BRITISH ACADEMY OF MANAGEMENT

BAM CONFERENCE

3RD-5TH SEPTEMBER
ASTON UNIVERSITY BIRMINGHAM UNITED KINGDOM

This paper is from the BAM2019 Conference Proceedings

About BAM

The British Academy of Management (BAM) is the leading authority on the academic field of management in the UK, supporting and representing the community of scholars and engaging with international peers.

http://www.bam.ac.uk/

An Analysis of Bank Efficiency and Corporate Governance

Using Random Forest Regression for Second Stage DEA

(Corresponding Author)
Keyur Thaker, Associate Professor, F&A Area
Indian Institute of Management, Indore, India
Rau Pithampur Road, Indore, MP, India 453331
E mail: keyurt@iimidr.ac.in, thakerkeyur@yahoo.com

Vincent Charles, Professor of Management Science and Director of Research
Buckingham Business School, University of Buckingham, Buckingham MK18 1EG, United Kingdom
E mail v.charles@bukingham.ac.uk

Abhay Pant, Assistant Professor
Jindal Global Business School, Sonipat, India
E mail f14abhayp@iimidr.ac.in

Abstract

Using the DEA at the first stage, our study examines the Indian bank's technical, cost and profit efficiency across public and private ownership groups for the 2008-16 time period. In the second stage, we examine impact of corporate governance in terms of board characteristics on bank efficiency. We first use the conventional OLS and Fractional regression and then demonstrate the use of methodological superior Random forest regression. We found that the average Technical efficiency was very high (0.957) while cost efficiency was marginally lower (0.947) and profit efficiency was much lower (0.790) for all the banks. Interestingly larger number of banks were on the frontier of profit efficiency followed by technical and cost efficiency. In the second stage, we find Board independence to be stronger predictor of cost efficiency; and all variables namely board size, meetings, gender diversity, board independence has been stronger predictors of profit efficiency. Furthermore, we observe that the RF model explained the highest variance for Profit Efficiency and tuning lead to further improvement. Study point out the limitations of the conventional OLS and fractional regressions and demonstrate the use of methodological superior random forest regression. Our study has important policy and well methodological implications. Moreover this is a pioneering study using the Random Forest to explore bank governance, and its impact on three different types of bank efficiency in India.

An Analysis of Bank Efficiency and Corporate Governance

Using Random Forest Regression for Second Stage DEA

Summarized Paper

Purpose:

The study have following purposes. Firstly to estimate the technical, cost and profit efficiency of Indian banks using Data Envelopment Analysis for a period of 2008-2016. Secondly to examine the difference and trends in the three efficiency across banks from the different ownership groups. Thirdly to investigate the relationship between corporate governance (CG) and bank efficiency. Using the board characteristics variables as proxy of the CG we study its impact on three different efficiency using the conventional bootstrapped OLS and Fractional logistic regression. Fourthly, recognizing the limitation of conventional regression methods in identifying determinants given the statistical properties of dependent variables we pioneer in demonstration of use of methodological superior technique, i.e. Random forest in the context banking sector. Finally we compare the results of Random Forest and Conventional Regressions and argue for the superiority of the RF in the given context.

Design:

In the first stage, we estimate technical, cost, and profit efficiency of 41 Indian (24 public and 17 private) banks using non parametric Data Envelopment Analysis, over the period 2008-2016. Following intermediation approach we consider funds, fixed assets and number of employees as input variables and advances, investment and other income as output variables. We take prices of inputs and outputs in estimation of profit efficiency while only prices of input in cost efficiency. We also examine the difference between and trends in the three types of efficiency scores across the different ownership groups. Then, in the second stage, we investigate the relationship between corporate governance (CG) and bank efficiency. Using the board characteristics variables as proxies for the CG, we study their impact on the three types of efficiency using the conventional bootstrapped OLS and fractional logistic regression. More specifically, we examine the impact of corporate governance (Board Size, Board Independence, Duality, Gender Diversity and Board Meetings), bank characteristics (Return on Assets, Size, and Equity by Total Assets), and other characteristics (Ownership and Years) on bank efficiency. We recognize, however, the limitation of conventional regression methods in identifying determinants given the statistical properties of the dependent variables; as such, we pioneer the use of a methodological superior technique, *i.e.*, random forest, in the context of the banking sector.

Findings:

The DEA analysis performed in the first stage revealed that the average Technical efficiency was very high (0.957) while cost efficiency was marginally lower (0.947) and profit efficiency was much lower (0.790) for all the banks. As can be seen in Table I, while average technical and cost efficiency have remained more or less same during the study period, the overall average profit efficiency of all banks have declined. Interestingly among all the efficiency frontier the highest number of banks were on the Profit Efficiency frontier. Furthermore, while average PE of Public Sector banks was lower than that of private banks, it showed better improvement over study period. Across all three efficiency, we observe private banks to obtain a higher score and have a larger number of frontier banks when

compared to the public sector banks. We observed higher standard deviation in profit efficiency across different banks as compared to technical and cost efficiency.

Table 1 about here

In the second stage, as can be seen in Table 2, we found that among the ten variables, Return on Assets (ROA) and Equity to Total Assets (EQA) to be stronger predictors of technical efficiency; Asset Size, Return on Assets, and Board independence to be stronger predictors of cost efficiency; and all variables namely board size, meetings, gender diversity, board independence has been stronger predictors. While duality and ownership did not explain profit efficiency. We find that Random Forest Regression performs better than OLS Bootstrap and Fractional Logit Bootstrap, providing more convincing results in terms of modelling performance, as well as variable importance. As can be seen in Table 3, we find that RF model explained highest variance for Profit Efficiency and tuning lead to further improvement.

Table 2 about here

Table 3 about here

Overall we also find that larger number of Corporate Governance variables namely board meetings, independence, board size and gender diversity explains Profit efficiency. PE is considered to be a more informational as it considers not only input output variables but also their prices and more appropriate measure of bank performance given the competitive environment and increasing importance of profits as compared to technical efficiency (which do not consider prices of input and output) and cost efficiency (only considers prices of input variables), across both conventional regression and random forest models. Amongst all the variables related to corporate governance we find that Duality was found significant at 6 instances across OLS, Fractional and RF models, followed by board independence and board meetings (5 instances). The table 4 provides the comparisons across three models and three measures of bank efficiency.

Table 4 about here

Implications:

In the context of lower profitability and rising non- performing assets (NPA) of public sector banks and dual regulations set up by Ministry of Finance and Reserve Bank of India (RBI), there is a need to rationalize the regulation and empower and augment the skills of the bank board and for enhanced and effective banks governance and performance. The Reserve Bank of India appointed Nayak committee to review the governance of boards and give recommendation. The committee submitted its report in May 2014 which shall form basis for set of rules for bank Governance that RBI would formulate. Amongst other recommendations they particularly emphasized on the number of board meetings, board composition, board compensation, board independence, board members

with specialist skills, quality of discussion and type of issues discussed in the board meetings, especially the business strategy, risk mgmt. related as important areas where the banks need to work on (Nayak et al 2014). Our study corroborates with some of the recommendations of the committee and also contributes to further understanding of impact of Bank Governance in terms of board characteristics on the bank performance and efficiency. For instance, we found that for profit efficiency, board characteristics play a significant role; therefore, policymakers and regulators should consider board features such as board size, board independence, gender diversity, and board meetings while framing guidelines for enhancing bank performance. At the same time our study found that, board independence significantly impacts bank cost efficiency, while none of the corporate governance variables significantly explain the technical efficiency. Overall indicating that the board characteristics has an important implications for the profit performance of Indian banks.

Our study also pointed out the limitations of the conventional OLS and fractional regressions and demonstrated the use of methodological superior random forest regression. Thus our study have important policy and well methodological implications.

Originality:

First, this study makes methodological contributions by employing machine learning based random forest regression to examine corporate governance and bank efficiency, which is a pioneering attempt. Second, there are hardly any studies that explore the linkages between corporate governance and bank efficiency in the Indian context and our study fills this void. Third, our study examines all three type of efficiency, namely cost, technical and profit efficiency for a 9 year long period while most past studies coved one or two types of efficiency within a shorter time frame. To the best of our knowledge this is the only study in the Indian context that covers a long time period and all three types of efficiency.

Limitations and Future Areas for Research:

While board meetings and composition are important and significant, future studies could examine the types of issues discussed and time spent on those issues, and the role of the directors in the board meetings, among others and its relations with bank efficiency. Our study could not address the qualitative aspects due not being in our scope and also partly due to unavailability of the data. . While Random forest is methodologically superior to conventional regression in number of ways some limitations are, not so strait forward interpretability of results compared to an individual tree and difficulty in clearly identifying most important variables. Our study being pioneer in examining corporate governance and bank efficiency and also use of random forest in banking context we expect it to trigger future research in this direction.

Reference:

P. J. Nayak, S. R., Shubhalakshmi Panse, Pratip Kar, Joydeep Sengupta, Harsh Vardhan, Somasekhar Sundaresan, Krishnamurthy Subramanian. (2014). Report of the Committee to Review Governance of Boards of Banks in India. Mumbai: Reserve Bank of India.

Table 1: Sector level statistics for TE, CE and PE

Sectors	Stats	2008	2009	2010	2011	2012	2013	2014	2015	2016
				Technical		Efficiency				
ALL	EB	13	10	12	11	9	11	12	13	13
	Mean	0.9443	0.9455	0.9584	0.9430	0.9584	0.9636	0.9706	0.9694	0.9642
	SD	0.0573	0.0462	0.0376	0.0450	0.0315	0.0302	0.0309	0.0330	0.0359
	CV*	0.0611	0.0492	0.0394	0.0480	0.0331	0.0316	0.0321	0.0342	0.0374
State	EB	6	5	4	3	4	4	5	6	5
	Mean	0.9473	0.9432	0.9528	0.9388	0.9557	0.9632	0.9704	0.9711	0.9553
	SD	0.0417	0.0414	0.0381	0.0368	0.0286	0.0255	0.0241	0.0269	0.0349
	CV*	0.0444	0.0444	0.0404	0.0396	0.0303	0.0268	0.0251	0.0280	0.0369
Private	EB	7	5	8	8	5	7	7	7	8
	Mean	0.9400	0.9488	0.9662	0.9490	0.9622	0.9641	0.9708	0.9670	0.9768
	SD	0.0754	0.0535	0.0364	0.0552	0.0358	0.0367	0.0395	0.0409	0.0344
	CV*	0.0814	0.0572	0.0383	0.0591	0.0378	0.0387	0.0412	0.0429	0.0357
				Cost		Efficiency				
ALL	EB	12	12	8	9	6	6	11	10	8
	Mean	0.9546	0.9505	0.9381	0.9277	0.9417	0.9555	0.9585	0.9541	0.9434
	SD	0.0464	0.0470	0.0573	0.0542	0.0471	0.0416	0.0420	0.0406	0.0468
	CV*	0.0489	0.0497	0.0615	0.0587	0.0504	0.0438	0.0441	0.0428	0.0499
State	EB	4	4	2	2	3	3	6	4	1
	Mean	0.9466	0.9458	0.9392	0.9254	0.9468	0.9566	0.9643	0.9581	0.9372
	SD	0.0412	0.0337	0.0378	0.0379	0.0340	0.0300	0.0287	0.0336	0.0443
	CV*	0.0440	0.0360	0.0407	0.0414	0.0363	0.0317	0.0301	0.0354	0.0478
Private	EB	8	8	6	7	3	3	5	6	7
	Mean	0.9659	0.9572	0.9366	0.9311	0.9345	0.9540	0.9502	0.9485	0.9521
	SD	0.0521	0.0617	0.0784	0.0724	0.0616	0.0551	0.0557	0.0495	0.0500
	CV*	0.0547	0.0654	0.0850	0.0789	0.0669	0.0586	0.0595	0.0529	0.0533
				Profit		Efficiency				
ALL	EB	14	12	13	11	10	13	13	17	14
	Mean	0.7866	0.8117	0.7711	0.7358	0.7666	0.8291	0.8299	0.8262	0.7603
	SD	0.2506	0.2232	0.2324	0.2193	0.2365	0.2145	0.2213	0.2234	0.2839
	CV*	0.3205	0.2766	0.3033	0.2998	0.3103	0.2603	0.2682	0.2720	0.3756
State	EB	7	5	5	3	4	6	5	9	5
	Mean	0.7116	0.7285	0.7308	0.6904	0.7471	0.8237	0.8206	0.8334	0.7447
	SD	0.2447	0.2178	0.2191	0.1818	0.1852	0.1873	0.1840	0.2002	0.2468
	CV*	0.3475	0.3021	0.3029	0.2661	0.2505	0.2297	0.2266	0.2427	0.3349
Private	EB	7	7	8	8	6	7	8	8	9
	Mean	0.8925	0.9292	0.8280	0.7999	0.7942	0.8367	0.8431	0.8159	0.7824
	SD	0.2248	0.1767	0.2453	0.2554	0.2985	0.2540	0.2709	0.2588	0.3362
	CV*	0.2556	0.1930	0.3007	0.3239	0.3814	0.3080	0.3260	0.3219	0.4360

Table 2: Bootstrap Regression

	TE[1]	CE[1]	PE[1]	TE[2]	CE[2]	PE[2]
Size[G]	0.0010059	-0.0001334	-0.0194**	-0.0165585	0.0098453	-0.1837**
	(0.0008396)	(0.0022149)	(0.0083566)	(0.0840642)	(0.0380198)	(0.0785735)
Indep[G]	0.0003**	-0.0001905	-0.0029**	0.0004843	0.0016864	-0.0505***
	(0.0001231)	(0.0003681)	(0.0014141)	(0.0141071)	(0.0061431)	(0.0183838)
Meetings[G]	-0.0004363	-0.0024**	-0.0052096	-0.0682*	-0.0394**	-0.1005***
	(0.0003443)	(0.0011384)	(0.0032763)	(0.0394119)	(0.0170128)	(0.0344792)
G-Div[G]	-0.0000370	0.0007602	-0.0038*	-0.0163499	0.0109769	-0.0159073
	(0.0002199)	(0.0007057)	(0.0019504)	(0.023061)	(0.0091582)	(0.0156167)
Duality[G]	-0.0105***	0.0523***	0.0872**	-0.9757**	0.5916***	0.7265**
	(0.0039376)	(0.0091229)	(0.0356195)	(0.4713115)	(0.1398428)	(0.3486627)
ROA[BC]	0.0126277	0.0230712	0.1565233	0.3539192	0.2568310	1.1694370
	(0.0029899)	(0.0085622)	(0.0236376)	(0.2715565)	(0.1226172)	(0.2171403)
Size[BC]	0.0002382	0.0150216	0.0818325	0.6437869	0.2792868	1.0984570
	(0.0014469)	(0.0042094)	(0.0126917)	(0.3706265)	(0.0862633)	(0.2907785)
EQA[BC]	0.0020183	0.0028035	0.0243519	0.6776551	0.0079324	0.4036947
	(0.0004485)	(0.0033142)	(0.0040431)	(0.2411413)	(0.0657606)	(0.1181632)
Ownership[BC]	-0.0244179	0.0284290	0.1424963	-0.4277097	0.4731351	3.9158000
	(0.0070707)	(0.0232847)	(0.0835663)	(0.9466662)	(0.4406857)	(1.365333)
YearOC	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R2	0.2197	0.2223	0.3374			
Wald χ^2	122.77	210.46	313.78			
RMSE	0.0236	0.0698	0.2345	3.7595	0.9236	4.4549
N	362	362	362	362	362	362

N - Number of observations, 1 - OLS Bootstrap, 2 - Fractional Logit Bootstrap, G - Governance, BC - Bank characteristics, and OC - Other Characteristics

Table 3: Random Forest Statistics

Type	RF	ntree	mtry	MSE	%VE	Cor	Min-Max Accuracy
TE	Initial	3000	3	0.00057	20.17	0.5564	0.9817
	Tunning	2840	2	0.00052	20.71	0.5685	0.9841
CE	Initial	3000	3	0.00529	24.88	0.5189	0.9374
	Tunning	1170	3	0.00433	24.97	0.5273	0.9380
PE	Initial	3000	3	0.05564	30.21	0.6647	0.7482
	Tunning	1610	8	0.05367	32.05	0.6444	0.7570

RF - Random Forest, ntree - Number of decision tress, mtry - Number of features in the dataset, %VE - Percentage of variation explained, and Cor - Correlation

Table 4: Factors Explained$

		Indepen-dence	Board Meetings	Gender Diversity	Board Size	Duality	*Total*
EFFICIENCY							
Technical	OLS	Yes				Yes	*2*
	Fractional		Yes			Yes	*2*
	Random Forest						*0*
Total							*4*
Cost	OLS		Yes			Yes	*2*
	Fractional		Yes			Yes	*2*
	Random Forest	Yes					*1*
Total							*5*
Profit	OLS	Yes		Yes	Yes	Yes	*4*
	Fractional	Yes	Yes		Yes	Yes	*4*
	Random Forest	Yes	Yes	Yes	Yes		*4*
Total							*12*
	Total	*5*	*5*	*2*	*3*	*6*	
OLS							*8*
Fractional							*8*
RF							*5*

$ Predicted in case of RF