

# Performance of financial distress prediction models: Evidence from bankruptcy petition filings in India

Manish K. Singh & Surbhi Bhatia

mks@iitd.ac.in  
surbhi.b@livemint.com

Finance Research Group (FRG), Indira Gandhi Institute of Development Research  
www.ifrogs.org

September 11, 2020

# Overview

Background

Literature

Data

Empirical analysis

Part I: Selection of relevant variables

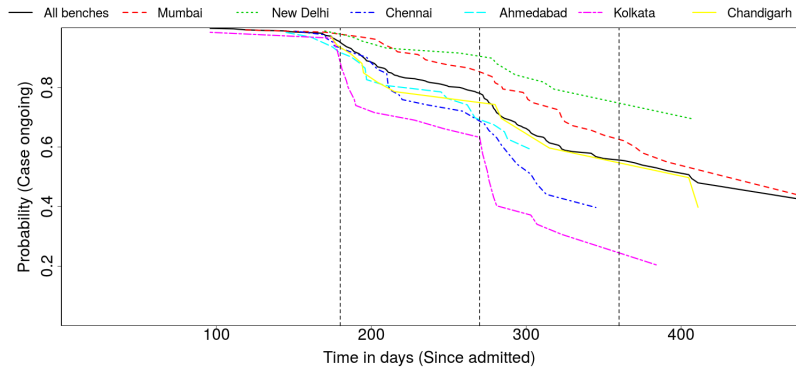
Part II: Performance of default prediction models

Conclusion

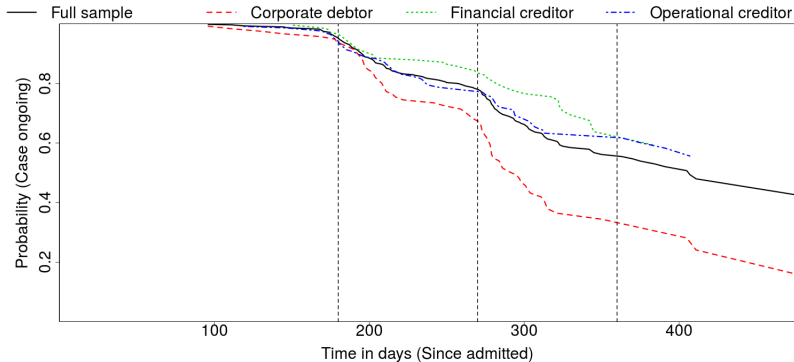
# Institutional background

- ▶ In May 2016, Insolvency and Bankruptcy Code (IBC) was passed
  - ▶ Replaced Sick Industrial Companies Act, 1985 (SICA), Recovery of Debt due to Banks and Financial Institutions Act, 1993 (RDDBFI), Corporate Debt Restructuring, 2001 (CDR), and Securitization and Reconstruction of Financial Assets and Enforcement of Security Interest, 2002 (SARFAESI);
  - ▶ Unified law for all non-financial firms.
- ▶ The National Company Law Tribunal (NCLT) and the National Company Law Appellate Tribunal (NCLAT) were vested with the power to adjudicate on matters related to bankruptcy
  - ▶ It laid down stipulated timeline for speedy disposal and resolution;
  - ▶ Created data infrastructure to store and disseminate case level records.

# Performance: By bench



# Performance: By type of creditor



# In search of a problem statement

- ▶ A rough breakdown of cases registered and their current status as on 30.06.2018: (based on hand-collected FRG dataset)
  - ▶ Total filed: 6,678
  - ▶ Decision pending (admit or dismiss): 3,379
  - ▶ Withdrawn: 477
  - ▶ Dismissed: 2,225
- ▶ Number of cases filed are not the same as number of cases admitted.
  - ▶ Multiple cases filed against the same debtor;
  - ▶ Issues with recording data (missing dates, names).

# Problem statement

How to estimate the expected workload pertaining to the IBC?

## 1. Provide full sample forecast

<b>Financial year</b>	<b>Actual # of cases</b>
2016-17	40 (Jan-Mar 2017)
2017-18	573
2018-19	116 (Apr-Jun 2018)
	<b>Expected # of cases</b>
2018-19	?
2019-20	?
2020-21	?
2021-22	?
2022-23	?

## 2. If possible, bench wise expected caseload

# An easy answer

Extrapolate based on the past trend

- ▶ Case admission rate - Case disposal rate;
- ▶ Current backlog.



# Our approach

Medium term estimate (1 to 5 years) based on

1. Expected probability of default for each firm  $i$ ,  $P(def_i)$ ;
2. Based on international comparison or value judgement, assume firm default above certain threshold ( $P(def_i) \geq T$ );
3. Bench ( $j$ ) jurisdiction;

$$E(Bench\_Workload_j) = Count\{Firm\_Location_{i,j}((P(def_i) \geq T))\}$$

where,  $i = 1, 2, 3, 4, \dots, N$  and  $N$  is the total number of firms in the economy.

Estimating probability of default of Indian firms,  $P(def)$

# Literature

## Default prediction

### 1. Variables that matter for distress prediction

- ▶ Using accounting-based financial ratios
  - ▶ Univariate studies: [Bureau of Business Research(1930)], [Patrick(1932)], [Smith and Winakor(1935)]
  - ▶ Multivariate studies: [Beaver(1966)], [Altman(1968)], [Ohlson(1979)]
- ▶ Using market-based indicators (stock return, bond yield, CDS)
  - ▶ [Merton(1974)], [Shumway(2001)]

### 2. Multivariate models to predict distress

- ▶ Linear discriminant analysis - [Altman(1968)]
- ▶ Logit models - [Ohlson(1979)]
- ▶ Hazard models - [Shumway(2001)]
- ▶ CART (Classification and regression tree) - [Li et al.(2010) Li, Sun and Wu]
- ▶ Neural network - [Shah and Murtaza(2000)]
- ▶ LASSO - [Tibshirani(1996)]
- ▶ ...

# Data

## List of defaulted manufacturing firms

- ▶ We use the Finance Research Group (FRG) hand collected dataset;
- ▶ Based on cases filed with the first nine NCLT benches;
- ▶ We identify, for the admitted cases, a unique list of debtor firms;
- ▶ We have information on when the case was admitted, by whom it was filed (whether creditor or debtor itself), the type of creditor (whether financial or operational), the status of the case (whether admitted or dismissed), and reason for dismissal.
- ▶ Total 714 unique firms in our sample;
- ▶ We remove all firms who have filed for bankruptcy under previous bankruptcy regime (BIFR, CDR or DRT); (370 firms)
- ▶ We match the remaining firms with ProwessIQ dataset; (170 firms)
- ▶ We consider only the manufacturing firms; (41 firms)
- ▶ *Note: Our full sample consists of total manufacturing firms in CMIE ProwessIQ dataset (4312 firms).*

# Correcting the sample bias: Matching

- ▶ Only 1% of firms in our sample have filed for bankruptcy. To balance this skewed sample, we do matching.
  - ▶ 1-to-1 matching;
  - ▶ Without replacement;
  - ▶ Nearest distance (using propensity score);
  - ▶ Based on:
    - ▶ Industry sub-classification within manufacturing;
    - ▶ Ownership = 1 for public limited companies, 2 for private limited companies and 3 otherwise;
    - ▶ Business House = 1 if the firm belongs to a business group and 0 otherwise;
    - ▶ Listed = 1 if the firm is listed on NSE/BSE and 0 otherwise;
    - ▶ Total assets;
    - ▶ Total sales;

Quality of match

# Data

## List of accounting variables

- ▶ Financial ratios from Altman-Z score model:
  - ▶ R1 for liquidity ( $\text{Working capital} / \text{Total assets}$ );
  - ▶ R2 for profitability ( $\text{Retained earnings} / \text{Total assets}$ );
  - ▶ R3 for efficiency ( $\text{Earnings before Interest and Tax (EBIT)} / \text{Total assets}$ );
  - ▶ R4 for leverage ( $\text{Market value of equity} / \text{Book value of total liabilities}$ ); and
  - ▶ R5 for asset turnover ( $\text{Sales} / \text{Total Assets}$ )
- ▶ Financial ratios from Sarkar and Thomas (2003) paper;
- ▶ A few other variables from CMIE ProwessIQ dataset;
- ▶ A few accounting ratios based on value judgement;
- ▶ In total, we have 23 ratios (19 accounting ratios while 4 identity variables)

# Accounting ratios

Description	Measure of
Raw material turnover	Asset turnover ratio
Total income/Compensation to employees	Asset utilisation ratio
Total sales/Net fixed asset	Asset utilisation ratio
Profit after tax/Total income	Net profit margin
PBDITA <sup>1</sup> /Total sales	Operating profit
PBDITA/Total income <sup>2</sup>	Profitability
PBDITA/Net Fixed Assets	Return over investments
Profit after tax/Total assets	Returns on total assets
Quick ratio	Short-term liquidity
Cash profit/Total income	Short-term liquidity
Current ratio	Short-term liquidity
Interest coverage ratio	Short-term liquidity
Cash to current liabilities	Short-term liquidity
Total outside liabilities/Total net worth	Tangibility
Total term liabilities/Tangible net worth	Tangibility

## Some derived indicators

R1A = Gross working capital/Total assets (Cost of sales method)	Liquidity
R1B = Net working capital/Total assets (Cost of sales method)	Liquidity
R1C = Net working capital/Total assets	Liquidity
R2 = Retained profits/Total assets	Profitability
R3 = (PBDITA - Depreciation and amortisation of fixed assets)/Total assets	Efficiency
R5 = Total sales/Total assets	Asset turnover
Non current liabilities/Total assets	Leverage

<sup>1</sup>Profit before depreciation, interest, tax and amortization

<sup>2</sup>It excludes prior period and extra-ordinary transactions.

# Empirical estimation

- ▶ Part I: Selection of relevant variables in the Indian context
  1. Univariate logit models;
  2. LASSO and elastic net;
  3. Principal Component Analysis.
- ▶ Part II: Default prediction models comparison
  1. Single variable analysis: Logit model estimates using leading indicators;
  2. Linear Discriminant Analysis (LDA);
  3. Logistic regression (Probit model);
  4. Least absolute shrinkage and selection operator (LASSO).

## Part I: Selection of relevant variables

1. Univariate logit models;
2. LASSO and elastic net;
3. Principal Component Analysis.



# Part I: Selection of relevant variables

## 1.1 Univariate logit model

Variable	Coefficient	p-value	HLteststats
R1C = Net working capital/Total assets	-0.02**	0.02	0.00
R2 = Retained profits/Total assets	-0.04**	0.02	0.00
R3 = PBDITA-DA/Total assets	-0.05**	0.02	0.00
R5 = Total sales/Total assets	-0.01***	0.02	0.00
Cash to current liabilities	-4.15*	0.07	0.00

*Note:*

\* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$

# Part I: Selection of relevant variables

## 1.2 LASSO and Elastic Net

Variable	Lasso	Elastic net
R1C = Net working capital/TA	-0.001	-0.002
R2 = Retained profits/TA		-0.0001
R3 = (PBDITA - Depreciation and amortisation of fixed assets)/TA	-0.001	-0.004
R5 = Total sales/TA	-0.001	-0.001
Cash to current liabilities		-0.192

*Note* - TA: Total Assets

# Part I: Selection of relevant variables

## 1.3 Principal Component Analysis

	PC1	PC2	PC3	PC4	PC5
Standard deviation	2.402	1.612	1.562	1.400	1.326
Proportion of Variance	0.262	0.118	0.111	0.089	0.080
Cumulative Proportion	0.262	0.380	0.491	0.580	0.660

# Part I: Selection of relevant variables

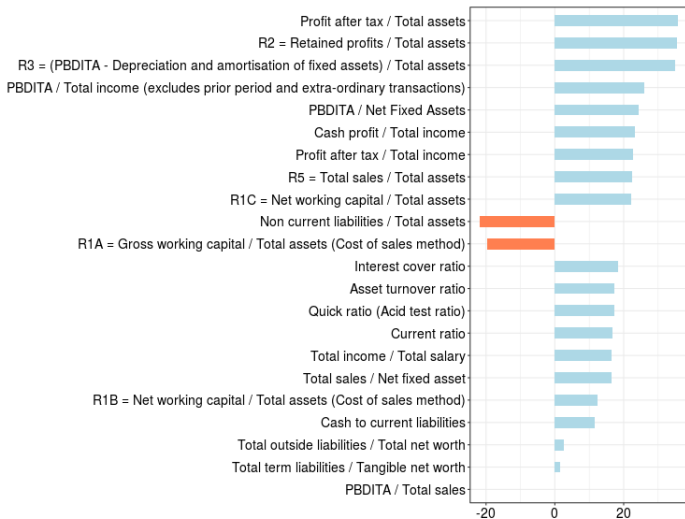
## 1.3 Principal Component Analysis

<i>Dependent variable: Default status</i>					
	(1)	(2)	(3)	(4)	(5)
PC1	-0.138** (0.064)	-0.138** (0.064)	-0.142** (0.064)	-0.146** (0.065)	-0.150** (0.068)
PC2	0.094 (0.095)	0.069 (0.099)	0.066 (0.101)	0.058 (0.101)	0.073 (0.110)
Constant	0.002 (0.142)	-0.081 (0.166)	0.123 (0.277)	0.030 (0.316)	0.273 (0.593)
Ownership	No	Yes	Yes	Yes	Yes
Listed	No	No	Yes	Yes	Yes
Entity type	No	No	No	Yes	Yes
Industry type	No	No	No	No	Yes
Observations	82	82	82	82	82
Log Likelihood	-53.599	-53.113	-52.706	-52.526	-50.732
Akaike Inf. Crit.	113.197	114.226	115.412	117.051	129.464

*Note:* \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$  The coefficients and standard error for other principal components is not reported here to save space but is available upon request to authors.

# Part I: Selection of relevant variables

## 1.3 Principal Component Analysis: Eigenvalues for PC1



# Part I: Selection of relevant variables

In summary

- ▶ Univariate logit model
  - ▶ **R1C, R2, R3, R5** and cash to current liability
- ▶ LASSO and elastic net
  - ▶ **R1C, R2, R3, R5** and cash to current liability
- ▶ Principal Component Analysis (PCA)
  - ▶ PAT/TA, **R2, R3**, PBDITA/Total Income (TI), PBDITA/Net Fixed Assets, Cash profit/TI, PAT/TI, **R5, R1C**, Non-current liabilities/TA, R1A

## Part II: Performance of default prediction models

1. Single variable analysis: Probit model estimates using leading indicators;
2. Linear Discriminant Analysis (LDA);
3. Logistic regression (Probit model);
4. Least absolute shrinkage and selection operator (LASSO).

# Part II: Performance of default prediction models

## 2.1 Single variable prediction

Variable	Actual	Predicted				Area under the curve	
		In-sample No	In-sample Yes	Full-sample No	Full-sample Yes	In-sample	Full-sample
R1C	No	<b>0.78</b>	0.22	0.62	0.38	0.72	0.65
	Yes	0.34	0.66				
R2	No	0.66	0.34	0.69	0.31	0.69	0.65
	Yes	0.42	0.59				
R3	No	0.73	0.27	0.69	0.31	0.70	0.68
	Yes	0.42	0.59				
R5	No	0.54	0.46	0.64	0.36	0.68	0.72
	Yes	0.34	0.66				
Cash to current liabilities	No	0.51	0.49	0.39	0.61	0.73	0.64
	Yes	0.17	<b>0.83</b>				



## Part II: Performance of default prediction models

### 2.2 Modified Altman z-score model: Linear discriminant analysis (LDA)

Description	LD1	Pr(>F)
R1C = Net working capital/Total assets	-0.0035***	0.01
R2 = Retained profits/Total assets	-0.0267**	0.02
R5 = Total sales/Total assets	-0.0074***	0.01
Cash to current liabilities	-2.5431**	0.03
Total sales/Net fixed asset	-0.0716**	0.04
Total term liabilities/Tangible net worth	0.1708*	0.09

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

$$\begin{aligned} \text{Z-score} = & 0.0035 \text{ R1C} + 0.0267 \text{ R2} + 0.0074 \text{ R5} \\ & + 2.5431 \text{ Cash to current liability} \\ & + 0.0716 \text{ Total sales/Net fixed asset} \\ & - 0.1708 \text{ Total term liability/Tangible net worth} \\ & (\text{Z-score range: } -5.79 \text{ to } 6.34) \end{aligned}$$

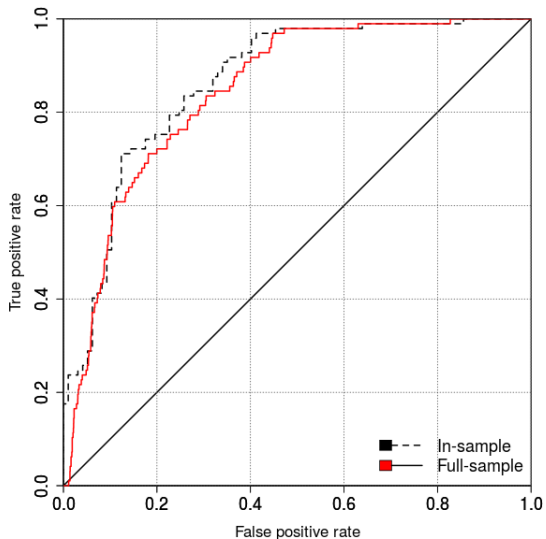
## Part II: Performance of default prediction models

### 2.2 Modified Altman z-score model: Linear discriminant analysis (LDA)

	<b>Predicted</b>			
	In-sample		Full-sample	
	No	Yes	No	Yes
<b>Actual</b>				
No	70.73	29.27	71.32	28.68
Yes	29.27	70.73	29.27	70.73
<b>AUC</b>	0.84		0.80	

# Part II: Performance of default prediction models

## 2.2 Modified Altman z-score model: Linear discriminant analysis (LDA)



## Part II: Performance of default prediction models

### 2.3 Probit model: List of significant variables

Variable	Coefficient	pvalue	HLteststats
R1C = Net working capital / Total assets	-0.010**	0.025	0.092
R2 = Retained profits / Total assets	-0.022**	0.015	0.521
R3 = (PBDITA - D&A) / Total assets	-0.027**	0.017	0.630
R5 = Total sales / Total assets	-0.005**	0.022	0.688
Total sales / Net fixed asset	-0.061*	0.073	0.637
Profit after tax / Total assets	-0.022**	0.027	0.374
Cash to current liabilities	-2.442*	0.068	0.141

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

# Part II: Performance of default prediction models

## 2.3 Probit model

- ▶ Probit (selected variables)

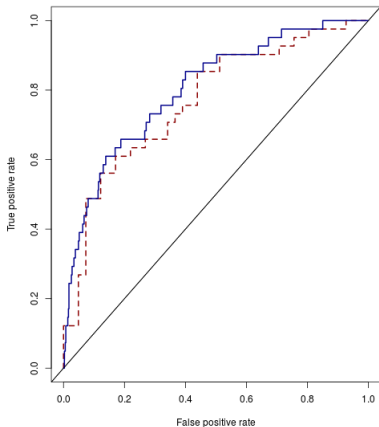
	Predicted			
	In-sample		Full-sample	
Actual	No	Yes	No	Yes
No	69	31	75	25
Yes	34	66	34	66
AUC	0.84		0.80	

- ▶ Probit (All variables)

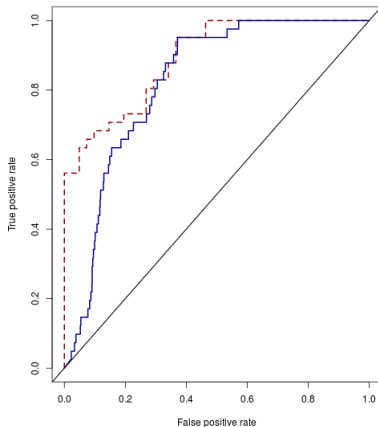
	Predicted			
	In-sample		Full-sample	
Actual	No	Yes	No	Yes
No	81	19	73	27
Yes	29	71	29	71
AUC	0.89		0.82	

# Part II: Performance of default prediction models

## 2.3 Probit model



Probit (Selected variables)



Probit (All variables)

## Part II: Performance of default prediction models

Multivariate prediction: Lasso and Elastic net

▶ LASSO: Accuracy 51%

		Actual	
		No	Yes
Prediction	No	41	40
	Yes	0	1

▶ Elastic net: Accuracy 64%

		Actual	
		No	Yes
Predicted	No	40	28
	Yes	1	13

▶ Ridge: Accuracy 53%

		Actual	
		No	Yes
Prediction	No	40	37
	Yes	1	4

## Part II: Performance of default prediction models

In summary

Table: Goodness of fit statistics

Variable	Area under the ROC	
	In-sample	Full-sample
R1C = Net working capital / Total assets	0.72	0.65
R2 = Retained profits / Total assets	0.69	0.65
R3 = (PBDITA - D&A) / Total assets	0.70	0.68
R5 = Total sales / Total assets	0.68	0.72
Cash to current liabilities	0.73	0.64
Modified Altman z-score (LDA)	0.84	0.80
Probit (Selected ratios)	0.84	0.80
Probit (All ratios)	0.89	0.82



# Conclusion

## 1. Baseline model (based on our analysis)

- ▶ How is our model different from the benchmark model?

- ▶ Altman z-score (1968):

$$Z = 0.012 R1 + 0.014 R2 + 0.033 R3 + 0.006 R4 + 0.99 R5$$

- ▶ Our model

$$Z = 0.0035 R1C + 0.0267 R2 + 0.74 R5 + 2.5431 C/CL + 0.0716 TS/NFA - 0.1708 TTL/TNW$$

where,

C/CL: Cash to current liability

TS/NFA: Total sales/Net fixed asset

TTL/TNW: Total term liability/Tangible net worth.

# Conclusion

## 2. On model accuracy

Table: Confusion matrix - In-sample model comparison

---

	<b>Predicted</b>			
	Altman Z		Our model	
	No	Yes	No	Yes
<b>Actual</b>				
No	97	3	71	29
Yes	6	94	29	71

---

# Future work

How to increase the classification accuracy of our model?

- ▶ Variable selection
  - ▶ Add additional variables: For example, accounting for the macroeconomic environment might help;
  - ▶ We would like to explore the multicollinearity in detail. The more number of variable we look for, the more is the data drop in terms of number of bankrupt firms in our sample.
  - ▶ Timely availability of data is of concern as well. Can we use 2,3,4,... years old data to predict distress! When firms release their most recent data is also of concern.
- ▶ Search for better methodology
  - ▶ A better matching technique;
  - ▶ More advanced classification techniques.

Thank You!

Questions


# Quality of match

Description	Mean	All firms (4,254)		Defaulted firms (41)			Matched firms (41)		
		SD	Median	Mean	SD	Median	Mean	SD	Median
<b>Matched variables</b>									
Total sales	10841.53	90934.91	1625.70	14718.76	45657.32	1727.60	6965.26	15094.70	1332.80
Total Assets (TA)	12504.36	101229.80	1499.60	53304.37	149839.23	5132.70	18812.71	75759.84	2622.50
<b>Other financial ratios</b>									
R1A = Gross working capital <sup>3</sup> / TA	57.00	146.69	48.57	67.21	55.07	51.07	60.80	39.13	56.18
R1B = Net working capital <sup>4</sup> / TA	29.64	157.46	25.94	20.23	92.80	25.26	26.60	47.39	30.16
R1C = Net working capital / TA	6.33	34.82	8.82	-25.29	66.64	-10.00	5.51	34.41	12.69
R2 = Retained profits / TA	0.32	12.79	1.93	-13.97	21.44	-6.28	-4.03	15.29	0.38
R3 = (PBDITA - DA <sup>5</sup> ) / TA	7.22	12.78	7.74	-4.72	15.94	0.05	3.67	14.09	4.72
R5 = Total sales / TA	130.53	98.81	113.01	57.13	55.76	35.25	101.56	96.04	91.75
Total sales / Net fixed asset	11.50	76.78	4.06	3.00	3.90	1.56	5.26	5.88	3.88
Total outside liabilities / Total NW	3.63	139.24	1.72	4.91	21.81	3.00	1.17	14.52	1.27
Cash to current liabilities	0.25	1.00	0.05	0.05	0.10	0.01	0.17	0.32	0.07
Cash profit / Total income	-10.52	192.70	4.47	-240.05	709.52	-15.87	-150.84	764.93	2.94
Current ratio	1.53	2.07	1.20	1.04	1.62	0.71	1.48	1.09	1.24
Interest cover ratio	109.42	4906.67	1.94	-6.15	30.06	-0.54	-71.99	614.54	1.11
Profit after tax / TA	1.26	11.57	2.23	-11.38	17.65	-5.47	-3.47	15.13	0.37
Profit after tax / Total income	-23.48	377.44	1.65	-376.37	1226.58	-17.22	-177.74	882.84	0.22
PBDITA / Total sales	-4.67	362.12	7.80	-0.56	377.39	4.77	3.23	27.03	6.96
PBDITA / Total income <sup>6</sup>	1.15	76.79	7.71	-83.38	259.10	1.48	-27.16	181.15	6.95
Quick ratio (Acid test ratio)	0.89	1.30	0.66	0.54	0.98	0.33	0.81	0.70	0.63
Asset turnover ratio	53.19	1874.68	7.56	6.90	13.12	3.15	7.03	7.60	4.43
Total income / Total salary	35.29	166.21	14.25	22.90	26.65	11.91	21.07	34.45	13.19
Total term liabilities / Tangible NW	0.09	47.61	0.21	2.16	7.31	0.53	-0.24	5.32	0.09

<sup>3</sup>Cost of sales method

<sup>4</sup>Cost of sales method

<sup>5</sup>Depreciation and amortisation of fixed assets.

<sup>6</sup>It excludes prior period and extra-ordinary transactions. 

# Test: Univariate analysis of difference in mean and variance

## Matched sample

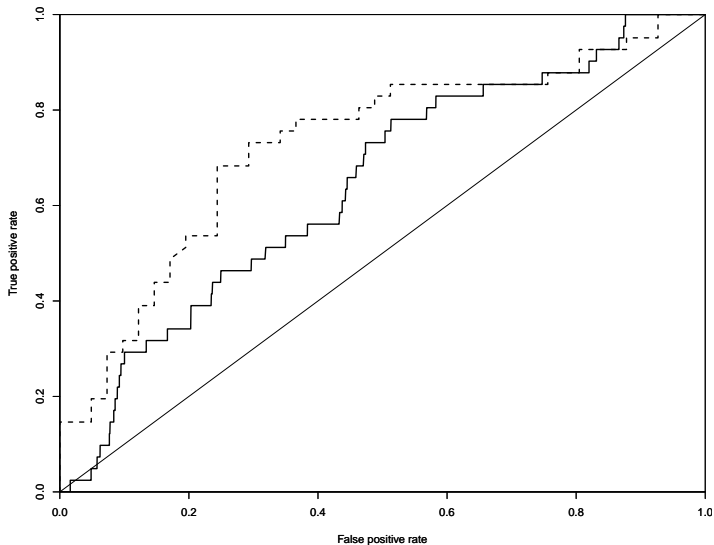
Description	Bartlett test	ANOVA (p-value)
R1A = Gross working capital / Total assets (Cost of sales method)	0.03**	0.55
R1B = Net working capital / Total assets (Cost of sales method)	0.00***	0.70
R1C = Net working capital / Total assets	0.00***	0.01***
R2 = Retained profits / Total assets	0.04**	0.02**
R3 = (PBDITA - Depreciation and amortisation of fixed assets) / Total assets	0.44	0.01
R5 = Total sales / Total assets	0.00***	0.01***
Cash to current liabilities	0.00***	0.03**
Cash profit / Total income	0.64	0.59
Current ratio	0.01***	0.15
Interest cover ratio	0.00***	0.50
Profit after tax / Total assets	0.34	0.03**
Profit after tax / Total income	0.04**	0.40
PBDITA / Total sales	0.00***	0.95
PBDITA / Total income (excludes prior period and extra-ordinary transactions)	0.03**	0.26
Quick ratio (Acid test ratio)	0.04**	0.16
Asset turnover ratio	0.00***	0.96
Total sales / Net fixed asset	0.01***	0.04**
Total outside liabilities / Total net worth	0.01***	0.36
Total income / Total salary	0.11	0.79
Total term liabilities / Tangible net worth	0.05**	0.09*
Non current liabilities / Total assets	0.00	0.83

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

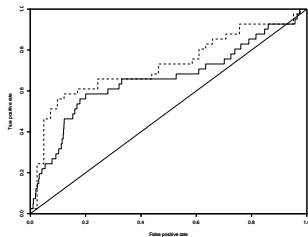
Back to [Matching](#)

# Leading indicators: ROCs

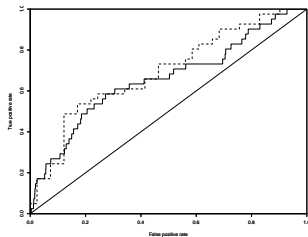


(a) Cash to current liabilities

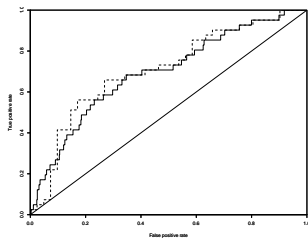
# Leading indicators: ROCs



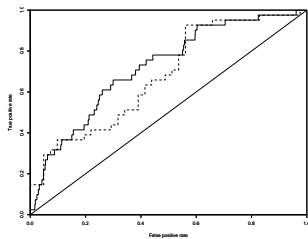
(b) R1C



(c) R2



(d) R3

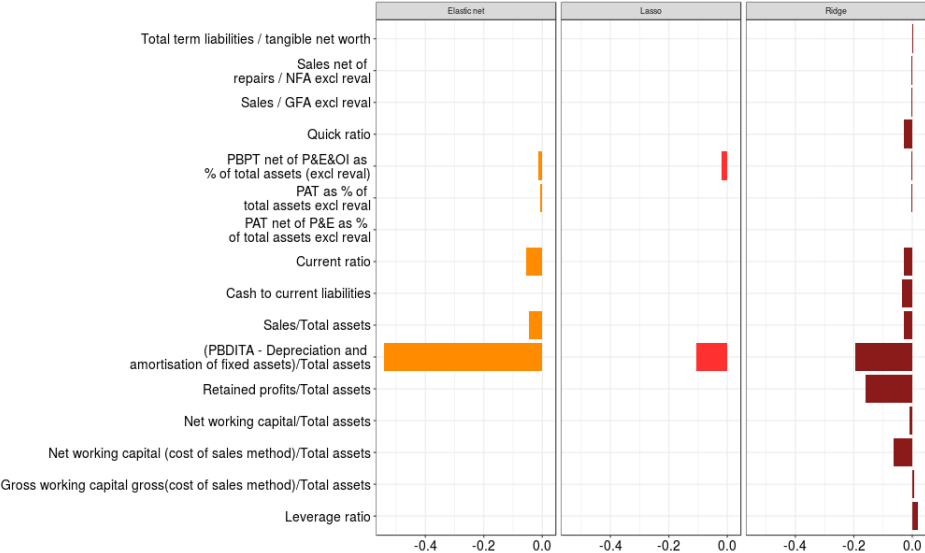


(e) R5









# LASSO



# References I

-  Altman, E., 1968.  
Financial ratios, discriminant analysis and the prediction of corporate bankruptcy.  
[The Journal of Finance 23, 589–609.](#)
-  Beaver, W.H., 1966.  
Financial ratios as predictors of failure.  
[Journal of Accounting Research 4, 71–111.](#)
-  Bureau of Business Research, 1930.  
A test analysis of unsuccessful industrial companies.  
Urbana, University of Illinois.  
URL: <https://books.google.co.in/books?id=F9fcPQAACAAJ>.
-  Li, H., Sun, J., Wu, J., 2010.  
Predicting business failure using classification and regression tree:  
An empirical comparison with popular classical statistical methods  
and top classification mining methods.  
[Expert Systems with Applications 37, 5895–5904.](#)

## References II



Merton, R.C., 1974.

On the pricing of corporate debt: The risk structure of interest rates.  
*The Journal of Finance* 29, 449–470.



Ohlson, J.A., 1979.

Financial ratios and the probabilistic prediction of bankruptcy.  
*Journal of Accounting Research* 18, 109–131.



Patrick, P., 1932.

A comparison of ratios of successful industrial enterprises with those of failed companies.  
*Certified Public Accountant* 2, 598–605.



Shah, J.R., Murtaza, M.B., 2000.

A neural network based clustering procedure for bankruptcy prediction.  
*American Business Review* 18, 80–86.



Shumway, T., 2001.

Forecasting bankruptcy more accurately: A simple hazard model.  
*The Journal of Business* 74, 101–124.

## References III



Smith, R., Winakor, A., 1935.

Changes in financial structure of unsuccessful industrial corporations.

Urbana: University of Illinois.



Tibshirani, R., 1996.

Regression shrinkage and selection via the LASSO.

Journal of the Royal Statistical Society. Series B (Methodological)  
58, 267–288.