ABSTRACT

The paper analyses the relative efficiency scores of 57 private, public and foreign scheduled commercial banks from eleven different variable returns to scale Data Envelopment Analysis (DEA) models with input orientation. The results of Decision Tree (DT) Analysis indicate that two out of the eleven models are the most appropriate to measure relative efficiency when intermediary approach is considered to evaluate the banks. This leads to the conclusion that the management of the banks in our study should focus their efforts and resources on the input-output factors from these two models to improve efficiency and increase profitability.

KEYWORDS: Data Envelopment Analysis; Data Envelopment Analysis in Banking; Efficiency of Banks; Banking in India; Decision Tree Technique

INTRODUCTION

Studies on efficiency measurement of financial institutions, banking in particular, are widely available. They mostly focus on institutions in developed counties. In recent years more research is being conducted employing Data Envelopment Analysis (DEA) with several combinations of various input and output factors to assess the technical efficiency of Decision Making Units (DMUs) in Indian banking sector (e.g., Bhattacharya et al., 1997; Saha and Ravishanker, 2000; Mukherjee et al., 2002; Sathye, 2003; Kumbhakar and Sarkar, 2005; Das and Ghosh, 2006; Sahoo and Tone, 2008, Kumar and Gulati, 2008, Zhao et al., 2008). While these studies have examined the technical efficiency of banks in India across time, the most challenging aspect of estimating efficiency using DEA methodology, when there are many input and output combinations to consider, is to select appropriate and relevant input and output factors. Only broad guidelines and screening procedures are mentioned in the literature for selecting the inputs and outputs in a DEA study. The choice of inputs and outputs largely affects the derived efficiency level and the decisions thereof. In the literature, there is no universal agreement on what constitutes inputs and outputs in banking services.
A reference to the literature on application of DEA to the Indian banking sector, reveals that the choice of input and output factors were primarily based on the judgment of the researcher(s) and/or based on the literature references. Thus, they lack the rigor and scientific justification pertaining to the appropriateness of or the efficacy of such input-output factors to measure the true technical efficiency of the DMUs/Banks. Differences in the selection of input-output factors may lead to different efficiency scores which eventually lead to erroneous evaluation of true efficiency of the banking sector. For instance, greater the number of inputs or outputs, more the number of DMUs with efficiency rating of 1 because the DMU can become too specialized to be evaluated with others. Moreover, in banking, certain variables can be considered as both inputs and outputs. Therefore, there is a need to apply scientific methods in selecting the appropriate input–output variables in a DEA study to avoid subjectivity and selection bias of the researcher. This paper, addresses this important issue.

In this context, this paper uses a Decision Tree (DT) approach for a better selection of input-output factors for DEA studies with reference to the Indian banking sector. The present study makes a significant contribution towards the existing literature in the following ways. First, this study evaluates selected DEA models (discussed in detail in the literature review section) with various input–output factors and investigates the relationship between the computed efficiency scores and a single performance criterion of the DMUs by using “Decision Tree” approach, as suggested by (Lim, 2008). In his study, Lim used this approach and validated it with DMUs from South Korean steel industry. This gives us an analytically based criterion for the selection of input-outputs. Second, the present study is based on the data for all private, public and foreign scheduled commercial banks in India for the period of 2000 to 2012.

LITERATURE REVIEW

Data Envelopment Analysis

Data Envelopment analysis (DEA), a non-parametric performance assessment methodology, was developed by Charnes, Cooper and Rhodes (Charnes et al., 1978) to measure the relative efficiencies of organizational units or decision making units (DMUs) under evaluation [in the present case the banks] from the data for the same set of inputs and outputs. This technique aims to measure how efficiently a DMU (bank) uses the resources (inputs) available to generate a set of outputs. The DEA approach applies linear programming techniques to construct an efficient production frontier based on best practices over the data set. Each DMU’s efficiency is then measured relative to this frontier. A majority of the studies related to efficiency in banking sector throughout the world have used DEA.

In the last few decades there have been significant developments in the basic DEA model (Charnes et al., 1978), - popularly referred to as CCR model - as several variations of it such as the BCC model (Banker, Charnes and Cooper, 1984), the additive model, Banker & Morey model (1986) to name a few, have emerged improvising on and supplementing the CCR model. For instance, the BCC model (see Appendix) allows for variable returns to scale (VRS) while constructing the efficient frontier thus discriminating efficiency of scale, from pure technical efficiency, while CCR model assumed constant returns to scale. Banker & Morey model involves qualitative inputs and outputs unlike other variations which allow only quantitative inputs and outputs.
Efficiency of Banks in India

There have been many studies pertaining to developed countries of efficiency measurement of financial institutions and banking in particular. Berger and Humphrey (1997) made an extensive survey on the then existing literature of financial institutions on the efficiency measurement and found that around 75% of the studies were focused on banking efficiency measurement particularly of US banking industry and only around five percent of the total studies examined the efficiency of banking sector in developing countries. The more recent study of Fethi and Pasiouras (2010) examined the efficiency studies only of banking industry in various research journals during the period of 1998-2008. They examined 179 studies and found that more than 75% of studies were from developed countries. In the Indian context, only a few studies are on banking efficiency measurement. In this literature review we focus mainly on the literature on efficiency measurement on Indian banking sector mentioned in Fethi and Pasiouras (2010) along with some other cited papers.

Bhattacharya et al. (1997) examined the efficiency of Indian banks using a two step procedure, DEA technique to determine the technical efficiency and then applying stochastic frontier approach to explain variation in calculated efficiency. They applied intermediation approach using two inputs (interest expenses and operating expenses) and three outputs (deposits, advances and investments) of 70 banks, for the period 1986-1991. They constructed one grand frontier on the entire data set for DEA analysis and found that the public sector banks were more efficient than the foreign banks, which in turn were marginally more efficient than private sector banks. After performing a regression analysis in the second stage, they concluded that public sector bank efficiency declined over time whereas that of the foreign banks improved over time. The performance of the Indian private sector banks remained almost unchanged. Saha and Ravishanker (2000) studied the 25 Public Sector Banks (PSB) in India for the period 1992 to 1995 and provided a ranking based on DEA efficiency score. They used number of branches, number of employees, establishment expenses and non-establishment expenses as inputs and deposits, advances, investments, spread, total income, interest income and non-interest income as outputs. They found that the efficiency, in general, of PSBs improved over the time period of their study.

Mukherjee et al. (2002) explored technical efficiency and benchmarked the performance of 68 commercial banks using DEA for the period 1996-1999. They observed that in India, public sector banks (PSBs) are more efficient than both private and foreign banks. Also, the performance of PSBs improved over the study period.

Sathye (2003) measured the productive efficiency of 94 Indian banks (27 public sector commercial banks, 33 private sector commercial banks and 34 foreign banks) by using variable returns to scale input oriented model of the DEA methodology. Two models were constructed to show how efficiency scores vary with change in inputs and outputs. The first model used interest expenses, non-interest expenses as inputs and net interest income and non-interest income as outputs. The second model used deposits and staff numbers as inputs and net loans and non-interest income as outputs.

The study showed that the mean efficiency scores of Indian PSBs were higher than that of the private sector and foreign commercial banks in India and Indian PSBs compared well with the world mean efficiency score. The study recommends that the existing policy of reducing non-
performing assets and rationalization of staff and branches may be continued to make the Indian banks internationally competitive.

Kumbhakar and Sarkar (2005) used stochastic frontier analysis (SFA) to evaluate the efficiency of public and private sector banks in India over the period 1986 to 2000. In Indian banking this translated into examining the effect of ownership, and especially the effect of the then deregulation measures. They found that the deregulation had led to an increase in the cost inefficiency of the Indian banks. The study also revealed that the private banks, on average, were generally more cost efficient than public banks.

Das and Ghosh (2006) examined the performance of Indian commercial banking sector during the post reform period of 1992–2002 using non-parametric Data Envelopment Analysis (DEA). Three different approaches viz., intermediation approach, value-added approach and operating approach were employed to differentiate how efficiency scores vary with changes in inputs and outputs such as bank size, ownership, capital adequacy ratio, non-performing loans and management quality. They found that medium-sized public sector banks performed reasonably well and were more likely to operate at higher levels of technical efficiency. The empirical results also showed that technically more efficient banks were those that had, on an average, less nonperforming loans.

Ali and Le (2006) examine the impact of various elements of economic reforms (ER) on the efficiency of banks in India during 1992–1998. Bank efficiency is measured using data envelopment analysis (DEA) and the relationship between the measured efficiency and various bank-specific characteristics and environmental factors associated with the economic reforms were examined using the OLS (ordinary least square) and the GMM (generalized method of moments) estimations. They found that the efficiency of the banking industry improved during the post-ERs era due to the improvement in the efficiency of all three ownership groups, namely: public sector banks; domestic private banks; and foreign banks. They also found a positive relationship between the level of competition and the efficiency of banks.

Sahoo and Tone (2008) studied the capacity utilization of Indian banking industry during 1997 to 2001 by using DEA. They adopted the intermediary approach to study the industry and used fixed assets, borrowed fund and labor as inputs along with investments, performing loan assets and non-interest income as outputs. Their empirical results found that increased competitive pressure after the liberalization of the banking industry helped to reduce excess capacity of the banking industry. Their study also highlighted that the short run cost was higher for the public sector banks than the private sector banks.

Kumar and Gulati (2008) evaluated the technical efficiency of 27 PSBs operating in India and provided a ranking of these banks based on those efficiency scores with the help of two popular data envelopment analysis (DEA) models, namely, CCR model and Andersen and Petersen’s super-efficiency model (Anderson and Peterson, 1993) for the financial year 2004-2005.

Their study found that only seven of the 27 banks were efficient with technical efficiency scores ranging from 0.632 to 1. Andhra Bank was the most efficient bank. The study also found that the banks affiliated with State Bank of India (SBI) group were more efficient than the other public sector banks. The regression results indicated that the exposure to off-balance sheet activities, staff productivity, market share and size were the major determinants of the technical efficiency.
Gupta et al. (2008) analyzed the performance of the Indian banking sector through non-parametric frontier methodology, DEA and found the determinants of productive efficiency through TOBIT model. Inputs (interest expenses and operating expenses) and outputs (interest income, fee based income and investment income) were measured in monetary value and efficiency scores determined for the period 1999-2003. The study found that SBI and its group had the highest efficiency, followed by private banks and the other nationalized banks. The results were consistent over the period, but efficiency differences diminished over period of time.

For measuring productive efficiency through TOBIT, the authors used five independent variables, profitability, productivity, size, regulatory measures and asset quality and found that the capital adequacy ratio had a significantly positive impact on the productive efficiency whereas assets size had no significant influence. Therefore, bank efficiency was independent of the size of the bank.

Zhao et al. (2008) examined the impact of regulatory reform on the performance of Indian commercial banks for a period of 1992 to 2004 using a balanced panel data set and employing a DEA-based Malmquist index of total factor productivity (TFP) change. The empirical results indicated that, after an initial adjustment phase, the Indian banking industry experienced sustained productivity growth, driven mainly by technological progress. Banks’ ownership structure seems to have an impact on bank efficiency but does not appear to have an effect on total factor productivity change. Study revealed that during the deregulation process foreign banks appeared to be technological innovators, thereby increasing even further the competitive pressure in the Indian banking industry.

**Decision Tree Technique**

Decision Tree (DT), is a non-parametric data mining technique used to discover patterns of meaningful relationships and rules from huge databases. In this technique, the data is systematically broken down to classify patterns discovered from the dataset and make rule based predictions (Berry and Linoff, 2000). DT has a structure of a tree representing a given decision problem with target and predictor variables. The tree continues to branch off into child nodes till the terminal node for each branch is reached. Each terminal node represents a decision rule which can be used to predict the value of the target variable from the given values of the predictor variables. DT can be a classification tree, if the target is a discrete variable or a regression tree, if the target is a continuous variable.

There are several DT algorithms namely CHAID (Kass, 1980), CART (Breiman et al, 1984; Ripley, 1996) and C4.5 (Quinlan, 1993). The choice of the algorithm depends upon the combination of factors such as the type of the target variable (discrete or continuous), the split type (binary or multi-way) and the split criterion (Gini index or entropy index or variance reduction).

The application of DT in conjunction with DEA has found many takers. Some studies have used it to predict/forecast, e.g., Lee and Park (2005) used it to identify segments of potentially profitable customers; Shon and Moon (2004) used it to forecast the degree of new technology commercialization; yet other studies used it to analyze the impact of select factors on
organizational efficiency, e.g., Seol et al. (2008) used it to analyze investment in information technology on organizational efficiency; Ali and Abdel (2010) used it to analyze the impact of internal and external factors on the productivity; and yet another study used it to scientifically identify the best input-output combination while applying DEA analysis, e.g. Lim (2008).

**METHODOLOGY**

This paper uses a combination of BCC (variable returns to scale, VRS) model of DEA along with Decision Tree using CART (Classification And Regression Tree) with binary split based on Gini Index to identify the best input-output factor combination from among eleven different input-output combinations found in the literature cited above. The study adopts the framework suggested by Lim (2008) for the selection of input-output factors.

**DEA Model and Banking Efficiency Studies**

The studies cited above and several others have employed various DEA models to measure the efficiency of banks in the Indian context. All these can be categorized into three broad categories based on their approach in defining a banking unit and its major role. They are as follows:

a. Production Approach – bank is defined as a typical production unit that use purchased inputs (physical assets like labor, space, material etc) to provide services to customers. Often deposits and various bank assets are taken as proxies for quantum of services provided (produced) as outputs (Benston, 1965; Das and Ghosh, 2006; Kumar and Gulati, 2008). This approach is employed in studying the branch level efficiency of a bank.

b. Intermediary Approach – banks, as financial institution are viewed as intermediating funds between savers and borrowers. They produce intermediation services through collection of deposits and other liabilities and their application in interest earning assets, such as loans and advances, securities and other investments (Das and Ghosh, 2006; Kumar and Gulati, 2008).

c. Value-adding agency approach – this approach identifies those balance sheet items as outputs that contribute to the bank value-added (Berger, Hanweck, and Humphrey, 1987; Berger and Humphery, 1992; Grifell-Tatje and Lovell, 1996; Kumbhakar and Sarkar, 2004; Das and Ghosh, 2006).

In all three cases, the input-output factors required will differ with reference to bank’s core business objective (see Table 1).

<table>
<thead>
<tr>
<th>Model No.</th>
<th>Author/s</th>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>M 1</td>
<td>Ataullah &amp; Le (2006)  (Intermediary approach)</td>
<td>Interest expenses, Operating expenses.</td>
<td>Interest income, operating income</td>
</tr>
<tr>
<td>M 2</td>
<td>Kumar &amp; Gulati (2008) (Intermediary approach)</td>
<td>Net fixed assets, Labor, Loanable fund</td>
<td>Interest spread, non-interest income</td>
</tr>
<tr>
<td>M 3</td>
<td>Sathye (2003) (Intermediary approach)</td>
<td>Interest expenses, non-interest expenses.</td>
<td>Net interest income, non-interest income</td>
</tr>
<tr>
<td>Model</td>
<td>Authors</td>
<td>Description</td>
<td>Inputs</td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>-------------</td>
<td>--------</td>
</tr>
<tr>
<td>M 4</td>
<td>Kumbhkar and Sarkar (2004)</td>
<td>Value-added approach</td>
<td>Labor, Capital, Operating cost</td>
</tr>
<tr>
<td>M 5</td>
<td>Das &amp; Ghosh (2006)</td>
<td>Production approach</td>
<td>Interest Expenses, Operating expenses, employee expenses</td>
</tr>
<tr>
<td>M 6</td>
<td>Bhattacharjee et al. (1997)</td>
<td>Intermediary approach</td>
<td>Interest expenses, Operating expenses</td>
</tr>
<tr>
<td>M 7</td>
<td>Gupta et al. (2008)</td>
<td>Intermediary approach</td>
<td>Interest expenses, Operating expenses</td>
</tr>
<tr>
<td>M 8</td>
<td>Zhao et al. (2008)</td>
<td>Intermediary approach</td>
<td>Total Operating costs</td>
</tr>
<tr>
<td>M 9</td>
<td>Sathye (2003)</td>
<td>Intermediary approach</td>
<td>Deposits, Staff numbers</td>
</tr>
<tr>
<td>M 10</td>
<td>Das &amp; Ghosh (2006)</td>
<td>Intermediary approach</td>
<td>Demand Deposits, savings Deposits, Fixed Deposits, Operating expenses, Labor expenses</td>
</tr>
<tr>
<td>M 11</td>
<td>Mukesh &amp; Charles (2008)</td>
<td>Intermediary approach</td>
<td>Total Expenses, Deposits</td>
</tr>
<tr>
<td>M 12</td>
<td>Zhao et al. (2008)</td>
<td>Intermediary approach</td>
<td>Total Operating costs</td>
</tr>
<tr>
<td>M 13</td>
<td>Sahoo &amp; Tone (2008)</td>
<td>Intermediary approach</td>
<td>Fixed Assets, Borrowed Fund, Labor</td>
</tr>
</tbody>
</table>

A set of eleven different input-output combinations representing intermediary approach (Table 1) from the literature that used fourteen different inputs and ten different outputs were selected for the first stage of analysis. Model M4 (value-added approach) and M5 (production approach) are not considered for the analysis.

Data for more than 127 scheduled commercial banks (excluding regional rural banks) for 24 different factors (Table 3) was collected for a period of 12 years, 2000-2012. Out of these, banks where data for all variables for all 11 models was available were retained for the further analysis. The final sample consists of 57 scheduled commercial banks (DMUs) across size and ownership (private, public and foreign ownership) operating in India. The mean of the available data for each input/output factor across the study period is used in the DEA analysis.

The next step was to set up a single performance measure that defines the major role or purpose of the DMUs. The efficiency scores from DEA should best explain such a measure. In other words, when DMUs from the same industry (field of operation) are evaluated using a particular combination of inputs-outputs the DEA efficiency scores should best reflect the purpose of the DMUs.

The paper chooses the intermediary approach to evaluate DMUs and hence the ‘average change in operating income’ is set as the single performance measure. The role of a bank as an intermediary is to channelize deposits and purchased inputs to create deposits and various...
Jain et al.  

Decision Tree Analysis for Selection of Factors in DEA

categories of bank assets (Kumar and Gulati, 2008). Banks pay interest on deposits and incur cost on other inputs and get returns on loans and advances, investment and other services. Hence, the purpose of a bank as an intermediary can best be defined as to ‘increase operating income’.

While ‘operating income’ or a proxy of it e.g., interest income, non-interest income is considered as an output in most of the DEA models using the intermediary approach it is not clear if there exists a certain production function that relates input factors to the single performance measure or explain the change in such measure. Further, in the general DEA application procedure outputs that are intermediately produced are considered instead of those which are produced ultimately (Lim, 2008).

Further, the choice of intermediary approach in our study is prompted by two reasons. One, this is the most commonly used approach in empirical studies (Kumar and Gulati, 2008). Two, as this paper uses bank level efficiency and according to Berger et al. (1997) the intermediary approach is best suited for analyzing bank level efficiency.

For the purpose of Decision Tree (DT) analysis, an average of annual change in operating income over the study period 2000-2012 is taken as a single performance measure and set as the target (dependent) variable while the efficiency scores from all eleven DEA models are taken as predictive (independent) variables. All the DMUs are divided into two groups- one with all DMUs that have an average change in operating income above the median and the other with those having an average change in operating income below the median. Each DMU in the former group is assigned ‘1’ and each DMU in the latter group is assigned ‘0’

In DT analysis, the classification and regression tree (CART) algorithm is used for tree building. As the target variable is a discrete variable, a binary split with Gini index as split criterion is chosen.

**DISCUSSION AND CONCLUSIONS**

Table 2 provides a summary of results from DT analysis. It is evident that model M1 plays the most decisive role in classifying the banks into two classes; the first corresponding to those banks whose average change in operating income is above the median and the second class below the median. In other words, input-output factors from model M1 are the most appropriate in measuring the true relative efficiency of a bank vis-a-vis other banks. It can be observed from Table 2 that when the efficiency score from model M1 is higher than 0.904 and that from model M9 is higher than 0.7745, the probability of a bank belonging to the first class reaches 100%. This implies that model M9 plays the next decisive role in classifying the banks into better or otherwise. These results suggest that the management of a bank should focus their efforts and resources on increasing the efficiency scores from model M1 in order to improving their ability to generate profits. The next best would be to focus on increasing the efficiency scores from model M9.

Further, when the efficiency scores from model M1 is greater than 0.904 and that model M9 is less than .7745 that the probability of the firm belonging to class one reaches 60%. However, when the efficiency scores from model M2 is lower (higher) than 0.113; the probability of a bank belonging to the first class reaches 80% (40%). Such a conclusion does not support the logic that greater the efficiency score better the profitability of a bank. Hence model M2 can be ignored.
It can be also observed that when the efficiency scores model M1 is less than 0.904 and from model M6 is lower (higher) than 0.833 the probability of a bank belonging to second class is 0% (20%). This implies that even when the efficiency scores from model M1 is low, the profitability of the bank can be improved by improving the efficiency score of model M6.

The above analysis supports the conclusion that the input-output factors from model M1 (i.e. Interest expenses and operating expenses as inputs, and interest income and operating income as outputs) followed by those from model M9 (i.e., deposits and staff members as inputs, and net loans and non-interest income as outputs) are the most appropriate in measuring the relative efficiency of banking units at least in the Indian context.

<table>
<thead>
<tr>
<th>Type</th>
<th>Pattern of efficiency scores</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M1&lt;.904; M6&lt;.833</td>
<td>0%</td>
</tr>
<tr>
<td>2</td>
<td>M1&lt;.904; M6&gt;=.833</td>
<td>20%</td>
</tr>
<tr>
<td>3</td>
<td>M1&gt;=.904; M9&lt;.7745; M2&lt;.113</td>
<td>80%</td>
</tr>
<tr>
<td>4</td>
<td>M1&gt;=.904; M9&lt;.7745; M2&gt;=.113</td>
<td>40%</td>
</tr>
<tr>
<td>5</td>
<td>M1&gt;=.904; M9&gt;=.7745</td>
<td>100%</td>
</tr>
</tbody>
</table>

APPENDIX

The input oriented DEA model with variable returns to scale (VRS) due to Banker, Charnes and Cooper (1984) is given below. The relative efficiency of a Decision Making Unit DMU0 is obtained from the following linear programming model:

\[
\min \theta_0 \quad \text{subject to}
\]

\[
\theta_0 x_{ijo} - \sum_{j=1}^{n} \lambda_j x_{ij} \geq 0, \quad i = 1, \ldots, m
\]

\[
\sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{rjo}, \quad r = 1, \ldots, s
\]

\[
\sum_{j=1}^{n} \lambda_j = 1, \quad j = 1, \ldots, n \quad \text{(Note - the CCR model does not have this constraint on } \lambda_j \text{)}
\]

\[
\lambda_j \geq 0, \quad j = 1, \ldots, n
\]

where, \( y_{rj} \) is the amount of the \( r \)-th output to DMU \( j \), \( x_{ij} \) is the amount of the \( i \)-th input to DMU \( j \), \( \lambda_j \) are the weights of DMU \( j \) and \( \theta_0 \) is the shrinkage factor.

The BCC model seeks a set of \( \lambda_j \) values which add up to 1, minimizes \( \theta_0 \) to \( \theta_0^* \) and identifies a point within the production possibility set which uses the lowest proportion \( \theta_0^* \) of input levels of DMU \( j \) while offering output levels which are at least as high as those of DMU \( j \). This point is a composite DMU corresponding to the linear combination of efficient DMUs:

\[
\sum_{j=1}^{n} \lambda_j x_{ij}, \quad \sum_{j=1}^{n} \lambda_j y_{rj} \quad \text{with } i = 1, \ldots, m \text{ and } r = 1, \ldots, s.
\]
It can be said that:

\[
\sum_{j=1}^{n} \lambda_{j}^{*} x_{ij}, \sum_{j=1}^{n} \lambda_{j}^{*} y_{rj} \text{ outperforms } (\theta_{0},x_{j0}, y_{j0}) \text{ when } \theta_{0}^{*} < 1
\]

REFERENCES


