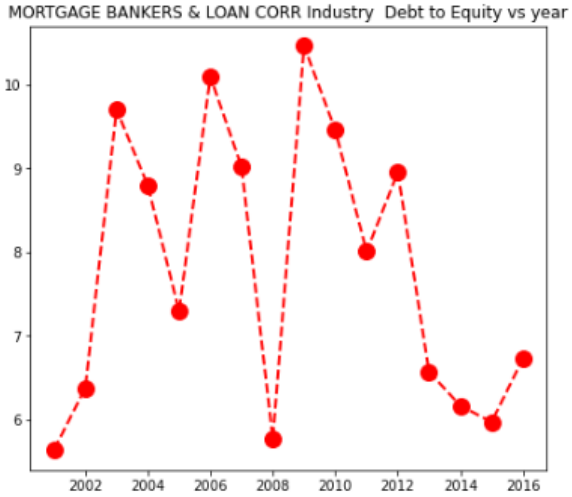
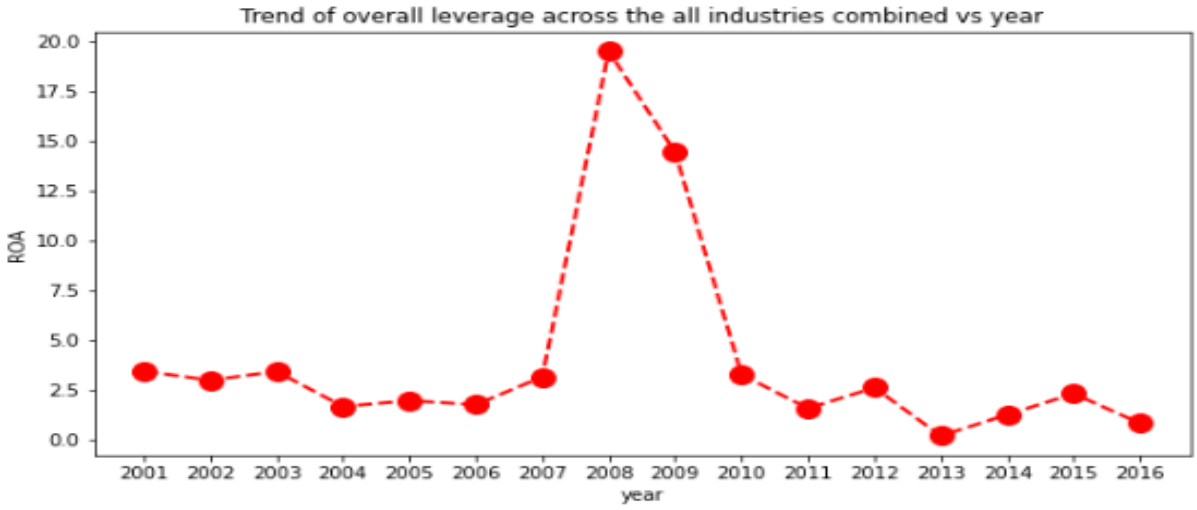


Figure 12 (ROA and Debt to equity ratio yearly trend of real estate dealers industry)

Figure 13 (ROA and Debt to equity ratio yearly trend of fiancial services industry)

Figure 14 (yearly trend of mean debt to equity of all given firms in sample)



III. Model for forcasting neat year ROA of firm

a) OLS method

since data is large enough and all exogenous varibles have a significant variation variation in data and assume all there is linear relationship between independent variables and dependent variables also correctly specified, so assumption1, assumption6, assumption7, assumption9 in OlS assumtion already holds. Since since we assume all independent varibles are exogenous so there is no covariance between independent variable and error term so assumption2 also holds.since,we droped NILT, ROE, SDTA,LDTA, STDEQ, LDEQ to tackle the problem of multicolinearity hance assumption8 also hold,so remaining exogenous fyear,sic, STDTA,TDTA, DEQ,NIRV, RVAT, ROA for OLS method. Mean of error term is 0.00 which implies assumption3 hold. For autocorrelation test ,we got Breusch–Godfrey (BG) Test statistics 2.92 and p-value is 0.711 and Durbin-Watson statistics is 2.003 which is about equal to 2 from both statistics we accept null hypothesis that model does not have autocorrelation problem,

hence Assumption5 holds. From figure 15 which present the distribution of error term which show clearly it follow normal distribution ,hence assumption10 holds. White's General Heteroscedasticity Test is used to test Heteroscedasticity Test, found test statistics 7384.48 with p-value 0.00 so we reject the null hypothesis hance model face problem of Heteroscedasticity hance assumption 4 does not hold. (Gujarati .D & Porter.D.2012)Hance in absence of homoscedasticity model's coefficients will be consistent, unbiased but will not be efficient(lowest variance).Since Heteroscedasticity is present in model, we will use robust standard errors in model that correct the presence of Heteroscedasticity while calculation of standard error of estimated coefficient for exogeneous variables.

Result of model is given in Table 5. The model is overall statistically significant because p-value of F-statistics is 0.00. R2 and adjR2 of model are 0.265. coefficient of TDTA, DEQ are statistically insignificant. Since coefficient of DEQ is statistically insignificant, which implies the variation in DEQ among firm have no impact on firm performance also a firm financing the capital from debt instrument which means owner or management of firm is confident enough to generate more return in future than without debt and debt proportion in total asset will not impact firm performance, it also supports the finding in Hypothesis3 but on the other hand STDTA is highly negatively statistically significant, which implies short term debt impact negatively to next year performance, it may be because more interest burden short run due to short term debt. coefficient of sic is statistically significant , which implies industry factor impact the firm performance, which means two firm having same financial ratio will have different ROA because of industry difference. Firm next year performance linked to current year activity like macroeconomics condition in economy will impact firm performance, since coefficient of fyear is statistically significant and fyear will incorporate all firm outside conditions which are overall business, economic environment, government policy . current year ROA is highly significant than all other exogenous variables , which implies the next year performance is highly dependent on current year performance of firms. RMSE for train dataset is 0.10 and RMSE for Test dataset 0.10 . coefficient of DEQ and TDTA are statistically insignificant. In Figure15 shown the error term distribution in train and test dataset when applied OLS method. Figure16 show the error term distribution using OLS method.

Table 5

(OLS method result)

Endogenous variable : next year ROA

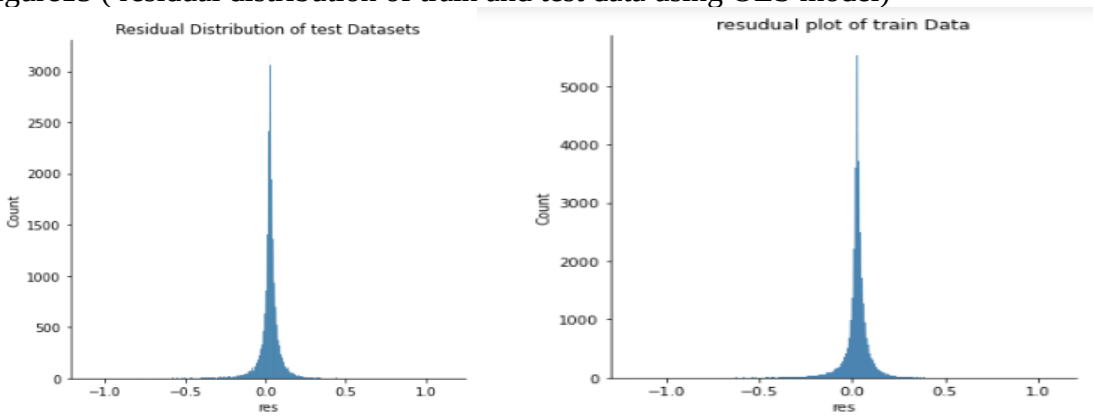
Total number of observations : 213978

Adj. R- squared : 0.265

Exogeneous variable	Coefficient	Robust Standard Error	Z-statistics	P-value
fyear	-0.0014	0.000	-7.274	0.000
sic	-0.0019	0.000	-8.502	0.000
STDTA	-0.0033	0.000	-11.578	0.000
TDTA	0.0003	0.000	0.961	0.337

DEQ	0.0005	0.000	1.549	0.121
NIRV	0.0020	0.000	5.093	0.000
RVAT	0.0035	0.000	14.498	0.000
ROA	0.0508	0.000	109.305	0.000

Figure15 (residual distribution of train and test data using OLS model)



b) **Decision tree method**

In decision tree model we used fyear,sic, NILT, ROE, LDTA, STDEQ, LDE, STDTA,TDTA, DEQ,NIRV, RVAT, ROA as exogeneous variables, since Decision tree model don't make any assumption for multicollinearity for exogeneous variable. We got RMSE for train data 0.08 and for test data 0.08. in figure16 , it represents importance of exogenous variable to forecast next year ROA.

From figure17, current year ROA is most important factor to determine the next year ROA of firm, then NILT, NIRV, RVAT, fyear, sic have importance to forecast next year ROA. From this finding we can conclude the current year performance or present cash flow is most import factor for future cash flow. In figure18 it represents the decision criteria for forecasting next year ROA of firm. RMSE for train data is 0.08 and RMSE for test data is 0.08.

Figure 16(show exogenous variables proportion importance in Decision Tree Model)

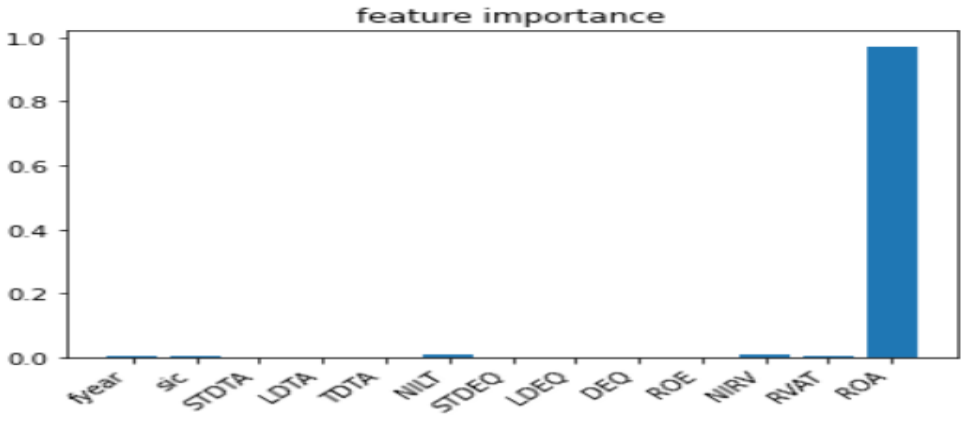


Figure 17(residual distribution of train and test data using Decision tree model)

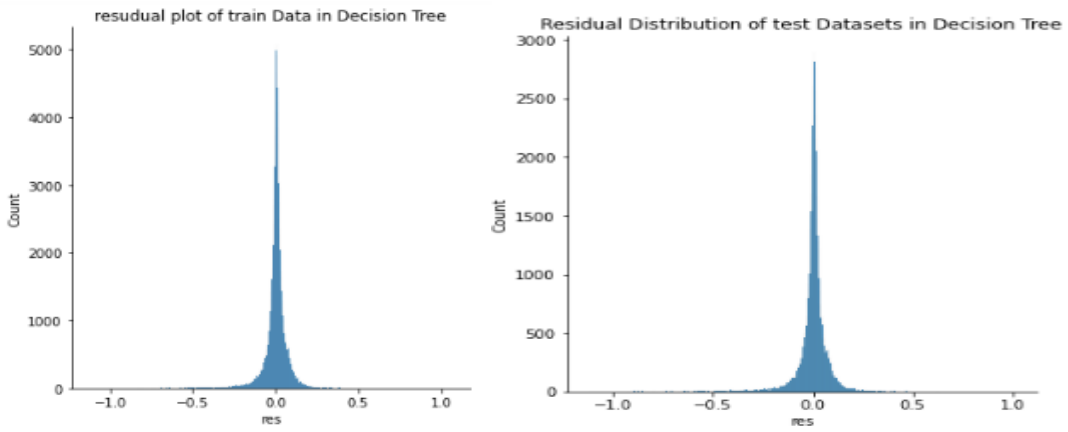
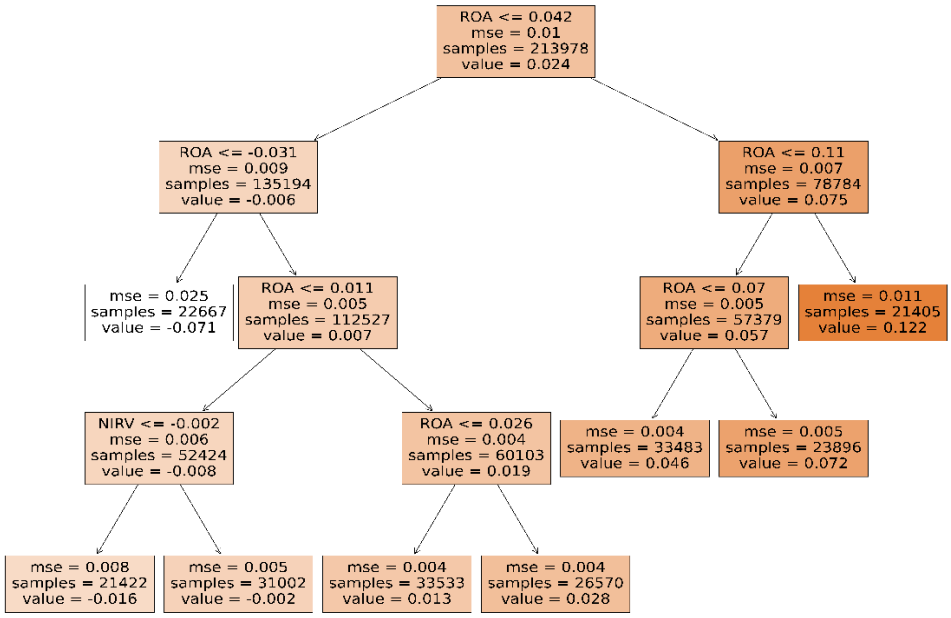


Figure 18 (Decision making flow chart in Decision Tree Model)



7. Conclusion

I. The mean of ROA of all firms in sample in period 2001-2008 is significantly higher than in period 2009-2016**.** which implies, the financial crisis 2008 may cause for some structural change in economy.

II. Real estate industry performance which yet did not recover to the level before financial crisis 2008 performance level. it might be because of real estate property price in period 2009-2016 is in saturation state while price was highly inflated before financial crisis.

III. Mortgage Bankers, commercial Bank ,Real estate dealers, Finance lessors , Finance Service industries performance recovered to the level in period 2001-2008.

Which may be because of during crisis, US government bailout packages in these most affected industries.

IV. Total debt proportion in total Asset does not affect significantly to the firm performance, while short term debt impact negatively to the firm's ROA .

V. Decision Tree model performed better than OLS model to forecast next year ROA of firm with respect to Root Mean square error value as a performance measure of model.

VI. Current year ROA of firm is the most important factor to forecast next year ROA of firm and impact positively to next year performance of firms. which tell us the present cash flow predict future cash flow.

VII. Financial year is significant to forecast firm performance. Which implies macro-economic condition (incorporated in Financial year) like government policy, interest rate and other economic activity which is not controlled by firm impact the future performance of firms.

VIII. Industry effect on next year firm performance is significant, which implies two firm from two different industries with same financial data in current year impact differently firm next year performance.

References

Horrigan, James O. (1965). Some empirical bases of financial ratio analysis. The Accounting Review, 40(3), 558–568.

Ross, Stehen A., Westerfield, Randolph W., & Jordan, Bradford D. (2003).Fundamentals of corporate finance (6th ed.). New York: The McGraw-Hill Companies.

Altman, Edward I. (1968). Financial ratios, discriminant analysis and the predication of corporate bankruptcy. The Journal of Finance, 23(4), 589–609.

Kumar, P. R., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review. European Journal of Operations Research, 180(1), 1–28.

Uyar, Ali, and Ersin Okumuş. "Finansal oranlar aracılığıyla küresel ekonomik krizin üretim şirketlerine etkilerinin analizi: İMKB'de bir uygulama." *Muhasebe ve Finansman Dergisi* 46 (2010): 146-156.

Alfaro, E., García, N., Gámez, M., & Elizondo, D. (2008). Bankruptcy forecasting: Anempirical comparison of AdaBoost and neural networks. Decision Support Systems, 45(1), 110–122.

Lee, K. C., Han, I., & Kwon, Y. (1996). Hybrid neural network models for bankruptcy predictions. Decision Support Systems, 18(1), 63–72.

Olson, D. L., Delen, D., & Meng, Y. (2012). Comparative analysis of data mining methods for bankruptcy prediction. Decision Support Systems, 52(2), 464–473.

Youn,H. & Gu,Z.(2010). Factors Affecting Return on Assets in the Korean Lodging Industry: A Preliminary Empirical Investigation ," Journal of Hospitality Financial Management: Vol. 15 : Iss. 2 , Article 3.

Quinlan J. (1986) Induction of decision trees. Machine learning, 1(1), pp.81-106 Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification andregression trees. New York: Chapman & Hall/CRC.

Wang, H., Jiang, Y., & Wang, H. (2009). Stock return prediction based on baggingdecision tree, In Proceedings of 2009 IEEE international conference on grey systems and intelligent services, November 10–12, Nanjing, China.

Sun, J., & Hui, X-F. (2006). An application of decision tree and genetic algorithms for financial ratios' dynamic selection and financial distress prediction, In Proceedings of the fifth international conference on machine learning and cybernetics, Dalian, 13–16 August

Yu, G., & Wenjuan, G. (2010). Decision tree method in financial analysis of listed logistics companie, In 2010 International conference on intelligent computation technology and automation.

Kim,M & Burnie,D.(2002). The Firm Size Effect and the Economic Cycle. The Journal of Financial Research • Voi. XXV No. 1 • Pages 111-124 • Spring 2002

Kass, G. (1980). An exploratory technique for investigating large quantities of categorical data. Applied Statistics, 29:2, 119–127

Panday,R & Diaz,J. (2019). Diaz, John Francis, and Rudresh Pandey. "Factors affecting return on assets of US technology and financial corporations." *Jurnal Manajemen dan Kewirausahaan* 21.2 (2019): 134-144.

Agnihotri, A. (2014). Impact of Strategy–Capital Structure on Firms' Overall Financial Performanc. John Wiley and Sons, Ltd.

Saeedi, A., and Mahmoodi, I. (2011). Capital Structure and Firm Performance: Evidene from Iranian Companies. International Research Journal of Finance and Economics, 70, 20-30.

Ebrati, M. R., Emadi, F., Balasang, R. S., and Safari, G. (2013). The Impact of Capital Structure on Firm Performance: Evidence from Tehran Stock Exchange. Australian Journal of Basic and Applied Sciences, 7 (4), 1-8.

Tailab, M.(2014). Analyzing Factors Effecting Profitability of Non-Financial U.S. Firms. Research Journal of Finance and Accounting Vol.5, No.22, 2014.

Delen,D.Kuzey,C.Uyar,A (2013). Delen, Dursun, Cemil Kuzey, and Ali Uyar. "Measuring firm performance using financial ratios: A decision tree approach." *Expert systems with applications* 40.10 (2013): 3970-3983.

Hossain,A & Nguyen, D. (2016). Hossain, Ashrafee Tanvir, and Dao Xuan Nguyen. "Capital Structure, Firm Performance and the Recent Financial Crisis." *Journal of Accounting & Finance (2158-3625)* 16.1 (2016).

Saleh,A. Halili,E.Zeitun,R & Salim,R.(2017). Saleh, Ali Salman, et al. "Global financial crisis, ownership structure and firm financial performance: An examination of listed firms in Australia." *Studies in Economics and Finance* (2017).